Compositional trend filtering[*]

Christopher Rieser, Peter Filzmoser

TU Wien
Institute of Statistics and Mathematical Methods in Economics
christopher.rieser@tuwien.ac.at
P.Filzmoser@tuwien.ac.at

Abstract

Trend filtering is known as the technique for detecting piecewise linear trends in univariate time series. This technique is extended to the setting of compositional data, which are multivariate data where only the relative information is of importance. According to this, we formulate the problem and present a procedure how to efficiently solve it. To show the usefulness of this method, we consider the number of COVID-19 infections in several European countries in a chosen time period.

Keywords: Trend filtering, compositional data, COVID-19

1. Introduction

Filtering linear trends of a univariate time series has been extensively investigated in the literature, and many methods exist for this purpose, see [7] or [11]. In the univariate context, estimating a piecewise linear trend and its change points can deliver valuable insights and serve as an analytical tool. The standard $l1$ linear trend filtering estimator for given measurements $y_t \in \mathbb{R}$ with equidistant time stamp $t = 1, \ldots, T$ is given as the solution of the following optimisation problem

$$\min_{a_t} \frac{1}{2} \sum_{t=1}^{T} \|y_t - a_t\|^2 + \frac{\lambda}{2} \sum_{t=3}^{T} |a_t - 2a_{t-1} + a_{t-2}|,$$

for a fixed positive tuning parameter λ. The term on the left controls the goodness of fit to the data, whereas the right penalty term enforces the parameters a_t to be close to a linear function in t. This follows from the fact that a lasso penalty [12] is used on the second differences, i.e. $a_t - 2a_{t-1} + a_{t-2}$, setting the latter for certain t – depending on λ – to zero. It is easy to see that when $a_{t_j} - 2a_{t_j-1} + a_{t_j-2} = 0$ holds for consecutive t_j, with $j = 1, \ldots, K$, we get $a_{t_j} = a + bt_j$, for fixed $a, b \in \mathbb{R}$. Therefore, by using a lasso penalty on the second differences we get that a_t is forced to become piecewise linear with growing λ. Recently, trend filtering has also been extended to many other contexts keeping the property that for a growing penalty parameter λ linear functions in the appropriate setting are selected; e.g. graphs [15], vector-valued graphs [13] and additive models [10].

To the best of our knowledge, so far, trend filtering has not been extended to compositional data. Compositional data are in its nature multivariate and strictly positive, and for this type of data it is the relative rather than the absolute information which is of interest. As an example we might consider chemical concentrations for $D \geq 2$ different elements. An observation is written as a D dimensional vector $\boldsymbol{x} = (x_1, \ldots, x_D)'$ with positive entries x_1 to x_D, where each entry is the concentration of a certain element. In a chemical setting it is the relative information between different elements which is of main interest. Relative meaning that the important information is contained in the ratios for two different elements, i.e. $\frac{x_i}{x_j}$, for $i \neq j \in \{1, \ldots, D\}$. In the ground breaking work [1], not only has it been made clear how a compositional view can be advantageous, but also the mathematical foundations of compositional data have been laid out.

In this work we consider fitting linear trends to a time series of compositional data, i.e. we look at the multivariate linear trend of compositional data with an equidistant time stamp t. As an example, consider the case where we compare the number of healthy individuals to infected ones in the whole population of a country. Denoting the number of infected individuals at each time t as κ_t and the total number of individuals in a country by P, the infected vs. non-infected individuals in a population over time can be described by the two dimensional time series $(\kappa_t, P - \kappa_t)$. If we are interested in analyzing how the number of healthy vs unhealthy individuals behaves we can see this as a compositional time series.

This perspective is relevant in many other contexts, e.g. comparing the performance of different stocks relative to each other. This explains the success compositional data analysis has had in past applications. The method we propose in this work guarantees to find an estimator of piecewise compositional linear trends. The trend estimates will be strictly positive and sum up to a given total.

This paper is organized as follows. In Section 2 we will review some important compositional data analysis concepts. In Section 3 we introduce compositional trend filtering, and in Section 4 we present a procedure for the computation. Section 5 shows an application of the presented method to the number of COVID-19 infected individuals in various European countries, and the final Section 6 concludes.

2. Compositional data analysis concepts

In the following, consider a composition $\boldsymbol{x} = (x_1, \ldots, x_D)'$ with D strictly positive entries, called compositional parts, which sum up to 1. This leads to the definition of compositional data as observations from the D part simplex $\mathcal{S}^D$,

$$\mathcal{S}^D := \left\{ \boldsymbol{x} = (x_1, \ldots, x_D)' \in \mathbb{R}_+^D, \ \sum_{i=1}^{D} x_i = 1 \right\}.$$

As established in [1], compositional data aims to capture the relative information between the entries of $\boldsymbol{x}$. This is done by identifying each point $\boldsymbol{x} \in \mathbb{R}_+^D$ with a whole ray starting at zero and going through $\boldsymbol{x}$, which means that we identify an element $\boldsymbol{x} \in \mathcal{S}^D$ with all elements $\gamma \boldsymbol{x}$ for any $\gamma > 0$. In other words, the constraint of sum equal to 1 can always be achieved by rescaling, see [5].

