Journal of Multivariate Analysis 198 (2023) 105209

Contents lists available at ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Extending compositional data analysis from a graph signal processing perspective

Christopher Rieser *, Peter Filzmoser

Institute of Statistics and Mathematical Methods in Economics, TU Wien, Vienna, Austria

ARTICLE INFO

Article history:
Received 6 September 2022
Received in revised form 13 June 2023
Accepted 13 June 2023
Available online 17 June 2023

AMS 2022 subject classifications:
62H99

Keywords:
Compositional data
Graph Laplacian
Graph signal processing
Graph theory
Log-ratio analysis

ABSTRACT

Traditional methods for the analysis of compositional data consider the log-ratios between all different pairs of variables with equal weight, typically in the form of aggregated contributions. This is not meaningful in contexts where it is known that a relationship only exists between very specific variables (e.g. for metabolomic pathways), while for other pairs a relationship does not exist. Modeling absence or presence of relationships is done in graph theory, where the vertices represent the variables, and the connections refer to relations. This paper links compositional data analysis with graph signal processing, and it extends the Aitchison geometry to a setting where only selected log-ratios can be considered. The presented framework retains the desirable properties of scale invariance and compositional coherence. A real data example from bioinformatics underlines the usefulness of this approach.

1. Introduction

Since the fundamental work of Aitchison [2], the field of compositional data analysis (CoDa) has received a lot of attention. Analyzing positive multivariate data from a compositional point of view shifts the focus from the Euclidean perspective of absolute quantities to the view of relative information. Many data sets have since then been recognized to be of compositional nature. Examples include Microbiome data Gloor et al. [20], omics data Quinn et al. [44], time-use data Dumuid et al. [12], economical data and many more. Good overviews of standard theory on CoDa can be found in Aitchison [3], Pawlowsky-Glahn and Egozcue [41] or Filzmoser et al. [16].

In the following we will denote the space of multivariate positive real values $\{(x_1, \dots, x_D)^\top \in \mathbb{R}^D \mid x_j > 0 \, \forall j \in \{1, \dots, D\}\}$ by $\mathbb{R}^D_+$ and define the D-part simplex as

$$\mathcal{S}^D := \left\{ (x_1, \dots, x_D)^\top \in \mathbb{R}^D_+ \,\middle|\, \sum_{j=1}^{D} x_j = 1 \right\} \subset \mathbb{R}^D_+ .$$

For two compositions $\boldsymbol{x} = (x_1, \dots, x_D)^\top, \boldsymbol{y} = (y_1, \dots, y_D)^\top \in \mathbb{R}^D_+$ and $\alpha \in \mathbb{R}$ the following two operations called perturbation and powering are defined,

- $\boldsymbol{x} \oplus \boldsymbol{y} := (x_1 y_1, \dots, x_D y_D)^\top$
- $\alpha \odot \boldsymbol{x} := (x_1^\alpha, \dots, x_D^\alpha)^\top$,

* Corresponding author.
E-mail address: christopher.rieser@tuwien.ac.at (C. Rieser).

https://doi.org/10.1016/j.jmva.2023.105209

as well as their scaled versions

- $\boldsymbol{x} \oplus_{\mathcal{A}} \boldsymbol{y} \coloneqq \frac{1}{\sum_{j=1}^{D} x_j y_j} \boldsymbol{x} \oplus \boldsymbol{y}$
- $\alpha \odot_{\mathcal{A}} \boldsymbol{x} \coloneqq \frac{1}{\sum_{j=1}^{D} x_j^{\alpha}} \alpha \odot \boldsymbol{x}$,

see Billheimer et al. [6], Pawlowsky-Glahn and Egozcue [40]. The D-part simplex $\mathcal{S}^D$ is equipped with an inner product, known as the Aitchison inner product,

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{A}} \coloneqq \frac{1}{2D} \sum_{i,j=1}^{D} \ln\left(\frac{x_i}{x_j}\right) \ln\left(\frac{y_i}{y_j}\right), \tag{1}$$

such that $(\mathcal{S}^D, \langle \cdot, \cdot \rangle_{\mathcal{A}}, \oplus_{\mathcal{A}}, \odot_{\mathcal{A}})$ is a Hilbert space with neutral element $\frac{1}{D}(1, \ldots, 1)^{\top} \in \mathbb{R}_+^D$ and norm $\|\boldsymbol{x}\|_{\mathcal{A}} \coloneqq \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\mathcal{A}}}$, see Pawlowsky-Glahn et al. [43]. A common tool for the analysis of compositional data is the clr (centered log-ratio)-map

$$\operatorname{clr} : \mathcal{S}^D \to \mathbb{R}^D, \quad \operatorname{clr}(\boldsymbol{x}) \coloneqq \left(\ln\left(\frac{x_1}{\sqrt[D]{\prod_{j=1}^{D} x_j}} \right), \ldots, \ln\left(\frac{x_D}{\sqrt[D]{\prod_{j=1}^{D} x_j}} \right) \right)^{\top}, \tag{2}$$

which can be shown to be distance preserving on $\mathcal{S}^D$, Pawlowsky-Glahn et al. [43]. Further, the clr-map has the following properties,

$$\operatorname{clr}(\boldsymbol{x} \oplus_{\mathcal{A}} \boldsymbol{y}) = \operatorname{clr}(\boldsymbol{x}) + \operatorname{clr}(\boldsymbol{y}), \tag{3}$$

$$\operatorname{clr}(\alpha \odot_{\mathcal{A}} \boldsymbol{x}) = \alpha \operatorname{clr}(\boldsymbol{x}), \tag{4}$$

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{A}} = \langle \operatorname{clr}(\boldsymbol{x}), \operatorname{clr}(\boldsymbol{y}) \rangle_2, \tag{5}$$

where $\langle \cdot, \cdot \rangle_2$ denotes the standard inner product in $\mathbb{R}^D$ Pawlowsky-Glahn et al. [43]. As the clr-map is not bijective onto $\mathbb{R}^D$, a modification has been considered by Egozcue et al. [15], called the ilr (isometric log-ratio)-map

$$\operatorname{ilr}_{\mathbf{V}} : \mathcal{S}^D \to \mathbb{R}^{D-1}, \quad \operatorname{ilr}_{\mathbf{V}}(\boldsymbol{x}) \coloneqq \mathbf{V}^{\top} \operatorname{clr}(\boldsymbol{x}), \tag{6}$$

where $\mathbf{V} \in \mathbb{R}^{D \times (D-1)}$ is a matrix with orthogonal columns spanning the $D-1$ dimensional subspace $\{\boldsymbol{z} \in \mathbb{R}^D \mid \sum_{j=1}^{D} z_j = 0\} \subset \mathbb{R}^D$. The ilr-map is not unique depending on the chosen basis, but is an isometric bijective map onto $\mathbb{R}^{D-1}$ that fulfills (3), (4) and (5). Thus, the purpose of a clr (ilr)-map is to transform compositional data to the standard Euclidean geometry for which classical tools in statistical data analysis are appropriate and designed. Note that the ℓ-th component of the clr-map can be written in terms of pairwise log-ratios, since

$$\ln\left(\frac{x_\ell}{\sqrt[D]{\prod_{j=1}^{D} x_j}} \right) = \frac{1}{D} \left(\ln \frac{x_\ell}{x_1} + \cdots + \ln \frac{x_\ell}{x_{\ell-1}} + \ln \frac{x_\ell}{x_{\ell+1}} + \cdots + \ln \frac{x_\ell}{x_D} \right),$$

for $\ell \in \{1, \ldots, D\}$, and thus clr (ilr)-maps consider the information of all pairwise log-ratios, and they receive equal weight in the analysis. This is not always desirable, because log-ratios of pairs which are not in a meaningful relationship might not be considered at all for the analysis.

In this paper we investigate the connections between CoDa and signal processing on graphs, and it will be shown that CoDa can be viewed as calculus on finite graphs. The goal is to extend tools of the latter by defining a scale invariant inner product that depends on the graphical structure. Subsequently, a scale invariant isometric bijective map shall be identified that depends on the graph structure and on weights, such that after transforming the original data one can work again in a Euclidean space. The main idea of the paper is thus to modify the Aitchison inner product in such a way that only certain log-ratios influence the geometry of the space. This can be done by looking at a weighted version of the Aitchison inner product (1), with weights w_{ij}:

$$\frac{1}{2} \sum_{i,j=1}^{D} \ln\left(\frac{x_i}{x_j}\right) \ln\left(\frac{y_i}{y_j}\right) w_{ij}.$$

These weights can be fixed before the analysis to only let the information of certain important log-ratios play a role, or they can be chosen in a data dependent way.

The idea of weighting in CoDa has been considered before, see Van den Boogaart et al. [7], Egozcue and Pawlowsky-Glahn [14], Hron et al. [27], Greenacre [22], Greenacre et al. [24] and Hron et al. [26]. Greenacre [22] also considers only keeping a few log-ratios which represent the data well and draws connections to graph theory. Our contribution differs, however, from the mentioned ones in that the underlying geometry is adapted through distinct w_{ij}, and that we are able after transforming the data to work in the standard Euclidean geometry.

Before introducing the new concepts in detail, we will recapture some important results from graph theory.

1.1. Some results from graph theory

In this subsection we use the notation $(f_1, \dots, f_D) \in \mathbb{R}^D$ and $(g_1, \dots, g_D) \in \mathbb{R}^D$ for two sets of variables, in order to avoid any confusion with the compositional case. We define a graph as a fixed pair $(\mathcal{V}, \mathbf{W})$, where $\mathcal{V} := \{1, \dots, D\}$ denotes a set of indices, and $\mathbf{W} = (w_{ij})_{1 \le i,j \le D} \in \mathbb{R}^{D \times D}$ is a symmetric matrix with zero diagonal and non-negative entries, corresponding to weights between indices. Graphs are useful to model the relation between variables $(f_1, \dots, f_D) \in \mathbb{R}^D$. The idea is that the bigger a weight w_{ij}, the bigger the relationship between the two variables f_i and f_j is. Whenever w_{ij} is zero there is no relation.

The edge-set of a graph is defined as $\mathcal{E} := \{(i,j) \mid w_{ij} \neq 0\} \subset \mathcal{V}^2$. We write in the following $i \sim j$ whenever $(i,j) \in \mathcal{E}$. We say that there exists a path from the vertex i to the vertex j if there are vertices $i_1, \dots, i_k$, with $i = i_1, j = i_k$, and $w_{i_1 i_2} \neq 0, w_{i_2 i_3} \neq 0, \dots, w_{i_{k-1} i_k} \neq 0$. A subset of indices $\{i_1, \dots, i_k\} \subset \mathcal{V}$ is called connected if there is a path from each vertex in the subset to another vertex in the subset. If the subset is equal to $\mathcal{V}$ then the graph is said to be connected, otherwise we say it is disconnected. The set of vertices $\mathcal{V}$ can always be written as a union of its connected components $\mathcal{V} = \cup_{m=1}^{M} \mathcal{V}_m$, with disjoint sets $\mathcal{V}_1, \dots, \mathcal{V}_M$, for any graph. Such a decomposition can be obtained by starting with one vertex, say $v_1 = 1$, and looking for all other vertices connected with v_1 through a path to obtain $\mathcal{V}_1$. Deleting $\mathcal{V}_1$ from $\mathcal{V}$ we can restart the procedure to get $\mathcal{V}_2$, and so on.

An important analytical tool for finite graphs is the so called Laplacian-matrix, defined as

$$\mathbf{L}_{\mathbf{W}} := \operatorname{diag}\left(\sum_{j=1}^{D} w_{1j}, \dots, \sum_{j=1}^{D} w_{Dj}\right) - \mathbf{W}, \tag{7}$$

where diag is the diagonal matrix of the corresponding entries. The definition of $\mathbf{L}_{\mathbf{W}}$ is motivated by the following key equality

$$\frac{1}{2} \sum_{(i,j) \in \mathcal{E}} (f_i - f_j)(g_i - g_j) w_{ij} = \boldsymbol{f}^\top \mathbf{L}_{\mathbf{W}} \boldsymbol{g}, \tag{8}$$

for any $\boldsymbol{f}, \boldsymbol{g} \in \mathbb{R}^D$, see Merris [37].

