Diplomarbeit

Sequential Bayesian Optical Flow Estimation Using Temporal Coherence

ausgeführt zur Erlangung des akademischen Grades
eines Diplom-Ingenieurs unter der Leitung von

Ao.Univ.Prof. Dipl.-Ing. Dr.techn. Franz Hlawatsch
Projektass.(FWF) Ing. Jan Dorazil

am

Institute of Telecommunications

eingereicht an der TU Wien
Fakultät für Elektrotechnik und Informationstechnik

von

Angelika Andrea Hermanek
01225032

Wien, September 29, 2021

Abstract

Optical flow (OF) estimation is an ambitious research field in computer vision that has met with growing interest and has produced an increasing number of publications during the last 35 years. Several algorithms for OF estimation have been developed, and their performance has been continuously improved. OF estimation has a large number of applications, such as the tracking of moving objects and the analysis of the dynamical behavior of objects in an image sequence.

In this thesis, we review the classical deterministic model for OF estimation and subsequently reformulate it in a probabilistic framework. We then extend the probabilistic OF model to account for temporal coherence. With the concept of optimal Bayesian filtering as our basis, we derive the information form of the Kalman filter and the variational Bayesian filter, and we formulate these filters in the specific context of OF estimation. Finally, we study and compare the accuracy and computational complexity of the two filters for synthetic and real image sequences. In particular, we discuss the potential benefit of the temporal coherence assumption for different types of data.

The original contributions of this thesis include the following:

- We develop a probabilistic OF model that combines temporal coherence with a non-linear brightness constancy constraint.
- We present a direct derivation of the information form of the Kalman filter and formulate the filter's prediction and update steps in the context of temporally coherent OF estimation.
- We derive a method for temporally coherent OF estimation based on variational Bayesian filtering.
- We provide simulation results that assess and compare the accuracy and computational complexity of the presented OF estimation methods for several synthetic and real image sequences, and we investigate if and how much temporal coherence improves the results of OF estimation.

Kurzfassung

Die Schätzung des Optical Flow (OF) ist ein anspruchvolles Forschungsgebiet der digitalen Bildverarbeitung, dem in den letzten 35 Jahren große Aufmerksamkeit gewidment wurde, verbunden mit einer zunehmenden Anzahl von Veröffentlichungen. Es wurden zahlreiche Algorithmen für die OF-Schätzung entwickelt, und die Leistungsfähigkeit dieser Algorithmen wurde kontinuierlich verbessert. Für die OF-Schätzung gibt es eine Vielzahl von Anwendungen, z.B. die Verfolgung von sich bewegenden Objekten und die Analyse des dynamischen Verhaltens von Objekten in einer Bildfolge.

In dieser Arbeit wird das klassische deterministische Modell für die OF-Schätzung erläutert und in einem probabilistischen Rahmen neu formuliert. Anschließend wird das probabilistische OF-Modell dahingehend erweitert, dass es die zeitliche Kohärenz von Bildfolgen berücksichtigt. Auf der Grundlage des optimalen Bayesschen Filters werden sodann die Informationsform des Kalman-Filters und das variationale Bayessche Filter hergeleitet und im spezifischen Kontext der OF-Schätzung formuliert. Schließlich werden die Genauigkeit und der Rechenaufwand der beiden Filter für verschiedene synthetische und reale Bildfolgen untersucht und verglichen, wobei auch der potenzielle Nutzen der Annahme zeitlicher Kohärenz für verschiedene Arten von Daten evaluiert wird.

Die Arbeit umfasst die folgenden Hauptbeiträge:

- Ein probabilistisches OF-Modell, das zeitliche Kohärenz mit einer nichtlinearen Modellierung konstanter Helligkeit kombiniert.
- Eine direkte Ableitung der Informationsform des Kalman-Filters sowie eine Formulierung der Prädiktions- und Korrekturschritte des Filters im Kontext einer zeitlich kohärenten OF-Schätzung.
- Eine Methode zur zeitlich kohärenten OF-Schätzung, die auf einem variationalen Bayesschen Filter beruht.
- Simulationsergebnisse, die die Genauigkeit und den Rechenaufwand der beschriebenen OF-Schätzmethoden für mehrere synthetische und reale Bildfolgen angeben und vergleichen. Dabei wird untersucht, ob und wie stark zeitliche Kohärenz die Ergebnisse der OF-Schätzung verbessert.

Acknowledgments

Dieses Dokument beschreibt mehr als der Titel es vermuten lässt. Es beschreibt den Abschluss eines Lebensabschnittes und den Beginn eines neuen. Einen Aufbruch in neue Möglichkeiten und neue Herausforderungen. Gleichzeitig steht es auch für ein Resümee und eine Würdigung der vergangenen Jahre. Ich möchte diese Gelegenheit nutzen, um mich bei allen zu bedanken, die mich auf diesem Weg begleitet haben und in diesen Zeilen mein Herz sprechen lassen.

Diese Danksagung richtet sich an alle, die mit mir gemeinsam Höhen und Tiefen überstanden haben. Die mich unterstützt, motiviert und manchmal auch zurechtgewiesen haben. Manche von euch waren meine ganze Studienzeit über an meiner Seite. Haben jeden Tag mit mir mitgefiebert, sich gemeinsam über Erfolge gefreut und über Fehlschläge geärgert. Ein jeder von euch war immer für mich da und hat mir geholfen, wo es nur möglich war. Ihr seid meine Freunde und meine Familie, ich danke euch aufrichtig von Herzen.

Während des Studiums habe ich auch viele neue und fantastische Leute kennengelernt, ohne die mein Leben nur halb so schön wäre, und die ich auch in der kommenden Zeit nicht vermissen will. Wir haben uns miteinander durch das Studium gekämpft, und das hat uns zusammen geschweißt. Ich denke es ist nicht vermessen zu sagen, dass aus der Studienzeit einmalige Freundschaften entstanden sind, die unendlich wertvoll sind. Auch ihr seid meine Freunde und Familie, ich danke euch aufrichtig von Herzen.

Selbstverständlich gilt mein Dank auch meinen Betreuer und meinem betreuenden Professor. Beide haben mich während der Ausführung dieser Arbeit tatkräftig unterstützt und motiviert. Ihnen ist es zu verdanken, dass diese Arbeit in der heutigen Form vorliegt und ich bin dankbar für alles, was ich während dieser Zeit lernen durfte.

Nun gibt es nichts Weiteres zu sagen. Die Menschen, denen ich diese Zeilen widme wissen, dass sie für sie gedacht sind. Dieses Dokument ist nicht nur für mich, es ist auch für euch. Dieses Dokument, und alles wofür es steht, ist für uns.

Contents

List of Figures

List of Abbreviations

AEE	Average Endpoint Error
CAVI	Coordinate Ascent Variational Inference
CG	Conjugate Gradient
EE	Endpoint Error
ELBO	Evidence Lower Bound
IKF	Information Kalman Filter
KF	Kalman Filter
KL	Kullback-Leibler
MF	Mean Field
ML	Maximum-Likelihood
OF	Optical Flow
pdf	probability density function
VBF	Variational Bayesian Filter
VI	Variational Inference

Chapter 1

Introduction

Optical flow (OF) estimation is a basic but non-trivial task in computer vision. In this thesis, we develop models and methods for sequential OF estimation using spatial and temporal coherence assumptions. Specifically, we establish a linear-Gaussian state-space model for spatially and temporally coherent OF. Based on this model, we develop an information Kalman filter and a variational Bayesian filter for sequential OF estimation. Subsequently, the two filters are compared in terms of accuracy and complexity.

This introductory chapter is organized as follows. In Section 1.1, we explain the concept of OF and mention some application areas. In Section 1.2, we give an overview of the state of the art and refer to related works. Finally, an outline of this thesis and a summary of contributions are provided in Section 1.3.

1.1 Background and Applications

Motion analysis in image sequences is one of the most important tasks in computer vision. It offers the possibility to track objects, quantify deformations, analyze the dynamical behavior of objects, and detect abnormal behavior. The task we consider in this thesis is the estimation of a dense motion field, i.e., the motion field in every pixel in an image sequence. The motion field is identified by the displacement of intensity patterns, called OF. This definition of OF is based on the assumption that the brightness of moving intensity patterns remains constant during the motion.

OF estimation has numerous applications in diverse domains. In multimedia, the video compression standard MPEG uses OF estimation to predict image frames [1]. Applications in biomedical imaging include the quantification of blood flow in arteries or in the heart [2], and the restoration of medical images [3–5]. In robotics, it can be used as an input to control systems for automated navigation [6] or autonomous car driving [7]. Other applications can be

found in facial expression analysis [8], gesture recognition [9], meteorology [10], fluid analysis and mechanics, oceanography, aerodynamics [11–13], and many other areas.

1.2 State of the Art

One of the most fundamental and earliest works in dense OF estimation is [14], which introduced a method for OF estimation by assuming that the brightness of an infinitesimally small point in the image is constant along the trajectory of that point. This assumption is known as the *brightness constancy constraint*. It was furthermore assumed in [14] that neighboring points in an image have similar velocities, i.e., the velocity field is spatially smooth. To obtain a tractable algorithm, the authors of [14] used a linearized version of the brightness constancy constraint, which is, however, only an approximation in general. This approximation tends to be poor for large displacements between successive image frames.

In [15], the authors presented an extension of the dense OF estimation model by introducing a *temporal coherence constraint* that penalizes large frame-to-frame deviations of the OF field. Temporal coherence represents the prior knowledge or the assumption that the movement of an observed object is smooth over time for the entire frame sequence. Accordingly, the OF estimation method proposed in [15] is based on multiple image frames and not only on the previous frame as in frame-to-frame estimation. More specifically, the information form of the Kalman filter was applied to the temporally coherent model, providing an OF *tracking* algorithm that is sequential and recursive in time.

In contrast to the synthetic images that were initially used to evaluate OF estimation methods, realistic scenery also includes transparency, depth discontinuities, independently moving objects, shadows, and specular reflections. Thus, assumptions regarding brightness constancy and spatial smoothness are violated, pushing the algorithms proposed in [14, 15] off their limits. To address this issue, a robust estimation framework for OF estimation was proposed in [16]. This framework relaxes the spatial and temporal coherence assumptions introduced in [14] and [15], respectively, and thus reduces the sensitivity to violations of these assumptions.

In [17], a non-linear formulation of the brightness constancy constraint, extended by a *gradient constancy constraint*, was presented. The resulting method is able to estimate even large displacements, and is moreover robust against illumination variations.

A different robustification of the brightness constancy assumption was described in [18], leading to an improved estimation performance. In addition, [18] presented new numerical algorithms that allow real-time OF estimation for small images.

In [19], the spatial and temporal coherence constraints were decoupled and a new trajectorial regularization that penalizes first-order OF variations along motion trajectories was introduced.

One drawback of this method is that the motion fields at different time instants refer to different images, and thus have to be repeatedly registered onto each other. These registrations can be avoided by registering multiple frames onto one reference frame, as described in [20].

An outlier-robust approach to spatially non-smooth OF estimation that solely requires data and dispenses with the need for an explicit regularization, smoothing, or an additional data term was described in [21]. The authors also presented enhancements in outlier filtering, such that the presented method is suitable to estimate OF for small objects.

In [22], a multi-frame OF estimation method that benefits from long-term temporal cues was presented. The method first warps the OF field from previous frames to the current frame, yielding multiple plausible estimates. Then, the complementary information carried by these estimates is fused into an aggregate OF field.

1.3 Contribution and Outline

The contribution of this thesis is fourfold. First, we extend the temporally coherent OF estimation model of [15] by the non-linear brightness constancy formulation of [17], and we present a novel probabilistic model formulation for OF estimation. Second, we present a direct derivation of the information form of the Kalman filter, and we formulate the filter algorithm for the problem of OF estimation. Third, we derive a method for temporally coherent OF estimation based on the variational Bayesian filtering framework of [23]. Although variational methods for OF estimation have been proposed previously [19, 24], the approach of [23] has not been used in the context of OF estimation so far. Finally, we study the accuracy and complexity of the presented methods for artificial and real image sequences.

This thesis is organized as follows:

- In Chapter 2, we start with a general description of OF. We then develop a Bayesian probabilistic model for OF estimation, and finally extend this model to the estimation of multi-frame OF sequences using temporal coherence assumptions.

- In Chapter 3, we discuss the concept of Bayesian filtering and its application to OF estimation. We first review the general non-linear state-space model and the general Bayesian filter. Subsequently, we describe the linear-Gaussian state-space model and derive the information form of the Kalman filter, which we finally formulate for the problem of OF estimation.

- In Chapter 4, we discuss variational Bayesian filtering and its application to OF estimation. We first review variational Bayesian inference and the mean field approximation. The application of these techniques to Bayesian filtering then leads to the variational

Bayesian filter. Finally, we formulate the variational Bayesian filter for the problem of OF estimation.

- In Chapter 5, we compare the information Kalman filter presented in Chapter 3 and the variational Bayesian filter presented in Chapter 4 in terms of accuracy and complexity. We validate the two filters on artificial and real image sequences and show that they are comparable in accuracy but differ in complexity. We further show that the inclusion of temporal coherence in OF estimation improves the estimation accuracy and reduces the computational complexity for some but not all image sequences.
- In Chapter 6, we summarize our findings and suggest possible directions of future work.

Chapter 2

Optical Flow Models

In this chapter, we discuss deterministic models for spatially coherent OF estimation and show how they can be reformulated in a probabilistic framework and extended to account for temporal coherence. In Section 2.1, we start with a deterministic description of OF based on the brightness constancy constraint, which is followed by a discretization, a linearization, and a reformulation in vector-matrix notation. In Section 2.2, we reformulate the deterministic model in a probabilistic framework. Finally, in Section 2.3, the probabilistic OF model is extended to account for temporal coherence in addition to spatial coherence. The probabilistic OF model developed in this chapter constitutes a basis for the estimation of entire OF sequences subject to both spatial and temporal coherence constraints.

2.1 Deterministic Optical Flow

Let us consider a differentiable space- and time-dependent function $\breve{F}(x, y, t)$ that can be interpreted as the brightness of an image at spatial position (x, y) and time t. We will think of this image as representing a time-dependent material continuum.

2.1.1 Brightness Constancy Constraint

The *brightness constancy constraint* [17] states that the brightness of a particular point in the material continuum is locally constant along its trajectory. Introducing

$$\breve{F}_{\mathrm{P}}(t) \triangleq \breve{F}(x_{\mathrm{P}}(t), y_{\mathrm{P}}(t), t), \tag{2.1}$$

where $(x_{\mathrm{P}}(t), y_{\mathrm{P}}(t))^T$ is the trajectory of the considered point (see Figure 2.1), the brightness constancy constraint postulates

$$\breve{F}_{\mathrm{P}}(t + \Delta t) = \breve{F}_{\mathrm{P}}(t) \tag{2.2}$$

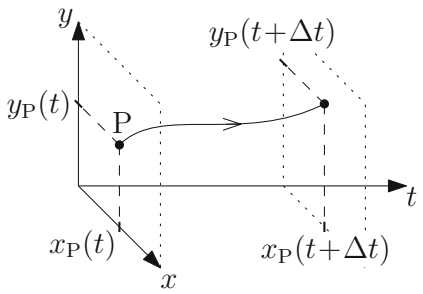


Fig. 2.1: *Movement of point "P" along its trajectory in the material continuum.*

for a sufficiently small Δt. Using the definition of $\breve{F}_P(t)$ in (2.1), we can equivalently write (2.2) as

$$\breve{F}(x_P(t+\Delta t), y_P(t+\Delta t), t+\Delta t) - \breve{F}(x_P(t), y_P(t), t) = 0. \tag{2.3}$$

Defining the functions $\breve{u}(x,y,t)$ and $\breve{v}(x,y,t)$ as

$$\breve{u}(x_P(t), y_P(t), t) \triangleq x_P(t+\Delta t) - x_P(t), \quad \breve{v}(x_P(t), y_P(t), t) \triangleq y_P(t+\Delta t) - y_P(t), \tag{2.4}$$

we have $x_P(t+\Delta t) = x_P(t) + \breve{u}(x_P(t), y_P(t), t)$ and $y_P(t+\Delta t) = y_P(t) + \breve{v}(x_P(t), y_P(t), t)$. By setting $x = x_P(t)$ and $y = y_P(t)$, we thus can rewrite (2.3) as

$$\breve{F}(x + \breve{u}(x,y,t), y + \breve{v}(x,y,t), t+\Delta t) - \breve{F}(x,y,t) = 0. \tag{2.5}$$

The displacement vector

$$\breve{\boldsymbol{w}}(x,y,t) \triangleq \begin{pmatrix} \breve{u}(x,y,t) \\ \breve{v}(x,y,t) \end{pmatrix} \tag{2.6}$$

of a point in the material continuum that at time t resides at position (x,y) is referred to as the *OF vector*. The task to be considered in later chapters is the estimation of this OF vector based on the brightness constancy constraint in equation (2.5) and further modeling assumptions.

2.1.2 Discretization

For reasons of practical feasibility, we reformulate the problem in a discretized setting. First, we discretize the image $\breve{F}(x,y,t)$ on a uniform grid,

$$x_i \triangleq (i-1)\Delta x, \quad i = 1, \ldots, N, \tag{2.7}$$

$$y_j \triangleq (j-1)\Delta y, \quad j = 1, \ldots, M, \tag{2.8}$$

$$t_k \triangleq (k-1)\Delta t, \quad k = 1, \ldots, L, \tag{2.9}$$

where, in particular, $x_1 = y_1 = t_1 = 0$. The discretized image is then obtained as

$$F[i,j,k] \triangleq \breve{F}(x_i, y_j, t_k) = \breve{F}((i-1)\Delta x, (j-1)\Delta y, (k-1)\Delta t), \tag{2.10}$$

for $i = 1, ..., N$, $j = 1, ..., M$, and $k = 1, ..., L$. The OF functions are similarly discretized by

$$u[i,j,k] \triangleq \breve{u}(x_i, y_j, t_k) \quad \text{and} \quad v[i,j,k] \triangleq \breve{v}(x_i, y_j, t_k), \tag{2.11}$$

for $i = 1, ..., N$, $j = 1, ..., M$, and $k = 1, ..., L$.

Let us evaluate equation (2.5) for $x = x_i$, $y = y_j$, and $t = t_k$. We obtain

$$\breve{F}(x_i + \breve{u}(x_i, y_j, t_k), y_j + \breve{v}(x_i, y_j, t_k), t_k + \Delta t) - \breve{F}(x_i, y_j, t_k) = 0. \tag{2.12}$$

Using (2.10) and (2.11), and noting that $t_k + \Delta t = t_{k+1}$, equation (2.12) can then be reformulated as

$$\breve{F}(x_i + u[i,j,k], y_j + v[i,j,k], t_{k+1}) - F[i,j,k] = 0, \tag{2.13}$$

or equivalently

$$\breve{F}(x_i(u[i,j,k]), y_j(v[i,j,k]), t_{k+1}) - F[i,j,k] = 0, \tag{2.14}$$

with

$$x_i(u[i,j,k]) \triangleq x_i + u[i,j,k] \quad \text{and} \quad y_j(v[i,j,k]) \triangleq y_j + v[i,j,k]. \tag{2.15}$$

2.1.3 Linearization

In view of the fact that $\breve{F}(x, y, t)$ is a non-linear function of x and y, the left-hand side of (2.14) is a non-linear function of the discrete OF functions $u[i,j,k]$ and $v[i,j,k]$. In order to simplify the numerical computation of the OF functions, we perform a linearization of (2.14) about discrete reference points $u_0[i,j,k]$ and $v_0[i,j,k]$. For simplicity, let us use the shorthands u, v, u_0, and v_0 for $u[i,j,k]$, $v[i,j,k]$, $u_0[i,j,k]$, and $v_0[i,j,k]$, respectively. The left-hand side of (2.14) is then approximated as

$$\begin{aligned}\breve{F}(x_i(u), y_j(v), t_{k+1}) \approx \breve{F}(x_i(u_0), y_j(v_0), t_{k+1}) &+ \frac{\partial}{\partial u}\breve{F}(x_i(u), y_j(v_0), t_{k+1})\Big|_{u_0}(u - u_0) \\ &+ \frac{\partial}{\partial v}\breve{F}(x_i(u_0), y_j(v), t_{k+1})\Big|_{v_0}(v - v_0).\end{aligned} \tag{2.16}$$

Using the chain rule, we obtain

$$\frac{\partial}{\partial u}\breve{F}(x_i(u), y_j(v_0), t_{k+1}) = \breve{F}_x(x_i(u), y_j(v_0), t_{k+1})\frac{\mathrm{d}x_i(u)}{\mathrm{d}u} \tag{2.17}$$

and

$$\frac{\partial}{\partial v}\breve{F}(x_i(u_0), y_j(v), t_{k+1}) = \breve{F}_y(x_i(u_0), y_j(v), t_{k+1})\frac{\mathrm{d}y_j(v)}{\mathrm{d}v}, \tag{2.18}$$

with

$$\breve{F}_x(x, y, t) \triangleq \frac{\partial \breve{F}(x, y, t)}{\partial x}, \quad \breve{F}_y(x, y, t) \triangleq \frac{\partial \breve{F}(x, y, t)}{\partial y}. \tag{2.19}$$

Since, according to (2.15), we have $\mathrm{d}x_i(u)/\mathrm{d}u = \mathrm{d}y_j(v)/\mathrm{d}v = 1$, equation (2.16) can be rewritten as

$$\begin{aligned}\breve{F}(x_i(u), y_j(v), t_{k+1}) \approx{}& \breve{F}(x_i(u_0), y_j(v_0), t_{k+1}) + \breve{F}_x(x_i(u_0), y_j(v_0), t_{k+1})(u - u_0)\\ &+ \breve{F}_y(x_i(u_0), y_j(v_0), t_{k+1})(v - v_0).\end{aligned} \tag{2.20}$$

Using this approximation in the left-hand side of (2.14), we obtain the desired linearized version of (2.14) as

$$\begin{aligned}&\breve{F}(x_i(u_0), y_j(v_0), t_{k+1}) + \breve{F}_x(x_i(u_0), y_j(v_0), t_{k+1})(u - u_0) + \breve{F}_y(x_i(u_0), y_j(v_0), t_{k+1})(v - v_0)\\ &\quad - F[i, j, k] \approx 0,\end{aligned} \tag{2.21}$$

or equivalently

$$\breve{F}[i, j, k+1] + \breve{F}_x[i, j, k+1](u - u_0) + \breve{F}_y[i, j, k+1](v - v_0) - F[i, j, k] \approx 0, \tag{2.22}$$

with the shorthand notation

$$\breve{F}[i, j, k+1] \triangleq \breve{F}(x_i(u_0[i, j, k]), y_j(v_0[i, j, k]), t_{k+1}). \tag{2.23}$$

Finally, to obtain a completely discrete formulation, we approximate the differential quotients $\breve{F}_x(x, y, t)$ and $\breve{F}_y(x, y, t)$ in (2.19), evaluated at $x = x_i(u_0)$, $y = y_j(v_0)$, and $t = t_{k+1}$, by finite difference quotients. Therefore, we formally replace $\breve{F}_x(x_i(u_0), y_j(v_0), t_{k+1})$ by

$$F_x[i, j, k+1] \triangleq \begin{cases} \dfrac{\breve{F}[i+1, j, k+1] - \breve{F}[i-1, j, k+1]}{2\,\Delta x}, & i = 2, \ldots, N-1,\\[2ex] \dfrac{\breve{F}[i+1, j, k+1] - \breve{F}[i, j, k+1]}{\Delta x}, & i = 1,\\[2ex] \dfrac{\breve{F}[i, j, k+1] - \breve{F}[i-1, j, k+1]}{\Delta x}, & i = N, \end{cases} \tag{2.24}$$

and $\breve{F}_y(x_i(u_0), y_j(v_0), t_{k+1})$ by

$$F_y[i,j,k+1] \triangleq \begin{cases} \dfrac{\breve{F}[i,j+1,k+1] - \breve{F}[i,j-1,k+1]}{2\,\Delta y}, & j = 2,\ldots,M-1, \\ \dfrac{\breve{F}[i,j+1,k+1] - \breve{F}[i,j,k+1]}{\Delta y}, & j = 1, \\ \dfrac{\breve{F}[i,j,k+1] - \breve{F}[i,j-1,k+1]}{\Delta y}, & j = M. \end{cases} \tag{2.25}$$