The simplex can be equipped with an addition, multiplication with a scalar, an inner product and a norm, which leads to the so-called Aitchison geometry on the simplex [1]. Consider the compositions $\boldsymbol{x} = (x_1, \ldots, x_D)'$ and $\boldsymbol{y} = (y_1, \ldots, y_D)'$:

- For $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}^D$, perturbation is defined by $\boldsymbol{x} \oplus \boldsymbol{y} := (x_1 y_1, \ldots, x_D y_D)'$

- For $\boldsymbol{x} \in \mathcal{S}^D$ and $\alpha \in \mathbb{R}$, powering is defined by $\alpha \odot \boldsymbol{x} := (x_1^{\alpha}, \ldots, x_D^{\alpha})'$

- For $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{S}^D$, the inner product is defined as

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_A := \frac{1}{2D} \sum_{i=1}^{D} \sum_{j=1}^{D} \log\left(\frac{x_i}{x_j}\right) \log\left(\frac{y_i}{y_j}\right).$$

Remark 2.1. The difference of $\boldsymbol{x}$ and $\boldsymbol{y}$ denoted by $\boldsymbol{x} \ominus \boldsymbol{y}$ is therefore $(\frac{x_1}{y_1}, \ldots, \frac{x_D}{y_D})'$. The norm can be defined in the usual manner using the inner product defined above, i.e $\|\boldsymbol{x}\|_A = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle_A}$.

Interestingly, one can construct isometric mappings from $\mathcal{S}^D$ into $\mathbb{R}^{D-1}$, using the so called (Centered Logratio Coefficients) clr-mapping, which is defined as

$$\mathrm{clr} : \mathcal{S}^D \to \mathbb{R}^D, \quad \mathrm{clr}\,(\boldsymbol{x}) := \left(\log\left(\frac{x_1}{\sqrt[D]{\prod_{i=1}^{D} x_i}} \right), \ldots, \log\left(\frac{x_D}{\sqrt[D]{\prod_{i=1}^{D} x_i}} \right) \right)'.$$

This mapping fulfills important properties regarding addition, multiplication by a scalar and the inner product, namely

$$\mathrm{clr}(\boldsymbol{x} \oplus \boldsymbol{y}) = \mathrm{clr}(\boldsymbol{x}) + \mathrm{clr}(\boldsymbol{y}) \tag{2.1}$$

$$\mathrm{clr}(\alpha \odot \boldsymbol{x}) = \alpha\,\mathrm{clr}(\boldsymbol{x}) \tag{2.2}$$

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle_A = \langle \mathrm{clr}(\boldsymbol{x}), \mathrm{clr}(\boldsymbol{y}) \rangle_E \tag{2.3}$$

where $\langle \cdot, \cdot \rangle_E$ denotes the standard inner product in $\mathbb{R}^D$. However, clr is not a one-to-one mapping onto $\mathbb{R}^D$ as for each $\boldsymbol{x} \in \mathcal{S}^D$ we have $\sum_{i=1}^{D} \mathrm{clr}\,(\boldsymbol{x})_i = 0$; meaning that the sum of the entries of $\mathrm{clr}\,(\boldsymbol{x})$ is always zero and therefore the image of clr lies in the subspace $\{z \in \mathbb{R}^D | \sum_{i=1}^{D} z_i = 0\}$.

Nevertheless, fixing $D - 1$ orthonormal basis vectors $\boldsymbol{v}_1, \ldots, \boldsymbol{v}_{D-1} \in \mathbb{R}^D$, denoted in the following in matrix form $\mathbf{V} \in \mathbb{R}^{D \times (D-1)}$, of the subspace $\{z \in \mathbb{R}^D | \sum_{i=1}^{D} z_i = 0\} \subset \mathbb{R}^D$ one can define an isometry from $\mathcal{S}^D$ to $\mathbb{R}^D$ by

$$\mathrm{ilr}_{\mathbf{V}} : \mathcal{S}^D \to \mathbb{R}^{D-1}, \quad \mathrm{ilr}_{\mathbf{V}}(\boldsymbol{x}) := \mathbf{V}' \, \mathrm{clr}(\boldsymbol{x}),$$

where $\mathbf{V}'$ denotes the transposed matrix, called Isometric Logratio. $\mathrm{ilr}_{\mathbf{V}}$ naturally preserves the properties (2.1), (2.2) and (2.3) and is an isometry. This mapping will be used in the following for trend filtering.

For a more thorough explanation of compositional data and different coordinate representations we refer to [5].

In the following we will speak of a compositional time series when talking about a time series $\boldsymbol{s}_t \in \mathcal{S}^D$, with time index $t = 1, \ldots, T$. It is interesting to note here that due to the compositional nature of $\boldsymbol{s}_t$ we can multiply the latter with any univariate time series $P_t \in \mathbb{R}_+$ such that $P_t \boldsymbol{s}_t$ still lies in $\mathcal{S}^D$ for each t. In a compositional setting, $P_t \boldsymbol{s}_t$ is therefore equivalent to $\boldsymbol{s}_t$. This means that if we would like to go back from a compositional view to absolute numbers one needs to prespecify P_t. We will see how to do that in the next section. In the following, we will write $\boldsymbol{x}_t = P_t \boldsymbol{s}_t$, where P_t is given by the user.

3. Compositional trend filtering

We will now show how to extend the linear univariate trend filtering framework to compositional time series $\boldsymbol{x}_t \in \mathcal{S}^D$ and discuss why the basic property of fitting piecewise linear trends is kept.