The properties of $\mathbf{L}_{\mathbf{W}}$ have been analyzed extensively. In this paper we will need the following standard results in spectral graph theory, see Mohar [38] for a proof.

Lemma 1. *Assume that* $\mathbf{W}$ *is symmetric with zero diagonal and non-negative entries, then:*

(i) $\mathbf{L}_{\mathbf{W}}$ *is a symmetric positive semi-definite matrix;*
(ii) The vector of all ones $\mathbf{1} := (1, \dots, 1)^\top \in \mathbb{R}^D$ *is always an eigenvector to zero of* $\mathbf{L}_{\mathbf{W}}$*, i.e.,* $\mathbf{L}_{\mathbf{W}} \mathbf{1} = \mathbf{0}$*;*
(iii) If $\mathcal{V}$ *is the union of more than one connected component,* $\mathcal{V} = \cup_{m=1}^{M} \mathcal{V}_m$*, then for each connected component* $\mathcal{V}_m$*, the vector in* $\mathbb{R}^D$ *with ones at position* $i \in \mathcal{V}_m$ *and otherwise zeros,* $\mathbf{1}_{i \in \mathcal{V}_m} \in \mathbb{R}^D$*, is an eigenvector to zero, i.e.,* $\mathbf{L}_{\mathbf{W}} \mathbf{1}_{i \in \mathcal{V}_m} = \mathbf{0}$*. The vectors* $\mathbf{1}_{i \in \mathcal{V}_m}$ *span the kernel of* $\mathbf{L}_{\mathbf{W}}$*;*
(iv) There exists a permutation matrix $\mathbf{P}$ *such that* $\mathbf{P}\mathbf{L}_{\mathbf{W}}\mathbf{P}^\top$ *is in block diagonal form with blocks* $\mathbf{L}_1, \dots, \mathbf{L}_M$ *for the weights* $\mathbf{P}\mathbf{W}\mathbf{P}^\top$*;*
(v) The Laplacian-matrix $\mathbf{L}_{\mathbf{W}}$ *has exactly* M *many zero eigenvalues.*

From now on we will assume for non-connected graphs that the Laplacian-matrix $\mathbf{L}_{\mathbf{W}}$ is always in block diagonal form. This can always be achieved by simply relabeling the vertices using Lemma 1.

For a more thorough introduction to graph theory we refer to Gross and Yellen [25] or Chung [9].

2. Compositional data on graphs

2.1. Graph simplex space, norms and inner products

Eq. (8) allows us to make the connection to compositional data: When taking the weight matrix $\mathbf{W} = \frac{1}{D}(\mathbf{1}\mathbf{1}^\top - \mathbf{I}_D)$, $w_{ij} = \frac{1}{D}$, where $\mathbf{I}_D$ denotes the identity matrix of dimension D, and $\mathbf{L}_{\mathcal{A}} := (1 - \frac{1}{D})\mathbf{I}_D - \frac{1}{D}(\mathbf{1}\mathbf{1}^\top - \mathbf{I}_D) = \mathbf{I}_D - \frac{1}{D}\mathbf{1}\mathbf{1}^\top$, we recover for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}^D$

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathcal{A}} = \ln(\boldsymbol{x})^\top \mathbf{L}_{\mathcal{A}} \ln(\boldsymbol{y}).$$

$\mathbf{L}_{\mathcal{A}}$ is known as the centering matrix, see Marden [35]. It has exactly one eigenvalue equal to zero, with eigenvector $\mathbf{1}$. All other eigenvalues are equal to 1. The eigenvector $\mathbf{1}$ corresponds to the null space of $\mathbf{L}_{\mathcal{A}}$ and so the scale invariance of the Aitchison inner product is a direct consequence of the Laplacian matrix $\mathbf{L}_{\mathcal{A}}$, as $\mathbf{L}_{\mathcal{A}} \ln(\boldsymbol{x} \oplus c\mathbf{1}) = \mathbf{L}_{\mathcal{A}}(\ln(\boldsymbol{x}) + \ln(c)\mathbf{1}) = \mathbf{L}_{\mathcal{A}} \ln(\boldsymbol{x})$ holds for any positive constant c. We can see that on $(\mathbb{R}_+^D, \oplus, \odot)$ the bilinear form $\ln(\boldsymbol{x})^\top \mathbf{L}_{\mathcal{A}} \ln(\boldsymbol{y})$ is not an inner product as we can only deduce from $\ln(\boldsymbol{x})^\top \mathbf{L}_{\mathcal{A}} \ln(\boldsymbol{x}) = 0$ that $\ln(\boldsymbol{x})$ is in the null space of $\mathbf{L}_{\mathcal{A}}$. Instead of considering quotient spaces and modified operations, an additional condition, such as $\sum_{j=1}^{D} x_j = 1$, is used in compositional data.

Given a graph $(\mathcal{V}, \mathbf{W})$, with a partition into connected components $\mathcal{V} = \cup_{m=1}^{M} \mathcal{V}_m$, this leads to the definition of the D-part Graph Simplex as

$$\mathcal{S}_{\mathbf{W}}^{D} := \left\{ (x_1, \dots, x_D)^\top \in \mathbb{R}_+^D \mid \sum_{j \in \mathcal{V}_m} x_j = \kappa_m, m \in \{1, \dots, M\} \right\}, \tag{9}$$

for some $\kappa_1, \dots, \kappa_M > 0$, and scaled versions of the perturbation and the powering operations such that the latter map into $\mathcal{S}_{\mathbf{W}}^{D}$:

- $(\boldsymbol{x} \oplus_{\mathbf{W}} \boldsymbol{y})_{i \in \mathcal{V}_m} := \frac{\kappa_m}{\sum_{j \in \mathcal{V}_m} x_j y_j} (x_i y_i)_{i \in \mathcal{V}_m}^\top$
- $(\alpha \odot_{\mathbf{W}} \boldsymbol{x})_{i \in \mathcal{V}_m} := \frac{\kappa_m}{\sum_{j \in \mathcal{V}_m} x_j^\alpha} (x_i^\alpha)_{i \in \mathcal{V}_m}^\top$

for all $m \in \{1, \dots, M\}$, where the subscript $i \in \mathcal{V}_m$ denotes the entries with index in $\mathcal{V}_m$. Note that in the definition of $\mathcal{S}_{\mathbf{W}}^{D}$ other conditions could be used, provided that the perturbation and the power operation is changed accordingly. The natural extension of the Aitchison inner product to a graph structure on $\ln(\boldsymbol{x})$ is to equip $\mathcal{S}_{\mathbf{W}}^{D}$ with the inner product $\ln(\boldsymbol{x})^\top \mathbf{L}_{\mathbf{W}} \ln(\boldsymbol{y})$, defining:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathbf{W}} := \langle \ln(\boldsymbol{x}), \mathbf{L}_{\mathbf{W}} \ln(\boldsymbol{y}) \rangle_2, \tag{10}$$

$$\|\boldsymbol{x}\|_{\mathbf{W}} := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\mathbf{W}}}, \tag{11}$$

for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}_+^D$. We have the following Lemma:

Lemma 2. *The space $(\mathcal{S}_{\mathbf{W}}^{D}, \oplus_{\mathbf{W}}, \odot_{\mathbf{W}})$ equipped with $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathbf{W}}$ is a Hilbert space.*

The proof can be found in Appendix A.

2.2. Negative weights

We can generalize the theory developed for the space $(\mathcal{S}_{\mathbf{W}}^{D}, \langle \cdot, \cdot \rangle_{\mathbf{W}}, \oplus_{\mathbf{W}}, \odot_{\mathbf{W}})$ even further. One important extension that comes to mind is to allow the weights w_{ij} to also take negative values as the following example shows:

Example 1. Assume that through expert knowledge we are interested in analyzing only certain weighted combinations, say $\sum_{j=1}^{D} \ln(\frac{x_i}{x_j}) w_{ij} = \ln\left(\frac{x_i^{d(i)}}{\prod_{j=1}^{D} x_j^{w_{ij}}} \right)$, where $d(i) := \sum_{j=1}^{D} w_{ij}$, for $i \in \{i_1, \dots, i_L\} \subset \{1, \dots, D\}$ and $w_{ij} \in \mathbb{R}$, not necessarily symmetric. We can collect these weighted combinations in a vector, $\widetilde{\mathbf{L}}_{\mathbf{W}} \ln(\boldsymbol{x})$, where $\widetilde{\mathbf{L}}_{\mathbf{W}} \in \mathbb{R}^{L \times D}$ is a rectangular matrix, $L \leq D$, with $(\widetilde{\mathbf{L}}_{\mathbf{W}})_{\ell i_\ell} = \sum_{j=1}^{D} w_{ij}$ for $\ell \in \{1, \dots, L\}$ and $(\widetilde{\mathbf{L}}_{\mathbf{W}})_{lj} = -w_{lj}$, for $j \neq i_l$. The matrix $\mathbf{L}_{\mathbf{W}} := \widetilde{\mathbf{L}}_{\mathbf{W}} (\widetilde{\mathbf{L}}_{\mathbf{W}})^\top$ is symmetric, positive semi-definite, has real valued entries and $\mathbf{1}$ in its null space, $\mathbf{L}_{\mathbf{W}} \mathbf{1} = \mathbf{0}$. $\widetilde{\mathbf{L}}_{\mathbf{W}}$ can be used as a building block for a scale-invariant Hilbert-space as explained below.

Dropping the assumption of non-negative weights w_{ij} but still assuming that $\mathbf{L}_{\mathbf{W}}$ is positive semi-definite, the goal is to again be able to define scale-invariant spaces. As before, the idea is to define a Hilbert-space based on a bilinear form

$$(\boldsymbol{x}, \boldsymbol{y}) \mapsto \langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\sim} := \langle \ln(\boldsymbol{x}), \mathbf{L}_{\mathbf{W}} \ln(\boldsymbol{y}) \rangle_2, \tag{12}$$

for a symmetric and positive semi-definite matrix $\mathbf{L}_{\mathbf{W}}$ with $\mathbf{L}_{\mathbf{W}} \mathbf{1} = \mathbf{0}$. Definition (12) does not define a proper inner product on $(\mathbb{R}_+^D, \oplus, \odot)$ as it is degenerated, i.e., $\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\sim} = 0$ is satisfied by infinitely many $\boldsymbol{x}$, and $\mathbb{R}_+^D$ equipped with the standard operations needs to be modified.

More formally, we start by defining the equivalence relation $\ln(\boldsymbol{x}) \sim \ln(\boldsymbol{y}) \Leftrightarrow \ln(\boldsymbol{x}) - \ln(\boldsymbol{y}) \in \mathrm{Ker}(\mathbf{L}_{\mathbf{W}})$, where Ker denotes the kernel of the linear operator $\mathbf{L}_{\mathbf{W}}$. This equivalence relation induces the equivalence classes $[\ln(\boldsymbol{x})] := \{\ln(\boldsymbol{y}) \mid \ln(\boldsymbol{x}) \sim \ln(\boldsymbol{y})\}$. In the following we write " $\equiv$ " whenever two elements belong to the same equivalence class. From the theory of quotient spaces, see Roman [46], we get that the space of equivalence classes, namely $\{[\ln(\boldsymbol{x})] \mid \boldsymbol{x} \in \mathbb{R}_+^D\}$, which can be also written as $[\ln(\boldsymbol{x})] = \{\ln(\boldsymbol{x}) + \boldsymbol{q} \mid \boldsymbol{q} \in \mathrm{Ker}(\mathbf{L}_{\mathbf{W}})\}$, is a linear vector space equipped with the operations $[\ln(\boldsymbol{x})] + [\ln(\boldsymbol{y})] := [\ln(\boldsymbol{x}) + \ln(\boldsymbol{y})]$ and $\alpha[\ln(\boldsymbol{x})] := [\alpha \ln(\boldsymbol{x})]$ for any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}_+^D$ and $\alpha \in \mathbb{R}$. Therefore to define a scale invariant Hilbert space on $\mathcal{S}_{\sim}^{D} := \{\exp([\ln(\boldsymbol{x})]) \mid \boldsymbol{x} \in \mathbb{R}_+^D\}$ we set for $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}_{\sim}^{D}$ and $\alpha \in \mathbb{R}$

$$\boldsymbol{x} \oplus_{\sim} \boldsymbol{y} := \exp([\ln(\boldsymbol{x})] + [\ln(\boldsymbol{y})]), \tag{13}$$

$$\alpha \odot_{\sim} \boldsymbol{x} := \exp(\alpha[\ln(\boldsymbol{x})]), \tag{14}$$

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\sim} := \langle [\ln(\boldsymbol{x})], \mathbf{L}_{\mathbf{W}} [\ln(\boldsymbol{y})] \rangle_2. \tag{15}$$

With this we again have

Lemma 3. *The space $(\mathcal{S}_{\sim}^{D}, \oplus_{\sim}, \odot_{\sim})$ equipped with $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\sim}$ is a Hilbert space.*

The proof can be found in Appendix A.

