ℓ_1 Common Trend Filtering

Hiroshi Yamada¹ · Ruoyi Bao²

Accepted: 18 March 2021 / Published online: 11 April 2021
© The Author(s) 2021, Corrected publication 2021

Abstract

The ℓ_1 trend filtering enables us to estimate a continuous piecewise linear trend of univariate time series. This filter and its variants have subsequently been applied in various fields, including astronomy, climatology, economics, electronics, environmental science, finance, and geophysics. Although the ℓ_1 trend filtering can estimate a continuous piecewise linear trend of univariate time series, it cannot estimate a common continuous piecewise linear trend of multiple time series. This paper develops a statistical procedure that enables us to estimate it, which is a multivariate extension of the ℓ_1 trend filtering. We provide an algorithm for estimating it and a clue to specify the tuning parameter of the procedure, both required for its application. We also (i) numerically illustrate how well the algorithm works, (ii) provide an empirical illustration, and (iii) introduce a generalization of our novel method.

Keywords ℓ_1 trend filtering · Common trend · Lasso regression · Reduced rank regression · Total variation denoising · Hodrick–Prescott filtering

Mathematics Subject Classification 62G05

The authors thank Eiji Kurozumi, Kazuhiko Hayakawa, Hiroaki Mukaidani, Heewon Park, Takashi Yamagata, and two anonymous referees for their valuable comments on an earlier version of this paper. The first author dedicates this research to Michio Hatanaka. The usual caveat applies. The Japan Society for the Promotion of Science supported this work through KAKENHI Grant Number 20H01484.

✉ Hiroshi Yamada
yamada@hiroshima-u.ac.jp

¹ School of Informatics and Data Science, Hiroshima University, 1-2-1 Kagamiyama, Higashi-Hiroshima 739-8525, Japan

² Graduate School of Humanities and Social Sciences, Hiroshima University, Higashi-Hiroshima, Japan

1 Introduction

The ℓ_1 trend filtering, which was developed by Steidl et al. (2006), Steidl (2006), Kim et al. (2009), Tibshirani (2014), and Guntuboyina et al. (2020), enables us to extract a continuous piecewise linear trend of univariate time series.¹ Figure 1 illustrates a continuous piecewise linear trend. The filter and its variants have been subsequently applied in various fields, including astronomy (Politsch et al. 2020), climatology (Khodadadi and McDonald 2019), economics (Yamada and Jin 2013; Yamada and Yoon 2014; Winkelried 2016; Yamada 2017; Klein 2018), electronics (Suo et al. 2019), environmental science (Brantley et al. 2019), finance (Mitra and Rohit 2018), and geophysics (Wu et al. 2018).

The ℓ_1 trend filtering is defined by replacing the squared ℓ_2 -norm penalty of the Hodrick–Prescott (HP) (1997) filtering with the ℓ_1 -norm penalty.² It is notable that, even though the modification seems to be somewhat minor, the ℓ_1 trend filtering provides a continuous piecewise linear trend, whereas the HP filtering provides a smooth trend. In econometrics, such a continuous piecewise linear trend was dealt with by Perron (1989) and Rappoport and Reichlin (1989) and it reflects the idea that ‘economic events that have large permanent effects are relatively rare’ (Hamilton 1994). Thus, it is possible to say that the ℓ_1 trend filtering is a method to obtain the trend considered by Perron (1989) and Rappoport and Reichlin (1989).

Although the ℓ_1 trend filtering can estimate a continuous piecewise linear trend of univariate time series, it cannot estimate a common continuous piecewise linear trend of multiple time series. In this paper, we develop a statistical procedure that enables us to estimate it, which is a multivariate extension of the ℓ_1 trend filtering. To explain more precisely, let $y_{i,t}$ be an observation of a univariate time series i at t , where $i = 1, \dots, n$ and $t = 1, \dots, T$, and suppose that it has a continuous piecewise linear trend $x_{i,t}$. As stated, the ℓ_1 trend filtering can be applied for estimating $x_{i,t}$ from $y_{i,t}$. In this paper, we consider the situation such that $x_{i,t}$ can be expressed as

$$x_{i,t} = a_i x_t, \quad i = 1, \dots, n, \quad t = 1, \dots, T, \quad (1)$$

where x_t is a continuous piecewise linear trend and a_i is a loading coefficient. Given that (1) can be represented as

¹ ‘ ℓ_1 trend filtering’ is the terminology used by Kim et al. (2009). The approach is a form of ℓ_1 -norm penalized least squares and may also be regarded as a type of generalized lasso regression (Tibshirani 1996; Kim et al. 2009; Tibshirani and Taylor 2011) and as a generalization of one-dimensional total variation denoising (Rudin et al. 1992; Steidl et al. 2006; Guntuboyina et al. 2020).

² In econometrics, we have observed growing interest in the HP filtering. Recent studies about it include those by de Jong and Sakarya (2016), Cornea-Madeira (2017), Hamilton (2018), Phillips and Jin (2020), Phillips and Shi (2020), Sakarya and de Jong (2020), and Yamada (2020). HP filtering has been used to extract the cyclical component of a univariate time series. For other such methods, see, e.g., Pollock (2016) and Michaelides et al. (2018). Also, it is a type of the Whittaker–Henderson (WH) method of graduation. For a historical survey of the WH method of graduation, see, e.g., Weinert (2007) and Phillips (2010).

Fig. 1 A continuous piecewise linear trend. The ten bullets depict kink points

$$\begin{bmatrix} x_{1,t} \\ \vdots \\ x_{n,t} \end{bmatrix} = \begin{bmatrix} a_1 x_t \\ \vdots \\ a_n x_t \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} x_t, \quad t = 1, \dots, T, \quad (2)$$

even though $y_{1,t}, \dots, y_{n,t}$ commonly have x_t , their linear combination $\beta_1 y_{1,t} + \dots + \beta_n y_{n,t}$ no longer has x_t if $[\beta_1, \dots, \beta_n]'$ is a vector that belongs to the orthogonal complement of the space spanned by $[a_1, \dots, a_n]'$. Hatanaka and Yamada (2003) referred to it as ‘co-trending.’³

In this paper, by extending the ℓ_1 trend filtering, we develop a novel method to estimate x_t and a_i from $y_{i,t}$. Recall that $i = 1, \dots, n$ and $t = 1, \dots, T$, where n (resp. T) represents the number of univariate time series (resp. observations). We refer to the novel filtering method as ‘ ℓ_1 common trend filtering.’ We provide an algorithm for estimating this and a clue to specify the tuning parameter of the procedure, both of which are required for its application. We also (i) numerically illustrate how well the algorithm works, (ii) provide an empirical illustration, and (iii) introduce a generalization of our novel method.

Here, we remark that (2) is not an unlikely model of trends in macroeconomic time series but has strong relevance. To explain more precisely, let $y_{1,t}, \dots, y_{n,t}$ be macroeconomic time series in natural logarithms and $e_{1,t}, \dots, e_{n,t}$ be such that

³ It is a similar concept to cointegration (Engle and Granger 1987). In the case of cointegration, variables have common stochastic trends. For more details, see, e.g., Hatanaka (1996).

$$y_{i,t} = x_{i,t} + e_{i,t}, \quad i = 1, \dots, n, \quad t = 1, \dots, T. \quad (3)$$

Let $g_{i,t} = \Delta y_{i,t} (= y_{i,t} - y_{i,t-1})$. Accordingly, $g_{i,t}$ for $i = 1, \dots, n$ represent the growth rates of the original time series. Then, (2) and (3) are equivalent to

$$\Delta g_{i,t} = a_i b_t + v_{i,t}, \quad i = 1, \dots, n, \quad t = 3, \dots, T, \quad (4)$$

with initial conditions such as $y_{i,1} = a_i x_1 + e_{i,1}$ and $\Delta y_{i,2} = a_i \Delta x_2 + \Delta e_{i,2}$, where $b_t = \Delta x_t - \Delta x_{t-1}$ and $v_{i,t} = \Delta e_{i,t} - \Delta e_{i,t-1}$. Recall that $\Delta g_{i,t}$ in (4) denotes the difference of growth rates of variable i at t . Given that x_t in (2) is a continuous piecewise linear trend, only a few of $b_3, \dots, b_T$ are not equal to 0. We may regard such nonzero b_t s in (4) as occasional permanent shocks that shift growth rates of multiple time series simultaneously and a_i for $i = 1, \dots, n$ in (4) represent individual reaction coefficients of the time series. A typical example of such occasional permanent shocks is the oil price shock in 1973. It is natural to consider that, at the time, the growth rates of macroeconomic time series changed simultaneously with their own reaction rates.

