

Dissertation

An RKHS Approach to Estimation with Sparsity Constraints

ausgeführt zum Zwecke der Erlangung des akademischen Grades eines
Doktors der technischen Wissenschaften

unter der Leitung von
Ao. Univ.-Prof. Dipl.-Ing. Dr. Franz Hlawatsch
Institute of Telecommunications

eingereicht an der Technischen Universität Wien
Fakultät für Elektrotechnik

von
Dipl.-Ing. Alexander Jung
Vorderer Ödhof 1/1
A-3062 Kirchstetten

Acknowledgements

First of all, I would like to thank Prof. Franz Hlawatsch, whose spirit of work will always be a motivating but unreachable ideal for me.

I am furthermore grateful to Prof. Helmut Bölcskei, who very kindly agreed to act as a referee.

I would like to express my thanks to my parents. Their support made my studies of electrical engineering possible.

Special thanks to Dipl.-Ing. Sebastian Schmutzhard for sharing his impressive expertise on the mathematical theory of reproducing kernel Hilbert spaces.

Finally and most importantly I would like to express my deep gratefulness for Julia, Helene and Emily for being with me. While continuously supporting and motivating me throughout the writing of this thesis, they regularly reminded me about what is really important in life.

To Julia, Helene and Emily

Abstract

The investigation of the effects of sparsity or sparsity constraints in signal processing problems has received considerable attention recently. Sparsity constraints refer to the a priori information that the object or signal of interest can be represented by using only few elements of a predefined dictionary. Within this thesis, sparsity refers to the fact that a vector to be estimated has only few nonzero entries.

One specific field concerned with sparsity constraints has become popular under the name “Compressed Sensing” (CS). Within CS, the sparsity is exploited in order to perform (nearly) lossless compression. Moreover, this compression is carried out jointly or simultaneously with the process of sensing a physical quantity.

In contrast to CS, one can alternatively use sparsity to enhance signal processing methods. Obviously, sparsity constraints can only improve the obtainable estimation performance since the constraints can be interpreted as an additional prior information about the unknown parameter vector which is to be estimated. Our main focus will be on this aspect of sparsity, i.e., we analyze how much we can gain in estimation performance due to the sparsity constraints.

We will study in detail two specific estimation problems that are already well investigated in the absence of sparsity constraints. First, we consider an estimation problem coined the “sparse linear model” (SLM). Here, the unknown parameter determines the mean of an observed Gaussian random vector. Second, we consider an estimation problem that we refer to as the “sparse parametric covariance model” (SPCM). In this case, the unknown parameter determines the covariance matrix of an observed Gaussian random vector.

The main results of this thesis will be lower bounds on the variance of estimators for these two sparse estimation problems. We will demonstrate that the mathematical framework of reproducing kernel Hilbert spaces (RKHS) allows a simple derivation of lower bounds on the estimator variance and, moreover, a natural comparison of estimation problems with and without sparsity constraints.

Zusammenfassung

Die Untersuchung der Auswirkungen von Spärlichkeitsbedingungen (SB) in verschiedenen Problemen der Signalverarbeitung ist ein wichtiges Ziel der aktuellen Forschung. Unter SB versteht man das Vorwissen, dass das Nutzsignal als Superposition weniger “Elementarsignale” dargestellt werden kann. In dieser Arbeit beziehen sich die SB auf das Vorwissen, dass ein zu schätzender Parametervektor wenige (im Vergleich zu seiner Länge) Einträge aufweist, die von Null verschieden sind. Diese Vektoren werden dann als “spärlich” (sparse) bezeichnet.

Solche SB spielen eine zentrale Rolle in der Theorie von “Compressed Sensing” (CS). Im Rahmen von CS werden die SB verwendet, um eine (nahezu) verlustfreie Datenkompression durchzuführen. Diese Datenkompression wird dabei mit dem Prozess des Messens einer physikalischen Größe kombiniert.

Im Gegensatz zu CS kann man die SB auch zur Verbesserung bestehender Signalverarbeitungsmethoden benutzen. In dieser Arbeit steht diese zweite Möglichkeit im Vordergrund. Ein wesentliches Ziel ist die Quantifizierung der durch SB bewirkten potenziellen Leistungsverbesserung.

Es werden zwei verschiedene Schätzprobleme mit SB im Detail analysiert. Im ersten Problem, dem “sparse linear model” (SLM), bestimmt der unbekannte spärliche Parametervektor in linearer Weise den Mittelwert eines beobachteten Gaußschen Zufallsvektors. Im zweiten Problem, hier als “sparse parametric covariance model” (SPCM) bezeichnet, bestimmt der unbekannte spärliche Parametervektor die Kovarianzmatrix eines beobachteten Gaußschen Zufallsvektors.

Die Hauptergebnisse dieser Arbeit sind untere Schranken für die Varianz von Schätzern für diese beiden Schätzprobleme. Es wird gezeigt, dass die Theorie der “reproducing kernel Hilbert spaces” (RKHS) eine elegante Ableitung von unteren Varianzschranken erlaubt. Die gefundenen Schranken gestatten einen natürlichen Vergleich von Schätzproblemen mit und ohne SB.

Contents

1	Introduction	1
1.1	Sparsity in Statistical Signal Processing	2
1.2	Outline of the Thesis	3
1.3	Contribution and Related Work	4
1.4	Notation	7
2	Elements of Classical Estimation Theory	11
2.1	Basic Concepts	11
2.2	Minimax Estimation	15
2.3	Minimum Variance Estimation	16
2.3.1	Vector-Valued vs. Scalar-Valued Parameter Function	17
2.3.2	Lower Bounds via Generalized Inequalities	18
2.3.3	Exchangeability of Prescribed Bias and Parameter Function	19
2.3.4	Existence and Uniqueness of Minimum Variance Estimators	20
2.3.5	Transformations of the Parameter Function	21
2.3.6	Relaxing The Bias Equality Constraints	22
2.4	Transformation of the Observation	22
2.4.1	Invertible Transformation	22
2.4.2	Non-invertible Transformation	23
2.5	Reducing the Parameter Set	24
2.6	Statistical Models Belonging to an Exponential Family	26
3	Mathematical Tools	29
3.1	Hilbert Spaces	29
3.1.1	Some Basic Definitions	29
3.1.2	Some Basic Facts	33
3.2	Reproducing Kernel Hilbert Spaces (RKHS)	38

3.3	Important Facts about RKHS	40
3.4	RKHSs with a Differentiable Kernel	44
4	The RKHS Approach to Minimum Variance Estimation	49
4.1	Introduction	49
4.2	A Hilbert Space of Estimators	50
4.3	The RKHS Associated to a Minimum Variance Problem	52
4.3.1	Definition	52
4.3.2	Main Result	53
4.3.3	Transforming the Parameter Function	55
4.3.4	Reducing the Parameter Set	57
4.3.5	Classes of Minimum Variance Problems with Continuously Varying Kernels	57
4.3.6	Sufficient Statistics from the RKHS Viewpoint	60
4.4	RKHS Interpretation of Known Variance Bounds	62
4.4.1	Barankin Bound	63
4.4.2	Cramér–Rao bound	65
4.4.3	Bhattacharyya Bound	74
4.4.4	Hammersley-Chapman-Robbins Bound	75
5	The Sparse Linear Model	79
5.1	Introduction	79
5.2	RKHS Associated with the SLM	82
5.3	Minimum Variance Estimation for the SLM	93
5.4	Lower Bounds on the Estimator Variance for the SLM	99
5.5	The Sparse Signal in Noise Model	109
5.6	Necessity of Strict Sparsity	123
5.7	The SLM Viewpoint on Compressed Sensing	127
5.8	Comparison of the Bounds with Existing Estimators	133
5.8.1	SLM Viewpoint on Fourier Analysis	133
5.8.2	Minimum Variance Analysis of Estimators for the SSNM	135
6	The Sparse Parametric Covariance Model	139
6.1	Introduction	139
6.2	The \mathcal{D} -restricted SPCM	142
6.3	Lower Bounds on the Estimator Variance for the SPCM	148
6.4	The Sparse Diagonalizable Parametric Covariance Model	153
6.4.1	The \mathcal{D} -restricted SDPCM	154

CONTENTS

ix

6.4.2	Variance Bounds for the SDPCM	155
6.4.3	Special Case: Unbiased Spectrum Estimation	163
6.5	Comparison of the Bounds with Existing Estimators	167
7	Conclusion and Outlook	171
	Appendix A	175
	Appendix B	187
	Appendix C	197
	Appendix D	201
	Appendix E	205

Chapter 1

Introduction

Sparsity or *sparsity constraints* refer to the a-priori information that the parameter or signal vector of interest can be represented by using only few elements of a predefined dictionary. In particular, within this thesis, sparsity refers to the fact that a vector has only few non-zero entries. This thesis is concerned with the theory of classical (i.e., non-Bayesian) estimation problems where it is known that the unknown parameter vector is sparse.

Given a general classical estimation problem,¹ in particular one with sparsity constraints, two main questions are:

- How to design accurate and efficient estimators?
- Given a specific estimator, how far is it away from being optimal?

This work will mainly address the second question, i.e., the assessment of the quality of a given estimator. To that end we have to define an optimality criterion. In this thesis we will use the estimator variance (cf. Chapter 2) as the primary performance criterion. If we have a lower bound on the set of all realizable values of the performance measure, we can assess the optimality of a given estimator by comparing its performance measure, i.e., its variance, with this lower bound. A central part of this work is concerned with the derivation of such lower bounds on the minimum achievable variance of estimators with a prescribed bias function. We will make these ideas precise in Chapter 2.

This thesis considers the application of the well-known theory of *reproducing kernel Hilbert spaces* (RKHS) to specific problems within classical estimation theory. A RKHS is a special type of Hilbert space which has specific properties that will be discussed in detail in Chapter 3. The application of RKHS theory to statistical signal processing proved to be very successful, in particular in the context of machine learning [1, 2]. While the application of RKHS to general

¹We will make the notion of a classical estimation problem precise in Chapter 2.

classical estimation theory dates back to the late 1950s [3], it seems that so far the RKHS framework has not been used for the study of estimation problems with sparsity constraints. It will be demonstrated in this thesis that RKHS theory is a well-suited tool for the analysis of the effect of sparsity constraints within classical estimation problems.

Using the theory of RKHS, we will study in detail two specific estimation problems that have already been well investigated in the absence of sparsity constraints. First, we consider an estimation problem coined the “sparse linear model” (SLM),² in which the unknown parameter determines the mean of an observed Gaussian random vector. Second, we consider the estimation problem coined the “sparse parametric covariance model” (SPCM), in which the unknown parameter determines the covariance matrix of an observed Gaussian random vector. We will show how the SLM and the SPCM can be obtained as special cases of a larger class of sparse estimation problems where the statistic of the observation is given by an *exponential family* (cf. Section 2.6). The main results of this thesis will be lower bounds on the variance of estimators for the sparse estimation problems that have a prescribed bias or mean.

1.1 Sparsity in Statistical Signal Processing

Consider a vector $\mathbf{x} \in \mathbb{R}^N$ in the N -dimensional real Euclidean space [4–6]. We then call the vector \mathbf{x} *sparse* if it has only few nonzero entries, i.e., $\|\mathbf{x}\|_0 \ll N$ where $\|\mathbf{x}\|_0$ denotes the number of nonzero entries of \mathbf{x} . More precisely, the set of all vectors $\mathbf{x} \in \mathbb{R}^N$ with *at most* S nonzero entries will be called the set of *strictly S -sparse vectors* and denoted by

$$\mathcal{X}_S \triangleq \{\mathbf{x} \in \mathbb{R}^N \mid \|\mathbf{x}\|_0 \leq S\}, \quad (1.1)$$

where $S \in [N] \triangleq \{1, \dots, N\}$, is called the *sparsity degree*. Beside the notion of strict sparsity, there are also weaker notions of sparsity or sparse signals. E.g., the class of *compressible* signals [7] consists of vectors whose sorted magnitudes of its entries exhibit a power-law decay so that the signal vector can be well approximated by a strictly S -sparse vector with a suitable sparsity degree S . Alternatively, we call a vector $\mathbf{x} \in \mathbb{R}^N$ *approximately S -sparse* if it consists of S “large” entries and $N - S$ “small” entries. However, for a rigorous analysis, one has to make the meaning of a “large” and a “small” entry precise. A specific approach to characterize the set of approximately S -sparse vectors is to measure the norm of the “tail,” i.e., the norm of the vector that is obtained from \mathbf{x} by setting the S largest (in magnitude) entries to zero. This quantification of approximate sparsity is used in [8] to characterize the performance of a signal processing algorithm. Another approach to model approximately sparse vectors will be discussed in detail in Section 5.6.

²Again we justify the unusual convention to name estimation problems after models by the fact, that an estimation problem is defined to a large extent by an observation model which connects the unknown parameter to be estimation with the observation.

If for a given application, e.g. radar [9] or channel estimation [10,11], it is known that the object of interest can be represented by a sparse vector, then we should exploit this a-priori information. This can be done in two alternative ways. On the one hand, we can use sparsity to perform a nearly lossless compression. On the other hand, we could use the sparsity information in order to improve the performance of existing signal processing schemes.

The former approach is taken within *Compressed Sensing* (CS) [12–14], i.e., CS refers to signal processing applications where the sparsity is exploited in order to perform a nearly lossless compression simultaneously with the sensing process. The output of the sensing process is then referred to as compressive measurements [15]. From a more formal viewpoint, the theory of CS is concerned with two fundamental problems:

- The first problem considered in CS is that of finding good methods for the joint compression and sensing process. In particular, if the sensing process is linear and modeled by a matrix-vector multiplication between a CS measurement matrix $\mathbf{M} \in \mathbb{R}^{M \times N}$ and the sparse signal vector $\mathbf{x} \in \mathbb{R}^N$, we would like to construct measurement matrices which yield both a good compression performance (i.e., a small number of compressive measurements z_m , with $\mathbf{z} = \mathbf{M}\mathbf{x}$, which is equivalent to the requirement $M \ll N$) and an accurate reconstruction of the signal vector \mathbf{x} from the compressive measurements $\mathbf{z} = \mathbf{M}\mathbf{x}$ at a later point. These two desiderata are contradicting goals and thus there is a tradeoff between compression performance and achievable reconstruction accuracy. So far, the best known performance guarantees are obtained by random constructions of the CS measurement matrix [12, 13].
- The second main problem faced in CS is how to exploit the sparsity of the sparse signal vector $\mathbf{x} \in \mathbb{R}^N$ in order to estimate or recover it in an efficient and accurate way from given compressive measurements $\mathbf{z} = \mathbf{M}\mathbf{x}$. Existing results show that the sparsity allows for computationally efficient and accurate estimation or recovery procedures [16–18].

By contrast to CS, one may choose not to perform any compression but instead use the full available amount of raw data and exploit the sparsity constraints in order to obtain a better performance of signal processing algorithms (compared to the performance obtained in the absence of any sparsity constraints). Our thesis is concerned with this rationale, i.e., we analyze how much one can potentially gain in performance using the a priori information represented by the sparsity constraints.

1.2 Outline of the Thesis

In Chapter 2, we present a brief review of some important concepts used within classical (non-Bayesian) estimation theory. We will discuss two major rationales within classical estimation

theory, i.e., the minimax and the minimum variance rationales. The mathematical tools used in later chapters are presented in Chapter 3. In Chapter 4, we introduce the RKHS approach to minimum variance estimation and give detailed interpretations of some well-known lower bounds on the estimator variance from the (geometric) viewpoint of RKHS. The two core chapters of this thesis are Chapter 5 and Chapter 6, where the RKHS approach to minimum variance estimation is specialized to two specific estimation problems, i.e., the SLM and the SPCM. In Chapter 5, we give a detailed characterization of the RKHS associated to the SLM and use this characterization for the derivation of novel lower bounds on the estimator variance. We will also show in Chapter 5 that the (linear) CS recovery problem can be interpreted as a specific instance of the SLM. In Chapter 6, we will introduce an RKHS approach to the SPCM, which is quite different from the RKHS approach to the SLM. This RKHS approach will then be used to derive novel lower bounds on the estimator variance for the SPCM.

1.3 Contribution and Related Work

This thesis develops the application of the theory of RKHS to classical estimation problems with sparsity constraints. The main part of the thesis is concerned with the analysis of two specific estimation problems.

The first problem concerns the SLM. The goal is the estimation of a sparse parameter vector of which a noisy and linearly distorted version is observed. The SLM is obtained from the well-known linear Gaussian model (LGM) [19–21] by adding sparsity constraints on the unknown parameter vector. A special case of the SLM has been considered in [22–24], where the authors investigate asymptotic minimax estimation. By contrast, we consider a finite-dimensional setting. Furthermore, we do not consider minimax estimation but a different estimation rationale, i.e., minimum variance estimation, where the bias or mean of an estimator is prescribed. We give results concerning the minimum variance that can be obtained by estimators with a prescribed bias. In this spirit, the authors of [25,26] derive lower bounds on the variance of estimators for the SLM with a locally prescribed bias. By locally we mean that the bias has to be defined only at a given, fixed parameter value and in a small neighborhood around this point. Their bound is an adaption of the well-known Cramér–Rao bound (CRB) [27] to the sparse setting and depends only on a first-order characterization (related to a truncated Taylor expansion of an analytic function) of the estimation problem and bias. In a similar setting, the authors of [28] prove the asymptotic achievability of the CRB for the CS recovery problem. Whereas [28] considers the SLM with a random system matrix, we consider a fixed deterministic system matrix as in [25].

By contrast to [25,26], we derive lower bounds on the estimator variance using a higher-order characterization of the estimation problem and estimator bias. This implies that our lower bounds

will be in general tighter, i.e., higher, than those found in [25, 26]. Another major difference is that the results of [25, 26] and, in a more general setting, those of [29] are stated in terms of generalized (matrix) inequalities [30] involving the covariance matrix of an estimator. However, in this work, we present bounds on the variance of the individual entries of an estimator. In principle both type of bounds are equivalent, i.e., we can in general reformulate our bounds in terms of generalized inequalities and vice versa. In particular, the bounds in [25, 26, 29] can be reformulated such that they can be compared with our bounds.

The second estimation problem considered in this thesis concerns the SPCM. The goal is to estimate a sparse parameter vector that determines the covariance matrix \mathbf{C} of a Gaussian (random) signal vector of which a noisy version is observed. While covariance estimation is an established topic of research (see, e.g., the recent work [31–33]), sparsity constraints are mostly placed directly on the inverse covariance matrix, i.e., the *precision matrix* \mathbf{C}^{-1} , by assuming that most entries of \mathbf{C}^{-1} are zero. This is mainly because the precision matrix of a Gaussian random vector is directly related to a graphical model [34] describing the statistical structure of the random vector. A sparse precision matrix \mathbf{C}^{-1} thus corresponds to a sparse graph, i.e., a graph with relatively few edges compared to the number of nodes.

From a methodological perspective, this thesis specializes the theory of RKHS for the application to minimum variance estimation with sparsity constraints. By sparsity constraint, we mean that the unknown parameter vector $\mathbf{x} \in \mathbb{R}^N$ to be estimated is known to be S -sparse, i.e., $\mathbf{x} \in \mathcal{X}_S$ with a known sparsity degree $S \in \mathbb{N}$. The application of Hilbert space, in particular RKHS, techniques to minimum variance estimation dates back to the seminal work of Parzen [3], which introduced the theory of RKHS to general estimation problems. The approach of [3] is further developed in [35] for specific estimation problems. We see our thesis as another specialization of the work [3] to estimation problems with sparsity constraints.

Differently from this “RKHS line” of work, a Hilbert space approach has been proposed for general minimum variance estimation problems in the recent paper [36]. The Hilbert space used in [36] for the derivation of lower bounds on the estimator variance is different from the RKHS used in this thesis and based on integral operators associated to a minimum variance estimation problem. Furthermore, the authors in [36] do not consider estimation with sparsity constraints.

The main contributions of this thesis (listed according to their order of appearance) are the following.

- Chapter 4: We show in Section 4.3.5 that under mild conditions on the underlying estimation problem and prescribed, the minimum achievable variance $L_{\mathcal{M}}$ of the minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$, viewed as a function of the parameter vector \mathbf{x}_0 , is lower semi-continuous.

- Chapter 4: We prove in Section 4.3.6, that the RKHS associated to a minimum variance problem remains unchanged if the original observation is replaced by any sufficient statistic. This result could be regarded as an RKHS analogue of the famous “Rao-Blackwell-Lehmann-Scheffé” (RBLs) theorem [20].³
- Chapter 4: We provide a detailed derivation of some well-known variance bounds using exclusively RKHS theory. This yields an intuitive geometric interpretation of these bounds.
- Section 5.2: We present a derivation and complete characterization of the RKHS associated with the SLM. This includes as a special case a novel characterization of the RKHS associated with the LGM. Our characterization will be quite different from those given [35] which already considered the LGM and associated RKHS.
- Section 5.3: We discuss minimum variance estimation for the SLM. We characterize the class of valid bias functions, i.e., those bias functions for which there exists at least one finite-variance estimator having this bias function. In particular, we show that there does not exist a finite-variance unbiased estimator of the support set of the unknown sparse parameter vector.
- Section 5.4: We interpret existing lower bounds on the estimator variance for the SLM from the RKHS viewpoint and then derive novel lower bounds on the estimator variance based on the RKHS derived in Section 5.2.
- Section 5.5: We focus on a special case of the SLM termed the *sparse signal in noise model* (SSNM). By exploiting the specific structure of the SSNM, it is possible to derive stronger results than for the general SLM. In particular, we obtain closed-form expressions for the minimum achievable variance (i.e., the Barankin bound) and the corresponding estimator which achieves this minimum variance (i.e., the locally minimum variance (LMV) estimator, which will be defined in Section 2.3) for a wide class of bias functions, which includes the bias function of the hard-thresholding (HT) estimator [37].
- Section 5.6: We show that the strict sparsity constraints of the SLM are necessary in order for the minimum achievable variance of the estimation problem to be strictly lower than that of the minimum achievable variance for the LGM.

³The RBLs theorem basically states that given an arbitrary estimator one can always find an estimator that (i) depends on the observation only via a sufficient statistic, (ii) has the same bias as the original estimator and (iii) its variance does not exceed those of the original estimator. Thus, using a sufficient statistic as the new observation means no loss of information in terms of minimum variance estimation.

- Section 5.7: Based on the RKHS approach, we present an SLM-based analysis of the CS recovery problem.
- Section 5.8: We compare the actual variance behaviors of some popular estimation schemes for the SLM and SSNM with the corresponding lower bounds.
- Section 6.1 and 6.2: We present a specific RKHS approach to the SPCM and discuss the differences between the RKHS approach to the SPCM and SLM.
- Section 6.3: We derive novel lower bounds on the estimator variance for the SPCM based on the RKHS defined in Section 6.2.
- Section 6.4: We focus on a special case of the SPCM termed the *sparse diagonalizable parametric covariance model* (SDPCM). By exploiting the specific structure of the SDPCM, it is possible to derive stronger results than for the general SPCM. In particular, we will show in Section 6.4.3 that for unbiased estimation for the SDPCM, the strict sparsity constraints are necessary in order for the minimum achievable variance to be strictly lower than for the case where no sparsity constraints are present.
- Section 6.5: We compare our lower bounds on the variance with the actual variance behavior of two specific estimation schemes for the SDPCM.

1.4 Notation

In the following we will define the notation used in this thesis. The notation related to probability theory will be introduced separately in Chapter 2.

The set of (nonnegative) real numbers is denoted by (\mathbb{R}_+) \mathbb{R} , the set of natural numbers is denoted by $\mathbb{N} = \{1, 2, 3, \dots\}$ and the set of nonnegative integers by $\mathbb{Z}_+ = \{0, 1, 2, \dots\}$. Given a nonnegative integer $L \in \mathbb{Z}_+$, we denote by $[L]$ the set of all numbers up to L , i.e., $[L] \triangleq \{1, \dots, L\}$ if $L > 0$ and for the special case $L = 0$ we set $[0] \triangleq \emptyset$ where \emptyset denotes the empty set. Given an arbitrary set \mathcal{A} , we define the Kronecker delta $\delta_{l,l'} \in \{0, 1\}$ for $l, l' \in \mathcal{A}$ by setting $\delta_{l,l'} = 0$ if and only if $l \neq l'$. Given the two sets \mathcal{A} and \mathcal{B} , we denote by $\mathcal{A} \setminus \mathcal{B}$ the set which consists of those elements of \mathcal{A} that are not an element of \mathcal{B} and by $\mathcal{A} \cup \mathcal{B}$ and $\mathcal{A} \cap \mathcal{B}$ the union and intersection, respectively of \mathcal{A} and \mathcal{B} [6]. Furthermore we denote by $|\mathcal{A}|$ the cardinality, i.e., the number of elements of a finite set \mathcal{A} . We denote by $\log x$ the natural logarithm of the positive real number x .

Uppercase boldface letters denote matrices and lowercase boldface letters denote vectors.⁴ The superscript T stands for transposition. Given a matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ we denote by $\text{vec}\{\mathbf{H}\} \in$

⁴Since a matrix can be viewed as a special case of a linear operator, we will use uppercase boldface letters also to denote general linear operators.

\mathbb{R}^{MN} the certain vector that is obtained by vertically stacking the columns of \mathbf{H} . For a square matrix $\mathbf{H} \in \mathbb{R}^{N \times N}$, we denote by $\text{Tr}\{\mathbf{H}\}$, $\det\{\mathbf{H}\}$ and \mathbf{H}^{-1} its trace, determinant and inverse (if it exists), respectively. The kernel or null space $\mathcal{N}(\mathbf{H})$ of a matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ is defined as $\mathcal{N}(\mathbf{H}) \triangleq \{\mathbf{x} \in \mathbb{R}^N | \mathbf{H}\mathbf{x} = \mathbf{0}\}$. We denote by $\text{span}(\mathbf{H})$ the column span of a matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$, i.e. $\text{span}(\mathbf{H}) \triangleq \{\mathbf{y} \in \mathbb{R}^M | \exists \mathbf{x} \in \mathbb{R}^N : \mathbf{y} = \mathbf{H}\mathbf{x}\}$.

The k th entry of a vector $\mathbf{x} \in \mathbb{R}^N$ and the entry in the k th row and l th column of a matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ are denoted by x_k and $(\mathbf{H})_{k,l}$, respectively. We designate the identity matrix of dimension $N \times N$ as \mathbf{I}_N or just as \mathbf{I} if the dimension is clear from the context. The k th column of the identity matrix will be denoted by \mathbf{e}_k . By $\mathbf{0}$ we denote a vector or matrix consisting only of zeros whereby the dimension of the vector or matrix should be clear from context. Given two vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^N$ and two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{M \times N}$ we denote by $\mathbf{a} \neq \mathbf{b}$ and $\mathbf{A} \neq \mathbf{B}$ the fact that the two vectors and matrices, respectively differ in at least one entry. The support, i.e. the set of indices that correspond to non-zero entries, and number of non-zero entries of a vector $\mathbf{x} \in \mathbb{R}^N$ are denoted by $\text{supp}(\mathbf{x})$ and $\|\mathbf{x}\|_0$, respectively. Given a vector $\mathbf{x} \in \mathbb{R}^N$, we define its p norm as $\|\mathbf{x}\|_p \triangleq \left(\sum_{k \in [N]} x_k^p\right)^{1/p}$. We denote by $\|\mathbf{a}\|_\infty \triangleq \max_{k \in [N]} |a_k|$ the largest magnitude of the entries of the vector $\mathbf{a} \in \mathbb{R}^N$. Given a square matrix $\mathbf{H} \in \mathbb{R}^{M \times M}$ we denote by $\|\mathbf{H}\|_p$ its matrix p -norm defined as $\|\mathbf{H}\|_p \triangleq \sup_{\|\mathbf{x}\|_p=1} \|\mathbf{H}\mathbf{x}\|_p$, where $\mathbf{x} \in \mathbb{R}^M$.

Given an index set $\mathcal{K} \subseteq [N]$, a vector $\mathbf{x} \in \mathbb{R}^N$, and a matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$, we denote by $\mathbf{x}^{\mathcal{K}} \in \mathbb{R}^N$ and $\mathbf{H}_{\mathcal{K}} \in \mathbb{R}^{M \times |\mathcal{K}|}$ the vector which is obtained from \mathbf{x} by zeroing all entries except those indexed by \mathcal{K} and the matrix that is obtained by selecting only the columns of \mathbf{H} indexed by \mathcal{K} , respectively.

The fact that a square matrix $\mathbf{R} \in \mathbb{R}^{N \times N}$ is positive semi-definite (psd), i.e.,

$$\mathbf{x}^T \mathbf{R} \mathbf{x} \geq 0 \quad \forall \mathbf{x} \in \mathbb{R}^N, \quad (1.2)$$

is denoted by $\mathbf{R} \geq \mathbf{0}$. Given two psd matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N \times N}$, we write $\mathbf{A} \geq \mathbf{B}$ if the matrix $\mathbf{A} - \mathbf{B}$, given elementwise as the difference between the elements of \mathbf{A} and \mathbf{B} , is psd, i.e., $\mathbf{A} - \mathbf{B} \geq \mathbf{0}$ [30]. If the inequality in (1.2) is replaced by a strict inequality, we call the matrix \mathbf{R} positive definite, in symbols $\mathbf{R} > \mathbf{0}$.

Given an arbitrary matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$, we call any decomposition of the form

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \quad (1.3)$$

with $\mathbf{U} \in \mathbb{R}^{M \times r}$, $\mathbf{V} \in \mathbb{R}^{N \times r}$, $\mathbf{\Sigma} \in \mathbb{R}^{r \times r}$ the *thin singular value decomposition* (SVD) [38] of \mathbf{H} if $\mathbf{U}^T \mathbf{U} = \mathbf{I}$, $\mathbf{V}^T \mathbf{V} = \mathbf{I}$ and the matrix $\mathbf{\Sigma}$ is a diagonal matrix with strictly positive real-valued main diagonal elements, i.e., $(\mathbf{\Sigma})_{k,k} > 0$. These main diagonal entries are called the *singular values* of \mathbf{H} . Note that it can be shown that every nonzero matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ has a thin SVD [38]. The

rank or column rank of \mathbf{H} , denoted by $\text{rank}(\mathbf{H})$, is defined as the unique dimension r of the matrix $\mathbf{\Sigma}$ appearing in (1.3). We also set $\text{rank}(\mathbf{0}) = 0$.

Given a psd matrix $\mathbf{H} \in \mathbb{R}^{N \times N}$, i.e., $\mathbf{H} \geq \mathbf{0}$, we call any decomposition of the form

$$\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T \quad (1.4)$$

with $\mathbf{U} \in \mathbb{R}^{N \times \text{rank}(\mathbf{H})}$, $\mathbf{\Sigma} \in \mathbb{R}^{\text{rank}(\mathbf{H}) \times \text{rank}(\mathbf{H})}$ the *thin eigenvalue decomposition* (EVD) (or *thin symmetric Schur decomposition*) [38] of \mathbf{H} if $\mathbf{U}^T\mathbf{U} = \mathbf{I}$ and the matrix $\mathbf{\Sigma}$ is a diagonal matrix with positive real-valued main diagonal elements, i.e., $(\mathbf{\Sigma})_{k,k} > 0$. These main diagonal elements are called the *eigenvalues* of \mathbf{H} . Note that it can be shown that every psd matrix $\mathbf{H} \in \mathbb{R}^{N \times N}$, where $\mathbf{H} \neq \mathbf{0}$, has a thin EVD [38].

Given a nonzero matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ and its thin SVD (1.3), we denote by $\mathbf{H}^\dagger \in \mathbb{R}^{N \times M}$ its Moore Penrose pseudo inverse [38] defined by $\mathbf{H}^\dagger \triangleq \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T$. Note that due to the properties of the thin SVD, the inverse $\mathbf{\Sigma}^{-1}$ is guaranteed to exist. For $\mathbf{H} = \mathbf{0}$, the Moore Penrose inverse is defined as $\mathbf{H}^\dagger = \mathbf{0}$.

A sequence of elements $f_l \in \mathcal{M}$, where $l \in \mathbb{N}$, that belong to a certain set \mathcal{M} is denoted by $\{f_l \in \mathcal{M}\}_{l \rightarrow \infty}$ or, if the underlying set \mathcal{M} is clear from the context, by $\{f_l\}_{l \rightarrow \infty}$.

We call a map or function $f(\cdot) : \mathcal{A} \rightarrow \mathcal{B}$ invertible if there exists a map $f^{-1}(\cdot)$, called the inverse map or inverse function, such that $f^{-1}(f(\mathbf{x})) = \mathbf{x}$ for every $\mathbf{x} \in \mathcal{A}$. A map $f(\cdot) : \mathcal{A} \rightarrow \mathcal{B}$ is called bijective if $f(\mathcal{A}) = \mathcal{B}$ and if it is invertible, i.e., there exists an inverse map $f^{-1}(\cdot)$ [5]. If a real-valued function $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ is identically zero on its domain, i.e., $f(\mathcal{D}) = \{0\}$, we will denote this by $f(\cdot) \equiv 0$. Given a function $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ defined on some domain \mathcal{D} and a subdomain $\mathcal{D}_1 \subseteq \mathcal{D}$, we denote by $f|_{\mathcal{D}_1}(\cdot)$ or $f(\cdot)|_{\mathcal{D}_1}$ its restriction to \mathcal{D}_1 , i.e., the specific function defined on \mathcal{D}_1 that coincides with $f(\cdot)$ on \mathcal{D}_1 . More explicitly, the function $f|_{\mathcal{D}_1}(\cdot) : \mathcal{D}_1 \rightarrow \mathbb{R}$ satisfies $f|_{\mathcal{D}_1}(\mathbf{x}) = f(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{D}_1$. For the special case when the subdomain consists of a single element, i.e., $\mathcal{D}_1 = \{\mathbf{x}\}$, we denote by $f(\cdot)|_{\mathbf{x}}$ the function value $f(\mathbf{x})$ at \mathbf{x} .

The closed ball centered at $\mathbf{x}_0 \in \mathbb{R}^N$ with radius $r > 0$ is denoted by $\mathcal{B}(\mathbf{x}_0, r) \triangleq \{\mathbf{x} \in \mathbb{R}^N \mid \|\mathbf{x} - \mathbf{x}_0\|_2 \leq r\}$.

Given a real-valued function $f(\cdot) : \mathcal{D} \subseteq \mathbb{R}^N \rightarrow \mathbb{R}$, we say that $f(\cdot)$ is continuous at the point $\mathbf{x}_0 \in \mathcal{D}$ if $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$. A function is said to be continuous if it is continuous at every point of its domain. Furthermore, we say that a function $f(\cdot) : \mathcal{D} \subseteq \mathbb{R}^N \rightarrow \mathbb{R}$ is lower (upper) semi-continuous at the point $\mathbf{x}_0 \in \mathcal{D}$ if for every $\varepsilon > 0$ there exists a radius r such that $f(\mathbf{x}) \geq f(\mathbf{x}_0) - \varepsilon$ ($f(\mathbf{x}) \leq f(\mathbf{x}_0) + \varepsilon$) for every $\mathbf{x} \in \mathcal{B}(\mathbf{x}_0, r)$ and denote this fact by $\liminf_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) \geq f(\mathbf{x}_0)$ ($\limsup_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) \leq f(\mathbf{x}_0)$) [5, 39], where

$$\liminf_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) \triangleq \sup_{r > 0} \left(\inf_{\mathbf{x} \in \mathcal{D} \cap (\mathcal{B}(\mathbf{x}_0, r) \setminus \{\mathbf{x}_0\})} f(\mathbf{x}) \right), \quad \limsup_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) \triangleq \inf_{r > 0} \left(\sup_{\mathbf{x} \in \mathcal{D} \cap (\mathcal{B}(\mathbf{x}_0, r) \setminus \{\mathbf{x}_0\})} f(\mathbf{x}) \right). \quad (1.5)$$

Given a real-valued function $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ and an element $\mathbf{x} \in \mathcal{D}$, we denote by $R(\cdot, \mathbf{x})$ the specific function $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ given by $f(\mathbf{x}') = R(\mathbf{x}', \mathbf{x})$ for every $\mathbf{x}' \in \mathcal{D}$. Given two functions $f(\cdot) : \mathcal{A} \rightarrow \mathcal{B}$, $g(\cdot) : \mathcal{C} \rightarrow \mathcal{D}$, we denote by $f(\cdot) = g(\cdot)$ the fact that $\mathcal{A} = \mathcal{C}$, $\mathcal{B} = \mathcal{D}$ and for every $\mathbf{x} \in \mathcal{A}$ we have $f(\mathbf{x}) = g(\mathbf{x})$.

Given a real-valued function $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ where $\mathcal{D} \subseteq \mathbb{R}^N$, we denote by $\frac{\partial^{\mathbf{e}_k} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}_k}}$ or $\frac{\partial f(\mathbf{x})}{\partial x_k}$ the partial derivative (if it exists) of $f(\cdot)$ with respect to the k th entry, i.e., $\frac{\partial^{\mathbf{e}_k} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}_k}}$ denotes the function $g(\mathbf{x}) = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{e}_k) - f(\mathbf{x})}{h}$ [5]. This definition is then recursively extended to an arbitrary multi-index $\mathbf{p} \in \mathbb{Z}_+^N$ [40] as follows: Denoting the partial derivative of order \mathbf{p} by $h(\mathbf{x})$, i.e., $h(\mathbf{x}) = \frac{\partial^{\mathbf{p}} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}}$, the partial derivative of order $\mathbf{p}' = \mathbf{p} + \mathbf{e}_k$ denoted by $\frac{\partial^{\mathbf{p}'} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}'}}$ is defined as $\frac{\partial^{\mathbf{e}_k} h(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}_k}}$. Strictly speaking, this definition is not consistent when used for an arbitrary function $f(\cdot)$, since the partial derivative of order \mathbf{p} might depend on the order of adding the \mathbf{e}_k . However, within this thesis we exclusively deal with functions $f(\cdot)$ for which this definition of the partial derivative of order $\mathbf{p} \in \mathbb{Z}_+^N$ is unambiguous. Given a real-valued function $g(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ where $\mathcal{D} \subseteq \mathbb{R}^N$, we denote by $\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \Big|_{\mathbf{x}_0}$ the vector $\mathbf{a} \in \mathbb{R}^N$ given elementwise by $a_k = \frac{\partial g(\mathbf{x})}{\partial x_k} \Big|_{\mathbf{x}_0}$. Given a real-valued function $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ where $\mathcal{D} \subseteq \mathbb{R}^N$, we denote by $\frac{\partial^{\mathbf{e}_k} R(\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{e}_k}}$ or $\frac{\partial R(\mathbf{x}_1, \mathbf{x}_2)}{\partial x_{2,k}}$ its partial derivative of order \mathbf{e}_k w.r.t. \mathbf{x}_2 , i.e., the function $g(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ defined as $g(\mathbf{x}_1, \mathbf{x}_2) = \lim_{h \rightarrow 0} \frac{R(\mathbf{x}_1, \mathbf{x}_2 + h\mathbf{e}_k) - R(\mathbf{x}_1, \mathbf{x}_2)}{h}$. The generalization to partial derivatives of the form $\frac{\partial^{\mathbf{p}_1} \partial^{\mathbf{p}_2} R(\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{x}_1^{\mathbf{p}_1} \partial \mathbf{x}_2^{\mathbf{p}_2}}$ with $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{Z}_+^N$ is a straightforward extension of the definition of $\frac{\partial^{\mathbf{e}_k} R(\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{e}_k}}$ obtained by considering the pair $(\mathbf{x}_1, \mathbf{x}_2)$ as a “super-vector” of length $2N$.

For a finite set \mathcal{T} , the sum $\sum_{l \in \mathcal{T}} f[l]$ and product $\prod_{l \in \mathcal{T}} f[l]$ of a real valued function $f[\cdot] : \mathcal{T} \rightarrow \mathbb{R}$ are defined in the usual sense. For the special case given by the empty set, i.e., $\mathcal{T} = \emptyset$ we set $\sum_{l \in \mathcal{T}} f[l] = 0$ and $\prod_{l \in \mathcal{T}} f[l] = 1$ respectively. Given an arbitrary set \mathcal{T} , we define the sum $\sum_{l \in \mathcal{T}} f[l]$ as the supremum of all finite sums $\sum_{l \in \mathcal{T}' \subseteq \mathcal{T}} f[l]$ where \mathcal{T}' is an arbitrary finite subset of \mathcal{T} [6, p. 83]. As noted in [6], the so defined sum is nothing but the Lebesgue integral with respect to the counting measure on \mathcal{T} .

We denote by $\ell^2(\mathcal{T})$ the set of all real-valued functions $f(\cdot) : \mathcal{T} \rightarrow \mathbb{R}$ defined on \mathcal{T} for which the sum $\sum_{l \in \mathcal{T}} |f[l]|^2$ is finite. It can be shown [6] that the set $\ell^2(\mathcal{T})$ is a normed vector space with norm $\|f[\cdot]\|_{\mathcal{T}} \triangleq \sqrt{\sum_{l \in \mathcal{T}} |f[l]|^2}$.

Given a nonnegative integer $p \in \mathbb{Z}_+$ we denote by $p!$ the factorial, defined recursively by $0! \triangleq 1$ and $p! \triangleq p(p-1)!$. Given a multi-index $\mathbf{p} \in \mathbb{Z}_+^N$ [40], i.e., an N -tuple of nonnegative integers, we denote by $\mathbf{p}!$ and $|\mathbf{p}|$ the product $\mathbf{p}! \triangleq \prod_{l \in [N]} p_l!$ and the sum $|\mathbf{p}| \triangleq \sum_{l \in [N]} p_l$, respectively. For a vector $\mathbf{x} \in \mathbb{R}^N$ and a multi-index $\mathbf{p} \in \mathbb{Z}_+^N$, we denote by $\mathbf{x}^{\mathbf{p}}$ the product $\mathbf{x}^{\mathbf{p}} \triangleq \prod_{l \in [N]} (x_l)^{p_l}$. Given two multi-indices $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{Z}_+^N$, we mean by $\mathbf{p}_1 \leq \mathbf{p}_2$ that this inequality holds separately for every single entry, i.e., $p_{1,l} \leq p_{2,l}$ for every $l \in [N]$.

Chapter 2

Elements of Classical Estimation Theory

Classical or non-Bayesian estimation theory [19–21, 41] is concerned with the problem of inferring the value of an unknown but deterministic parameter $\mathbf{x} \in \mathbb{R}^N$ based on the observation of a random vector $\mathbf{y} \in \mathbb{R}^M$.¹ By contrast to Bayesian estimation theory [19–21, 41], we do not model the parameter \mathbf{x} as a random vector that has a certain statistical characterization. Instead, it is assumed that the parameter vector has a fixed value that is however unknown.

2.1 Basic Concepts

The elements of a classical estimation problem are the following.

- **Parameter set.** We assume that the parameter \mathbf{x} is an element of the parameter set \mathcal{X} , i.e.,

$$\mathbf{x} \in \mathcal{X}. \tag{2.1}$$

Note that the knowledge of (2.1) also expresses some prior information. Intuitively speaking, the smaller the parameter set \mathcal{X} , the more prior information is available. In the extreme case (which is practically irrelevant) of a singleton $\mathcal{X} = \{\mathbf{x}_0\}$, i.e., the parameter set consists of a single element, the estimation problem becomes obviously trivial. A main part of this thesis is concerned with the question of how much certain reductions of the parameter set \mathcal{X} help in terms of estimation quality.

- **Statistical model.** In order to infer the value of the parameter \mathbf{x} from an observed random quantity \mathbf{y} , there has to be some relationship between \mathbf{x} and \mathbf{y} . Within classical estimation theory, this relationship is modeled by a family (a set) of probability density functions (pdf)

¹Of course \mathbf{x} and \mathbf{y} do not necessarily have to be real-valued vectors, but we will assume this without loss of generality for our purposes.

denoted by $f(\mathbf{y}; \mathbf{x})$ and called the statistical model. If the true parameter vector is \mathbf{x}_0 , then the probability of the event $\{\mathbf{y} \in \mathcal{A}\}$, where $\mathcal{A} \subset \mathbb{R}^M$ is an arbitrary measurable² set, is given by

$$\mathbf{P}\{\mathbf{y} \in \mathcal{A}\} = \int_{\mathbf{y} \in \mathcal{A}} f(\mathbf{y}; \mathbf{x}_0) d\mathbf{y}. \quad (2.2)$$

We will consider only statistical models for which one can define a pdf $f(\mathbf{y}; \mathbf{x})$ w.r.t. to the Lebesgue measure in \mathbb{R}^M such that the relation (2.2) can be interpreted in the sense of integration w.r.t. the Lebesgue measure [6]. However, the main concepts developed in this thesis could be in principle also applied to a more general case where the statistics of the observation \mathbf{y} are described by a *probability measure* [42, 43].³ We will use the symbol $\mathbf{E}_{\mathbf{x}_0}\{\mathbf{T}(\mathbf{y})\}$, where $\mathbf{T}(\mathbf{y}) \in \mathbb{R}^K$ is a (possibly random) function of the observation, for the expectation (cf. [42, 43]) of $\mathbf{T}(\mathbf{y})$ w.r.t. the pdf $f(\mathbf{y}; \mathbf{x}_0)$, i.e.,

$$\mathbf{E}_{\mathbf{x}_0}\{\mathbf{T}(\mathbf{y})\} \triangleq \int_{\mathbf{z}, \mathbf{y}} \mathbf{z} f(\mathbf{z} = \mathbf{T}(\mathbf{y}) | \mathbf{y}) f(\mathbf{y}; \mathbf{x}_0) d\mathbf{z} d\mathbf{y}. \quad (2.3)$$

Note that the conditional pdf $f(\mathbf{z} = \mathbf{T}(\mathbf{y}) | \mathbf{y})$ of the function value $\mathbf{z} = \mathbf{T}(\mathbf{y})$, given the observation \mathbf{y} , does not depend on the parameter vector \mathbf{x}_0 . This corresponds to the fact that the random function $\mathbf{T}(\cdot)$ is statistically independent of the observation \mathbf{y} for every parameter vector $\mathbf{x} \in \mathcal{X}$. For the special case of a deterministic mapping $\mathbf{T}(\mathbf{y})$, the definition (2.3) reduces to

$$\mathbf{E}_{\mathbf{x}_0}\{\mathbf{T}(\mathbf{y})\} \triangleq \int_{\mathbf{y}} \mathbf{T}(\mathbf{y}) f(\mathbf{y}; \mathbf{x}_0) d\mathbf{y}. \quad (2.4)$$

- **Estimation problem.** In some situations, we are not interested in the parameter \mathbf{x} directly but rather the value $\mathbf{g}(\mathbf{x})$ of a vector-valued function

$$\mathbf{g}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P \quad (2.5)$$

of the true parameter. Consider, e.g., the situation where the vector \mathbf{x} represents a transmitted signal and one is not interested in the signal itself but only if a signal has been transmitted, i.e., if the signal energy $g(\mathbf{x}) = \|\mathbf{x}\|_2^2$ is above a threshold.

Let us formalize the notion of a *classical estimation problem* by

Definition 2.1.1. A (classical) estimation problem \mathcal{E} is a triplet $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$ consisting of the parameter set $\mathcal{X} \subseteq \mathbb{R}^N$, a statistical model $f(\mathbf{y}; \mathbf{x})$, and the function $\mathbf{g}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$ of which we would like to estimate the value $\mathbf{g}(\mathbf{x})$.

²We consider the Lebesgue measure for the Euclidean space \mathbb{R}^N .

³A pdf is nothing but a convenient representation of probability measures that satisfy some technical conditions [42, 43].

- **Estimator.** The estimation or inference of the value $\mathbf{g}(\mathbf{x})$ based on the observation \mathbf{y} is performed by applying an estimator function $\hat{\mathbf{g}}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^P$ to the observation \mathbf{y} . For a specific realization of the observation \mathbf{y} , the function value $\hat{\mathbf{g}}(\mathbf{y}) \in \mathbb{R}^P$ is an estimate of the value $\mathbf{g}(\mathbf{x})$. Note that the estimator function value $\hat{\mathbf{g}}(\mathbf{y})$ is a random quantity since it is obtained by a mapping applied to the random vector \mathbf{y} . In this thesis, we deal mainly with estimators defined via a deterministic mapping $\hat{\mathbf{g}}(\cdot)$. However, it is also possible to consider random or randomized estimators which are defined by a conditional pdf $f(\hat{\mathbf{g}}|\mathbf{y})$, i.e., even if we make the same observation \mathbf{y} twice, the corresponding outputs of the randomized estimator may be different. Such randomized estimators may be useful especially if the parameter set \mathcal{X} is finite.

Implicit in the notation $f(\hat{\mathbf{g}}|\mathbf{y})$ for the conditional pdf of the estimator outcome $\hat{\mathbf{g}}$ given the observation \mathbf{y} , is the requirement that it *cannot* depend on the parameter vector \mathbf{x} , which determines the statistics of \mathbf{y} via the statistical model $f(\mathbf{y}; \mathbf{x})$ (see [41, p. 33]). Strictly speaking, the correct notation for the conditional pdf of the estimator would be $f(\hat{\mathbf{g}}|\mathbf{y}; \mathbf{x})$, but since its values cannot depend on \mathbf{x} we are allowed to write $f(\hat{\mathbf{g}}|\mathbf{y}; \mathbf{x}) = f(\hat{\mathbf{g}}|\mathbf{y})$ with a suitable conditional pdf $f(\hat{\mathbf{g}}|\mathbf{y})$. Equivalently, a randomized estimator $\hat{\mathbf{g}}(\cdot)$ is a random function (a realization of $\hat{\mathbf{g}}(\cdot)$ is then a deterministic mapping $\mathbb{R}^M \rightarrow \mathbb{R}^P$), which is statistically independent of the observation \mathbf{y} for every $\mathbf{x} \in \mathcal{X}$.

We emphasize that all concepts and facts developed in this thesis that relate to the notion of an estimator, hold for deterministic as well as random(ized) estimators.

- **Mean squared error.** We also need a measure to compare the quality of different estimators $\hat{\mathbf{g}}(\cdot)$. A popular choice for the performance measure that will also be used in this work is the *mean squared error* (MSE)

$$\varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x}) \triangleq \mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{g}}(\mathbf{y}) - \mathbf{g}(\mathbf{x})\|_2^2\}. \quad (2.6)$$

- **Bias.** Given an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$, we define the *bias* $\mathbf{b}(\hat{\mathbf{g}}(\cdot); \mathbf{x})$ of an estimator $\hat{\mathbf{g}}(\cdot)$ as

$$\mathbf{b}(\hat{\mathbf{g}}(\cdot); \mathbf{x}) \triangleq \mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{g}}(\mathbf{y})\} - \mathbf{g}(\mathbf{x}). \quad (2.7)$$

An estimator $\hat{\mathbf{g}}(\cdot)$ whose bias vanishes identically, i.e.,

$$\mathbf{b}(\hat{\mathbf{g}}(\cdot); \mathbf{x}) = \mathbf{0} \quad \forall \mathbf{x} \in \mathcal{X} \quad (2.8)$$

is called *unbiased*. The bias of an estimator represents the “systematic” or deterministic error which is incurred by the estimator.

- **Variance and stochastic power.** Given an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$, we define the *variance* $v(\hat{\mathbf{g}}(\cdot); \mathbf{x})$ of an estimator $\hat{\mathbf{g}}(\cdot)$ as

$$v(\hat{\mathbf{g}}(\cdot); \mathbf{x}) \triangleq \mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{g}}(\mathbf{y}) - \mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{g}}(\mathbf{y})\}\|_2^2\}. \quad (2.9)$$

Similarly, we define the *stochastic power* $P(\hat{\mathbf{g}}(\cdot); \mathbf{x})$ as

$$P(\hat{\mathbf{g}}(\cdot); \mathbf{x}) \triangleq \mathbb{E}_{\mathbf{x}}\{\|\hat{\mathbf{g}}(\mathbf{y})\|_2^2\}. \quad (2.10)$$

The following relation between the estimator's mean, variance, and stochastic power can be verified easily:

$$P(\hat{\mathbf{g}}(\cdot); \mathbf{x}) = v(\hat{\mathbf{g}}(\cdot); \mathbf{x}) + \|\mathbb{E}_{\mathbf{x}}\{\hat{\mathbf{g}}(\mathbf{y})\}\|_2^2. \quad (2.11)$$

Note that two estimators that differ only on a set $\mathcal{A} \subseteq \mathbb{R}^M$ of measure zero yield the same bias, variance, MSE, and stochastic power at every $\mathbf{x} \in \mathcal{X}$. Therefore, when we say that two estimators are identical, we mean actually that they may differ only on a set of measure zero.

- **Bias - variance tradeoff.** The MSE $\varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x})$ of an estimator $\hat{\mathbf{g}}(\cdot)$ can be decomposed as the sum of two nonnegative terms:

$$\varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x}) = \|\mathbf{b}(\hat{\mathbf{g}}(\cdot); \mathbf{x})\|_2^2 + v(\hat{\mathbf{g}}(\cdot); \mathbf{x}), \quad (2.12)$$

where we used the estimator bias as defined in (2.7) and the estimator variance as defined in (2.9). For a small MSE $\varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x})$, according to (2.12), it is desirable to construct estimators that have both a small bias and a small variance. However, in general these are conflicting desiderata, i.e., an estimator with a small variance tends to have a large bias and vice versa. Therefore, to obtain a small MSE, one has to carefully balance between these two terms. It might be, e.g., that if one allows a small bias the variance can be reduced significantly, resulting in a smaller MSE as compared to unbiased estimators [44]. However, there are at least two situations where unbiased estimators are preferable. First, if the statistical model $f(\mathbf{y}; \mathbf{x})$, viewed as a function of \mathbf{y} , is highly concentrated around its mean, then the bias effectively determines the MSE and thus a small MSE requires the estimator to be effectively unbiased. The second situation where unbiasedness is advantageous is when one can observe a large number N of independent identically distributed (i.i.d.) realizations $\{\mathbf{y}_l\}_{l \in [N]}$ of the observation \mathbf{y} . It can be shown that, without loss of generality, the estimation can then be based on the sample mean $\bar{\mathbf{y}} = \frac{1}{N} \sum_{l \in [N]} \mathbf{y}_l$.⁴ Under mild conditions on the statistical model

⁴This is due to the fact that the sample mean $\bar{\mathbf{y}}$ is a sufficient statistic for the problem of estimating \mathbf{x} from the observation $\{\mathbf{y}_l\}_{l \in [N]}$ (cf. Section 4.3.6).

$f(\mathbf{y}; \mathbf{x})$, it follows from the central limit theorem [45] that the pdf of $\bar{\mathbf{y}}$ becomes more and more concentrated around its mean as N increases and therefore we have the same situation as before, i.e., the MSE is dominated by the bias and thus a small MSE requires an estimator to be effectively unbiased.

- **Optimality of estimators.** Given a specific estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$, it is natural to look for an *optimum* estimator $\hat{\mathbf{g}}(\cdot)$. Since our primary objective is a small MSE $\varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x})$, we could be tempted to search for the specific estimator that minimizes the MSE, i.e., to find the specific estimator (if it exists) that solves the problem

$$\min_{\hat{\mathbf{g}}(\cdot)} \varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x}) \quad (2.13)$$

for all $\mathbf{x} \in \mathcal{X}$ *simultaneously*.

However, it turns out that in general such an optimality criterion is not meaningful since there does not exist a single estimator that minimizes the MSE at all parameter vectors $\mathbf{x} \in \mathcal{X}$ simultaneously [41, 44], i.e., the minimization problem in (2.13) has in general no solution. This can be verified easily since the minimum MSE obtainable at any fixed parameter vector \mathbf{x}_0 is zero and achieved by the dumb estimator $\hat{\mathbf{g}}(\mathbf{y}) = \mathbf{g}(\mathbf{x}_0)$ which always yields a constant value since it completely ignores the observation \mathbf{y} . Therefore, if there existed an optimum estimator in the sense of minimum MSE, it would have to yield zero MSE for all parameter vectors \mathbf{x} and this is in general impossible. We thus have to consider alternative optimality criteria, two of which will be discussed in the following.

2.2 Minimax Estimation

A widely used notion of estimator optimality is based on the concept of *robustness*. By a robust estimator we mean an estimator that performs well under any operating condition [46, 47]. For our scope, this means that we would like to keep control of the worst-case MSE of an estimator $\hat{\mathbf{g}}(\cdot)$ for a given estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$. This worst-case MSE is defined as

$$R_{\mathcal{E}}(\hat{\mathbf{g}}(\cdot)) \triangleq \sup_{\mathbf{x} \in \mathcal{X}} \varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x}). \quad (2.14)$$

The optimal estimator in terms of robustness, called the *minimax estimator*, is defined as the solution to the problem [41]

$$\operatorname{arginf}_{\hat{\mathbf{g}}(\cdot)} R_{\mathcal{E}}(\hat{\mathbf{g}}(\cdot)) = \operatorname{arginf}_{\hat{\mathbf{g}}(\cdot)} \sup_{\mathbf{x} \in \mathcal{X}} \varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x}). \quad (2.15)$$

The infimum value $R_{\mathcal{E}} \triangleq \inf_{\hat{\mathbf{g}}(\cdot)} R_{\mathcal{E}}(\hat{\mathbf{g}}(\cdot))$ is called the *minimax risk* of the estimation problem \mathcal{E} . Finding minimax estimators or even characterizing the minimax risk (e.g., by lower bounds)

is often difficult. However, for the estimation problem corresponding to the *sparse linear model*, which will be discussed in detail in Chapter 5, some recent work (see e.g. [48–51]) derives sharp lower bounds on the minimax risk and presents practical estimation schemes that are close to the minimax estimator. In this regard, also the authors of [52, 53] present performance evaluations of existing estimation schemes revealing that they are close to the minimax estimator.

2.3 Minimum Variance Estimation

An optimality criterion different from the minimax criterion is based on the bias-variance decomposition of the MSE in (2.12). In particular, it is common to fix the estimator bias, i.e., to require that

$$\mathbf{b}(\hat{\mathbf{g}}(\cdot); \mathbf{x}) \stackrel{!}{=} \mathbf{c}(\mathbf{x}) \quad \forall \mathbf{x} \in \mathcal{X}, \quad (2.16)$$

with a given function $\mathbf{c}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$ (called the prescribed bias function), and then look for estimators that minimize the MSE.

If, for a given estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$, one considers only estimators with the same prescribed bias function $\mathbf{c}(\mathbf{x})$, minimizing the MSE $\varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x})$ is completely equivalent to minimizing the variance $v(\hat{\mathbf{g}}(\cdot); \mathbf{x})$. It is often impossible to perform a uniform minimization of the variance, and therefore we will pursue a local approach: We consider a specific parameter vector $\mathbf{x}_0 \in \mathcal{X}$ and try to minimize the variance $v(\hat{\mathbf{g}}(\cdot); \mathbf{x}_0)$ at \mathbf{x}_0 of estimators $\hat{\mathbf{g}}(\cdot)$ whose bias is equal to the prescribed function $\mathbf{c}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$, i.e., $\mathbf{b}(\hat{\mathbf{g}}(\cdot); \mathbf{x}_0) = \mathbf{c}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$.

We make some important definitions:

Definition 2.3.1. A “*minimum variance estimation problem*” or “*minimum variance problem*” \mathcal{M} is the triplet

$$\mathcal{M} \triangleq (\mathcal{E}, \mathbf{c}(\mathbf{x}), \mathbf{x}_0) \quad (2.17)$$

consisting of an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$, a prescribed bias function $\mathbf{c}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$, and a fixed parameter vector $\mathbf{x}_0 \in \mathcal{X}$ at which we will try to minimize the variance of estimators whose bias is equal to $\mathbf{c}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$.

Definition 2.3.2. Given a minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$, we denote by $\mathcal{F}(\mathcal{M})$ the set of all (possibly randomized) estimators with the prescribed bias function $\mathbf{c}(\mathbf{x})$ and with finite variance at \mathbf{x}_0 , i.e.,

$$\mathcal{F}(\mathcal{M}) \triangleq \{\hat{\mathbf{g}}(\cdot) \mid v(\hat{\mathbf{g}}(\cdot); \mathbf{x}_0) < \infty, \mathbf{b}(\hat{\mathbf{g}}(\cdot); \mathbf{x}) = \mathbf{c}(\mathbf{x}) \forall \mathbf{x} \in \mathcal{X}\}. \quad (2.18)$$

We will refer to the set $\mathcal{F}(\mathcal{M})$ as the set of allowed estimators for the minimum variance problem \mathcal{M} .

The theoretically achievable performance of any estimator for a minimum variance problem is characterized by

Definition 2.3.3. *Given a minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$, the minimum achievable variance at the parameter vector \mathbf{x}_0 is defined as*

$$L_{\mathcal{M}} \triangleq \inf_{\hat{\mathbf{g}}(\cdot) \in \mathcal{F}(\mathcal{M})} v(\hat{\mathbf{g}}(\cdot); \mathbf{x}_0). \quad (2.19)$$

If there is no allowed estimator for \mathcal{M} , i.e., the set $\mathcal{F}(\mathcal{M})$ is empty, we define $L_{\mathcal{M}} \triangleq \infty$.

In the literature, the minimum achievable variance is sometimes referred to as the *Barankin bound* (in particular if unbiased estimation is considered).

The most important notion of estimator optimality in the context of minimum variance estimation is stated in

Definition 2.3.4 (Locally Minimum Variance Estimation). *Given a minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$, we call any (possibly randomized) estimator $\hat{\mathbf{g}}^{(\mathbf{x}_0)}(\cdot) \in \mathcal{F}(\mathcal{M})$ whose variance at \mathbf{x}_0 attains $L_{\mathcal{M}}$, i.e.,*

$$v(\hat{\mathbf{g}}^{(\mathbf{x}_0)}(\cdot); \mathbf{x}_0) = L_{\mathcal{M}}, \quad (2.20)$$

a locally minimum variance (LMV) estimator for \mathcal{M} . For the special case where the prescribed bias function is identically zero, i.e., $\mathbf{c}(\cdot) \equiv \mathbf{0}$, such an estimator is called a locally minimum variance unbiased (LMVU) estimator.

While an LMV estimator in general performs well (has a small variance) only locally around a specific parameter vector \mathbf{x}_0 , there may be estimators which have the prescribed bias $\mathbf{c}(\cdot)$ and moreover perform well (have a small variance) everywhere, i.e., for any parameter value $\mathbf{x} \in \mathcal{X}$:

Definition 2.3.5 (Uniformly Minimum Variance Estimation). *Given an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$, a prescribed bias function $\mathbf{c}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$ and a parameter vector $\mathbf{x}_0 \in \mathcal{X}$, we denote by $\mathcal{M}(\mathbf{x}_0)$ the specific minimum variance problem that is given by $\mathcal{M}(\mathbf{x}_0) = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$. Then, we call any estimator $\hat{\mathbf{g}}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^P$ which is an LMV estimator for $\mathcal{M}(\mathbf{x}_0)$ simultaneously for every $\mathbf{x}_0 \in \mathcal{X}$, a uniformly minimum variance (UMV) estimator for the given estimation problem \mathcal{E} and bias function $\mathbf{c}(\cdot)$. For the special case where the bias function is identically zero, i.e., $\mathbf{c}(\cdot) \equiv \mathbf{0}$, such an estimator is called a uniformly minimum variance unbiased (UMVU) estimator for the estimation problem \mathcal{E} .*

2.3.1 Vector-Valued vs. Scalar-Valued Parameter Function

So far, we considered the estimation of the value $\mathbf{g}(\mathbf{x})$ of a vector-valued function $\mathbf{g}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$ of the parameter vector \mathbf{x} . However, the minimum variance estimation of a P -dimensional

vector-valued function $\mathbf{g}(\mathbf{x}) = (g_1(\mathbf{x}), \dots, g_P(\mathbf{x}))^T$ can be reduced to separately estimating the P scalar functions $\{g_k(\mathbf{x})\}_{k \in [P]}$. This is possible because on the one hand we have that the variance $v(\hat{\mathbf{g}}(\cdot); \mathbf{x})$ can be decomposed as $v(\hat{\mathbf{g}}(\cdot); \mathbf{x}) = \sum_{k \in [P]} v(\hat{g}_k(\cdot); \mathbf{x})$, and on the other hand the bias constraint decomposes elementwise, i.e., $\mathbf{b}(\hat{\mathbf{g}}(\cdot); \mathbf{x}) = \mathbf{c}(\mathbf{x})$ if and only if $b(\hat{g}_k(\cdot); \mathbf{x}) = c_k(\mathbf{x})$ for every $k \in [P]$. Thus, given a minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$ with prescribed bias $\mathbf{c}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$, we have that the minimum achievable variance at \mathbf{x}_0 is given as

$$L_{\mathcal{M}} = \sum_{k \in [P]} L_{\mathcal{M}_k}, \quad (2.21)$$

where $L_{\mathcal{M}_k}$ denotes the minimum achievable variance for the “scalar” minimum variance problem $\mathcal{M}_k = (\mathcal{E}_k, c_k(\cdot), \mathbf{x}_0)$, with \mathcal{E}_k denoting the associated scalar estimation problems, i.e., $\mathcal{E}_k \triangleq (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g_k(\cdot))$. Here, $g_k(\cdot)$ and $c_k(\cdot)$ denote the k th component of the parameter function $\mathbf{g}(\cdot)$ and the prescribed bias function $\mathbf{c}(\cdot)$, respectively. Furthermore, if the estimators $\{\hat{g}_k(\cdot)\}_{k \in [P]}$ are LMV estimators for the minimum variance problems \mathcal{M}_k , then the vector-valued estimator given by $\hat{\mathbf{g}}(\cdot) = (\hat{g}_1(\cdot), \dots, \hat{g}_P(\cdot))^T$ is the LMV estimator for \mathcal{M} , and vice versa.

Therefore, in the context of minimum variance estimation, we can restrict ourselves without loss of generality to estimation problems $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ with a scalar-valued parameter function $g(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$, i.e., we can assume $P = 1$ in (2.5). Note that the restriction to scalar-valued parameter functions is in general not possible in the context of minimax estimation since the optimization problem (2.15) does not decompose in general.

2.3.2 Lower Bounds via Generalized Inequalities

Some results on minimum variance estimation (e.g., [25, 29]) are stated in terms of generalized (matrix) inequalities [30] involving the estimator covariance matrix $\text{cov}_{\mathbf{x}_0}\{\hat{\mathbf{g}}(\mathbf{y})\} \in \mathbb{R}^{P \times P}$ defined as

$$\text{cov}_{\mathbf{x}_0}\{\hat{\mathbf{g}}(\mathbf{y})\} \triangleq \mathbf{E}_{\mathbf{x}_0}\{[\hat{\mathbf{g}}(\mathbf{y}) - \mathbf{E}_{\mathbf{x}_0}\{\hat{\mathbf{g}}(\mathbf{y})\}][\hat{\mathbf{g}}(\mathbf{y}) - \mathbf{E}_{\mathbf{x}_0}\{\hat{\mathbf{g}}(\mathbf{y})\}]^T\}. \quad (2.22)$$

Note that $\text{cov}_{\mathbf{x}_0}\{\hat{\mathbf{g}}(\mathbf{y})\}$ is a psd matrix by its very definition. A typical result is then a lower bound of the form

$$\text{cov}_{\mathbf{x}_0}\{\hat{\mathbf{g}}(\mathbf{y})\} \geq \mathbf{L}, \quad (2.23)$$

where the lower bound $\mathbf{L} \in \mathbb{R}^{P \times P}$ is a psd matrix that depends of course on the underlying estimation problem. By contrast, we derive exclusively lower bounds on the variance $v(\hat{\mathbf{g}}(\cdot); \mathbf{x}_0)$, which is always a scalar quantity.

However, since $v(\hat{\mathbf{g}}(\cdot); \mathbf{x}_0) = \text{Tr}\{\text{cov}_{\mathbf{x}_0}\{\hat{\mathbf{g}}(\mathbf{y})\}\}$, we have that any bound of the form (2.23) induces also a lower bound on $v(\hat{\mathbf{g}}(\cdot); \mathbf{x}_0)$ via

$$v(\hat{\mathbf{g}}(\cdot); \mathbf{x}_0) = \text{Tr}\{\text{cov}_{\mathbf{x}_0}\{\hat{\mathbf{g}}(\mathbf{y})\}\} \geq \text{Tr}\{\mathbf{L}\}, \quad (2.24)$$

where the last inequality follows from the fact that for two psd matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{P \times P}$, we have $\mathbf{A} \geq \mathbf{B} \Rightarrow \text{Tr}\{\mathbf{A}\} \geq \text{Tr}\{\mathbf{B}\}$ [30].

2.3.3 Exchangeability of Prescribed Bias and Parameter Function

A minimum variance problem \mathcal{M} depends on the prescribed bias $c(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$ and the parameter function $g(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$. However, the minimum achievable variance $L_{\mathcal{M}}$ as well as the corresponding LMV estimator do not depend *independently* on $c(\cdot)$ and $g(\cdot)$, as stated in

Theorem 2.3.1. *Consider two estimation problems $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and $\mathcal{E}' = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g'(\cdot))$ which share a common parameter set \mathcal{X} and statistical model $f(\mathbf{y}; \mathbf{x})$. We have that any estimator $\hat{g}(\cdot)$ for \mathcal{E} with bias $c(\cdot)$ is also an estimator for \mathcal{E}' with bias $c'(\cdot) = c(\cdot) + g(\cdot) - g'(\cdot)$. For any two minimum variance problems $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$, $\mathcal{M}' = (\mathcal{E}', c'(\cdot), \mathbf{x}_0)$ associated with \mathcal{E} and \mathcal{E}' , respectively, with a common fixed parameter vector $\mathbf{x}_0 \in \mathcal{X}$ but different prescribed bias functions $c(\cdot)$ and $c'(\cdot)$, we have that if*

$$g(\cdot) + c(\cdot) = g'(\cdot) + c'(\cdot), \quad (2.25)$$

then the minimum achievable variance is the same for \mathcal{M} and \mathcal{M}' , i.e.,

$$L_{\mathcal{M}} = L_{\mathcal{M}'}. \quad (2.26)$$

Furthermore, if an estimator $\hat{g}(\cdot)$ is an LMV estimator for \mathcal{M} , it is also an LMV estimator for \mathcal{M}' .

Proof. Consider an estimator $\hat{g}(\cdot)$ for \mathcal{E} with bias $c(\cdot)$, i.e., $b(\hat{g}(\cdot); \mathbf{x}) = c(\mathbf{x})$ for any $\mathbf{x} \in \mathcal{X}$. This implies by (2.7) that $\mathbb{E}_{\mathbf{x}_0}\{\hat{g}(\mathbf{y})\} = g(\mathbf{x}) + c(\mathbf{x})$ for any $\mathbf{x} \in \mathcal{X}$. However, if we use the same estimator for \mathcal{E}' we have that its bias is $b(\hat{g}(\cdot); \mathbf{x}) = \mathbb{E}_{\mathbf{x}_0}\{\hat{g}(\mathbf{y})\} - g'(\mathbf{x}) = g(\mathbf{x}) + c(\mathbf{x}) - g'(\mathbf{x})$. Furthermore, since both estimation problems \mathcal{E} and \mathcal{E}' share the same statistical model $f(\mathbf{y}; \mathbf{x})$, the variance $v(\hat{g}(\cdot); \mathbf{x})$ (cf. (2.9)) is the same for both estimation problems. It follows that if the two minimum variance problems \mathcal{M} and \mathcal{M}' satisfy (2.25), the associated sets of allowed estimators coincide, i.e., $\mathcal{F}(\mathcal{M}) = \mathcal{F}(\mathcal{M}')$, and moreover for each allowed estimator, the corresponding value of the objective in (2.19) is the same for \mathcal{M} and \mathcal{M}' , i.e., $L_{\mathcal{M}} = L_{\mathcal{M}'}$. \square

According to the definition of the minimum achievable variance in (2.19), it is possible that for a given minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$ there is no LMV estimator, i.e., no estimator whose bias is equal to $\mathbf{c}(\mathbf{x})$ for all $\mathbf{x} \in \mathcal{X}$ and whose variance at \mathbf{x}_0 is finite and equal to $L_{\mathcal{M}}$. This motivates

Definition 2.3.6. *Given a minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$, we call the prescribed bias function $\mathbf{c}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$ valid for \mathcal{M} if $L_{\mathcal{M}}$ is finite and there exists at least one estimator with bias $\mathbf{c}(\mathbf{x})$ whose variance at \mathbf{x}_0 equals $L_{\mathcal{M}}$.*

Similarly, we have [3, 35]

Definition 2.3.7. *Given a minimum variance problem with zero bias, $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot) \equiv 0, \mathbf{x}_0)$, with associated estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$, we call a parameter function $\mathbf{g}(\mathbf{x})$ estimable for \mathcal{M} if $L_{\mathcal{M}}$ is finite and there exists at least one unbiased estimator for $g(\mathbf{x})$ whose variance at \mathbf{x}_0 equals $L_{\mathcal{M}}$.*

Due to Theorem 2.3.1, we have

Corollary 2.3.2. *Given the minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$ associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$, the prescribed bias function $c(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$ is valid for \mathcal{M} if and only if the parameter function $g'(\cdot) = g(\cdot) + c(\cdot)$ of the estimation problem $\mathcal{E}' = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ is estimable for the minimum variance problem $\mathcal{M}' = (\mathcal{E}', c'(\cdot) \equiv 0, \mathbf{x}_0)$.*

2.3.4 Existence and Uniqueness of Minimum Variance Estimators

As already mentioned above, it is not guaranteed that there exists an LMV estimator for a given minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$ associated to an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$. Discussions on the conditions for the existence of an LMV estimator can be found in the seminal works [54, 55]. It is interesting to note that there need not exist an LMV estimator even if the set $\mathcal{F}(\mathcal{M})$ is nonempty, i.e., even if there is at least one estimator $\hat{\mathbf{g}}(\cdot)$ with bias $\mathbf{c}(\mathbf{x})$ and whose variance at \mathbf{x}_0 is finite. An example of a minimum variance problem where this is the case can be found in [54, p. 407]. The difficulty in establishing the existence of an LMV estimator, i.e., an allowed estimator whose variance at \mathbf{x}_0 attains the infimum in (2.19), results from the nontrivial topological characterization of the set $\mathcal{F}(\mathcal{M})$ of allowed estimators. While one can easily verify that this set of estimators is an affine subset [30, p. 21] of a specific Hilbert space, it is in general not true that this set is closed, i.e., there might exist a Cauchy sequence of estimators belonging to $\mathcal{F}(\mathcal{M})$ which has no limit that belongs to $\mathcal{F}(\mathcal{M})$.⁵ We will discuss this issue in more detail in Section 4.2. In any case, if a LMV estimator exists it is unique, as stated in

Theorem 2.3.3. *Consider a minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$. If an LMV estimator exists for \mathcal{M} , it is unique.*

Proof. The uniqueness of an LMVU estimator, i.e., the uniqueness of an LMV estimator for the special case of a minimum variance problem with zero bias, $\mathbf{c}(\mathbf{x}) = \mathbf{0}$, has been proven in the seminal works [54, 55]. The statement follows then for an arbitrary prescribed bias function $\mathbf{c}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$ by Theorem 2.3.1. \square

A direct consequence of Theorem 2.3.3, which is also proved in [45] is stated in

⁵We will introduce the concepts of closedness, limits, and Cauchy sequences in Chapter 3.

Theorem 2.3.4. *Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$ and a prescribed bias function $\mathbf{c}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$. If an UMV estimator exists for \mathcal{E} and $\mathbf{c}(\cdot)$, it is unique.*

Proof. This result follows straightforwardly from Theorem 2.3.3 since the UMV is necessarily also the unique LMV estimator for any minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$, which is obtained from \mathcal{E} and $\mathbf{c}(\cdot)$ by choosing an arbitrary value for $\mathbf{x}_0 \in \mathcal{X}$. If there would be two different UMV estimators, they would be at the same time two different LMV estimators for \mathcal{M} which contradicts Theorem 2.3.3. \square

2.3.5 Transformations of the Parameter Function

Consider an estimation problem $\mathcal{E}_1 = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g_1(\cdot))$ and a modified estimation problem $\mathcal{E}_2 = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g_2(\cdot))$ that is obtained from \mathcal{E}_1 by using the parameter function $g_2(\mathbf{x})$ instead of $g_1(\mathbf{x})$. If the two parameter functions are related by $g_2(\mathbf{x}) = g_1(\mathbf{x}) + a$, where $a \in \mathbb{R}$ is a fixed constant, we have obviously that if an estimator $\hat{g}_1(\cdot)$ is an unbiased estimator for the estimation problem \mathcal{E}_1 , then the estimator $\hat{g}_2(\cdot) \triangleq \hat{g}_1(\cdot) + a$ is an unbiased estimator for the estimation problem \mathcal{E}_2 (cf. Theorem 2.3.1). Moreover, this estimator has the same variance as the original estimator, i.e., $v_{\mathcal{E}_2}(\hat{g}_2(\cdot); \mathbf{x}) = v_{\mathcal{E}_1}(\hat{g}_1(\cdot); \mathbf{x})$ for every $\mathbf{x} \in \mathcal{X}$, where $v_{\mathcal{E}}(\hat{g}(\cdot); \mathbf{x})$ denotes the variance $v(\hat{g}(\cdot); \mathbf{x})$ at \mathbf{x} of an estimator $\hat{g}(\cdot)$ that uses an observation \mathbf{y} whose statistics are determined by the statistical model $f(\mathbf{y}; \mathbf{x})$ of the estimation problem \mathcal{E} . By a slight generalization of this simple example, we arrive at

Theorem 2.3.5. *Consider two estimation problems $\mathcal{E}_1 = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}_1(\cdot))$, $\mathcal{E}_2 = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}_2(\cdot))$ and the associated minimum variance problems $\mathcal{M}_1 = (\mathcal{E}_1, \mathbf{c}(\cdot), \mathbf{x}_0)$, $\mathcal{M}_2 = (\mathcal{E}_2, \mathbf{c}(\cdot), \mathbf{x}_0)$ that are identical except for the parameter functions $\mathbf{g}_1(\mathbf{x})$ and $\mathbf{g}_2(\mathbf{x})$, respectively. If the parameter functions are related by*

$$\mathbf{g}_2(\cdot) = \mathbf{U}\mathbf{g}_1(\cdot) + \mathbf{a} \quad (2.27)$$

where $\mathbf{a} \in \mathbb{R}^P$ is a constant vector and $\mathbf{U} \in \mathbb{R}^{P \times P}$ is an orthonormal matrix, i.e., $\mathbf{U}^T \mathbf{U} = \mathbf{I}$, then we have that

$$L_{\mathcal{M}_1} = L_{\mathcal{M}_2}. \quad (2.28)$$

Moreover, if an estimator $\hat{\mathbf{g}}_1(\cdot)$ is the LMV estimator for \mathcal{M}_1 , then the estimator $\hat{\mathbf{g}}_2(\cdot) \triangleq \mathbf{U}\hat{\mathbf{g}}_1(\cdot) + \mathbf{a}$ is the LMV estimator for \mathcal{M}_2 .

Proof. The statement is proven by showing that there is a one-to-one correspondence between the sets $\mathcal{F}(\mathcal{M}_1)$ and $\mathcal{F}(\mathcal{M}_2)$. Indeed, to any estimator $\hat{\mathbf{g}}_1(\cdot) \in \mathcal{F}(\mathcal{M}_1)$, we can associate uniquely the estimator $\hat{\mathbf{g}}_2(\cdot) \triangleq \mathbf{U}\hat{\mathbf{g}}_1(\cdot) + \mathbf{a} \in \mathcal{F}(\mathcal{M}_2)$. Moreover, the value of the objective in (2.19), i.e., $v(\hat{\mathbf{g}}(\cdot); \mathbf{x}_0)$, is the same for $\hat{\mathbf{g}}_1(\cdot)$ and $\hat{\mathbf{g}}_2(\cdot)$. \square

2.3.6 Relaxing The Bias Equality Constraints

Instead of requiring the estimator bias to equal a given function $\mathbf{c}(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{X}$, one may pose inequality constraints on the bias. This is pursued by the authors of [56] and [57], who consider minimum variance estimation with inequality constraints placed on the bias gradient. More specifically, they require the norm of the bias gradient matrix to be upper bounded by a prescribed maximum value.

2.4 Transformation of the Observation

In some situations it might be desirable not to work with the observation \mathbf{y} of a given estimation problem directly, but to perform some “preprocessing” yielding a modified or transformed observation \mathbf{z} which is easier to process for some reason. As an example for such a preprocessing consider e.g. a compression or dimension reduction as is done within CS by multiplying the observation $\mathbf{y} \in \mathbb{R}^M$ with a CS measurement matrix $\mathbf{M} \in \mathbb{R}^{L \times M}$ (typically $L \ll M$) to yield a small number of compressive measurements z_k , with $\mathbf{z} = \mathbf{M}\mathbf{y}$. However, in general we cannot expect that the achievable estimation accuracy remains unchanged if we use the transformed observation. The next two subsections discuss the influence of a transformation of the observation, in the context of minimum variance estimation, for the two cases where the transformation is invertible and where it is not.

2.4.1 Invertible Transformation – Invariance of Classical Estimation Problems

Consider the problem of estimating the value $g(\mathbf{x})$ of the parameter function $g(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$ using the observation $\mathbf{y} \in \mathbb{R}^M$. Then, intuitively, the estimation problem does not become any harder if we use not directly \mathbf{y} but a transformed observation $\mathbf{z} = \mathbf{T}(\mathbf{y})$, where $\mathbf{T}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^K$ denotes an invertible deterministic function (this requires that $K \geq M$), as the new observation. Indeed, we have

Theorem 2.4.1. *Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$ with a statistical model $f(\mathbf{y}; \mathbf{x})$ and an invertible deterministic mapping $\mathbf{T}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^K$. We denote by \mathcal{E}' the estimation problem that is obtained from \mathcal{E} by using the vector $\mathbf{z} = \mathbf{T}(\mathbf{y})$ as the observation, i.e., $\mathcal{E}' = (\mathcal{X}, f(\mathbf{z}; \mathbf{x}), \mathbf{g}(\cdot))$ where $f(\mathbf{z}; \mathbf{x})$ denotes the pdf of the new observation $\mathbf{z} = \mathbf{T}(\mathbf{y})$. We then have that for any two minimum variance problems $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$, $\mathcal{M}' = (\mathcal{E}', \mathbf{c}(\cdot), \mathbf{x}_0)$ which share a common prescribed bias function $\mathbf{c}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$ and a fixed parameter vector $\mathbf{x}_0 \in \mathcal{X}$,*

$$L_{\mathcal{M}} = L_{\mathcal{M}'}. \quad (2.29)$$

Moreover, the minimax risks of \mathcal{E} and \mathcal{E}' satisfy

$$R_{\mathcal{E}} = R_{\mathcal{E}'}. \quad (2.30)$$

Proof. This statement is proven by showing that the optimization problems (2.19) and (2.15) which define the quantities $L_{\mathcal{M}}$, $L_{\mathcal{M}'}$ and $R_{\mathcal{E}}$, $R_{\mathcal{E}'}$ respectively are equivalent for \mathcal{E} , \mathcal{E}' and \mathcal{M} , \mathcal{M}' . Indeed, for any estimator $\hat{\mathbf{g}}(\mathbf{y})$ for \mathcal{E} and \mathcal{M} that uses the observation \mathbf{y} we have that the specific estimator $\hat{\mathbf{g}}'(\mathbf{z}) \triangleq \hat{\mathbf{g}}(\mathbf{T}^{-1}(\mathbf{z}))$ that uses the observation $\mathbf{z} = \mathbf{T}(\mathbf{y})$ has the same bias and variance at any $\mathbf{x} \in \mathcal{X}$. The same goes the other way, i.e., given any estimator $\hat{\mathbf{g}}'(\mathbf{z})$ for \mathcal{E}' and \mathcal{M}' that uses the observation \mathbf{z} we have that the specific estimator $\hat{\mathbf{g}}(\mathbf{y}) \triangleq \hat{\mathbf{g}}'(\mathbf{T}(\mathbf{y}))$ that uses the observation \mathbf{y} has the same bias and variance at any $\mathbf{x} \in \mathcal{X}$. Thus we have that $\hat{\mathbf{g}}'(\mathbf{z}) \in \mathcal{F}(\mathcal{M}') \Leftrightarrow \hat{\mathbf{g}}(\mathbf{y}) \in \mathcal{F}(\mathcal{M})$, i.e., $b(\hat{\mathbf{g}}(\mathbf{y}); \mathbf{x}) = b(\hat{\mathbf{g}}'(\mathbf{z}); \mathbf{x})$ for every $\mathbf{x} \in \mathcal{X}$, and $v(\hat{\mathbf{g}}(\mathbf{y}); \mathbf{x}_0) = v(\hat{\mathbf{g}}'(\mathbf{z}); \mathbf{x}_0)$. \square

2.4.2 Non-invertible Transformation – Data Processing Inequality for Classical Estimation

As discussed in the previous section, replacing the observation \mathbf{y} by $\mathbf{z} = \mathbf{T}(\mathbf{y})$ where $\mathbf{T}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^K$ is an arbitrary (but known) invertible deterministic function is irrelevant w.r.t. minimum variance and minimax estimation. Loosely speaking, an invertible transformation is just a “relabeling” of the observed data, i.e., the realization of the observation, which does not incur any gain or loss of information.

However, if the transformation $\mathbf{T}(\cdot)$ applied to the observation is not invertible or even random, the situation changes. While in general the achievable performance for minimum variance and minimax estimators is different for the transformed estimation problem \mathcal{E}' whose statistical model is given by $f(\mathbf{z}; \mathbf{x})$, we still have a relation between the modified and the original estimation problem. This relation is due to the fact that any estimator $\hat{\mathbf{g}}(\cdot)$ that uses the new observation \mathbf{z} is also an (possibly random) estimator for the original estimation problem with the observation \mathbf{y} since $\hat{\mathbf{g}}(\mathbf{z}) = \hat{\mathbf{g}}(\mathbf{T}(\mathbf{y}))$. More precisely, we have

Theorem 2.4.2. *Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$ with a statistical model $f(\mathbf{y}; \mathbf{x})$ and a (possibly random) function $\mathbf{T}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^K$. If the function $\mathbf{T}(\cdot)$ is random we assume that it is statistically independent of \mathbf{y} for every $\mathbf{x} \in \mathcal{X}$. We denote by \mathcal{E}' the estimation problem that is obtained from \mathcal{E} by using the vector $\mathbf{z} = \mathbf{T}(\mathbf{y})$ as the observation, i.e., $\mathcal{E}' = (\mathcal{X}, f(\mathbf{z}; \mathbf{x}), \mathbf{g}(\cdot))$ where $f(\mathbf{z}; \mathbf{x})$ denotes the pdf of the new observation $\mathbf{z} = \mathbf{T}(\mathbf{y})$. We then have that for any two minimum variance problems given by $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$, $\mathcal{M}' = (\mathcal{E}', \mathbf{c}(\cdot), \mathbf{x}_0)$ which share a common prescribed bias function $\mathbf{c}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$ and a fixed parameter vector $\mathbf{x}_0 \in \mathcal{X}$,*

$$L_{\mathcal{M}} \leq L_{\mathcal{M}'}. \quad (2.31)$$

Moreover, the minimax risks of \mathcal{E} and \mathcal{E}' satisfy

$$R_{\mathcal{E}} \leq R_{\mathcal{E}'}. \quad (2.32)$$

Proof. The statement follows from the fact that for any estimator $\hat{\mathbf{g}}'(\mathbf{z})$ for \mathcal{E}' and \mathcal{M}' that uses the observation $\mathbf{z} = \mathbf{T}(\mathbf{y})$, we have that the specific (possibly random) estimator $\hat{\mathbf{g}}(\mathbf{y}) \triangleq \hat{\mathbf{g}}'(\mathbf{T}(\mathbf{y}))$ for \mathcal{E} and \mathcal{M} that uses the observation \mathbf{y} has the same bias and variance at any $\mathbf{x} \in \mathcal{X}$.

In particular, assume that $L_{\mathcal{M}'} < L_{\mathcal{M}}$ which would imply, via Definition 2.3.3, that there must be an allowed estimator $\hat{\mathbf{g}}'_0(\mathbf{z})$ for \mathcal{M}' whose variance at \mathbf{x}_0 is strictly smaller than $L_{\mathcal{M}}$, i.e., $v(\hat{\mathbf{g}}'_0(\mathbf{z}); \mathbf{x}_0) < L_{\mathcal{M}}$. We can then construct an allowed estimator for \mathcal{M} as $\hat{\mathbf{g}}_0(\mathbf{y}) \triangleq \hat{\mathbf{g}}'_0(\mathbf{T}(\mathbf{y}))$ which has the same variance $v(\hat{\mathbf{g}}'_0(\mathbf{z}); \mathbf{x}_0) < L_{\mathcal{M}}$ at \mathbf{x}_0 . This, however, is a contradiction due to the definition of $L_{\mathcal{M}}$. A similar argument holds for the minimax risks of the estimation problems \mathcal{E} and \mathcal{E}' .

Thus the infimum in (2.19) and (2.15) cannot decrease by moving from \mathcal{E} to \mathcal{E}' or from \mathcal{M} to \mathcal{M}' , respectively. \square

2.5 Reducing the Parameter Set

The following definition will prove handy:

Definition 2.5.1. Consider a minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$ associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$ and an arbitrary subset $\mathcal{X}' \subseteq \mathcal{X}$ of the parameter set of \mathcal{M} . We then denote by

$$\mathcal{M}|_{\mathcal{X}'} \triangleq \left(\mathcal{E}', \mathbf{c}(\cdot)|_{\mathcal{X}'}, \mathbf{x}_0 \right) \quad (2.33)$$

the specific minimum variance problem associated with the estimation problem

$$\mathcal{E}' \triangleq \left(\mathcal{X}', f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot)|_{\mathcal{X}'} \right) \quad (2.34)$$

which is identical to \mathcal{M} except for the parameter set.

Given an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$, it may be known beforehand that the true parameter vector \mathbf{x} belongs to a subset $\mathcal{X}' \subseteq \mathcal{X}$ of the “nominal” parameter set \mathcal{X} . Intuitively, this a priori knowledge should be exploitable in order to improve the performance of estimators for $\mathbf{g}(\mathbf{x})$. A precise statement in this regard for minimax estimation is

Theorem 2.5.1. Given two estimation problems $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$, $\mathcal{E}' = (\mathcal{X}', f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$ which differ only in their parameter sets \mathcal{X} and \mathcal{X}' , we have

$$\mathcal{X}' \subseteq \mathcal{X} \Rightarrow R_{\mathcal{E}'} \leq R_{\mathcal{E}}, \quad (2.35)$$

i.e., reducing the parameter set can never increase the minimax risk of an estimation problem.

Proof. For any estimator $\hat{\mathbf{g}}(\cdot)$, we have due to $\mathcal{X}' \subseteq \mathcal{X}$ the following relation for the worst MSE's associated with \mathcal{E} and \mathcal{E}' :

$$R_{\mathcal{E}'}(\hat{\mathbf{g}}(\cdot)) = \sup_{\mathbf{x} \in \mathcal{X}'} \varepsilon(\hat{\mathbf{g}}(\cdot), \mathbf{x}) \leq \sup_{\mathbf{x} \in \mathcal{X}} \varepsilon(\hat{\mathbf{g}}(\cdot), \mathbf{x}) = R_{\mathcal{E}}(\hat{\mathbf{g}}(\cdot)). \quad (2.36)$$

Since this inequality holds for every estimator and since $R_{\mathcal{E}}(\hat{\mathbf{g}}(\cdot)) \triangleq \sup_{\mathbf{x} \in \mathcal{X}} \varepsilon(\hat{\mathbf{g}}(\cdot); \mathbf{x})$, we obtain (2.35). \square

Compared to minimax estimation, the effect of reducing the parameter set of an estimation problem in the context of minimum variance estimation is more subtle [25, 26, 29, 58–60]. As one may intuitively expect, the achievable performance as quantified by the minimum achievable variance cannot become worse. However, this improved performance bound may not be achieved by incorporating the a-priori information of a reduced parameter set directly into the design of an estimator. E.g., if we are given a minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$ associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), \mathbf{g}(\cdot))$ and it is known beforehand that the true parameter vector \mathbf{x} is contained within $\mathcal{X}' \subseteq \mathcal{X}$, then it might be reasonable to constrain an estimator $\hat{\mathbf{g}}(\mathbf{y})$ for \mathcal{M} to take on values only within the set $\mathbf{g}(\mathcal{X}') \subseteq \mathbb{R}^P$. However, for minimum variance estimation it may not be optimal to place such an additional constraint on an estimator. The real effect of a reduced parameter set for minimum variance estimation is that the set of allowed estimators $\mathcal{F}(\mathcal{M})$ becomes larger if the parameter set \mathcal{X} is reduced to \mathcal{X}' since the bias constraint defining $\mathcal{F}(\mathcal{M})$ (see (2.18)) has to be satisfied only on the smaller set \mathcal{X}' instead of the original parameter set \mathcal{X} . Since the set of allowed estimators can only become larger, the minimum achievable variance (cf. (2.19)) can never increase, as is stated in

Theorem 2.5.2. *Consider a minimum variance problem $\mathcal{M} = (\mathcal{E}, \mathbf{c}(\cdot), \mathbf{x}_0)$ and an arbitrary subset $\mathcal{X}' \subseteq \mathcal{X}$ of the parameter set of \mathcal{M} . For any modified minimum variance problem $\mathcal{M}' = \mathcal{M}|_{\mathcal{X}'}$, associated with the smaller parameter set $\mathcal{X}' \subseteq \mathcal{X}$, we have*

$$L_{\mathcal{M}'} \leq L_{\mathcal{M}}, \quad (2.37)$$

i.e., a reduction of the parameter set can never result in an increase of the minimum achievable variance.

Furthermore, if we consider a minimum variance problem \mathcal{M} with zero bias, $\mathbf{c}(\cdot) \equiv 0$, and an arbitrary parameter function $\mathbf{g}(\cdot) : \mathcal{X} \rightarrow \mathbb{R}^P$, we have that if the parameter function $\mathbf{g}(\cdot)|_{\mathcal{X}'}$ is not estimable for \mathcal{M}' , then the parameter function $\mathbf{g}(\cdot)$ is also not estimable for \mathcal{M} .

Proof. This result follows from (2.19) by the obvious fact that under the condition $\mathcal{X}' \subseteq \mathcal{X}$ we have that $\mathcal{F}(\mathcal{M}) \subseteq \mathcal{F}(\mathcal{M}')$, i.e., any allowed estimator for \mathcal{M} is necessarily also an allowed estimator for \mathcal{M}' . \square

2.6 Statistical Models Belonging to an Exponential Family

An important class of estimation problems is obtained by assuming that the statistical model $f(\mathbf{y}; \mathbf{x})$ belongs to an exponential family [34, 41, 61]. An exponential family is a class of pdfs that is associated with two fixed vector-valued functions $\Phi(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^K$, $\mathbf{u}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}^K$ of the observation \mathbf{y} and the parameter vector \mathbf{x} , respectively:

$$f^{(\Phi, \mathbf{u})}(\mathbf{y}; \mathbf{x}) = \exp\left([\Phi(\mathbf{y})]^T \mathbf{u}(\mathbf{x}) - A^{(\Phi)}(\mathbf{x})\right) h(\mathbf{y}) \quad (2.38)$$

where $h(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}$ is a fixed nonnegative weight function. Associated with an exponential family is the *natural parameter space* [61] denoted $N^{(\Phi)}$ and defined as

$$N^{(\Phi)} \triangleq \left\{ \mathbf{x} \in \mathbb{R}^N \mid \int_{\mathbf{y}} \exp\left([\Phi(\mathbf{y})]^T \mathbf{u}(\mathbf{x})\right) h(\mathbf{y}) d\mathbf{y} < \infty \right\}. \quad (2.39)$$

The function $A^{(\Phi)}(\mathbf{x}) : N^{(\Phi)} \rightarrow \mathbb{R}$ is defined by the requirement that $\int_{\mathbf{y}} f^{(\Phi, \mathbf{u})}(\mathbf{y}; \mathbf{x}) d\mathbf{y} = 1$ for every $\mathbf{x} \in N^{(\Phi)}$, i.e.,

$$A^{(\Phi)}(\mathbf{x}) \triangleq \log \int_{\mathbf{y}} \exp\left([\Phi(\mathbf{y})]^T \mathbf{u}(\mathbf{x})\right) h(\mathbf{y}) d\mathbf{y}. \quad (2.40)$$

In the literature, the function $\Phi(\cdot)$ is known as the *potential function* or *sufficient statistic* and the function $A^{(\Phi)}(\cdot)$ is known as the *log partition function* or *cumulant (generating) function* [34, 61].

It can be shown [61] that under some weak technical requirements on the parameter function $\mathbf{u}(\cdot)$, every exponential family can be reduced to a standard form, with parameter function $\mathbf{u}'(\mathbf{x}) = \mathbf{x}$. In particular, any estimation problem with parameter set \mathcal{X} and statistical model $f^{(\Phi, \mathbf{u})}(\mathbf{y}; \mathbf{x})$ is equivalent to an estimation problem with parameter set $\mathcal{X}' = \mathbf{u}(\mathcal{X})$ and statistical model $f^{(\Phi, \mathbf{u}'(\mathbf{x}=\mathbf{x}))}(\mathbf{y}; \mathbf{x})$.

By definition, every exponential family in standard form, i.e., $\mathbf{u}(\mathbf{x}) = \mathbf{x}$, is completely characterized by the log partition function $A^{(\Phi)}(\cdot) : N^{(\Phi)} \rightarrow \mathbb{R}$ which has some interesting properties. In particular, for an exponential family in standard form, according to [34, 61] for every \mathbf{x}_0 in the interior of $N^{(\Phi)}$, i.e., there exists a radius $r > 0$ such that $\mathcal{B}(\mathbf{x}_0, r) \subseteq N^{(\Phi)}$, the partial derivatives $\left. \frac{\partial^{\mathbf{p}} A^{(\Phi)}(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \right|_{\mathbf{x}=\mathbf{x}_0}$ up to any order $\mathbf{p} \in \mathbb{Z}_+^N$ exist and are given by the cumulants of the pdf $f^{(\Phi)}(\mathbf{y}; \mathbf{x}_0)$. In particular, we obtain for the first order derivatives

$$\left. \frac{\partial^{e_l} A^{(\Phi)}(\mathbf{x})}{\partial \mathbf{x}^{e_l}} \right|_{\mathbf{x}=\mathbf{x}_0} = \mathbb{E}_{\mathbf{x}_0} \{\Phi_l(\mathbf{y})\}. \quad (2.41)$$

Another useful fact is that the partial derivatives $\left. \frac{\partial^{\mathbf{p}} f(\mathbf{y}; \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \right|_{\mathbf{x}=\mathbf{x}_0}$ exist for any order $\mathbf{p} \in \mathbb{Z}_+^N$ at every \mathbf{x}_0 in the interior of $N^{(\Phi)}$. Moreover, it holds that [34, 61]

$$\mathbb{E}_{\mathbf{x}_0} \left\{ \left(\frac{1}{f(\mathbf{y}; \mathbf{x})} \frac{\partial^{\mathbf{p}} f(\mathbf{y}; \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \right)^2 \right\} < \infty \quad (2.42)$$

for any multi-index $\mathbf{p} \in \mathbb{Z}_+^N$.

Maybe the most important instance of a pdf belonging to an exponential family is the multivariate normal distribution $f(\mathbf{y})$ of a Gaussian random vector $\mathbf{y} \in \mathbb{R}^M$ with mean $\boldsymbol{\mu} \in \mathbb{R}^M$ and the positive definite covariance matrix $\mathbf{C} \in \mathbb{R}^{M \times M}$, i.e., $f(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^M \det\{\mathbf{C}\}}} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \mathbf{C}^{-1}(\mathbf{y} - \boldsymbol{\mu})\right)$. This distribution is obtained for the choice $\mathbf{x} = \boldsymbol{\mu}$ from the exponential family in standard form, i.e., $\mathbf{u}(\mathbf{x}) = \mathbf{x}$, with sufficient statistic $\boldsymbol{\Phi}(\mathbf{y}) = \mathbf{C}^{-1}\mathbf{y}$, weight function $h(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^M \det\{\mathbf{C}\}}} \exp\left(-\frac{1}{2}\mathbf{y}^T \mathbf{C}^{-1}\mathbf{y}\right)$ and cumulant function $A^{(\boldsymbol{\Phi})}(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{C}^{-1}\mathbf{x}$. If a random vector $\mathbf{y} \in \mathbb{R}^M$ is distributed normally with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C} \in \mathbb{R}^{M \times M}$ we denote this by $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$.

Finally, we note that the concept of exponential families provides a natural link between classical estimation theory and the theory of graphical models. Indeed, the structure of the sufficient statistics translates directly to a specific connectivity structure of the associated graphical model [31, 34].

Chapter 3

Mathematical Tools

This chapter is a self-contained presentation of the mathematical tools that will be required later on. All the mathematical concepts discussed in this chapter are based on sets of real-valued functions $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ which share a common domain \mathcal{D} . We define the multiplication of a function with a real number and the sum of two functions over the same domain in the usual pointwise sense. The following presentation is based mainly on [5, 6] for general Hilbert space theory and [3, 62] for the theory of RKHS.

3.1 Hilbert Spaces

3.1.1 Some Basic Definitions

Definition 3.1.1. *An inner-product function space \mathcal{I} over a domain \mathcal{D} is a linear space of real-valued functions $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$. By a linear space we mean that*

$$f(\cdot), g(\cdot) \in \mathcal{I} \text{ and } a, b \in \mathbb{R} \quad \Rightarrow \quad af(\cdot) + bg(\cdot) \in \mathcal{I}, \quad (3.1)$$

i.e., any linear combination of two elements of \mathcal{I} is also an element of \mathcal{I} . Moreover, an inner product is defined on \mathcal{I} , i.e., a mapping that associates to two functions $f(\cdot), g(\cdot) \in \mathcal{I}$ the real number $\langle f(\cdot), g(\cdot) \rangle_{\mathcal{I}} \in \mathbb{R}$. This inner product has the properties

- (a) $\langle f(\cdot), g(\cdot) \rangle_{\mathcal{I}} = \langle g(\cdot), f(\cdot) \rangle_{\mathcal{I}}$
 - (b) $\langle af(\cdot) + bh(\cdot), g(\cdot) \rangle_{\mathcal{I}} = a\langle f(\cdot), g(\cdot) \rangle_{\mathcal{I}} + b\langle h(\cdot), g(\cdot) \rangle_{\mathcal{I}} \quad \forall a, b \in \mathbb{R} \text{ and } \forall f(\cdot), h(\cdot), g(\cdot) \in \mathcal{I}$
 - (c) $\langle f(\cdot), f(\cdot) \rangle_{\mathcal{I}} \geq 0 \quad \forall f(\cdot) \in \mathcal{I}$
 - (d) $\langle f(\cdot), f(\cdot) \rangle_{\mathcal{I}} = 0 \quad \text{only if } f(\cdot) \equiv 0.$
- (3.2)

In what follows we will need the famous *Cauchy-Schwarz inequality* [6] as stated in

Theorem 3.1.1. Consider an function space \mathcal{I}' together with a linear mapping that assigns to any two functions $f(\cdot), g(\cdot) \in \mathcal{I}'$ the real number $\langle f(\cdot), g(\cdot) \rangle_{\mathcal{I}'}$. If the mapping is such that $\langle f(\cdot), g(\cdot) \rangle_{\mathcal{I}'}$ has the properties (a)-(c) of (3.2), then we have that

$$|\langle f(\cdot), g(\cdot) \rangle_{\mathcal{I}'}| \leq \sqrt{\langle f(\cdot), f(\cdot) \rangle_{\mathcal{I}'} \langle g(\cdot), g(\cdot) \rangle_{\mathcal{I}'}} \quad (3.3)$$

where the equality sign holds if and only if there is a number $c \in \mathbb{R}$ such that $\langle f(\cdot) - cg(\cdot), f(\cdot) - cg(\cdot) \rangle_{\mathcal{I}'} = 0$.