We define the trend filtering estimator of a compositional time series $\boldsymbol{x}_t$ as:

$$(\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_T)' := \underset{\boldsymbol{a}_t \in \mathcal{S}^D}{\arg\min} \, \frac{1}{2} \sum_{t=1}^{T} \|\boldsymbol{x}_t \ominus \boldsymbol{a}_t\|_A^2 + \frac{\lambda}{2} \sum_{t=3}^{T} \|\Delta^2 \boldsymbol{a}_t\|_A \qquad (3.1)$$

where $\Delta^2 \boldsymbol{a}_t$ denotes $\boldsymbol{a}_t \ominus 2\boldsymbol{a}_{t-1} \oplus \boldsymbol{a}_{t-2}$, for a fixed $\lambda > 0$. This means that we fit T vectors $\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_T \in \mathcal{S}^D$ to the observed data $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_T$, taking into account a given level of smoothness controlled by the penalty term. When λ goes to infinity we get $\Delta^2 \boldsymbol{a}_t = 0$ which can be shown to be equal to $\boldsymbol{a}_t = \boldsymbol{a} \oplus (t \odot \boldsymbol{b})$, for all t, for some $\boldsymbol{a}$ and $\boldsymbol{b}$ in $\mathcal{S}^D$; i.e. $\boldsymbol{a}_t$ is a linear function in the compositional sense. For $\lambda < \infty$ we will see that we usually get piecewise linear trends in the compositional sense.

Remark 3.1. Problem (3.1) can be extended to fit higher order polynomial trends by defining for all t firstly $\Delta^1 \boldsymbol{x}_t := \boldsymbol{x}_t \ominus \boldsymbol{x}_{t-1}$ resp. $\Delta^2 \boldsymbol{x}_t := \boldsymbol{x}_t \ominus 2\boldsymbol{x}_{t-1} \oplus \boldsymbol{x}_{t-2}$ and then higher order k-th finite differences incrementally by $\Delta^k \boldsymbol{x}_t := \Delta(\Delta^{k-1} \boldsymbol{x}_t)$.

To solve Problem (3.1) we will use isometric logratios. As mentioned in the last section, the mapping ilr$_{\mathbf{V}}$ is, for any fixed basis $\mathbf{V}$, an isometry and therefore the terms $\|\boldsymbol{x}_t \ominus \boldsymbol{a}_t\|_A^2$ resp. $\|\Delta^2 \boldsymbol{a}_t\|_A$ translate for any t into $\|\mathrm{ilr}(\boldsymbol{x}_t) - \mathrm{ilr}(\boldsymbol{a}_t)\|_E^2$ resp. $\|\Delta^2 \mathrm{ilr}(\boldsymbol{a}_t)\|_E$, where in the latter Δ^2 is defined in the same way as before but for the usual addition and multiplication in $\mathbb{R}^{D-1}$.

Therefore we get that it suffices to solve the optimisation problem

$$(\hat{\boldsymbol{u}}_1, \ldots, \hat{\boldsymbol{u}}_T)' := \underset{\boldsymbol{u}_t \in \mathbb{R}^{D-1}}{\arg\min} \; \frac{1}{2} \sum_{t=1}^{T} \|\mathrm{ilr}(\boldsymbol{x}_t) - \boldsymbol{u}_t\|_E^2 + \frac{\lambda}{2} \sum_{t=3}^{T} \|\Delta^2 \boldsymbol{u}_t\|_E. \qquad (3.2)$$

It is easy to see that the latter is strictly convex and therefore has a unique solution $(\hat{\boldsymbol{u}}_1, \ldots, \hat{\boldsymbol{u}}_T)$. By using the inverse of ilr$_{\mathbf{V}}$, we can recover the solution to (3.1) by defining $\hat{\boldsymbol{a}}_t := \mathrm{ilr}_{\mathbf{V}}^{-1}(\hat{\boldsymbol{u}}_t)$ for every t. This also shows that the solution to (3.1) is unique and not dependent on the choice of $\mathbf{V}$ as any solution to (3.1) transformed by ilr$_{\mathbf{V}}$ is necessarily the unique one to (3.2). Therefore, the fit in ilr coordinates for any matrix $\mathbf{V}$, as well as in any special coordinate system like balances or symmetric pivot coordinates is immediately available [5].

Note that in problem (3.2) we use the $l2$ norm as a penalty on the second differences. This corresponds to a group lasso penalty, see [16], on the latter. The interpretation of this is that we look for times t where the trends of $\boldsymbol{x}_t$ change at once for all components together as the penalty automatically sets $\Delta^2 \boldsymbol{a}_t$ for certain t to zero.

Remark 3.2. It is also possible to extend this approach to compositional data with an index t_l, for $l = 1, \ldots, T$, marking rather a position in space than time. In such a case the spacing of t_l is not equidistant. However, a generalisation is straight-forward by taking the penalty terms $\left\| \frac{\Delta \boldsymbol{a}_{t_l}}{t_l - t_{l-1}} \ominus \frac{\Delta \boldsymbol{a}_{t_{l-1}}}{t_{l-1} - t_{l-2}} \right\|_A$ instead; compare with the univariate usual trend filtering extension to non-equidistant points discussed in [9].

We also want to remark that the estimator defined in (3.1) changes accordingly under rescaling of each coordinate. That is, assume that each coordinate of the time series $(x_i)_t$ is rescaled by α_i, meaning that we look at $\alpha_i(x_i)_t$. Then if $\hat{\boldsymbol{a}}_t$ is the solution to (3.1) for the data $\boldsymbol{x}_t$, $\hat{\boldsymbol{a}}_t \oplus \alpha$ is the solution to (3.1) for the data $\boldsymbol{x}_t \oplus \alpha$. The proof is trivial, as $\Delta^2 \boldsymbol{\alpha} = 0$ and because $\boldsymbol{x}_t \ominus \hat{\boldsymbol{a}}_t = (\boldsymbol{x}_t \oplus \boldsymbol{\alpha}) \ominus (\hat{\boldsymbol{a}}_t \oplus \boldsymbol{\alpha})$. Therefore, rescaling the data rescales the estimator accordingly. This is interesting when looking at log-ratios and other compositional tools. As, for example, when analysing the trend in log-ratio coordinates we get that after rescaling, the latter is only shifted by a positive constant.