With these replacements, equation (2.22) finally becomes

$$F_x[i,j,k+1](u-u_0) + F_y[i,j,k+1](v-v_0) + F_t[i,j,k] \approx 0, \tag{2.26}$$

where

$$F_t[i,j,k] \triangleq \breve{F}[i,j,k+1] - F[i,j,k]. \tag{2.27}$$

2.1.4 Vector-Matrix Formulation

Next, we develop a vector-matrix formulation of the discretized brightness constancy equation (2.26). For any $k = 1, ..., L$, we define the vector $\boldsymbol{f}_k$ of dimension NM by stacking the columns of $F[i,j,k]$ (viewed as a matrix with row index j and column index i), i.e.,

$$(\boldsymbol{f}_k)_{(i-1)M+j} \triangleq F[i,j,k] = \breve{F}(x_i, y_j, t_k), \tag{2.28}$$

for $i = 1,\ldots,N$ and $j = 1,\ldots,M$. Analogously, we define the vectors $\boldsymbol{f}_{x,k+1}$, $\boldsymbol{f}_{y,k+1}$, $\boldsymbol{f}_{t,k}$, $\boldsymbol{u}_k$, $\boldsymbol{v}_k$, $\boldsymbol{u}_{k,0}$, and $\boldsymbol{v}_{k,0}$ as stacked versions of $F_x[i,j,k+1]$, $F_y[i,j,k+1]$, $F_t[i,j,k]$, $u[i,j,k]$, $v[i,j,k]$, $u_0[i,j,k]$, and $v_0[i,j,k]$, respectively. We also define the stacked vectors

$$\boldsymbol{w}_k \triangleq \begin{pmatrix} \boldsymbol{u}_k \\ \boldsymbol{v}_k \end{pmatrix} \quad \text{and} \quad \boldsymbol{w}_{k,0} \triangleq \begin{pmatrix} \boldsymbol{u}_{k,0} \\ \boldsymbol{v}_{k,0} \end{pmatrix}. \tag{2.29}$$

Finally, we define the diagonal matrices of dimension $NM \times NM$ that have the entries of the vectors $\boldsymbol{f}_{x,k+1}$ and $\boldsymbol{f}_{x,k+1}$ on their main diagonal, i.e.,

$$\boldsymbol{F}_{x,k+1} \triangleq \operatorname{diag}\{\boldsymbol{f}_{x,k+1}\} \quad \text{and} \quad \boldsymbol{F}_{y,k+1} \triangleq \operatorname{diag}\{\boldsymbol{f}_{y,k+1}\}, \tag{2.30}$$

and we combine $\boldsymbol{F}_{x,k+1}$ and $\boldsymbol{F}_{y,k+1}$ into the $NM \times 2NM$ matrix

$$\boldsymbol{G}_{k+1} \triangleq (\boldsymbol{F}_{x,k+1}\ \boldsymbol{F}_{y,k+1}). \tag{2.31}$$

Equation (2.26) can now be written in vector-matrix form as

$$\boldsymbol{F}_{x,k+1}(\boldsymbol{u}_k - \boldsymbol{u}_{k,0}) + \boldsymbol{F}_{y,k+1}(\boldsymbol{v}_k - \boldsymbol{v}_{k,0}) + \boldsymbol{f}_{t,k} \approx \boldsymbol{0}, \tag{2.32}$$

or more compactly as

$$\boldsymbol{G}_{k+1}(\boldsymbol{w}_k - \boldsymbol{w}_{k,0}) + \boldsymbol{f}_{t,k} \approx \boldsymbol{0}. \tag{2.33}$$

2.1.5 Spatial Regularization

An estimate $\hat{\boldsymbol{w}}_k$ of the OF vector $\boldsymbol{w}_k$ for a given k can, in principle, be obtained by minimizing the squared norm of the left-hand side of (2.33), since (2.33) implies that that squared norm is close to zero. That is,

$$\hat{\boldsymbol{w}}_k = \underset{\boldsymbol{w}_k}{\operatorname{argmin}}\ E_{\mathrm{d}}^{(k)}(\boldsymbol{w}_k), \tag{2.34}$$

with

$$E_{\mathrm{d}}^{(k)}(\boldsymbol{w}_k) \triangleq \|\boldsymbol{G}_{k+1}(\boldsymbol{w}_k - \boldsymbol{w}_{k,0}) + \boldsymbol{f}_{t,k}\|^2 = \|\boldsymbol{G}_{k+1}\boldsymbol{w}_k - \boldsymbol{d}_k\|^2, \tag{2.35}$$

and

$$\boldsymbol{d}_k \triangleq \boldsymbol{G}_{k+1}\boldsymbol{w}_{k,0} - \boldsymbol{f}_{t,k}. \tag{2.36}$$

According to [25, Section 11], the solution to this least squares problem is obtained as a solution of the normal equation

$$\boldsymbol{G}_{k+1}^T\boldsymbol{G}_{k+1}\boldsymbol{w}_k = \boldsymbol{G}_{k+1}^T\boldsymbol{d}_k. \tag{2.37}$$

However, since according to (2.31) $\boldsymbol{G}_{k+1}$ is of dimension $NM \times 2NM$, the rank of the $2NM \times 2NM$ Gram matrix $\boldsymbol{G}_{k+1}^T\boldsymbol{G}_{k+1}$ is at most NM. Therefore $\boldsymbol{G}_{k+1}^T\boldsymbol{G}_{k+1}$ is rank deficient and the normal equation (2.37) is under-determined. Thus, the minimization in (2.34) is an ill-posed problem.

As in [26], we address this issue by enforcing spatial smoothness of the vectors $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$. This can be done by augmenting $E_{\mathrm{d}}^{(k)}(\boldsymbol{w}_k)$ in (2.35) by a regularization term

$$E_{\mathrm{r}}(\boldsymbol{w}_k) \triangleq \|\boldsymbol{u}_{H,k}\|^2 + \|\boldsymbol{v}_{H,k}\|^2, \tag{2.38}$$

where $\boldsymbol{u}_{H,k}$ and $\boldsymbol{v}_{H,k}$ are defined by

$$\boldsymbol{u}_{H,k} \triangleq \boldsymbol{H}\boldsymbol{u}_k \quad \text{and} \quad \boldsymbol{v}_{H,k} \triangleq \boldsymbol{H}\boldsymbol{v}_k. \tag{2.39}$$

The matrix $\boldsymbol{H}$ represents a linear filter matrix that passes the "non-smooth" components of the OF vectors $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$. For spatially smooth OF vectors $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$, the filtered vectors $\boldsymbol{u}_{H,k}$ and

$\boldsymbol{v}_{H,k}$ have small values, and thus the regularization term $E_{\mathrm{r}}(\boldsymbol{w}_k)$ will be small. In this thesis, following [27], $\boldsymbol{H}$ is defined as the discrete counterpart of the gradient operator ∇, i.e.,

$$\boldsymbol{H} \triangleq \begin{pmatrix} \boldsymbol{H}_x \\ \boldsymbol{H}_y \end{pmatrix}. \tag{2.40}$$

Here,

$$\boldsymbol{H}_x \triangleq \frac{1}{\Delta x} \boldsymbol{S}_N \otimes \mathbf{I}_M, \tag{2.41}$$

with the $N \times N$ matrix

$$\boldsymbol{S}_N \triangleq \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ & & & -1 & 1 \end{pmatrix}, \tag{2.42}$$

and

$$\boldsymbol{H}_y \triangleq \frac{1}{\Delta y} \mathbf{I}_N \otimes \boldsymbol{S}_M, \tag{2.43}$$

where $\boldsymbol{S}_M$ is constructed equivalently to $\boldsymbol{S}_N$ but with N replaced by M. Note that $\otimes$ denotes the Kronecker product. Expanding (2.38) yields

$$E_{\mathrm{r}}(\boldsymbol{w}_k) = (\boldsymbol{H}\boldsymbol{u}_k)^T \boldsymbol{H}\boldsymbol{u}_k + (\boldsymbol{H}\boldsymbol{v}_k)^T \boldsymbol{H}\boldsymbol{v}_k = \boldsymbol{u}_k^T \boldsymbol{\Lambda}_H \boldsymbol{u}_k + \boldsymbol{v}_k^T \boldsymbol{\Lambda}_H \boldsymbol{v}_k, \tag{2.44}$$

where

$$\boldsymbol{\Lambda}_H \triangleq \boldsymbol{H}^T \boldsymbol{H} = \begin{pmatrix} \boldsymbol{H}_x^T & \boldsymbol{H}_y^T \end{pmatrix} \begin{pmatrix} \boldsymbol{H}_x \\ \boldsymbol{H}_y \end{pmatrix} = \boldsymbol{H}_x^T \boldsymbol{H}_x + \boldsymbol{H}_y^T \boldsymbol{H}_y. \tag{2.45}$$

By setting

$$\boldsymbol{M}_H \triangleq \begin{pmatrix} \boldsymbol{\Lambda}_H & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Lambda}_H \end{pmatrix}, \tag{2.46}$$

expression (2.44) can be written compactly as

$$E_{\mathrm{r}}(\boldsymbol{w}_k) = \boldsymbol{w}_k^T \boldsymbol{M}_H \boldsymbol{w}_k. \tag{2.47}$$

Augmenting $E_{\mathrm{d}}^{(k)}(\boldsymbol{w}_k)$ by $E_{\mathrm{r}}(\boldsymbol{w}_k)$, the overall objective function is then defined by (cf. (2.35) and (2.47))

$$E_{\mathrm{d\text{-}r}}^{(k)}(\boldsymbol{w}_k) \triangleq E_{\mathrm{d}}^{(k)}(\boldsymbol{w}_k) + \alpha E_{\mathrm{r}}(\boldsymbol{w}_k) = \|\boldsymbol{G}_{k+1}(\boldsymbol{w}_k - \boldsymbol{w}_{k,0}) + \boldsymbol{f}_{t,k}\|^2 + \alpha \boldsymbol{w}_k^T \boldsymbol{M}_H \boldsymbol{w}_k, \tag{2.48}$$

where $\alpha > 0$ is a parameter that balances the influence of $E_{\mathrm{d}}^{(k)}(\boldsymbol{w}_k)$ and $E_{\mathrm{r}}(\boldsymbol{w}_k)$. The regularized minimization – replacing (2.34) – is then formulated as

$$\hat{\boldsymbol{w}}_{k,\mathrm{d\text{-}r}} = \underset{\boldsymbol{w}_k}{\operatorname{argmin}}\, E_{\mathrm{d\text{-}r}}^{(k)}(\boldsymbol{w}_k) = \underset{\boldsymbol{w}_k}{\operatorname{argmin}} \{\|\boldsymbol{G}_{k+1}(\boldsymbol{w}_k - \boldsymbol{w}_{k,0}) + \boldsymbol{f}_{t,k}\|^2 + \alpha \boldsymbol{w}_k^T \boldsymbol{M}_H \boldsymbol{w}_k\}, \tag{2.49}$$

for $k = 1, ..., L-1$. The solution to this minimization problem can be shown to be

$$\hat{\boldsymbol{w}}_{k,\text{d-r}} = (\boldsymbol{G}_{k+1}^T \boldsymbol{G}_{k+1} + \alpha \boldsymbol{M}_H)^{-1} \boldsymbol{G}_{k+1}^T \boldsymbol{d}_k, \tag{2.50}$$

where again $\boldsymbol{d}_k = \boldsymbol{G}_{k+1}\boldsymbol{w}_{k,0} - \boldsymbol{f}_{t,k}$ (see (2.36)).

2.2 Probabilistic Optical Flow

In this section, we reformulate the optimization problem in (2.49) in a probabilistic setting.

2.2.1 Measurement Model

We can rewrite (2.33) as

$$\boldsymbol{G}_{k+1}\boldsymbol{w}_k \approx \boldsymbol{d}_k, \tag{2.51}$$

with the NM-dimensional "data vector" $\boldsymbol{d}_k = \boldsymbol{G}_{k+1}\boldsymbol{w}_{k,0} - \boldsymbol{f}_{t,k}$. This can furthermore be formulated as the exact relation

$$\boldsymbol{d}_k = \boldsymbol{G}_{k+1}\boldsymbol{w}_k + \boldsymbol{n}_k, \tag{2.52}$$

where $\boldsymbol{n}_k = \boldsymbol{d}_k - \boldsymbol{G}_{k+1}\boldsymbol{w}_k$ reflects the approximation error. We can interpret (2.52) as a *measurement model* with "data vector" $\boldsymbol{d}_k$ and "measurement noise" $\boldsymbol{n}_k$. Note that (2.51) implies that $\boldsymbol{n}_k$ is small.

We now pass from the deterministic OF description developed so far to a probabilistic description by modeling the measurement noise $\boldsymbol{n}_k$ as a random vector. More specifically, we assume that $\boldsymbol{n}_k$ is an iid (across entries and across time k) zero-mean Gaussian random vector with variance λ^{-1}, i.e.,

$$\boldsymbol{n}_k \sim \mathcal{N}(\boldsymbol{0}, \lambda^{-1}\mathbf{I}_{NM}). \tag{2.53}$$

From (2.52) and (2.53), the conditional probability distribution function (pdf) of $\boldsymbol{d}_k$ is obtained as

$$p(\boldsymbol{d}_k|\boldsymbol{w}_k) = \mathcal{N}(\boldsymbol{d}_k; \boldsymbol{G}_{k+1}\boldsymbol{w}_k, \lambda^{-1}\mathbf{I}_{NM}) = \Big(\frac{\lambda}{2\pi}\Big)^{NM/2} \exp\Big(-\frac{\lambda}{2}\|\boldsymbol{d}_k - \boldsymbol{G}_{k+1}\boldsymbol{w}_k\|^2\Big). \tag{2.54}$$

We furthermore assume that $\boldsymbol{n}_k$ is statistically independent of $\boldsymbol{w}_k$ and also of all $\boldsymbol{w}_{k'}$ with $k' \neq k$.

The maximum-likelihood (ML) estimate of the OF vector could in theory be obtained by maximizing (2.54) with respect to $\boldsymbol{w}_k$ for a fixed (observed) data vector $\boldsymbol{d}_k$. This is easily verified to reduce to the minimization in (2.34), which was recognized earlier to be an an ill-posed problem.

2.2.2 Augmented Measurement Model

We address this issue by incorporating a probabilistic formulation of the spatial regularization discussed in Section 2.1.5. Whereas regularization terms are usually incorporated in probabilistic formulations via a suitable definition of the prior pdf, we here follow [15] and introduce additional "pseudo-measurements" $\boldsymbol{\epsilon}_{u,k}$ and $\boldsymbol{\epsilon}_{v,k}$, defined as noisy versions of the filtered OF vectors $\boldsymbol{u}_{H,k}$ and $\boldsymbol{v}_{H,k}$, respectively, i.e.,

$$\boldsymbol{\epsilon}_{u,k} \triangleq \boldsymbol{u}_{H,k} + \boldsymbol{m}_{u,k} \quad \text{and} \quad \boldsymbol{\epsilon}_{v,k} \triangleq \boldsymbol{v}_{H,k} + \boldsymbol{m}_{v,k}, \tag{2.55}$$

where $\boldsymbol{u}_{H,k} = \boldsymbol{H}\boldsymbol{u}_k$ and $\boldsymbol{v}_{H,k} = \boldsymbol{H}\boldsymbol{v}_k$ as introduced in (2.39). The measurement noise vectors $\boldsymbol{m}_{u,k}$ and $\boldsymbol{m}_{v,k}$ are assumed to be mutually independent, independent of $\boldsymbol{n}_k$, zero-mean Gaussian, iid across entries and across time with "small" variance β^{-1}, i.e.,

$$\boldsymbol{m}_{u,k} \sim \mathcal{N}(\boldsymbol{0}, \beta^{-1}\mathbf{I}_{2NM}) \quad \text{and} \quad \boldsymbol{m}_{v,k} \sim \mathcal{N}(\boldsymbol{0}, \beta^{-1}\mathbf{I}_{2NM}). \tag{2.56}$$

We now combine the OF measurement model (2.52) and the "pseudo-measurement" model (2.55) into the augmented measurement model

$$\tilde{\boldsymbol{d}}_k = \tilde{\boldsymbol{G}}_{k+1}\boldsymbol{w}_k + \tilde{\boldsymbol{n}}_k, \tag{2.57}$$

with the augmented data vector $\tilde{\boldsymbol{d}}_k$, the augmented measurement matrix $\tilde{\boldsymbol{G}}_{k+1}$, and the augmented measurement noise vector $\tilde{\boldsymbol{n}}_k$ defined as

$$\tilde{\boldsymbol{d}}_k \triangleq \begin{pmatrix} \boldsymbol{d}_k \\ \boldsymbol{\epsilon}_{u,k} \\ \boldsymbol{\epsilon}_{v,k} \end{pmatrix}, \quad \tilde{\boldsymbol{G}}_{k+1} \triangleq \begin{pmatrix} \boldsymbol{F}_{x,k+1} & \boldsymbol{F}_{y,k+1} \\ \boldsymbol{H} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H} \end{pmatrix}, \quad \tilde{\boldsymbol{n}}_k \triangleq \begin{pmatrix} \boldsymbol{n}_k \\ \boldsymbol{m}_{u,k} \\ \boldsymbol{m}_{v,k} \end{pmatrix}. \tag{2.58}$$

Using (2.53), (2.56), and the independence of $\boldsymbol{n}_k$, $\boldsymbol{m}_{u,k}$, and $\boldsymbol{m}_{v,k}$, we see that $\tilde{\boldsymbol{n}}_k$ is zero-mean Gaussian, iid across entries and across time. In particular,

$$\tilde{\boldsymbol{n}}_k \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{Q}^{-1}) \tag{2.59}$$

with the augmented precision matrix

$$\boldsymbol{Q} \triangleq \begin{pmatrix} \lambda\mathbf{I}_{NM} & \boldsymbol{0} \\ \boldsymbol{0} & \beta\mathbf{I}_{4NM} \end{pmatrix}. \tag{2.60}$$

From (2.57) and (2.59), the conditional pdf of the augmented data vector $\tilde{\boldsymbol{d}}_k$ is therefore obtained

as

$$p(\tilde{\boldsymbol{d}}_k|\boldsymbol{w}_k) = \mathcal{N}(\tilde{\boldsymbol{d}}_k; \tilde{\boldsymbol{G}}_{k+1}\boldsymbol{w}_k, \boldsymbol{Q}^{-1}) \\ = \Big(\frac{\lambda}{2\pi}\Big)^{NM/2}\Big(\frac{\beta}{2\pi}\Big)^{2NM} \exp\Big(-\frac{1}{2}(\tilde{\boldsymbol{d}}_k - \tilde{\boldsymbol{G}}_{k+1}\boldsymbol{w}_k)^T\boldsymbol{Q}(\tilde{\boldsymbol{d}}_k - \tilde{\boldsymbol{G}}_{k+1}\boldsymbol{w}_k)\Big). \tag{2.61}$$

For a fixed augmented data vector $\tilde{\boldsymbol{d}}_k$ i.e., fixed data vector $\boldsymbol{d}_k$ and fixed pseudo-measurements $\boldsymbol{\epsilon}_{u,k}$ and $\boldsymbol{\epsilon}_{v,k}$, equation (2.61) is the likelihood function corresponding to the augmented measurement model.

2.2.3 Probabilistic Formulation of Spatial Regularization

Let us recall that the constraint of spatial smoothness of the vectors $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$ was expressed in Section 2.1.5 via our assumption that $\boldsymbol{u}_{H,k}$ and $\boldsymbol{v}_{H,k}$ have a small norm (cf. (2.38)). Because of (2.55), we can express this by setting the pseudo-measurements $\boldsymbol{\epsilon}_{u,k}$ and $\boldsymbol{\epsilon}_{v,k}$ to zero, i.e., $\boldsymbol{\epsilon}_{u,k} = \boldsymbol{\epsilon}_{v,k} = \boldsymbol{0}$. Indeed, (2.55) here yields

$$\boldsymbol{u}_{H,k} = -\boldsymbol{m}_{u,k} \quad \text{and} \quad \boldsymbol{v}_{H,k} = -\boldsymbol{m}_{v,k}. \tag{2.62}$$

Because β^{-1} is assumed small, $\boldsymbol{m}_{u,k}$ and $\boldsymbol{m}_{v,k}$ are small with high probability, and thus (2.62) implies that $\boldsymbol{u}_{H,k}$ and $\boldsymbol{v}_{H,k}$ will be small with high probability. The augmented data vector $\tilde{\boldsymbol{d}}_k$ is then given by

$$\tilde{\boldsymbol{d}}_k = \begin{pmatrix} \boldsymbol{d}_k \\ \boldsymbol{0}_{NM} \\ \boldsymbol{0}_{NM} \end{pmatrix}. \tag{2.63}$$

Using this augmented data vector, we can furthermore reformulate the augmented likelihood function $p(\tilde{\boldsymbol{d}}_k|\boldsymbol{w}_k)$ in (2.61). To this end, we first expand the quadratic form in the argument of the exponential function in (2.61), i.e.,

$$(\tilde{\boldsymbol{d}}_k - \tilde{\boldsymbol{G}}_{k+1}\boldsymbol{w}_k)^T\boldsymbol{Q}(\tilde{\boldsymbol{d}}_k - \tilde{\boldsymbol{G}}_{k+1}\boldsymbol{w}_k) = \tilde{\boldsymbol{d}}_k^T\boldsymbol{Q}\tilde{\boldsymbol{d}}_k - \tilde{\boldsymbol{d}}_k^T\boldsymbol{Q}\tilde{\boldsymbol{G}}_{k+1}\boldsymbol{w}_k - \boldsymbol{w}_k^T\tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{d}}_k^T \\ + \boldsymbol{w}_k^T\tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{G}}_{k+1}\boldsymbol{w}_k. \tag{2.64}$$

Using (2.58) and (2.60), the term $\tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{G}}_{k+1}$ in (2.64) can be rewritten as

$$\tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{G}}_{k+1} = \begin{pmatrix} \boldsymbol{F}_{x,k+1}^T & \boldsymbol{H}^T & \boldsymbol{0} \\ \boldsymbol{F}_{y,k+1}^T & \boldsymbol{0} & \boldsymbol{H}^T \end{pmatrix}\begin{pmatrix} \lambda\mathbf{I}_{NM} & \boldsymbol{0} \\ \boldsymbol{0} & \beta\mathbf{I}_{4NM} \end{pmatrix}\begin{pmatrix} \boldsymbol{F}_{x,k+1} & \boldsymbol{F}_{y,k+1} \\ \boldsymbol{H} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{H} \end{pmatrix} = \beta\boldsymbol{M}_H + \lambda\boldsymbol{G}_{k+1}^T\boldsymbol{G}_{k+1}, \tag{2.65}$$

where also (2.31), (2.45), and (2.46) were used. Furthermore, using (2.63), the term $\tilde{\boldsymbol{G}}_{k+1}^T \boldsymbol{Q} \tilde{\boldsymbol{d}}_k$ in (2.64) can be rewritten as

$$\tilde{\boldsymbol{G}}_{k+1}^T \boldsymbol{Q} \tilde{\boldsymbol{d}}_k = \begin{pmatrix} \boldsymbol{F}_{x,k+1}^T & \boldsymbol{H}^T & \boldsymbol{0} \\ \boldsymbol{F}_{y,k+1}^T & \boldsymbol{0} & \boldsymbol{H}^T \end{pmatrix} \begin{pmatrix} \lambda \mathbf{I}_{NM} & \boldsymbol{0} \\ \boldsymbol{0} & \beta \mathbf{I}_{4NM} \end{pmatrix} \begin{pmatrix} \boldsymbol{d}_k \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} = \lambda \boldsymbol{G}_{k+1}^T \boldsymbol{d}_k. \tag{2.66}$$

Finally, we rewrite the term $\tilde{\boldsymbol{d}}_k^T \boldsymbol{Q} \tilde{\boldsymbol{d}}_k$ in (2.64) as

$$\tilde{\boldsymbol{d}}_k^T \boldsymbol{Q} \tilde{\boldsymbol{d}}_k = \begin{pmatrix} \boldsymbol{d}_k^T & \boldsymbol{0} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \lambda \mathbf{I}_{NM} & \boldsymbol{0} \\ \boldsymbol{0} & \beta \mathbf{I}_{4NM} \end{pmatrix} \begin{pmatrix} \boldsymbol{d}_k \\ \boldsymbol{0} \\ \boldsymbol{0} \end{pmatrix} = \lambda \boldsymbol{d}_k^T \boldsymbol{d}_k. \tag{2.67}$$

Thus, (2.64) becomes

$$\begin{aligned} (\tilde{\boldsymbol{d}}_k - \tilde{\boldsymbol{G}}_{k+1} \boldsymbol{w}_k)^T \boldsymbol{Q} (\tilde{\boldsymbol{d}}_k - \tilde{\boldsymbol{G}}_{k+1} \boldsymbol{w}_k) &= \lambda (\boldsymbol{d}_k^T \boldsymbol{d}_k - \boldsymbol{d}_k^T \boldsymbol{G}_{k+1} \boldsymbol{w}_k - \boldsymbol{w}_k^T \boldsymbol{G}_{k+1}^T \boldsymbol{d}_k + \boldsymbol{w}_k^T \boldsymbol{G}_{k+1}^T \boldsymbol{G}_{k+1} \boldsymbol{w}_k) \\ &\quad + \beta \boldsymbol{w}_k^T \boldsymbol{M}_H \boldsymbol{w}_k \\ &= \lambda (\boldsymbol{d}_k - \boldsymbol{G}_{k+1} \boldsymbol{w}_k)^T (\boldsymbol{d}_k - \boldsymbol{G}_{k+1} \boldsymbol{w}_k) + \beta \boldsymbol{w}_k^T \boldsymbol{M}_H \boldsymbol{w}_k \\ &= \lambda \| \boldsymbol{d}_k - \boldsymbol{G}_{k+1} \boldsymbol{w}_k \|^2 + \beta \boldsymbol{w}_k^T \boldsymbol{M}_H \boldsymbol{w}_k. \end{aligned} \tag{2.68}$$

Defining $\boldsymbol{\epsilon}_k \triangleq (\boldsymbol{\epsilon}_{u,k}^T \ \boldsymbol{\epsilon}_{v,k}^T)^T$, the augmented likelihood function $p(\tilde{\boldsymbol{d}}_k | \boldsymbol{w}_k)$ in (2.61) can finally be reformulated as

$$p(\boldsymbol{d}_k, \boldsymbol{\epsilon}_k = \boldsymbol{0} | \boldsymbol{w}_k) = \Big(\frac{\lambda}{2\pi}\Big)^{NM/2} \Big(\frac{\beta}{2\pi}\Big)^{2NM} \exp\Big(-\frac{\lambda}{2} \| \boldsymbol{d}_k - \boldsymbol{G}_{k+1} \boldsymbol{w}_k \|^2 \Big) \exp\Big(-\frac{\beta}{2} \boldsymbol{w}_k^T \boldsymbol{M}_H \boldsymbol{w}_k \Big). \tag{2.69}$$

Therefore we see that λ^{-1}, the variance of $\boldsymbol{n}_k$, reflects the influence of the data term, i.e., for large values of λ^{-1} that influence increases. Similarly, β^{-1}, the variance of $\boldsymbol{m}_{u,k}$ and $\boldsymbol{m}_{v,k}$, reflects the influence of the spatial regularization term, i.e., for large values of β^{-1} that influence increases.