Proof. [6, p. 77] □

For every inner product space \mathcal{I} , one can naturally define an *induced norm* $\|f(\cdot)\|_{\mathcal{I}}$ via

$$\|f(\cdot)\|_{\mathcal{I}} \triangleq \sqrt{\langle f(\cdot), f(\cdot) \rangle_{\mathcal{I}}} \quad (3.4)$$

As can be verified easily, the induced norm satisfies the axioms of a norm, i.e.,

- Positive definiteness:

$$\forall f(\cdot) \in \mathcal{H} : \quad \|f(\cdot)\|_{\mathcal{I}} \geq 0, \quad \|f(\cdot)\|_{\mathcal{I}} = 0 \Leftrightarrow f(\cdot) \equiv 0. \quad (3.5)$$

- Homogeneity:

$$a \in \mathbb{R}, f(\cdot) \in \mathcal{I} \Rightarrow \|af(\cdot)\|_{\mathcal{I}} = |a|\|f(\cdot)\|_{\mathcal{I}}. \quad (3.6)$$

- Triangle inequality:

$$\|f(\cdot) + g(\cdot)\|_{\mathcal{I}} \leq \|f(\cdot)\|_{\mathcal{I}} + \|g(\cdot)\|_{\mathcal{I}}. \quad (3.7)$$

It may happen that for a given linear function space \mathcal{I}' one can define naturally a real-valued function $\langle f(\cdot), g(\cdot) \rangle_{\mathcal{I}'}$ of pairs of elements of the function space, that satisfies all requirements for an inner product in (3.2) except condition (d).¹ It is then customary (cf. [6]) to consider two functions $f(\cdot), g(\cdot) \in \mathcal{I}'$ as identical if $\langle f(\cdot) - g(\cdot), f(\cdot) - g(\cdot) \rangle_{\mathcal{I}'} = 0$. Technically speaking, one constructs an inner-product space whose elements are the equivalence classes

$$[f(\cdot)] \triangleq \left\{ f'(\cdot) \in \mathcal{I}' \mid \langle f(\cdot) - f'(\cdot), f(\cdot) - f'(\cdot) \rangle_{\mathcal{I}'} = 0 \right\} \quad (3.8)$$

where $f(\cdot) \in \mathcal{I}'$. Based on Theorem 3.1.1, one can then verify that this inner-product space is well defined, i.e., if $f'(\cdot) \in [f(\cdot)]$ and $g'(\cdot) \in [g(\cdot)]$ then $\langle f(\cdot), g(\cdot) \rangle_{\mathcal{I}'} = \langle f'(\cdot), g'(\cdot) \rangle_{\mathcal{I}'}$. However, we will follow the usual convention of considering the space \mathcal{I}' itself as an inner-product space [6] and denote by a function $f(\cdot)$ not only the function itself but the whole equivalence class $[f(\cdot)]$.

¹Maybe the best known example of this situation is given by the set of square-integrable real-valued functions.

Therefore, if we write $f(\cdot) \equiv 0$, we mean not necessarily that $f(\mathbf{x}) = 0$ for every $\mathbf{x} \in \mathcal{D}$ but that the function $f(\cdot)$ is in the same class as the function which is identically zero. However, within this thesis we will work primarily with a special kind of Hilbert spaces, i.e., a RKHS, and in that case we have that every class $[f(\cdot)]$ consists of one and only one function, i.e., $[f(\cdot)] = \{f(\cdot)\}$.

The induced norm $\|\cdot\|_{\mathcal{I}}$ also defines the notion of convergence in an inner-product function space \mathcal{I} .

Definition 3.1.2. A sequence $\{f_l(\cdot)\}_{l \rightarrow \infty}$ of functions $f_l(\cdot)$ belonging to an inner-product function space \mathcal{I} , i.e., $f_l(\cdot) \in \mathcal{I}$, is called convergent to a function $f(\cdot) \in \mathcal{I}$ if to any $\varepsilon > 0$ there exists a $n_0 \in \mathbb{N}$ such that

$$n \geq n_0 \Rightarrow \|f_n(\cdot) - f(\cdot)\|_{\mathcal{I}} \leq \varepsilon. \quad (3.9)$$

The function $f(\cdot)$ is then called the limit of the sequence $\{f_l(\cdot)\}_{l \rightarrow \infty}$ and denoted by $\lim_{l \rightarrow \infty} f_l(\cdot)$.

A necessary condition for a sequence to be convergent is that it is a Cauchy sequence:

Definition 3.1.3. A sequence $\{f_l(\cdot) \in \mathcal{I}\}_{l \rightarrow \infty}$ of functions $f_l(\cdot)$ belonging to an inner-product function space \mathcal{I} is called a Cauchy sequence if to any $\varepsilon > 0$ there exists a $n_0 \in \mathbb{N}$ such that

$$n_1, n_2 \geq n_0 \Rightarrow \|f_{n_1}(\cdot) - f_{n_2}(\cdot)\|_{\mathcal{I}} \leq \varepsilon. \quad (3.10)$$

Now we are in the position to define the central mathematical structure for our thesis, i.e., a Hilbert space:

Definition 3.1.4. A function Hilbert space \mathcal{H} is an inner-product function space that has the additional property of being complete, i.e., every Cauchy sequence of functions belonging to \mathcal{H} converges to a function belonging to \mathcal{H} .

Every inner-product space \mathcal{I} can be completed in a canonical way such that it becomes a Hilbert space [62]. In what follows, we will state the most important concepts and facts about Hilbert spaces, as required for our purposes.

Definition 3.1.5. Given a function Hilbert space \mathcal{H} , and an index set \mathcal{T} , we call a subset $\mathcal{S} = \{g_l(\cdot)\}_{l \in \mathcal{T}} \subseteq \mathcal{H}$ dense, if any element of \mathcal{H} can be approximated arbitrary well by an element of the set \mathcal{S} , i.e., for any $f(\cdot) \in \mathcal{H}$ and any $\varepsilon > 0$, there exists a function $g(\cdot) \in \mathcal{S}$ such that

$$\|g(\cdot) - f(\cdot)\|_{\mathcal{H}} \leq \varepsilon. \quad (3.11)$$

Definition 3.1.6. A function Hilbert space \mathcal{H} is said to be separable if it contains a countable subset which is dense in \mathcal{H} .

Within this thesis, we exclusively work with Hilbert spaces that are separable.

Definition 3.1.7. Given a function Hilbert space \mathcal{H} , a set of functions $\mathcal{S} = \{g_l(\cdot)\}_{l \in \mathcal{T}}$ with an arbitrary index set \mathcal{T} is called complete if the following implication holds:

$$\langle g_l(\cdot), f(\cdot) \rangle_{\mathcal{H}} = 0 \quad \forall l \in \mathcal{T} \quad \Rightarrow \quad f(\cdot) \equiv 0. \quad (3.12)$$

Definition 3.1.8. Consider a function Hilbert space \mathcal{H} and a subset $\mathcal{U} \subseteq \mathcal{H}$. If the subset is such that any linear combination of two elements of \mathcal{U} is still in \mathcal{U} , i.e.,

$$f(\cdot), g(\cdot) \in \mathcal{U} \quad \text{and} \quad a, b \in \mathbb{R} \quad \Rightarrow \quad af(\cdot) + bg(\cdot) \in \mathcal{U}, \quad (3.13)$$

and if moreover every Cauchy sequence $\{f_l(\cdot) \in \mathcal{U}\}_{l \rightarrow \infty}$ converges to a function belonging to \mathcal{U} , then we call \mathcal{U} a subspace of the Hilbert space \mathcal{H} .

Note that according to this definition, we have that any subspace of a Hilbert space is itself a Hilbert space. Indeed, a subspace is a linear space since by (3.13) any linear combination of two elements are again in the subspace. A subspace is also an inner-product space since an inner product is given naturally by the inner product defined on the larger Hilbert space, in which the subspace is contained. Finally, by definition, a subspace is an inner-product space in which every Cauchy converges to a function of the same subspace, i.e., it is a complete inner-product space, i.e., a Hilbert space.

Definition 3.1.9. Given a function Hilbert space \mathcal{H} that contains an inner-product space \mathcal{I} (which is defined for the same inner product as that of \mathcal{H}), i.e., $\mathcal{I} \subseteq \mathcal{H}$, we define the closure of \mathcal{I} relative to \mathcal{H} , denoted by $\text{cl}\{\mathcal{I}\}$, as the subspace of \mathcal{H} that consists of all functions belonging to \mathcal{I} as well as any limit of a Cauchy sequence of functions belonging to \mathcal{I} .

Definition 3.1.10. Given an inner-product function space \mathcal{I} (in particular, a Hilbert space), an arbitrary index set \mathcal{T} , and a set $\{v_l(\cdot)\}_{l \in \mathcal{T}}$ of functions $v_l(\cdot) \in \mathcal{I}$, we define the linear span of $\{v_l(\cdot)\}_{l \in \mathcal{T}}$, denoted by $\text{span}\{v_l(\cdot)\}_{l \in \mathcal{T}}$, as the set of all finite linear combinations of the functions $v_l(\cdot)$, i.e.,

$$\text{span}\{v_l(\cdot)\}_{l \in \mathcal{T}} \triangleq \left\{ v(\cdot) \in \mathcal{I} \mid v(\cdot) = \sum_{j \in [J]} a_j v_{l_j}(\cdot) \quad \text{with } J \in \mathbb{N}, a_j \in \mathbb{R}, l_j \in \mathcal{T} \right\}. \quad (3.14)$$

Definition 3.1.11. Given a function Hilbert space \mathcal{H} and an arbitrary index set \mathcal{T} , we say that the set of functions $\{f_l(\cdot) \in \mathcal{H}\}_{l \in \mathcal{T}}$ spans \mathcal{H} if the linear span $\text{span}\{f_l(\cdot)\}_{l \in \mathcal{T}}$ is dense in the Hilbert space \mathcal{H} .

Definition 3.1.12. Given a function Hilbert space \mathcal{H} , we call any set of functions $\{g_l(\cdot) \in \mathcal{H}\}_{l \in \mathcal{T}}$ an orthonormal basis (ONB) for the Hilbert space \mathcal{H} if the functions $\{g_l(\cdot)\}_{l \in \mathcal{T}}$ span the Hilbert space \mathcal{H} and the functions $g_l(\cdot)$ are orthonormal, i.e.,

$$\langle g_k(\cdot), g_l(\cdot) \rangle_{\mathcal{H}} = \delta_{k,l}. \quad (3.15)$$

Definition 3.1.13. Given a function Hilbert space \mathcal{H} and a subspace $\mathcal{U} \subseteq \mathcal{H}$, we define the orthogonal projection of a function $f(\cdot) \in \mathcal{H}$ onto the subspace \mathcal{U} , denoted by $\mathbf{P}_{\mathcal{U}}f(\cdot)$, as the element of \mathcal{U} given by

$$\mathbf{P}_{\mathcal{U}}f(\cdot) \triangleq \operatorname{argmin}_{g(\cdot) \in \mathcal{U}} \|g(\cdot) - f(\cdot)\|_{\mathcal{H}}. \quad (3.16)$$

From its definition, it is clear that the orthogonal projection $\mathbf{P}_{\mathcal{U}}f(\cdot)$ is the best approximation of $f(\cdot)$ by an element of \mathcal{U} .

Definition 3.1.14. We call two Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 with associated inner products $\langle \cdot, \cdot \rangle_{\mathcal{H}_1}$ and $\langle \cdot, \cdot \rangle_{\mathcal{H}_2}$ isometric if there exists a bijective linear map $J[\cdot] : \mathcal{H}_1 \rightarrow \mathcal{H}_2$, i.e.,

$$f(\cdot), g(\cdot) \in \mathcal{H}_1 \text{ and } a, b \in \mathbb{R} \Rightarrow J[af(\cdot) + bg(\cdot)] = aJ[f(\cdot)] + bJ[g(\cdot)], \quad (3.17)$$

which preserves inner products, i.e.,

$$f(\cdot), g(\cdot) \in \mathcal{H}_1 \Rightarrow \langle f(\cdot), g(\cdot) \rangle_{\mathcal{H}_1} = \langle J[f(\cdot)], J[g(\cdot)] \rangle_{\mathcal{H}_2}. \quad (3.18)$$

Any bijective linear map $J[\cdot]$ for which (3.18) holds is called a congruence (or congruence map) from \mathcal{H}_1 to \mathcal{H}_2 .

3.1.2 Some Basic Facts

Theorem 3.1.2. Consider a function Hilbert space \mathcal{H} and a subset $\mathcal{L} \subseteq \mathcal{H}$ which is dense in \mathcal{H} . Then we can express the norm $\|f(\cdot)\|_{\mathcal{H}}$ of any function $f(\cdot) \in \mathcal{H}$ as

$$\|f(\cdot)\|_{\mathcal{H}} = \sup_{\substack{g(\cdot) \in \mathcal{L} \\ \|g(\cdot)\|_{\mathcal{H}} > 0}} \frac{\langle f(\cdot), g(\cdot) \rangle_{\mathcal{H}}}{\|g(\cdot)\|_{\mathcal{H}}}. \quad (3.19)$$

Proof. By Theorem 3.1.1 we have that $|\langle f(\cdot), g(\cdot) \rangle_{\mathcal{H}}| \leq \|f(\cdot)\|_{\mathcal{H}} \|g(\cdot)\|_{\mathcal{H}}$, which implies that

$$\sup_{\substack{g(\cdot) \in \mathcal{L} \\ \|g(\cdot)\|_{\mathcal{H}} > 0}} \frac{\langle f(\cdot), g(\cdot) \rangle_{\mathcal{H}}}{\|g(\cdot)\|_{\mathcal{H}}} \leq \|f(\cdot)\|_{\mathcal{H}}. \quad (3.20)$$

From this, the validity of (3.19) is obvious for $\|f(\cdot)\|_{\mathcal{H}} = 0$.

It remains to prove (3.19) for the case $\|f(\cdot)\|_{\mathcal{H}} > 0$. However, in this case we can choose a $g(\cdot) \in \mathcal{L}$ with $\|g(\cdot)\|_{\mathcal{H}} > 0$ such that $\frac{\langle f(\cdot), g(\cdot) \rangle_{\mathcal{H}}}{\|g(\cdot)\|_{\mathcal{H}}}$ is arbitrarily close to $\|f(\cdot)\|_{\mathcal{H}}$. Indeed, since \mathcal{L} is dense in \mathcal{H} we can find for any sufficiently small $\varepsilon > 0$ a function $g(\cdot) \in \mathcal{L}$ such that $\|f(\cdot) - g(\cdot)\|_{\mathcal{H}} \leq \varepsilon$ and $\|g(\cdot)\|_{\mathcal{H}} > 0$. This specific function $g(\cdot)$ yields

$$\frac{\langle f(\cdot), g(\cdot) \rangle_{\mathcal{H}}}{\|f(\cdot)\|_{\mathcal{H}}} = \frac{\langle f(\cdot), g(\cdot) - f(\cdot) + f(\cdot) \rangle_{\mathcal{H}}}{\|f(\cdot)\|_{\mathcal{H}}} = \frac{\langle f(\cdot), g(\cdot) - f(\cdot) \rangle_{\mathcal{H}}}{\|f(\cdot)\|_{\mathcal{H}}} + \frac{\langle f(\cdot), f(\cdot) \rangle_{\mathcal{H}}}{\|f(\cdot)\|_{\mathcal{H}}}$$

$$\stackrel{(a)}{\leq} \frac{\|f(\cdot)\|_{\mathcal{H}} \|g(\cdot) - f(\cdot)\|_{\mathcal{H}}}{\|f(\cdot)\|_{\mathcal{H}}} + \frac{\langle f(\cdot), f(\cdot) \rangle_{\mathcal{H}}}{\|f(\cdot)\|_{\mathcal{H}}} \leq \varepsilon + \|f(\cdot)\|_{\mathcal{H}}, \quad (3.21)$$

where step (a) is due to Theorem 3.1.1. Similarly,

$$\frac{\langle f(\cdot), g(\cdot) \rangle_{\mathcal{H}}}{\|f(\cdot)\|_{\mathcal{H}}} = \frac{\langle f(\cdot), g(\cdot) - f(\cdot) + f(\cdot) \rangle_{\mathcal{H}}}{\|f(\cdot)\|_{\mathcal{H}}} = \frac{\langle f(\cdot), g(\cdot) - f(\cdot) \rangle_{\mathcal{H}}}{\|f(\cdot)\|_{\mathcal{H}}} + \frac{\langle f(\cdot), f(\cdot) \rangle_{\mathcal{H}}}{\|f(\cdot)\|_{\mathcal{H}}} \geq -\varepsilon + \|f(\cdot)\|_{\mathcal{H}}. \quad (3.22)$$

From (3.21) and (3.22) we conclude that

$$\sup_{\substack{g(\cdot) \in \mathcal{L} \\ \|g(\cdot)\|_{\mathcal{H}} > 0}} \frac{\langle f(\cdot), g(\cdot) \rangle_{\mathcal{H}}}{\|g(\cdot)\|_{\mathcal{H}}} \geq \|f(\cdot)\|_{\mathcal{H}}. \quad (3.23)$$

Combining (3.20) and (3.23) yields (3.19). \square

Theorem 3.1.3. *Consider a function Hilbert space \mathcal{H} , a finite index set \mathcal{T} (which can be assumed without loss of generality to be of the form $\mathcal{T} = [L]$ with some $L \in \mathbb{N}$), and the set of functions $\{f_l(\cdot) \in \mathcal{H}\}_{l \in \mathcal{T}}$. We then have that the linear span $\text{span}\{f_l(\cdot)\}_{l \in \mathcal{T}}$ is a subspace of \mathcal{H} .*

Proof. The linear span $\text{span}\{f_l(\cdot)\}_{l \in \mathcal{T}}$, for a finite index set \mathcal{T} , together with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ is a finite dimensional inner-product function space, i.e., a special case of an abstract finite dimensional inner-product space. From this, the statement follows from the well-known fact, that any finite dimensional inner-product space is complete [6, 63]. \square

Theorem 3.1.4. *Consider a function Hilbert space \mathcal{H} and an ONB $\{g_l(\cdot)\}_{l \in \mathcal{T}}$ for it, where \mathcal{T} is an arbitrary index set. We then have for any $f(\cdot) \in \mathcal{H}$*

$$\|f(\cdot)\|_{\mathcal{H}}^2 = \sum_{l \in \mathcal{T}} |\langle g_l(\cdot), f(\cdot) \rangle_{\mathcal{H}}|^2, \quad (3.24)$$

and

$$f(\cdot) = \sum_{l \in \mathcal{T}} \langle f(\cdot), g_l(\cdot) \rangle_{\mathcal{H}} g_l(\cdot). \quad (3.25)$$

Proof. [6] \square

The meaning of (3.25) in case of an infinite index set \mathcal{T} is that the function $f(\cdot)$ can be approximated arbitrarily well by the sums $\sum_{l \in \mathcal{T}'} \langle f(\cdot), g_l(\cdot) \rangle_{\mathcal{H}} g_l(\cdot)$, where $\mathcal{T}' \subseteq \mathcal{T}$ is an arbitrary finite subset of \mathcal{T} , i.e.,

$$\inf_{\mathcal{T}' \subseteq \mathcal{T}} \left\| f(\cdot) - \sum_{l \in \mathcal{T}'} \langle f(\cdot), g_l(\cdot) \rangle_{\mathcal{H}} g_l(\cdot) \right\|_{\mathcal{H}} = 0. \quad (3.26)$$

Theorem 3.1.5. Consider two Hilbert spaces \mathcal{H}_1 and \mathcal{H}_2 as well as two sets of functions $\mathcal{A} \triangleq \{f_j(\cdot) \in \mathcal{H}_1\}_{j \in \mathcal{T}}$ and $\mathcal{B} \triangleq \{g_j(\cdot) \in \mathcal{H}_2\}_{j \in \mathcal{T}}$, indexed by the same set \mathcal{T} , that span \mathcal{H}_1 and \mathcal{H}_2 respectively, i.e., $\text{span}\{\mathcal{A}\}$ is dense in \mathcal{H}_1 and $\text{span}\{\mathcal{B}\}$ is dense in \mathcal{H}_2 . Then, if

$$t, s \in \mathcal{T} \Rightarrow \langle f_t(\cdot), f_s(\cdot) \rangle_{\mathcal{H}_1} = \langle g_t(\cdot), g_s(\cdot) \rangle_{\mathcal{H}_2}, \quad (3.27)$$

the two Hilbert spaces $\mathcal{H}_1, \mathcal{H}_2$ are isometric. A sufficient condition for a linear mapping $J[\cdot] : \mathcal{H}_1 \rightarrow \mathcal{H}_2$, which satisfies

$$t \in \mathcal{T} \Rightarrow J[f_t(\cdot)] = g_t(\cdot), \quad (3.28)$$

to be a congruence from \mathcal{H}_1 to \mathcal{H}_2 is that

$$J[f(\cdot)] = \lim_{l \rightarrow \infty} J[f_l(\cdot)], \quad (3.29)$$

for any Cauchy sequence $\{f_l(\cdot) \in \text{span}\{\mathcal{A}\}\}_{l \rightarrow \infty}$ with limit $f(\cdot) \in \mathcal{H}_1$.

Proof. [3] □

Theorem 3.1.6. Consider a function Hilbert space \mathcal{H} , an arbitrary function $f(\cdot) \in \mathcal{H}$, and a subspace $\mathcal{U} \subseteq \mathcal{H}$. We then have for any $g(\cdot) \in \mathcal{U}$

$$\langle f(\cdot) - \mathbf{P}_{\mathcal{U}}f(\cdot), g(\cdot) \rangle_{\mathcal{H}} = 0, \quad (3.30)$$

which implies that

$$\|\mathbf{P}_{\mathcal{U}}f(\cdot)\|_{\mathcal{H}}^2 = \|f(\cdot)\|_{\mathcal{H}}^2 - \|f(\cdot) - \mathbf{P}_{\mathcal{U}}f(\cdot)\|_{\mathcal{H}}^2 \leq \|f(\cdot)\|_{\mathcal{H}}^2. \quad (3.31)$$

Proof. [6] □

Theorem 3.1.7. Consider a function Hilbert space \mathcal{H} and a subspace $\mathcal{U} \subseteq \mathcal{H}$ for which an ONB $\{v_l(\cdot)\}_{l \in \mathcal{T}}$ with a finite index set \mathcal{T} exists. Then the orthogonal projection and squared norm of a function $f(\cdot) \in \mathcal{H}$ onto the subspace \mathcal{U} is given by

$$\mathbf{P}_{\mathcal{U}}f(\cdot) = \sum_{l \in \mathcal{T}} \langle f(\cdot), v_l(\cdot) \rangle_{\mathcal{H}} v_l(\cdot), \quad (3.32)$$

and

$$\|\mathbf{P}_{\mathcal{U}}f(\cdot)\|_{\mathcal{H}}^2 = \sum_{l \in \mathcal{T}} |\langle f(\cdot), v_l(\cdot) \rangle_{\mathcal{H}}|^2, \quad (3.33)$$

respectively.

Proof. [6] □

Theorem 3.1.8. Consider a function Hilbert space \mathcal{H} and a finite number of functions $v_l(\cdot) \in \mathcal{H}$, $l \in [L]$. Then the squared norm of the orthogonal projection of a function $f(\cdot) \in \mathcal{H}$ on the subspace $\mathcal{U} \triangleq \text{span}\{v_l(\cdot)\}_{l \in [L]}$ (cf. Theorem 3.1.3) is given by

$$\|\mathbf{P}_{\mathcal{U}}f(\cdot)\|_{\mathcal{H}}^2 = \mathbf{c}^T \mathbf{G}^\dagger \mathbf{c} \quad (3.34)$$

where the vector $\mathbf{c} \in \mathbb{R}^L$ and matrix $\mathbf{G} \in \mathbb{R}^{L \times L}$ are defined elementwise by $c_l = \langle f(\cdot), v_l(\cdot) \rangle_{\mathcal{H}}$ and $(\mathbf{G})_{m,n} = \langle v_m(\cdot), v_n(\cdot) \rangle_{\mathcal{H}}$ respectively.

Proof. According to Theorem 3.1.6, we can decompose any $f(\cdot) \in \mathcal{H}$ as a sum $f(\cdot) = \mathbf{P}_{\mathcal{U}}f(\cdot) + h(\cdot)$, where $\langle h(\cdot), g(\cdot) \rangle_{\mathcal{H}} = 0$ for every $g(\cdot) \in \mathcal{U}$. This implies that for any $l \in [L]$ we have

$$\langle f(\cdot), v_l(\cdot) \rangle_{\mathcal{H}} = \langle \mathbf{P}_{\mathcal{U}}f(\cdot), v_l(\cdot) \rangle_{\mathcal{H}}. \quad (3.35)$$

Since, per definition, the orthogonal projection $\mathbf{P}_{\mathcal{U}}f(\cdot)$ belongs to the subspace $\mathcal{U} \triangleq \text{span}\{v_l(\cdot)\}_{l \in [L]}$, we can represent it as

$$\mathbf{P}_{\mathcal{U}}f(\cdot) = \sum_{l \in [L]} d_l v_l(\cdot), \quad (3.36)$$

with suitable coefficients $d_l \in \mathbb{R}$. The squared norm of the projection can then, due to the properties (3.2), be expressed as

$$\|\mathbf{P}_{\mathcal{U}}f(\cdot)\|_{\mathcal{H}}^2 = \mathbf{d}^T \mathbf{G} \mathbf{d}, \quad (3.37)$$

where \mathbf{G} as defined above and the vector $\mathbf{d} \in \mathbb{R}^L$ is obtained by stacking the coefficients d_l . From (3.36) and the linearity of the inner product one can verify using (3.35), that the vector \mathbf{d} and the above defined vector $\mathbf{c} \in \mathbb{R}^L$ are related by

$$\mathbf{c} = \mathbf{G} \mathbf{d}. \quad (3.38)$$

Using the identity $\mathbf{G} \mathbf{G}^\dagger \mathbf{G} = \mathbf{G}$ [38], we finally obtain

$$\mathbf{c}^T \mathbf{G}^\dagger \mathbf{c} \stackrel{(3.38)}{=} \mathbf{d}^T \mathbf{G} \mathbf{G}^\dagger \mathbf{G} \mathbf{d} = \mathbf{d}^T \mathbf{G} \mathbf{d} \stackrel{(3.37)}{=} \|\mathbf{P}_{\mathcal{U}}f(\cdot)\|_{\mathcal{H}}^2.$$

□

We will need in the following a slight variation of Theorem 3.1.8 stated in

Theorem 3.1.9. Consider a function Hilbert space \mathcal{H} and two sets of functions $\{v_l(\cdot) \in \mathcal{H}\}_{l \in [L_1]}$ and $\{w_{l'}(\cdot) \in \mathcal{H}\}_{l' \in [L_2]}$ that are mutually orthogonal, i.e., $\langle v_l(\cdot), w_{l'}(\cdot) \rangle_{\mathcal{H}} = 0$ for any pair of indices $l \in [L_1]$, $l' \in [L_2]$. Then the squared norm of the orthogonal projection of a function $f(\cdot) \in \mathcal{H}$ on the subspace $\mathcal{U} \triangleq \text{span}\{\{v_l(\cdot)\}_{l \in [L_1]} \cup \{w_{l'}(\cdot)\}_{l' \in [L_2]}\}$ is given by

$$\|\mathbf{P}_{\mathcal{U}}f(\cdot)\|_{\mathcal{H}}^2 = \mathbf{c}_1^T \mathbf{G}_1^\dagger \mathbf{c}_1 + \mathbf{c}_2^T \mathbf{G}_2^\dagger \mathbf{c}_2, \quad (3.39)$$

where the vectors $\mathbf{c}_1 \in \mathbb{R}^{L_1}$, $\mathbf{c}_2 \in \mathbb{R}^{L_2}$ and matrices $\mathbf{G}_1 \in \mathbb{R}^{L_1 \times L_1}$, $\mathbf{G}_2 \in \mathbb{R}^{L_2 \times L_2}$ are defined elementwise by $c_{1,l} \triangleq \langle f(\cdot), v_l(\cdot) \rangle_{\mathcal{H}}$, $c_{2,l} \triangleq \langle f(\cdot), w_l(\cdot) \rangle_{\mathcal{H}}$ and $(\mathbf{G}_1)_{m,n} = \langle v_m(\cdot), v_n(\cdot) \rangle_{\mathcal{H}}$, $(\mathbf{G}_2)_{m,n} = \langle w_m(\cdot), w_n(\cdot) \rangle_{\mathcal{H}}$, respectively.

Proof. We merge the two sets $\{v_l(\cdot) \in \mathcal{H}\}_{l \in [L_1]}$ and $\{w_l(\cdot) \in \mathcal{H}\}_{l \in [L_2]}$ into one large set $\mathcal{A} \triangleq \{u_l(\cdot)\}_{l \in [L_1+L_2]}$ such that $u_l(\cdot) = v_l(\cdot)$ for $l \in [L_1]$ and $u_l(\cdot) = w_{l-L_1}(\cdot)$ for $l \in [L_1+L_2] \setminus [L_1]$ and then define the matrix $\mathbf{G} \in \mathbb{R}^{(L_1+L_2) \times (L_1+L_2)}$ elementwise by $(\mathbf{G})_{k,l} \triangleq \langle u_l(\cdot), u_k(\cdot) \rangle_{\mathcal{H}}$. Obviously, we have $\text{span}\{\mathcal{A}\} = \mathcal{U}$. Note that \mathbf{G} is a block diagonal matrix, i.e., $\mathbf{G} = \begin{pmatrix} \mathbf{G}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2 \end{pmatrix}$, which implies that also its pseudo-inverse is block diagonal, i.e., $\mathbf{G}^\dagger = \begin{pmatrix} \mathbf{G}_1^\dagger & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2^\dagger \end{pmatrix}$ (cf. [64, Theorem 4.2.15]). Similarly, we define the vector $\mathbf{c} \in \mathbb{R}^{L_1+L_2}$ via $c_l = \langle f(\cdot), u_l(\cdot) \rangle_{\mathcal{H}}$ and observe that $\mathbf{c} = (\mathbf{c}_1^T \quad \mathbf{c}_2^T)^T$. The application of Theorem 3.1.8, remember that $\mathcal{U} = \text{span}\{\mathcal{A}\}$, yields

$$\|\mathbf{P}_{\mathcal{U}}f(\cdot)\|_{\mathcal{H}}^2 = \mathbf{c}^T \mathbf{G}^\dagger \mathbf{c} = (\mathbf{c}_1^T \quad \mathbf{c}_2^T) \begin{pmatrix} \mathbf{G}_1^\dagger & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_2^\dagger \end{pmatrix} \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \end{pmatrix} = \mathbf{c}_1^T \mathbf{G}_1^\dagger \mathbf{c}_1 + \mathbf{c}_2^T \mathbf{G}_2^\dagger \mathbf{c}_2.$$

□

Theorem 3.1.10. Consider a function Hilbert space with an ONB $\{g_l(\cdot) \in \mathcal{H}\}_{l \in \mathcal{T}}$ with an arbitrary index set \mathcal{T} . We then have that the Hilbert space \mathcal{H} is isometric to the Hilbert space $\ell^2(\mathcal{T})$ with inner product $\langle f[\cdot], g[\cdot] \rangle_{\ell^2(\mathcal{T})} \triangleq \sum_{l \in \mathcal{T}} f[l]g[l]$. Furthermore, a congruence is given by $\mathbf{J}[\cdot] : \mathcal{H} \rightarrow \ell^2(\mathcal{T}) : f(\cdot) \in \mathcal{H} \mapsto h[\cdot] \in \ell^2(\mathcal{T})$ where $h[l] = \langle f(\cdot), g_l(\cdot) \rangle_{\mathcal{H}}$. Any element $f(\cdot) \in \mathcal{H}$ of the Hilbert space \mathcal{H} can be written as

$$f(\cdot) = \sum_{l \in \mathcal{T}} a[l]g_l(\cdot) \tag{3.40}$$

with a coefficient sequence $a[l] \in \ell^2(\mathcal{T})$, and we have

$$\langle f(\cdot), g(\cdot) \rangle_{\mathcal{H}} = \langle \mathbf{J}[f(\cdot)], \mathbf{J}[g(\cdot)] \rangle_{\ell^2(\mathcal{T})}. \tag{3.41}$$

Conversely, for any coefficient sequence $a[l] \in \ell^2(\mathcal{T})$ the function given by (3.40) is an element of \mathcal{H} .

Proof. [6, p. 85] □

The meaning of (3.40) in case of an infinite index set \mathcal{T} is that the function $f(\cdot)$ can be approximated arbitrarily well by the sums $\sum_{l \in \mathcal{T}'} a[l]g_l(\cdot)$, where $\mathcal{T}' \subseteq \mathcal{T}$ is an arbitrary finite subset of \mathcal{T} , i.e.,

$$\inf_{\substack{\mathcal{T}' \subseteq \mathcal{T} \\ a[l] \in \mathbb{R}}} \left\| f(\cdot) - \sum_{l \in \mathcal{T}'} a[l]g_l(\cdot) \right\|_{\mathcal{H}} = 0. \tag{3.42}$$

3.2 Reproducing Kernel Hilbert Spaces (RKHS)

A RKHS \mathcal{H} is a special kind of Hilbert space consisting of real-valued functions $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ defined over a specific domain \mathcal{D} (which will always be a subset of \mathbb{R}^N within this thesis). The special thing about a RKHS \mathcal{H} is that the point evaluations $f_{\mathbf{y}}[\cdot] : \mathcal{H} \rightarrow \mathbb{R} : g(\cdot) \mapsto g(\mathbf{y})$ are continuous functionals [62]. According to [3], we have

Definition 3.2.1. *Given a function Hilbert space \mathcal{H} over the domain \mathcal{D} (i.e., the Hilbert space \mathcal{H} consists of functions $f : \mathcal{D} \rightarrow \mathbb{R}$) and a function $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$, we call \mathcal{H} a RKHS with reproducing kernel $R(\cdot, \cdot)$ if the following two statements are valid:*

- The function $f(\cdot) \triangleq R(\cdot, \mathbf{x})$ obtained by fixing the second argument of the kernel to an arbitrary $\mathbf{x} \in \mathcal{D}$, i.e., $f(\mathbf{x}') = R(\mathbf{x}', \mathbf{x})$, is an element of the RKHS \mathcal{H} , i.e.,

$$\mathbf{x} \in \mathcal{D} \Rightarrow R(\cdot, \mathbf{x}) \in \mathcal{H}. \quad (3.43)$$

- The inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ on the RKHS \mathcal{H} satisfies the following “reproducing property”:

$$f(\cdot) \in \mathcal{H}, \mathbf{x} \in \mathcal{D} \Rightarrow \langle f(\cdot), R(\cdot, \mathbf{x}) \rangle_{\mathcal{H}} = f(\mathbf{x}). \quad (3.44)$$

We make the notion of a kernel function precise by

Definition 3.2.2. *Given an arbitrary set \mathcal{D} , we call a real-valued function $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ a “kernel function over the domain \mathcal{D} ” if it is symmetric, i.e.,*

$$\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D} \Rightarrow R(\mathbf{x}_1, \mathbf{x}_2) = R(\mathbf{x}_2, \mathbf{x}_1) \quad (3.45)$$

and for any finite set $\{\mathbf{x}_k\}_{k \in [K]}$ of points $\mathbf{x}_k \in \mathcal{D}$, the matrix $\mathbf{R} \in \mathbb{R}^{K \times K}$ defined elementwise by $(\mathbf{R})_{k,l} = R(\mathbf{x}_k, \mathbf{x}_l)$ is psd, i.e.,

$$\mathbf{R} \geq \mathbf{0}. \quad (3.46)$$

As shown in [62], a necessary and sufficient condition for the existence of a (necessarily unique) reproducing kernel for a given Hilbert space \mathcal{H} is that the point evaluations $f_{\mathbf{y}}[\cdot] : \mathcal{H} \rightarrow \mathbb{R} : g(\cdot) \mapsto g(\mathbf{y})$ are continuous functionals. However, for our purposes, it is important to observe that in the opposite direction, i.e., given a kernel function $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$, one can always associate a RKHS to it [2, 62, 65, 66].

Definition 3.2.3. *Given a kernel function $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ over a domain \mathcal{D} , we associate with it the linear function space denoted by $\mathcal{L}(R)$ and defined as*

$$\mathcal{L}(R) \triangleq \text{span}\{R(\cdot, \mathbf{x})\}_{\mathbf{x} \in \mathcal{D}}. \quad (3.47)$$

That is, any $f(\cdot) \in \mathcal{L}(R)$ can be written as a finite sum $f(\cdot) = \sum_{l \in [L]} a_l R(\cdot, \mathbf{x}_l)$ with certain coefficients $a_l \in \mathbb{R}$ and points $\mathbf{x}_l \in \mathcal{D}$.

Theorem 3.2.1. *Given an arbitrary kernel function $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$, we have that on the linear space $\mathcal{L}(R)$, one can naturally define an inner product denoted by $\langle f(\cdot), g(\cdot) \rangle_{\mathcal{L}(R)}$. This inner product satisfies the basic properties (3.2) and is completely specified by the requirement*

$$\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D} \Rightarrow \langle R(\cdot, \mathbf{x}_1), R(\cdot, \mathbf{x}_2) \rangle_{\mathcal{L}(R)} = R(\mathbf{x}_1, \mathbf{x}_2). \quad (3.48)$$

Given two arbitrary functions $f(\cdot), g(\cdot) \in \mathcal{L}(R)$ and their representations $f(\cdot) = \sum_{l \in [L_1]} a_l R(\cdot, \mathbf{x}_{1,l})$, $g(\cdot) = \sum_{l \in [L_2]} b_l R(\cdot, \mathbf{x}_{2,l})$ with suitable coefficients $a_l, b_l \in \mathbb{R}$ and points $\mathbf{x}_{1,l}, \mathbf{x}_{2,l} \in \mathcal{D}$, their inner product is given by

$$\langle f(\cdot), g(\cdot) \rangle_{\mathcal{L}(R)} = \sum_{l \in [L_1]} \sum_{l' \in [L_2]} a_l b_{l'} R(\mathbf{x}_{2,l'}, \mathbf{x}_{1,l}). \quad (3.49)$$

Moreover, if two functions $f(\cdot), g(\cdot)$ are such that $\langle f(\cdot) - g(\cdot), f(\cdot) - g(\cdot) \rangle_{\mathcal{L}(R)} = 0$, we have that $f(\mathbf{x}) = g(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{D}$.

Proof. The properties (a) and (b) in (3.2) are evident from (3.49). Note that (3.49) contains as a special case the reproducing property (3.44) since for $g(\cdot) = R(\cdot, \mathbf{x}_2)$, i.e., $L' = 1$ and $b_1 = 1$, we obtain

$$\langle f(\cdot), g(\cdot) \rangle_{\mathcal{L}(R)} = \sum_{l \in [L_1]} a_l R(\mathbf{x}_2, \mathbf{x}_{1,l}) = f(\mathbf{x}_2). \quad (3.50)$$

As already discussed, the property (d) in (3.2) is always satisfied if we define $f(\cdot) \equiv 0$ to mean that $\langle f(\cdot) - n(\cdot), f(\cdot) - n(\cdot) \rangle_{\mathcal{L}(R)} = 0$, where $n(\cdot)$ denotes the zero function, i.e., $n(\mathbf{x}) = 0$ for every $\mathbf{x} \in \mathcal{D}$.

The property (c) in (3.2), i.e., $\langle f(\cdot), f(\cdot) \rangle_{\mathcal{L}(R)} \geq 0$ for any $f(\cdot) = \sum_{l \in [L_1]} a_l R(\cdot, \mathbf{x}_l) \in \mathcal{L}(R)$ follows straightforwardly from (3.46) and the formula (3.49) evaluated for $f(\cdot) = g(\cdot)$. Indeed,

$$\langle f(\cdot), f(\cdot) \rangle_{\mathcal{L}(R)} = \sum_{l \in [L_1]} \sum_{l' \in [L_1]} a_l a_{l'} R(\mathbf{x}_{1,l'}, \mathbf{x}_{1,l}) = \mathbf{a}^T \mathbf{R} \mathbf{a} \geq 0, \quad (3.51)$$

with the vector $\mathbf{a} \in \mathbb{R}^{L_1}$ whose entries are the coefficients a_l and psd matrix $\mathbf{R} \in \mathbb{R}^{L_1 \times L_1}$ whose elements are given by $(\mathbf{R})_{m,n} \triangleq R(\mathbf{x}_{1,m}, \mathbf{x}_{1,n})$.

Finally, we have that $\langle f(\cdot) - g(\cdot), f(\cdot) - g(\cdot) \rangle_{\mathcal{L}(R)} = 0$ implies $f(\mathbf{x}) = g(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{D}$. In particular, $f(\cdot) \equiv 0$ implies that $f(\mathbf{x}) = 0$ for every $\mathbf{x} \in \mathcal{D}$. Indeed, we have by (3.44) (which is a special case of (3.49)) and Theorem 3.1.1 that for any $\mathbf{x} \in \mathcal{D}$

$$\begin{aligned} |f(\mathbf{x}) - g(\mathbf{x})| &\stackrel{(3.44)}{=} |\langle f(\cdot) - g(\cdot), R(\cdot, \mathbf{x}) \rangle_{\mathcal{L}(R)}| \\ &\stackrel{(3.3)}{\leq} \sqrt{\langle f(\cdot) - g(\cdot), f(\cdot) - g(\cdot) \rangle_{\mathcal{L}(R)}} \sqrt{\langle R(\cdot, \mathbf{x}), R(\cdot, \mathbf{x}) \rangle_{\mathcal{L}(R)}} \\ &\stackrel{(3.48)}{=} \sqrt{\langle f(\cdot) - g(\cdot), f(\cdot) - g(\cdot) \rangle_{\mathcal{L}(R)}} \sqrt{R(\mathbf{x}, \mathbf{x})} = 0, \end{aligned}$$

i.e., $f(\mathbf{x}) = g(\mathbf{x})$ for any $\mathbf{x} \in \mathcal{D}$. □

With a slight abuse of notation, we will use from now on the same symbol $\mathcal{L}(R)$ for both the linear space given by (3.47) and the inner-product space that is given by this linear space together with the inner product $\langle f(\cdot), g(\cdot) \rangle_{\mathcal{L}(R)}$ defined by (3.48). It should be clear from the context what precise meaning of the symbol $\mathcal{L}(R)$ is used.

Based on the inner-product space $\mathcal{L}(R)$, one can show

Theorem 3.2.2. *Consider an arbitrary kernel function $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ over a domain \mathcal{D} , and let \mathcal{C} denote the set of all functions $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ that are the pointwise limits of a Cauchy sequence in the inner-product space $\mathcal{L}(R)$. Then, the set of functions $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ denoted by $\mathcal{H}(R)$ and defined as*

$$\mathcal{H}(R) \triangleq \mathcal{L}(R) \cup \mathcal{C} \quad (3.52)$$

forms a RKHS with kernel $R(\cdot, \cdot)$. Moreover, the linear space $\mathcal{L}(R)$ is dense in the Hilbert space $\mathcal{H}(R)$, i.e., the set $\{R(\cdot, \mathbf{x}) \in \mathcal{L}(R)\}_{\mathbf{x} \in \mathcal{D}}$ spans the RKHS $\mathcal{H}(R)$. The inner product on the Hilbert space $\mathcal{H}(R)$ is completely determined by the basic properties (3.2) and by (3.44).

Proof. [3, 62, 65] □

Given a kernel $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$, we have that the RKHS $\mathcal{H}(R)$ constructed in Theorem 3.2.2 is the unique RKHS associated with this kernel, as stated in

Theorem 3.2.3. *Given an arbitrary kernel function $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ over a domain \mathcal{D} , there exists a unique RKHS which satisfies (3.43) and (3.44). This unique RKHS coincides with $\mathcal{H}(R)$ constructed in Theorem 3.2.2.*

Proof. The statement about the existence is proven trivially by Theorem 3.2.2. A proof of the uniqueness can be found in [3, 62]. □

Example 3.2.4. *A well-known example of a RKHS is the Euclidean space \mathbb{R}^N endowed with the inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k \in [N]} x_k y_k$. This is a RKHS of functions over the domain $\mathcal{D} = [N]$, which can be represented by length N vectors. This RKHS is associated to the kernel $R(\cdot, \cdot) : [N] \times [N] \rightarrow \mathbb{R} : R(k, l) = \delta_{k,l}$, i.e., the kernel can be represented by the identity matrix \mathbf{I}_N . For any $k \in [N]$ we have $\mathbf{e}_k = R(\cdot, k) \in \mathbb{R}^N$ which verifies (3.43). Next, for any vector $\mathbf{a} \in \mathbb{R}^N$ it holds that $a_k = \mathbf{a}^T \mathbf{e}_k = \langle \mathbf{a}, R(\cdot, k) \rangle$ verifying (3.44). Finally, one can verify (3.48) since $\langle R(\cdot, k), R(\cdot, l) \rangle = \mathbf{e}_k^T \mathbf{e}_l = \delta_{k,l} = R(k, l)$.*

3.3 Important Facts about RKHS

Theorem 3.3.1. *Consider a RKHS $\mathcal{H}(R)$ and a set $\{v_l(\cdot) \in \mathcal{H}(R)\}_{l \in \mathcal{T}}$, where \mathcal{T} is an arbitrary index set. If the set $\{v_l(\cdot) \in \mathcal{H}(R)\}_{l \in \mathcal{T}}$ is dense in $\mathcal{L}(R)$, then it is also dense in the RKHS $\mathcal{H}(R)$.*

Proof. Consider an arbitrary $f(\cdot) \in \mathcal{H}(R)$. Since by Theorem 3.2.2 we have that the linear space $\mathcal{L}(R)$ is dense in $\mathcal{H}(R)$, we can find for any $\varepsilon > 0$ a function $g(\cdot) \in \mathcal{L}(R)$ such that $\|g(\cdot) - f(\cdot)\|_{\mathcal{H}(R)} \leq \varepsilon$. Furthermore, since the set $\{v_l(\cdot) \in \mathcal{H}(R)\}_{l \in \mathcal{T}}$ is dense in $\mathcal{L}(R)$, we can also find a function $v_l(\cdot)$ such that $\|g(\cdot) - v_l(\cdot)\|_{\mathcal{H}(R)} \leq \varepsilon$. By the triangle inequality (3.7), we have

$$\begin{aligned} \|f(\cdot) - v_l(\cdot)\|_{\mathcal{H}(R)} &= \|(f(\cdot) - g(\cdot)) + (g(\cdot) - v_l(\cdot))\|_{\mathcal{H}(R)} \\ &\leq \|f(\cdot) - g(\cdot)\|_{\mathcal{H}(R)} + \|g(\cdot) - v_l(\cdot)\|_{\mathcal{H}(R)} \leq 2\varepsilon. \end{aligned} \quad (3.53)$$

Thus, we have shown that any $f(\cdot) \in \mathcal{H}(R)$ can be approximated arbitrary well by a function $v_l(\cdot)$. \square

Theorem 3.3.2. *Consider a RKHS $\mathcal{H}(R)$ over the domain \mathcal{D} . If a sequence $\{f_l(\cdot) \in \mathcal{H}(R)\}_{l \rightarrow \infty}$ converges to a function $f(\cdot) \in \mathcal{H}(R)$, i.e., $\lim_{l \rightarrow \infty} f_l(\cdot) = f(\cdot)$, then the sequence converges also pointwise to $f(\cdot)$, i.e., for every $\mathbf{x} \in \mathcal{D}$ we have that $\lim_{l \rightarrow \infty} f_l(\mathbf{x}) = f(\mathbf{x})$. Moreover, this pointwise convergence is uniform on every subset $\mathcal{D}' \subseteq \mathcal{D}$ on which $R(\mathbf{x}, \mathbf{x})$ is uniformly bounded, i.e., there exists a constant C such that $R(\mathbf{x}, \mathbf{x}) \leq C$ for every $\mathbf{x} \in \mathcal{D}'$.*

Proof. [62] \square

A direct consequence of Theorem 3.3.2 which is of major importance for our purposes is stated in

Theorem 3.3.3. *Consider a RKHS $\mathcal{H}(R)$ over the domain \mathcal{D} and an arbitrary Cauchy sequence $\{f_l(\cdot) \in \mathcal{H}(R)\}_{l \rightarrow \infty}$. If the sequence converges pointwise to the function $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$, i.e., for every $\mathbf{x} \in \mathcal{D}$ we have $\lim_{l \rightarrow \infty} f_l(\mathbf{x}) = f(\mathbf{x})$, then the sequence converges to $f(\cdot)$, i.e., $\lim_{l \rightarrow \infty} f_l(\cdot) = f(\cdot)$ and $f(\cdot) \in \mathcal{H}(R)$.*

Proof. This result is proven by contradiction: Denote the limit of the Cauchy sequence by $f'(\cdot)$, i.e., $f'(\cdot) = \lim_{l \rightarrow \infty} f_l(\cdot)$. Note that the limit exists since per definition any Cauchy sequence in a Hilbert space has a limit. Now let us assume that the limit $f'(\cdot)$ is different from $f(\cdot)$. Since $f'(\cdot)$ and $f(\cdot)$ are different, there must be some point $\mathbf{x}' \in \mathcal{D}$ for which $f'(\mathbf{x}') \neq f(\mathbf{x}')$. However, since the sequence $\{f_l(\cdot)\}_{l \rightarrow \infty}$ converges to $f'(\cdot)$ in RKHS norm, we have according to Theorem 3.3.2 that the sequence $\{f_l(\mathbf{x}')\}_{l \rightarrow \infty}$ must converge to $f'(\mathbf{x}')$. But this is a contradiction to the fact that $\{f_l(\mathbf{x}')\}_{l \rightarrow \infty}$ converges to $f(\mathbf{x}')$. \square

Theorem 3.3.4. *Consider two RKHSs $\mathcal{H}(R_1)$ and $\mathcal{H}(R_2)$ defined over the domains \mathcal{D}_1 and \mathcal{D}_2 , respectively, as well as two sets $\mathcal{A} = \{f_j(\cdot) \in \mathcal{H}(R_1)\}_{j \in \mathcal{T}}$ and $\mathcal{B} = \{g_j(\cdot) \in \mathcal{H}(R_2)\}_{j \in \mathcal{T}}$, indexed by the same index set \mathcal{T} , that span $\mathcal{H}(R_1)$ and $\mathcal{H}(R_2)$, respectively. Assume that these two sets are such that*

$$t, s \in \mathcal{T} \Rightarrow \langle f_t(\cdot), f_s(\cdot) \rangle_{\mathcal{H}(R_1)} = \langle g_t(\cdot), g_s(\cdot) \rangle_{\mathcal{H}(R_2)}. \quad (3.54)$$

Then a sufficient condition for a linear mapping $K[\cdot] : \mathcal{H}(R_1) \rightarrow \mathcal{H}(R_2)$, satisfying $K[f_t(\cdot)] = g_t(\cdot)$ for every $t \in \mathcal{T}$, to be a congruence from $\mathcal{H}(R_1)$ to $\mathcal{H}(R_2)$ is that for any sequence $\{f_l(\cdot) \in \text{span}\{\mathcal{A}\}\}_{l \rightarrow \infty}$ such that the pointwise limit function $f(\mathbf{x}) = \lim_{k \rightarrow \infty} f_k(\mathbf{x})$ exists and belongs to $\mathcal{H}(R_1)$, we have that

$$\forall \mathbf{x} \in \mathcal{D}_2 : K[f(\cdot)](\mathbf{x}) = \lim_{l \rightarrow \infty} K[f_l(\cdot)](\mathbf{x}). \quad (3.55)$$

Proof. This result is proven by Theorem 3.1.5. Obviously we have that $K[\cdot]$ satisfies (3.28) and it only remains to verify (3.29). To that end, consider an arbitrary Cauchy sequence $\{f_l(\cdot) \in \text{span}\{\mathcal{A}\}\}_{l \rightarrow \infty}$ with limit $f(\cdot) \in \mathcal{H}(R_1)$ and the associated sequence $\{K[f_l(\cdot)] \in \text{span}\{\mathcal{B}\}\}_{l \rightarrow \infty}$ which due to (3.54) is also a Cauchy sequence with some limit $g(\cdot) \in \mathcal{H}(R_2)$. By Theorem 3.3.2, we have that the sequence $\{f_l(\cdot) \in \text{span}\{\mathcal{A}\}\}_{l \rightarrow \infty}$ converges pointwise to $f(\cdot)$, which in turn implies via (3.55) that the Cauchy sequence $\{K[f_l(\cdot)](\cdot) \in \text{span}\{\mathcal{B}\}\}_{l \rightarrow \infty}$ converges pointwise to the function $K[f(\cdot)](\cdot)$. Since by Theorem 3.3.3 and $\lim_{l \rightarrow \infty} K[f_l(\cdot)] = g(\cdot)$, we have that $K[f(\cdot)] = g(\cdot)$, we have verified (3.29). \square

A main theme of this thesis is how a certain RKHS \mathcal{H}_1 defined over domain \mathcal{D}_1 is related to another RKHS \mathcal{H}_2 that is obtained by restricting the functions $f(\cdot) \in \mathcal{H}_1$ to the subset $\mathcal{D}_2 \subseteq \mathcal{D}_1$. The most important result concerning this question is stated in

Theorem 3.3.5. *Consider a RKHS $\mathcal{H}(R_1)$ associated with the kernel $R_1(\cdot, \cdot) : \mathcal{D}_1 \times \mathcal{D}_1 \rightarrow \mathbb{R}$, whose elements are functions with domain \mathcal{D}_1 . Then the set of functions that is obtained by restricting each function $f(\cdot) \in \mathcal{H}(R_1)$ to the subdomain $\mathcal{D}_2 \subseteq \mathcal{D}_1$ coincides with the RKHS $\mathcal{H}(R_2)$ associated with the kernel $R_2(\cdot, \cdot) : \mathcal{D}_2 \times \mathcal{D}_2 \rightarrow \mathbb{R}$ that is the restriction of the kernel $R_1(\cdot, \cdot) : \mathcal{D}_1 \times \mathcal{D}_1 \rightarrow \mathbb{R}$ to the subdomain $\mathcal{D}_2 \times \mathcal{D}_2$, i.e., $R_2(\cdot, \cdot) = R_1(\cdot, \cdot)|_{\mathcal{D}_2 \times \mathcal{D}_2}$. Furthermore, we have the relation*

$$\|f(\cdot)\|_{\mathcal{H}(R_2)} = \min_{\substack{g(\cdot) \in \mathcal{H}(R_1) \\ g(\cdot)|_{\mathcal{D}_2} = f(\cdot)}} \|g(\cdot)\|_{\mathcal{H}(R_1)}. \quad (3.56)$$

Thus, the norm of any element $f(\cdot) \in \mathcal{H}(R_2)$ is equal to the minimal norm of a function $g(\cdot) \in \mathcal{H}(R_1)$ that coincides with $f(\cdot)$ on the restricted domain \mathcal{D}_2 .

Proof. [62] \square

In what follows, we will need

Theorem 3.3.6. *Consider a RKHS $\mathcal{H}(R)$ whose kernel $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ is given pointwise by*

$$R(\mathbf{x}_1, \mathbf{x}_2) = \sum_{k \in \mathcal{T}} g_k(\mathbf{x}_1) g_k(\mathbf{x}_2) \quad (3.57)$$

with an arbitrary index set \mathcal{T} . The functions $g_k(\cdot) \in \mathcal{H}(R)$ are assumed to be orthonormal, i.e.,

$$\langle g_k(\cdot), g_l(\cdot) \rangle_{\mathcal{H}(R)} = \delta_{k,l}. \quad (3.58)$$

Then we have that the set $\{g_k(\cdot) \in \mathcal{H}(R)\}_{k \in \mathcal{T}}$ forms an ONB for the RKHS $\mathcal{H}(R)$.

Proof. The proof is a slight modification of the derivation of [67, Theorem 3.12]. First, we note that the sum $\sum_{k \in \mathcal{T}} g_k(\mathbf{x}_1)g_k(\mathbf{x}_2)$ can be understood in the usual sense, i.e., as the supremum over all finite sums $\sum_{k \in \mathcal{T}'} g_k(\mathbf{x}_1)g_k(\mathbf{x}_2)$ with a finite set $\mathcal{T}' \subseteq \mathcal{T}$. This is possible via the Cauchy-Schwarz inequality for the Hilbert space $\ell^2(\mathcal{T})$ (cf. Theorem 3.1.1 and [6]) since the sequences $a[k] \triangleq g_k(\mathbf{x}_1)$, $b[k] \triangleq g_k(\mathbf{x}_2)$ both belong to $\ell^2(\mathcal{T})$ for any $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{D}$.

We now show that any function $f(\cdot) \in \mathcal{H}(R)$ can be approximated arbitrarily well by a finite linear combination of the functions $\{g_k(\cdot) \in \mathcal{H}(R)\}_{k \in \mathcal{T}}$, i.e., the set spans the RKHS $\mathcal{H}(R)$ which implies that $\{g_k(\cdot) \in \mathcal{H}(R)\}_{k \in \mathcal{T}}$ is an ONB for $\mathcal{H}(R)$. To this end, according to Theorem 3.3.1, it is already sufficient to show that any function $f(\cdot) \in \mathcal{L}(R)$ in the linear span $\mathcal{L}(R)$ can be approximated arbitrarily well.

Consider an arbitrary function $f(\cdot) \in \mathcal{L}(R)$ (cf. (3.47)), i.e., $f(\cdot) = \sum_{l \in [L]} a_l R(\cdot, \mathbf{x}_l)$ with some $a_l \in \mathbb{R}$ and $\mathbf{x}_l \in \mathcal{D}$, $l \in [L]$. We then have that

$$\begin{aligned}
\|f(\cdot)\|_{\mathcal{H}(R)}^2 &= \left\langle \sum_{l \in [L]} a_l R(\cdot, \mathbf{x}_l), \sum_{l \in [L]} a_l R(\cdot, \mathbf{x}_l) \right\rangle_{\mathcal{H}(R)} \\
&= \sum_{l, l' \in [L]} a_l a_{l'} R(\mathbf{x}_l, \mathbf{x}_{l'}) \stackrel{(3.57)}{=} \sum_{l, l' \in [L]} a_l a_{l'} \sum_{k \in \mathcal{T}} g_k(\mathbf{x}_l) g_k(\mathbf{x}_{l'}) \\
&= \sum_{l, l' \in [L]} a_l a_{l'} \sum_{k \in \mathcal{T}} \left\langle g_k(\cdot), R(\cdot, \mathbf{x}_l) \right\rangle_{\mathcal{H}(R)} \left\langle g_k(\cdot), R(\cdot, \mathbf{x}_{l'}) \right\rangle_{\mathcal{H}(R)} \\
&\stackrel{(a)}{=} \sum_{k \in \mathcal{T}} \left\langle g_k(\cdot), \sum_{l \in [L]} a_l R(\cdot, \mathbf{x}_l) \right\rangle_{\mathcal{H}(R)} \left\langle g_k(\cdot), \sum_{l' \in [L]} a_{l'} R(\cdot, \mathbf{x}_{l'}) \right\rangle_{\mathcal{H}(R)} \\
&= \sum_{k \in \mathcal{T}} \left\langle g_k(\cdot), f(\cdot) \right\rangle_{\mathcal{H}(R)} \left\langle g_k(\cdot), f(\cdot) \right\rangle_{\mathcal{H}(R)} \\
&= \sum_{k \in \mathcal{T}} |\langle g_k(\cdot), f(\cdot) \rangle_{\mathcal{H}(R)}|^2, \tag{3.59}
\end{aligned}$$

where the change of summation order in (a) can be validated by an argument based on the Cauchy-Schwarz inequality for the Hilbert space $\ell^2(\mathcal{T})$ (cf. Theorem 3.1.1 and [6]). Let us now introduce the function $s_{\mathcal{T}'}(\cdot) \triangleq \sum_{k \in \mathcal{T}'} \langle f(\cdot), g_k(\cdot) \rangle_{\mathcal{H}(R)} g_k(\cdot) \in \text{span}\{g_k(\cdot)\}_{k \in \mathcal{T}}$ for any finite set $\mathcal{T}' \subseteq \mathcal{T}$. It can be verified that according to Theorem 3.1.7, the function $s_{\mathcal{T}'}(\cdot)$ coincides with the orthogonal projection of $f(\cdot)$ on the subspace

$$\mathcal{U}^{(\mathcal{T}')} \triangleq \text{span}\{g_k(\cdot)\}_{k \in \mathcal{T}'} \subseteq \text{span}\{g_k(\cdot)\}_{k \in \mathcal{T}} \subseteq \mathcal{H}(R), \tag{3.60}$$

i.e., $\mathbf{P}_{\mathcal{U}(\mathcal{T}')} f(\cdot) = s_{\mathcal{T}'}(\cdot)$. Obviously, an ONB for the subspace $\mathcal{U}(\mathcal{T}')$ is given by the function set $\text{span}\{g_k(\cdot)\}_{k \in \mathcal{T}'}$. Therefore, by Theorem 3.1.6 and Theorem 3.1.7, we have

$$\begin{aligned}
\|f(\cdot) - s_{\mathcal{T}'}(\cdot)\|_{\mathcal{H}(R)}^2 &= \|f(\cdot) - \mathbf{P}_{\mathcal{U}(\mathcal{T}')} f(\cdot)\|_{\mathcal{H}(R)}^2 \stackrel{(3.31)}{=} \|f(\cdot)\|_{\mathcal{H}(R)}^2 - \|\mathbf{P}_{\mathcal{U}(\mathcal{T}')} f(\cdot)\|_{\mathcal{H}(R)}^2 \\
&\stackrel{(3.33)}{=} \|f(\cdot)\|_{\mathcal{H}}^2 - \sum_{k \in \mathcal{T}'} |\langle f(\cdot), g_k(\cdot) \rangle_{\mathcal{H}(R)}|^2 \\
&\stackrel{(3.59)}{=} \sum_{k \in \mathcal{T}} |\langle f(\cdot), g_k(\cdot) \rangle_{\mathcal{H}(R)}|^2 - \sum_{k \in \mathcal{T}'} |\langle f(\cdot), g_k(\cdot) \rangle_{\mathcal{H}(R)}|^2 \\
&= \sum_{k \in \mathcal{T} \setminus \mathcal{T}'} |\langle f(\cdot), g_k(\cdot) \rangle_{\mathcal{H}(R)}|^2. \tag{3.61}
\end{aligned}$$

From this it follows that

$$\inf_{\substack{|\mathcal{T}'| \in \mathbb{N} \\ \mathcal{T}' \subseteq \mathcal{T}}} \|f(\cdot) - s_{\mathcal{T}'}(\cdot)\|_{\mathcal{H}(R)} = 0,$$

i.e., any function $f(\cdot) \in \mathcal{L}(R)$ can be approximated arbitrary well by an element of $\text{span}\{g_k(\cdot)\}_{k \in \mathcal{T}}$ due to (3.60). Thus, we have that $\text{span}\{g_k(\cdot)\}_{k \in \mathcal{T}}$ is dense in the linear space $\mathcal{L}(R)$ and by Theorem 3.3.1 it is also dense in $\mathcal{H}(R)$. This means in turn that the set $\{g_k(\cdot)\}_{k \in \mathcal{T}}$ of orthonormal functions spans the RKHS $\mathcal{H}(R)$, i.e., is an ONB for $\mathcal{H}(R)$. \square

3.4 RKHSs with a Differentiable Kernel

The main part of this thesis is concerned exclusively with RKHSs over a domain $\mathcal{D} \subseteq \mathbb{R}^N$ that are associated with a “differentiable” kernel [2, 68]. It will turn out that under specific conditions, the functions of a RKHS associated with a differentiable kernel are characterized completely by their behavior in an arbitrarily small neighborhood around a single point belonging to the domain \mathcal{D} . Note that it will always be implicitly assumed that the standard topology, induced by the inner product $\mathbf{x}^T \mathbf{y}$, is used for \mathbb{R}^N .

For a precise definition of a differentiable kernel, consider an index set $\mathcal{K} \subseteq [N]$ and a vector $\mathbf{x}_c \in \mathbb{R}^N$, and define the “ ε - \mathcal{K} -neighborhood” of \mathbf{x}_c by

$$\mathcal{N}_{\mathbf{x}_c}^{\mathcal{K}}(\varepsilon) \triangleq \{\mathbf{x}_c + \mathbf{a} \mid \mathbf{a} \in \mathbb{R}_+^N, \text{supp}(\mathbf{a}) = \mathcal{K}, \|\mathbf{a}\|_{\infty} \leq \varepsilon\}. \tag{3.62}$$

Definition 3.4.1. A kernel $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ over a domain $\mathcal{D} \subseteq \mathbb{R}^N$ is said to be differentiable up to order m if for any \mathbf{x}_c and \mathcal{K} for which there exists an $\varepsilon > 0$ such that $\mathcal{N}_{\mathbf{x}_c}^{\mathcal{K}}(\varepsilon) \subseteq \mathcal{D}$, the partial derivatives

$$\left. \frac{\partial^{\mathbf{p}_1} \partial^{\mathbf{p}_2} R(\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{x}_1^{\mathbf{p}_1} \partial \mathbf{x}_2^{\mathbf{p}_2}} \right|_{\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}_c} \tag{3.63}$$

exist for any orders $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{Z}_+^N$ with $\text{supp}(\mathbf{p}_1), \text{supp}(\mathbf{p}_2) \subseteq \mathcal{K}$ and $\|\mathbf{p}_1\|_\infty, \|\mathbf{p}_2\|_\infty \leq m$, and moreover this partial derivatives are continuous functions of $(\mathbf{x}_1, \mathbf{x}_2)$ (viewed as a vector in \mathbb{R}^{2N}). We will call a RKHS $\mathcal{H}(R)$ differentiable, if it is associated to a differentiable kernel $R(\cdot, \cdot)$.

An important property of the RKHS associated to a differentiable kernel is stated in

Theorem 3.4.1. *Let $\mathcal{D} \subseteq \mathbb{R}^N$, and consider a RKHS $\mathcal{H}(R)$ associated to a kernel $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ which is differentiable up to order m . Then for any $\mathbf{p} \in \mathbb{Z}_+^N$ with $\|\mathbf{p}\|_\infty \leq m$, the function $g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ defined by*

$$g_{\mathbf{x}_c}^{(\mathbf{p})}(\mathbf{x}) \triangleq \left. \frac{\partial^{\mathbf{p}} R(\mathbf{x}, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}}} \right|_{\mathbf{x}_2 = \mathbf{x}_c}, \quad (3.64)$$

where $\mathbf{x}_c \in \mathcal{D}$ is such that $\mathcal{N}_{\mathbf{x}_c}^{\text{supp}(\mathbf{p})}(\varepsilon) \subseteq \mathcal{D}$ (for a suitable $\varepsilon > 0$), is an element of $\mathcal{H}(R)$, i.e.,

$$g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot) \in \mathcal{H}(R). \quad (3.65)$$

Furthermore, the inner product of the function $g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)$ with an arbitrary function $f(\cdot) \in \mathcal{H}(R)$ is given by

$$\left\langle g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot), f(\cdot) \right\rangle_{\mathcal{H}(R)} = \left. \frac{\partial^{\mathbf{p}} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \right|_{\mathbf{x} = \mathbf{x}_c}. \quad (3.66)$$

Proof. [68] □

We will also need the following slight variation of Theorem 3.4.1.

Corollary 3.4.2. *Consider a RKHS $\mathcal{H}(R)$ associated to a kernel $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$, $\mathcal{D} \subseteq \mathbb{R}^N$, which is differentiable up to order $m \in \mathbb{N}$. Furthermore, consider a matrix $\mathbf{A} \in \mathbb{R}^{N \times D}$, a vector $\mathbf{z}_c \in \mathbb{R}^D$, and a multi-index $\mathbf{p} \in \mathbb{Z}_+^D$ with $\|\mathbf{p}\|_\infty \leq m$ that satisfy $\mathbf{A}\mathcal{N}_{\mathbf{z}_c}^{\text{supp}(\mathbf{p})}(\varepsilon) \subseteq \mathcal{D}$ for some small $\varepsilon > 0$. Then for any function $h(\cdot) : \mathbb{R}^D \rightarrow \mathbb{R}$ for which the partial derivatives of any order $\mathbf{p}' \in \mathbb{Z}_+^D$ with $\|\mathbf{p}'\|_\infty \leq m$ exist at the point $\mathbf{z}_c \in \mathbb{R}^D$, the function $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ defined by*

$$f(\mathbf{x}) \triangleq \left. \frac{\partial^{\mathbf{p}} (R(\mathbf{x}, \mathbf{A}\mathbf{z}_2)h(\mathbf{z}_2))}{\partial \mathbf{z}_2^{\mathbf{p}}} \right|_{\mathbf{z}_2 = \mathbf{z}_c}, \quad (3.67)$$

is an element of $\mathcal{H}(R)$, i.e.,

$$f(\cdot) \in \mathcal{H}(R). \quad (3.68)$$

Proof. The statement follows from Theorem 3.4.1 by induction on the order \mathbf{p} by the chain rule and product rule of differentiation [5]. We will restrict ourselves here to the begin of the induction, i.e., we consider $\mathbf{p} = \mathbf{e}_k$ with $k \in [D]$ which yields

$$f(\mathbf{x}) = \left. \frac{\partial^{\mathbf{e}_k} (R(\mathbf{x}, \mathbf{A}\mathbf{z}_2)h(\mathbf{z}_2))}{\partial \mathbf{z}_2^{\mathbf{e}_k}} \right|_{\mathbf{z}_2 = \mathbf{z}_c} = \left. \frac{\partial^{\mathbf{e}_k} R(\mathbf{x}, \mathbf{A}\mathbf{z}_2)}{\partial \mathbf{z}_2^{\mathbf{e}_k}} \right|_{\mathbf{z}_2 = \mathbf{z}_c} \left. \frac{\partial^{\mathbf{e}_k} h(\mathbf{z}_2)}{\partial \mathbf{z}_2^{\mathbf{e}_k}} \right|_{\mathbf{z}_2 = \mathbf{z}_c}$$

$$= \left[\sum_{k' \in [D]} \frac{\partial^{e_{k'}} R(\mathbf{x}, x_2)}{\partial \mathbf{x}_2^{e_{k'}}} \Big|_{\mathbf{x}_2 = \mathbf{A} \mathbf{z}_c} (\mathbf{A})_{k', k} \right] \frac{\partial^{e_k} h(\mathbf{z}_2)}{\partial \mathbf{z}_2^{e_k}} \Big|_{\mathbf{z}_2 = \mathbf{z}_c}. \quad (3.69)$$

By assumption, we have $\frac{\partial^{e_{k'}} R(\cdot, x_2)}{\partial \mathbf{x}_2^{e_{k'}}} \Big|_{\mathbf{x}_2 = \mathbf{A} \mathbf{z}_c} \in \mathcal{H}(R)$ which implies, since $f(\cdot)$ is a linear combination of the functions $\frac{\partial^{e_{k'}} R(\cdot, x_2)}{\partial \mathbf{x}_2^{e_{k'}}} \Big|_{\mathbf{x}_2 = \mathbf{A} \mathbf{z}_c}$ and $\mathcal{H}(R)$ is a Hilbert space (in particular a linear space), that $f(\cdot) \in \mathcal{H}(R)$. \square

In the next chapters, we will exclusively consider differentiable kernels that are such that the function $h(\mathbf{x}) \triangleq R(\mathbf{x}, \mathbf{x})$ is bounded on every bounded subset $\mathcal{D}' \subseteq \mathcal{D}$. Note that that a differentiable kernel up to an order $m \geq 1$ is always continuous, i.e., $\lim_{(\mathbf{x}, \mathbf{x}') \rightarrow (\mathbf{x}_0, \mathbf{x}'_0)} R(\mathbf{x}, \mathbf{x}') = R(\mathbf{x}_0, \mathbf{x}'_0)$, since the partial derivatives $\frac{\partial^{e_k} R(\mathbf{x}, \mathbf{x}_2)}{\partial x_{2,k}}$ (and $\frac{\partial^{e_k} R(\mathbf{x}, \mathbf{x}_2)}{\partial x_k}$ by symmetry of the kernel) are continuous and thus the kernel $R(\cdot, \cdot)$, viewed as a function with domain \mathbb{R}^{2N} is differentiable in the usual sense of multivariable calculus [5, Theorem 9.21]. We have that the functions belonging to a RKHS whose kernel is continuous and bounded in the above sense must be continuous:

Theorem 3.4.3. *Consider the RKHS $\mathcal{H}(R)$ associated with a continuous kernel $R(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$, with $\mathcal{D} \subseteq \mathbb{R}^N$, for which $h(\mathbf{x}) \triangleq R(\mathbf{x}, \mathbf{x})$ is bounded for $\mathbf{x} \in \mathcal{D}'$, where \mathcal{D}' is any set of the form $\mathcal{D}' = \mathcal{D} \cap \mathcal{B}(\mathbf{x}_c, r)$ with some center $\mathbf{x}_c \in \mathbb{R}^N$ and radius $r > 0$. Then every function $f(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$ belonging to $\mathcal{H}(R)$ is continuous.*

Proof. Given an arbitrary but fixed function $f(\cdot) \in \mathcal{H}(R)$ and point $\mathbf{x}_0 \in \mathcal{D}$, we have to show that for every $\varepsilon > 0$ we can find a $\delta > 0$ such that $|f(\mathbf{x}_0) - f(\mathbf{x})| < \varepsilon$ for any $\mathbf{x} \in \mathcal{D}$ with $\|\mathbf{x} - \mathbf{x}_0\|_2 < \delta$.

First we observe that any function $f_{\mathbf{x}}(\cdot) = R(\cdot, \mathbf{x})$ with $\mathbf{x} \in \mathcal{D}$ is trivially continuous by our assumption that $R(\cdot, \cdot)$ is continuous. Since a finite linear combination of continuous functions is still continuous, we also have that any function $w(\cdot) \in \mathcal{L}(R)$ is continuous.

Now consider the ball $\mathcal{B}(\mathbf{x}_0, r)$ centered at \mathbf{x}_0 and with an arbitrary radius $r > 0$. According to our assumptions, there exists a constant M such that $R(\mathbf{x}, \mathbf{x}) < M$ for every $\mathbf{x} \in \mathcal{D} \cap \mathcal{B}(\mathbf{x}_0, r)$. According to Theorem 3.2.2, the linear space $\mathcal{L}(R)$ is dense in $\mathcal{H}(R)$, i.e., we can find a function $w(\cdot) \in \mathcal{L}(R)$ such that $\|w(\cdot) - f(\cdot)\|_{\mathcal{H}(R)} \leq \varepsilon / (4\sqrt{M})$ for an arbitrary but fixed $\varepsilon > 0$. Since $w(\cdot)$ is continuous, we can find a radius $\delta \leq r$ such that $|w(\mathbf{x}_0) - w(\mathbf{x})| < \varepsilon/2$ for every $\mathbf{x} \in \mathcal{D} \cap \mathcal{B}(\mathbf{x}_0, \delta)$. Now we have for any $\mathbf{x} \in \mathcal{D} \cap \mathcal{B}(\mathbf{x}_0, \delta)$ that

$$\begin{aligned} |f(\mathbf{x}_0) - f(\mathbf{x})| &= |w(\mathbf{x}_0) - w(\mathbf{x}) + [f(\mathbf{x}_0) - w(\mathbf{x}_0)] + [w(\mathbf{x}) - f(\mathbf{x})]| \\ &\stackrel{(a)}{=} |w(\mathbf{x}_0) - w(\mathbf{x}) + \langle f(\cdot) - w(\cdot), R(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(R)} + \langle f(\cdot) - w(\cdot), R(\cdot, \mathbf{x}) \rangle_{\mathcal{H}(R)}| \\ &\stackrel{(b)}{\leq} |w(\mathbf{x}_0) - w(\mathbf{x})| + |\langle f(\cdot) - w(\cdot), R(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(R)}| + |\langle f(\cdot) - w(\cdot), R(\cdot, \mathbf{x}) \rangle_{\mathcal{H}(R)}| \end{aligned}$$

$$\begin{aligned}
& \stackrel{(c)}{\leq} \varepsilon/2 + \|f(\cdot) - w(\cdot)\|_{\mathcal{H}(R)} \|R(\cdot, \mathbf{x}_0)\|_{\mathcal{H}(R)} + \|f(\cdot) - w(\cdot)\|_{\mathcal{H}(R)} \|R(\cdot, \mathbf{x})\|_{\mathcal{H}(R)} \\
& \stackrel{(d)}{=} \varepsilon/2 + \|f(\cdot) - w(\cdot)\|_{\mathcal{H}(R)} \sqrt{R(\mathbf{x}_0, \mathbf{x}_0)} + \|f(\cdot) - w(\cdot)\|_{\mathcal{H}(R)} \sqrt{R(\mathbf{x}, \mathbf{x})} \\
& \leq \varepsilon/2 + \frac{\varepsilon}{4\sqrt{M}} \sqrt{M} + \frac{\varepsilon}{4\sqrt{M}} \sqrt{M} \\
& = \varepsilon/2 + \varepsilon/4 + \varepsilon/4 = \varepsilon, \tag{3.70}
\end{aligned}$$

where (a) and (d) is due to the reproducing property (3.44), (b) follows from the inequality $|a+b| \leq |a| + |b|$ for $a, b \in \mathbb{R}$ and the step (c) is obtained by an application of the Cauchy-Schwarz inequality (cf. Theorem 3.1.1). \square

While Theorem 3.4.3 gives a sufficient condition on the kernel of a RKHS such that it consists only of continuous functions, a necessary and sufficient condition for this to be the case in a more general setting can be found in [66, Theorem 5].

Chapter 4

The RKHS Approach to Minimum Variance Estimation

4.1 Introduction

This chapter reviews the RKHS approach to minimum variance estimation as introduced in [3,35]. Using the RKHS approach, we will derive two fundamental results, which (to the best of the authors knowledge) seem to be novel. The first result makes a characterization of the minimum achievable variance $L_{\mathcal{M}}$ viewed as a function of the fixed parameter vector \mathbf{x}_0 of a minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$, when \mathbf{x}_0 is varied. In particular, we show in Section 4.3.5 that $L_{\mathcal{M}}$, viewed as a function of \mathbf{x}_0 , is lower semi-continuous. The second result is related to the concept of sufficient statistics. In particular, we prove in Section 4.3.6, that the RKHS associated to a minimum variance problem remains unchanged if the original observation is replaced by any sufficient statistic. Within this chapter we will also use the RKHS approach to give a detailed geometric derivation and interpretation of some well-known lower bounds on the variance of estimators with a given bias. Although it has been already mentioned in [3], that virtually any known lower bound on the variance can be interpreted geometrically using RKHS theory, a detailed elaboration of this interpretation for some well-known bounds seems to be missing in the literature.

From a mathematical viewpoint, the RKHS approach to minimum variance estimation deals with two specific Hilbert spaces which are naturally associated to a given minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$ associated with the estimation problem $(\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$.

The first Hilbert space, which is a RKHS and denoted $\mathcal{H}(\mathcal{M})$, consists of all functions $\gamma(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$ that are obtained from a valid bias function $c(\cdot)$ for \mathcal{M} by $\gamma(\cdot) = c(\cdot) + g(\cdot)$, where $g(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$ is the parameter function of \mathcal{E} . Thus, the RKHS $\mathcal{H}(\mathcal{M})$ consists of all functions

$\gamma(\cdot)$ for which there exists an estimator for \mathcal{M} with finite variance at \mathbf{x}_0 and whose mean function is $\gamma(\cdot)$. The function $\gamma(\cdot)$ will be referred to as the prescribed mean function of the minimum variance problem \mathcal{M} .

The second Hilbert space, denoted $\mathcal{L}(\mathcal{M})$, which is naturally associated to the minimum variance problem \mathcal{M} consists of estimators $\hat{g}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}$ with a finite variance at \mathbf{x}_0 but whose bias is not necessarily equal to the prescribed bias $c(\cdot)$ of \mathcal{M} .

It will turn out that the Hilbert spaces $\mathcal{H}(\mathcal{M})$ and $\mathcal{L}(\mathcal{M})$ are isometric. Moreover, a specific congruence between these two Hilbert spaces will give us a tool for determining the LMV estimator (if it exists) for the minimum variance problem \mathcal{M} .

In what follows, we will place an important constraint on the class of minimum variance problems that are considered. More precisely, we assume that for any considered minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$ with associated estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$, the following holds:

Postulate 4.1.1. For any $\mathbf{x} \in \mathcal{X}$,

$$\mathbf{E}_{\mathbf{x}_0} \left\{ \left(\frac{f(\mathbf{y}; \mathbf{x})}{f(\mathbf{y}; \mathbf{x}_0)} \right)^2 \right\} < \infty. \quad (4.1)$$

Unless otherwise stated, we will implicitly assume from now on that any minimum variance problem that is considered satisfies (4.1). We note that many works on minimum variance estimation introduce a variant of Postulate 4.1.1 [3, 35, 54, 55].

4.2 A Hilbert Space of Estimators

An estimator $\hat{g}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}$ for an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ is nothing but a (possibly random) mapping that maps the observation space \mathbb{R}^M to the range of the parameter function, i.e., \mathbb{R} .¹ However, for a given minimum variance problem \mathcal{M} , only the allowed estimators $\mathcal{F}(\mathcal{M})$ (see Definition 2.3.2) make sense. As mentioned in Section 2.3.4, one can easily verify that the set $\mathcal{F}(\mathcal{M})$ is an affine set of estimator functions, i.e., if $\hat{g}_1(\cdot), \hat{g}_2(\cdot) \in \mathcal{F}(\mathcal{M})$ then also any affine combination (cf. [30, p. 21]) $\hat{g}_3(\cdot) \triangleq a\hat{g}_1(\cdot) + (1-a)\hat{g}_2(\cdot)$, where $a \in \mathbb{R}$ is arbitrary, yields an allowed estimator, i.e., $\hat{g}_3(\cdot) \in \mathcal{F}(\mathcal{M})$. Moreover, every allowed estimator $\hat{g}(\cdot) \in \mathcal{F}(\mathcal{M})$ belongs to the Hilbert space $\mathcal{P}(\mathcal{M})$ associated with \mathcal{M} that is defined as

$$\mathcal{P}(\mathcal{M}) \triangleq \{\hat{g}(\cdot) \mid P(\hat{g}(\cdot); \mathbf{x}_0) < \infty\}. \quad (4.2)$$

(We recall that $P(\hat{g}(\cdot); \mathbf{x}) = v(\hat{g}(\cdot); \mathbf{x}) + |\mathbf{E}_{\mathbf{x}}\{\hat{g}(\mathbf{y})\}|^2$.) To put it formally, we have [3]

¹We will assume a scalar parameter function $g(\mathbf{x})$, i.e., $P = 1$ throughout this chapter (cf. Section 2.1).

Theorem 4.2.1. *Given a minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$, the set of estimator functions $\mathcal{P}(\mathcal{M})$ defined in (4.2) forms a function Hilbert space together with the inner product $\langle \hat{g}_1(\cdot), \hat{g}_2(\cdot) \rangle_{\text{RV}}$ defined by*

$$\langle \hat{g}_1(\cdot), \hat{g}_2(\cdot) \rangle_{\text{RV}} \triangleq \mathbb{E}_{\mathbf{x}_0} \{ \hat{g}_1(\mathbf{y}) \hat{g}_2(\mathbf{y}) \}. \quad (4.3)$$

The Hilbert space $\mathcal{P}(\mathcal{M})$ contains the set $\mathcal{F}(\mathcal{M})$ of allowed estimators for \mathcal{M} as defined in (2.18), i.e.,

$$\mathcal{F}(\mathcal{M}) \subseteq \mathcal{P}(\mathcal{M}). \quad (4.4)$$

Proof. By definition, an (possibly random) estimator $\hat{g}_1(\cdot)$ with finite stochastic power at \mathbf{x}_0 is nothing but a random variable with a finite stochastic power at \mathbf{x}_0 . Therefore, the Hilbert space structure of $\mathcal{P}(\mathcal{M})$ with the inner product $\langle \hat{g}_1(\cdot), \hat{g}_2(\cdot) \rangle_{\text{RV}}$ follows from elementary probability- or measure theory (see, e.g., [3, 6, 42, 43]).

Furthermore, by the definition of $\mathcal{F}(\mathcal{M})$ and (2.11), we have that the stochastic power at \mathbf{x}_0 of any allowed estimator $\hat{g}(\cdot) \in \mathcal{F}(\mathcal{M})$ satisfies

$$P(\hat{g}(\cdot); \mathbf{x}_0) = v(\hat{g}(\cdot); \mathbf{x}_0) + [b(\hat{g}(\cdot); \mathbf{x}_0) + g(\mathbf{x}_0)]^2 = v(\hat{g}(\cdot); \mathbf{x}_0) + [c(\mathbf{x}_0) + g(\mathbf{x}_0)]^2 < \infty,$$

verifying (4.4). □

It will turn out that, in the context of minimum variance estimation, the set $\mathcal{P}(\mathcal{M})$ can be further reduced without affecting the minimum achievable variance (cf. (2.19)). In particular, we will show that for a given minimum variance problem \mathcal{M} and associated estimation problem \mathcal{E} that satisfy (4.1), it is sufficient to consider only estimators which belong to the subspace $\mathcal{L}(\mathcal{M}) \subseteq \mathcal{P}(\mathcal{M})$ defined by

$$\mathcal{L}(\mathcal{M}) \triangleq \text{cl}\{\text{span}\{\rho_{\mathcal{M}}(\cdot, \mathbf{x})\}_{\mathbf{x} \in \mathcal{X}}\}, \quad (4.5)$$

where $\rho_{\mathcal{M}}(\cdot, \cdot) : \mathbb{R}^M \times \mathcal{X} \rightarrow \mathbb{R}$ is the likelihood ratio of the minimum variance problem \mathcal{M} defined as

$$\rho_{\mathcal{M}}(\mathbf{y}, \mathbf{x}) \triangleq \frac{f(\mathbf{y}; \mathbf{x})}{f(\mathbf{y}; \mathbf{x}_0)}. \quad (4.6)$$

Note that the condition (4.1) of our Postulate 4.1.1 can be written as

$$\mathbb{E}_{\mathbf{x}_0} \{ (\rho_{\mathcal{M}}(\mathbf{y}; \mathbf{x}))^2 \} < \infty. \quad (4.7)$$

The closure in (4.5) relative to the Hilbert space $\mathcal{P}(\mathcal{M})$ (see Definition 3.1.9) exists, since we have that

$$\text{span}\{\rho_{\mathcal{M}}(\cdot, \mathbf{x})\}_{\mathbf{x} \in \mathcal{X}} \subseteq \mathcal{P}(\mathcal{M}). \quad (4.8)$$

This can be easily verified by noting that for any estimator of the form $\hat{g}(\cdot) \triangleq \rho_{\mathcal{M}}(\cdot, \mathbf{x})$, where $\mathbf{x} \in \mathcal{X}$, it holds that

$$P(\hat{g}(\cdot); \mathbf{x}_0) = \mathbb{E}_{\mathbf{x}_0} \{ (\hat{g}(\mathbf{y}))^2 \} = \mathbb{E}_{\mathbf{x}_0} \{ (\rho_{\mathcal{M}}(\mathbf{y}, \mathbf{x}))^2 \} < \infty, \quad (4.9)$$

where the inequality follows from (4.1).

We emphasize the fact that in contrast to the set of allowed estimators $\mathcal{F}(\mathcal{M})$, the set $\mathcal{L}(\mathcal{M})$ is by its very definition a function Hilbert space with associated inner product $\langle \hat{g}_1(\cdot), \hat{g}_2(\cdot) \rangle_{\text{RV}}$. This is a very important fact since, as we will see in Section 4.3.2, it guarantees that if the set $\mathcal{F}(\mathcal{M})$ is nonempty, the infimum in (2.19) can be achieved by a suitable estimator, i.e., there exists an LMV estimator if there is at least one estimator with the prescribed bias and finite variance at \mathbf{x}_0 .

4.3 The RKHS Associated to a Minimum Variance Problem

In this section, we introduce the section Hilbert space, beside $\mathcal{L}(\mathcal{M})$ defined in (4.5), which is naturally associated with a given minimum variance problem. This second Hilbert space has a particular structure, i.e., it is a RKHS.

4.3.1 Definition

Consider a minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$, associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$, that satisfies the condition of Postulate 4.1.1. We then have for any pair $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$ that the corresponding likelihood ratios $\rho_{\mathcal{M}}(\cdot, \mathbf{x}_1)$, $\rho_{\mathcal{M}}(\cdot, \mathbf{x}_2)$ belong to the Hilbert space $\mathcal{P}(\mathcal{M})$ with inner product $\langle \rho_{\mathcal{M}}(\cdot, \mathbf{x}_1), \rho_{\mathcal{M}}(\cdot, \mathbf{x}_2) \rangle_{\text{RV}}$.

Therefore, one can naturally associate with the minimum variance problem \mathcal{M} a real-valued function $R_{\mathcal{M}}(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ defined as [3]

$$\begin{aligned} R_{\mathcal{M}}(\mathbf{x}_1, \mathbf{x}_2) &\triangleq \mathbf{E}_{\mathbf{x}_0} \{ \rho_{\mathcal{M}}(\mathbf{y}, \mathbf{x}_1) \rho_{\mathcal{M}}(\mathbf{y}, \mathbf{x}_2) \} = \langle \rho_{\mathcal{M}}(\cdot, \mathbf{x}_1), \rho_{\mathcal{M}}(\cdot, \mathbf{x}_2) \rangle_{\text{RV}} \\ &= \int_{\mathbf{y}} \frac{f(\mathbf{y}; \mathbf{x}_1) f(\mathbf{y}; \mathbf{x}_2)}{f(\mathbf{y}; \mathbf{x}_0)} d\mathbf{y}. \end{aligned} \quad (4.10)$$

An easily verified and useful observation is that $R_{\mathcal{M}}(\mathbf{x}_0, \mathbf{x}) = 1$ for every $\mathbf{x} \in \mathcal{X}$.

The function $R_{\mathcal{M}}(\cdot, \cdot)$ has an important property as stated in

Theorem 4.3.1. *The function $R_{\mathcal{M}}(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is a kernel function with domain \mathcal{X} in the sense of Definition 3.2.2. Moreover, for every $\mathbf{x} \in \mathcal{X}$ it holds that*

$$R_{\mathcal{M}}(\mathbf{x}, \mathbf{x}_0) = R_{\mathcal{M}}(\mathbf{x}_0, \mathbf{x}) = 1, \quad (4.11)$$

where \mathbf{x}_0 is the fixed parameter vector associated with the minimum variance problem \mathcal{M} .

Proof. Consider an arbitrary finite set $\{\mathbf{x}_l\}_{l \in [L]}$ consisting of $L \in \mathbb{N}$ points $\mathbf{x}_l \in \mathcal{X}$. The matrix $\mathbf{R} \in \mathbb{R}^{L \times L}$ defined elementwise by $(\mathbf{R})_{l,l'} \triangleq R_{\mathcal{M}}(\mathbf{x}_l, \mathbf{x}_{l'})$ can be verified to coincide with the

correlation matrix $\mathbf{E}_{\mathbf{x}_0}\{\mathbf{r}\mathbf{r}^T\}$ of the random vector $\mathbf{r} \in \mathbb{R}^L$ given elementwise by $r_l \triangleq \rho_{\mathcal{M}}(\mathbf{y}, \mathbf{x}_l)$, i.e., $\mathbf{R} = \mathbf{E}_{\mathbf{x}_0}\{\mathbf{r}\mathbf{r}^T\}$. Thus, since any correlation matrix of a real-valued random vector is psd [45], we have that \mathbf{R} is psd for any finite set $\{\mathbf{x}_l\}_{l \in [L]}$. Furthermore, it is straightforward to verify that $R_{\mathcal{M}}(\mathbf{x}_1, \mathbf{x}_2) = R_{\mathcal{M}}(\mathbf{x}_2, \mathbf{x}_1)$ for any pair $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}$. Thus, the two axioms of Definition 3.2.2 are satisfied.

The identity (4.11) follows from the symmetry of $R_{\mathcal{M}}(\cdot, \cdot)$ and

$$R_{\mathcal{M}}(\mathbf{x}, \mathbf{x}_0) = \int_{\mathbf{y}} \frac{f(\mathbf{y}; \mathbf{x})f(\mathbf{y}; \mathbf{x}_0)}{f(\mathbf{y}; \mathbf{x}_0)} d\mathbf{y} = \int_{\mathbf{y}} f(\mathbf{y}; \mathbf{x}) d\mathbf{y} = 1. \quad (4.12)$$

□

We are now in the position to make

Definition 4.3.1. *Given a minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$ associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$, we associate with it the RKHS $\mathcal{H}(\mathcal{M})$ that is given via the kernel $R_{\mathcal{M}}(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ and Theorem 3.2.2.*

Thus, the RKHS $\mathcal{H}(\mathcal{M})$ is given by $\mathcal{H}(\mathcal{M}) = \mathcal{L}(R_{\mathcal{M}}) \cup \mathcal{C}$ where $\mathcal{L}(R_{\mathcal{M}})$ is the inner-product space given by Definition 3.2.3 and \mathcal{C} denotes the set of functions $f(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$, that are pointwise limits of a Cauchy sequence in the inner-product space $\mathcal{L}(R_{\mathcal{M}})$.

4.3.2 Main Result

Besides the RKHS $\mathcal{H}(\mathcal{M})$, a second Hilbert space naturally associated to a minimum variance problem \mathcal{M} is the Hilbert space $\mathcal{L}(\mathcal{M})$ defined in (4.5). Since both Hilbert spaces, $\mathcal{H}(\mathcal{M})$ and $\mathcal{L}(\mathcal{M})$, are associated with the same minimum variance problem \mathcal{M} , it does not come as a big surprise that these two Hilbert spaces are closely related to each other. In particular, we have [3,35]

Theorem 4.3.2. *Consider a minimum variance problem \mathcal{M} with the associated Hilbert spaces $\mathcal{H}(\mathcal{M})$ and $\mathcal{L}(\mathcal{M})$ given by Definition 4.3.1 and (4.5) respectively. We then have that these two Hilbert spaces are isometric. Moreover, there exists a unique congruence $\mathbf{J}[\cdot] : \mathcal{H}(\mathcal{M}) \rightarrow \mathcal{L}(\mathcal{M})$ from $\mathcal{H}(\mathcal{M})$ to $\mathcal{L}(\mathcal{M})$ that satisfies*

$$\mathbf{J}[R_{\mathcal{M}}(\cdot, \mathbf{x})] = \rho_{\mathcal{M}}(\cdot, \mathbf{x}) \quad (4.13)$$

for every $\mathbf{x} \in \mathcal{X}$. Finally, if two elements $f(\cdot) \in \mathcal{H}(\mathcal{M})$ and $g(\cdot) \in \mathcal{L}(\mathcal{M})$ satisfy $\mathbf{J}[f(\cdot)] = g(\cdot)$, we have the identity

$$f(\mathbf{x}) = \mathbf{E}_{\mathbf{x}}\{g(\mathbf{y})\} \quad (4.14)$$

for every $\mathbf{x} \in \mathcal{X}$.

Proof. [3] □

If a minimum variance problem \mathcal{M} is such that the associated kernel $R_{\mathcal{M}}(\cdot, \cdot)$ is differentiable, then one can show

Theorem 4.3.3. *Consider a minimum variance problem \mathcal{M} with parameter set $\mathcal{X} \subseteq \mathbb{R}^N$ that is such that the associated kernel $R_{\mathcal{M}}(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is differentiable up to order m (see Definition 3.4.1). We denote by $J[\cdot]$ the congruence from $\mathcal{H}(\mathcal{M})$ to $\mathcal{L}(\mathcal{M})$ defined by (4.13). Then, given a matrix $\mathbf{A} \in \mathbb{R}^{N \times D}$, a vector $\mathbf{z}_c \in \mathbb{R}^D$, and a multi-index $\mathbf{p} \in \mathbb{Z}_+^D$ with $\|\mathbf{p}\|_\infty \leq m$ that satisfy $\mathbf{A}\mathcal{N}_{\mathbf{z}_c}^{\text{supp}(\mathbf{p})}(\varepsilon) \subseteq \mathcal{X}$ for a sufficiently small $\varepsilon > 0$, we have that the image $J[f(\cdot)]$ of the function*

$$f(\cdot) \triangleq \frac{\partial^{\mathbf{p}}(R_{\mathcal{M}}(\cdot, \mathbf{A}\mathbf{z}_2)h(\mathbf{z}_2))}{\partial \mathbf{z}_2^{\mathbf{p}}} \Big|_{\mathbf{z}_2=\mathbf{z}_c} \in \mathcal{H}(\mathcal{M}) \quad (4.15)$$

is given by

$$J[f(\cdot)] = \frac{\partial^{\mathbf{p}}(\rho_{\mathcal{M}}(\cdot, \mathbf{A}\mathbf{z})h(\mathbf{z}))}{\partial \mathbf{z}^{\mathbf{p}}} \Big|_{\mathbf{z}=\mathbf{z}_c}. \quad (4.16)$$

Here, the function $h(\cdot) : \mathbb{R}^D \rightarrow \mathbb{R}$ is arbitrary except that it is assumed that its partial derivatives of any order $\mathbf{p} \in \mathbb{Z}_+^D$ with $\|\mathbf{p}\|_\infty \leq m$ exist at \mathbf{z}_c .

Proof. This statement can be shown by induction on \mathbf{p} . We restrict ourselves to the proof for the case $\mathbf{p} = \mathbf{e}_k$ where $k \in [D]$. We can then write the function $f(\cdot)$ defined in (4.15) as the limit of the Cauchy sequence

$$\left\{ f_l(\cdot) \triangleq \frac{1}{l} \left[R_{\mathcal{M}}\left(\cdot, \mathbf{A}\left(\mathbf{z}_c + \frac{1}{l}\mathbf{e}_k\right)\right)h\left(\mathbf{z}_c + \frac{1}{l}\mathbf{e}_k\right) - R_{\mathcal{M}}(\cdot, \mathbf{A}\mathbf{z}_c)h(\mathbf{z}_c) \right] \right\}_{l \rightarrow \infty},$$

i.e., $f(\cdot) = \lim_{l \rightarrow \infty} f_l(\cdot)$. Since, almost trivially by definition (cf. Definition 3.1.14), any congruence is a continuous mapping (in particular, the limit operation commutes with the application of a congruence) we have that $J[f(\cdot)] = \lim_{l \rightarrow \infty} J[f_l(\cdot)]$, which by Theorem 4.3.2 yields

$$\begin{aligned} J[f(\cdot)] &= \lim_{l \rightarrow \infty} \frac{1}{l} \left[\rho_{\mathcal{M}}\left(\cdot, \mathbf{A}\left(\mathbf{z}_c + \frac{1}{l}\mathbf{e}_k\right)\right)h\left(\mathbf{z}_c + \frac{1}{l}\mathbf{e}_k\right) - \rho_{\mathcal{M}}(\cdot, \mathbf{A}\mathbf{z}_c)h(\mathbf{z}_c) \right] \\ &= \frac{\partial^{\mathbf{e}_k}(\rho_{\mathcal{M}}(\cdot, \mathbf{A}\mathbf{z})h(\mathbf{z}))}{\partial \mathbf{z}^{\mathbf{e}_k}} \Big|_{\mathbf{z}=\mathbf{z}_c}. \end{aligned} \quad (4.17)$$

□

The connection between the theory of RKHS and minimum variance estimation is summarized in [3, 35]

Theorem 4.3.4. *Given a minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$ associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$, satisfying (4.1) such that the associated RKHS $\mathcal{H}(\mathcal{M})$ exists. Then, we have that:*

- *There exists at least one allowed estimator $\hat{g}(\cdot)$, i.e., the set $\mathcal{F}(\mathcal{M})$ defined in (2.18) is nonempty, if and only if*

$$\gamma(\cdot) \in \mathcal{H}(\mathcal{M}) \quad (4.18)$$

where $\gamma(\cdot) : \mathcal{X} \rightarrow \mathbb{R} : \gamma(\mathbf{x}) \triangleq c(\mathbf{x}) + g(\mathbf{x})$ denotes the prescribed mean function of the minimum variance problem \mathcal{M} .

- *If (4.18) holds, then the minimum achievable variance $L_{\mathcal{M}}$ (see (2.19)) is given by*

$$L_{\mathcal{M}} = \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 - [\gamma(\mathbf{x}_0)]^2, \quad (4.19)$$

and the corresponding LMV estimator $\hat{g}^{(\mathbf{x}_0)}(\cdot)$ (see Definition 2.3.4) is obtained as

$$\hat{g}^{(\mathbf{x}_0)}(\cdot) = \mathbf{J}[\gamma(\cdot)], \quad (4.20)$$

where $\mathbf{J}[\cdot] : \mathcal{H}(\mathcal{M}) \rightarrow \mathcal{L}(\mathcal{M})$ denotes the congruence from $\mathcal{H}(\mathcal{M})$ to $\mathcal{L}(\mathcal{M})$ which is defined by (4.13).

Proof. [3]

□

The relation (4.19) together with the definition of the minimum achievable variance $L_{\mathcal{M}}$ (see (2.19)) gives us a powerful tool for the derivation of lower bounds on the estimator variance if the estimator's bias is known. Indeed, any lower bound on $\|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2$ induces via (4.19) and (2.19) a lower bound on the variance $v(\hat{g}(\cdot); \mathbf{x}_0)$ at \mathbf{x}_0 of any estimator $\hat{g}(\cdot)$ with bias equal to $c(\mathbf{x})$ (or equivalently with mean function equal to $\gamma(\mathbf{x}) = c(\mathbf{x}) + g(\mathbf{x})$).

As already observed in [3], a specific class of lower bounds on the variance $v(\hat{g}(\cdot); \mathbf{x}_0)$ can be readily obtained via Theorem 3.1.6 by projecting $\gamma(\cdot) \in \mathcal{H}(\mathcal{M})$ on a subspace $\mathcal{U} \subseteq \mathcal{H}(\mathcal{M})$ of the RKHS $\mathcal{H}(\mathcal{M})$. Indeed, denoting by $\mathbf{P}_{\mathcal{U}}\gamma(\cdot)$ the orthogonal projection of the function $\gamma(\cdot) \in \mathcal{H}(\mathcal{M})$ on the subspace \mathcal{U} , we have by Theorem 3.1.6 the lower bound

$$\|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 \geq \|\mathbf{P}_{\mathcal{U}}\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2. \quad (4.21)$$

4.3.3 Transforming the Parameter Function

A RKHS that is associated to a minimum variance problem has a specific property as stated in

Lemma 4.3.5. Consider a minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$ associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and with prescribed bias $c(\cdot) \equiv 0$. Then, if the parameter function $g(\cdot)$ is estimable, i.e., $g(\cdot) \in \mathcal{H}(\mathcal{M})$ (due to Theorem 4.3.4 and Definition 2.3.7), we have that any other parameter function $g'(\cdot) = g(\cdot) + a$ with $a \in \mathbb{R}$ is also estimable, i.e., $g'(\cdot) \in \mathcal{H}(\mathcal{M})$, and moreover we have that

$$\|g'(\cdot)\|_{\mathcal{H}(R)}^2 = \|g(\cdot)\|_{\mathcal{H}(R)}^2 + 2ag(\mathbf{x}_0) + a^2. \quad (4.22)$$

The minimum achievable variance remains unchanged after replacing the parameter function $g(\cdot)$ with $g'(\cdot)$, i.e.,

$$L_{\mathcal{M}} = L_{\mathcal{M}'}, \quad (4.23)$$

where \mathcal{M}' denotes the minimum variance problem that is obtained from \mathcal{M} by changing the parameter function from $g(\cdot)$ to $g'(\cdot)$.

Proof. Consider the specific function $v(\cdot) = R_{\mathcal{M}}(\cdot, \mathbf{x}_0)$ which obviously belongs to the RKHS $\mathcal{H}(\mathcal{M})$. By (4.11), we have that $v(\mathbf{x}) = 1$ for every $\mathbf{x} \in \mathcal{X}$, and in turn via the reproducing property (3.44) that $\langle v(\cdot), v(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} = \langle v(\cdot), R(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M})} = v(\mathbf{x}_0) = 1$. Thus, we can represent $g'(\cdot) = g(\cdot) + a$ as $g'(\cdot) = g(\cdot) + av(\cdot)$. The squared norm of $g'(\cdot)$ can then be developed as

$$\begin{aligned} \|g'(\cdot)\|_{\mathcal{H}(M)}^2 &= \|g(\cdot) + av(\cdot)\|_{\mathcal{H}(R)}^2 = \langle g(\cdot) + av(\cdot), g(\cdot) + av(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} \\ &= \|g(\cdot)\|_{\mathcal{H}(M)}^2 + 2a\langle g(\cdot), v(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} + a^2\langle v(\cdot), v(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} \\ &= \|g(\cdot)\|_{\mathcal{H}(M)}^2 + 2a\langle g(\cdot), R(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M})} + a^2 \underbrace{\langle v(\cdot), v(\cdot) \rangle_{\mathcal{H}(\mathcal{M})}}_{=1} \\ &\stackrel{(3.44)}{=} \|g(\cdot)\|_{\mathcal{H}(M)}^2 + 2ag(\mathbf{x}_0) + a^2. \end{aligned} \quad (4.24)$$

Based on (4.24), and using (4.19), the minimum achievable variance for \mathcal{M}' can be calculated as

$$\begin{aligned} L_{\mathcal{M}'} &= \|g'(\cdot)\|_{\mathcal{H}(\mathcal{M}')}^2 - (g'(\mathbf{x}_0))^2 \\ &= \|g'(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 - (g'(\mathbf{x}_0))^2 \\ &\stackrel{(4.24)}{=} \|g(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 + 2ag(\mathbf{x}_0) + a^2 - (g(\mathbf{x}_0) + a)^2 \\ &= \|g(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 + 2ag(\mathbf{x}_0) + a^2 - (g(\mathbf{x}_0))^2 - 2ag(\mathbf{x}_0) - a^2 \\ &= \|g(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 - (g(\mathbf{x}_0))^2 = L_{\mathcal{M}}, \end{aligned} \quad (4.25)$$

where we used the fact, that the two kernels $R_{\mathcal{M}}(\cdot, \cdot)$, $R_{\mathcal{M}'}(\cdot, \cdot)$ and therefore also the two RKHSs $\mathcal{H}(\mathcal{M})$ and $\mathcal{H}(\mathcal{M}')$ coincide. This follows from (4.10), since the parameter sets and statistical models of \mathcal{M} and \mathcal{M}' are identical. \square

We note that Lemma 4.3.5 is the RKHS-based formulation of a special case of Theorem 2.3.5.