Finally, when we want to go back to a non-compositional view, we simply multiply the estimator $\hat{\boldsymbol{a}}_t$ with P_t. Choices of P_t are case dependent. If a multivariate smooth to the original data is also of interest then a smoothed version of $\sum_{i=1}^{p}(\boldsymbol{x}_t)_i$ can make sense, see Section 5.

4. Computational considerations

4.1. The ADMM approach

As problem (3.2) has a very special structure, we use an ADMM (Alternating Direction Method of Multipliers) approach. The latter is very easy to implement and has been, with a slight modification, used for the univariate case [9], as well as for multivariate piecewise constant trend filtering before, see [14]. Naturally other, approaches could be used here, such as the Primal-Dual Interior-Point Method, as briefly mentioned in [7]. For a more thorough introduction to ADMM we refer to [3].

For given functions f, g, matrices $\mathbf{A}, \mathbf{B}$ and a vector $\boldsymbol{c}$, the augmented Lagragian of the optimisation problem

$$\min_{\boldsymbol{x}, \boldsymbol{y}} f(\boldsymbol{x}) + g(\boldsymbol{y}) \quad \text{s.t} \quad \mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} = \boldsymbol{c} \tag{4.1}$$

is defined as:

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}) := f(\boldsymbol{x}) + g(\boldsymbol{y}) + \boldsymbol{\theta}'(\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{c}) + \frac{\rho}{2} \|\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{c}\|_E^2,$$

where $\boldsymbol{\theta}$ denotes the dual variable, and $\rho > 0$ is fixed. Solving (4.1) is then done by the following iterative scheme,

$$\boldsymbol{x} \leftarrow \arg\min_{\boldsymbol{x}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}), \ \ \boldsymbol{y} \leftarrow \arg\min_{\boldsymbol{y}} \mathcal{L}(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\theta}), \ \ \boldsymbol{\theta} \leftarrow \boldsymbol{\theta} + \rho(\mathbf{A}\boldsymbol{x} + \mathbf{B}\boldsymbol{y} - \boldsymbol{c}),$$

with given starting vectors for $\boldsymbol{x}, \boldsymbol{y}$ and $\boldsymbol{\theta}$.

To be able to use ADMM for (3.2) we firstly need to reformulate the problem. For this, denote by $\mathcal{I}_D$ the unit matrix of size D and by $\mathcal{O}$ a matrix of only zeros. We define the second difference matrix as

$$\mathbf{D}^2 := \begin{bmatrix} \mathcal{I}_D & -2\mathcal{I}_D & \mathcal{I}_D & \mathcal{O} & \dots & \mathcal{O} \\ \mathcal{O} & \mathcal{I}_D & -2\mathcal{I}_D & \mathcal{I}_D & \mathcal{O} & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \mathcal{O} \\ \mathcal{O} & \mathcal{O} & \mathcal{O} & \mathcal{I}_D & -2\mathcal{I}_D & \mathcal{I}_D \end{bmatrix} \in \mathbb{R}^{(D-1)(T-2) \times (D-1)T}.$$

It is easy to see that for the stacked vector of all $\boldsymbol{u}_t$, i.e. $\boldsymbol{u} := (\boldsymbol{u}_1, \dots, \boldsymbol{u}_T)'$, we have

$$\mathbf{D}^2\boldsymbol{u} = (\boldsymbol{u}_3 - 2\boldsymbol{u}_2 + \boldsymbol{u}_2, \dots, \boldsymbol{u}_T - 2\boldsymbol{u}_{T-1} + \boldsymbol{u}_{T-2})'$$

and therefore problem (3.2) can be written as

$$\arg\min_{\boldsymbol{u}_t \in \mathbb{R}^{D-1}} \frac{1}{2} \sum_{t=1}^{T} \|\mathrm{ilr}_{\mathbf{V}}(\boldsymbol{x}_t) - \boldsymbol{u}_t\|_E^2 + \frac{\lambda}{2} \sum_{t=3}^{T} \|\boldsymbol{\eta}_{t-2}\|_E$$
$$\text{s.t. } \mathbf{D}^2\boldsymbol{u} = \boldsymbol{\eta} \in \mathbb{R}^{(D-1)(T-2)},$$

where $\boldsymbol{\eta}_l$ denotes the subvector $(\eta_{(l-1)(D-1)+1}, \ldots, \eta_{l(D-1)})'$ of $\boldsymbol{\eta}$. In this form we can use ADMM as explained before, where f is simply the first sum and g the second one.

As outlined above, denote by $\boldsymbol{\theta} \in \mathbb{R}^{(p-1)(T-k)}$ the dual variable. The augmented Lagrangian is then given by

$$\mathcal{L}(\boldsymbol{u}, \boldsymbol{\eta}, \boldsymbol{\theta}) := \frac{1}{2} \sum_{t=1}^{T} \left\| \mathrm{ilr}_{\mathbf{V}}(\boldsymbol{x}_t) - \boldsymbol{u}_t \right\|_E^2 + \frac{\lambda}{2} \sum_{t=3}^{T} \left\| \boldsymbol{\eta}_{t-2} \right\|_E$$
$$+ \boldsymbol{\theta}'(\mathbf{D}^2 \boldsymbol{u} - \boldsymbol{\eta}) + \frac{\rho}{2} \left\| \mathbf{D}^2 \boldsymbol{u} - \boldsymbol{\eta} \right\|_E^2.$$