The ML estimate of $\boldsymbol{w}_k$ is now obtained by maximizing the logarithm of (2.69) with respect to $\boldsymbol{w}_k$ for a fixed (observed) $\boldsymbol{d}_k$ and $\boldsymbol{\epsilon}_k$ set to zero, i.e.,

$$\hat{\boldsymbol{w}}_{k,\mathrm{ML}} = \underset{\boldsymbol{w}_k}{\operatorname{argmax}} \ln p(\boldsymbol{d}_k, \boldsymbol{\epsilon}_k = \boldsymbol{0} | \boldsymbol{w}_k). \tag{2.70}$$

By (2.69), we have

$$\ln p(\boldsymbol{d}_k, \boldsymbol{\epsilon}_k = \boldsymbol{0} | \boldsymbol{w}_k) = -\frac{\lambda}{2} \| \boldsymbol{d}_k - \boldsymbol{G}_{k+1} \boldsymbol{w}_k \|^2 - \frac{\beta}{2} \boldsymbol{w}_k^T \boldsymbol{M}_H \boldsymbol{w}_k + c, \tag{2.71}$$

where c does not depend on $\boldsymbol{w}_k$. By setting $\alpha \triangleq \frac{\beta}{\lambda}$, equation (2.70) can thus be rewritten as

$$\begin{aligned}\hat{\boldsymbol{w}}_{k,\mathrm{ML}} &= \underset{\boldsymbol{w}_k}{\operatorname{argmax}} \{-\|\boldsymbol{d}_k - \boldsymbol{G}_{k+1}\boldsymbol{w}_k\|^2 - \alpha \boldsymbol{w}_k^T \boldsymbol{M}_H \boldsymbol{w}_k\} \\ &= \underset{\boldsymbol{w}_k}{\operatorname{argmin}} \{\|\boldsymbol{G}_{k+1}\boldsymbol{w}_k - \boldsymbol{d}_k\|^2 + \alpha \boldsymbol{w}_k^T \boldsymbol{M}_H \boldsymbol{w}_k\}. \end{aligned} \tag{2.72}$$

Recalling that $\boldsymbol{d}_k = \boldsymbol{G}_{k+1}\boldsymbol{w}_{k,0} - \boldsymbol{f}_{t,k}$, this is seen to be identical to (2.49), and thus the ML estimate $\hat{\boldsymbol{w}}_{k,\mathrm{ML}}$ is recognized to be equal to our least squares estimate $\hat{\boldsymbol{w}}_{k,\text{d-r}}$ in (2.49), (2.50). We have thus reformulated the deterministic OF model and least squares framework for OF estimation of Section 2.1 as a probabilistic OF model and an ML framework for OF estimation.

2.3 Temporal Evolution Model

So far, we considered the estimation of $\boldsymbol{w}_k$ for every frame k separately. In this section, we consider the estimation of the entire OF vector sequence $\boldsymbol{w}_{1:L-1} \triangleq \left(\boldsymbol{w}_1^T, \ldots, \boldsymbol{w}_{L-1}^T\right)^T$. This offers the opportunity to exploit temporal dependencies (temporal coherence) in addition to the already stated spatial dependencies. To model temporal coherence, we assume that $\boldsymbol{w}_{1:L-1}$ is a Markov chain, i.e.,

$$p(\boldsymbol{w}_k|\boldsymbol{w}_{1:k-1}) = p(\boldsymbol{w}_k|\boldsymbol{w}_{k-1}), \quad k = 1, ..., L-1. \tag{2.73}$$

The transition pdf $p(\boldsymbol{w}_k|\boldsymbol{w}_{k-1})$ in (2.73) can be expressed in the form of an *evolution model*. The *Eulerian evolution model* [28] is based on the assumption that the functions $\breve{u}(x, y, t)$ and $\breve{v}(x, y, t)$ defined in (2.4) evolve slowly over time. This can be expressed by setting

$$\frac{\partial \breve{u}(x, y, t)}{\partial t} = \breve{r}^{(u)}(x, y, t) \quad \text{and} \quad \frac{\partial \breve{v}(x, y, t)}{\partial t} = \breve{r}^{(v)}(x, y, t), \tag{2.74}$$

where $\breve{r}^{(u)}(x, y, t)$ and $\breve{r}^{(v)}(x, y, t)$ are mutually independent random fields that have zero mean and a "small" variance for all x, y, t. In the discrete domain, we approximate the temporal derivatives by finite difference quotients, resulting in

$$\frac{1}{\Delta t}(\boldsymbol{u}_k - \boldsymbol{u}_{k-1}) = \boldsymbol{r}_k^{(u)} \quad \text{and} \quad \frac{1}{\Delta t}(\boldsymbol{v}_k - \boldsymbol{v}_{k-1}) = \boldsymbol{r}_k^{(v)}, \quad \text{for } k = 2, ..., L-1. \tag{2.75}$$

Here, $\boldsymbol{r}_k^{(u)}$ and $\boldsymbol{r}_k^{(v)}$ are random vector sequences that are mutually independent, zero-mean Gaussian, independent of $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$, iid across k and across elements, and with a "small" variance γ^{-1}, i.e.,

$$\boldsymbol{r}_k^{(u)} \sim \mathcal{N}(\boldsymbol{0}, \gamma^{-1}\mathbf{I}_{NM}) \quad \text{and} \quad \boldsymbol{r}_k^{(v)} \sim \mathcal{N}(\boldsymbol{0}, \gamma^{-1}\mathbf{I}_{NM}). \tag{2.76}$$

We note that γ, the inverse variance of $\boldsymbol{r}_k^{(u)}$ and $\boldsymbol{r}_k^{(v)}$, quantifies the temporal coherence of the OF sequences $\boldsymbol{u}_k$ and $\boldsymbol{v}_k$, i.e., for large values of γ the OF sequences exhibit a strong temporal coherence.

Hereafter, for simplicity, we set $\Delta t = 1$. Equation (2.75) can then be written in the form of an evolution model as

$$\boldsymbol{u}_k = \boldsymbol{u}_{k-1} + \boldsymbol{r}_k^{(u)} \quad \text{and} \quad \boldsymbol{v}_k = \boldsymbol{v}_{k-1} + \boldsymbol{r}_k^{(v)}, \tag{2.77}$$

or more compactly as

$$\boldsymbol{w}_k = \boldsymbol{w}_{k-1} + \boldsymbol{r}_k, \tag{2.78}$$

where

$$\boldsymbol{r}_k \triangleq \begin{pmatrix} \boldsymbol{r}_k^{(v)} \\ \boldsymbol{r}_k^{(v)} \end{pmatrix}. \tag{2.79}$$

Because of (2.76) and the fact that $\boldsymbol{r}_k^{(u)}$ and $\boldsymbol{r}_k^{(v)}$ are independent, we have

$$\boldsymbol{r}_k \sim \mathcal{N}(\boldsymbol{0}, \gamma^{-1}\mathbf{I}_{2NM}), \tag{2.80}$$

or equivalently

$$p(\boldsymbol{r}_k) = \mathcal{N}(\boldsymbol{r}_k; \boldsymbol{0}, \gamma^{-1}\mathbf{I}_{2NM}) = \Big(\frac{\gamma}{2\pi}\Big)^{NM} \exp\Big(-\frac{\gamma}{2}\|\boldsymbol{w}_k\|^2\Big). \tag{2.81}$$

Finally, by (2.78), the transition pdf involved in (2.73) is obtained as

$$p(\boldsymbol{w}_k|\boldsymbol{w}_{k-1}) = \mathcal{N}(\boldsymbol{w}_k; \boldsymbol{w}_{k-1}, \gamma^{-1}\mathbf{I}_{2MN}) = \Big(\frac{\gamma}{2\pi}\Big)^{NM} \exp\Big(-\frac{\gamma}{2}\|\boldsymbol{w}_k - \boldsymbol{w}_{k-1}\|^2\Big). \tag{2.82}$$

Chapter 3

Bayesian Filtering

In this chapter, we discuss the concept of Bayesian filtering and its application to OF estimation. We start by introducing the general non-linear state-space model in Section 3.1, followed by a presentation of the general Bayesian filter in Section 3.2. In Section 3.3, we consider a linear-Gaussian state-space model, which leads to a tractable form of the general Bayesian filter—the information Kalman filter (Section 3.4). In Section 3.5, we apply the information Kalman filter to the problem of OF estimation.

3.1 State-space Model

Let us consider a non-linear, non-Gaussian, time-varying state-space model [25], where the *state vector* $\boldsymbol{\theta}_k \in \mathbb{R}^n$ evolves over time k according to the *state-transition equation*

$$\boldsymbol{\theta}_k = \boldsymbol{a}_k(\boldsymbol{\theta}_{k-1}, \boldsymbol{\tau}_k), \quad k = 1, 2, \ldots, \tag{3.1}$$

and the observed *data vector* $\boldsymbol{x}_k \in \mathbb{R}^m$ is related to the state $\boldsymbol{\theta}_k$ via the *measurement equation*

$$\boldsymbol{x}_k = \boldsymbol{b}_k(\boldsymbol{\theta}_k, \boldsymbol{\varphi}_k), \quad k = 1, 2, \ldots. \tag{3.2}$$

Here $\boldsymbol{a}_k(\cdot,\cdot)$ and $\boldsymbol{b}_k(\cdot,\cdot)$ are arbitrary non-linear, time-dependent functions, and the *driving noise* $\boldsymbol{\tau}_k$ and the *measurement noise* $\boldsymbol{\varphi}_k$ are temporally and mutually independent random vector processes with known pdfs $p(\boldsymbol{\tau}_k)$ and $p(\boldsymbol{\varphi}_k)$ respectively; they are furthermore independent of the state process $\boldsymbol{\theta}_k$. The initial state $\boldsymbol{\theta}_0$ is a random vector with an arbitrary pdf $p(\boldsymbol{\theta}_0)$. The state-space model formulated by (3.1) and (3.2), together with the statistical assumptions, determines the *transition pdf* $p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})$ and the *likelihood function* $p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)$.

Two conditional statistical independence properties follow from the assumptions above [23]. First, the current state vector $\boldsymbol{\theta}_k$ is conditionally independent of all past data vectors $\boldsymbol{x}_{k-1}$ given

$\boldsymbol{\theta}_{k-1}$, i.e.,

$$p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}, \boldsymbol{x}_{1:k-1}) = p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}), \quad k = 1, 2, \ldots. \tag{3.3}$$

Second, the data vector $\boldsymbol{x}_k$ is conditionally independent of all past data vectors $\boldsymbol{x}_{k-1}$ given $\boldsymbol{\theta}_k$, i.e.,

$$p(\boldsymbol{x}_k|\boldsymbol{\theta}_k, \boldsymbol{x}_{1:k-1}) = p(\boldsymbol{x}_k|\boldsymbol{\theta}_k), \quad k = 1, 2, \ldots. \tag{3.4}$$

3.2 General Bayesian Filter

The task we consider is estimation of the state $\boldsymbol{\theta}_k$ from all data vectors up to time k, $\boldsymbol{x}_{1:k}$, for $k = 1, 2, \ldots$ In the Bayesian framework, this essentially amounts to calculating the *posterior pdf* $p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$. Indeed, using the posterior pdf, many different Bayesian estimators of $\boldsymbol{\theta}_k$ can be constructed. We will consider in particular the minimum mean square error (MMSE) estimator, which is given by

$$\hat{\boldsymbol{\theta}}_{k,\mathrm{MMSE}} \triangleq \mathrm{E}\{\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}\} = \int_{\mathbb{R}^n} \boldsymbol{\theta}_k p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}) \, \mathrm{d}\boldsymbol{\theta}_k, \quad k = 1, 2, \ldots. \tag{3.5}$$

Based on the conditional independence properties in (3.3) and (3.4), the posterior pdf can be calculated recursively by performing the following two steps at each k [23, Section 2.3]:

1. In the *prediction step*, the *predicted posterior pdf* $p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k-1})$ is calculated from the previous posterior pdf $p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})$ and the transition pdf $p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})$ by

$$p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k-1}) = \int_{\mathbb{R}^n} p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}) p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1}) \, \mathrm{d}\boldsymbol{\theta}_{k-1}, \quad k = 1, 2, \ldots. \tag{3.6}$$

2. In the *update step*, the posterior pdf $p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$ is calculated from the predicted posterior pdf $p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k-1})$ and the likelihood function $p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)$ by

$$p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}) = \frac{p(\boldsymbol{x}_k|\boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k-1})}{p(\boldsymbol{x}_k|\boldsymbol{x}_{1:k-1})}, \tag{3.7}$$

where

$$p(\boldsymbol{x}_k|\boldsymbol{x}_{1:k-1}) = \int_{\mathbb{R}^n} p(\boldsymbol{x}_k|\boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k-1}) \, \mathrm{d}\boldsymbol{\theta}_k, \quad k = 1, 2, \ldots. \tag{3.8}$$

This recursion is initialized with $p(\boldsymbol{\theta}_0|\boldsymbol{x}_{1:0}) = p(\boldsymbol{\theta}_0)$. The overall recursive algorithm is sometimes called the *general Bayesian filter*, since it applies to the general (i.e., possibly non-linear and non-Gaussian) state-space model.

3.3 Linear Gaussian State-space Model

Assuming that the functions $\boldsymbol{a}_k(\cdot,\cdot)$ and $\boldsymbol{b}_k(\cdot,\cdot)$ in (3.1) and (3.2) are linear and $\boldsymbol{\tau}_k$ and $\boldsymbol{\varphi}_k$ are Gaussian distributed, we obtain the linear-Gaussian state-space model [25] with state-transition equation

$$\boldsymbol{\theta}_k = \boldsymbol{A}_k\boldsymbol{\theta}_{k-1} + \boldsymbol{\tau}_k, \quad k = 1, 2, \ldots, \tag{3.9}$$

and measurement equation

$$\boldsymbol{x}_k = \boldsymbol{B}_k\boldsymbol{\theta}_k + \boldsymbol{\varphi}_k, \quad k = 1, 2, \ldots. \tag{3.10}$$

Here $\boldsymbol{A}_k$ and $\boldsymbol{B}_k$ are matrices of size $n \times n$ and $m \times n$, respectively. Furthermore, the driving noise $\boldsymbol{\tau}_k$ and the measurement noise $\boldsymbol{\varphi}_k$ are temporally and mutually independent, Gaussian, zero-mean vector processes that are also independent of the state process $\boldsymbol{\theta}_k$, with covariance matrices $\boldsymbol{\Sigma}_{\tau,k}$ and $\boldsymbol{\Sigma}_{\varphi,k}$ respectively, i.e.,

$$p(\boldsymbol{\tau}_k) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\tau,k}), \quad p(\boldsymbol{\varphi}_k) = \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_{\varphi,k}). \tag{3.11}$$

We also assume that the initial state vector $\boldsymbol{\theta}_0$ is Gaussian distributed with mean $\boldsymbol{\mu}_0$ and covariance matrix $\boldsymbol{\Sigma}_0$, i.e.,

$$p(\boldsymbol{\theta}_0) = \mathcal{N}(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0). \tag{3.12}$$

The transition pdf is therefore Gaussian with mean $\boldsymbol{A}_k\boldsymbol{\theta}_{k-1}$ and covariance $\boldsymbol{\Sigma}_{\tau,k}$, i.e.,

$$p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}) = (2\pi)^{-n/2}\det(\boldsymbol{\Sigma}_{\tau,k})^{-1/2}\exp\Big(-\frac{1}{2}(\boldsymbol{\theta}_k - \boldsymbol{A}_k\boldsymbol{\theta}_{k-1})^T\boldsymbol{\Sigma}_{\tau,k}^{-1}(\boldsymbol{\theta}_k - \boldsymbol{A}_k\boldsymbol{\theta}_{k-1})\Big), \tag{3.13}$$

and the likelihood function (viewed as the conditional pdf of $\boldsymbol{x}_k$) is Gaussian with mean $\boldsymbol{B}_k\boldsymbol{\theta}_k$ and covariance $\boldsymbol{\Sigma}_{\varphi,k}$, i.e.,

$$p(\boldsymbol{x}_k|\boldsymbol{\theta}_k) = (2\pi)^{-m/2}\det(\boldsymbol{\Sigma}_{\varphi,k})^{-1/2}\exp\Big(-\frac{1}{2}(\boldsymbol{x}_k - \boldsymbol{B}_k\boldsymbol{\theta}_k)^T\boldsymbol{\Sigma}_{\varphi,k}^{-1}(\boldsymbol{x}_k - \boldsymbol{B}_k\boldsymbol{\theta}_k)\Big). \tag{3.14}$$

3.4 Information Kalman Filter

For the linear-Gaussian state-space model in (3.9) and (3.10), the prediction and update steps of the general Bayesian filter, given in (3.6) and (3.7), respectively, reduce to the Kalman filter (KF) equations [29]. For practical reasons, which will be explained in Section 3.5, we consider the information form of the Kalman filter (IKF) [15, 29]. In the IKF, the posterior pdf is parametrized by the precision matrix rather than the covariance matrix.

By choosing the prior pdf $p(\boldsymbol{\theta}_0)$ to be Gaussian, it follows from (3.9) and (3.10) that the posterior pdf $p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$ will be Gaussian for arbitrary k. Denoting the mean and precision matrix of the posterior pdf as $\boldsymbol{\mu}_{k|k}$ and $\boldsymbol{J}_{k|k}$, respectively, we have

$$\begin{aligned} p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}) &= \mathcal{N}(\boldsymbol{\mu}_{k|k}, \boldsymbol{J}_{k|k}^{-1}) \\ &= (2\pi)^{-n/2} \det(\boldsymbol{J}_{k|k})^{1/2} \exp\Big(-\frac{1}{2}(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k})^T \boldsymbol{J}_{k|k}(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k})\Big). \end{aligned} \tag{3.15}$$

Note that $\boldsymbol{\mu}_{k|k}$ and $\boldsymbol{J}_{k|k}$ depend on $\boldsymbol{x}_{1:k}$ In the following, we will derive the IKF equations, which provide a recursive algorithm for calculating $\boldsymbol{\mu}_{k|k}$ and $\boldsymbol{J}_{k|k}$ for $k = 1, 2, \ldots$.

3.4.1 Prediction Step

Inserting (3.13) and (3.15) with k replaced by $k-1$, we recognize that (3.6) can be written (up to a constant factor that does not depend on $\boldsymbol{\theta}_k$) as

$$\begin{aligned} p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k-1}) &\propto \int_{\mathbb{R}^n} \exp\Big(-\frac{1}{2}(\boldsymbol{\theta}_k - \boldsymbol{A}_k\boldsymbol{\theta}_{k-1})^T \boldsymbol{J}_{\tau,k}(\boldsymbol{\theta}_k - \boldsymbol{A}_k\boldsymbol{\theta}_{k-1})\Big) \\ &\quad \times \exp\Big(-\frac{1}{2}(\boldsymbol{\theta}_{k-1} - \boldsymbol{\mu}_{k-1|k-1})^T \boldsymbol{J}_{k-1|k-1}(\boldsymbol{\theta}_{k-1} - \boldsymbol{\mu}_{k-1|k-1})\Big)\, \mathrm{d}\boldsymbol{\theta}_{k-1} \\ &\propto \int_{\mathbb{R}^n} \exp\Big(-\frac{1}{2}\Phi_k(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1})\Big)\, \mathrm{d}\boldsymbol{\theta}_{k-1}, \end{aligned} \tag{3.16}$$

with

$$\begin{aligned} \Phi_k(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) &\triangleq (\boldsymbol{\theta}_k - \boldsymbol{A}_k\boldsymbol{\theta}_{k-1})^T \boldsymbol{J}_{\tau,k}(\boldsymbol{\theta}_k - \boldsymbol{A}_k\boldsymbol{\theta}_{k-1}) \\ &\quad + (\boldsymbol{\theta}_{k-1} - \boldsymbol{\mu}_{k-1|k-1})^T \boldsymbol{J}_{k-1|k-1}(\boldsymbol{\theta}_{k-1} - \boldsymbol{\mu}_{k-1|k-1}), \end{aligned} \tag{3.17}$$

where $\boldsymbol{J}_{\tau,k} \triangleq \boldsymbol{\Sigma}_{\tau,k}^{-1}$. Expanding (3.17) and gathering all terms in $\boldsymbol{\theta}_{k-1}$ yields

$$\begin{aligned} \Phi_k(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) &\overset{\mathrm{c}}{=} \boldsymbol{\theta}_k^T \boldsymbol{J}_{\tau,k}\boldsymbol{\theta}_k + \boldsymbol{\theta}_{k-1}^T(\boldsymbol{J}_{k-1|k-1} + \boldsymbol{A}_k^T \boldsymbol{J}_{\tau,k}\boldsymbol{A}_k)\boldsymbol{\theta}_{k-1} \\ &\quad -2\boldsymbol{\theta}_{k-1}^T(\boldsymbol{A}_k^T \boldsymbol{J}_{\tau,k}\boldsymbol{\theta}_k + \boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}), \end{aligned} \tag{3.18}$$

where "$\overset{\mathrm{c}}{=}$" denotes equality up to an additive constant (which does not depend on $\boldsymbol{\theta}_k$ or $\boldsymbol{\theta}_{k-1}$). Note that $\boldsymbol{\mu}_{k-1|k-1}^T \boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}$ was absorbed into the additive constant.