This paper is organized as follows. Section 2 introduces the novel filtering method and provides its reduced-rank-regression (RRR) representations. Section 3 discusses a numerical computation method for x_t and a_i in (1). Section 4 provides a clue to specify the tuning parameter of the procedure required for its application. Section 5 numerically illustrates how well the novel statistical procedure works. Section 6 includes an empirical illustration. Section 7 mentions a generalization of our method. Section 8 concludes the paper.

Notations Let $\mathbf{y}_i = [y_{i,1}, \dots, y_{i,T}]'$, $\mathbf{x}_i = [x_{i,1}, \dots, x_{i,T}]'$, $\mathbf{x} = [x_1, \dots, x_T]'$, $\mathbf{Y} = [\mathbf{y}_1, \dots, \mathbf{y}_n] \in \mathbb{R}^{T \times n}$, $\mathbf{I}_m$ is the identity matrix of order m , $\mathbf{J} = [\mathbf{0}, \mathbf{I}_{T-2}] \in \mathbb{R}^{(T-2) \times T}$, $\mathbf{a} = [a_1, \dots, a_n]'$, and $\mathbf{D} \in \mathbb{R}^{(T-2) \times T}$ be the second order difference matrix such that $\mathbf{D}\mathbf{x}_i = [\Delta^2 x_{i,3}, \dots, \Delta^2 x_{i,T}]'$. Explicitly, $\mathbf{D}$ is the $(T-2) \times T$ Toeplitz matrix of which the first and last rows are $[1, -2, 1, 0, \dots, 0]$ and $[0, \dots, 0, 1, -2, 1]$, respectively. In addition, let

$$\mathbf{X} = \begin{bmatrix} 1 & 1 & 0 & \cdots & 0 & 0 \\ 1 & 2 & 0 & \cdots & 0 & 0 \\ 1 & 3 & 1 & \ddots & \vdots & \vdots \\ 1 & 4 & 2 & \ddots & 0 & \vdots \\ \vdots & \vdots & \vdots & \ddots & 1 & 0 \\ 1 & T & T-2 & \cdots & 2 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{\Pi} & \mathbf{\Psi} \end{bmatrix} \in \mathbb{R}^{T \times T}. \quad (5)$$

Finally, for a vector $\boldsymbol{\gamma} = [\gamma_1, \dots, \gamma_m]'$, $\|\boldsymbol{\gamma}\|_2^2 = \boldsymbol{\gamma}'\boldsymbol{\gamma} = \sum_{t=1}^m \gamma_t^2$, $\|\boldsymbol{\gamma}\|_1 = \sum_{t=1}^m |\gamma_t|$, $\|\boldsymbol{\gamma}\|_\infty = \max\{|\gamma_1|, \dots, |\gamma_m|\}$, and, for a matrix $\boldsymbol{\Gamma} \in \mathbb{R}^{r \times s}$ whose (i, j) entry is denoted by γ_{ij} , $\|\boldsymbol{\Gamma}\|_F^2 = \sum_{i=1}^r \sum_{j=1}^s \gamma_{ij}^2$.

A small note (i) The null space $\mathbf{D}$ is identical to the column space of $\mathbf{\Pi}$ and accordingly $\mathbf{D}\mathbf{\Pi} = \mathbf{0}$, (ii) $\mathbf{\Psi}$ is a right inverse of $\mathbf{D}$, i.e., $\mathbf{D}\mathbf{\Psi} = \mathbf{I}_{T-2}$ (Paige and

Trindade 2010), (iii) $\det(\mathbf{X}) = 1$ and thus $\mathbf{X}$ is nonsingular, and (iv) given (1), we have $[\mathbf{x}_1, \dots, \mathbf{x}_n] = [a_1\mathbf{x}, \dots, a_n\mathbf{x}] = \mathbf{x}\mathbf{a}' \in \mathbb{R}^{T \times n}$.

2 ℓ_1 Common Trend Filtering

2.1 ℓ_1 Trend Filtering

The ℓ_1 trend filtering is defined by

$$\min_{x_{i,1}, \dots, x_{i,T} \in \mathbb{R}} \sum_{t=1}^T (y_{i,t} - x_{i,t})^2 + \psi \sum_{t=3}^T |\Delta^2 x_{i,t}|, \quad (6)$$

where $\psi > 0$ is a tuning parameter. In matrix notation, it is expressed as

$$\min_{\mathbf{x}_i \in \mathbb{R}^T} \|\mathbf{y}_i - \mathbf{x}_i\|_2^2 + \psi \|\mathbf{D}\mathbf{x}_i\|_1. \quad (7)$$

2.2 ℓ_1 Common Trend Filtering

In this paper, we extend the ℓ_1 trend filtering so that we may estimate a common continuous piecewise linear trend of multiple time series, $y_{1,t}, \dots, y_{n,t}$. The filtering method we introduce in this paper is:

$$\min_{\substack{x_1, \dots, x_T \in \mathbb{R} \\ a_1, \dots, a_n \in \mathbb{R}}} \sum_{i=1}^n \sum_{t=1}^T (y_{i,t} - a_i x_t)^2 + \lambda \sum_{t=3}^T |\Delta^2 x_t|, \quad \text{s.t. } \sum_{i=1}^n a_i^2 = 1, \quad (8)$$

where $\lambda > 0$ is a tuning parameter. We refer to the filtering method described by (8) as ℓ_1 common trend filtering. In matrix notation, the filtering is expressed as

$$\min_{\substack{\mathbf{x} \in \mathbb{R}^T \\ a_1, \dots, a_n \in \mathbb{R}}} \sum_{i=1}^n \|\mathbf{y}_i - a_i \mathbf{x}\|_2^2 + \lambda \|\mathbf{D}\mathbf{x}\|_1, \quad \text{s.t. } \sum_{i=1}^n a_i^2 = 1. \quad (9)$$

Furthermore, given that $\sum_{i=1}^n \|\mathbf{y}_i - a_i \mathbf{x}\|_2^2 = \|\mathbf{Y} - \mathbf{x}\mathbf{a}'\|_F^2$ and $\sum_{i=1}^n a_i^2 = \|\mathbf{a}\|_2^2$, (9) can be represented by

$$\min_{\substack{\mathbf{x} \in \mathbb{R}^T \\ \mathbf{a} \in \mathbb{R}^n}} \|\mathbf{Y} - \mathbf{x}\mathbf{a}'\|_F^2 + \lambda \|\mathbf{D}\mathbf{x}\|_1, \quad \text{s.t. } \|\mathbf{a}\|_2^2 = 1. \quad (10)$$

This is an ℓ_1 -norm penalized RRR.