From the latter we can easily deduct the ADMM updates by optimising in each variable at once holding the others fixed. Optimizing in $\boldsymbol{u}$ is simple and can be done by setting the derivative to zero. Optimizing in $\boldsymbol{\eta}$ alone is a group Lasso problem with non-overlapping groups. Denoting the proximal operator of the group lasso with non-overlapping groups as $\mathcal{P}_{\frac{\lambda}{\rho}}$, see [8], and writting from now on $\overline{\boldsymbol{\theta}} := \frac{\boldsymbol{\theta}}{\rho}$, we get the following updates:

$$\boldsymbol{u} \leftarrow (\mathcal{I}_{(D-1)T} + \rho \mathbf{D}^{2\prime} \mathbf{D}^2)^{-1} (\mathrm{ilr}_{\mathbf{V}}(\boldsymbol{x}_t) - \rho \mathbf{D}^{2\prime}(\overline{\boldsymbol{\theta}} - \boldsymbol{\eta})) \qquad (4.2)$$

$$\boldsymbol{\eta} \leftarrow \mathcal{P}_{\frac{\lambda}{\rho}}(\mathbf{D}^2 \boldsymbol{u} + \overline{\boldsymbol{\theta}}) \qquad (4.3)$$

$$\overline{\boldsymbol{\theta}} \leftarrow \overline{\boldsymbol{\theta}} + \mathbf{D}^2 \boldsymbol{u} - \boldsymbol{\eta}. \qquad (4.4)$$

As we usually like to solve problem (3.1) for a whole set of given λ's, e.g. $\lambda_1, \ldots, \lambda_L$, it seems sensible to use as starting vectors $(\boldsymbol{u}, \boldsymbol{\eta}, \overline{\boldsymbol{\theta}})$ for the above iteration a warm start scheme, meaning that when solving problem (3.1) for λ_i we use the solutions obtained by (4.2)–(4.4) belonging to λ_{i-1} as a start for (4.2)–(4.4) belonging to λ_i.

4.2. Cross Validation (CV)

The optimal λ from a set of $\{\lambda_1, \ldots, \lambda_L\}$ can be chosen by K-fold CV. More precisely, denote with $\mathcal{F}_1, \ldots, \mathcal{F}_K$ the folds, i.e. a non-overlapping partition of $1, \ldots, T$ into K sets. In the case of trend filtering these should be chosen in an interleaved way, which means that for any time point t belonging to a certain fold $\mathcal{F}_k$, the elements of the neighbouring time points belong to other folds.

For each pair $(\lambda_i, \mathcal{F}_k)$ we calculate $CV(\lambda_i, \mathcal{F}_k) := \sum_{t \in \mathcal{F}_k} \left\| \mathrm{ilr}_{\mathbf{V}}(\boldsymbol{x}_t) - \hat{\boldsymbol{u}}_t^{-\mathcal{F}_k} \right\|^2$, where $\hat{\boldsymbol{u}}_t^{-\mathcal{F}_k}$ denotes the prediction, at $t \in \mathcal{F}_k$, of the estimator calculated on the subset of observations with indices $\{1, \ldots, T\} \setminus \mathcal{F}_k$.

To chose λ, one can take the argmin in $\{\lambda_1, \ldots, \lambda_L\}$ of

$$CV(\lambda) := \frac{1}{T} \sum_{k=1}^{K} CV(\lambda, \mathcal{F}_k).$$

Denoting with $\lambda^\star$ the argmin of the latter, another popular choice is to use the one standard error rule to choose the optimal λ, as often done for usual univariate trend filtering. Thus we choose the optimal λ as the maximal λ satisfying $CV(\lambda) \le CV(\lambda^\star) + \sigma(\lambda^\star)$, where $\sigma(\lambda)$ is the standard deviation of the data points $CV(\lambda, \mathcal{F}_1), \ldots, CV(\lambda, \mathcal{F}_K)$, see [4].

5. Coronavirus data

To illustrate the utility of the method presented above we will look at the number of COVID-19 infections in 9 different countries in the time period from 2020-03-01 until 2020-07-31. This data set is publicly available at `https://ourworldindata.org/coronavirus-testing`.

In Figure 1 we show the absolute number of reported infections per 100000 inhabitants. As for a very few days the reported number of positive cases are zero, we used time series imputation from the R package forecast [6] to replace the zeros by positive numbers. Most countries show a very similar pattern and reached their maximum number of cases around April; except for Sweden which reached it in June/July. We can see that some countries like Sweden, Spain and Belgium had much higher peaks. However, we can also see that the periods of high values around the peaks also differs a lot among the countries.
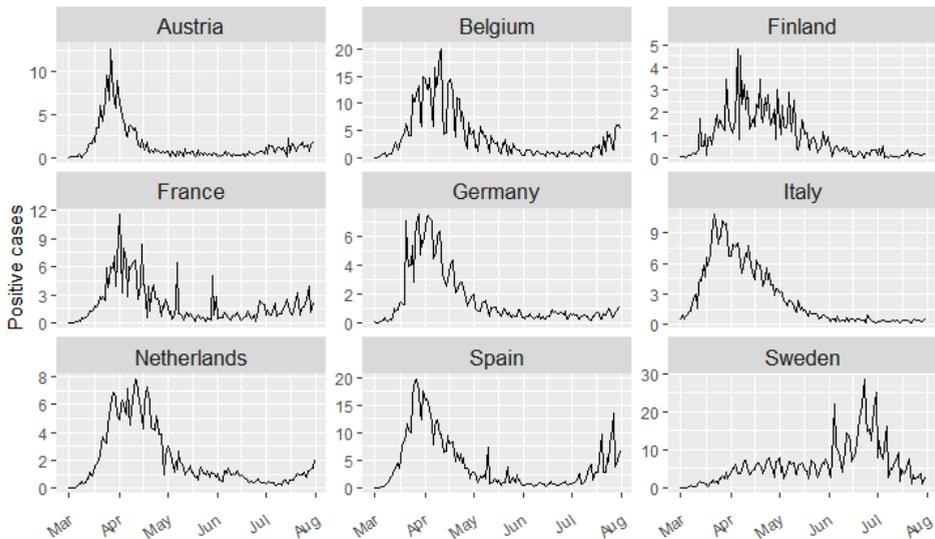

Figure 1. Number of reported COVID-19 infections per 100000 inhabitants from 2020-03-01 to 2020-07-31 for different countries.