4.3.4 Reducing the Parameter Set

The effect of a reduction of the parameter set \mathcal{X} of a minimum variance problem on its associated minimum achievable variance can be analyzed conveniently using the RKHS framework. Indeed, consider a minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$ associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and the modified minimum variance problem $\mathcal{M}' = \mathcal{M}|_{\mathcal{X}'}$ obtained for the new parameter set $\mathcal{X}' \subseteq \mathcal{X}$.

We then have that

$$L_{\mathcal{M}} \stackrel{(a)}{=} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 - (\gamma(\mathbf{x}_0))^2 \stackrel{(b)}{\geq} \|\gamma(\cdot)|_{\mathcal{X}'}\|_{\mathcal{H}(\mathcal{M}')}^2 - (\gamma(\mathbf{x}_0))^2 = L_{\mathcal{M}'}, \quad (4.26)$$

where step (a) follows from Theorem 4.3.4, step (b) is due to Theorem 3.3.5 and $\gamma(\cdot) : \mathcal{X} \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = c(\mathbf{x}) + g(\mathbf{x})$ is the prescribed mean function of \mathcal{M} . The main aspect of (4.26) (which was already mentioned in Section 2.5) is that a reduction of the parameter set \mathcal{X} can never result in a worse achievable performance, i.e., in a higher minimum achievable variance.

In Chapter 5 and Chapter 6, we will study two specific estimation problems and the associated minimum variance problems. These estimation problems are obtained from well-known estimation problems by a reduction of the original parameter set $\mathcal{X} = \mathbb{R}^N$. This reduction will be quite substantial, since the reduced parameter sets will have measure zero w.r.t. to the Lebesgue measure in \mathbb{R}^N . However, using RKHS theory, we will show that this large reduction is necessary in order to obtain strictly lower values for the minimum achievable variance (see Section 5.6 and 6.4.3). This agrees with the observation that pure or strict inequality constraints on the parameter set \mathcal{X} of an estimation problem have no influence on the Cramér–Rao bound when evaluated at a point in the interior of \mathcal{X} [29, p. 1292].

4.3.5 Classes of Minimum Variance Problems with Continuously Varying Kernels

In some applications, it is of interest to characterize not only a single minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$ associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$, but rather to characterize the global properties of a whole class of minimum variance problems $\{\mathcal{M}(\mathbf{x}_0)\}_{\mathbf{x}_0 \in \mathcal{X}}$. This class is obtained by varying the parameter vector $\mathbf{x}_0 \in \mathcal{X}$ used for the definition of the minimum variance problem. Therefore, let us consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ with a parameter set $\mathcal{X} \subseteq \mathbb{R}^N$ that is such that the kernel $R_{\mathcal{M}(\mathbf{x}_0)}(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is pointwise

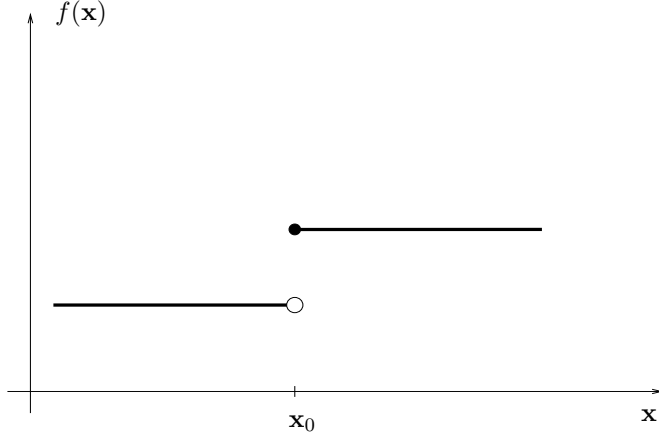


Figure 4.1: Graph of a function that is lower semi-continuous at \mathbf{x}_0 . The solid dot indicates the function value $f(\mathbf{x}_0)$.

continuous w.r.t. the parameter \mathbf{x}_0 , i.e.,

$$\lim_{\mathbf{x}'_0 \rightarrow \mathbf{x}_0} R_{\mathcal{M}(\mathbf{x}'_0)}(\mathbf{x}_1, \mathbf{x}_2) = R_{\mathcal{M}(\mathbf{x}_0)}(\mathbf{x}_1, \mathbf{x}_2), \quad \forall \mathbf{x}_0, \mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}. \quad (4.27)$$

In this case, the dependence of the minimum achievable variance $L_{\mathcal{M}(\mathbf{x}_0)}$ on $\mathbf{x}_0 \in \mathcal{X}$ is characterized by

Theorem 4.3.6. *Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ with parameter set $\mathcal{X} \subseteq \mathbb{R}^N$, and the class of minimum variance problems given by $\{\mathcal{M}(\mathbf{x}_0) \triangleq (\mathcal{E}, c(\cdot), \mathbf{x}_0)\}_{\mathbf{x}_0 \in \mathcal{X}}$ with a prescribed bias function $c(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$ that is valid for every $\mathcal{M}(\mathbf{x}_0)$ with $\mathbf{x}_0 \in \mathcal{X}$. We denote by $\gamma(\cdot) : \mathcal{X} \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = c(\mathbf{x}) + g(\mathbf{x})$ the prescribed mean function. Then, if $\gamma(\mathbf{x})$ is a continuous function of \mathbf{x} and the kernel $R_{\mathcal{M}(\mathbf{x})}$ is pointwise continuous w.r.t. \mathbf{x} , i.e., it satisfies (4.27), the minimum achievable variance $L_{\mathcal{M}(\mathbf{x})}$ exists, i.e. is finite, for every \mathbf{x} . Furthermore, viewed as a function of \mathbf{x} , $L_{\mathcal{M}(\mathbf{x})}$ is lower semi-continuous (see Figure 4.1).*

Proof. The finiteness of $L_{\mathcal{M}(\mathbf{x})}$ for every $\mathbf{x} \in \mathcal{X}$ follows trivially from the definition of a valid bias function (see Definition 2.3.6) and Theorem 4.3.4, since it is assumed that the prescribed bias function $c(\cdot)$ is valid for $\mathcal{M}(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{X}$.

Now, observe that

$$\begin{aligned} L_{\mathcal{M}(\mathbf{x})} &\stackrel{(a)}{=} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 - (\gamma(\mathbf{x}))^2 \\ &\stackrel{(b)}{=} \sup_{\substack{f(\cdot) \in \mathcal{L}(R_{\mathcal{M}}) \\ \|f(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 > 0}} \frac{(\langle \gamma(\cdot), f(\cdot) \rangle_{\mathcal{H}(\mathcal{M})})^2}{\|f(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2} - (\gamma(\mathbf{x}))^2 \end{aligned}$$

$$\begin{aligned}
& \stackrel{(c)}{=} \sup_{\substack{\mathcal{D}=\{\mathbf{x}_1, \dots, \mathbf{x}_L\} \\ L \in \mathbb{N}, \mathbf{x}_l \in \mathcal{X} \\ \mathbf{a} \in \mathcal{A}_{\mathcal{D}}}} \frac{(\langle \gamma(\cdot), \sum_{l \in [L]} a_l R(\cdot, \mathbf{x}_l) \rangle_{\mathcal{H}(\mathcal{M})})^2}{\sum_{l, l' \in [L]} a_l a_{l'} R_{\mathcal{M}(\mathbf{x})}(\mathbf{x}_l, \mathbf{x}_{l'})} - (\gamma(\mathbf{x}))^2 \\
& = \sup_{\substack{\mathcal{D}=\{\mathbf{x}_1, \dots, \mathbf{x}_L\} \\ L \in \mathbb{N}, \mathbf{x}_l \in \mathcal{X} \\ \mathbf{a} \in \mathcal{A}_{\mathcal{D}}}} \frac{(\sum_{l \in [L]} a_l \gamma(\mathbf{x}_l))^2}{\sum_{l, l' \in [L]} a_l a_{l'} R_{\mathcal{M}(\mathbf{x})}(\mathbf{x}_l, \mathbf{x}_{l'})} - (\gamma(\mathbf{x}))^2, \tag{4.28}
\end{aligned}$$

where we used $\mathcal{A}_{\mathcal{D}} \triangleq \{\mathbf{a} \in \mathbb{R}^L \mid \sum_{l, l' \in [L]} a_l a_{l'} R_{\mathcal{M}(\mathbf{x})}(\mathbf{x}_l, \mathbf{x}_{l'}) > 0\}$. The step (a) follows from Theorem 4.3.4, step (b) is due to Theorem 3.1.2 since the linear space $\mathcal{L}(R_{\mathcal{M}})$ (cf. Definition 3.2.3) is dense in the RKHS $\mathcal{H}(\mathcal{M})$ by Theorem 3.2.2. The step (c) follows from the fact that any function $f(\cdot)$ belonging to the linear space $\mathcal{L}(R_{\mathcal{M}})$ can be written as a finite linear combination $f(\cdot) = \sum_{l \in [L]} a_l R(\cdot, \mathbf{x}_l)$ with suitable coefficients $a_l \in \mathbb{R}$ and points $\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_L\} \subseteq \mathcal{X}$, which can be used in combination with the reproducing property (3.44) to express the squared norm of $f(\cdot)$ as $\|f(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 = \sum_{l, l' \in [L]} a_l a_{l'} R_{\mathcal{M}(\mathbf{x})}(\mathbf{x}_l, \mathbf{x}_{l'})$. By (4.28), we have

$$L_{\mathcal{M}(\mathbf{x})} = \sup_{\substack{\mathcal{D}=\{\mathbf{x}_1, \dots, \mathbf{x}_L\} \\ L \in \mathbb{N}, \mathbf{x}_l \in \mathcal{X} \\ \mathbf{a} \in \mathcal{A}_{\mathcal{D}}}} h_{\mathcal{D}, \mathbf{a}}(\mathbf{x}), \tag{4.29}$$

with the function $h_{\mathcal{D}, \mathbf{a}}(\mathbf{x}) : \mathcal{X} \rightarrow \mathbb{R}$ defined as

$$h_{\mathcal{D}, \mathbf{a}}(\mathbf{x}) \triangleq \frac{(\sum_{l \in [L]} a_l \gamma(\mathbf{x}_l))^2}{\sum_{l, l' \in [L]} a_l a_{l'} R_{\mathcal{M}(\mathbf{x})}(\mathbf{x}_l, \mathbf{x}_{l'})} - (\gamma(\mathbf{x}))^2, \tag{4.30}$$

where $L \triangleq |\mathcal{D}|$. For any finite set $\mathcal{D} = \{\mathbf{x}_l \in \mathcal{X}\}_{l \in [L]}$ and $\mathbf{a} \in \mathcal{A}_{\mathcal{D}}$, it follows from the continuity of $R_{\mathcal{M}(\mathbf{x})}(\cdot, \cdot)$ and $\gamma(\mathbf{x})$ w.r.t. \mathbf{x} , that the function $h_{\mathcal{D}, \mathbf{a}}(\mathbf{x})$ is continuous in a neighborhood around any point $\mathbf{x}_0 \in \mathcal{X}$. Thus, for any point $\mathbf{x}_0 \in \mathcal{X}$, there exists a radius $\delta_0 > 0$ such that the function $h_{\mathcal{D}, \mathbf{a}}(\mathbf{x})$ is continuous on $\mathcal{B}(\mathbf{x}_0, \delta_0) \cap \mathcal{X}$.

We will show by contradiction that the function $L_{\mathcal{M}(\mathbf{x})}$ given by (4.29) must be lower semicontinuous. To that end, let us assume that the minimum achievable variance $L_{\mathcal{M}(\mathbf{x})}$ is not lower-semicontinuous at a specific parameter vector $\mathbf{x}_0 \in \mathcal{X}$, i.e., we have that $\liminf_{\mathbf{x} \rightarrow \mathbf{x}_0} L_{\mathcal{M}(\mathbf{x})} \leq L_{\mathcal{M}(\mathbf{x}_0)} - \varepsilon_0$ with a specific $\varepsilon_0 > 0$. This implies that for any radius $r > 0$, there exists at least one parameter vector $\mathbf{x}' \in \mathcal{X} \cap \mathcal{B}(\mathbf{x}_0, r)$ such that $L_{\mathcal{M}(\mathbf{x}')} < L_{\mathcal{M}(\mathbf{x}_0)} - \varepsilon_0/2$, i.e.,

$$\forall r > 0 : \exists \mathbf{x}' \in \mathcal{X} \cap \mathcal{B}(\mathbf{x}_0, r) \Rightarrow L_{\mathcal{M}(\mathbf{x}')} < L_{\mathcal{M}(\mathbf{x}_0)} - \varepsilon_0/2. \tag{4.31}$$

Now, we have that due to (4.29) there must be a finite subset $\mathcal{D}_0 \subseteq \mathcal{X}$ and a vector $\mathbf{a}_0 \in \mathcal{A}_{\mathcal{D}_0}$ such that

$$h_{\mathcal{D}_0, \mathbf{a}_0}(\mathbf{x}_0) \geq L_{\mathcal{M}(\mathbf{x}_0)} - \varepsilon_0/4, \tag{4.32}$$

since otherwise $h_{\mathcal{D},\mathbf{a}}(\mathbf{x}) \leq L_{\mathcal{M}(\mathbf{x}_0)} - \varepsilon/4$ for every \mathcal{D} , \neg , which would imply via (4.29) the contradiction $L_{\mathcal{M}(\mathbf{x}_0)} \leq L_{\mathcal{M}(\mathbf{x}_0)} - \varepsilon/4 < L_{\mathcal{M}(\mathbf{x}_0)}$. Furthermore, $h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x})$ is continuous at \mathbf{x}_0 within $\mathcal{B}(\mathbf{x}_0, \delta_0) \cap \mathcal{X}$, implying the existence of a small radius $r_0 > 0$ (where $r_0 < \delta_0$) such that for any $\mathbf{x} \in \mathcal{X} \cap \mathcal{B}(\mathbf{x}_0, r_0)$ we have

$$h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x}) \geq h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x}_0) - \varepsilon_0/4. \quad (4.33)$$

By combining (4.32) and (4.33), we see that there exists a radius $r_0 > 0$ (with $r_0 < \delta_0$) such that for any $\mathbf{x} \in \mathcal{X} \cap \mathcal{B}(\mathbf{x}_0, r_0)$ we have

$$h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x}) \stackrel{(4.33)}{\geq} h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x}_0) - \varepsilon_0/4 \stackrel{(4.32)}{\geq} L_{\mathcal{M}(\mathbf{x}_0)} - \varepsilon_0/4 - \varepsilon_0/4 = L_{\mathcal{M}(\mathbf{x}_0)} - \varepsilon_0/2. \quad (4.34)$$

This lower bound on the specific function $h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x})$ implies also a lower bound on the supremum in (4.29), i.e., for those parameter vectors \mathbf{x} for which the bound in (4.34) is in force, we have simultaneously the lower bound

$$L_{\mathcal{M}(\mathbf{x})} \stackrel{(4.29)}{=} \sup_{\substack{\mathcal{D}=\{\mathbf{x}_1,\dots,\mathbf{x}_L\} \\ L \in \mathbb{N}, \mathbf{x}_l \in \mathcal{X} \\ \mathbf{a} \in \mathcal{A}_{\mathcal{D}}}} h_{\mathcal{D},\mathbf{a}}(\mathbf{x}) \geq h_{\mathcal{D}_0,\mathbf{a}_0}(\mathbf{x}) \stackrel{(4.34)}{\geq} L_{\mathcal{M}(\mathbf{x}_0)} - \varepsilon_0/2. \quad (4.35)$$

Putting together the pieces, we have that there exists a radius $r_0 > 0$ (with $r_0 < \delta_0$) such that for any $\mathbf{x} \in \mathcal{X} \cap \mathcal{B}(\mathbf{x}_0, r_0)$ we have $L_{\mathcal{M}(\mathbf{x})} \geq L_{\mathcal{M}(\mathbf{x}_0)} - \varepsilon_0/2$, i.e.,

$$\forall \mathbf{x} \in \mathcal{X} \cap \mathcal{B}(\mathbf{x}_0, r_0) \quad \Rightarrow \quad L_{\mathcal{M}(\mathbf{x})} \geq L_{\mathcal{M}(\mathbf{x}_0)} - \varepsilon_0/2. \quad (4.36)$$

Since this fact contradicts (4.31), we have shown that $L_{\mathcal{M}(\mathbf{x})}$ must be lower-semicontinuous at every point $\mathbf{x} \in \mathcal{X}$. \square

4.3.6 Sufficient Statistics from the RKHS Viewpoint

Consider a minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$ associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$. In some cases, the observation $\mathbf{y} \in \mathbb{R}^M$, whose statistic is given by the statistical model $f(\mathbf{y}; \mathbf{x})$ of the estimation problem, carries irrelevant information and can therefore be compressed in some sense. This compression or extraction of useful information (i.e., useful for the problem of estimating $g(\mathbf{x})$) can be performed by applying a (possibly randomized) mapping $\mathbf{T}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^K$ to the observation \mathbf{y} , yielding a modified observation $\mathbf{z} \in \mathbb{R}^K$, $\mathbf{z} = \mathbf{T}(\mathbf{y})$. A compression is achieved if $K < M$. We will make the weak technical assumption that the mapping $\mathbf{T}(\cdot)$ is such that the modified observation $\mathbf{z} = \mathbf{T}(\mathbf{y})$, which is a random vector, possesses a pdf [42, 43, 45]. For clarity, we will denote the evaluation of the pdf of the random vector $\mathbf{z} = \mathbf{T}(\mathbf{y})$ at the specific vector $\mathbf{z}' \in \mathbb{R}^K$ by $f_{\mathbf{z}}(\mathbf{z}'; \mathbf{x})$.

Definition 4.3.2 ([41]). A modified observation $\mathbf{z} = \mathbf{T}(\mathbf{y}) \in \mathbb{R}^K$ is called a sufficient statistic for the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ if the conditional distribution $f(\mathbf{y}|\mathbf{z}; \mathbf{x})$ of \mathbf{y} given \mathbf{z} does not depend on \mathbf{x} , i.e., if

$$f(\mathbf{y}|\mathbf{z}; \mathbf{x}) = f(\mathbf{y}|\mathbf{z}), \quad (4.37)$$

with a suitable (conditional) pdf $f(\mathbf{y}|\mathbf{z})$.

The verification if a given modified observation is a sufficient statistic can be based on

Theorem 4.3.7. Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and a modified observation $\mathbf{z} = \mathbf{T}(\mathbf{y})$ for \mathcal{E} for which there exists a pdf. The modified observation \mathbf{z} is a sufficient statistic for \mathcal{E} if and only if we can factor the pdf $f(\mathbf{y}; \mathbf{x})$ of the observation \mathbf{y} as

$$f(\mathbf{y}; \mathbf{x}) = f_{\mathbf{z}}(\mathbf{T}(\mathbf{y}); \mathbf{x})h(\mathbf{y}), \quad (4.38)$$

where $h(\cdot)$ is a nonnegative function which does not depend on \mathbf{x} and $f_{\mathbf{z}}(\cdot; \mathbf{x})$ denotes the pdf of the random vector $\mathbf{z} = \mathbf{T}(\mathbf{y})$.

Proof. [20, 41, 69] □

We note that Theorem 4.3.7 is a variant of a famous result in classical estimation theory known as the ‘‘Neyman-Fisher factorization theorem’’ [20]. However, the Neyman-Fisher factorization theorem is more general than Theorem 4.3.7 since it does not require the existence of a well-defined pdf of the modified observation $\mathbf{z} = \mathbf{T}(\mathbf{y})$.

Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and a sufficient statistic $\mathbf{z} = \mathbf{T}(\mathbf{y})$, i.e., we can factor the statistical model $f(\mathbf{y}; \mathbf{x})$ as in (4.38). We can then define a new estimation problem \mathcal{E}' by using the sufficient statistic as the observation, i.e., we define $\mathcal{E}' \triangleq (\mathcal{X}, f(\mathbf{z}; \mathbf{x}), g(\cdot))$. If we consider the two minimum variance problems \mathcal{M} , \mathcal{M}' that are obtained from \mathcal{E} and \mathcal{E}' and a common prescribed bias function $c(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$ and parameter vector $\mathbf{x}_0 \in \mathcal{X}$, we have that the associated RKHSs are identical:

Theorem 4.3.8. Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ with a sufficient statistic $\mathbf{z} = \mathbf{T}(\mathbf{y})$ and the modified estimation problem $\mathcal{E}' = (\mathcal{X}, f(\mathbf{z}; \mathbf{x}), g(\cdot))$ that is obtained from \mathcal{E} by using the sufficient statistic \mathbf{z} as the observation. Then we have for any fixed parameter vector $\mathbf{x}_0 \in \mathcal{X}$ and any prescribed bias function $c(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$, that the RKHSs associated to the two minimum variance problems $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$ and $\mathcal{M}' = (\mathcal{E}', c(\cdot), \mathbf{x}_0)$ are identical, i.e.,

$$\mathcal{H}(\mathcal{M}) = \mathcal{H}(\mathcal{M}'). \quad (4.39)$$

In particular, this implies that the minimum achievable variances of \mathcal{M} and \mathcal{M}' are equal, i.e.,

$$L_{\mathcal{M}} = L_{\mathcal{M}'}. \quad (4.40)$$

Proof. The statement is proven by showing that the kernels $R_{\mathcal{M}}(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ and $R_{\mathcal{M}'}(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ associated with \mathcal{M} and \mathcal{M}' , respectively, are identical. Indeed, by denoting the value of the joint pdf and the conditional pdf of the two random vectors \mathbf{y} and \mathbf{z} evaluated at $\mathbf{z} = \mathbf{z}'$ by $f(\mathbf{y}, \mathbf{z}'; \mathbf{x})$ and $f(\mathbf{y}|\mathbf{z}'; \mathbf{x})$ respectively, we have by (4.10) that

$$\begin{aligned} R_{\mathcal{M}}(\mathbf{x}_1, \mathbf{x}_2) &= \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{f(\mathbf{y}; \mathbf{x}_1) f(\mathbf{y}; \mathbf{x}_2)}{f(\mathbf{y}; \mathbf{x}_0) f(\mathbf{y}; \mathbf{x}_0)} \right\} \stackrel{(a)}{=} \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{f_{\mathbf{z}}(\mathbf{T}(\mathbf{y}); \mathbf{x}_1) f_{\mathbf{z}}(\mathbf{T}(\mathbf{y}); \mathbf{x}_2)}{f_{\mathbf{z}}(\mathbf{T}(\mathbf{y}); \mathbf{x}_0) f_{\mathbf{z}}(\mathbf{T}(\mathbf{y}); \mathbf{x}_0)} \right\} \\ &\stackrel{(b)}{=} \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{f_{\mathbf{z}}(\mathbf{z}; \mathbf{x}_1) f_{\mathbf{z}}(\mathbf{z}; \mathbf{x}_2)}{f_{\mathbf{z}}(\mathbf{z}; \mathbf{x}_0) f_{\mathbf{z}}(\mathbf{z}; \mathbf{x}_0)} \right\} = R_{\mathcal{M}'}(\mathbf{x}_1, \mathbf{x}_2). \end{aligned} \quad (4.41)$$

Here, the step (a) follows from the factorization (4.38) and (b) follows from a fundamental property of the expectation operation² [42, 43, 45]. Since $R_{\mathcal{M}}(\cdot, \cdot) = R_{\mathcal{M}'}(\cdot, \cdot)$, we have by Theorem 3.2.3 that $\mathcal{H}(\mathcal{M}) = \mathcal{H}(\mathcal{M}')$. Finally, the relation (4.40) follows then by (4.19) of Theorem 4.3.4. \square

We already mentioned in Theorem 2.4.1 of Section 2.4.1 that any transformation of the observation, by an invertible map is irrelevant to minimum variance estimation. Indeed, as can be verified easily by Theorem 4.3.7, any modified observation that is obtained by an invertible map $\mathbf{T}(\cdot)$ is trivially a sufficient statistic, and hence Theorem 4.3.8 applies.

Furthermore, we note that Theorem 4.3.8 agrees with a famous result within classical (non-Bayesian) estimation theory, that is known as the ‘‘Rao-Blackwell-Lehmann-Scheffé’’ theorem [20]. This theorem states that given an estimator with a specific bias function, and given a sufficient statistic, one can always find another estimator, depending on the observation only via the sufficient statistic, with the same bias function and a variance that does not exceed the variance of the original estimator. This fact is consistent with Theorem 4.3.8, in particular with (4.40).

4.4 RKHS Interpretation of Known Variance Bounds

In this section, we will interpret some well-known lower bounds on the estimator variance either as a specific evaluation of the squared norm $\|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2$ in (4.19) or as specific lower bounds on $\|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2$.

Throughout this section, we will consider only minimum variance problems with prescribed bias $c(\cdot) \equiv 0$, i.e, we consider unbiased estimation, which means that $\gamma(\cdot) = g(\cdot)$. However, by Section 2.3.3, in particular by Theorem 2.3.1, this is no real restriction since we do not constrain a priori the parameter functions $g(\cdot)$ of the considered estimation problems.

²For a random vector $\mathbf{y} \in \mathbb{R}^M$ and a function $\mathbf{t}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^K$, the expectation $\mathbb{E}_{\mathbf{x}_0}\{\mathbf{t}(\mathbf{y})\}$ is equal to the expectation $\mathbb{E}_{\mathbf{x}_0}\{\mathbf{b}\}$, where $\mathbf{b} \in \mathbb{R}^K$ is the random vector whose realization \mathbf{b}' is obtained from \mathbf{y}' , the realization of the random vector \mathbf{y} , by $\mathbf{b}' = \mathbf{t}(\mathbf{y}')$.

4.4.1 Barankin Bound

The term ‘‘Barankin bound’’ refers to the following result [55, 70]:

Theorem 4.4.1 (Barankin bound). *Consider a minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot), \mathbf{x}_0)$ associated with the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$, where $c(\mathbf{x}) \equiv 0$ and for which Postulate 4.1.1 is fulfilled. Then, for any estimator $\hat{g}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}$ which is unbiased and has finite variance at \mathbf{x}_0 , we have that its variance at \mathbf{x}_0 is lower bounded by*

$$v(\hat{g}(\cdot); \mathbf{x}_0) \geq \sup_{\substack{\mathcal{D}=\{\mathbf{x}_1, \dots, \mathbf{x}_L\} \\ L \in \mathbb{N}, \mathbf{x}_l \in \mathcal{X} \\ \mathbf{a} \in \mathcal{B}_{\mathcal{D}}}} \frac{\left| \sum_{l \in [L]} a_l h(\mathbf{x}_l) \right|^2}{\mathbb{E}_{\mathbf{x}_0} \left\{ \left(\sum_{l \in [L]} a_l \rho_{\mathcal{M}}(\mathbf{y}, \mathbf{x}_l) \right)^2 \right\}}, \quad (4.42)$$

where we used

$$h(\cdot) : \mathcal{X} \rightarrow \mathbb{R} : h(\mathbf{x}) \triangleq g(\mathbf{x}) - g(\mathbf{x}_0), \quad (4.43)$$

$\rho_{\mathcal{M}}(\mathbf{y}, \mathbf{x}_l)$ as defined in (4.6), and $\mathcal{B}_{\mathcal{D}} \triangleq \{\mathbf{a} \in \mathbb{R}^L \mid \mathbb{E}_{\mathbf{x}_0} \{ (\sum_{l \in [L]} a_l \rho_{\mathcal{M}}(\mathbf{y}, \mathbf{x}_l))^2 \} > 0\}$. Moreover, if there exists at least one unbiased estimator with a finite variance at \mathbf{x}_0 , we have that the supremum in (4.42) is equal to the minimum achievable variance at \mathbf{x}_0 , i.e.,

$$L_{\mathcal{M}} = \sup_{\substack{\mathcal{D}=\{\mathbf{x}_1, \dots, \mathbf{x}_L\} \\ L \in \mathbb{N}, \mathbf{x}_l \in \mathcal{X} \\ \mathbf{a} \in \mathcal{B}_{\mathcal{D}}}} \frac{\left| \sum_{l \in [L]} a_l h(\mathbf{x}_l) \right|^2}{\mathbb{E}_{\mathbf{x}_0} \left\{ \left(\sum_{l \in [L]} a_l \rho_{\mathcal{M}}(\mathbf{y}, \mathbf{x}_l) \right)^2 \right\}}, \quad (4.44)$$

which is then called the Barankin bound for the minimum variance problem \mathcal{M} .

We note that the RKHS interpretation of the Barankin bound has been already discussed in [71]. However, the following derivation will be self-contained and more detailed than those presented in [71].

First, we have that by Theorem 4.3.4, the existence of at least one unbiased estimator with finite variance at \mathbf{x}_0 , i.e., the existence of at least one allowed estimator, is equivalent to the fact that the function $g(\cdot)$ (which coincides with the prescribed mean function $\gamma(\mathbf{x}) = g(\mathbf{x}) + c(\mathbf{x})$ for \mathcal{M}) belongs to the RKHS $\mathcal{H}(\mathcal{M})$, i.e., $g(\cdot) \in \mathcal{H}(\mathcal{M})$. However, since $h(\cdot) = g(\cdot) + a$ with $a = -g(\mathbf{x}_0)$, we have by Lemma 4.3.5, that $g(\cdot) \in \mathcal{H}(\mathcal{M})$ if and only if $h(\cdot) \in \mathcal{H}(\mathcal{M})$. Thus, there exists at least one unbiased estimator with finite variance at \mathbf{x}_0 if and only if $h(\cdot) \in \mathcal{H}(\mathcal{M})$. In what follows we will verify (4.44) in the case $h(\cdot) \in \mathcal{H}(\mathcal{M})$, which in turn shows that the supremum in (4.42) is finite if $h(\cdot) \in \mathcal{H}(\mathcal{M})$ (due to Theorem 4.3.4), i.e., if there exists at least one unbiased estimator with finite variance at \mathbf{x}_0 . The sufficiency of the condition $h(\cdot) \in \mathcal{H}(\mathcal{M})$ for the existence

of a finite supremum in (4.44) is also mentioned in [66, Lemma 4], where it is moreover shown that the condition $h(\cdot) \in \mathcal{H}(\mathcal{M})$ is also necessary for the existence of a finite supremum in (4.44).

Let us now show that for the case $h(\cdot) \in \mathcal{H}(\mathcal{M})$ (which is equivalent to $g(\cdot) \in \mathcal{H}(\mathcal{M})$), the relation (4.44) of Theorem 4.4.1 is just a reformulation of (4.19) in Theorem 4.3.4 which is based on Theorem 3.1.2. First, we note that by Theorem 4.3.4 and Lemma 4.3.5, we have that

$$\begin{aligned} L_{\mathcal{M}} &= \|g(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 - (g(\mathbf{x}_0))^2 \stackrel{(4.43)}{=} \|h(\cdot) + g(\mathbf{x}_0)\|_{\mathcal{H}(\mathcal{M})}^2 - (g(\mathbf{x}_0))^2 \\ &\stackrel{(4.22)}{=} \|h(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 + 2g(\mathbf{x}_0) \underbrace{h(\mathbf{x}_0)}_{=0} + (g(\mathbf{x}_0))^2 - (g(\mathbf{x}_0))^2 = \|h(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2. \end{aligned} \quad (4.45)$$

By a similar derivation as used for (4.28), we obtain

$$\begin{aligned} \|h(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 &= \sup_{\substack{f(\cdot) \in \mathcal{L}(R_{\mathcal{M}}) \\ \|f(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 > 0}} \frac{(\langle h(\cdot), f(\cdot) \rangle_{\mathcal{H}(\mathcal{M})})^2}{\|f(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2} \\ &= \sup_{\substack{\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_L\} \\ L \in \mathbb{N}, \mathbf{x}_l \in \mathcal{X} \\ \mathbf{a} \in \mathcal{A}_{\mathcal{D}}}} \frac{(\langle h(\cdot), \sum_{l \in [L]} a_l R(\cdot, \mathbf{x}_l) \rangle_{\mathcal{H}(\mathcal{M})})^2}{\sum_{l, l' \in [L]} a_l a_{l'} R_{\mathcal{M}(\mathbf{x})}(\mathbf{x}_l, \mathbf{x}_{l'})} - (\gamma(\mathbf{x}))^2 \\ &= \sup_{\substack{\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_L\} \\ L \in \mathbb{N}, \mathbf{x}_l \in \mathcal{X} \\ \mathbf{a} \in \mathcal{A}_{\mathcal{D}}}} \frac{(\sum_{l \in [L]} a_l h(\mathbf{x}_l))^2}{\sum_{l, l' \in [L]} a_l a_{l'} R_{\mathcal{M}(\mathbf{x})}(\mathbf{x}_l, \mathbf{x}_{l'})}, \end{aligned} \quad (4.46)$$

where we used $\mathcal{A}_{\mathcal{D}} \triangleq \{\mathbf{a} \in \mathbb{R}^L \mid \sum_{l, l' \in [L]} a_l a_{l'} R_{\mathcal{M}(\mathbf{x})}(\mathbf{x}_l, \mathbf{x}_{l'}) > 0\}$. Combining (4.46) and (4.45) yields

$$L_{\mathcal{M}} = \sup_{\substack{\mathcal{D} = \{\mathbf{x}_1, \dots, \mathbf{x}_L\} \\ L \in \mathbb{N}, \mathbf{x}_l \in \mathcal{X} \\ \mathbf{a} \in \mathcal{A}_{\mathcal{D}}}} \frac{|\sum_{l \in [L]} a_l h(\mathbf{x}_l)|^2}{\sum_{l, l' \in [L]} a_l a_{l'} R_{\mathcal{M}(\mathbf{x})}(\mathbf{x}_l, \mathbf{x}_{l'})}. \quad (4.47)$$

But this is equal to (4.44) since for any finite $\mathcal{D} \subseteq \mathcal{X}$, of size $|\mathcal{D}| = L$, and arbitrary coefficient vector $\mathbf{a} \in \mathbb{R}^L$, we straightforwardly have that

$$\sum_{k, k' \in [L]} a_k a_{k'} R_{\mathcal{M}}(\mathbf{x}_k, \mathbf{x}'_{k'}) = \mathbb{E}_{\mathbf{x}_0} \left\{ \left(\sum_{k \in [L]} a_k \rho_{\mathcal{M}}(\mathbf{y}, \mathbf{x}_k) \right)^2 \right\}. \quad (4.48)$$

This also implies that $\mathcal{A}_{\mathcal{D}} = \mathcal{B}_{\mathcal{D}}$. The lower bound in (4.42) follows then directly from (4.44) and the definition of $L_{\mathcal{M}}$ in (2.19).

4.4.2 Cramér–Rao bound

The Cramér–Rao bound (CRB) [20, 27, 72] is probably the best known lower bound on the variance of estimators for a given estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ which satisfies some weak regularity conditions as discussed presently. Since the CRB applies to any unbiased estimator (the CRB applies also for biased estimators since it depends solely on the mean function of a given estimator) for \mathcal{E} , the CRB also yields a lower bound on the minimum achievable variance $L_{\mathcal{M}}$ for the minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot) \equiv 0, \mathbf{x}_0)$ associated with the estimation problem \mathcal{E} . We will consider the CRB for two different settings which are defined by the structure of the parameter set \mathcal{X} associated with the estimation problem \mathcal{E} . These two instances of the CRB are termed the “unconstrained CRB” and the “constrained CRB”, respectively.

The validity of the CRB requires some weak technical conditions to be placed on the estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and associated minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot) \equiv 0, \mathbf{x}_0)$, which are summarized in

Postulate 4.4.2. *The estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and associated minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot) \equiv 0, \mathbf{x}_0)$ are such that:*

- *The parameter set $\mathcal{X} \subseteq \mathbb{R}^N$ contains an N -dimensional open ball $\mathcal{B}(\mathbf{x}_0, r)$ centered at $\mathbf{x}_0 \in \mathcal{X}$ and with a positive radius $r > 0$, i.e.,*

$$\mathcal{B}(\mathbf{x}_0, r) \subseteq \mathcal{X}. \quad (4.49)$$

- *For every $\mathbf{x} \in \mathcal{X}$, the partial derivatives $\frac{\partial^{\mathbf{p}} f(\mathbf{y}; \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}}$ exist and moreover*

$$\mathbb{E}_{\mathbf{x}_0} \left\{ \left(\frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}} f(\mathbf{y}; \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \right)^2 \right\} < \infty \quad (4.50)$$

for every multi-index $\mathbf{p} \in \mathbb{Z}_+^N$ with $\|\mathbf{p}\|_{\infty} \leq m$, where $m \in \mathbb{N}$ is called the maximum differentiation order.

- *For any function $h(\mathbf{y}) : \mathbb{R}^M \rightarrow \mathbb{R}$ such that $\mathbb{E}_{\mathbf{x}}\{h(\mathbf{y})\}$ exists³ we have that for all $\mathbf{x} \in \mathcal{B}(\mathbf{x}_0, r)$ the expectation operation commutes with partial differentiation, i.e.,*

$$\frac{\partial^{\mathbf{p}} \int_{\mathbf{y}} h(\mathbf{y}) f(\mathbf{y}; \mathbf{x}) d\mathbf{y}}{\partial \mathbf{x}^{\mathbf{p}}} = \int_{\mathbf{y}} h(\mathbf{y}) \frac{\partial^{\mathbf{p}} f(\mathbf{y}; \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} d\mathbf{y} \quad (4.51)$$

or equivalently, expressed in terms of expectations, we have

$$\frac{\partial^{\mathbf{p}} \mathbb{E}_{\mathbf{x}}\{h(\mathbf{y})\}}{\partial \mathbf{x}^{\mathbf{p}}} = \mathbb{E}_{\mathbf{x}} \left\{ h(\mathbf{y}) \frac{1}{f(\mathbf{y}; \mathbf{x})} \frac{\partial^{\mathbf{p}} f(\mathbf{y}; \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \right\} \quad (4.52)$$

for every multi-index $\mathbf{p} \in \mathbb{Z}_+^N$ with $\|\mathbf{p}\|_{\infty} \leq m$ provided that the right hand side of (4.51) and (4.52) is finite.

³Note that this implies that $\mathbb{E}_{\mathbf{x}}\{|h(\mathbf{y})|\} < \infty$ [42, 43, 73].

- The parameter function $g(\cdot) : \mathcal{X} \rightarrow \mathbb{R}$ is such that the partial derivatives $\frac{\partial^{\mathbf{p}} g(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}}|_{\mathbf{x}=\mathbf{x}_0}$ exist at \mathbf{x}_0 for every multi-index $\mathbf{p} \in \mathbb{Z}_+^N$ with $\|\mathbf{p}\|_\infty \leq m$.
- The expectation

$$\mathbb{E}_{\mathbf{x}_0} \left\{ \frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_1} f(\mathbf{y}; \mathbf{x}_1)}{\partial \mathbf{x}_1^{\mathbf{p}_1}} \frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_2} f(\mathbf{y}; \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}_2}} \right\} \quad (4.53)$$

depends continuously on the parameter vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \cap \mathcal{B}(\mathbf{x}_0, r)$ for every pair of multi-indices $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{Z}_+^N$ with $\|\mathbf{p}_1\|_\infty, \|\mathbf{p}_2\|_\infty \leq m$.

It can be shown that Postulate 4.4.2 is satisfied in particular for estimation problems where the statistical model is given by an exponential family (see Section 2.6).

In the following, we will need

Theorem 4.4.3. Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and associated minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot) \equiv 0, \mathbf{x}_0)$ for which Postulate 4.1.1 and Postulate 4.4.2 with maximum differentiation order m , and radius r hold true. Then, the kernel $R_{\mathcal{M}}(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is differentiable up to order m .

Proof. First, note that the condition (4.50) implies that

$$\mathbb{E}_{\mathbf{x}_0} \left\{ \frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}} f(\mathbf{y}; \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \right\} \leq \sqrt{\mathbb{E}_{\mathbf{x}_0} \left\{ \left(\frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}} f(\mathbf{y}; \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \right)^2 \right\}} < \infty \quad (4.54)$$

due to the Cauchy-Schwarz inequality (cf. Theorem 3.1.1) for the Hilbert space $\mathcal{P}(\mathcal{M})$ (see (4.2) and Theorem 4.2.1) with inner product $\langle \cdot, \cdot \rangle_{\text{RV}}$ and the two specific estimators $\hat{g}_1(\mathbf{y}) = \frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}} f(\mathbf{y}; \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}}$ and $\hat{g}_2(\mathbf{y}) = 1$ where obviously $\langle \hat{g}_2(\cdot), \hat{g}_2(\cdot) \rangle_{\text{RV}} = 1$.

Then, for any any two multi-indices $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{Z}_+^N$ with $\|\mathbf{p}_1\|_\infty, \|\mathbf{p}_2\|_\infty \leq m$ and parameter vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \cap \mathcal{B}(\mathbf{x}_0, r)$ we have

$$\begin{aligned} & \infty > \sqrt{\mathbb{E}_{\mathbf{x}_0} \left\{ \left(\frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_1} f(\mathbf{y}; \mathbf{x}_1)}{\partial \mathbf{x}_1^{\mathbf{p}_1}} \right)^2 \right\} \mathbb{E}_{\mathbf{x}_0} \left\{ \left(\frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_2} f(\mathbf{y}; \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}_2}} \right)^2 \right\}} \\ & \stackrel{(b)}{\geq} \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_1} f(\mathbf{y}; \mathbf{x}_1)}{\partial \mathbf{x}_1^{\mathbf{p}_1}} \frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_2} f(\mathbf{y}; \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}_2}} \right\} \\ & = \int_{\mathbf{y}} \frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_1} f(\mathbf{y}; \mathbf{x}_1)}{\partial \mathbf{x}_1^{\mathbf{p}_1}} \frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_2} f(\mathbf{y}; \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}_2}} f(\mathbf{y}; \mathbf{x}_0) d\mathbf{y} \\ & = \int_{\mathbf{y}} \frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_1} f(\mathbf{y}; \mathbf{x}_1)}{\partial \mathbf{x}_1^{\mathbf{p}_1}} \frac{\partial^{\mathbf{p}_2} f(\mathbf{y}; \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}_2}} d\mathbf{y} \end{aligned}$$

$$\begin{aligned}
& \stackrel{(c)}{=} \frac{\partial^{\mathbf{p}_2} \int_{\mathbf{y}} \frac{f(\mathbf{y}; \mathbf{x}_2)}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_1} f(\mathbf{y}; \mathbf{x}_1)}{\partial \mathbf{x}_1^{\mathbf{p}_1}} d\mathbf{y}}{\partial \mathbf{x}_2^{\mathbf{p}_2}} \\
& \stackrel{(d)}{=} \frac{\partial^{\mathbf{p}_2} \partial^{\mathbf{p}_1} \int_{\mathbf{y}} \frac{f(\mathbf{y}; \mathbf{x}_1) f(\mathbf{y}; \mathbf{x}_2)}{f(\mathbf{y}; \mathbf{x}_0)} d\mathbf{y}}{\partial \mathbf{x}_1^{\mathbf{p}_1} \partial \mathbf{x}_2^{\mathbf{p}_2}} = \frac{\partial^{\mathbf{p}_1} \partial^{\mathbf{p}_2} R_{\mathcal{M}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{x}_1^{\mathbf{p}_1} \partial \mathbf{x}_2^{\mathbf{p}_2}}, \tag{4.55}
\end{aligned}$$

where (a) is due to (4.50), step (b) is due to the Cauchy-Schwarz inequality (cf. Theorem 3.1.1) for the Hilbert space $\mathcal{P}(\mathcal{M})$ (see (4.2) and Theorem 4.2.1), step (c) follows from (4.52) with $h(\mathbf{y}) = \frac{1}{f(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_1} f(\mathbf{y}; \mathbf{x}_1)}{\partial \mathbf{x}_1^{\mathbf{p}_1}}$ and (d) follows by another application of (4.52) for the choice $h(\mathbf{y}) = \frac{f(\mathbf{y}; \mathbf{x}_2)}{f(\mathbf{y}; \mathbf{x}_0)}$. The application of (4.52) is allowed in step (c) and (d) since a finite value of the right hand side of (4.51) and (4.52) is guaranteed by (4.54) and (4.1), respectively. Thus, we have shown that the partial derivative $\frac{\partial^{\mathbf{p}_1} \partial^{\mathbf{p}_2} R_{\mathcal{M}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{x}_1^{\mathbf{p}_1} \partial \mathbf{x}_2^{\mathbf{p}_2}}$ exists for $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{Z}_+^N$ with $\|\mathbf{p}_1\|_\infty, \|\mathbf{p}_2\|_\infty \leq m$ and parameter vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X} \cap \mathcal{B}(\mathbf{x}_0, r)$. Since moreover, these partial derivatives are continuous functions of \mathbf{x}_1 and \mathbf{x}_2 , due the continuity of the function (4.53) in Postulate 4.4.2, we have that the kernel $R_{\mathcal{M}}(\cdot, \cdot) : \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ is differentiable up to order m . \square

Unconstrained CRB

The following form of the unconstrained CRB has been presented in [56, 74]:

Theorem 4.4.4. *Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and associated minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot) \equiv 0, \mathbf{x}_0)$ that satisfy the conditions of Postulate 4.1.1 and Postulate 4.4.2 with $m = 1$. Then the variance $v(\hat{g}(\cdot); \mathbf{x}_0)$ at \mathbf{x}_0 of any unbiased estimator $\hat{g}(\cdot)$ with finite variance at \mathbf{x}_0 is lower bounded by*

$$v(\hat{g}(\cdot); \mathbf{x}_0) \geq \left(\left. \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} \right)^T \mathbf{J}_{\mathbf{x}_0}^\dagger \left. \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} \tag{4.56}$$

where $\mathbf{J}_{\mathbf{x}_0} \in \mathbb{R}^{N \times N}$, known as the Fisher information matrix associated to the estimation problem \mathcal{E} , is given elementwise by

$$(\mathbf{J}_{\mathbf{x}_0})_{k,l} \triangleq \mathbb{E}_{\mathbf{x}_0} \left\{ \left. \frac{\partial \log f(\mathbf{y}; \mathbf{x})}{\partial x_k} \right|_{\mathbf{x}=\mathbf{x}_0} \left. \frac{\partial \log f(\mathbf{y}; \mathbf{x})}{\partial x_l} \right|_{\mathbf{x}=\mathbf{x}_0} \right\}. \tag{4.57}$$

Proof. See [20, 56]. \square

We now provide an RKHS-based derivation of Theorem 4.4.4 by showing that under the validity of Postulate 4.4.2, the bound in (4.56) can be obtained via an orthogonal projection on a specific subspace of the RKHS $\mathcal{H}(\mathcal{M})$. To this end, we define the subspace $\mathcal{U}_{\text{CR}} \subseteq \mathcal{H}(\mathcal{M})$ as

$$\mathcal{U}_{\text{CR}} \triangleq \text{span} \{ \{v_0(\cdot)\} \cup \{v_l(\cdot)\}_{l \in [N]} \}, \tag{4.58}$$

with the functions $v_0(\cdot) \triangleq R_{\mathcal{M}}(\cdot, \mathbf{x}_0) \in \mathcal{H}(\mathcal{M})$ and

$$v_l(\cdot) \triangleq \left. \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{x})}{\partial x_l} \right|_{\mathbf{x}=\mathbf{x}_0} \in \mathcal{H}(\mathcal{M}), \quad l \in [N]. \quad (4.59)$$

By Theorem 4.4.3, the kernel $R_{\mathcal{M}}(\cdot, \cdot)$ is differentiable up to order $m = 1$. Therefore, the functions $\{v_l(\cdot)\}_{l \in [N]}$ belong to $\mathcal{H}(\mathcal{M})$ due to Theorem 3.4.1. In particular, the definition of the functions $v_l(\cdot)$ is of the same form as those in (3.64). The inner product between any $v_l(\cdot)$ with $l \in [N]$ and $v_0(\cdot)$ is given by

$$\begin{aligned} \langle v_0(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} &\stackrel{(a)}{=} \langle R_{\mathcal{M}}(\cdot, \mathbf{x}_0), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} = v_l(\mathbf{x}_0) \\ &\stackrel{(b)}{=} \left. \frac{\partial R_{\mathcal{M}}(\mathbf{x}_0, \mathbf{x})}{\partial x_l} \right|_{\mathbf{x}=\mathbf{x}_0} \stackrel{(c)}{=} \left. \frac{\partial}{\partial x_l} \int_{\mathbf{y}} \frac{f(\mathbf{y}; \mathbf{x}) f(\mathbf{y}; \mathbf{x}_0)}{f(\mathbf{y}; \mathbf{x}_0)} d\mathbf{y} \right|_{\mathbf{x}=\mathbf{x}_0} \\ &= \left. \frac{\partial}{\partial x_l} \int_{\mathbf{y}} f(\mathbf{y}; \mathbf{x}) d\mathbf{y} \right|_{\mathbf{x}=\mathbf{x}_0} = \left. \frac{\partial}{\partial x_l} 1 \right|_{\mathbf{x}=\mathbf{x}_0} = 0. \end{aligned} \quad (4.60)$$

Here, we used for the step (a) the reproducing property (3.44), for (b) the definition of the functions $v_l(\cdot)$ and for (c) the definition of the kernel $R_{\mathcal{M}}(\cdot, \cdot)$ (4.10).

Now we consider the projection $\mathbf{P}_{\mathcal{U}_{\text{CR}}} g(\cdot)$ of the parameter function $g(\cdot)$ onto the subspace \mathcal{U}_{CR} . Since Theorem 4.4.4 makes a statement about an unbiased estimator of $g(\cdot)$ with a finite variance at \mathbf{x}_0 , we can assume that $g(\cdot)$ is estimable, i.e., we have $g(\cdot) \in \mathcal{H}(\mathcal{M})$ by Theorem 4.3.4. (Note that $\gamma(\cdot) = g(\cdot)$ in Theorem 4.3.4.) Due to (4.60), we have that subset \mathcal{U}_{CR} (cf. (4.58)) is spanned by the union of the two mutually orthogonal sets $\{v_0(\cdot)\}$ and $\{v_l(\cdot)\}_{l \in [N]}$. Therefore, we can apply Theorem 3.1.9 to obtain the squared norm of the projection $\mathbf{P}_{\mathcal{U}_{\text{CR}}} g(\cdot)$ as

$$\begin{aligned} \|\mathbf{P}_{\mathcal{U}_{\text{CR}}} g(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 &= \langle g(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} (\langle v_0(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M})})^{-1} \langle g(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} + \mathbf{a}^T \mathbf{V}^\dagger \mathbf{a} \\ &= (g(\mathbf{x}_0))^2 + \mathbf{a}^T \mathbf{V}^\dagger \mathbf{a}, \end{aligned} \quad (4.61)$$

where, according to the reproducing property (3.44), we used $\langle g(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} = g(\mathbf{x}_0)$ and

$$\langle v_0(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} = \langle R_{\mathcal{M}}(\cdot, \mathbf{x}_0), R_{\mathcal{M}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M})} = R(\mathbf{x}_0, \mathbf{x}_0) \stackrel{(4.11)}{=} 1. \quad (4.62)$$

The vector $\mathbf{a} \in \mathbb{R}^N$ and the matrix $\mathbf{V} \in \mathbb{R}^{N \times N}$ in (4.61) are defined elementwise as $a_l = \langle g(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})}$ and $(\mathbf{V})_{k,l} = \langle v_k(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})}$, respectively. Note that due to Theorem 3.4.1 (using $g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot) = v_l(\cdot)$ for $\mathbf{p} = \mathbf{e}_l$, $\mathbf{x}_c = \mathbf{x}_0$ and $f(\cdot) = g(\cdot)$),

$$a_l = \langle v_l(\cdot), g(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} = \left. \frac{\partial g(\mathbf{x})}{\partial x_l} \right|_{\mathbf{x}=\mathbf{x}_0}, \quad l \in [N]. \quad (4.63)$$

Furthermore,

$$\begin{aligned}
(\mathbf{V})_{k,l} &= \langle v_k(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} \\
&\stackrel{(a)}{=} \left. \frac{\partial^2 R_{\mathcal{M}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial x_{1,k} \partial x_{2,l}} \right|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}_0} = \left. \frac{\partial^2}{\partial x_{1,k} \partial x_{2,l}} \int_{\mathbf{y}} \frac{f(\mathbf{y}; \mathbf{x}_1) f(\mathbf{y}; \mathbf{x}_2)}{f(\mathbf{y}; \mathbf{x}_0)} d\mathbf{y} \right|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}_0} \\
&\stackrel{(b)}{=} \int_{\mathbf{y}} \left. \frac{\partial^2}{\partial x_{1,k} \partial x_{2,l}} \frac{f(\mathbf{y}; \mathbf{x}_1) f(\mathbf{y}; \mathbf{x}_2)}{f(\mathbf{y}; \mathbf{x}_0)} \right|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}_0} d\mathbf{y} \\
&= \int_{\mathbf{y}} \left. \frac{\partial}{\partial x_{1,k}} \frac{f(\mathbf{y}; \mathbf{x}_1)}{f(\mathbf{y}; \mathbf{x}_0)} \right|_{\mathbf{x}_1=\mathbf{x}_0} \left. \frac{\partial}{\partial x_{2,l}} \frac{f(\mathbf{y}; \mathbf{x}_2)}{f(\mathbf{y}; \mathbf{x}_0)} \right|_{\mathbf{x}_2=\mathbf{x}_0} f(\mathbf{y}; \mathbf{x}_0) d\mathbf{y} \\
&= \int_{\mathbf{y}} \left. \frac{\partial}{\partial x_{1,k}} \log f(\mathbf{y}; \mathbf{x}_1) \right|_{\mathbf{x}_1=\mathbf{x}_0} \left. \frac{\partial}{\partial x_{2,l}} \log f(\mathbf{y}; \mathbf{x}_2) \right|_{\mathbf{x}_2=\mathbf{x}_0} f(\mathbf{y}; \mathbf{x}_0) d\mathbf{y} \\
&= \mathbb{E}_{\mathbf{x}_0} \left\{ \left. \frac{\partial}{\partial x_{1,k}} \log f(\mathbf{y}; \mathbf{x}_1) \right|_{\mathbf{x}_1=\mathbf{x}_0} \left. \frac{\partial}{\partial x_{2,l}} \log f(\mathbf{y}; \mathbf{x}_2) \right|_{\mathbf{x}_2=\mathbf{x}_0} \right\}, \tag{4.64}
\end{aligned}$$

where step (a) follows from Theorem 3.4.1 (using $g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)$ for the two choices $\mathbf{p} = \mathbf{e}_k$ and $\mathbf{p} = \mathbf{e}_l$, respectively, and $\mathbf{x}_c = \mathbf{x}_0$) and (b) is obtained by two applications of (4.51) in Postulate 4.4.2 (for the first application we use $h(\mathbf{y}) = \frac{f(\mathbf{y}; \mathbf{x}_1)}{f(\mathbf{y}; \mathbf{x}_0)}$ and differentiate w.r.t. $x_{2,l}$; and for the second application we set $h(\mathbf{y}) = \frac{\partial}{\partial x_{2,l}} \frac{f(\mathbf{y}; \mathbf{x}_2)}{f(\mathbf{y}; \mathbf{x}_0)} \Big|_{\mathbf{x}_2=\mathbf{x}_0}$ and differentiate w.r.t. $x_{1,k}$). Thus we have $\mathbf{V} = \mathbf{J}_{\mathbf{x}_0}$, where $\mathbf{J}_{\mathbf{x}_0}$ is the Fisher information matrix as defined in (4.57). As a consequence, by combining (4.19) of Theorem 4.3.4, (4.21), (4.61) and (4.63) we obtain

$$\begin{aligned}
L_{\mathcal{M}} &\stackrel{(4.19)}{=} \|g(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 - (g(\mathbf{x}_0))^2 \stackrel{(4.21)}{\geq} \|\mathbf{P}_{\mathcal{U}_{\text{CR}}} g(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 - (g(\mathbf{x}_0))^2 \\
&\stackrel{(4.61),(4.63)}{=} (g(\mathbf{x}_0))^2 + \left(\left. \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} \right)^T \mathbf{J}_{\mathbf{x}_0}^\dagger \left. \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} - (g(\mathbf{x}_0))^2 = \left(\left. \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} \right)^T \mathbf{J}_{\mathbf{x}_0}^\dagger \left. \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}_0}. \tag{4.65}
\end{aligned}$$

This coincides with the unconstrained CRB in (4.56), since we have $v(\hat{g}(\cdot); \mathbf{x}_0) \geq L_{\mathcal{M}}$ by the definition of $L_{\mathcal{M}}$ in (2.19).

Constrained CRB

Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ with associated minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot) \equiv 0, \mathbf{x}_0)$. We assume that the parameter set \mathcal{X} is defined via a set of equality constraints, i.e.,

$$\mathcal{X} = \{\mathbf{x} \in \mathbb{R}^N \mid \mathbf{f}(\mathbf{x}) = \mathbf{0}\}, \tag{4.66}$$

where $\mathbf{f}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}^L$ is a continuously differentiable vector-valued function. We also assume that $\mathbf{f}(\cdot)$ is such that the set \mathcal{X} is nonempty and that its Jacobian $\mathbf{F}(\mathbf{x}) \triangleq \frac{\partial \mathbf{f}(\mathbf{x})}{\partial \mathbf{x}} \in \mathbb{R}^{L \times N}$, given elementwise by $(\mathbf{F}(\mathbf{x}))_{m,n} \triangleq \frac{\partial f_m(\mathbf{x})}{\partial x_n}$, has rank L whenever $\mathbf{f}(\mathbf{x}) = \mathbf{0}$, i.e., for every $\mathbf{x} \in \mathcal{X}$. Such parameter sets are considered e.g. in [58–60].

As discussed in Section 2.5, there are two fundamentally different ways in how one can exploit the prior knowledge that is given by a specific form of the parameter set (e.g., such as (4.66)). For the problem of estimating the parameter vector itself, i.e., $\mathbf{g}(\mathbf{x}) = \mathbf{x}$, the authors of [58] require an estimator $\hat{\mathbf{g}}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}^N$ to take on values only in \mathcal{X} , i.e., $\hat{\mathbf{g}}(\mathbf{y}) \in \mathcal{X}$ for all $\mathbf{y} \in \mathbb{R}^M$. This is reasonable since if it is known that the unknown parameter vector \mathbf{x} is an element \mathcal{X} , then the value of an estimator of \mathbf{x} should also be an element of \mathcal{X} .

However, in this work, we will focus on another approach that is also considered by the authors of [59]. Within this second approach, we do not constrain the values of the estimator itself to be an element of \mathcal{X} but instead we only require the bias of the estimator to equal the prescribed bias for parameter vectors $\mathbf{x} \in \mathcal{X}$. In this approach, the a priori knowledge of a certain type of parameter set \mathcal{X} helps us in that we must place fewer bias constraints on an estimator. Thus, the set of allowed estimators $\mathcal{F}(\mathcal{M})$ in (2.18) tends to become larger as the parameter set \mathcal{X} is getting smaller, i.e., more “informative”.

The constrained CRB is a generalization of the unconstrained CRB to estimation problems with a parameter set \mathcal{X} of the form (4.66). A specific formulation of the constrained CRB is given in [59]:

Theorem 4.4.5. *Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and associated minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot) \equiv 0, \mathbf{x}_0)$ that satisfy Postulate 4.1.1 and Postulate 4.4.2 with $m = 1$ and whose parameter set \mathcal{X} satisfies (4.66). Then the variance $v(\hat{g}(\cdot); \mathbf{x}_0)$ at $\mathbf{x}_0 \in \mathcal{X}$ of any unbiased estimator $\hat{g}(\cdot)$ with a finite variance at \mathbf{x}_0 is lower bounded by*

$$v(\hat{g}(\cdot); \mathbf{x}_0) \geq \left(\left. \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} \right)^T \mathbf{U}(\mathbf{x}_0) (\mathbf{U}^T(\mathbf{x}_0) \mathbf{J}_{\mathbf{x}_0} \mathbf{U}(\mathbf{x}_0))^\dagger \mathbf{U}^T(\mathbf{x}_0) \left. \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}_0} \quad (4.67)$$

where $\mathbf{J}_{\mathbf{x}_0}$ is the FIM as defined in (4.57) and $\mathbf{U}(\mathbf{x}_0) \in \mathbb{R}^{N \times (N-L)}$ is any matrix whose column vectors form an ONB for the null space $\mathcal{N}(\mathbf{F}(\mathbf{x}_0))$ of the Jacobian $\mathbf{F}(\mathbf{x}_0)$, i.e.,

$$\mathbf{F}(\mathbf{x}_0) \mathbf{U}(\mathbf{x}_0) = \mathbf{0}, \quad \mathbf{U}^T(\mathbf{x}_0) \mathbf{U}(\mathbf{x}_0) = \mathbf{I}_{(N-L)}. \quad (4.68)$$

Proof. [59] □

Note that if the function $\mathbf{f}(\mathbf{x})$ in (4.66) is given as $\mathbf{f}(\mathbf{x}) \equiv \mathbf{0}$, one can verify that the matrix $\mathbf{U}(\mathbf{x}_0)$ used in Theorem 4.4.5 can be chosen as $\mathbf{U}(\mathbf{x}_0) = \mathbf{I}$ in which case the constrained CRB (4.67) reduces to the unconstrained CRB (4.56). We now re-derive the bound (4.67) using the

RKHS $\mathcal{H}(\mathcal{M})$ associated to the minimum variance problem \mathcal{M} of Theorem 4.4.5. This derivation will be very similar to the RKHS-based derivation of the unconstrained CRB as performed above. However, there are some modifications necessary due to the additional assumption on the parameter set \mathcal{X} given by (4.66). These modifications are based mainly on the methods used in [60].

In particular, will show that the bound (4.67) can be obtained via an orthogonal projection on a finite-dimensional subspace $\mathcal{U}_{\text{CCRB}} \subseteq \mathcal{H}(\mathcal{M})$ which is defined by

$$\mathcal{U}_{\text{CCRB}} \triangleq \text{span} \{ \{v_0(\cdot)\} \cup \{v_l(\cdot)\}_{l \in [N-L]} \}, \quad (4.69)$$

where $v_0(\cdot) \triangleq R_{\mathcal{M}}(\cdot, \mathbf{x}_0) \in \mathcal{H}(\mathcal{M})$ and the functions $\{v_l(\cdot)\}_{l \in [N-L]} \in \mathcal{H}(\mathcal{M})$ will be specified in the following. We first note that given the fixed vector \mathbf{x}_0 associated to the minimum variance problem \mathcal{M} , we have according to the implicit function theorem (see [60, Theorem 3.3] or [5]) that there exists a continuously differentiable and bijective map $\mathbf{r}(\cdot)$, which moreover has a continuously differentiable inverse⁴ $\mathbf{r}^{-1}(\cdot)$, from an open set $\mathbb{O} \subseteq \mathbb{R}^{N-L}$ into an open set $\mathbb{P} \subseteq \mathcal{X}$ which also contains \mathbf{x}_0 , i.e.,

$$\mathbf{r}(\cdot) : \mathbb{O} \subseteq \mathbb{R}^{N-L} \rightarrow \mathbb{P} \subseteq \mathcal{X} \quad (4.70)$$

where $\mathbf{x}_0 \in \mathbb{P}$. Note that any function value $\mathbf{r}(\boldsymbol{\theta})$ is an element of the parameter set, i.e., $\mathbf{r}(\boldsymbol{\theta}) \in \mathcal{X}$. The Jacobian of $\mathbf{r}(\cdot)$ at an arbitrary $\boldsymbol{\theta} \in \mathbb{O}$ will be denoted by $\mathbf{G}(\boldsymbol{\theta}) \in \mathbb{R}^{N \times (N-L)}$, i.e.,

$$(\mathbf{G}(\boldsymbol{\theta}))_{m,n} = \frac{\partial r_m(\boldsymbol{\theta})}{\partial \theta_n}. \quad (4.71)$$

We now define the functions $v_l(\cdot) \in \mathcal{H}(\mathcal{M})$ in (4.69) as

$$v_l(\cdot) \triangleq \left. \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{r}(\boldsymbol{\theta}))}{\partial \theta_l} \right|_{\boldsymbol{\theta}=\mathbf{r}^{-1}(\mathbf{x}_0)}, \quad l \in [N-L]. \quad (4.72)$$

The fact that the so-defined functions $v_l(\cdot)$ belong to $\mathcal{H}(\mathcal{M})$ can again be verified straightforwardly by Theorem 4.4.3 together with Theorem 3.4.1. Indeed, using the chain rule for differentiation [5, Theorem 9.15], we have

$$\begin{aligned} v_l(\cdot) &= \left. \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{r}(\boldsymbol{\theta}))}{\partial \theta_l} \right|_{\boldsymbol{\theta}=\mathbf{r}^{-1}(\mathbf{x}_0)} \\ &= \sum_{l' \in [N]} \left. \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{x})}{\partial x_{l'}} \right|_{\mathbf{x}=\mathbf{x}_0} \left. \frac{\partial r_{l'}(\boldsymbol{\theta})}{\partial \theta_l} \right|_{\boldsymbol{\theta}=\mathbf{r}^{-1}(\mathbf{x}_0)}, \quad l \in [N-L], \end{aligned} \quad (4.73)$$

which is a linear combination of the functions $\left. \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{x})}{\partial x_{l'}} \right|_{\mathbf{x}=\mathbf{x}_0}$ which are of the same form (for the choice $\mathbf{p} = \mathbf{e}_{l'}$ and $\mathbf{x}_c = \mathbf{x}_0$) as the functions $g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)$ defined in Theorem 3.4.1,

⁴Such a function is called a diffeomorphism [60].

The inner product between any $v_l(\cdot)$ with $l \in [N - L]$ and $v_0(\cdot)$ is given by (cf. our derivation in (4.60))

$$\begin{aligned} \langle v_0(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} &= \left. \frac{\partial R_{\mathcal{M}}(\mathbf{x}_0, \mathbf{r}(\boldsymbol{\theta}))}{\partial \theta_l} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left. \frac{\partial}{\partial \theta_l} \int_{\mathbf{y}} \frac{f(\mathbf{y}; \mathbf{r}(\boldsymbol{\theta})) f(\mathbf{y}; \mathbf{x}_0)}{f(\mathbf{y}; \mathbf{x}_0)} d\mathbf{y} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \\ &= \left. \frac{\partial}{\partial \theta_l} \int_{\mathbf{y}} f(\mathbf{y}; \mathbf{r}(\boldsymbol{\theta})) d\mathbf{y} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left. \frac{\partial}{\partial \theta_l} 1 \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = 0, \end{aligned} \quad (4.74)$$

where we introduced the shorthand $\boldsymbol{\theta}_0 \triangleq \mathbf{r}^{-1}(\mathbf{x}_0)$.

Since Theorem 4.4.5 makes a statement about an unbiased estimator of the parameter function $g(\cdot)$ that has a finite variance at \mathbf{x}_0 , we can assume that the parameter function $g(\cdot)$ is estimable, i.e., we have $g(\cdot) \in \mathcal{H}(\mathcal{M})$ by Theorem 4.3.4 (where again $\gamma(\cdot) = g(\cdot)$).

According to Theorem 3.1.8 and (4.74), we can express the squared norm of the orthogonal projection $\mathbf{P}_{\mathcal{U}_{\text{CCRB}}} g(\cdot)$ as

$$\|\mathbf{P}_{\mathcal{U}_{\text{CCRB}}} g(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 = (\langle g(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M})})^2 + \mathbf{a}^T \mathbf{V}^\dagger \mathbf{a} = (g(\mathbf{x}_0))^2 + \mathbf{a}^T \mathbf{V}^\dagger \mathbf{a}, \quad (4.75)$$

where the vector $\mathbf{a} \in \mathbb{R}^{N-L}$ and the matrix $\mathbf{V} \in \mathbb{R}^{(N-L) \times (N-L)}$ are defined elementwise as $a_l \triangleq \langle g(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})}$ and $(\mathbf{V})_{k,l} \triangleq \langle v_k(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})}$, respectively. Due to the derivative-reproducing property (3.66) and the chain rule of differentiation [5, Theorem 9.15], we obtain

$$\begin{aligned} a_l &= \langle g(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} = \left\langle g(\cdot), \left. \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{r}(\boldsymbol{\theta}))}{\partial \theta_l} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right\rangle_{\mathcal{H}(\mathcal{M})} \\ &\stackrel{(4.73)}{=} \left\langle g(\cdot), \sum_{\nu \in [N]} \left. \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{x})}{\partial x_\nu} \right|_{\mathbf{x}=\mathbf{x}_0} \left. \frac{\partial r_\nu(\boldsymbol{\theta})}{\partial \theta_l} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right\rangle_{\mathcal{H}(\mathcal{M})} \\ &= \sum_{\nu \in [N]} \left. \frac{\partial r_\nu(\boldsymbol{\theta})}{\partial \theta_l} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \left\langle g(\cdot), \left. \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{x})}{\partial x_\nu} \right|_{\mathbf{x}=\mathbf{x}_0} \right\rangle_{\mathcal{H}(\mathcal{M})} \\ &\stackrel{(3.66)}{=} \sum_{\nu \in [N]} \left. \frac{\partial r_\nu(\boldsymbol{\theta})}{\partial \theta_l} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \left. \frac{\partial g(\mathbf{x})}{\partial x_\nu} \right|_{\mathbf{x}_0} \stackrel{(4.71)}{=} \sum_{\nu \in [N]} (\mathbf{G}(\boldsymbol{\theta}_0))_{\nu, l} \left. \frac{\partial g(\mathbf{x})}{\partial x_\nu} \right|_{\mathbf{x}_0}, \end{aligned} \quad (4.76)$$

, i.e., in vector notation:

$$\mathbf{a} = \mathbf{G}^T(\boldsymbol{\theta}_0) \left. \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \right|_{\mathbf{x}_0}. \quad (4.77)$$

Furthermore, by using the chain rule for derivation once again, we obtain

$$(\mathbf{V})_{k,l} = \langle v_k(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} = \left\langle \left. \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{r}(\boldsymbol{\theta}))}{\partial \theta_k} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \left. \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{r}(\boldsymbol{\theta}))}{\partial \theta_l} \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right\rangle_{\mathcal{H}(\mathcal{M})}$$

$$\begin{aligned}
&= \left\langle \sum_{l' \in [N]} \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{x})}{\partial x_{l'}} \Big|_{\mathbf{x}=\mathbf{x}_0} \frac{\partial r_{l'}(\boldsymbol{\theta})}{\partial \theta_k} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \sum_{l'' \in [N]} \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{x})}{\partial x_{l''}} \Big|_{\mathbf{x}=\mathbf{x}_0} \frac{\partial r_{l''}(\boldsymbol{\theta})}{\partial \theta_l} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right\rangle_{\mathcal{H}(\mathcal{M})} \\
&= \sum_{l', l'' \in [N]} \frac{\partial r_{l'}(\boldsymbol{\theta})}{\partial \theta_k} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \frac{\partial r_{l''}(\boldsymbol{\theta})}{\partial \theta_l} \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \left\langle \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{x})}{\partial x_{l'}} \Big|_{\mathbf{x}=\mathbf{x}_0}, \frac{\partial R_{\mathcal{M}}(\cdot, \mathbf{x})}{\partial x_{l''}} \Big|_{\mathbf{x}=\mathbf{x}_0} \right\rangle_{\mathcal{H}(\mathcal{M})} \\
&\stackrel{(a)}{=} \sum_{l', l'' \in [N]} (\mathbf{G}(\boldsymbol{\theta}_0))_{l', k} (\mathbf{G}(\boldsymbol{\theta}_0))_{l'', l} \frac{\partial^2 R_{\mathcal{M}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial x_{1, l'} \partial x_{2, l''}} \Big|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}_0} \\
&= (\mathbf{G}(\boldsymbol{\theta}_0)^T \mathbf{W} \mathbf{G}(\boldsymbol{\theta}_0))_{k, l}, \tag{4.78}
\end{aligned}$$

where step (a) is due to Theorem 3.4.1 (using $g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)$ for the choices $\mathbf{p} = \mathbf{e}_{l'}$, $\mathbf{p} = \mathbf{e}_{l''}$ and $\mathbf{x}_c = \mathbf{x}_0$), and the matrix $\mathbf{W} \in \mathbb{R}^{N \times N}$ is given elementwise via

$$(\mathbf{W})_{k, l} \triangleq \frac{\partial^2 R_{\mathcal{M}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial x_{1, k} \partial x_{2, l}} \Big|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}_0} = \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{\partial}{\partial x_{1, k}} \log f(\mathbf{y}; \mathbf{x}_1) \Big|_{\mathbf{x}_1=\mathbf{x}_0} \frac{\partial}{\partial x_{2, l}} \log f(\mathbf{y}; \mathbf{x}_2) \Big|_{\mathbf{x}_2=\mathbf{x}_0} \right\}. \tag{4.79}$$

Here, we reused for (4.79) the derivation of (4.64). According to (4.79), the matrix \mathbf{W} coincides with the FIM $\mathbf{J}_{\mathbf{x}_0}$ as defined in (4.57), i.e., $\mathbf{W} = \mathbf{J}_{\mathbf{x}_0}$. Putting together the pieces, we obtain

$$\begin{aligned}
L_{\mathcal{M}} &\stackrel{(4.19)}{=} \|g(\cdot)\|_2^2 - (g(\mathbf{x}_0))^2 \stackrel{(4.21)}{\geq} \|\mathbf{P}_{\mathcal{U}_{\text{CCRB}}} g(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 - (g(\mathbf{x}_0))^2 \\
&\stackrel{(4.75)}{=} \mathbf{a}^T \mathbf{V}^\dagger \mathbf{a} \stackrel{(4.78), (4.77)}{=} \left(\frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \Big|_{\mathbf{x}_0} \right)^T \mathbf{G}(\boldsymbol{\theta}_0) (\mathbf{G}^T(\boldsymbol{\theta}_0) \mathbf{J}_{\mathbf{x}_0} \mathbf{G}(\boldsymbol{\theta}_0))^\dagger \mathbf{G}^T(\boldsymbol{\theta}_0) \frac{\partial g(\mathbf{x})}{\partial \mathbf{x}} \Big|_{\mathbf{x}_0}. \tag{4.80}
\end{aligned}$$

As shown in [60, p. 29], the Jacobian $\mathbf{G}(\boldsymbol{\theta}_0)$ in (4.71) and the matrix $\mathbf{U}(\mathbf{x}_0)$ appearing in Theorem 4.4.5 are related by an invertible linear transformation, i.e., there exists a nonsingular matrix $\mathbf{S} \in \mathbb{R}^{(N-L) \times (N-L)}$ (which may depend on \mathbf{x}_0) such that

$$\mathbf{G}(\boldsymbol{\theta}_0) \mathbf{S} = \mathbf{U}(\mathbf{x}_0). \tag{4.81}$$

However, one can always find a map $\mathbf{r}(\cdot)$ such that $\mathbf{S} = \mathbf{I}$, in which case $\mathbf{G}(\boldsymbol{\theta}_0) = \mathbf{U}(\mathbf{x}_0)$ [60, p. 50]. Indeed, suppose that the map $\mathbf{r}(\cdot)$ is such that $\mathbf{S} \neq \mathbf{I}$ and therefore $\mathbf{G}(\boldsymbol{\theta}_0) \neq \mathbf{U}(\mathbf{x}_0)$. We can then use for the definition of the functions $v_l(\cdot)$ in (4.72) instead of $\mathbf{r}(\cdot)$ a new map $\mathbf{r}'(\cdot)$ defined as $\mathbf{r}'(\boldsymbol{\theta}) = \mathbf{r}(\mathbf{S}\boldsymbol{\theta})$, whose Jacobian $\mathbf{G}'(\boldsymbol{\theta}_0)$ satisfies

$$\mathbf{G}'(\boldsymbol{\theta}_0) \stackrel{(a)}{=} \mathbf{G}(\boldsymbol{\theta}_0) \mathbf{S} \stackrel{(4.81)}{=} \mathbf{U}(\mathbf{x}_0), \tag{4.82}$$

where step (a) is due to the chain rule for differentiation [5, Theorem 9.15].

Thus, since we can assume that $\mathbf{G}(\boldsymbol{\theta}_0) = \mathbf{U}(\mathbf{x}_0)$, we have by the definition of $L_{\mathcal{M}}$ in (2.19) that the bound (4.80) coincides with the bound (4.67).

4.4.3 Bhattacharyya Bound

The CRB of a given minimum variance problem depends only on the first-order partial derivatives of the statistical model $f(\mathbf{y}, \mathbf{x})$ w.r.t. to the parameter \mathbf{x} . By incorporating also higher-order partial derivatives, one obtains the Bhattacharyya bound [75,76]. Like the CRB the Bhattacharyya bound, is a lower bound on the variance of any unbiased estimator for a given estimation problem (which is required to satisfy some regularity conditions). The Bhattacharyya bound may be formulated in our setting as follows.

Theorem 4.4.6. *Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ with associated minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot) \equiv 0, \mathbf{x}_0)$ that satisfy the conditions of Postulate 4.1.1 and Postulate 4.4.2 with maximum order of differentiation equal to m . For any set of $L \in \mathbb{N}$ multi-indices $\{\mathbf{p}_l \in \mathbb{Z}_+^N\}_{l \in [L]}$ with $\|\mathbf{p}_l\|_\infty \leq m$, we have that the variance $v(\hat{g}(\cdot); \mathbf{x}_0)$ at \mathbf{x}_0 of any unbiased estimator $\hat{g}(\cdot)$ with finite variance at \mathbf{x}_0 is lower bounded by*

$$v(\hat{g}(\cdot); \mathbf{x}_0) \geq \mathbf{b}^T \mathbf{B}^\dagger \mathbf{b}, \quad (4.83)$$

where the vector $\mathbf{b} \in \mathbb{R}^L$ and the matrix $\mathbf{B} \in \mathbb{R}^{L \times L}$ are given elementwise by

$$b_l \triangleq \left. \frac{\partial^{\mathbf{p}_l} g(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}_l}} \right|_{\mathbf{x}_0} \quad (4.84)$$

and

$$(\mathbf{B})_{k,l} \triangleq \mathbb{E}_{\mathbf{x}_0} \left\{ \left. \frac{1}{f^2(\mathbf{y}; \mathbf{x}_0)} \frac{\partial^{\mathbf{p}_k} f(\mathbf{y}; \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}_k}} \right|_{\mathbf{x}=\mathbf{x}_0} \left. \frac{\partial^{\mathbf{p}_l} f(\mathbf{y}; \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}_l}} \right|_{\mathbf{x}=\mathbf{x}_0} \right\}, \quad (4.85)$$

respectively.

A derivation of Theorem 4.4.6 can be based on the RKHS $\mathcal{H}(\mathcal{M})$. This derivation is completely analogous to that of Theorem 4.4.4. The only difference is that instead of the subspace \mathcal{U}_{CR} defined via (4.58) and (4.59), which is used for the derivation of Theorem 4.4.4, one has to use the subspace $\mathcal{U}_{\text{BHATT}} \triangleq \text{span}\{\{v_0(\cdot)\} \cup \{v_l(\cdot)\}_{l \in [L]}\}$. Here, we used the functions $v_0(\cdot) \triangleq R_{\mathcal{M}}(\cdot, \mathbf{x}_0) \in \mathcal{H}(\mathcal{M})$ and $v_l(\cdot) \triangleq \left. \frac{\partial^{\mathbf{p}_l} R_{\mathcal{M}}(\cdot, \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}_l}} \right|_{\mathbf{x}=\mathbf{x}_0} \in \mathcal{H}(\mathcal{M})$ for $l \in [L]$. The fact that the functions $\{v_l(\cdot)\}_{l \in [L]}$ belong to the RKHS $\mathcal{H}(\mathcal{M})$ may be verified by Theorem 4.4.3 together with Theorem 3.4.1.

While the RKHS interpretation of the Bhattacharyya bound has already been presented in [35] for a specific estimation problem, the above derivation holds for general estimation problems. Finally, note that the unconstrained CRB in Theorem 4.4.4 is obtained as a special case of the Bhattacharyya bound for the choices $L = N$, $m = 1$ and multi-indices $\{\mathbf{p}_l = \mathbf{e}_l\}_{l \in [L]}$ in Theorem 4.4.6.

4.4.4 Hammersley-Chapman-Robbins Bound

Despite its popularity, a drawback of the CRB (and also of the Bhattacharyya bound) is that it only exploits the local structure of an estimation problem \mathcal{E} around a specific point $\mathbf{x}_0 \in \mathcal{X}$ [75]. As an illustrative example, consider two different estimation problems $\mathcal{E}_1 = (\mathcal{X}_1, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and $\mathcal{E}_2 = (\mathcal{X}_2, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ that share the same statistical model $f(\mathbf{y}; \mathbf{x})$ and parameter function $g(\cdot)$ but are defined for two different parameter sets \mathcal{X}_1 and \mathcal{X}_2 . Let us assume that both parameter sets are open balls centered at \mathbf{x}_0 but with different radii r_1 and r_2 , i.e., $\mathcal{X}_1 = \mathcal{B}(\mathbf{x}_0, r_1)$ and $\mathcal{X}_2 = \mathcal{B}(\mathbf{x}_0, r_2)$ with $r_1 \neq r_2$. Then the CRB at \mathbf{x}_0 will be the same for both estimation problems irrespective of the precise values of r_1, r_2 . Thus, the CRB ignores in some sense a part of the information that is contained in the parameter set \mathcal{X} .

On the other hand, the Barankin bound exploits the full information carried by the parameter set \mathcal{X} since it is the tightest possible lower bound on the estimator variance. However, in the general case, it is difficult to evaluate the exact Barankin bound in (4.44).

The Hammersley-Chapman-Robbins bound (HCRB) [29, 77, 78] is a lower bound on the estimator variance which also incorporates the “global” structure of the estimation problem, i.e., it depends on the parameter set \mathcal{X} not only through its local structure as does the CRB. However, it can be evaluated much more easily than the Barankin bound. In our context, the HCRB can be stated as follows [29]:

Theorem 4.4.7. *Consider an estimation problem $\mathcal{E} = (\mathcal{X}, f(\mathbf{y}; \mathbf{x}), g(\cdot))$ and associated minimum variance problem $\mathcal{M} = (\mathcal{E}, c(\cdot) \equiv 0, \mathbf{x}_0)$ that satisfy Postulate 4.1.1. Furthermore, consider a finite set of “test-points” $\{\mathbf{x}_l \in \mathcal{X}\}_{l \in [L]}$. Then the variance $v(\hat{g}(\cdot); \mathbf{x}_0)$ at $\mathbf{x}_0 \in \mathcal{X}$ of any unbiased estimator $\hat{g}(\cdot)$ with a finite variance at \mathbf{x}_0 is lower bounded by*

$$v(\hat{g}(\cdot); \mathbf{x}_0) \geq \mathbf{m}^T \mathbf{V}^\dagger \mathbf{m}, \quad (4.86)$$

where the vector $\mathbf{m} \in \mathbb{R}^L$ and the matrix $\mathbf{V} \in \mathbb{R}^{L \times L}$ are defined elementwise by

$$m_l \triangleq g(\mathbf{x}_l) - g(\mathbf{x}_0) \quad (4.87)$$

and

$$(\mathbf{V})_{l,l'} \triangleq \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{[f(\mathbf{y}; \mathbf{x}_l) - f(\mathbf{y}; \mathbf{x}_0)][f(\mathbf{y}; \mathbf{x}_{l'}) - f(\mathbf{y}; \mathbf{x}_0)]}{f^2(\mathbf{y}; \mathbf{x}_0)} \right\}, \quad (4.88)$$

respectively.

Proof. [29] □

We now show how the bound in (4.86) can be obtained via the RKHS $\mathcal{H}(\mathcal{M})$. First, note that since Theorem 4.4.7 makes a statement about an existing unbiased estimator of $g(\mathbf{x})$ with finite

variance at \mathbf{x}_0 , we can assume that $g(\cdot)$ is estimable, i.e., $g(\cdot) \in \mathcal{H}(\mathcal{M})$ according to Theorem 4.3.4. Then, let us define the subspace

$$\mathcal{U}_{\text{HCRB}} \triangleq \text{span} \left\{ \{v_0(\cdot)\} \cup \{v_l(\cdot)\}_{l \in [L]} \right\}, \quad (4.89)$$

where L is the number of test-points $\{\mathbf{x}_l\}_{l \in [L]}$ used in Theorem 4.4.7. The subspace $\mathcal{U}_{\text{HCRB}}$ is spanned by the functions $v_0(\cdot) \triangleq R_{\mathcal{M}}(\cdot, \mathbf{x}_0)$ and $\{v_l(\cdot) \triangleq R_{\mathcal{M}}(\cdot, \mathbf{x}_l) - R_{\mathcal{M}}(\cdot, \mathbf{x}_0)\}_{l \in [L]}$. Note that we obviously have by Theorem 3.2.2 that these functions belong to $\mathcal{H}(\mathcal{M})$.

The function $v_0(\cdot) \in \mathcal{H}(\mathcal{M})$ is orthogonal to any $v_l(\cdot) \in \mathcal{H}(\mathcal{M})$ with $l \in [L]$, i.e., the two sets $\{v_0(\cdot)\}$ and $\{v_l(\cdot)\}_{l \in [L]}$ are mutually orthogonal, since by the reproducing property (3.44) we have

$$\begin{aligned} \langle v_0(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} &= \langle R_{\mathcal{M}}(\cdot, \mathbf{x}_0), R_{\mathcal{M}}(\cdot, \mathbf{x}_l) - R_{\mathcal{M}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M})} \\ &= \langle R_{\mathcal{M}}(\cdot, \mathbf{x}_0), R_{\mathcal{M}}(\cdot, \mathbf{x}_l) \rangle_{\mathcal{H}(\mathcal{M})} - \langle R_{\mathcal{M}}(\cdot, \mathbf{x}_0), R_{\mathcal{M}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M})} \\ &= R_{\mathcal{M}}(\mathbf{x}_l, \mathbf{x}_0) - R_{\mathcal{M}}(\mathbf{x}_0, \mathbf{x}_0) \stackrel{(4.11)}{=} 1 - 1 = 0. \end{aligned} \quad (4.90)$$

Now we define the matrix $\mathbf{V} \in \mathbb{R}^{L \times L}$ elementwise by

$$\begin{aligned} (\mathbf{V})_{l,l'} &\triangleq \langle v_l(\cdot), v_{l'}(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} = \left\langle R_{\mathcal{M}}(\cdot, \mathbf{x}_l) - R_{\mathcal{M}}(\cdot, \mathbf{x}_0), R_{\mathcal{M}}(\cdot, \mathbf{x}_{l'}) - R_{\mathcal{M}}(\cdot, \mathbf{x}_0) \right\rangle_{\mathcal{H}(\mathcal{M})} \\ &\stackrel{(3.44)}{=} R_{\mathcal{M}}(\mathbf{x}_{l'}, \mathbf{x}_l) - R_{\mathcal{M}}(\mathbf{x}_{l'}, \mathbf{x}_0) - R_{\mathcal{M}}(\mathbf{x}_l, \mathbf{x}_0) + R_{\mathcal{M}}(\mathbf{x}_0, \mathbf{x}_0). \end{aligned} \quad (4.91)$$

The inner products of $g(\cdot)$ with the vectors $v_l(\cdot)$ are calculated (also using the reproducing property (3.44)) as

$$\langle g(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} = g(\mathbf{x}_0), \quad \text{and} \quad \langle g(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} = g(\mathbf{x}_l) - g(\mathbf{x}_0) = m_l \quad \text{for } l \in [L]. \quad (4.92)$$

By projecting the prescribed mean function $\gamma(\cdot) = g(\cdot) + c(\cdot) = g(\cdot)$ of the minimum variance problem \mathcal{M} on the subspace $\mathcal{U}_{\text{HCRB}}$, we obtain

$$\begin{aligned} L_{\mathcal{M}} &\stackrel{(4.19)}{=} \|g(\cdot)\|_2^2 - (g(\mathbf{x}_0))^2 \stackrel{(4.21)}{\geq} \|\mathbf{P}_{\mathcal{U}_{\text{HCRB}}} g(\cdot)\|_{\mathcal{H}(\mathcal{M})}^2 - (g(\mathbf{x}_0))^2 \\ &\stackrel{(a)}{=} \langle g(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} (\langle v_0(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M})})^{-1} \langle g(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} + \mathbf{m}^T \mathbf{V}^\dagger \mathbf{m} - (g(\mathbf{x}_0))^2 \\ &\stackrel{(b)}{=} (g(\mathbf{x}_0))^2 + \mathbf{m}^T \mathbf{V}^\dagger \mathbf{m} - (g(\mathbf{x}_0))^2 = \mathbf{m}^T \mathbf{V}^\dagger \mathbf{m}, \end{aligned} \quad (4.93)$$

where for step (a) we used (4.92) and Theorem 3.1.9, which applies since $\mathcal{U}_{\text{HCRB}}$ is defined as the linear span of the union of the two mutually orthogonal function sets $\{v_0(\cdot)\}$ and $\{v_l(\cdot)\}_{l \in [L]}$. The step (b) follows from

$$\langle v_0(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M})} = \langle R_{\mathcal{M}}(\cdot, \mathbf{x}_0), R_{\mathcal{M}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M})} \stackrel{(3.44)}{=} R_{\mathcal{M}}(\mathbf{x}_0, \mathbf{x}_0) \stackrel{(4.11)}{=} 1. \quad (4.94)$$

and (4.92).

The bound in (4.86) follows from (4.93), the definition of $L_{\mathcal{M}}$ in (2.19) and the fact that

$$\begin{aligned}
(\mathbf{V})_{l,l'} &\stackrel{(4.91)}{=} R_{\mathcal{M}}(\mathbf{x}_{l'}, \mathbf{x}_l) - R_{\mathcal{M}}(\mathbf{x}_{l'}, \mathbf{x}_0) - R_{\mathcal{M}}(\mathbf{x}_l, \mathbf{x}_0) + R_{\mathcal{M}}(\mathbf{x}_0, \mathbf{x}_0) \\
&= \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{f(\mathbf{y}; \mathbf{x}_{l'}) f(\mathbf{y}; \mathbf{x}_l)}{f(\mathbf{y}; \mathbf{x}_0) f(\mathbf{y}; \mathbf{x}_0)} \right\} - \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{f(\mathbf{y}; \mathbf{x}_{l'}) f(\mathbf{y}; \mathbf{x}_0)}{f(\mathbf{y}; \mathbf{x}_0) f(\mathbf{y}; \mathbf{x}_0)} \right\} - \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{f(\mathbf{y}; \mathbf{x}_l) f(\mathbf{y}; \mathbf{x}_0)}{f(\mathbf{y}; \mathbf{x}_0) f(\mathbf{y}; \mathbf{x}_0)} \right\} \\
&\quad + \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{f(\mathbf{y}; \mathbf{x}_0) f(\mathbf{y}; \mathbf{x}_0)}{f(\mathbf{y}; \mathbf{x}_0) f(\mathbf{y}; \mathbf{x}_0)} \right\} \\
&= \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{f(\mathbf{y}; \mathbf{x}_{l'}) - f(\mathbf{y}; \mathbf{x}_0)}{f(\mathbf{y}; \mathbf{x}_0)} \frac{f(\mathbf{y}; \mathbf{x}_l) - f(\mathbf{y}; \mathbf{x}_0)}{f(\mathbf{y}; \mathbf{x}_0)} \right\} \\
&= \mathbb{E}_{\mathbf{x}_0} \left\{ \frac{[f(\mathbf{y}; \mathbf{x}_l) - f(\mathbf{y}; \mathbf{x}_0)][f(\mathbf{y}; \mathbf{x}_{l'}) - f(\mathbf{y}; \mathbf{x}_0)]}{f^2(\mathbf{y}; \mathbf{x}_0)} \right\}. \tag{4.95}
\end{aligned}$$

Chapter 5

The Sparse Linear Model

5.1 Introduction

In this chapter, we study the problem of estimating a nonrandom parameter vector $\mathbf{x} \in \mathbb{R}^N$ which is known to be strictly S -sparse, i.e., at most S of its entries are nonzero, where $S \in [N]$ (typically $S \ll N$). We thus have

$$\mathbf{x} \in \mathcal{X}_S, \quad \text{with } \mathcal{X}_S = \{\mathbf{x}' \in \mathbb{R}^N \mid \|\mathbf{x}'\|_0 \leq S\}. \quad (5.1)$$

While the sparsity degree $S \in \mathbb{N}$ is assumed to be known, the set of positions of the nonzero entries of \mathbf{x} , i.e., the support $\text{supp}(\mathbf{x})$, is unknown.

The estimation of \mathbf{x} is based on the observed vector $\mathbf{y} \in \mathbb{R}^M$ given by

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (5.2)$$

with a known system matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ and additive white Gaussian noise (AWGN) $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ with a known positive (i.e., nonzero) noise variance $\sigma^2 > 0$. The matrix \mathbf{H} is arbitrary, typically, it will be assumed to satisfy the requirement

$$\text{spark}(\mathbf{H}) > S, \quad (5.3)$$

where $\text{spark}(\mathbf{H})$ denotes the minimum number of linearly dependent columns of \mathbf{H} [17].¹ Note that we also allow $M < N$ (this case is relevant to compressed sensing methods as discussed in Section 5.7); however, condition (5.3) implies that $M \geq S$. Furthermore, note that the requirement (5.3) is weaker than the standard requirement [25]

$$\text{spark}(\mathbf{H}) > 2S, \quad (5.4)$$

¹For a matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ with full column rank, i.e., $\text{rank}(\mathbf{H}) = N$, we define $\text{spark}(\mathbf{H}) = N$.

which is reasonable since otherwise we can find two different parameter vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_S$ for which the pdf of the observation is identical, i.e., $f(\mathbf{y}; \mathbf{x}_1) = f(\mathbf{y}; \mathbf{x}_2)$ for every \mathbf{y} , which means that it is completely impossible to distinguish between \mathbf{x}_1 and \mathbf{x}_2 only by looking at the observation \mathbf{y} .

More formally, by combining the observation model (5.2) with the requirement (5.1), we define the *sparse linear model* (SLM) as the specific estimation problem denoted \mathcal{E}_{SLM} and given as

$$\mathcal{E}_{\text{SLM}} \triangleq (\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k), \quad (5.5)$$

where $k \in [N]$ is an arbitrary but fixed index (the precise choice of k will be made explicit whenever necessary) and with the statistical model

$$f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}) = \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right). \quad (5.6)$$

One may argue that considering the parameter function $g(\mathbf{x}) = x_k$ is too restrictive, since in practice one may be interested in the whole parameter vector \mathbf{x} rather than only a single entry x_k . However, as discussed in Section 2.3.1, in the context of minimum variance estimation, we have that estimation of the parameter vector \mathbf{x} is completely equivalent to the N separate estimation problems

$$\left\{ (\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k) \right\}_{k \in [N]}. \quad (5.7)$$

By comparing (5.6) with (2.38), we see that the statistical model of the SLM is an instance of the standard exponential family, with sufficient statistic $\Phi(\mathbf{y}) = \frac{1}{\sigma^2}\mathbf{H}^T\mathbf{y}$, parameter function $\mathbf{u}(\mathbf{x}) = \mathbf{x}$, cumulant function $A^{(\Phi)}(\mathbf{x}) = \frac{1}{2\sigma^2}\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}$, and weight $h(\mathbf{y}) = \exp\left(-\frac{1}{2\sigma^2}\mathbf{y}^T\mathbf{y}\right)$.

The SLM is very similar to a well-known estimation problem, i.e., the *linear Gaussian model* (LGM) [19–21, 56], which is given by

$$\mathcal{E}_{\text{LGM}} \triangleq (\mathbb{R}^N, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k). \quad (5.8)$$

Indeed, the only difference between the LGM \mathcal{E}_{LGM} and the SLM \mathcal{E}_{SLM} is the parameter set. The SLM is obtained from the LGM by reducing the parameter set from $\mathcal{X} = \mathbb{R}^N$ to the set of S -sparse vectors, i.e., $\mathcal{X} = \mathcal{X}_S$. If the sparsity degree S is equal to the dimension of the observation, i.e., $S = N$, then the SLM coincides with the LGM.

The SLM is relevant, e.g., for sparse channel estimation [11] where the sparse parameter vector \mathbf{x} represents the taps of a linear time-invariant channel and the system matrix \mathbf{H} represents the training signal. Generally, the SLM can be used for any type of sparse deconvolution [37] like in optical coherence tomography (OCT) [79], where the sparse vector \mathbf{x} models the layer structure of the human eye and the matrix \mathbf{H} represents the properties of the measurement device.

An important special case of the SLM is given by $\mathbf{H} = \mathbf{I}$ (so that $M = N$), i.e.,

$$\mathbf{y} = \mathbf{x} + \mathbf{n}, \quad (5.9)$$

where again $\mathbf{x} \in \mathcal{X}_S$ and $\mathbf{n} \in \mathbb{R}^N$ represents AWGN, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ with known variance $\sigma^2 > 0$. This will be referred to as the *sparse signal in noise model* (SSNM). Formally, the SSNM is defined as the estimation problem

$$\mathcal{E}_{\text{SSNM}} \triangleq (\mathcal{X}_S, f_{\mathbf{H}=\mathbf{I}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k), \quad (5.10)$$

i.e., with the statistical model (5.6) for the specific choice $\mathbf{H} = \mathbf{I}$ for the system matrix. The SSNM can be used, e.g., for channel estimation [11] when the channel consists only of few significant taps and an orthogonal training signal is used [80]. Another potential application for the SSNM is image denoising using an orthonormal wavelet basis [23]. Note that due to Theorem 2.4.2, any lower bound on the minimum achievable variance of a minimum variance problem $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}_0)$ entails a lower bound on the minimum achievable variance for a minimum variance problem for any estimation problem with parameter set \mathcal{X}_S and observation model

$$\mathbf{y} = \mathbf{H}(\mathbf{x} + \mathbf{n}), \quad (5.11)$$

where \mathbf{n} denotes AWGN, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$ and $\mathbf{H} \in \mathbb{R}^{M \times N}$ is an arbitrary (possibly random) matrix. However, if the system matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ in (5.2) is a deterministic orthonormal matrix, i.e., $\mathbf{H}^T \mathbf{H} = \mathbf{I}$ (which implies that $M \geq N$), then we can transform the observation \mathbf{y} in an invertible manner to obtain the modified observation $\mathbf{y}' = \mathbf{H}^T \mathbf{y} = \mathbf{x} + \mathbf{n}'$, where $\mathbf{n}' \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$. The modified observation corresponds then to an SSNM and moreover, by Theorem 2.4.1, is equivalent to the original SLM since the transformation of the observation is invertible. Thus, any SLM with an orthonormal system matrix is, from the viewpoint of minimum variance estimation, completely equivalent to the SSNM (cf. [81, 82]).

The SLM has been considered in [16], where the authors analyze the performance of specific estimation schemes that are based on convex optimization problems. In a minimax estimation context, the SLM has been studied in [48, 49], where the authors derive bounds on the minimax risk (see Section 2.2) and also discuss estimators that come close to these bounds. An asymptotic analysis of minimax estimation for the SSNM is given in the seminal work [22, 23]. In our context, i.e., minimum variance estimation, lower bounds on the minimum achievable estimation variance for the SLM have been studied recently. In particular, the CRB for the SLM was derived and analyzed in [25, 26]. In [81], lower and upper bounds on the minimum achievable variance of unbiased estimators were derived for the SSNM by using techniques that differ from the RKHS approach that will be used in this thesis. A remarkable property of the lower bounds presented in [25] and [81] is the fact that they exhibit a discontinuity when passing from the case $\|\mathbf{x}\|_0 = S$ to the case $\|\mathbf{x}\|_0 < S$.

In this chapter, we use the mathematical framework of RKHS to derive novel lower bounds on the variance of estimators for the SLM. These bounds hold for estimators with a given differentiable bias function. For the special case of the SSNM, we obtain a lower bound for unbiased estimators which is tighter than the bounds in [25, 26, 81] and, moreover, is a continuous function of \mathbf{x} . As we will show, RKHS theory relates the lower bound for the SLM to that obtained for a specific LGM. The RKHS framework has been previously applied to estimation [3, 35] but, to the best of our knowledge, not to the SLM.

This chapter is organized as follows: In Section 5.2, we introduce and discuss the RKHS associated to minimum variance problems that arise from the SLM. Based on this RKHS, we discuss in Section 5.3 some fundamental facts about minimum variance estimation for the SLM. In Section 5.4, we use RKHS theory first to reinterpret existing lower bounds from the geometric RKHS perspective and then to derive novel lower bounds on the variance of estimators for the SLM which have a prescribed bias, or, equivalently, a prescribed mean function. The important special case of the SLM given by the SSNM is discussed in Section 5.5, where we derive closed-form expressions for the minimum achievable variance, i.e., the Barankin bound and for the corresponding LMV estimator. A discussion on the necessity of strict sparsity requirements is presented in Section 5.6. The SLM viewpoint on the CS recovery problem is presented in Section 5.7. Finally, we compare the theoretical bounds with the actual variance behavior of some popular estimation schemes in Section 5.8.

The key results of this chapter have been presented in part in [81–84].

5.2 RKHS Associated with the SLM

We will denote any minimum variance problem that is obtained from \mathcal{E}_{SLM} (5.5) by additionally fixing a prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$ and parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$, by

$$\mathcal{M}_{\text{SLM}} \triangleq (\mathcal{E}_{\text{SLM}}, c(\cdot), \mathbf{x}_0). \quad (5.12)$$

It should be clear from the context which bias function $c(\cdot)$ and fixed parameter vector \mathbf{x}_0 is used within the minimum variance problem \mathcal{M}_{SLM} . Note that the restriction to the specific parameter function $g(\mathbf{x}) = x_k$ in the SLM (5.5) is no real restriction, since we do not constrain the prescribed bias $c(\cdot)$ used for the minimum variance problem \mathcal{M}_{SLM} and, according to Theorem 2.3.1, it is no loss of generality if one fixes either the prescribed bias or the parameter function of a minimum variance problem, but not both.