2.3 Another Representation

Let $\mathbf{b} \in \mathbb{R}^T$ be a column vector such as $\mathbf{x} = \mathbf{X}\mathbf{b}$. Given that (i) $\mathbf{X}$ is nonsingular and (ii) $\mathbf{D}\mathbf{X} = \mathbf{J}$ (Paige and Trindade 2010), we obtain another RRR representation of (10):

$$\min_{\substack{\mathbf{a} \in \mathbb{R}^n \\ \mathbf{b} \in \mathbb{R}^T}} f(\mathbf{a}, \mathbf{b}) = \|\mathbf{Y} - \mathbf{X}\mathbf{b}\mathbf{a}'\|_{\text{F}}^2 + \lambda \|\mathbf{J}\mathbf{b}\|_1, \quad \text{s.t. } \|\mathbf{a}\|_2^2 = 1. \quad (11)$$

We remark that when $\mathbf{J}\mathbf{b}$ is sparse, $\mathbf{x} = \mathbf{X}\mathbf{b}$ represents a continuous piecewise linear trend.⁴ In addition, interestingly, (11) is similar to Eq. (8) in Chen and Huang (2012).⁵

3 Numerical Solution

3.1 Two Key Results

Given $\|\mathbf{a}\|_2^2 = 1$, the objective function in (11) can be represented as

$$\|\mathbf{Y} - \mathbf{X}\mathbf{b}\mathbf{a}'\|_{\text{F}}^2 = \text{tr}(\mathbf{Y}'\mathbf{Y}) + \text{tr}(\mathbf{X}\mathbf{b}\mathbf{b}'\mathbf{X}') - 2\text{tr}\{\mathbf{Y}'(\mathbf{X}\mathbf{b}\mathbf{a}')\}. \quad (12)$$

3.1.1 The Case Where $\mathbf{b} \in \mathbb{R}^T$ is Given

Suppose that $\mathbf{b} \in \mathbb{R}^T$ is given. Then, $\mathbf{X}\mathbf{b} (= \mathbf{x}) \in \mathbb{R}^T$ is a known column vector. Because both $\text{tr}(\mathbf{Y}'\mathbf{Y})$ and $\text{tr}(\mathbf{X}\mathbf{b}\mathbf{b}'\mathbf{X}')$ in (12) do not depend on $\mathbf{a}$, when $\mathbf{b} \in \mathbb{R}^T$ is given, (11) reduces to

$$\max_{\mathbf{a} \in \mathbb{R}^n} \text{tr}\{\mathbf{Y}'(\mathbf{X}\mathbf{b}\mathbf{a}')\}, \quad \text{s.t. } \|\mathbf{a}\|_2^2 = 1. \quad (13)$$

We remark that, given $\mathbf{x} = \mathbf{X}\mathbf{b}$, it follows that $\text{tr}\{\mathbf{Y}'(\mathbf{X}\mathbf{b}\mathbf{a}')\} = \sum_{i=1}^n \mathbf{y}'_i(a_i\mathbf{x})$, which is quite reasonable as an objective function for estimating $a_1, \dots, a_n$. Moreover, letting $\boldsymbol{\phi} = \mathbf{Y}'\mathbf{X}\mathbf{b} \in \mathbb{R}^n$, it follows that $\text{tr}\{\mathbf{Y}'(\mathbf{X}\mathbf{b}\mathbf{a}')\} = \boldsymbol{\phi}'\mathbf{a}$. Therefore, instead of (13), we may consider the following constrained maximization problem:

$$\max_{\mathbf{a} \in \mathbb{R}^n} g(\mathbf{a}) = \boldsymbol{\phi}'\mathbf{a}, \quad \text{s.t. } \|\mathbf{a}\|_2^2 = 1. \quad (14)$$

Given $\|\mathbf{a}\|_2 = 1$, by the Cauchy–Schwarz inequality, we obtain $g(\mathbf{a}) = \boldsymbol{\phi}'\mathbf{a} \leq |\boldsymbol{\phi}'\mathbf{a}| \leq \|\boldsymbol{\phi}\|_2 \|\mathbf{a}\|_2 = \|\boldsymbol{\phi}\|_2$, from which we have $g(\mathbf{a}) < g(\hat{\mathbf{a}})$ if $\mathbf{a} \neq \hat{\mathbf{a}}$, where $\hat{\mathbf{a}} = (\boldsymbol{\phi}'\boldsymbol{\phi})^{-1/2} \boldsymbol{\phi} = (\mathbf{b}'\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}\mathbf{b})^{-1/2} \mathbf{Y}'\mathbf{X}\mathbf{b}$. Consequently, given $\mathbf{b} \in \mathbb{R}^T$, we have the following inequality:

⁴ See Section A.1 in the Appendix.

⁵ See Section A.2 in the Appendix.

$$f(\hat{\mathbf{a}}, \mathbf{b}) < f(\mathbf{a}, \mathbf{b}) \quad (15)$$

if $\mathbf{a} \neq \hat{\mathbf{a}}$.

3.1.2 The Case Where $\mathbf{a} \in \mathbb{R}^n$ is Given

Next, suppose that $\mathbf{a} \in \mathbb{R}^n$ such that $\|\mathbf{a}\|_2^2 = 1$ is given. Then, (11) reduces to

$$\min_{\mathbf{b} \in \mathbb{R}^T} \|\mathbf{Y} - \mathbf{X}\mathbf{b}\mathbf{a}'\|_{\text{F}}^2 + \lambda \|\mathbf{J}\mathbf{b}\|_1. \quad (16)$$

Let $\mathbf{A}_\perp \in \mathbb{R}^{n \times (n-1)}$ be a matrix of an orthonormal basis of the orthogonal complement of the space spanned by $\mathbf{a}$. Then, given $\|\mathbf{a}\|_2^2 = 1$, $[\mathbf{a}, \mathbf{A}_\perp] \in \mathbb{R}^{n \times n}$ is an orthogonal matrix. Thus, given that $(\mathbf{Y} - \mathbf{X}\mathbf{b}\mathbf{a}')[\mathbf{a}, \mathbf{A}_\perp] = [\mathbf{Y}\mathbf{a} - \mathbf{X}\mathbf{b}, \mathbf{Y}\mathbf{A}_\perp]$, it follows that $\|\mathbf{Y} - \mathbf{X}\mathbf{b}\mathbf{a}'\|_{\text{F}}^2 = \|\mathbf{Y}\mathbf{a} - \mathbf{X}\mathbf{b}\|_2^2 + \|\mathbf{Y}\mathbf{A}_\perp\|_{\text{F}}^2$. Given that $\|\mathbf{Y}\mathbf{A}_\perp\|_{\text{F}}^2$ does not depend on $\mathbf{b}$, (16) becomes

$$\min_{\mathbf{b} \in \mathbb{R}^T} \|\mathbf{Y}\mathbf{a} - \mathbf{X}\mathbf{b}\|_2^2 + \lambda \|\mathbf{J}\mathbf{b}\|_1. \quad (17)$$

We remark here that (17) can be represented as

$$\min_{\mathbf{x} \in \mathbb{R}^T} \|\mathbf{Y}\mathbf{a} - \mathbf{x}\|_2^2 + \lambda \|\mathbf{D}\mathbf{x}\|_1. \quad (18)$$

See, e.g., Eq. (9) in Kim et al. (2009).⁶

As (17) is a problem whose objective function is coercive and strictly convex over $\mathbb{R}^T$, it has a unique global minimizer. Thus, denoting the solution by $\hat{\mathbf{b}}$, we have the following result. Given $\mathbf{a} \in \mathbb{R}^n$ such that $\|\mathbf{a}\|_2^2 = 1$, we have the following inequality:

$$f(\mathbf{a}, \hat{\mathbf{b}}) < f(\mathbf{a}, \mathbf{b}) \quad (19)$$

if $\mathbf{b} \neq \hat{\mathbf{b}}$.