To gain more insight into how one country performed in comparison to the whole group of countries we use the method described above. We estimate the multivariate trend filtering fits $\hat{\boldsymbol{a}}_t$ for $\boldsymbol{x}_t$, as described above, for λ chosen by $K = 10$

fold cross validation and the one standard error rule, see Subsection 4.2, with $\lambda \in \{1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30\}$. In Figure 2, we plot the first pivot coordinate $\sqrt{\frac{8}{9}} \log \left(\frac{(\boldsymbol{x}_t)_i}{\sqrt[8]{\prod_{j \neq i}^{9} (\boldsymbol{x}_t)_j}} \right)$, where i is the index corresponding to the i-th country in $\boldsymbol{x}_t$, and the estimated trend. Pivot coordinates can be used to compare the fit (or number of cases) of one country to the geometric mean of the fit (or number of cases) of the rest of the group, containing all relative information of a specific part to the remaining parts in the considered composition [5].
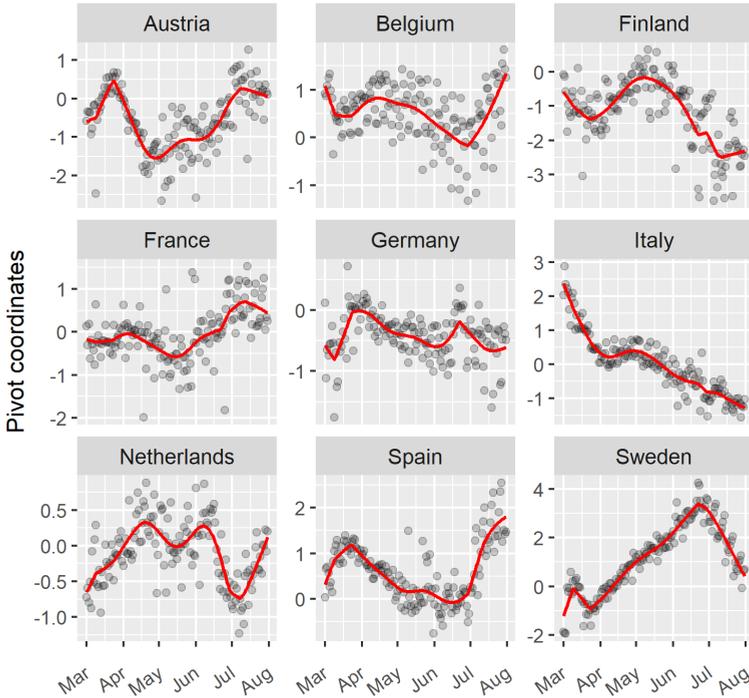

Figure 2. Positive COVID-19 cases in first pivot coordinate per country. The black dots are the measured number of cases in the first pivot coordinate for each country. Equally, the red lines are its compositional trend filtering fits for $\lambda = 10$.

It can be seen in Figure 2 that, compared to the whole group, Germany and Finland have had a comparatively low number of cases, as the values are mainly negative. Finland experienced a trend change in the middle/end of March with rising numbers compared to the geometric mean of the rest of the group, whereas Germany experienced a downward one. It is interesting to see that many countries have had a change in trend since mid/end June - e.g. Belgium, Germany and Sweden. For instance, since the absolute numbers in Spain and Belgium have been rising quickly in July, more than in other countries, also a relative increase com-

pared to the other countries is visible. The contrary behaviour can be seen for Sweden, with a decay in relative number in the beginning of July. It is also surprising to see that Italy, which had the highest infection numbers at the beginning all over Europe, has constantly been improving compared to the other countries, and is by the end of June doing better than average. The Netherlands have at the end of July had average numbers however, its trend might have been alarming.

Furthermore, for a fixed country it might also be interesting to see how the trend of positive cases behaves compared to one other country. For this we display in Figure 3 the log-ratios of Austria and each country present in the composition.
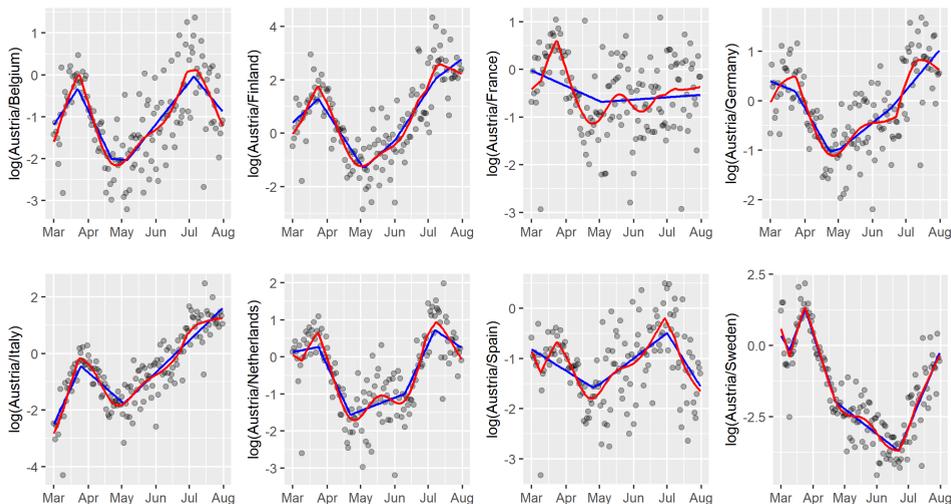

Figure 3. Log-ratios of Austria and all other countries of the composition. The black points are the observations, in red we display the compositional trends and in blue the regular univariate trend-filtering estimates. λ was chosen in both cases by cross validation with the one standard error rule.