Note that in the case of compositional data, $\mathbf{L_W} = \mathbf{L}_\mathcal{A}$, (15) is equivalent to the Aitchison inner product (1) and (13) as well as (14) is proportional to the scaled versions of perturbation and powering in the former. By definition of the quotient space any two elements are the same if their difference lies in the kernel. For compositional data this means $\ln(\boldsymbol{x}) \equiv \ln(\boldsymbol{x}) + \alpha\mathbf{1}$ for any $\alpha \in \mathbb{R}$. The usual condition $\sum_{j=1}^D x_j = 1$ as used in the definition of $\mathcal{S}^D$, is simply a matter of fixing a representative for each equivalence class $[\ln(\boldsymbol{x})]$, but others could be chosen as well. The same goes for the graphical extension represented so far in this paper.

The quotient space nature of regular compositional data has been investigated extensively in Barceló-Vidal et al. [4]. The graphical approach presented here is thus a natural extension. Note, however, that if we allow also for negative weights then, depending on the kernel of the Laplacian, we might allow for more than invariance under rescaling on subgraphs, as the latter might contain more than the constant vectors $\mathbf{1}_{\mathcal{V}_m}$.

We finish this subsection by noting that for any matrix $\mathbf{L_W}$ that is symmetric and fulfills $\mathbf{L_W 1} = \mathbf{0}$ we have the following identity

$$\langle \ln(\boldsymbol{x}), \mathbf{L_W} \ln(\boldsymbol{y})\rangle_2 = \frac{1}{2}\sum_{i,j=1}^D \ln\left(\frac{x_i}{x_j}\right)\ln\left(\frac{y_i}{y_j}\right)w_{ij}, \tag{16}$$

where $\mathbf{W}$ is defined through the decomposition $\mathbf{L_W} = \operatorname{diag}(\mathbf{W1}) - \mathbf{W}$. The proof is the same as the proof for (8).

2.3. Compositional normal distribution and total variance

In classical CoDa $\boldsymbol{x}$ is assumed to be normally distributed if and only if for some fixed $\mathbf{V}$ the ilr-transformed data follows a multivariate normal, i.e., $\operatorname{ilr}_\mathbf{V}(\boldsymbol{x}) \sim \mathcal{N}(\boldsymbol{\eta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} \in \mathbb{R}^{(D-1)\times(D-1)}$ is a positive definite matrix and $\boldsymbol{\eta} \in \mathbb{R}^{D-1}$, see Pawlowsky-Glahn et al. [42]. $\operatorname{ilr}_\mathbf{V}(\boldsymbol{x})$ can be written as $\mathbf{V}^\top \mathbf{L}_\mathcal{A} \ln(\boldsymbol{x})$, and so we conclude that $\ln(\boldsymbol{x})$ follows a degenerate multivariate normal distribution with covariance matrix $(\mathbf{L}_\mathcal{A}\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{V}^\top\mathbf{L}_\mathcal{A})^+$ and mean $\mathbf{V}\boldsymbol{\eta}$, $\ln(\boldsymbol{x}) \sim \mathcal{N}(\mathbf{V}\boldsymbol{\eta}, (\mathbf{L}_\mathcal{A}\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{V}^\top\mathbf{L}_\mathcal{A})^+)$, see also the next section, where the superscript $+$ indicates the Moore–Penrose inverse, Ben-Israel and Greville [5]. The matrix $\mathbf{L} := \mathbf{L}_\mathcal{A}\mathbf{V}\boldsymbol{\Sigma}^{-1}\mathbf{V}^\top\mathbf{L}_\mathcal{A}$ is symmetric and positive semi-definite. Furthermore, $\mathbf{1}$ is in its nullspace. Any symmetric matrix $\mathbf{L}$ with $\mathbf{L1} = \mathbf{0}$ can be decomposed into $\mathbf{L} = \operatorname{diag}(\mathbf{W1}) - \mathbf{W}$ for a symmetric matrix $\mathbf{W}$, with possibly negative entries w_{ij}, for $i \neq j$, and zero diagonal. Therefore we define the graph compositional normal distribution more generally:

Definition 1. A random vector $\boldsymbol{x} \in \mathcal{S}^D_\sim$ is said to be graph compositionally normal distributed, with $\mathbb{E}(\ln(\boldsymbol{x})) = \boldsymbol{\mu}$, if and only if

$$\ln(\boldsymbol{x}) \sim \mathcal{N}(\boldsymbol{\mu}, (\mathbf{L_W})^+),$$

where $\mathbf{L_W}$ is symmetric, positive semi-definite, has real valued entries and fulfills $\mathbf{L_W 1} = \mathbf{0}$.

Note that this definition includes the classical CoDa case, but also allows for more general $\mathbf{L_W}$ as seen in the previous chapters.

The total variance of a classical compositional random vector $\boldsymbol{x} \in \mathcal{S}^D$, with $\mathbb{E}(\ln(\boldsymbol{x})) = \boldsymbol{\mu}$, denoted by $\operatorname{totvar}(\boldsymbol{x})$, is an appropriate measure of global dispersion, see Hron and Kubáček [29], and defined as

$$\operatorname{totvar}(\boldsymbol{x}) := \frac{1}{2D}\sum_{i,j=1}^D \mathbb{E}\left(\left(\ln(x_i^\mu) - \ln(x_j^\mu)\right)^2\right),$$

where $\boldsymbol{x}^\mu$ denotes the centered variable $\boldsymbol{x}^\mu := \exp(\ln(\boldsymbol{x}) - \boldsymbol{\mu})$.

We define the graph version of the total variance in an analogous way.

Definition 2. For a random vector $\boldsymbol{x} \in \mathcal{S}^D_\sim$, with $\mathbb{E}(\ln(\boldsymbol{x})) = \boldsymbol{\mu}$, we define the graph total variance

$$\operatorname{totvar}_{\mathbf{L_W}}(\boldsymbol{x}) := \mathbb{E}(\langle \ln(\boldsymbol{x}), \mathbf{L_W}\ln(\boldsymbol{y})\rangle_2) = \frac{1}{2}\sum_{i,j=1}^D \mathbb{E}\left(\left(\ln(x_i^\mu) - \ln(x_j^\mu)\right)^2\right)w_{ij}, \tag{17}$$

where the second equality is a consequence of (16).

2.4. Centered and isometric log-ratio transforms

In the classical compositional setting, where $\mathbf{L_W} = \mathbf{L}_\mathcal{A}$, the clr and the $\operatorname{ilr}_\mathbf{V}$ mappings play an important role in establishing isometry of $\mathcal{S}^D$ to $\mathbb{R}^{D-1}$, as well as in the interpretation of results. In the notation of the Laplacian-matrix the clr map in classical CoDa is given by

$$\operatorname{clr}(\boldsymbol{x}) = \mathbf{L}_\mathcal{A}\ln(\boldsymbol{x}).$$

It is well known that $\mathbf{L}_{\mathcal{A}}$ is idempotent, $\mathbf{L}_{\mathcal{A}} = \mathbf{L}_{\mathcal{A}}^2$, from which follows that clr is distance preserving, see (5). This motivates us to define the weighted clr map as

$$\operatorname{clr}(\boldsymbol{x})_{\mathbf{W}} := \mathbf{P}_{\mathbf{W}} \ln(\boldsymbol{x}),$$

for any matrix $\mathbf{P}_{\mathbf{W}}$ that fulfills $\mathbf{P}_{\mathbf{W}}^\top \mathbf{P}_{\mathbf{W}} = \mathbf{L}_{\mathbf{W}}$. From the definition of $\operatorname{clr}(\cdot)_{\mathbf{W}}$ we immediately have $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\mathbf{W}} = \langle \mathbf{L}_{\mathbf{W}} \ln(\boldsymbol{x}), \mathbf{L}_{\mathbf{W}} \ln(\boldsymbol{y}) \rangle_2$ for any $\boldsymbol{x}, \boldsymbol{y}$. In the case of non-negative as well as negative weights, as described in Sections 2.1 and 2.2, the matrix $\mathbf{P}_{\mathbf{W}} = \mathbf{L}_{\mathbf{W}}^{\frac{1}{2}}$ is a natural choice. However, more interpretable choices might be appropriate depending on the setting. In Example 1 and the subsequent discussion we saw that any rectangular matrix $\widetilde{\mathbf{L}}_{\mathbf{W}}$ with the property $\widetilde{\mathbf{L}}_{\mathbf{W}} \mathbf{1} = \mathbf{0}$ can be used as a building block to define a Hilbert space starting from $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_\sim := \langle \ln(\boldsymbol{x}), \mathbf{L}_{\mathbf{W}} \ln(\boldsymbol{y}) \rangle_2$ with $\mathbf{L}_{\mathbf{W}} = (\widetilde{\mathbf{L}}_{\mathbf{W}})^\top \widetilde{\mathbf{L}}_{\mathbf{W}}$. In this example $\mathbf{P}_{\mathbf{W}} = \widetilde{\mathbf{L}}_{\mathbf{W}}$ seems to be the more appropriate choice as the coordinates of $\mathbf{P}_{\mathbf{W}} \ln(\boldsymbol{x})$ amount to $\sum_{j=1}^D \ln(\frac{x_i}{x_j}) w_{ij}$.

Although $\operatorname{clr}(\cdot)_{\mathbf{W}}$ is distance preserving it is not bijective. In the classical CoDa approach the ilr map (6) is often used to map data into the Euclidean space $\mathbb{R}^{D-1}$ where common statistical methods can be used. Additionally, depending on the choice of $\mathbf{V}$ one obtains very interpretable maps, see Filzmoser et al. [16] or Fišerová and Hron [17], that are of the form

$$(\operatorname{ilr}(\boldsymbol{x})_{\mathbf{V}})_j := \sqrt{\frac{D-j}{D-j+1}} \ln\Big(\frac{x_j}{\sqrt[D-j]{\prod_{i=j+1}^D x_i}} \Big). \tag{18}$$

Here the jth coordinate $(\operatorname{ilr}(\boldsymbol{x})_{\mathbf{V}})_j$ only incorporates information of x_i with $i \geq j$ and x_1 only appears in the first coordinate, which can thus be interpreted as carrying the main information about the latter. As the order of variables is arbitrary, simply permuting the variables leads to maps with first coordinate only containing any chosen variable. In the graph setting we would also like to obtain interpretable isometries. The idea is to use a modified version of the Cholesky decomposition for positive semi-definite matrices to write $\mathbf{L}_{\mathcal{A}} = \mathbf{C}^\top \mathbf{C}$, with $\mathbf{C}$ being an upper triangular matrix with non-negative diagonal elements. The jth entry of the mapping $\boldsymbol{x} \mapsto \mathbf{C} \ln(\boldsymbol{x})$ contains only the information of $\ln(x_j), \dots, \ln(x_D)$ as a weighted sum similar to (18).

Without loss of generality we will assume in the following for ease of notation that the graph leading to $\mathbf{L}_{\mathbf{W}}$ is connected. This is no restriction as the two Lemmata which follow can otherwise simply be applied to each subgraph separately.

