*Article*

# Some Results on $\ell_1$ Polynomial Trend Filtering

**Hiroshi Yamada * and Ruixue Du**

Graduate School of Social Sciences, Hiroshima University, 1-2-1 Kagamiyama, Higashi-Hiroshima 739-8525, Japan; tozuisetsu@yahoo.co.jp
* Correspondence: yamada@hiroshima-u.ac.jp; Tel.: +81-82-424-7214; Fax: +81-82-424-7212

check for
updates

**Abstract:** $\ell_1$ polynomial trend filtering, which is a filtering method described as an $\ell_1$-norm penalized least-squares problem, is promising because it enables the estimation of a piecewise polynomial trend in a univariate economic time series without prespecifying the number and location of knots. This paper shows some theoretical results on the filtering, one of which is that a small modification of the filtering provides not only identical trend estimates as the filtering but also extrapolations of the trend beyond both sample limits.

## 1. Introduction

The $\ell_1$-norm penalized least-squares problem, defined as:

$$\min_{x_1,\dots,x_T} \sum_{t=1}^{T} (y_t - x_t)^2 + \lambda \sum_{t=3}^{T} |\Delta^2 x_t|, \tag{1}$$

where $y_1, \dots, y_T$ are observed time-series data, was developed by Kim et al. (2009), who called it $\ell_1$ trend filtering.[1] Here, $\lambda > 0$ is a tuning parameter and $\Delta$ denotes the backward difference operator such that $\Delta x_t = x_t - x_{t-1}$. Accordingly, $\Delta^2 x_t = \Delta(\Delta x_t) = x_t - 2x_{t-1} + x_{t-2}$. Recall that $\sum_{t=3}^{T} |\Delta^2 x_t|$ in (1) is $\ell_1$-norm of $[\Delta^2 x_3, \dots, \Delta^2 x_T]^\top$. Unlike Hodrick and Prescott (1997) filtering, which is defined as the following squared $\ell_2$-norm penalized least-squares problem:

$$\min_{x_1,\dots,x_T} \sum_{t=1}^{T} (y_t - x_t)^2 + \psi \sum_{t=3}^{T} (\Delta^2 x_t)^2, \tag{2}$$

where $\psi > 0$ is a smoothing/tuning parameter, the solution of $\ell_1$ trend filtering becomes a continuous piecewise linear trend. The relationship between HP filtering and $\ell_1$ trend filtering corresponds to that between ridge regression of Hoerl and Kennard (1970) and Lasso (least absolute shrinkage and selection operator) regression of Tibshirani (1996)/BPDN (basis pursuit denoising) of Chen et al. (1998). Econometric applications of $\ell_1$ trend filtering include Yamada and Jin (2013), Yamada and Yoon (2014), Winkelried (2016), and Yamada (2017a).

---

[1] $\ell_1$ trend filtering is supported in several standard software packages such as MATLAB, R, Python, and EViews.

It has been well-known that HP filtering is a form of the Whittaker–Henderson (WH) method of graduation, which is defined as:

$$\min_{x_1,\ldots,x_T} \sum_{t=1}^{T}(y_t - x_t)^2 + \psi \sum_{t=p+1}^{T}(\Delta^p x_t)^2. \tag{3}$$

For historical surveys of WH filtering, see Weinert (2007), Phillips (2010), and Nocon and Scott (2012). Likewise, as shown in Kim et al. (2009), Tibshirani and Taylor (2011), and Tibshirani (2014), $\ell_1$ trend filtering may be generalized as:

$$\min_{x_1,\ldots,x_T} \sum_{t=1}^{T}(y_t - x_t)^2 + \lambda \sum_{t=p+1}^{T}|\Delta^p x_t|. \tag{4}$$

We refer to it as $\ell_1$ polynomial trend filtering.[2] This filtering method is promising because it enables us to estimate a piecewise $(p-1)$-th order polynomial trend of a univariate economic time series without prespecifying the number and location of knots. For more details, see Yamada (2017b).

Let $\hat{x}_1,\ldots,\hat{x}_T$ denote the solution of (3) and define $\hat{x}_{T+1},\ldots,\hat{x}_{T+h}$, where $h$ denotes the length of extrapolation by:

$$\Delta^p \hat{x}_{T+j} = 0, \quad (j = 1,\ldots,h). \tag{5}$$

Recently, Yamada and Du (2018) introduced the following three modifications of the WH method of graduation:[3]

$$\text{(a)} \min_{x_1,\ldots,x_{T+h}} \sum_{t=1}^{T}(y_t - x_t)^2 + \psi \sum_{t=p+1}^{T+h}(\Delta^p x_t)^2, \tag{6}$$

$$\text{(b)} \min_{x_1,\ldots,x_{T+h}} \sum_{t=1}^{T+h}(y_t - x_t)^2 + \psi \sum_{t=p+1}^{T+h}(\Delta^p x_t)^2, \tag{7}$$

$$\text{(c)} \min_{x_1,\ldots,x_{T+h}} \sum_{t=1}^{T+h}(y_t - x_t)^2 + \psi \sum_{t=p+1}^{T}(\Delta^p x_t)^2, \tag{8}$$