We will now show that $\Phi_k(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1})$ can be written (up to an additive constant) as the sum of a quadratic form in $\boldsymbol{\theta}_{k-1}$ and a quadratic form in $\boldsymbol{\theta}_k$. To this end, we first complete the square with respect to $\boldsymbol{\theta}_{k-1}$, which introduces new terms in $\boldsymbol{\theta}_k$. Defining

$$\boldsymbol{P}_k \triangleq \boldsymbol{J}_{k-1|k-1} + \boldsymbol{A}_k^T \boldsymbol{J}_{\tau,k}\boldsymbol{A}_k, \tag{3.19}$$

we can rewrite (3.18) as

$$\begin{aligned}\Phi_k(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) &\overset{\mathrm{c}}{=} \boldsymbol{\theta}_{k-1}^T \boldsymbol{P}_k \boldsymbol{\theta}_{k-1} - 2\boldsymbol{\theta}_{k-1}^T (\boldsymbol{A}_k^T \boldsymbol{J}_{\tau,k} \boldsymbol{\theta}_k + \boldsymbol{J}_{k-1|k-1} \boldsymbol{\mu}_{k-1|k-1}) + \boldsymbol{\theta}_k^T \boldsymbol{J}_{\tau,k} \boldsymbol{\theta}_k \\ &\overset{\mathrm{c}}{=} (\boldsymbol{\theta}_{k-1} - \boldsymbol{m}_k)^T \boldsymbol{P}_k (\boldsymbol{\theta}_{k-1} - \boldsymbol{m}_k) + \boldsymbol{\theta}_k^T \boldsymbol{J}_{\tau,k} \boldsymbol{\theta}_k \\ &\quad - (\boldsymbol{A}_k^T \boldsymbol{J}_{\tau,k} \boldsymbol{\theta}_k + \boldsymbol{J}_{k-1|k-1} \boldsymbol{\mu}_{k-1|k-1})^T \boldsymbol{P}_k^{-1} (\boldsymbol{A}_k^T \boldsymbol{J}_{\tau,k} \boldsymbol{\theta}_k + \boldsymbol{J}_{k-1|k-1} \boldsymbol{\mu}_{k-1|k-1}),\end{aligned} \tag{3.20}$$

with

$$\boldsymbol{m}_k \triangleq \boldsymbol{P}_k^{-1} (\boldsymbol{A}_k^T \boldsymbol{J}_{\tau,k} \boldsymbol{\theta}_k + \boldsymbol{J}_{k-1|k-1} \boldsymbol{\mu}_{k-1|k-1}). \tag{3.21}$$

Next, we expand (3.20) and gather all terms in $\boldsymbol{\theta}_k$. We obtain

$$\begin{aligned}\Phi_k(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) &\overset{\mathrm{c}}{=} (\boldsymbol{\theta}_{k-1} - \boldsymbol{m}_k)^T \boldsymbol{P}_k (\boldsymbol{\theta}_{k-1} - \boldsymbol{m}_k) + \boldsymbol{\theta}_k^T (\boldsymbol{J}_{\tau,k} - \boldsymbol{J}_{\tau,k} \boldsymbol{A}_k \boldsymbol{P}_k^{-1} \boldsymbol{A}_k^T \boldsymbol{J}_{\tau,k}) \boldsymbol{\theta}_k \\ &\quad - 2\boldsymbol{\theta}_k^T \boldsymbol{J}_{\tau,k} \boldsymbol{A}_k \boldsymbol{P}_k^{-1} \boldsymbol{J}_{k-1|k-1} \boldsymbol{\mu}_{k-1|k-1},\end{aligned} \tag{3.22}$$

where $\boldsymbol{\mu}_{k-1|k-1}^T \boldsymbol{J}_{k-1|k-1} \boldsymbol{P}_k^{-1} \boldsymbol{J}_{k-1|k-1} \boldsymbol{\mu}_{k-1|k-1}$ was absorbed into the additive constant. Defining

$$\boldsymbol{J}_{k|k-1} \triangleq \boldsymbol{J}_{\tau,k} - \boldsymbol{J}_{\tau,k} \boldsymbol{A}_k \boldsymbol{P}_k^{-1} \boldsymbol{A}_k^T \boldsymbol{J}_{\tau,k}, \tag{3.23}$$

we can rewrite (3.22) as

$$\Phi_k(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) \overset{\mathrm{c}}{=} \boldsymbol{\theta}_k^T \boldsymbol{J}_{k|k-1} \boldsymbol{\theta}_k - 2\boldsymbol{\theta}_k^T \boldsymbol{J}_{\tau,k} \boldsymbol{A}_k \boldsymbol{P}_k^{-1} \boldsymbol{J}_{k-1|k-1} \boldsymbol{\mu}_{k-1|k-1} + (\boldsymbol{\theta}_{k-1} - \boldsymbol{m}_k)^T \boldsymbol{P}_k (\boldsymbol{\theta}_{k-1} - \boldsymbol{m}_k). \tag{3.24}$$

Completing the square with respect to $\boldsymbol{\theta}_k$ then yields

$$\Phi_k(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}) \overset{\mathrm{c}}{=} (\boldsymbol{\theta}_{k-1} - \boldsymbol{m}_k)^T \boldsymbol{P}_k (\boldsymbol{\theta}_{k-1} - \boldsymbol{m}_k) + (\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k-1})^T \boldsymbol{J}_{k|k-1} (\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k-1}), \tag{3.25}$$

with

$$\boldsymbol{\mu}_{k|k-1} \triangleq \boldsymbol{J}_{k|k-1}^{-1} \boldsymbol{J}_{\tau,k} \boldsymbol{A}_k \boldsymbol{P}_k^{-1} \boldsymbol{J}_{k-1|k-1} \boldsymbol{\mu}_{k-1|k-1}. \tag{3.26}$$

Thus, we showed that $\Phi_k(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1})$ is (up to an additive constant) the sum of a quadratic form in $\boldsymbol{\theta}_{k-1}$ and one in $\boldsymbol{\theta}_k$. Inserting (3.25) into (3.16) then yields

$$\begin{aligned}p(\boldsymbol{\theta}_k | \boldsymbol{x}_{1:k-1}) \propto \exp\Big(&- \frac{1}{2} (\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k-1})^T \boldsymbol{J}_{k|k-1} (\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k-1}) \Big) \\ &\times \int_{\mathbb{R}^n} \exp\Big(- \frac{1}{2} (\boldsymbol{\theta}_{k-1} - \boldsymbol{m}_k)^T \boldsymbol{P}_k (\boldsymbol{\theta}_{k-1} - \boldsymbol{m}_k) \Big) \mathrm{d}\boldsymbol{\theta}_{k-1}.\end{aligned} \tag{3.27}$$

Here the integral is constant w.r.t. $\boldsymbol{\theta}_k$ and can thus be absorbed into the multiplicative constant. Hence, the predicted posterior pdf is recognized to be Gaussian with mean $\boldsymbol{\mu}_{k|k-1}$ and precision

matrix $\boldsymbol{J}_{k|k-1}$, i.e.,

$$p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k-1}) = \mathcal{N}(\boldsymbol{\mu}_{k|k-1}, \boldsymbol{J}_{k|k-1}^{-1}) \\ = (2\pi)^{-n/2} \det(\boldsymbol{J}_{k|k-1})^{1/2} \exp\Big(-\frac{1}{2}(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k-1})^T \boldsymbol{J}_{k|k-1}(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k-1})\Big). \quad (3.28)$$

The expression for the predicted mean in (3.26) can be simplified as follows. By inserting (3.23) into (3.26), we obtain

$$\boldsymbol{\mu}_{k|k-1} = (\boldsymbol{J}_{\tau,k} - \boldsymbol{J}_{\tau,k}\boldsymbol{A}_k\boldsymbol{P}_k^{-1}\boldsymbol{A}_k^T\boldsymbol{J}_{\tau,k})^{-1}\boldsymbol{J}_{\tau,k}\boldsymbol{A}_k\boldsymbol{P}_k^{-1}\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}. \quad (3.29)$$

Using the matrix inversion lemma [29, Section 6.3] $(\boldsymbol{Z}^{-1} - \boldsymbol{H}\boldsymbol{R}^{-1}\boldsymbol{H}^T)^{-1}\boldsymbol{H}\boldsymbol{R}^{-1} = \boldsymbol{Z}\boldsymbol{H}(\boldsymbol{H}^T\boldsymbol{Z}\boldsymbol{H} - \boldsymbol{R})^{-1}$ with $\boldsymbol{Z}^{-1} = \boldsymbol{J}_{\tau,k}$, $\boldsymbol{H} = \boldsymbol{J}_{\tau,k}\boldsymbol{A}_k$, and $\boldsymbol{R} = \boldsymbol{P}_k$ in (3.29), we obtain

$$\boldsymbol{\mu}_{k|k-1} = \boldsymbol{J}_{\tau,k}^{-1}\boldsymbol{J}_{\tau,k}\boldsymbol{A}_k(\boldsymbol{A}_k^T\boldsymbol{J}_{\tau,k}\boldsymbol{J}_{\tau,k}^{-1}\boldsymbol{J}_{\tau,k}\boldsymbol{A}_k + \boldsymbol{P}_k)^{-1}\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}. \quad (3.30)$$

Noting that $\boldsymbol{J}_{\tau,k}^{-1}\boldsymbol{J}_{\tau,k} = \mathbf{I}$ and inserting (3.19), we obtain further

$$\boldsymbol{\mu}_{k|k-1} = \boldsymbol{A}_k(\boldsymbol{A}_k^T\boldsymbol{J}_{\tau,k}\boldsymbol{A}_k + \boldsymbol{J}_{k-1|k-1} - \boldsymbol{A}_k^T\boldsymbol{J}_{\tau,k}\boldsymbol{A}_k)^{-1}\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1} \\ = \boldsymbol{A}_k\boldsymbol{J}_{k-1|k-1}^{-1}\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1} \\ = \boldsymbol{A}_k\boldsymbol{\mu}_{k-1|k-1}. \quad (3.31)$$

Furthermore, inserting (3.19) into (3.23) yields

$$\boldsymbol{J}_{k|k-1} = \boldsymbol{J}_{\tau,k} - \boldsymbol{J}_{\tau,k}\boldsymbol{A}_k(\boldsymbol{J}_{k-1|k-1} + \boldsymbol{A}_k^T\boldsymbol{J}_{\tau,k}\boldsymbol{A}_k)^{-1}\boldsymbol{A}_k^T\boldsymbol{J}_{\tau,k}. \quad (3.32)$$

To summarize, the prediction step for the linear-Gaussian model reduces to the mapping $(\boldsymbol{\mu}_{k-1|k-1}, \boldsymbol{J}_{k-1|k-1}) \to (\boldsymbol{\mu}_{k|k-1}, \boldsymbol{J}_{k|k-1})$ as described by (3.31) and (3.32).

3.4.2 Update Step

Next we calculate the posterior pdf $p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$ in (3.7). Noticing that $p(\boldsymbol{x}_k|\boldsymbol{x}_{1:k-1})$ in (3.7) is not a function of $\boldsymbol{\theta}_k$, we can write

$$p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}) \propto p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k-1}). \quad (3.33)$$

Inserting (3.14) and (3.28) yields

$$p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}) \propto \exp\Big(-\frac{1}{2}\Phi_k(\boldsymbol{\theta}_k)\Big), \quad (3.34)$$

where

$$\Phi_k(\boldsymbol{\theta}_k) \triangleq (\boldsymbol{x}_k - \boldsymbol{B}_k\boldsymbol{\theta}_k)^T \boldsymbol{J}_{\varphi,k}(\boldsymbol{x}_k - \boldsymbol{B}_k\boldsymbol{\theta}_k) + (\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k-1})^T \boldsymbol{J}_{k|k-1}(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k-1}), \tag{3.35}$$

with $\boldsymbol{J}_{\varphi,k} \triangleq \boldsymbol{\Sigma}_{\varphi,k}^{-1}$. By expanding the right-hand side and gathering all terms in $\boldsymbol{\theta}_k$, we obtain

$$\Phi_k(\boldsymbol{\theta}_k) \overset{c}{=} \boldsymbol{\theta}_k^T(\boldsymbol{J}_{k|k-1} + \boldsymbol{B}_k^T\boldsymbol{J}_{\varphi,k}\boldsymbol{B}_k)\boldsymbol{\theta}_k - 2\boldsymbol{\theta}_k^T(\boldsymbol{B}_k^T\boldsymbol{J}_{\varphi,k}\boldsymbol{x}_k + \boldsymbol{J}_{k|k-1}\boldsymbol{\mu}_{k|k-1}), \tag{3.36}$$

where $\boldsymbol{x}_k^T\boldsymbol{J}_{\varphi,k}\boldsymbol{x}_k$ and $\boldsymbol{\mu}_{k|k-1}^T\boldsymbol{J}_{k|k-1}\boldsymbol{\mu}_{k|k-1}$ were absorbed into the additive constant. Defining

$$\boldsymbol{J}_{k|k} \triangleq \boldsymbol{J}_{k|k-1} + \boldsymbol{B}_k^T\boldsymbol{J}_{\varphi,k}\boldsymbol{B}_k, \tag{3.37}$$

we can rewrite (3.36) as

$$\Phi_k(\boldsymbol{\theta}_k) \overset{c}{=} \boldsymbol{\theta}_k^T\boldsymbol{J}_{k|k}\boldsymbol{\theta}_k - 2\boldsymbol{\theta}_k^T(\boldsymbol{B}_k^T\boldsymbol{J}_{\varphi,k}\boldsymbol{x}_k + \boldsymbol{J}_{k|k-1}\boldsymbol{\mu}_{k|k-1}). \tag{3.38}$$

Completing the square with respect to $\boldsymbol{\theta}_k$ and dropping some terms that do not depend on $\boldsymbol{\theta}_k$, we further obtain

$$\Phi_k(\boldsymbol{\theta}_k) \overset{c}{=} (\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k})^T\boldsymbol{J}_{k|k}(\boldsymbol{\theta}_k - \boldsymbol{\mu}_{k|k}), \tag{3.39}$$

with

$$\boldsymbol{\mu}_{k|k} \triangleq \boldsymbol{J}_{k|k}^{-1}(\boldsymbol{B}_k^T\boldsymbol{J}_{\varphi,k}\boldsymbol{x}_k + \boldsymbol{J}_{k|k-1}\boldsymbol{\mu}_{k|k-1}). \tag{3.40}$$

Inserting (3.39) into (3.34), we recognize that the posterior pdf is Gaussian with mean $\boldsymbol{\mu}_{k|k}$ and precision matrix $\boldsymbol{J}_{k|k}$, i.e.,

$$p(\boldsymbol{\theta}_k|x_{1:k}) = \mathcal{N}(\boldsymbol{\mu}_{k|k}, \boldsymbol{J}_{k|k}^{-1}). \tag{3.41}$$

We thus see that the update step for the linear-Gaussian model reduces to the mapping $(\boldsymbol{\mu}_{k|k-1}, \boldsymbol{J}_{k|k-1}) \to (\boldsymbol{\mu}_{k|k}, \boldsymbol{J}_{k|k})$ as described by (3.37) and (3.40).

3.4.3 Summary of the Information Kalman Filter Algorithm

We can now summarize the IKF algorithm. For $k = 1, 2, \ldots$, convert $(\boldsymbol{\mu}_{k-1|k-1}, \boldsymbol{J}_{k-1|k-1})$ into $(\boldsymbol{\mu}_{k|k}, \boldsymbol{J}_{k|k})$ via the following two steps:

1. Prediction step (see (3.32) and (3.31)):

$$\boldsymbol{J}_{k|k-1} = \boldsymbol{J}_{\tau,k} - \boldsymbol{J}_{\tau,k}\boldsymbol{A}_k(\boldsymbol{J}_{k-1|k-1} + \boldsymbol{A}_k^T\boldsymbol{J}_{\tau,k}\boldsymbol{A}_k)^{-1}\boldsymbol{A}_k^T\boldsymbol{J}_{\tau,k}, \tag{3.42}$$

$$\boldsymbol{\mu}_{k|k-1} = \boldsymbol{A}_k\boldsymbol{\mu}_{k-1|k-1}. \tag{3.43}$$

2. Update step (see (3.37) and (3.40)):

$$\boldsymbol{J}_{k|k} = \boldsymbol{J}_{k|k-1} + \boldsymbol{B}_k^T \boldsymbol{J}_{\varphi,k} \boldsymbol{B}_k, \tag{3.44}$$

$$\boldsymbol{\mu}_{k|k} = \boldsymbol{J}_{k|k}^{-1} (\boldsymbol{B}_k^T \boldsymbol{J}_{\varphi,k} \boldsymbol{x}_k + \boldsymbol{J}_{k|k-1} \boldsymbol{\mu}_{k|k-1}). \tag{3.45}$$

This recursion is initialized with $\boldsymbol{J}_{0|0} = \boldsymbol{J}_0 = \boldsymbol{\Sigma}_0^{-1}$ and $\boldsymbol{\mu}_{0|0} = \boldsymbol{\mu}_0$ (see (3.12)).

3.5 Information Kalman Filter for Optical Flow Estimation

Looking at the model for OF estimation in (2.78) and (2.57), we recognize that these equations constitute a special case of the linear-Gaussian state-space model presented in Section 3.3 with $\boldsymbol{\theta}_k = \boldsymbol{w}_k$, $\boldsymbol{\tau}_k = \boldsymbol{r}_{k-1}$, $\boldsymbol{x}_k = \tilde{\boldsymbol{d}}_k$, $\boldsymbol{\varphi}_k = \tilde{\boldsymbol{n}}_k$, $\boldsymbol{A}_k = \mathbf{I}_{2MN}$, $\boldsymbol{B}_k = \tilde{\boldsymbol{G}}_{k+1}$, $\boldsymbol{\Sigma}_{\tau,k} = \gamma^{-1}\mathbf{I}_{2MN}$ (thus, $\boldsymbol{J}_{\tau,k} = \gamma \mathbf{I}_{2MN}$), and $\boldsymbol{\Sigma}_{\varphi,k} = \boldsymbol{Q}^{-1}$ (thus, $\boldsymbol{J}_{\varphi,k} = \boldsymbol{Q}$). Hence, the IKF of Section 3.4 can be directly applied.

3.5.1 Prediction and Update Steps

The k-th recursion for $k = 1, \ldots, L-1$ is summarized in the following two steps:

1. Prediction step (see (3.42), (3.43)):

$$\boldsymbol{J}_{k|k-1} = \gamma \mathbf{I}_{2MN} - \gamma^2 (\boldsymbol{J}_{k-1|k-1} + \gamma \mathbf{I}_{2MN})^{-1}, \tag{3.46}$$

$$\boldsymbol{\mu}_{k|k-1} = \boldsymbol{\mu}_{k-1|k-1}. \tag{3.47}$$

2. Update step (see (3.44), (3.45)):

$$\boldsymbol{J}_{k|k} = \boldsymbol{J}_{k|k-1} + \tilde{\boldsymbol{G}}_{k+1}^T \boldsymbol{Q} \tilde{\boldsymbol{G}}_{k+1}, \tag{3.48}$$

$$\boldsymbol{\mu}_{k|k} = \boldsymbol{J}_{k|k}^{-1} (\boldsymbol{J}_{k|k-1} \boldsymbol{\mu}_{k|k-1} + \tilde{\boldsymbol{G}}_{k+1}^T \boldsymbol{Q} \tilde{\boldsymbol{d}}_k). \tag{3.49}$$

Recalling from (2.65) and (2.66) that

$$\tilde{\boldsymbol{G}}_{k+1}^T \boldsymbol{Q} \tilde{\boldsymbol{G}}_{k+1} = \beta \boldsymbol{M}_H + \lambda \boldsymbol{G}_{k+1}^T \boldsymbol{G}_{k+1} \tag{3.50}$$

and

$$\tilde{\boldsymbol{G}}_{k+1}^T \boldsymbol{Q} \tilde{\boldsymbol{d}}_k = \lambda \boldsymbol{G}_{k+1}^T \boldsymbol{d}_k, \tag{3.51}$$

the update step (3.48), (3.49) becomes

$$\boldsymbol{J}_{k|k} = \boldsymbol{J}_{k|k-1} + \beta \boldsymbol{M}_H + \lambda \boldsymbol{G}_{k+1}^T \boldsymbol{G}_{k+1}, \tag{3.52}$$

$$\boldsymbol{\mu}_{k|k} = \boldsymbol{J}_{k|k}^{-1}(\boldsymbol{J}_{k|k-1}\boldsymbol{\mu}_{k|k-1} + \lambda \boldsymbol{G}_{k+1}^T \boldsymbol{d}_k). \tag{3.53}$$

This recursion is initialized with arbitrary $\boldsymbol{J}_{0|0} = \boldsymbol{J}_0$ and $\boldsymbol{\mu}_{0|0} = \boldsymbol{\mu}_0$. The MMSE estimate of the OF vector $\boldsymbol{w}$ is obtained as the mean of the posterior pdf (cf. (3.5)), i.e.,

$$\hat{\boldsymbol{w}}_{k,\mathrm{MMSE}} = \boldsymbol{\mu}_{k|k} = \boldsymbol{J}_{k|k}^{-1}(\boldsymbol{J}_{k|k-1}\boldsymbol{\mu}_{k|k-1} + \lambda \boldsymbol{G}_{k+1}^T \boldsymbol{d}_k). \tag{3.54}$$

3.5.2 Reduced-Parameter Formulation

In the following, we will show that the IKF given by (3.46), (3.47), (3.52), and (3.53) can be reformulated in terms of only the two parameters

$$\eta \triangleq \frac{\gamma}{\lambda} \quad \text{and} \quad \psi \triangleq \frac{\beta}{\lambda}, \tag{3.55}$$

rather than the three parameters β, γ, and λ. Let us we rewrite $\boldsymbol{J}_{k|k-1}$ in the prediction step (cf. (3.46)) as

$$\boldsymbol{J}_{k|k-1} = \lambda \bar{\boldsymbol{J}}_{k|k-1} \tag{3.56}$$

with

$$\bar{\boldsymbol{J}}_{k|k-1} \triangleq \frac{1}{\lambda}\boldsymbol{J}_{k|k-1} = \frac{\gamma}{\lambda}\mathbf{I}_{2MN} - \frac{\gamma^2}{\lambda}(\boldsymbol{J}_{k-1|k-1} + \gamma \mathbf{I}_{2MN})^{-1}. \tag{3.57}$$

Similarly, we rewrite $\boldsymbol{J}_{k|k}$ in the update step (cf. (3.52)) as

$$\boldsymbol{J}_{k|k} = \lambda \bar{\boldsymbol{J}}_{k|k}, \tag{3.58}$$

with

$$\begin{aligned}
\bar{\boldsymbol{J}}_{k|k} &\triangleq \frac{1}{\lambda}\boldsymbol{J}_{k|k} \\
&= \frac{1}{\lambda}\boldsymbol{J}_{k|k-1} + \frac{\beta}{\lambda}\boldsymbol{M}_H + \boldsymbol{G}_{k+1}^T \boldsymbol{G}_{k+1} \\
&= \frac{1}{\lambda}\boldsymbol{J}_{k|k-1} + \psi \boldsymbol{M}_H + \boldsymbol{G}_{k+1}^T \boldsymbol{G}_{k+1} \\
&= \bar{\boldsymbol{J}}_{k|k-1} + \psi \boldsymbol{M}_H + \boldsymbol{G}_{k+1}^T \boldsymbol{G}_{k+1},
\end{aligned} \tag{3.59}$$

where we used (3.56) in the last line. Inserting (3.58), with k replaced by $k-1$ into (3.57), we then obtain

$$\begin{aligned}\bar{\boldsymbol{J}}_{k|k-1} &= \frac{\gamma}{\lambda}\mathbf{I}_{2MN} - \frac{\gamma^2}{\lambda}(\lambda\bar{\boldsymbol{J}}_{k-1|k-1} + \gamma\mathbf{I}_{2MN})^{-1} \\ &= \frac{\gamma}{\lambda}\mathbf{I}_{2MN} - \Big(\frac{\gamma}{\lambda}\Big)^2(\bar{\boldsymbol{J}}_{k-1|k-1} + \frac{\gamma}{\lambda}\mathbf{I}_{2MN})^{-1} \\ &= \eta\mathbf{I}_{2MN} - \eta^2(\bar{\boldsymbol{J}}_{k-1|k-1} + \eta\mathbf{I}_{2MN})^{-1}. \end{aligned} \tag{3.60}$$

Next, we consider $\boldsymbol{\mu}_{k|k}$ in the update step (cf. (3.53)), which is given as

$$\begin{aligned}\boldsymbol{\mu}_{k|k} &= \lambda\boldsymbol{J}_{k|k}^{-1}\Big(\frac{1}{\lambda}\boldsymbol{J}_{k|k-1}\boldsymbol{\mu}_{k|k-1} + \boldsymbol{G}_{k+1}^T\boldsymbol{d}_k\Big) \\ &= \Big(\frac{1}{\lambda}\boldsymbol{J}_{k|k}\Big)^{-1}\Big(\frac{1}{\lambda}\boldsymbol{J}_{k|k-1}\boldsymbol{\mu}_{k|k-1} + \boldsymbol{G}_{k+1}^T\boldsymbol{d}_k\Big) \\ &= \bar{\boldsymbol{J}}_{k|k}^{-1}(\bar{\boldsymbol{J}}_{k|k-1}\boldsymbol{\mu}_{k|k-1} + \boldsymbol{G}_{k+1}^T\boldsymbol{d}_k). \end{aligned} \tag{3.61}$$

Thus, we can finally summarize reduced-parameter our reformulation of the IKF:

1. Prediction step (see (3.60), (3.47)):

$$\bar{\boldsymbol{J}}_{k|k-1} = \eta\mathbf{I}_{2MN} - \eta^2(\bar{\boldsymbol{J}}_{k-1|k-1} + \eta\mathbf{I}_{2MN})^{-1}, \tag{3.62}$$

$$\boldsymbol{\mu}_{k|k-1} = \boldsymbol{\mu}_{k-1|k-1}. \tag{3.63}$$

2. Update step (see (3.59), (3.61)):

$$\bar{\boldsymbol{J}}_{k|k} = \bar{\boldsymbol{J}}_{k|k-1} + \psi\boldsymbol{M}_H + \boldsymbol{G}_{k+1}^T\boldsymbol{G}_{k+1}, \tag{3.64}$$

$$\boldsymbol{\mu}_{k|k} = \bar{\boldsymbol{J}}_{k|k}^{-1}(\bar{\boldsymbol{J}}_{k|k-1}\boldsymbol{\mu}_{k|k-1} + \boldsymbol{G}_{k+1}^T\boldsymbol{d}_k). \tag{3.65}$$

Assuming that the initialization of the IKF recursion does not involve λ, i.e., $\boldsymbol{\mu}_{0|0}$ and $\boldsymbol{J}_{0|0}$ are chosen such that they do not depend on λ (e.g., $\boldsymbol{\mu}_{0|0} = \mathbf{0}_{2MN}$ and $\boldsymbol{J}_{0|0} = \mathbf{0}_{2MN\times 2MN}$), we see that the IKF is formulated in terms of η and ψ rather than β, γ, and λ. This formulation of the IKF is finally recognized to be identical to the IKF given in (3.46), (3.47), (3.52), and (3.53) for $\lambda = 1$, which will therefore be used in the following.