Lemma 4 (*Graph Isometric Log-ratio Map (GILR1)*). *Let $\mathbf{P}$ be a permutation matrix of the nodes $\mathcal{V}$. Computing the eigen-decomposition of $\mathbf{P}\mathbf{L}_{\mathbf{W}}\mathbf{P}^\top$, that is $\mathbf{P}\mathbf{L}_{\mathbf{W}}\mathbf{P}^\top = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^\top$, followed by a QR decomposition $\mathbf{U}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{U}^\top = \mathbf{Q}\mathbf{R}$, with $\mathbf{Q}, \mathbf{R} \in \mathbb{R}^{D \times D}$, we set $\mathbf{C} = \mathbf{R}_{-\{\ell_1,\dots,\ell_k\}}$, where $\mathbf{R}_{-\{\ell_1,\dots,\ell_k\}}$ is the matrix $\mathbf{R}$ with zero rows, given by the indices $\ell_1, \dots, \ell_M$, deleted. Then the map*

$$\boldsymbol{x} \mapsto \mathbf{C}\mathbf{P} \ln(\boldsymbol{x}), \tag{19}$$

is a linear bijective isometry from $(\mathcal{S}_\sim^D, \oplus_\sim, \odot_\sim)$, equipped with $\langle \cdot, \cdot \rangle_\sim$, to $(\mathbb{R}^{D-M}, +, \cdot)$ equipped with $\langle \cdot, \cdot \rangle_2$. Additionally the matrix $\mathbf{C}$ is an upper triangular matrix that fulfills $\mathbf{C}\mathbf{1} = \mathbf{0}$.

A proof to this Lemma can be found in Appendix A. The role of the permutation matrix in Lemma 4 is to be able to put focus on a coordinate of interest similar to the ilr map (18). After the choice of a permutation matrix $\mathbf{P}$, which can be identified with the mapping $\pi : \{1, \dots, D\} \to \{1, \dots, D\}$, we see that for (19), the first element is $\ln\Big(\frac{x_{\pi(1)}^{c_{11}}}{\prod_{i=2}^D x_{\pi(i)}^{-c_{1i}}}\Big) = \ln\Big(\prod_{i=2}^D \big(\frac{x_{\pi(1)}}{x_{\pi(i)}}\big)^{-c_{1i}}\Big)$, the second element $\ln\Big(\frac{x_{\pi(2)}^{c_{22}}}{\prod_{i=3}^D x_{\pi(i)}^{-c_{2i}}}\Big) = \ln\Big(\prod_{i=3}^D \big(\frac{x_{\pi(2)}}{x_{\pi(i)}}\big)^{-c_{1i}}\Big)$ and so on; where the equalities hold because from $\mathbf{C}\mathbf{1} = 0$ follows $c_{ii} = -\sum_{j=i+1}^D c_{ij}$.

Although the map in (19) is a bijective isometry, the rows of $\mathbf{C}\mathbf{P}$ are not orthogonal. A linear bijective isometry with orthogonal rows can however be constructed in the following way.

Lemma 5 (*Graph Isometric Log-ratio Map 2 (GILR2)*). *Taking the eigen-decomposition of $\mathbf{L}_{\mathbf{W}}$, $\mathbf{L}_{\mathbf{W}} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^\top$, where the diagonal elements $\lambda_1, \dots, \lambda_D$ of $\boldsymbol{\Sigma}$ are ordered from biggest to smallest and denoting the ith column of $\mathbf{U}$ by $\boldsymbol{u}_i$ and M the number of zero eigenvalues, we can define a bijective isometric map by*

$$\boldsymbol{x} \mapsto (\sqrt{\lambda_i} \langle \boldsymbol{u}_i, \ln(\boldsymbol{x}) \rangle_2)_{i \in 1,\dots,D-M}^\top \in \mathbb{R}^{D-M}. \tag{20}$$

Its inverse is given by $\boldsymbol{z} \mapsto \sum_{i=1}^{D-M} z_i \frac{1}{\sqrt{\lambda_i}} \boldsymbol{u}_i$.

A proof can be found in Appendix A.

3. The weights w_{ij}

In many cases there will not be any expert knowledge about the weights w_{ij}. In this section we will look at different models to find such weights. In the context of graphical models, it has been shown that if a random vector $\boldsymbol{x}$ follows a

multivariate Gaussian, $\boldsymbol{x} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, with positive definite $\boldsymbol{\Sigma}$, that the precision matrix $\boldsymbol{\Sigma}^{-1}$ reveals a graph structure in form of conditional independence [32]. For ease of notation we assume in the following that $\boldsymbol{x}$ is replaced with its centered version $\boldsymbol{x} - \boldsymbol{\mu}$. In Yuan and Lin [48] and Friedman et al. [18], a penalized log-likelihood problem is solved to find an estimate $\hat{\boldsymbol{\Sigma}}^{-1}$

$$\hat{\boldsymbol{\Sigma}}^{-1} := \underset{\substack{\mathbf{A} \in \mathbb{R}^{d \times d} \\ \mathbf{A}^\top = \mathbf{A}, \mathbf{A} \, p.d.}}{\arg\min} \ln(|\, \mathbf{A} \,|) - \operatorname{tr}\left(\mathbf{A}(N^{-1}\mathbf{X}^\top \mathbf{X})\right) + \lambda \sum_{i,j=1}^{D} |\, A_{ij} \,|, \tag{21}$$

where p.d. is short for positive definite, $\mathbf{X} \in \mathbb{R}^{N \times D}$ is the data matrix and λ is a parameter controlling the sparsity of the entries of A_{ij} which give the edge weights. Another common approach to finding a graph structure and corresponding weights, besides (21), is to solve a series of problems of the form, see Meinshausen and Bühlmann [36]:

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^{D-1}} \frac{1}{N} \left\| \boldsymbol{X}_i - \mathbf{X}^{-i} \boldsymbol{\beta}_i \right\|_2^2 + \lambda \sum_{j=1}^{D-1} |\, (\boldsymbol{\beta}_i)_j \,|, \tag{22}$$

for $i \in \{1, \ldots, D\}$ and $\lambda \geq 0$, where $\boldsymbol{X}_i$ denotes the ith column of the data matrix $\mathbf{X}$ and $\mathbf{X}^{-i}$ the latter with the ith column deleted. To obtain a weight matrix a post-processing step is applied by setting each weight to $\tilde{w}_{ij} := |\, \frac{1}{2}(\beta_{ij} + \beta_{ji}) \,|$ for $i \neq j$ and $\tilde{w}_{ii} := 0$ for $i = j$. For compositional data an approach for finding an underlying graph structure has been proposed in Kurtz et al. [31] and is based on solving (21) or (22) after clr transforming the data. In the area of graph signal processing various other methods for finding meaningful weights have been developed in a non-compositional context, see Dong et al. [11], Kalofolias [30] and Egilmez et al. [13]. The weights are found such that a given signal, is considered smooth for the found weights, whereas smoothness is measured by $\sum_{i,j=1}^{D}(x_i - x_j)^2 w_{ij}$. Finally, in [22] an algorithm for finding important log-ratios was explored. However, there was no connection made to weights that could be used as building blocks for a geometry.

All the mentioned methods have in common that they lead to bigger weights between two variables $\ln(x_i)$ and $\ln(x_j)$ the bigger their dependence is. Depending on the goals in mind and the specific setting, it might be more interesting to find weights that put a larger emphasis on log-ratios that explain a lot of variance, as we believe that those are of major interest. Log-ratios which have low variance should get small weights. Note that this is, in its core, similar to principal component analysis. We propose to maximize the weighted total variance as defined in (17), $\operatorname{totvar}_{\mathbf{L_W}}(\cdot)$, with an additional sparsity penalty and under convex constraints. Defining the following sets

$$\Omega^{\text{sym}} := \{\mathbf{L_W} = \operatorname{diag}(\mathbf{W1}) - \mathbf{W} \mid \operatorname{diag}(\mathbf{W}) = \mathbf{0},\ \mathbf{L_W} = \mathbf{L_W}^\top\}, \tag{23}$$

$$\Omega^{\text{pos}} := \{\mathbf{L_W} = \operatorname{diag}(\mathbf{W1}) - \mathbf{W} \mid \operatorname{diag}(\mathbf{W}) = \mathbf{0},\ w_{ij} \geq 0\ \forall i \neq j\}, \tag{24}$$

$$\Omega^{2} := \{\mathbf{L_W} = \operatorname{diag}(\mathbf{W1}) - \mathbf{W} \mid \operatorname{diag}(\mathbf{W}) = \mathbf{0},\ \|\mathbf{L_W}\|_2 \leq 1\}, \tag{25}$$

we suggest to solve either

$$\max_{\mathbf{L_W} \in \mathbb{R}^{D \times D}} \frac{1}{N} \sum_{n=1}^{N} \langle \ln(\boldsymbol{x}_n), \mathbf{L_W} \ln(\boldsymbol{x}_n) \rangle_2 - \lambda \sum_{i,j=1}^{D} |\, w_{ij} \,| \ \text{ s.t. } \mathbf{L_W} \in \Omega^2 \cup \Omega \tag{26}$$

or

$$\max_{\mathbf{L_W} \in \mathbb{R}^{D \times D}} \frac{1}{N} \sum_{n=1}^{N} \langle \ln(\boldsymbol{x}_n), \mathbf{L_W}^\top \mathbf{L_W} \ln(\boldsymbol{x}_n) \rangle_2 - \lambda \sum_{i,j=1}^{D} |\, w_{ij} \,| \ \text{ s.t. } \mathbf{L_W} \in \Omega^2 \cup \Omega, \tag{27}$$

where $\lambda > 0$ is a tuning parameter that controls the sparsity and Ω is $\Omega^{\text{sym}} \cup \Omega^{\text{pos}}$ for the first problem (26) and either empty or a combination of Ω^{sym} and Ω^{pos} for the second problem (27). The constraint $\|\mathbf{L_W}\|_2 \leq 1$ is always present and avoids graphs that have nodes with too many edges similar than done in Dong et al. [11] or Kalofolias [30]. Both problems correspond to the models described in Section 2.1 resp. 2.2 and depend on the modeling choice. If one is rather interested in the natural choice of the weighted clr as described in Section 2.4 then problem (27) seems to be more appropriate. Note that problem (27) is not convex and thus harder to solve than (26). To solve both problems we first define a matrix $\mathbf{S}$ and a vector $\boldsymbol{w}$ such that $\operatorname{vec}(\mathbf{L_W}) = \mathbf{S}\boldsymbol{w}$, where vec is the vectorization of a matrix, i.e., each column of a matrix is stacked into one big vector. Then it is easy to see that the sample weighted total variance in both cases can be written as $\langle \operatorname{vec}(\boldsymbol{\Gamma}), \mathbf{S}\boldsymbol{w} \rangle_2$ with $\boldsymbol{\Gamma} = \sum_{n=1}^{N} \ln(\boldsymbol{x}_n) \ln(\boldsymbol{x}_n)^\top$ in the first case and $\boldsymbol{\Gamma} = \mathbf{L_W}(\sum_{n=1}^{N} \ln(\boldsymbol{x}_n) \ln(\boldsymbol{x}_n)^\top)$ in the second. In the second case the strategy is thus to fix one $\mathbf{L_W}$ in the $\mathbf{L_W}\mathbf{L_W}^\top$ term of the objective function and optimize in the other, similar to ideas as described in sparse principal component analysis in Witten et al. [47]. Therefore we need to solve problems of the type

$$\max_{\boldsymbol{w} \in \mathbb{R}^{D(D-1)}} \langle \frac{1}{N} \operatorname{vec}(\boldsymbol{\Gamma}), \mathbf{S}\boldsymbol{w} \rangle_2 - \lambda \|\boldsymbol{w}\|_1 \ \text{ s.t. } \ \|\mathbf{S}\boldsymbol{w}\|_2 \leq 1 \text{ and } \boldsymbol{w} \in \Phi, \tag{28}$$

where Φ is a set of linear (in)equalities corresponding to Ω^{sym} and Ω^{pos}. To solve (28) we decouple the constraint $\|\mathbf{S}\boldsymbol{w}\|_2 \leq 1$ by rewriting it as

$$\min_{\boldsymbol{w}\in\mathbb{R}^{D(D-1)}} -\langle \frac{1}{N}\operatorname{vec}(\boldsymbol{\Gamma}), \mathbf{S}\boldsymbol{w}\rangle_2 + \lambda\|\boldsymbol{w}\|_1 + \chi(\boldsymbol{w}\in\Phi) + \chi(\|\boldsymbol{\eta}\|_2 \leq 1), \tag{29}$$

$$\mathbf{S}\boldsymbol{w} = \boldsymbol{\eta} \tag{30}$$

where χ is a function that assigns $+\infty$ if the condition is not satisfied and zero else. As the objective function of problem (29) is a sum of two separate functions in $\boldsymbol{w}$ and $\boldsymbol{\eta}$ with linear constraints it is amendable to the ADMM (alternating direction method of multipliers) algorithm, see Boyd et al. [8]. The ADMM algorithm applied to this problem has the following substeps, Algorithm 1, that are executed till convergence.