---

[2]  (4) where $p = 1$ has been known as total variation denoising in signal processing, which may be regarded as a form of the fused Lasso by Tibshirani et al. (2005). Harchaoui and Lévy-Leduc (2010) proposed using the filtering to detect multiple change points. (4) may be regarded as a form of the generalized Lasso by Tibshirani and Taylor (2011). In addition, we note that there exist some pioneering works on the filtering that uses the $\ell_1$-norm penalty. (Miller 1946, sct. 1.7) mentioned that $\sum_{t=p+1}^{T}|\Delta^p x_t|$ could be an alternative measure of smoothness to $\sum_{t=p+1}^{T}(\Delta^p x_t)^2$, Schuette (1978) introduced a filtering, defined as:

$$\min_{x_1,\ldots,x_T} \sum_{t=1}^{T}|y_t - x_t| + \lambda \sum_{t=p+1}^{T}|\Delta^p x_t|,$$

and Koenker et al. (1994) presented $\ell_1$-norm penalized quantile smoothing spline. Incidentally, Schuette (1978) and Koenker et al. (1994) motivate us to consider a penalized quantile regression that is obtainable by replacing the quadratic loss function in (4) by the check loss function:

$$\min_{x_1,\ldots,x_T} \sum_{t=1}^{T}\rho_\tau(y_t - x_t) + \lambda \sum_{t=p+1}^{T}|\Delta^p x_t|,$$

where, letting $\tau \in (0,1)$,

$$\rho_\tau(u) = \begin{cases} \tau|u| & (u \geq 0), \\ (1-\tau)|u| & (u < 0), \end{cases}$$

which is suggested by (Kim et al. 2009, sct. 7.3).

[3]  See also Yamada (2017c).

where $y_{T+j} = \widehat{x}_{T+j}$ for $j = 1, \ldots, h$. Denote the solution of (a), (b), and (c) by $\widehat{x}_t^{(i)}$ for $i = a, b, c$ and $t = 1, \ldots, T + h$. Yamada and Du (2018) showed that, for $i = a, b, c$ and $t = 1, \ldots, T + h$, it follows that:

$$\widehat{x}_t^{(i)} = \widehat{x}_t. \tag{9}$$

Among the above results, $\widehat{x}_t^{(a)} = \widehat{x}_t$ is of practical use because it provides not only a smoothed series identical to that of the WH graduation, but also an extrapolation beyond the sample limit of current data. Also, $\widehat{x}_t^{(b)} = \widehat{x}_t$ is of interest because it shows that $\widehat{x}_{T+1}, \ldots, \widehat{x}_{T+h}$ based on (5) are useless to reduce the end-point problem of the WH graduation.[4] In addition, Yamada and Du (2018) proved that, for $i = a, b, c$ and $t = 1, \ldots, T + h$:

$$\lim_{\psi \to \infty} \widehat{x}_t^{(i)} = \widehat{\beta}_0 t^0 + \cdots + \widehat{\beta}_{p-1} t^{p-1}, \tag{10}$$

where $(\widehat{\beta}_0, \ldots, \widehat{\beta}_{p-1}) = \arg\min_{\beta_0, \ldots, \beta_{p-1}} \sum_{t=1}^{T} (y_t - \beta_0 t^0 - \cdots - \beta_{p-1} t^{p-1})^2$.

In this paper, we present three modifications of $\ell_1$ polynomial trend filtering and show that they provide not only identical trend estimates as $\ell_1$ polynomial trend filtering, but also extrapolations of the trend beyond both sample limits. In addition, we show some other results on the modified filtering. We also provide a MATLAB function for calculating the solution of one of the modified filtering methods.

The paper is organized as follows. In Section 2, we present three modifications of $\ell_1$ polynomial trend filtering. In Section 3, we state the main results of the paper. In Section 4, we make some remarks on the results provided in Section 3. Section 5 provides some concluding remarks.

*Notation.* Let $\mathbf{y} = [y_1, \ldots, y_T]^\top$ and $\mathbf{I}_T$ be the $T \times T$ identity matrix. For an $n$-dimensional column vector, $\boldsymbol{\eta} = [\eta_1, \ldots, \eta_n]^\top$, $\|\boldsymbol{\eta}\|_1 = \sum_{i=1}^{n} |\eta_i|$, $\|\boldsymbol{\eta}\|_2^2 = \sum_{i=1}^{n} \eta_i^2$, and $\|\boldsymbol{\eta}\|_\infty = \max(|\eta_1|, \ldots, |\eta_n|)$. $\mathbf{D}_n$ is the $(n - p) \times n$ $p$-th order difference matrix such that $\mathbf{D}_n \boldsymbol{\eta} = [\Delta^p \eta_{p+1}, \ldots, \Delta^p \eta_n]^\top$. We denote $\mathbf{D}_T$ by $\mathbf{D}$. $\boldsymbol{\Pi}_{g+T+h}$ is a $(g + T + h) \times p$ Vandermonde matrix, defined by

$$\boldsymbol{\Pi}_{g+T+h} = \begin{bmatrix} (1-g)^0 & (1-g)^1 & \cdots & (1-g)^{p-1} \\ \vdots & \vdots & & \vdots \\ 1^0 & 1^1 & \cdots & 1^{p-1} \\ \vdots & \vdots & & \vdots \\ T^0 & T^1 & \cdots & T^{p-1} \\ \vdots & \vdots & & \vdots \\ (T+h)^0 & (T+h)^1 & \cdots & (T+h)^{p-1} \end{bmatrix},$$

and we denote $\boldsymbol{\Pi}_{0+T+0}$, which is a $T \times p$ matrix, by $\boldsymbol{\Pi}$.