We plot the compositional trend filtering fit described by the method above, which was also used for Figure 2, in log-ratios in red. Additionally, we show the univariate non-compositional trend filtering fits to each log-ratio pair in blue with λ obtained by cross validation with the one standard error rule as implemented in the R package [2]. We can see from the log-ratio between Austria and Germany that since July Austria has increasingly more cases than Germany. This upward trend has already started at the beginning of May when Austria still had less positive cases. At the same time the trend for the log-ratio between Austria and Italy started to change. This means that the positive cases per 100000 inhabitants in Austria is growing at a faster rate compared to each, Germany and Italy, since the beginning of July. Something similar holds for the pair Austria and Finland. At the end of July we can see that the cases in Austria compared to Italy or Finland

show a similar trend and numbers – around two. This is surprising as Italy notably started off with the highest infection numbers. At last, since July, Belgium and Spain seem do be doing worse and worse compared to Austria, indicated by a very steep downward trend. The non-compostional univariate trend filtering estimates (in blue) show a slightly different picture. The compositional approach differs sometimes vastly, e.g. for the log-ratio Austria-France, and seems to follow the data better where the non-compositional approach oversmooths, e.g. the log-ratio Austria-Spain in March, end of April and end of June. We also note that the compositional nature leads to trends which on the one hand are at times very smooth, and on the other, also display rapid changes.
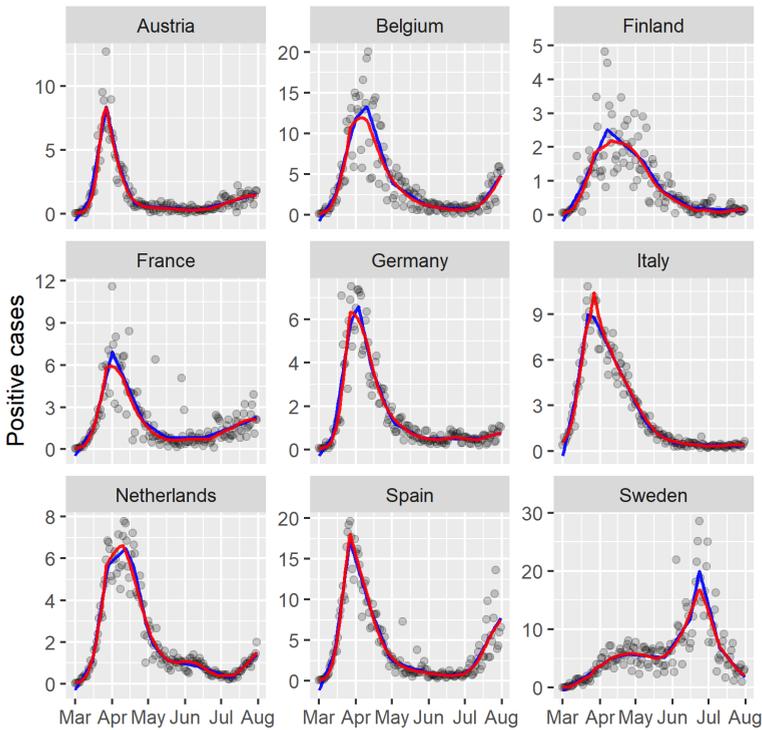

Figure 4. The black dots show for every country the recorded number of positive cases per 100000 inhabitants. The red lines display the estimator $P_t \hat{\boldsymbol{a}}_t$. The blue lines display the trends estimated by non-compositional trend filtering.

The fit $(\hat{\boldsymbol{u}}_1, \ldots, \hat{\boldsymbol{u}}_T)'$ can also be back-transformed to $(\hat{\boldsymbol{a}}_1, \ldots, \hat{\boldsymbol{a}}_T)'$ into the simplex, and the elements per time point sum up to 1. If we want to present those back-transformed fits together with the absolute numbers of infections per 100000 inhabitants, we need to multiply with the total P_t per time point. However, as we want to keep the smoothness of the fits, we will multiply with a smoothed total, as

described at the end of Section 3. To do this we sum up at each time point for all countries the number of positive cases, log-transform the latter and divide it by the absolute maximum, fit a smoother – in our case a univariate trend filter from the R package genlasso [2] with $\lambda = 1$ – and transform it back by multiplying it first with the absolute maximum of the log-transform and then taking its exponential. In Figure 4 we display the fit $P_t\hat{\boldsymbol{a}}_t$ to the original points in red. In blue we show the univariate non-compositional trend filtering fits for $\lambda = 1$, where the data were first divided by the maximum, then the trend filtering fit was obtained, and lastly, the latter was again multiplied by the maximum. Dividing the data by the absolute maximum before fitting puts the λ on a similar footing for comparison with $P_t\hat{\boldsymbol{a}}_t$. Note that the estimator $P_t\hat{\boldsymbol{a}}_t$ always sums up to the total P_t, as we are taking a compositional approach. It is interesting to see that for most countries the highest number of cases had been reached around mid/end of March with a change in trend since then. Also note that the fitted trend to Italy in the first half of April seems to be larger than one would expect from the data in Euclidean space. Examining the data at this time more closely we can see that the number of positive cases actually is very high for two consecutive days, before suddenly dropping and so the compostional fit reflects this better than the non-compositional one. Lastly, the non-compositional approach gives a negative estimate at the beginning of the measurements and shows a slightly different behaviour at the peaks.