3.5.3 Sparsity of $\boldsymbol{J}_{k|k}$

In OF estimation, we prefer the IKF over the conventional KF because the covariance matrices arising in the conventional KF are typically dense and therefore operations involving these matrices are costly, whereas the precision matrices $\boldsymbol{J}_{k|k}$ and $\boldsymbol{J}_{k|k-1}$ arising in the IKF are sparse. In particular, when using the IKF rather the conventional KF, the sparsity of the precision

matrices is preserved in the update step (3.52), (3.53). Indeed, assuming that $\boldsymbol{J}_{k|k-1}$ is sparse and using the fact that $\boldsymbol{M}_H$ and $\boldsymbol{G}_{k+1}$ are sparse by definition (see (2.30), (2.31), (2.45), and (2.46)), we conclude from (3.52) that $\boldsymbol{J}_{k|k}$ is also sparse. Unfortunately, a similar result does not hold directly for the prediction step, because (3.46) does not yield a sparse $\boldsymbol{J}_{k|k-1}$ even if $\boldsymbol{J}_{k-1|k-1}$ is sparse. For a dense matrix $\boldsymbol{J}_{k|k-1}$, the update step (3.52) will then yield a dense matrix $\boldsymbol{J}_{k|k}$, and henceforth the IKF will be similarly complex as the conventional KF.

To address this problem, we replace (3.46) by a reduced order-approximation that is obtained by expanding the matrix inverse by a series as follows [15]:

$$(\boldsymbol{J}_{k-1|k-1} + \gamma \mathbf{I}_{2MN})^{-1} = \boldsymbol{\Psi}^{-1} - \boldsymbol{\Psi}^{-1}\boldsymbol{\Omega}\boldsymbol{\Psi}^{-1} + \boldsymbol{\Psi}^{-1}\boldsymbol{\Omega}\boldsymbol{\Psi}^{-1}\boldsymbol{\Omega}\boldsymbol{\Psi}^{-1} - \ldots, \tag{3.66}$$

where $\boldsymbol{\Psi}$ and $\boldsymbol{\Omega}$ are chosen such that

$$\boldsymbol{\Psi} + \boldsymbol{\Omega} = \boldsymbol{J}_{k-1|k-1} + \gamma \mathbf{I}_{2MN}. \tag{3.67}$$

This series converges if all eigenvalues of $\boldsymbol{\Psi}^{-1}\boldsymbol{\Omega}$ lie inside the unit disk, or equivalently if $\boldsymbol{\Psi} + \boldsymbol{\Omega}$ is strictly diagonally dominant [30]. Convergence is especially fast if the eigenvalues of $\boldsymbol{\Psi}^{-1}\boldsymbol{\Omega}$ are close to zero. In this thesis, we investigate two alternative definitions of $\boldsymbol{\Psi}$:

1. $\boldsymbol{\Psi}$ is a block diagonal matrix whose 2×2 diagonal blocks are identical to the corresponding diagonal blocks of the original matrix $\boldsymbol{J}_{k-1|k-1} + \gamma \mathbf{I}_{2MN}$ (as proposed in [15]);

2. $\boldsymbol{\Psi}$ is a diagonal matrix whose main diagonal is identical to the corresponding main diagonal of the original matrix $\boldsymbol{J}_{k-1|k-1} + \gamma \mathbf{I}_{2MN}$ (as proposed in [30]).

As a consequence, $\boldsymbol{\Psi}^{-1}$ is sparse. By (3.67), matrix $\boldsymbol{\Omega}$ is then given by the remaining elements of $\boldsymbol{J}_{k-1|k-1} + \gamma \mathbf{I}_{2MN}$, i.e.,

$$\boldsymbol{\Omega} = \boldsymbol{J}_{k-1|k-1} + \gamma \mathbf{I}_{2MN} - \boldsymbol{\Psi}. \tag{3.68}$$

The series in (3.66) may be truncated after any number of terms depending on the desired level of accuracy. However, a large number of terms results in a relatively dense approximation, thus there is a tradeoff between accuracy and computational complexity. We observed that using the first two terms typically yields a good approximation of the inverse. Therefore we approximate (3.46) according to

$$\boldsymbol{J}_{k|k-1} \approx \gamma \mathbf{I}_{2MN} - \gamma^2(\boldsymbol{\Psi}^{-1} + \boldsymbol{\Psi}^{-1}\boldsymbol{\Omega}\boldsymbol{\Psi}^{-1}), \tag{3.69}$$

which, unlike the original expression, is sparse and thus preserves the sparse structure of $\boldsymbol{J}_{k-1|k-1}$ (because $\boldsymbol{\Psi}^{-1}$ is sparse).

We note that besides the presented approach, other techniques for sparse approximate matrix inversion can be found in the literature, see e.g. [31], [32].

3.5.4 Efficient Iterative Computation the OF Estimate

The sparsity of $\boldsymbol{J}_{k|k}$ and $\boldsymbol{J}_{k|k-1}$ can also be exploited for an efficient computation of the OF estimate $\hat{\boldsymbol{w}}_{k,\mathrm{MMSE}}$. We can reformulate (3.54) (with $\lambda = 1$) as $\boldsymbol{J}_{k|k}\hat{\boldsymbol{w}}_{k,\mathrm{MMSE}} = \boldsymbol{J}_{k|k-1}\boldsymbol{\mu}_{k|k-1} + \boldsymbol{G}_{k+1}^T\boldsymbol{d}_k$, i.e., $\hat{\boldsymbol{w}}_{k,\mathrm{MMSE}}$ is the solution to the linear equation

$$\boldsymbol{J}_{k|k}\boldsymbol{w}_k = \boldsymbol{b}_k, \tag{3.70}$$

with $\boldsymbol{b}_k \triangleq \boldsymbol{J}_{k|k-1}\boldsymbol{\mu}_{k|k-1}+\boldsymbol{G}_{k+1}^T\boldsymbol{d}_k$. An iterative solution of this linear equation can be obtained by the *conjugate gradient* (CG) algorithm with Jacobi preconditioning [33]. This method produces a sequence of approximate solutions that converge to the solution $\boldsymbol{w}_k = \boldsymbol{J}_{k|k}^{-1}\boldsymbol{b}_k$ in a reasonable rate. Since the precision matrix $\boldsymbol{J}_{k|k}$ only appears as matrix/vector multiplications, (3.70) can be solved efficiently if $\boldsymbol{J}_{k|k}$ is sparse. Weaknesses of this iterative method are typically, the rate of convergence, the computational complexity per step, the required storage, and the pattern of memory access.

Chapter 4

Variational Bayesian Filtering

In this chapter, we discuss variational Bayesian filtering and its application to OF estimation. First, in Section 4.1, we introduce variational Bayesian inference. Next, we present the mean field approximation in Section 4.2. In Section 4.3, we apply variational inference and the mean field approximation to the general Bayesian filter, which leads to the formulation of the variational Bayesian filter. Finally, in Section 4.4, we apply the variational Bayesian filter to the problem of OF estimation.

4.1 Variational Inference

The principle of *variational inference* (VI) is to approximate the posterior pdf $p(\boldsymbol{\theta}|\boldsymbol{x})$ by a tractable *variational pdf* $q(\boldsymbol{\theta})$ [34]. This pdf is found in a set of variational pdfs, referred to as the *variational family*, such that it is closest to the posterior pdf in terms of a metric.

4.1.1 Kullback-Leibler Divergence

One commonly used metric that measures "closeness" of two pdfs is the *Kullback-Leibler* (KL) divergence [34], which is defined as

$$\mathrm{KL}(p(\boldsymbol{\theta})\|q(\boldsymbol{\theta})) \triangleq \mathrm{E}_p\left\{\ln \frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right\} = \int_{\mathbb{R}^n} p(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{\theta})}{q(\boldsymbol{\theta})} \,\mathrm{d}\boldsymbol{\theta}, \tag{4.1}$$

or equivalently

$$\mathrm{KL}(p(\boldsymbol{\theta})\|q(\boldsymbol{\theta})) = \mathrm{E}_p\{\ln p(\boldsymbol{\theta})\} - \mathrm{E}_p\{\ln q(\boldsymbol{\theta})\}. \tag{4.2}$$

In terms of the KL divergence, the VI problem can be formulated as the optimization problem [34]

$$\widehat{q}(\boldsymbol{\theta}) = \underset{q(\boldsymbol{\theta})\in\mathcal{Q}}{\operatorname{argmin}} \mathrm{KL}(q(\boldsymbol{\theta})\|p(\boldsymbol{\theta}|\boldsymbol{x})), \tag{4.3}$$

where $\mathcal{Q}$ is the variational family.

4.1.2 Evidence Lower Bound

Given the joint pdf $p(\boldsymbol{x}, \boldsymbol{\theta})$ of the data vector $\boldsymbol{x}$ and parameter vector $\boldsymbol{\theta}$, we define the *evidence lower bound* (ELBO) for any variational pdf $q(\boldsymbol{\theta})$ as [34]

$$\mathcal{L}(q) \triangleq \mathrm{E}_q\left\{\ln \frac{p(\boldsymbol{x}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})}\right\} = \int_{\mathbb{R}^n} q(\boldsymbol{\theta}) \ln \frac{p(\boldsymbol{x}, \boldsymbol{\theta})}{q(\boldsymbol{\theta})} \, \mathrm{d}\boldsymbol{\theta} = \mathrm{E}_q\{\ln p(\boldsymbol{x}, \boldsymbol{\theta})\} - \mathrm{E}_q\{\ln q(\boldsymbol{\theta})\}. \tag{4.4}$$

Expanding the KL divergence in (4.3), using (4.2) and the identity

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}, \boldsymbol{\theta})}{p(\boldsymbol{x})}, \tag{4.5}$$

we obtain

$$\begin{aligned} \mathrm{KL}(q(\boldsymbol{\theta}) \| p(\boldsymbol{\theta}|\boldsymbol{x})) &= \mathrm{E}_q\{\ln q(\boldsymbol{\theta})\} - \mathrm{E}_q\{\ln p(\boldsymbol{\theta}|\boldsymbol{x})\} \\ &= \mathrm{E}_q\{\ln q(\boldsymbol{\theta})\} - \mathrm{E}_q\{\ln p(\boldsymbol{x}, \boldsymbol{\theta})\} + \mathrm{E}_q\{\ln p(\boldsymbol{x})\} \\ &= -\mathcal{L}(q) + \ln p(\boldsymbol{x}). \end{aligned} \tag{4.6}$$

The minimization in (4.3) can now be rewritten as

$$\widehat{q}(\boldsymbol{\theta}) = \underset{q(\boldsymbol{\theta}) \in \mathcal{Q}}{\operatorname{argmin}} \{-\mathcal{L}(q) + \ln p(\boldsymbol{x})\} = \underset{q(\boldsymbol{\theta}) \in \mathcal{Q}}{\operatorname{argmin}} \{-\mathcal{L}(q)\} = \underset{q(\boldsymbol{\theta}) \in \mathcal{Q}}{\operatorname{argmax}} \mathcal{L}(q). \tag{4.7}$$

This shows that the VI problem (4.3) can be reformulated as a maximization of the ELBO. This is an essential result in VI, because the optimization can be performed without having direct access to the posterior pdf $p(\boldsymbol{\theta}|\boldsymbol{x})$ (cf. (4.4)).

4.2 Mean Field Approximation

The most important variational family is based on partitioning the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^n$ into $m \leq n$ disjoint sub-vectors, i.e.,

$$\boldsymbol{\theta} = \begin{pmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \\ \vdots \\ \boldsymbol{\theta}_m \end{pmatrix}, \tag{4.8}$$

with the sub-vectors $\boldsymbol{\theta}_j \in \mathbb{R}^{n_j}$ where $j = 1, \ldots, m$ and $\sum_{j=1}^{m} n_j = n$. Let $\mathcal{Q}$ be the variational family whose members $q(\boldsymbol{\theta})$ factorize into the variational pdfs of the sub-vectors $\boldsymbol{\theta}_j$, i.e.,

$$q(\boldsymbol{\theta}) = \prod_{j=1}^{m} q_j(\boldsymbol{\theta}_j). \tag{4.9}$$

Thus, according to the variational approximation $q(\boldsymbol{\theta})$ in (4.9), the sub-vectors $\boldsymbol{\theta}_j$ are statistically independent and distributed according to $q_j(\boldsymbol{\theta}_j)$, and each $q_j(\boldsymbol{\theta}_j)$ is a marginal pdf of $q(\boldsymbol{\theta})$. The variational family $\mathcal{Q}$ defined by (4.9) is termed the *mean field* (MF) variational family [34].

Combining the VI optimization problem defined in (4.7) with the MF variational family defined in (4.9), we recognize that this optimization problem is equivalent to the joint optimization of the ELBO with respect to all $q_j(\boldsymbol{\theta}_j), j = 1, \ldots, m$. To find an approximate solution for this problem, we use an iterative algorithm that optimizes the ELBO with respect to each factor $q_j(\boldsymbol{\theta}_j)$ individually, while the other factors $q_i(\boldsymbol{\theta}_i)$ with $i \neq j$ are fixed. The sequence of optimization steps for $j = 1, \ldots, m$ is repeated iteratively several times, i.e., the iteration index ℓ ranges from 1 to some maximum value and every iteration ℓ has m sub-steps labeled by the sub-step index j. Then in the current iteration, the jth sub-step updates $q_j(\boldsymbol{\theta}_j)$ by solving

$$q_j^*(\boldsymbol{\theta}_j) = \underset{q_j(\boldsymbol{\theta_j}) \in \mathcal{Q}_j}{\operatorname{argmax}} \mathcal{L}(q). \tag{4.10}$$

Here $\mathcal{Q}_j$ is the set of all possible pdfs for a n_j-dimensional random vector and the $q(\boldsymbol{\theta})$ involved in $\mathcal{L}(q)$ is given by (4.9) where all factors $q_i(\boldsymbol{\theta}_i)$ with $i \neq j$ are equal to the result of the most recent update from the previous or current iteration. This iterative algorithm is known as the *coordinate ascent variational inference* (CAVI) algorithm [34].

The optimization problem (4.10) can be solved analytically in terms of an expectation and a normalization. It is shown in [34] that

$$q_j^*(\boldsymbol{\theta}_j) \propto \exp(\mathrm{E}_{q_{\sim j}}\{\ln p(\boldsymbol{x}, \boldsymbol{\theta})\}), \tag{4.11}$$

where

$$\mathrm{E}_{q_{\sim j}}\{\ln p(\boldsymbol{x}, \boldsymbol{\theta})\} = \int_{\mathbb{R}^{n_{\sim j}}} \ln p(\boldsymbol{x}, \boldsymbol{\theta}) q_{\sim j}(\boldsymbol{\theta}_{\sim j}) \, \mathrm{d}\boldsymbol{\theta}_{\sim j}. \tag{4.12}$$

Here, $\boldsymbol{\theta}_{\sim j}$ denotes the parameter vector $\boldsymbol{\theta}$ without the sub-vector $\boldsymbol{\theta}_j$, and $n_{\sim j} \triangleq n - n_j$ is the dimension of $\boldsymbol{\theta}_{\sim j}$. Furthermore, $q_{\sim j}(\boldsymbol{\theta}_{\sim j})$ denotes the posterior pdf $q(\boldsymbol{\theta})$ with the jth factor $q_j(\boldsymbol{\theta}_j)$ removed, i.e.,

$$q_{\sim j}(\boldsymbol{\theta}_{\sim j}) \triangleq \prod_{i \neq j} q_i(\boldsymbol{\theta}_i), \tag{4.13}$$

wherein each factor $q_i(\boldsymbol{\theta}_i), i \neq j$ equals the result of the most recent update from the current or

previous iteration.

4.3 The Variational Bayesian Filter

In the general Bayesian filter considered in Section 3.2, the prediction step (cf. (3.6)) consists of a high-dimensional integral which, in general, is difficult to evaluate. For the linear-Gaussian state-space model, this integral becomes tractable, and the closed-form solution of the prediction step is given by (3.42) and (3.43). However, as discussed in Section 3.5, this is still problematic within the IKF framework because the matrix inversion in (3.42) produces a dense precision matrix, which renders the subsequent update step computationally infeasible. As we show below, using the VI principle presented in Section 4.2, it is possible to derive a *variational Bayesian filter* (VBF) in which the matrix inversion of (3.42) is avoided altogether. If we apply the VBF to the problem of OF estimation, we find that no dense matrices emerge in the calculation, and that the algorithm is computationally feasible.

4.3.1 VI Approach to Bayesian Filtering

The principal idea behind the VBF is to approximate the joint posterior pdf $p(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})$ by a variational pdf from the MF family [23], i.e.,

$$p(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k}) \approx q(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k}) = q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k}). \tag{4.14}$$

Therefore, in the case of the VBF, the role of the posterior pdf $p(\boldsymbol{\theta}|\boldsymbol{x})$ that is approximated in the generic VI approach of Section 4.1 is played by the joint posterior pdf $p(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})$. The application of the product rule (with "background condition" $\boldsymbol{x}_{1:k-1}$) yields

$$p(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k}) = p(\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\boldsymbol{x}_k, \boldsymbol{x}_{1:k-1}) = \frac{p(\boldsymbol{x}_k, \boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})}{p(\boldsymbol{x}_k|\boldsymbol{x}_{1:k-1})}. \tag{4.15}$$

Using the chain rule and the assumption (3.3) along with the assumption that the data vector $\boldsymbol{x}_k$ is conditionally independent of all past data vectors $\boldsymbol{x}_{1:k-1}$ and the previous estimate $\boldsymbol{\theta}_{k-1}$ given $\boldsymbol{\theta}_k$, i.e.,

$$p(\boldsymbol{x}_k|\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}, \boldsymbol{x}_{1:k-1}) = p(\boldsymbol{x}_k|\boldsymbol{\theta}_k), \quad k = 1, 2, \ldots, \tag{4.16}$$

we obtain for the numerator in (4.15)

$$\begin{aligned} p(\boldsymbol{x}_k, \boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1}) &= p(\boldsymbol{x}_k|\boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}, \boldsymbol{x}_{1:k-1})p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}, \boldsymbol{x}_{1:k-1})p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1}) \\ &= p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1}). \end{aligned} \tag{4.17}$$

Note that the joint pdf $p(\boldsymbol{x}_k, \boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})$ here plays the role of the joint pdf $p(\boldsymbol{x}, \boldsymbol{\theta})$ considered in the generic VI approach. Comparing the factorization in (4.14) with (4.8) and (4.9), we see that the full parameter vector is

$$\boldsymbol{\theta}' = \begin{pmatrix} \boldsymbol{\theta}'_1 \\ \boldsymbol{\theta}'_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{\theta}_k \\ \boldsymbol{\theta}_{k-1} \end{pmatrix}, \tag{4.18}$$

and the variational factors $q_j(\boldsymbol{\theta}'_j)$ for $j = 1, 2$ are identified as

$$q_1(\boldsymbol{\theta}'_1) = q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}) \quad \text{and} \quad q_2(\boldsymbol{\theta}'_2) = q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k}). \tag{4.19}$$

Note also that the pdfs $q_{\sim j}(\boldsymbol{\theta}'_{\sim j})$ are given by

$$q_{\sim 1}(\boldsymbol{\theta}'_{\sim 1}) = q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k}) \quad \text{and} \quad q_{\sim 2}(\boldsymbol{\theta}'_{\sim 2}) = q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}). \tag{4.20}$$

We now use the CAVI algorithm to calculate the variational factors $q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$ and $q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})$. By applying (4.11) with the joint pdf $p(\boldsymbol{\theta}', \boldsymbol{x}_k|\boldsymbol{x}_{1:k-1})$ given by (4.17) and the posterior pdf $p(\boldsymbol{\theta}'|\boldsymbol{x}_{1:k})$ given by (4.14), we obtain update equations for our variational factors $q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$ and $q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})$. These update equations will be derived next.

4.3.2 Variational Filtering and Smoothing pdfs

For $q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$, which we refer to as the *variational filtering pdf*, we obtain (cf. (4.11), (4.14), and (4.17))

$$\begin{aligned} q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}) &\propto \exp\left(\mathrm{E}_{q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{x}_k, \boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})\}\right) && (4.21) \\ &\propto \exp\left(\mathrm{E}_{q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{x}_k|\boldsymbol{\theta}_k) p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1}) p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})\}\right) \\ &= \exp\Big(\mathrm{E}_{q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)\} + \mathrm{E}_{q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})\} && (4.22) \\ &\quad + \mathrm{E}_{q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})\}\Big). \end{aligned}$$

Absorbing the term $\mathrm{E}_{q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})\}$ into the multiplicative constant (since it is not a function of $\boldsymbol{\theta}_k$), and using the equality $\mathrm{E}_{q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)\} = \ln\, p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)$, which holds since $p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)$ does not involve $\boldsymbol{\theta}_{k-1}$, we obtain

$$\begin{aligned} q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}) &\propto \exp(\ln\, p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)) \exp\left(\mathrm{E}_{q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})\}\right) \\ &= p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)\Gamma_1(\boldsymbol{\theta}_k), \end{aligned} \tag{4.23}$$

with

$$\Gamma_1(\boldsymbol{\theta}_k) \triangleq \exp\left(\mathrm{E}_{q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})\}\right). \tag{4.24}$$

For $q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})$, which we refer to as the *variational smoothing pdf*, we obtain (cf. (4.11), (4.14), and (4.17))

$$q^*(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k}) \propto \exp\Big(\mathrm{E}_{q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{x}_k, \boldsymbol{\theta}_k, \boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})\}\Big) \tag{4.25}$$

$$\propto \exp\Big(\mathrm{E}_{q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})\}\Big)$$

$$= \exp\Big(\mathrm{E}_{q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)\} + \mathrm{E}_{q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})\} \tag{4.26}$$

$$+ \mathrm{E}_{q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})\}\Big).$$

Absorbing $\mathrm{E}_{q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)\}$ into the multiplicative constant (since it is not a function of $\boldsymbol{\theta}_{k-1}$), and using the equality $\mathrm{E}_{q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})\} = \ln\, p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})$, we obtain

$$q^*(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k}) \propto \exp(\ln\, p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1}))\exp\Big(\mathrm{E}_{q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})\}\Big)$$

$$= p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})\Gamma_2(\boldsymbol{\theta}_{k-1}), \tag{4.27}$$

with

$$\Gamma_2(\boldsymbol{\theta}_{k-1}) \triangleq \exp\Big(\mathrm{E}_{q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})\}\Big). \tag{4.28}$$

4.3.3 The Variational Bayesian Filter Recursion

In the VBF, the posterior pdf $p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$ of $\boldsymbol{\theta}_k$ is approximated by the variational filtering pdf $q(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$. Thus, also $p(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})$ involved in (4.27) is replaced by its approximation $q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})$. That is, the update equation (4.27) is replaced by $q^*(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k}) \propto q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})$ $\times\Gamma_2(\boldsymbol{\theta}_{k-1})$. The VBF recursion for the general state-space model in Section 3.1 is thus given by (4.23) and by this modified version of (4.27):

$$q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}) \propto p(\boldsymbol{x}_k|\boldsymbol{\theta}_k)\Gamma_1(\boldsymbol{\theta}_k), \tag{4.29}$$

$$q^*(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k}) \propto q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})\Gamma_2(\boldsymbol{\theta}_{k-1}), \tag{4.30}$$

with (cf. (4.24), (4.28))

$$\Gamma_1(\boldsymbol{\theta}_k) = \exp\Big(\mathrm{E}_{q^*(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})\}\Big), \tag{4.31}$$

$$\Gamma_2(\boldsymbol{\theta}_{k-1}) = \exp\Big(\mathrm{E}_{q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})}\{\ln\, p(\boldsymbol{\theta}_k|\boldsymbol{\theta}_{k-1})\}\Big). \tag{4.32}$$

Thus, at time step k, $q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})$ is converted into $q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$ and into $q^*(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})$. According to the CAVI algorithm, these conversions are repeated in an iterative manner. This corresponds to performing the following steps:

1. Initialize $q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$ arbitrarily.

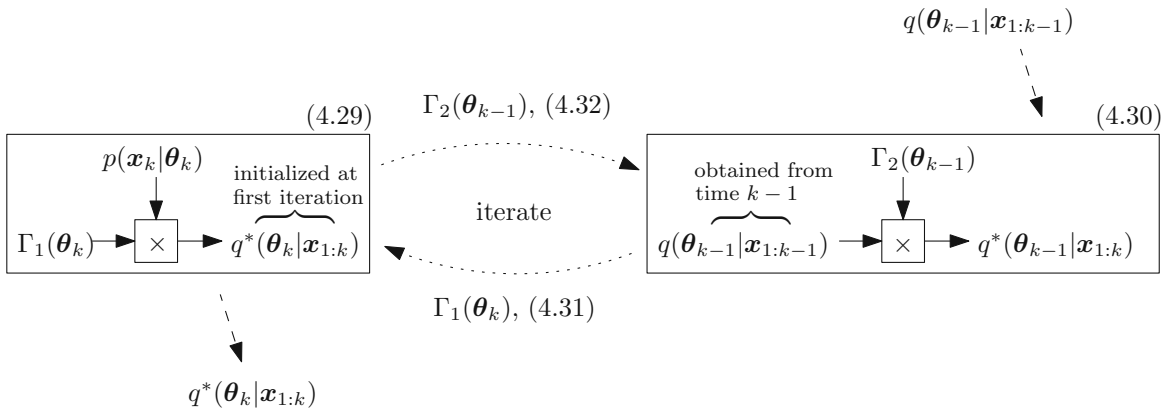


Fig. 4.1: *CAVI iteration performed by the VBF recursion at time k, converting $q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})$ into $q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$. The block $\times$ calculates a normalized pdf from the product of the input functions. Dotted arrows indicate calculating an expectation and passing the result, and dashed arrows indicate passing the result of one recursion to the next.*

2. Calculate $\Gamma_2(\boldsymbol{\theta}_{k-1})$ using (4.32).
3. Calculate $q^*(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k})$ using (4.30). This involves $q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})$, which was calculated at the previous time step $k-1$.
4. Calculate $\Gamma_1(\boldsymbol{\theta}_k)$ using (4.31).
5. Update $q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$ using (4.29).
6. Repeat steps 2–5 until convergence is reached [23, Section 7.2].
7. Use $q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$ as the variational pdf approximating the posterior pdf $p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$.
8. Propagate $q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$ to the next time step $k+1$, where it will take the place of $q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1})$ used at time step k (cf. step 3).