## 2. Three Modifications of $\ell_1$ Polynomial Trend Filtering

Let $\widetilde{x}_1, \ldots, \widetilde{x}_T$ denote the solution of (4) and define $\widetilde{x}_{1-g}, \ldots, \widetilde{x}_{1-1}$ and $\widetilde{x}_{T+1}, \ldots, \widetilde{x}_{T+h}$, where $g$ and $h$ denote the length of extrapolations:

$$\Delta^p \widetilde{x}_{p+1-i} = 0, \quad (i = 1, \ldots, g), \tag{11}$$

$$\Delta^p \widetilde{x}_{T+j} = 0, \quad (j = 1, \ldots, h). \tag{12}$$

---

4    An argument similar to this is given by (Mohr 2005, p. 20).

For example, $\widetilde{x}_{T+1}, \ldots, \widetilde{x}_{T+h}$, defined by (12) for $p = 1, 2, 3$, are explicitly expressed as follows:

$$(p = 1) \quad \widetilde{x}_{T+j} = \widetilde{x}_T, \quad (j = 1, \ldots, h), \tag{13}$$

$$(p = 2) \quad \widetilde{x}_{T+j} = \widetilde{x}_T + j(\Delta \widetilde{x}_T), \quad (j = 1, \ldots, h), \tag{14}$$

$$(p = 3) \quad \widetilde{x}_{T+j} = \widetilde{x}_T + j(\Delta \widetilde{x}_T) + \frac{j(j+1)}{2}(\Delta^2 \widetilde{x}_T), \quad (j = 1, \ldots, h). \tag{15}$$

For a proof of (15), see the Appendix A.

Consider the following three modifications of $\ell_1$ polynomial trend filtering:

$$(d) \quad \min_{x_{1-g}, \ldots, x_{T+h}} \sum_{t=1}^{T} (y_t - x_t)^2 + \lambda \sum_{t=p+1-g}^{T+h} |\Delta^p x_t|, \tag{16}$$

$$(e) \quad \min_{x_{1-g}, \ldots, x_{T+h}} \sum_{t=1-g}^{T+h} (y_t - x_t)^2 + \lambda \sum_{t=p+1-g}^{T+h} |\Delta^p x_t|, \tag{17}$$

$$(f) \quad \min_{x_{1-g}, \ldots, x_{T+h}} \sum_{t=1-g}^{T+h} (y_t - x_t)^2 + \lambda \sum_{t=p+1}^{T} |\Delta^p x_t|, \tag{18}$$

where $y_{1-i} = \widetilde{x}_{1-i}$ for $i = 1, \ldots, g$ and $y_{T+j} = \widetilde{x}_{T+j}$ for $j = 1, \ldots, h$. Note that (16) is equivalent to $\ell_1$ polynomial trend filtering if $g = h = 0$. We denote the solution of (d), (e), and (f) by $\widetilde{x}_t^{(i)}$ for $i = d, e, f$ and $t = 1 - g, \ldots, T + h$.

Among (16)–(18), the objective function of (16) may be represented in matrix notation as:

$$\|y - Sx_{g+T+h}\|_2^2 + \lambda \|D_{g+T+h} x_{g+T+h}\|_1, \tag{19}$$

where $S = [0, I_T, 0]$ is a $T \times (g + T + h)$ matrix and $x_{g+T+h}$ is a $(g + T + h)$-dimensional column vector. Let $\widetilde{x}_{g+T+h}^{(d)} = [\widetilde{x}_g^{(d)\top}, \widetilde{x}^{(d)\top}, \widetilde{x}_h^{(d)\top}]^\top$, where $\widetilde{x}_g^{(d)} = [\widetilde{x}_{1-g}^{(d)}, \ldots, \widetilde{x}_{1-1}^{(d)}]^\top$, $\widetilde{x}^{(d)} = [\widetilde{x}_1^{(d)}, \ldots, \widetilde{x}_T^{(d)}]^\top$, and $\widetilde{x}_h^{(d)} = [\widetilde{x}_{T+1}^{(d)}, \ldots, \widetilde{x}_{T+h}^{(d)}]^\top$. The MATLAB function for calculating $\widetilde{x}_g^{(d)}$, $\widetilde{x}^{(d)}$, and $\widetilde{x}_h^{(d)}$, which depends on CVX developed by Grant and Boyd (2013), is as follows:

```
function [x_g,x,x_h]=m_l1_pt_filtering(y,lambda,p,g,h)
% y: T-dimensional column vector
% lambda: positive real number
% p, g, h: positive integer
% x_g: g-dimensional column vector
% x: T-dimensional column vector
% x_h: h-dimensional column vector
   T=length(y);
   S=[sparse(T,g),speye(T),sparse(T,h)];
   D=diff(speye(g+T+h),p);
   cvx_begin
      variables z(g+T+h)
      minimize(sum((y-S*z).^2)+lambda*norm(D*z,1))
   cvx_end
   x_g=z(1:g); x=z(g+1:g+T); x_h=z(g+T+1:g+T+h);
end
```