Algorithm 1

- **Input**: $\operatorname{vec}(\Gamma)$ and $\lambda > 0$
- **Initialize:** $\boldsymbol{u}, \boldsymbol{\eta} \in \mathbb{R}^{D^2}$ and $\rho > 0$

1: **while** not converged **do**
2: $\quad \boldsymbol{w} \leftarrow \operatorname{argmin}_{\boldsymbol{w}\in\Phi} -\frac{1}{N}\langle\operatorname{vec}(\Gamma), \mathbf{S}\boldsymbol{w}\rangle_2 + \lambda\|\boldsymbol{w}\|_1 + \frac{\rho}{2}\|\boldsymbol{\eta}+\boldsymbol{u}-\mathbf{S}\boldsymbol{w}\|_2^2$
3: $\quad \boldsymbol{\eta} \leftarrow \operatorname{argmin}_{\boldsymbol{\eta}} \chi(\|\boldsymbol{\eta}\|_2 \leq 1) + \frac{\rho}{2}\|\boldsymbol{\eta}+\boldsymbol{u}-\mathbf{S}\boldsymbol{w}\|_2^2$
4: $\quad \boldsymbol{u} \leftarrow \boldsymbol{u}+\boldsymbol{\eta}-\mathbf{S}\boldsymbol{w}$
5: **end while**

Solving the problem in step 2 in Algorithm 1 amounts to solving a Lasso problem, as the objective function can be reformulated, modulo a constant not dependent on $\boldsymbol{w}$, as $\frac{1}{2N}\left\|\boldsymbol{\eta}+\boldsymbol{u}+\frac{1}{\rho}\operatorname{vec}(\boldsymbol{\Gamma})-\mathbf{S}\boldsymbol{w}\right\|_2^2 + \frac{\lambda}{\rho}\|\boldsymbol{w}\|_1$. Possible positivity constraints in Φ can be handled easily by state of the art algorithms as coordinated descent, see Friedman et al. [19], and a possible symmetric constraint corresponds to linearly reparametrizing $\boldsymbol{w}$ by a quantity of half its dimension. Note also that the matrix $\mathbf{S}$ is extremely sparse and known beforehand. Solving the problem in step 3 in Algorithm 1 simply amounts to a projection of $\boldsymbol{u}-\mathbf{S}\boldsymbol{w}$ onto the unit ball. The convergence of the ADMM algorithm is highly dependent on the choice of ρ. From the above reformulation into a Lasso problem in step 2 of Algorithm 1 the appearance of ρ in a quotient with λ suggests setting $\rho = \lambda$, which we found to be a good choice. We also see that the norm in step 3 could be replaced by any other norm leading to a projection problem. As mentioned, Algorithm 1 can be used to directly solve (26) using $\boldsymbol{\Gamma} = \sum_{n=1}^{N}\ln(\boldsymbol{x}_n)\ln(\boldsymbol{x}_n)^\top$ and a $\lambda > 0$ as input. For solving Problem (27) we can use Algorithm 2.

Algorithm 2

- **Input**: $\Sigma = \sum_{n=1}^{N}\ln(\boldsymbol{x}_n)\ln(\boldsymbol{x}_n)^\top$ and $\lambda > 0$.
- **Initialize**: $w_{ii} = 0$ and $w_{ij} = \frac{1}{D}$ for $i \neq j$. Set $\mathbf{L}_{\mathbf{W}} = \operatorname{diag}(\mathbf{W1}) - \mathbf{W}$.

1: **while** not converged **do**
2: $\quad$ Update $\Gamma = \mathbf{L}_{\mathbf{W}}\Sigma$
3: $\quad$ Get $\boldsymbol{w}$ by using Algorithm 1 with input (Γ, λ)
4: $\quad$ Update $\mathbf{L}_{\mathbf{W}} = \operatorname{vec}^{-1}(\mathbf{S}\boldsymbol{w})$
5: **end while**

An implementation in Fortran and R of Algorithms 1 and 2 as well as for the maps of Lemmas 4 and 5 for different structural constraints of $\mathbf{L}_{\mathbf{W}}$ is available at https://github.com/Kristats/Graph-CoDa-Public.git.

We conducted a small simulation study to examine the suggested method. For different dimensions $D \in \{5, 10, 20, 30\}$, and number of samples $N = 5D$ we first generated a variation matrix $\mathbf{T} \in \mathbb{R}^{D\times D}$. To do so we pick at random about thirty percent of the rows, and as $\mathbf{T}$ is symmetric the same columns, and set the entries of each to a separate uniformly distributed number between $[1, \frac{1}{D}]$. The rest of the rows and columns are set to some uniformly distributed numbers between $[0, \frac{1}{D^2}]$. To obtain an appropriate covariance matrix $\boldsymbol{\Sigma}^{\text{clr}}$ in clr-coordinates we use the relation $\boldsymbol{\Sigma}^{\text{clr}} = -\frac{1}{2}\mathbf{L}_{\mathcal{A}}\mathbf{T}\mathbf{L}_{\mathcal{A}}$, see Aitchison [2]. A data matrix $\mathbf{X} \in \mathbb{R}^{N\times D}$ is then generated in clr-coordinates according to a multivariate normal distribution with mean sampled at random from a univariate standard normal and the singular covariance $\boldsymbol{\Sigma}^{\text{clr}}$. This process guarantees that the simulated data has some high variation log-ratios and many low variation ones. Each simulation setting was performed 100 times. Table 1 shows the median computation times, in seconds, and the mean of the best possible F-score, concerning the high variation ratios, that can be found in the path of lambdas on average. In Table 1 sym stands for problem (26) with $\Omega = \Omega^{\text{sym}} \cup \Omega^{\text{pos}}$, nsym for problem (27) with empty Ω, symp as well as nsymp for the parallelized versions in lambda with ten cores and step for the algorithm used in Lin et al. [33] and implemented in the easyCODA R package [23]. We can see that the F-scores are very similar for the first four methods (sym,nsym,symp,nsymp) and constantly beat the step algorithm by a large margin. The computation time grows very quickly for the non-parallelized algorithms (sym,nsym) and becomes unfeasible with growing dimension. However, we can see that the parallelized versions with ten cores lead to a significant speed up. As the F-scores for the parallelized method of problem (26) are very similar to the other ones, with the step algorithm excluded, with much lower computation times, we can conclude that (symp) should be the method of choice in this setting.

Table 1
Mean F-scores and median computation time in seconds of the algorithms presented in this paper compared with the step algorithm as implemented in the easyCODA R package [23].

	$D = 5$		$D = 10$		$D = 20$		$D = 30$	
	F-score	Time	F-score	Time	F-score	Time	F-score	Time
sym	0.83	0.99 s	0.69	10.85 s	0.66	93.97 s	0.68	321.16 s
nsym	0.84	0.75 s	0.72	14.75 s	0.69	203.44 s	0.68	909.76 s
symp	0.83	5.00 s	0.69	5.80 s	0.66	18.11 s	0.67	66.49 s
nsymp	0.84	5.03 s	0.72	6.00 s	0.69	25.89 s	0.67	127.87 s
step	0.46	0.04 s	0.27	0.4 s	0.13	6.86 s	0.09	46.29 s

4. Nugent data set

To illustrate the proposed framework we look at a dataset consisting of vaginal bacterial communities data, i.e., microbiome data, which is best treated as compositional, see Gloor et al. [20] or Lubbe et al. [34]. This data set consists of $N = 388$ samples of 16 s ribosomal RNA sequences recorded as $D = 84$ different operational taxonomic units (OTUs), see Ravel et al. [45]. As a preprocessing step only OTUs present in at least 5% of the samples were kept and subsequently the 82% of zeros of the remaining resulting data matrix $\mathbf{X} \in \mathbb{R}_+^{N\times D}$ were replaced by a value of 0.5, compare to Lubbe et al. [34]. Additionally to $\mathbf{X}$, a response $\boldsymbol{y} \in \{0, \dots, 10\}^D$ of 11 categories was recorded indicating the risk of bacterial vaginosis, from low risk, given by 0, to high risk, given by 10. We scale $\boldsymbol{y}$ and center column-wise the log-transformed data matrix $\ln(\mathbf{X})$. Following this, we solve problem (27) by using algorithm 2 for two different settings, once with $\Omega = \Omega^{\text{sym}} \cup \Omega^{\text{pos}}$ and once with $\Omega = \Omega^{pos}$, to obtain for a range of different λ solutions $\mathbf{L}_{\mathbf{W}}(\lambda)$. In each setting we propose to continue with $\mathbf{L}_{\mathbf{W}}(\lambda_{min})$ where λ_{min} is chosen such that the following BIC type criterion is minimized over the whole range of possible λ

$$BIC(\lambda) = -\,\mathrm{R}^2(\lambda) + \mathrm{df}(\lambda)\frac{\ln(N)}{N},$$

with $\mathrm{R}^2(\lambda)$ denoting the R-squared coefficient regressing $\ln(\boldsymbol{x}_n)$ onto $\mathbf{L}_{\mathbf{W}}(\lambda)\ln(\boldsymbol{x}_n)$, $n \in \{1, \dots, N\}$, and $\mathrm{df}(\lambda)$ denoting the number of non-zero weights of $\mathbf{L}_{\mathbf{W}}(\lambda)$, similar to ideas in the robust and sparse principal component setting [10]. Fig. 1 shows the number of non-zero weights, $\mathrm{df}(\lambda)$, plotted against the explained information, $\mathrm{R}^2(\lambda)$, of the solutions for both model choices, $\Omega = \Omega^{\text{sym}} \cup \Omega^{\text{pos}}$ in black and $\Omega = \Omega^{pos}$ in blue, for all λ (dots) as well as for the λ_{min} (vertical lines). Both models show the typical behavior that with increasing non-zero weights more information of the data is retained. The non-symmetric solutions always explain more information (blue dots) and the final solution $\mathbf{L}_{\mathbf{W}}(\lambda_{min})$ retains more information with only a very slight increase of non-zero weights (vertical lines).

In Fig. 2 we visualize $\mathbf{L}_{\mathbf{W}}(\lambda_{min})$ for the $\Omega = \Omega^{pos}$ case only by a directed graph where the nodes without any connections are not shown. The line width of an arrow leaving a node i and pointing to a node j is proportional to the weight w_{ij}. We can see that Lactobacillus crispatus as well as Lactobacillus iners are central nodes with many high ingoing and outgoing edges. Thus they seem to play a central role in explaining much of the information in the data. This is in accordance with Ravel et al. [45]. Furthermore, Atopobium vaginae and Prevotella timonensis seem to be also central phylotypes to explaining these data. In Fig. 3 which shows the diagonal elements of $\mathbf{L}_{\mathbf{W}}(\lambda_{min})$, i.e., the sum of outgoing edge weights $\sum_{j=1}^{D} w_{ij}$, these phylotypes are also clearly the most influential ones.