These iteration steps of the CAVI algorithm are visualized in Figure 4.1. At time $k=1$, the VBF algorithm is initialized with an arbitrary pdf $q(\boldsymbol{\theta}_{k-1}|\boldsymbol{x}_{1:k-1}) = q(\boldsymbol{\theta}_0|\boldsymbol{x}_{1:0}) = q(\boldsymbol{\theta}_0)$.

To obtain an estimate of $\boldsymbol{\theta}_k$ from $q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$, we consider the MMSE estimator (see Section 3.2) based on $q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})$, i.e.,

$$\hat{\boldsymbol{\theta}}_{k,\mathrm{VBF}} = \mathrm{E}_{q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})}\{\boldsymbol{\theta}_k\} = \int_{\mathbb{R}^n} \boldsymbol{\theta}_k \; q^*(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}) \; \mathrm{d}\boldsymbol{\theta}_k, \tag{4.33}$$

which approximates $\hat{\boldsymbol{\theta}}_{k,\mathrm{MMSE}} = \mathrm{E}\{\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k}\} = \mathrm{E}_{p(\boldsymbol{\theta}_k|\boldsymbol{x}_{1:k})}\{\boldsymbol{\theta}_k\}$.

4.4 Variational Bayesian Filtering for Optical Flow Estimation

Let us now apply the VBF of the previous section to the linear-Gaussian state-space model in (2.57) and (2.78). The roles of $\boldsymbol{\theta}_k$ and $\boldsymbol{x}_k$ are now played by $\boldsymbol{w}_k$, and $\tilde{\boldsymbol{d}}_k$, respectively.

4.4.1 Calculation of $\Gamma_1(\boldsymbol{w}_k)$ and $\Gamma_2(\boldsymbol{w}_{k-1})$

Formally replacing $\boldsymbol{\theta}_k$ with $\boldsymbol{w}_k$ and $\boldsymbol{x}_k$ with $\tilde{\boldsymbol{d}}_k$, equation (4.31) becomes

$$\Gamma_1(\boldsymbol{w}_k) = \exp\Big(\mathrm{E}_{q(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k})}\{\ln p(\boldsymbol{w}_k|\boldsymbol{w}_{k-1})\}\Big), \tag{4.34}$$

and (4.32) is given by

$$\Gamma_2(\boldsymbol{w}_{k-1}) = \exp\Big(\mathrm{E}_{q(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k})}\{\ln p(\boldsymbol{w}_k|\boldsymbol{w}_{k-1})\}\Big). \tag{4.35}$$

It is seen that $\Gamma_1(\boldsymbol{w}_k)$ and $\Gamma_2(\boldsymbol{w}_{k-1})$ are obtained by evaluating the expectation of the logarithm of the transition pdf $p(\boldsymbol{w}_k|\boldsymbol{w}_{k-1})$ in (2.82) with respect to $q(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k})$ and $q(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k})$, respectively. The logarithm of (2.82) can be written as

$$\ln p(\boldsymbol{w}_k|\boldsymbol{w}_{k-1}) \stackrel{\mathrm{c}}{=} -\frac{\gamma}{2}(\boldsymbol{w}_k - \boldsymbol{w}_{k-1})^T(\boldsymbol{w}_k - \boldsymbol{w}_{k-1}) = -\frac{\gamma}{2}(\boldsymbol{w}_k^T\boldsymbol{w}_k - 2\boldsymbol{w}_k^T\boldsymbol{w}_{k-1} + \boldsymbol{w}_{k-1}^T\boldsymbol{w}_{k-1}). \tag{4.36}$$

Inserting the result into (4.34), we obtain

$$\begin{aligned}\Gamma_1(\boldsymbol{w}_k) &\propto \exp\Big(-\frac{\gamma}{2}\,\mathrm{E}_{q(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k})}\{\boldsymbol{w}_k^T\boldsymbol{w}_k - 2\boldsymbol{w}_k^T\boldsymbol{w}_{k-1} + \boldsymbol{w}_{k-1}^T\boldsymbol{w}_{k-1}\}\Big)\\ &= \exp\Big(-\frac{\gamma}{2}(\boldsymbol{w}_k^T\boldsymbol{w}_k - 2\boldsymbol{w}_k^T\,\mathrm{E}_{q(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k})}\{\boldsymbol{w}_{k-1}\} + \mathrm{E}_{q(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k})}\{\boldsymbol{w}_{k-1}^T\boldsymbol{w}_{k-1}\})\Big)\\ &\propto \exp\Big(-\frac{\gamma}{2}(\boldsymbol{w}_k^T\boldsymbol{w}_k - 2\boldsymbol{w}_k^T\boldsymbol{\mu}_{k-1|k})\Big),\end{aligned} \tag{4.37}$$

with

$$\boldsymbol{\mu}_{k-1|k} \triangleq \mathrm{E}_{q(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k})}\{\boldsymbol{w}_{k-1}\}. \tag{4.38}$$

Here, the term $\mathrm{E}_{q(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k})}\{\boldsymbol{w}_{k-1}^T\boldsymbol{w}_{k-1}\}$ was absorbed into the multiplicative constant since it is not a function of $\boldsymbol{w}_k$. Similarly, by inserting (4.36) into (4.35), we obtain

$$\begin{aligned}\Gamma_2(\boldsymbol{w}_{k-1}) &\propto \exp\Big(-\frac{\gamma}{2}\,\mathrm{E}_{q(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k})}\{\boldsymbol{w}_k^T\boldsymbol{w}_k - 2\boldsymbol{w}_k^T\boldsymbol{w}_{k-1} + \boldsymbol{w}_{k-1}^T\boldsymbol{w}_{k-1}\}\Big)\\ &= \exp\Big(-\frac{\gamma}{2}(\mathrm{E}_{q(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k})}\{\boldsymbol{w}_k^T\boldsymbol{w}_k\} - 2\,\mathrm{E}_{q(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k})}\{\boldsymbol{w}_k^T\}\boldsymbol{w}_{k-1} + \boldsymbol{w}_{k-1}^T\boldsymbol{w}_{k-1})\Big)\\ &\propto \exp\Big(-\frac{\gamma}{2}(\boldsymbol{w}_{k-1}^T\boldsymbol{w}_{k-1} - 2\boldsymbol{\mu}_{k|k}^T\boldsymbol{w}_{k-1})\Big),\end{aligned} \tag{4.39}$$

with

$$\boldsymbol{\mu}_{k|k} \triangleq \mathrm{E}_{q(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k})}\{\boldsymbol{w}_k\}. \tag{4.40}$$

Here, the term $\mathrm{E}_{q(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k})}\{\boldsymbol{w}_k^T\boldsymbol{w}_k\}$ was absorbed into the multiplicative constant since it is not a function of $\boldsymbol{w}_{k-1}$.

4.4.2 Variational Filtering pdf

The update equation (4.29) for the variational filtering pdf $q(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k})$ now reads

$$q^*(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k}) \propto p(\tilde{\boldsymbol{d}}_k|\boldsymbol{w}_k)\Gamma_1(\boldsymbol{w}_k). \tag{4.41}$$

Expanding (2.61), we see that the augmented likelihood $p(\tilde{\boldsymbol{d}}_k|\boldsymbol{w}_k)$ function is given by

$$p(\tilde{\boldsymbol{d}}_k|\boldsymbol{w}_k) \propto \exp\Big(-\frac{1}{2}(\tilde{\boldsymbol{d}}_k^T\boldsymbol{Q}\tilde{\boldsymbol{d}}_k - 2\boldsymbol{w}_k^T\tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{d}}_k + \boldsymbol{w}_k^T\tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{G}}_{k+1}\boldsymbol{w}_k)\Big), \tag{4.42}$$

where we used the fact that $\boldsymbol{Q}$ is symmetric (cf. (2.60)). Inserting (4.42) and (4.37) into (4.41), we obtain

$$q^*(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k}) \propto \exp\Big(-\frac{\gamma}{2}(\boldsymbol{w}_k^T\boldsymbol{w}_k - 2\boldsymbol{w}_k^T\boldsymbol{\mu}_{k-1|k}) - \frac{1}{2}(\boldsymbol{w}_k^T\tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{G}}_{k+1}\boldsymbol{w}_k - 2\boldsymbol{w}_k^T\tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{d}}_k)\Big). \tag{4.43}$$

Note that the term $\tilde{\boldsymbol{d}}_k^T\boldsymbol{Q}\tilde{\boldsymbol{d}}_k$ was dropped since it is not a function of $\boldsymbol{w}_k$. Rewriting (4.43) as

$$q^*(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k}) \propto \exp\Big(-\frac{1}{2}(\boldsymbol{w}_k^T(\gamma\mathbf{I}_{2MN} + \tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{G}}_{k+1})\boldsymbol{w}_k - 2\boldsymbol{w}_k^T(\gamma\boldsymbol{\mu}_{k-1|k} + \tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{d}}_k))\Big), \tag{4.44}$$

and completing the square with respect to $\boldsymbol{w}_k$, the updated variational filtering pdf is recognized to be Gaussian i.e.,

$$\begin{aligned} q^*(\boldsymbol{w}_k|\tilde{\boldsymbol{d}}_{1:k}) &= \mathcal{N}(\boldsymbol{w}_k; \boldsymbol{\mu}_{k|k}^*, \boldsymbol{J}_{k|k}^{*-1}) \\ &= (2\pi)^{-NM}\ \det(\boldsymbol{J}_{k|k}^*)^{1/2} \exp\Big(-\frac{1}{2}(\boldsymbol{w}_k - \boldsymbol{\mu}_{k|k}^*)^T\boldsymbol{J}_{k|k}^*(\boldsymbol{w}_k - \boldsymbol{\mu}_{k|k}^*)\Big), \end{aligned} \tag{4.45}$$

with precision matrix

$$\boldsymbol{J}_{k|k}^* = \gamma\mathbf{I}_{2MN} + \tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{G}}_{k+1} \tag{4.46}$$

and mean

$$\boldsymbol{\mu}_{k|k}^* = \boldsymbol{J}_{k|k}^{*-1}(\tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{d}}_k + \gamma\boldsymbol{\mu}_{k-1|k}). \tag{4.47}$$

We recall from (2.65) and (2.66) that

$$\tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{G}}_{k+1} = \lambda\boldsymbol{G}_{k+1}^T\boldsymbol{G}_{k+1} + \beta\boldsymbol{M}_H \tag{4.48}$$

and

$$\tilde{\boldsymbol{G}}_{k+1}^T\boldsymbol{Q}\tilde{\boldsymbol{d}}_k = \lambda\boldsymbol{G}_{k+1}^T\boldsymbol{d}_k. \tag{4.49}$$

Thus, the expressions (4.46) and (4.47) can be rewritten as

$$\boldsymbol{J}^*_{k|k} = \gamma \mathbf{I}_{2MN} + \lambda \boldsymbol{G}^T_{k+1}\boldsymbol{G}_{k+1} + \beta \boldsymbol{M}_H \tag{4.50}$$

and

$$\boldsymbol{\mu}^*_{k|k} = \boldsymbol{J}^{*-1}_{k|k}\Big(\lambda \boldsymbol{G}^T_{k+1}\boldsymbol{d}_k + \gamma \boldsymbol{\mu}_{k-1|k}\Big), \tag{4.51}$$

respectively. Note that $\boldsymbol{J}^*_{k|k}$ can be calculated directly, whereas $\boldsymbol{\mu}^*_{k|k}$ is a function of $\boldsymbol{\mu}_{k-1|k}$, which was obtained in the previous iteration of the CAVI algorithm (cf. (4.55)).

4.4.3 Variational Smoothing pdf

The update equation (4.30) for the variational smoothing pdf $q(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k})$ reads

$$q^*(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k}) \propto q(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k-1})\Gamma_2(\boldsymbol{w}_{k-1}). \tag{4.52}$$

Inserting (4.39) into (4.52) yields

$$q^*(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k}) \propto q(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k-1}) \exp\Big(-\frac{\gamma}{2}(\boldsymbol{w}^T_{k-1}\boldsymbol{w}_{k-1} - 2\boldsymbol{\mu}^T_{k|k}\boldsymbol{w}_{k-1})\Big). \tag{4.53}$$

Identifying the first factor in (4.53) as the variational filtering pdf from recursion $k-1$, i.e., $q^*(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k-1})$ as given by (4.45) with k replaced by $k-1$, we can rewrite (4.53) as

$$q^*(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k}) \propto \exp\Big(-\frac{1}{2}(\boldsymbol{w}^T_{k-1}(\gamma\mathbf{I}_{2MN}+\boldsymbol{J}_{k-1|k-1})\boldsymbol{w}_{k-1} - 2\boldsymbol{w}^T_{k-1}(\gamma\boldsymbol{\mu}_{k|k}+\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}))\Big).$$

Note that the term $\boldsymbol{\mu}^T_{k-1|k-1}\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}$ was dropped because it is not a function of $\boldsymbol{w}_{k-1}$. By rearranging terms and completing the square with respect to $\boldsymbol{w}_{k-1}$, the updated variational smoothing pdf is recognized to be Gaussian i.e.,

$$q^*(\boldsymbol{w}_{k-1}|\tilde{\boldsymbol{d}}_{1:k}) = \mathcal{N}(\boldsymbol{w}_{k-1}; \boldsymbol{\mu}^*_{k-1|k}, \boldsymbol{J}^{*-1}_{k-1|k}),$$

with precision matrix

$$\boldsymbol{J}^*_{k-1|k} = \gamma \mathbf{I}_{2MN} + \boldsymbol{J}_{k-1|k-1} \tag{4.54}$$

and mean

$$\boldsymbol{\mu}^*_{k-1|k} = \boldsymbol{J}^{*-1}_{k-1|k}(\gamma\boldsymbol{\mu}_{k|k} + \boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}). \tag{4.55}$$

Note that $\boldsymbol{J}^*_{k-1|k}$ can be calculated directly, whereas $\boldsymbol{\mu}^*_{k-1|k}$ is a function of $\boldsymbol{\mu}_{k|k}$, which was obtained in the previous sub-step of the current iteration of the CAVI algorithm (cf. (4.51)).

4.4.4 Closed-form Solution for $\boldsymbol{\mu}_{k|k}$

The expressions (4.51) of $\boldsymbol{\mu}_{k|k}$ and (4.55) of $\boldsymbol{\mu}_{k-1|k}$ establish a circular dependency, because the expression of $\boldsymbol{\mu}_{k|k}$ involves $\boldsymbol{\mu}_{k-1|k}$ and the expression of $\boldsymbol{\mu}_{k-1|k}$ involves $\boldsymbol{\mu}_{k|k}$. According to the CAVI algorithm, the calculation of $\boldsymbol{\mu}_{k|k}$ and $\boldsymbol{\mu}_{k-1|k}$ would be performed iteratively by alternating between (4.51) and (4.55), where $\boldsymbol{\mu}_{k-1|k}$ and $\boldsymbol{\mu}_{k|k}$ are always set to the most recent value. However, as we will show next, we can avoid this iterative scheme and obtain a closed-form solution for $\boldsymbol{\mu}_{k|k}$ and $\boldsymbol{\mu}_{k-1|k}$.

Substituting the expression of $\boldsymbol{\mu}^*_{k-1|k}$ in (4.55) for $\boldsymbol{\mu}_{k-1|k}$ in (4.51), and using $\boldsymbol{J}^*_{k-1|k} = \boldsymbol{J}_{k-1|k}$ and $\boldsymbol{J}^*_{k|k} = \boldsymbol{J}_{k|k}$, yields

$$\begin{aligned}\boldsymbol{\mu}^*_{k|k} &= \boldsymbol{J}^{-1}_{k|k}(\lambda \boldsymbol{G}^T_{k+1}\boldsymbol{d}_k + \gamma \boldsymbol{J}^{-1}_{k-1|k}(\gamma \boldsymbol{\mu}_{k|k} + \boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}))\\ &= \gamma^2 \boldsymbol{J}^{-1}_{k|k}\boldsymbol{J}^{-1}_{k-1|k}\boldsymbol{\mu}_{k|k} + \lambda \boldsymbol{J}^{-1}_{k|k}(\boldsymbol{G}^T_{k+1}\boldsymbol{d}_k + \gamma \boldsymbol{J}^{-1}_{k-1|k}\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}).\end{aligned}$$

Setting $\boldsymbol{\mu}^*_{k|k} = \boldsymbol{\mu}_{k|k}$ and grouping terms, we obtain further

$$(\mathbf{I}_{2MN} - \gamma^2 \boldsymbol{J}^{-1}_{k|k}\boldsymbol{J}^{-1}_{k-1|k})\boldsymbol{\mu}_{k|k} = \boldsymbol{J}^{-1}_{k|k}(\lambda \boldsymbol{G}^T_{k+1}\boldsymbol{d}_k + \gamma \boldsymbol{J}^{-1}_{k-1|k}\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}),$$

which yields the explicit solution

$$\boldsymbol{\mu}_{k|k} = (\mathbf{I}_{2MN} - \gamma^2 \boldsymbol{J}^{-1}_{k|k}\boldsymbol{J}^{-1}_{k-1|k})^{-1}\boldsymbol{J}^{-1}_{k|k}(\lambda \boldsymbol{G}^T_{k+1}\boldsymbol{d}_k + \gamma \boldsymbol{J}^{-1}_{k-1|k}\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}). \tag{4.56}$$

To avoid nested inversions of time-varying matrices in (4.56), we first rewrite this expression as

$$\boldsymbol{\mu}_{k|k} = (\mathbf{I}_{2MN} - \gamma^2 \boldsymbol{J}^{-1}_{k|k}\boldsymbol{J}^{-1}_{k-1|k})^{-1}\boldsymbol{J}^{-1}_{k|k}\boldsymbol{J}^{-1}_{k-1|k}(\boldsymbol{J}_{k-1|k}\lambda \boldsymbol{G}^T_{k+1}\boldsymbol{d}_k + \gamma \boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}).$$

Applying now the relation $\boldsymbol{A}^{-1}\boldsymbol{B}^{-1} = (\boldsymbol{B}\boldsymbol{A})^{-1}$ twice, we obtain

$$\begin{aligned}\boldsymbol{\mu}_{k|k} &= (\mathbf{I}_{2MN} - \gamma^2(\boldsymbol{J}_{k-1|k}\boldsymbol{J}_{k|k})^{-1})^{-1}(\boldsymbol{J}_{k-1|k}\boldsymbol{J}_{k|k})^{-1}(\boldsymbol{J}_{k-1|k}\lambda \boldsymbol{G}^T_{k+1}\boldsymbol{d}_k + \gamma \boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1})\\ &= (\boldsymbol{J}_{k-1|k}\boldsymbol{J}_{k|k}(\mathbf{I}_{2MN} - \gamma^2(\boldsymbol{J}_{k-1|k}\boldsymbol{J}_{k|k})^{-1}))^{-1}(\boldsymbol{J}_{k-1|k}\lambda \boldsymbol{G}^T_{k+1}\boldsymbol{d}_k + \gamma \boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1})\\ &= (\boldsymbol{J}_{k-1|k}\boldsymbol{J}_{k|k} - \gamma^2\mathbf{I}_{2MN})^{-1}(\boldsymbol{J}_{k-1|k}\lambda \boldsymbol{G}^T_{k+1}\boldsymbol{d}_k + \gamma \boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}).\end{aligned} \tag{4.57}$$

The VBF for OF estimation is now given by (4.50) and by (4.57) with the right-hand side of (4.54) substituted for $\boldsymbol{J}_{k-1|k}$, i.e.,

$$\boldsymbol{J}_{k|k} = \gamma \mathbf{I}_{2MN} + \lambda \boldsymbol{G}^T_{k+1}\boldsymbol{G}_{k+1} + \beta \boldsymbol{M}_H, \tag{4.58}$$

$$\boldsymbol{\mu}_{k|k} = ((\gamma \mathbf{I}_{2MN} + \boldsymbol{J}_{k-1|k-1})\boldsymbol{J}_{k|k} - \gamma^2\mathbf{I}_{2MN})^{-1}((\gamma \mathbf{I}_{2MN} + \boldsymbol{J}_{k-1|k-1})\lambda \boldsymbol{G}^T_{k+1}\boldsymbol{d}_k + \gamma \boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}), \tag{4.59}$$

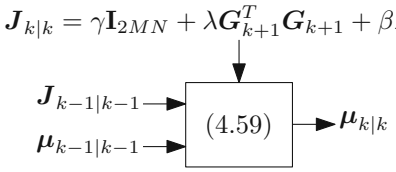


Fig. 4.2: *Structure of the VBF algorithm for sequential OF estimation based on the linear-Gaussian state-space model at time k. The precision matrix $\boldsymbol{J}_{k|k}$ is calculated by (4.58) and does not involve results from the previous time $k-1$. The mean $\boldsymbol{\mu}_{k|k}$ is calculated by (4.59) and involves $\boldsymbol{J}_{k|k}$ as well as $\boldsymbol{\mu}_{k-1|k-1}$ and $\boldsymbol{J}_{k-1|k-1}$.*

where $k = 1, \ldots, L-1$. Thus, in contrast to the IKF described in Section 3.5, the precision matrix $\boldsymbol{J}_{k|k}$ is calculated according to (4.58) independently for every iteration, and not recursively using $\boldsymbol{J}_{k-1|k-1}$, i.e., the precision matrix of the previous iteration. On the other hand, according to (4.59), the calculation of the mean $\boldsymbol{\mu}_{k|k}$ requires the precision matrix of the current iteration step, $\boldsymbol{J}_{k|k}$, as well as the mean and precision matrix of the previous iteration, $\boldsymbol{\mu}_{k-1|k-1}$ and $\boldsymbol{J}_{k-1|k-1}$. The recursion for $\boldsymbol{\mu}_{k|k}$ is initialized at time $k = 1$ with arbitrary $\boldsymbol{J}_{k-1|k-1} = \boldsymbol{J}_{0|0} = \boldsymbol{J}_0$ and $\boldsymbol{\mu}_{k-1|k-1} = \boldsymbol{\mu}_{0|0} = \boldsymbol{\mu}_0$. Note that the VBF algorithm is not structured into a prediction step and an update step like the IKF. The structure of the VBF algorithm is visualized in Figure 4.2.

To obtain an estimate of $\boldsymbol{w}_k$ from $q(\boldsymbol{w}_k|\boldsymbol{x}_{1:k}) = \mathcal{N}(\boldsymbol{w}_k; \boldsymbol{\mu}_{k|k}, \boldsymbol{J}_{k|k}^{-1})$, we consider the MMSE estimator (see Section 3.2) based on $q(\boldsymbol{w}_k|\boldsymbol{x}_{1:k})$, i.e.,

$$\begin{aligned}\hat{\boldsymbol{w}}_{k,\mathrm{VBF}} = \mathrm{E}_{q(\boldsymbol{w}_k|\boldsymbol{x}_{1:k})}\{\boldsymbol{w}_k\} = \boldsymbol{\mu}_{k|k} = \big((\gamma\mathbf{I}_{2MN} + \boldsymbol{J}_{k-1|k-1})\boldsymbol{J}_{k|k} - \gamma^2\mathbf{I}_{2MN}\big)^{-1} \\ \times \big((\gamma\mathbf{I}_{2MN} + \boldsymbol{J}_{k-1|k-1})\lambda\boldsymbol{G}_{k+1}^T\boldsymbol{d}_k + \gamma\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}\big)\end{aligned} \tag{4.60}$$

(see (4.59)), which approximates $\hat{\boldsymbol{w}}_{k,\mathrm{MMSE}} = \mathrm{E}\{\boldsymbol{w}_k|\boldsymbol{x}_{1:k}\} = \mathrm{E}_{p(\boldsymbol{w}_k|\boldsymbol{x}_{1:k})}\{\boldsymbol{w}_k\}$.