## 3. Main Results

**Theorem 1.** *Denote the solution of (d), (e), and (f) by $\widetilde{x}_t^{(i)}$ for $i = \mathrm{d}, \mathrm{e}, \mathrm{f}$. For $i = \mathrm{d}, \mathrm{e}, \mathrm{f}$, and $t = 1 - g, \dots, T + h$, it follows that:*

$$\widetilde{x}_t^{(i)} = \widetilde{x}_t, \tag{20}$$

*where $\widetilde{x}_1, \dots, \widetilde{x}_T$ are the solution of* (4) *and $\widetilde{x}_{1-g}, \dots, \widetilde{x}_{1-1}$ and $\widetilde{x}_{T+1}, \dots, \widetilde{x}_{T+h}$ are defined by* (11) *and* (12).

**Proof.** Because the objective function of (4) is coercive and strictly convex with respect to $x_1, \dots, x_T$, $\widetilde{x}_1, \dots, \widetilde{x}_T$ are the unique global minimizer of the function. It follows that:

$$\sum_{t=1}^{T}(y_t - x_t)^2 + \lambda \sum_{t=p+1}^{T} |\Delta^p x_t| \geq \sum_{t=1}^{T}(y_t - \widetilde{x}_t)^2 + \lambda \sum_{t=p+1}^{T} |\Delta^p \widetilde{x}_t|, \tag{21}$$

where the equality holds only if $x_t = \widetilde{x}_t$ for $t = 1, \dots, T$.[5] In addition, from (11) and (12), $y_{1-i} = \widetilde{x}_{1-i}$ for $i = 1, \dots, g$, and $y_{T+j} = \widetilde{x}_{T+j}$ for $j = 1, \dots, h$, we have the following inequalities:

$$\lambda \sum_{t=p+1-g}^{p+1-1} |\Delta^p x_t| \geq 0 = \lambda \sum_{t=p+1-g}^{p+1-1} |\Delta^p \widetilde{x}_t|, \tag{22}$$

$$\lambda \sum_{t=T+1}^{T+h} |\Delta^p x_t| \geq 0 = \lambda \sum_{t=T+1}^{T+h} |\Delta^p \widetilde{x}_t|, \tag{23}$$

$$\sum_{t=1-g}^{1-1} (y_t - x_t)^2 \geq 0 = \sum_{t=1-g}^{1-1} (y_t - \widetilde{x}_t)^2, \tag{24}$$

$$\sum_{t=T+1}^{T+h} (y_t - x_t)^2 \geq 0 = \sum_{t=T+1}^{T+h} (y_t - \widetilde{x}_t)^2. \tag{25}$$

Combining (21)–(23) yields

$$\sum_{t=1}^{T}(y_t - x_t)^2 + \lambda \sum_{t=p+1-g}^{T+h} |\Delta^p x_t| \geq \sum_{t=1}^{T}(y_t - \widetilde{x}_t)^2 + \lambda \sum_{t=p+1-g}^{T+h} |\Delta^p \widetilde{x}_t|, \tag{26}$$

where the equality in (26) holds only if $x_t = \widetilde{x}_t$ for $t = 1 - g, \dots, T + h$, which proves that $\widetilde{x}_t^{(\mathrm{d})} = \widetilde{x}_t$ for $t = 1 - g, \dots, T + h$. Likewise, combining (21)–(25) proves that $\widetilde{x}_t^{(\mathrm{e})} = \widetilde{x}_t$ for $t = 1 - g, \dots, T + h$ and combining (21), (24) and (25) proves that $\widetilde{x}_t^{(\mathrm{f})} = \widetilde{x}_t$ for $t = 1 - g, \dots, T + h$. □

As an illustration of the above theorem, we give a numerical example. Consider the case where $T = 5$, $g = 1$, and $h = 2$. Suppose that we obtained

$$\widetilde{x}_1 = 3, \quad \Delta \widetilde{x}_2 = 2, \quad [\Delta^2 \widetilde{x}_3, \Delta^2 \widetilde{x}_4, \Delta^2 \widetilde{x}_5]^\top = [0, -1, 0]^\top$$

by applying $\ell_1$ polynomial trend filtering of order 2 (i.e., $\ell_1$ trend filtering) to a $T$-dimensional time-series data.[6] Because $2 = \Delta \widetilde{x}_2 = \Delta \widetilde{x}_3 \neq \Delta \widetilde{x}_4 = \Delta \widetilde{x}_5 = 1$, the line plot of $(t, \widetilde{x}_t)$ for $t = 1, \dots, 5$ becomes a continuous piecewise linear line such that $(3, \widetilde{x}_3)$ is a knot. $\widetilde{x}_t$ for $t = 1, \dots, 5$ are explicitly

---

[5] In the objective function of (4), $\sum_{t=1}^{T}(y_t - x_t)^2$ is coercive because it is a quadratic function whose Hessian matrix is positive definite. See, e.g., (Beck 2014, Lemma 2.42).