The estimation problem \mathcal{E}_{SLM} satisfies the requirements of Postulate 4.1.1, since it can be easily verified that $\mathbf{E}_{\mathbf{x}_0} \left\{ \left[\frac{f_{\mathbf{H}}(\mathbf{y}; \mathbf{x})}{f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}_0)} \right]^2 \right\} < \infty$ for any $\mathbf{x} \in \mathcal{X}_S$. Therefore, according to Section 4.3.1,

we can associate with the minimum variance problem \mathcal{M}_{SLM} an RKHS $\mathcal{H}(\mathcal{M}_{\text{SLM}})$ whose kernel $R_{\mathcal{M}_{\text{SLM}}}(\cdot, \cdot) : \mathcal{X}_S \times \mathcal{X}_S \rightarrow \mathbb{R}$ is given as (cf. (4.10)):

$$\begin{aligned}
R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_1, \mathbf{x}_2) &= \mathbf{E}_{\mathbf{x}_0} \{ \rho_{\mathcal{M}_{\text{SLM}}}(\mathbf{y}, \mathbf{x}_1) \rho_{\mathcal{M}_{\text{SLM}}}(\mathbf{y}, \mathbf{x}_2) \} = \mathbf{E}_{\mathbf{x}_0} \left\{ \frac{f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}_1) f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}_2)}{f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}_0) f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}_0)} \right\} \\
&= \int_{\mathbf{y}} \frac{f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}_1) f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}_2)}{f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}_0)} d\mathbf{y} \\
&= \int_{\mathbf{y}} \frac{1}{(2\pi\sigma^2)^{M/2}} \frac{\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}_1\|_2^2\right) \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}_2\|_2^2\right)}{\exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}_0\|_2^2\right)} d\mathbf{y} \\
&= \int_{\mathbf{y}} \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left(-\frac{1}{2\sigma^2}\left[\|\mathbf{y}\|_2^2 - 2\mathbf{y}^T \mathbf{H}(\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{x}_0) + \|\mathbf{H}\mathbf{x}_1\|_2^2 + \|\mathbf{H}\mathbf{x}_2\|_2^2 - \|\mathbf{H}\mathbf{x}_0\|_2^2\right]\right) d\mathbf{y} \\
&= \int_{\mathbf{y}} \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left(-\frac{1}{2\sigma^2}\left[\|\mathbf{y} - \mathbf{H}(\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{x}_0)\|^2 + 2\mathbf{x}_0^T \mathbf{H}^T \mathbf{H}(\mathbf{x}_1 + \mathbf{x}_2) \right. \right. \\
&\quad \left. \left. - 2\mathbf{x}_1^T \mathbf{H}^T \mathbf{H}\mathbf{x}_2 - 2\mathbf{x}_0^T \mathbf{H}^T \mathbf{H}\mathbf{x}_0\right]\right) d\mathbf{y} \\
&= \exp\left(-\frac{1}{2\sigma^2}\left[+2\mathbf{x}_0^T \mathbf{H}^T \mathbf{H}(\mathbf{x}_1 + \mathbf{x}_2) - 2\mathbf{x}_1^T \mathbf{H}^T \mathbf{H}\mathbf{x}_2 - 2\mathbf{x}_0^T \mathbf{H}^T \mathbf{H}\mathbf{x}_0\right]\right) \times \\
&\quad \underbrace{\int_{\mathbf{y}} \frac{1}{(2\pi\sigma^2)^{M/2}} \exp\left(-\frac{1}{2\sigma^2}\left[\|\mathbf{y} - \mathbf{H}(\mathbf{x}_1 + \mathbf{x}_2 - \mathbf{x}_0)\|^2\right]\right) d\mathbf{y}}_{=1} \\
&= \exp\left(\frac{1}{\sigma^2}(\mathbf{x}_1 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H}(\mathbf{x}_2 - \mathbf{x}_0)\right). \tag{5.13}
\end{aligned}$$

Obviously, the kernel $R_{\mathcal{M}_{\text{SLM}}}(\cdot, \cdot)$ is differentiable in the sense of Definition 3.4.1 up to any order $m \in \mathbb{N}$, and therefore Theorem 3.4.1 applies to the RKHS $\mathcal{H}(\mathcal{M}_{\text{SLM}})$ (note that this RKHS consists of functions $f(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$).

Let us now consider the SLM for the specific case $S = N$, for which the SLM coincides with the LGM. We will denote by

$$\mathcal{M}_{\text{LGM}} \triangleq (\mathcal{E}_{\text{LGM}}, \tilde{c}(\cdot), \mathbf{x}_0) \tag{5.14}$$

any minimum variance problem arising from the LGM \mathcal{E}_{LGM} by fixing a prescribed bias function $\tilde{c}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}$ and parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$.² The corresponding RKHS will be denoted by

²Note that we require $\mathbf{x}_0 \in \mathcal{X}_S$ even if the parameter set of the LGM is $\mathbb{R}^N \supseteq \mathcal{X}_S$.

$\mathcal{H}(\mathcal{M}_{\text{LGM}})$, i.e., $\mathcal{H}(\mathcal{M}_{\text{LGM}})$ is the RKHS associated to the kernel $R_{\mathcal{M}_{\text{LGM}}}(\cdot, \cdot) : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$:

$$R_{\mathcal{M}_{\text{LGM}}}(\mathbf{x}_1, \mathbf{x}_2) = \exp\left(\frac{1}{\sigma^2}(\mathbf{x}_1 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H}(\mathbf{x}_2 - \mathbf{x}_0)\right). \quad (5.15)$$

Note that the functional form of $R_{\mathcal{M}_{\text{LGM}}}(\mathbf{x}_1, \mathbf{x}_2)$ is identical with that of $R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_1, \mathbf{x}_2)$ (see (5.13)). However, these two kernel functions differ in their domain, which is $\mathbb{R}^N \times \mathbb{R}^N$ and $\mathcal{X}_S \times \mathcal{X}_S$, respectively.

As can be verified easily, the restriction $\mathcal{M}_{\text{LGM}}|_{\mathcal{X}_S}$ of the minimum variance problem \mathcal{M}_{LGM} given by (5.14) coincides with the minimum variance problem \mathcal{M}_{SLM} given by (5.12) with prescribed bias $c(\cdot) = \tilde{c}(\cdot)|_{\mathcal{X}_S}$. We highlight the fact that we prescribe the estimator bias within minimum variance estimation for the SLM only for the smaller set \mathcal{X}_S , i.e., we require $b(\hat{x}_k(\cdot); \mathbf{x}) = c(\mathbf{x})$ only for $\mathbf{x} \in \mathcal{X}_S \subseteq \mathbb{R}^N$, whereas for the LGM we prescribe the bias for every $\mathbf{x} \in \mathbb{R}^N$.

Theorem 5.2.1. *Consider the minimum variance problem $\mathcal{M}_{\text{LGM}} = (\mathcal{E}_{\text{LGM}}, \tilde{c}(\cdot), \mathbf{x}_0)$ with an arbitrary but fixed choice for σ , M , N , \mathbf{H} , prescribed bias $\tilde{c}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}$, and parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$. We then have that for any sparsity degree $S \in [N]$, the RKHS $\mathcal{H}(\mathcal{M}_{\text{SLM}})$ associated with $\mathcal{M}_{\text{SLM}} = \mathcal{M}_{\text{LGM}}|_{\mathcal{X}_S}$ consists of the restrictions of all functions $f(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}$ belonging to the RKHS $\mathcal{H}(\mathcal{M}_{\text{LGM}})$ to the subdomain $\mathcal{X}_S \subseteq \mathbb{R}^N$, i.e.,*

$$\mathcal{H}(\mathcal{M}_{\text{SLM}}) = \left\{ f(\cdot)|_{\mathcal{X}_S} \mid f(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}}) \right\}. \quad (5.16)$$

Moreover, the norm of any element $f(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SLM}})$ is given by

$$\|f(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} = \min_{\substack{g(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}}) \\ g(\cdot)|_{\mathcal{X}_S} = f(\cdot)}} \|g(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}})} \quad (5.17)$$

Proof. This result is obtained from the application of Theorem 3.3.5 to the specific RKHSs $\mathcal{H}(\mathcal{M}_{\text{SLM}})$ and $\mathcal{H}(\mathcal{M}_{\text{LGM}})$ since the associated kernels obviously satisfy $R_{\mathcal{M}_{\text{SLM}}}(\cdot, \cdot) = R_{\mathcal{M}_{\text{LGM}}}(\cdot, \cdot)|_{\mathcal{X}_S \times \mathcal{X}_S}$. \square

The RKHS $\mathcal{H}(\mathcal{M}_{\text{LGM}})$ associated with the minimum variance problem \mathcal{M}_{LGM} has already been analyzed in [35]. However, we will now present an alternative characterization of the RKHS $\mathcal{H}(\mathcal{M}_{\text{LGM}})$ that is based on a congruence with a specific RKHS that has some pleasing properties.

Theorem 5.2.2. *Consider the RKHS $\mathcal{H}(\mathcal{M}_{\text{LGM}})$ associated with $\mathcal{M}_{\text{LGM}} \triangleq (\mathcal{E}_{\text{LGM}}, \tilde{c}(\cdot), \mathbf{x}_0)$ with an arbitrary system matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$, not necessarily satisfying (5.3), and denote its thin SVD by $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$. Then we have that any function $f(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}})$ is invariant w.r.t. translations by vectors $\mathbf{x}' \in \mathbb{R}^N$ that belong to the null-space of \mathbf{H} , i.e.,*

$$f(\mathbf{x}) = f(\mathbf{x} + \mathbf{x}') \quad (5.18)$$

for any $\mathbf{x}' \in \mathcal{N}(\mathbf{H})$, $\mathbf{x} \in \mathbb{R}^N$. Furthermore, the RKHS $\mathcal{H}(\mathcal{M}_{\text{LGM}})$ is isometric to the RKHS $\mathcal{H}(R_g^{(D)})$ for $D = \text{rank}(\mathbf{H})$, which is defined via the kernel $R_g^{(D)}(\cdot, \cdot) : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$:

$$R_g^{(D)}(\mathbf{x}_1, \mathbf{x}_2) \triangleq \exp(\mathbf{x}_1^T \mathbf{x}_2). \quad (5.19)$$

The mapping $\mathsf{K}_g[\cdot] : \mathcal{H}(R_g^{(D)}) \rightarrow \mathcal{H}(\mathcal{M}_{\text{LGM}})$ given by

$$f(\cdot) \mapsto \tilde{f}(\cdot) = \mathsf{K}_g[f(\cdot)] : \tilde{f}(\mathbf{x}') = f\left(\frac{1}{\sigma} \tilde{\mathbf{H}}^\dagger \mathbf{x}'\right) \exp\left(\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} (\mathbf{x}')^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right), \quad (5.20)$$

where $\tilde{\mathbf{H}} \triangleq \mathbf{V}\Sigma^{-1}$, is a congruence from $\mathcal{H}(R_g^{(D)})$ to $\mathcal{H}(\mathcal{M}_{\text{LGM}})$. The inverse mapping $\mathsf{K}_g^{-1}[\cdot] : \mathcal{H}(\mathcal{M}_{\text{LGM}}) \rightarrow \mathcal{H}(R_g^{(D)})$ given by

$$f(\cdot) \mapsto \tilde{f}(\cdot) = \mathsf{K}_g^{-1}[f(\cdot)] : \tilde{f}(\mathbf{x}') = f(\sigma \tilde{\mathbf{H}} \mathbf{x}') \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 + \frac{1}{\sigma} (\mathbf{x}')^T \tilde{\mathbf{H}}^\dagger \mathbf{x}_0\right), \quad (5.21)$$

is a congruence from $\mathcal{H}(\mathcal{M}_{\text{LGM}})$ to $\mathcal{H}(R_g^{(D)})$.

Proof. Consider the function $f_{\mathbf{x}_1}(\cdot) = R_{\mathcal{M}_{\text{LGM}}}(\cdot, \mathbf{x}_1)$, where $\mathbf{x}_1 \in \mathbb{R}^N$ is arbitrary but fixed. Then, for any $\mathbf{x}' \in \mathcal{N}(\mathbf{H})$ and $\mathbf{x} \in \mathbb{R}^N$,

$$\begin{aligned} f_{\mathbf{x}_1}(\mathbf{x} + \mathbf{x}') &= \exp\left(\frac{1}{\sigma^2} (\mathbf{x} + \mathbf{x}' - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} (\mathbf{x}_1 - \mathbf{x}_0)\right) \\ &= \exp\left(\frac{1}{\sigma^2} (\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} (\mathbf{x}_1 - \mathbf{x}_0) + \underbrace{((\mathbf{x}_1 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} \mathbf{x}')^T}_{=0}\right) \\ &= \exp\left(\frac{1}{\sigma^2} (\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} (\mathbf{x}_1 - \mathbf{x}_0)\right) = f_{\mathbf{x}_1}(\mathbf{x}). \end{aligned} \quad (5.22)$$

From this, it follows straightforwardly that any function $f(\cdot) \in \text{span}\{R_{\mathcal{M}_{\text{LGM}}}(\cdot, \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^N}$ satisfies (5.18). Consider now a general $f(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}})$ and fix an arbitrary $\mathbf{x} \in \mathbb{R}^N$ and $\mathbf{x}' \in \mathcal{N}(\mathbf{H})$. Then, for any $f'(\cdot) \in \text{span}\{R_{\mathcal{M}_{\text{LGM}}}(\cdot, \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^N}$, we have

$$\begin{aligned} f(\mathbf{x} + \mathbf{x}') - f(\mathbf{x}) &= [f(\mathbf{x} + \mathbf{x}') - f'(\mathbf{x} + \mathbf{x}')] - [f(\mathbf{x}) - f'(\mathbf{x})] + f'(\mathbf{x} + \mathbf{x}') - f'(\mathbf{x}) \\ &\stackrel{(a)}{=} [f(\mathbf{x} + \mathbf{x}') - f'(\mathbf{x} + \mathbf{x}')] - [f(\mathbf{x}) - f'(\mathbf{x})] \end{aligned} \quad (5.23)$$

where (a) follows from the fact that $f'(\cdot)$ satisfies (5.18). Since according to Theorem 3.2.2, the linear space $\text{span}\{R_{\mathcal{M}_{\text{LGM}}}(\cdot, \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^N}$ is dense in $\mathcal{H}(\mathcal{M}_{\text{LGM}})$ we can choose for any $\varepsilon > 0$ a function $f'(\cdot) \in \text{span}\{R_{\mathcal{M}_{\text{LGM}}}(\cdot, \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^N}$ such that $\|f'(\cdot) - f(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}})} \leq \varepsilon$ which implies via Theorem 3.1.1 and the reproducing property (3.44) that

$$|f(\mathbf{x} + \mathbf{x}') - f'(\mathbf{x} + \mathbf{x}')| \stackrel{(3.44)}{=} \left| \langle f(\cdot) - f'(\cdot), R_{\mathcal{M}_{\text{LGM}}}(\cdot, \mathbf{x} + \mathbf{x}') \rangle_{\mathcal{H}(\mathcal{M}_{\text{LGM}})} \right|$$

$$\begin{aligned}
& \stackrel{(3.3)}{\leq} \|f'(\cdot) - f(\cdot)\|_{\mathcal{M}_{\text{LGM}}} \|R_{\mathcal{M}_{\text{LGM}}}(\cdot, \mathbf{x} + \mathbf{x}')\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}})} \leq \varepsilon \|R_{\mathcal{M}_{\text{LGM}}}(\cdot, \mathbf{x} + \mathbf{x}')\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}})} \\
& \stackrel{(3.44)}{=} \varepsilon \sqrt{R_{\mathcal{M}_{\text{LGM}}}(\mathbf{x} + \mathbf{x}', \mathbf{x} + \mathbf{x}')}, \tag{5.24}
\end{aligned}$$

and similarly that

$$|f(\mathbf{x}) - f'(\mathbf{x})| \leq \varepsilon \sqrt{R_{\mathcal{M}_{\text{LGM}}}(\mathbf{x}, \mathbf{x})}. \tag{5.25}$$

Thus, we can make the last expression of (5.23) arbitrarily small by a suitable choice for $f'(\cdot)$. This concludes the proof of (5.18).

For the following, we note that based on the thin SVD $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ of \mathbf{H} , we have

$$\mathbf{H}^T \mathbf{H} = \mathbf{V}\mathbf{\Sigma}^2 \mathbf{V}^T = (\tilde{\mathbf{H}}^\dagger)^T \tilde{\mathbf{H}}^\dagger. \tag{5.26}$$

Consider the two sets of functions given by

$$\mathcal{A} \triangleq \left\{ g_{\mathbf{x}}(\cdot) \triangleq \exp\left(\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma} \mathbf{x}^T \mathbf{H}^T \mathbf{H}\mathbf{x}_0\right) R_g^{(D)}(\cdot, \tilde{\mathbf{H}}^\dagger \mathbf{x}) \right\}_{\mathbf{x} \in \mathbb{R}^N} \tag{5.27}$$

and

$$\mathcal{B} \triangleq \{f_{\mathbf{x}}(\cdot) \triangleq R_{\mathcal{M}_{\text{LGM}}}(\cdot, \sigma \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^N}. \tag{5.28}$$

We have obviously that

$$\mathcal{B} = \{R_{\mathcal{M}_{\text{LGM}}}(\cdot, \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^N} \tag{5.29}$$

and in turn

$$\text{span}\{\mathcal{B}\} = \text{span}\{R_{\mathcal{M}_{\text{LGM}}}(\cdot, \mathbf{x})\}_{\mathbf{x} \in \mathbb{R}^N}. \tag{5.30}$$

Furthermore, since for any $\mathbf{x} \in \mathbb{R}^D$ and $\mathbf{x}' \triangleq \mathbf{V}\mathbf{\Sigma}\mathbf{x} \in \mathbb{R}^N$ (note that $\tilde{\mathbf{H}}^\dagger \mathbf{x}' = \mathbf{x}$) we have

$$R_g^{(D)}(\cdot, \tilde{\mathbf{H}}^\dagger \mathbf{x}') = R_g^{(D)}(\cdot, \mathbf{x}), \tag{5.31}$$

it holds that

$$\left\{ R_g^{(D)}(\cdot, \mathbf{x}) \right\}_{\mathbf{x} \in \mathbb{R}^D} \subseteq \left\{ R_g^{(D)}(\cdot, \tilde{\mathbf{H}}^\dagger \mathbf{x}') \right\}_{\mathbf{x}' \in \mathbb{R}^N}. \tag{5.32}$$

Similarly, since any $\mathbf{x}' \in \mathbb{R}^N$ and $\mathbf{x} \triangleq \tilde{\mathbf{H}}^\dagger \mathbf{x}' \in \mathbb{R}^D$ also satisfy (5.31) it holds that

$$\left\{ R_g^{(D)}(\cdot, \tilde{\mathbf{H}}^\dagger \mathbf{x}') \right\}_{\mathbf{x}' \in \mathbb{R}^N} \subseteq \left\{ R_g^{(D)}(\cdot, \mathbf{x}) \right\}_{\mathbf{x} \in \mathbb{R}^D}. \tag{5.33}$$

Combining (5.32) and (5.33) yields

$$\left\{ R_g^{(D)}(\cdot, \mathbf{x}) \right\}_{\mathbf{x} \in \mathbb{R}^D} = \left\{ R_g^{(D)}(\cdot, \tilde{\mathbf{H}}^\dagger \mathbf{x}') \right\}_{\mathbf{x}' \in \mathbb{R}^N}, \tag{5.34}$$

and in turn

$$\text{span}\{\mathcal{A}\} = \text{span}\left\{R_g^{(D)}(\cdot, \mathbf{x})\right\}_{\mathbf{x} \in \mathbb{R}^D}, \quad (5.35)$$

since the weights $\exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}^T(\mathbf{H}^T\mathbf{H})\mathbf{x}_0\right)$ appearing in the definition of \mathcal{A} have no influence on the set $\text{span}\{\mathcal{A}\}$. Thus, we have shown that \mathcal{A} and \mathcal{B} span the two RKHSs $\mathcal{H}(R_g^{(D)})$ and $\mathcal{H}(\mathcal{M}_{\text{LGM}})$, respectively.

For any two vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^N$:

$$\begin{aligned} \langle f_{\mathbf{x}_1}(\cdot), f_{\mathbf{x}_2}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{LGM}})} &= \langle R_{\mathcal{M}_{\text{LGM}}}(\cdot, \sigma\mathbf{x}_1), R_{\mathcal{M}_{\text{LGM}}}(\cdot, \sigma\mathbf{x}_2) \rangle_{\mathcal{H}(\mathcal{M}_{\text{LGM}})} \\ &\stackrel{(3.44)}{=} R_{\mathcal{M}_{\text{LGM}}}(\sigma\mathbf{x}_1, \sigma\mathbf{x}_2) \\ &= \exp\left(\frac{1}{\sigma^2}(\sigma\mathbf{x}_1 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} (\sigma\mathbf{x}_2 - \mathbf{x}_0)\right) \\ &= \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_1^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0 + \mathbf{x}_1^T \mathbf{H}^T \mathbf{H} \mathbf{x}_2 + \frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_2^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) \\ &= \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_1^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) \exp(\mathbf{x}_1^T \mathbf{H}^T \mathbf{H} \mathbf{x}_2) \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_2^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) \\ &\stackrel{(5.26)}{=} \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_1^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) \exp\left(\mathbf{x}_1^T (\tilde{\mathbf{H}}^\dagger)^T \tilde{\mathbf{H}}^\dagger \mathbf{x}_2\right) \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_2^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) \\ &\stackrel{(5.19)}{=} \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_1^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) R_g^{(D)}(\tilde{\mathbf{H}}^\dagger \mathbf{x}_1, \tilde{\mathbf{H}}^\dagger \mathbf{x}_2) \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_2^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) \\ &\stackrel{(3.44)}{=} \left\langle \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_1^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) R_g^{(D)}(\cdot, \tilde{\mathbf{H}}^\dagger \mathbf{x}_1), \right. \\ &\quad \left. \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_2^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) R_g^{(D)}(\cdot, \tilde{\mathbf{H}}^\dagger \mathbf{x}_2) \right\rangle_{\mathcal{H}(R_g^{(D)})} \\ &= \langle g_{\mathbf{x}_1}(\cdot), g_{\mathbf{x}_2}(\cdot) \rangle_{\mathcal{H}(R_g^{(D)})}. \end{aligned} \quad (5.36)$$

Now, for an arbitrary $\mathbf{x} \in \mathbb{R}^N$, let us denote by $h_{\mathbf{x}}(\cdot) \triangleq \mathbf{K}_g[g_{\mathbf{x}}(\cdot)] \in \mathcal{H}(\mathcal{M}_{\text{LGM}})$ the image of the function $g_{\mathbf{x}}(\cdot) \triangleq \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}^T(\mathbf{H}^T\mathbf{H})\mathbf{x}_0\right) R_g^{(D)}(\cdot, \tilde{\mathbf{H}}^\dagger \mathbf{x}) \in \mathcal{H}(R_g^{(D)})$ (see (5.27)) under the mapping $\mathbf{K}_g[\cdot]$ defined in (5.20). We have

$$h_{\mathbf{x}}(\mathbf{x}') = g_{\mathbf{x}}\left(\frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger \mathbf{x}'\right) \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2}(\mathbf{x}')^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right)$$

$$\begin{aligned}
& \stackrel{(5.27)}{=} \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0\right) R_g^{(D)}\left(\frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger\mathbf{x}', \tilde{\mathbf{H}}^\dagger\mathbf{x}\right) \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2}(\mathbf{x}')^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0\right) \\
& \stackrel{(5.19)}{=} \exp\left(\frac{1}{\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0\right) \exp\left(\frac{1}{\sigma}(\mathbf{x}')^T(\tilde{\mathbf{H}}^\dagger)^T\tilde{\mathbf{H}}^\dagger\mathbf{x}\right) \exp\left(-\frac{1}{\sigma^2}(\mathbf{x}')^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0\right) \\
& \stackrel{(5.26)}{=} \exp\left(\frac{1}{\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0 + \frac{1}{\sigma}(\mathbf{x}')^T\mathbf{H}^T\mathbf{H}\mathbf{x} - \frac{1}{\sigma^2}(\mathbf{x}')^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0\right) \\
& = \exp\left(\frac{1}{\sigma^2}(\mathbf{x}' - \mathbf{x}_0)^T\mathbf{H}^T\mathbf{H}(\sigma\mathbf{x} - \mathbf{x}_0)\right) \stackrel{(5.13)}{=} R_{\mathcal{M}_{\text{LGM}}}(\mathbf{x}', \sigma\mathbf{x}) = f_{\mathbf{x}}(\mathbf{x}'). \tag{5.37}
\end{aligned}$$

The fact that the mapping $\mathsf{K}_g[\cdot]$ defined in (5.20) is a congruence from $\mathcal{H}(R_g^{(D)})$ to $\mathcal{H}(\mathcal{M}_{\text{LGM}})$ follows then from Theorem 3.3.4, since for every argument $\mathbf{x} \in \mathbb{R}^N$ the function value $\mathsf{K}_g[g(\cdot)](\mathbf{x})$ depends continuously on the function value $g\left(\frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger\mathbf{x}\right)$, which implies that the image $\mathsf{K}_g[g(\cdot)]$ of a function $g(\cdot) \in \mathcal{H}(R_g^{(D)})$ which is the pointwise limit of a sequence $\{g_l(\cdot) \in \text{span}\{\mathcal{A}\}\}_{l \rightarrow \infty}$ is the pointwise limit of the functions $\mathsf{K}_g[g_l(\cdot)] \in \mathcal{H}(\mathcal{M}_{\text{LGM}})$. \square

In the following, we will make use of the obvious facts summarized in

Lemma 5.2.3. *The kernel $R_g^{(D)}(\cdot, \cdot) : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$ defined in (5.19) is differentiable up to any order $m \in \mathbb{N}$ (see Definition 3.4.1). We have that for any vector $\mathbf{x}_c \in \mathbb{R}^D$, index set $\mathcal{K} \subseteq [D]$, and $\varepsilon > 0$, the ε - \mathcal{K} -neighborhood of \mathbf{x}_c (see (3.62)) belongs to the domain of $R_g^{(D)}(\cdot, \cdot)$, i.e., $\mathcal{N}_{\mathbf{x}_c}^{\mathcal{K}}(\varepsilon) \subseteq \mathbb{R}^D$.*

The RKHS $\mathcal{H}(R_g^{(D)})$ is entirely characterized by

Theorem 5.2.4. *The RKHS $\mathcal{H}(R_g^{(D)})$ is separable and it contains the functions $g^{(\mathbf{p})}(\cdot) : \mathbb{R}^D \rightarrow \mathbb{R}$ given by*

$$g^{(\mathbf{p})}(\mathbf{x}) \triangleq \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} R_g^{(D)}(\mathbf{x}, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}}} \Big|_{\mathbf{x}_2=\mathbf{0}} = \frac{1}{\sqrt{\mathbf{p}!}} \mathbf{x}^{\mathbf{p}}, \tag{5.38}$$

where $\mathbf{p} \in \mathbb{Z}_+^D$ is an arbitrary multi-index, i.e.,

$$\mathbf{p} \in \mathbb{Z}_+^D \Rightarrow g^{(\mathbf{p})}(\cdot) \in \mathcal{H}(R_g^{(D)}), \tag{5.39}$$

with $\mathbf{p}! \triangleq \prod_{l \in [D]} p_l!$ and $\mathbf{x}^{\mathbf{p}} \triangleq \prod_{l \in [D]} (x_l)^{p_l}$. The inner product of an arbitrary function $f(\cdot) \in \mathcal{H}(R_g^{(D)})$ with $g^{(\mathbf{p})}(\cdot)$ is given by

$$\langle f(\cdot), g^{(\mathbf{p})}(\cdot) \rangle_{\mathcal{H}(R_g^{(D)})} = \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}=\mathbf{0}}. \tag{5.40}$$

Moreover, the set $\{g^{(\mathbf{p})}(\cdot) \in \mathcal{H}(R_g^{(D)})\}_{\mathbf{p} \in \mathbb{Z}_+^D}$ forms an ONB for $\mathcal{H}(R_g^{(D)})$. Thus, a function $f(\cdot) : \mathbb{R}^D \rightarrow \mathbb{R}$ belongs to $\mathcal{H}(R_g^{(D)})$, i.e., $f(\cdot) \in \mathcal{H}(R_g^{(D)})$ if and only if it can be written pointwise as

$$f(\mathbf{x}) = \sum_{\mathbf{p} \in \mathbb{Z}_+^D} a[\mathbf{p}]g^{(\mathbf{p})}(\mathbf{x}) = \sum_{\mathbf{p} \in \mathbb{Z}_+^D} a[\mathbf{p}] \frac{1}{\sqrt{\mathbf{p}!}} \mathbf{x}^{\mathbf{p}}, \quad (5.41)$$

with a unique coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$.

Proof. The separability of $\mathcal{H}(R_g^{(D)})$ follows from [66, Theorem 7], since obviously the domain \mathbb{R}^D is separable and the kernel $R_g^{(D)}(\cdot, \cdot)$ is continuous. The validity of (5.39) and (5.40) follows from Lemma 5.2.3 and Theorem 3.4.1. By the power (Taylor) series representation of the exponential function [6], we have that the kernel $R_g^{(D)}(\cdot, \cdot)$ can be written pointwise as

$$R_g^{(D)}(\mathbf{x}_1, \mathbf{x}_2) = \exp(\mathbf{x}_1^T \mathbf{x}_2) = \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{1}{\mathbf{p}!} \mathbf{x}_1^{\mathbf{p}} \mathbf{x}_2^{\mathbf{p}} = \sum_{\mathbf{p} \in \mathbb{Z}_+^D} g^{(\mathbf{p})}(\mathbf{x}_1)g^{(\mathbf{p})}(\mathbf{x}_2). \quad (5.42)$$

This implies via Theorem 3.3.6 that the set $\{g^{(\mathbf{p})}(\cdot) \in \mathcal{H}(R_g^{(D)})\}_{\mathbf{p} \in \mathbb{Z}_+^D}$ forms an ONB for $\mathcal{H}(R_g^{(D)})$, since it can be easily verified by (5.40) that $\langle g^{(\mathbf{p})}(\cdot), g^{(\mathbf{p}')(\cdot)} \rangle_{\mathcal{H}(R_g^{(D)})} = \delta_{\mathbf{p}, \mathbf{p}'}$. The series representation in (5.41) finally follows from Theorem 3.1.10. \square

A useful consequence of Theorem 5.2.4 is obtained by Theorem 3.1.10:

Corollary 5.2.5. *The mapping $\mathcal{K}_a[\cdot] : \ell^2(\mathbb{Z}_+^D) \rightarrow \mathcal{H}(R_g^{(D)})$:*

$$\mathcal{K}_a[a[\mathbf{p}]](\mathbf{x}) = \sum_{\mathbf{p} \in \mathbb{Z}_+^D} a[\mathbf{p}]g^{(\mathbf{p})}(\mathbf{x}) = \sum_{\mathbf{p} \in \mathbb{Z}_+^D} a[\mathbf{p}] \frac{1}{\sqrt{\mathbf{p}!}} \mathbf{x}^{\mathbf{p}}, \quad \mathbf{x} \in \mathbb{R}^D, \quad (5.43)$$

is a congruence from $\ell^2(\mathbb{Z}_+^D)$ to $\mathcal{H}(R_g^{(D)})$.

Proof. The linearity of the mapping $\mathcal{K}_a[\cdot]$ is obvious. Since the functions $\{g^{(\mathbf{p})}(\cdot) \in \mathcal{H}(R_g^{(D)})\}_{\mathbf{p} \in \mathbb{Z}_+^D}$ form an ONB for $\mathcal{H}(R_g^{(D)})$, we have by Theorem 3.1.10 that any function $f(\cdot) \in \mathcal{H}(R_g^{(D)})$ can be written as a sum $\sum_{\mathbf{p} \in \mathbb{Z}_+^D} a[\mathbf{p}] \frac{1}{\sqrt{\mathbf{p}!}} \mathbf{x}^{\mathbf{p}}$, i.e., as the image $\mathcal{K}_a[a[\mathbf{p}]]$ of some coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$. We also have that the mapping preserves inner products, i.e., given two functions $f(\cdot) = \sum_{\mathbf{p} \in \mathbb{Z}_+^D} a[\mathbf{p}]g^{(\mathbf{p})}(\mathbf{x})$, $f'(\cdot) = \sum_{\mathbf{p} \in \mathbb{Z}_+^D} a'[\mathbf{p}]g^{(\mathbf{p})}(\mathbf{x})$ we have

$$\langle f(\cdot), f'(\cdot) \rangle_{\mathcal{H}(R_g^{(D)})} \stackrel{(3.41)}{=} \langle a[\mathbf{p}], a'[\mathbf{p}] \rangle_{\ell^2(\mathbb{Z}_+^D)} = \sum_{\mathbf{p} \in \mathbb{Z}_+^D} a[\mathbf{p}]a'[\mathbf{p}]. \quad (5.44)$$

This relation implies that two different coefficient sequences $a[\mathbf{p}]$ and $a'[\mathbf{p}]$, i.e., $\|a[\mathbf{p}] - a'[\mathbf{p}]\|_{\ell^2(\mathbb{Z}_+^D)} \neq 0$ cannot yield the same image under the mapping $\mathcal{K}_a[\cdot]$. To summarize, we have that the mapping $\mathcal{K}_a[\cdot]$ is linear, inner-product preserving, and bijective, i.e., a congruence (see Definition 3.1.14). \square

A specific class of functions that belong to $\mathcal{H}(R_g^{(D)})$ is presented in

Theorem 5.2.6. Consider a coefficient sequence $a[\mathbf{p}] : \mathbb{Z}_+^D \rightarrow \mathbb{R}$ which satisfies the inequality

$$|a[\mathbf{p}]| \leq C^{|\mathbf{p}|}, \quad (5.45)$$

where $C \in \mathbb{R}_+$ is an arbitrary constant. The RKHS $\mathcal{H}(R_g^{(D)})$ contains any function $f(\cdot) : \mathbb{R}^D \rightarrow \mathbb{R}$ given by

$$f(\mathbf{x}) = \exp(\mathbf{x}^T \mathbf{x}_1) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\mathbf{p}!} \mathbf{x}^{\mathbf{p}}, \quad (5.46)$$

where $\mathbf{x}_1 \in \mathbb{R}^D$ is arbitrary.

Proof. Since $\exp(\mathbf{x}^T \mathbf{x}_1) = \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{\mathbf{x}_1^{\mathbf{p}}}{\mathbf{p}!} \mathbf{x}^{\mathbf{p}}$ and $\sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\mathbf{p}!} \mathbf{x}^{\mathbf{p}}$ are power series that converge everywhere in \mathbb{R}^D , we can by [40] write their product $f(\cdot)$ also as a power series in the form

$$f(\mathbf{x}) = \sum_{\mathbf{p} \in \mathbb{Z}_+^D} c[\mathbf{p}] \mathbf{x}^{\mathbf{p}}, \quad (5.47)$$

where the coefficients $c[\mathbf{p}]$ are obtained as [40]

$$c[\mathbf{p}] = \sum_{\mathbf{n} \leq \mathbf{p}} \frac{1}{\mathbf{n}!(\mathbf{p} - \mathbf{n})!} a[\mathbf{p} - \mathbf{n}] \mathbf{x}_1^{\mathbf{n}}. \quad (5.48)$$

Now we show that the coefficients $d[\mathbf{p}] \triangleq c[\mathbf{p}] \sqrt{\mathbf{p}!}$ are square summable, i.e., $d[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$ which then implies via (5.47) and Theorem 5.2.4 that $f(\cdot) \in \mathcal{H}(R_g^{(D)})$.

Indeed, using (5.45), (5.48) and the fact that for any $\mathbf{x} \in \mathbb{R}^D$ and multi-index $\mathbf{p} \in \mathbb{Z}_+^D$ it holds that

$$\mathbf{x}^{\mathbf{p}} = \prod_{l \in [D]} x_l^{p_l} \leq \prod_{l \in [D]} \|\mathbf{x}\|_{\infty}^{p_l} = \|\mathbf{x}\|_{\infty}^{|\mathbf{p}|}, \quad (5.49)$$

we obtain the bound

$$\begin{aligned} |d[\mathbf{p}]| &= \left| \sum_{\mathbf{n} \leq \mathbf{p}} \frac{\sqrt{\mathbf{p}!}}{\mathbf{n}!(\mathbf{p} - \mathbf{n})!} a[\mathbf{p} - \mathbf{n}] \mathbf{x}_1^{\mathbf{n}} \right| \leq \sum_{\mathbf{n} \leq \mathbf{p}} \frac{\sqrt{\mathbf{p}!}}{\mathbf{n}!(\mathbf{p} - \mathbf{n})!} \left| a[\mathbf{p} - \mathbf{n}] \mathbf{x}_1^{\mathbf{n}} \right| \\ &= \frac{1}{\sqrt{\mathbf{p}!}} \sum_{\mathbf{n} \leq \mathbf{p}} \frac{\mathbf{p}!}{\mathbf{n}!(\mathbf{p} - \mathbf{n})!} \left| a[\mathbf{p} - \mathbf{n}] \mathbf{x}_1^{\mathbf{n}} \right| \stackrel{(5.49)}{\leq} \frac{1}{\sqrt{\mathbf{p}!}} \sum_{\mathbf{n} \leq \mathbf{p}} \frac{\mathbf{p}!}{\mathbf{n}!(\mathbf{p} - \mathbf{n})!} |a[\mathbf{p} - \mathbf{n}]| \|\mathbf{x}_1\|_{\infty}^{|\mathbf{n}|} \\ &\leq \frac{1}{\sqrt{\mathbf{p}!}} \sum_{\mathbf{n} \leq \mathbf{p}} \frac{\mathbf{p}!}{\mathbf{n}!(\mathbf{p} - \mathbf{n})!} C^{|\mathbf{p} - \mathbf{n}|} \|\mathbf{x}_1\|_{\infty}^{|\mathbf{n}|} = \frac{1}{\sqrt{\mathbf{p}!}} \sum_{\mathbf{n} \leq \mathbf{p}} \prod_{l \in [D]} \frac{p_l!}{n_l!(p_l - n_l)!} C^{p_l - n_l} \|\mathbf{x}_1\|_{\infty}^{n_l} \\ &= \frac{1}{\sqrt{\mathbf{p}!}} \sum_{\mathbf{n} \leq \mathbf{p}} \prod_{l \in [D]} \binom{p_l}{n_l} C^{p_l - n_l} \|\mathbf{x}_1\|_{\infty}^{n_l} = \frac{1}{\sqrt{\mathbf{p}!}} \prod_{l \in [D]} (C + \|\mathbf{x}_1\|_{\infty})^{p_l} \end{aligned}$$

$$= \frac{1}{\sqrt{\mathbf{p}!}} (C + \|\mathbf{x}_1\|_\infty)^{|\mathbf{p}|}, \quad (5.50)$$

which implies that

$$(d[\mathbf{p}])^2 \leq \frac{1}{\mathbf{p}!} (C + \|\mathbf{x}_1\|_\infty)^{2|\mathbf{p}|}. \quad (5.51)$$

By the ratio test [5, p. 66]) we have that

$$A \triangleq \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{1}{\mathbf{p}!} (C + \|\mathbf{x}_1\|_\infty)^{2|\mathbf{p}|} < \infty, \quad (5.52)$$

since

$$\lim_{p_l \rightarrow \infty} \frac{\frac{1}{(p_l+1)!} (C + \|\mathbf{x}_1\|_\infty)^{2(p_l+1)}}{\frac{1}{p_l!} (C + \|\mathbf{x}_1\|_\infty)^{2p_l}} = \lim_{p_l \rightarrow \infty} \frac{1}{p_l + 1} (C + \|\mathbf{x}_1\|_\infty)^2 = 0. \quad (5.53)$$

Now, consider an arbitrary finite index set $\mathcal{T} \subseteq \mathbb{Z}_+^D$ and denote the largest entry of any element of \mathcal{T} by $k(\mathcal{T}) \triangleq \max_{\mathbf{p}' \in \mathcal{T}} \|\mathbf{p}'\|_\infty$. Given the index set \mathcal{T} , we define a larger index set \mathcal{T}' , i.e.,

$$\mathcal{T}' \triangleq \{\mathbf{p} \in \mathbb{Z}_+^D \mid p_l \leq k(\mathcal{T}) \text{ for every } l \in [D]\}, \quad (5.54)$$

which obviously contains \mathcal{T} , i.e., $\mathcal{T} \subseteq \mathcal{T}'$. Then, we can bound the sum $\sum_{\mathbf{p} \in \mathcal{T}} (d[\mathbf{p}])^2$ as

$$\begin{aligned} \sum_{\mathbf{p} \in \mathcal{T}} (d[\mathbf{p}])^2 &\leq \sum_{\mathbf{p} \in \mathcal{T}'} (d[\mathbf{p}])^2 \stackrel{(5.51)}{\leq} \sum_{\mathbf{p} \in \mathcal{T}'} \frac{1}{\mathbf{p}!} (C + \|\mathbf{x}_1\|_\infty)^{2|\mathbf{p}|} = \prod_{l \in [D]} \sum_{p_l \in [k(\mathcal{T})]} \frac{1}{p_l!} (C + \|\mathbf{x}_1\|_\infty)^{2p_l} \\ &\leq \prod_{l \in [D]} \sum_{p_l \in \mathbb{Z}_+} \frac{1}{p_l!} (C + \|\mathbf{x}_1\|_\infty)^{2p_l} \stackrel{(5.52)}{=} A^D, \end{aligned} \quad (5.55)$$

i.e., we can upper bound any finite sum $\sum_{\mathbf{p} \in \mathcal{T}} (d[\mathbf{p}])^2$ by the finite quantity A^D (which does not depend on the index set \mathcal{T}). Thus, we have verified that $d[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$. \square

Note that Theorem 5.2.6 trivially implies that $\mathcal{H}(R_g^{(D)})$ contains any function $f(\cdot) : \mathbb{R}^D \rightarrow \mathbb{R}$ that can be written as

$$f(\mathbf{x}) = \exp(\mathbf{x}^T \mathbf{x}_1) \sum_{l \in [L]} a_l \mathbf{x}^{p_l}, \quad (5.56)$$

with arbitrary $\mathbf{x}_1 \in \mathbb{R}^D$, $L \in \mathbb{N}$, $a_l \in \mathbb{R}$, and $\mathbf{p}_l \in \mathbb{Z}_+^D$.

We will also make use of

Theorem 5.2.7. *For an arbitrary parameter vector $\mathbf{x}_c \in \mathbb{R}^D$ and multi-index $\mathbf{p} \in \mathbb{Z}_+^D$, consider the function $g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot) : \mathbb{R}^D \rightarrow \mathbb{R}$ given by*

$$g_{\mathbf{x}_c}^{(\mathbf{p})}(\mathbf{x}) \triangleq \left. \frac{\partial^{\mathbf{p}} R_g^{(D)}(\mathbf{x}, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}}} \right|_{\mathbf{x}_2 = \mathbf{x}_c}. \quad (5.57)$$

Then $g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)$ belongs to the RKHS $\mathcal{H}(R_g^{(D)})$, i.e.,

$$\mathbf{p} \in \mathbb{Z}_+^D, \mathbf{x}_c \in \mathbb{R}^D \quad \Rightarrow \quad g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot) \in \mathcal{H}(R_g^{(D)}). \quad (5.58)$$

The inner product of an arbitrary function $f(\cdot) \in \mathcal{H}(R_g^{(D)})$ with $g_{\mathbf{x}_c}^{(\mathbf{p})}(\mathbf{x})$ is given by

$$\langle f(\cdot), g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot) \rangle_{\mathcal{H}(R_g^{(D)})} = \left. \frac{\partial^{\mathbf{p}} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \right|_{\mathbf{x}=\mathbf{x}_c}. \quad (5.59)$$

Moreover, the set $\{g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)\}_{\mathbf{p} \in \mathbb{Z}_+^D}$ is complete for $\mathcal{H}(R_g^{(D)})$, which implies in particular that if two functions $f(\cdot), f'(\cdot) \in \mathcal{H}(R_g^{(D)})$ take on the same values on $\mathcal{B}(\mathbf{x}_c, r) \subseteq \mathbb{R}^D$ with an arbitrary radius $r > 0$, then they are identical, i.e., $f(\cdot) = f'(\cdot)$.

Proof. The validity of (5.58) and (5.59) follows from Theorem 3.4.1 by the fact that $R_g^{(D)}(\cdot, \cdot)$ is a differentiable kernel. To prove the completeness (see Definition 3.1.7) of the set $\{g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)\}_{\mathbf{p} \in \mathbb{Z}_+^D}$, we consider an arbitrary function $f(\cdot) \in \mathcal{H}(R_g^{(D)})$ and note that by (5.41) in Theorem 5.2.4, it can be represented as a power series which converges everywhere in \mathbb{R}^D . According to [40, Proposition 1.2.3], we can then represent the function as

$$f(\mathbf{x}) = \sum_{\mathbf{p}: |\mathbf{p}| \leq L} \left. \frac{\partial^{\mathbf{p}} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \right|_{\mathbf{x}=\mathbf{x}_c} \frac{1}{\mathbf{p}!} (\mathbf{x} - \mathbf{x}_c)^{\mathbf{p}} + R_L(\mathbf{x}), \quad (5.60)$$

where the remainder term $R_L(\mathbf{x})$ is such that $\lim_{L \rightarrow \infty} |R_L(\mathbf{x})| = 0$ for any $\mathbf{x} \in \mathbb{R}^D$. By using (5.59), we can rewrite (5.60) as

$$f(\mathbf{x}) = \sum_{\mathbf{p}: |\mathbf{p}| \leq L} \frac{\langle f(\cdot), g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot) \rangle_{\mathcal{H}(R_g^{(D)})}}{\mathbf{p}!} (\mathbf{x} - \mathbf{x}_c)^{\mathbf{p}} + R_L(\mathbf{x}). \quad (5.61)$$

From this, we conclude that if $\langle f(\cdot), g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot) \rangle_{\mathcal{H}(R_g^{(D)})} = 0$ for all $\mathbf{p} \in \mathbb{Z}_+^D$, the function $f(\cdot)$ necessarily has to be zero for all arguments, i.e., $f(\cdot) \equiv 0$. Thus, according to Definition 3.1.7, $\{g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)\}_{\mathbf{p} \in \mathbb{Z}_+^D}$ is complete for $\mathcal{H}(R_g^{(D)})$.

Finally, consider two functions $f(\cdot), f'(\cdot) \in \mathcal{H}(R_g^{(D)})$ that take on the same values on $\mathcal{B}(\mathbf{x}_c, r) \subseteq \mathbb{R}^D$ with an arbitrary radius $r > 0$, i.e., the difference $f(\cdot) - f'(\cdot)$ is zero for all $\mathbf{x} \in \mathcal{B}(\mathbf{x}_c, r)$ and in turn the partial derivative $\left. \frac{\partial^{\mathbf{p}} [f(\mathbf{x}) - f'(\mathbf{x})]}{\partial \mathbf{x}^{\mathbf{p}}} \right|_{\mathbf{x}=\mathbf{x}_c}$ is zero for every multi-index \mathbf{p} . However, by (5.59), this means that the inner product $\langle f(\cdot) - f'(\cdot), g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot) \rangle_{\mathcal{H}(R_g^{(D)})}$ is equal to zero for every $\mathbf{p} \in \mathbb{Z}_+^D$. Since the set $\{g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)\}_{\mathbf{p} \in \mathbb{Z}_+^D}$ is complete for $\mathcal{H}(R_g^{(D)})$, this implies that $f(\cdot) - f'(\cdot) \equiv 0$, i.e., $f(\cdot) = f'(\cdot)$. □

5.3 Minimum Variance Estimation for the SLM

In the next section, we will derive lower bounds on the minimum achievable variance $L_{\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)}$ for the class of minimum variance problems $\{\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)\}_{\mathbf{x}_0 \in \mathcal{X}_S}$. Here,

$$\mathcal{M}_{\text{SLM}}(\mathbf{x}_0) \triangleq (\mathcal{E}_{\text{SLM}}, c(\cdot), \mathbf{x}_0), \quad (5.62)$$

for a fixed choice of the SLM parameters σ , S , M , N , \mathbf{H} and prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$. For the special case $S = N$, i.e., where the SLM coincides with the LGM,

$$\mathcal{M}_{\text{LGM}}(\mathbf{x}_0) \triangleq (\mathcal{E}_{\text{LGM}}, \tilde{c}(\cdot), \mathbf{x}_0). \quad (5.63)$$

In this section, we present some fundamental properties of $L_{\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)}$.

The following statement gives a detailed characterization of the class of valid bias functions for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$ and moreover states explicit expressions of the minimum achievable variance and corresponding LMV estimator for a valid bias function.

Theorem 5.3.1. *Consider the SLM $\mathcal{E}_{\text{SLM}} = (\mathcal{X}_S, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k)$ with specific values of k , S , M , N , $\mathbf{H} \in \mathbb{R}^{M \times N}$, and denote the thin SVD of \mathbf{H} by $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$.*

1. *For a specific choice for $\mathbf{x}_0 \in \mathcal{X}_S$, a prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$ is valid for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0) \triangleq (\mathcal{E}_{\text{SLM}}, c(\cdot), \mathbf{x}_0)$, i.e., $L_{\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)} < \infty$, if and only if it can be written as*

$$c(\mathbf{x}) = \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2}\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0\right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{1}{\sqrt{\mathbf{p}!}} a[\mathbf{p}] \left(\frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger \mathbf{x}\right)^{\mathbf{p}} - x_k, \quad \mathbf{x} \in \mathcal{X}_S, \quad (5.64)$$

with a suitable coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$, $D = \text{rank}(\mathbf{H})$, and $\tilde{\mathbf{H}} \triangleq \mathbf{V}\mathbf{\Sigma}^{-1}$. If $c(\cdot)$ is valid for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$, then it is necessarily continuous.

2. *The minimum achievable variance for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$ is obtained in terms of the minimum squared $\ell^2(\mathbb{Z}_+^D)$ norm $\|a[\cdot]\|_{\ell^2(\mathbb{Z}_+^D)} \triangleq \sqrt{\sum_{\mathbf{p} \in \mathbb{Z}_+^D} (a[\mathbf{p}])^2}$ among all coefficient sequences $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$ that are consistent with (5.64), i.e.,*

$$L_{\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)} = \min_{\substack{a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D) \\ a[\mathbf{p}] \text{ consistent with (5.64)}}} \|a[\cdot]\|_{\ell^2(\mathbb{Z}_+^D)}^2 - [c(\mathbf{x}_0) + x_{0,k}]^2. \quad (5.65)$$

3. *Furthermore, given any valid bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$, we have that for every coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$ which is consistent with (5.64), the estimator $\hat{g}(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}$ given by*

$$\hat{g}(\cdot) \triangleq \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2\right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} [\rho_{\mathcal{M}_{\text{LGM}}(\mathbf{x}_0)}(\cdot, \sigma\tilde{\mathbf{H}}\mathbf{x}) \exp(\frac{1}{\sigma}\mathbf{x}_0^T\mathbf{H}^T\mathbf{H}\tilde{\mathbf{H}}\mathbf{x})]}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}=\mathbf{0}} \quad (5.66)$$

is an allowed estimator for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$, i.e., $\hat{g}(\cdot) \in \mathcal{F}(\mathcal{M}_{\text{SLM}}(\mathbf{x}_0))$.

4. Finally, the estimator obtained from (5.66) by the specific coefficient sequence $a_0[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$ that has the minimum $\ell^2(\mathbb{Z}_+^D)$ norm $\|a[\cdot]\|_{\ell^2(\mathbb{Z}_+^D)}$ among all coefficient sequences that are consistent with (5.64), is the LMV estimator for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$.

Proof. 1. By Theorem 4.3.4, we have that the prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$ is valid for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$ if and only if $c(\mathbf{x}) + x_k \in \mathcal{H}(\mathcal{M}_{\text{SLM}}(\mathbf{x}_0))$ which by Theorem 5.2.1 is the case if and only if there exists a function $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0))$ such that $\gamma(\mathbf{x}) = c(\mathbf{x}) + x_k$ for every $\mathbf{x} \in \mathcal{X}_S$. Furthermore, by Theorem 5.2.2, $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0))$ if and only if the function $\gamma(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}$ is the image under the congruence $\mathbf{K}_g[f(\cdot)]$ (see (5.20)) of a function $f(\cdot) \in \mathcal{H}(R_g^{(D)})$ (with $D = \text{rank}(\mathbf{H})$), i.e.,

$$\gamma(\mathbf{x}) = f\left(\frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger\mathbf{x}'\right) \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2}(\mathbf{x}')^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0\right), \quad (5.67)$$

with a suitable function $f(\cdot) \in \mathcal{H}(R_g^{(D)})$. By Theorem 5.2.4, a function $f(\cdot) : \mathbb{R}^D \rightarrow \mathbb{R}$ belongs to $\mathcal{H}(R_g^{(D)})$ if and only if it can be written as (5.41) with a unique coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$. The condition (5.64) for a bias function to be valid follows then by combining (5.67) with (5.41).

The continuity of any valid bias function $c(\cdot)$ can be verified by Theorem 3.4.3, since the kernel $R_{\mathcal{M}_{\text{SLM}}}$ (see (5.13)) is obviously continuous.

2. Given a valid bias function $c(\mathbf{x})$, and using the shorthand $\tilde{\gamma}(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \tilde{\gamma}(\mathbf{x}) = c(\mathbf{x}) + x_k$, the minimum achievable variance is then obtained by Theorem 4.3.4 and (5.17) in Theorem 5.2.1 as

$$\begin{aligned} L_{\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)} &\stackrel{(4.19)}{=} \|\tilde{\gamma}(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SLM}}(\mathbf{x}_0))}^2 - [\tilde{\gamma}(\mathbf{x}_0)]^2 \\ &\stackrel{(5.17)}{=} \min_{\substack{g(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0)) \\ g(\cdot)|_{\mathcal{X}_S} = \tilde{\gamma}(\cdot)}} \|g(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0))}^2 - [\tilde{\gamma}(\mathbf{x}_0)]^2 \end{aligned} \quad (5.68)$$

Using Theorem 5.2.2 and Corollary 5.2.5, we can rewrite (5.68) as

$$\begin{aligned} L_{\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)} &= \min_{\substack{g(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0)) \\ g(\cdot)|_{\mathcal{X}_S} = \tilde{\gamma}(\cdot)}} \|g(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0))}^2 - [\tilde{\gamma}(\mathbf{x}_0)]^2 \\ &= \min_{\substack{f(\cdot) \in \mathcal{H}(R_g^{(D)}) \\ \mathbf{K}_g[f(\cdot)]|_{\mathcal{X}_S} = \tilde{\gamma}(\cdot)}} \left\| \mathbf{K}_g[f(\cdot)] \right\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0))}^2 - [\tilde{\gamma}(\mathbf{x}_0)]^2 \end{aligned}$$

$$\begin{aligned}
&= \min_{\substack{a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D) \\ \mathcal{K}_g[\mathcal{K}_a[a[\mathbf{p}]]]|_{\mathcal{X}_S} = \tilde{\gamma}(\cdot)}} \left\| \mathcal{K}_g[\mathcal{K}_a[a[\mathbf{p}]]] \right\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0))}^2 - [\tilde{\gamma}(\mathbf{x}_0)]^2 \\
&= \min_{\substack{a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D) \\ \mathcal{K}_g[\mathcal{K}_a[a[\mathbf{p}]]]|_{\mathcal{X}_S} = \tilde{\gamma}(\cdot)}} \|a[\mathbf{p}]\|_{\ell^2(\mathbb{Z}_+^D)}^2 - [c(\mathbf{x}_0) + x_{0,k}]^2. \tag{5.69}
\end{aligned}$$

Since

$$\begin{aligned}
\mathcal{K}_g[\mathcal{K}_a[a[\mathbf{p}]]] &\stackrel{(5.43)}{=} \mathcal{K}_g \left[\sum_{\mathbf{p} \in \mathbb{Z}_+^N} \frac{1}{\sqrt{\mathbf{p}!}} a[\mathbf{p}] \mathbf{x}^{\mathbf{p}} \right] \\
&\stackrel{(5.20)}{=} \exp \left(\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0 \right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{1}{\sqrt{\mathbf{p}!}} a[\mathbf{p}] \left(\frac{1}{\sigma} \tilde{\mathbf{H}}^\dagger \mathbf{x} \right)^{\mathbf{p}}, \tag{5.70}
\end{aligned}$$

we have that a coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$ satisfies

$$\mathcal{K}_g[\mathcal{K}_a[a[\mathbf{p}]]]|_{\mathcal{X}_S} = \tilde{\gamma}(\mathbf{x}) = c(\mathbf{x}) + x_k \tag{5.71}$$

if and only if it is consistent with (5.64). Based on this observation, the relation (5.65) follows then from (5.69).

3. Now consider a coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$ which is consistent (5.64). This implies that the function

$$\tilde{\gamma}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R} : \tilde{\gamma}(\mathbf{x}) \triangleq \exp \left(\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0 \right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{1}{\sqrt{\mathbf{p}!}} a[\mathbf{p}] \left(\frac{1}{\sigma} \tilde{\mathbf{H}}^\dagger \mathbf{x} \right)^{\mathbf{p}} \tag{5.72}$$

satisfies

$$\tilde{\gamma}(\cdot)|_{\mathcal{X}_S} = c(\mathbf{x}) + x_k. \tag{5.73}$$

Since by (5.70)

$$\tilde{\gamma}(\cdot) = \mathcal{K}_g[\mathcal{K}_a[a[\mathbf{p}]]], \tag{5.74}$$

we have that the function $\tilde{\gamma}(\cdot)$ belongs to $\mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0))$, i.e., we can define the estimator

$$\hat{g}^{a[\mathbf{p}]}(\cdot) \triangleq \mathcal{J}[\tilde{\gamma}(\cdot)] = \mathcal{J}[\mathcal{K}_g[\mathcal{K}_a[a[\mathbf{p}]]]] \tag{5.75}$$

as the image of $\tilde{\gamma}(\cdot)$ under the congruence $\mathcal{J}[\cdot] : \mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0)) \rightarrow \mathcal{L}(\mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0)))$ defined in (4.13). By Theorem 4.3.2 and Theorem 4.3.4, the estimator $\hat{g}^{a[\mathbf{p}]}(\cdot)$ has variance at \mathbf{x}_0 equal to

$$v(\hat{g}^{a[\mathbf{p}]}(\cdot); \mathbf{x}_0) = \|\tilde{\gamma}(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0))}^2 - [\tilde{\gamma}(\mathbf{x}_0)]^2$$

$$= \|\mathbf{K}_g[\mathbf{K}_a[a[\mathbf{p}]]]\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}}(\mathbf{x}_0))}^2 - [\bar{\gamma}(\mathbf{x}_0)]^2 = \|a[\mathbf{p}]\|_{\ell^2(\mathbb{Z}_+^D)}^2 - [\bar{\gamma}(\mathbf{x}_0)]^2 < \infty \quad (5.76)$$

and mean function equal to $\bar{\gamma}(\mathbf{x})$. In particular, the bias of $\hat{g}^{a[\mathbf{p}]}(\cdot)$ is equal to $c(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{X}_S$. This implies that the estimator $\hat{g}^{a[\mathbf{p}]}(\cdot)$ is an allowed estimator for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$, i.e., $\hat{g}^{a[\mathbf{p}]}(\cdot) \in \mathcal{F}(\mathcal{M}_{\text{SLM}}(\mathbf{x}_0))$

Let us now show that the estimator $\hat{g}^{a[\mathbf{p}]}(\cdot)$ coincides with the estimator (5.66), when using the same coefficient sequence $a[\mathbf{p}]$. Based on the identity

$$\tilde{\mathbf{H}}^T \mathbf{H}^T \mathbf{H} = \Sigma^{-1} \mathbf{V}^T \mathbf{V} \Sigma^2 \mathbf{V}^T = \Sigma \mathbf{V}^T = \tilde{\mathbf{H}}^\dagger, \quad (5.77)$$

we observe that

$$\begin{aligned} & \exp\left(\frac{1}{\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) \left(\frac{1}{\sigma} \tilde{\mathbf{H}}^\dagger \mathbf{x}\right)^{\mathbf{p}} \\ &= \exp\left(\frac{1}{\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) \frac{\partial^{\mathbf{p}} \exp\left(\frac{1}{\sigma} (\mathbf{x}')^T \tilde{\mathbf{H}}^\dagger \mathbf{x}\right)}{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \\ &\stackrel{(5.77)}{=} \exp\left(\frac{1}{\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0\right) \frac{\partial^{\mathbf{p}} \exp\left(\frac{1}{\sigma} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \tilde{\mathbf{H}} \mathbf{x}'\right)}{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \\ &= \frac{\partial^{\mathbf{p}} \exp\left(\frac{1}{\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \mathbf{x}_0 + \frac{1}{\sigma} \mathbf{x}^T \mathbf{H}^T \mathbf{H} \tilde{\mathbf{H}} \mathbf{x}'\right)}{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \\ &= \frac{\partial^{\mathbf{p}} \exp\left(\frac{1}{\sigma^2} (\mathbf{x} - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} (\sigma \tilde{\mathbf{H}} \mathbf{x}' - \mathbf{x}_0) + \frac{1}{\sigma} \mathbf{x}_0^T \mathbf{H}^T \mathbf{H} \tilde{\mathbf{H}} \mathbf{x}'\right)}{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \\ &\stackrel{(5.15)}{=} \frac{\partial^{\mathbf{p}} [R_{\mathcal{M}_{\text{LGM}}(\mathbf{x}_0)}(\mathbf{x}, \sigma \tilde{\mathbf{H}} \mathbf{x}') \exp\left(\frac{1}{\sigma} \mathbf{x}_0^T \mathbf{H}^T \mathbf{H} \tilde{\mathbf{H}} \mathbf{x}'\right)]}{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}}. \end{aligned} \quad (5.78)$$

This allows us to rewrite (5.72) as

$$\bar{\gamma}(\cdot) = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2\right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}}!} \frac{\partial^{\mathbf{p}} \left(R_{\text{LGM}}(\mathbf{x}, \sigma \tilde{\mathbf{H}} \mathbf{x}') \exp\left(\frac{1}{\sigma} \mathbf{x}_0^T \mathbf{H}^T \mathbf{H} \tilde{\mathbf{H}} \mathbf{x}'\right) \right)}{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}}, \quad (5.79)$$

from which it follows by an application of Theorem 4.3.3 that

$$\hat{g}^{a[\mathbf{p}]}(\cdot) = \mathbf{J}[\bar{\gamma}(\cdot)]$$

$$\begin{aligned}
&= \mathbb{J} \left[\exp \left(-\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 \right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} \left(R_{\text{LGM}}(\mathbf{x}, \sigma \tilde{\mathbf{H}}\mathbf{x}') \exp \left(\frac{1}{\sigma} \mathbf{x}_0^T \mathbf{H}^T \mathbf{H} \tilde{\mathbf{H}}\mathbf{x}' \right) \right)}{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \right] \\
&= \exp \left(-\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 \right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \mathbb{J} \left[\frac{\partial^{\mathbf{p}} \left(R_{\text{LGM}}(\mathbf{x}, \sigma \tilde{\mathbf{H}}\mathbf{x}') \exp \left(\frac{1}{\sigma} \mathbf{x}_0^T \mathbf{H}^T \mathbf{H} \tilde{\mathbf{H}}\mathbf{x}' \right) \right)}{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \right] \\
&= \left(-\frac{1}{2\sigma^2} \|\mathbf{H}\mathbf{x}_0\|_2^2 \right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} [\rho_{\mathcal{M}_{\text{LGM}}(\mathbf{x}_0)}(\cdot, \sigma \tilde{\mathbf{H}}\mathbf{x}) \exp \left(\frac{1}{\sigma} \mathbf{x}_0^T \mathbf{H}^T \mathbf{H} \tilde{\mathbf{H}}\mathbf{x} \right)]}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}=\mathbf{0}}, \quad (5.80)
\end{aligned}$$

which coincides with (5.66).

4. Note that by definition, the coefficient sequence $a_0[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^D)$ is a minimizer of (5.69). Since the estimator (5.66) obtained for the coefficients $a_0[\mathbf{p}]$ coincides with the estimator $\hat{g}^{a_0[\mathbf{p}]}(\cdot) = \mathbb{J}[\mathbf{K}_g[\mathbf{K}_a[a_0[\mathbf{p}]]]]$, which is an allowed estimator for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$ whose variance at \mathbf{x}_0 equals (cf. (5.76))

$$\begin{aligned}
\|a_0[\mathbf{p}]\|_{\ell^2(\mathbb{Z}_+^D)}^2 - [\bar{\gamma}(\mathbf{x}_0)]^2 &= \|a_0[\mathbf{p}]\|_{\ell^2(\mathbb{Z}_+^D)}^2 - [c(\mathbf{x}_0) + x_{0,k}]^2 \\
&\stackrel{(5.69)}{=} L_{\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)}. \quad (5.81)
\end{aligned}$$

□

One specific implication of Theorem 5.3.1 concerning the problem of sparsity pattern detection [85–87] is stated in

Theorem 5.3.2. *There exists no estimator $\hat{g}(\mathbf{y})$ of an injective real-valued function $g(\text{supp}(\mathbf{x}))$ of the support of $\mathbf{x} \in \mathcal{X}_S$ which uses only the observation (5.2) of the SLM \mathcal{E}_{SLM} , is unbiased for every $\mathbf{x} \in \mathcal{X}_S$, and has a finite variance at any $\mathbf{x}_0 \in \mathcal{X}_S$.*

Proof. Consider the SLM \mathcal{E}_{SLM} , but instead of estimating the parameter function $g(\mathbf{x}) = x_k$ (for some fixed index $k \in [N]$), we are interested in unbiased estimation of an injective real-valued function $g(\text{supp}(\mathbf{x}))$ of the support $\text{supp}(\mathbf{x})$. According to Section 2.3.3, this is equivalent to a minimum variance problem associated with the ordinary SLM \mathcal{E}_{SLM} , i.e., with the parameter function $g(\mathbf{x}) = x_k$, and using the prescribed bias function $c'(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : c'(\mathbf{x}) = g(\text{supp}(\mathbf{x})) - x_k$, which is obviously discontinuous. Indeed, consider the set $\{\mathbf{x}^{(a)} \triangleq a\mathbf{e}_1\}_{a \in \mathbb{R}} \subseteq \mathcal{X}_S$ consisting of parameter vectors of the form $\mathbf{x}^{(a)} = (a, 0, \dots, 0)^T$, i.e., with exactly one nonzero entry whose value is equal to the number $a \in \mathbb{R}$. We then have $\lim_{a \rightarrow 0} c'(\mathbf{x}^{(a)}) = g(\{1\})$ but $c'(\mathbf{x}^{(0)}) = g(\emptyset)$, where $g(\emptyset) \neq g(\{1\})$ since $g(\text{supp}(\mathbf{x}))$ is assumed to be injective.

However, since the prescribed bias function $c'(\cdot)$ is discontinuous, we have by Theorem 5.3.1 that this bias function is not valid for any minimum variance problem $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$ with $\mathbf{x}_0 \in \mathcal{X}_S$, i.e., there exists no estimator with the bias $c'(\cdot)$ and finite variance at any point $\mathbf{x}_0 \in \mathcal{X}_S$. Thus, by Theorem 2.3.2, we have that the parameter function $g(\text{supp}(\mathbf{x}))$ is not estimable for the SLM \mathcal{E}_{SLM} at any $\mathbf{x}_0 \in \mathcal{X}_S$, which is just an equivalent formulation of the statement to be proved. \square

It is important to note that Theorem 5.3.2 asserts the impossibility of a finite variance estimator of the support of the parameter vector \mathbf{x} which is unbiased for all $\mathbf{x} \in \mathcal{X}_S$. However, even if it might seem at first sight, Theorem 5.3.2 does not contradict [85, 86] since the authors of [85, 86] intentionally reduce the parameter set \mathcal{X}_S of the SLM to a smaller parameter set \mathcal{X}' which includes only vectors \mathbf{x} with exactly S non-zeros, i.e., $\|\mathbf{x}\|_0 = S$, and whose smallest nonzero magnitude is not smaller than a given threshold θ_{\min} , i.e., $\min_{k \in \text{supp}(\mathbf{x})} |x_k| \geq \theta_{\min}$.

The shape of $L_{\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)}$, viewed as a function of $\mathbf{x}_0 \in \mathcal{X}_S$, is characterized by

Theorem 5.3.3. *Consider the SLM \mathcal{E}_{SLM} with arbitrary values of S , M , N , and \mathbf{H} and a prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$ that is valid for any $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$ (cf. (5.62)) with $\mathbf{x}_0 \in \mathcal{X}_S$. Then the minimum achievable variance $L_{\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)}$ exists, i.e., is finite for every \mathbf{x}_0 , and viewed as a function of \mathbf{x}_0 is lower semi-continuous.*

Proof. This fact follows from Theorem 4.3.6 by noting that the kernel $R_{\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)}(\cdot, \cdot)$ (see (5.13)) obviously satisfies (4.27). \square

Finally, we give two sufficient conditions on the prescribed bias $c(\cdot)$ to be valid for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$ with arbitrary $\mathbf{x}_0 \in \mathcal{X}_S$:

Theorem 5.3.4. *Consider the minimum variance problem $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$ with system matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ and a prescribed bias function $c(\mathbf{x}) : \mathcal{X}_S \rightarrow \mathbb{R}$ which is given as*

$$c(\mathbf{x}) = \exp(\mathbf{x}_1^T \tilde{\mathbf{H}}^\dagger \mathbf{x}) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\mathbf{p}!} \left(\frac{1}{\sigma} \tilde{\mathbf{H}}^\dagger \mathbf{x} \right)^{\mathbf{p}} - x_k, \quad (5.82)$$

with an arbitrary vector $\mathbf{x}_1 \in \mathbb{R}^D$. Then, if the coefficients satisfy $|a[\mathbf{p}]| \leq C^{|\mathbf{p}|}$ with an arbitrary constant $C \in \mathbb{R}_+$, the bias $c(\cdot)$ is valid for any $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$ with $\mathbf{x}_0 \in \mathcal{X}_S$.

Proof. By Theorem 4.3.4, we have that a prescribed bias function $c(\mathbf{x})$ is valid for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$, if (and only if) the prescribed mean function $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$:

$$\gamma(\mathbf{x}) = c(\mathbf{x}) + x_k = \exp(\mathbf{x}_1^T \tilde{\mathbf{H}}^\dagger \mathbf{x}) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\mathbf{p}!} \left(\frac{1}{\sigma} \tilde{\mathbf{H}}^\dagger \mathbf{x} \right)^{\mathbf{p}} \quad (5.83)$$

belongs to the RKHS $\mathcal{H}(\mathcal{M}_{\text{SLM}}(\mathbf{x}_0))$. According to Theorem 5.2.6 the function $f(\mathbf{x}) : \mathbb{R}^D \rightarrow \mathbb{R}$:

$$f(\mathbf{x}) = \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2\right) \exp\left(\left(\sigma\mathbf{x}_1 + \frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger\mathbf{x}_0\right)^T\mathbf{x}\right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\mathbf{p}!} \mathbf{x}^{\mathbf{p}}, \quad (5.84)$$

with the same $\mathbf{x}_1 \in \mathbb{R}^D$ and coefficients $a[\mathbf{p}]$ as in (5.82), belongs to the RKHS $\mathcal{H}(R_g^{(D)})$ since $|a[\mathbf{p}]| \leq C^{|\mathbf{p}|}$. The image $K_g[f(\cdot)] \in \mathcal{H}(\mathcal{M}_{\text{LGM}})$ under the congruence (5.20) is then obtained as

$$\begin{aligned} K_g[f(\cdot)](\mathbf{x}) &= f\left(\frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger\mathbf{x}\right) \exp\left(\frac{1}{2\sigma^2}\|\mathbf{H}\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2}\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0\right) \\ &= \exp\left(\sigma\left(\mathbf{x}_1 + \frac{1}{\sigma^2}\tilde{\mathbf{H}}^\dagger\mathbf{x}_0\right)^T\frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger\mathbf{x}\right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\mathbf{p}!} \left(\frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger\mathbf{x}\right)^{\mathbf{p}} \exp\left(-\frac{1}{\sigma^2}\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0\right) \\ &= \exp\left(-\frac{1}{\sigma^2}\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0 + \mathbf{x}_1\tilde{\mathbf{H}}^\dagger\mathbf{x} + \frac{1}{\sigma^2}\mathbf{x}_0^T(\tilde{\mathbf{H}}^\dagger)^T\tilde{\mathbf{H}}^\dagger\mathbf{x}\right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\mathbf{p}!} \left(\frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger\mathbf{x}\right)^{\mathbf{p}} \\ &\stackrel{(5.26)}{=} \exp\left(-\frac{1}{\sigma^2}\mathbf{x}^T\mathbf{H}^T\mathbf{H}\mathbf{x}_0 + \mathbf{x}_1\tilde{\mathbf{H}}^\dagger\mathbf{x} + \frac{1}{\sigma^2}\mathbf{x}_0^T\mathbf{H}^T\mathbf{H}\mathbf{x}\right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\mathbf{p}!} \left(\frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger\mathbf{x}\right)^{\mathbf{p}} \\ &= \exp\left(\mathbf{x}_1\tilde{\mathbf{H}}^\dagger\mathbf{x}\right) \sum_{\mathbf{p} \in \mathbb{Z}_+^D} \frac{a[\mathbf{p}]}{\mathbf{p}!} \left(\frac{1}{\sigma}\tilde{\mathbf{H}}^\dagger\mathbf{x}\right)^{\mathbf{p}}. \end{aligned} \quad (5.85)$$

Comparing (5.83) with (5.85) reveals that the prescribed mean function $\gamma(\cdot)$, which corresponds to the prescribed bias in (5.82), is the restriction of $K_g[f(\cdot)](\mathbf{x}) \in \mathcal{H}(\mathcal{M}_{\text{LGM}})$ to the subdomain \mathcal{X}_S , i.e., $\gamma(\cdot) = K_g[f(\cdot)]|_{\mathcal{X}_S}$, which implies by Theorem 5.2.1 that $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SLM}}(\mathbf{x}_0))$ and in turn that the prescribed bias $c(\mathbf{x})$ is valid for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$ with $\mathbf{x}_0 \in \mathcal{X}_S$. \square

By the definition of a valid bias function in Definition 2.3.6, we have trivially that any given estimator with finite variance for all $\mathbf{x} \in \mathbb{R}^N$ induces a valid bias function:

Lemma 5.3.5. *Consider an estimator $\hat{x}_k(\mathbf{y})$ for the LGM \mathcal{E}_{LGM} that has finite variance everywhere, i.e., $v(\hat{x}_k(\cdot); \mathbf{x}) < \infty$ for every $\mathbf{x} \in \mathbb{R}^N$. Then the bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : c(\mathbf{x}) = b(\hat{x}_k(\cdot); \mathbf{x})$ is valid for $\mathcal{M}_{\text{SLM}}(\mathbf{x}_0)$ (where σ , M , N and \mathbf{H} are the same as for \mathcal{E}_{LGM}) with arbitrary $\mathbf{x}_0 \in \mathcal{X}_S$.*

5.4 Lower Bounds on the Estimator Variance for the SLM

In this section we will derive lower bounds on the minimum achievable variance $L_{\mathcal{M}_{\text{SLM}}}$ where the parameters σ , S , M , N , \mathbf{H} , $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$, and $\mathbf{x}_0 \in \mathcal{X}_S$ of the minimum variance problem

$\mathcal{M}_{\text{SLM}} = (\mathcal{E}_{\text{SLM}}, c(\cdot), \mathbf{x}_0)$ are assumed arbitrary but fixed. Furthermore, we will assume from now on that the prescribed bias function $c(\cdot)$ is valid. This is no real restriction since the lower bounds that will be derived are finite and therefore trivially apply also formally if $c(\cdot)$ is not valid, in which case $L_{\mathcal{M}_{\text{SLM}}} = \infty$ by definition (see Definition 2.3.3).

The various bounds that will be presented are similar in that they are based on a projection of the prescribed mean function $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \gamma(\mathbf{x}) \triangleq c(\mathbf{x}) + x_k$ (which satisfies $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SLM}})$ if $c(\cdot)$ is valid) onto some subspace \mathcal{U} of the RKHS $\mathcal{H}(\mathcal{M}_{\text{SLM}})$. Indeed (as already discussed after Theorem 4.3.4), by Theorem 4.3.4 and Theorem 3.1.6, we have for an arbitrary subspace $\mathcal{U} \subseteq \mathcal{H}(\mathcal{M}_{\text{SLM}})$ the bound

$$L_{\mathcal{M}_{\text{SLM}}} = \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SLM}})}^2 - [\gamma(\mathbf{x}_0)]^2 \geq \|\mathbf{P}_{\mathcal{U}}\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SLM}})}^2 - [\gamma(\mathbf{x}_0)]^2. \quad (5.86)$$

The first bound is based on a generalization of the CRB and has been previously presented in [25]:

Theorem 5.4.1. *If the prescribed bias $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$ is such that the partial derivatives $\frac{\partial c(\mathbf{x})}{\partial x_l} \Big|_{\mathbf{x}=\mathbf{x}_0}$ exist for $l \in [N]$, then*

$$L_{\mathcal{M}_{\text{SLM}}} \geq \sigma^2 \mathbf{b}^T (\mathbf{H}^T \mathbf{H})^\dagger \mathbf{b} \quad \text{when } \|\mathbf{x}_0\|_0 < S \quad (5.87)$$

$$L_{\mathcal{M}_{\text{SLM}}} \geq \sigma^2 \mathbf{b}_{\mathbf{x}_0}^T (\mathbf{H}_{\mathbf{x}_0}^T \mathbf{H}_{\mathbf{x}_0})^\dagger \mathbf{b}_{\mathbf{x}_0} \quad \text{when } \|\mathbf{x}_0\|_0 = S, \quad (5.88)$$

where $\mathbf{b} \in \mathbb{R}^N$ is defined elementwise by $b_l \triangleq \delta_{k,l} + \frac{\partial c(\mathbf{x})}{\partial x_l} \Big|_{\mathbf{x}=\mathbf{x}_0}$ and $\mathbf{b}_{\mathbf{x}_0} \in \mathbb{R}^S$, $\mathbf{H}_{\mathbf{x}_0} \in \mathbb{R}^{M \times S}$ denote the restrictions to the entries and columns of \mathbf{b} and \mathbf{H} , respectively, which are indexed by $\text{supp}(\mathbf{x}_0) = (i_1, \dots, i_S)$, i.e., $(\mathbf{b}_{\mathbf{x}_0})_j = b_{i_j}$.

Proof. Since $c(\cdot)$ is assumed valid, we have by Theorem 4.3.4 that the prescribed mean function $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = c(\mathbf{x}) + x_k$ belongs to the RKHS $\mathcal{H}(\mathcal{M}_{\text{SLM}})$.

For the case $\|\mathbf{x}_0\|_0 < S$, consider the subspace $\mathcal{U}_1 \triangleq \text{span} \{ \{v_0(\cdot)\} \cup \{v_l(\cdot)\}_{l \in [N]} \}$ spanned by the functions $v_0(\cdot) \triangleq R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0)$ and

$$v_l(\cdot) \triangleq \frac{\partial^{e_l} R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_2) \Big|_{\mathbf{x}_2=\mathbf{x}_0}}{\partial \mathbf{x}_2^{e_l}}, \quad l \in [N]. \quad (5.89)$$

We have trivially $v_0(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SLM}})$ and by Theorem 3.4.1 we have also that $v_l(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SLM}})$ for $l \in [N]$. In a completely analogous manner as the RKHS-based derivation of Theorem 4.4.4 in Section 4.4.2, one can show that $\langle v_0(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} = 0$ for $l \in [N]$ (see (4.60)), $\langle v_l(\cdot), v_{l'}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} = \sigma^2 (\mathbf{H}^T \mathbf{H})_{l,l'}$ for $l, l' \in [N]$ (see (4.64)), and $\langle v_l(\cdot), \gamma(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} = b_l$ for $l \in [N]$ (see (4.63)). The bound (5.87) is then obtained via Theorem 4.3.4, Theorem 3.1.6, and Theorem 3.1.8 by projecting $\gamma(\cdot)$ on the subspace \mathcal{U}_1 (see (4.65)).

In an almost identical manner one can prove the bound (5.88) for the case $\|\mathbf{x}_0\|_0 = S$. Indeed, the only difference is that instead of using the subspace \mathcal{U}_1 one has to use the subspace $\mathcal{U}_2 \triangleq \text{span} \{ \{v_0(\cdot)\} \cup \{v_l(\cdot)\}_{l \in \text{supp}(\mathbf{x}_0)} \}$. \square

The formulation of the bound in [25] slightly differs from Theorem 5.4.1 in that the authors of [25] present a bound on the variance $v(\hat{\mathbf{x}}(\cdot); \mathbf{x}_0)$ of a vector-valued estimator that estimates \mathbf{x} itself and not only the k th coefficient x_k as we consider. However, according to Section 2.3.1, by summing the bound in Theorem 5.4.1 over all individual indices $k \in [N]$, one obtains the bound presented in [25]. This is just another consequence of the fact that minimum variance estimation of the parameter vector \mathbf{x} is equivalent to separate minimum variance estimation of the coefficients x_k .

We note two important aspects of Theorem 5.4.1: First, if the matrix \mathbf{H} has full column rank, i.e., $\text{rank}(\mathbf{H}) = N$, and we consider unbiased estimation, i.e., $c(\cdot) \equiv 0$, then the bound in (5.87) coincides with the variance of the well-known least-squares (LS) estimator [20, 21] given as

$$\hat{x}_{k,\text{LS}}(\mathbf{y}) \triangleq \mathbf{e}_k^T (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}, \quad (5.90)$$

which implies that in this case the inequality in (5.87) becomes an equality, i.e., $L_{\mathcal{M}_{\text{SLM}}} = \sigma^2 \mathbf{b}^T (\mathbf{H}^T \mathbf{H})^\dagger \mathbf{b}$. Indeed, a straightforward calculation reveals that $\hat{x}_{k,\text{LS}}(\cdot)$ is unbiased and has variance

$$v(\hat{x}_{k,\text{LS}}(\cdot); \mathbf{x}_0) = \sigma^2 \mathbf{e}_k^T (\mathbf{H}^T \mathbf{H})^\dagger \mathbf{e}_k = \sigma^2 \mathbf{e}_k^T (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{e}_k, \quad (5.91)$$

where the last equality follows because $\text{rank}(\mathbf{H}) = N$. Thus, under these conditions, the LS estimator is the LMVU estimator for the SLM at every parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$ with $\|\mathbf{x}_0\|_0 < S$. Since (by Theorem 2.3.3) it is the unique LMVU estimator for \mathbf{x}_0 with $\|\mathbf{x}_0\|_0 < S$, we have that if there existed a UMVU estimator, it necessarily would be the ordinary LS estimator $\hat{x}_{k,\text{LS}}(\cdot)$. Note that the LS estimator does not exploit the sparsity assumption expressed by the parameter set \mathcal{X}_S , and it has the constant variance (5.91) for every $\mathbf{x} \in \mathcal{X}_S$. Thus, even if there existed a UMVU estimator it would not be able to take advantage of the sparsity assumption. However, as already shown in [81, 82] and discussed in Section 5.5, there does not exist a UMVU estimator for the SLM in general.

A second important aspect of Theorem 5.4.1 is that the lower bound given by (5.87) and (5.88) is not a continuous function of \mathbf{x}_0 in general. Indeed, for the case $\mathbf{H} = \mathbf{I}$ and $c(\cdot) \equiv 0$ considered in [81, 82], it can be verified that the bound is a strictly upper semi-continuous function of \mathbf{x}_0 . Consider e.g. $\mathcal{M}_{\text{SLM}} = (\mathcal{E}_{\text{SLM}}, c(\cdot), \mathbf{x}_0)$ with $\mathbf{H} = \mathbf{I}$, $S = 1$, $M = N = 2$, $k = 2$, $c(\cdot) \equiv 0$, and a parameter vector \mathbf{x}_0 of the form $\mathbf{x}_0 = a(1, 0)^T$ where $a \in \mathbb{R}_+$. The corresponding bound given by Theorem 5.4.1 is equal to 0 for all $a > 0$ but equals 1 for $a = 0$. However, since by Theorem 5.3.3 the minimum achievable variance $L_{\mathcal{M}_{\text{SLM}}}$ has to be a *lower* semi-continuous function of \mathbf{x}_0 ,

we have that the bound in Theorem 5.4.1 cannot be tight, i.e., we have a strict inequality in (5.87) or (5.88) in general.

While the bound in Theorem 5.4.1 was based on a generalization of the CRB (see Section 4.4.2), the authors of [81] present a lower bound on $L_{\mathcal{M}_{\text{SLM}}}$ for $\mathbf{H} = \mathbf{I}$ and $c(\cdot) \equiv 0$ that is based on a generalization of the HCRB (see Section 4.4.4):

Theorem 5.4.2. *Consider the minimum variance problem $\mathcal{M}_{\text{SLM}} = (\mathcal{E}_{\text{SLM}}, c(\cdot), \mathbf{x}_0)$ with $\mathbf{H} = \mathbf{I}$ and $c(\cdot) \equiv 0$. Then we have the bound*

$$L_{\mathcal{M}_{\text{SLM}}} \geq \sigma^2 \quad \text{when } |\text{supp}(\mathbf{x}_0) \cup \{k\}| < S + 1 \quad (5.92)$$

$$L_{\mathcal{M}_{\text{SLM}}} \geq \sigma^2 \frac{N - S - 1}{N - S} \exp(-\xi_0^2/\sigma^2) \quad \text{when } |\text{supp}(\mathbf{x}_0) \cup \{k\}| = S + 1, \quad (5.93)$$

where ξ_0 denotes the value of the S -largest (in magnitude) entry of \mathbf{x}_0 .

Proof. The bias function $c(\cdot) \equiv 0$ is assumed to be valid, i.e., the mean function $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = x_k$ belongs to $\mathcal{H}(\mathcal{M}_{\text{SLM}})$.

First consider the case where $|\text{supp}(\mathbf{x}_0) \cup \{k\}| = S + 1$ (which can only occur if $N > S$). In order to show (5.93) we denote by j_0 and ξ_0 respectively the index and value of the S -largest (in magnitude) entry of \mathbf{x}_0 . We then introduce the subspace $\mathcal{U}_1^{(t)} \triangleq \text{span} \{v_0(\cdot) \cup \{v_l^{(t)}(\cdot)\}_{l \in [N]}\}$ which is parametrized by $t \in \mathbb{R}$ and spanned by the vectors $v_0(\cdot) \triangleq R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0)$, $v_l^{(t)}(\cdot) \triangleq R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0 - \mathbf{e}_j \xi_0 + t \mathbf{e}_l) - R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0)$ for $l \in [N] \setminus \text{supp}(\mathbf{x}_0)$ and $v_l^{(t)}(\cdot) \triangleq R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0 + t \mathbf{e}_l) - R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0)$ for $l \in \text{supp}(\mathbf{x}_0)$. Similarly to the RKHS-based derivation of the HCRB in Theorem 4.4.7, one can show by a projection of the mean function $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SLM}})$ on the subspace $\mathcal{U}^{(t)}$ that (see (4.93))

$$L_{\mathcal{M}_{\text{SLM}}} \geq t^2 \mathbf{e}_k^T \mathbf{V}^\dagger \mathbf{e}_k, \quad (5.94)$$

where $\mathbf{V} \in \mathbb{R}^{N \times N}$ is defined elementwise as

$$\begin{aligned} (\mathbf{V})_{m,n} &\triangleq \langle v_m^{(t)}(\cdot), v_n^{(t)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \\ &= \langle R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0 + t \mathbf{e}_m) - R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0), R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0 + t \mathbf{e}_n) - R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \\ &\stackrel{(3.44)}{=} R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_0 + t \mathbf{e}_m, \mathbf{x}_0 + t \mathbf{e}_n) - R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_0, \mathbf{x}_0 + t \mathbf{e}_m) \\ &\quad - R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_0 + t \mathbf{e}_n, \mathbf{x}_0) + R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_0, \mathbf{x}_0) \\ &\stackrel{(4.11)}{=} \exp(t^2 \mathbf{e}_n^T \mathbf{e}_m / \sigma^2) - 1 - 1 + 1 \\ &= \delta_{m,n} \exp(t^2 / \sigma^2) - 1. \end{aligned} \quad (5.95)$$

As shown in [82], the bound in (5.93) is obtained from (5.94) as the limit $\lim_{t \rightarrow 0} t^2 \mathbf{e}_k^T \mathbf{V}^\dagger \mathbf{e}_k$.

Now we consider the remaining case where $|\text{supp}(\mathbf{x}_0) \cup \{k\}| < S + 1$. Instead of $\mathcal{U}_1^{(t)}$, we will now use the subspace $\mathcal{U}_2^{(t)} \triangleq \{v_0(\cdot), v^{(t)}(\cdot)\}$, where $v^{(t)}(\cdot) \triangleq R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0 + t\mathbf{e}_k) - R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0)$. In an analogous manner as before, one obtains

$$L_{\mathcal{M}_{\text{SLM}}} \geq \frac{t^2}{\|v^{(t)}(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SLM}})}^2}. \quad (5.96)$$

The squared norm of $v^{(t)}(\cdot)$ can be calculated as

$$\begin{aligned} \|v^{(t)}(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SLM}})}^2 &= \langle v^{(t)}(\cdot), v^{(t)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \\ &\stackrel{(3.44)}{=} R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_0 + t\mathbf{e}_k, \mathbf{x}_0 + t\mathbf{e}_k) - R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_0, \mathbf{x}_0 + t\mathbf{e}_k) \\ &\quad - R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_0 + t\mathbf{e}_k, \mathbf{x}_0) + R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_0, \mathbf{x}_0) \\ &\stackrel{(4.11)}{=} \exp(t^2 \mathbf{e}_k^T \mathbf{e}_k / \sigma^2) - 1 - 1 + 1 \\ &= \exp(t^2 / \sigma^2) - 1. \end{aligned} \quad (5.97)$$

Thus, one straightforwardly obtains the bound $L_{\mathcal{M}_{\text{SLM}}} \geq \sigma^2$ in (5.92) by the limit

$$\lim_{t \rightarrow 0} t^2 / [\exp(t^2 / \sigma^2) - 1]. \quad (5.98)$$

Since the variance of the specific unbiased estimator $\hat{g}(\mathbf{y}) = y_k$ for the SLM is equal to σ^2 , we have that $L_{\mathcal{M}_{\text{SLM}}} = \sigma^2$. \square

The bound given in Theorem 5.4.2 has been presented in a slightly different form in [81, 82], where a lower bound on the variance $v(\hat{\mathbf{x}}(\cdot), \mathbf{x}_0)$ of an unbiased estimator of the parameter vector \mathbf{x} itself, instead of the single coefficient x_k , has been derived.³ As discussed in Section 2.3.1, minimum variance estimation of the parameter vector is equivalent to separate minimum variance estimation of the coefficients x_k . In particular, the bound given in [81, 82] is obtained by summing the bound in Theorem 5.4.2 over all indices $k \in [N]$.

While in general the bound in Theorem 5.4.2 is tighter, i.e., higher than the bound of Theorem 5.4.1 (when specialized to $c(\cdot) \equiv 0$ and $\mathbf{H} = \mathbf{I}$), it is again upper semi-continuous. This again implies by Theorem 5.3.3 that the bound cannot be tight, i.e., we have a strict inequality in (5.93)

³More precisely, a lower bound on the MSE of any unbiased estimator of \mathbf{x} was presented in [81, 82]. However, the variance of an unbiased estimator is equal to its MSE.

in general. However, for the case $|\text{supp}(\mathbf{x}_0) \cup \{k\}| < S + 1$,⁴ the bound in (5.92) is tight since, as already mentioned above, it is achieved by the ordinary LS estimator $\hat{x}_{k,\text{LS}}(\mathbf{y}) = y_k$.

An improved lower bound on $L_{\mathcal{M}_{\text{SLM}}}$ has been presented in [83]:

Theorem 5.4.3. *Consider the minimum variance problem $\mathcal{M}_{\text{SLM}} = (\mathcal{E}_{\text{SLM}}, c(\cdot), \mathbf{x}_0)$ with a system matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ satisfying (5.3) and with prescribed mean function $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = c(\mathbf{x}) + x_k$. We assume that the prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$ is such that the partial derivatives $\frac{\partial^{e_l} c(\mathbf{x})}{\partial \mathbf{x}^{e_l}}$ exist for all $l \in [N]$. Then, for an arbitrary set $\mathcal{K} = \{i_1, \dots, i_{|\mathcal{K}|}\} \subseteq [N]$ consisting of no more than S different indices, i.e., $|\mathcal{K}| \leq S$, we have the bound*

$$L_{\mathcal{M}_{\text{SLM}}} \geq \exp \left(-\frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{H}\mathbf{x}_0\|_2^2 \right) \left[\sigma^2 \mathbf{r}_{\mathbf{x}_0}^T (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{r}_{\mathbf{x}_0} + \gamma^2(\tilde{\mathbf{x}}_0) \right] - \gamma^2(\mathbf{x}_0), \quad (5.99)$$

where $\mathbf{P}_{\mathcal{K}} \triangleq \mathbf{H}_{\mathcal{K}} \mathbf{H}_{\mathcal{K}}^\dagger \in \mathbb{R}^{M \times M}$, $\mathbf{r}_{\mathbf{x}_0} \in \mathbb{R}^{|\mathcal{K}|}$ is defined elementwise as $r_{\mathbf{x}_0, l} \triangleq \frac{\partial^{e_l} c(\mathbf{x})}{\partial \mathbf{x}^{e_l}} \Big|_{\mathbf{x}=\tilde{\mathbf{x}}_0}$, and $\tilde{\mathbf{x}}_0 \in \mathbb{R}^N$ is defined as the unique (due to (5.3)) vector with $\text{supp}(\tilde{\mathbf{x}}_0) = \mathcal{K}$ that solves

$$\mathbf{H}\tilde{\mathbf{x}}_0 = \mathbf{H}_{\mathcal{K}} \mathbf{H}_{\mathcal{K}}^\dagger \mathbf{H}\mathbf{x}_0. \quad (5.100)$$

Proof. First we note that since \mathbf{H} is assumed to satisfy (5.3), the matrix $\mathbf{H}_{\mathcal{K}} \in \mathbb{R}^{M \times |\mathcal{K}|}$ has full column rank, which implies that $\mathbf{H}_{\mathcal{K}}^\dagger = (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{H}_{\mathcal{K}}^T$ [38]. Thus,

$$\begin{aligned} \mathbf{H}_{\mathcal{K}}^T \mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0) &= \mathbf{H}_{\mathcal{K}}^T (\mathbf{H}_{\mathcal{K}} \mathbf{H}_{\mathcal{K}}^\dagger \mathbf{H}\mathbf{x}_0 - \mathbf{H}\mathbf{x}_0) = \mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}} \mathbf{H}_{\mathcal{K}}^\dagger \mathbf{H}\mathbf{x}_0 - \mathbf{H}_{\mathcal{K}}^T \mathbf{H}\mathbf{x}_0 \\ &= \mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}} (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{H}_{\mathcal{K}}^T \mathbf{H}\mathbf{x}_0 - \mathbf{H}_{\mathcal{K}}^T \mathbf{H}\mathbf{x}_0 = \mathbf{H}_{\mathcal{K}}^T \mathbf{H}\mathbf{x}_0 - \mathbf{H}_{\mathcal{K}}^T \mathbf{H}\mathbf{x}_0 = \mathbf{0}, \end{aligned} \quad (5.101)$$

which implies in turn that $\text{span}(\mathbf{H}_{\mathcal{K}})$ is orthogonal to $\mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)$. Let us then consider the subspace

$$\mathcal{U}_3 \triangleq \text{span} \{v_0(\cdot) \cup \{v_l(\cdot)\}_{l \in \mathcal{K}}\} \quad (5.102)$$

spanned by the functions $v_0(\cdot) \triangleq R_{\mathcal{M}_{\text{SLM}}}(\cdot, \tilde{\mathbf{x}}_0)$ and $v_l(\cdot) \triangleq \frac{\partial^{e_l} R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_2)}{\partial \mathbf{x}_2^{e_l}} \Big|_{\mathbf{x}_2=\tilde{\mathbf{x}}_0}$, which by Theorem 3.4.1 belong to the RKHS $\mathcal{H}(\mathcal{M}_{\text{SLM}})$. The inner product between any $v_l(\cdot)$ with $l \in \mathcal{K}$ and $v_0(\cdot)$ is given due to Theorem 3.4.1 by

$$\begin{aligned} \langle v_0(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} &= \frac{\partial R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial x_{2, l}} \Big|_{\mathbf{x}_1=\mathbf{x}_2=\tilde{\mathbf{x}}_0} \\ &= \frac{1}{\sigma^2} (\tilde{\mathbf{x}}_0 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} \mathbf{e}_l \exp \left(\frac{1}{\sigma^2} \|\mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)\|_2^2 \right) \stackrel{(a)}{=} 0, \end{aligned} \quad (5.103)$$

⁴Note that the case $|\text{supp}(\mathbf{x}_0) \cup \{k\}| < S + 1$ can only occur if either $\|\mathbf{x}\|_0 < S$ or $\|\mathbf{x}\|_0 = S$ and additionally $k \in \text{supp}(\mathbf{x}_0)$.

where (a) follows from the fact that $\mathbf{H}\mathbf{e}_l \in \text{span}(\mathbf{H}_{\mathcal{K}})$ and therefore, as shown by (5.101), has to be orthogonal to $\mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)$, i.e., $(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} \mathbf{e}_l = 0$. Now, we define the matrix $\mathbf{V} \in \mathbb{R}^{|\mathcal{K}| \times |\mathcal{K}|}$ elementwise via

$$\begin{aligned}
(\mathbf{V})_{m,n} &\triangleq \langle v_{i_m}(\cdot), v_{i_n}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \\
&\stackrel{(a)}{=} \left. \frac{\partial R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial x_{1,i_n} \partial x_{2,i_m}} \right|_{\mathbf{x}_1 = \mathbf{x}_2 = \tilde{\mathbf{x}}_0} \\
&= \frac{1}{\sigma^2} \left. \frac{\partial [(\mathbf{x}_1 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} \mathbf{e}_{i_m} \exp(\frac{1}{\sigma^2}(\mathbf{x}_1 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0))] }{\partial x_{1,i_n}} \right|_{\mathbf{x}_1 = \tilde{\mathbf{x}}_0} \\
&= \frac{1}{\sigma^2} \mathbf{e}_{i_n}^T \mathbf{H}^T \mathbf{H} \mathbf{e}_{i_m} \exp\left(\frac{1}{\sigma^2}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)\right) \\
&\quad + \frac{1}{\sigma^4} \left((\tilde{\mathbf{x}}_0 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} \mathbf{e}_{i_m} \right) \mathbf{e}_{i_n}^T \mathbf{H}^T \mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0) \exp\left(\frac{1}{\sigma^2}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)\right) \\
&\stackrel{(b)}{=} \frac{1}{\sigma^2} \mathbf{e}_{i_n}^T \mathbf{H}^T \mathbf{H} \mathbf{e}_{i_m} \exp\left(\frac{1}{\sigma^2}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)\right), \tag{5.104}
\end{aligned}$$

where (a) follows again from Theorem 3.4.1 and (b) follows from the fact that $\mathbf{H}\mathbf{e}_{i_m} \in \text{span}(\mathbf{H}_{\mathcal{K}})$, implying via (5.101) that $(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} \mathbf{e}_{i_m} = 0$. From (5.104), we have by elementwise comparison that

$$\mathbf{V} = \frac{1}{\sigma^2} \mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}} \exp\left(\frac{1}{\sigma^2} \|\mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)\|_2^2\right). \tag{5.105}$$

Since furthermore

$$\|\mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)\|_2^2 \stackrel{(5.100)}{=} \|\mathbf{H}\mathbf{x}_0 - \mathbf{H}_{\mathcal{K}} \mathbf{H}_{\mathcal{K}}^\dagger \mathbf{H}\mathbf{x}_0\|_2^2 = \|(\mathbf{I} - \mathbf{H}_{\mathcal{K}} \mathbf{H}_{\mathcal{K}}^\dagger) \mathbf{H}\mathbf{x}_0\|_2^2 = \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H}\mathbf{x}_0\|_2^2, \tag{5.106}$$

the bound in (5.99) follows from Theorem 4.3.4 and Theorem 3.1.9 (note that due to (5.3) we have that $(\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} = (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^\dagger$) via (5.104), (5.103) and the inner products

$$\langle v_0(\cdot), \gamma(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} = \gamma(\tilde{\mathbf{x}}_0) \tag{5.107}$$

and

$$\langle v_l(\cdot), \gamma(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} = \left. \frac{\partial^{e_{i_l}} \gamma(\mathbf{x})}{\partial \mathbf{x}^{e_{i_l}}} \right|_{\mathbf{x} = \tilde{\mathbf{x}}_0} = r_{\mathbf{x}_0, l} \tag{5.108}$$

(cf. Theorem 3.4.1):

$$L_{\mathcal{M}_{\text{SLM}}} \stackrel{(4.19)}{=} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SLM}})}^2 - [\gamma(\mathbf{x}_0)]^2 \stackrel{(4.21)}{\geq} \|\mathbf{P}_{\mathcal{U}_3} \gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SLM}})}^2 - [\gamma(\mathbf{x}_0)]^2$$

$$\begin{aligned}
& \stackrel{(3.39),(5.108)}{=} \langle \gamma(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \left(\langle v_0(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \right)^{-1} \underbrace{\langle \gamma(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})}}_{\stackrel{(5.107)}{=} \gamma(\tilde{\mathbf{x}}_0)} \\
& + \mathbf{r}_{\mathbf{x}_0}^T \mathbf{V}^\dagger \mathbf{r}_{\mathbf{x}_0} - [\gamma(\mathbf{x}_0)]^2 \\
& \stackrel{(a)}{=} \exp \left(-\frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H} \mathbf{x}_0\|_2^2 \right) [\gamma(\tilde{\mathbf{x}}_0)]^2 + \mathbf{r}_{\mathbf{x}_0}^T \mathbf{V}^\dagger \mathbf{r}_{\mathbf{x}_0} - [\gamma(\mathbf{x}_0)]^2 \\
& \stackrel{(5.105)}{=} \exp \left(-\frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H} \mathbf{x}_0\|_2^2 \right) \left[\sigma^2 \mathbf{r}_{\mathbf{x}_0}^T (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{r}_{\mathbf{x}_0} + \gamma^2(\tilde{\mathbf{x}}_0) \right] - [\gamma(\mathbf{x}_0)]^2, \quad (5.109)
\end{aligned}$$

where step (a) is due to

$$\begin{aligned}
\langle v_0(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} &= \langle R_{\mathcal{M}_{\text{SLM}}}(\cdot, \tilde{\mathbf{x}}_0), R_{\mathcal{M}_{\text{SLM}}}(\cdot, \tilde{\mathbf{x}}_0) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \stackrel{(3.44)}{=} R_{\mathcal{M}_{\text{SLM}}}(\tilde{\mathbf{x}}_0, \tilde{\mathbf{x}}_0) \\
&= \exp \left(\frac{1}{\sigma^2} (\tilde{\mathbf{x}}_0 - \mathbf{x}_0)^T \mathbf{H}^T \mathbf{H} (\tilde{\mathbf{x}}_0 - \mathbf{x}_0) \right) = \exp \left(\frac{1}{\sigma^2} \|\mathbf{H}(\tilde{\mathbf{x}}_0 - \mathbf{x}_0)\|_2^2 \right) \\
&\stackrel{(5.106)}{=} \exp \left(\frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H} \mathbf{x}_0\|_2^2 \right).
\end{aligned}$$

□

Note that the matrix $\mathbf{P}_{\mathcal{K}} = \mathbf{H}_{\mathcal{K}} \mathbf{H}_{\mathcal{K}}^\dagger$ defined in Theorem 5.4.3 is an orthogonal projection matrix [38] on the subspace $\mathcal{U}_{\mathcal{K}} \triangleq \text{span}(\mathbf{H}_{\mathcal{K}}) \subseteq \mathbb{R}^M$; consequently, $(\mathbf{I} - \mathbf{P}_{\mathcal{K}})$ is an orthogonal projection matrix on the orthogonal complement [4, 38] of $\mathcal{U}_{\mathcal{K}}$ and the norm $\|(\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H} \mathbf{x}_0\|$ thus represents the distance between the point $\mathbf{H} \mathbf{x}_0$ and the subspace $\mathcal{U}_{\mathcal{K}}$ [6]. Therefore, the factor $\exp(-\frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H} \mathbf{x}_0\|_2^2)$ appearing in the bound (5.99) can be interpreted as a measure of the distance between the point $\mathbf{H} \mathbf{x}_0$ and the subspace $\mathcal{U}_{\mathcal{K}}$. In general, the bound (5.99) is tighter, i.e., higher, if \mathcal{K} is chosen such that this distance is small, i.e., $\|(\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H} \mathbf{x}_0\|_2^2$ is small.

One of the appealing properties of the bound (5.99) is that it is continuous for the special case $\mathbf{H} = \mathbf{I}$ and $c(\cdot) \equiv 0$, irrespective of what index k is chosen for the parameter function in the SLM (see (5.5)). We also have for this specific instance of \mathcal{M}_{SLM} that the bound (5.99) is tighter, i.e., higher, than the bounds given in Theorem 5.4.1 and Theorem 5.4.2.

The quantity $\sigma^2 \mathbf{r}_{\mathbf{x}_0}^T (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{r}_{\mathbf{x}_0}$ appearing in the bound (5.99) has a remarkable interpretation as the unconstrained CRB (see Section 4.4.2) for a minimum variance problem associated with the LGM with $N = |\mathcal{K}|$. In fact, for $k \in \text{supp}(\mathbf{x}_0)$, the factor $\exp(-\frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H} \mathbf{x}_0\|_2^2)$ can be made equal to 1 by choosing $\mathcal{K} = \text{supp}(\mathbf{x}_0)$. On the other hand, consider $k \notin \text{supp}(\mathbf{x}_0)$, and chose $\mathcal{K} = \{\{k\} \cup \mathcal{L}\}$, where \mathcal{L} denotes the indices of the $S - 1$ largest (in magnitude) entries of

\mathbf{x}_0 . The bound (5.99) then indicates a transition from a “low”-SNR regime, where

$$\exp\left(-\frac{1}{\sigma^2}\|(\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{H}\mathbf{x}_0\|_2^2\right) \approx 1,$$

to a “high”-SNR regime, where

$$\exp\left(-\frac{1}{\sigma^2}\|(\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{H}\mathbf{x}_0\|_2^2\right) \approx 0.$$

In the low-SNR regime, the bound (5.99) is equal to the CRB of a LGM with $N = |\mathcal{K}|$. In the high-SNR regime, the bound becomes approximately equal to 0; this means that the nonzero entries x_k with $k \notin \text{supp}(\mathbf{x})$ can be estimated with small variance.