4.4.5 Reduced-Parameter Formulation

As in the IKF (see Section 3.5), the VBF can be reformulated as a function of $\eta = \frac{\gamma}{\lambda}$ and $\psi = \frac{\beta}{\lambda}$, rather than β, γ, and λ. To this end, we rewrite $\boldsymbol{J}_{k|k}$ in (4.58) as

$$\boldsymbol{J}_{k|k} = \lambda\bar{\boldsymbol{J}}_{k|k}, \tag{4.61}$$

with

$$\bar{\boldsymbol{J}}_{k|k} \triangleq \frac{1}{\lambda}\boldsymbol{J}_{k|k} = \frac{\gamma}{\lambda}\mathbf{I}_{2MN} + \boldsymbol{G}_{k+1}^T\boldsymbol{G}_{k+1} + \frac{\beta}{\lambda}\boldsymbol{M}_H = \eta\mathbf{I}_{2MN} + \boldsymbol{G}_{k+1}^T\boldsymbol{G}_{k+1} + \psi\boldsymbol{M}_H. \tag{4.62}$$

Inserting (4.62) into (4.59) then yields

$$\begin{aligned}
\boldsymbol{\mu}_{k|k} &= \left((\gamma\mathbf{I}_{2MN} + \lambda\bar{\boldsymbol{J}}_{k-1|k-1})\lambda\bar{\boldsymbol{J}}_{k|k} - \gamma^2\mathbf{I}_{2MN}\right)^{-1} \\
&\quad \times \left((\gamma\mathbf{I}_{2MN} + \lambda\bar{\boldsymbol{J}}_{k-1|k-1})\lambda\boldsymbol{G}_{k+1}^T\boldsymbol{d}_k + \gamma\lambda\bar{\boldsymbol{J}}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}\right) \\
&= \left(\lambda^2\left(\left(\frac{\gamma}{\lambda}\mathbf{I}_{2MN} + \bar{\boldsymbol{J}}_{k-1|k-1}\right)\bar{\boldsymbol{J}}_{k|k} - \left(\frac{\gamma}{\lambda}\right)^2\mathbf{I}_{2MN}\right)\right)^{-1} \\
&\quad \times \lambda^2\left(\left(\frac{\gamma}{\lambda}\mathbf{I}_{2MN} + \bar{\boldsymbol{J}}_{k-1|k-1}\right)\boldsymbol{G}_{k+1}^T\boldsymbol{d}_k + \frac{\gamma}{\lambda}\bar{\boldsymbol{J}}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}\right) \\
&= \frac{1}{\lambda^2}\left(\left(\frac{\gamma}{\lambda}\mathbf{I}_{2MN} + \bar{\boldsymbol{J}}_{k-1|k-1}\right)\bar{\boldsymbol{J}}_{k|k} - \left(\frac{\gamma}{\lambda}\right)^2\mathbf{I}_{2MN}\right)^{-1} \\
&\quad \times \lambda^2\left(\left(\frac{\gamma}{\lambda}\mathbf{I}_{2MN} + \bar{\boldsymbol{J}}_{k-1|k-1}\right)\boldsymbol{G}_{k+1}^T\boldsymbol{d}_k + \frac{\gamma}{\lambda}\bar{\boldsymbol{J}}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}\right) \\
&= \left((\eta\mathbf{I}_{2MN} + \bar{\boldsymbol{J}}_{k-1|k-1})\bar{\boldsymbol{J}}_{k|k} - \eta^2\mathbf{I}_{2MN}\right)^{-1}(\eta\mathbf{I}_{2MN} + \bar{\boldsymbol{J}}_{k-1|k-1})\boldsymbol{G}_{k+1}^T\boldsymbol{d}_k + \eta\bar{\boldsymbol{J}}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}).
\end{aligned} \tag{4.63}$$

Thus, we summarize our reduced-parameter reformulation of the VBF as

$$\bar{\boldsymbol{J}}_{k|k} = \eta\mathbf{I}_{2MN} + \boldsymbol{G}_{k+1}^T\boldsymbol{G}_{k+1} + \psi\boldsymbol{M}_H, \tag{4.64}$$

$$\boldsymbol{\mu}_{k|k} = \left((\eta\mathbf{I}_{2MN} + \bar{\boldsymbol{J}}_{k-1|k-1})\bar{\boldsymbol{J}}_{k|k} - \eta^2\mathbf{I}_{2MN}\right)^{-1}(\eta\mathbf{I}_{2MN} + \bar{\boldsymbol{J}}_{k-1|k-1})\boldsymbol{G}_{k+1}^T\boldsymbol{d}_k + \eta\bar{\boldsymbol{J}}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}). \tag{4.65}$$

Assuming that $\boldsymbol{\mu}_{0|0}$ is chosen such that it does not depend on λ (e.g., $\boldsymbol{\mu}_{0|0} = \mathbf{0}_{2MN}$), we see that the VBF is a function of η and ψ rather than β, γ, and λ. This formulation of the VBF is identical to the VBF given in (4.58) and (4.59) for $\lambda = 1$, which will therefore be considered in the following.

4.4.6 Efficient Iterative Computation of the OF Estimate

Looking at the expression of the precision matrix $\boldsymbol{J}_{k|k}$ of the VBF given in (4.58), and using that $\boldsymbol{G}_{k+1}$ and $\boldsymbol{M}_H$ are sparse by definition (see (2.30), (2.31), and (2.45), (2.46)), we see that $\boldsymbol{J}_{k|k}$ is sparse for $k = 1, 2, \ldots$. Similarly to the IKF, the sparsity of $\boldsymbol{J}_{k|k}$ can be exploited for an efficient computation of the OF estimate $\hat{\boldsymbol{w}}_{k,\mathrm{VBF}}$. We can rewrite (4.60) (with $\lambda = 1$) as

$$\left((\gamma\mathbf{I}_{2MN} + \boldsymbol{J}_{k-1|k-1})\boldsymbol{J}_{k|k} - \gamma^2\mathbf{I}_{2MN}\right)\hat{\boldsymbol{w}}_{k,\mathrm{VBF}} = (\gamma\mathbf{I}_{2MN} + \boldsymbol{J}_{k-1|k-1})\boldsymbol{G}_{k+1}^T\boldsymbol{d}_k + \gamma\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}. \tag{4.66}$$

Thus, $\hat{\boldsymbol{w}}_{k,\mathrm{VBF}}$ is the solution to the linear equation $\boldsymbol{A}_k\boldsymbol{w}_k = \boldsymbol{b}_k$ with $\boldsymbol{A}_k = (\gamma\mathbf{I}_{2MN} + \boldsymbol{J}_{k-1|k-1})\boldsymbol{J}_{k|k} - \gamma^2\mathbf{I}_{2MN}$ and $\boldsymbol{b}_k = (\gamma\mathbf{I}_{2MN} + \boldsymbol{J}_{k-1|k-1})\boldsymbol{G}_{k+1}^T\boldsymbol{d}_k + \gamma\boldsymbol{J}_{k-1|k-1}\boldsymbol{\mu}_{k-1|k-1}$. Since $\boldsymbol{J}_{k-1|k-1}$ and $\boldsymbol{J}_{k|k}$ are sparse, $\boldsymbol{A}_k$ will also be sparse. Thus, as for the IKF in Section 3.5, we can use the CG algorithm with Jacobi preconditioning [33] to solve (4.66) efficiently.

Chapter 5

Simulation Results

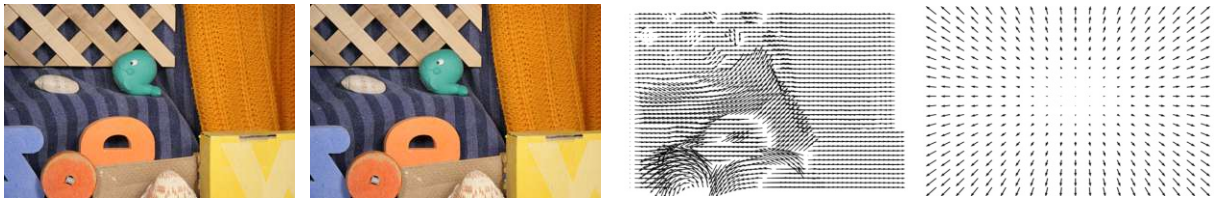
In this chapter, we evaluate the performance of the IKF and the VBF for OF estimation, derived in Section 3.5 and Section 4.4, respectively, on artificial and real image sequences. In Section 5.1, we present visualization methods and a quantitative metric, which we will use in Section 5.3 to assess the accuracy of the OF estimates. In Section 5.2, we introduce the datasets underlying our experiments. Finally, in Section 5.3, we present and discuss the results of our experiments.

5.1 Visualization and Error Metrics

In order to compare the performance of different OF estimation algorithms, multiple approaches for evaluation of the OF estimates have been developed [35]. In this section, we describe methods for visual representation of the OF that enable a qualitative evaluation, and we define error metrics that allow a quantitative evaluation.

5.1.1 Visualization

In the literature, two methods for visualizing the OF field have been established. First, the visualization using arrows [35], which is shown in Figure 5.1. For a given frame index k, this method directly shows the OF vectors $(u[i,j,k]\ v[i,j,k])^T$ on a uniform grid, where the origin of the (i,j)-th vector is shifted to (x_i, y_j). This method offers an easy interpretation of the OF field. However, for high-resolution images, the arrows are overlapping, which necessitates subsampling of the OF field in order to obtain a clear visualization. Second, the visualization using color coding [35], which associates hue to orientation and saturation to magnitude, as shown in Figure 5.2. This method allows the visualization of a dense OF field without subsampling, and is thus better suited for high-resolution images.



a) b) c) d)

Fig. 5.1: *OF visualization using arrows. The OF field corresponding to the temporally successive images depicted in a) and b) is visualized in c) using arrows that indicate the direction and magnitude of the movement. For reference, in d) we show the arrow pattern corresponding to a diverging OF field with the OF magnitude increasing proportionally to the distance from the center.*

a) b) c) d)

Fig. 5.2: *OF visualization using color coding. The OF field corresponding to the temporally successive images depicted in a) and b) is visualized in c) using color coding. For reference, in d) we show the color pattern corresponding to a diverging OF field, with the OF magnitude increasing proportionally to the distance from the center. Note that OF orientation and OF magnitude are represented by hue and saturation, respectively.*

5.1.2 Error Metrics

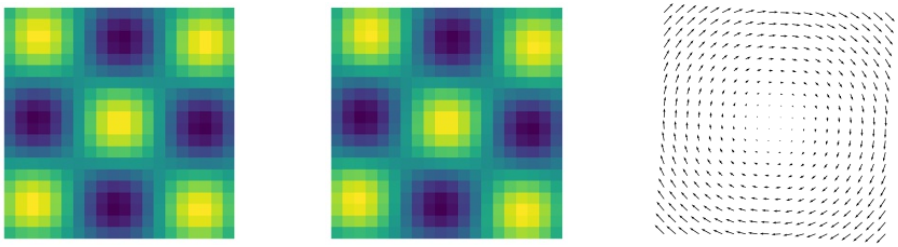
For a quantitative evaluation of the OF estimates, the following error metric is commonly used in the literature. Let $(\hat{u}[i,j,k]\ \hat{v}[i,j,k])^T$ (cf. (2.11)) denote the estimated OF vector for a pixel at position (i,j) and time k, and let $(\bar{u}[i,j,k]\ \bar{v}[i,j,k])^T$ denote the corresponding ground-truth OF vector. The *endpoint error* (EE) at position (i,j) and time k is then defined as the Euclidian distance between the estimated OF vector and the the ground-truth OF vector, i.e. [35]

$$\mathrm{EE}(i,j,k) \triangleq \sqrt{(\hat{u}[i,j,k]-\bar{u}[i,j,k])^2+(\hat{v}[i,j,k]-\bar{v}[i,j,k])^2}. \tag{5.1}$$

The EE is rather insensitive to OF estimation errors with small magnitude but strongly penalizes OF estimation errors with large magnitude.

5.2 Datasets

Next, we describe three datasets that we will use to evaluate the performance of our methods. First, we present synthetic image sequences that show moving brightness patterns. We consider the rotation of the whole image, and small brightness patterns ("objects") that move in front of a static background. These image sequences will also be used to validate the concept of temporal



a) b) c)

Fig. 5.3: *Two successive image frames of the rotation dataset are shown in a) and b). The corresponding OF is shown in c).*

coherence introduced in Section 2.3. Then, we consider the real image sequences provided by the college of Middlebury.[1] This dataset is freely available and widely used for evaluating OF estimation algorithms.

5.2.1 Rotation Dataset

Our first synthetic dataset, termed the rotation dataset, was generated according to the equations stated in [27, Section 5]. The first image, at time $k = 1$, was constructed by evaluating the spatial function

$$F[x, y, 1] = \frac{1}{2}(\cos(\pi x)\cos(\pi y) + 1) \tag{5.2}$$

for (x, y) taken from a uniformly spaced 20×20 grid within $[-1, 1] \times [-1, 1]$. Then, the subsequent images for $k = 2, 3, \ldots$ were generated recursively as

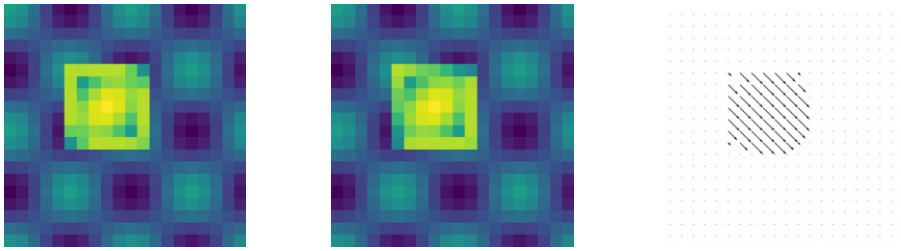
$$F[x, y, k] = F[x, y, k-1] - F_x[x, y, k-1]u[x, y] - F_y[x, y, k-1]v[x, y], \tag{5.3}$$

where $F_x[x, y, k]$ and $F_y[x, y, k]$ are defined in (2.24) and (2.25), respectively. Using (5.2) and (5.3), we generated image sequences consisting of 20 frames. Here, the OF functions are constant with respect to time k and given by

$$u[x, y] = -\pi \sin(\pi x)\cos(\pi y), \quad v[x, y] = -\pi \cos(\pi x)\sin(\pi y). \tag{5.4}$$

Since the OF functions do not change with time, the rotation angle is constant for all frames. Two successive image frames along with the corresponding ground-truth OF are shown in Figure 5.3.

[1] `https://vision.middlebury.edu/flow/data/`



a) b) c)

Fig. 5.4: *Two successive image frames of the moving object dataset are shown in a) and b). The corresponding OF is shown in c).*

5.2.2 Moving Object Dataset

The moving object dataset features a square object of 7×7 pixels with a distinguishable brightness pattern that moves linearly in front of a static background. The brightness pattern $\boldsymbol{Z}$ is given by

$$Z[i,j] \triangleq \sin\left(([j-1]n+i)\frac{\pi}{n^2}\right) + \sin\left(([i-1]n+j)\frac{\pi}{n^2}\right), \tag{5.5}$$

for $i = 1,\ldots,n$ and $j = 1,\ldots,n$, where $n = 7$. In order to increase the contrast to the background, the object contains a one-pixel border constituted by pixels with constant brightness. The background is equal to the first frame of the rotation dataset. Bilinear interpolation was used to enable displacements that do not conform to the pixel grid. We consider a frame-to-frame displacement of 0.3 pixels in the direction from the upper left corner to the lower right corner. Two successive frames along with the corresponding ground-truth OF are shown in Figure 5.4. We generated image sequences of 20 frames where all images were constructed on a grid of 20×20 pixels.

5.2.3 Middlebury Dataset

The Middlebury dataset contains synthetic and real image sequences. The ground-truth OF for two successive frames is publicly available only for some of the sequences. Two successive frames from the "Rubber Whale" sequence of the Middlebury dataset with the corresponding ground-truth OF were shown in Figures 5.1 and 5.2.

5.3 Results and Discussion

Next, we present and discuss our simulation results. In particular, we also discuss benefits and drawbacks of using temporal coherence, introduced in Section 2.3, compared to using spatial regularization alone. As described in Sections 2.2 and 2.3, the influence of spatial regularization and of temporal coherence is quantified by the parameters β and γ, respectively, where larger

values correspond to a stronger influence. We evaluate the accuracy and complexity of the IKF and the VBF on the synthetic datasets presented in the previous section, and on those sequences of the Middlebury dataset where the ground truth is available. To study the effect of spatial regularization and temporal coherence, we vary the parameter tuple (β, γ) on a uniform grid within $[0.1, 1] \times [0.01, 5]$. The lower limit of this grid was chosen such that the linear equations solved in the execution of our methods (see Sections 3.5.4 and 4.4.6) are still well-conditioned, and the upper limit was chosen such that the influence of the spatial and temporal regularization can be investigated over a sufficiently large value range. For each dataset and each parameter tuple (β, γ), we evaluate:

- the average EE (AEE) with the averaging performed over all pixels and all frames;
- the runtime of the methods averaged over 15 execution runs for the synthetic datasets and three execution runs for the Middlebury dataset;
- and the number of iterations of the CG method that is needed to achieve a difference of 10^{-5} or smaller between two subsequent iteration steps (see Sections 3.5.4 and 4.4.6), averaged over 15 execution runs for the synthetic datasets and three execution runs for the Middlebury dataset.

In addition, we present the optimum values of the parameter tuple (β, γ) (minimizing the AEE of the IKF or the VBF) and the corresponding minimal AEE for the considered sequences of the Middlebury dataset.

For the CG method, the final threshold is chosen as 10^{-5} because using a smaller value would lead to an excessive computational complexity. In order to improve the estimation accuracy of the IKF, the update step (see (3.52) and (3.53)) is executed m times for each time k. In the m-th execution, the linearization point $\boldsymbol{w}_{k,0}^{(m)}$ (see (2.36)) is set to the estimate of the $(m-1)$-th execution, i.e.,

$$\boldsymbol{w}_{k,0}^{(m)} = \hat{\boldsymbol{w}}_k^{(m-1)}. \tag{5.6}$$

Looking at (2.36), we see that $\boldsymbol{w}_{k,0}^{(m)}$ is used to calculate the data term $\boldsymbol{d}_k^{(m)}$ from which the estimate $\hat{\boldsymbol{w}}_k^{(m)}$ is obtained (cf. (3.54)). This repeated execution is initialized by setting

$$\boldsymbol{w}_{k,0}^{(0)} = \hat{\boldsymbol{w}}_{k-1}, \tag{5.7}$$

and terminated when

$$|\hat{\boldsymbol{w}}_k^{(m)} - \hat{\boldsymbol{w}}_k^{(m-1)}| \leq 0.001 \quad \text{or} \quad m = 30. \tag{5.8}$$

The threshold in (5.8) is chosen as 0.001 because using a smaller value does not affect the accuracy of the estimates significantly. For the same reason, we use the restriction to $m \leq 30$.

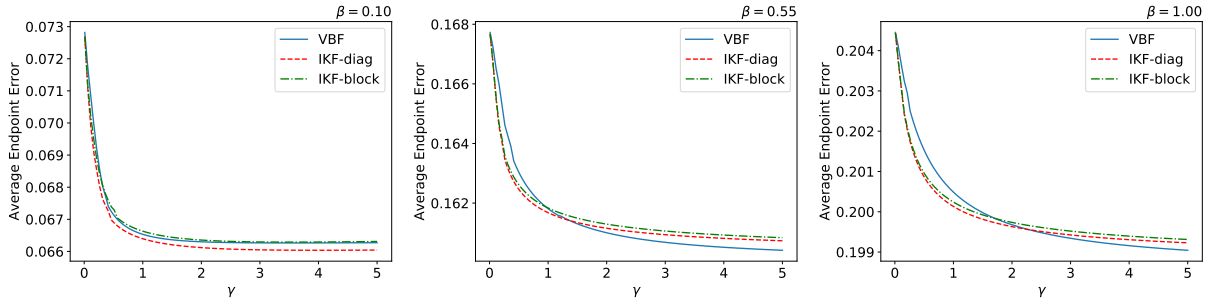


Fig. 5.5: *AEE vs. γ for the rotation dataset and a) $\beta = 0.1$, b) $\beta = 0.55$, and c) $\beta = 1$.*

Similarly, the calculation of the VBF estimate (see (4.59)) is repeated multiple times in each recursion, with the linearization point updated accordingly. As discussed in Section 3.5, we consider both a block-diagonal and a diagonal definition of $\mathbf{\Psi}$. We will refer to the IKF using the block-diagonal definition as "IKF-block" and to the IKF using the diagonal definition as "IKF-diag".

5.3.1 Rotation Dataset

We first present our results for the rotation dataset. In Figure 5.5, the AEE of the three methods (IKF-block, IKF-diag, and VBF) is plotted versus the temporal coherence parameter γ for three different values of the spatial regularization parameter β. We see that for all three methods and values of β, the AEE decreases with increasing γ. Thus, temporal coherence typically improves the accuracy of the OF estimates. For $\beta = 0.1$, IKF-diag yields slightly more accurate estimates than IKF-block and VBF, whereas the AEEs of IKF-block and VBF are comparable. As the spatial regularization parameter β increases, VBF achieves a slightly higher accuracy than the IKF methods for γ sufficiently large, while the results of IKF-diag and IKF-block are similar.

Figure 5.6 shows the average runtime versus γ. Comparing the two IKF methods, we see that IKF-block is considerately slower than IKF-diag. This is intuitive because $\mathbf{\Psi}$ is denser for IKF-block than for IKF-diag, and thus the matrix inversion in the prediction step (cf. (3.46)) takes longer. The runtimes of VBF and IKF-diag are comparable. We furthermore see that the runtime of all methods initially decreases for increasing values of γ but then stays approximately constant for γ larger than a certain value below 1.

Finally, Figure 5.7 depicts the average number of CG iterations versus γ. It can be seen that for all three methods, the number of CG iterations decreases with increasing γ, although for IKF-block and IKF-diag it effectively stays constant for γ above a certain value below 1. Thus, temporal coherence typically leads to a reduced number of CG iterations, which in turn results in a reduced runtime. It can also be seen that for all filter methods, the number of CG iterations increases with a higher value of β, i.e., with a stronger spatial regularization. This may be due

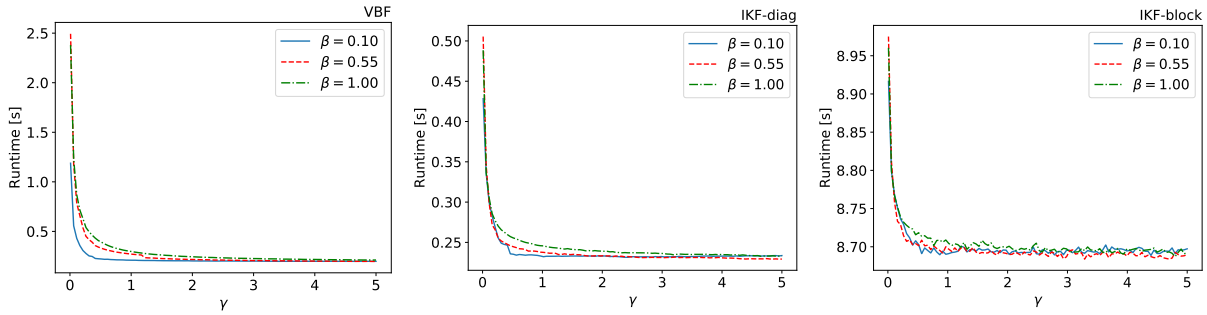


Fig. 5.6: *Average runtime vs. γ for the rotation dataset and a) VBF, b) IKF-diag, and c) IKF-block.*

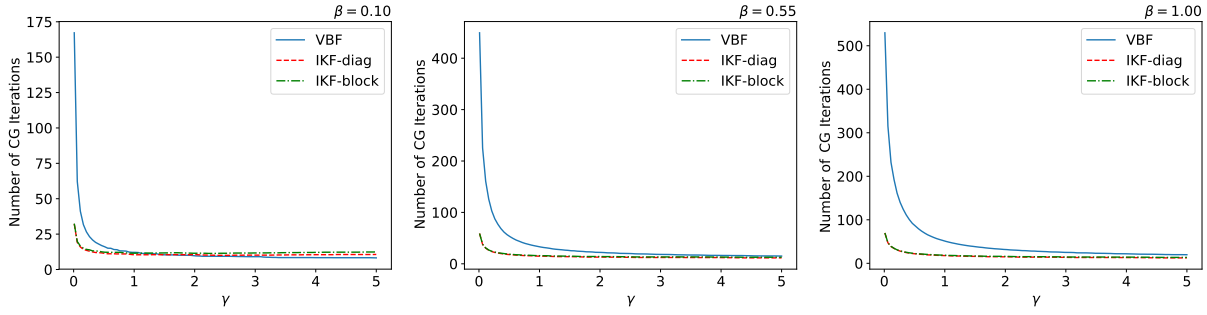


Fig. 5.7: *Average number of CG iterations vs. γ for the rotation dataset and a) $\beta = 0.1$, b) $\beta = 0.55$, and c) $\beta = 1$.*

to the fact that for a stronger spatial regularization, the system matrix of the equation system iteratively solved by the CG method is more different from a diagonal matrix.