[6] In the case, $[\Delta^2 \widetilde{x}_3, \Delta^2 \widetilde{x}_4, \Delta^2 \widetilde{x}_5]^\top$ is expected to become sparse, as in the numerical example, because $\sum_{i=3}^{5} |\Delta^2 x_t|$ is included as a penalty.

$[\widetilde{x}_1, \widetilde{x}_2, \widetilde{x}_3, \widetilde{x}_4, \widetilde{x}_5]^\top = [3, 5, 7, 8, 9]^\top$. Then, from the above theorem, in the case, $\widetilde{x}_t^{(i)}$ for $i = $ d, e, f and $t = 1 - 1, \ldots, 5 + 2$ are as follows:

$$[\widetilde{x}_{1-1}^{(i)}, \widetilde{x}_1^{(i)}, \widetilde{x}_2^{(i)}, \widetilde{x}_3^{(i)}, \widetilde{x}_4^{(i)}, \widetilde{x}_5^{(i)}, \widetilde{x}_{5+1}^{(i)}, \widetilde{x}_{5+2}^{(i)}]^\top = [1, 3, 5, 7, 8, 9, 10, 11]^\top.$$

**Theorem 2.** *If $\lambda \geq 2\|(DD^\top)^{-1}Dy\|_\infty$, for $i = $ d, e, f and $t = 1 - g, \ldots, T + h$, it follows that*

$$\widetilde{x}_t^{(i)} = \widehat{\beta}_0 t^0 + \cdots + \widehat{\beta}_{p-1} t^{p-1}, \tag{27}$$

*where $(\widehat{\beta}_0, \ldots, \widehat{\beta}_{p-1}) = \arg\min_{\beta_0, \ldots, \beta_{p-1}} \sum_{t=1}^T (y_t - \beta_0 t^0 - \cdots - \beta_{p-1} t^{p-1})^2$.*

**Proof.** Because $D_{g+T+h}$ is a $(g + T + h - p) \times (g + T + h)$ $(p + 1)$-diagonal Toeplitz matrix, such that:

$$D_{g+T+h} = \begin{bmatrix} a_0 & \cdots & a_p & 0 & \cdots & 0 \\ 0 & \ddots & & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & & \ddots & 0 \\ 0 & \cdots & 0 & a_0 & \cdots & a_p \end{bmatrix},$$

where $a_k = (-1)^{p-k}\binom{p}{k}$ for $k = 0, \ldots, p$, it may be expressed as

$$D_{g+T+h} = \begin{bmatrix} G_1 & G_2 & 0 \\ 0 & D & 0 \\ 0 & H_1 & H_2 \end{bmatrix},$$

where $G_1$ is a $g \times g$ upper triangular matrix, $G_2$ is a $g \times T$ matrix, $H_1$ is an $h \times T$ matrix, and $H_2$ is an $h \times h$ unit lower-triangular matrix. For example, when $p = 3$, $g = h = 2$, and $T = 5$:

$$D_{2+5+2} = \begin{bmatrix} -1 & 3 & -3 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 3 & -3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 3 & -3 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 3 & -3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 3 & -3 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & -1 & 3 & -3 & 1 \end{bmatrix}. \tag{28}$$

Let $\widetilde{x}_g = [\widetilde{x}_{1-g}, \ldots, \widetilde{x}_{1-1}]^\top$, $\widetilde{x} = [\widetilde{x}_1, \ldots, \widetilde{x}_T]^\top$, $\widetilde{x}_h = [\widetilde{x}_{T+1}, \ldots, \widetilde{x}_{T+h}]^\top$, and $\widetilde{x}_{g+T+h} = [\widetilde{x}_g^\top, \widetilde{x}^\top, \widetilde{x}_h^\top]^\top$, which is a $(g + T + h)$-dimensional column vector. Then, by definition of $\widetilde{x}_g$ and $\widetilde{x}_h$, it follows that:

$$G_1 \widetilde{x}_g + G_2 \widetilde{x} = 0, \tag{29}$$
$$H_1 \widetilde{x} + H_2 \widetilde{x}_h = 0, \tag{30}$$

which leads to:

$$D_{g+T+h}\widetilde{x}_{g+T+h} = \begin{bmatrix} 0 \\ D\widetilde{x} \\ 0 \end{bmatrix}. \tag{31}$$