3.2 A Numerical Algorithm

Based on the above two inequalities, (15) and (19), we introduce a numerical algorithm. Given $\hat{\mathbf{a}}_1 \in \mathbb{R}^n$ and $\hat{\mathbf{b}}_1 \in \mathbb{R}^T$, for $i \in \mathbb{N}$, we define $\hat{\mathbf{a}}_{i+1}$ and $\hat{\mathbf{b}}_{i+1}$ by

$$\hat{\mathbf{a}}_{i+1} = (\hat{\mathbf{b}}_i' \mathbf{X}' \mathbf{Y} \mathbf{Y}' \mathbf{X} \hat{\mathbf{b}}_i)^{-1/2} \mathbf{Y}' \mathbf{X} \hat{\mathbf{b}}_i, \quad (20)$$

$$\hat{\mathbf{b}}_{i+1} = \arg \min_{\mathbf{b}_{i+1} \in \mathbb{R}^T} \|\mathbf{Y} \hat{\mathbf{a}}_{i+1} - \mathbf{X} \mathbf{b}_{i+1}\|_2^2 + \lambda \|\mathbf{J} \mathbf{b}_{i+1}\|_1. \quad (21)$$

Then, we have the following result.

⁶ The dual problem of (18) and its implication are given in Section A.3 in the Appendix.

Lemma 3.1 For $i \in \mathbb{N}$, it follows that $f(\hat{\mathbf{a}}_{i+1}, \hat{\mathbf{b}}_{i+1}) \leq f(\hat{\mathbf{a}}_i, \hat{\mathbf{b}}_i)$, where the equality holds only if $\hat{\mathbf{a}}_i = \hat{\mathbf{a}}_{i+1}$ and $\hat{\mathbf{b}}_i = \hat{\mathbf{b}}_{i+1}$.

Proof (i) From (15) and (20), we have $f(\hat{\mathbf{a}}_{i+1}, \hat{\mathbf{b}}_i) \leq f(\mathbf{a}, \hat{\mathbf{b}}_i)$ for any $\mathbf{a} \in \mathbb{R}^n$ and we thus have $f(\hat{\mathbf{a}}_{i+1}, \hat{\mathbf{b}}_i) \leq f(\hat{\mathbf{a}}_i, \hat{\mathbf{b}}_i)$, where the equality holds only if $\hat{\mathbf{a}}_i = \hat{\mathbf{a}}_{i+1}$. (ii) Likewise, from (19) and (21), we have $f(\hat{\mathbf{a}}_{i+1}, \hat{\mathbf{b}}_{i+1}) \leq f(\hat{\mathbf{a}}_{i+1}, \mathbf{b})$ for any $\mathbf{b} \in \mathbb{R}^T$ and we thus have $f(\hat{\mathbf{a}}_{i+1}, \hat{\mathbf{b}}_{i+1}) \leq f(\hat{\mathbf{a}}_{i+1}, \hat{\mathbf{b}}_i)$, where the equality holds only if $\hat{\mathbf{b}}_i = \hat{\mathbf{b}}_{i+1}$. (iii) Combining these inequalities, we have $f(\hat{\mathbf{a}}_{i+1}, \hat{\mathbf{b}}_{i+1}) \leq f(\hat{\mathbf{a}}_{i+1}, \hat{\mathbf{b}}_i) \leq f(\hat{\mathbf{a}}_i, \hat{\mathbf{b}}_i)$, which leads to $f(\hat{\mathbf{a}}_{i+1}, \hat{\mathbf{b}}_{i+1}) \leq f(\hat{\mathbf{a}}_i, \hat{\mathbf{b}}_i)$, where the equality holds only if $\hat{\mathbf{a}}_i = \hat{\mathbf{a}}_{i+1}$ and $\hat{\mathbf{b}}_i = \hat{\mathbf{b}}_{i+1}$. $\square$

Given Lemma 3.1, we have the following result.

Proposition 3.2 Let $f_i = f(\hat{\mathbf{a}}_i, \hat{\mathbf{b}}_i)$ for $i \in \mathbb{N}$. A sequence of real numbers, $(f_i)_{i \in \mathbb{N}}$, has a finite limit.

Proof From Lemma 3.1, $(f_i)_{i \in \mathbb{N}}$ is a nonincreasing sequence. In addition, as $f_i \geq 0$ for $i \in \mathbb{N}$, it is bounded below. Consequently, it has a finite limit. $\square$

Proposition 3.2 implies that the objective function in (11) converges by alternatively minimizing it over $\mathbf{a}$ and $\mathbf{b}$. Denote $\mathbf{a}$ and $\mathbf{b}$ such that the objective function in (11) is converged by $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$. A Matlab user-defined function for estimating $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ from $\mathbf{Y}$ and λ , `ll_common_trend_filter`, is provided in the supplementary material.

4 A Clue for Specifying the Tuning Parameter

Applying the ℓ_1 common trend filtering requires the specification of its tuning parameter. In this section, we provide a clue for specifying it.

Consider the following convex problem:

$$\min_{\mathbf{b} \in \mathbb{R}^T} h(\mathbf{b}) = \|\mathbf{Y}\mathbf{a} - \mathbf{X}\mathbf{b}\|_2^2, \quad \text{s.t. } \|\mathbf{J}\mathbf{b}\|_1 \leq c, \quad (22)$$

where $c > 0$ and $\mathbf{a} \in \mathbb{R}^n$ is a given vector. Given that $\mathbf{X}$ is nonsingular and $h(\mathbf{b}) \geq 0$, if $\mathbf{b} = \mathbf{X}^{-1}\mathbf{Y}\mathbf{a}$ is feasible, i.e., $\|\mathbf{D}\mathbf{Y}\mathbf{a}\|_1 \leq c$, then $\mathbf{b} = \mathbf{X}^{-1}\mathbf{Y}\mathbf{a}$ is the solution of the above convex problem. Here, recall $\mathbf{J}\mathbf{X}^{-1} = \mathbf{D}$. In the case, $h(\mathbf{X}^{-1}\mathbf{Y}\mathbf{a}) = 0$. If it is not feasible, i.e., $\|\mathbf{D}\mathbf{Y}\mathbf{a}\|_1 > c$, then $\mathbf{b} = \mathbf{X}^{-1}\mathbf{Y}\mathbf{a}$ cannot be the solution. In the case, the solution, denoted by $\tilde{\mathbf{b}}$, locates at the boundary. Thus, we have $\|\mathbf{J}\tilde{\mathbf{b}}\|_1 = c$.

More precisely, concerning $\tilde{\mathbf{b}}$, we have the following results.

Proposition 4.1 If $0 < c < \|\mathbf{D}\mathbf{Y}\mathbf{a}\|_1$, then (i) $\tilde{\mathbf{b}}$ equals $\hat{\mathbf{b}}$ estimated by (16)/(17) with

$$\lambda = \frac{2(\tilde{X}\tilde{\mathbf{b}})'(\mathbf{Y}\mathbf{a} - \tilde{X}\tilde{\mathbf{b}})}{\|\tilde{X}\tilde{\mathbf{b}}\|_1} \quad (23)$$

and (ii) $\|\tilde{X}\tilde{\mathbf{b}}\|_1 = c$ holds.