Next, we randomly split the data set a hundred times into a training and a test data set (80%–20%) and fit a compositional linear model and a linear model in the weighted case

$$\min_{\boldsymbol{a}} \sum_{n=1}^{N} (y_n - \langle \ln(\boldsymbol{a}), \mathbf{L}_{\mathbf{W}}^{\top}\mathbf{L}_{\mathbf{W}} \ln(\boldsymbol{x}_n)\rangle_2)^2, \tag{31}$$

on the training set, where we dropped the λ_{min} in the notation. To solve (31) we map the covariate data $\mathbf{X}$ using the GILR2 map (20) and then use the standard linear model in the Euclidean space. Fig. 4 shows the boxplots for the mean squared error on the test set for these three different models. We can see that the classical compositional model performs worse than the weighted cases. Again the weighted model with only positivity constraints performs best. Interpreting linear models in the classical compositional context is in general harder due to the constrained nature of the data. However, the so called pivot coordinates (18) can be used to interpret resulting coefficients, see Hron et al. [28] or Filzmoser et al. [16]. In the graph setting we can analogously use the map (19). Choosing a permutation matrix $\mathbf{P}$ that permutes the covariate vector such that any chosen x_j is on the first place, i.e., $(\mathbf{P}\ln(\boldsymbol{x}))_1 = \ln(x_j)$, allows to interpret the first compositional part $(\mathbf{CP}\ln(\boldsymbol{x}))_1$ as a quantity that carries only information about the jth coordinate, as mentioned in Section 2.4. Multiplying each coefficient, $(\mathbf{CP}\ln(\boldsymbol{a}))_1$, by the standard deviation of former permits us to interpret the size as the effect a covariate has on the outcome. Fig. 5 shows these scaled coefficients. We can see that Lactobacillus crispatus and Atopobium vaginae as well as Prevotella timonensis and Lactobacillus jensenii contribute most to a lower risk of bacterial vaginosis. Higher risk can mostly be attributed to Lactobacillus gasseri.

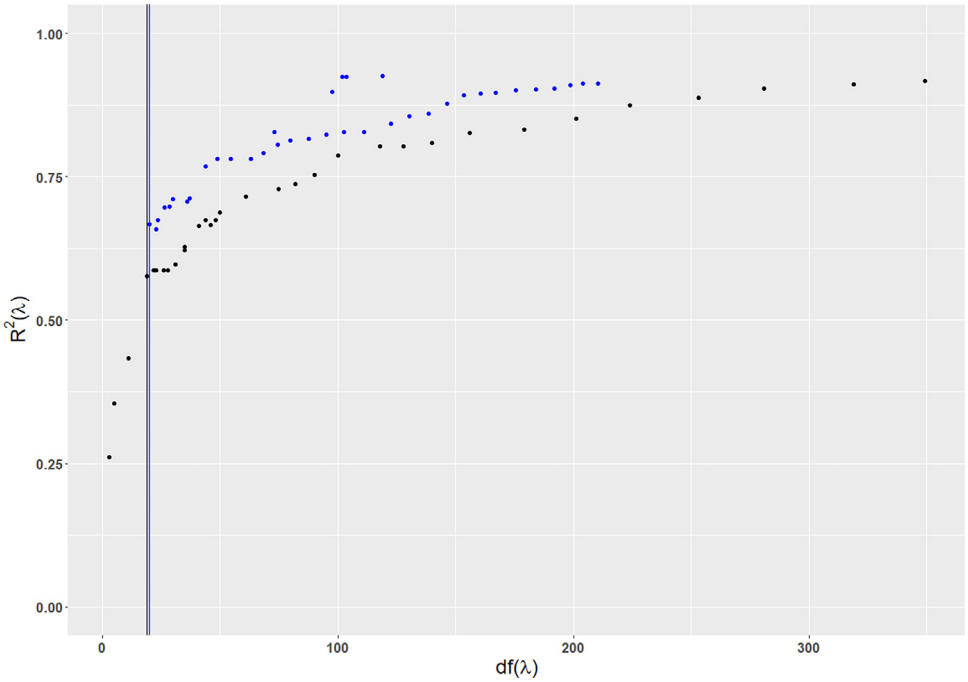

Fig. 1. Plot of $\mathrm{df}(\lambda)$ versus $R^2(\lambda)$ for the solutions $\mathbf{L}_{\mathbf{W}}(\lambda)$ of problem (27) with $\Omega = \Omega^{\mathrm{sym}} \cup \Omega^{\mathrm{pos}}$ (in black) and $\Omega = \Omega^{pos}$ (in blue). The vertical lines show the minima of $BIC(\lambda)$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

5. Conclusions

This paper is an attempt to link compositional data analysis with the concepts of signal processing on graphs. We started from the observation that the Aitchison norm is equally influenced by all pairwise log-ratios between the variables, regardless if these are meaningful from the problem context or not. Modifying the Aitchison norm by putting a non-negative weight onto each (squared) pairwise log-ratio led to a norm that gives different log-ratios a different impact. With this change comes a semi-inner product induced by the Laplacian matrix, well known in graph theory, and a different geometry that still satisfies the most important property for compositional data analysis, namely scale invariance. This radically differs from previous approaches that did consider weighting, for example by constructing an ilr map with the first coordinate being a weighted sum of log-ratios, and the others being such that they form an appropriate basis, without changing the underlying (Aitchison) geometry.

The framework we propose is very flexible and it includes many extensions that allow for different additional modeling choices, such as only considering a low number of interesting subsets of weighted balances — that is, coordinates that consist of a few variables in the numerator as well as denominator. We could show that such modeling choices do not lead to any loss in interpretability and that in fact mappings, such as clr and ilr, which are of central importance in compositional data analysis, have an analog in the graph setting. To find appropriate weights we resorted to an algorithm that selects log-ratios and their weighting based on explained variance of the whole data set.

To show its utility, we applied the proposed methodology to a real data set. In this data set we looked at a regression problem by modeling the risk of getting bacterial vaginosis as being depended on different phylotypes. We showed that the graph perspective leads to a considerably better model which can also be used to gain different interpretable insights by first reducing the dimension and then visualizing the resulting model as a graph. The risk was mainly driven by only a few central phylotypes having strong connections to others.

In the future we intend to look into the performance and possible extensions of the methods presented in this paper to different settings such as classification and regression. Also, we plan to investigate further the problem of finding weights under deviations from the assumptions to more robust cases.

CRediT authorship contribution statement

Christopher Rieser: Conceptualization, Methodology, Software, Data preparation, Visualization, Writing – original draft, Writing and revision. **Peter Filzmoser:** Supervision, Reviewing and editing.

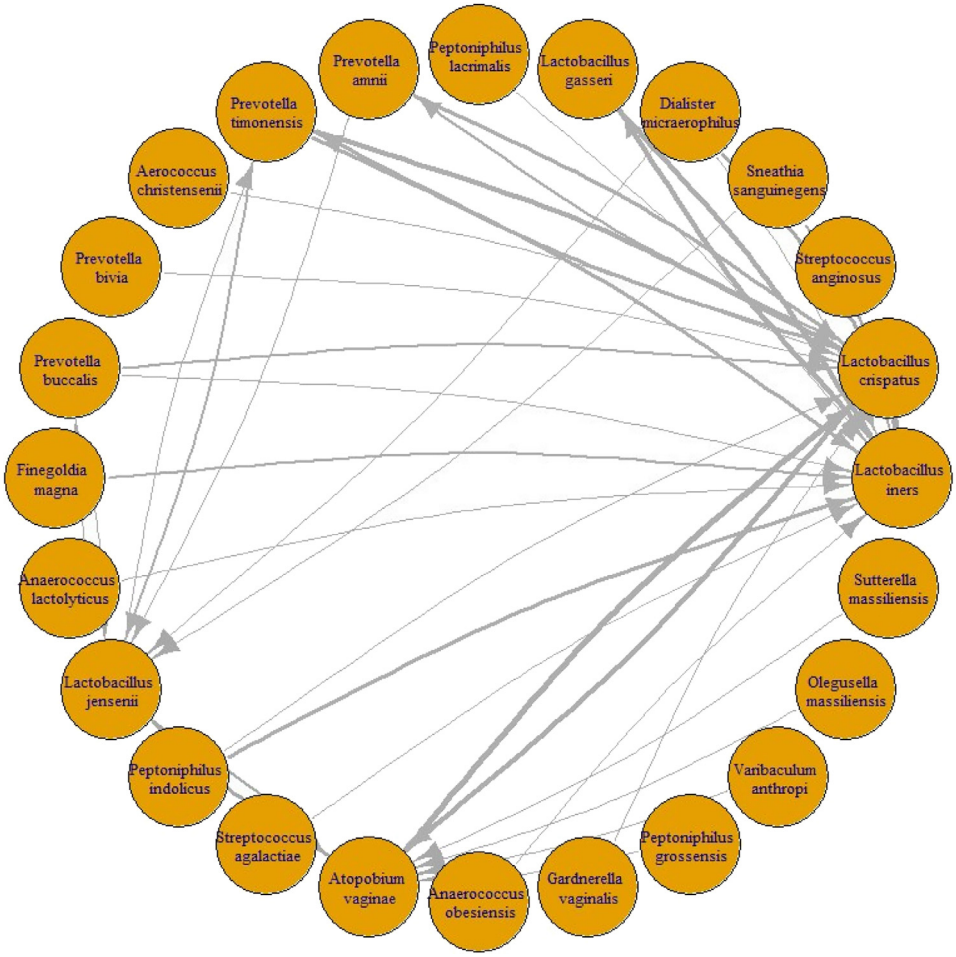

Fig. 2. Directed graph visualization of $\mathbf{L}_{\mathbf{W}}(\lambda_{min})$ with positive weights and non-symmetric structure. The arrow width is proportional to the according weight.

Acknowledgments

This research was supported by the Austrian Science Fund (FWF) under the grant number P 32819 Einzelprojekte. Additionally, we would like to thank the Editor, Associate Editor and the referees for all the valuable comments and suggestions that helped to improve this paper.

Appendix A. Scale-invariant q-norms

An important tool in graph theory is the incidence matrix $\mathbf{d}_{\mathbf{W}} \in \mathbb{R}^{|\mathcal{E}|\times|V|}$. For fixed positive weights $\mathbf{W}$ it is defined as

$$(\mathbf{d}_{\mathbf{W}})_{e,l} := \begin{cases} w_{\ell j}, & e = (\ell, j), \\ -w_{i\ell}, & e = (i, \ell), \\ 0, & \text{else.} \end{cases}$$

The incidence matrix defines a graph theoretic analog to usual differentiation, see Ostrovskii [39], or Grady and Polimeni [21], along weighted edges. For a fixed edge $e = (\ell, j)$ we get $(\mathbf{d}_{\mathbf{W}}\boldsymbol{f})_e = (w_{\ell j}(f_\ell - f_j))$ which measures the weighted difference between values in the nodes ℓ and j. A standard result in graph theory is the equality $\mathbf{d}^\top_{\mathbf{W}^{\frac{1}{2}}}\mathbf{d}_{\mathbf{W}^{\frac{1}{2}}} = 2\mathbf{L}_{\mathbf{W}}$, where the superscript $\frac{1}{2}$ is understood as a coordinate-wise operation. With the latter property we can write Eq. (8) also as

$$\frac{1}{2}\sum_{(i,j)\in\mathcal{E}}(f_i - f_j)(g_i - g_j)w_{ij} = \boldsymbol{f}^\top\mathbf{L}_{\mathbf{W}}\boldsymbol{g} = \frac{1}{2}\langle \mathbf{d}_{\mathbf{W}^{\frac{1}{2}}}\boldsymbol{f}, \mathbf{d}_{\mathbf{W}^{\frac{1}{2}}}\boldsymbol{g}\rangle_2. \tag{32}$$

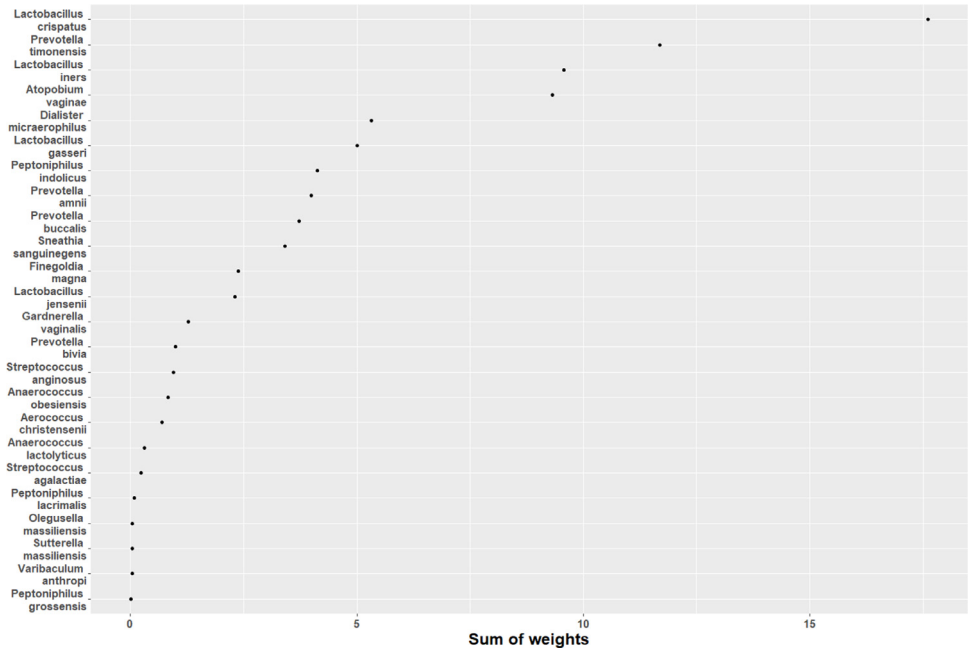

Fig. 3. Plot of the diagonal elements of $\mathbf{L_W}(\lambda_{min})$ with positive weights and non-symmetric structure for the retained nodes ordered from highest to lowest.