From Kim et al. (2009), if $\lambda \geq 2\|(DD^\top)^{-1}Dy\|_\infty$, it follows that $\widetilde{x} = \Pi\widehat{\beta}$, where $\widehat{\beta} = (\Pi^\top\Pi)^{-1}\Pi^\top y$. Recalling that $D\Pi = 0$, we obtain $D_{g+T+h}\widetilde{x}_{g+T+h} = 0$ if $\lambda \geq 2\|(DD^\top)^{-1}Dy\|_\infty$, which implies that $\widetilde{x}_{g+T+h}$ may be represented as $\Pi_{g+T+h}\gamma$. Because $\widetilde{x} = \Pi\widehat{\beta}$, $\gamma$ must equal $\widehat{\beta}$. Therefore, if $\lambda \geq 2\|(DD^\top)^{-1}Dy\|_\infty$, then $\widetilde{x}_{g+T+h} = \Pi_{g+T+h}\widehat{\beta}$. $\square$

**Theorem 3.** *Suppose that* $y = \Pi\alpha$*, where* $\alpha \neq 0$ *is a p-dimensional column vector. Then, for* $i = d, e, f$*, it follows that:*

$$\widetilde{x}^{(i)}_{g+T+h} = \Pi_{g+T+h}\alpha, \tag{32}$$

*where* $\widetilde{x}^{(i)}_{g+T+h} = [\widetilde{x}^{(i)}_{1-g}, \ldots, \widetilde{x}^{(i)}_{T+h}]^\top$.

**Proof.** If $y = \Pi\alpha$, it follows that: $\widetilde{x} = \Pi\alpha$. Accordingly, $D_{g+T+h}\widetilde{x}_{g+T+h} = 0$, which indicates that $\widetilde{x}_{g+T+h}$ may be represented as $\Pi_{g+T+h}\gamma$. Because $\widetilde{x} = \Pi\alpha$ if $y = \Pi\alpha$, $\gamma$ must equal $\alpha$. Therefore, we obtain $\widetilde{x}_{g+T+h} = \Pi_{g+T+h}\alpha$ if $y = \Pi\alpha$. $\square$

**Corollary 1.** *Let* $\widetilde{x}^{(i)}_{g+T+h} = [\widetilde{x}^{(i)}_{1-g}, \ldots, \widetilde{x}^{(i)}_{T+h}]^\top$ *for* $i = d, e, f$.

(i) *Denote the* $(j+1)$*-th column of* $\Pi$ *and that of* $\Pi_{g+T+h}$*, respectively, by* $\tau_j$ *and by* $\tau_{g+T+h,j}$ *for* $j = 0, \ldots, p-1$*. If* $y = \tau_j$*, then* $\widetilde{x}^{(i)}_{g+T+h} = \tau_{g+T+h,j}$ *for any* $\lambda > 0$*.*

(ii) *Let* $z$ *be a T-dimensional column vector. If* $y = \Pi(\Pi^\top\Pi)^{-1}\Pi^\top z$*, then* $\widetilde{x}^{(i)}_{g+T+h} = \Pi_{g+T+h}(\Pi^\top\Pi)^{-1}\Pi^\top z$ *for any* $\lambda > 0$*.*

## 4. Some Remarks on the Main Results

First, we make a remark on Theorem 1. Because $|G_1| = (-1)^{g \cdot p}$, from (29), $\widetilde{x}_g$ may be expressed with $\widetilde{x}$ as $\widetilde{x}_g = -G_1^{-1}G_2\widetilde{x}$. Likewise, because $|H_2| = 1$, from (30), $\widetilde{x}_h$ may be expressed with $\widetilde{x}$ as $\widetilde{x}_h = -H_2^{-1}H_1\widetilde{x}$. Thus, the modified $\ell_1$ polynomial trend filtering, (16), may be characterized as a filtering that calculates

$$\begin{bmatrix} -G_1^{-1}G_2 \\ I_T \\ -H_2^{-1}H_1 \end{bmatrix} \widetilde{x} \tag{33}$$

from $y$.[7] In addition, from Kim et al. (2009), it follows that $\widetilde{x} \to y$ as $\lambda \to 0$. Therefore, we obtain:

$$\widetilde{x}^{(d)}_{g+T+h} \to \begin{bmatrix} -G_1^{-1}G_2 \\ I_T \\ -H_2^{-1}H_1 \end{bmatrix} y, \quad (\lambda \to 0). \tag{34}$$

Second, we provide a remark on Theorems 2 and 3. Yamada (2017b) recently showed that:

$$\widetilde{x} = \Pi\widehat{\beta} + X\widehat{\phi}, \tag{35}$$

---

[7] Let us calculate $-H_2^{-1}H_1\widetilde{x}$ for the case where $p = 3$, $g = h = 2$, and $T = 5$. From (28), it follows that

$$-H_1\widetilde{x} = \begin{bmatrix} \widetilde{x}_{T-2} - 3\widetilde{x}_{T-1} + 3\widetilde{x}_T \\ \widetilde{x}_{T-1} - 3\widetilde{x}_T \end{bmatrix} = \begin{bmatrix} \widetilde{x}_T + (\Delta\widetilde{x}_T) + (\Delta^2\widetilde{x}_T) \\ -2\widetilde{x}_T - (\Delta\widetilde{x}_T) \end{bmatrix}.$$