Proof See Section A.4 in the Appendix. $\square$

Proposition 4.1 implies that we can obtain $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ by specifying c in (22) instead of λ in (17). A Matlab user-defined function for calculating $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ from $\mathbf{Y}$ and c , `l1_common_trend_filter_c`, is provided in the supplementary material. Here, we point out that specifying c is much easier than specifying λ . We do not have any useful information for specifying λ , whereas we have such an information for specifying c . As stated in Proposition 4.1, we may estimate $\hat{\mathbf{b}}$ such that

$$c = \|\hat{X}\hat{\mathbf{b}}\|_1 = \sum_{t=3}^T |\hat{b}_t|. \quad (24)$$

We may utilize this relation for specifying c . See, e.g., (A.1). In the case,

$$x_t - x_{t-1} = \begin{cases} b_2, & \text{if } t = 2, 3, \\ b_2 + b_4, & \text{if } t = 4, 5, \\ b_2 + b_4 + b_6, & \text{if } t = 6, 7. \end{cases} \quad (25)$$

Thus, we may specify rough range of c from the plots of first differences of multiple time series.⁷

Finally, given $X\mathbf{b} = \mathbf{x}$ and $X\mathbf{b} = \mathbf{D}\mathbf{x}$, we remark that the convex problem (22) may be replaced with the following convex problem:

$$\min_{\mathbf{x} \in \mathbb{R}^T} \|\mathbf{Y}\mathbf{a} - \mathbf{x}\|_2^2, \quad \text{s.t. } \|\mathbf{D}\mathbf{x}\|_1 \leq c, \quad (26)$$

where $c > 0$.

5 Numerical Illustrations

In this section, we numerically illustrate how well the algorithm described in the last section works. Figure 1 plots generated $X\mathbf{b}$, where $T = 100$, $b_1 = 10$, $b_2 = 0.5$, and $\sum_{t=3}^T |b_t| = 2.7243$. Ten bullets in the figure depict the kink points and accordingly, 10 entries of $\mathbf{J}\mathbf{b} = [b_3, \dots, b_T]'$ are not equal to 0. Using $X\mathbf{b}$ shown in Fig. 1, we generated $\mathbf{y}_1$, $\mathbf{y}_2$, and $\mathbf{y}_3$ by

$$[\mathbf{y}_1, \mathbf{y}_2, \mathbf{y}_3] = X\mathbf{b}\mathbf{a}' + \mathbf{E}, \quad (27)$$

where $\mathbf{a} = [1, 0.6, 0.2]'$ and $\text{vec}(\mathbf{E}) \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_{3T})$ with $\sigma = 1$. Figure 2 plots them.

⁷ See Sect. 6 for an empirical illustration.

Fig. 2 y_1 , y_2 , and y_3 respectively denote y_1 , y_2 , and y_3 generated by (27). The dashed lines on y_1 , y_2 , and y_3 plot Xb , $0.6Xb$, and $0.2Xb$, respectively. Here, Xb is identical to the solid line depicted in Fig. 1

For obtaining $\hat{a}$ and $\hat{b}$, we used `11_common_trend_filter_c`, which is presented in the supplementary material, with $c = 3$. This required three iterations for convergence. As a result, we obtained

$$\hat{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \end{bmatrix} = \begin{bmatrix} 0.8455 \\ 0.5066 \\ 0.1688 \end{bmatrix}, \quad \hat{a}/\hat{a}_1 = \begin{bmatrix} \hat{a}_1/\hat{a}_1 \\ \hat{a}_2/\hat{a}_1 \\ \hat{a}_3/\hat{a}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.5991 \\ 0.1996 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 0.6 \\ 0.2 \end{bmatrix}, \quad (28)$$

and $\hat{b}$ such that $\|J\hat{b}\|_1 = 3$, the latter of which is consistent with Proposition 4.1(ii). The value of λ for obtaining $\hat{b}$ from $Y\hat{a}$ is 10.3730.

Figure 3 illustrates the results. The solid line in Fig. 3 plots the estimated $\hat{a}_1 X \hat{b}$. The dashed line in the figure plots $a_1 X b$. Note that, given $a_1 = 1$, it is identical to the solid line depicted in Fig. 1. From the figure, we can see that $\hat{a}_1 X \hat{b}$ looks very much like $a_1 X b$. Figure 4 also illustrates the results. The solid lines in Fig. 4 are identical to those in Fig. 2. The dashed lines on y_1 , y_2 and y_3 respectively plot $\hat{a}_1 X \hat{b}$, $\hat{a}_2 X \hat{b}$, and $\hat{a}_3 X \hat{b}$. Again, the figure shows that our novel procedure works well.

As a supplementary examination, we generated an additional data set, repeated the same analysis, and revealed similar results. For example, we obtained

Fig. 3 The dashed line plots $a_1 X \mathbf{b}$. Since $a_1 = 1$, it is identical to the solid line depicted in Fig. 1. The solid line plots estimated $\hat{a}_1 X \hat{\mathbf{b}}$ by the algorithm shown in Sect. 3.2 with $c = 3$

$$\hat{\mathbf{a}} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \end{bmatrix} = \begin{bmatrix} 0.8462 \\ 0.5060 \\ 0.1670 \end{bmatrix}, \quad \hat{\mathbf{a}}/\hat{a}_1 = \begin{bmatrix} \hat{a}_1/\hat{a}_1 \\ \hat{a}_2/\hat{a}_1 \\ \hat{a}_3/\hat{a}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.5980 \\ 0.1974 \end{bmatrix} \approx \begin{bmatrix} 1 \\ 0.6 \\ 0.2 \end{bmatrix}. \quad (29)$$

See also Figures B.1–B.4 in the supplementary material.

6 An Empirical Illustration

Figure 5, which is identical to Figure 1.2 of Hatanaka and Yamada (2003), plots two quarterly macroeconomic time series. More precisely, the upper [resp. lower] panel of the figure depicts the natural logarithm of Japanese M2+CD [resp. real gross domestic product (GDP)], from the first quarter of 1980 to the third quarter of 2001. From the figure, we can observe that these two time series seem to contain a common piecewise linear trend such that a major kink point is located at around 1991, which corresponds to the peak of the Japanese asset price bubble. Actually, the statistical procedure developed by Hatanaka and Yamada (2003) detected a common piecewise linear trend. [See Section 8.5 of Hatanaka and Yamada (2003).]

Figure 6 depicts the corresponding demeaned series. We estimated a common piecewise linear trend of these demeaned data. Denote it by $X \hat{\mathbf{b}}$. For the estimation, we used `l1_common_trend_filter_c` with $c = 0.018$. We specified the value of c by reference to the plots of time series in first differences shown in Fig. 7. The

Fig. 4 The solid lines are identical to those of Fig. 2. The dashed lines on y_1 , y_2 and y_3 respectively plot estimated $\hat{a}_1 X \hat{\mathbf{b}}$, $\hat{a}_2 X \hat{\mathbf{b}}$, and $\hat{a}_3 X \hat{\mathbf{b}}$ by the algorithm shown in Section 3.2 with $c = 3$

upper panel (resp. lower panel) of Fig. 8 depicts $X \hat{\mathbf{b}}$ (resp. $\hat{b}_2, \dots, \hat{b}_T$, where $T = 87$). From the panels, we may observe that (i) a major kink point is located at around 1991 and (ii) $\sum_{t=3}^T |\hat{b}_t|$ equals the value of c . The solid lines in Fig. 9 are identical to those plotted in Fig. 6. The dashed line in the upper (resp. lower) panel plots $\hat{a}_1 X \hat{\mathbf{b}}$ (resp. $\hat{a}_2 X \hat{\mathbf{b}}$), where

$$\hat{\mathbf{a}} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix} = \begin{bmatrix} 0.8962 \\ 0.4437 \end{bmatrix}, \quad \hat{\mathbf{a}}/\hat{a}_1 = \begin{bmatrix} \hat{a}_1/\hat{a}_1 \\ \hat{a}_2/\hat{a}_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0.4951 \end{bmatrix}.$$

Finally, the dashed lines in Fig. 10 are the mean-restored estimated piecewise linear trends. The solid lines in the figure are identical to those in Fig. 5.