We note that in [83], we presented a version of the bound (5.99), that is obtained by choosing the index set \mathcal{K} for which the right hand side of (5.99) is largest. We note also that the proof of the bound (5.99) used here is slightly more direct than that sketched in [83].

By a slight modification of the proof of Theorem 5.4.3, we obtain

Theorem 5.4.4. *Consider the minimum variance problem $\mathcal{M}_{\text{SLM}} = (\mathcal{E}_{\text{SLM}}, c(\cdot), \mathbf{x}_0)$ with a system matrix \mathbf{H} satisfying (5.3) and with prescribed mean function $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = c(\mathbf{x}) + x_k$. We assume that $c(\cdot)$ is valid and such that the partial derivatives $\frac{\partial^{e_l} c(\mathbf{x})}{\partial \mathbf{x}^{e_l}}$ exist for all $l \in [N]$. Then, for an arbitrary set $\mathcal{K} = \{i_1, \dots, i_{|\mathcal{K}|}\} \subseteq [N]$ consisting of no more than S different indices, i.e., $|\mathcal{K}| \leq S$, we have the bound*

$$L_{\mathcal{M}_{\text{SLM}}} \geq \exp\left(-\frac{1}{\sigma^2}\|(\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{H}\mathbf{x}_0\|_2^2\right) \sigma^2 \mathbf{r}_{\mathbf{x}_0}^T (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{r}_{\mathbf{x}_0}, \quad (5.110)$$

where $\mathbf{P}_{\mathcal{K}} \in \mathbb{R}^{M \times M}$ and $\mathbf{r}_{\mathbf{x}_0} \in \mathbb{R}^{|\mathcal{K}|}$ are defined as in Theorem 5.4.3.

Proof. We use the same derivation as in the proof of Theorem 5.4.3, but with the subspace \mathcal{U}_3 in (5.102) replaced by the subspace

$$\mathcal{U}_4 \triangleq \text{span} \{v_0(\cdot) \cup \{v_l(\cdot)\}_{l \in \mathcal{K}}\} \quad (5.111)$$

which is spanned by the functions $v_0(\cdot) \triangleq R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0)$ and $v_l(\cdot) \triangleq \frac{\partial^{e_l} R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_2)}{\partial \mathbf{x}_2^{e_l}} \Big|_{\mathbf{x}_2 = \tilde{\mathbf{x}}_0}$. The difference between \mathcal{U}_3 and \mathcal{U}_4 is the definition of the function $v_0(\cdot)$. Therefore, only the equations in the proof of Theorem 5.4.3 which involve $v_0(\cdot)$ change. However, we still have that $v_l(\cdot)$ with $l \in \mathcal{K}$ is orthogonal to $v_0(\cdot)$ (see (5.103)), since

$$\begin{aligned} \langle v_0(\cdot), v_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} &= \left\langle R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0), \frac{\partial^{e_l} R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_2)}{\partial \mathbf{x}_2^{e_l}} \Big|_{\mathbf{x}_2 = \tilde{\mathbf{x}}_0} \right\rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \\ &\stackrel{(3.44)}{=} \frac{\partial^{e_l} R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_0, \mathbf{x}_2)}{\partial \mathbf{x}_2^{e_l}} \Big|_{\mathbf{x}_2 = \tilde{\mathbf{x}}_0} \end{aligned}$$

$$\stackrel{(4.11)}{=} \frac{\partial^{\mathbf{e}_i} 1}{\partial \mathbf{x}_2^{\mathbf{e}_i}} \Big|_{\mathbf{x}_2 = \tilde{\mathbf{x}}_0} = 0. \quad (5.112)$$

We also obtain the inner products

$$\begin{aligned} \langle v_0(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} &= \langle R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0), R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \\ &\stackrel{(3.44)}{=} R_{\mathcal{M}_{\text{SLM}}}(\mathbf{x}_0, \mathbf{x}_0) \stackrel{(4.11)}{=} 1, \end{aligned} \quad (5.113)$$

and

$$\langle \gamma(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} = \langle \gamma(\cdot), R_{\mathcal{M}_{\text{SLM}}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \stackrel{(3.44)}{=} \gamma(\mathbf{x}_0). \quad (5.114)$$

Therefore, the relations of (5.109) up to the last expression before step (a) are still valid, i.e.,

$$\begin{aligned} L_{\mathcal{M}_{\text{SLM}}} &\stackrel{(4.19)}{=} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SLM}})}^2 - [\gamma(\mathbf{x}_0)]^2 \stackrel{(4.21)}{\geq} \|\mathbf{P}_{\mathcal{U}_4} \gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SLM}})}^2 - [\gamma(\mathbf{x}_0)]^2 \\ &\stackrel{(3.39), (5.108)}{=} \langle \gamma(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \underbrace{\left(\langle v_0(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})} \right)^{-1}}_{\stackrel{(5.113)}{=} 1} \underbrace{\langle \gamma(\cdot), v_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SLM}})}}_{\stackrel{(5.114)}{=} \gamma(\mathbf{x}_0)} \\ &\quad + \mathbf{r}_{\mathbf{x}_0}^T \mathbf{V}^\dagger \mathbf{r}_{\mathbf{x}_0} - [\gamma(\mathbf{x}_0)]^2 \\ &= [\gamma(\mathbf{x}_0)]^2 + \mathbf{r}_{\mathbf{x}_0}^T \mathbf{V}^\dagger \mathbf{r}_{\mathbf{x}_0} - [\gamma(\mathbf{x}_0)]^2 \\ &\stackrel{(5.105)}{=} \exp\left(-\frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H} \mathbf{x}_0\|_2^2\right) \sigma^2 \mathbf{r}_{\mathbf{x}_0}^T (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{r}_{\mathbf{x}_0}. \end{aligned} \quad (5.115)$$

□

The difference between the bound (5.110) and the bound (5.99) is equal to

$$\gamma^2(\mathbf{x}_0) - \exp\left(-\frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H} \mathbf{x}_0\|_2^2\right) \gamma^2(\tilde{\mathbf{x}}_0), \quad (5.116)$$

which depends on the choice of the index set \mathcal{K} . If for some index set \mathcal{K} (note that $\mathbf{r}_{\mathbf{x}_0}$ depends on \mathcal{K}), the prescribed bias is such that $\gamma^2(\tilde{\mathbf{x}}_0) \approx \gamma^2(\mathbf{x}_0)$,⁵ then the difference (5.116) is approximately nonnegative since

$$\exp\left(-\frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H} \mathbf{x}_0\|_2^2\right) \leq 1.$$

This in turn, implies that the bound (5.110) is tighter, i.e. higher, in this case.

⁵This is the case e.g. if $c(\cdot) \equiv 0$, i.e. for unbiased estimation, and the columns of $\mathbf{H}_{\mathcal{K}}$ are nearly orthonormal.

5.5 The Sparse Signal in Noise Model

We now specialize the results obtained for minimum variance estimation associated with the general SLM \mathcal{E}_{SLM} to the special case of the SLM given by the SSNM $\mathcal{E}_{\text{SSNM}}$ (5.10). The SSNM is obtained from the SLM by choosing $\mathbf{H} = \mathbf{I}$ (which implies that $M = N$). We denote by $\mathcal{M}_{\text{SSNM}} \triangleq (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}_0)$ any minimum variance problem that is associated with $\mathcal{E}_{\text{SSNM}}$, a prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$, and a parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$. The specific choices for $c(\cdot)$ and \mathbf{x}_0 should be clear from the context. The following discussions and results are presented in part in [84].

In principle, we can characterize the RKHS $\mathcal{H}(\mathcal{M}_{\text{SSNM}})$, which is associated with the kernel

$$R_{\mathcal{M}_{\text{SSNM}}}(\cdot, \cdot) : \mathcal{X}_S \times \mathcal{X}_S \rightarrow \mathbb{R} : R_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_1, \mathbf{x}_2) \triangleq \exp\left(\frac{1}{\sigma^2}(\mathbf{x}_1 - \mathbf{x}_0)^T(\mathbf{x}_2 - \mathbf{x}_0)\right), \quad (5.117)$$

using Theorem 5.2.2 and Theorem 5.2.1 specialized to $\mathbf{H} = \mathbf{I}$. However, it will turn out that a more convenient characterization is possible by exploiting the specific structure of $\mathcal{H}(\mathcal{M}_{\text{SLM}})$ induced by the choice $\mathbf{H} = \mathbf{I}$:

Theorem 5.5.1. *Consider the minimum variance problem $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}_0)$ with some $\sigma, S, N, \mathbf{x}_0 \in \mathcal{X}_S$, and $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$. The RKHS $\mathcal{H}(\mathcal{M}_{\text{SSNM}})$ is isometric to the RKHS $\mathcal{H}(R_e)$ associated with the kernel $R_e(\cdot, \cdot) : \mathcal{X}_S \times \mathcal{X}_S \rightarrow \mathbb{R}$:*

$$R_e(\mathbf{x}_1, \mathbf{x}_2) = \exp(\mathbf{x}_1^T \mathbf{x}_2). \quad (5.118)$$

A congruence $\mathsf{K}_e[\cdot] : \mathcal{H}(\mathcal{M}_{\text{SSNM}}) \rightarrow \mathcal{H}(R_e)$ is given by

$$f(\cdot) \mapsto \tilde{f}(\cdot) = \mathsf{K}_e[f(\cdot)] : \tilde{f}(\mathbf{x}') = f(\sigma \mathbf{x}') \nu_{\mathbf{x}_0}(\mathbf{x}') \quad (5.119)$$

with $\nu_{\mathbf{x}_0}(\mathbf{x}) \triangleq \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 + \frac{1}{\sigma}\mathbf{x}^T \mathbf{x}_0\right)$. The RKHS $\mathcal{H}(R_e)$ is differentiable up to any order and contains the functions

$$g^{(\mathbf{p})}(\mathbf{x}) \triangleq \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} R_e(\mathbf{x}, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}}} \Big|_{\mathbf{x}_2=0} = \frac{1}{\sqrt{\mathbf{p}!}} \mathbf{x}^{\mathbf{p}}, \quad (5.120)$$

i.e.,

$$g^{(\mathbf{p})}(\cdot) \in \mathcal{H}(R_e), \quad (5.121)$$

where $\mathbf{p} \in \mathbb{Z}_+^N$ is an arbitrary multi-index with $\|\mathbf{p}\|_0 \leq S$, i.e., $\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S$. The inner product of an arbitrary function $f(\cdot) \in \mathcal{H}(R_e)$ with any function $g^{(\mathbf{p})}(\cdot)$ is given by

$$\langle f(\cdot), g^{(\mathbf{p})}(\cdot) \rangle_{\mathcal{H}(R_e)} = \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} f(\mathbf{x})}{\mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}=0}. \quad (5.122)$$

Moreover, the set $\{g^{(\mathbf{p})}(\cdot)\}_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S}$ is an ONB for $\mathcal{H}(R_e)$ and therefore any function $f(\cdot) \in \mathcal{H}(R_e)$ can be written pointwise as

$$f(\mathbf{x}) = \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} a[\mathbf{p}] g^{(\mathbf{p})}(\mathbf{x}) = \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} a[\mathbf{p}] \frac{1}{\sqrt{\mathbf{p}!}} \mathbf{x}^{\mathbf{p}}, \quad (5.123)$$

with a coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)$. Conversely, for any coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)$ the function given by (5.123) belongs to $\mathcal{H}(R_e)$.

Proof. The derivation of the congruence in (5.119) is completely analogous to the proof of Theorem 5.2.2 specialized to $\mathbf{H} = \mathbf{I}$:

Let us define the function sets (similar to (5.27) and (5.28))

$$\mathcal{A}_1 \triangleq \left\{ g_{\mathbf{x}}(\cdot) \triangleq R_{\mathcal{M}_{\text{SSNM}}}(\cdot, \sigma \mathbf{x}) \right\}_{\mathbf{x} \in \mathcal{X}_S} \quad (5.124)$$

and

$$\mathcal{B}_1 \triangleq \left\{ f_{\mathbf{x}}(\cdot) \triangleq \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}^T \mathbf{x}_0\right) R_e(\cdot, \mathbf{x}) \right\}_{\mathbf{x} \in \mathcal{X}_S}. \quad (5.125)$$

Obviously, we have that the sets \mathcal{A}_1 and \mathcal{B}_1 span $\mathcal{H}(\mathcal{M}_{\text{SSNM}})$ and $\mathcal{H}(R_e)$, respectively, i.e., $\text{span}\{\mathcal{A}_1\} = \mathcal{H}(\mathcal{M}_{\text{SSNM}})$ and $\text{span}\{\mathcal{B}_1\} = \mathcal{H}(R_e)$. Similar to (5.36) one can verify that

$$\begin{aligned} \langle g_{\mathbf{x}_1}(\cdot), g_{\mathbf{x}_2}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})} &= \langle R_{\mathcal{M}_{\text{SSNM}}}(\cdot, \sigma \mathbf{x}_1), R_{\mathcal{M}_{\text{SSNM}}}(\cdot, \sigma \mathbf{x}_2) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})} \\ &\stackrel{(3.44)}{=} R_{\mathcal{M}_{\text{SSNM}}}(\sigma \mathbf{x}_1, \sigma \mathbf{x}_2) \\ &= \exp\left(\frac{1}{\sigma^2}(\sigma \mathbf{x}_1 - \mathbf{x}_0)^T(\sigma \mathbf{x}_2 - \mathbf{x}_0)\right) \\ &= \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_1^T \mathbf{x}_0 + \mathbf{x}_1^T \mathbf{x}_2 + \frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_2^T \mathbf{x}_0\right) \\ &= \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_1^T \mathbf{x}_0\right) \exp(\mathbf{x}_1^T \mathbf{x}_2) \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_2^T \mathbf{x}_0\right) \\ &\stackrel{(5.118)}{=} \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_1^T \mathbf{x}_0\right) R_e(\mathbf{x}_1, \mathbf{x}_2) \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_2^T \mathbf{x}_0\right) \\ &\stackrel{(3.44)}{=} \left\langle \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_1^T \mathbf{x}_0\right) R_e(\cdot, \mathbf{x}_1), \right. \\ &\quad \left. \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}_2^T \mathbf{x}_0\right) R_e(\cdot, \mathbf{x}_2) \right\rangle_{\mathcal{H}(R_e)} \\ &= \langle f_{\mathbf{x}_1}(\cdot), f_{\mathbf{x}_2}(\cdot) \rangle_{\mathcal{H}(R_e)}. \end{aligned} \quad (5.126)$$

for any two vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_S$. Similar to (5.37), let us, for an arbitrary $\mathbf{x} \in \mathcal{X}_S$, denote by $h_{\mathbf{x}}(\cdot) \triangleq \mathcal{K}_e[g_{\mathbf{x}}(\cdot)] \in \mathcal{H}(R_e)$ the image of the function $g_{\mathbf{x}}(\cdot) = R_{\mathcal{M}_{\text{SSNM}}}(\cdot, \sigma \mathbf{x})$ (see (5.124)) under

the mapping $\mathsf{K}_e[\cdot]$ defined in (5.119). We have

$$\begin{aligned}
h_{\mathbf{x}}(\mathbf{x}') &= g_{\mathbf{x}}(\sigma\mathbf{x}') \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 + \frac{1}{\sigma}(\mathbf{x}')^T \mathbf{x}_0\right) \\
&\stackrel{(5.124)}{=} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 + \frac{1}{\sigma}(\mathbf{x}')^T \mathbf{x}_0\right) R_{\mathcal{M}_{\text{SSNM}}}(\sigma\mathbf{x}', \sigma\mathbf{x}) \\
&\stackrel{(5.117)}{=} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 + \frac{1}{\sigma}(\mathbf{x}')^T \mathbf{x}_0\right) \exp\left(\frac{1}{\sigma^2}(\sigma\mathbf{x}' - \mathbf{x}_0)^T (\sigma\mathbf{x} - \mathbf{x}_0)\right) \\
&= \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 + \frac{1}{\sigma}(\mathbf{x}')^T \mathbf{x}_0 + (\mathbf{x}')^T \mathbf{x} + \frac{1}{\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}(\mathbf{x}')^T \mathbf{x}_0 - \frac{1}{\sigma}\mathbf{x}^T \mathbf{x}_0\right) \\
&= \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}^T \mathbf{x}_0\right) \exp((\mathbf{x}')^T \mathbf{x}) \\
&\stackrel{(5.118)}{=} \exp\left(\frac{1}{2\sigma^2}\|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma}\mathbf{x}^T \mathbf{x}_0\right) R_e(\mathbf{x}', \mathbf{x}) \stackrel{(5.125)}{=} f_{\mathbf{x}}(\mathbf{x}'), \tag{5.127}
\end{aligned}$$

i.e.,

$$\mathsf{K}_e[g_{\mathbf{x}}(\cdot)] = f_{\mathbf{x}}(\cdot). \tag{5.128}$$

for any $\mathbf{x} \in \mathcal{X}_S$. The fact that the mapping $\mathsf{K}_e[\cdot]$ defined in (5.119) is a congruence from $\mathcal{H}(\mathcal{M}_{\text{SSNM}})$ to $\mathcal{H}(R_e)$ follows then from Theorem 3.3.4, since for every argument $\mathbf{x} \in \mathcal{X}_S$ the function value $\mathsf{K}_e[g(\cdot)](\mathbf{x})$ depends continuously on the function value $g(\sigma\mathbf{x})$, which implies that the image $\mathsf{K}_e[g(\cdot)]$ of a function $g(\cdot) \in \mathcal{H}(R_e)$ which is the pointwise limit of a sequence $\{g_l(\cdot) \in \text{span}\{\mathcal{A}\}\}_{l \rightarrow \infty}$, is the pointwise limit of the functions $\mathsf{K}_e[g_l(\cdot)] \in \mathcal{H}(\mathcal{M}_{\text{SSNM}})$.

From Theorem 3.4.1, we obtain (5.121) and (5.122).

The functions $\{g^{(\mathbf{p})}(\cdot)\}_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S}$ are orthogonal since for two difference indices $\mathbf{p}_1, \mathbf{p}_2 \in \mathbb{Z}_+^N \cap \mathcal{X}_S$ there must be at least one index $l \in [N]$ such that $p_{1,l} > p_{2,l}$ or $p_{1,l} < p_{2,l}$. By the symmetry of the inner product we can restrict ourselves without loss of generality to the case $p_{1,l} > p_{2,l}$, and compute the inner product

$$\begin{aligned}
\langle g^{(\mathbf{p}_2)}(\cdot), g^{(\mathbf{p}_1)}(\cdot) \rangle_{\mathcal{H}(R_e)} &= \frac{1}{\sqrt{\mathbf{p}_1!} \sqrt{\mathbf{p}_2!}} \frac{\partial^{\mathbf{p}_1} \mathbf{x}^{\mathbf{p}_2}}{\mathbf{x}^{\mathbf{p}_1}} \Big|_{\mathbf{x}=0} \\
&= \frac{1}{\sqrt{\mathbf{p}_1!} \sqrt{\mathbf{p}_2!}} \left[\prod_{l' \in [N] \setminus \{l\}} \frac{\partial^{p_{1,l'}} \mathbf{x}^{p_{2,l'}}}{\mathbf{x}^{p_{1,l'}}} \Big|_{\mathbf{x}=0} \right] \underbrace{\frac{\partial^{p_{1,l}} \mathbf{x}^{p_{2,l}}}{\mathbf{x}^{p_{1,l}}} \Big|_{\mathbf{x}=0}}_{=0} \\
&= 0. \tag{5.129}
\end{aligned}$$

Since moreover

$$\langle g^{(\mathbf{p})}(\cdot), g^{(\mathbf{p})}(\cdot) \rangle_{\mathcal{H}(R_e)} = \frac{1}{\mathbf{p}!} \frac{\partial^{\mathbf{p}} \mathbf{x}^{\mathbf{p}}}{\mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}=0} = \frac{\mathbf{p}!}{\mathbf{p}!} = 1, \quad (5.130)$$

we have that $\{g^{(\mathbf{p})}(\cdot)\}_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S}$ is an orthonormal set, i.e.,

$$\langle g^{(\mathbf{p}_2)}(\cdot), g^{(\mathbf{p}_1)}(\cdot) \rangle_{\mathcal{H}(R_e)} = \delta_{\mathbf{p}_1, \mathbf{p}_2}. \quad (5.131)$$

The fact that the set $\{g^{(\mathbf{p})}(\cdot)\}_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S}$ is an ONB follows then from Theorem 3.3.6 and the identity

$$\begin{aligned} R_e(\mathbf{x}_1, \mathbf{x}_2) &\stackrel{(5.118)}{=} \exp(\mathbf{x}_1^T \mathbf{x}_2) = \prod_{l \in [N]} \exp(x_{1,l} x_{2,l}) = \prod_{l \in [N]} \sum_{p_l \in \mathbb{Z}_+} \frac{1}{p_l!} (x_{1,l} x_{2,l})^{p_l} \\ &= \sum_{\mathbf{p} \in \mathbb{Z}_+^N} \frac{1}{\mathbf{p}!} \mathbf{x}_1^{\mathbf{p}} \mathbf{x}_2^{\mathbf{p}} \stackrel{(a)}{=} \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} \frac{\mathbf{x}_1^{\mathbf{p}}}{\sqrt{\mathbf{p}!}} \frac{\mathbf{x}_2^{\mathbf{p}}}{\sqrt{\mathbf{p}!}} \stackrel{(5.120)}{=} \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} g^{(\mathbf{p})}(\mathbf{x}_1) g^{(\mathbf{p})}(\mathbf{x}_2), \end{aligned} \quad (5.132)$$

which is valid for every $\mathbf{x}_1, \mathbf{x}_2 \in \mathcal{X}_S$ (for $\mathbf{x}_1, \mathbf{x}_2 \notin \mathcal{X}_S$ the step (a) is not valid).

Finally, the series representation in (5.123) follows from Theorem 3.1.10. \square

Note that while it might seem at first sight that (5.119) is just the specialization of (5.20) to the case $\mathbf{H} = \mathbf{I}$, their meanings are rather different, since the domain of $R_{\mathcal{M}_{\text{SSNM}}}(\cdot, \cdot)$ and $R_e(\cdot, \cdot)$ is $\mathcal{X}_S \times \mathcal{X}_S$ and not $\mathbb{R}^N \times \mathbb{R}^N$, which is the domain of $R_{\mathcal{M}_{\text{LGM}}}(\cdot, \cdot)$ and $R_g^{(N)}(\cdot, \cdot)$ in the special case $\mathbf{H} = \mathbf{I}$. Also note, that the kernel $R_e(\cdot, \cdot)$ in Theorem 5.5.1 is the restriction to the sub-domain $\mathcal{X}_S \times \mathcal{X}_S$ of the kernel $R_g^{(N)}(\cdot, \cdot) : \mathbb{R}^D \times \mathbb{R}^D \rightarrow \mathbb{R}$ defined in Theorem 5.2.2, i.e., $R_e(\cdot, \cdot) = R_g^{(N)}(\cdot, \cdot)|_{\mathcal{X}_S \times \mathcal{X}_S}$. From this we have that the RKHSs $\mathcal{H}(R_e)$ and $\mathcal{H}(R_g^{(N)})$ are related via Theorem 3.3.5.

Based on Theorem 5.5.1, we have the following general characterization of the minimum achievable variance and the corresponding LMV estimator for the minimum variance problem $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}_0)$:

Theorem 5.5.2. *For the minimum variance problem $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}_0)$, the prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$ is valid if and only if there exists a (necessarily unique) coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)$ such that*

$$\gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}) = \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} a[\mathbf{p}] \frac{\mathbf{x}^{\mathbf{p}}}{\sqrt{\mathbf{p}!}} \quad (5.133)$$

for every $\mathbf{x} \in \mathcal{X}_S$, where $\nu_{\mathbf{x}_0}(\mathbf{x})$ as defined in Theorem 5.5.1 and $\gamma(\mathbf{x}) = c(\mathbf{x}) + x_k$ denotes the prescribed mean function. If the prescribed bias $c(\cdot)$ is valid for $\mathcal{M}_{\text{SSNM}}$, then the minimum achievable variance is given by

$$L_{\mathcal{M}_{\text{SSNM}}} = \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} (a_{\mathbf{x}_0}[\mathbf{p}])^2 - (\gamma(\mathbf{x}_0))^2 = \|a_{\mathbf{x}_0}[\mathbf{p}]\|_{\ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)}^2 - (\gamma(\mathbf{x}_0))^2 \quad (5.134)$$

with coefficients

$$a_{\mathbf{x}_0}[\mathbf{p}] \triangleq \langle \gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}), g^{(\mathbf{p})}(\mathbf{x}) \rangle_{\mathcal{H}(R_e)} = \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} (\gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}))}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}=\mathbf{0}}. \quad (5.135)$$

The corresponding LMV estimator $\hat{x}_k^{(\mathbf{x}_0)}(\cdot)$ is obtained as

$$\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) = \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} \frac{a_{\mathbf{x}_0}[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} \psi_{\mathbf{x}_0}(\mathbf{x}; \mathbf{y})}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}=\mathbf{0}} \quad (5.136)$$

where $\psi_{\mathbf{x}_0}(\mathbf{x}; \mathbf{y}) \triangleq \exp\left(\frac{1}{\sigma^2} \mathbf{y}^T (\sigma \mathbf{x} - \mathbf{x}_0) + \frac{1}{\sigma} \mathbf{x}_0^T \mathbf{x} - \frac{1}{2} \|\mathbf{x}\|_2^2\right)$, i.e., $b(\hat{x}_k^{(\mathbf{x}_0)}(\cdot); \mathbf{x}) = c(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{X}_S$ and $v(\hat{x}_k^{(\mathbf{x}_0)}(\cdot); \mathbf{x}_0) = L\mathcal{M}_{\text{SSNM}}$.

Proof. By Theorem 4.3.4, we have that the prescribed bias function $c(\cdot)$ is valid for $\mathcal{M}_{\text{SSNM}}$ if and only if the prescribed mean $\gamma(\cdot)$ belongs to $\mathcal{H}(\mathcal{M}_{\text{SSNM}})$. By Theorem 5.5.1, this is the case if and only if

$$\gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}) \in \mathcal{H}(R_e). \quad (5.137)$$

which, again by Theorem 5.5.1, is the case if and only if there exists a coefficient sequence $a[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)$ that satisfies (5.133) for every $\mathbf{x} \in \mathcal{X}_S$.

Let us assume from now on that the prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$ is valid for $\mathcal{M}_{\text{SSNM}}$, i.e., (5.133) and (5.137) hold. By Theorem 4.3.4, Theorem 5.5.1, and Theorem 3.1.4, we obtain

$$\begin{aligned} L\mathcal{M}_{\text{SSNM}} &\stackrel{(4.19)}{=} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2 - [\gamma(\mathbf{x}_0)]^2 \\ &\stackrel{(5.119)}{=} \|\gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x})\|_{\mathcal{H}(R_e)}^2 - [\gamma(\mathbf{x}_0)]^2 \\ &\stackrel{(3.24)}{=} \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} \left(\langle \gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}), g^{(\mathbf{p})}(\mathbf{x}) \rangle_{\mathcal{H}(R_e)} \right)^2 - [\gamma(\mathbf{x}_0)]^2 \\ &\stackrel{(5.122)}{=} \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} \frac{1}{\mathbf{p}!} \left(\frac{\partial^{\mathbf{p}} (\gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}))}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}=\mathbf{0}} \right)^2 - [\gamma(\mathbf{x}_0)]^2, \end{aligned} \quad (5.138)$$

which verifies (5.134). We then have by (5.133) that

$$\gamma(\mathbf{x}) = \exp\left(\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{x}_0\right) \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} \frac{a_0[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \left(\frac{\mathbf{x}}{\sigma}\right)^{\mathbf{p}}, \quad (5.139)$$

with a suitable coefficient sequence $a_0[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)$. The image $\mathbb{K}_e[\gamma(\cdot)] \in \mathcal{H}(R_e)$ of the function $\gamma(\cdot)$ given by (5.139) is obtained as

$$\mathbb{K}_e[\gamma(\cdot)] = \gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}) = \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} \frac{a_0[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \mathbf{x}^{\mathbf{p}}, \quad (5.140)$$

i.e.,

$$\mathbf{K}_e[\gamma(\cdot)] = \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} \frac{a_0[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \mathbf{x}^{\mathbf{p}} \stackrel{(5.120)}{=} \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} a_0[\mathbf{p}] g^{(\mathbf{p})}(\cdot). \quad (5.141)$$

Since the functions $\{g^{(\mathbf{p})}(\cdot)\}_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S}$ are orthonormal (cf. (5.131)), we have

$$\begin{aligned} a_0[\mathbf{p}] &\stackrel{(5.141)}{=} \langle \mathbf{K}_e[\gamma(\cdot)], g^{(\mathbf{p})}(\mathbf{x}) \rangle_{\mathcal{H}(R_e)} \\ &\stackrel{(5.119)}{=} \langle \gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}), g^{(\mathbf{p})}(\mathbf{x}) \rangle_{\mathcal{H}(R_e)} \stackrel{(5.122)}{=} \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} (\gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}))}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}=\mathbf{0}}, \end{aligned} \quad (5.142)$$

implying also their uniqueness.

Using

$$\begin{aligned} &\exp\left(\frac{1}{\sigma^2} \|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{x}_0\right) \left(\frac{1}{\sigma} \mathbf{x}\right)^{\mathbf{p}} \\ &= \exp\left(\frac{1}{\sigma^2} \|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{x}_0\right) \frac{\partial^{\mathbf{p}} \exp\left(\frac{1}{\sigma} (\mathbf{x}')^T \mathbf{x}\right)}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \\ &= \frac{\partial^{\mathbf{p}} \exp\left(\frac{1}{\sigma^2} \|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{x}_0 + \frac{1}{\sigma} \mathbf{x}^T \mathbf{x}'\right)}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \\ &= \frac{\partial^{\mathbf{p}} \exp\left(\frac{1}{\sigma^2} (\mathbf{x} - \mathbf{x}_0)^T (\sigma \mathbf{x}' - \mathbf{x}_0) + \frac{1}{\sigma} \mathbf{x}_0^T \mathbf{x}'\right)}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \\ &\stackrel{(5.117)}{=} \frac{\partial^{\mathbf{p}} [R_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}, \sigma \mathbf{x}') \exp\left(\frac{1}{\sigma} \mathbf{x}_0^T \mathbf{x}'\right)]}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}}, \end{aligned} \quad (5.143)$$

we can further develop (5.139) as

$$\begin{aligned} \gamma(\mathbf{x}) &= \exp\left(\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2 - \frac{1}{\sigma^2} \mathbf{x}^T \mathbf{x}_0\right) \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} \frac{a_0[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \left(\frac{\mathbf{x}}{\sigma}\right)^{\mathbf{p}} \\ &= \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2\right) \sum_{\mathbf{p} \in \mathbb{Z}_+^N} \frac{a_0[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} [R_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}, \sigma \mathbf{x}') \exp\left(\frac{1}{\sigma} \mathbf{x}_0^T \mathbf{x}'\right)]}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}}. \end{aligned} \quad (5.144)$$

This implies via Corollary 3.4.2, Theorem 4.3.4, Theorem 4.3.3 and the identity

$$\rho_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \mathbf{x}) = \exp\left(\frac{1}{\sigma^2} \mathbf{y}^T (\mathbf{x} - \mathbf{x}_0) - \frac{1}{2\sigma^2} (\|\mathbf{x}\|_2^2 - \|\mathbf{x}_0\|_2^2)\right), \quad (5.145)$$

that the estimator defined by (5.136) is the LMV estimator for $\mathcal{M}_{\text{SSNM}}$:

$$\begin{aligned}
\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) &\triangleq \hat{g}^{(\mathbf{x}_0)}(\mathbf{y}) \stackrel{(4.20)}{=} \mathbf{J}[\gamma(\cdot)] \\
&\stackrel{(5.144)}{=} \mathbf{J} \left[\exp \left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2 \right) \sum_{\mathbf{p} \in \mathbb{Z}_+^N} \frac{a_0[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} [R_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}, \sigma \mathbf{x}') \exp(\frac{1}{\sigma} \mathbf{x}_0^T \mathbf{x}')] }{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \right] \\
&= \exp \left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2 \right) \sum_{\mathbf{p} \in \mathbb{Z}_+^N} \frac{a_0[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \mathbf{J} \left[\frac{\partial^{\mathbf{p}} [R_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}, \sigma \mathbf{x}') \exp(\frac{1}{\sigma} \mathbf{x}_0^T \mathbf{x}')] }{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \right] \\
&\stackrel{(4.16)}{=} \exp \left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2 \right) \sum_{\mathbf{p} \in \mathbb{Z}_+^N} \frac{a_0[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} [\rho_{\mathcal{M}_{\text{SSNM}}}(\cdot, \sigma \mathbf{x}') \exp(\frac{1}{\sigma} \mathbf{x}_0^T \mathbf{x}')] }{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \\
&\stackrel{(5.145)}{=} \sum_{\mathbf{p} \in \mathbb{Z}_+^N} \frac{a_0[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} [\exp(\frac{1}{\sigma^2} \mathbf{y}^T (\sigma \mathbf{x}' - \mathbf{x}_0) - \frac{1}{2} \|\mathbf{x}'\|_2^2 + \frac{1}{\sigma} \mathbf{x}_0^T \mathbf{x}')] }{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}} \\
&= \sum_{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S} \frac{a_{\mathbf{x}_0}[\mathbf{p}]}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} \psi_{\mathbf{x}_0}(\mathbf{x}'; \mathbf{y})}{\partial \mathbf{x}'^{\mathbf{p}}} \Big|_{\mathbf{x}'=\mathbf{0}}. \tag{5.146}
\end{aligned}$$

□

The statement of Theorem 5.5.2 is stronger than that of Theorem 5.3.1 when specialized to $\mathbf{H} = \mathbf{I}$, because it contains explicit expressions for the minimum achievable variance $L_{\mathcal{M}_{\text{SSNM}}}$ and the corresponding LMV estimator $\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y})$. This is possible because of the isometry between the RKHS $\mathcal{H}(\mathcal{M}_{\text{SSNM}})$ and the RKHS $\mathcal{H}(R_e)$, i.e., without the intermediate step via the LGM and Theorem 3.3.5, as used (indirectly through Theorem 5.2.1) in Theorem 5.3.1.

The expression (5.134) nicely demonstrates the influence of the sparsity constraints on minimum variance estimation for the SSNM. Indeed, let us consider the SSNM without any sparsity constraints, i.e., where $S = N$ and the SSNM reduces to the LGM with $\mathbf{H} = \mathbf{I}$. For a fixed prescribed bias $c_0(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}$ and parameter vector $\mathbf{x}_0 \in \mathcal{X}_{S'}$ with some sparsity degree $S' < N$, denote the minimum achievable variance for $\mathcal{M}_{\text{non-sparse}} = (\mathcal{E}_{\text{SSNM}}, c_0(\cdot), \mathbf{x}_0)$ by $L_{\text{non-sparse}}$. Then consider the SSNM with sparsity degree $S' < N$, i.e., where sparsity constraints are present, and denote the minimum achievable variance for $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c_0(\cdot)|_{\mathcal{X}_{S'}}, \mathbf{x}_0) = \mathcal{M}_{\text{non-sparse}}|_{\mathcal{X}_{S'}}$ by $L_{S'}$. Then we have

$$L_{\text{non-sparse}} - L_{S'} = \sum_{\mathbf{p} \in \mathbb{Z}_+^N \setminus (\mathbb{Z}_+^N \cap \mathcal{X}_{S'})} \frac{1}{\mathbf{p}!} \left(\frac{\partial^{\mathbf{p}} [(c_0(\sigma \mathbf{x}) + \sigma x_k) \nu_{\mathbf{x}_0}(\mathbf{x})]}{\partial \mathbf{x}^{\mathbf{p}}} \Big|_{\mathbf{x}=\mathbf{0}} \right)^2. \tag{5.147}$$

By reducing the sparsity degree S' , the set $(\mathbb{Z}_+^N \cap \mathcal{X}_{S'})$ becomes smaller. This in turn implies that number of summands in (5.147) is increased and therefore, since the summands in (5.147) are non-negative, the minimum achievable variance $L_{S'}$ becomes smaller.

Based on Theorem 5.5.2, we have the following obvious result concerning the existence of a UMV estimator for the SSNM:

Lemma 5.5.3. *Consider the SSNM $\mathcal{E}_{\text{SSNM}} = (\mathcal{X}_S, f_{\mathbf{H}=\mathbf{I}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k)$ and a prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$. Then there exists a UMV estimator for $\mathcal{E}_{\text{SSNM}}$ and $c(\cdot)$ if and only if the LMV $\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y})$ for the minimum variance problem $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}_0)$ given explicitly by (5.136) does not depend on \mathbf{x}_0 .*

In what follows, we will need the l th order (probabilists') Hermite polynomial $H_l(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ defined as [88]

$$H_l(x) = (-1)^l e^{x^2/2} \frac{d^l}{dx^l} e^{-x^2/2}. \quad (5.148)$$

We now specialize Theorem 5.5.2 to a specific class of bias functions $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$ that we call “diagonal” bias functions. A diagonal bias function $c(\mathbf{x})$ depends on the parameter \mathbf{x} only through the k th component (where k is the same index as used for the parameter function of the SSNM $\mathcal{E}_{\text{SSNM}} = (\mathcal{X}_S, f_{\mathbf{H}=\mathbf{I}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k)$), i.e., $c(\mathbf{x}) = \tilde{c}(x_k)$ with some function $\tilde{c}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ that may vary with the index k used for the definition of the parameter function of $\mathcal{E}_{\text{SSNM}}$. Similarly, we say that an estimator $\hat{x}_k(\mathbf{y})$ for the SSNM is diagonal if it depends on the observation \mathbf{y} only via the k th entry, i.e., $\hat{x}_k(\mathbf{y}) = \hat{x}_k(y_k)$. Obviously, the bias function of a diagonal estimator for the SSNM is diagonal. The well-known hard- and soft-thresholding estimators are diagonal estimators. However, the maximum likelihood (ML) estimator for the SSNM is not diagonal.

Making the weak assumption that $\tilde{c}(x_k) : \mathbb{R} \rightarrow \mathbb{R}$ can be represented by a power-series which converges everywhere on \mathbb{R} , we obtain

Theorem 5.5.4. *Consider the minimum variance problem $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}_0)$ for some choice of σ , S , M , N , \mathbf{x}_0 and for a diagonal prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$, i.e., $c(\mathbf{x}) = \tilde{c}(x_k)$. Moreover, we assume that the function $\tilde{c}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ is such that the prescribed mean function $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = \tilde{c}(x_k) + x_k$ can be written as a power series centered at \mathbf{x}_0 and which converges everywhere, i.e.,*

$$\gamma(\mathbf{x}) = \sum_{l \in \mathbb{Z}_+} \frac{m_l}{l!} (x_k - x_{0,k})^l \quad (5.149)$$

with suitable coefficients $m_l \in \mathbb{R}$. We then have that the bias $c(\cdot)$ is valid for $\mathcal{M}_{\text{SSNM}}$ if and only if

$$\sum_{l \in \mathbb{Z}_+} \frac{(m_l)^2 \sigma^{2l}}{l!} < \infty. \quad (5.150)$$

If the coefficients m_l satisfy (5.150), then we have for the case where $|\text{supp}(\mathbf{x}_0) \cup \{k\}| < S+1$ that

$$L_{\mathcal{M}_{\text{SSNM}}} = \sum_{l \in \mathbb{N}} \frac{(m_l)^2}{l!} \sigma^{2l}, \quad (5.151)$$

with the corresponding LMV estimator

$$\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) = \sum_{l \in \mathbb{Z}_+} \frac{m_l}{l!} \sigma^l \text{H}_l \left(\frac{y_k - x_{0,k}}{\sigma} \right), \quad (5.152)$$

i.e., $b(\hat{x}_k^{(\mathbf{x}_0)}(\cdot); \mathbf{x}) = \gamma(\mathbf{x}) - x_k = \tilde{c}(x_k)$ for every $\mathbf{x} \in \mathcal{X}_S$ and $v(\hat{x}_k^{(\mathbf{x}_0)}(\cdot); \mathbf{x}_0) = L_{\mathcal{M}_{\text{SSNM}}}$. For the complementary case where $|\text{supp}(\mathbf{x}_0) \cup \{k\}| = S+1$, we have, by denoting the indices of the support of \mathbf{x}_0 as $\text{supp}(\mathbf{x}_0) = \{i_1, \dots, i_S\}$, that

$$L_{\mathcal{M}_{\text{SSNM}}} = \left(\sum_{l \in \mathbb{Z}_+} \frac{m_l^2 \sigma^{2l}}{l!} \right) \left[\sum_{j \in [S]} \exp \left(-\frac{x_{0,i_j}^2}{\sigma^2} \right) \prod_{j' \in [j-1]} \left[1 - \exp \left(-\frac{x_{0,i_{j'}}^2}{\sigma^2} \right) \right] \right] - [\gamma(\mathbf{x}_0)]^2, \quad (5.153)$$

with the corresponding LMV estimator

$$\begin{aligned} \hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) = & \left[\sum_{l \in \mathbb{Z}_+} \frac{m_l}{l!} \sigma^l \text{H}_l \left(\frac{y_k}{\sigma} \right) \right] \sum_{j \in [S]} \exp \left(-\frac{x_{0,i_j}^2 + 2y_{i_j} x_{0,i_j}}{2\sigma^2} \right) \prod_{j' \in [j-1]} \left[1 - \exp \left(-\frac{x_{0,i_{j'}}^2 + 2y_{i_{j'}} x_{0,i_{j'}}}{2\sigma^2} \right) \right]. \end{aligned} \quad (5.154)$$

Proof. Appendix A. □

A slight reformulation of Theorem 5.5.4 yields

Corollary 5.5.5. *Consider the minimum variance problem $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}_0)$ for some choice of σ , S , N , $\mathbf{x}_0 \in \mathcal{X}_S$, and for a diagonal prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$, i.e., $c(\mathbf{x}) = \tilde{c}(x_k)$. Moreover we assume that the bias function $c(\cdot)$ is valid and such that the prescribed mean function $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = \tilde{c}(x_k) + x_k$ can be written as a power series centered at $x_{0,k}$ and which converges everywhere, i.e., we have $\gamma(\mathbf{x}) = \sum_{l \in \mathbb{Z}_+} \frac{m_l}{l!} (x_k - x_{0,k})^l$ with coefficients m_l satisfying (5.150). Then we obtain for the minimum achievable variance*

$$L_{\mathcal{M}_{\text{SSNM}}} = t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_0) \sum_{l \in \mathbb{Z}_+} \frac{m_l^2 \sigma^{2l}}{l!} - [\gamma(\mathbf{x}_0)]^2, \quad (5.155)$$

and for the corresponding LMV estimator

$$\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) = h_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \mathbf{x}_0) \sum_{l \in \mathbb{Z}_+} \frac{m_l}{l!} \sigma^l \text{H}_l \left(\frac{y_k - x_{0,k}}{\sigma} \right), \quad (5.156)$$

where

$$t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}) \triangleq \begin{cases} \sum_{j \in [S]} \exp\left(-\frac{x_{i_j}^2}{\sigma^2}\right) \prod_{j' \in [j-1]} \left[1 - \exp\left(-\frac{x_{i_{j'}}^2}{\sigma^2}\right)\right] \leq 1, & \text{when } |\text{supp}(\mathbf{x}_0) \cup \{k\}| = S + 1, \\ 1, & \text{when } |\text{supp}(\mathbf{x}_0) \cup \{k\}| < S + 1 \end{cases} \quad (5.157)$$

$$h_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \mathbf{x}) \triangleq \begin{cases} \sum_{j \in [S]} \exp\left(-\frac{x_{i_j}^2 + 2y_{i_j} x_{i_j}}{2\sigma^2}\right) \prod_{j' \in [j-1]} \left[1 - \exp\left(-\frac{x_{i_{j'}}^2 + 2y_{i_{j'}} x_{i_{j'}}}{2\sigma^2}\right)\right], & \text{when } |\text{supp}(\mathbf{x}_0) \cup \{k\}| = S + 1, \\ 1, & \text{when } |\text{supp}(\mathbf{x}_0) \cup \{k\}| < S + 1 \end{cases} \quad (5.158)$$

with an arbitrary index set $\mathcal{I} = \{i_1, \dots, i_S\}$ such that $|\mathcal{I}| = S$, $k \notin \mathcal{I}$ and $\text{supp}(\mathbf{x}_0) \setminus \{k\} \subseteq \mathcal{I}$.

Proof. The relation (5.155) follows from (5.151) and (5.153) since $\gamma(\mathbf{x}_0) = m_0$, which implies that

$$\sum_{l \in \mathbb{Z}_+} \frac{m_l^2 \sigma^{2l}}{l!} - [\gamma(\mathbf{x}_0)]^2 = \sum_{l \in \mathbb{N}} \frac{m_l^2 \sigma^{2l}}{l!}. \quad (5.159)$$

The relation (5.156) follows from (5.152) and (5.154), since for the case $|\text{supp}(\mathbf{x}_0) \cup \{k\}| = S + 1$ we must have $k \notin \text{supp}(\mathbf{x}_0)$, i.e., $x_{0,k} = 0$. Indeed, if $k \in \text{supp}(\mathbf{x})$ we would have $\text{supp}(\mathbf{x}_0) \cup \{k\} = \text{supp}(\mathbf{x}_0)$, and since the cardinality of the support $\text{supp}(\mathbf{x}_0)$ cannot be larger than S ($\mathbf{x}_0 \in \mathcal{X}_S$), this would mean that $|\text{supp}(\mathbf{x}_0) \cup \{k\}| = |\text{supp}(\mathbf{x}_0)| \leq S < S + 1$.

The inequality $t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}) \leq 1$ in (5.157) can be verified by observing that

$$\begin{aligned} & \sum_{j \in [S]} \exp\left(-\frac{x_{i_j}^2}{\sigma^2}\right) \prod_{j' \in [j-1]} \left(1 - \exp\left(-\frac{x_{i_{j'}}^2}{\sigma^2}\right)\right) \\ &= \left[\prod_{j'' \in [S]} \exp\left(-\frac{x_{i_{j''}}^2}{\sigma^2}\right) \right] \sum_{j \in [S]} \left[\prod_{j''' \in [S]} \exp\left(\frac{x_{i_{j'''}}^2}{\sigma^2}\right) \right] \exp\left(-\frac{x_{i_j}^2}{\sigma^2}\right) \left[\prod_{j' \in [j-1]} \left(1 - \exp\left(-\frac{x_{i_{j'}}^2}{\sigma^2}\right)\right) \right] \\ &= \left[\prod_{j'' \in [S]} \exp\left(-\frac{x_{i_{j''}}^2}{\sigma^2}\right) \right] \sum_{j \in [S]} \left[\prod_{j''' \in [S] \setminus [j]} \exp\left(\frac{x_{i_{j'''}}^2}{\sigma^2}\right) \right] \left[\prod_{j' \in [j-1]} \left(\exp\left(\frac{x_{i_{j'}}^2}{\sigma^2}\right) - 1\right) \right] \\ &\stackrel{(a)}{=} \left[\prod_{j'' \in [S]} \exp\left(-\frac{x_{i_{j''}}^2}{\sigma^2}\right) \right] \sum_{j \in [S]} \sum_{\substack{\mathbf{k} \in \mathbb{Z}_+^S \\ k_j = 0 \\ k_{j''} > 0, j'' \in [S] \setminus [j]}} \prod_{j' \in [S]} \left(\frac{x_{i_{j'}}^2}{\sigma^2}\right)^{k_{j'}} \end{aligned}$$

$$\leq \left[\prod_{j'' \in [S]} \exp\left(-\frac{x_{i_{j''}}^2}{\sigma^2}\right) \right] \sum_{\mathbf{k} \in \mathbb{Z}_+^S} \prod_{j' \in [S]} \left(\frac{x_{i_{j'}}^2}{\sigma^2}\right)^{k_{j'}}$$

$$\stackrel{(b)}{=} \left[\prod_{j'' \in [S]} \exp\left(-\frac{x_{i_{j''}}^2}{\sigma^2}\right) \right] \left[\prod_{j \in [S]} \exp\left(\frac{x_{i_j}^2}{\sigma^2}\right) \right] = 1,$$

where the step (a) and (b) can be verified by the distributive law and the series representation $\exp\left(\frac{x_{i_j}^2}{\sigma^2}\right) = \sum_{l \in \mathbb{Z}_+} \frac{1}{l!} \left(\frac{x_{i_j}^2}{\sigma^2}\right)^l$. \square

Note that the coefficients m_l appearing in the representation (5.149) of the prescribed mean function depend on the parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$ associated with the minimum variance problem $\mathcal{M}_{\text{SSNM}}$, because the power series representing $\gamma(\mathbf{x})$ is centered at \mathbf{x}_0 .

If the prescribed bias function $c(\cdot)$ is the actual bias function $b(\hat{x}'_k(\cdot); \mathbf{x})$ of a given diagonal estimator $\hat{x}'_k(\mathbf{y}) = \hat{x}'_k(y_k)$ with finite variance at \mathbf{x}_0 , i.e., $\hat{x}'_k(\cdot) \in \mathcal{P}(\mathcal{M}_{\text{SSNM}})$ (see (4.2)), the coefficients m_l appearing in Corollary 5.5.5 and Theorem 5.5.4 have a particular interpretation.

In what follows, we need the following result:

Lemma 5.5.6. *Consider the minimum variance problem $\mathcal{M}_{\text{SSNM}}(\mathbf{x}) = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x})$ and the associated function Hilbert space $\mathcal{P}(\mathcal{M}_{\text{SSNM}})$ of finite-variance estimators with inner product*

$$\begin{aligned} \langle \hat{g}_1(\cdot), \hat{g}_2(\cdot) \rangle_{\text{RV}} &\triangleq \mathbf{E}_{\mathbf{x}} \{ \hat{g}_1(\mathbf{y}) \hat{g}_2(\mathbf{y}) \} \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \int_{\mathbf{y} \in \mathbb{R}^N} \hat{g}_1(\mathbf{y}) \hat{g}_2(\mathbf{y}) \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{x}\|_2^2\right) d\mathbf{y}. \end{aligned} \quad (5.160)$$

Then the subset $\mathcal{D}(\mathbf{x}) \subseteq \mathcal{P}(\mathcal{M}_{\text{SSNM}})$ which consists of all diagonal estimators $\hat{g}(\mathbf{y}) = \hat{g}(y_k)$ is a Hilbert subspace with induced inner product

$$\begin{aligned} \langle \hat{g}_1(\cdot), \hat{g}_2(\cdot) \rangle_{\mathcal{D}(\mathbf{x})} &\triangleq \mathbf{E}_{\mathbf{x}_0} \{ \hat{g}_1(y_k) \hat{g}_2(y_k) \} \\ &= \frac{1}{\sqrt{2\pi}\sigma} \int_{y_k \in \mathbb{R}} \hat{g}_1(y_k) \hat{g}_2(y_k) \exp\left(-\frac{1}{2\sigma^2} (y_k - x_k)^2\right) dy_k. \end{aligned} \quad (5.161)$$

A specific ONB for $\mathcal{D}(\mathbf{x})$ is given by $\{h^{(l)}(\cdot)\}_{l \in \mathbb{Z}_+}$ with functions $h^{(l)}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}$ defined as

$$h^{(l)}(\mathbf{y}) \triangleq \frac{1}{\sqrt{l!}} \text{H}_l\left(\frac{y_k - x_k}{\sigma}\right). \quad (5.162)$$

Proof. [89] \square

We have the following straightforward consequence of Corollary 5.5.5:

Corollary 5.5.7. Consider the minimum variance problem $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}_0)$ for some choice of σ , S , N , $\mathbf{x}_0 \in \mathcal{X}_S$, and assume that the prescribed bias function $c(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R}$ is given as the actual bias function of a diagonal estimator $\hat{x}_k(\mathbf{y}) = \hat{x}_k(y_k)$, i.e., $c(\mathbf{x}) = b(\hat{x}_k(\cdot); \mathbf{x})$. The estimator $\hat{x}_k(\mathbf{y})$ is assumed to have finite variance and stochastic power at all parameter vectors $\mathbf{x} \in \mathcal{X}_S$, i.e., $\hat{x}_k(\cdot) \in \mathcal{D}(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{X}_S$. Then, the prescribed mean function $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = c(\mathbf{x}) + x_k = \mathbf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\}$ can be written as a power series (5.149) which converges everywhere. The coefficients m_l in (5.149) are given by scaled expansion coefficients of $\hat{x}_k(\mathbf{y})$ w.r.t. the ONB $\{h^{(l)}(\cdot)\}_{l \in \mathbb{Z}_+}$ defined in (5.162), i.e.,

$$m_l = \frac{\sqrt{l!}}{\sigma^l} \langle \hat{x}_k(\cdot), h^{(l)}(\cdot) \rangle_{\mathcal{D}(\mathcal{M}_{\text{SSNM}})}. \quad (5.163)$$

The prescribed bias $c(\cdot)$ is valid by construction. The minimum achievable variance and corresponding LMV estimator are obtained as

$$L_{\mathcal{M}_{\text{SSNM}}} = v(\hat{x}_k(\cdot); \mathbf{x}_0) t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_0) + [t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_0) - 1] [\gamma(\mathbf{x}_0)]^2, \quad (5.164)$$

$$\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) = \hat{x}_k(y_k) h_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \mathbf{x}_0), \quad (5.165)$$

respectively, with the functions $t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x})$, $h_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \mathbf{x})$ as defined in (5.157), and (5.158), respectively.

Proof. Since the functions $\{h^{(l)}(\cdot)\}_{l \in \mathbb{Z}_+}$ form an ONB for $\mathcal{D}(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{X}_S$, we have by Theorem 3.1.4:

$$\begin{aligned} \sum_{l \in \mathbb{Z}_+} \left[\langle \hat{x}_k(\cdot), h^{(l)}(\cdot) \rangle_{\mathcal{D}(\mathbf{x})} \right]^2 &\stackrel{(3.24)}{=} \|\hat{x}_k(\cdot)\|_{\mathcal{D}(\mathbf{x})}^2 \stackrel{(5.161)}{=} P(\hat{x}_k(\cdot); \mathbf{x}) = v(\hat{x}_k(\cdot); \mathbf{x}) + [\mathbf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\}]^2 \\ &= v(\hat{x}_k(\cdot); \mathbf{x}) + [\gamma(\mathbf{x})]^2. \end{aligned} \quad (5.166)$$

and

$$\sum_{l \in \mathbb{Z}_+} \langle \hat{x}_k(\cdot), h^{(l)}(\cdot) \rangle_{\mathcal{D}(\mathbf{x})} h^{(l)}(\cdot) = \hat{x}_k(\cdot). \quad (5.167)$$

The prescribed mean function can be expressed as

$$\gamma(\mathbf{x}) = \mathbf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\} = \frac{1}{\sqrt{2\pi}\sigma} \int_{y_k \in \mathbb{R}} \hat{x}_k(y_k) \exp\left(-\frac{1}{2\sigma^2}(y_k - x_k)^2\right) dy_k. \quad (5.168)$$

Note that $\gamma(\mathbf{x})$ depends only on x_k , i.e., $\gamma(\mathbf{x}) = \gamma(x_k)$. Differentiating the mean function l times w.r.t. x_k yields

$$\frac{\partial^{l\mathbf{e}_k} \gamma(\mathbf{x})}{\partial \mathbf{x}^{l\mathbf{e}_k}} \stackrel{(5.168)}{=} \frac{1}{\sqrt{2\pi}\sigma} \frac{\partial^l \int_{y_k \in \mathbb{R}} \hat{x}_k(y_k) \exp\left(-\frac{1}{2\sigma^2}(y_k - x_k)^2\right) dy_k}{\partial x_k^l}$$

$$\begin{aligned}
&\stackrel{(a)}{=} \frac{1}{\sqrt{2\pi}\sigma} \int_{y_k \in \mathbb{R}} \hat{x}_k(y_k) \frac{\partial^l \exp\left(-\frac{1}{2\sigma^2}(y_k - x_k)^2\right)}{\partial x_k^l} dy_k \\
&\stackrel{(b)}{=} \frac{1}{\sqrt{2\pi}\sigma^{l+1}} \int_{y_k \in \mathbb{R}} \hat{x}_k(y_k) \text{H}_l\left(\frac{y_k - x_k}{\sigma}\right) \exp\left(-\frac{1}{2\sigma^2}(y_k - x_k)^2\right) dy_k \\
&= \frac{\sqrt{l!}}{\sigma^l} \langle \hat{x}_k(\cdot), h^{(l)}(\cdot) \rangle_{\mathcal{D}(\mathbf{x})}, \tag{5.169}
\end{aligned}$$

where (a) follows from changing the order of differentiation and integration (cf. [41, Theorem 1.5.8]), which be verified, e.g., by a standard argument using the dominated convergence theorem [5, 61, 73]. The step (b) can be verified by a calculation similar to (A.9).

Since (by the assumption $\hat{x}_k(\cdot) \in \mathcal{D}(\mathbf{x})$ for every $\mathbf{x} \in \mathcal{X}_S$) it holds that $P(\hat{x}_k(\cdot); \mathbf{x}) < \infty$ for every $\mathbf{x} \in \mathcal{X}_S$, we have from (5.166) that the inner products $\langle \hat{x}_k(\cdot), h^{(l)}(\cdot) \rangle_{\mathcal{D}(\mathbf{x})}$ must be bounded by some constant C (strictly speaking, the sequence of the inner products has to decay faster than $1/l$). This in turn yields by (5.169), that the l th order partial derivative of the mean function $\gamma(\mathbf{x})$ is bounded by $C\frac{\sqrt{l!}}{\sigma^l}$, which implies by Taylor's theorem [5] that the mean function can be written as a power series in the form (5.149) with coefficients

$$m_l = \frac{\sqrt{l!}}{\sigma^l} \langle \hat{x}_k(\cdot), h^{(l)}(\cdot) \rangle_{\mathcal{D}(\mathbf{x}_0)}. \tag{5.170}$$

The power series is centered at \mathbf{x}_0 and converges everywhere on \mathcal{X}_S .

The relation (5.164) follows then from (5.155) via (5.166) and (5.170) as

$$\begin{aligned}
L_{\mathcal{M}_{\text{SSNM}}} &\stackrel{(5.155)}{=} t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_0) \sum_{l \in \mathbb{Z}_+} \frac{m_l^2 \sigma^{2l}}{l!} - [\gamma(\mathbf{x}_0)]^2 \\
&\stackrel{(5.170)}{=} t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_0) \sum_{l \in \mathbb{Z}_+} \left[\frac{\sqrt{l!}}{\sigma^l} \langle \hat{x}_k(\cdot), h^{(l)}(\cdot) \rangle_{\mathcal{D}(\mathbf{x}_0)} \right]^2 \frac{\sigma^{2l}}{l!} - [\gamma(\mathbf{x}_0)]^2 \\
&= t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_0) \sum_{l \in \mathbb{Z}_+} [\langle \hat{x}_k(\cdot), h^{(l)}(\cdot) \rangle_{\mathcal{D}(\mathbf{x}_0)}]^2 - [\gamma(\mathbf{x}_0)]^2 \\
&\stackrel{(5.166)}{=} t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_0) v(\hat{x}_k(\cdot); \mathbf{x}_0) + t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_0) [\gamma(\mathbf{x}_0)]^2 - [\gamma(\mathbf{x}_0)]^2 \\
&\stackrel{(5.166)}{=} v(\hat{x}_k(\cdot); \mathbf{x}_0) t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_0) + [t_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_0) - 1] [\gamma(\mathbf{x}_0)]^2. \tag{5.171}
\end{aligned}$$

Finally, the relation (5.165) follows from (5.156) by (5.167) as

$$\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) \stackrel{(5.156)}{=} h_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \mathbf{x}_0) \sum_{l \in \mathbb{Z}_+} \frac{m_l}{l!} \sigma^l \text{H}_l\left(\frac{y_k - x_{0,k}}{\sigma}\right)$$

$$\begin{aligned}
& \stackrel{(5.162)}{=} h_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \mathbf{x}_0) \sum_{l \in \mathbb{Z}_+} \frac{m_l}{\sqrt{l!}} \sigma^l h^{(l)}(\mathbf{y}) \\
& \stackrel{(5.170)}{=} h_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \mathbf{x}_0) \sum_{l \in \mathbb{Z}_+} \langle \hat{x}_k(\cdot), h^{(l)}(\cdot) \rangle_{\mathcal{D}(\mathbf{x}_0)} h^{(l)}(\mathbf{y}) \\
& \stackrel{(5.167)}{=} h_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \mathbf{x}_0) \hat{x}_k(\mathbf{y}). \tag{5.172}
\end{aligned}$$

□

Note that Corollary 5.5.7 also applies to unbiased estimation where $c(\cdot) \equiv 0$ and $\gamma(\mathbf{x}) = x_k$, since these are the bias and mean functions of the ordinary LS estimator $\hat{x}_{\text{LS},k}(\mathbf{y}) = y_k$, which obviously is diagonal and has finite variance at every $\mathbf{x} \in \mathcal{X}_S$.

Remarkably, according to Corollary 5.5.7 (in particular, due to (5.157) and (5.158)), every diagonal estimator $\hat{x}_k(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}$ for the SSNM with finite variance at $\mathbf{x}_0 \in \mathcal{X}_S$ is necessarily the LMV estimator for the minimum variance problem $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\mathbf{x}) = b(\hat{x}_k(\cdot); \mathbf{x}), \mathbf{x}_0)$ whenever $|\text{supp}(\mathbf{x}_0) \cup \{k\}| < S + 1$, irrespective of the sparsity degree S . Thus, in this case, the sparsity information does not help anything with regard to minimum variance estimation, i.e., the estimator $\hat{x}_k(\cdot)$ is the LMV estimator for any parameter set \mathcal{X}_S including the non-sparse case $\mathcal{X} = \mathbb{R}^N$ and the variance $v(\hat{x}_k(\cdot); \mathbf{x}_0)$ coincides with the minimum achievable variance $L_{\mathcal{M}_{\text{SSNM}}}$, i.e., the Barankin bound.

However, whenever $|\text{supp}(\mathbf{x}_0) \cup \{k\}| = S + 1$, we have by Corollary 5.5.7 and the inequality in (5.157) that there are estimators with the same bias and mean as $\hat{x}_k(\cdot)$ but with a smaller variance at \mathbf{x}_0 . The LMV estimator $\hat{x}_k^{(\mathbf{x}_0)}(\cdot)$ that achieves the minimum variance at \mathbf{x}_0 is obtained by a multiplication of the estimator $\hat{x}_k(\mathbf{y}) = \hat{x}_k(y_k)$ with a ‘‘correction’’ factor $h_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \mathbf{x}_0)$ that does not depend on y_k . Note that by a straightforward calculation one can show that the LMV estimator in (5.165) is robust against deviations from the nominal parameter \mathbf{x}_0 . In particular, that estimator has always finite mean and variance for any parameter vector $\mathbf{x} \in \mathbb{R}^N$ and an observation $\mathbf{y} = \mathbf{x} + \mathbf{n}$ that follows the statistical model of the LGM with $\mathbf{H} = \mathbf{I}$ and an arbitrary noise variance σ^2 . We verify this claim separately for the two complementary cases $|\text{supp}(\mathbf{x}_0) \cup \{k\}| = S + 1$ and $|\text{supp}(\mathbf{x}_0) \cup \{k\}| < S + 1$. First consider the case $|\text{supp}(\mathbf{x}_0) \cup \{k\}| = S + 1$, where, as already observed above, necessarily

$$\{k\} \notin \text{supp}(\mathbf{x}_0) = \{i_1, \dots, i_S\}, \tag{5.173}$$

i.e., $x_{0,k} = 0$ and moreover $|\text{supp}(\mathbf{x}_0)| = \|\mathbf{x}_0\|_0 = S$. Given an arbitrary vector $\mathbf{x} \in \mathbb{R}^N$, the stochastic power $P(\hat{x}_k^{(\mathbf{x}_0)}(\cdot); \mathbf{x})$ at \mathbf{x} of the LMV estimator in (5.165) is then obtained as

$$P(\hat{x}_k^{(\mathbf{x}_0)}(\cdot); \mathbf{x}) = \mathbf{E}_{\mathbf{x}} \left\{ \left[\hat{x}_k^{(\mathbf{x}_0)}(\cdot) \right]^2 \right\}$$

$$\begin{aligned}
& \stackrel{(5.165)}{=} \mathbf{E}_{\mathbf{x}} \left\{ \left[\hat{x}_k(y_k) h_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \mathbf{x}_0) \right]^2 \right\} \\
& \stackrel{(5.158)}{=} \mathbf{E}_{\mathbf{x}} \left\{ \left[\hat{x}_k(y_k) \sum_{j \in [S]} \alpha_{i_j} \prod_{j' \in [j-1]} (1 - \alpha_{i_{j'}}) \right]^2 \right\} \tag{5.174}
\end{aligned}$$

$$\stackrel{(5.158)}{=} \mathbf{E}_{\mathbf{x}} \left\{ \left[\hat{x}_k(y_k) \sum_{j \in [S]} \alpha_{i_j} \prod_{j' \in [j-1]} (1 - \alpha_{i_{j'}}) \right]^2 \right\} \tag{5.175}$$

$$\begin{aligned}
& \stackrel{(5.173)}{=} \underbrace{\mathbf{E}_{\mathbf{x}} \left\{ \left[\hat{x}_k(y_k) \right]^2 \right\}}_{< \infty} \mathbf{E}_{\mathbf{x}} \left\{ \left[\sum_{j \in [S]} \alpha_{i_j} \prod_{j' \in [j-1]} (1 - \alpha_{i_{j'}}) \right]^2 \right\} \\
& = \mathbf{E}_{\mathbf{x}} \left\{ \left[\hat{x}_k(y_k) \right]^2 \right\} \mathbf{E}_{\mathbf{x}} \left\{ \sum_{j, j'' \in [S]} \alpha_{i_j} \alpha_{i_{j''}} \left[\prod_{j' \in [j-1]} (1 - \alpha_{i_{j'}}) \right] \left[\prod_{j''' \in [j''-1]} (1 - \alpha_{i_{j'''}}) \right] \right\} < \infty \tag{5.176}
\end{aligned}$$

where we introduced the shorthand $\alpha_i \triangleq \exp\left(-\frac{x_{0,i}^2 + 2y_i x_{0,i}}{2\sigma^2}\right)$ and the last step follows by collecting terms, and the fact that

$$\mathbf{E}_{\mathbf{x}} \left\{ \prod_{i \in \text{supp}(\mathbf{x}_0)} [\alpha_i]^{n_i} \right\} = \prod_{i \in \text{supp}(\mathbf{x}_0)} \mathbf{E}_{\mathbf{x}} \{ [\alpha_i]^{n_i} \} < \infty, \tag{5.177}$$

for any nonnegative exponents $n_i \in \mathbb{Z}_+$. Similarly, one can also show that the mean of $\hat{x}_k^{(\mathbf{x}_0)}(\cdot)$ is finite at every $\mathbf{x} \in \mathbb{R}^N$, i.e., $\mathbf{E}_{\mathbf{x}} \left\{ \hat{x}_k^{(\mathbf{x}_0)}(\cdot) \right\} < \infty$.

5.6 Necessity of Strict Sparsity

In what follows, we consider the LGM \mathcal{E}_{LGM} with a fixed choice of σ , M , N , \mathbf{H} , and a fixed choice of the sparsity degree S , and the associated minimum variance problem $\mathcal{M}_{\text{LGM}} = (\mathcal{E}_{\text{LGM}}, \tilde{c}(\cdot), \mathbf{x}_0)$ with parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$ and prescribed bias $\tilde{c}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}$. Since the SLM \mathcal{E}_{SLM} for $S < N$ is obtained by reducing the parameter set $\mathcal{X} = \mathbb{R}^N$ of the LGM \mathcal{E}_{LGM} to the set of S -sparse vectors \mathcal{X}_S , the set of allowed estimators $\mathcal{F}(\mathcal{M}_{\text{SLM}})$ (see (2.18)) for the minimum variance problem $\mathcal{M}_{\text{SLM}} = \mathcal{M}_{\text{LGM}}|_{\mathcal{X}_S}$ is in general larger than the set of allowed estimators $\mathcal{F}(\mathcal{M}_{\text{LGM}})$ for \mathcal{M}_{LGM} , i.e., $\mathcal{F}(\mathcal{M}_{\text{SLM}}) \supseteq \mathcal{F}(\mathcal{M}_{\text{LGM}})$.

It is this increase of the set of allowed estimators that causes the minimum achievable variance $L_{\mathcal{M}_{\text{SLM}}}$ (for $S < N$) to be strictly smaller than $L_{\mathcal{M}_{\text{LGM}}}$ in general. This fact agrees with the discussion of Section 4.3.4 according to which we have $L_{\mathcal{M}_{\text{SLM}}} \leq L_{\mathcal{M}_{\text{LGM}}}$. An important exception

to this rule is the special case of an affine prescribed bias, i.e., $c(\mathbf{x}) = \mathbf{a}^T \mathbf{x} + b$ with a fixed vector $\mathbf{a} \in \mathbb{R}^N$ and number $b \in \mathbb{R}$, and a full column rank system matrix \mathbf{H} . In this case, the lower bound in Theorem 5.4.1 for the case $\|\mathbf{x}_0\|_0 < S$ coincides with $L_{\mathcal{M}_{\text{LGM}}}$. Therefore, the minimum achievable variance $L_{\mathcal{M}_{\text{SLM}}}$ for the SLM coincides with the minimum achievable variance $L_{\mathcal{M}_{\text{LGM}}}$ for the LGM (since $L_{\mathcal{M}_{\text{SLM}}} \leq L_{\mathcal{M}_{\text{LGM}}}$).

Note that the reduction from the parameter set $\mathcal{X} = \mathbb{R}^N$ of the LGM \mathcal{E}_{LGM} to the parameter set $\mathcal{X} = \mathcal{X}_S$ of the SLM \mathcal{E}_{SLM} with $S < N$ is rather large, since the Lebesgue measure of \mathcal{X}_S (with $S < N$) is zero w.r.t. \mathbb{R}^N [73].

Intuitively, it should make not a big difference if we relax the constraint $\mathbf{x} \in \mathcal{X}_S$ and include also *approximately* S -sparse vectors in the parameter set. Thus, we prescribe the estimator bias not only for strictly but also for approximately S -sparse vectors. A specific definition of approximate S -sparsity is used, e.g., in [22, 23, 48], where the authors model the set of approximately S -sparse vectors by an ℓ_q ball of radius S , where $0 \leq q \leq 1$. The ℓ_q ball $\mathcal{B}_q(S)$ of radius S is defined by

$$\mathcal{B}_q(S) \triangleq \left\{ \mathbf{x}' \in \mathbb{R}^N \mid \|\mathbf{x}'\|_q^q \leq S \right\}. \quad (5.178)$$

Note that the set \mathcal{X}_S of strictly sparse vectors is obtained as the limiting case when $q = 0$, i.e., $\mathcal{X}_S = \mathcal{B}_0(S)$.

In Figure 5.1, we illustrate $\mathcal{B}_q(S)$ for $N = 2$, $S = 1$ and various q . Note also that in contrast to \mathcal{X}_S , the parameter sets $\mathcal{B}_q(S)$ are bounded for $q > 0$, i.e., for every $q > 0$ and S there exists a radius r such that $\mathcal{B}_q(S) \subseteq \mathcal{B}(\mathbf{0}, r) \subseteq \mathbb{R}^N$.

Let us now define the minimum variance problem $\mathcal{M}^{(q)} \triangleq \mathcal{M}_{\text{LGM}}|_{\mathcal{B}_q(S)}$, which is obtained from \mathcal{M}_{LGM} by reducing the parameter set from $\mathcal{X} = \mathbb{R}^N$ to $\mathcal{X} = \mathcal{B}_q(S)$. Note that $\mathcal{M}_{\text{SLM}} = \mathcal{M}^{(0)}$. Then, one might expect that for $q \rightarrow 0$, i.e., when $\mathcal{B}_q(S) \rightarrow \mathcal{X}_S$, we should have $L_{\mathcal{M}^{(q)}} \rightarrow L_{\mathcal{M}_{\text{SLM}}}$.

Despite the intuition that a change of the parameter set from \mathcal{X}_S to $\mathcal{B}_q(S)$ with small $q > 0$, i.e., considering $\mathcal{M}^{(q)}$ instead of \mathcal{M}_{SLM} , should make not a big difference, the next result tells us that for minimum variance estimation it makes a very big difference if we use \mathcal{X}_S or $\mathcal{B}_q(S)$, no matter how small q is.

Theorem 5.6.1. *Consider a parameter set $\mathcal{X} \subseteq \mathbb{R}^N$ and the minimum variance problem $\mathcal{M}' = \mathcal{M}_{\text{LGM}}|_{\mathcal{X}} = (\mathcal{E}', \tilde{c}(\cdot)|_{\mathcal{X}}, \mathbf{x}_0 \in \mathcal{X})$ associated with the estimation problem $\mathcal{E}' = (\mathcal{X}, f_{\mathbf{H}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k)$. Note that the minimum variance problem \mathcal{M}' is identical to $\mathcal{M}_{\text{LGM}} = (\mathcal{E}_{\text{LGM}}, \tilde{c}(\cdot), \mathbf{x}_0)$ except for the parameter set. We assume that the prescribed bias function $\tilde{c}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}$ is valid for \mathcal{M}_{LGM} . Then, if the parameter set \mathcal{X} contains an open ball $\mathcal{B}(\mathbf{x}_c, r)$ with some radius $r > 0$ and center*

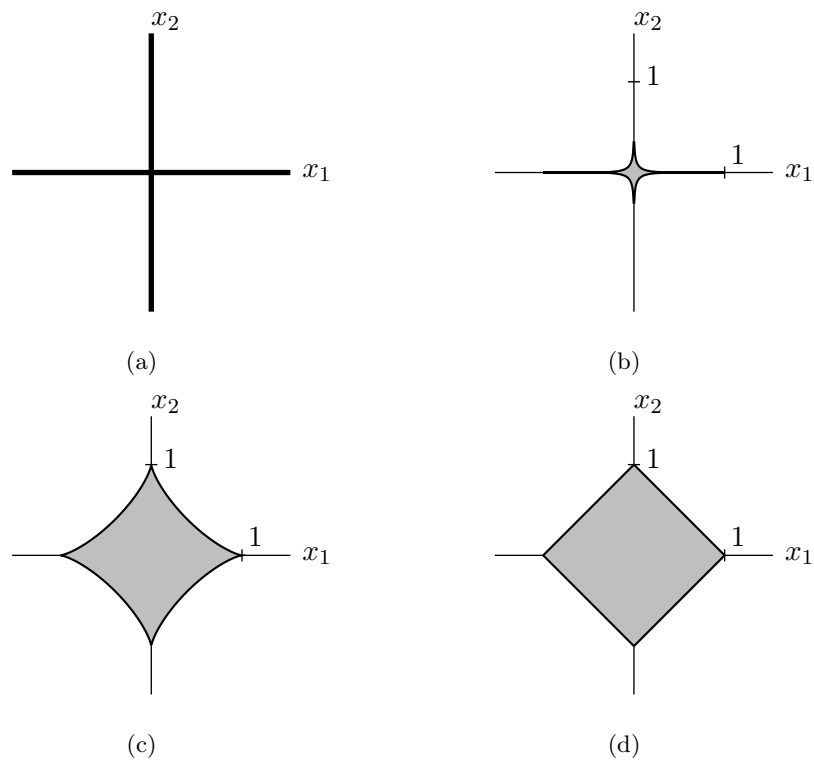


Figure 5.1: Different parameter sets consisting of strictly or approximately 1-sparse vectors $\mathbf{x} = (x_1, x_2)^T \in \mathbb{R}^2$. (a) $\mathcal{X} = \mathcal{B}_0(1) = \mathcal{X}_1$. (b) $\mathcal{X} = \mathcal{B}_{0.25}(1)$. (c) $\mathcal{X} = \mathcal{B}_{0.75}(1)$. (d) $\mathcal{X} = \mathcal{B}_1(1)$.

$\mathbf{x}_c \in \mathcal{X}$, i.e.,

$$\mathcal{B}(\mathbf{x}_c, r) \subseteq \mathcal{X}, \quad (5.179)$$

we have that

$$L_{\mathcal{M}'} = L_{\mathcal{M}_{\text{LGM}}}. \quad (5.180)$$

Proof. We will denote the prescribed mean functions of the minimum variance problems \mathcal{M}' and \mathcal{M}_{LGM} by $\gamma'(\cdot) : \mathcal{X} \rightarrow \mathbb{R} : \gamma'(\mathbf{x}) \triangleq \tilde{c}(\mathbf{x}) + x_k$ and $\gamma(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R} : \gamma(\mathbf{x}) \triangleq \tilde{c}(\mathbf{x}) + x_k$, respectively. These two mean functions are related by

$$\gamma'(\cdot) = \gamma(\cdot)|_{\mathcal{X}}. \quad (5.181)$$

Note that the assumption that $\tilde{c}(\cdot)$ is valid for \mathcal{M}_{LGM} implies that the bias function $\tilde{c}(\cdot)|_{\mathcal{X}}$ is valid for \mathcal{M}' (due to Theorem 3.3.5). We therefore have that $\gamma'(\cdot) \in \mathcal{H}(\mathcal{M}')$ and $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}})$ (due to Theorem 4.3.4).

Since the kernel $R_{\mathcal{M}'}(\cdot, \cdot)$ is the restriction of the kernel $R_{\mathcal{M}_{\text{LGM}}}(\cdot, \cdot)$ to the subdomain $\mathcal{X} \times \mathcal{X}$, i.e., $R_{\mathcal{M}'}(\cdot, \cdot) = R_{\mathcal{M}_{\text{LGM}}}(\cdot, \cdot)|_{\mathcal{X} \times \mathcal{X}}$, we have by Theorem 4.3.4 and Theorem 3.3.5 that

$$\begin{aligned} L_{\mathcal{M}'} &\stackrel{(4.19)}{=} \|\gamma'(\cdot)\|_{\mathcal{H}(\mathcal{M}')}^2 - [\gamma'(\mathbf{x}_0)]^2 \\ &\stackrel{(3.56)}{=} \min_{\substack{\gamma''(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}}) \\ \gamma''(\cdot)|_{\mathcal{X}} = \gamma'(\cdot)}} \|\gamma''(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}})}^2 - [\gamma'(\mathbf{x}_0)]^2 \\ &\stackrel{(5.181)}{=} \min_{\substack{\gamma''(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}}) \\ \gamma''(\cdot)|_{\mathcal{X}} \equiv 0}} \|\gamma''(\cdot) + \gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}})}^2 - [\gamma'(\mathbf{x}_0)]^2. \end{aligned} \quad (5.182)$$

Using Theorem 5.2.2, we can reformulate (5.182) as

$$L_{\mathcal{M}'} = \min_{\substack{\gamma''(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}}) \\ \gamma''(\cdot)|_{\mathcal{X}} \equiv 0}} \|\mathbf{K}_g^{-1}[\gamma''(\cdot) + \gamma(\cdot)]\|_{\mathcal{H}(R_g^{(D)})}^2 - [\gamma'(\mathbf{x}_0)]^2 \quad (5.183)$$

with $D = \text{rank}(\mathbf{H})$.

Consider then an arbitrary function $\gamma''(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}})$ which vanishes on \mathcal{X} , i.e., $\gamma''(\mathbf{x}) = 0$ for every $\mathbf{x} \in \mathcal{X}$. Due to (5.179), the function $\mathbf{K}_g^{-1}[\gamma''(\cdot)](\mathbf{x}')$ vanishes for every $\mathbf{x}' \in \mathcal{B}(\tilde{\mathbf{H}}\mathbf{x}_c, r_1) \subseteq \mathbb{R}^D$ with a suitable radius $r_1 > 0$ and with center $\tilde{\mathbf{H}}\mathbf{x}_c$, where $\tilde{\mathbf{H}}$ is defined as in Theorem 5.2.2. From this, it follows via Theorem 5.2.7 that for every $\gamma''(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}})$ that satisfies $\gamma''(\cdot)|_{\mathcal{X}} \equiv 0$ we have

$$\mathbf{K}_g^{-1}[\gamma''(\cdot) + \gamma(\cdot)] = \mathbf{K}_g^{-1}[\gamma''(\cdot)] + \mathbf{K}_g^{-1}[\gamma(\cdot)] = \mathbf{K}_g^{-1}[\gamma(\cdot)]. \quad (5.184)$$

This, in turn, implies that

$$\begin{aligned}
L_{\mathcal{M}'} &= \min_{\substack{\gamma''(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{LGM}}) \\ \gamma''(\cdot)|_{\mathcal{X}} \equiv 0}} \|\mathbf{K}_g^{-1}[\gamma''(\cdot) + \gamma(\cdot)]\|_{\mathcal{H}(R_g^{(D)})}^2 - [\gamma'(\mathbf{x}_0)]^2 \\
&= \|\mathbf{K}_g^{-1}[\gamma(\cdot)]\|_{\mathcal{H}(R_g^{(D)})}^2 - [\gamma'(\mathbf{x}_0)]^2 \\
&\stackrel{(a)}{=} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{LGM}})}^2 - [\gamma(\mathbf{x}_0)]^2 \stackrel{(4.19)}{=} L_{\mathcal{M}_{\text{LGM}}},
\end{aligned} \tag{5.185}$$

where for the step (a) we used again Theorem 5.2.2. \square

Since we have obviously that the parameter set $\mathcal{X} = \mathcal{B}_q(S)$ satisfies (5.179) for any $q > 0$, Theorem 5.6.1 implies that $L_{\mathcal{M}^{(q)}} = L_{\mathcal{M}_{\text{LGM}}}$ for any $q > 0$. Thus, the minimum achievable variance for $\mathcal{M}^{(q)}$ with $q > 0$ is always equal to that of the minimum variance problem \mathcal{M}_{LGM} , which does not include any sparsity constraints since its parameter set is given by $\mathcal{X} = \mathbb{R}^N$. Furthermore, since in general (see (5.147)) $L_{\mathcal{M}_{\text{LGM}}} > L_{\mathcal{M}_{\text{SLM}}}$, we have also that $L_{\mathcal{M}^{(q)}}$ does not converge to $L_{\mathcal{M}_{\text{SLM}}}$ as q goes to zero in general.

5.7 The SLM Viewpoint on Compressed Sensing

The compressive measurement process of a compressed sensing (CS) application is often modeled by a linear measurement equation of the form [8, 16, 17, 37, 53]

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{5.186}$$

where $\mathbf{y} \in \mathbb{R}^M$ denotes the compressive measurements; $\mathbf{H} \in \mathbb{R}^{M \times N}$ (where typically $M \ll N$) denotes the CS measurement matrix; the signal or parameter vector $\mathbf{x} \in \mathbb{R}^N$ is assumed to be sparse, i.e., $\mathbf{x} \in \mathcal{X}_S$ with known sparsity degree S (where typically $S \ll N$); and \mathbf{n} represents some additive measurement noise. If the vector $\mathbf{n} \in \mathbb{R}^M$ in (5.186) is AWGN, i.e., $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$, then the CS measurement equation (5.186) is identical to the observation model of the SLM (see (5.2)).

Any CS recovery method, e.g., the ‘‘Basis Pursuit’’ (BP) [16, 90] or the ‘‘Orthogonal Matching Pursuit’’ (OMP) [17, 91] to name two well-known instances,⁶ can be interpreted as an estimator $\hat{\mathbf{x}}(\mathbf{y})$ that aims to estimate the sparse vector \mathbf{x} using the observation \mathbf{y} given by (5.186).

In general, it is infeasible to characterize the analytical properties of the measurement matrix \mathbf{H} in an exact manner (e.g., to compute its thin SVD), due to the large dimension (typically $M \geq 10$ and $N \geq 100$). Instead, one can use an incomplete characterization via the concept of (mutual) coherence and the restricted isometry property [8, 16, 17, 53], which we will now define.

⁶A comprehensive overview is provided at <http://dsp.rice.edu/cs>.

Definition 5.7.1. The coherence of a matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ is defined as

$$\mu(\mathbf{H}) \triangleq \max_{i \neq j} |\mathbf{h}_j^T \mathbf{h}_i|, \quad (5.187)$$

where $\mathbf{h}_i \in \mathbb{R}^M$ denotes the i th column of \mathbf{H} .

Definition 5.7.2. A matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ is said to satisfy the restricted isometry property (RIP) of order K with RIP constant $\delta_K \in \mathbb{R}_+$ if for every index set $\mathcal{I} \subseteq [N]$ of size K , i.e., $|\mathcal{I}| = K$, we have

$$(1 - \delta_K) \|\mathbf{z}\|_2^2 \leq \|\mathbf{H}_{\mathcal{I}} \mathbf{z}\|_2^2 \leq (1 + \delta_K) \|\mathbf{z}\|_2^2 \quad (5.188)$$

for every $\mathbf{z} \in \mathbb{R}^K$.

Note that it can be straightforwardly shown that $\delta_K \leq \delta_{K'}$ if $K' \geq K$.

While the coherence of a matrix can be calculated more efficiently than the RIP constants δ_K [53], it is a coarser description of the matrix \mathbf{H} than the RIP constants. However, as stated in [53], we have the following relation between these two concepts:

Theorem 5.7.1. For a matrix \mathbf{H} with coherence $\mu(\mathbf{H})$, the RIP constant δ_K of order K satisfies

$$\delta_K \leq (K - 1)\mu(\mathbf{H}). \quad (5.189)$$

Proof. [53] □

We now specialize Theorem 5.4.4 to the CS scenario of the SLM, i.e., where the system matrix \mathbf{H} of the SLM is a CS measurement matrix with known RIP constants.

Theorem 5.7.2. Consider the minimum variance problem $\mathcal{M}_{\text{SLM}} = (\mathcal{E}_{\text{SLM}}, c(\cdot), \mathbf{x}_0)$ with some choice of σ , S , M , N , $\mathbf{x}_0 \in \mathcal{X}_S$ and with system matrix \mathbf{H} , where $\mathbf{H} \in \mathbb{R}^{M \times N}$ is a CS measurement matrix satisfying the RIP of order S with RIP constant $\delta_S < 1$. We denote the prescribed mean by $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = c(\mathbf{x}) + x_k$. Then, for an arbitrary set $\mathcal{K} = \{i_1, \dots, i_{|\mathcal{K}|}\} \subseteq [N]$ consisting of no more than S different indices, i.e., $|\mathcal{K}| \leq S$, for which the partial derivatives $\frac{\partial^{e_{i_l}} \gamma(\mathbf{x})}{\partial \mathbf{x}^{e_{i_l}}}$ exist, we have that

$$L_{\mathcal{M}_{\text{SLM}}} \geq \exp \left(- \frac{1 + \delta_S}{\sigma^2} \left\| \mathbf{x}_0^{\text{supp}(\mathbf{x}_0) \setminus \mathcal{K}} \right\|_2^2 \right) \sigma^2 \mathbf{r}_{\mathbf{x}_0}^T (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{r}_{\mathbf{x}_0}, \quad (5.190)$$

where the vector $\mathbf{r}_{\mathbf{x}_0} \in \mathbb{R}^{|\mathcal{K}|}$ is defined elementwise as $r_{\mathbf{x}_0, l} \triangleq \frac{\partial^{e_{i_l}} \gamma(\mathbf{x})}{\partial \mathbf{x}^{e_{i_l}}} \Big|_{\mathbf{x}=\tilde{\mathbf{x}}_0}$ with $\tilde{\mathbf{x}}_0$ as defined by (5.100) in Theorem 5.4.3.

Proof. Let $\mathbf{P}_{\mathcal{K}} \in \mathbb{R}^{M \times |\mathcal{K}|}$ denote the orthogonal projection matrix on the subspace $\text{span}(\mathbf{H}_{\mathcal{K}}) \subseteq \mathbb{R}^M$. We have that $\mathbf{P}_{\mathcal{K}} = \mathbf{H}_{\mathcal{K}} (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{H}_{\mathcal{K}}^T$ since \mathbf{H} satisfies the RIP of order S , which implies that $\mathbf{H}_{\mathcal{K}}$ has full column rank [38]. We also have

$$(\mathbf{I} - \mathbf{P}_{\mathcal{K}})^T \mathbf{H}_{\mathcal{K}} = (\mathbf{I} - \mathbf{P}_{\mathcal{K}}) \mathbf{H}_{\mathcal{K}} = \mathbf{0},$$

i.e., for every vector $\mathbf{x}' \in \text{span}(\mathbf{H}_{\mathcal{K}})$ we have that $(\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{x}' = \mathbf{0}$. This implies that

$$(\mathbf{I} - \mathbf{P}_{\mathcal{K}})^T \mathbf{H}\mathbf{x}_0 = (\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{H}(\mathbf{x}_0^{\text{supp}(\mathbf{x}_0) \setminus \mathcal{K}} + \mathbf{x}_0^{\mathcal{K}}) = (\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{H}\mathbf{x}_0^{\text{supp}(\mathbf{x}_0) \setminus \mathcal{K}}, \quad (5.191)$$

since $\mathbf{H}\mathbf{x}_0^{\mathcal{K}} \in \text{span}(\mathbf{H}_{\mathcal{K}})$. Based on (5.191) and using the shorthand $\mathbf{x}'_0 \triangleq \mathbf{x}_0^{\text{supp}(\mathbf{x}_0) \setminus \mathcal{K}}$, we have

$$\begin{aligned} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{H}\mathbf{x}_0\|_2^2 &= \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{H}\mathbf{x}'_0\|_2^2 \\ &= \|\mathbf{H}\mathbf{x}'_0\|_2^2 - 2(\mathbf{x}'_0)^T \mathbf{H}^T \mathbf{P}_{\mathcal{K}} \mathbf{H}\mathbf{x}'_0 + (\mathbf{x}'_0)^T \mathbf{H}^T \mathbf{P}_{\mathcal{K}}^2 \mathbf{H}\mathbf{x}'_0 \\ &\stackrel{(a)}{=} \|\mathbf{H}\mathbf{x}'_0\|_2^2 - (\mathbf{x}'_0)^T \mathbf{H}^T \mathbf{P}_{\mathcal{K}} \mathbf{H}\mathbf{x}'_0 \\ &\stackrel{(b)}{\leq} \|\mathbf{H}\mathbf{x}'_0\|_2^2 \\ &\stackrel{(c)}{\leq} (1 + \delta_S) \|\mathbf{x}'_0\|_2^2, \end{aligned} \quad (5.192)$$

where the step (a) follows from $(\mathbf{P}_{\mathcal{K}})^2 = \mathbf{P}_{\mathcal{K}}$ since $\mathbf{P}_{\mathcal{K}}$ is an orthogonal projection matrix [38], the step (b) follows from the fact that $\mathbf{P}_{\mathcal{K}}$ is psd, and (c) is due to $\|\mathbf{x}'_0\|_0 \leq S$ and the assumption that \mathbf{H} satisfies the RIP of order S with RIP constant $\delta_S < 1$. The statement follows then as a reformulation of (5.110) from Theorem 5.4.4 using (5.192), i.e.,

$$\begin{aligned} L_{\mathcal{M}_{\text{SLM}}} &\stackrel{(5.110)}{\geq} \exp\left(-\frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{H}\mathbf{x}_0\|_2^2\right) \sigma^2 \mathbf{r}_{\mathbf{x}_0}^T (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{r}_{\mathbf{x}_0} \\ &\stackrel{(5.192)}{\geq} \exp\left(-\frac{1 + \delta_S}{\sigma^2} \|\mathbf{x}_0^{\text{supp}(\mathbf{x}_0) \setminus \mathcal{K}}\|_2^2\right) \sigma^2 \mathbf{r}_{\mathbf{x}_0}^T (\mathbf{H}_{\mathcal{K}}^T \mathbf{H}_{\mathcal{K}})^{-1} \mathbf{r}_{\mathbf{x}_0}. \end{aligned} \quad (5.193)$$

□

If we want to use Theorem 5.7.2 for comparing the actual variance behavior of a given CS recovery scheme (using the CS measurement matrix \mathbf{H}), which is an estimator $\hat{x}_k(\cdot)$ for the SLM with system matrix \mathbf{H} , with the theoretically minimum variance achievable for the bias of the estimator $\hat{x}_k(\cdot)$, we have to ensure that the first-order partial derivatives of the mean function $\mathbb{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\}$ of the given estimator $\hat{x}_k(\cdot)$ exist. The next lemma states that this is indeed the case for a very broad class of estimators $\hat{x}_k(\cdot)$ resulting from a CS recovery scheme. It is essentially a more rigorous formulation of the relation (19) in [56].

Lemma 5.7.3. *Consider the SLM \mathcal{E}_{SLM} with system matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ and an estimator $\hat{x}_k(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}$ which may arise from a CS recovery scheme like BP or OMP. We denote the mean*

function of $\hat{x}_k(\cdot)$ by $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\}$. If the estimator $\hat{x}_k(\mathbf{y})$ is a continuous function of \mathbf{y} , and moreover for every $\mathbf{y} \in \mathbb{R}^M$ we have

$$|\hat{x}_k(\mathbf{y})| \leq C\|\mathbf{y}\|_2^L \quad (5.194)$$

with some constants $C, L \in \mathbb{R}_+$, then the partial derivatives $\frac{\partial^{e_l} \gamma(\mathbf{x})}{\partial \mathbf{x}^{e_l}}$ exist for every $l \in [N]$ and are given by

$$\frac{\partial^{e_l} \gamma(\mathbf{x})}{\partial \mathbf{x}^{e_l}} = \frac{1}{\sigma^2} \mathbb{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})(\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{H}\mathbf{e}_l\}. \quad (5.195)$$

Proof. We have

$$\gamma(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\} = \frac{1}{(2\pi\sigma^2)^{M/2}} \int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right) d\mathbf{y}, \quad (5.196)$$

where the existence of the integral follows from the dominated convergence theorem [5, 6, 73] since (i) the function

$$\hat{x}_k(\mathbf{y}) \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right)$$

is continuous and therefore measurable and (ii) it is upper bounded in magnitude (dominated) by the function $C\|\mathbf{y}\|_2^L \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right)$, which is obviously integrable. We further have

$$\begin{aligned} \frac{\partial^{e_l} \gamma(\mathbf{x})}{\partial \mathbf{x}^{e_l}} &\stackrel{(5.196)}{=} \frac{\partial^{e_l}}{\partial \mathbf{x}^{e_l}} \frac{1}{(2\pi\sigma^2)^{M/2}} \int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right) d\mathbf{y} \\ &\stackrel{(a)}{=} \frac{1}{(2\pi\sigma^2)^{M/2}} \int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) \frac{\partial^{e_l} \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right)}{\partial \mathbf{x}^{e_l}} d\mathbf{y} \\ &\stackrel{(b)}{=} \frac{1}{(2\pi\sigma^2)^{M/2}} \int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) \frac{1}{\sigma^2} (\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{H}\mathbf{e}_l \exp\left(-\frac{1}{2\sigma^2}\|\mathbf{y} - \mathbf{H}\mathbf{x}\|_2^2\right) d\mathbf{y} \\ &= \frac{1}{\sigma^2} \mathbb{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})(\mathbf{y} - \mathbf{H}\mathbf{x})^T \mathbf{H}\mathbf{e}_l\}, \end{aligned} \quad (5.197)$$

where step (a) follows from a change of the order of differentiation and integration (see [41, Theorem 1.5.8]), and step (b) is due to the chain rule for differentiation [5]. The existence of the last integral in (5.197) can again be verified by the continuity of the integrand and the validity of the bound (5.194) on the estimator $\hat{x}_k(\mathbf{y})$. \square

For the application of Lemma 5.7.3, we have to verify for a given CS recovery scheme, i.e., for a given estimator $\hat{\mathbf{x}}(\mathbf{y})$, that its entries $\hat{x}_k(\mathbf{y})$ satisfy the bound (5.194) and are continuous functions of \mathbf{y} . While showing (5.194) is rather straightforward for the BP and OMP, the rigorous proof of

continuity is subtle. Consider, e.g., the BP recovery method [90], which is defined as the solution to the problem

$$\min_{\mathbf{x} \in \mathbb{R}^N} \|\mathbf{x}\|_1 \quad \text{s.t.} \quad \mathbf{H}\mathbf{x} = \mathbf{y}. \quad (5.198)$$

Since there may be multiple solutions to this problem, there are many possible estimator functions $\hat{\mathbf{x}}_{\text{BP}}(\cdot)$ corresponding to the BP. However, some popular implementations (e.g., the primal dual algorithm presented in [30, 92]) of the estimator $\hat{\mathbf{x}}_{\text{BP}}(\mathbf{y}) : \mathbb{R}^M \rightarrow \mathbb{R}^N$ yield a continuous function of the observation \mathbf{y} . The resulting approximate BP estimator is then continuous “by design.” As can be verified easily, the statement and proof of Lemma 5.7.3 remain unchanged if one replaces the requirement of a continuous estimator function with that of a Lebesgue measurable estimator function. It seems then fairly unlikely that any reasonable and practically implemented CS recovery scheme does not yield an estimator that is both measurable and bounded as in (5.194).

We note that condition (5.3) is necessary for a matrix \mathbf{H} to have the RIP of order S with a constant $\delta_S < 1$.⁷ Moreover, for CS applications, one favors in general measurement matrices \mathbf{H} that have RIP constants close to zero, i.e., $\delta_S \approx 0$ [8, 18, 53, 93, 94]. However, in this case the bound in (5.190) is nearly identical to the bound (5.110) for the SLM with $\mathbf{H} = \mathbf{I}$, given explicitly as

$$L_{\mathcal{M}_{\text{SLM}}} \geq \exp\left(-\frac{1}{\sigma^2} \|(\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{x}_0\|_2^2\right) \sigma^2 \mathbf{r}_{\mathbf{x}_0}^T \mathbf{r}_{\mathbf{x}_0}. \quad (5.199)$$

Note that for $\mathbf{H} = \mathbf{I}$, implying that

$$\mathbf{P}_{\mathcal{K}} = \mathbf{H}_{\mathcal{K}}\mathbf{H}_{\mathcal{K}}^\dagger = \sum_{l \in \mathcal{K}} \mathbf{e}_l \mathbf{e}_l^T, \quad (5.200)$$

the effect of the multiplication $(\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{x}_0$ is a zeroing of all entries of the vector \mathbf{x}_0 whose indices belong to \mathcal{K} , i.e.,

$$(\mathbf{I} - \mathbf{P}_{\mathcal{K}})\mathbf{x}_0 = \mathbf{x}_0^{\text{supp}(\mathbf{x}_0) \setminus \mathcal{K}}. \quad (5.201)$$

Thus, if one uses a “good” CS measurement matrix, i.e., with RIP constant δ_S close to zero, then Theorem 5.7.2 suggests that the minimum achievable variance is close to the minimum achievable variance of another instance of the SLM with the same S , and N but with system the matrix being the identity matrix \mathbf{I} instead of the measurement matrix \mathbf{H} . This means that in terms of achievable estimation accuracy, there is no loss of information incurred by applying the CS measurement matrix $\mathbf{H} \in \mathbb{R}^{M \times N}$ and thereby reducing the signal dimension from N to M , where $M \ll N$ in general. This agrees with the fact that if a CS measurement matrix satisfies the RIP with RIP constants that are sufficiently small, then one can recover - e.g., by using the BP -

⁷Indeed, assume that $\text{spark}(\mathbf{H}) \leq S$ which means that there exists an index set $\mathcal{I} \subseteq [N]$ consisting of S indices, i.e., $|\mathcal{I}| = S$ such that the columns of $\mathbf{H}_{\mathcal{I}}$ are linearly dependent. This implies that there is a coefficient vector $\mathbf{z}_0 \in \mathbb{R}^S$ with $\mathbf{z}_0 \neq \mathbf{0}$ (i.e., $\|\mathbf{z}_0\|_2^2 > 0$) such that $\mathbf{H}_{\mathcal{I}}\mathbf{z}_0 = \mathbf{0}$ and in turn $\|\mathbf{H}_{\mathcal{I}}\mathbf{z}_0\|_2^2 = 0$. Therefore, there cannot exist a constant $\delta_{\mathcal{K}} < 1$ that satisfies the RIP condition (5.188) for every $\mathbf{z} \in \mathbb{R}^S$.

the sparse parameter vector $\mathbf{x} \in \mathcal{X}_S$ from the compressed observation \mathbf{y} given by (5.186) up to a reconstruction error that is solely determined by the noise in (5.186) [8, 50].

The CS measurement model in (5.186) assumes that the noise component is added after the measurement process. However, in some applications it may be more accurate to model the effect of the noise before the measurement process takes place, i.e., to use the following measurement model:

$$\mathbf{y}' = \mathbf{H}(\mathbf{x} + \mathbf{n}), \quad (5.202)$$

where the meanings of \mathbf{H} , \mathbf{x} , and \mathbf{n} are the same as for (5.186). Note that the observation \mathbf{y}' is obtained by applying a mapping, i.e., the matrix multiplication by \mathbf{H} , to the observation $\mathbf{y} = \mathbf{x} + \mathbf{n}$ of the SSNM. Therefore, if the CS measurement process is modeled by (5.202), then we can analyze the resulting minimum variance problem of estimating the sparse vector $\mathbf{x} \in \mathcal{X}_S$ from the observation \mathbf{y}' by using the results on the SSNM in Section 5.5, in particular Theorem 5.5.2, together with Theorem 2.4.2. We thus obtain the following result:

Theorem 5.7.4. *Consider the observation model (5.202) and an estimator $\hat{x}_k(\mathbf{y}') : \mathbb{R}^M \rightarrow \mathbb{R}$, which may arise from a CS recovery scheme, of x_k . The estimator $\hat{x}_k(\mathbf{y}')$ can use only the observation \mathbf{y}' and is assumed to have finite variance at $\mathbf{x}_0 \in \mathcal{X}_S$, i.e., $v(\hat{x}_k(\cdot); \mathbf{x}_0) < \infty$. We assume that the mean of this estimator exists for all parameter vectors $\mathbf{x} \in \mathcal{X}_S$ and denote its mean function by $\gamma(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = \mathbf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y}')\}$. Then, we have that*

$$v(\hat{x}_k(\cdot); \mathbf{x}_0) \geq \|a_{\mathbf{x}_0}[\mathbf{p}]\|_{\ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)}^2 - [\gamma(\mathbf{x}_0)]^2, \quad (5.203)$$

where the coefficient sequence $a_{\mathbf{x}_0}[\mathbf{p}] \in \ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)$ are given by (5.135).

Proof. Let us denote by \mathcal{M}' the minimum variance problem obtained for the observation model (5.202), the parameter set \mathcal{X}_S , the fixed parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$, the parameter function $g(\mathbf{x}) = x_k$, and prescribed bias function $c(\cdot) \triangleq \gamma(\cdot) - x_k = \mathbf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y}')\} - x_k$. Since $c(\cdot)$ is identical with the bias function $b(\hat{x}_k(\cdot), \mathbf{x})$ of the estimator $\hat{x}_k(\cdot)$ which is assumed to have finite variance at \mathbf{x}_0 , we have that $c(\cdot)$ is valid for \mathcal{M}' and $v(\hat{x}_k(\cdot); \mathbf{x}_0) \geq L_{\mathcal{M}'}$.

The minimum variance problem \mathcal{M}' is obtained from the minimum variance problem $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}_0)$ (which uses the same index $k \in [N]$ in its parameter function (see (5.10)) as used within \mathcal{M}') by a matrix multiplication of the observation of the SSNM with the matrix \mathbf{H} . Therefore, by Theorem 2.4.2, we have that $c(\cdot)$ is valid for $\mathcal{M}_{\text{SSNM}}$. Indeed, since $c(\cdot)$ is valid for \mathcal{M}' we have that $L_{\mathcal{M}'}$ is finite and this implies via (2.31) that $L_{\mathcal{M}_{\text{SSNM}}}$ is finite, i.e., $c(\cdot)$ must be valid for $\mathcal{M}_{\text{SSNM}}$. Since $c(\cdot)$ is valid for $\mathcal{M}_{\text{SSNM}}$ we have that $L_{\mathcal{M}_{\text{SSNM}}}$ is given by (5.134) in Theorem 5.5.2. The bound (5.203) follows then as

$$v(\hat{x}_k(\cdot); \mathbf{x}_0) \geq L_{\mathcal{M}'} \stackrel{(2.31)}{\geq} L_{\mathcal{M}_{\text{SSNM}}} \stackrel{(5.134)}{=} \|a_{\mathbf{x}_0}[\mathbf{p}]\|_{\ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)}^2 - [\gamma(\mathbf{x}_0)]^2. \quad (5.204)$$

□

It is important to note that the statement and proof of Theorem 5.7.4 are also valid for the case where \mathbf{H} in (5.202) is a random matrix, which is independent of the noise \mathbf{n} . This is practically relevant since it is often the case that CS measurement matrices are modeled as random matrices.