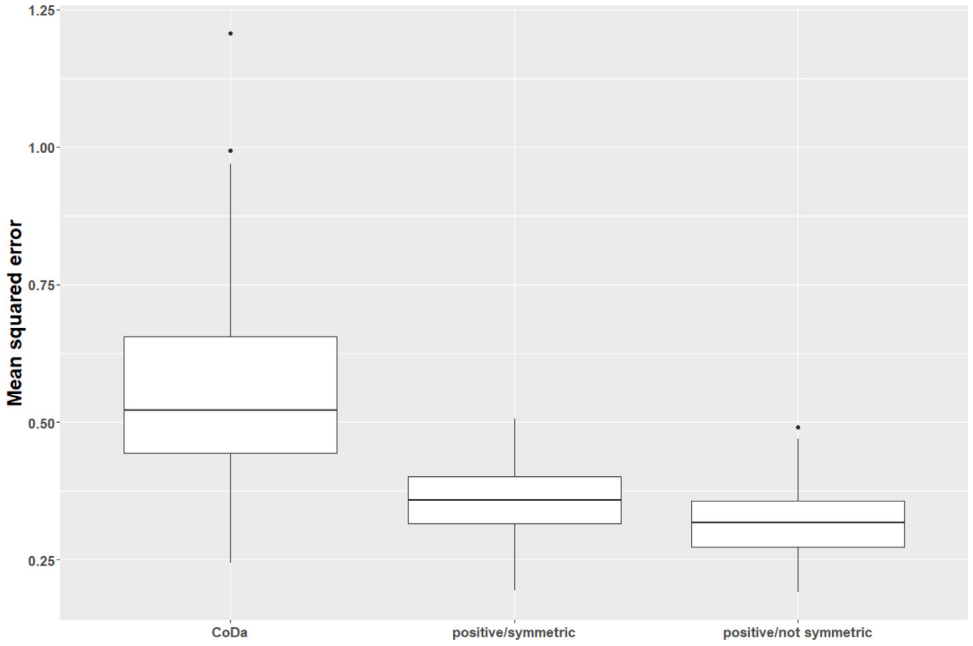

Fig. 4. Boxplots of mean squared errors on the test set for the classical compositional model (far left) and for the weighted one (31) with $\mathbf{L_W}$, constrained by $\Omega = \Omega^{\text{sym}} \cup \Omega^{\text{pos}}$ (middle) and $\Omega = \Omega^{pos}$ (far right), as found by the $BIC(\lambda)$ criterion.

The bilinear form $(\boldsymbol{f}, \boldsymbol{g}) \mapsto \langle \boldsymbol{f}, \mathbf{L_W}\boldsymbol{g}\rangle_2$ can be thought of as the graph analog to the bilinear form $(f, g) \mapsto \int f'(x)g'(x)dx$, for functions $f, g : \mathbb{R} \to \mathbb{R}$ in a suitable function space, giving rise to the Sobolev semi-norm in functional analysis, see Adams and Fournier [1]. Adding $\int f(x)g(x)dx$ to $\int f'(x)g'(x)dx$ turns it into an inner product.

Eq. (32) motivates us to define scale-invariant q-norms. The standard q-norm on the Euclidean space $-\boldsymbol{f} \in \mathbb{R}^D$ is given as

$$\|\boldsymbol{f}\|_q := \begin{cases} \left(\sum_{i=1}^{D} \mid f_i \mid^q\right)^{\frac{1}{q}} & 1 \leq q < \infty, \\ \max_{i=1,\ldots,D} \mid f_i \mid & q = \infty. \end{cases}$$

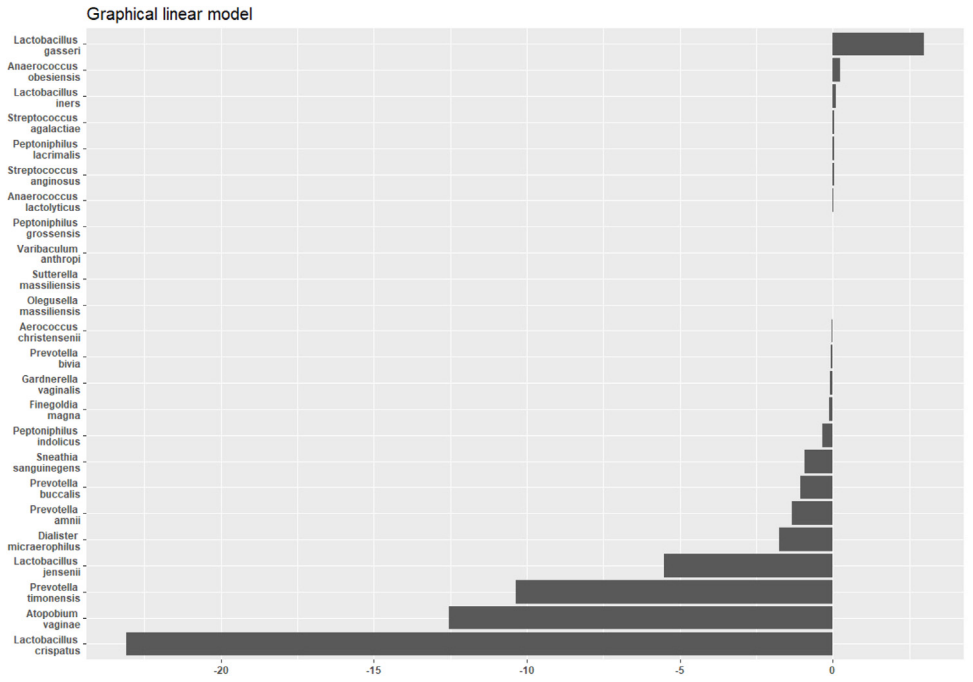

Fig. 5. Transformed estimated coefficients of problem (31) by the map (19) with according permutation and rescaling.

For any $\boldsymbol{x} \in \mathbb{R}^D_+$ we define

$$\|\boldsymbol{x}\|_{q,\alpha} := \begin{cases} \left\| \mathbf{d}_{\mathbf{w}^{\frac{1}{q}}} \ln(\boldsymbol{x}) \right\|_q & 1 \le q < \infty, \\ \|\mathbf{d}_{\mathbf{w}} \ln(\boldsymbol{x})\|_\infty & q = \infty. \end{cases}$$

It can be shown that $\|\cdot\|_{q,}$ is a norm on $(\mathcal{S}^D_{\mathbf{w}}, \oplus_{\mathbf{w}}, \odot_{\mathbf{w}})$. The proof is in Appendix B.

Appendix B. Proofs

Proof of Lemma 2.
As $(\mathcal{S}^D_{\mathbf{w}}, \oplus_{\mathbf{w}}, \odot_{\mathbf{w}})$ can be seen as M separate D-part simplices, $\mathcal{S}^m$, $m \in \{1, \dots, M\}$ – not taking into account any inner products for the moment – we conclude that $(\mathcal{S}^D_{\mathbf{w}}, \oplus_{\mathbf{w}}, \odot_{\mathbf{w}})$ is also a vector space. What is left is to show is that from $\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\mathbf{w}} = 0$ we can conclude $\boldsymbol{x} = \sum_{m=1}^M \frac{\kappa_m}{|\mathcal{V}_m|} \mathbf{1}_{i \in \mathcal{V}_m}$. Using the eigenvalue decomposition of $\mathbf{L}_{\mathbf{W}} = \mathbf{U}\Sigma\mathbf{U}^\top$ we can write, $\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\mathbf{w}} = \ln(\boldsymbol{x})^\top \mathbf{U}\Sigma\mathbf{U}^\top \ln(\boldsymbol{x}) = \sum_{j=1}^D \lambda_j (\langle \boldsymbol{u}_j, \ln(\boldsymbol{x}) \rangle)^2$, with $\lambda_j \ge 0$ and $\lambda_j = 0$ if and only if an eigenvalue of $\mathbf{L}_{\mathbf{W}}$ is zero. Therefore $\langle \boldsymbol{u}_j, \ln(\boldsymbol{x}) \rangle$ must be zero whenever $\lambda_j > 0$ which is equivalent to saying that $\ln(\boldsymbol{x})$ is in the kernel. We know from Lemma 1 that the kernel of $\mathbf{L}_{\mathbf{W}}$ is spanned by $\mathbf{1}_{i \in \mathcal{V}_m}$, so there exist $c_1, \dots, c_m \in \mathbb{R}$ such that $\ln(\boldsymbol{x}) = \sum_{m=1}^M c_m \mathbf{1}_{i \in \mathcal{V}_m}$. From the latter we get $\ln(\boldsymbol{x}_{\mathcal{V}_m}) = c_m \mathbf{1}$, where $\boldsymbol{x}_{\mathcal{V}_m}$ denotes the entries of $\boldsymbol{x}$ with index in $\mathcal{V}_m$, and thus $\boldsymbol{x}_{\mathcal{V}_m} = \exp(c_m)\mathbf{1}$. As, by definition, $\sum_{i \in \mathcal{V}_m} x_i = \kappa_m$ holds, we conclude $\boldsymbol{x}_{\mathcal{V}_m} = \frac{\kappa_m}{|\mathcal{V}_m|}\mathbf{1}$, $\boldsymbol{x} = \sum_{m=1}^M \frac{\kappa_m}{|\mathcal{V}_m|} \mathbf{1}_{i \in \mathcal{V}_m}$. □

Proof of Lemma 3. To prove that $(\mathcal{S}^D_{\sim}, \oplus_{\sim}, \odot_{\sim})$ equipped with $\langle \boldsymbol{x}, \boldsymbol{y} \rangle_{\sim}$ is a Hilbert space we only need to check that (15) is indeed an inner product. Linearity follows trivially from the definitions (13) and (14) as the log is taken in (15). The property $\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\sim} \ge 0$ holds for any $\boldsymbol{x} \in \mathcal{S}^D_{\sim}$ as $\mathbf{L}_{\mathbf{W}}$ is assumed to be positive semi-definite. When $\langle \boldsymbol{x}, \boldsymbol{x} \rangle_{\sim} = 0$ we get by taking the eigen-decomposition of $\mathbf{L}_{\mathbf{W}} = \mathbf{U}\Sigma\mathbf{U}^\top$ and writing (15) as $\sum_{i=1}^{D-M} (\langle \boldsymbol{u}_i, [\ln(\boldsymbol{x})] \rangle_2)^2$ that $[\ln(\boldsymbol{x})]$ is orthogonal to the eigenvectors of the non-zero eigenvalues of $\boldsymbol{u}_i$, and therefore $[\ln(\boldsymbol{x})] \in \mathrm{Ker}(\mathbf{L}_{\mathbf{W}})$, $\ln(\boldsymbol{x}) \in [\mathbf{0}]$. □