7 A Generalization

In this section, we mention a generalization of our method briefly. Let $\mathbf{D}_p \in \mathbb{R}^{(T-p) \times T}$ be the p -th order difference matrix such that $\mathbf{D}_p \mathbf{x}_i = [\Delta^p x_{i,p+1}, \dots, \Delta^p x_{i,T}]'$. Explicitly, $\mathbf{D}_p$ is a Toeplitz matrix as follows:

Fig. 5 Japanese $\log(\text{M2+CD})$ and $\log(\text{real GDP})$ from the first quarter of 1980 to the third quarter of 2001. This figure is identical to Figure 1.2 in Hatanaka and Yamada (2003)

$$\mathbf{D}_p = \begin{bmatrix} a_0 & \cdots & a_p & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & a_0 & \cdots & a_p \end{bmatrix}, \quad (30)$$

where $a_k = (-1)^{p-k} \binom{p}{k}$ for $k = 0, \dots, p$. Accordingly, $\mathbf{D}_2$ equals $\mathbf{D}$. Then, without any difficulty, we may extend our procedure to

$$\min_{\mathbf{x} \in \mathbb{R}^T} \sum_{i=1}^n \|\mathbf{y}_i - a_i \mathbf{x}\|_2^2 + \lambda_p \|\mathbf{D}_p \mathbf{x}\|_1, \quad \text{s.t.} \quad \sum_{i=1}^n a_i^2 = 1, \quad (31)$$

$$a_1, \dots, a_n \in \mathbb{R}$$

where $p \in \mathbb{N}$ and $\lambda_p > 0$ is a tuning parameter. We refer to (31) as ‘ ℓ_1 common polynomial trend filtering.’ The solution of the problem represents a continuous piecewise $(p-1)$ -th order polynomial trend. Note that the corresponding $\mathbf{X}_p = [\mathbf{\Pi}_p, \mathbf{\Psi}_p]$ such that $\mathbf{D}_p \mathbf{X}_p = [\mathbf{D}_p \mathbf{\Pi}_p, \mathbf{D}_p \mathbf{\Psi}_p] = [\mathbf{0}, \mathbf{I}_{T-p}] \in \mathbb{R}^{(T-p) \times T}$ is given in Yamada (2015).

Fig. 6 Japanese $\log(\text{M2+CD})$ and $\log(\text{real GDP})$ from the first quarter of 1980 to the third quarter of 2001. Each time series data point reduced by its mean

8 Concluding Remarks

In this paper, we developed an extension of the ℓ_1 trend filtering. The ℓ_1 trend filtering can estimate a continuous piecewise linear trend of univariate time series. However, it cannot estimate a common continuous piecewise linear trend of multiple time series, which the novel statistical procedure developed in this paper enables. We provided an algorithm and a clue to specify the tuning parameter of the procedure, both of which are required for its application. We also numerically illustrated how well the algorithm works, provided an empirical illustration, and introduced a generalization of our novel method.

Fig. 7 First differences of two time series plotted in Fig. 6

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10614-021-10114-9>.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

A Appendix

A.1 $x = Xb$ when Jb is sparse

We illustrate how $x = Xb$ represents a continuous piecewise linear trend when $Jb = [b_3, \dots, b_T]'$ is sparse. Consider the case where $T = 7$ and $b = [b_1, b_2, 0, b_4, 0, b_6, 0]'$. In the case, x_t ($t = 1, \dots, 7$) are expressed as follows:

Fig. 8 The upper panel plots an estimated common piecewise linear trend of 1D signals shown in Fig. 6. This is obtained by `ll_common_trend_filter_c.m` with $c = 0.018$. The lower panel plots $\hat{b}_2, \dots, \hat{b}_T$, where $T = 87$. We may observe that (i) a major kink point is located at around 1991 and (ii) $\sum_{t=3}^T |\hat{b}_t|$ equals the value of c

$$x_t = \begin{cases} b_1 + b_2 t, & \text{if } t = 1, 2, 3, \\ b_1 + b_2 t + b_4(t - 3), & \text{if } t = 4, 5, \\ b_1 + b_2 t + b_4(t - 3) + b_6(t - 5), & \text{if } t = 6, 7. \end{cases} \quad (\text{A.1})$$

Moreover, these three trends are connected at $t = 3$ and $t = 5$ as

$$\begin{aligned} b_1 + b_2 \cdot 3 &= b_1 + b_2 \cdot 3 + \underbrace{b_4 \cdot (3 - 3)}_{=0}, \\ b_1 + b_2 \cdot 5 + b_4 \cdot (5 - 3) &= b_1 + b_2 \cdot 5 + b_4 \cdot (5 - 3) + \underbrace{b_6 \cdot (5 - 5)}_{=0}. \end{aligned}$$

Thus, $x_1, \dots, x_7$ are on the following continuous piecewise linear function of t :

$$\alpha(t) = b_1 + b_2 t + b_4(t - 3)_+ + b_6(t - 5)_+, \quad (1 \leq t \leq 7), \quad (\text{A.2})$$

where $(\zeta)_+ = \zeta$ if $\zeta > 0$ and $(\zeta)_+ = 0$ if $\zeta \leq 0$. In this function, $(3, \alpha(3))$ and $(6, \alpha(6))$ are kink points. Accordingly, $\{x_1, x_2, x_3\}$ are on the same straight line whose slope is b_2 , $\{x_3, x_4, x_5\}$ are on the same straight line whose slope is $b_2 + b_4$, and $\{x_5, x_6, x_7\}$ are on the same straight line whose slope is $b_2 + b_4 + b_6$.

Fig. 9 The solid lines are identical to those plotted in Fig. 6. The dashed lines are estimated piecewise linear trend from the common piecewise linear trend shown in Fig. 8. More precisely, denoting the estimated common piecewise linear trend by $\widehat{X\hat{b}}$, the dashed line in the upper (resp. lower) panel plots $0.8962X\hat{b}$ (resp. $0.4437X\hat{b}$)

A.2 (11) and Chen and Huang (2012)

We point out a similarity between the ℓ_1 -norm penalized RRR (11) and Eq. (8) in Chen and Huang (2012):

$$\begin{aligned} \min_{\substack{\mathbf{A} \in \mathbb{R}^{n \times r} \\ \mathbf{B} \in \mathbb{R}^{p \times r}}} & \|\mathbf{Y} - \mathbf{ZBA}'\|_{\text{F}}^2 + \sum_{i=1}^p \lambda_i \|\mathbf{B}^{(i)}\|_2, \quad \text{s.t. } \mathbf{A}'\mathbf{A} = \mathbf{I}_r, \end{aligned} \quad (\text{A.3})$$

where $\mathbf{Z} \in \mathbb{R}^{T \times p}$ is a matrix of full column rank, $\mathbf{B}^{(i)}$ denotes the i th row of $\mathbf{B}$, and $\lambda_i > 0$ ($i = 1, \dots, p$) are tuning parameters. To clarify the similarity, let $r = 1$, $p = T$, and $\lambda_i = \lambda$ for $i = 1, \dots, T$ in (A.3). Then, $\mathbf{A} \in \mathbb{R}^n$, $\mathbf{B} \in \mathbb{R}^T$, and $\mathbf{Z} \in \mathbb{R}^{T \times T}$, and we denote them by $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{X}$, respectively. Given $\|\mathbf{B}^{(i)}\|_2 = |b_i|$ in this setting, where $\mathbf{b} = [b_1, \dots, b_T]'$, (A.3) finally becomes

Fig. 10 The dashed lines are the mean-restored estimated piecewise linear trends. The solid lines are identical to those in Fig. 5

$$\min_{\substack{\mathbf{a} \in \mathbb{R}^n \\ \mathbf{b} \in \mathbb{R}^T}} \|\mathbf{Y} - \mathbf{X}\mathbf{b}\mathbf{a}'\|_F^2 + \lambda \|\mathbf{b}\|_1, \quad \text{s.t. } \|\mathbf{a}\|_2^2 = 1, \quad (\text{A.4})$$

which is not identical to (11), but they are similar.