6. Summary and conclusions

In this paper we presented a new method for fitting piecewise compositional linear trends to compositional time series. The method we proposed is a direct extension of univariate trend filtering to the compositional setting, which is multivariate by definition. This was done by reformulating the optimization problem in the appropriate Aitchison geometry on the simplex. We showed how to derive the problem in the usual Euclidean geometry, expressed in ilr coordinates, and that the solution does not depend on the specific choice of ilr coordinates. We proceeded by describing how to efficiently solve the problem through an ADMM algorithm.

To show the usefulness of our method, we looked at the number of COVID-19 cases in different European countries in the period from March to July 2020. Namely, once the compositional trend was fitted, we explored the trends in Pivot coordinates and log-ratio representations. This gave insights into how the COVID-19 infections in some countries behaved compared to the compositional mean of other countries, during the said time period.

The fitted trends have been back-transformed to the simplex. The results have to sum up to 1 per time point, and the values cannot be negative – which is often a desirable property. After multiplication by a smoothed total for every country, the total infection numbers can be compared to the smooth line, which still represents relative information to all other countries, and deviations might indicate interesting phenomena (higher variability, etc.).

Additionally to the compositional time series case presented in this paper, one

could look at various extensions. For example, instead of looking at compositional data with a time stamp, one could also look at compositional data with a geographic stamp. That means that we would look at a random variable $\boldsymbol{x}_{(u,s)} \in \mathcal{S}^{D-1}$ with (u,s) being a geographic location. Such a case might be interesting for geochemical exploration where at certain geographic points we measure the concentration of elements. For the latter a shift in concentrations might correspond to interesting areas, thus leading to piecewise constant trend filtering in the compositional sense. Another extension might consider robustification of the proposed method. Outliers are quite common in any statistical setting and compositional data are just as much affected by such. Therefore, investigating a compositionally robust definition of trend filtering might be interesting.

References

[1] J. AITCHISON: *The statistical analysis of compositional data*, Journal of the Royal Statistical Society: Series B (Methodological) 44.2 (1982), pp. 139–160,
DOI: https://doi.org/10.1111/j.2517-6161.1982.tb01195.x.

[2] T. B. ARNOLD, R. J. TIBSHIRANI: *Package 'genlasso'*, Statistics 39.3 (2020), pp. 1335–1371.

[3] S. BOYD, N. PARIKH, E. CHU: *Distributed optimization and statistical learning via the alternating direction method of multipliers*, Now Publishers Inc, 2011,
DOI: https://doi.org/10.1561/9781601984616.

[4] L. BREIMAN, J. FRIEDMAN, C. J. STONE, R. A. OLSHEN: *Classification and Regression Trees*, CRC Press, 1984.

[5] P. FILZMOSER, K. HRON, M. TEMPL: *Applied Compositional Data Analysis*, Switzerland: Springer Nature (2018),
DOI: https://doi.org/10.1007/978-3-319-96422-5.

[6] R. HYNDMAN, G. ATHANASOPOULOS, C. BERGMEIR, G. CACERES, L. CHHAY, M. O'HARA-WILD, F. PETROPOULOS, S. RAZBASH, E. WANG, F. YASMEEN: *forecast: Forecasting functions for time series and linear models*, R package version 8.13, 2020,
URL: https://pkg.robjhyndman.com/forecast/.

[7] S. J. KIM, K. KOH, S. BOYD, D. GORINEVSKY: *l_1 trend filtering*, SIAM Review 51.2 (2009), pp. 339–360,
DOI: https://doi.org/10.1137/070690274.

[8] N. PARIKH, S. BOYD: *Proximal algorithms*, Foundations and trends in optimization 1.3 (2014), pp. 127–239,
DOI: https://doi.org/10.1561/2400000003.

[9] A. RAMDAS, R. J. TIBSHIRANI: *Fast and flexible ADMM algorithms for trend filtering*, Journal of Computational and Graphical Statistics 25.3 (2016), pp. 839–858.

[10] V. SADHANALA, R. J. TIBSHIRANI: *Additive models with trend filtering*, arXiv preprint arXiv:1702.05037 (2017).

[11] R. J. TIBSHIRANI: *Adaptive piecewise polynomial estimation via trend filtering*, The Annals of Statistics 42.1 (2014), pp. 285–323,
DOI: https://doi.org/10.1214/13-AOS1189.

[12] R. TIBSHIRANI: *Regression shrinkage and selection via the lasso*, Journal of the Royal Statistical Society: Series B (Methodological) 58.1 (1996), pp. 267–288,
DOI: https://doi.org/10.1111/j.2517-6161.1996.tb02080.x.

[13] R. Varma, H. Lee, J. Kovačević, Y. Chi: *Vector-valued graph trend filtering with non-convex penalties*, IEEE Transactions on Signal and Information Processing over Networks 6 (2019), pp. 48–62,
DOI: `https://doi.org/10.1109/TSIPN.2019.2957717`.

[14] B. Wahlberg, S. Boyd, M. Annergren, Y. Wang: *An ADMM algorithm for a class of total variation regularized estimation problems*, IFAC Proceedings Volumes 45.16 (2012), pp. 83–88,
DOI: `https://doi.org/10.3182/20120711-3-BE-2027.00310`.

[15] Y.-X. Wang, J. Sharpnack, A. J. Smola, R. J. Tibshirani: *Trend filtering on graphs*, The Journal of Machine Learning Research 17.1 (2016), pp. 3651–3691.

[16] M. Yuan, Y. Lin: *Model selection and estimation in regression with grouped variables*, Journal of the Royal Statistical Society: Series B (Statistical Methodology) 68.1 (2006), pp. 49–67,
DOI: `https://doi.org/10.1111/j.1467-9868.2005.00532.x`.