5.8 Comparison of the Bounds with Existing Estimators

Let us now compare the lower bounds derived so far with the actual variance behavior of some well-known estimators. In what follows, we will denote by $L_{\mathcal{K},a}^{(c(\cdot),k)}(\mathbf{x}_0)$, $L_{\mathcal{K},b}^{(c(\cdot),k)}(\mathbf{x}_0)$ the bounds (5.99) and (5.110) obtained by Theorem 5.4.3 and Theorem 5.4.4, respectively, using the index set \mathcal{K} and prescribed bias function $c(\cdot)$.

Given an estimator $\hat{\mathbf{x}}(\cdot)$ whose bias is equal to $\mathbf{c}(\mathbf{x})$ and mean equal to $\boldsymbol{\gamma}(\mathbf{x}) \triangleq \mathbf{c}(\mathbf{x}) + \mathbf{x}$, we obtain due to (2.21), a lower bound on the estimator variance $v(\hat{\mathbf{x}}(\cdot); \mathbf{x}_0)$ by summing the quantities $L_{\mathcal{K},a/b}^{(c_k(\cdot),k)}(\mathbf{x}_0)$, i.e.,

$$v(\hat{\mathbf{x}}(\cdot); \mathbf{x}_0) \geq L_{\boldsymbol{\gamma}(\cdot), \mathbf{x}_0}^{a/b} \triangleq \sum_{k \in [N]} L_{\mathcal{K}_k, a/b}^{(c_k(\cdot), k)}(\mathbf{x}_0). \quad (5.205)$$

Note that in (5.205), we allow the index $\mathcal{K} = \mathcal{K}_k$ in (5.99), (5.110) to vary with the index k .

5.8.1 SLM Viewpoint on Fourier Analysis

The first experiment is inspired by [20, Example 4.2 – ‘Fourier Analysis’] and considers the general SLM with $\sigma^2 = 1$ and system matrix

$$\mathbf{H} = \begin{pmatrix} \cos(\theta_1 0) & \dots & \cos(\theta_L 0) & \sin(\theta_1 0) & \dots & \sin(\theta_L 0) \\ \cos(\theta_1 1) & \dots & \cos(\theta_L 1) & \sin(\theta_1 1) & \dots & \sin(\theta_L 1) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \cos(\theta_1 (M-1)) & \dots & \cos(\theta_L (M-1)) & \sin(\theta_1 (M-1)) & \dots & \sin(\theta_L (M-1)) \end{pmatrix}, \quad (5.206)$$

i.e., $\mathbf{H} \in \mathbb{R}^{M \times N}$ with $N = 2L$. Note that this system matrix corresponds to a finite-length discrete Fourier transform evaluated at the frequencies⁸ $\{\theta_l\}_{l \in [L]}$. In our simulation we choose $M = 128$, $L = 8$, and $\theta_l = 0.2 + 3.9 \cdot 10^{-3}(l-1)$. Our choice of θ_l corresponds to a frequency resolution of $\Delta\theta = 3.9 \times 10^{-3}$; this is about half of the DFT frequency resolution, which is given by $\Delta\theta = \frac{1}{128} \approx 7.8 \times 10^{-3}$. We assume that the parameter vector can have at maximum 4 nonzero entries, i.e., we consider the SLM with sparsity degree $S = 4$.

⁸Here, we consider the normalized frequency $\theta \in [0, 1]$ of a discrete-time harmonic signal $x[n] = \sin(2\pi\theta n)$.

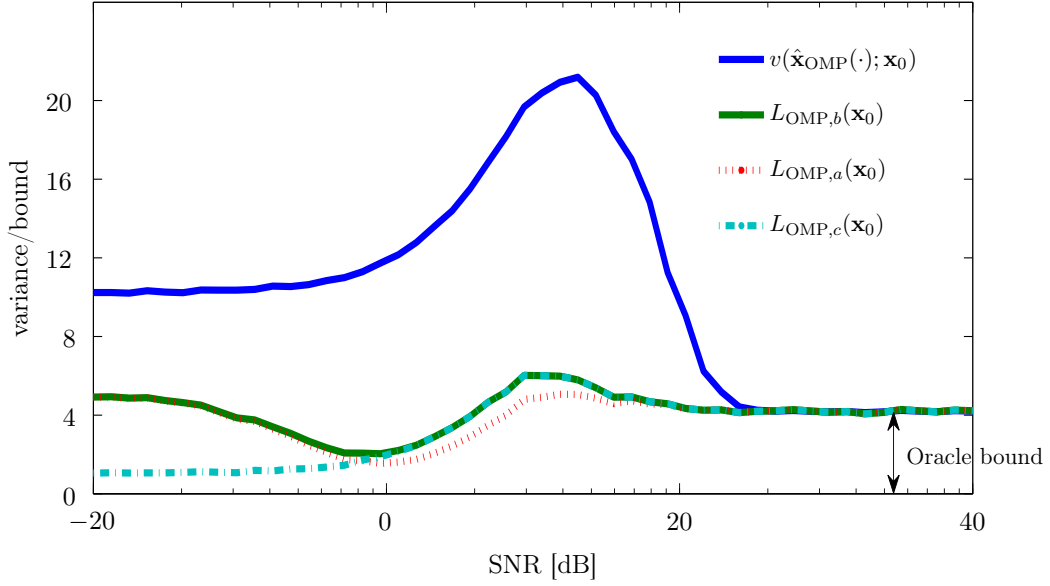


Figure 5.2: Variance of the OMP estimator and corresponding lower bounds versus the SNR, for the SLM with $N=16$, $M=128$, $\sigma^2=1$ and $S=4$.

In Fig. 5.2, we compare the variance of the OMP estimator $\hat{\mathbf{x}}_{\text{OMP}}(\mathbf{y})$, which is obtained by applying OMP (see [17,91]) with S iterations to the observation, with the lower bounds $L_{\text{OMP},a/b}(\mathbf{x}_0)$ defined as (cf. (5.205))

$$L_{\text{OMP},a/b}(\mathbf{x}_0) = \sum_{k \in [N]} L_{\mathcal{K}_k, a/b}^{(c_k(\cdot), k)}(\mathbf{x}_0), \quad (5.207)$$

where $c_k(\cdot)$ is chosen as $c_k(\cdot) = \gamma_{k, \text{OMP}}(\mathbf{x}) - x_k$, with $\gamma_{k, \text{OMP}}(\mathbf{x}) \triangleq \mathbf{E}_{\mathbf{x}} \{ \hat{x}_{k, \text{OMP}}(\mathbf{y}) \}$ being the actual mean function of the OMP estimator. We also plotted the lower bound $L_{\text{OMP},c}(\mathbf{x}_0)$ obtained by summing the lower bound (5.87), (5.88) (using $\gamma_{k, \text{OMP}} - x_k$ as prescribed bias function) in Theorem 5.4.1 over all indices $k \in [N]$. By (2.21), we have that $v(\hat{\mathbf{x}}_{\text{OMP}}(\cdot); \mathbf{x}_0) \geq L_{\text{OMP},c}(\mathbf{x}_0)$.

The variance $v(\hat{\mathbf{x}}_{\text{OMP}}(\cdot); \mathbf{x}_0)$ and the lower bounds $L_{\text{OMP},a/b/c}(\mathbf{x}_0)$ are evaluated for parameter vectors $\mathbf{x}_0 = \sqrt{\text{SNR}} \sigma \mathbf{x}_1$, where $\mathbf{x}_1 \in \{0, 1\}^{16}$, $\text{supp}(\mathbf{x}_1) = \{3, 6, 67, 70\}$, and $\sqrt{\text{SNR}}$ varies between 0.1 and 100. The variance $v(\hat{\mathbf{x}}_{\text{OMP}}(\cdot); \mathbf{x}_0)$ of the OMP estimator is estimated by means of Monte Carlo simulation. For the evaluation of the bound $L_{\text{OMP},a/b/c}(\mathbf{x}_0)$ (cf. (5.99), (5.110), (5.87), (5.88)), we have to compute the first-order partial derivatives of the mean function $\gamma_{k, \text{OMP}}(\mathbf{x})$. We estimated these partial derivatives by means of Lemma 5.7.3 and a Monte Carlo simulation. This method of estimating the derivative of an estimator bias or mean function is also used and explained in detail in [56]. For simplicity, the index sets \mathcal{K}_k in (5.207) are chosen as $\mathcal{K}_k = \text{supp}(\mathbf{x}_0)$ for $k \in \text{supp}(\mathbf{x}_0)$ and $\mathcal{K}_k = \{k\}$ for $k \notin \text{supp}(\mathbf{x}_0)$. This is the simplest non-trivial choice that yields

a bound $L_{\text{OMP},b}(\mathbf{x}_0)$ that is tighter than the bound $L_{\text{OMP},c}(\mathbf{x}_0)$, which is based on the sparse CRB in Theorem 5.4.1 (cf. [25]). In Fig. 5.2, we also indicated the ‘‘oracle CRB,’’ which is defined as the CRB under the assumption that one knows the support of \mathbf{x}_0 , i.e., the oracle bound is the CRB of a linear Gaussian model with system matrix $\mathbf{H}_{\text{supp}(\mathbf{x}_0)}$ and is thus given by [20]

$$\sigma^2 \text{Tr} \{ [\mathbf{H}_{\text{supp}(\mathbf{x}_0)}^T \mathbf{H}_{\text{supp}(\mathbf{x}_0)}]^{-1} \} \approx 4.19 \times \sigma^2. \quad (5.208)$$

As can be seen from Fig. 5.2, there are two SNR regimes regarding the variance of the OMP estimator: Below 20 dB, i.e., in the low-SNR regime the variance of $\hat{\mathbf{x}}_{\text{OMP}}(\cdot)$ is significantly higher than the oracle CRB and also significantly higher than the lower bounds $L_{\text{OMP},a/b/c}(\mathbf{x}_0)$. This suggests that there might exist estimators with a reduced variance but the same bias and mean as the OMP estimator. Above 20 dB, i.e., in the high-SNR regime the variance $\hat{\mathbf{x}}_{\text{OMP}}(\cdot)$ shows a fast convergence towards the oracle CRB as well as to the bounds $L_{\text{OMP},a/b/c}(\mathbf{x}_0)$, which is due to the fact that for SNR values above 20 dB the OMP estimator is able to detect the support of \mathbf{x}_0 with very high probability. Note also that the curves of $L_{\text{OMP},a/b}(\mathbf{x}_0)$ agrees with the discussion around (5.117), i.e., that the bound $L_{\text{OMP},b}(\mathbf{x}_0)$ tends to be higher than $L_{\text{OMP},a}(\mathbf{x}_0)$ in general.

5.8.2 Minimum Variance Analysis of Estimators for the SSNM

Let us now consider the SSNM in (5.9) for $N = 50$, $S = 5$, and $\sigma^2 = 1$. We will compute the lower bound on the estimator variance given in (5.205) and compare it with the variance of two established estimators, namely, the maximum likelihood (ML) estimator and the hard-thresholding (HT) estimator. The ML estimator is given by

$$\hat{\mathbf{x}}_{\text{ML}}(\mathbf{y}) \triangleq \underset{\mathbf{x}' \in \mathcal{X}_S}{\text{argmax}} f(\mathbf{y}; \mathbf{x}') = \mathbf{P}_S(\mathbf{y}),$$

where the operator \mathbf{P}_S retains the S largest (in magnitude) entries and zeros out all others. The HT estimator $\hat{\mathbf{x}}_{\text{HT}}(\mathbf{y})$ is given by

$$\hat{x}_{\text{HT},k}(\mathbf{y}) = \hat{x}_{\text{HT},k}(y_k) = \begin{cases} y_k, & |y_k| \geq T \\ 0, & \text{else,} \end{cases} \quad (5.209)$$

where T is a fixed threshold. Note that in the limiting case where $T = 0$, the HT estimator coincides with the least square (LS) estimator [20,21,41] for the SSNM $\hat{\mathbf{x}}_{\text{LS}}(\mathbf{y}) = \mathbf{y}$. The prescribed bias $c_k(\cdot)$ in (5.205) is chosen as k th entry of the bias function of the ML and HT estimator, respectively.

The mean and variance of the HT estimator are given by

$$\mathbf{E}_{\mathbf{x}} \{ \hat{x}_{\text{HT},k}(\mathbf{y}) \} = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{y_k \in \mathbb{R}^N \setminus [-T, T]} y_k \exp \left(-\frac{1}{2\sigma^2} (y_k - x_k)^2 \right) dy_k \quad (5.210)$$

and

$$\begin{aligned} v(\hat{x}_{\text{HT}(\cdot),k}; \mathbf{x}) &\stackrel{(2.11)}{=} P(\hat{x}_{\text{HT},k}(\cdot); \mathbf{x}) - [\mathbf{E}_{\mathbf{x}}\{\hat{x}_{\text{HT},k}(\mathbf{y})\}]^2 \\ &= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{y_k \in \mathbb{R} \setminus [-T, T]} y_k^2 \exp\left(-\frac{1}{2\sigma^2}(y_k - x_k)^2\right) dy_k - [\mathbf{E}_{\mathbf{x}}\{\hat{x}_{\text{HT},k}(\mathbf{y})\}]^2, \end{aligned} \quad (5.211)$$

respectively. The integrals are calculated using numerical integration. For the evaluation of the mean and variance of the ML estimator, we used the (complicated) closed-form expressions derived in [82].

Let us denote by $c_{\text{ML},k}(\mathbf{x}) \triangleq b(\hat{x}_{\text{ML},k}(\cdot), \mathbf{x})$ and $c_{\text{HT},k}(\mathbf{x}) \triangleq b(\hat{x}_{\text{HT},k}(\cdot), \mathbf{x})$ the bias function of the k th entry of the ML and HT estimator, respectively. Using these bias functions, we compare the variance of the HT and ML estimators with the lower bounds

$$L_{\text{ML},a/b}(\mathbf{x}_0) \triangleq \sum_{k \in [N]} L_{\mathcal{K}_k, a/b}^{(c_{\text{ML},k}(\cdot), k)}(\mathbf{x}_0) \quad (5.212)$$

and

$$L_{\text{HT},a/b}(\mathbf{x}_0) \triangleq \sum_{k \in [N]} L_{\mathcal{K}_k, a/b}^{(c_{\text{HT},k}(\cdot), k)}(\mathbf{x}_0), \quad (5.213)$$

respectively. As before, the index set $\mathcal{K}_k \subseteq [N]$ is allowed to vary with the index k . In particular, we choose $\mathcal{K}_k = \text{supp}(\mathbf{x})$ for $k \in \text{supp}(\mathbf{x})$ and $\mathcal{K}_k = \{k\} \cup \{\text{supp}(\mathbf{x}_0) \setminus \{j_0\}\}$ for $k \notin \text{supp}(\mathbf{x}_0)$, where j_0 denotes the index of the S -largest (in magnitude) entry of \mathbf{x}_0 . It can be verified easily that for these choices of the index sets \mathcal{K}_k , the bounds for the HT estimator, $L_{\mathcal{K}_k, a}^{(c_{\text{HT},k}(\cdot), k)}(\mathbf{x}_0)$ and $L_{\mathcal{K}_k, b}^{(c_{\text{HT},k}(\cdot), k)}(\mathbf{x}_0)$, as well as the bounds for the ML estimator, $L_{\mathcal{K}_k, a}^{(c_{\text{ML},k}(\cdot), k)}(\mathbf{x}_0)$ and $L_{\mathcal{K}_k, b}^{(c_{\text{ML},k}(\cdot), k)}(\mathbf{x}_0)$ coincide for the SSNM (i.e., when $\mathbf{H} = \mathbf{I}$). Thus, we have in this case that $L_{\text{HT},a}(\mathbf{x}_0) = L_{\text{HT},b}(\mathbf{x}_0)$ and $L_{\text{ML},a}(\mathbf{x}_0) = L_{\text{ML},b}(\mathbf{x}_0)$.

In order to evaluate the bounds $L_{\mathcal{K}_k, a/b}^{(c_{\text{ML},k}(\cdot), k)}(\mathbf{x}_0)$ and $L_{\mathcal{K}_k, a/b}^{(c_{\text{HT},k}(\cdot), k)}(\mathbf{x}_0)$, we need to compute the first-order partial derivatives of the mean function. This is accomplished by using Lemma 5.7.3 for the HT estimator and by using a finite-difference quotient approximation [56] for the ML estimator, i.e.,

$$\frac{\partial \mathbf{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}(\mathbf{y})\}}{\partial x_l} \approx \frac{\mathbf{E}_{\mathbf{x}+\Delta \mathbf{e}_l}\{\hat{x}_{\text{ML},k}(\mathbf{y})\} - \mathbf{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}(\mathbf{y})\}}{\Delta}, \quad (5.214)$$

where $\Delta \in \mathbb{R}_+$ is a small stepsize and the expectations are calculated using numerical integration.

We generated parameter vectors $\mathbf{x}_0 = \sqrt{\text{SNR}}\sigma \mathbf{x}_1$, where $\mathbf{x}_1 \in \{0, 1\}^{50}$, $\text{supp}(\mathbf{x}_1) = [S]$, and SNR varies between 0.01 and 100. (The fixed choice $\text{supp}(\mathbf{x}_0) = [S]$ is justified by the fact that neither the variances of the ML and HT estimators nor the corresponding variance bounds depend on the location of $\text{supp}(\mathbf{x}_0)$.) In Fig. 5.3, we plot the variances $v(\hat{\mathbf{x}}_{\text{ML}}(\cdot); \mathbf{x}_0)$ and $v(\hat{\mathbf{x}}_{\text{HT}}(\cdot); \mathbf{x}_0)$ (the latter for four different choices of T in (5.209)) along with the corresponding bounds $L_{\text{ML},a/b}(\mathbf{x}_0)$

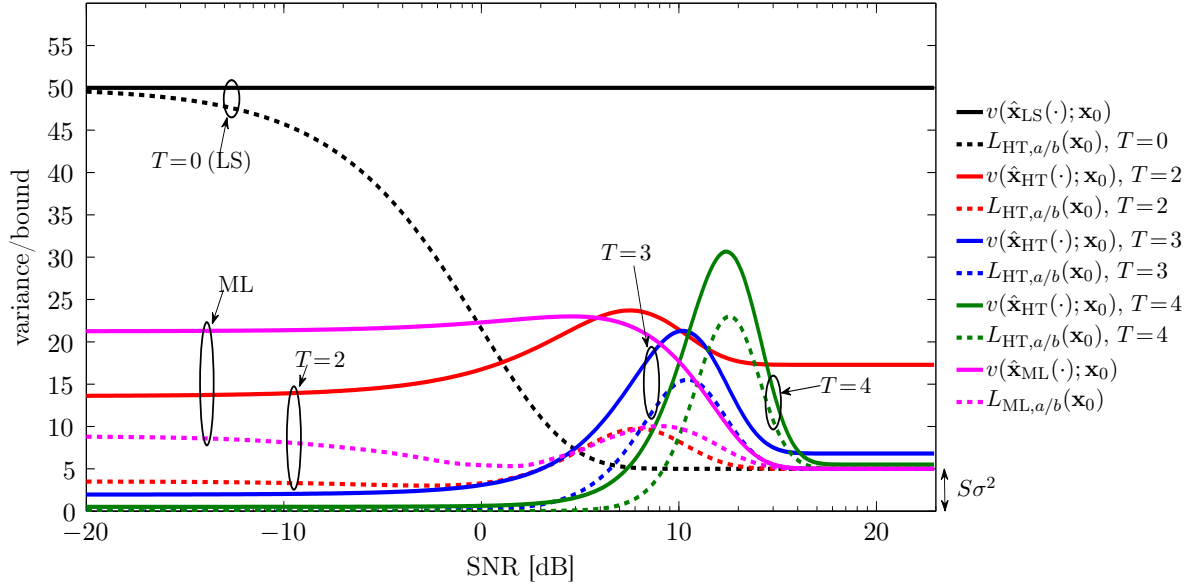


Figure 5.3: Variance of the ML and HT estimators and corresponding lower bounds versus the SNR, for the SSNM with $N=50$, $S=5$ and $\sigma^2 = 1$.

and $L_{\text{HT},a/b}(\mathbf{x}_0)$ (cf. (5.212) and (5.212)), as a function of SNR. The variances are obtained by numerical integration from (5.211) for the HT estimator and from the closed-form expression for the variance of the ML estimator presented in [82]. It is seen that for SNR larger than about 18 dB, all variances and bounds are effectively equal (for the HT estimator, this is true if T is not too small). However, in the medium-SNR range, the variances of the ML and HT estimators are significantly higher than the corresponding lower bounds. We can conclude that there *might* exist estimators with the same mean as that of the ML or HT estimator but a smaller variance. On the other hand, a positive statement regarding the existence of an estimator that has the same mean as the ML or HT estimator but a *uniformly* lower variance cannot be based on our analysis.

However, for the special case of diagonal estimators for the SSNM, such as the HT estimator, Theorem 5.5.4 and Corollary 5.5.7 make a positive statement about the existence of estimators that have *locally* a smaller variance than the HT estimator. In particular, we can use Corollary 5.5.7 to obtain the LMV estimator and corresponding minimum achievable variance at a parameter vector $\mathbf{x}_0 \in \mathcal{X}_S$ for the given bias function of the HT estimator.

As before, we generated parameter vectors $\mathbf{x}_0 = \sqrt{\text{SNR}}\sigma\mathbf{x}_1$, where $\mathbf{x}_1 \in \{0, 1\}^{50}$, $\text{supp}(\mathbf{x}_1) = [S]$, and SNR varies between 0.01 and 100. (The fixed choice $\text{supp}(\mathbf{x}_0) = [S]$ is justified by the fact that neither the variance of the HT estimators nor the corresponding minimum achievable variance depend on the location of $\text{supp}(\mathbf{x}_0)$.) In Fig. 5.4, we plot the variances $v(\hat{\mathbf{x}}_{\text{HT}}(\cdot); \mathbf{x}_0)$

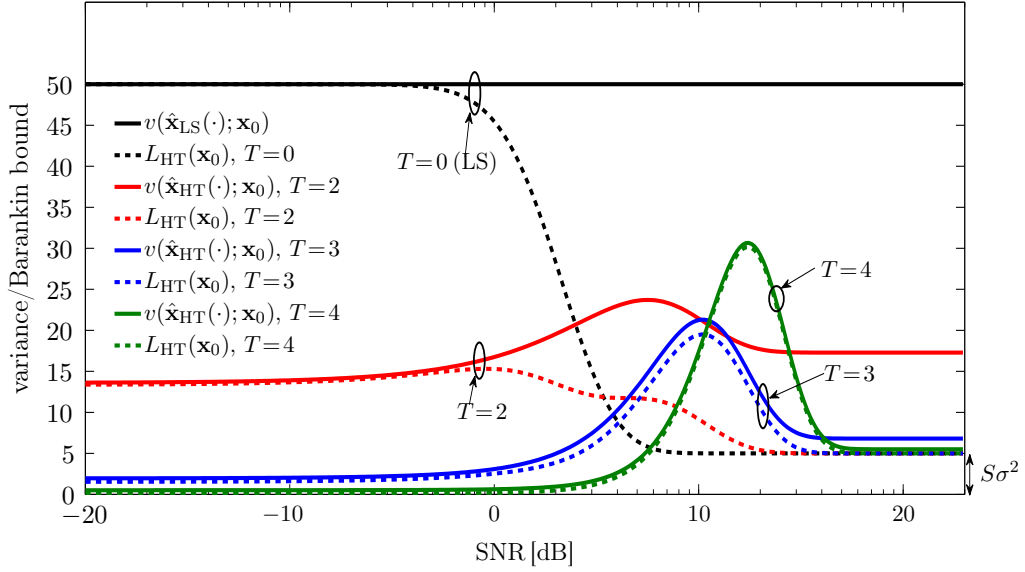


Figure 5.4: Variance of the HT estimator, $v(\hat{\mathbf{x}}_{\text{HT}}(\cdot); \mathbf{x}_0)$, for different T (solid lines) and corresponding minimum achievable variance (Barankin bound) $L_{\text{HT}}(\mathbf{x}_0)$ (dashed lines) versus the SNR, for the SSNM with $N = 50$, $S = 5$, and $\sigma^2 = 1$.

(obtained by numerical integration using (5.211)) for four different choices of T in (5.209), as a function of SNR. We also plot the corresponding minimum achievable variance (Barankin bound) $L_{\text{HT}}(\mathbf{x}_0) \triangleq \sum_{k \in [N]} L_{\mathcal{M}_k}$ applying to estimators for the SSNM that have the same bias (and mean function) as the HT estimator. Note that, by definition, $L_{\text{HT}}(\mathbf{x}_0) \triangleq \sum_{k \in [N]} L_{\mathcal{M}_k}$ is the maximally tight lower bound on the variance of any estimator that has the same bias (and mean function) as the HT estimator. This bound is achieved by the LMV estimator given component-wise by (5.165) with $\hat{x}_k(y_k) = \hat{x}_{\text{HT},k}(y_k)$. The minimum achievable variance $L_{\text{HT}}(\mathbf{x}_0)$ is obtained by summing the minimum achievable variances $L_{\mathcal{M}_k}$ of the scalar minimum variance problems \mathcal{M}_k which coincide with $\mathcal{M}_{\text{SSNM}}$ using the parameter function $g(\mathbf{x}) = x_k$ and prescribed bias $c(\mathbf{x}) = b(\hat{x}_{\text{HT},k}(\mathbf{y}), \mathbf{x})$. We obtained $L_{\mathcal{M}_k}$ by an application of Corollary 5.5.7 since the estimator $\hat{x}_{\text{HT},k}(\mathbf{y})$ is diagonal and has finite variance at every $\mathbf{x}_0 \in \mathcal{X}_S$.

It is seen that for small T (in particular, for $T=0$ where the HT estimator reduces to the LS estimator), the Barankin bound $L_{\text{HT}}(\mathbf{x}_0)$ is significantly below the corresponding variance curve. However, as T increases, the gap between variance and Barankin bound $L_{\text{HT}}(\mathbf{x}_0)$ becomes smaller; in particular, the two curves are already indistinguishable for $T=4$. For high SNR, the Barankin bound $L_{\text{HT}}(\mathbf{x}_0)$ converges to $S\sigma^2$ for any value of T ; this equals the variance of an oracle estimator that knows the support of \mathbf{x} .

Chapter 6

The Sparse Parametric Covariance Model

6.1 Introduction

In the previous chapter, we considered the problem of estimating a sparse parameter vector using a noisy and linearly distorted (by the system matrix \mathbf{H}) observation. Now, we consider the fundamentally different problem of estimating a sparse parameter vector which determines the covariance matrix of a Gaussian signal vector of which a noisy version is observed.

More specifically, we consider a Gaussian signal vector $\mathbf{s} \in \mathbb{R}^M$, i.e., $\mathbf{s} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$ with the psd covariance matrix $\mathbf{C} \in \mathbb{R}^{M \times M}$, embedded in white Gaussian noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_M)$. The observed vector is

$$\mathbf{y} = \mathbf{s} + \mathbf{n}, \quad (6.1)$$

where \mathbf{s} and \mathbf{n} are independent and the signal mean $\boldsymbol{\mu} \in \mathbb{R}^M$ and noise variance $\sigma^2 > 0$ are known. These assumptions imply that the observation \mathbf{y} is a Gaussian random vector with mean $\boldsymbol{\mu}$ and covariance matrix $\mathbf{C} + \sigma^2 \mathbf{I}$, i.e., $\mathbf{y} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C} + \sigma^2 \mathbf{I})$. In what follows, we assume that $\boldsymbol{\mu} = \mathbf{0}$ since a nonzero $\boldsymbol{\mu}$ can always be removed from \mathbf{s} by subtracting it from \mathbf{y} .

The signal covariance matrix $\mathbf{C} \in \mathbb{R}^{M \times M}$ is unknown. We will parametrize it in a linear manner by

$$\mathbf{C} = \mathbf{C}(\mathbf{x}) = \sum_{k \in [N]} x_k \mathbf{C}_k, \quad (6.2)$$

with unknown nonrandom coefficients $x_k \in \mathbb{R}_+$ and known psd “basis matrices” $\mathbf{C}_k \in \mathbb{R}^{M \times M}$, i.e., $\mathbf{C}_k \geq \mathbf{0}$ with ranks $r_k \triangleq \text{rank}(\mathbf{C}_k)$. In what follows, we will use

$$R \triangleq \sum_{k \in [N]} r_k \quad (6.3)$$

to denote the maximum overall dimension of the signal space. We use the term “maximum overall dimension” since the dimension of the signal space is not equal to R in general. It is only equal to R if the basis matrices \mathbf{C}_k fulfill additional requirements. One such additional requirement would be that the basis matrices span disjoint orthogonal subspaces of \mathbb{R}^M . Due to the parameterization (6.2), the estimation of the covariance matrix \mathbf{C} reduces to the estimation of the coefficient vector $\mathbf{x} \triangleq (x_1, \dots, x_N)^T \in \mathbb{R}_+^N$.

Let

$$\tilde{\mathbf{C}}(\mathbf{x}) \triangleq \mathbf{C}(\mathbf{x}) + \sigma^2 \mathbf{I} \quad (6.4)$$

denote the covariance matrix of the Gaussian observation vector \mathbf{y} (cf. (6.1)), i.e., $\text{cov}\{\mathbf{y}\} = \tilde{\mathbf{C}}(\mathbf{x})$. Due to the assumption that $\sigma^2 > 0$, one can verify by Lemma 6.2.2 that $\|\tilde{\mathbf{C}}(\mathbf{x})\|_2 \geq \sigma^2$, i.e., the matrix $\tilde{\mathbf{C}}(\mathbf{x})$ is positive definite ($\tilde{\mathbf{C}}(\mathbf{x}) > \mathbf{0}$) for every parameter vector $\mathbf{x} \in \mathbb{R}_+^N$. Another useful fact which can be verified easily (see, e.g., [38]) is that it holds

$$\|\tilde{\mathbf{C}}^{-1}(\mathbf{x})\|_2 \leq \sigma^{-2} \quad (6.5)$$

for every $\mathbf{x} \in \mathbb{R}_+^N$.

The parameter vector \mathbf{x} is assumed to be S -sparse, i.e.,

$$\mathbf{x} \in \mathcal{X}_{S,+} \triangleq \mathcal{X}_S \cap \mathbb{R}_+^N = \{\mathbf{x}' \in \mathbb{R}_+^N \mid \|\mathbf{x}'\|_0 \leq S\}. \quad (6.6)$$

While the sparsity degree $S \in [N]$ is assumed known, the support of \mathbf{x} , i.e., $\text{supp}(\mathbf{x})$ is unknown. Typically, $S \ll N$.

We now define the sparse parametric covariance model (SPCM) $\mathcal{E}_{\text{SPCM}}$ as the estimation problem given as

$$\mathcal{E}_{\text{SPCM}} = (\mathcal{X}_{S,+}, f_{\text{SPCM}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x})), \quad (6.7)$$

with an arbitrary parameter function $g(\cdot) : \mathcal{X}_{S,+} \rightarrow \mathbb{R}$ and the statistical model

$$f_{\text{SPCM}}(\mathbf{y}; \mathbf{x}) = \frac{1}{(2\pi)^{M/2} [\det\{\tilde{\mathbf{C}}(\mathbf{x})\}]^{1/2}} \exp\left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}) \mathbf{y}\right). \quad (6.8)$$

The statistical model of the SPCM belongs to an exponential family (see Section 2.6). In particular, it is obtained for the sufficient statistic $\Phi(\mathbf{y}) = -\frac{1}{2} \text{vec}\{\mathbf{y}\mathbf{y}^T\}$, parameter function $\mathbf{u}(\mathbf{x}) = \text{vec}\{(\mathbf{C}(\mathbf{x}) + \sigma^2 \mathbf{I})^{-1}\}$, weight function $h(\mathbf{y}) = \frac{1}{\sqrt{(2\pi)^M}}$, and cumulant function $A^{(\Phi)}(\mathbf{x}) = \frac{1}{2} \log(\det\{\mathbf{C}(\mathbf{x}) + \sigma^2 \mathbf{I}\})$.

Note that by contrast to the SLM, the entries of the parameter vector $\mathbf{x} \in \mathcal{X}_{S,+}$ are non-negative. Another difference from the SLM is that we allow for a general parameter function $g(\cdot) : \mathcal{X}_{S,+} \rightarrow \mathbb{R}$ instead of fixing the parameter function to $g(\mathbf{x}) = x_k$ as for the SLM. However, according to Theorem 2.3.1, this latter difference is only a notational one as long as we allow for

arbitrary prescribed bias functions $c(\cdot) : \mathcal{X}_{S,+} \rightarrow \mathbb{R}$ for the minimum variance problems arising from the SLM \mathcal{E}_{SLM} .

Since the basis matrices \mathbf{C}_k are assumed to be psd, we can factor them in the form $\mathbf{C}_k = \mathbf{H}_k \mathbf{H}_k^T$, where the factors $\mathbf{H}_k \in \mathbb{R}^{M \times r_k}$ are not unique [38, 64]. One possible choice of \mathbf{H}_k is obtained via the thin EVD $\mathbf{C}_k = \mathbf{U} \boldsymbol{\Sigma} \mathbf{U}^T$ by setting $\mathbf{H}_k = \mathbf{U} \sqrt{\boldsymbol{\Sigma}}$. If we consider a specific factorization of the basis matrices and stack the resulting factors \mathbf{H}_k into the matrix

$$\mathbf{H} \triangleq \begin{pmatrix} \mathbf{H}_1 & \mathbf{H}_2 & \dots & \mathbf{H}_N \end{pmatrix} \in \mathbb{R}^{M \times R}, \quad (6.9)$$

we can rewrite the parameterization (6.2) as

$$\mathbf{C} = \mathbf{C}(\mathbf{x}) = \mathbf{H} \mathbf{D}(\mathbf{x}) \mathbf{H}^T. \quad (6.10)$$

Here, $\mathbf{D}(\mathbf{x}) \in \mathbb{R}^{R \times R}$ is a diagonal matrix which has on its main diagonal the entries of $\mathbf{x} \in \mathbb{R}^N$, where each entry x_k appears r_k times, i.e.,

$$\mathbf{D}(\mathbf{x}) = \begin{pmatrix} x_1 \mathbf{I}_{r_1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & x_2 \mathbf{I}_{r_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & x_N \mathbf{I}_{r_N} \end{pmatrix}. \quad (6.11)$$

A partial characterization of the set of basis matrices can be made based on

Definition 6.1.1. *Given the SPCM $\mathcal{E}_{\text{SPCM}}$, we say that the set of basis matrices $\{\mathbf{C}_k \in \mathbb{R}^{M \times M}\}_{k \in [N]}$ has RIP of order K with constant $\delta_K \in [0, 1]$ if there exist factorizations $\mathbf{C}_k = \mathbf{H}_k \mathbf{H}_k^T$ such that the singular values of any matrix $\mathbf{H}^{(\mathcal{I})} = \begin{pmatrix} \mathbf{H}_{i_1} & \mathbf{H}_{i_2} & \dots & \mathbf{H}_{i_K} \end{pmatrix} \in \mathbb{R}^{M \times \sum_{k \in \mathcal{I}} r_k}$, where $\mathcal{I} = \{i_1, \dots, i_K\} \subseteq [N]$ is an arbitrary set of K different indices, are located in the interval $[1 - \delta_K, 1 + \delta_K]$ and moreover any matrix $\mathbf{H}^{(\mathcal{I})}$ has full column rank, i.e., $\text{rank}(\mathbf{H}^{(\mathcal{I})}) = \sum_{k \in \mathcal{I}} r_k$.*

The SPCM and estimation of \mathbf{x} are relevant, e.g., in time-frequency analysis [95, 96], where the basis matrices \mathbf{C}_k correspond to disjoint time-frequency regions and x_k represents the mean signal energy in the k th time-frequency region. Such a sparsity-based time-frequency analysis may be used, e.g., for cognitive radio scene analysis [97]. Another application that fits our scope would be compressive spectrum estimation [98, 99], where classical (non-stationary) spectrum estimation techniques are combined with novel CS methods.

An important special case of the SPCM is obtained if the basis matrices \mathbf{C}_k appearing in the parameterization (6.2) are orthogonal projection matrices on orthogonal subspaces of \mathbb{R}^M , i.e.,

$$\mathbf{C}_k \mathbf{C}_{k'} = \delta_{k,k'} \mathbf{C}_k, \quad \forall k, k' \in [N]. \quad (6.12)$$

Note that condition (6.12) implies that the basis matrices satisfy the RIP of any order $K \in [N]$ with constant $\delta_K = 0$. The so-obtained special case of the SPCM will be termed the *sparse diagonalizable parametric covariance model* (SDPCM). This naming is due to the fact that for the SDPCM, i.e., when the basis matrices \mathbf{C}_k in (6.2) satisfy (6.12), the covariance matrix $\mathbf{C}(\mathbf{x})$ given by (6.2) can be diagonalized by a signal transformation $\mathbf{s}' = \mathbf{U}\mathbf{s}$, with an orthonormal matrix $\mathbf{U} \in \mathbb{R}^{M \times M}$, i.e., $\mathbf{U}^T \mathbf{U} = \mathbf{I}$, that does not depend on the parameter vector \mathbf{x} . We will denote by $\mathcal{E}_{\text{SDPCM}}$ the special case of the estimation problem $\mathcal{E}_{\text{SPCM}}$ in (6.7) with basis matrices satisfying (6.12). More specifically,

$$\mathcal{E}_{\text{SDPCM}} = (\mathcal{X}_{S,+}, f_{\text{SDPCM}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x})), \quad (6.13)$$

with an arbitrary parameter function $g(\cdot) : \mathcal{X}_{S,+} \rightarrow \mathbb{R}$ and the statistical model

$$f_{\text{SDPCM}}(\mathbf{y}; \mathbf{x}) = \frac{1}{(2\pi)^{M/2} [\det\{\tilde{\mathbf{C}}(\mathbf{x})\}]^{1/2}} \exp\left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}) \mathbf{y}\right), \quad (6.14)$$

where $\tilde{\mathbf{C}}(\mathbf{x})$ is defined as in (6.4) but with basis matrices \mathbf{C}_k satisfying (6.12).

The key results of this chapter have been presented in part in [100].

6.2 The \mathcal{D} -restricted SPCM

As already observed in [35], the RKHS-based analysis of minimum variance estimation for estimation problems where the parameter determines the covariance matrix of the observation (e.g., the SPCM) is more involved than for those problems where the parameter determines the mean (e.g., the SLM). In particular, any minimum variance problem $\mathcal{M}_{\text{SPCM}} = (\mathcal{E}_{\text{SPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ associated with the SPCM and parameter vector $\mathbf{x}_0 \in \mathcal{X}_{S,+}$ does not fulfill Postulate 4.1.1. This precludes a direct RKHS approach to the SPCM.

One specific reason for the minimum variance problem $\mathcal{M}_{\text{SPCM}}$ to fail in satisfying Postulate 4.1.1 is that the space $\mathcal{L}(\mathcal{M}_{\text{SPCM}})$ (cf. (4.5)) with the inner product $\langle \cdot, \cdot \rangle_{\text{RV}}$ (see (4.3)) fails to be a Hilbert space. An ad-hoc approach in order to facilitate an RKHS-based analysis of the minimum variance problem $\mathcal{M}_{\text{SPCM}}$ could be to reduce the space $\mathcal{L}(\mathcal{M}_{\text{SPCM}})$ (consisting of linear combinations of the elementary estimator functions $\hat{g}_{\mathbf{x}}(\cdot) = \rho_{\mathcal{M}}(\cdot, \mathbf{x})$) to obtain a Hilbert space $\mathcal{L}'(\mathcal{M}_{\text{SPCM}}) \subseteq \mathcal{L}(\mathcal{M}_{\text{SPCM}})$. However, the resulting RKHS associated with $\mathcal{L}'(\mathcal{M}_{\text{SPCM}})$ depends on the completely arbitrary reduction of $\mathcal{L}(\mathcal{M}_{\text{SPCM}})$ to $\mathcal{L}'(\mathcal{M}_{\text{SPCM}})$, and a convenient analytical description of the resulting RKHS can be hardly expected in general. Therefore, as discussed presently, we will pursue a different RKHS approach to the minimum variance problem $\mathcal{M}_{\text{SPCM}} = (\mathcal{E}_{\text{SPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$.

Specifically, in order to derive lower bounds on the minimum achievable variance $L_{\mathcal{M}_{\text{SPCM}}}$, we

can further restrict the parameter set of $\mathcal{E}_{\text{SPCM}}$ to obtain the so called “ \mathcal{D} -restricted” SPCM

$$\mathcal{E}_{\text{SPCM}}^{(\mathcal{D})} \triangleq (\mathcal{D}, f_{\text{SPCM}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x})) \quad (6.15)$$

with a certain set $\mathcal{D} \subseteq \mathcal{X}_{S,+}$. Indeed, consider the minimum variance problem $\mathcal{M}_{\text{SPCM}} = (\mathcal{E}_{\text{SPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ associated with the SPCM $\mathcal{E}_{\text{SPCM}}$ and the “ \mathcal{D} -restricted” minimum variance problem

$$\mathcal{M}_{\mathcal{D},\text{SPCM}} \triangleq \mathcal{M}_{\text{SPCM}}|_{\mathcal{D}} = (\mathcal{E}_{\text{SPCM}}^{(\mathcal{D})}, c(\cdot) \equiv 0, \mathbf{x}_0) \quad (6.16)$$

associated with the \mathcal{D} -restricted SPCM $\mathcal{E}_{\text{SPCM}}^{(\mathcal{D})}$ and the same parameter vector \mathbf{x}_0 as used for $\mathcal{M}_{\text{SPCM}}$. We do not require the set \mathcal{D} to contain \mathbf{x}_0 , i.e., we may have $\mathbf{x}_0 \notin \mathcal{D}$. According to Theorem 2.5.2 (cf. also Section 4.3.4), we have then the inequality

$$L_{\mathcal{M}_{\text{SPCM}}} \geq L_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}. \quad (6.17)$$

The idea is now to choose the set $\mathcal{D} \subseteq \mathcal{X}_{S,+}$ such that $\mathcal{M}_{\mathcal{D},\text{SPCM}}$ fulfills Postulate 4.1.1, so that the RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})$ as in Definition 4.3.1 exists for the minimum variance problem $\mathcal{M}_{\mathcal{D},\text{SPCM}}$. Based on (6.17), we can then find lower bounds for $L_{\mathcal{M}_{\text{SPCM}}}$ by deriving lower bounds on $L_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}$ using the RKHS theory for $\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})$.

Postulate 4.1.1 can be written for the SPCM as (see (6.8))

$$\begin{aligned} \mathbb{E}_{\mathbf{x}_0} \left\{ \left(\frac{f_{\text{SPCM}}(\mathbf{y}; \mathbf{x})}{f_{\text{SPCM}}(\mathbf{y}; \mathbf{x}_0)} \right)^2 \right\} &= \int_{\mathbf{y}} \frac{f_{\text{SPCM}}(\mathbf{y}; \mathbf{x}) f_{\text{SPCM}}(\mathbf{y}; \mathbf{x})}{f_{\text{SPCM}}(\mathbf{y}; \mathbf{x}_0)} d\mathbf{y} \\ &= \frac{1}{(2\pi)^{M/2}} \frac{[\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2}}{\det \{\tilde{\mathbf{C}}(\mathbf{x})\}} \int_{\mathbf{y}} \exp \left(-\frac{1}{2} \mathbf{y}^T [2\tilde{\mathbf{C}}^{-1}(\mathbf{x}) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)] \mathbf{y} \right) d\mathbf{y} < \infty. \end{aligned} \quad (6.18)$$

Since by (6.4) we have $\tilde{\mathbf{C}}(\mathbf{x}) > \mathbf{0}$ and therefore $\det \{\tilde{\mathbf{C}}(\mathbf{x})\} > 0$ for every $\mathbf{x} \in \mathcal{X}_{S,+}$, we can conclude from (6.18) that in order to satisfy Postulate 4.1.1, it is necessary and sufficient to choose the set \mathcal{D} such that $(2\tilde{\mathbf{C}}^{-1}(\mathbf{x}) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0))$ is positive definite, i.e.,

$$2\tilde{\mathbf{C}}^{-1}(\mathbf{x}) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) > \mathbf{0} \quad \forall \mathbf{x} \in \mathcal{D}. \quad (6.19)$$

We will now give a sufficient condition for the set \mathcal{D} to fulfill (6.19). To that end, the following two technical results will prove handy. Remember that given a square matrix $\mathbf{H} \in \mathbb{R}^{M \times M}$, we denote by $\|\mathbf{H}\|_2$ its matrix 2-norm defined as $\|\mathbf{H}\|_2 \triangleq \sup_{\|\mathbf{x}\|_p=1} \|\mathbf{H}\mathbf{x}\|_2$.

Lemma 6.2.1. *Consider an invertible matrix $\mathbf{A} \in \mathbb{R}^{M \times M}$ and a matrix $\mathbf{E} \in \mathbb{R}^{M \times M}$. Then if $\|\mathbf{E}\|_2 < \frac{1}{\|\mathbf{A}^{-1}\|_2}$, we have that $\mathbf{A} + \mathbf{E}$ is invertible and*

$$\|(\mathbf{A} + \mathbf{E})^{-1} - \mathbf{A}^{-1}\|_2 \leq \frac{\|\mathbf{A}^{-1}\|_2^2 \|\mathbf{E}\|_2}{1 - \|\mathbf{A}^{-1}\|_2 \|\mathbf{E}\|_2}. \quad (6.20)$$

Proof. The statement follows by combining [38, Theorem 2.3.4] with the fact that for two matrices $\mathbf{N}, \mathbf{M} \in \mathbb{R}^{M \times M}$ we have $\|\mathbf{NM}\|_2 \leq \|\mathbf{N}\|_2 \|\mathbf{M}\|_2$ [38, p. 55]. \square

In what follows, we denote the k th largest eigenvalue of a symmetric matrix $\mathbf{C} \in \mathbb{R}^{L \times L}$, with $k \in [L]$, by $\lambda_k(\mathbf{C})$, i.e.,

$$\lambda_L(\mathbf{C}) \leq \dots \leq \lambda_2(\mathbf{C}) \leq \lambda_1(\mathbf{C}). \quad (6.21)$$

The eigenvalues of a symmetric matrix are nonnegative (positive) if and only if the matrix is positive semi-definite (positive definite) [38, 101]. It can be verified that $\|\mathbf{C}\|_2 = \max\{|\lambda_1(\mathbf{C})|, |\lambda_L(\mathbf{C})|\}$ [38, p. 394] and in turn

$$|\lambda_k(\mathbf{C})| \leq \|\mathbf{C}\|_2. \quad (6.22)$$

Moreover, for any matrix $\mathbf{C} \in \mathbb{R}^{L \times L}$ it holds that [38, p. 310]

$$\det\{\mathbf{C}\} = \prod_{k \in [L]} \lambda_k(\mathbf{C}). \quad (6.23)$$

Lemma 6.2.2. ([38, Theorem 8.1.5]) *Consider two symmetric matrices $\mathbf{A}, \mathbf{E} \in \mathbb{R}^{L \times L}$, i.e., $\mathbf{A}^T = \mathbf{A}$ and $\mathbf{E}^T = \mathbf{E}$. We then have the double inequality*

$$\lambda_k(\mathbf{A}) + \lambda_L(\mathbf{E}) \leq \lambda_k(\mathbf{A} + \mathbf{E}) \leq \lambda_k(\mathbf{A}) + \lambda_1(\mathbf{E}), \quad \text{for every } k \in [L]. \quad (6.24)$$

Proof. [38] \square

The next result shows that \mathcal{D} can be chosen as the intersection of $\mathcal{X}_{S,+}$ and a ball $\mathcal{B}(\mathbf{x}_c, r)$, with a suitable center $\mathbf{x}_c \in \mathcal{X}_{S,+}$ and a sufficiently small radius r :

Theorem 6.2.3. *Consider the minimum variance problem $\mathcal{M}_{\text{SPCM}} = (\mathcal{E}_{\text{SPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$. Any set \mathcal{D} given as*

$$\mathcal{D} = \mathcal{X}_{S,+} \cap \mathcal{B}(\mathbf{x}_c, r), \quad (6.25)$$

with $\mathbf{x}_c \in \mathcal{X}_{S,+}$ such that

$$2\tilde{\mathbf{C}}^{-1}(\mathbf{x}_c) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) > \mathbf{0}, \quad (6.26)$$

and a sufficiently small radius $r > 0$, satisfies (6.19). Two specific choices for \mathbf{x}_c which satisfy (6.26) are $\mathbf{x}_c = \mathbf{x}_0$ and $\mathbf{x}_c = \mathbf{0}$.

Proof. Let us consider a parameter vector $\mathbf{x}_c \in \mathcal{X}_{S,+}$ such that (6.26) is satisfied. We can then write $\tilde{\mathbf{C}}(\mathbf{x}) = \tilde{\mathbf{C}}(\mathbf{x}_c) + \mathbf{E}(\mathbf{x})$, where

$$\mathbf{E}(\mathbf{x}) \triangleq \tilde{\mathbf{C}}(\mathbf{x}) - \tilde{\mathbf{C}}(\mathbf{x}_c) = \sum_{k \in [N]} \mathbf{C}_k \bar{x}_k \quad (6.27)$$

is a symmetric matrix, i.e. $\mathbf{E}^T(\mathbf{x}) = \mathbf{E}(\mathbf{x})$, with $\bar{x}_k \triangleq x_k - x_{c,k}$. For every $\mathbf{x} \in \mathcal{D} = \mathcal{X}_{S,+} \cap \mathcal{B}(\mathbf{x}_c, r)$ (implying that $\|\mathbf{x} - \mathbf{x}_c\|_2 \leq r$), we have that

$$|\bar{x}_k| = |x_k - x_{c,k}| = \sqrt{|x_k - x_{c,k}|^2} \leq \sqrt{\sum_{k \in [N]} |x_k - x_{c,k}|^2} = \|\mathbf{x} - \mathbf{x}_c\|_2 \leq r. \quad (6.28)$$

This implies, via the triangle inequality for the matrix 2-norm, that

$$\|\mathbf{E}(\mathbf{x})\|_2 \leq N q r, \quad (6.29)$$

with $q \triangleq \max_{k \in [N]} \|\mathbf{C}_k\|_2$. Indeed, we have

$$\begin{aligned} \|\mathbf{E}(\mathbf{x})\|_2 &\stackrel{(6.27)}{=} \left\| \sum_{k \in [N]} \mathbf{C}_k \bar{x}_k \right\|_2 \leq \sum_{k \in [N]} \|\mathbf{C}_k \bar{x}_k\|_2 \\ &= \sum_{k \in [N]} |\bar{x}_k| \|\mathbf{C}_k\|_2 \stackrel{(6.28)}{\leq} r \sum_{k \in [N]} \|\mathbf{C}_k\|_2 \leq r N \max_{k \in [N]} \|\mathbf{C}_k\|_2. \end{aligned} \quad (6.30)$$

Let us now choose a radius r_0 such that $N q r_0 < \frac{1}{\|\tilde{\mathbf{C}}^{-1}(\mathbf{x}_c)\|}$ which implies via (6.29) that

$$\|\mathbf{E}(\mathbf{x})\|_2 \leq N q r_0 < \frac{1}{\|\tilde{\mathbf{C}}^{-1}(\mathbf{x}_c)\|} \quad (6.31)$$

for every $\mathbf{x} \in \mathcal{X}_{S,+} \cap \mathcal{B}(\mathbf{x}_c, r_0)$. The fulfillment of condition (6.31) allows us then to invoke Lemma 6.2.1, using $\mathbf{A} = \tilde{\mathbf{C}}(\mathbf{x}_c)$ and $\mathbf{E} = \mathbf{E}(\mathbf{x})$, which yields that for every $\mathbf{x} \in \mathcal{X}_{S,+} \cap \mathcal{B}(\mathbf{x}_c, r_0)$ we have

$$\tilde{\mathbf{C}}^{-1}(\mathbf{x}) = \tilde{\mathbf{C}}^{-1}(\mathbf{x}_c) + \mathbf{E}_0, \quad (6.32)$$

with a symmetric matrix $\mathbf{E}_0 \in \mathbb{R}^{M \times M}$ satisfying

$$\|\mathbf{E}_0\|_2 = \|\tilde{\mathbf{C}}^{-1}(\mathbf{x}) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_c)\|_2 \stackrel{(6.20), (6.27)}{\leq} \frac{\|\tilde{\mathbf{C}}^{-1}(\mathbf{x}_c)\|_2^2 \|\mathbf{E}(\mathbf{x})\|_2}{1 - \|\tilde{\mathbf{C}}^{-1}(\mathbf{x}_c)\|_2 \|\mathbf{E}(\mathbf{x})\|_2}$$

$$\stackrel{(6.31),(6.5)}{\leq} \frac{Nqr_0\sigma^{-4}}{1 - \sigma^{-2}Nqr_0}. \quad (6.33)$$

By choosing r_0 small enough we can, according to (6.33), make $\|\mathbf{E}_0\|_2$ arbitrarily small. Observe that by assumption the matrix $\mathbf{F} \triangleq 2\tilde{\mathbf{C}}^{-1}(\mathbf{x}_c) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) \in \mathbb{R}^{M \times M}$ is positive definite, which implies that its eigenvalues $\lambda_k(\mathbf{F})$ are positive, i.e., they satisfy $\lambda_k(\mathbf{F}) \geq \varepsilon_0$ with some positive constant $\varepsilon_0 > 0$. We can then choose r_0 such that $\|\mathbf{E}_0\|_2 < \varepsilon_0/2$ (implying that $|\lambda_M(\mathbf{E}_0)| < \varepsilon_0/2$) for every $\mathbf{x} \in \mathcal{X}_{S,+} \cap \mathcal{B}(\mathbf{x}_c, r_0)$, which implies by Lemma 6.2.2 that the eigenvalues $\lambda_k(\mathbf{G})$ of the symmetric matrix

$$\mathbf{G} \triangleq 2\tilde{\mathbf{C}}^{-1}(\mathbf{x}) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) \stackrel{(6.32)}{=} 2\tilde{\mathbf{C}}^{-1}(\mathbf{x}_c) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) + 2\mathbf{E}_0 = \mathbf{F} + 2\mathbf{E}_0. \quad (6.34)$$

satisfy

$$\lambda_k(\mathbf{G}) = \lambda_k(\mathbf{F} + 2\mathbf{E}_0) \stackrel{(6.24)}{\geq} \lambda_k(\mathbf{F}) + \lambda_M(2\mathbf{E}_0) = \lambda_k(\mathbf{F}) + 2\lambda_M(\mathbf{E}_0) \geq \varepsilon_0 - 2\|\mathbf{E}_0\|_2 > 0, \quad (6.35)$$

i.e., all eigenvalues of the matrix $\mathbf{G} = 2\tilde{\mathbf{C}}^{-1}(\mathbf{x}) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)$ are strictly positive. This in turn means that the matrix $2\tilde{\mathbf{C}}^{-1}(\mathbf{x}) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)$ is positive definite for every $\mathbf{x} \in \mathcal{X}_{S,+} \cap \mathcal{B}(\mathbf{x}_c, r_0)$.

Finally, for $\mathbf{x}_c = \mathbf{x}_0$, the matrix $2\tilde{\mathbf{C}}^{-1}(\mathbf{x}_c) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) = \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)$ is obviously positive definite since the matrix $\tilde{\mathbf{C}}(\mathbf{x}_0)$ is positive definite. Also, for $\mathbf{x}_c = \mathbf{0}$, the matrix

$$2\tilde{\mathbf{C}}^{-1}(\mathbf{x}_c) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) \stackrel{(6.4)}{=} 2[\mathbf{C}(\mathbf{0}) + \sigma^2\mathbf{I}]^{-1} - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) \stackrel{(6.2)}{=} 2\sigma^{-2}\mathbf{I} - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) \quad (6.36)$$

is positive definite since its eigenvalues obey

$$\begin{aligned} \lambda_k(2\tilde{\mathbf{C}}^{-1}(\mathbf{x}_c) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)) &\stackrel{(6.36)}{=} \lambda_k(2\sigma^{-2}\mathbf{I} - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)) \stackrel{(6.24)}{\geq} \lambda_k(2\sigma^{-2}\mathbf{I}) + \lambda_M(-\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)) \\ &= 2\sigma^{-2} + \lambda_M(-\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)) \stackrel{(6.5)}{\geq} 2\sigma^{-2} - \sigma^{-2} = \sigma^{-2} > 0, \end{aligned} \quad (6.37)$$

i.e., are all strictly positive. \square

In what follows, we will always assume that the set \mathcal{D} is of the form (6.25). Note that any set $\mathcal{D} \subseteq \mathbb{R}^N$ of the form (6.25) is closed and bounded and therefore *compact* [5, Theorem 2.41]. If the set \mathcal{D} is chosen such that (6.19) is fulfilled, we obtain the kernel $R_{\mathcal{M}_{\mathcal{D}}, \text{SPCM}}(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ (see (4.10) and (6.8)):

$$\begin{aligned} R_{\mathcal{M}_{\mathcal{D}}, \text{SPCM}}(\mathbf{x}_1, \mathbf{x}_2) &\triangleq \mathbf{E}_{\mathbf{x}_0} \{ \rho_{\mathcal{M}_{\text{SPCM}}}(\mathbf{y}, \mathbf{x}_1) \rho_{\mathcal{M}_{\text{SPCM}}}(\mathbf{y}, \mathbf{x}_2) \} = \int_{\mathbf{y} \in \mathbb{R}^M} \frac{f_{\text{SPCM}}(\mathbf{y}; \mathbf{x}_1) f_{\text{SPCM}}(\mathbf{y}; \mathbf{x}_2)}{f_{\text{SPCM}}(\mathbf{y}; \mathbf{x}_0)} d\mathbf{y} \\ &= \frac{(2\pi)^{-M/2} [\det \{ \tilde{\mathbf{C}}(\mathbf{x}_0) \}]^{1/2}}{[\det \{ \tilde{\mathbf{C}}(\mathbf{x}_1) \} \det \{ \tilde{\mathbf{C}}(\mathbf{x}_2) \}]^{1/2}} \int_{\mathbf{y} \in \mathbb{R}^M} \exp \left(-\frac{1}{2} \mathbf{y}^T \left(\tilde{\mathbf{C}}^{-1}(\mathbf{x}_1) + \tilde{\mathbf{C}}^{-1}(\mathbf{x}_2) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) \right) \mathbf{y} \right) d\mathbf{y} \end{aligned}$$

$$\begin{aligned}
&= [\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} [\det \{\tilde{\mathbf{C}}(\mathbf{x}_1)\tilde{\mathbf{C}}(\mathbf{x}_2)(\tilde{\mathbf{C}}^{-1}(\mathbf{x}_1) + \tilde{\mathbf{C}}^{-1}(\mathbf{x}_2) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0))\}]^{-1/2} \\
&= [\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} [\det \{\tilde{\mathbf{C}}(\mathbf{x}_1) + \tilde{\mathbf{C}}(\mathbf{x}_2) - \tilde{\mathbf{C}}(\mathbf{x}_1)\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\tilde{\mathbf{C}}(\mathbf{x}_2)\}]^{-1/2}, \tag{6.38}
\end{aligned}$$

with $\tilde{\mathbf{C}}(\mathbf{x})$ as defined in (6.4). Obviously, according to elementary linear algebra [38], the kernel $R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\mathbf{x}_1, \mathbf{x}_2)$ is differentiable in the sense of Definition 3.4.1 up to any order m and therefore also continuous. Moreover, since we assume that the set \mathcal{D} is of the form (6.25), the function

$$k(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R} : k(\mathbf{x}) = R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\mathbf{x}, \mathbf{x}) = [\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} [\det \{2\tilde{\mathbf{C}}(\mathbf{x}) - \tilde{\mathbf{C}}(\mathbf{x})\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\tilde{\mathbf{C}}(\mathbf{x})\}]^{-1/2} \tag{6.39}$$

is bounded over its domain \mathcal{D} . In order to verify the boundedness, we observe that (i) every set of the form (6.25) is closed¹ and (ii) the fact that $\det \{2\tilde{\mathbf{C}}^{-1}(\mathbf{x}) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\}$ is a continuous function of \mathbf{x} . It follows then from the extreme value theorem [5, Theorem 4.16] and the compactness, i.e., boundedness and closedness, of \mathcal{D} that there must exist a positive constant $\varepsilon > 0$ such that

$$\det \{2\tilde{\mathbf{C}}^{-1}(\mathbf{x}) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\} \geq \varepsilon \tag{6.40}$$

for every vector $\mathbf{x} \in \mathcal{D}$. This implies that

$$\begin{aligned}
k(\mathbf{x}) &= [\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} [\det \{2\tilde{\mathbf{C}}(\mathbf{x}) - \tilde{\mathbf{C}}(\mathbf{x})\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\tilde{\mathbf{C}}(\mathbf{x})\}]^{-1/2} \\
&= [\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} \left[[\det \{\tilde{\mathbf{C}}(\mathbf{x})\}]^2 \det \{2\tilde{\mathbf{C}}^{-1}(\mathbf{x}) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\} \right]^{-1/2} \\
&\stackrel{\det\{\mathbf{A}^{-1}\} = \det^{-1}\{\mathbf{A}\}}{=} [\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} \det \{\tilde{\mathbf{C}}^{-1}(\mathbf{x})\} [\det \{2\tilde{\mathbf{C}}^{-1}(\mathbf{x}) - \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\}]^{-1/2} \\
&\stackrel{(6.5), (6.40)}{\leq} [\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} \sigma^{-2M} \varepsilon^{-1/2}, \tag{6.41}
\end{aligned}$$

i.e., the function $k(\mathbf{x})$ is bounded by the constant (w.r.t. \mathbf{x}) $[\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} \sigma^{-2M} \varepsilon^{-1/2}$. Therefore, we can apply Theorem 3.4.3 to derive

Theorem 6.2.4. *Consider the minimum variance problem $\mathcal{M}_{\text{SPCM}} = (\mathcal{E}_{\text{SPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0 \in \mathcal{X}_{S,+})$ and a set $\mathcal{D} = \mathcal{X}_{S,+} \cap \mathcal{B}(\mathbf{x}_c, r)$ satisfying (6.19) so that $R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \cdot) : \mathcal{D} \times \mathcal{D} \rightarrow \mathbb{R}$ exists. Then, if the parameter function $g(\cdot) : \mathcal{X}_{S,+} \rightarrow \mathbb{R}$ of the associated SPCM $\mathcal{E}_{\text{SPCM}}$ has a discontinuity on the set \mathcal{D} , i.e., there exists a point $\mathbf{x} \in \mathcal{D}$ at which the function $g(\cdot)$ is discontinuous, there exists no allowed estimator (i.e., unbiased and with finite variance at \mathbf{x}_0) for $\mathcal{M}_{\text{SPCM}}$.*

Proof. Consider the \mathcal{D} -restricted minimum variance problem $\mathcal{M}' = \mathcal{M}_{\text{SPCM}}|_{\mathcal{D}}$ with the set \mathcal{D} of the form (6.25). The RKHS $\mathcal{H}(\mathcal{M}')$ associated to the kernel $R_{\mathcal{M}'}$ (see (6.38)) exists since the set \mathcal{D}

¹We implicitly assume the Hilbert space structure in \mathbb{R}^N , which is induced by the inner product $\langle \mathbf{x}, \mathbf{y} \rangle \triangleq \mathbf{x}^T \mathbf{y}$

is assumed to satisfy (6.19). Moreover, according to our discussion above, Theorem 3.4.3 applies to the RKHS $\mathcal{H}(\mathcal{M}')$. We have by Theorem 3.4.3 that $\mathcal{H}(\mathcal{M}')$ consists exclusively of continuous functions. Therefore, it cannot contain any function with a discontinuity within the set \mathcal{D} . By Theorem 4.3.4, this implies that any discontinuous function is not estimable for \mathcal{M}' . However, since \mathcal{M}' is obtained by a reduction of the parameter set underlying $\mathcal{M}_{\text{SPCM}}$, it follows from Theorem 2.5.2 that also for the minimum variance problem $\mathcal{M}_{\text{SPCM}}$ every discontinuous function is not estimable, i.e., there exists no unbiased estimator of a discontinuous (on \mathcal{D}) function with finite variance at $\mathbf{x}_0 \in \mathcal{X}_{S,+}$. \square

As for the SLM (cf. Theorem 5.3.2), we have also for the SPCM that there exists no finite-variance unbiased estimator of an injective function of the support $\text{supp}(\mathbf{x})$:

Corollary 6.2.5. *There exists no estimator $\hat{g}(\mathbf{y})$ of an injective real-valued function $g(\text{supp}(\mathbf{x}))$ of the support of $\mathbf{x} \in \mathcal{X}_{S,+}$ which uses only the observation (6.1) of the SPCM $\mathcal{E}_{\text{SPCM}}$, is unbiased for every $\mathbf{x} \in \mathcal{X}_S$, and has a finite variance at any $\mathbf{x}_0 \in \mathcal{X}_{S,+}$.*

Proof. Consider a set $\mathcal{D} = \mathcal{X}_{S,+} \cap \mathcal{B}(\mathbf{0}, r)$ with a sufficiently small radius $r > 0$ such that, according to Theorem 6.2.3, the set \mathcal{D} fulfills (6.19), and with an arbitrary parameter vector $\mathbf{x}_0 \in \mathcal{X}_{S,+}$. As verified in the proof of Theorem 5.3.2, any injective function $g(\text{supp}(\mathbf{x}))$ of the support must have at least one discontinuity on the set \mathcal{D} . Therefore, by Theorem 6.2.4, there does not exist an allowed estimator for the minimum variance problem $\mathcal{M}_{\text{SPCM}} = (\mathcal{E}_{\text{SPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0 \in \mathcal{X}_{S,+})$ associated with the SPCM with parameter function $g(\text{supp}(\mathbf{x}))$. By the definition of an allowed estimator, this means that there exists no estimator of $g(\text{supp}(\mathbf{x}))$ that is unbiased for every $\mathbf{x} \in \mathcal{X}_{S,+}$ and has finite variance at \mathbf{x}_0 . \square

6.3 Lower Bounds on the Estimator Variance for the SPCM

Let us consider the minimum variance problem $\mathcal{M}_{\text{SPCM}} = (\mathcal{E}_{\text{SPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ associated with the SPCM, for a parameter vector $\mathbf{x}_0 \in \mathcal{X}_{S,+}$. Note again that the prescribed bias $c(\cdot)$ is assumed to vanish on $\mathcal{X}_{S,+}$, i.e., we consider only unbiased estimation. However, by Theorem 2.3.1, this is no real restriction since we allow for general parameter functions $g(\mathbf{x})$ in the SPCM. We then define a second minimum variance problem $\mathcal{M}_{\mathcal{D},\text{SPCM}} \triangleq \mathcal{M}_{\text{SPCM}}|_{\mathcal{D}} = (\mathcal{E}_{\text{SPCM}}^{(\mathcal{D})}, c(\cdot) \equiv 0, \mathbf{x}_0)$ associated to the \mathcal{D} -restricted SPCM, which is identical to $\mathcal{M}_{\text{SPCM}}$ except for the parameter set $\mathcal{D} \subseteq \mathcal{X}_{S,+}$.

We will now derive lower bounds on the minimum achievable variance $L_{\mathcal{M}_{\text{SPCM}}}$ via (6.17). We choose \mathcal{D} such that Theorem 6.2.3 applies and therefore the RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})$, associated with the kernel $R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \cdot)$ given in (6.38), exists. We will assume that the parameter function $g(\cdot)$ is estimable for $\mathcal{M}_{\text{SPCM}}$, which implies via Theorem 2.5.2 that the restricted parameter function $g(\cdot)|_{\mathcal{D}}$ is estimable for $\mathcal{M}_{\mathcal{D},\text{SPCM}}$. This assumption is justified by the fact that the bounds that we

will derive are always finite, i.e., they trivially apply also in the case when $g(\cdot)|_{\mathcal{D}}$ is not estimable for $\mathcal{M}_{\mathcal{D},\text{SPCM}}$, for which $L_{\mathcal{M}_{\text{SPCM}}} = \infty$. These bounds will be obtained by projecting the restricted parameter function $g|_{\mathcal{D}}(\cdot) : \mathcal{D} \rightarrow \mathbb{R}$, which is equal to the prescribed mean function $\gamma(\mathbf{x}) = c(\mathbf{x}) + g(\mathbf{x})$ of $\mathcal{M}_{\mathcal{D},\text{SPCM}}$ since $c(\mathbf{x}) = 0$, on suitable subspaces \mathcal{U} of the RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})$. Our choice of \mathcal{U} will be inspired by the derivation of the lower bounds for the SLM in Chapter 5. Hence, there will be a natural correspondence between the bounds for the SPCM and the bounds for the SLM presented in Chapter 5.

For the derivation of the bounds, we will need an identity which is stated in

Lemma 6.3.1. *Consider the function $f(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R} : f(\mathbf{x}) = \det\{\mathbf{H}(\mathbf{x})\}$, where the matrix-valued function $\mathbf{H}(\cdot) : \mathbb{R}^N \rightarrow \mathbb{R}^{M \times M}$ is such that there exist the partial derivatives $\frac{\partial^{\mathbf{p}}(\mathbf{H}(\mathbf{x}))_{m,n}}{\partial \mathbf{x}^{\mathbf{p}}}|_{\mathbf{x}=\mathbf{x}_0}$ for any order $\mathbf{p} \in \mathbb{Z}_+^N$ with $\|\mathbf{p}\|_{\infty} \leq 1$ as well as the inverse $\mathbf{H}^{-1}(\mathbf{x}_0)$. Then we have*

$$\left. \frac{\partial f(\mathbf{x})}{\partial x_k} \right|_{\mathbf{x}=\mathbf{x}_0} = \det\{\mathbf{H}(\mathbf{x}_0)\} \text{Tr} \{ \mathbf{H}^{-1}(\mathbf{x}_0) (\partial \mathbf{H}(\mathbf{x}_0))^T \}, \quad (6.42)$$

where $\partial_k \mathbf{H}(\mathbf{x}_0)$ is the matrix whose entry in the m th row and n th column is given by $\left. \frac{\partial (\mathbf{H}(\mathbf{x}))_{m,n}}{\partial x_k} \right|_{\mathbf{x}=\mathbf{x}_0}$.

Proof. [20, p. 73] □

The first lower bound on $L_{\mathcal{M}_{\text{SPCM}}}$ will be the ‘‘SPCM analogue’’ of the SLM bound given by Theorem 5.4.1.

Theorem 6.3.2. *Consider the minimum variance problem $\mathcal{M}_{\text{SPCM}} = (\mathcal{E}_{\text{SPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ associated with the SPCM with arbitrary psd basis matrices $\{\mathbf{C}_k \in \mathbb{R}^{M \times M}\}_{k \in [N]}$. If the parameter function $g(\cdot) : \mathcal{X}_{S,+} \rightarrow \mathbb{R}$ is such that the partial derivatives $\left. \frac{\partial g(\mathbf{x})}{\partial x_l} \right|_{\mathbf{x}=\mathbf{x}_0}$ exist for $l \in [N]$, we have*

$$L_{\mathcal{M}_{\text{SPCM}}} \geq \mathbf{b}^T \mathbf{J}^\dagger \mathbf{b} \quad \text{when } \|\mathbf{x}_0\|_0 < S \quad (6.43)$$

$$L_{\mathcal{M}_{\text{SPCM}}} \geq \mathbf{b}_{\mathbf{x}_0}^T \mathbf{J}_{\mathbf{x}_0}^\dagger \mathbf{b}_{\mathbf{x}_0} \quad \text{when } \|\mathbf{x}_0\|_0 = S. \quad (6.44)$$

Here, $\mathbf{b} \in \mathbb{R}^N$ is defined elementwise by $b_l \triangleq \left. \frac{\partial g(\mathbf{x})}{\partial x_l} \right|_{\mathbf{x}=\mathbf{x}_0}$, and $\mathbf{b}_{\mathbf{x}_0} \in \mathbb{R}^S$ denotes the restriction to the entries of \mathbf{b} indexed by $\text{supp}(\mathbf{x}_0) = (i_1, \dots, i_S)$, i.e., $(\mathbf{b}_{\mathbf{x}_0})_j = b_{i_j}$. Furthermore, the matrix $\mathbf{J} \in \mathbb{R}^{N \times N}$ is given elementwise by

$$(\mathbf{J})_{m,n} = \frac{1}{2} \text{Tr} \{ \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) \mathbf{C}_m \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) \mathbf{C}_n \}, \quad (6.45)$$

and the matrix $\mathbf{J}_{\mathbf{x}_0} \in \mathbb{R}^{S \times S}$ is given elementwise by

$$(\mathbf{J}_{\mathbf{x}_0})_{m,n} = \frac{1}{2} \text{Tr} \{ \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) \mathbf{C}_{i_m} \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) \mathbf{C}_{i_n} \}. \quad (6.46)$$

Proof. Consider the minimum variance problem $\mathcal{M}_{\mathcal{D},\text{SPCM}} = \mathcal{M}_{\text{SPCM}}|_{\mathcal{D}}$, where the set \mathcal{D} is chosen as in (6.25) with $\mathbf{x}_c = \mathbf{x}_0$, so that Theorem 6.2.3 applies, i.e., the RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})$ exists. We can assume that $g(\cdot)$ is estimable for $\mathcal{M}_{\text{SDPCM}}$, which implies via Theorem 2.5.2 that the restriction $g(\cdot)|_{\mathcal{D}}$ is estimable for $\mathcal{M}_{\mathcal{D},\text{SDPCM}}$. Thus, according to Theorem 4.3.4, the prescribed mean function $\gamma(\cdot) : \mathcal{D} \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = c(\mathbf{x}) + g(\mathbf{x}) = g(\mathbf{x})$ belongs to the RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})$.

For the case $\|\mathbf{x}_0\|_0 < S$, in order to prove (6.43), consider the subspace $\mathcal{U}_1 \triangleq \text{span}\{w_0(\cdot) \cup \{w_l(\cdot)\}_{l \in [N]}\}$ spanned by the functions $w_0(\cdot) \triangleq R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \mathbf{x}_0)$ and

$$w_l(\cdot) \triangleq \left. \frac{\partial^{e_l} R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \mathbf{x}_2)}{\partial \mathbf{x}_2^{e_l}} \right|_{\mathbf{x}_2 = \mathbf{x}_0} \quad (6.47)$$

for $l \in [N]$. We have trivially $w_0(\cdot) \in \mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})$, and by Theorem 3.4.1 under the condition $\|\mathbf{x}_0\|_0 < S$, we have also that $w_l(\cdot) \in \mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})$ for $l \in [N]$. In a completely analogous manner as the RKHS-based derivation of Theorem 4.4.4 (see Section 4.4.2), one can show that $\langle w_0(\cdot), w_l(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})} = 0$ (cf. (4.60)), $\langle w_0(\cdot), \gamma(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})} = \gamma(\mathbf{x}_0)$, and $\langle w_l(\cdot), \gamma(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})} = b_l$ (cf. (4.63)) for $l \in [N]$ (note that $\gamma(\cdot) = g(\cdot)$). For $l, l' \in [N]$, we obtain by (6.38), Lemma 6.3.1, and the obvious fact $\frac{\partial^{e_l}}{\partial \mathbf{x}^{e_l}} \tilde{\mathbf{C}}(\mathbf{x}) = \mathbf{C}_l$ that

$$\begin{aligned} \langle w_l(\cdot), w_{l'}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})} &= \left\langle \left. \frac{\partial^{e_l} R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \mathbf{x}_2)}{\partial \mathbf{x}_2^{e_l}} \right|_{\mathbf{x}_2 = \mathbf{x}_0}, \left. \frac{\partial^{e_{l'}} R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \mathbf{x}_2)}{\partial \mathbf{x}_2^{e_{l'}}} \right|_{\mathbf{x}_2 = \mathbf{x}_0} \right\rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})} \\ &\stackrel{(3.66)}{=} \left. \frac{\partial^{e_l} \partial^{e_{l'}} R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{x}_1^{e_l} \partial \mathbf{x}_2^{e_{l'}}} \right|_{\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}_0} \\ &\stackrel{(6.38)}{=} \left. \frac{\partial^{e_l} \partial^{e_{l'}} [\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} [\det \{\tilde{\mathbf{C}}(\mathbf{x}_1) + \tilde{\mathbf{C}}(\mathbf{x}_2) - \tilde{\mathbf{C}}(\mathbf{x}_1)\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\tilde{\mathbf{C}}(\mathbf{x}_2)\}]^{-1/2}}{\partial \mathbf{x}_1^{e_l} \partial \mathbf{x}_2^{e_{l'}}} \right|_{\mathbf{x}_1 = \mathbf{x}_2 = \mathbf{x}_0} \\ &\stackrel{(a)}{=} -\frac{1}{2} [\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} \times \\ &\quad \left. \frac{\partial^{e_l} \text{Tr}\{\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)(\mathbf{I} - \tilde{\mathbf{C}}(\mathbf{x}_1)\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0))\mathbf{C}_{l'}\} [\det \{\tilde{\mathbf{C}}(\mathbf{x}_1) + \tilde{\mathbf{C}}(\mathbf{x}_0) - \tilde{\mathbf{C}}(\mathbf{x}_1)\}]^{-1/2}}{\partial \mathbf{x}_1^{e_l}} \right|_{\mathbf{x}_1 = \mathbf{x}_0} \\ &= -\frac{1}{2} \left. \frac{\partial^{e_l} \text{Tr}\{\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)(\mathbf{I} - \tilde{\mathbf{C}}(\mathbf{x}_1)\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0))\mathbf{C}_{l'}\}}{\partial \mathbf{x}_1^{e_l}} \right|_{\mathbf{x}_1 = \mathbf{x}_0} \\ &= \frac{1}{2} \text{Tr}\{\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\mathbf{C}_l\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\mathbf{C}_{l'}\} \stackrel{(6.45)}{=} (\mathbf{J})_{l,l'}. \end{aligned} \quad (6.48)$$

Here, the step (a) is due to Lemma 6.3.1 using $\mathbf{H}(\mathbf{x}_2) = \tilde{\mathbf{C}}(\mathbf{x}_1) + \tilde{\mathbf{C}}(\mathbf{x}_2) - \tilde{\mathbf{C}}(\mathbf{x}_1)\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\tilde{\mathbf{C}}(\mathbf{x}_2)$ and the chain rule for differentiation [5]. The bound (6.43) is then obtained by projecting $\gamma(\cdot)$ on

the subspace \mathcal{U}_1 as derived in the following. Using Theorem 4.3.4, and Theorem 3.1.6,

$$L_{\mathcal{M}_{\text{SPCM}}} \stackrel{(4.19)}{=} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})}^2 - \underbrace{[\gamma(\mathbf{x}_0)]^2}_{=g(\mathbf{x}_0)} \stackrel{(3.31)}{\geq} \|\mathbf{P}_{\mathcal{U}_1}\gamma(\cdot)\|^2 - (g(\mathbf{x}_0))^2. \quad (6.49)$$

According to Theorem 3.1.9, this becomes further

$$\begin{aligned} L_{\mathcal{M}_{\text{SPCM}}} &\stackrel{(3.39)}{\geq} \frac{\langle w_0(\cdot), \gamma(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})}^2}{\langle w_0(\cdot), w_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})}} + \mathbf{b}^T \mathbf{J}^\dagger \mathbf{b} - [g(\mathbf{x}_0)]^2 \\ &= \frac{\langle R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \mathbf{x}_0), \gamma(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})}^2}{\|R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \mathbf{x}_0)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})}^2} + \mathbf{b}^T \mathbf{J}^\dagger \mathbf{b} - [g(\mathbf{x}_0)]^2 \\ &\stackrel{(a)}{=} \frac{[g(\mathbf{x}_0)]^2}{1} + \mathbf{b}^T \mathbf{J}^\dagger \mathbf{b} - [g(\mathbf{x}_0)]^2 \\ &= \mathbf{b}^T \mathbf{J}^\dagger \mathbf{b}, \end{aligned} \quad (6.50)$$

which is (6.43). In (6.50), the step (a) follows from the facts that

$$\langle R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \mathbf{x}_0), \gamma(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})} \stackrel{(3.44)}{=} \gamma(\mathbf{x}_0) = g(\mathbf{x}_0) \quad (6.51)$$

and

$$\begin{aligned} \|R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \mathbf{x}_0)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})}^2 &= \langle R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \mathbf{x}_0), R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})} \\ &\stackrel{(3.44)}{=} R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\mathbf{x}_0, \mathbf{x}_0) \stackrel{(4.11)}{=} 1. \end{aligned} \quad (6.52)$$

In an almost identical manner, one can prove the bound (6.44) for the case $\|\mathbf{x}_0\|_0 = S$. The only difference is that instead of using the subspace \mathcal{U}_1 (which does not exist in this case), one has to use the subspace $\mathcal{U}_2 \triangleq \text{span}\{w_0(\cdot) \cup \{w_l(\cdot)\}_{l \in \text{supp}(\mathbf{x}_0)}\}$, with $w_l(\cdot)$ as defined in (6.47). \square

The bounds (6.43), (6.44) have a particular interpretation as the CRB of a non-sparse estimation problem which is identical to the SPCM except for the sparsity constraint. More precisely, the bounds (6.43), (6.44) are the CRB for a general non-sparse estimation problem with a Gaussian observation $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}(\mathbf{x}))$. Here, the covariance matrix $\mathbf{C}(\mathbf{x})$ depends arbitrarily on the parameter vector \mathbf{x} , which is not required to be sparse [20, p. 47].

A specific similarity of Theorem 6.3.2 to Theorem 5.4.1 becomes apparent when the corresponding lower bounds are evaluated for the special case of the SPCM with $g(\mathbf{x}) = x_k$ where $k \notin \text{supp}(\mathbf{x}_0)$ and the basis matrices $\mathbf{C}_l = \mathbf{e}_l \mathbf{e}_l^T$. This choice yields an instance of the SDPCM

(that will be discussed presently). For this special case, we observe a discontinuity between the two domains of \mathcal{X} corresponding to $\|\mathbf{x}\|_0 = S$ and $\|\mathbf{x}\|_0 < S$. In fact, while the bound in (6.43) for $\|\mathbf{x}_0\|_0 < S$ is then equal to $2\sigma^4$, the bound (6.44) for $\|\mathbf{x}_0\|_0 = S$ is equal to 0.

In what follows, we will need a standard result in matrix calculus:

Lemma 6.3.3. *Consider a matrix-valued function $\mathbf{H}(\cdot) : \mathbb{R} \rightarrow \mathbb{R}^{M \times M}$ such that the derivatives of any component function $f_{m,n}(x) \triangleq (\mathbf{H}(x))_{m,n}$ exist to a sufficient order. We denote by $\partial\mathbf{H}(x) \in \mathbb{R}^{M \times M}$ the specific matrix whose entries are given by $\frac{\partial f_{m,n}(x)}{\partial x}$. We then have the identity*

$$\partial\mathbf{H}^{-1}(x) = -\mathbf{H}^{-1}(x)\partial\mathbf{H}(x)\mathbf{H}^{-1}(x). \quad (6.53)$$

Proof. [38] □

Whereas Theorem 6.3.2 places no assumption on the basis matrices $\{\mathbf{C}_k\}_{k \in [N]}$ of the SPCM, the next result requires the basis matrices to satisfy the RIP of order $S+1$ with a sufficiently small constant δ_{S+1} :

Theorem 6.3.4. *Consider the minimum variance problem $\mathcal{M}_{\text{SPCM}} = (\mathcal{E}_{\text{SPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ associated with the SPCM with sparsity degree S and basis matrices $\{\mathbf{C}_k\}_{k \in [N]}$, with $\text{rank } r_k \triangleq \text{rank}(\mathbf{C}_k)$, that satisfy the RIP of order $S+1$ with RIP constant $\delta_{S+1} < 1/32$. If the parameter function $g(\cdot) : \mathcal{X}_{S,+} \rightarrow \mathbb{R}$ is such that the partial derivatives $b_l \triangleq \left. \frac{\partial g(\mathbf{x})}{\partial x_l} \right|_{\mathbf{x}=\mathbf{x}_0}$ exist for $l \in [N] \setminus \text{supp}(\mathbf{x}_0)$, we have*

$$L_{\mathcal{M}_{\text{SPCM}}} \geq \frac{2\sigma^4 b_l^2}{r_l} \frac{(\sigma^2 \beta)^{Q/2} (\sigma^2 (1 + 5\delta_{S+1}))^{-r_l/2}}{\prod_{k \in \text{supp}(\mathbf{x}_0)} (x_{0,k} + \sigma^2 (1 + 5\delta_{S+1}))^{r_k/2}} \beta^2 \left[r_l \frac{(12\delta_{S+1})^2}{2} + (12\delta_{S+1})^2 + \beta(6\delta_{S+1} + 1) \right]^{-1} \quad (6.54)$$

for any $l \in [N] \setminus \text{supp}(\mathbf{x}_0)$. Here,

$$\beta \triangleq \frac{1}{(1 + \delta_{S+1})^2} \left(2 - \frac{(1 + \delta_{S+1})^4}{(1 - \delta_{S+1})^4} \right), \quad (6.55)$$

and $Q \triangleq \sum_{k \in \text{supp}(\mathbf{x}_0)} r_k + r_l$.

Proof. Appendix B. □

Note that Theorem 6.3.4 yields a lower bound on the minimum achievable variance $L_{\mathcal{M}_{\text{SPCM}}}$ that is continuous with respect to the parameter vector \mathbf{x}_0 . This is in contrast to the bound given by Theorem 6.3.2 which yields a discontinuous bound in general.

However, as can be verified easily, for the special case of the SPCM given by the SDPCM with parameter function $g(\mathbf{x}) = x_k$ where $k \notin \text{supp}(\mathbf{x}_0)$, the lower bound (6.54) of Theorem 6.3.4 is in general looser, i.e., lower, than the bound (6.43), (6.44) of Theorem 6.3.2.

In the following, we specialize our RKHS approach for the general SPCM to the SDPCM and derive a lower bound that is both continuous with respect to \mathbf{x}_0 and, for the parameter function $g(\mathbf{x}) = x_k$ with $k \notin \text{supp}(\mathbf{x}_0)$, at least as tight as the bound given by (6.43) and (6.44).

6.4 The Sparse Diagonalizable Parametric Covariance Model

We now specialize and extend the results found for the general SPCM to the special case termed the *sparse diagonalizable parametric covariance model* (SDPCM). The SDPCM is obtained from the SPCM by restricting to basis matrices that are orthogonal projection matrices on orthogonal subspaces and, thus, satisfy (6.12), i.e., $\mathbf{C}_k \mathbf{C}_{k'} = \delta_{k,k'} \mathbf{C}_k$ for all $k, k' \in [N]$. This special case, which can be seen as the SPCM analogue of the SSNM,² will allow for a slightly more precise analysis via RKHS theory than was possible for the general SPCM. In particular, we will derive lower bounds that are tighter than the bound given by Theorem 6.3.2 when specialized to the SDPCM. We will use the generic notation $\mathcal{M}_{\text{SDPCM}} = (\mathcal{E}_{\text{SDPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ for a minimum variance problem that is obtained from the SDPCM $\mathcal{E}_{\text{SDPCM}} = (\mathcal{X}_{S,+}, f_{\text{SDPCM}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}))$, the identically vanishing prescribed bias (i.e., we consider unbiased estimation) and a parameter vector $\mathbf{x}_0 \in \mathcal{X}_{S,+}$.

Since the basis matrices $\mathbf{C}_k \in \mathbb{R}^{M \times M}$ are assumed psd, the condition (6.12) is equivalent to requiring the basis matrices to be orthogonal projection matrices on orthogonal subspaces of \mathbb{R}^M . The condition (6.12) evaluated for $k' = k$, i.e., $\mathbf{C}_k^2 = \mathbf{C}_k$, is the defining requirement for the psd matrix \mathbf{C}_k to be an orthogonal projection matrix (cf. [38]), i.e., it can be written as

$$\mathbf{C}_k = \sum_{i \in [r_k]} \mathbf{u}_{m_{k,i}} \mathbf{u}_{m_{k,i}}^T. \quad (6.56)$$

Here, $\{\mathbf{u}_m \in \mathbb{R}^M\}_{m \in [R]}$ is an orthonormal set of vectors, i.e., $\mathbf{u}_m^T \mathbf{u}_{m'} = \delta_{m,m'}$, and the sets $\mathcal{U}_k \triangleq \{\mathbf{u}_{m_{k,i}}\}_{i \in [r_k]}$ are disjoint, so that they span orthogonal subspaces of \mathbb{R}^M . Indeed, assume that for two different basis matrices $\mathbf{C}_k, \mathbf{C}_{k'}$ ($k \neq k'$), there would be at least two vectors $\mathbf{a} \in \mathcal{U}_k, \mathbf{b} \in \mathcal{U}_{k'}$ which are nonorthogonal, i.e., $\mathbf{a}^T \mathbf{b} \neq 0$. Then, using (6.56), we would obtain $\mathbf{C}_k \mathbf{C}_{k'} \mathbf{b} = \mathbf{C}_k \mathbf{b} \neq \mathbf{0}$,

²Indeed, by a suitable transformation of the observation vector \mathbf{y} of the SDPCM, we have that within the SDPCM and the SSNM, each entry x_k of the parameter vector determines the statistics of an individual subset of the entries of the observation vector. The difference between the SDPCM and the SSNM is then how the parameter vector entry x_k determines the statistics of the corresponding observation vector entries. In the SSNM x_k determines the mean, while in the SDPCM x_k determines the variance. Another difference is that for the the SSNM, the subset of observation vector entries that correspond to x_k consist only of the single entry y_k , while for the SDPCM these subsets may contain more than one entry of the observation vector.

which is a contradiction to (6.12). From the representation (6.56), it also follows that for the SDPCM, there must be $M \geq R = \sum_{k \in [N]} r_k$ (in particular, $M \geq N$).

Using the representation (6.56) of the basis matrices, we can interpret the observation model of the SDPCM in (6.1) as a latent variable model

$$\mathbf{y} = \sum_{k \in [N]} \mathbf{s}_k + \mathbf{n}, \quad (6.57)$$

with $\mathbf{s}_k = \sum_{i \in [r_k]} \xi_{m_{k,i}} \mathbf{u}_{m_{k,i}}$, where the $\xi_{m_{k,i}}$ are independent zero-mean Gaussian random variables with variance x_k for all i , i.e., $\xi_{m_{k,i}} \sim \mathcal{N}(0, x_k)$. This is similar to the latent variable model used in probabilistic principal component analysis [102] except that our “factors” \mathbf{u}_m are fixed.

Another useful implication of (6.56) is based on a fundamental result in linear algebra [4] which states that we can add vectors $\{\mathbf{v}_l \in \mathbb{R}^M\}_{l \in [L]}$, with $L = M - R$, to the set $\{\mathbf{u}_m \in \mathbb{R}^M\}_{m \in [R]}$ such that their union $\mathcal{U} \triangleq \{\mathbf{v}_l\}_{l \in [L]} \cup \{\mathbf{u}_m\}_{m \in [R]}$ forms an ONB for \mathbb{R}^M . This implies in particular that we can represent the covariance matrix of the observation $\tilde{\mathbf{C}}(\mathbf{x}) = \mathbf{C}(\mathbf{x}) + \sigma^2 \mathbf{I}$ (cf. (6.4)), as

$$\tilde{\mathbf{C}}(\mathbf{x}) = \mathbf{U}[\mathbf{D}'(\mathbf{x}) + \sigma^2 \mathbf{I}]\mathbf{U}^T, \quad (6.58)$$

where the columns of the orthonormal matrix $\mathbf{U} \in \mathbb{R}^{M \times M}$ are the vectors belonging to \mathcal{U} . The diagonal matrix $\mathbf{D}'(\mathbf{x}) \in \mathbb{R}^{M \times M}$ has a main diagonal consisting of two parts: the first part of length R contains the entries x_k where each entry is repeated r_k times, while the second part consists of $M - R$ zeros, i.e.,

$$\mathbf{D}'(\mathbf{x}) = \begin{pmatrix} x_1 \mathbf{I}_{r_1} & \mathbf{0} & \dots & \dots & \dots & \dots & \mathbf{0} \\ \mathbf{0} & x_2 \mathbf{I}_{r_2} & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & x_N \mathbf{I}_{r_N} & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \mathbf{0} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \dots & \dots & \dots & \mathbf{0} \end{pmatrix}. \quad (6.59)$$

6.4.1 The \mathcal{D} -restricted SDPCM

As for the SPCM, we cannot associate a RKHS to the minimum variance problem $\mathcal{M}_{\text{SDPCM}}$ directly. Instead, we will work with the RKHS associated to the \mathcal{D} -restricted minimum variance problem $\mathcal{M}_{\mathcal{D}, \text{SDPCM}} = \mathcal{M}_{\text{SDPCM}}|_{\mathcal{D}}$, with a set $\mathcal{D} \subseteq \mathcal{X}_{S,+}$ that satisfies (6.19).

For the SDPCM, the condition (6.19) can be rewritten due to (6.58) as

$$2(x_i + \sigma^2)^{-1} - (x_{0,i} + \sigma^2)^{-1} > 0 \quad \forall \mathbf{x} \in \mathcal{D}, \forall i \in [N]. \quad (6.60)$$

This is because a matrix $\mathbf{A} \in \mathbb{R}^{M \times M}$ is positive definite, i.e., $\mathbf{A} > \mathbf{0}$, if and only if $\mathbf{VAV}^T > \mathbf{0}$ for any orthonormal matrix $\mathbf{V} \in \mathbb{R}^{M \times M}$, i.e., satisfying $\mathbf{V}^T \mathbf{V} = \mathbf{I}$. The largest set $\mathcal{D} \subseteq \mathcal{X}_{S,+}$ which satisfies (6.60) is obviously given by

$$\mathcal{D}_0 \triangleq \{\mathbf{x} \in \mathcal{X}_{S,+} \mid x_i < 2x_{0,i} + \sigma^2\}. \quad (6.61)$$

Note that the set \mathcal{D}_0 depends on \mathbf{x}_0 .

Since the set \mathcal{D}_0 is the largest set satisfying (6.60), the corresponding minimum achievable variance $L_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}$ will yield via (6.17) the tightest lower bound on $L_{\mathcal{M}_{\text{SDPCM}}}$ among all bounds that are obtained from $L_{\mathcal{M}_{\mathcal{D}, \text{SDPCM}}}$ with a set \mathcal{D} that satisfies (6.60).

6.4.2 Variance Bounds for the SDPCM

We will now derive a lower bound on the minimum achievable variance $L_{\mathcal{M}_{\text{SDPCM}}}$ by lower bounding $L_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}$, i.e., the minimum achievable variance of $\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}$, using the RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})$ which is associated to the kernel $R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\cdot, \cdot) : \mathcal{D}_0 \times \mathcal{D}_0 \rightarrow \mathbb{R}$:

$$R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\mathbf{x}_1, \mathbf{x}_2) \triangleq \mathbb{E}_{\mathbf{x}_0} \left\{ \rho_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\mathbf{y}, \mathbf{x}_1) \rho_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\mathbf{y}, \mathbf{x}_2) \right\}. \quad (6.62)$$

Using (6.38) and (6.58) and the identity $\det\{\mathbf{AB}\} = \det\{\mathbf{A}\} \det\{\mathbf{B}\}$ valid for any two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{M \times M}$ [38, p. 51], we obtain

$$\begin{aligned} R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\mathbf{x}_1, \mathbf{x}_2) &= [\det\{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} [\det\{\tilde{\mathbf{C}}(\mathbf{x}_1) + \tilde{\mathbf{C}}(\mathbf{x}_2) - \tilde{\mathbf{C}}(\mathbf{x}_1)\tilde{\mathbf{C}}^{-1}(\mathbf{x}_0)\tilde{\mathbf{C}}(\mathbf{x}_2)\}]^{-1/2} \\ &= \frac{\prod_{k \in [N]} (x_{0,k} + \sigma^2)^{r_k}}{\prod_{k \in [N]} [(x_{0,k} + \sigma^2)^2 - (x_{1,k} - x_{0,k})(x_{2,k} - x_{0,k})]^{r_k/2}}. \end{aligned} \quad (6.63)$$

Note that the kernel $R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\cdot, \cdot)$ given by (6.63) is obviously differentiable up to any order m .

The following result partly characterizes the RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})$.

Theorem 6.4.1. *The RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})$ is differentiable up to any order. Let $\mathcal{K} \subseteq [N]$ be an arbitrary set of S different indices, i.e., $|\mathcal{K}| = S$, and let $\mathbf{p} \in \mathbb{Z}_+^N$ be an arbitrary multi-index with $\text{supp}(\mathbf{p}) \subseteq \mathcal{K}$. Then the function $g^{(\mathbf{p}, \mathcal{K})}(\cdot) : \mathcal{D}_0 \rightarrow \mathbb{R}$ defined as*

$$\begin{aligned} g^{(\mathbf{p}, \mathcal{K})}(\mathbf{x}) &\triangleq \left. \frac{\partial^{\mathbf{p}} R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\mathbf{x}, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}}} \right|_{\mathbf{x}_2 = \mathbf{x}_0^{\mathcal{K}}} \\ &= \prod_{k \in \text{supp}(\mathbf{p})} \frac{c_k (x_k - x_{0,k})^{p_k}}{(x_{0,k} + \sigma^2)^{2p_k}} \prod_{k \in \text{supp}(\mathbf{x}_0) \setminus \mathcal{K}} \frac{(x_{0,k} + \sigma^2)^{r_k}}{[(x_{0,k} + \sigma^2)^2 + (x_k - x_{0,k})x_{0,k}]^{r_k/2}}, \end{aligned} \quad (6.64)$$

where $c_k \triangleq \prod_{l \in [p_k]} (r_k/2 + (l-1))$, is an element of $\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})$, i.e.,

$$g^{(\mathbf{p}, \mathcal{K})}(\cdot) \in \mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}). \quad (6.65)$$

The inner product of an arbitrary function $f(\cdot) \in \mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})$ with $g^{(\mathbf{p}, \mathcal{K})}(\cdot)$ is given by

$$\langle f(\cdot), g^{(\mathbf{p}, \mathcal{K})}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})} = \left. \frac{\partial^{\mathbf{p}} f(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}} \right|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}}. \quad (6.66)$$

Proof. The statement follows straightforwardly from Theorem 3.4.1, since the kernel $R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\cdot, \cdot)$ is differentiable up to any order. \square

Based on Theorem 6.4.1, we will now derive a lower bound on the minimum achievable variance $L_{\mathcal{M}_{\text{SDPCM}}}$:

Theorem 6.4.2. Consider the minimum variance problem $\mathcal{M}_{\text{SDPCM}} = (\mathcal{E}_{\text{SDPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ with sparsity degree S . Let $\{\mathbf{p}_l \in \mathbb{Z}_+^N\}_{l \in [L]}$ be a set of multi-indices such that $\text{supp}(\mathbf{p}_l) \subseteq \mathcal{K}$, where $\mathcal{K} \subseteq [N]$ is an arbitrary set of S different indices, i.e., $|\mathcal{K}| = S$. If the parameter function $g(\cdot) : \mathcal{X}_{S,+} \rightarrow \mathbb{R}$ associated with $\mathcal{E}_{\text{SDPCM}}$ is such that the partial derivatives $\left. \frac{\partial^{\mathbf{p}_l} g(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}_l}} \right|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}}$ exist for all $l \in [L]$, we have

$$L_{\mathcal{M}_{\text{SDPCM}}} \geq \sum_{l \in [L]} \frac{1}{q_l(\mathbf{x}_0)} \left[\left. \frac{\partial^{\mathbf{p}_l} g(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}_l}} \right|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}} \right]^2 - [g(\mathbf{x}_0)]^2, \quad (6.67)$$

where $q_l(\mathbf{x}_0) \triangleq \left. \frac{\partial^{\mathbf{p}_l} \partial^{\mathbf{p}_l} R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{x}_1^{\mathbf{p}_l} \partial \mathbf{x}_2^{\mathbf{p}_l}} \right|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}_0^{\mathcal{K}}}$. Furthermore, there exists an unbiased estimator

$\hat{g}(\cdot)$ whose variance at \mathbf{x}_0 achieves this bound, i.e., $v(\hat{g}(\cdot); \mathbf{x}_0) = \sum_{l \in [L]} \frac{1}{q_l(\mathbf{x}_0)} \left[\left. \frac{\partial^{\mathbf{p}_l} g(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}_l}} \right|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}} \right]^2 - [g(\mathbf{x}_0)]^2$, if and only if it can be written as

$$\hat{g}(\cdot) = \sum_{l \in [L]} a_l \left. \frac{\partial^{\mathbf{p}_l} \rho_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\cdot, \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}_l}} \right|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}}, \quad (6.68)$$

with suitable (non-random) coefficients $a_l \in \mathbb{R}$.

Proof. Since the bound in (6.67) is always finite, we assume without loss of generality that $g(\cdot)$ is estimable for $\mathcal{M}_{\text{SDPCM}}$, which implies via Theorem 2.5.2 that the restriction $g(\cdot)|_{\mathcal{D}_0}$ is estimable for $\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}$. Thus we can assume by Theorem 4.3.4 that the prescribed mean function $\gamma(\cdot) : \mathcal{D}_0 \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = c(\mathbf{x}) + g(\mathbf{x}) = g(\mathbf{x})$ belongs to the RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})$.

Let us consider the subspace $\mathcal{U} \triangleq \text{span} \{g^{(\mathbf{p}_l, \mathcal{K})}(\cdot)\}_{l \in [L]} \subseteq \mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})$ spanned by the functions $g^{(\mathbf{p}_l, \mathcal{K})}(\cdot) \in \mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})$ as defined in Theorem 6.4.1. For the inner products between $g^{(\mathbf{p}_l, \mathcal{K})}(\cdot)$ and $g^{(\mathbf{p}_{l'}, \mathcal{K})}(\cdot)$, we obtain

$$\langle g^{(\mathbf{p}_l, \mathcal{K})}(\cdot), g^{(\mathbf{p}_{l'}, \mathcal{K})}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})} \stackrel{(6.66)}{=} \left. \frac{\partial^{\mathbf{p}_{l'}} g^{(\mathbf{p}_l, \mathcal{K})}(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}_{l'}}} \right|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}}$$

$$\begin{aligned}
&\stackrel{(6.64)}{=} \frac{\partial^{\mathbf{p}l'}}{\partial \mathbf{x}^{\mathbf{p}l'}} \prod_{k \in \text{supp}(\mathbf{p}l)} \frac{c_k(x_k - x_{0,k})^{p_k}}{(x_{0,k} + \sigma^2)^{2p_k}} \prod_{k \in \text{supp}(\mathbf{x}_0) \setminus \mathcal{K}} \frac{(x_{0,k} + \sigma^2)^{r_k}}{[(x_{0,k} + \sigma^2)^2 + (x_k - x_{0,k})x_{0,k}]^{r_k/2}} \Big|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}} \\
&\stackrel{(a)}{=} \left[\prod_{k \in \text{supp}(\mathbf{x}_0) \setminus \mathcal{K}} \frac{(x_{0,k} + \sigma^2)^{r_k}}{[(x_{0,k} + \sigma^2)^2 - x_{0,k}^2]^{r_k/2}} \right] \frac{\partial^{\mathbf{p}l'}}{\partial \mathbf{x}^{\mathbf{p}l'}} \prod_{k \in \text{supp}(\mathbf{p}l)} \frac{c_k(x_k - x_{0,k})^{p_k}}{(x_{0,k} + \sigma^2)^{2p_k}} \Big|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}}, \quad (6.69)
\end{aligned}$$

where step (a) is due to the fact that $\text{supp}(\mathbf{p}l), \text{supp}(\mathbf{p}l') \subseteq \mathcal{K}$. Evaluating (6.69) for $l \neq l'$ reveals that the functions $\{g^{(\mathbf{p}l, \mathcal{K})}(\cdot)\}_{l \in [L]}$ are orthogonal, i.e.

$$\begin{aligned}
\langle g^{(\mathbf{p}l, \mathcal{K})}(\cdot), g^{(\mathbf{p}l', \mathcal{K})}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}} &= \delta_{l,l'} \langle g^{(\mathbf{p}l, \mathcal{K})}(\cdot), g^{(\mathbf{p}l, \mathcal{K})}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}} \\
&\stackrel{(6.66)}{=} \delta_{l,l'} \frac{\partial^{\mathbf{p}l'} g^{(\mathbf{p}l, \mathcal{K})}(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}l'}} \Big|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}} \\
&= \delta_{l,l'} \frac{\partial^{\mathbf{p}l} \partial^{\mathbf{p}l} R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{x}_1^{\mathbf{p}l} \partial \mathbf{x}_2^{\mathbf{p}l}} \Big|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}_0^{\mathcal{K}}} \\
&= \delta_{l,l'} q_l(\mathbf{x}_0). \quad (6.70)
\end{aligned}$$

Therefore, an ONB for the subspace $\mathcal{U} \subseteq \mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}$ is given by the set $\left\{ \frac{1}{\sqrt{q_l(\mathbf{x}_0)}} g^{(\mathbf{p}l, \mathcal{K})}(\cdot) \right\}_{l \in [L]}$.