Accordingly, we obtain:

$$-H_2^{-1}H_1\widetilde{x} = \begin{bmatrix} 1 & 0 \\ 3 & 1 \end{bmatrix} \begin{bmatrix} \widetilde{x}_T + (\Delta\widetilde{x}_T) + (\Delta^2\widetilde{x}_T) \\ -2\widetilde{x}_T - (\Delta\widetilde{x}_T) \end{bmatrix} = \begin{bmatrix} \widetilde{x}_T + (\Delta\widetilde{x}_T) + (\Delta^2\widetilde{x}_T) \\ \widetilde{x}_T + 2(\Delta\widetilde{x}_T) + 3(\Delta^2\widetilde{x}_T) \end{bmatrix},$$

which is consistent with (15).

where $X = D^\top (DD^\top)^{-1}$ and $\widetilde{\phi}$, which is a $(T - p)$-dimensional column vector, is the solution of the following Lasso regression/BPDN:

$$\min_{\phi} \|y - X\phi\|_2^2 + \lambda\|\phi\|_1. \tag{36}$$

Because $X^\top \Pi = 0$, $\Pi\widehat{\beta} + X\widetilde{\phi}$ in (35) represents an orthogonal decomposition of $\widetilde{x}$. Here, we show that we may prove Theorems 2 and 3 by using (35) and (36). Premultiplying (35) by $D$ yields $D\widetilde{x} = \widetilde{\phi}$. We accordingly obtain:

$$D_{g+T+h}\widetilde{x}_{g+T+h} = \begin{bmatrix} \mathbf{0} \\ \widetilde{\phi} \\ \mathbf{0} \end{bmatrix}. \tag{37}$$

(i) From (Osborne et al. 2000, p. 324), if $\lambda \geq 2\|X^\top y\|_\infty$, then $\widetilde{\phi} = \mathbf{0}$. Therefore, we obtain $\widetilde{x} = \Pi\widehat{\beta}$ and $D_{g+T+h}\widetilde{x}_{g+T+h} = \mathbf{0}$, which proves Theorem 2.

(ii) If $y = \Pi\alpha$, where $\alpha \neq \mathbf{0}$, then $X^\top y = \mathbf{0}$, which implies that $\lambda > 2\|X^\top y\|_\infty = 0$. Again, from Osborne et al. (2000), we obtain $\widetilde{\phi} = \mathbf{0}$ if $y = \Pi\alpha$. Therefore, if $y = \Pi\alpha$, it follows that $\widetilde{x} = \Pi\widehat{\beta} = \Pi\alpha$ and $D_{g+T+h}\widetilde{x}_{g+T+h} = \mathbf{0}$, which proves Theorem 3.

Third, we give an example of Corollary 1 (i). For the case where $y = [1, \ldots, 5]^\top$ and $p = g = h = 2$, it follows that $\widetilde{x}_{2+5+2}^{(d)} = [-1, 0, 1, \ldots, 5, 6, 7]^\top$ for any $\lambda > 0$.

## 5. Concluding Remarks

The $\ell_1$ polynomial trend filtering method is a promising piecewise polynomial curve-fitting method because it does not require prespecifying the number and location of knots. We have shown some theoretical results on this method. One of them is that a small modification of the filtering provides identical trend estimates and also extrapolations of the trend beyond both sample limits. Another is that $\widetilde{x}_{T+1}, \ldots, \widetilde{x}_{T+h}$ based on (12) are useless to improve the trend estimates of $\ell_1$ polynomial trend filtering. We also provided a MATLAB function for calculating the solution of one of the modified filtering methods. The main results of the paper are summarized in Theorems 1–3 and Corollary 1.

Finally, we remark that applying the modified $\ell_1$ polynomial trend filtering (16)–(18) requires specification of the value of $\lambda$. For this purpose, the methods proposed in Yamada and Yoon (2016) and Yamada (2018) are applicable.

## Appendix A. Proof of (15)

Because $\Delta^3\widetilde{x}_{T+j} = \Delta^2\widetilde{x}_{T+j} - \Delta^2\widetilde{x}_{T+j-1}$, from $\Delta^3\widetilde{x}_{T+j} = 0$ for $j = 1, \ldots, h$, we obtain $\Delta^2\widetilde{x}_{T+k} = \Delta^2\widetilde{x}_T$ for $k = 1, \ldots, h$. Then, because $\sum_{k=1}^l (\Delta^2\widetilde{x}_{T+k}) = l(\Delta^2\widetilde{x}_T)$ for $l = 1, \ldots, h$ and $\sum_{k=1}^l (\Delta^2\widetilde{x}_{T+k}) = \Delta\widetilde{x}_{T+l} - \Delta\widetilde{x}_T$, it follows that

$$\Delta\widetilde{x}_{T+l} = \Delta\widetilde{x}_T + l(\Delta^2\widetilde{x}_T), \quad (l = 1, \ldots, h).$$

Furthermore, because $\sum_{l=1}^{j}(\Delta \widetilde{x}_{T+l}) = j(\Delta \widetilde{x}_T) + (\sum_{l=1}^{j} l)(\Delta^2 \widetilde{x}_T)$ for $j = 1, \ldots, h$ and $\sum_{l=1}^{j}(\Delta \widetilde{x}_{T+l}) = \widetilde{x}_{T+j} - \widetilde{x}_T$, we finally obtain:

$$\widetilde{x}_{T+j} = \widetilde{x}_T + j(\Delta \widetilde{x}_T) + \frac{j(j+1)}{2}(\Delta^2 \widetilde{x}_T), \quad (j = 1, \ldots, h).$$

## References

Beck, Amir. 2014. *Introduction to Nonlinear Optimization Theory, Algorithms, and Applications with MATLAB*. Philadelphia: SIAM.