Proof of Lemma 4. From the decomposition of $\mathbf{P}\mathbf{L}_{\mathbf{W}}\mathbf{P}^\top$ we have

$$\mathbf{P}\mathbf{L}_{\mathbf{W}}\mathbf{P}^\top = \mathbf{U}\Sigma\mathbf{U}^\top = (\mathbf{U}\Sigma^{\frac{1}{2}}\mathbf{U}^\top)^\top \mathbf{U}\Sigma^{\frac{1}{2}}\mathbf{U}^\top = \mathbf{R}^\top\mathbf{Q}^\top\mathbf{Q}\mathbf{R} = \mathbf{R}^\top_{-\{\ell_1,\dots,\ell_k\}}\mathbf{R}_{-\{\ell_1,\dots,\ell_k\}}.$$

We then have

$$\ln(\boldsymbol{x})^\top \mathbf{L}_{\mathbf{W}} \ln(\boldsymbol{y}) = (\mathbf{P}\ln(\boldsymbol{x}))^\top (\mathbf{P}\mathbf{L}_{\mathbf{W}}\mathbf{P}^\top)(\mathbf{P}\ln(\boldsymbol{y})) = (\mathbf{P}\ln(\boldsymbol{x}))^\top (\mathbf{R}^\top_{-\{\ell_1,\dots,\ell_k\}}\mathbf{R}_{-\{\ell_1,\dots,\ell_k\}})(\mathbf{P}\ln(\boldsymbol{y})) = (\mathbf{C}\mathbf{P}\ln(\boldsymbol{x}))^\top (\mathbf{C}\mathbf{P}\ln(\boldsymbol{y}))$$

From this follows that $\boldsymbol{x} \mapsto \mathbf{CP}\ln(\boldsymbol{x})$(19) preserves distances and inner products and is thus injective. As the map (19) has rank $D - M$ by definition it is also (19) is surjective onto $\mathbb{R}^{D-M}$. □

Proof of Lemma 5. The proof of this Lemma follows directly from the eigen-decomposition. The mapping (20) is per definition an isometry and therefore also injective. Surjectivity follows as $\boldsymbol{u}_i$ form an orthonormal system. The inverse can easily be checked by plugging in one expression into the other. □

Proof that $\|\cdot\|_{q,\alpha}$ is a norm on $(\mathcal{S}^D_{\mathbf{W}}, \oplus_{\mathbf{W}}, \odot_{\mathbf{W}})$. As mentioned in the proof of Lemma 2, $(\mathcal{S}^D_{\mathbf{W}}, \oplus_{\mathbf{W}}, \odot_{\mathbf{W}})$ is a vector space. Because the standard q-norm $\|\cdot\|_q$ is a norm on $\mathbb{R}^D$ all conditions for $\|\cdot\|_{q,\alpha}$ being a norm, but one are trivially fulfilled, as $\boldsymbol{x} \mapsto \ln(\boldsymbol{x})$ maps into a subset of $\mathbb{R}^D$. We only need to proof that from $\|\boldsymbol{x}\|_q = 0$ we can deduce that $\boldsymbol{x}$ is the neutral element. From $\|\boldsymbol{x}\|_q = 0$ we get for each connected subgraph that the pairwise log-ratios to the corresponding edges are zero. As the subgraphs are connected this is equals to $\boldsymbol{x}_{i\in\mathcal{V}_m} = c_m\mathbf{1}$. Again we can conclude, as $\sum_{i\in\mathcal{V}_m} x_i = \kappa_m$ holds, $\boldsymbol{x} = \sum_{m=1}^{M} \frac{\kappa_m}{|\mathcal{V}_m|}\mathbf{1}_{i\in\mathcal{V}_m}$. □

References

[1] R. Adams, J. Fournier, Sobolev Spaces. Vol. 140, second ed., Elsevier, Amsterdam, 2003.
[2] J. Aitchison, The statistical analysis of compositional data, J. R. Stat. Soc. Ser. B Stat. Methodol. 44 (2) (1982) 139–177.
[3] J. Aitchison, The Statistical Analysis of Compositional Data, Chapman & Hall, London, 1986.
[4] C. Barceló-Vidal, J.A. Martín-Fernández, V. Pawlowsky-Glahn, Mathematical foundations of compositional data analysis, in: Proceedings of the Sixth Annual Conference of the International Association for Mathematical Geology, 2001, pp. 1–20.
[5] A. Ben-Israel, T.N. Greville, Generalized Inverses: Theory and Applications. Vol. 15, second ed., Springer Science & Business Media, New York, 2003.
[6] D. Billheimer, P. Guttorp, W.F. Fagan, Statistical interpretation of species composition, J. Amer. Statist. Assoc. 96 (456) (2001) 1205–1214.
[7] K.G. Van den Boogaart, J.J. Egozcue, V. Pawlowsky-Glahn, Bayes Hilbert spaces, Aust. New Zealand J. Stat. 56 (2) (2014) 171–194.
[8] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (1) (2011) 1–122.
[9] F.R. Chung, Spectral Graph Theory, in: CBMS Regional Conference Series, vol. 92, American Mathematical Society, 1997.
[10] C. Croux, P. Filzmoser, H. Fritz, Robust sparse principal component analysis, Technometrics 55 (2) (2013) 202–214.
[11] X. Dong, D. Thanou, P. Frossard, P. Vandergheynst, Learning Laplacian matrix in smooth graph signal representations, IEEE Trans. Signal Process. 64 (23) (2016) 6160–6173.
[12] D. Dumuid, Ž. Pedišić, T.E. Stanford, J.-A. Martín-Fernández, K. Hron, C.A. Maher, L.K. Lewis, T. Olds, The compositional isotemporal substitution model: A method for estimating changes in a health outcome for reallocation of time between sleep, physical activity and sedentary behaviour, Stat. Methods Med. Res. 28 (3) (2019) 846–857.
[13] H.E. Egilmez, E. Pavez, A. Ortega, Graph learning from data under Laplacian and structural constraints, IEEE J. Sel. Top. Sign. Proces. 11 (6) (2017) 825–841.
[14] J.J. Egozcue, V. Pawlowsky-Glahn, Changing the reference measure in the simplex and its weighting effects, Austrian J. Stat. 45 (4) (2016) 25–44.
[15] J.J. Egozcue, V. Pawlowsky-Glahn, G. Mateu-Figueras, C. Barcelo-Vidal, Isometric logratio transformations for compositional data analysis, Math. Geol. 35 (3) (2003) 279–300.
[16] P. Filzmoser, K. Hron, M. Templ, Applied Compositional Data Analysis: With Worked Examples in R, Springer Nature Switzerland, 2018, p. 280.
[17] E. Fišerová, K. Hron, On the interpretation of orthonormal coordinates for compositional data, Math. Geosci. 43 (4) (2011) 455–468.
[18] J. Friedman, T. Hastie, R. Tibshirani, Sparse inverse covariance estimation with the graphical lasso, Biostatistics 9 (3) (2007) 432–441.
[19] J. Friedman, T. Hastie, R. Tibshirani, Regularization paths for generalized linear models via coordinate descent, J. Stat. Softw. 33 (1) (2010) 1.
[20] G.B. Gloor, J.M. Macklaim, V. Pawlowsky-Glahn, J.J. Egozcue, Microbiome datasets are compositional: And this is not optional, Front. Microbiol. 8 (2017) 2224.
[21] L.J. Grady, J.R. Polimeni, Discrete Calculus: Applied Analysis on Graphs for Computational Science, Springer-Verlag London Limited, 2010.
[22] M. Greenacre, Variable selection in compositional data analysis using pairwise logratios, Math. Geosci. 51 (5) (2019) 649–682.
[23] M. Greenacre, EasyCODA: Compositional data analysis in practice, 2020, R package version 0.34.3.
[24] M. Greenacre, E. Grunsky, J. Bacon-Shone, A comparison of isometric and amalgamation logratio balances in compositional data analysis, Comput. Geosci. 148 (2021) 104621.
[25] J.L. Gross, J. Yellen, Graph Theory and Its Applications, second ed., Chapman & Hall /CRC, Boca Raton, 2006.
[26] K. Hron, M. Engle, P. Filzmoser, E. Fišerová, Weighted symmetric pivot coordinates for compositional data with geochemical applications, Math. Geosci. 53 (2021) 655–674.
[27] K. Hron, P. Filzmoser, P. de Caritat, E. Fišerová, A. Gardlo, Weighted pivot coordinates for compositional data and their application to geochemical mapping, Math. Geosci. 49 (2017) 797–814.
[28] K. Hron, P. Filzmoser, K. Thompson, Linear regression with compositional explanatory variables, J. Appl. Stat. 39 (5) (2012) 1115–1128.
[29] K. Hron, L. Kubáček, Statistical properties of the total variation estimator for compositional data, Metrika 74 (2) (2011) 221–230.
[30] V. Kalofolias, How to learn a graph from smooth signals, in: Proceedings of the 19th International Conference on Artificial Intelligence and Statistics, PMLR, 2016, pp. 920–929.
[31] Z.D. Kurtz, C.L. Müller, E.R. Miraldi, D.R. Littman, M.J. Blaser, R.A. Bonneau, Sparse and compositionally robust inference of microbial ecological networks, PLoS Comput. Biol. 11 (5) (2015).
[32] S.L. Lauritzen, Graphical Models, in: Oxford Statistical Science Series, Oxford University Press, Oxford, 1996.
[33] W. Lin, P. Shi, R. Feng, H. Li, Variable selection in regression with compositional covariates, Biometrika 101 (4) (2014) 785–797.
[34] S. Lubbe, P. Filzmoser, M. Templ, Comparison of zero replacement strategies for compositional data with large numbers of zeros, Chemometr. Intell. Lab. Syst. 210 (2021) 104248.
[35] J.I. Marden, Analyzing and Modeling Rank Data, Chapman and Hall, Boca Raton, 1995.
[36] N. Meinshausen, P. Bühlmann, High-dimensional graphs and variable selection with the lasso, Ann. Statist. 34 (3) (2006) 1436–1462.
[37] R. Merris, Laplacian matrices of graphs: A survey, Linear Algebra Appl. 197–198 (1994) 143–176.
[38] B. Mohar, The Laplacian spectrum of graphs, in: Graph Theory, Combinatorics, and Applications, vol. 2, Wiley, 1991, pp. 871–898.
[39] M. Ostrovskii, Sobolev spaces on graphs, Quaest. Math. 28 (4) (2005) 501–523.

[40] V. Pawlowsky-Glahn, J.J. Egozcue, Geometric approach to statistical analysis on the simplex, Stoch. Environ. Res. Risk Assess. 15 (2001) 384–398.
[41] V. Pawlowsky-Glahn, J.J. Egozcue, Compositional data and their analysis: An introduction, Compos. Data Anal. Geosci.:Theory Pract. Spec. Publ. 264 (1) (2006) 1–10.
[42] V. Pawlowsky-Glahn, J.J. Egozcue, R. Tolosana Delgado, Lecture Notes on Compositional Data Analysis, Universitat de Girona, 2007.
[43] V. Pawlowsky-Glahn, J.J. Egozcue, R. Tolosana-Delgado, Modeling and Analysis of Compositional Data, in: Statistics in Practice, John Wiley & Sons, Chichester, 2015.
[44] T.P. Quinn, I. Erb, G. Gloor, C. Notredame, M.F. Richardson, T.M. Crowley, A field guide for the compositional analysis of any-omics data, GigaScience 8 (9) (2019) giz107.
[45] J. Ravel, P. Gajer, Z. Abdo, G.M. Schneider, S.S. Koenig, S.L. McCulle, S. Karlebach, R. Gorle, J. Russell, C.O. Tacket, et al., Vaginal microbiome of reproductive-age women, Proc. Natl. Acad. Sci. USA 108 (Supplement 1) (2011) 4680–4687.
[46] S. Roman, Advanced Linear Algebra, second ed., Springer New York, 2005.
[47] D.M. Witten, R. Tibshirani, T. Hastie, A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis, Biostatistics 10 (3) (2009) 515–534.
[48] M. Yuan, Y. Lin, Model selection and estimation in the Gaussian graphical model, Biometrika 94 (1) (2007) 19–35.