5.3.2 Moving Object Dataset

Next, we present results for the moving object dataset. In Figure 5.8, the AEE versus γ is plotted for three different values of β. We see that for all values of β, the AEE of the VBF attains its minimum for some $\gamma \in [1, 3]$ and increases as γ moves away from that point. This result indicates that temporal coherence can improve the accuracy of the estimation if γ is set correctly. For γ too small, the effect of temporal coherence is negligible, whereas for γ too large, the estimated OF field retains a "trailing trace" after the moving object as depicted for VBF in Figure 5.9. This latter effect increases the overall error. The IKF methods show a similar behavior for $\beta = 0.55$ and $\beta = 1$, whereas for $\beta = 0.1$ the AEE monotonously decreases with growing γ within the displayed range $\gamma \in [0, 5]$. In fact, for $\beta = 0.1$, the minimum AEE of the IKF methods is obtained around $\gamma = 10$. For $\beta = 0.1$, VBF achieves a lower AEE than the IKF methods for $\gamma < 3$ and a higher AEE for $\gamma > 3.5$. For $\beta = 0.55$ and $\beta = 1$, on the other hand, IKF-diag always yields a lower AEE than IKF-block.

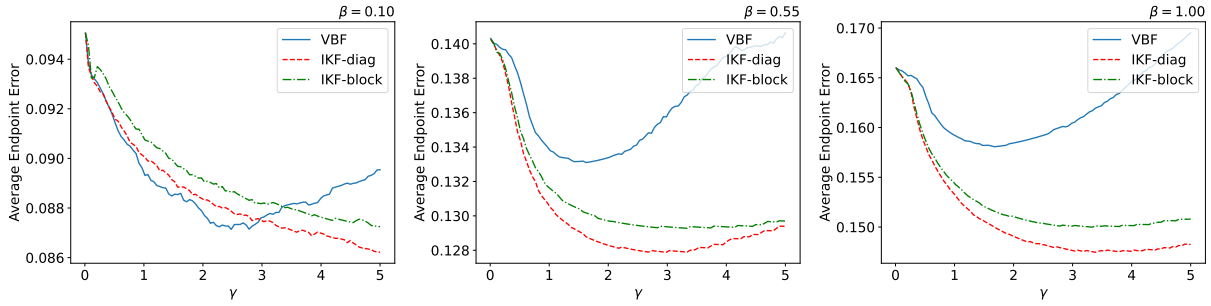


Fig. 5.8: *AEE vs. γ for the moving object dataset and a) $\beta = 0.1$, b) $\beta = 0.55$, and c) $\beta = 1$.*

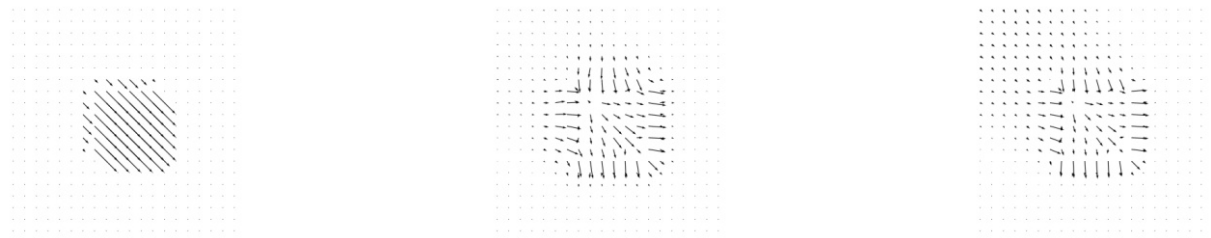


Fig. 5.9: *OF fields for the moving object dataset. a) Ground truth, b) and c) estimated OF fields obtained with VBF using $\beta = 0.1$ and b) $\gamma = 2.5$ and c) $\gamma = 5$.*

In Figure 5.10, the average runtime is plotted versus γ. For all values of β, the runtime of all methods decreases with growing γ for γ above a certain threshold, indicating that temporal coherence here reduces the runtime. For smaller values of γ, the runtime partly increases with growing γ. Furthermore, the runtime is larger for a higher value of β. As for the rotation dataset, IKF-block is slower than IKF-diag due to the higher density of $\mathbf{\Psi}$. For $\beta = 0.55$ and $\beta = 1$, the runtime of VBF is larger than the runtime of IKF-diag, whereas for $\beta = 0.1$, the runtime of the two methods is comparable.

In Figure 5.11, the average number of CG iterations is plotted versus γ. The results are

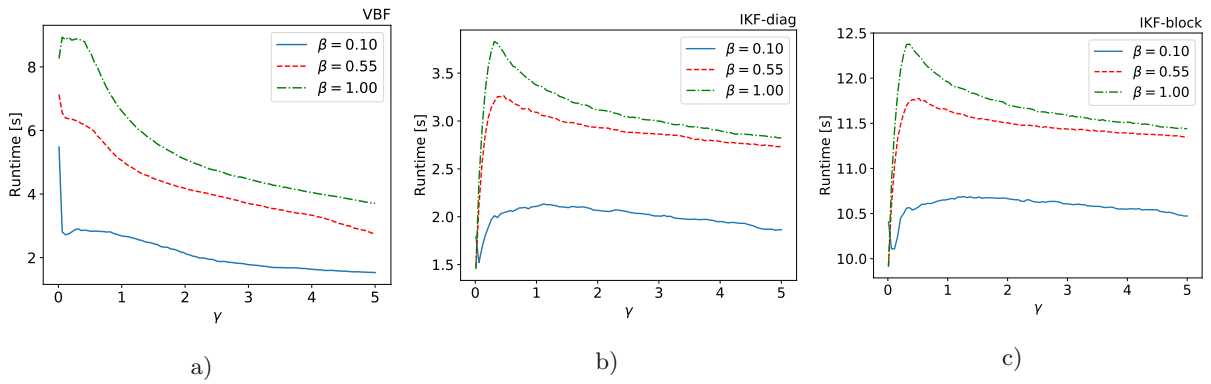
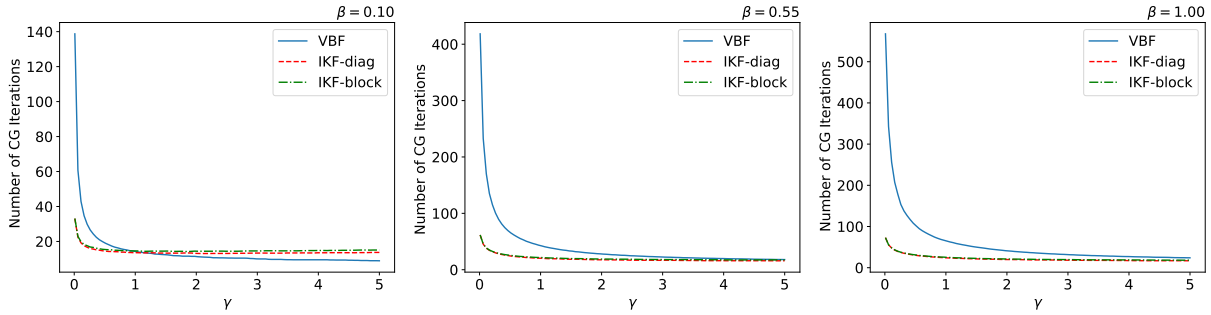


Fig. 5.10: *Average runtime vs. γ for the moving object dataset and a) VBF, b) IKF-diag, and c) IKF-block.*



a) b) c)

Fig. 5.11: *Average number of CG iterations vs. γ for the moving object dataset and a) $\beta = 0.1$, b) $\beta = 0.55$, and c) $\beta = 1$.*

similar to those obtained for the rotation dataset.

5.3.3 Middlebury Dataset

Finally, we analyze the effect of temporal coherence on the accuracy and runtime of our methods using the "Hydrangea" sequence of the Middlebury dataset. We found that the conclusions drawn from this analysis apply also to all the remaining sequences of the Middlebury dataset for which a ground truth is available. Thus, for the other sequences, we only show the AEE for the optimal values of the parameters β and γ.

In Figure 5.12, the AEE is plotted versus γ for three different values of β. We see that for all three methods, the AEE initially increases with increasing γ and remains constant for larger values of γ for $\beta = 0.55$ and $\beta = 1$. For $\beta = 0.1$, the AEE of the IKF methods slightly decreases for growing values of γ whereas the AEE of VBF still increases. Thus, contrary to the result obtained for the rotation and moving object datasets, temporal coherence reduces the accuracy of the OF estimates. This is probably caused by large displacements between two subsequent frames in the sequence. For $\beta = 0.1$, IKF-block achieves a smaller AEE than IKF-diag, while VBF yields the highest AEE. As the spatial regularization parameter β increases, VBF achieves a lower AEE than the IKF methods, whereas IKF-diag and IKF-block yield the same AEE.

Figure 5.13 shows the average runtime versus γ. Comparing the two IKF methods, we again see that IKF-block is slower than IKF-diag, due to the higher density of $\mathbf{\Psi}$ as explained in Section 5.3.1. In contrast to the other considered datasets, VBF now is slower than the IKF methods for all values of β. We conjecture that this is because, using the considered values of β and γ, the system matrix of the equation system solved by the CG method is more poorly conditioned for VBF than for the IKF methods. We furthermore see that for all values of β, the runtime of VBF decreases for $\gamma < 1.2$ and remains constant as γ exceeds this value; thus, temporal coherence results in a reduced runtime of VBF. IKF-diag and IKF-block show this behavior only for $\beta = 0.55$ and $\beta = 1$. For $\beta = 0.1$, the runtime initially decreases with growing

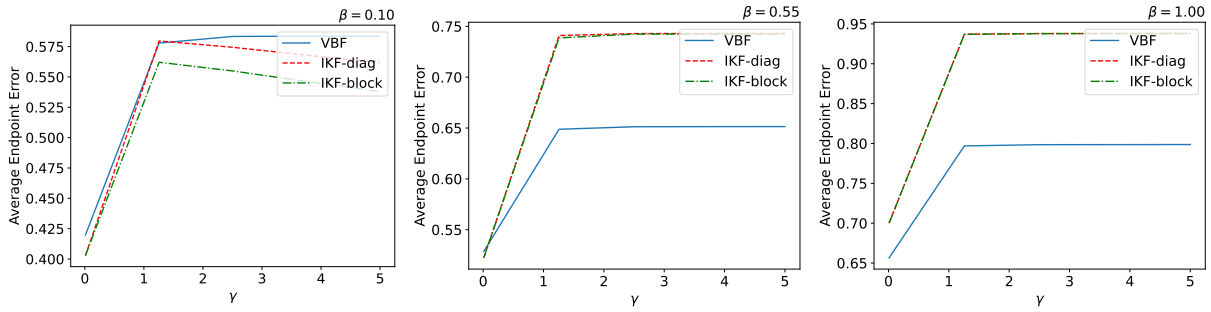


Fig. 5.12: *AEE vs. γ for the "Hydrangea" sequence of the Middlebury dataset and a) $\beta = 0.1$, b) $\beta = 0.55$, and c) $\beta = 1$.*

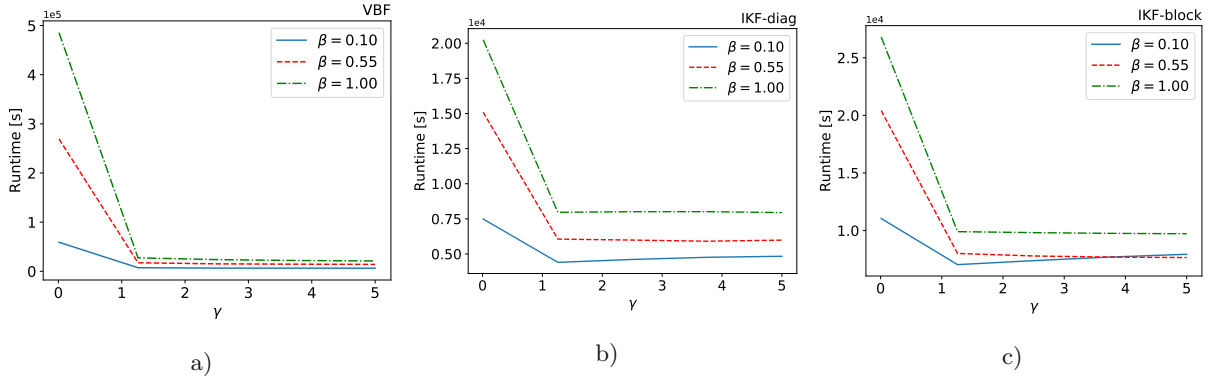
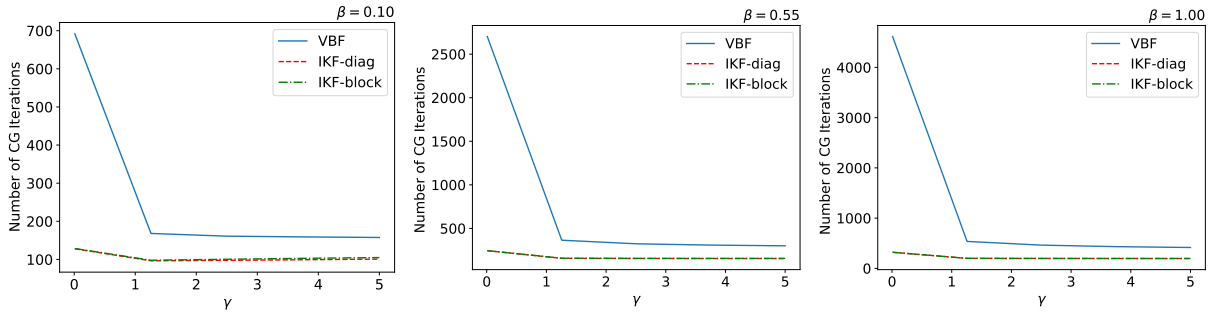


Fig. 5.13: *Average runtime vs. γ for the "Hydrangea" sequence of the Middlebury dataset and a) VBF, b) IKF-diag, and c) IKF-block.*

γ and then slightly increases for $\gamma > 1.2$.

In Figure 5.14, the average number of CG iterations is plotted versus γ. The results are again similar to those we obtained for the other datasets.

Finally, we present the optimal values of the parameters β and γ along with the corresponding AEE for those sequences of the Middlebury dataset where the ground truth is available. These results were obtained by means of the grid search described at the beginning of this section. We found for each image sequence individually that the parameter values $\beta = 0.1$ and $\gamma = 0.01$ (the lower bound for γ used in our grid search) produce the lowest AEE. This indicates that, as mentioned earlier, temporal coherence is not beneficial for the Middlebury dataset. For these optimal parameter values, the absolute AEE and the AEE relative to the maximum magnitude of the ground truth OF field of the corresponding sequence are given for each considered sequence and for each of the three methods in Table 5.1. It can be seen that the AEE values obtained by the three methods tend to be very similar, with the exception of the image sequences "Urban2", "Urban3", and "Grove2", where VBF achieves a lower AEE than the IKF methods.



a) b) c)

Fig. 5.14: *Average number of CG iterations vs. γ for the "Hydrangea" sequence of the Middlebury dataset and a) $\beta = 0.1$, b) $\beta = 0.55$, and c) $\beta = 1$.*

Sequence	Filter Method and AEE		
	IKF-diag	IKF-block	VBF
Rubber Whale	0.414 (8.97 %)	0.414 (8.97 %)	0.411 (8.9 %)
Venus	0.838 (8.93 %)	0.838 (8.93 %)	0.838 (8.93 %)
Hydrangea	0.402 (3.61 %)	0.402 (3.61 %)	0.419 (3.76 %)
Dimetrodon	0.220 (4.71 %)	0.220 (4.71 %)	0.220 (4.71 %)
Urban 2	4.987 (22.47 %)	4.975 (22.42 %)	3.591 (16.18 %)
Urban 3	2.933 (16.65 %)	2.932 (16.64 %)	2.814 (15.97 %)
Grove 2	0.354 (7.03 %)	0.354 (7.03 %)	0.296 (5.88 %)
Grove3	1.115 (5.99 %)	1.115 (5.99 %)	1.102 (5.92 %)

Tab. 5.1: *Minimum AEE obtained for the optimal parameter values $\beta = 0.1$ and $\gamma = 0.01$ for the considered sequences of the Middlebury dataset.*

5.3.4 Summary

To investigate the effect of using temporal coherence in OF estimation, we evaluated the performance of the IKF, both with the block-diagonal and the diagonal approximation, and of the VBF on two synthetic datasets and certain sequences of the Middlebury dataset. To this end, we varied the parameter tuple (β, γ) over a uniformly spaced grid and calculated the AEE, the average runtime, and the average number of CG iterations for all three methods. We showed that for some datasets, temporal coherence leads to a higher accuracy and a lower runtime when compared to using spatial regularization alone, but this is not true for all datasets. We also showed that the considered methods have similar accuracy but differ in computational complexity.

In addition, for the synthetic datasets, we also measured the sparsity of the precision matrices calculated in our methods. As discussed in Sections 3.5 and 4.4, the sparsity measure is 99.3% for IKF-diag and VBF and 99.6% for IKF-block.

Chapter 6

Conclusions

In this thesis, we studied Bayesian probabilistic models and methods for optical flow (OF) estimation with both spatial regularization and temporal coherence. We first reviewed the classical deterministic OF model based on the brightness constancy constraint and a spatial coherence assumption. Next, we reformulated the deterministic model in a probabilistic setting and subsequently extended the resulting probabilistic model to account for temporal coherence of the OF field. We then described the optimal Bayesian state estimator (filter) for a general non-linear and non-Gaussian state-space model. Restricting to a linear-Gaussian state-space model, we next derived the information form of the Kalman filter from the general Bayesian filter and specialized the filter equations to the problem of OF estimation. As an alternative approach to OF estimation, we considered variational Bayesian inference and discussed the mean field approximation and the coordinate ascent algorithm. The application of variational Bayesian inference to the filtering problem led to the formulation of the variational Bayesian filter and its specialization to OF estimation.

We evaluated the accuracy and computational complexity of the information Kalman filter and the variational Bayesian filter for artificially generated data and a commonly used real dataset. Our experimental results show that accounting for temporal coherence in OF estimation can yield improvements in terms of accuracy and computational complexity, but this is not generally true.

To model temporal coherence, we used the Eulerian evolution model. It is possible that other evolution models, such as the Langrangian evolution model proposed in [28], allow an improved exploitation of temporal coherence in OF estimation. This appears to be an interesting direction of future research. Furthermore, we conjecture that performance improvements can also be obtained by modeling the measurement noise vectors $\boldsymbol{n}_k$, $\boldsymbol{m}_{u,k}$, and $\boldsymbol{m}_{v,k}$ in (2.53) and (2.56) using a "robust" distribution like the Laplacian or Student's t-distribution, rather than the Gaussian distribution.

Bibliography

[1] M. Jakubowski and G. Pastuszak, "Block-based motion estimation algorithms — A survey," *Opto-Electronics Review*, vol. 21, no. 1, pp. 86–102, 2013.

[2] T.-C. Huang, C.-K. Chang, C.-H. Liao, and Y.-J. Ho, "Quantification of blood flow in internal cerebral artery by optical flow method on digital subtraction angiography in comparison with time-of-flight magnetic resonance angiography," *PloS one*, vol. 8, no. 1, p. e54678, 2013.

[3] B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, and N. Navab, "Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming," in *Proc. Biennial International Conference on Information Processing in Medical Imaging*, (Rolduc Abbey, Kerkrade, Netherlands), pp. 408–420, 2007.

[4] W. Shi, X. Zhuang, L. Pizarro, W. Bai, H. Wang, K.-P. Tung, P. Edwards, and D. Rueckert, "Registration using sparse free-form deformations," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Nice, France), pp. 659–666, 2012.

[5] J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, *et al.*, "A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations," in *Proc. International Conference on Medical Image Computing and Computer-Assisted Intervention*, (Utrecht, Netherlands), pp. 573–581, 2001.

[6] A. Giachetti, M. Campani, and V. Torre, "The use of optical flow for road navigation," *IEEE Transactions on Robotics and Automation*, vol. 14, no. 1, pp. 34–48, 1998.

[7] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? The KITTI vision benchmark suite," in *Proc. 2012 IEEE Conference on Computer Vision and Pattern Recognition*, (Providence, Rhode Island, USA), pp. 3354–3361, IEEE, 2012.

[8] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *International Journal of Computer Vision*, vol. 25, no. 1, pp. 23–48, 1997.

[9] R. Cutler and M. Turk, "View-based interpretation of real-time optical flow for gesture recognition," in *Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition*, (Nara, Japan), pp. 416–421, 1998.

[10] P. Héas, E. Mémin, N. Papadakis, and A. Szantai, "Layered estimation of atmospheric mesoscale dynamics from satellite imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 12, pp. 4087–4104, 2007.

[11] T. Corpetti, É. Mémin, and P. Pérez, "Dense estimation of fluid flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 365–380, 2002.

[12] D. Heitz, E. Mémin, and C. Schnörr, "Variational fluid flow measurements from image sequences: Synopsis and perspectives," *Experiments in Fluids*, vol. 48, no. 3, pp. 369–393, 2010.

[13] T. Liu and L. Shen, "Fluid flow and optical flow," *Journal of Fluid Mechanics*, vol. 614, p. 253, 2008.

[14] B. K. Horn and B. G. Schunck, "Determining optical flow," in *Techniques and Applications of Image Understanding*, vol. 281, pp. 319–331, International Society for Optics and Photonics, 1981.

[15] T. M. Chin, W. C. Karl, and A. S. Willsky, "Probabilistic and sequential computation of optical flow using temporal coherence," *IEEE Transactions on Image Processing*, vol. 3, no. 6, pp. 773–788, 1994.

[16] M. J. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[17] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *Proc. European Conference on Computer Vision*, (Prague, Czech Republic), pp. 25–36, 2004.

[18] A. Bruhn and J. Weickert, "Towards ultimate motion estimation: Combining highest accuracy with real-time performance," in *Proc. Tenth IEEE International Conference on Computer Vision (ICCV'05)*, vol. 1, (Beijing, China), pp. 749–755, 2005.

[19] A. Salgado and J. Sánchez, "Temporal constraints in large optical flow estimation," in *Proc. International Conference on Computer Aided Systems Theory*, (Las Palmas de Gran Canaria, Spain), pp. 709–716, 2007.

[20] S. Volz, A. Bruhn, L. Valgaerts, and H. Zimmer, "Modeling temporal coherence for optical flow," in *Proc. 2011 International Conference on Computer Vision*, (Barcelona, Spain), pp. 1116–1123, 2011.

[21] C. Bailer, B. Taetz, and D. Stricker, "Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation," in *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, (Santiago, Chile), December 2015.

[22] Z. Ren, O. Gallo, D. Sun, M.-H. Yang, E. B. Sudderth, and J. Kautz, "A fusion approach for multi-frame optical flow estimation," in *Proc. 2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, (Waikoloa Village, HI, USA), pp. 2077–2086, 2019.

[23] V. Šmídl and A. Quinn, *The Variational Bayes Method in Signal Processing*. Springer Science & Business Media, 2006.

[24] J. Weickert, A. Bruhn, T. Brox, and N. Papenberg, "A survey on variational optic flow methods for small displacements," in *Mathematical Models for Registration and Applications to Medical Imaging*, pp. 103–136, Springer, 2006.

[25] S. M. Kay, *Fundamentals of Statistical Signal Processing*. Prentice Hall, 1993.

[26] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *International Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.

[27] J. Sun, F. J. Quevedo, and E. Bollt, "Bayesian optical flow with uncertainty quantification," *Inverse Problems*, vol. 34, no. 10, p. 105008, 2018.

[28] T. M. Chin, *Dynamic estimation in computational vision*. PhD thesis, Massachusetts Institute of Technology, 1992.

[29] B. D. Anderson and J. B. Moore, *Optimal Filtering*. Dover, 2005.

[30] T. M. Chin, W. C. Karl, and A. S. Willsky, "Sequential filtering for multi-frame visual reconstruction," *Signal Processing*, vol. 28, no. 3, pp. 311–333, 1992.

[31] A. Asif and J. M. Moura, "Block matrices with l-block-banded inverse: Inversion algorithms," *IEEE Transactions on Signal Processing*, vol. 53, no. 2, pp. 630–642, 2005.

[32] M. J. Grote and T. Huckle, "Parallel preconditioning with sparse approximate inverses," *SIAM Journal on Scientific Computing*, vol. 18, no. 3, pp. 838–853, 1997.

[33] G. H. Golub and C. F. Van Loan, *Matrix Computations*. The Johns Hopkins University Press, fourth ed., 1996.

[34] D. M. Blei, A. Kucukelbir, and J. D. McAuliffe, "Variational inference: A review for statisticians," *Journal of the American Statistical Association*, vol. 112, no. 518, pp. 859–877, 2017.

[35] D. Fortun, P. Bouthemy, and C. Kervrann, "Optical flow modeling and computation: A survey," *Computer Vision and Image Understanding*, vol. 134, pp. 1–21, 2015.

Hiermit erkläre ich, dass die vorliegende Arbeit gemäß dem Code of Conduct – Regeln zur Sicherung guter wissenschaftlicher Praxis (in der aktuellen Fassung des jeweiligen Mitteilungsblattes der TU Wien), insbesondere ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel, angefertigt wurde. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet.
Die Arbeit wurde bisher weder im In– noch im Ausland in gleicher oder in ähnlicher Form in anderen Prüfungsverfahren vorgelegt.

Wien, September 29, 2021 *Angelika Andrea Hermanek*