Using the inner products (recall that $\gamma(\cdot) = g(\cdot)$)

$$\langle \gamma(\cdot), g^{(\mathbf{p}l, \mathcal{K})}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}} \stackrel{(6.66)}{=} \frac{\partial^{\mathbf{p}l} g(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}l}} \Big|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}}, \quad (6.71)$$

the bound in (6.67) follows by projecting $\gamma(\cdot)$ onto the subspace \mathcal{U} , since

$$\begin{aligned}
L_{\mathcal{M}_{\text{SDPCM}}} &\stackrel{(6.17)}{\geq} L_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}} \stackrel{(4.19)}{=} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}}^2 - \underbrace{[\gamma(\mathbf{x}_0)]^2}_{=g(\mathbf{x}_0)} \\
&\stackrel{(3.31)}{\geq} \|\mathbf{P}_{\mathcal{U}} \gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}}^2 - [g(\mathbf{x}_0)]^2 \\
&\stackrel{(3.34)}{=} \sum_{l \in [L]} \frac{\langle \gamma(\cdot), g^{(\mathbf{p}l, \mathcal{K})}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}}^2}{\langle g^{(\mathbf{p}l, \mathcal{K})}(\cdot), g^{(\mathbf{p}l, \mathcal{K})}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}}} - [g(\mathbf{x}_0)]^2 \\
&\stackrel{(6.70), (6.71)}{=} \sum_{l \in [L]} \frac{1}{q_l(\mathbf{x}_0)} \left[\frac{\partial^{\mathbf{p}l} g(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}l}} \Big|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}} \right]^2 - [g(\mathbf{x}_0)]^2. \quad (6.72)
\end{aligned}$$

Let us now show that if an unbiased estimator for $\mathcal{M}_{\text{SDPCM}}$ achieves the bound (6.67), i.e., its variance at \mathbf{x}_0 equals $\sum_{l \in [L]} \frac{1}{q_l(\mathbf{x}_0)} \left[\frac{\partial^{\mathbf{p}l} g(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}l}} \Big|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}} \right]^2 - [g(\mathbf{x}_0)]^2$ (which also implies that $g(\cdot)$ is

estimable for $\mathcal{M}_{\text{SDPCM}}$, it must be necessarily of the form (6.68). To that end, we note that according to the above derivation of (6.72), the bound (6.67) holds also for $L_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}$, i.e.,

$$L_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}} \geq \sum_{l \in [L]} \frac{1}{q_l(\mathbf{x}_0)} \left[\frac{\partial^{\mathbf{p}^l} g(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}^l}} \Big|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}} \right]^2 - [g(\mathbf{x}_0)]^2. \quad (6.73)$$

Therefore, if an unbiased estimator exists whose variance at \mathbf{x}_0 achieves the bound (6.67), it also achieves the bound (6.73), and thus is the unique LMV estimator (which is guaranteed to exist in this case) for $\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}$. The variance at \mathbf{x}_0 of this LMV estimator, which is of course also the minimum achievable variance $L_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}$, attains the bound (6.67) if and only if the projection $\mathbf{P}_{\mathcal{U}}\gamma(\cdot)$ coincides with $\gamma(\cdot)$, since this is necessary and sufficient for the second inequality in (6.72) to become an equality. This in turn, is the case if and only if $\gamma(\cdot) \in \mathcal{U}$, i.e., the prescribed mean function $\gamma(\cdot)$ can be written as the linear basis expansion $\sum_{l \in [L]} a_l g^{(\mathbf{p}^l, \mathcal{K})}(\cdot)$ with coefficients $a_l \in \mathbb{R}$. However, in this case we can express the unique LMV estimator $\hat{g}^{(\mathbf{x}_0)}(\cdot)$ for $\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}$ via Theorem 4.3.3 and Theorem 4.3.4 as

$$\hat{g}^{(\mathbf{x}_0)}(\cdot) \stackrel{(4.20)}{=} \mathbf{J}[\gamma(\cdot)] = \sum_{l \in [L]} a_l \mathbf{J} \left[g^{(\mathbf{p}^l, \mathcal{K})}(\cdot) \right] \stackrel{(6.64), (4.16)}{=} \sum_{l \in [L]} a_l \frac{\partial^{\mathbf{p}^l} \rho_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\cdot, \mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}^l}} \Big|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}}. \quad (6.74)$$

Thus, if an estimator achieves the bound (6.67), it is necessarily of the form (6.68).

Conversely, if there exists an estimator $\hat{g}(\cdot)$ of the form (6.68) we have by Theorem 4.3.2 and Theorem 4.3.3 that

$$\begin{aligned} \gamma(\mathbf{x}) &\stackrel{(a)}{=} \mathbf{E}_{\mathbf{x}} \{ \hat{g}(\mathbf{y}) \} = \mathbf{E}_{\mathbf{x}} \left\{ \sum_{l \in [L]} a_l \frac{\partial^{\mathbf{p}^l} \rho_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\mathbf{x}, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}^l}} \Big|_{\mathbf{x}_2=\mathbf{x}_0^{\mathcal{K}}} \right\} \\ &= \mathbf{E}_{\mathbf{x}} \left\{ \sum_{l \in [L]} a_l \frac{\partial^{\mathbf{p}^l} \rho_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\mathbf{x}, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}^l}} \Big|_{\mathbf{x}_2=\mathbf{x}_0^{\mathcal{K}}} \right\} \stackrel{(6.64), (4.16)}{=} \mathbf{E}_{\mathbf{x}} \left\{ \sum_{l \in [L]} a_l \mathbf{J}^{-1} \left[g^{(\mathbf{p}^l, \mathcal{K})}(\cdot) \right] \right\} \\ &\stackrel{(4.14)}{=} \sum_{l \in [L]} a_l g^{(\mathbf{p}^l, \mathcal{K})}(\mathbf{x}) \in \mathcal{U}, \end{aligned} \quad (6.75)$$

where (a) follows from the assumption that $\hat{g}(\cdot)$ is unbiased (note that we consider unbiased estimation since $c(\cdot) \equiv 0$). Thus, the existence of an estimator $\hat{g}(\cdot)$ of the form (6.68) implies that $\gamma(\cdot) \in \mathcal{U}$, which, as already shown above, is necessary and sufficient for the variance of the estimator in (6.68), i.e., $v(\hat{g}(\cdot), \mathbf{x}_0)$ to achieve the bound (6.67). \square

Note that by Theorem 2.3.1, unbiased estimation of a parameter function $g(\mathbf{x})$ is equivalent to biased estimation of x_k itself with bias $g(\mathbf{x}) - x_k$. Therefore, we have due to the definition of the minimum achievable variance the following implication of Theorem 6.4.2 for the variance of a biased estimator $\hat{x}_k(\mathbf{y})$ of x_k :

Corollary 6.4.3. Consider an estimator $\hat{x}_k(\mathbf{y})$ which uses the observation \mathbf{y} of a SDPCM with sparsity degree S and denote its mean function by $m_k(\cdot) : \mathcal{X}_{S,+} \rightarrow \mathbb{R} : m_k(\mathbf{x}) \triangleq \mathbf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\}$. Let $\{\mathbf{p}_l \in \mathbb{Z}_+^N\}_{l \in [L]}$ be a set of multi-indices such that $\text{supp}(\mathbf{p}_l) \subseteq \mathcal{K}$, where $\mathcal{K} \subseteq [N]$ is an arbitrary set of S different indices, i.e., $|\mathcal{K}| = S$. If for a fixed parameter vector $\mathbf{x}_0 \in \mathcal{X}_{S,+}$, the partial derivatives $\left. \frac{\partial^{\mathbf{p}_l} m_k(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}_l}} \right|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}}$ of the mean function exist, the variance $v(\hat{x}_k(\cdot); \mathbf{x}_0)$ at \mathbf{x}_0 is lower bounded by

$$v(\hat{x}_k(\cdot); \mathbf{x}_0) \geq \sum_{l \in [L]} \frac{1}{q_l(\mathbf{x}_0)} \left[\left. \frac{\partial^{\mathbf{p}_l} m_k(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}_l}} \right|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}}} \right]^2 - [m_k(\mathbf{x}_0)]^2 \quad (6.76)$$

Let us now present another lower bound on $L_{\mathcal{M}_{\text{SDPCM}}}$ which depends only on the first-order partial derivatives of the parameter function $g(\cdot)$ associated with the minimum variance problem $\mathcal{M}_{\text{SDPCM}} = (\mathcal{E}_{\text{SDPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$. This bound is derived by a slight modification of the proof of Theorem 6.4.2.

Theorem 6.4.4. Consider the minimum variance problem $\mathcal{M}_{\text{SDPCM}} = (\mathcal{E}_{\text{SDPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ associated with the SDPCM $\mathcal{E}_{\text{SDPCM}} = (\mathcal{X}_{S,+}, f_{\text{SDPCM}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}))$ with sparsity degree S and denote the value and index of the S -largest entry of $\mathbf{x}_0 \in \mathcal{X}_{S,+}$ by ξ_0 and j_0 , respectively. We furthermore define the sets $\mathcal{K}_l \triangleq \{l\} \cup \{\text{supp}(\mathbf{x}_0) \setminus \{j_0\}\}$ for $l \in [N] \setminus \text{supp}(\mathbf{x}_0)$ and $\mathcal{K}_l \triangleq \text{supp}(\mathbf{x}_0)$ for $l \in \text{supp}(\mathbf{x}_0)$. If the parameter function $g(\cdot) : \mathcal{X}_{S,+} \rightarrow \mathbb{R}$ associated with $\mathcal{E}_{\text{SDPCM}}$ is such that the partial derivatives $b_l \triangleq \left. \frac{\partial^{\mathbf{e}_l} g(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}_l}} \right|_{\mathbf{x}=\mathbf{x}_0^{\mathcal{K}_l}}$ exist, we have

$$L_{\mathcal{M}_{\text{SDPCM}}} \geq 2 \sum_{l \in \text{supp}(\mathbf{x}_0)} \frac{(x_{0,l} + \sigma^2)^2}{r_l} b_l^2 + 2\sigma^4 \left[1 - \frac{\xi_0^2}{(\xi_0 + \sigma^2)^2} \right]^{\frac{r_{j_0}}{2}} \sum_{l \in [N] \setminus \text{supp}(\mathbf{x}_0)} \frac{b_l^2}{r_l}. \quad (6.77)$$

Proof. As for the derivation of Theorem 6.4.2, we can assume that the prescribed mean function $\gamma(\cdot) : \mathcal{D}_0 \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = g(\mathbf{x})$ belongs to the RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})$. We then consider the subspace $\mathcal{U} \triangleq \text{span} \{ \{w_0(\cdot)\} \cup \{g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot)\}_{l \in [N]} \} \subseteq \mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})$ which is spanned by the functions $w_0(\cdot) \triangleq R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\cdot, \mathbf{x}_0)$ and $g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \in \mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})$ as defined in Theorem 6.4.1, i.e., $g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) = \left. \frac{\partial^{\mathbf{p}} R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\cdot, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{e}_l}} \right|_{\mathbf{x}_2=\mathbf{x}_0^{\mathcal{K}_l}}$.

The function $w_0(\cdot)$ is orthogonal to any function $g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot)$ for $l \in [N]$, since

$$\begin{aligned} \langle g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot), w_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})} &= \langle g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot), R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})} \\ &\stackrel{(3.44)}{=} \left. \frac{\partial^{\mathbf{e}_l} R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\mathbf{x}_0, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{e}_l}} \right|_{\mathbf{x}_2=\mathbf{x}_0^{\mathcal{K}_l}} \\ &\stackrel{(4.11)}{=} \left. \frac{\partial^{\mathbf{e}_l} 1}{\partial \mathbf{x}_2^{\mathbf{e}_l}} \right|_{\mathbf{x}_2=\mathbf{0}} = 0. \end{aligned} \quad (6.78)$$

The inner products between the functions $\{g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot)\}_{l \in [N]}$ are calculated according to (6.66) as

$$\begin{aligned} \langle g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot), g^{(\mathbf{e}_{l'}, \mathcal{K}_{l'})}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}} &\stackrel{(6.66)}{=} \frac{\partial^{\mathbf{e}_l}}{\partial \mathbf{x}_1^{\mathbf{e}_l}} g^{(\mathbf{e}_{l'}, \mathcal{K}_{l'})}(\mathbf{x}_1) \Big|_{\mathbf{x}_1 = \mathbf{x}_0^{\mathcal{K}_l}} \\ &\stackrel{(6.64)}{=} \frac{\partial^{\mathbf{e}_l}}{\partial \mathbf{x}_1^{\mathbf{e}_l}} \frac{r_{l'}(x_{1,l'} - x_{0,l'})}{2(x_{0,l'} + \sigma^2)^2} \prod_{k \in \text{supp}(\mathbf{x}_0) \setminus \mathcal{K}_{l'}} \frac{(x_{0,k} + \sigma^2)^{r_k}}{[(x_{0,k} + \sigma^2)^2 + (x_{1,k} - x_{0,k})x_{0,k}]^{r_k/2}} \Big|_{\mathbf{x}_1 = \mathbf{x}_0^{\mathcal{K}_l}}, \end{aligned} \quad (6.79)$$

which is most conveniently evaluated by considering separately the case where $l, l' \in \text{supp}(\mathbf{x}_0)$ and the complementary case where either l or l' (or both) belong to $[N] \setminus \text{supp}(\mathbf{x}_0)$. Let us begin with the latter case, where due to the symmetry of the inner product, we can assume without loss of generality that $l \in [N] \setminus \text{supp}(\mathbf{x}_0)$ and $l' \in [N]$ is arbitrary. In this case, (6.79) yields

$$\begin{aligned} &\langle g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot), g^{(\mathbf{e}_{l'}, \mathcal{K}_{l'})}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}} \\ &= \frac{\partial^{\mathbf{e}_l}}{\partial \mathbf{x}_1^{\mathbf{e}_l}} \frac{r_{l'}(x_{1,l'} - x_{0,l'})}{2(x_{0,l'} + \sigma^2)^2} \prod_{k \in \text{supp}(\mathbf{x}_0) \setminus \mathcal{K}_{l'}} \frac{(x_{0,k} + \sigma^2)^{r_k}}{[(x_{0,k} + \sigma^2)^2 + (x_{1,k} - x_{0,k})x_{0,k}]^{r_k/2}} \Big|_{\mathbf{x}_1 = \mathbf{x}_0^{\mathcal{K}_l}} \\ &\stackrel{l \notin \text{supp}(\mathbf{x}_0)}{=} \left[\prod_{k \in \text{supp}(\mathbf{x}_0) \setminus \mathcal{K}_{l'}} \frac{(x_{0,k} + \sigma^2)^{r_k}}{[(x_{0,k} + \sigma^2)^2 + (x_{1,k} - x_{0,k})x_{0,k}]^{r_k/2}} \Big|_{\mathbf{x}_1 = \mathbf{x}_0^{\mathcal{K}_l}} \right] \frac{\partial^{\mathbf{e}_l}}{\partial \mathbf{x}_1^{\mathbf{e}_l}} \frac{r_{l'}(x_{1,l'} - x_{0,l'})}{2(x_{0,l'} + \sigma^2)^2} \Big|_{\mathbf{x}_1 = \mathbf{x}_0^{\mathcal{K}_l}} \\ &= \delta_{l,l'} \langle g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot), g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})}. \end{aligned} \quad (6.80)$$

In the former case, where $l, l' \in \text{supp}(\mathbf{x}_0)$ and therefore $\mathcal{K}_l = \mathcal{K}_{l'} = \text{supp}(\mathbf{x}_0)$, we obtain by (6.79) that

$$\begin{aligned} &\langle g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot), g^{(\mathbf{e}_{l'}, \mathcal{K}_{l'})}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}} \\ &= \frac{\partial^{\mathbf{e}_l}}{\partial \mathbf{x}_1^{\mathbf{e}_l}} \frac{r_{l'}(x_{1,l'} - x_{0,l'})}{2(x_{0,l'} + \sigma^2)^2} \prod_{k \in \text{supp}(\mathbf{x}_0) \setminus \mathcal{K}_{l'}} \frac{(x_{0,k} + \sigma^2)^{r_k}}{[(x_{0,k} + \sigma^2)^2 + (x_{1,k} - x_{0,k})x_{0,k}]^{r_k/2}} \Big|_{\mathbf{x}_1 = \mathbf{x}_0^{\mathcal{K}_l}} \\ &\stackrel{\text{supp}(\mathbf{x}_0) \setminus \mathcal{K}_{l'} = \emptyset}{=} \frac{\partial^{\mathbf{e}_l}}{\partial \mathbf{x}_1^{\mathbf{e}_l}} \frac{r_{l'}(x_{1,l'} - x_{0,l'})}{2(x_{0,l'} + \sigma^2)^2} \Big|_{\mathbf{x}_1 = \mathbf{x}_0^{\mathcal{K}_l}} \\ &= \delta_{l,l'} \langle g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot), g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}})}. \end{aligned} \quad (6.81)$$

By combining (6.80) and (6.81), we obtain

$$\langle g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot), g^{(\mathbf{e}_{l'}, \mathcal{K}_{l'})}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}} = \delta_{l,l'} q_l(\mathbf{x}_0), \quad (6.82)$$

with

$$\begin{aligned}
q_l(\mathbf{x}_0) &\triangleq \langle g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot), g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}} \\
&= \frac{\partial^{\mathbf{e}_l}}{\partial \mathbf{x}_1^{\mathbf{e}_l}} \frac{r_l(x_{1,l} - x_{0,l})}{2(x_{0,l} + \sigma^2)^2} \prod_{k \in \text{supp}(\mathbf{x}_0) \setminus \mathcal{K}_l} \frac{(x_{0,k} + \sigma^2)^{r_k}}{[(x_{0,k} + \sigma^2)^2 + (x_{1,k} - x_{0,k})x_{0,k}]^{r_k/2}} \Big|_{\mathbf{x}_1 = \mathbf{x}_0^{\mathcal{K}_l}} \\
&= \left[\prod_{k \in \text{supp}(\mathbf{x}_0) \setminus \mathcal{K}_l} \frac{(x_{0,k} + \sigma^2)^{r_k}}{[(x_{0,k} + \sigma^2)^2 - x_{0,k}^2]^{r_k/2}} \right] \frac{\partial^{\mathbf{e}_l}}{\partial \mathbf{x}_1^{\mathbf{e}_l}} \frac{r_l(x_{1,l} - x_{0,l})}{2(x_{0,l} + \sigma^2)^2} \Big|_{\mathbf{x}_1 = \mathbf{x}_0^{\mathcal{K}_l}} \\
&= \begin{cases} \frac{r_l}{2(x_{0,l} + \sigma^2)^2}, & \text{if } l \in \text{supp}(\mathbf{x}_0) \\ \frac{r_l}{2\sigma^4} \left[\frac{(\xi_0 + \sigma^2)^2}{(\xi_0 + \sigma^2)^2 - \xi_0^2} \right]^{r_{j_0}/2} & \text{if } l \in [N] \setminus \text{supp}(\mathbf{x}_0). \end{cases} \tag{6.83}
\end{aligned}$$

Using the inner products (recall that $\gamma(\cdot) = g(\cdot)$)

$$\langle \gamma(\cdot), g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}} \stackrel{(6.66)}{=} \frac{\partial^{\mathbf{e}_l} g(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}_l}} \Big|_{\mathbf{x} = \mathbf{x}_0^{\mathcal{K}_l}} = b_l, \tag{6.84}$$

the bound (6.77) follows then by projecting $\gamma(\cdot)$ onto the subspace \mathcal{U} :

$$\begin{aligned}
L_{\mathcal{M}_{\text{SDPCM}}} &\stackrel{(6.17)}{\geq} L_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}} \stackrel{(4.19)}{=} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}}^2 - \underbrace{[\gamma(\mathbf{x}_0)]^2}_{=g(\mathbf{x}_0)} \\
&\stackrel{(3.31)}{\geq} \|\mathbf{P}\mathcal{U}\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}}^2 - [g(\mathbf{x}_0)]^2 \\
&\stackrel{(3.39)}{=} \frac{\langle \gamma(\cdot), w_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SPCM})}}^2}{\langle w_0(\cdot), w_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SPCM})}}} + \sum_{l \in [N]} \frac{\langle \gamma(\cdot), g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SPCM})}}^2}{\langle g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot), g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SPCM})}}} - [g(\mathbf{x}_0)]^2 \\
&= \frac{\langle \gamma(\cdot), R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\cdot, \mathbf{x}_0)(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SPCM})}}^2}{\langle R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\cdot, \mathbf{x}_0), R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SPCM})}}} \\
&\quad + \sum_{l \in [N]} \frac{\langle \gamma(\cdot), g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SPCM})}}^2}{\langle g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot), g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SPCM})}}} - [g(\mathbf{x}_0)]^2 \\
&\stackrel{(3.44), (6.83)}{=} \frac{[\gamma(\mathbf{x}_0)]^2}{R_{\mathcal{M}_{\mathcal{D}_0, \text{SDPCM}}}(\mathbf{x}_0, \mathbf{x}_0)} + \sum_{l \in [N]} \frac{\langle \gamma(\cdot), g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SPCM})}}^2}{q_l(\mathbf{x}_0)} - [g(\mathbf{x}_0)]^2
\end{aligned}$$

$$\begin{aligned}
& \stackrel{(4.11)}{=} \frac{[\gamma(\mathbf{x}_0)]^2}{1} + \sum_{l \in [N]} \frac{\langle \gamma(\cdot), g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}}^2}{q_l(\mathbf{x}_0)} - [g(\mathbf{x}_0)]^2 \\
& \gamma(\cdot) \stackrel{=} {=} g(\cdot) \sum_{l \in [N]} \frac{\langle \gamma(\cdot), g^{(\mathbf{e}_l, \mathcal{K}_l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}_0, \text{SDPCM})}}^2}{q_l(\mathbf{x}_0)} \\
& \stackrel{(6.83), (6.84)}{=} 2 \sum_{l \in \text{supp}(\mathbf{x}_0)} \frac{(x_{0,l} + \sigma^2)^2}{r_l} b_l^2 + 2\sigma^4 \left[1 - \frac{\xi_0^2}{(\xi_0 + \sigma^2)^2} \right]^{r_{j_0}/2} \sum_{l \in [N] \setminus \text{supp}(\mathbf{x}_0)} \frac{b_l^2}{r_l}. \quad (6.85)
\end{aligned}$$

□

In addition to being easily evaluated, the bound in (6.77) has two appealing properties. First, it is continuous with respect to \mathbf{x}_0 , and second, it is always at least as tight as the bound given by Theorem 6.3.2 when specialized to the SDPCM. Indeed, the bounds (6.43) and (6.44) of Theorem 6.3.2 applied to the SDPCM yield $L_{\mathcal{M}_{\text{SDPCM}}} \geq \sum_{l \in [N]} 2(x_{0,l} + \sigma^2)^2 \frac{b_l^2}{r_l}$ for $\|\mathbf{x}_0\|_0 < S$ and $L_{\mathcal{M}_{\text{SDPCM}}} \geq \sum_{l \in \text{supp}(\mathbf{x}_0)} 2(x_{0,l} + \sigma^2)^2 \frac{b_l^2}{r_l}$ when $\|\mathbf{x}_0\|_0 = S$. Comparing these expressions with (6.77) (note that $\xi_0 = 0$ when $\|\mathbf{x}_0\|_0 < S$), it can be shown that Theorem 6.4.4 yields always a lower bound that is at least as tight, i.e., at least as high as the bound given by Theorem 6.3.2 applied to the SDPCM.

Similar to Corollary 6.4.3, we have the following corollary of Theorem 6.4.4, which considers estimators for the SDPCM with an arbitrary bias function.

Corollary 6.4.5. *Consider an estimator $\hat{x}_k(\mathbf{y})$ which uses the observation \mathbf{y} of a SDPCM with sparsity degree S and denote its mean function by $m_k(\cdot) : \mathcal{X}_{S,+} \rightarrow \mathbb{R} : m_k(\mathbf{x}) \triangleq \mathbf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\}$. Let us fix a parameter vector $\mathbf{x}_0 \in \mathcal{X}_{S,+}$ and denote the value and index of the S largest entry of \mathbf{x}_0 by ξ_0 and j_0 , respectively. Furthermore, we define the sets $\mathcal{K}_l \triangleq \{\{l\} \cup \{\text{supp}(\mathbf{x}_0) \setminus \{j_0\}\}\}$ for $l \in [N] \setminus \text{supp}(\mathbf{x}_0)$ and $\mathcal{K}_l \triangleq \text{supp}(\mathbf{x}_0)$ for $l \in \text{supp}(\mathbf{x}_0)$. Then, if $m_k(\cdot)$ is such that the partial derivatives $b_l \triangleq \frac{\partial^{\mathbf{e}_l} m_k(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}_l}} \Big|_{\mathbf{x}=\mathbf{x}_0}$ exist, we have*

$$v(\hat{x}_k(\cdot); \mathbf{x}_0) \geq 2 \sum_{l \in \text{supp}(\mathbf{x}_0)} \frac{(x_{0,l} + \sigma^2)^2}{r_l} b_l^2 + 2\sigma^4 \left[1 - \frac{\xi_0^2}{(\xi_0 + \sigma^2)^2} \right]^{\frac{r_{j_0}}{2}} \sum_{l \in [N] \setminus \text{supp}(\mathbf{x}_0)} \frac{b_l^2}{r_l}. \quad (6.86)$$

If we want to use Corollary 6.4.3 or Corollary 6.4.5 for the comparison of the actual variance behavior of a given estimation scheme $\hat{x}_k(\cdot)$ with the lower bounds (6.76) and (6.86), respectively, we have to ensure that the partial derivatives of the mean function $m_k(\mathbf{x}) = \mathbf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\}$ of the given estimator $\hat{x}_k(\cdot)$ exist. That this is indeed the case for a very broad class of estimators $\hat{x}_k(\cdot)$ is stated in

Lemma 6.4.6. Consider an estimator $\hat{x}_k(\mathbf{y})$ which uses the observation \mathbf{y} of a SDPCM, and whose mean is $m_k(\mathbf{x})$. If $\hat{x}_k(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}$ is a Lebesgue-measurable function, and moreover

$$|\hat{x}_k(\mathbf{y})| \leq C \|\mathbf{y}\|_2^L \quad \forall \mathbf{y} \in \mathbb{R}^M, \quad (6.87)$$

with arbitrary but fixed constants $C, L \in \mathbb{R}$, then the partial derivatives $\frac{\partial^{\mathbf{p}} m_k(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}}$ exist for every multi-index $\mathbf{p} \in \mathbb{Z}_+^N \in [N]$. For the case $\mathbf{p} = \mathbf{e}_l$ with $l \in [N]$, the partial derivatives are given explicitly by

$$\frac{\partial^{\mathbf{e}_l} m_k(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{e}_l}} = -\frac{r_l}{2(x_l + \sigma^2)} m_k(\mathbf{x}) + \frac{1}{2(x_l + \sigma^2)^2} \mathbb{E}_{\mathbf{x}} \{ \mathbf{y}^T \mathbf{C}_l \mathbf{y} \hat{x}_k(\mathbf{y}) \}. \quad (6.88)$$

Proof. Appendix C. □

6.4.3 Special Case: Unbiased Spectrum Estimation

Let us now consider the special case of the SDPCM obtained for the parameter function $g(\mathbf{x}) = x_k$, where $k \in [N]$ is an arbitrary but fixed index. Since the entries x_k of the parameter vector \mathbf{x} can be interpreted as mean powers of the signal vector \mathbf{s} within the subspace $\text{span}(\mathbf{C}_k) \subseteq \mathbb{R}^M$ (cf. (6.2)), the resulting estimation problem is that of unbiased nonstationary spectrum estimation [103, 104] if the basis matrices \mathbf{C}_k correspond to well-localized regions in a time-frequency domain. The specialization of Theorem 6.4.4 to the parameter function $g(\mathbf{x}) = x_k$ yields

Corollary 6.4.7. Consider the minimum variance problem $\mathcal{M}_{\text{SDPCM}} = (\mathcal{E}_{\text{SDPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ associated with the SDPCM $\mathcal{E}_{\text{SDPCM}} = (\mathcal{X}_{S,+}, f_{\text{SDPCM}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k)$. For any estimator $\hat{x}_k(\cdot) : \mathbb{R}^M \rightarrow \mathbb{R}$ that is unbiased, i.e., $\mathbb{E}_{\mathbf{x}} \{ \hat{x}_k(\cdot) \} = x_k$ for every $\mathbf{x} \in \mathcal{X}_{S,+}$, we have the following lower bound on its variance at \mathbf{x}_0 :

$$v(\hat{x}_k(\cdot); \mathbf{x}_0) \geq \begin{cases} \frac{2}{r_k} (x_{0,k} + \sigma^2)^2, & k \in \text{supp}(\mathbf{x}_0) \\ \frac{2}{r_k} \sigma^4 \left[1 - \frac{\xi_0^2}{(\xi_0 + \sigma^2)^2} \right]^{\frac{r_{j_0}}{2}}, & k \notin \text{supp}(\mathbf{x}_0), \end{cases} \quad (6.89)$$

where ξ_0 and j_0 denote the value and index, respectively, of the S -largest entry of \mathbf{x}_0 .

Proof. This Corollary follows straightforwardly from Theorem 6.4.4 and the relation $b_l = \delta_{l,k}$ where b_l as defined in Theorem 6.4.4. □

The lower bound (6.89) can be achieved at least for the case $|\{k\} \cup \text{supp}(\mathbf{x}_0)| < S + 1$. In this case, the estimator given by (cf. (6.56))

$$\hat{x}_k(\mathbf{y}) = \beta_k(\mathbf{y}) - \sigma^2, \quad \text{with } \beta_k(\mathbf{y}) \triangleq \frac{1}{r_k} \sum_{i \in [r_k]} (\mathbf{u}_{m_k, i}^T \mathbf{y})^2, \quad (6.90)$$

is unbiased and its variance achieves the bound (6.89). Indeed, we have for the mean

$$\begin{aligned}
\mathbf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\} &= \mathbf{E}_{\mathbf{x}}\{\beta_k(\mathbf{y}) - \sigma^2\} = \mathbf{E}_{\mathbf{x}}\left\{\frac{1}{r_k} \sum_{i \in [r_k]} (\mathbf{u}_{m_k,i}^T \mathbf{y})^2\right\} - \sigma^2 \\
&= \frac{1}{r_k} \sum_{i \in [r_k]} \mathbf{E}_{\mathbf{x}}\{(\mathbf{u}_{m_k,i}^T \mathbf{y})^2\} - \sigma^2 = \frac{1}{r_k} \sum_{i \in [r_k]} (x_k + \sigma^2) - \sigma^2 \\
&= \frac{1}{r_k} r_k (x_k + \sigma^2) - \sigma^2 = x_k,
\end{aligned} \tag{6.91}$$

where we used the fact

$$\mathbf{u}_{m_k,i}^T \mathbf{y} \sim \mathcal{N}(0, \mathbf{u}_{m_k,i}^T \tilde{\mathbf{C}}(\mathbf{x}) \mathbf{u}_{m_k,i}) = \mathcal{N}(0, x_k + \sigma^2), \tag{6.92}$$

which follows from $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \tilde{\mathbf{C}}(\mathbf{x}))$, (6.4) and (6.56). Using the shorthand $z_i \triangleq \mathbf{u}_{m_k,i}^T \mathbf{y}$, the variance at an arbitrary parameter vector $\mathbf{x}_0 \in \mathcal{X}_{S,+}$ can be calculated as

$$\begin{aligned}
v(\hat{x}_k(\cdot); \mathbf{x}_0) &= v(\beta_k(\cdot); \mathbf{x}_0) \\
&= P(\beta_k(\cdot); \mathbf{x}_0) - [\mathbf{E}_{\mathbf{x}_0}\{\beta_k(\mathbf{y})\}]^2 \\
&= \mathbf{E}_{\mathbf{x}_0}\left\{\left[\frac{1}{r_k} \sum_{i \in [r_k]} z_i^2\right]^2\right\} - \left[\frac{1}{r_k} \sum_{i \in [r_k]} \mathbf{E}_{\mathbf{x}_0}\{z_i^2\}\right]^2 \\
&= \frac{1}{r_k^2} \sum_{i,i' \in [r_k]} \mathbf{E}_{\mathbf{x}_0}\{z_i^2 z_{i'}^2\} - \frac{1}{r_k^2} \left[\sum_{i \in [r_k]} \mathbf{E}_{\mathbf{x}_0}\{z_i^2\}\right]^2.
\end{aligned} \tag{6.93}$$

Using the fact that the random variables $\{z_i\}_{i \in [r_k]}$ are i.i.d. with $z_i \sim \mathcal{N}(0, x_k + \sigma^2)$, we obtain further

$$\begin{aligned}
v(\hat{x}_k(\cdot); \mathbf{x}_0) &= \frac{1}{r_k^2} \left[\sum_{i \in [r_k]} \mathbf{E}_{\mathbf{x}_0}\{z_i^4\} + \sum_{\substack{i,i' \in [r_k] \\ i \neq i'}} \mathbf{E}_{\mathbf{x}_0}\{z_i^2\} \mathbf{E}_{\mathbf{x}_0}\{z_{i'}^2\} - \left[\sum_{i \in [r_k]} \mathbf{E}_{\mathbf{x}_0}\{z_i^2\}\right]^2 \right] \\
&= \frac{1}{r_k^2} \left[\sum_{i \in [r_k]} \mathbf{E}_{\mathbf{x}_0}\{z_i^4\} + \left[\sum_{i \in [r_k]} \mathbf{E}_{\mathbf{x}_0}\{z_i^2\}\right]^2 - \sum_{i \in [r_k]} [\mathbf{E}_{\mathbf{x}_0}\{z_i^2\}]^2 - \left[\sum_{i \in [r_k]} \mathbf{E}_{\mathbf{x}_0}\{z_i^2\}\right]^2 \right] \\
&= \frac{1}{r_k^2} \left[\sum_{i \in [r_k]} \mathbf{E}_{\mathbf{x}_0}\{z_i^4\} - \sum_{i \in [r_k]} [\mathbf{E}_{\mathbf{x}_0}\{z_i^2\}]^2 \right]
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{r_k^2} \left[\sum_{i \in [r_k]} 3(x_k + \sigma^2)^2 - \sum_{i \in [r_k]} (x_k + \sigma^2)^2 \right] \\
&= \frac{2}{r_k} (x_k + \sigma^2)^2,
\end{aligned} \tag{6.94}$$

where we used a well-known identity for the fourth-order central moment of a zero-mean Gaussian random variable [45]. The estimator in (6.90) does not use the sparsity information and does not depend on \mathbf{x}_0 . Moreover, since this estimator is the unique LMVU estimator at any $\|\mathbf{x}_0\|_0 < S$, we have that if a UMVU estimator existed for the unbiased spectrum estimation problem $\mathcal{E}_{\text{SDPCM}} = (\mathcal{X}_{S,+}, f_{\text{SDPCM}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k)$, it would necessarily coincide with the nonsparse estimator in (6.90) and, thus, it would ignore the sparsity information $\mathbf{x} \in \mathcal{X}_{S,+}$. This is an instance of the general rule that unbiased estimation is usually not optimal for estimation problems with sparsity constraints.

Let us define a “signal-to-noise ratio” (SNR) quantity as $\text{SNR}(\mathbf{x}_0) \triangleq \xi_0/\sigma^2$. For $\text{SNR}(\mathbf{x}_0) \ll 1$, where $\left[1 - \frac{\xi_0^2}{(\xi_0 + \sigma^2)^2}\right]^{\frac{r_{j_0}}{2}} \approx 1$, the lower bound (6.89) in Corollary 6.4.7 is approximately $\frac{2}{r_k} (x_{0,k} + \sigma^2)^2$ for any k , which does not depend on S and moreover equals the variance of the unbiased estimator (6.90). Since that estimator does not exploit any sparsity information, Corollary 6.4.7 suggests that, in the low-SNR regime, unbiased estimators cannot exploit the prior information that \mathbf{x} is S -sparse, i.e., $\mathbf{x} \in \mathcal{X}_{S,+}$. However, in the high-SNR regime ($\text{SNR}(\mathbf{x}_0) \rightarrow \infty$), (6.89) becomes $\frac{2}{r_k} (x_{0,k} + \sigma^2)^2$ for $k \in \text{supp}(\mathbf{x}_0)$ and 0 for $k \notin \text{supp}(\mathbf{x}_0)$, which can be shown to equal the variance of the oracle estimator that knows $\text{supp}(\mathbf{x}_0)$ (this oracle estimator yields $\hat{x}_k = x_{0,k} = 0$ for all $k \notin \text{supp}(\mathbf{x}_0)$). The transition of the lower bound (6.89) from the low-SNR regime to the high-SNR regime has a polynomial characteristic; it is thus much slower than the exponential transition of the analogous lower bound for the SLM given by Theorem 5.4.3. This slow transition suggests that the optimal unbiased estimator for low SNR—which ignores the sparsity information—will also be a nearly optimal unbiased estimator over a relatively wide SNR range. This further suggests that, for unbiased covariance estimation based on the SDPCM, prior information of sparsity is not as helpful as for minimum variance estimation for the SLM.

Based on the specific estimator given by (6.90), we can also show that for the SDPCM, similarly to the SLM (see Section 5.6), the strict sparsity requirement $\mathbf{x} \in \mathcal{X}_{S,+}$ is necessary in order to allow for the existence of unbiased spectrum estimators $\hat{x}_k(\cdot)$ whose variance at \mathbf{x}_0 is smaller than the variance of the nonsparse estimator in (6.90). We will also show that the variance of the estimator in (6.90) is simultaneously the theoretical minimum variance that can be achieved without any sparsity constraints.

Indeed, consider any minimum variance problem $\mathcal{M}' = (\mathcal{E}', c(\cdot) \equiv 0, \mathbf{x}_0)$ associated with the estimation problem $\mathcal{E}' = (\mathcal{X}', f_{\text{SPCM}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k)$ (with arbitrary $k \in [N]$), where the sta-

tistical model $f_{\text{SPCM}}(\mathbf{y}; \mathbf{x})$ (cf. (6.8)) is defined using basis matrices \mathbf{C}_k satisfying (6.12). Thus, the minimum variance problem \mathcal{M}' is identical to the minimum variance problem $\mathcal{M}_{\text{SDPCM}} = (\mathcal{E}_{\text{SDPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ with $\mathcal{E}_{\text{SDPCM}} = (\mathcal{X}_{S,+}, f_{\text{SDPCM}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k)$ except for the parameter set which is \mathcal{X}' instead of $\mathcal{X}_{S,+}$. Let us assume that the parameter set \mathcal{X}' contains an open ball $\mathcal{B}(\mathbf{x}_0, r)$, i.e., $\mathcal{B}(\mathbf{x}_0, r) \subseteq \mathcal{X}'$ with some radius $r > 0$. We can then apply Theorem 4.4.4, i.e., the ordinary unconstrained CRB for a general nonsparse estimation problem with Gaussian observation $\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}(\mathbf{x}))$ [20], to obtain the lower bound $v(\hat{x}_k(\cdot); \mathbf{x}_0) \geq \frac{2}{r_k}(x_{0,k} + \sigma^2)^2$, which is achieved by the estimator in (6.90). From this, we conclude that if the parameter set is chosen in particular as $\mathcal{X}' = \mathbb{R}_+^N \cap \mathcal{B}_q(S)$ with $q > 0$ (cf. (5.178)), i.e., consisting of approximately S -sparse vectors, then the UMVU estimator is always given by the estimator in (6.90), whose variance is in general strictly larger than the bound (6.89) for the case $k \notin \text{supp}(\mathbf{x}_0)$. In particular for the case $\|\mathbf{x}_0\|_0 = S$, i.e., $\xi_0 > 0$ we have

$$\frac{2}{r_k} \sigma^4 \underbrace{\left[1 - \frac{\xi_0^2}{(\xi_0 + \sigma^2)^2}\right]^{\frac{r_{j_0}}{2}}}_{<1} < \frac{2}{r_k} (x_{0,k} + \sigma^2)^2. \quad (6.95)$$

However, for the SDPCM with the strictly sparse parameter set $\mathcal{X}_{S,+}$, there exist in the general case unbiased estimators $\hat{x}_k(\cdot)$ for $\mathcal{M}_{\text{SDPCM}} = (\mathcal{E}_{\text{SDPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ with $\mathcal{E}_{\text{SDPCM}} = (\mathcal{X}_{S,+}, f_{\text{SDPCM}}(\mathbf{y}; \mathbf{x}), g(\mathbf{x}) = x_k)$, that have a variance that is strictly smaller than that of the estimator (6.90) at a certain parameter vector $\mathbf{x}_0 \in \mathcal{X}_{S,+}$. In particular, consider the simplest configuration of the SDPCM where $\mathbf{C}_k = \sum_{i \in [r_k]} \mathbf{e}_{j_{k,i}} \mathbf{e}_{j_{k,i}}^T$, $N > 1$, and $S = 1$. For this instance of the SDPCM, one can find the LMVU estimator $\hat{x}_k^{(\mathbf{x}_0)}(\cdot)$ for the minimum variance problem $\mathcal{M}_{\text{SDPCM}} = (\mathcal{E}_{\text{SDPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$ (with parameter function $g(\mathbf{x}) = x_k$):

Theorem 6.4.8. *Consider the minimum variance problem $\mathcal{M}_{\text{SDPCM}} = (\mathcal{E}_{\text{SDPCM}}, c(\cdot) \equiv 0, \mathbf{x}_0)$, with $\mathbf{x}_0 \in \mathcal{X}_{S,+}$, $N > 1$, $S = 1$, and $g(\mathbf{x}) = x_k$. Denoting by ξ_0 and j_0 the value and index, respectively, of the only nonzero entry of \mathbf{x}_0 , the LMVU estimator for $\mathcal{M}_{\text{SDPCM}}$ can be written as*

$$\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) = \begin{cases} \beta_k(\mathbf{y}) - \sigma^2, & k = j_0 \\ \alpha(\mathbf{y}; \mathbf{x}_0) (\beta_k(\mathbf{y}) - \sigma^2), & k \neq j_0, \end{cases} \quad (6.96)$$

where

$$\alpha(\mathbf{y}; \mathbf{x}_0) \triangleq a(\mathbf{x}_0) \exp(-r_{j_0} b(\mathbf{x}_0) \beta_{j_0}(\mathbf{y})) \quad (6.97)$$

with $a(\mathbf{x}_0) \triangleq \left[\frac{2\xi_0 + \sigma^2}{\xi_0 + \sigma^2}\right]^{r_{j_0}/2}$, $b(\mathbf{x}_0) \triangleq \frac{1}{2} \left(\frac{1}{\sigma^2} - \frac{1}{\xi_0 + \sigma^2}\right)$, and $\beta_k(\mathbf{y})$ as defined in (6.90).

Proof. Appendix D. □

6.5 Comparison of the Bounds with Existing Estimators

We will now compare the lower bound (6.76) (in Corollary 6.4.3) on the variance of estimators for the SDPCM, with the actual variance behavior of two practical estimation schemes. The first is an ad-hoc adaptation of the hard-thresholding (HT) estimator [23] to SDPCM-based covariance estimation. It is defined componentwise as (cf. (6.90))

$$\hat{x}_{\text{HT},k}(\mathbf{y}) \triangleq \frac{1}{r_k} \sum_{i \in [r_k]} \varphi_T^2(\mathbf{u}_{m_{k,i}}^T \mathbf{y}) - \sigma^2,$$

where $\varphi_T: \mathbb{R} \rightarrow \mathbb{R}$ denotes the hard-thresholding function with threshold $T \geq 0$, i.e.,

$$\varphi_T(y) = \begin{cases} y, & |y| \geq T \\ 0, & \text{else.} \end{cases} \quad (6.98)$$

The second standard method is the maximum likelihood (ML) estimator

$$\hat{\mathbf{x}}_{\text{ML}}(\mathbf{y}) \triangleq \underset{\mathbf{x}' \in \mathcal{X}_{S,+}}{\operatorname{argmax}} f_{\text{SDPCM}}(\mathbf{y}; \mathbf{x}').$$

For the ML estimator, we obtain

$$\begin{aligned} \hat{\mathbf{x}}_{\text{ML}}(\mathbf{y}) &\triangleq \underset{\mathbf{x}' \in \mathcal{X}_{S,+}}{\operatorname{argmax}} f_{\text{SDPCM}}(\mathbf{y}; \mathbf{x}') \\ &\stackrel{(6.14)}{=} \underset{\mathbf{x}' \in \mathcal{X}_{S,+}}{\operatorname{argmax}} \frac{1}{(2\pi)^{M/2} [\det\{\tilde{\mathbf{C}}(\mathbf{x}')\}]^{1/2}} \exp\left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}') \mathbf{y}\right) \\ &\stackrel{(a)}{=} \underset{\mathbf{x}' \in \mathcal{X}_{S,+}}{\operatorname{argmax}} \log \left[\frac{1}{[\det\{\tilde{\mathbf{C}}(\mathbf{x}')\}]^{1/2}} \exp\left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}') \mathbf{y}\right) \right] \\ &= \underset{\mathbf{x}' \in \mathcal{X}_{S,+}}{\operatorname{argmax}} \left\{ -\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}') \mathbf{y} - \frac{1}{2} \log [\det\{\tilde{\mathbf{C}}(\mathbf{x}')\}] \right\} \\ &= \underset{\mathbf{x}' \in \mathcal{X}_{S,+}}{\operatorname{argmax}} \left\{ -\mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}') \mathbf{y} - \log [\det\{\tilde{\mathbf{C}}(\mathbf{x}')\}] \right\}, \end{aligned} \quad (6.99)$$

where the step (a) follows from the fact that the logarithm is monotonically increasing on its domain $\mathbb{R}_+ \setminus \{0\}$, i.e., it preserves the position of the maximum. Using (6.58) and (6.59), we obtain further

$$\hat{\mathbf{x}}_{\text{ML}}(\mathbf{y}) = \underset{\mathbf{x}' \in \mathcal{X}_{S,+}}{\operatorname{argmax}} \left\{ -\sum_{k \in [N]} r_k \left[\frac{\beta_k(\mathbf{y})}{x'_k + \sigma^2} + \log(x'_k + \sigma^2) \right] \right\}, \quad (6.100)$$

with $\beta_k(\mathbf{y})$ as defined in (6.90).

We have

Theorem 6.5.1. *The ML estimator for the SDPCM is given componentwise as*

$$\hat{x}_{\text{ML},k}(\mathbf{y}) = \begin{cases} \beta_k(\mathbf{y}) - \sigma^2, & k \in \mathcal{L}_1 \cap \mathcal{L}_2 \\ 0, & \text{else,} \end{cases} \quad (6.101)$$

where $\beta_k(\mathbf{y})$ is defined in (6.90), \mathcal{L}_1 consists of the S indices $k \in [N]$ for which $r_k[\beta_k(\mathbf{y})/\sigma^2 - \log(\beta_k(\mathbf{y})/\sigma^2) - 1]$ is largest, and \mathcal{L}_2 consists of all indices k for which $\beta_k(\mathbf{y}) \geq \sigma^2$.

Proof. Appendix E. □

In what follows, we will denote by $L^{(m_k(\cdot))}(\mathbf{x}_0)$ the bound (6.76) in Corollary 6.4.3 for an estimator $\hat{x}_k(\cdot)$ with mean function $\mathbf{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\} = m_k(\cdot)$, using the index set $\mathcal{K} = \text{supp}(\mathbf{x}_0)$ when $k \in \text{supp}(\mathbf{x}_0)$ and $\mathcal{K} = \{\text{supp}(\mathbf{x}_0) \setminus \{j_0\}\} \cup \{k\}$ else, where $j_0 \in [N]$ denotes the index of the S -largest entry of \mathbf{x}_0 . We furthermore use the choices $L=2$, $\mathbf{p}_1 = \mathbf{0}$, and $\mathbf{p}_2 = \mathbf{e}_k$ for the application of Corollary 6.4.3. In order to evaluate the bound $L^{(m_k(\cdot))}(\mathbf{x}_0)$, we need to compute the first-order partial derivatives of the mean function $m_k(\cdot)$. This is accomplished by using Lemma 6.4.6 for the HT estimator and by using a finite-difference quotient approximation [56] for the ML estimator, i.e.,

$$\frac{\partial \mathbf{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}(\mathbf{y})\}}{\partial x_l} \approx \frac{\mathbf{E}_{\mathbf{x}+\Delta \mathbf{e}_l}\{\hat{x}_{\text{ML},k}(\mathbf{y})\} - \mathbf{E}_{\mathbf{x}}\{\hat{x}_{\text{ML},k}(\mathbf{y})\}}{\Delta}, \quad (6.102)$$

where $\Delta \in \mathbb{R}_+$ is a small stepsize and the expectations are calculated using numerical integration.

Given an estimator $\hat{\mathbf{x}}(\cdot)$ whose mean is equal to $\mathbf{m}(\mathbf{x}) \triangleq \mathbf{E}_{\mathbf{x}}\{\hat{\mathbf{x}}(\cdot)\}$, we obtain due to (2.21) a lower bound on the estimator variance $v(\hat{\mathbf{x}}(\cdot); \mathbf{x}_0)$ by summing the quantities $L^{(m_k(\cdot))}(\mathbf{x}_0)$, i.e.,

$$v(\hat{\mathbf{x}}(\cdot); \mathbf{x}_0) \geq L^{(\mathbf{m}(\cdot))}(\mathbf{x}_0) \triangleq \sum_{k \in [N]} L^{(m_k(\cdot))}(\mathbf{x}_0). \quad (6.103)$$

For a numerical evaluation, we considered the SDPCM with $N=50$, $S=5$, and $\mathbf{C}_k = \mathbf{e}_k \mathbf{e}_k^T$. We generated parameter vectors $\mathbf{x}_0 = \text{SNR} \cdot \sigma^2 \mathbf{x}_1$, where $\mathbf{x}_1 \in \{0, 1\}^{50}$, $\text{supp}(\mathbf{x}_1) = [S]$, and SNR varies between 0.01 and 100. (The fixed choice $\text{supp}(\mathbf{x}_0) = [S]$ is justified by the fact that neither the variances of the ML and HT estimators nor the corresponding variance bounds depend on the location of $\text{supp}(\mathbf{x}_0)$.) In Fig. 6.1, we show the variance at \mathbf{x}_0 , $v(\hat{\mathbf{x}}(\cdot); \mathbf{x}_0) = \sum_{k \in [N]} v(\hat{x}_k(\cdot); \mathbf{x}_0)$ (computed by means of numerical integration), for the HT estimator using various choices of T and for the ML estimator. The variance is plotted versus SNR. Along with each variance curve, we display the corresponding lower bound $L^{(\mathbf{m}(\cdot))}(\mathbf{x}_0)$, using for $\mathbf{m}(\cdot)$ the mean function of the respective estimator (HT or ML). (The mean functions of the HT and ML estimators were computed by means of numerical integration.) In Fig. 6.1, all variances and bounds are normalized by $2(\text{SNR}+1)^2\sigma^4$, which is the variance of an optimum unbiased oracle estimator for the SDPCM with $S=1$ and the nonzero entry of \mathbf{x}_0 being equal to $\text{SNR} \cdot \sigma^2$. The computation of the oracle

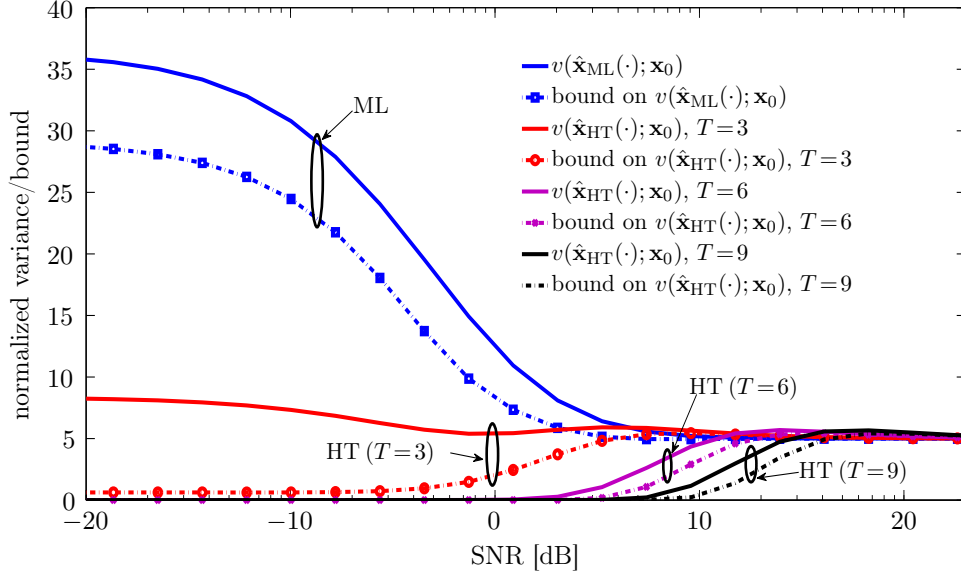


Figure 6.1: Normalized variance of the ML and HT estimators and corresponding lower bounds versus the SNR, for the SDPCM with $N=50$, $S=5$, $\sigma^2=1$, and $\mathbf{C}_k = \mathbf{e}_k \mathbf{e}_k^T$.

estimator presupposes the knowledge of the position of the nonzero entry of \mathbf{x}_0 , i.e., it knows $\text{supp}(\mathbf{x}_0)$.

It can be seen from Fig. 6.1 that in the high-SNR regime, for both estimators, the gap between the variance and the corresponding lower bound is quite small. This indicates that the variances of both estimators are nearly minimal (for the respective bias functions). Indeed, for high SNR, all variance curves approach the variance of an optimum unbiased oracle estimator that knows the position of the nonzero entries of \mathbf{x}_0 , i.e., it knows $\text{supp}(\mathbf{x}_0)$. The variance of this oracle estimator is given by $2S(\text{SNR} + 1)^2\sigma^4$ (this expression is only valid for parameter vectors \mathbf{x}_0 with exactly S nonzero entries whose values are all equal to $\text{SNR} \cdot \sigma^2$). However, in the low-SNR regime, the variances of the estimators tend to be significantly higher than the bounds. This means that there *may* be estimators with the same bias function as that of the HT or ML estimator but a lower variance. However, the actual existence of such estimators is not shown by our analysis.

Chapter 7

Conclusion and Outlook

This thesis specialized the RKHS approach to minimum variance estimation [3, 35] to estimation problems with sparsity constraints on the unknown parameter vector to be estimated. After a brief introduction to the main concepts of classical (non-Bayesian) estimation and Hilbert space theory - with emphasis on RKHS - we reviewed the RKHS approach to minimum variance estimation. This approach is based on two isometric Hilbert spaces, one of them being a RKHS, which can be associated in a natural manner to a minimum variance problem. We presented two fundamental facts about minimum variance estimation using the RKHS approach, which seem to be novel, to the best of the author's knowledge. The first fact is that the minimum achievable variance, viewed as a function of the parameter vector \mathbf{x}_0 at which the variance is minimized, is a lower semi-continuous function. The second fact is that the RKHS associated to a minimum variance problem remains unchanged if the observation of the minimum variance problem is replaced by a sufficient statistic. This fact is closely related in spirit to the Rao-Blackwell-Lehmann-Scheffé theorem of classical estimation.

The core of this thesis are Chapter 5 and Chapter 6, in which we considered classical estimation problems with sparsity constraints. In Chapter 5, we considered the SLM, where the unknown sparse parameter vector is linearly distorted and corrupted by additive white Gaussian noise. The SLM can be used to model the recovery problem in a linear compressed sensing scheme. We analyzed the minimum variance estimation problem using the RKHS approach and derived novel lower bounds on the estimator variance for a prescribed estimator bias. These lower bounds are obtained by an orthogonal projection of the prescribed mean function onto a subspace of the RKHS that is associated to the SLM. The bounds share the characteristic that they vary between two extreme cases. On the one hand, we have a low-SNR regime where the entries of the true parameter vector are small compared with the noise variance. Here, our bounds predict that the a-priori information of a sparse parameter vector does not help much for the SLM if the

estimator bias is approximately zero. However, by allowing a nonzero bias it is possible to reduce the variance. On the other hand, we have a high-SNR regime where the nonzero entries of the sparse parameter vector are all considerable larger than the noise variance. Here, our bounds coincide with the CRB of an associated LGM that is obtained from the SLM if the support of the unknown parameter vector is known. Our bounds exhibit a very steep transition between these two regimes. In general, this transition has an exponential decay. For the special case of the SLM given by the SSNM, we strengthened the results and derived closed-form expressions for the minimum achievable variance and corresponding LMV estimator. This includes the (unbiased) Barankin bound and LMVU estimator for the SSNM. We also compared our lower bounds on the variance as well as the expressions for the minimum achievable variance to the actual variance behavior of popular estimation schemes for the SLM and SSNM.

In Chapter 6, we considered the SPCM, where the unknown sparse parameter vector determines the covariance matrix of a random signal vector of which a noisy version is observed. The SPCM is relevant, e.g., in the context of spectrum estimation for scene analysis in cognitive radio systems. We discussed the fact that the RKHS approach is more involved for the SPCM than for the SLM. We presented one particular method of associating a RKHS to the SPCM. Based on this RKHS, we derived lower bounds on the estimator variance for estimators with a given bias function. As for the SLM, these bounds were obtained by projecting the corresponding mean function onto a subspace of the RKHS. For the special case of the SPCM which is given by the SDPCM, we strengthened our results and derived the minimum achievable variance and corresponding LMVU estimator for the case where the parameter vector is known to have at most one nonzero entry. As for the SLM and SSNM, also the lower bounds for the SPCM and SDPCM exhibit a transition from a low-SNR regime to a high-SNR regime, and the bounds for the high-SNR regime again coincide with the CRB of an associated estimation problem without sparsity constraints. However, a remarkable difference between the bounds for the SDPCM and for the SSNM is that the transition of the SDPCM bounds is only polynomial in the SNR whereas that of the SSNM bounds is exponential. This further suggests that in terms of minimum variance estimation, for covariance estimation based on the SDPCM, prior information of sparsity is not as helpful as for estimating the parameter vector of the SSNM.

Topics for future research that may be addressed using this thesis as a starting point include the following.

- **Asymptotic Analysis.** Can we find (classes of) estimators that come close to the lower variance bounds asymptotically, i.e., when we consider not only one observation but an ensemble of i.i.d. observations whose number tends to infinity? Intuitively, the class of ML estimators should achieve the variance bounds asymptotically in the unbiased case. However, a rigorous proof of this claim seems to be nontrivial for the sparse parameter set \mathcal{X}_S . Indeed,

most approaches to the analysis of the asymptotic behavior of ML estimators assume that the parameter set is an open subset of \mathbb{R}^N [41, 69, 105], which is not the case for the parameter set \mathcal{X}_S . Another popular class of estimators is given by the M-estimators or penalized maximum likelihood estimators, for which a characterization of their asymptotic behavior is available (cf. [57, 105, 106]). Moreover, under mild conditions, the class of M-estimators allows an efficient implementation via convex optimization techniques.

- **Finite-Rate-of-Innovation Signals.** Another desirable generalization of our results would be to consider estimation of finite-rate-of-innovation signals, which are in some sense the continuous-time analogue of sparse vectors in the finite-dimensional setting [107].
- **Block Sparsity.** A popular generalization of the concept of strict sparsity that we used in our thesis is block or group sparsity [108, 109], where a block structure is pre-specified for the parameter vector $\mathbf{x} \in \mathbb{R}^N$. Block sparsity then refers to the fact that only few blocks contain at least one nonzero entry. The traditional notion of strict sparsity is reobtained when the block size is equal to one. Block sparsity could be useful, e.g., in the context of spectrum estimation for sparse underspread processes [98, 99] if the process energy occurs in clusters in the time-frequency plane.
- **Sparse Exponential Family.** The SLM and SPCM are estimation problems with a statistical model that is a special case of the exponential family. It would be interesting to study estimation problems with sparsity constraints whose statistical model is the general exponential family. In particular, it would be interesting to study how the cumulant function of the exponential family relates to the RKHS and its geometric properties.
- **Random System Matrix and Basis Matrices.**

Another class of estimation problems that may be interesting to consider is obtained by a modification of the SLM and SPCM in which the system matrix \mathbf{H} and basis matrices \mathbf{C}_k , respectively, are modeled as random matrices. In particular, a characterization of the RKHS for a SLM with a random system matrix \mathbf{H} , corresponding to the compressed sensing measurement model, would be desirable. Similarly, one could investigate the RKHS associated to a SPCM with random basis matrices \mathbf{C}_k . The resulting RKHS could be used to analyze the compressive spectrum estimators proposed in [98, 99].

Appendix A

Proof of Theorem 5.5.4

Proof. We give two different (though similar) proofs for the case $|\text{supp}(\mathbf{x}_0) \cup \{k\}| < S + 1$ and the complementary case $|\text{supp}(\mathbf{x}_0) \cup \{k\}| = S + 1$.

Case I: $|\text{supp}(\mathbf{x}_0) \cup \{k\}| < S + 1$:

Consider the functions

$$h^{(l)}(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : h^{(l)}(\mathbf{x}) = \frac{\sigma^l}{\sqrt{l!}} \frac{\partial^{l\mathbf{e}_k} R_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}, \mathbf{x}_2)}{\partial \mathbf{x}_2^{l\mathbf{e}_k}} \Big|_{\mathbf{x}_2=\mathbf{x}_0} \stackrel{(5.117)}{=} \frac{1}{\sigma^l \sqrt{l!}} (x_{1,k} - x_{0,k})^l, \quad (\text{A.1})$$

with $l \in \mathbb{Z}_+$. By Theorem 3.4.1 (note that $\mathcal{N}_{\mathbf{x}_0}^{\text{supp}(l\mathbf{e}_k)} \subseteq \mathcal{X}_S$) we have that the functions $\{h^{(l)}(\cdot)\}_{l \in \mathbb{Z}_+}$ belong to $\mathcal{H}(\mathcal{M}_{\text{SSNM}})$ and moreover are orthonormal, since

$$\begin{aligned} \langle h^{(l)}(\cdot), h^{(l')}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})} &\stackrel{(\text{A.1})}{=} \left\langle \frac{\sigma^l}{\sqrt{l!}} \frac{\partial^{l\mathbf{e}_k} R_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}, \mathbf{x}_2)}{\partial \mathbf{x}_2^{l\mathbf{e}_k}} \Big|_{\mathbf{x}_2=\mathbf{x}_0}, \frac{\sigma^{l'}}{\sqrt{l'!}} \frac{\partial^{l'\mathbf{e}_k} R_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}, \mathbf{x}_2)}{\partial \mathbf{x}_2^{l'\mathbf{e}_k}} \Big|_{\mathbf{x}_2=\mathbf{x}_0} \right\rangle_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})} \\ &\stackrel{(3.66)}{=} \frac{\sigma^l}{\sqrt{l!}} \frac{\sigma^{l'}}{\sqrt{l'!}} \frac{\partial^{l\mathbf{e}_k} \partial^{l'\mathbf{e}_k} R_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{x}_1^{l\mathbf{e}_k} \partial \mathbf{x}_2^{l'\mathbf{e}_k}} \Big|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}_0} \\ &\stackrel{(5.117)}{=} \frac{\sigma^l}{\sqrt{l!}} \frac{\sigma^{l'}}{\sqrt{l'!}} \frac{\partial^{l\mathbf{e}_k} \partial^{l'\mathbf{e}_k} \exp\left(\frac{1}{\sigma^2}(\mathbf{x}_1 - \mathbf{x}_0)^T(\mathbf{x}_2 - \mathbf{x}_0)\right)}{\partial \mathbf{x}_1^{l\mathbf{e}_k} \partial \mathbf{x}_2^{l'\mathbf{e}_k}} \Big|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{x}_0} \\ &= \frac{\sigma^l}{\sqrt{l!}} \frac{\sigma^{l'}}{\sqrt{l'!}} \frac{1}{\sigma^{2l'}} \frac{\partial^{l\mathbf{e}_k} (x_{1,k} - x_{0,k})^{l'}}{\partial \mathbf{x}_1^{l\mathbf{e}_k}} \Big|_{\mathbf{x}_1=\mathbf{x}_0} = \delta_{l,l'}. \end{aligned} \quad (\text{A.2})$$

Obviously, then the function

$$s_N(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : s_N(\mathbf{x}) \triangleq \sum_{l \in [N]} \frac{m_l \sigma^l}{\sqrt{l!}} h^{(l)}(\mathbf{x}) + m_0 h^{(0)}(\mathbf{x}) \quad (\text{A.3})$$

also belongs to $\mathcal{H}(\mathcal{M}_{\text{SSNM}})$ and has squared norm $\|s_N(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2 = m_0^2 + \sum_{l \in [N]} \frac{m_l^2}{l!} \sigma^{2l}$.

Now let us assume that (5.150) is satisfied. Since (for $N > N'$)

$$\|s_N(\cdot) - s_{N'}(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2 = \sum_{l \in [N] \setminus [N']} \frac{m_l^2}{l!} \sigma^{2l}, \quad (\text{A.4})$$

we have from (5.150) that the sequence $\{s_N(\cdot)\}_{N \rightarrow \infty}$ is a Cauchy sequence and therefore converges in the RKHS $\mathcal{H}(\mathcal{M}_{\text{SSNM}})$. Since, by (A.1) and (5.149), this sequence moreover converges pointwise to the prescribed mean function $\gamma(\mathbf{x})$, we have by Theorem 3.3.3 that

$$\lim_{N \rightarrow \infty} s_N(\cdot) = \gamma(\cdot), \quad (\text{A.5})$$

which implies that $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SSNM}})$ and in turn that the prescribed bias function $c(\cdot)$ is valid. We have then by Theorem 4.3.4 that

$$L_{\mathcal{M}_{\text{SSNM}}} = \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2 - [\gamma(\mathbf{x}_0)]^2 = \lim_{N \rightarrow \infty} \|s_N(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2 - m_0^2 = \sum_{l \in \mathbb{N}} \frac{m_l^2}{l!} \sigma^{2l}, \quad (\text{A.6})$$

where we have used that $\gamma(\mathbf{x}_0) = m_0$ (according to (5.149)). Thus, we have verified the sufficiency of (5.150) for a bias function to be valid and for the expression (5.151) for the minimum achievable variance.

It remains to verify the necessity of (5.150) for a bias function to be valid. We will prove this by contradiction. Assume that (5.150) is not fulfilled, i.e., the sums $\sum_{l \in \mathcal{T}} \frac{m_l^2}{l!} \sigma^{2l}$ with a finite index set $\mathcal{T} \subseteq \mathbb{Z}_+$ can be arbitrarily large, but the bias function $c(\cdot)$ is valid, i.e., $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SSNM}})$. By Theorem 3.1.6, we then have for any subspace $\mathcal{U} \triangleq \text{span}\{h^{(l)}(\cdot)\}_{l \in \mathcal{T}}$ with a finite set $\mathcal{T} \subseteq \mathbb{Z}_+$ that¹

$$\begin{aligned} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2 &\stackrel{(3.31)}{\geq} \|\mathbf{P}_{\mathcal{U}}\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2 \\ &\stackrel{(3.33)}{=} \sum_{l \in \mathcal{T}} \langle \gamma(\cdot), h^{(l)}(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2 \\ &\stackrel{(3.66)}{=} \sum_{l \in \mathcal{T}} \frac{\sigma^{2l}}{l!} \left[\frac{\partial^{l\mathbf{e}_k} \gamma(\cdot)}{\partial \mathbf{x}^{l\mathbf{e}_k}} \Big|_{\mathbf{x}=\mathbf{x}_0} \right]^2 \\ &\stackrel{(5.149)}{=} \sum_{l \in \mathcal{T}} \frac{m_l^2}{l!} \sigma^{2l}, \end{aligned} \quad (\text{A.7})$$

where we also used Theorem 3.1.7 and Theorem 3.4.1. However, since (5.150) is not satisfied, i.e., we can make $\sum_{l \in \mathcal{T}} \frac{m_l^2}{l!} \sigma^{2l}$ arbitrarily large by using a suitable finite index set \mathcal{T} , it follows from (A.7) that the squared norm $\|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2$ is unbounded, which is impossible because $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SSNM}})$.

¹Note that, trivially, $\{h^{(l)}(\cdot)\}_{l \in \mathcal{T}}$ is an ONB for \mathcal{U} .

Finally, we will show (5.152). If the prescribed bias is valid, we can by (A.5) represent the corresponding mean as

$$\gamma(\cdot) = \lim_{N \rightarrow \infty} s_N(\cdot) = \sum_{l \in \mathbb{Z}_+} \frac{m_l}{\sqrt{l!}} \sigma^l h^{(l)}(\cdot). \quad (\text{A.8})$$

From this, we obtain the LMV estimator in (5.152) from (A.1), Theorem 3.4.1, Theorem 4.3.3, Theorem 4.3.4 and the identity

$$\begin{aligned} & \frac{m_l \sigma^{2l}}{l!} \frac{\partial^l \exp\left(-\frac{1}{2\sigma^2}(2y(x_0 - x) + x^2 - x_0^2)\right)}{\partial x^l} \Big|_{x=x_0} \\ &= \frac{m_l \sigma^{2l}}{l!} \frac{\partial^l \exp\left(-\frac{1}{2\sigma^2}((x - y)^2 - y^2 + 2yx_0 - x_0^2)\right)}{\partial x^l} \Big|_{x=x_0} \\ &= \frac{m_l \sigma^{2l}}{l!} \exp\left(\frac{1}{2\sigma^2}(y - x_0)^2\right) \frac{\partial^l \exp\left(-\frac{1}{2\sigma^2}(x - y)^2\right)}{\partial x^l} \Big|_{x=x_0} \\ & \left(x' \triangleq \frac{x - y}{\sigma}\right) \\ &= \frac{m_l \sigma^l}{l!} \exp\left(\frac{1}{2\sigma^2}(x_0 - y)^2\right) \frac{\partial^l \exp\left(-\frac{1}{2}x'^2\right)}{\partial x'^l} \Big|_{x' = \frac{x_0 - y}{\sigma}} \\ &= \frac{m_l \sigma^l}{l!} (-1)^l \text{H}_l\left(\frac{x_0 - y}{\sigma}\right) = \frac{m_l \sigma^l}{l!} \text{H}_l\left(\frac{y - x_0}{\sigma}\right). \end{aligned} \quad (\text{A.9})$$

Indeed, we have

$$\begin{aligned} \hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) &\stackrel{(4.20)}{=} \text{J}[\gamma(\cdot)] \stackrel{(A.8)}{=} \text{J}\left[\sum_{l \in \mathbb{Z}_+} \frac{m_l}{\sqrt{l!}} \sigma^l h^{(l)}(\cdot)\right] = \sum_{l \in \mathbb{Z}_+} \frac{m_l}{\sqrt{l!}} \sigma^l \text{J}[h^{(l)}(\cdot)] \\ &\stackrel{(A.1)}{=} \sum_{l \in \mathbb{Z}_+} \frac{m_l}{l!} \sigma^{2l} \text{J}\left[\frac{\partial^{l\mathbf{e}_k} R_{\text{MSSNM}}(\cdot, \mathbf{x})}{\partial \mathbf{x}^{l\mathbf{e}_k}} \Big|_{\mathbf{x}=\mathbf{x}_0}\right] \\ &\stackrel{(4.16)}{=} \sum_{l \in \mathbb{Z}_+} \frac{m_l}{l!} \sigma^{2l} \frac{\partial^{l\mathbf{e}_k} \rho_{\text{MSSNM}}(\mathbf{y}, \mathbf{x})}{\partial \mathbf{x}^{l\mathbf{e}_k}} \Big|_{\mathbf{x}=\mathbf{x}_0} \\ &\stackrel{(5.145)}{=} \sum_{l \in \mathbb{Z}_+} \frac{m_l \sigma^{2l}}{l!} \frac{\partial^{l\mathbf{e}_k} \exp\left(\frac{1}{\sigma^2} \mathbf{y}^T (\mathbf{x} - \mathbf{x}_0) - \frac{1}{2\sigma^2} (\|\mathbf{x}\|_2^2 - \|\mathbf{x}_0\|_2^2)\right)}{\partial \mathbf{x}^{l\mathbf{e}_k}} \Big|_{\mathbf{x}=\mathbf{x}_0} \\ &\stackrel{(A.9)}{=} \sum_{l \in \mathbb{Z}_+} \frac{m_l \sigma^l}{l!} \text{H}_l\left(\frac{y_k - x_{0,k}}{\sigma}\right). \end{aligned} \quad (\text{A.10})$$

Case II: $|\text{supp}(\mathbf{x}_0) \cup \{k\}| = S + 1$:

We will assume without loss of generality² that $k = S + 1$ and $\text{supp}(\mathbf{x}_0) = [S]$. The proof for Case II is similar to that for Case I. However, a considerable difference is that the functions $h^{(l)}(\cdot)$ defined in (A.1) do not belong to the RKHS $\mathcal{H}(\mathcal{M}_{\text{SSNM}})$ in this case.

Consider the functions

$$v^{(l)}(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : v^{(l)}(\mathbf{x}) \triangleq \frac{1}{\sigma^l l!} x_k^l \nu_{\mathbf{x}_0}(\mathbf{x}), \quad (\text{A.11})$$

where $l \in \mathbb{Z}_+$ and $\nu_{\mathbf{x}_0}(\mathbf{x}) = \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2 + \frac{1}{\sigma} \mathbf{x}^T \mathbf{x}_0\right)$ as defined in Theorem 5.5.2.

Using the Taylor expansion $\exp(x) = \sum_{l \in \mathbb{Z}_+} \frac{x^l}{l!}$ of the exponential function [6], we obtain

$$\begin{aligned} v^{(l)}(\mathbf{x}) &= \frac{1}{\sigma^l l!} x_k^l \nu_{\mathbf{x}_0}(\mathbf{x}) = \frac{1}{\sigma^l l!} \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{2\sigma^2}\right) x_k^l \sum_{\mathbf{p} \in \mathbb{Z}_+^S} \frac{1}{\mathbf{p}!} \prod_{j \in [S]} \left(\frac{x_{0,j} x_j}{\sigma}\right)^{p_j} \\ &\stackrel{(a)}{=} \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{2\sigma^2}\right) \sum_{\mathbf{p} \in \mathcal{A}^{(l)}} \frac{1}{\mathbf{p}!} \tilde{\mathbf{x}}_0^{\mathbf{p}} \mathbf{x}^{\mathbf{p}} \\ &= \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{2\sigma^2}\right) \sum_{\mathbf{p} \in \mathcal{A}^{(l)}} \frac{\tilde{\mathbf{x}}_0^{\mathbf{p}}}{\sqrt{\mathbf{p}!}} g^{(\mathbf{p})}(\mathbf{x}), \end{aligned} \quad (\text{A.12})$$

where $\tilde{\mathbf{x}}_0 \triangleq \frac{1}{\sigma}(x_{0,1}, \dots, x_{0,S}, 1, \dots, 1)^T \in \mathbb{R}^N$, $\mathcal{A}^{(l)} \triangleq \{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S \mid p_k = l, \text{supp}(\mathbf{p}) \subseteq [S + 1]\}$, and the functions $g^{(\mathbf{p})}(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : g^{(\mathbf{p})}(\mathbf{x}) = \frac{1}{\sqrt{\mathbf{p}!}} \frac{\partial^{\mathbf{p}} R_e(\mathbf{x}, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{p}}} \Big|_{\mathbf{x}_2=0}$ as defined in (5.120) of Theorem 5.5.1). The step (a) in (A.12) follows from the fact that the function $v^{(l)}(\mathbf{x})$ is defined only on the domain \mathcal{X}_S on which every monomial $\mathbf{x}^{\mathbf{p}}$ ($\mathbf{p} \in \mathbb{Z}_+^N$) with $\|\mathbf{p}\|_0 > S$ vanishes.

By Theorem 5.5.1, the functions $g^{(\mathbf{p})}(\cdot)$ belong to $\mathcal{H}(R_e)$ and are orthonormal, i.e.,

$$\langle g^{(\mathbf{p})}(\cdot), g^{(\mathbf{p}')(\cdot)} \rangle_{\mathcal{H}(R_e)} = \delta_{\mathbf{p}, \mathbf{p}'}. \quad (\text{A.13})$$

Since the coefficient sequence of (A.12) satisfies $\frac{\tilde{\mathbf{x}}_0^{\mathbf{p}}}{\sqrt{\mathbf{p}!}} \in \ell^2(\mathbb{Z}_+^N \cap \mathcal{X}_S)$, we then have also that $v^{(l)}(\cdot) \in \mathcal{H}(R_e)$.

Based on the orthonormality of the functions $\{g^{(\mathbf{p})}(\cdot)\}_{\mathbf{p} \in \mathcal{A}^{(l)}}$, we can calculate the squared norm of $v^{(l)}(\cdot)$ as

$$\|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)}^2 = \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{\sigma^2}\right) \sum_{\mathbf{p} \in \mathcal{A}^{(l)}} \frac{\tilde{\mathbf{x}}_0^{2\mathbf{p}}}{\mathbf{p}!}$$

²Indeed, consider the minimum variance problem $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}_0)$ with observation $\mathbf{y} \in \mathbb{R}^N$ and an arbitrary support $\text{supp}(\mathbf{x}_0)$ and index $k \in [N] \setminus \text{supp}(\mathbf{x}_0)$. According to Section 2.4.1, we then have that the minimum variance problem \mathcal{M}' that arises from $\mathcal{M}_{\text{SSNM}}$ by applying an invertible permutation matrix $\mathbf{P} \in \{0, 1\}^{N \times N}$ to the observation \mathbf{y} , is completely equivalent to $\mathcal{M}_{\text{SSNM}}$. By a suitable choice of the permutation matrix \mathbf{P} , the modified minimum variance problem \mathcal{M}' coincides with $\mathcal{M}_{\text{SSNM}} = (\mathcal{E}_{\text{SSNM}}, c(\cdot), \mathbf{x}'_0)$ where $\text{supp}(\mathbf{x}'_0) = [S]$ and $k = S + 1$.

$$\begin{aligned}
&= \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{\sigma^2}\right) \sum_{\mathbf{p} \in \mathcal{A}^{(l)}} \prod_{j \in [N]} \frac{\tilde{x}_{0,j}^{2p_j}}{p_j!} \\
&\stackrel{(a)}{=} \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{\sigma^2}\right) \sum_{m \in [S]} \sum_{\mathbf{p} \in \mathcal{A}_m^{(l)}} \prod_{j \in [N]} \frac{\tilde{x}_{0,j}^{2p_j}}{p_j!}, \tag{A.14}
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{A}_m^{(l)} &\triangleq \{\mathbf{p} \in \mathbb{Z}_+^N \cap \mathcal{X}_S \mid p_n > 0 \text{ for } n \in [m-1], p_m = 0, p_k = l, \text{supp}(\mathbf{p}) \subseteq [S+1]\} \\
&= \{\mathbf{p} \in \mathbb{Z}_+^N \mid p_n > 0 \text{ for } n \in [m-1], p_m = 0, p_n \geq 0 \text{ for } n \in [S] \setminus [m], p_k = l, \\
&\quad \text{supp}(\mathbf{p}) \subseteq [S+1]\}. \tag{A.15}
\end{aligned}$$

The step (a) in (A.14) follows from the fact that the sets $\{\mathcal{A}_m^{(l)}\}_{m \in [S]}$ form a partition of the set $\mathcal{A}^{(l)}$, i.e., the sets are disjoint ($\mathcal{A}_m^{(l)} \cap \mathcal{A}_{m'}^{(l)} = \emptyset$ for $m \neq m'$) and their union equals $\mathcal{A}^{(l)}$, i.e., $\mathcal{A}^{(l)} = \bigcup_{m \in [S]} \mathcal{A}_m^{(l)}$. Further developing (A.14) yields

$$\begin{aligned}
\|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)}^2 &= \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{\sigma^2}\right) \sum_{m \in [S]} \sum_{\mathbf{p} \in \mathcal{A}_m^{(l)}} \left(\prod_{n \in [m-1]} \frac{\tilde{x}_{0,n}^{2p_n}}{p_n!} \right) \frac{\tilde{x}_{0,m}^{2p_m}}{p_m!} \left(\prod_{n' \in [S] \setminus [m]} \frac{\tilde{x}_{0,n'}^{2p_{n'}}}{p_{n'}!} \right) \frac{\tilde{x}_{0,k}^{2p_k}}{p_k!} \\
&\stackrel{p_m=0, p_k=l}{=} \frac{1}{\sigma^{2l} l!} \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{\sigma^2}\right) \sum_{m \in [S]} \left(\prod_{n \in [m-1]} \sum_{p_n \in \mathbb{N}} \frac{\tilde{x}_{0,n}^{2p_n}}{p_n!} \right) \left(\prod_{n' \in [S] \setminus [m]} \sum_{p_{n'} \in \mathbb{Z}_+} \frac{\tilde{x}_{0,n'}^{2p_{n'}}}{p_{n'}!} \right) \\
&= \frac{1}{\sigma^{2l} l!} \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{\sigma^2}\right) \sum_{m \in [S]} \prod_{n \in [m-1]} \left[\exp\left(\frac{x_{0,n}^2}{\sigma^2}\right) - 1 \right] \prod_{n' \in [S] \setminus [m]} \exp\left(\frac{x_{0,n'}^2}{\sigma^2}\right) \\
&= \frac{1}{\sigma^{2l} l!} \sum_{m \in [S]} \exp\left(-\frac{x_{0,m}^2}{\sigma^2}\right) \prod_{n \in [m-1]} \left[1 - \exp\left(-\frac{x_{0,n}^2}{\sigma^2}\right) \right]. \tag{A.16}
\end{aligned}$$

It can also be verified by the orthonormality (cf. Theorem 5.5.1) of the functions $g^{(\mathbf{p})}(\cdot)$, that the functions $\{v^{(l)}(\cdot)\}_{l \in \mathbb{Z}_+}$ are orthogonal, i.e., $\langle v^{(l)}(\cdot), v^{(l')}(\cdot) \rangle_{\mathcal{H}(R_e)} = \delta_{l,l'} \|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)}^2$.

Now consider the function

$$w_N(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : w_N(\mathbf{x}) \triangleq \sum_{l \in [N]} m_l \sigma^{2l} v^{(l)}(\mathbf{x}) + m_0 v^{(0)}(\mathbf{x}), \tag{A.17}$$

which obviously belongs to the RKHS $\mathcal{H}(R_e)$ as the functions $v^{(l)}(\cdot)$ do. In the following we will use the relation

$$\|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)}^2 = \frac{1}{\sigma^{2l} l!} \|v^{(0)}(\cdot)\|_{\mathcal{H}(R_e)}^2, \tag{A.18}$$

which can be verified easily by (A.16). Based on (A.16) and the orthogonality of the functions $\{v^{(l)}(\cdot)\}_{l \in \mathbb{Z}_+}$ we obtain

$$\begin{aligned} \|w_N(\cdot)\|_{\mathcal{H}(R_e)}^2 &= m_0^2 \|v^{(0)}(\cdot)\|_{\mathcal{H}(R_e)}^2 + \sum_{l \in [N]} m_l^2 \sigma^{4l} \|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)}^2 \\ &\stackrel{(A.18)}{=} \left(m_0^2 + \sum_{l \in [N]} \frac{m_l^2}{l!} \sigma^{2l} \right) \|v^{(0)}(\cdot)\|_{\mathcal{H}(R_e)}^2. \end{aligned} \quad (A.19)$$

If (5.150) is satisfied, we have by (A.19) that the sequence $\{w_N(\cdot)\}_{N \rightarrow \infty}$ is a Cauchy sequence and therefore converges in the RKHS $\mathcal{H}(R_e)$. Indeed, by a similar calculation as used in (A.19), we have for $N' > N$ that

$$\|w_{N'}(\cdot) - w_N(\cdot)\|_{\mathcal{H}(R_e)}^2 = \|v^{(0)}(\cdot)\|_{\mathcal{H}(R_e)}^2 \sum_{l \in [N'] \setminus [N]} \frac{m_l^2}{l!} \sigma^{2l}, \quad (A.20)$$

which must become arbitrarily small for N', N sufficiently large since otherwise the sum in (5.150) cannot be finite. Moreover we have that the sequence $\{w_N(\cdot)\}_{N \rightarrow \infty}$ converges pointwise to the function $\tilde{\gamma}(\cdot) : \mathcal{X}_S \rightarrow \mathbb{R} : \tilde{\gamma}(\mathbf{x}) \triangleq \gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}) = \mathbf{K}_e[\gamma(\cdot)]$. Here, we denote by $\mathbf{K}_e[\cdot] : \mathcal{H}(\mathcal{M}_{\text{SSNM}}) \rightarrow \mathcal{H}(R_e)$ the congruence defined by (5.119) in Theorem 5.5.1. This pointwise convergence can be seen from

$$\begin{aligned} \lim_{N \rightarrow \infty} w_N(\mathbf{x}) &\stackrel{(A.17)}{=} \lim_{N \rightarrow \infty} \sum_{l \in [N]} m_l \sigma^{2l} v^{(l)}(\mathbf{x}) + m_0 v^{(0)}(\mathbf{x}) \\ &\stackrel{(A.11)}{=} \lim_{N \rightarrow \infty} \sum_{l \in [N]} m_l \sigma^{2l} \frac{1}{\sigma^l l!} x_k^l \nu_{\mathbf{x}_0}(\mathbf{x}) + m_0 \nu_{\mathbf{x}_0}(\mathbf{x}) \\ &= \nu_{\mathbf{x}_0}(\mathbf{x}) \sum_{l \in \mathbb{Z}_+} \frac{m_l \sigma^l}{l!} x_k^l \\ &\stackrel{(5.149)}{=} \gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}). \end{aligned} \quad (A.21)$$

Therefore, we have by Theorem 3.3.3 that

$$\lim_{N \rightarrow \infty} w_N(\cdot) = \tilde{\gamma}(\cdot) \quad (A.22)$$

and therefore $\tilde{\gamma}(\cdot) \in \mathcal{H}(R_e)$. However, since $\tilde{\gamma}(\cdot)$ is the image of the prescribed mean function $\gamma(\cdot)$ under the congruence $\mathbf{K}_e[\cdot] : \mathcal{H}(\mathcal{M}_{\text{SSNM}}) \rightarrow \mathcal{H}(R_e)$, we have that $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SSNM}})$, i.e., that the prescribed bias function $c(\cdot)$ is valid.

We then have also by Theorem 4.3.4 and Theorem 5.5.1 that

$$\begin{aligned}
L_{\mathcal{M}_{\text{SSNM}}} &\stackrel{(4.19)}{=} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2 - [\gamma(\mathbf{x}_0)]^2 \\
&\stackrel{(5.119)}{=} \|\gamma(\sigma\mathbf{x})\nu_{\mathbf{x}_0}(\mathbf{x})\|_{\mathcal{H}(R_e)}^2 - [\gamma(\mathbf{x}_0)]^2 \\
&= \|\tilde{\gamma}(\cdot)\|_{\mathcal{H}(R_e)}^2 - (\gamma(\mathbf{x}_0))^2 \\
&\stackrel{(A.19),(A.22)}{=} \lim_{N \rightarrow \infty} \left(m_0^2 + \sum_{l \in [N]} \frac{m_l^2}{l!} \sigma^{2l} \right) \|v^{(0)}(\cdot)\|_{\mathcal{H}(R_e)}^2 - [\gamma(\mathbf{x}_0)]^2 \\
&\stackrel{(A.16)}{=} \left(\sum_{l \in \mathbb{Z}_+} \frac{m_l^2}{l!} \sigma^{2l} \right) \sum_{m \in [S]} \exp\left(-\frac{x_{0,m}^2}{\sigma^2}\right) \prod_{n \in [m-1]} \left[1 - \exp\left(-\frac{x_{0,n}^2}{\sigma^2}\right) \right] - [\gamma(\mathbf{x}_0)]^2,
\end{aligned} \tag{A.23}$$

which is equal to (5.153).

In order to verify the expression (5.154) for the LMV estimator, we first consider the function

$$r^{(l)}(\cdot) \triangleq \mathbf{K}_e^{-1}[v^{(l)}(\cdot)]. \tag{A.24}$$

From (A.12), (5.119), and (5.120) it follows that

$$\begin{aligned}
r^{(l)}(\mathbf{x}) &= \mathbf{K}_e^{-1}[v^{(l)}(\cdot)] \\
&\stackrel{(A.12)}{=} \mathbf{K}_e^{-1} \left[\exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{2\sigma^2}\right) \sum_{\mathbf{p} \in \mathcal{A}^{(l)}} \frac{\tilde{\mathbf{x}}_0^{\mathbf{p}}}{\sqrt{\mathbf{p}!}} g^{(\mathbf{p})}(\mathbf{x}) \right] \\
&\stackrel{(5.119)}{=} \frac{1}{\nu_{\mathbf{x}_0}\left(\frac{\mathbf{x}}{\sigma}\right)} \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{2\sigma^2}\right) \sum_{\mathbf{p} \in \mathcal{A}^{(l)}} \frac{\tilde{\mathbf{x}}_0^{\mathbf{p}}}{\mathbf{p}!} g^{(\mathbf{p})}\left(\frac{\mathbf{x}}{\sigma}\right) \\
&\stackrel{(5.120)}{=} \frac{1}{\nu_{\mathbf{x}_0}\left(\frac{\mathbf{x}}{\sigma}\right)} \exp\left(-\frac{\|\mathbf{x}_0\|_2^2}{2\sigma^2}\right) \sum_{\mathbf{p} \in \mathcal{A}^{(l)}} \frac{\tilde{\mathbf{x}}_0^{\mathbf{p}}}{\mathbf{p}!} \frac{\partial^{\mathbf{p}} R_e\left(\frac{\mathbf{x}}{\sigma}, \mathbf{x}_2\right)}{\partial \mathbf{x}_2^{\mathbf{p}}} \Big|_{\mathbf{x}_2=0} \\
&\stackrel{(5.118)}{=} \exp\left(-\frac{1}{\sigma^2} \mathbf{x}^T \mathbf{x}_0\right) \sum_{\mathbf{p} \in \mathcal{A}^{(l)}} \frac{\tilde{\mathbf{x}}_0^{\mathbf{p}}}{\mathbf{p}!} \frac{\partial^{\mathbf{p}} \exp\left(\frac{1}{\sigma} \mathbf{x}^T \mathbf{x}_2\right)}{\partial \mathbf{x}_2^{\mathbf{p}}} \Big|_{\mathbf{x}_2=0} \\
&\stackrel{(5.117)}{=} \sum_{\mathbf{p} \in \mathcal{A}^{(l)}} \frac{\tilde{\mathbf{x}}_0^{\mathbf{p}}}{\mathbf{p}!} \frac{\partial^{\mathbf{p}} [R_{\mathcal{M}_{\text{SSNM}}}(\mathbf{x}, \sigma\mathbf{x}_2) \exp\left(-\frac{1}{\sigma^2} \|\mathbf{x}_0\|_2^2 + \frac{1}{\sigma} \mathbf{x}_2^T \mathbf{x}_0\right)]}{\partial \mathbf{x}_2^{\mathbf{p}}} \Big|_{\mathbf{x}_2=0}.
\end{aligned} \tag{A.25}$$

The image of $r^{(l)}(\cdot)$ under the congruence $J[\cdot] : \mathcal{H}(\mathcal{M}_{\text{SSNM}}) \rightarrow \mathcal{L}(\mathcal{M}_{\text{SSNM}})$ as defined in Theorem 4.3.2 can be calculated due to the expansion (A.25) and Theorem 4.3.3 as

$$\begin{aligned} J[r^{(l)}(\cdot)](\mathbf{y}) &\stackrel{(4.16)}{=} \sum_{\mathbf{p} \in \mathcal{A}^{(l)}} \frac{\tilde{\mathbf{x}}_0^{\mathbf{p}}}{\mathbf{p}!} \frac{\partial^{\mathbf{p}} [\rho_{\mathcal{M}_{\text{SSNM}}}(\mathbf{y}, \sigma \mathbf{x}_2) \exp(-\frac{1}{\sigma^2} \|\mathbf{x}_0\|_2^2 + \frac{1}{\sigma} \mathbf{x}_2^T \mathbf{x}_0)]}{\partial \mathbf{x}_2^{\mathbf{p}}} \Big|_{\mathbf{x}_2=0} \\ &\stackrel{(5.145)}{=} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2\right) \sum_{\mathbf{p} \in \mathcal{A}^{(l)}} \frac{\tilde{\mathbf{x}}_0^{\mathbf{p}}}{\mathbf{p}!} \frac{\partial^{\mathbf{p}} \exp\left(\frac{1}{\sigma^2} \mathbf{y}^T (\sigma \mathbf{x}_2 - \mathbf{x}_0) + \frac{1}{\sigma} \mathbf{x}_2^T \mathbf{x}_0 - \frac{1}{2} \|\mathbf{x}_2\|_2^2\right)}{\partial \mathbf{x}_2^{\mathbf{p}}} \Big|_{\mathbf{x}_2=0}. \end{aligned} \quad (\text{A.26})$$

Using the partition $\mathcal{A}^{(l)} = \bigcup_{m \in [S]} \mathcal{A}_m^{(l)}$, this becomes

$$\begin{aligned} J[r^{(l)}(\cdot)](\mathbf{y}) &= \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2\right) \sum_{m \in [S]} \sum_{\mathbf{p} \in \mathcal{A}_m^{(l)}} \frac{\tilde{\mathbf{x}}_0^{\mathbf{p}}}{\mathbf{p}!} \frac{\partial^{\mathbf{p}} \exp\left(\frac{1}{\sigma^2} \mathbf{y}^T (\sigma \mathbf{x}_2 - \mathbf{x}_0) + \frac{1}{\sigma} \mathbf{x}_2^T \mathbf{x}_0 - \frac{1}{2} \|\mathbf{x}_2\|_2^2\right)}{\partial \mathbf{x}_2^{\mathbf{p}}} \Big|_{\mathbf{x}_2=0} \\ &= \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2\right) \\ &\quad \times \sum_{m \in [S]} \sum_{\mathbf{p} \in \mathcal{A}_m^{(l)}} \prod_{n \in [N]} \frac{\tilde{x}_{0,n}^{p_n}}{p_n!} \frac{\partial^{p_n} \exp\left(\frac{1}{\sigma^2} y_n (\sigma x_{2,n} - x_{0,n}) + \frac{1}{\sigma} x_{2,n} x_{0,n} - \frac{1}{2} x_{2,n}^2\right)}{\partial x_{2,n}^{p_n}} \Big|_{x_{2,n}=0} \\ &= \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2\right) \sum_{m \in [S]} \sum_{\mathbf{p} \in \mathcal{A}_m^{(l)}} \frac{\tilde{x}_{0,m}^{p_m}}{p_m!} \frac{\partial^{p_m} \exp\left(\frac{1}{\sigma^2} y_m (\sigma x_{2,m} - x_{0,m}) + \frac{1}{\sigma} x_{2,m} x_{0,m} - \frac{1}{2} x_{2,m}^2\right)}{\partial x_{2,m}^{p_m}} \Big|_{x_{2,m}=0} \\ &\quad \times \left[\prod_{n \in [m-1]} \sum_{p_n \in \mathbb{N}} \frac{\tilde{x}_{0,n}^{p_n}}{p_n!} \frac{\partial^{p_n} \exp\left(\frac{1}{\sigma^2} y_n (\sigma x_{2,n} - x_{0,n}) + \frac{1}{\sigma} x_{2,n} x_{0,n} - \frac{1}{2} x_{2,n}^2\right)}{\partial x_{2,n}^{p_n}} \Big|_{x_{2,n}=0} \right] \\ &\quad \times \left[\prod_{n' \in [S] \setminus [m]} \sum_{p_{n'} \in \mathbb{Z}_+} \frac{\tilde{x}_{0,n'}^{p_{n'}}}{p_{n'}!} \frac{\partial^{p_{n'}} \exp\left(\frac{1}{\sigma^2} y_{n'} (\sigma x_{2,n'} - x_{0,n'}) + \frac{1}{\sigma} x_{2,n'} x_{0,n'} - \frac{1}{2} x_{2,n'}^2\right)}{\partial x_{2,n'}^{p_{n'}}} \Big|_{x_{2,n'}=0} \right] \\ &\quad \times \frac{\tilde{x}_{0,k}^{p_k}}{p_k!} \frac{\partial^{p_k} \exp\left(\frac{1}{\sigma^2} y_k (\sigma x_{2,k} - x_{0,k}) + \frac{1}{\sigma} x_{2,k} x_{0,k} - \frac{1}{2} x_{2,k}^2\right)}{\partial x_{2,k}^{p_k}} \Big|_{x_{2,k}=0} \\ &= \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2\right) \sum_{m \in [S]} \exp\left(-\frac{1}{\sigma^2} y_m x_{0,m}\right) \\ &\quad \times \left[\prod_{n \in [m-1]} \sum_{p_n \in \mathbb{N}} \frac{\tilde{x}_{0,n}^{p_n}}{p_n!} \frac{\partial^{p_n} \exp\left(\frac{1}{\sigma^2} y_n (\sigma x_{2,n} - x_{0,n}) + \frac{1}{\sigma} x_{2,n} x_{0,n} - \frac{1}{2} x_{2,n}^2\right)}{\partial x_{2,n}^{p_n}} \Big|_{x_{2,n}=0} \right] \end{aligned}$$

$$\begin{aligned}
& \times \left[\prod_{n' \in [S] \setminus [m]} \sum_{p_{n'} \in \mathbb{Z}_+} \frac{\tilde{x}_{0,n'}^{p_{n'}}}{p_{n'}!} \frac{\partial^{p_{n'}} \exp \left(\frac{1}{\sigma^2} y_{n'} (\sigma x_{2,n'} - x_{0,n'}) + \frac{1}{\sigma} x_{2,n'} x_{0,n'} - \frac{1}{2} x_{2,n'}^2 \right)}{\partial x_{2,n'}^{p_{n'}}} \Big|_{x_{2,n'}=0} \right] \\
& \times \frac{1}{l! \sigma^l} \frac{\partial^l \exp \left(\frac{1}{\sigma} y_k x_{2,k} - \frac{1}{2} x_{2,k}^2 \right)}{\partial x_{2,k}^l} \Big|_{x_{2,k}=0}. \tag{A.27}
\end{aligned}$$

Using the Taylor expansion of the exponential function [6], we obtain further

$$\begin{aligned}
\mathbf{J}[r^{(l)}(\cdot)](\mathbf{y}) &= \exp \left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2 \right) \sum_{m \in [S]} \exp \left(-\frac{1}{\sigma^2} y_m x_{0,m} \right) \\
& \times \prod_{n \in [m-1]} \left[\exp \left(\frac{1}{2\sigma^2} x_{0,n}^2 \right) - \exp \left(-\frac{1}{\sigma^2} y_n x_{0,n} \right) \right] \prod_{n' \in [S] \setminus [m]} \exp \left(\frac{1}{2\sigma^2} x_{0,n'}^2 \right) \\
& \times \frac{1}{l! \sigma^l} \frac{\partial^l \exp \left(\frac{1}{\sigma} y_k x_{2,k} - \frac{1}{2} x_{2,k}^2 \right)}{\partial x_{2,k}^l} \Big|_{x_{2,k}=0}. \tag{A.28}
\end{aligned}$$

Using the identity

$$\begin{aligned}
\frac{\partial^l \exp \left(\frac{1}{\sigma} y x - \frac{1}{2} x^2 \right)}{\partial x^l} \Big|_{x=0} &= \frac{\partial^l \exp \left(-\frac{1}{2} \left(x - \frac{y}{\sigma} \right)^2 + \frac{1}{2} \frac{y^2}{\sigma^2} \right)}{\partial x^l} \Big|_{x=0} \\
& \left(x' \triangleq \frac{y}{\sigma} - x \right) \\
& = \exp \left(\frac{1}{2} \frac{y^2}{\sigma^2} \right) \frac{\partial^l \exp \left(-\frac{1}{2} x'^2 \right)}{\partial x'^l} \Big|_{x'=-\frac{y}{\sigma}} = \mathbf{H}_l \left(\frac{y}{\sigma} \right), \tag{A.29}
\end{aligned}$$

we finally have that

$$\begin{aligned}
\mathbf{J}[r^{(l)}(\cdot)](\mathbf{y}) &\stackrel{(A.28)}{=} \exp \left(-\frac{1}{2\sigma^2} \|\mathbf{x}_0\|_2^2 \right) \sum_{m \in [S]} \exp \left(-\frac{1}{\sigma^2} y_m x_{0,m} \right) \\
& \times \prod_{n \in [m-1]} \left[\exp \left(\frac{1}{2\sigma^2} x_{0,n}^2 \right) - \exp \left(-\frac{1}{\sigma^2} y_n x_{0,n} \right) \right] \prod_{n' \in [S] \setminus [m]} \exp \left(\frac{1}{2\sigma^2} x_{0,n'}^2 \right) \\
& \times \frac{1}{l! \sigma^l} \mathbf{H}_l \left(\frac{y_k}{\sigma} \right) \\
& = \frac{1}{l! \sigma^l} \mathbf{H}_l \left(\frac{y_k}{\sigma} \right) \sum_{m \in [S]} \exp \left(-\frac{1}{2\sigma^2} (x_{0,m}^2 + 2y_m x_{0,m}) \right) \prod_{n \in [m-1]} \left[1 - \exp \left(-\frac{1}{2\sigma^2} (x_{0,n}^2 + 2y_n x_{0,n}) \right) \right], \tag{A.30}
\end{aligned}$$

By Theorem 4.3.4, we have that if the prescribed bias function $c(\cdot)$ is valid, i.e., $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SSNM}})$ and in turn $\tilde{\gamma}(\cdot) \in \mathcal{H}(R_e)$, the corresponding LMV estimator is given by

$$\begin{aligned}
\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y}) &\stackrel{(4.20)}{=} \mathbf{J}[\gamma(\cdot)] = \mathbf{J}[\mathbf{K}_e^{-1}[\tilde{\gamma}(\cdot)]] \stackrel{(A.22)}{=} \mathbf{J}\left[\mathbf{K}_e^{-1}\left[\lim_{N \rightarrow \infty} w_N(\cdot)\right]\right] \stackrel{(A.17)}{=} \mathbf{J}\left[\mathbf{K}_e^{-1}\left[\lim_{N \rightarrow \infty} \sum_{l \in [N] \cup \{0\}} m_l \sigma^{2l} v^{(l)}(\cdot)\right]\right] \\
&= \lim_{N \rightarrow \infty} \sum_{l \in [N] \cup \{0\}} m_l \sigma^{2l} \mathbf{J}[\mathbf{K}_e^{-1}[v^{(l)}(\cdot)]] \stackrel{(A.24)}{=} \lim_{N \rightarrow \infty} \sum_{l \in [N] \cup \{0\}} m_l \sigma^{2l} \mathbf{J}[r^{(l)}(\cdot)] \\
&\stackrel{(A.30)}{=} \sum_{l \in \mathbb{Z}_+} \frac{m_l}{l!} \sigma^l \mathbf{H}_l\left(\frac{y_k}{\sigma}\right) \sum_{m \in [S]} \exp\left(-\frac{1}{2\sigma^2}(x_{0,m}^2 + 2y_m x_{0,m})\right) \\
&\quad \times \prod_{n \in [m-1]} \left[1 - \exp\left(-\frac{1}{2\sigma^2}(x_{0,n}^2 + 2y_n x_{0,n})\right)\right], \tag{A.31}
\end{aligned}$$

which is equal to (5.154).