A.3 The dual problem of (18)

The dual problem of (18) is

$$\min_{\mathbf{v} \in \mathbb{R}^{T-2}} \|\mathbf{Y}\mathbf{a} - \mathbf{D}'\mathbf{v}\|_2^2, \quad \text{s.t. } 2\|\mathbf{v}\|_\infty \leq \lambda. \quad (\text{A.5})$$

See, e.g., Proposition 4.1 in Steidl et al. (2006) and Eq. (13) in Kim et al. (2009). The problem (A.5) is of use not only because it can be applied for solving (18) [and accordingly (17)] but also because we may see that $\hat{\mathbf{v}} = (\mathbf{D}\mathbf{D}')^{-1}\mathbf{D}\mathbf{Y}\mathbf{a}$ is feasible if $2\|(\mathbf{D}\mathbf{D}')^{-1}\mathbf{D}\mathbf{Y}\mathbf{a}\|_\infty \leq \lambda$ and in the case it follows that $\hat{\mathbf{x}} = \mathbf{Y}\mathbf{a} - \mathbf{D}'\hat{\mathbf{v}} = \mathbf{\Pi}(\mathbf{\Pi}'\mathbf{\Pi})^{-1}\mathbf{\Pi}'\mathbf{Y}\mathbf{a}$, which represents a linear trend estimated by ordinary least squares. Here, $\hat{\mathbf{v}}$ [resp. $\hat{\mathbf{x}}$] denotes a unique global minimizer of (A.5) [resp. (18)].

A.4 Proof of Proposition 4.1

For proving Proposition 4.1, we give the following two lemmata:

Lemma A.1 *There exists $\mu \geq 0$ such that:*

$$\begin{aligned} (\text{stationarity}) \quad & -2\mathbf{X}'(\mathbf{Y}\mathbf{a} - \mathbf{X}\tilde{\mathbf{b}}) + \mu\mathbf{J}'\tilde{\mathbf{v}} = \mathbf{0}, \\ (\text{complementary slackness}) \quad & \mu(\|\mathbf{J}\tilde{\mathbf{b}}\|_1 - c) = 0, \end{aligned}$$

where $\mathbf{J}'\tilde{\mathbf{v}}$ in the stationarity condition denotes a subgradient vector of $\|\mathbf{J}\mathbf{b}\|_1$ at $\mathbf{b} = \tilde{\mathbf{b}}$. Explicitly, $\tilde{\mathbf{v}} = [\tilde{v}_3, \dots, \tilde{v}_T]'$ is defined by $\tilde{v}_t \in \{1\}$ if $\tilde{b}_t > 0$, $\tilde{v}_t \in \{-1\}$ if $\tilde{b}_t < 0$, and $\tilde{v}_t \in [-1, 1]$ if $\tilde{b}_t = 0$, for $t = 3, \dots, T$.

Proof See Section A.5.1. □

Lemma A.2 *If $c < \|\mathbf{D}\mathbf{Y}\mathbf{a}\|_1$, then $\mathbf{X}\tilde{\mathbf{b}} \neq \mathbf{Y}\mathbf{a}$ follows.*

Proof See Section A.5.2. □

Now, we are ready to give a proof of Proposition 4.1. From Lemma A.2, if $c < \|\mathbf{D}\mathbf{Y}\mathbf{a}\|_1$, then $(\mathbf{Y}\mathbf{a} - \mathbf{X}\tilde{\mathbf{b}})$ in the stationarity condition in Lemma A.1 is not equal to $\mathbf{0}$. In addition, $\mathbf{X}'$ is of full column rank. Hence, if $c < \|\mathbf{D}\mathbf{Y}\mathbf{a}\|_1$, then we have $\mu\mathbf{J}'\tilde{\mathbf{v}} = 2\mathbf{X}'(\mathbf{Y}\mathbf{a} - \mathbf{X}\tilde{\mathbf{b}}) \neq \mathbf{0}$, which leads to $\mu \neq 0$. Thus, given $\mu \geq 0$, it follows that $\mu > 0$. From the complementary slackness condition in Lemma A.1, it follows that $\|\mathbf{J}\tilde{\mathbf{b}}\|_1 = c > 0$. Given that $(\mathbf{J}\tilde{\mathbf{b}})' \tilde{\mathbf{v}} = \sum_{t=3}^T |\tilde{b}_t| = \|\mathbf{J}\tilde{\mathbf{b}}\|_1$, premultiplying the stationarity condition by $\tilde{\mathbf{b}}'$ yields $-2(\mathbf{X}\tilde{\mathbf{b}})'(\mathbf{Y}\mathbf{a} - \mathbf{X}\tilde{\mathbf{b}}) + \mu\|\mathbf{J}\tilde{\mathbf{b}}\|_1 = 0$. Then, given $\|\mathbf{J}\tilde{\mathbf{b}}\|_1 = c > 0$, we obtain (23). In addition, given $\mu > 0$, from the stationarity condition, it follows that $\tilde{\mathbf{b}}$ equals $\hat{\mathbf{b}}$ estimated with $\lambda = \mu$.

A.5 Miscellaneous proofs

A.5.1 Proof of Lemma A.1

As $\|\mathbf{J}\mathbf{b}\|_1 - c = -c < 0$ if $\mathbf{b} = \mathbf{0}$, Slater's condition is satisfied. Then, from Schechter (1977) [and Theorem 1.2 in Weir (1988)], there exists $\mu \geq 0$ such that

$$\mathbf{0} \in \partial\|\mathbf{Y}\mathbf{a} - \mathbf{X}\tilde{\mathbf{b}}\|_2^2 + \mu \cdot \partial(\|\mathbf{J}\tilde{\mathbf{b}}\|_1 - c), \quad \mu(\|\mathbf{J}\tilde{\mathbf{b}}\|_1 - c) = 0, \quad (\text{A.6})$$

where $\partial\|\mathbf{Y}\mathbf{a} - \mathbf{X}\tilde{\mathbf{b}}\|_2^2$ [resp. $\partial(\|\mathbf{J}\tilde{\mathbf{b}}\|_1 - c)$] in (A.6) denotes the subdifferential of $\|\mathbf{Y}\mathbf{a} - \mathbf{X}\mathbf{b}\|_2^2$ [resp. $(\|\mathbf{J}\mathbf{b}\|_1 - c)$] at $\mathbf{b} = \tilde{\mathbf{b}} (= [\tilde{b}_1, \dots, \tilde{b}_T]')$. More precisely, (i) $\partial\|\mathbf{Y}\mathbf{a} - \mathbf{X}\tilde{\mathbf{b}}\|_2^2 = \{-2\mathbf{X}'(\mathbf{Y}\mathbf{a} - \mathbf{X}\tilde{\mathbf{b}})\}$, which is a singleton because $\|\mathbf{Y}\mathbf{a} - \mathbf{X}\mathbf{b}\|_2^2$ is a differentiable function of $\mathbf{b} \in \mathbb{R}^T$, and (ii) by Proposition B.24(e) in Bertsekas (1999) and Osborne et al. (2000), $\partial\|\mathbf{J}\tilde{\mathbf{b}}\|_1 = \mathbf{J}'\tilde{\boldsymbol{\xi}}$, where $\tilde{\boldsymbol{\xi}} = [\tilde{\xi}_3, \dots, \tilde{\xi}_T]'$. Here, $\tilde{\xi}_t = \{1\}$ if $\tilde{b}_t > 0$, $\tilde{\xi}_t = \{-1\}$ if $\tilde{b}_t < 0$, and $\tilde{\xi}_t = [-1, 1]$ if $\tilde{b}_t = 0$, for $t = 3, \dots, T$.