Chen, Scott Shaobing, David L. Donoho, and Michael A. Saunders. 1998. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing* 20: 33–61. [CrossRef]

Grant, M., and Stephen Boyd. 2013. CVX: Matlab Software for Disciplined Convex Programming, Version 2.0 Beta. Available online: http://cvxr.com/cvx (accessed on 9 July 2018).

Harchaoui, Zaïd, and Céline Lévy-Leduc. 2010. Multiple change-point estimation with a total variation penalty. *Journal of the American Statistical Association* 105: 1480–93. [CrossRef]

Hodrick, Robert J., and Edward C. Prescott. 1997. Postwar U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking* 29: 1–16. [CrossRef]

Hoerl, Arthur E., and Robert W. Kennard. 1970. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* 12: 55–67. [CrossRef]

Kim, Seung-Jean, Kwangmoo Koh, Stephen Boyd, and Dimitry Gorinevsky. 2009. $\ell_1$ trend filtering. *SIAM Review* 52: 339–60. [CrossRef]

Koenker, Roger, Pin Ng, and Stephen Portnoy. 1994. Quantile smoothing splines. *Biometrika* 81: 673–80. [CrossRef]

Miller, Morton D. 1946. *Elements of Graduation*. Philadelphia: Actuarial Society of America and American Institute of Actuaries.

Mohr, Matthias F. 2005. A trend-Cycle(-Season) Filter. European Central Bank Working Paper, No. 499. Frankfurt am Main, Germany: European Central Bank.

Nocon, Alicja S., and William F. Scott. 2012. An extension of the Whittaker–Henderson method of graduation. *Scandinavian Actuarial Journal* 2012: 70–79. [CrossRef]

Osborne, Michael R., Brett Presnell, and Berwin A. Turlach. 2000. On the lasso and its dual. *Journal of Computation and Graphical Statistics* 9: 319–37.

Phillips, Peter C. B. 2010. Two New Zealand pioneer econometricians. *New Zealand Economic Papers* 44: 1–26. [CrossRef]

Schuette, Donald R. 1978. A linear programming approach to graduation. *Transactions of Society of Actuaries* 30: 407–31.

Tibshirani, Robert. 1996. Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B* 58: 267–88.

Tibshirani, Robert, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. 2005. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistics Society: Series B* 67: 91–108. [CrossRef]

Tibshirani, Ryan J., and Jonathan Taylor. 2011. The solution path of the generalized lasso. *Annals of Statistics* 39: 1335–71. [CrossRef]

Tibshirani, Ryan J. 2014. Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics* 42: 285–323. [CrossRef]

Winkelried, Diego. 2016. Piecewise linear trends and cycles in primary commodity prices. *Journal of International Money and Finance* 64: 196–213. [CrossRef]

Weinert, Howard . 2007. Efficient computation for Whittaker–Henderson smoothing. *Computational Statistics and Data Analysis* 52: 959–74. [CrossRef]

Yamada, Hiroshi. 2017a. Estimating the trend in US real GDP using the $\ell_1$ trend filtering. *Applied Economics Letters* 24: 713–16. [CrossRef]

Yamada, Hiroshi. 2017b. A trend filtering method closely related to $\ell_1$ trend filtering. *Empirical Economics*. [CrossRef]

Yamada, Hiroshi. 2017c. A small but practically useful modification to the Hodrick–Prescott filtering: A note. *Communications in Statistics–Theory and Methods* 46: 8430–34. [CrossRef]

Yamada, Hiroshi. 2018. A new method for specifying the tuning parameter of $\ell_1$ trend filtering. *Studies in Nonlinear Dynamics and Econometrics*. [CrossRef]

Yamada, Hiroshi, and Ruixue Du. 2018. A modification of the Whittaker–Henderson method of graduation. *Communications in Statistics–Theory and Methods*, forthcoming.

Yamada, Hiroshi, and Lan Jin. 2013. Japan's output gap estimation and $\ell_1$ trend filtering. *Empirical Economics* 45: 81–88. [CrossRef]

Yamada, Hiroshi, and Gawon Yoon. 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. [CrossRef]

Yamada, Hiroshi, and Gawon Yoon. 2016. Selecting the tuning parameter of the $\ell_1$ trend filter. *Studies in Nonlinear Dynamics and Econometrics* 20: 97–105. [CrossRef]