It remains to verify the necessity of the condition (5.150) for a bias function $c(\cdot)$ to be valid. To that end, assume that (5.150) is not fulfilled, i.e., the sums $\sum_{l \in \mathcal{T}} \frac{m_l^2}{l!} \sigma^{2l}$ can be made arbitrary large by using a suitable finite index set $\mathcal{T} \subseteq \mathbb{Z}_+$, but the bias function $c(\cdot)$ is valid, i.e., $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SSNM}})$. However, $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SSNM}})$ implies, due to the congruence $\mathbf{K}_e[\cdot]$ in (5.119) of Theorem 5.5.1, that we also have $\tilde{\gamma}(\cdot) = \gamma(\sigma \mathbf{x}) \nu_{\mathbf{x}_0}(\mathbf{x}) = \mathbf{K}_e[\gamma(\cdot)] \in \mathcal{H}(R_e)$. By Theorem 3.1.6, we then have for any subspace $\mathcal{U} \triangleq \text{span}\{v^{(l)}(\cdot)\}_{l \in \mathcal{T}}$ that³

$$\begin{aligned}
\|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2 &= \|\mathbf{K}_e[\gamma(\cdot)]\|_{\mathcal{H}(R_e)}^2 = \|\tilde{\gamma}(\cdot)\|_{\mathcal{H}(R_e)}^2 \stackrel{(3.31)}{\geq} \|\mathbf{P}_{\mathcal{U}} \tilde{\gamma}(\cdot)\|_{\mathcal{H}(R_e)}^2 \\
&\stackrel{(3.33)}{=} \sum_{l \in \mathcal{T}} \left[\left\langle \tilde{\gamma}(\cdot), v^{(l)}(\cdot) / \|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)} \right\rangle_{\mathcal{H}(R_e)} \right]^2 \\
&\stackrel{(A.22)}{=} \sum_{l \in \mathcal{T}} \left[\left\langle \lim_{N \rightarrow \infty} w_N(\cdot), v^{(l)}(\cdot) / \|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)} \right\rangle_{\mathcal{H}(R_e)} \right]^2 \\
&\stackrel{(a)}{=} \sum_{l \in \mathcal{T}} \left[\lim_{N \rightarrow \infty} \left\langle w_N(\cdot), v^{(l)}(\cdot) / \|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)} \right\rangle_{\mathcal{H}(R_e)} \right]^2 \\
&\stackrel{(A.17)}{=} \sum_{l \in \mathcal{T}} \left[\frac{m_l \sigma^{2l}}{\|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)}} \left\langle v^{(l)}(\cdot), v^{(l)}(\cdot) \right\rangle_{\mathcal{H}(R_e)} \right]^2
\end{aligned}$$

³An ONB for \mathcal{U} is given by $\{v^{(l)}(\cdot) / \|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)}\}_{l \in \mathcal{T}}$ since the functions $v^{(l)}(\cdot)$ are orthogonal.

$$\begin{aligned}
&= \sum_{l \in \mathcal{T}} \left[m_l \sigma^{2l} \|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)} \right]^2 \\
&\stackrel{(A.18)}{=} \sum_{l \in \mathcal{T}} \left[\frac{m_l \sigma^l}{\sqrt{l!}} \|v^{(0)}(\cdot)\|_{\mathcal{H}(R_e)} \right]^2 = \sum_{l \in \mathcal{T}} \frac{m_l^2}{l!} \sigma^{2l} \|v^{(0)}(\cdot)\|_{\mathcal{H}(R_e)}^2, \tag{A.32}
\end{aligned}$$

where in (a) the change of the order of taking the limit $\lim_{N \rightarrow \infty}$ and computing the inner product with $v^{(l)}(\cdot)/\|v^{(l)}(\cdot)\|_{\mathcal{H}(R_e)}$, respectively, is due to the fact that the inner product is a continuous mapping (see, e.g., [6, Theorem 4.6]). However, since (5.150) is not satisfied, it follows from (A.32) that the squared norm $\|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\text{SSNM}})}^2$ is unbounded, which is impossible because $\gamma(\cdot) \in \mathcal{H}(\mathcal{M}_{\text{SSNM}})$. \square

Appendix B

Proof of Theorem 6.3.4

Proof. The statement is derived by considering the minimum variance problem $\mathcal{M}_{\mathcal{D},\text{SPCM}} = \mathcal{M}_{\text{SPCM}}|_{\mathcal{D}}$, where the set \mathcal{D} is chosen as

$$\mathcal{D} = (\mathcal{B}(\mathbf{0}, r) \cup \mathcal{B}(\mathbf{x}_0, r)) \cap \{\mathbf{x} \in \mathcal{X}_{S,+} \mid \text{supp}(\mathbf{x}) \subseteq \mathcal{K}\}, \quad (\text{B.1})$$

where

$$\mathcal{K} \triangleq \{\{l\} \cup \text{supp}(\mathbf{x}_0)\} = \{i_1, \dots, i_K\} \quad (\text{B.2})$$

with $K \leq S + 1$ and an arbitrary but (for the rest of the proof) fixed index $l \in [N] \setminus \text{supp}(\mathbf{x}_0)$. Furthermore, the radius r is sufficiently small such that Theorem 6.2.3 applies for the two sets $\mathcal{B}(\mathbf{0}, r) \cap \mathcal{X}_{S,+}$ and $\mathcal{B}(\mathbf{x}_0, r) \cap \mathcal{X}_{S,+}$ separately. Therefore, the condition (6.19) is also satisfied for their union $(\mathcal{B}(\mathbf{0}, r) \cup \mathcal{B}(\mathbf{x}_0, r)) \cap \mathcal{X}_{S,+}$ and in turn also for the set \mathcal{D} given in (B.1), since this set is contained in the union. Thus, the set \mathcal{D} given in (B.1) satisfies (6.19) and therefore the RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})$ exists. Since we can assume that $g(\cdot)$ is estimable for $\mathcal{M}_{\text{SDPCM}}$, which implies via Theorem 2.5.2 that the restriction $g(\cdot)|_{\mathcal{D}}$ is estimable for $\mathcal{M}_{\mathcal{D},\text{SDPCM}}$, we have that the prescribed mean function $\gamma(\cdot) : \mathcal{D} \rightarrow \mathbb{R} : \gamma(\mathbf{x}) = c(\mathbf{x}) + g(\mathbf{x}) = g(\mathbf{x})$ belongs to the RKHS $\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})$.

The outline of the rest of the proof is as follows: We will construct a subspace

$$\mathcal{U} \triangleq \text{span}\{w_0(\cdot), w_1(\cdot)\} \subseteq \mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}}) \quad (\text{B.3})$$

that is spanned by two orthogonal functions $w_0(\cdot)$ and $w_1(\cdot)$, respectively. We can then invoke Theorem 4.3.4 and Theorem 3.1.6 to lower bound the minimum achievable variance $L_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}$ by projecting the mean function $\gamma(\cdot)$ on the subspace \mathcal{U} . However, we will not compute the squared norm of the projection $\mathbf{P}_{\mathcal{U}}\gamma(\cdot)$ explicitly, but lower bound it by deriving an upper bound on the squared norm of the function $w_1(\cdot)$. The so obtained lower bound on $L_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}$ yields then via (6.17) a lower bound on $L_{\mathcal{M}_{\text{SPCM}}}$.

Consider the matrix $\mathbf{H}^{\mathcal{K}} \triangleq (\mathbf{H}_{i_1}, \dots, \mathbf{H}_{i_K}) \in \mathbb{R}^{M \times Q}$, where $Q \triangleq \sum_{k \in \mathcal{K}} r_k$, and its thin SVD

$$\mathbf{H}^{\mathcal{K}} = \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^T. \quad (\text{B.4})$$

Due to the assumption that the basis matrices $\{\mathbf{C}_k\}_{k \in [N]}$ satisfy the RIP of order $S + 1$, the diagonal entries of the diagonal matrix $\boldsymbol{\Sigma}_1 \in \mathbb{R}^{Q \times Q}$, i.e., the singular values of $\mathbf{H}^{\mathcal{K}}$, belong to the interval $[1 - \delta_{S+1}, 1 + \delta_{S+1}]$. This also implies that the matrix $\mathbf{H}^{\mathcal{K}}$ has full column rank, i.e., $\text{rank}(\mathbf{H}^{\mathcal{K}}) = Q$, and that the matrix $\boldsymbol{\Sigma}_1 \mathbf{V}_1^T \in \mathbb{R}^{Q \times Q}$ appearing in the thin SVD (B.4) is invertible. For every $\mathbf{x} \in \mathcal{D}$, we have due to (B.4) that

$$\begin{aligned} \tilde{\mathbf{C}}(\mathbf{x}) &\stackrel{(6.4), (6.10)}{=} \mathbf{H} \mathbf{D}^{\mathcal{K}}(\mathbf{x}) \mathbf{H}^T + \sigma^2 \mathbf{I} = \sum_{k \in \mathcal{K}} \mathbf{H}_k x_k \mathbf{H}_k^T + \sigma^2 \mathbf{I} \\ &= \mathbf{H}^{\mathcal{K}} \mathbf{D}^{\mathcal{K}}(\mathbf{x}) (\mathbf{H}^{\mathcal{K}})^T + \sigma^2 \mathbf{I} \\ &= \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^T \mathbf{D}^{\mathcal{K}}(\mathbf{x}) \mathbf{V}_1 \boldsymbol{\Sigma}_1 \mathbf{U}_1^T + \sigma^2 \mathbf{P}^{\mathcal{K}} + \sigma^2 (\mathbf{I} - \mathbf{P}^{\mathcal{K}}) \\ &= \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^T \mathbf{D}^{\mathcal{K}}(\mathbf{x}) \mathbf{V}_1 \boldsymbol{\Sigma}_1 \mathbf{U}_1^T + \sigma^2 \mathbf{U}_1 \mathbf{U}_1^T + \sigma^2 (\mathbf{I} - \mathbf{P}^{\mathcal{K}}) \\ &= \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^T \mathbf{C}^{\mathcal{K}}(\mathbf{x}) \mathbf{V}_1 \boldsymbol{\Sigma}_1 \mathbf{U}_1^T + (\mathbf{I} - \mathbf{P}^{\mathcal{K}}) \sigma^2. \end{aligned} \quad (\text{B.5})$$

Here, we used

$$\mathbf{D}^{\mathcal{K}}(\mathbf{x}) \triangleq \begin{pmatrix} x_{i_1} \mathbf{I}_{r_{i_1}} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & x_{i_2} \mathbf{I}_{r_{i_2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \mathbf{0} \\ \mathbf{0} & \dots & \mathbf{0} & x_{i_K} \mathbf{I}_{r_{i_K}} \end{pmatrix}, \quad (\text{B.6})$$

$$\mathbf{C}^{\mathcal{K}}(\mathbf{x}) \triangleq \mathbf{D}^{\mathcal{K}}(\mathbf{x}) + \sigma^2 \mathbf{V}_1 \boldsymbol{\Sigma}_1^{-2} \mathbf{V}_1^T = \mathbf{D}^{\mathcal{K}}(\mathbf{x}) + \sigma^2 \mathbf{I} + \mathbf{E}_1, \quad (\text{B.7})$$

$$\mathbf{E}_1 \triangleq \sigma^2 (\mathbf{V}_1 \boldsymbol{\Sigma}_1^{-2} \mathbf{V}_1^T - \mathbf{I}), \quad (\text{B.8})$$

and

$$\mathbf{P}^{\mathcal{K}} \triangleq \mathbf{H}^{\mathcal{K}} (\mathbf{H}^{\mathcal{K}})^{\dagger} = \mathbf{U}_1 \boldsymbol{\Sigma}_1 \mathbf{V}_1^T \mathbf{V}_1 \boldsymbol{\Sigma}_1^{-1} \mathbf{U}_1^T = \mathbf{U}_1 \mathbf{U}_1^T \quad (\text{B.9})$$

denotes the orthogonal projection matrix on the subspace $\text{span}(\mathbf{H}^{\mathcal{K}}) \subseteq \mathbb{R}^M$. Note that

$$(\mathbf{I} - \mathbf{P}^{\mathcal{K}}) \mathbf{U}_1 = \mathbf{U}_1 - \mathbf{U}_1 \mathbf{U}_1^T \mathbf{U}_1 = \mathbf{U}_1 - \mathbf{U}_1 = \mathbf{0}. \quad (\text{B.10})$$

Since, by (B.5), the matrix $\tilde{\mathbf{C}}(\mathbf{x})$ is the sum of two matrices which act (viewed as a linear transformation) independently on the two complementary subsets $\text{span}(\mathbf{U}_1)$, $\text{span}(\mathbf{I} - \mathbf{P}^{\mathcal{K}}) \subseteq \mathbb{R}^M$ we

have by [4, 38] that

$$\det \{\tilde{\mathbf{C}}(\mathbf{x})\} = \det \{\boldsymbol{\Sigma}_1 \mathbf{V}_1^T \mathbf{C}^{\mathcal{K}}(\mathbf{x}) \mathbf{V}_1 \boldsymbol{\Sigma}_1\} \sigma^{2(M-Q)}. \quad (\text{B.11})$$

The 2-norm of the matrix \mathbf{E}_1 can be bounded as

$$\|\mathbf{E}_1\|_2 \stackrel{(a)}{\leq} \sigma^2 \frac{4\delta_{S+1}}{1-2\delta_{S+1}} \stackrel{(b)}{\leq} \sigma^2 5\delta_{S+1}. \quad (\text{B.12})$$

Here, step (b) follows from $\delta_{S+1} < 1/32$. In order to verify step (a) in (B.12), note that for an arbitrary vector $\mathbf{x} \in \mathbb{R}^N$ with $\|\mathbf{x}\|_2 = 1$, we obtain (note that the diagonal elements of $\boldsymbol{\Sigma}_1^{-2}$ belong to the interval $[(1+\delta_{S+1})^{-2}, (1-\delta_{S+1})^{-2}]$)

$$\mathbf{x}^T \mathbf{E}_1 \mathbf{x} = \sigma^2 \left(\underbrace{\mathbf{x}^T \mathbf{V}_1 \boldsymbol{\Sigma}_1^{-2} \mathbf{V}_1^T \mathbf{x}}_{\in [(1+\delta_{S+1})^{-2}, (1-\delta_{S+1})^{-2}]} - \underbrace{\mathbf{x}^T \mathbf{x}}_{=1} \right) \in \sigma^2 [(1+\delta_{S+1})^{-2} - 1, (1-\delta_{S+1})^{-2} - 1]. \quad (\text{B.13})$$

From (B.13) it follows via the EVD of the symmetric matrix \mathbf{E}_1 [38] that the eigenvalues $\lambda_k(\mathbf{E}_1)$ belong to the interval $\sigma^2 [(1+\delta_{S+1})^{-2} - 1, (1-\delta_{S+1})^{-2} - 1]$, i.e.,

$$\lambda_k(\mathbf{E}_1) \in \sigma^2 [(1+\delta_{S+1})^{-2} - 1, (1-\delta_{S+1})^{-2} - 1]. \quad (\text{B.14})$$

This further implies that

$$\begin{aligned} \|\mathbf{E}_1\|_2 &= \max \{ |\lambda_Q(\mathbf{E}_1)|, |\lambda_1(\mathbf{E}_1)| \} \leq |\lambda_Q(\mathbf{E}_1)| + |\lambda_1(\mathbf{E}_1)| \\ &\stackrel{(\text{B.14})}{\leq} \sigma^2 [(1-\delta_{S+1})^{-2} - 1] - \sigma^2 [(1+\delta_{S+1})^{-2} - 1] \\ &= \sigma^2 [(1-\delta_{S+1})^{-2} - (1+\delta_{S+1})^{-2}] = \sigma^2 \frac{(1+\delta_{S+1})^2 - (1-\delta_{S+1})^2}{(1-\delta_{S+1})^2 (1+\delta_{S+1})^2} \\ &= \sigma^2 \frac{4\delta_{S+1}}{(1-\delta_{S+1})^2 (1+\delta_{S+1})^2} \leq \sigma^2 \frac{4\delta_{S+1}}{(1-\delta_{S+1})^2} = \sigma^2 \frac{4\delta_{S+1}}{1-2\delta_{S+1} + \delta_{S+1}^2} \\ &\leq \sigma^2 \frac{4\delta_{S+1}}{1-2\delta_{S+1}}, \end{aligned} \quad (\text{B.15})$$

where step (a) follows from the fact that \mathbf{E}_1 is symmetric, i.e., $\mathbf{E}_1^T = \mathbf{E}_1$ (cf. [38]). It is important to note that the matrix \mathbf{E}_1 appearing in (B.7) does not depend on \mathbf{x} .

From (B.5), we have via elementary linear algebra [38] that the kernel $R_{\mathcal{M}_{\mathcal{D}, \text{SPCM}}}(\cdot, \cdot)$ in (6.38) can be rewritten as

$$R_{\mathcal{M}_{\mathcal{D}, \text{SPCM}}}(\mathbf{x}_1, \mathbf{x}_2) = [\det \{\tilde{\mathbf{C}}(\mathbf{x}_0)\}]^{1/2} [\det \{\tilde{\mathbf{C}}(\mathbf{x}_1) + \tilde{\mathbf{C}}(\mathbf{x}_2) - \tilde{\mathbf{C}}(\mathbf{x}_1) \tilde{\mathbf{C}}^{-1}(\mathbf{x}_0) \tilde{\mathbf{C}}(\mathbf{x}_2)\}]^{-1/2}$$

$$\begin{aligned}
& \stackrel{(B.5),(B.11)}{=} [\det \{ \boldsymbol{\Sigma}_1 \mathbf{V}_1^T \mathbf{C}^{\mathcal{K}}(\mathbf{x}_0) \mathbf{V}_1 \boldsymbol{\Sigma}_1 \}]^{1/2} \times \\
& \quad [\det \{ \boldsymbol{\Sigma}_1 \mathbf{V}_1^T (\mathbf{C}^{\mathcal{K}}(\mathbf{x}_1) + \mathbf{C}^{\mathcal{K}}(\mathbf{x}_2) - \mathbf{C}^{\mathcal{K}}(\mathbf{x}_1) [\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1} \mathbf{C}^{\mathcal{K}}(\mathbf{x}_2)) \mathbf{V}_1 \boldsymbol{\Sigma}_1 \}]^{-1/2} \\
& \stackrel{(a)}{=} [\det \{ \mathbf{C}^{\mathcal{K}}(\mathbf{x}_0) \}]^{1/2} [\det \{ \mathbf{C}^{\mathcal{K}}(\mathbf{x}_1) + \mathbf{C}^{\mathcal{K}}(\mathbf{x}_2) - \mathbf{C}^{\mathcal{K}}(\mathbf{x}_1) [\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1} \mathbf{C}^{\mathcal{K}}(\mathbf{x}_2) \}]^{-1/2},
\end{aligned} \tag{B.16}$$

where (a) is due to the identity $\det \{ \mathbf{M} \mathbf{N} \} = \det \{ \mathbf{M} \} \det \{ \mathbf{N} \}$ for two square matrices \mathbf{M}, \mathbf{N} [4].

Let us now define the subspace $\mathcal{U} \triangleq \text{span}\{w_0(\cdot), w_1(\cdot)\} \subseteq \mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})$ spanned by the two functions $w_0(\cdot) \triangleq R_{\mathcal{M}_{\mathcal{D}, \text{SPCM}}}(\cdot, \mathbf{x}_0) \in \mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})$ and

$$w_1(\cdot) \triangleq \left. \frac{\partial^{\mathbf{e}_l} R_{\mathcal{M}_{\mathcal{D}, \text{SPCM}}}(\cdot, \mathbf{x}_2)}{\partial \mathbf{x}_2^{\mathbf{e}_l}} \right|_{\mathbf{x}_2=\mathbf{0}}, \tag{B.17}$$

where $l \in [N] \setminus \text{supp}(\mathbf{x}_0)$ is the specific index which is used in the definition of the set \mathcal{K} in (B.2). By Theorem 3.4.1 (using the function $g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)$ with $\mathbf{p} = \mathbf{e}_l$ and $\mathbf{x}_c = \mathbf{0}$), we have that $w_1(\cdot) \in \mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})$ and moreover

$$\langle w_0(\cdot), w_1(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})} \stackrel{(3.66)}{=} \left. \frac{\partial^{\mathbf{e}_l} R_{\mathcal{M}_{\mathcal{D}, \text{SPCM}}}(\mathbf{x}_2, \mathbf{x}_0)}{\partial \mathbf{x}_2^{\mathbf{e}_l}} \right|_{\mathbf{x}_2=\mathbf{0}} \stackrel{(4.11)}{=} \left. \frac{\partial^{\mathbf{e}_l} 1}{\partial \mathbf{x}_2^{\mathbf{e}_l}} \right|_{\mathbf{x}_2=\mathbf{0}} = 0. \tag{B.18}$$

Therefore, an ONB for \mathcal{U} is given by $\left\{ \frac{w_0(\cdot)}{\|w_0(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}}, \frac{w_1(\cdot)}{\|w_1(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}} \right\}$.

By projecting the prescribed mean function $\gamma(\mathbf{x}) = g(\mathbf{x}) + c(\mathbf{x}) = g(\mathbf{x})$ of $\mathcal{M}_{\mathcal{D}, \text{SPCM}}$ onto the subspace \mathcal{U} , we obtain

$$\begin{aligned}
L_{\mathcal{M}_{\text{SPCM}}} & \stackrel{(6.17)}{\geq} L_{\mathcal{M}_{\mathcal{D}, \text{SPCM}}} \stackrel{(4.19)}{=} \|\gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}^2 - [\gamma(\mathbf{x}_0)]^2 \stackrel{(4.21)}{\geq} \|\mathbf{P}_{\mathcal{U}} \gamma(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}^2 - [\gamma(\mathbf{x}_0)]^2 \\
& \stackrel{(a)}{=} \frac{\langle \gamma(\cdot), w_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}^2}{\langle w_0(\cdot), w_0(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}} + \frac{\langle \gamma(\cdot), w_1(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}^2}{\langle w_1(\cdot), w_1(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}} - [\gamma(\mathbf{x}_0)]^2 \\
& = \frac{\langle \gamma(\cdot), R_{\mathcal{M}_{\mathcal{D}, \text{SPCM}}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}^2}{\langle R_{\mathcal{M}_{\mathcal{D}, \text{SPCM}}}(\cdot, \mathbf{x}_0), R_{\mathcal{M}_{\mathcal{D}, \text{SPCM}}}(\cdot, \mathbf{x}_0) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}} + \frac{\langle \gamma(\cdot), w_1(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}^2}{\langle w_1(\cdot), w_1(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}} - [\gamma(\mathbf{x}_0)]^2 \\
& \stackrel{(3.44),(4.11)}{=} \frac{[\gamma(\mathbf{x}_0)]^2}{1} + \frac{\langle \gamma(\cdot), w_1(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}^2}{\langle w_1(\cdot), w_1(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}} - [\gamma(\mathbf{x}_0)]^2 \\
& \stackrel{\gamma(\cdot) \equiv g(\cdot)}{=} \frac{[g(\mathbf{x}_0)]^2}{1} + \frac{\langle g(\cdot), w_1(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}^2}{\langle w_1(\cdot), w_1(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}} - [g(\mathbf{x}_0)]^2 \\
& \stackrel{(b)}{=} b_l^2 \|w_1(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D}, \text{SPCM}})}^{-2},
\end{aligned} \tag{B.19}$$

where $b_l \triangleq \frac{\partial g(\mathbf{x})}{\partial x_l} \Big|_{\mathbf{x}=\mathbf{0}}$; step (a) follows from Theorem 3.1.9, and step (b) is due to Theorem 3.4.1 (using the function $g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)$ with $\mathbf{p}=\mathbf{e}_l$ and $\mathbf{x}_c=\mathbf{0}$).

The last and central step of the proof derives an upper bound on the squared norm $\|w_1(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})}^2$, which combined with (B.19) yields the bound in (6.54). By combining Theorem 3.4.1 (using the function $g_{\mathbf{x}_c}^{(\mathbf{p})}(\cdot)$ with $\mathbf{p}=\mathbf{e}_l$ and $\mathbf{x}_c=\mathbf{0}$), Lemma 6.3.1, and Lemma 6.3.3 we obtain

$$\begin{aligned}
\|w_1(\cdot)\|_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})}^2 &= \langle w_1(\cdot), w_1(\cdot) \rangle_{\mathcal{H}(\mathcal{M}_{\mathcal{D},\text{SPCM}})} \stackrel{(3.66)}{=} \frac{\partial^{\mathbf{e}_l} \partial^{\mathbf{e}_l} R_{\mathcal{M}_{\mathcal{D},\text{SPCM}}}(\mathbf{x}_1, \mathbf{x}_2)}{\partial \mathbf{x}_1^{\mathbf{e}_l} \partial \mathbf{x}_2^{\mathbf{e}_l}} \Big|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{0}} \\
&\stackrel{(B.16)}{=} [\det \{\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)\}]^{1/2} \frac{\partial^{\mathbf{e}_l} \partial^{\mathbf{e}_l} [\det \{\mathbf{C}^{\mathcal{K}}(\mathbf{x}_1) + \mathbf{C}^{\mathcal{K}}(\mathbf{x}_2) - \mathbf{C}^{\mathcal{K}}(\mathbf{x}_1)[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{C}^{\mathcal{K}}(\mathbf{x}_2)\}]^{-1/2}}{\partial \mathbf{x}_1^{\mathbf{e}_l} \partial \mathbf{x}_2^{\mathbf{e}_l}} \Big|_{\mathbf{x}_1=\mathbf{x}_2=\mathbf{0}} \\
&\stackrel{(6.42),(B.22)}{=} -\frac{1}{2} [\det \{\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)\}]^{1/2} \\
&\quad \times \frac{\partial^{\mathbf{e}_l}}{\partial \mathbf{x}_1^{\mathbf{e}_l}} \frac{\text{Tr} \{ [\mathbf{C}^{\mathcal{K}}(\mathbf{x}_1) + \mathbf{C}^{\mathcal{K}}(\mathbf{0}) - \mathbf{C}^{\mathcal{K}}(\mathbf{x}_1)[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{C}^{\mathcal{K}}(\mathbf{0})]^{-1} [\mathbf{I} - \mathbf{C}^{\mathcal{K}}(\mathbf{x}_1)[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{P}_l] \}}{[\det \{\mathbf{C}^{\mathcal{K}}(\mathbf{x}_1) + \mathbf{C}^{\mathcal{K}}(\mathbf{0}) - \mathbf{C}^{\mathcal{K}}(\mathbf{x}_1)[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{C}^{\mathcal{K}}(\mathbf{0})\}]^{1/2}} \Big|_{\mathbf{x}_1=\mathbf{0}} \\
&\stackrel{(6.53),(B.22)}{=} \left[\frac{\det \{\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)\}}{\det \{\mathbf{N}\}} \right]^{1/2} \\
&\quad \times \left[\frac{1}{4} \text{Tr} \{ \mathbf{N}^{-1} [\mathbf{I} - \mathbf{C}^{\mathcal{K}}(\mathbf{0})[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{P}_l] \} \text{Tr} \{ \mathbf{N}^{-1} \mathbf{P}_l [\mathbf{I} - [\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{C}^{\mathcal{K}}(\mathbf{0})] \} \right. \\
&\quad + \frac{1}{2} \text{Tr} \{ \mathbf{N}^{-1} \mathbf{P}_l [\mathbf{I} - [\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{C}^{\mathcal{K}}(\mathbf{0})] \mathbf{N}^{-1} [\mathbf{I} - \mathbf{C}^{\mathcal{K}}(\mathbf{0})[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{P}_l] \} \\
&\quad \left. + \frac{1}{2} \text{Tr} \{ \mathbf{N}^{-1} \mathbf{P}_l [\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1} \mathbf{P}_l \} \right], \tag{B.20}
\end{aligned}$$

where

$$\mathbf{N} \triangleq 2\mathbf{C}^{\mathcal{K}}(\mathbf{0}) - \mathbf{C}^{\mathcal{K}}(\mathbf{0})[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{C}^{\mathcal{K}}(\mathbf{0}) \tag{B.21}$$

and

$$\mathbf{P}_l \triangleq \sum_{l' \in \mathcal{I}_l} \mathbf{e}_{l'} \mathbf{e}_{l'}^T = \frac{\partial^{\mathbf{e}_l}}{\partial \mathbf{x}^{\mathbf{e}_l}} \mathbf{C}^{\mathcal{K}}(\mathbf{x}), \tag{B.22}$$

with the index set

$$\mathcal{I}_l \triangleq \left\{ \sum_{k \in [l-1]} r_k + 1, \sum_{k \in [l-1]} r_k + 2, \dots, \sum_{k \in [l-1]} r_k + r_l - 1 \right\}. \tag{B.23}$$

The index set \mathcal{I}_l consists of the indices of the main diagonal entries of $\mathbf{D}^{\mathcal{K}}(\mathbf{x})$ in (6.11) that correspond to x_l as defined by (6.11). Note that \mathbf{P}_l is a psd orthogonal projection matrix, i.e., $\mathbf{P}_l\mathbf{P}_l = \mathbf{P}_l$.

The various terms in (B.20) can be bounded by using Lemma 6.2.1, the assumption $\delta_{S+1} < 1/32$, and the Definition in (B.7), as follows. First, observe that by Lemma 6.2.2, the eigenvalues $\lambda_{l'_k}(\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0))$ of the matrix $\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)$ (cf. (B.7)), that correspond to the k th largest entry $x_{0,k}$ of \mathbf{x}_0 , are upper bounded as

$$\begin{aligned} \lambda_{l'_k}(\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)) &\stackrel{(6.24)}{\leq} \lambda_{l'_k}(\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2\mathbf{I}) + \lambda_1(\mathbf{E}_1) \\ &\stackrel{(6.22)}{\leq} \lambda_{l'_k}(\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2\mathbf{I}) + \|\mathbf{E}_1\|_2 = x_{0,k} + \sigma^2 + \|\mathbf{E}_1\|_2 \\ &\stackrel{(B.12)}{\leq} x_{0,k} + \sigma^2(1 + 5\delta_{S+1}). \end{aligned} \quad (\text{B.24})$$

Note that there are r_k eigenvalues $\lambda_{l'_k}(\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0))$ that correspond to $x_{0,k}$. From this it follows that

$$\begin{aligned} [\det \{\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)\}]^{1/2} &\stackrel{(6.23)}{=} \left[\prod_{k \in \mathcal{K}} \prod_{l'_k} \lambda_{l'_k}(\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)) \right]^{1/2} \stackrel{(B.24)}{\leq} \prod_{k \in \mathcal{K}} [x_{0,k} + \sigma^2(1 + 5\delta_{S+1})]^{r_k/2} \\ &\stackrel{(B.2)}{=} [\sigma^2(1 + 5\delta_{S+1})]^{r_l/2} \prod_{k \in \text{supp}(\mathbf{x}_0)} [x_{0,k} + \sigma^2(1 + 5\delta_{S+1})]^{r_k/2}. \end{aligned} \quad (\text{B.25})$$

Similar to the bound (B.24), we also have

$$\begin{aligned} \lambda_{l'}(\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)) &\stackrel{(B.7)}{=} \lambda_{l'}(\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2\mathbf{V}_1\mathbf{\Sigma}_1^{-2}\mathbf{V}_1^T) \\ &\stackrel{(6.24)}{\geq} \lambda_{l'}(\sigma^2\mathbf{V}_1\mathbf{\Sigma}_1^{-2}\mathbf{V}_1^T) + \lambda_Q(\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0)) \\ &\geq \lambda_{l'}(\sigma^2\mathbf{V}_1\mathbf{\Sigma}_1^{-2}\mathbf{V}_1^T) = \sigma^2\lambda_{l'}(\mathbf{\Sigma}_1^{-2}) \geq \sigma^2(1 + \delta_{S+1})^{-2}, \end{aligned} \quad (\text{B.26})$$

and

$$\begin{aligned} \lambda_{l'}(\mathbf{C}^{\mathcal{K}}(\mathbf{0})) &\stackrel{(B.7)}{=} \lambda_{l'}(\mathbf{D}^{\mathcal{K}}(\mathbf{0}) + \sigma^2\mathbf{V}_1\mathbf{\Sigma}_1^{-2}\mathbf{V}_1^T) \\ &\stackrel{(B.6)}{=} \lambda_{l'}(\sigma^2\mathbf{V}_1\mathbf{\Sigma}_1^{-2}\mathbf{V}_1^T) = \sigma^2\lambda_{l'}(\mathbf{\Sigma}_1^{-2}) \in \sigma^2[(1 + \delta_{S+1})^{-2}, (1 - \delta_{S+1})^{-2}], \end{aligned} \quad (\text{B.27})$$

where the last step in (B.26) and (B.27), respectively, follow from the fact that the basis matrices satisfy the RIP of order $S + 1$. From (B.26) and we can conclude that

$$\|[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\|_2 \leq (1 + \delta_{S+1})^2\sigma^{-2} \quad (\text{B.28})$$

and in turn

$$\|\mathbf{C}^{\mathcal{K}}(\mathbf{0})[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{C}^{\mathcal{K}}(\mathbf{0})\|_2 \leq \|\mathbf{C}^{\mathcal{K}}(\mathbf{0})\|_2^2 \|\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)\|_2^{-1} \stackrel{(\text{B.28}),(\text{B.27})}{\leq} \sigma^2 \frac{(1 + \delta_{S+1})^2}{(1 - \delta_{S+1})^4}. \quad (\text{B.29})$$

This yields further

$$\begin{aligned} \lambda_{l'}(\mathbf{N}) &\stackrel{(6.24)}{\geq} \lambda_{l'}(2\mathbf{C}^{\mathcal{K}}(\mathbf{0})) + \lambda_Q(-\mathbf{C}^{\mathcal{K}}(\mathbf{0})[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{C}^{\mathcal{K}}(\mathbf{0})) \\ &\stackrel{(6.22)}{\geq} \lambda_{l'}(2\mathbf{C}^{\mathcal{K}}(\mathbf{0})) - \|\mathbf{C}^{\mathcal{K}}(\mathbf{0})[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{C}^{\mathcal{K}}(\mathbf{0})\|_2 \\ &\stackrel{(\text{B.29})}{\geq} \lambda_{l'}(2\mathbf{C}^{\mathcal{K}}(\mathbf{0})) - \sigma^2 \frac{(1 + \delta_{S+1})^2}{(1 - \delta_{S+1})^4} \\ &\stackrel{(\text{B.27})}{\geq} 2\sigma^2(1 + \delta_{S+1})^{-2} - \sigma^2 \frac{(1 + \delta_{S+1})^2}{(1 - \delta_{S+1})^4} \\ &= \sigma^2 \frac{1}{(1 + \delta_{S+1})^2} \left(2 - \frac{(1 + \delta_{S+1})^4}{(1 - \delta_{S+1})^4} \right) = \sigma^2 \beta, \end{aligned} \quad (\text{B.30})$$

and in turn

$$[\det \{\mathbf{N}\}]^{1/2} \stackrel{(6.23)}{=} \left[\prod_{k \in \mathcal{K}} \prod_{l'_k} \lambda_{l'_k}(\mathbf{N}) \right]^{1/2} \stackrel{(\text{B.30})}{\geq} (\sigma^2 \beta)^{Q/2}. \quad (\text{B.31})$$

Note that (B.30) also implies that

$$\|\mathbf{N}^{-1}\|_2 = \frac{1}{\min_{l' \in [Q]} |\lambda_{l'}(\mathbf{N})|} \leq \sigma^2 \beta. \quad (\text{B.32})$$

Since we have (cf. (B.6)) $\|(\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I})^{-1}\|_2 \leq 1/\sigma^2$, it holds that

$$\|(\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I})^{-1} \cdot \mathbf{E}_1\|_2 \leq \|(\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I})^{-1}\|_2 \cdot \|\mathbf{E}_1\|_2 \stackrel{(\text{B.12})}{\leq} 5\delta_{S+1} \stackrel{\delta_{S+1} < 1/32}{<} 1. \quad (\text{B.33})$$

This allows us to involve Lemma 6.2.1, with the choices $\mathbf{A} = \mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I}$ and $\mathbf{E} = \mathbf{C}^{\mathcal{K}}(\mathbf{x}_0) - [\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I}] = \mathbf{E}_1$ (cf. (B.7)), to obtain

$$[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1} = [\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I}]^{-1} + \mathbf{E}' \quad (\text{B.34})$$

with an error term \mathbf{E}' that satisfies

$$\begin{aligned} \|\mathbf{E}'\|_2 &= \|[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1} - [\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I}]^{-1}\|_2 \\ &\stackrel{(6.20)}{\leq} \frac{\|[\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I}]^{-1}\|_2^2 \|\mathbf{E}_1\|_2}{1 - \|[\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I}]^{-1}\|_2 \|\mathbf{E}_1\|_2} \leq \frac{1}{\sigma^4} \frac{\|\mathbf{E}_1\|_2}{1 - \|\mathbf{E}_1\|_2/\sigma^2} \end{aligned}$$

$$\stackrel{\text{(B.12)}}{\leq} \frac{1}{\sigma^4} \frac{\sigma^2 5\delta_{S+1}}{1 - 5\delta_{S+1}} \stackrel{\delta_{S+1} < 1/32}{\leq} \frac{1}{\sigma^2} \frac{5\delta_{S+1}}{1 - 5/32} \leq \frac{1}{\sigma^2} 6\delta_{S+1}. \quad (\text{B.35})$$

In what follows, we will also need the identity [38]

$$\text{Tr}\{\mathbf{A}\} = \sum_{l' \in [Q]} \mathbf{u}_{l'}^T \mathbf{A} \mathbf{u}_{l'} \quad (\text{B.36})$$

valid for any matrix $\mathbf{A} \in \mathbb{R}^{Q \times Q}$ and any set $\{\mathbf{u}_{l'} \in \mathbb{R}^T\}_{l' \in [Q]}$ of vectors that form an ONB for \mathbb{R}^Q . We then obtain

$$\begin{aligned} & \left| \text{Tr} \left\{ \mathbf{N}^{-1} [\mathbf{I} - \mathbf{C}^{\mathcal{K}}(\mathbf{0}) [\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}] \mathbf{P}_l \right\} \right| \\ & \stackrel{\text{(B.36)}}{=} \left| \sum_{l' \in [N]} \mathbf{e}_{l'}^T \mathbf{N}^{-1} [\mathbf{I} - \mathbf{C}^{\mathcal{K}}(\mathbf{0}) [\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}] \mathbf{P}_l \mathbf{e}_{l'} \right| \stackrel{\text{(B.22)}}{=} \left| \sum_{l' \in \mathcal{I}_l} \mathbf{e}_{l'}^T \mathbf{N}^{-1} [\mathbf{I} - \mathbf{C}^{\mathcal{K}}(\mathbf{0}) [\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}] \mathbf{e}_{l'} \right| \\ & \stackrel{\text{(B.7), (B.34)}}{=} \left| \sum_{l' \in \mathcal{I}_l} \mathbf{e}_{l'}^T \mathbf{N}^{-1} \left[\mathbf{I} - (\sigma^2 \mathbf{I} + \mathbf{E}_1) [(\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I})^{-1} + \mathbf{E}'] \right] \mathbf{e}_{l'} \right| \\ & = \left| \sum_{l' \in \mathcal{I}_l} \mathbf{e}_{l'}^T \mathbf{N}^{-1} (\mathbf{I} - \sigma^2 (\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I})^{-1} - \mathbf{E}_1 (\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I})^{-1} - \sigma^2 \mathbf{E}' - \mathbf{E}_1 \mathbf{E}') \mathbf{e}_{l'} \right|. \quad (\text{B.37}) \end{aligned}$$

Using the assumption that $l \notin \text{supp}(\mathbf{x}_0)$, implying that $(\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I})^{-1} \mathbf{e}_{l'} = \mathbf{e}_{l'} / \sigma^2$ for every $l' \in \mathcal{I}_l$, this further becomes

$$\begin{aligned} & \left| \text{Tr} \left\{ \mathbf{N}^{-1} [\mathbf{I} - \mathbf{C}^{\mathcal{K}}(\mathbf{0}) [\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}] \mathbf{P}_l \right\} \right| \\ & = \left| \sum_{l' \in \mathcal{I}_l} \mathbf{e}_{l'}^T \mathbf{N}^{-1} [\mathbf{e}_{l'} - \mathbf{e}_{l'} - \mathbf{E}_1 (\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I})^{-1} \mathbf{e}_{l'} - \sigma^2 \mathbf{E}' \mathbf{e}_{l'} - \mathbf{E}_1 \mathbf{E}' \mathbf{e}_{l'}] \right| \\ & = \left| \sum_{l' \in \mathcal{I}_l} \mathbf{e}_{l'}^T \mathbf{N}^{-1} [-\mathbf{E}_1 (\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I})^{-1} \mathbf{e}_{l'} - \sigma^2 \mathbf{E}' \mathbf{e}_{l'} - \mathbf{E}_1 \mathbf{E}' \mathbf{e}_{l'}] \right| \\ & \leq \sum_{l' \in \mathcal{I}_l} \left| \mathbf{e}_{l'}^T \mathbf{N}^{-1} [-\mathbf{E}_1 (\mathbf{D}^{\mathcal{K}}(\mathbf{x}_0) + \sigma^2 \mathbf{I})^{-1} \mathbf{e}_{l'} - \sigma^2 \mathbf{E}' \mathbf{e}_{l'} - \mathbf{E}_1 \mathbf{E}' \mathbf{e}_{l'}] \right| \\ & \stackrel{(b)}{\leq} r_l \frac{1}{\sigma^2 \beta} (5\delta_{S+1} + 6\delta_{S+1} + 30\delta_{S+1}^2) \stackrel{\delta_{S+1} < 1/32}{\leq} r_l \frac{1}{\sigma^2 \beta} 12\delta_{S+1}, \quad (\text{B.38}) \end{aligned}$$

where (b) follows from the inequality $\|\mathbf{A}\mathbf{B}\|_2 \leq \|\mathbf{A}\|_2 \|\mathbf{B}\|_2$, which is valid for any two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{Q \times Q}$ [38]. For the step (b) we also used (B.12), (B.32), and (B.35).

Similar to (B.38), one can also derive the inequalities

$$\begin{aligned}
|\operatorname{Tr}\{\mathbf{N}^{-1}\mathbf{P}_l[\mathbf{I}-[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{C}^{\mathcal{K}}(\mathbf{0})]\}\}| &\leq r_l\frac{1}{\sigma^2\beta}12\delta_{S+1} \\
|\operatorname{Tr}\{\mathbf{N}^{-1}\mathbf{P}_l[\mathbf{I}-[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{C}^{\mathcal{K}}(\mathbf{0})]\mathbf{N}^{-1}[\mathbf{I}-\mathbf{C}^{\mathcal{K}}(\mathbf{0})[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}]\mathbf{P}_l\}\}| &\leq r_l(12\delta_{S+1})^2\frac{1}{\sigma^4\beta^2} \\
|\operatorname{Tr}\{\mathbf{N}^{-1}\mathbf{P}_l[\mathbf{C}^{\mathcal{K}}(\mathbf{x}_0)]^{-1}\mathbf{P}_l\}\}| &\leq r_l\frac{1}{\beta\sigma^4}(6\delta_{S+1}+1). \tag{B.39}
\end{aligned}$$

The bound in (6.54) follows then finally by inserting the bounds (B.25), (B.31), (B.38), (B.39) into (B.20) and using (B.19). \square

Appendix C

Proof of Lemma 6.4.6

Proof. For the following, we restrict ourselves to the case where $\mathbf{p} = \mathbf{e}_l$ with an arbitrary index $l \in [N]$ and introduce the function $\mathbf{z}(\mathbf{x}) : \mathbb{R}_+^N \rightarrow \mathbb{R}_+^N$ defined elementwise by $z_k(\mathbf{x}) = \frac{1}{x_k + \sigma^2}$. Moreover, we define the modified mean function $\tilde{m}_k(\mathbf{z})$ by requiring that $\tilde{m}_k(\mathbf{z}(\mathbf{x})) = m_k(\mathbf{x})$. We have

$$\begin{aligned}
\tilde{m}_k(\mathbf{z}(\mathbf{x})) &= m_k(\mathbf{x}) = \mathbb{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\} = \int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) f_{\text{SDPCM}}(\mathbf{y}; \mathbf{x}) d\mathbf{y} \\
&\stackrel{(6.14)}{=} \frac{1}{(2\pi)^{M/2} [\det\{\tilde{\mathbf{C}}(\mathbf{x})\}]^{1/2}} \int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) \exp\left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}) \mathbf{y}\right) d\mathbf{y} \\
&\stackrel{(6.56), (6.58)}{=} \frac{1}{(2\pi)^{M/2} [\det\{\tilde{\mathbf{C}}(\mathbf{x})\}]^{1/2}} \\
&\quad \times \int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) \exp\left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k(\mathbf{x}) \sum_{i \in [r_k]} [\mathbf{u}_{m_k, i}^T \mathbf{y}]^2 + \frac{1}{\sigma^2} \mathbf{y}^T \mathbf{P}_n \mathbf{y} \right]\right) d\mathbf{y} \\
&\stackrel{(6.58)}{=} \frac{1}{(2\pi)^{M/2}} \frac{1}{\prod_{k \in [N]} (x_k + \sigma^2)^{r_k/2}} \frac{1}{\sigma^{M-R}} \\
&\quad \times \int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) \exp\left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k(\mathbf{x}) \sum_{i \in [r_k]} [\mathbf{u}_{m_k, i}^T \mathbf{y}]^2 + \frac{1}{\sigma^2} \mathbf{y}^T \mathbf{P}_n \mathbf{y} \right]\right) d\mathbf{y} \\
&= \frac{1}{(2\pi)^{M/2} \sigma^{M-R}} \prod_{k \in [N]} (z_k(\mathbf{x}))^{r_k/2}
\end{aligned}$$

$$\begin{aligned}
& \times \int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) \exp \left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k(\mathbf{x}) \sum_{i \in [r_k]} [\mathbf{u}_{m_k, i}^T \mathbf{y}]^2 + \frac{1}{\sigma^2} \mathbf{y}^T \mathbf{P}_n \mathbf{y} \right] \right) d\mathbf{y} \\
& \stackrel{\tilde{\mathbf{y}} = \mathbf{U}\mathbf{y}}{=} \frac{1}{(2\pi)^{M/2} \sigma^{M-R}} \prod_{k \in [N]} (z_k(\mathbf{x}))^{r_k/2} \\
& \times \int_{\tilde{\mathbf{y}} \in \mathbb{R}^M} \hat{x}_k(\mathbf{U}^T \tilde{\mathbf{y}}) \exp \left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k(\mathbf{x}) \sum_{i \in [r_k]} \tilde{y}_{m_i, k}^2 + \frac{1}{\sigma^2} \sum_{k \in [M] \setminus [R]} \tilde{y}_k^2 \right] \right) d\tilde{\mathbf{y}}, \quad (\text{C.1})
\end{aligned}$$

where \mathbf{P}_n denotes the projection matrix on the subspace of \mathbb{R}^M which is orthogonal to the subspace $\bigcup_{k \in [N]} \text{span}(\mathbf{C}_k)$ with \mathbf{C}_k being the basis matrices of the SDPCM. The matrix \mathbf{U} , used for the substitution $\mathbf{y} \rightarrow \tilde{\mathbf{y}}$, is the orthonormal matrix that appears in (6.58). The existence of the integrals in (C.1) follows from the dominated convergence theorem [5, 6, 73] since the function $\hat{x}_k(\mathbf{y}) \exp \left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}) \mathbf{y} \right)$ is measurable, since $\hat{x}_k(\mathbf{y})$ is assumed measurable [73]. Moreover, because of (6.87), the above function is upper bounded in magnitude (dominated) by the function $C \|\mathbf{y}\|_2^L \exp \left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}) \mathbf{y} \right)$, which is obviously integrable. Let us now calculate the partial derivative $\frac{\partial^{e_i} \tilde{m}_k(\mathbf{z})}{\partial \mathbf{z}^{e_i}}$.

$$\begin{aligned}
& \frac{\partial^{e_i} \tilde{m}_k(\mathbf{z})}{\partial \mathbf{z}^{e_i}} \stackrel{(\text{C.1})}{=} \frac{1}{(2\pi)^{M/2} \sigma^{M-R}} \frac{\partial^{e_i}}{\partial \mathbf{z}^{e_i}} \prod_{k \in [N]} z_k^{r_k/2} \\
& \times \int_{\tilde{\mathbf{y}} \in \mathbb{R}^M} \hat{x}_k(\mathbf{U}^T \tilde{\mathbf{y}}) \exp \left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k \sum_{i \in [r_k]} \tilde{y}_{m_i, k}^2 + \frac{1}{\sigma^2} \sum_{k \in [M] \setminus [R]} \tilde{y}_k^2 \right] \right) d\tilde{\mathbf{y}} \\
& \stackrel{(a)}{=} \frac{1}{(2\pi)^{M/2} \sigma^{M-R}} \\
& \left\{ \left[\frac{\partial^{e_i}}{\partial \mathbf{z}^{e_i}} \prod_{k \in [N]} z_k^{r_k/2} \right] \int_{\tilde{\mathbf{y}} \in \mathbb{R}^M} \hat{x}_k(\mathbf{U}^T \tilde{\mathbf{y}}) \exp \left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k \sum_{i \in [r_k]} \tilde{y}_{m_i, k}^2 + \frac{1}{\sigma^2} \sum_{k \in [M] \setminus [R]} \tilde{y}_k^2 \right] \right) d\tilde{\mathbf{y}} \right. \\
& \left. + \prod_{k \in [N]} (z_k)^{r_k/2} \left[\frac{\partial^{e_i}}{\partial \mathbf{z}^{e_i}} \int_{\tilde{\mathbf{y}} \in \mathbb{R}^M} \hat{x}_k(\mathbf{U}^T \tilde{\mathbf{y}}) \exp \left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k \sum_{i \in [r_k]} \tilde{y}_{m_i, k}^2 + \frac{1}{\sigma^2} \sum_{k \in [M] \setminus [R]} \tilde{y}_k^2 \right] \right) d\tilde{\mathbf{y}} \right] \right\} \\
& \stackrel{(b)}{=} \frac{1}{(2\pi)^{M/2} \sigma^{M-R}}
\end{aligned}$$

$$\begin{aligned}
& \left\{ \left[\frac{\partial^{e_l}}{\partial \mathbf{z}^{e_l}} \prod_{k \in [N]} z_k^{r_k/2} \right] \int_{\tilde{\mathbf{y}} \in \mathbb{R}^M} \hat{x}_k(\mathbf{U}^T \tilde{\mathbf{y}}) \exp \left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k \sum_{i \in [r_k]} \tilde{y}_{m_i, k}^2 + \frac{1}{\sigma^2} \sum_{k \in [M] \setminus [R]} \tilde{y}_k^2 \right] \right) d\tilde{\mathbf{y}} \right. \\
& \left. + \prod_{k \in [N]} (z_k)^{r_k/2} \left[\int_{\tilde{\mathbf{y}} \in \mathbb{R}^M} \hat{x}_k(\mathbf{U}^T \tilde{\mathbf{y}}) \frac{\partial^{e_l}}{\partial \mathbf{z}^{e_l}} \exp \left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k \sum_{i \in [r_k]} \tilde{y}_{m_i, k}^2 + \frac{1}{\sigma^2} \sum_{k \in [M] \setminus [R]} \tilde{y}_k^2 \right] \right) d\tilde{\mathbf{y}} \right] \right\} \\
& = \frac{1}{(2\pi)^{M/2} \sigma^{M-R}} \\
& \quad \times \left\{ \left[\int_{\tilde{\mathbf{y}} \in \mathbb{R}^M} \hat{x}_k(\mathbf{U}^T \tilde{\mathbf{y}}) \exp \left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k \sum_{i \in [r_k]} \tilde{y}_{m_i, k}^2 + \frac{1}{\sigma^2} \sum_{k \in [M] \setminus [R]} \tilde{y}_k^2 \right] \right) d\tilde{\mathbf{y}} \right] \frac{r_l}{2z_l} \prod_{k \in [N]} z_k^{r_k/2} \right. \\
& \left. + \left[\int_{\tilde{\mathbf{y}} \in \mathbb{R}^M} \frac{-\sum_{i \in [r_l]} \tilde{y}_{m_i, l}^2}{2} \hat{x}_k(\mathbf{U}^T \tilde{\mathbf{y}}) \exp \left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k \sum_{i \in [r_k]} \tilde{y}_{m_i, k}^2 + \frac{1}{\sigma^2} \sum_{k \in [M] \setminus [R]} \tilde{y}_k^2 \right] \right) d\tilde{\mathbf{y}} \right] \prod_{k \in [N]} z_k^{r_k/2} \right\}, \tag{C.2}
\end{aligned}$$

where (a) follows from the product rule for differentiation [5] and (b) results from a change of the order of differentiation and integration [41, Theorem 1.5.8], which can be verified, e.g., by a standard argument using the dominated convergence theorem [5, 61, 73]. From (C.2), we obtain by the chain rule for differentiation [5, Theorem 9.15] that

$$\begin{aligned}
\frac{\partial^{e_l} m_k(\mathbf{x})}{\partial \mathbf{x}^{e_l}} &= \frac{\partial^{e_l} \tilde{m}_k(\mathbf{z})}{\partial \mathbf{z}^{e_l}} \frac{\partial z_l(\mathbf{x})}{\partial x_l} = \frac{\partial^{e_l} \tilde{m}_k(\mathbf{z})}{\partial \mathbf{z}^{e_l}} \frac{-1}{(x_l + \sigma^2)^2} \\
&\stackrel{(C.2)}{=} \frac{1}{(2\pi)^{M/2} \sigma^{M-R}} \cdot \frac{-1}{(x_l + \sigma^2)^2} \\
& \quad \times \left\{ \left[\int_{\tilde{\mathbf{y}} \in \mathbb{R}^M} \hat{x}_k(\mathbf{U}^T \tilde{\mathbf{y}}) \exp \left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k \sum_{i \in [r_k]} \tilde{y}_{m_i, k}^2 + \frac{1}{\sigma^2} \sum_{k \in [M] \setminus [R]} \tilde{y}_k^2 \right] \right) d\tilde{\mathbf{y}} \right] \frac{r_l}{2z_l} \prod_{k \in [N]} z_k^{r_k/2} \right. \\
& \left. + \left[\int_{\tilde{\mathbf{y}} \in \mathbb{R}^M} \frac{-\sum_{i \in [r_l]} \tilde{y}_{m_i, l}^2}{2} \hat{x}_k(\mathbf{U}^T \tilde{\mathbf{y}}) \exp \left(-\frac{1}{2} \left[\sum_{k \in [N]} z_k \sum_{i \in [r_k]} \tilde{y}_{m_i, k}^2 + \frac{1}{\sigma^2} \sum_{k \in [M] \setminus [R]} \tilde{y}_k^2 \right] \right) d\tilde{\mathbf{y}} \right] \prod_{k \in [N]} z_k^{r_k/2} \right\} \\
&\stackrel{(6.56), (6.58)}{=} \frac{1}{(2\pi)^{M/2} \sigma^{M-R}} \cdot \frac{-1}{(x_l + \sigma^2)^2} \\
& \quad \times \left\{ \left[\int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) \exp \left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}) \mathbf{y} \right) d\mathbf{y} \right] \frac{r_l}{2z_l} \prod_{k \in [N]} z_k^{r_k/2} \right.
\end{aligned}$$

$$\begin{aligned}
& + \left[\int_{\mathbf{y} \in \mathbb{R}^M} \frac{-\mathbf{y}^T \mathbf{C}_l \mathbf{y}}{2} \hat{x}_k(\mathbf{y}) \exp\left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}) \mathbf{y}\right) d\mathbf{y} \right] \prod_{k \in [N]} z_k^{r_k/2} \Big\} \\
& \stackrel{(6.58)}{=} \frac{1}{(2\pi)^{M/2} [\det\{\tilde{\mathbf{C}}(\mathbf{x})\}]^{1/2}} \cdot \frac{-1}{(x_l + \sigma^2)^2} \\
& \times \left\{ \left[\int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) \exp\left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}) \mathbf{y}\right) d\mathbf{y} \right] \frac{r_l}{2z_l} \right. \\
& \quad \left. + \left[\int_{\mathbf{y} \in \mathbb{R}^M} \frac{-\mathbf{y}^T \mathbf{C}_l \mathbf{y}}{2} \hat{x}_k(\mathbf{y}) \exp\left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}) \mathbf{y}\right) d\mathbf{y} \right] \right\} \\
& = \frac{1}{(2\pi)^{M/2} [\det\{\tilde{\mathbf{C}}(\mathbf{x})\}]^{1/2}} \\
& \times \left\{ \left[\int_{\mathbf{y} \in \mathbb{R}^M} \hat{x}_k(\mathbf{y}) \exp\left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}) \mathbf{y}\right) d\mathbf{y} \right] \frac{-r_l}{2(x_l + \sigma^2)} \right. \\
& \quad \left. + \frac{-1}{(x_l + \sigma^2)^2} \int_{\mathbf{y} \in \mathbb{R}^M} \frac{-\mathbf{y}^T \mathbf{C}_l \mathbf{y}}{2} \hat{x}_k(\mathbf{y}) \exp\left(-\frac{1}{2} \mathbf{y}^T \tilde{\mathbf{C}}^{-1}(\mathbf{x}) \mathbf{y}\right) d\mathbf{y} \right\} \\
& = -\frac{r_l}{2(x_l + \sigma^2)} \mathbb{E}_{\mathbf{x}}\{\hat{x}_k(\mathbf{y})\} + \frac{1}{2(x_l + \sigma^2)^2} \mathbb{E}_{\mathbf{x}}\{\mathbf{y}^T \mathbf{C}_l \mathbf{y} \hat{x}_k(\mathbf{y})\} \\
& = -\frac{r_l}{2(x_l + \sigma^2)} m_k(\mathbf{x}) + \frac{1}{2(x_l + \sigma^2)^2} \mathbb{E}_{\mathbf{x}}\{\mathbf{y}^T \mathbf{C}_l \mathbf{y} \hat{x}_k(\mathbf{y})\}.
\end{aligned} \tag{C.3}$$

The existence of the higher order partial derivatives $\frac{\partial^{\mathbf{p}} m_k(\mathbf{x})}{\partial \mathbf{x}^{\mathbf{p}}}$ can then be verified recursively. \square

Appendix D

Proof of Theorem 6.4.8

Proof. We prove the statement separately for the two complementary cases where $k = j_0$ and the case where $k \neq j_0$. First, consider the case $k = j_0$, where we have obviously that $|\{k\} \cup \text{supp}(\mathbf{x}_0)| = |\{k\}| = 1 < S + 1$. This implies that the estimator (6.96) equals the estimator (6.90), which is unbiased (cf. (6.91)) and whose variance achieves the bound (6.89) (for the case $k \in \text{supp}(\mathbf{x}_0)$) of Corollary 6.4.7 (cf. (6.93)). Therefore, the estimator (6.96) must be the LMVU for the case $k = j_0$.

Next, we consider the complementary case $k \neq j_0$. We calculate the mean $\mathbb{E}_x\{\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y})\}$ separately for parameter vectors \mathbf{x} such that $\text{supp}(\mathbf{x}) = j_0$ and parameter vectors such that $\text{supp}(\mathbf{x}) \neq j_0$.

- $\text{supp}(\mathbf{x}) \neq j_0$: We obtain for the mean of $\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y})$

$$\begin{aligned}
 \mathbb{E}_x\{\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y})\} &= \mathbb{E}_x\{\alpha(\mathbf{y}; \mathbf{x}_0)(\beta_k(\mathbf{y}) - \sigma^2)\} \stackrel{(a)}{=} \mathbb{E}_x\{\alpha(\mathbf{y}; \mathbf{x}_0)\}\mathbb{E}_x\{\beta_k(\mathbf{y}) - \sigma^2\} \\
 &\stackrel{(6.91)}{=} x_k \mathbb{E}_x\{\alpha(\mathbf{y}; \mathbf{x}_0)\} = x_k \mathbb{E}_x\{a(\mathbf{x}_0) \exp(-r_{j_0} b(\mathbf{x}_0) \beta_{j_0}(\mathbf{y}))\} \\
 &= x_k a(\mathbf{x}_0) \mathbb{E}_x\{\exp(-r_{j_0} b(\mathbf{x}_0) \beta_{j_0}(\mathbf{y}))\} \\
 &\stackrel{(6.90)}{=} x_k a(\mathbf{x}_0) \mathbb{E}_x\left\{\exp(-r_{j_0} b(\mathbf{x}_0) \frac{1}{r_{j_0}} \sum_{i \in [r_{j_0}]} (\mathbf{u}_{m_{j_0}, i}^T \mathbf{y})^2)\right\} \\
 &\stackrel{(b)}{=} x_k a(\mathbf{x}_0) \prod_{i \in [r_{j_0}]} \mathbb{E}_x\{\exp(-b(\mathbf{x}_0) z_i^2)\}, \tag{D.1}
 \end{aligned}$$

where $z_i \triangleq \mathbf{u}_{m_{j_0}, i}^T \mathbf{y}$ (cf. (6.56)). Note that the random variables $\{z_i\}_{i \in [r_{j_0}]}$ are i.i.d. with $z_i \sim \mathcal{N}(0, \sigma^2)$, which can be verified by (6.58) and the fact that we consider the case $\text{supp}(\mathbf{x}) \neq j_0$. The steps (a) and (b) are due to the statistical independence of the random variables

$\{\beta_l(\mathbf{y})\}_{l \in [N]}$, which can be verified by (6.58). We now use the identity

$$\mathbb{E}\{\exp(-Bz^2)\} = [(2B + c^{-2})c^2]^{-1/2}, \quad (\text{D.2})$$

which is valid for a Gaussian random variable $z \sim \mathcal{N}(0, c^2)$ with arbitrary variance c^2 [45], to obtain from (D.1) that

$$\begin{aligned} \mathbb{E}_{\mathbf{x}}\{\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y})\} &= x_k a(\mathbf{x}_0) \prod_{i \in [r_{j_0}]} [(2b(\mathbf{x}_0) + \sigma^{-2})\sigma^2]^{-1/2} \\ &= x_k a(\mathbf{x}_0) [(2b(\mathbf{x}_0) + \sigma^{-2})\sigma^2]^{-r_{j_0}/2} \\ &= x_k a(\mathbf{x}_0) [(2(\sigma^{-2} - (\xi_0 + \sigma^2)^{-1})/2 + \sigma^{-2})\sigma^2]^{-r_{j_0}/2} \\ &= x_k a(\mathbf{x}_0) [1 - \sigma^2(\xi_0 + \sigma^2)^{-1} + 1]^{-r_{j_0}/2} \\ &= x_k a(\mathbf{x}_0) [2 - \sigma^2(\xi_0 + \sigma^2)^{-1}]^{-r_{j_0}/2} \\ &= x_k a(\mathbf{x}_0) [2(\xi_0 + \sigma^2) - \sigma^2]^{-r_{j_0}/2} (\xi_0 + \sigma^2)^{r_{j_0}/2} \\ &= x_k a(\mathbf{x}_0) (2\xi_0 + \sigma^2)^{-r_{j_0}/2} (\xi_0 + \sigma^2)^{r_{j_0}/2} = x_k. \end{aligned} \quad (\text{D.3})$$

- $\text{supp}(\mathbf{x}) = j_0$: In this case, we have $k \neq \text{supp}(\mathbf{x})$ (since $k \neq j_0$ is assumed) and thus $x_k = 0$. Hence, we obtain for the mean of $\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y})$

$$\begin{aligned} \mathbb{E}_{\mathbf{x}}\{\hat{x}_k^{(\mathbf{x}_0)}(\mathbf{y})\} &= \mathbb{E}_{\mathbf{x}}\{\alpha(\mathbf{y}; \mathbf{x}_0) (\beta_k(\mathbf{y}) - \sigma^2)\} \\ &\stackrel{(a)}{=} \mathbb{E}_{\mathbf{x}}\{\alpha(\mathbf{y}; \mathbf{x}_0)\} \mathbb{E}_{\mathbf{x}}\{\beta_k(\mathbf{y}) - \sigma^2\} \stackrel{(b)}{=} 0 = x_k, \end{aligned} \quad (\text{D.4})$$

where (a) follows from the statistical independence of the random variables $\{\beta_l(\mathbf{y})\}_{l \in [N]}$ and (b) is due to $\mathbb{E}_{\mathbf{x}}\{\beta_k(\mathbf{y}) - \sigma^2\} = x_k = 0$ (again see (6.91)).

To summarize, we have shown that for the case $k \neq j_0$, the estimator (6.96) is unbiased for every $\mathbf{x} \in \mathcal{X}_{S=1,+}$.

Finally, we show that the variance at \mathbf{x}_0 of the estimator (6.96) achieves the lower bound (6.89) (for the case $k \notin \text{supp}(\mathbf{x}_0)$) of Corollary 6.4.7. Indeed, we obtain under the assumption $k \neq j_0 = \text{supp}(\mathbf{x}_0)$ and using the shorthand $z_i \triangleq \mathbf{u}_{m_{j_0}, i}^T \mathbf{y}$

$$v(\hat{x}_k^{(\mathbf{x}_0)}(\cdot); \mathbf{x}_0) \stackrel{(a)}{=} P(\hat{x}_k^{(\mathbf{x}_0)}(\cdot); \mathbf{x}_0)$$

$$\begin{aligned}
&= \mathbf{E}_{\mathbf{x}_0} \{ \alpha^2(\mathbf{y}; \mathbf{x}_0) (\beta_k(\mathbf{y}) - \sigma^2)^2 \} \\
&\stackrel{(b)}{=} \mathbf{E}_{\mathbf{x}_0} \{ \alpha^2(\mathbf{y}; \mathbf{x}_0) \} \mathbf{E}_{\mathbf{x}_0} \{ (\beta_k(\mathbf{y}) - \sigma^2)^2 \}. \tag{D.5}
\end{aligned}$$

Here, for step (a) we used the fact that the estimator $\hat{x}_k^{(\mathbf{x}_0)}(\cdot)$ is unbiased as has been proven just before and step (b) is due the statistical independence of the random variables $\{\beta_l(\mathbf{y})\}_{l \in [N]}$. We have $\mathbf{E}_{\mathbf{x}_0} \{ (\beta_k(\mathbf{y}) - \sigma^2)^2 \} = \frac{2\sigma^4}{r_k}$ because of (6.94). Thus, we obtain from (D.5) that

$$\begin{aligned}
v(\hat{x}_k^{(\mathbf{x}_0)}(\cdot); \mathbf{x}_0) &= \frac{2\sigma^4}{r_k} \mathbf{E}_{\mathbf{x}_0} \{ \alpha^2(\mathbf{y}; \mathbf{x}_0) \} \\
&\stackrel{(a)}{=} \frac{2\sigma^4}{r_k} a^2(\mathbf{x}_0) \prod_{i \in [r_{j_0}]} \mathbf{E}_{\mathbf{x}_0} \{ \exp(-2b(\mathbf{x}_0)z_i^2) \} \\
&\stackrel{(D.2)}{=} \frac{2\sigma^4}{r_k} a^2(\mathbf{x}_0) \prod_{i \in [r_{j_0}]} [2[\sigma^{-2} - (\xi_0 + \sigma^2)^{-1}] + (\xi_0 + \sigma^2)^{-1}]^{-1/2} (\xi_0 + \sigma^2)^{-1/2} \\
&= \frac{2\sigma^4}{r_k} a^2(\mathbf{x}_0) \prod_{i \in [r_{j_0}]} [2\sigma^{-2}(\xi_0 + \sigma^2) - 1]^{-1/2} \\
&= \frac{2\sigma^4}{r_k} a^2(\mathbf{x}_0) [2\sigma^{-2}(\xi_0 + \sigma^2) - 1]^{-r_{j_0}/2} \\
&= \frac{2\sigma^4}{r_k} a^2(\mathbf{x}_0) \sigma^{r_{j_0}} (2\xi_0 + \sigma^2)^{-r_{j_0}/2} \\
&= \frac{2\sigma^4}{r_k} \left[\frac{2\xi_0 + \sigma^2}{\xi_0 + \sigma^2} \right]^{r_{j_0}} \sigma^{r_{j_0}} (2\xi_0 + \sigma^2)^{-r_{j_0}/2} \\
&= \frac{2\sigma^4}{r_k} \frac{(2\xi_0\sigma^2 + \sigma^4)^{r_{j_0}/2}}{(\xi_0 + \sigma^2)^{r_{j_0}}} \\
&= \frac{2\sigma^4}{r_k} \frac{[(\xi_0 + \sigma^2)^2 - \xi_0^2]^{r_{j_0}/2}}{(\xi_0 + \sigma^2)^{r_{j_0}}} \\
&= \frac{2}{r_k} \sigma^4 \left[1 - \frac{\xi_0^2}{(\xi_0 + \sigma^2)^2} \right]^{\frac{r_{j_0}}{2}}, \tag{D.6}
\end{aligned}$$

which is (6.89) (for the case $k \notin \text{supp}(\mathbf{x}_0)$). Here, the step (a) is due to the fact that the random variables $\{z_i\}_{i \in [r_{j_0}]}$ are i.i.d. with $z_i \sim \mathcal{N}(0, \sigma^2 + \xi_0)$. \square

Appendix E

Proof of Theorem 6.5.1

Proof. Consider the function

$$h_k(x) : \mathbb{R}_+ \rightarrow \mathbb{R} : h_k(x) \triangleq -r_k \left[\frac{\beta_k(\mathbf{y})}{x + \sigma^2} + \log(x + \sigma^2) \right]. \quad (\text{E.1})$$

This function is continuous on its domain and differentiable at every point $x > 0$, with derivative

$$h'_k(x) = r_k \left[\frac{\beta_k(\mathbf{y})}{(x + \sigma^2)^2} - \frac{1}{x + \sigma^2} \right]. \quad (\text{E.2})$$

As can be verified easily, if $\beta_k(\mathbf{y}) < \sigma^2$, then the derivative in (E.2) is always negative, which implies via the mean value theorem [5] that the function $h_k(x)$ is monotonically decreasing and attains its maximum at $x = 0$. Thus, we have

$$\beta_k(\mathbf{y}) < \sigma^2 \quad \Rightarrow \quad \max_{x \in \mathbb{R}_+} h_k(x) = h_k(0) = -r_k \left[\frac{\beta_k(\mathbf{y})}{\sigma^2} + \log(\sigma^2) \right]. \quad (\text{E.3})$$

On the other hand, if $\beta_k(\mathbf{y}) \geq \sigma^2$, then we can distinguish two cases: (i) for $x \in [0, \beta_k(\mathbf{y}) - \sigma^2]$, the derivative $h'_k(x)$ is nonnegative, implying that $h_k(x)$ is monotonically increasing; (ii) for $x \geq \beta_k(\mathbf{y}) - \sigma^2$, the derivative is negative, implying that $h_k(x)$ is monotonically decreasing. These two facts imply that the maximum of $h_k(x)$ must occur at $x = \beta_k(\mathbf{y}) - \sigma^2$, i.e.,

$$\beta_k(\mathbf{y}) \geq \sigma^2 \quad \Rightarrow \quad \max_{x \in \mathbb{R}_+} h_k(x) = h_k(\beta_k(\mathbf{y}) - \sigma^2) = -r_k [1 + \log(\beta_k(\mathbf{y}))]. \quad (\text{E.4})$$

In what follows we will use the shorthand

$$H(\mathbf{x}) \triangleq \sum_{k \in [N]} h_k(x_k), \quad (\text{E.5})$$

which allows us to rewrite the ML estimator (cf. (6.100)) as

$$\hat{\mathbf{x}}_{\text{ML}}(\mathbf{y}) = \operatorname{argmax}_{\mathbf{x}' \in \mathcal{X}_{S,+}} \left\{ \sum_{k \in [N]} h_k(x'_k) \right\} = \operatorname{argmax}_{\mathbf{x}' \in \mathcal{X}_{S,+}} H(\mathbf{x}'). \quad (\text{E.6})$$

Let us now show by contradiction that $\text{supp}(\hat{\mathbf{x}}_{\text{ML}}(\mathbf{y})) \subseteq \mathcal{L}_2$, i.e., $\hat{x}_{k,\text{ML}}(\mathbf{y}) = 0$ for every index $k \notin \mathcal{L}_2$. Let $\tilde{\mathbf{x}} \triangleq \hat{\mathbf{x}}_{\text{ML}}(\mathbf{y})$, and assume that the ML estimate $\tilde{\mathbf{x}}$ has a nonzero entry $\tilde{x}_{k'}$ with index $k' \notin \mathcal{L}_2$. Consider then a new parameter vector $\mathbf{x}_2 \in \mathcal{X}_{S,+}$ which is obtained from $\tilde{\mathbf{x}}$ by zeroing the entry with index k' and retaining the rest. We can then write the difference of the objective values $H(\mathbf{x})$ for the parameter vectors $\tilde{\mathbf{x}}$ and \mathbf{x}_2 as

$$\begin{aligned}
H(\tilde{\mathbf{x}}) - H(\mathbf{x}_2) &\stackrel{\text{(E.5)}}{=} \left\{ \sum_{k \in [N]} h_k(\tilde{x}_k) \right\} - \left\{ \sum_{k \in [N]} h_k(x_{2,k}) \right\} \\
&= \left\{ \sum_{k \in [N] \setminus \{k'\}} h_k(\tilde{x}_k) \right\} + h_{k'}(\tilde{x}_{k'}) - \left\{ \sum_{k \in [N] \setminus \{k'\}} h_k(x_{2,k}) \right\} - h_{k'}(x_{2,k'}) \\
&= h_{k'}(\tilde{x}_{k'}) - h_{k'}(x_{2,k'}) \\
&= h_{k'}(\tilde{x}_{k'}) - h_{k'}(0). \tag{E.7}
\end{aligned}$$

Because $k' \notin \mathcal{L}_2$ we have $\beta_{k'}(\mathbf{y}) < \sigma^2$ and in turn that the derivative (E.2) (for $k = k'$) is strictly negative for all $x \in \mathbb{R}_+$. It then follows from (E.7) and (E.3) that

$$H(\tilde{\mathbf{x}}) - H(\mathbf{x}_2) < 0. \tag{E.8}$$

Thus, we have that the new parameter vector \mathbf{x}_2 yields a larger value of the objective $H(\mathbf{x})$ than $\tilde{\mathbf{x}}$. However, this contradicts the assumption that $\tilde{\mathbf{x}}$ is the ML estimator, i.e., the parameter vector in $\mathcal{X}_{S,+}$ that maximizes $H(\mathbf{x})$. Thus, we have shown that $\text{supp}(\hat{\mathbf{x}}_{\text{ML}}(\mathbf{y})) \subseteq \mathcal{L}_2$, i.e., $\hat{x}_{k,\text{ML}}(\mathbf{y}) = 0$ for every index $k \notin \mathcal{L}_2$.

Let \mathcal{I} be a given set of indices such that $\mathcal{I} \subseteq \mathcal{L}_2$ and $|\mathcal{I}| \leq S$, and let us denote by $H_{\mathcal{I}}$ the maximum value of $H(\mathbf{x}')$ under the additional constraint that $\text{supp}(\mathbf{x}') \subseteq \mathcal{I}$, i.e.,

$$\begin{aligned}
H_{\mathcal{I}} &\triangleq \max_{\substack{\mathbf{x}' \in \mathbb{R}_+^N \\ \text{supp}(\mathbf{x}') \subseteq \mathcal{I}}} H(\mathbf{x}') \\
&= \max_{\substack{\mathbf{x}' \in \mathbb{R}_+^N \\ \text{supp}(\mathbf{x}') \subseteq \mathcal{I}}} \sum_{k \in [N]} h_k(x'_k) \\
&= \max_{\substack{\mathbf{x}' \in \mathbb{R}_+^N \\ \text{supp}(\mathbf{x}') \subseteq \mathcal{I}}} \sum_{k \notin \mathcal{I}} h_k(0) + \sum_{k \in \mathcal{I}} h_k(x'_k) \\
&\stackrel{\text{(E.4)}}{=} \sum_{k \notin \mathcal{I}} h_k(0) + \sum_{k \in \mathcal{I}} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] \right\}. \tag{E.9}
\end{aligned}$$

(E.10)

Up to now, we have verified that the ML estimate $\tilde{\mathbf{x}}$ must satisfy $\text{supp}(\tilde{\mathbf{x}}) \subseteq \mathcal{L}_2$. We will now show that the vector \mathbf{x}_3 given by (6.101) yields the ML estimate, i.e., $H(\tilde{\mathbf{x}}) = H(\mathbf{x}_3)$. Note that there may be multiple ML estimates $\tilde{\mathbf{x}}$ which yield the global maximum in (6.100); we will show that \mathbf{x}_3 yields one of these solutions. By the definition of \mathbf{x}_3 ,

$$\begin{aligned}
H(\mathbf{x}_3) &\stackrel{\text{(E.5)}}{=} \sum_{k \in [N]} h_k(x_{3,k}) \\
&\stackrel{\text{(6.101)}}{=} \sum_{k \notin \mathcal{L}_1 \cap \mathcal{L}_2} h_k(0) + \sum_{k \in \mathcal{L}_1 \cap \mathcal{L}_2} h_k(\beta_k(\mathbf{y}) - \sigma^2) \\
&\stackrel{\text{(E.1)}}{=} \sum_{k \notin \mathcal{L}_1 \cap \mathcal{L}_2} h_k(0) + \sum_{k \in \mathcal{L}_1 \cap \mathcal{L}_2} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] \right\} \\
&\stackrel{\text{(E.9)}}{=} H_{\mathcal{L}_1 \cap \mathcal{L}_2}. \tag{E.11}
\end{aligned}$$

Consider then an arbitrary parameter vector $\mathbf{x}' \in \mathcal{X}_{S,+}$ which satisfies $\text{supp}(\mathbf{x}') \subseteq \mathcal{L}_2$ and denote its support by $\mathcal{I}' \triangleq \text{supp}(\mathbf{x}')$. Obviously, we have that $\mathcal{I}' \subseteq \mathcal{L}_2$ and $|\mathcal{I}'| \leq S$ which implies that

$$|\mathcal{I}'| \leq \min\{S, |\mathcal{L}_2|\}. \tag{E.12}$$

The difference of the objective values $H(\mathbf{x})$ for the parameter vectors \mathbf{x}_3 and \mathbf{x}' then satisfies

$$\begin{aligned}
H(\mathbf{x}_3) - H(\mathbf{x}') &\stackrel{\text{(E.11)}}{=} H_{\mathcal{L}_1 \cap \mathcal{L}_2} - H(\mathbf{x}') \stackrel{\text{(E.9)}}{\geq} H_{\mathcal{L}_1 \cap \mathcal{L}_2} - H_{\mathcal{I}'} \\
&\stackrel{\text{(E.9)}}{=} \sum_{k \in (\mathcal{L}_1 \cap \mathcal{L}_2)} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] \right\} + \sum_{k \notin (\mathcal{L}_1 \cap \mathcal{L}_2)} h_k(0) \\
&\quad - \sum_{k \in \mathcal{I}'} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] \right\} - \sum_{k \notin \mathcal{I}'} h_k(0) \\
&= \sum_{k \in (\mathcal{L}_1 \cap \mathcal{L}_2 \cap \mathcal{I}')} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] \right\} + \sum_{k \in (\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] \right\} \\
&\quad + \sum_{k \in [N] \setminus ((\mathcal{L}_1 \cap \mathcal{L}_2) \cup \mathcal{I}')} h_k(0) + \sum_{k \in \mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)} h_k(0)
\end{aligned}$$

$$\begin{aligned}
& - \sum_{k \in (\mathcal{L}_1 \cap \mathcal{L}_2 \cap \mathcal{I}')} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] \right\} - \sum_{k \in \mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] \right\} \\
& - \sum_{k \in [N] \setminus ((\mathcal{L}_1 \cap \mathcal{L}_2) \cup \mathcal{I}')} h_k(0) - \sum_{k \in (\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'} h_k(0) \\
& \stackrel{(E.9)}{=} \sum_{k \in (\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] - h_k(0) \right\} \\
& - \sum_{k \in \mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] - h_k(0) \right\} \\
& = \sum_{k \in (\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] + r_k \left[\frac{\beta_k(\mathbf{y})}{\sigma^2} + \log(\sigma^2) \right] \right\} \\
& - \sum_{k \in \mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)} \left\{ -r_k [1 + \log(\beta_k(\mathbf{y}))] + r_k \left[\frac{\beta_k(\mathbf{y})}{\sigma^2} + \log(\sigma^2) \right] \right\} \\
& = \sum_{k \in (\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'} r_k \left[\frac{\beta_k(\mathbf{y})}{\sigma^2} - \log\left(\frac{\beta_k(\mathbf{y})}{\sigma^2}\right) - 1 \right] \\
& - \sum_{k' \in \mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)} r_{k'} \left[\frac{\beta_{k'}(\mathbf{y})}{\sigma^2} - \log\left(\frac{\beta_{k'}(\mathbf{y})}{\sigma^2}\right) - 1 \right] \tag{E.13}
\end{aligned}$$

Note that by definition $|\mathcal{L}_1 \cap \mathcal{L}_2| = \min\{S, |\mathcal{L}_2|\}$, implying due to (E.12) that

$$|\mathcal{L}_1 \cap \mathcal{L}_2| \geq |\mathcal{I}'|, \tag{E.14}$$

and in turn

$$|(\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'| \geq |\mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)|. \tag{E.15}$$

From (E.15), we have that there exists an injective index map $\Omega(k') : \mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2) \rightarrow (\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'$, such that for each summand (with index $k' \in \mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)$) of the second summand in (E.13), we can associate a distinct summand (with index $\Omega(k') \in (\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'$) of the first sum in (E.13). However, for any $\Omega(k') \in (\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'$ (implying that $\Omega(k') \in \mathcal{L}_1 \cap \mathcal{L}_2$) and $k' \in \mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)$ (implying that $k' \in \mathcal{L}_2 \setminus \mathcal{L}_1$, since we assume that $\mathcal{I}' \subseteq \mathcal{L}_2$) we have

$$r_{\Omega(k')} \left[\frac{\beta_{\Omega(k')}(\mathbf{y})}{\sigma^2} - \log\left(\frac{\beta_{\Omega(k')}(\mathbf{y})}{\sigma^2}\right) - 1 \right] \geq r_{k'} \left[\frac{\beta_{k'}(\mathbf{y})}{\sigma^2} - \log\left(\frac{\beta_{k'}(\mathbf{y})}{\sigma^2}\right) - 1 \right]. \tag{E.16}$$

Consider the inequality

$$x - (\log(x) + 1) \geq 0, \quad (\text{E.17})$$

valid for every $x \geq 1$, which can be verified by inspecting the derivative of the left hand side. This inequality implies that for any index $k \in \mathcal{L}_2$ (i.e., $\beta_k(\mathbf{y}) \geq \sigma^2$) we have

$$r_k \left[\frac{\beta_k(\mathbf{y})}{\sigma^2} - \log \left(\frac{\beta_k(\mathbf{y})}{\sigma^2} \right) - 1 \right] \geq 0. \quad (\text{E.18})$$

Let us introduce the shorthand $\mathcal{I}'' \triangleq [(\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'] \setminus \Omega(\mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2))$, and observe that

$$\mathcal{I}'' \cup \Omega(\mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)) = (\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}', \quad (\text{E.19})$$

and

$$\mathcal{I}'' \cap \Omega(\mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)) = \emptyset, \quad (\text{E.20})$$

which is due to the obvious fact that $\Omega(\mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)) \subseteq (\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'$. We then have from (E.13) that

$$\begin{aligned} H(\mathbf{x}_3) - H(\mathbf{x}') &\stackrel{(\text{E.13})}{=} \sum_{k \in (\mathcal{L}_1 \cap \mathcal{L}_2) \setminus \mathcal{I}'} r_k \left[\frac{\beta_k(\mathbf{y})}{\sigma^2} - \log \left(\frac{\beta_k(\mathbf{y})}{\sigma^2} \right) - 1 \right] \\ &\quad - \sum_{k' \in \mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)} r_{k'} \left[\frac{\beta_{k'}(\mathbf{y})}{\sigma^2} - \log \left(\frac{\beta_{k'}(\mathbf{y})}{\sigma^2} \right) - 1 \right] \\ &\stackrel{(\text{E.19}), (\text{E.20})}{=} \sum_{k \in \mathcal{I}''} r_k \left[\frac{\beta_k(\mathbf{y})}{\sigma^2} - \log \left(\frac{\beta_k(\mathbf{y})}{\sigma^2} \right) - 1 \right] \\ &\quad + \sum_{k \in \Omega(\mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2))} r_k \left[\frac{\beta_k(\mathbf{y})}{\sigma^2} - \log \left(\frac{\beta_k(\mathbf{y})}{\sigma^2} \right) - 1 \right] \\ &\quad - \sum_{k' \in \mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)} r_{k'} \left[\frac{\beta_{k'}(\mathbf{y})}{\sigma^2} - \log \left(\frac{\beta_{k'}(\mathbf{y})}{\sigma^2} \right) - 1 \right] \\ &= \sum_{k \in \mathcal{I}''} r_k \left[\frac{\beta_k(\mathbf{y})}{\sigma^2} - \log \left(\frac{\beta_k(\mathbf{y})}{\sigma^2} \right) - 1 \right] \\ &\quad + \sum_{k' \in \mathcal{I}' \setminus (\mathcal{L}_1 \cap \mathcal{L}_2)} \left\{ r_{\Omega(k')} \left[\frac{\beta_{\Omega(k')}(\mathbf{y})}{\sigma^2} - \log \left(\frac{\beta_{\Omega(k')}(\mathbf{y})}{\sigma^2} \right) - 1 \right] - r_{k'} \left[\frac{\beta_{k'}(\mathbf{y})}{\sigma^2} - \log \left(\frac{\beta_{k'}(\mathbf{y})}{\sigma^2} \right) - 1 \right] \right\} \end{aligned}$$

$$\begin{aligned}
& \stackrel{\text{(E.16)}}{\geq} \sum_{k \in \mathcal{I}''} r_k \left[\frac{\beta_k(\mathbf{y})}{\sigma^2} - \log \left(\frac{\beta_k(\mathbf{y})}{\sigma^2} \right) - 1 \right] \\
& \stackrel{\text{(E.18)}}{\geq} 0. \tag{E.21}
\end{aligned}$$

Thus, we have verified that $H(\mathbf{x}_3) \geq H(\mathbf{x}')$ for all $\mathbf{x}' \in \mathcal{X}_{S,+}$ such that $\text{supp}(\mathbf{x}') \subseteq \mathcal{L}_2$. This means that the vector \mathbf{x}_3 given by (6.101) yields the global maximum in (6.100), i.e., it yields an ML estimate. \square

Bibliography

- [1] F. Cucker and S. Smale, “On the mathematical foundations of learning,” *Bulletin of the American Mathematical Society*, vol. 39, pp. 1–49, 2002.
- [2] H.-W. Sun and D.-X. Zhou, “Reproducing kernel Hilbert spaces associated with analytic translation-invariant Mercer kernels,” *J. Fourier Anal. Appl.*, vol. 14, no. 1, pp. 89–101, 2008.
- [3] E. Parzen, “Statistical inference on time series by Hilbert space methods, I.” Appl. Math. Stat. Lab., Stanford University, Stanford, CA, Tech. Rep. 23, Jan. 1959.
- [4] P. R. Halmos, *Finite-dimensional vector spaces*. New York: Springer-Verlag, 1974.
- [5] W. Rudin, *Principles of Mathematical Analysis*, 3rd ed. New York: McGraw-Hill, 1976.
- [6] ———, *Real and Complex Analysis*, 3rd ed. New York: McGraw-Hill, 1987.
- [7] V. Cevher, “Approximate distributions for compressible signals,” in *Information Theory Workshop, 2009. ITW 2009. IEEE*, Oct. 2009.
- [8] E. J. Candès, J. Romberg, and T. Tao, “Stable signal recovery from incomplete and inaccurate measurements,” *Comm. Pure Appl. Math.*, vol. 59, no. 8, March 2006.
- [9] M. A. Herman and T. Strohmer, “High-resolution radar via compressed sensing,” *IEEE Trans. Signal Processing*, vol. 57, no. 6, pp. 2275–2284, 2009.
- [10] G. Tauböck and F. Hlawatsch, “A compressed sensing technique for OFDM channel estimation in mobile environments: Exploiting channel sparsity for reducing pilots,” in *Proc. IEEE ICASSP-2008*, Las Vegas, NV, March/Apr. 2008, pp. 2885–2888.
- [11] C. Carbonelli, S. Vedantam, and U. Mitra, “Sparse channel estimation with zero tap detection,” *IEEE Trans. Wireless Comm.*, vol. 6, no. 5, pp. 1743–1763, May 2007.

- [12] D. L. Donoho, “Compressed sensing,” *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.
- [13] E. Candès and M. Wakin, “An introduction to compressive sampling,” *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 21–30, March 2008.
- [14] R. Baraniuk, “Compressive sensing [lecture notes],” *Signal Processing Magazine, IEEE*, vol. 24, no. 4, pp. 118–121, Jul. 2007.
- [15] M. Davenport, P. Boufounos, M. Wakin, and R. Baraniuk, “Signal processing with compressive measurements,” *IEEE J. Sel. Topics Sig. Proc.*, vol. 4, no. 2, pp. 445–460, Apr. 2010.
- [16] J. A. Tropp, “Just relax: Convex programming methods for identifying sparse signals in noise,” *IEEE Trans. Inf. Theory*, vol. 50, no. 3, pp. 1030–1051, March 2004.
- [17] —, “Greed is Good: Algorithmic results for sparse approximation,” *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.
- [18] D. Needell and J. A. Tropp, “CoSaMP: Iterative signal recovery from incomplete and inaccurate samples,” *Appl. Comp. Harmonic Anal.*, vol. 26, pp. 301–321, 2008.
- [19] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.
- [20] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [21] L. L. Scharf, *Statistical Signal Processing*. Reading (MA): Addison Wesley, 1991.
- [22] D. Donoho and I. M. Johnstone, “Minimax risk over ℓ_p balls for ℓ_q -error,” *Probab. Theory Relat. Fields*, vol. 99, pp. 277–303, 1994.
- [23] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaptation by wavelet shrinkage,” *Biometrika*, vol. 81, pp. 425–455, 1994.
- [24] D. Donoho and I. M. Johnstone, “Minimax estimation via wavelet shrinkage,” *Ann. Statist.*, vol. 26, pp. 879–921, 1998.
- [25] Z. Ben-Haim and Y. C. Eldar, “The Cramér–Rao bound for estimating a sparse parameter vector,” *IEEE Trans. Signal Processing*, vol. 58, pp. 3384–3389, June 2010.

- [26] ———, “Performance bounds for sparse estimation with random noise,” in *Proc. IEEE-SP Workshop Statist. Signal Process.*, Cardiff, Wales (UK), Aug. 2009, pp. 225–228.
- [27] H. Cramér, “A contribution to the theory of statistical estimation,” *Skand. Akt. Tidskr.*, vol. 29, pp. 85–94, 1945.
- [28] B. Babadi, N. Kalouptsidis, and V. Tarokh, “Asymptotic achievability of the Cramér–Rao bound for noisy compressive sampling,” *IEEE Trans. Signal Processing*, vol. 57, no. 3, pp. 1233–1236, March 2009.
- [29] J. D. Gorman and A. O. Hero, “Lower bounds for parametric estimation with constraints,” *IEEE Trans. Inf. Theory*, vol. 36, no. 6, pp. 1285–1301, Nov. 1990.
- [30] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge Univ. Press, 2004.
- [31] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero III, “Covariance estimation in decomposable Gaussian graphical models,” *IEEE Trans. Signal Processing*, vol. 58, no. 3, pp. 1482–1492, March 2010.
- [32] P. Rütimann and P. Bühlmann, “High dimensional sparse covariance estimation via directed acyclic graphs,” *Electron. J. Statist.*, vol. 3, pp. 1133–1160, 2009.
- [33] Y. Chen, A. Wiesel, Y. C. Eldar, and A. O. Hero III, “Shrinkage algorithms for mmse covariance estimation,” *IEEE Trans. Signal Processing*, to appear.
- [34] M. J. Wainwright and M. I. Jordan, *Graphical models, exponential families, and variational inference*, ser. Foundations and Trends in Machine Learning. now Publishers Inc., 2008, vol. 1, no. 4.
- [35] D. D. Duttweiler and T. Kailath, “RKHS approach to detection and estimation problems – Part V: Parameter estimation,” *IEEE Trans. Inf. Theory*, vol. 19, no. 1, pp. 29–37, Jan. 1973.
- [36] K. Todros and J. Tabrikian, “General classes of performance lower bounds for parameter estimation – Part I: Non-bayesian bounds for unbiased estimators,” *IEEE Trans. Inf. Theory*, vol. 56, no. 10, pp. 5045–5063, Oct. 2010.
- [37] S. G. Mallat, *A Wavelet Tour of Signal Processing – The Sparse Way*, 3rd ed. San Diego, CA: Academic Press, 2009.

- [38] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins University Press, 1996.
- [39] B. R. Gelbaum and J. M. Olmsted, *Counterexamples in Analysis*. Mineola, NY: Dover Publications, 2003.
- [40] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*, 2nd ed. Boston: Birkhäuser, 2002.
- [41] E. L. Lehmann and G. Casella, *Theory of Point Estimation*, 2nd ed. New York: Springer, 2003.
- [42] R. B. Ash, *Probability and Measure Theory*, 2nd ed. New York: Academic Press, 2000.
- [43] P. Billingsley, *Probability and Measure*, 3rd ed. New York: Wiley, 1995.
- [44] Y. C. Eldar, *Rethinking Biased Estimation: Improving Maximum Likelihood and the Cramér–Rao Bound*, ser. Foundations and Trends in Signal Processing. now Publishers Inc., 2008, vol. 1, no. 4.
- [45] A. Papoulis and S. U. Pillai, *Probability, Random Variables, and Stochastic Processes*, 4th ed. New York: Mc-Graw Hill, 2002.
- [46] S. Verdu and H. Poor, “On minimax robustness: A general approach and applications,” *IEEE Trans. Inf. Theory*, vol. 30, no. 2, pp. 328 – 340, March 1984.
- [47] Z. Ben-Haim and Y. C. Eldar, “Blind minimax estimation,” *IEEE Trans. Inf. Theory*, vol. 53, no. 9, pp. 3145–3157, Sep. 2007.
- [48] G. Raskutti, M. J. Wainwright, and B. Yu, “Minimax rates of estimation for high-dimensional linear regression over ℓ_q -balls,” *ArXiv e-prints*, Oct. 2009, presented in part at the Allerton Conference on Control, Communication and Computing, September 2009, Monticello, IL.
- [49] N. Verzelen, “Minimax risks for sparse regressions: Ultra-high-dimensional phenomenons,” *ArXiv e-prints*, Aug. 2010.
- [50] E. Candès and T. Tao, “The Dantzig selector: Statistical estimation when p is much larger than n ,” pp. 2313–2351, 2007, with discussion.
- [51] N. Meinshausen and B. Yu, “Lasso-type recovery of sparse representations from high dimensional data,” *Annals of Statistics*, pp. 246–270, 2009.

- [52] S. Negahban, P. Ravikumar, M. J. Wainwright, and B. Yu, “A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers,” Department of Statistics, UC Berkeley, Berkeley, CA, Tech. Rep. 797, Oct. 2010.
- [53] Z. Ben-Haim, Y. Eldar, and M. Elad, “Coherence-based performance guarantees for estimating a sparse vector under random noise,” *IEEE Trans. Signal Processing*, vol. 58, no. 10, pp. 5030–5043, Oct. 2010.
- [54] C. Stein, “Unbiased estimates with minimum variance,” *Ann. Math. Statist.*, vol. 21, no. 3, pp. 406–415, 1950.
- [55] E. W. Barankin, “Locally best unbiased estimates,” *Ann. Math. Statist.*, vol. 20, no. 4, pp. 477–501, 1949.
- [56] A. O. Hero III, J. Fessler, and M. Usman, “Exploring estimator bias-variance tradeoffs using the uniform CR Bound,” *IEEE Trans. Signal Processing*, vol. 44, no. 8, pp. 2026–2041, Aug. 1996.
- [57] Y. Eldar, “Minimum variance in biased estimation: Bounds and asymptotically optimal estimators,” *IEEE Trans. Signal Processing*, vol. 52, no. 7, pp. 1915–1930, Jul. 2004.
- [58] P. Stoica and B. C. Ng, “On the Cramér–Rao bound under parametric constraints,” *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 177–179, Jul. 1998.
- [59] Z. Ben-Haim and Y. Eldar, “On the constrained Cramér–Rao bound with a singular Fisher information matrix,” *Signal Processing Letters, IEEE*, vol. 16, no. 6, pp. 453–456, June 2009.
- [60] T. J. Moore, “A theory of Cramér–Rao bounds for constrained parametric models,” Ph.D. dissertation, University of Maryland, 2010.
- [61] L. D. Brown, *Fundamentals of Statistical Exponential Families*, ser. LECTURE NOTES - MONOGRAPH SERIES. Hayward, California: Institute of Mathematical Statistics, 1986.
- [62] N. Aronszajn, “Theory of reproducing kernels,” *Trans. Am. Math. Soc.*, vol. 68, no. 3, pp. 337–404, May 1950.
- [63] P. R. Halmos, *Finite-Dimensional Vector Spaces*. New York: Springer, 1987.
- [64] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. Cambridge, UK: Cambridge Univ. Press, 1991.

- [65] A. W. van der Vaart and J. H. van Zanten, “Reproducing kernel Hilbert spaces of Gaussian priors,” *IMS Collections*, vol. 3, pp. 200–222, 2008.
- [66] M. Hein and O. Bousquet, “Kernels, associated structures and generalizations,” Max Planck Institute for Biological Cybernetics, Tech. Rep. 127, 2004.
- [67] V. I. Paulsen, “An introduction to the theory of reproducing kernel Hilbert spaces,” Houston, TX, 2006. [Online]. Available: <http://www.math.uh.edu/~vern/rkhs.pdf>
- [68] D.-X. Zhou, “Derivative reproducing properties for kernel methods in learning theory,” *J. Fourier Anal. Appl.*, pp. 456–463, 2008.
- [69] I. A. Ibragimov and R. Z. Has’minskii, *Statistical Estimation. Asymptotic Theory*. New York: Springer, 1981.
- [70] R. McAulay and E. Hofstetter, “Barankin bounds on parameter estimation,” *IEEE Trans. Inf. Theory*, vol. 17, no. 6, pp. 669 – 676, Nov. 1971.
- [71] J. Albuquerque, “The Barankin bound: A geometric interpretation,” *IEEE Trans. Inf. Theory*, vol. 19, no. 4, pp. 559–561, Jul. 1973.
- [72] C. R. Rao, “Information and accuracy attainable in the estimation of statistical parameters,” *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–91, 1945.
- [73] P. R. Halmos, *Measure Theory*. New York: Springer, 1974.
- [74] A. O. Hero III, M. Usman, A. Sauve, and J. Fessler, “Recursive algorithms for computing the Cramér–Rao Bound,” *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 803–807, March 1997.
- [75] J. S. Abel, “A bound on mean-square-estimate error,” *IEEE Trans. Inf. Theory*, vol. 39, no. 5, pp. 1675–1680, Sep. 1993.
- [76] A. Bhattacharyya, “On some analogues of the amount of information and their use in statistical estimation,” *Shankya: The Indian Journal of Statistics (1933-1960)*, vol. 8, no. 1, pp. 1–14, 201–211, 211,218, 1946.
- [77] D. G. Chapman and H. Robbins, “Minimum variance estimation without regularity assumptions,” *Ann. Math. Statist.*, vol. 22, no. 4, pp. 581–586, Dec. 1951.
- [78] J. M. Hammersley, “On estimating restricted parameters,” *J. Roy. Statist. Soc. B*, vol. 12, no. 2, pp. 192–240, 1950.

- [79] G. Kail, C. Novak, B. Hofer, and F. Hlawatsch, “Blind monte carlo detection-estimation method for optical coherence tomography,” in *Proc. IEEE ICASSP-2009*, Taipei, Taiwan, Apr. 2009, pp. 493–496.
- [80] M. Dong and L. Tong, “Optimal design and placement of pilot symbols for channel estimation,” *IEEE Trans. Signal Processing*, vol. 50, no. 12, pp. 3055–3069, Dec 2002.
- [81] A. Jung, Z. Ben-Haim, F. Hlawatsch, and Y. C. Eldar, “On unbiased estimation of sparse vectors corrupted by Gaussian noise,” in *Proc. IEEE ICASSP-2010*, Dallas, TX, March 2010, pp. 3990–3993.
- [82] ———, “Unbiased estimation of a sparse vector in white Gaussian noise,” *submitted to IEEE Trans. Inf. Theory*, 2010.
- [83] S. Schmutzhard, A. Jung, F. Hlawatsch, Z. Ben-Haim, and Y. C. Eldar, “A lower bound on the estimator variance for the sparse linear model,” in *Proc. 44th Asilomar Conf. Signals, Systems, Computers*, Pacific Grove, CA, Nov. 2010.
- [84] S. Schmutzhard, A. Jung, and F. Hlawatsch, “Minimum variance estimation for the sparse signal in noise model,” in *Proc. IEEE ISIT-2011*, Jul.-Aug. 2011.
- [85] A. Hormati, A. Karbasi, S. Mohajer, and M. Vetterli, “An estimation theoretic approach for sparsity pattern recovery in the noisy setting,” *submitted to IEEE Trans. Inf. Theory*, 2009.
- [86] A. Karbasi, A. Hormati, S. Mohajer, and M. Vetterli, “Support recovery in compressed sensing: An estimation theoretic approach,” in *Proc. IEEE ISIT-2009*, Jun.-Jul. 2009, pp. 679–683.
- [87] M. J. Wainwright, “Information-theoretic bounds on sparsity recovery in the high-dimensional and noisy setting,” in *Proc. IEEE ISIT-2007*, June 2007, pp. 961–965.
- [88] M. Abramowitz and I. A. Stegun, Eds., *Handbook of Mathematical Functions*. New York: Dover, 1965.
- [89] G. Szegő, *Orthogonal Polynomials*. Providence, Rhode Island: American Mathematical Society, 1939.
- [90] S. S. Chen, D. L. Donoho, Michael, and A. Saunders, “Atomic decomposition by basis pursuit,” *SIAM Journal on Scientific Computing*, vol. 20, pp. 33–61, 1998.
- [91] J. Tropp and A. Gilbert, “Signal recovery from random measurements via orthogonal matching pursuit,” *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.

- [92] E. Candès and J. Romberg, “ ℓ_1 -MAGIC: Recovery of sparse signals via convex programming,” Caltech, Tech. Rep., Oct. 2005.
- [93] D. Needell and R. Vershynin, “Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit,” arXiv:0712.1360v1 [math.NA], Dec. 2007.
- [94] M. Davenport and M. Wakin, “Analysis of orthogonal matching pursuit using the restricted isometry property,” *IEEE Trans. Inf. Theory*, vol. 56, no. 9, pp. 4395–4401, Sept. 2010.
- [95] P. Flandrin, *Time-Frequency/Time-Scale Analysis*. San Diego (CA): Academic Press, 1999.
- [96] F. Hlawatsch, *Time-Frequency Analysis and Synthesis of Linear Signal Spaces: Time-Frequency Filters, Signal Detection and Estimation, and Range-Doppler Estimation*. Boston (MA): Kluwer, 1998.
- [97] S. Haykin, “Cognitive radio: Brain-empowered wireless communication,” *IEEE J. Sel. Areas Comm.*, vol. 23, pp. 201–220, Feb. 2005.
- [98] A. Jung, G. Tauböck, and F. Hlawatsch, “Compressive nonstationary spectral estimation using parsimonious random sampling of the ambiguity function,” in *Proc. IEEE-SP Workshop Statist. Signal Process.*, Cardiff, Wales, UK, Aug.-Sep. 2009, pp. 642–645.
- [99] —, “Compressive spectral estimation for nonstationary random processes,” in *Proc. IEEE ICASSP-2009*, Taipei, Taiwan, April 2009, pp. 3029–3032.
- [100] A. Jung, S. Schmutzhard, F. Hlawatsch, and A. O. Hero III, “Performance bounds for sparse parametric covariance estimation in Gaussian models,” in *Proc. IEEE ICASSP-2011*, Prague, Czech Republic, May 2011.
- [101] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, UK: Cambridge Univ. Press, 1985.
- [102] M. E. Tipping and C. M. Bishop, “Probabilistic principal component analysis,” *J. Roy. Stat. Soc. Ser. B*, vol. 61, pp. 611–622, 1999.
- [103] B. Boashash, Ed., *Time Frequency Signal Analysis and Processing: A Comprehensive Reference*. Amsterdam: Elsevier, 2003.
- [104] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Englewood Cliffs (NJ): Prentice Hall, 1997.
- [105] A. van der Vaart, *Asymptotic Statistics*. Cambridge, UK: Cambridge Univ. Press, 1998.

- [106] P. J. Huber, *Robust Statistics*. New York: Wiley, 1981.
- [107] Z. Ben-Haim, T. Michaeli, and Y. C. Eldar, “Performance bounds and design criteria for estimating finite rate of innovation signals,” *ArXiv e-prints*, Sep. 2010.
- [108] Y. Eldar, P. Kuppinger, and H. Bölcskei, “Block-sparse signals: Uncertainty relations and efficient recovery,” *IEEE Trans. Signal Processing*, vol. 58, no. 6, pp. 3042–3054, June 2010.
- [109] D. Eiwen, G. Tauböck, F. Hlawatsch, and H. G. Feichtinger, “Group sparsity methods for compressive channel estimation in doubly dispersive multicarrier systems,” in *Proc. IEEE SPAWC 2010*, Marrakech, Morocco, Jun. 2010.