Accordingly, (A.6) can be represented as $-2X'(Ya - X\tilde{b}) + \mu J'\tilde{v} = \mathbf{0}$, where $\tilde{v} = [\tilde{v}_3, \dots, \tilde{v}_T]'$ and $\tilde{v}_t$ is an element of $\tilde{\xi}_t$ for $t = 3, \dots, T$.

A.5.2 Proof of Lemma A.2

Let $\kappa = \tilde{b} - X^{-1}Ya$. Given $c < \|DYa\|_1$, $X^{-1}Ya$ is infeasible. Accordingly, if $c < \|DYa\|_1$, then we have $\kappa \neq \mathbf{0}$, which leads to $X\kappa = X\tilde{b} - Ya \neq \mathbf{0}$.

References

- Bertsekas, D. P. (1999). *Nonlinear programming* (2nd ed.). Belmont: Athena Scientific.
- Brantley, H. L., Guinness, J., & Chi, E. C. (2019). Baseline drift estimation for air quality data using quantile trend filtering, forthcoming. In *Annals of Applied Statistics*.
- Chen, L., & Huang, J. Z. (2012). Sparse reduced-rank regression for simultaneous dimension reduction and variable selection. *Journal of the American Statistical Association*, 107, 1533–1545.
- Cornea-Madeira, A. (2017). The explicit formula for the Hodrick–Prescott filter in a finite sample. *Review of Economics and Statistics*, 99(2), 314–318.
- de Jong, R. M., & Sakarya, N. (2016). The econometrics of the Hodrick–Prescott filter. *Review of Economics and Statistics*, 98(2), 310–317.
- Engle, R. F., & Granger, C. W. J. (1987). Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2), 251–276.
- Guntuboyina, A., Lieu, D., Chatterjee, S., & Sen, B. (2020). Adaptive risk bounds in univariate total variation denoising and trend filtering. *Annals of Statistics*, 48(1), 205–229.
- Hamilton, J. D. (1994). *Time series analysis*. Princeton: Princeton University Press.
- Hamilton, J. D. (2018). Why you should never use the Hodrick–Prescott filter. *Review of Economics and Statistics*, 100, 831–843.
- Hatanaka, M. (1996). *Time-series-based econometrics: Unit roots and co-integrations*. New York: Oxford University Press.
- Hatanaka, M., & Yamada, H. (2003). *Co-trending: A statistical system analysis of economic trends*. Tokyo: Springer.
- Hodrick, R. J., & Prescott, E. C. (1997). Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29(1), 1–16.
- Kim, S., Koh, K., Boyd, S., & Gorinevsky, D. (2009). ℓ_1 trend filtering. *SIAM Review*, 51(2), 339–360.
- Klein, T. (2018). Trends and contagion in WTI and Brent crude oil spot and futures markets—the role of OPEC in the last decade. *Energy Economics*, 75, 636–646.
- Khodadadi, A., & McDonald, D. J. (2019). Algorithms for estimating trends in global temperature volatility. (arXiv: 1805.07376v2).
- Michaelides, P. G., Tsonas, E. G., Vouldis, A. T., Konstantakis, K. N., & Patrinos, P. (2018). A semi-parametric non-linear neural network filter: Theory and empirical evidence. *Computational Economics*, 51, 637–675.
- Mitra, S., & Rohit, A. (2018). Momentum trading with the ell_1 -filter: Are the markets efficient?, *International Review of Finance*.
- Osborne, M. R., Presnell, B., & Turlach, B. A. (2000). On the LASSO and its dual. *Journal of Computation and Graphical Statistics*, 9(2), 319–337.
- Paige, R. L., & Trindade, A. A. (2010). The Hodrick–Prescott filter: A special case of penalized spline smoothing. *Electronic Journal of Statistics*, 4, 856–874.
- Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57(6), 1361–1401.
- Phillips, P. C. B. (2010). Two New Zealand pioneer econometricians. *New Zealand Economic Papers*, 44(1), 1–26.
- Phillips, P. C. B., & Jin, S. (2020). Business cycles, trend elimination, and the HP filter. *International Economic Review*, first online 02 December 2020. <https://doi.org/10.1111/ieere.12494>.

- Phillips, P. C. B., & Shi, Z. (2020). Boosting: Why you can use the HP filter, *International Economic Review*, first online 01 December 2020. <https://doi.org/10.1111/iere.12495>.
- Politsch, C. A., Cisewski-Kehe, J., Croft, R. A. C., & Wasserman, L. (2020). Trend filtering - I. A modern statistical tool for time-domain astronomy and astronomical spectroscopy. *Monthly Notices of the Royal Astronomical Society*, 492(3), 4005–4018.
- Pollock, D. S. G. (2016). Econometric filters. *Computational Economics*, 48, 669–691.
- Rappoport, P., & Reichlin, L. (1989). Segmented trends and non-stationary time series. *Economic Journal*, 99, 168–177.
- Rudin, L. I., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60, 259–268.
- Sakarya, N., & de Jong, R. M. (2020). A property of the Hodrick–Prescott filter and its application. *Econometric Theory*, 36(5), 840–870.
- Schechter, M. (1977). A subgradient duality theorem. *Journal of Mathematical Analysis and Applications*, 61(3), 850–855.
- Steidl, G. (2006). A note on the dual treatment of higher-order regularization functionals. *Computing*, 76, 135–148.
- Steidl, G., Didas, S., & Neumann, J. (2006). Splines in higher order TV regularization. *International Journal of Computer Vision*, 70, 241–255.
- Suo, C., Li, Z., Sun, Y., & Han, Y. (2019). Application of L1 trend filtering technology on the current time domain spectroscopy of dielectrics. *Electronics*, 8(9), 1046.
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B*, 58(1), 267–288.
- Tibshirani, R. J. (2014). Adaptive piecewise polynomial estimation via trend filtering. *Annals of Statistics*, 42(1), 285–323.
- Tibshirani, R. J., & Taylor, J. (2011). The solution path of the generalized lasso. *Annals of Statistics*, 39(3), 1335–1371.
- Weinert, H. L. (2007). Efficient computation for Whittaker–Henderson smoothing. *Computational Statistics and Data Analysis*, 52(2), 959–974.
- Weir, T. (1988). Subgradient duality using Fritz John conditions. *Journal of Information and Optimization Sciences*, 9(2), 287–296.
- Winkelried, D. (2016). Piecewise linear trends and cycles in primary commodity prices. *Journal of International Money and Finance*, 64, 196–213.
- Wu, D., Yan, H., & Yuan, S. (2018). L1 regularization for detecting offsets and trend change points in GNSS time series. *GPS Solutions*, 22, 88.
- Yamada, H. (2015). Ridge regression representations of the generalized Hodrick–Prescott filter. *Journal of the Japan Statistical Society*, 45(2), 121–128.
- Yamada, H. (2017). Estimating the trend in US real GDP using the ℓ_1 trend filtering. *Applied Economics Letters*, 24(10), 713–716.
- Yamada, H. (2020). A smoothing method that looks like the Hodrick–Prescott filter. *Econometric Theory*, 36(5), 961–981.
- Yamada, H., & Jin, L. (2013). Japan’s output gap estimation and ℓ_1 trend filtering. *Empirical Economics*, 45(1), 81–88.
- Yamada, H., & Yoon, G. (2014). When Grilli and Yang meet Prebisch and Singer: Piecewise linear trends in primary commodity prices. *Journal of International Money and Finance*, 42, 193–207.