

RobustTrend: A Huber Loss with a Combined First and Second Order Difference Regularization for Time Series Trend Filtering

Qingsong Wen, Jingkun Gao, Xiaomin Song, Liang Sun and Jian Tan

Machine Intelligence Technology, Alibaba Group, Bellevue, Washington 98004, USA

{qingsong.wen, jingkun.g, xiaomin.song, liang.sun, j.tan}@alibaba-inc.com

Abstract

Extracting the underlying trend signal is a crucial step to facilitate time series analysis like forecasting and anomaly detection. Besides noise signal, time series can contain not only outliers but also abrupt trend changes in real-world scenarios. To deal with these challenges, we propose a robust trend filtering algorithm based on robust statistics and sparse learning. Specifically, we adopt the Huber loss to suppress outliers, and utilize a combination of the first order and second order difference on the trend component as regularization to capture both slow and abrupt trend changes. Furthermore, an efficient method is designed to solve the proposed robust trend filtering based on majorization minimization (MM) and alternative direction method of multipliers (ADMM). We compared our proposed robust trend filter with other nine state-of-the-art trend filtering algorithms on both synthetic and real-world datasets. The experiments demonstrate that our algorithm outperforms existing methods.

1 Introduction

The explosion of time series data generated by the Internet of Things (IoT) and many other sources has made the time series analytics more challenging. Trend filtering, which estimates the underlying trend from the time series, is an important task arising in a variety of real-world applications [Kim *et al.*, 2009; Craigmile and Percival, 2006]. Specifically, we assume the time series consists of the trend component and a more rapidly changing random component. The goal of trend filtering is to extract the trend component accurately and efficiently. The extracted trend component can reveal certain traits of a time series and is crucial in many applications such as time series forecasting and anomaly detection. [?]

Many algorithms have been proposed for trend filtering, including moving average smoother, kernel smoothing, smoothing splines etc [Wu *et al.*, 2007; Wu and Huang, 2009; Fried *et al.*, 2018; Afanasyev and Fedorova, 2016]. Among them, one of the most classical and widely used trend filters is the Hodrick-Prescott (H-P) filtering [Hodrick and Prescott, 1997] and its variants. In the H-P trend filtering, the trend

component is obtained by solving an optimization problem which minimizes the weighted sum of the residual size and the smoothness of the trend component. On the one hand, the extracted trend component is expected to close to the original time series, but meanwhile the smoothness constraint is imposed. Specifically, the sum-of-squares loss function is used to measure the difference between the trend and the original time series, and the smoothness of the trend component is measured using the second order difference of the trend. The drawback of H-P filtering is that it is not robust to outliers. Based on the H-P filtering, the ℓ_1 trend filtering [Kim *et al.*, 2009] is proposed, which basically substitutes the ℓ_1 -norm for the ℓ_2 -norm in the regularization term in H-P trend filtering. Note that the ℓ_1 trend filtering assumes that the underlying trend is piecewise linear, and the kinks, knots, or changes in slope of the estimated trend can be interpreted as abrupt changes or events.

Although the sum-of-squares loss function in both the H-P trend filtering and the ℓ_1 trend filtering enjoys its popularity, it cannot handle heavy tailed error or outliers. One approach to deal with these situations is the least absolute deviation (LAD) loss [Wang *et al.*, 2007] as it mitigates the loss incurred by big errors. However, the LAD loss is not adapted for small errors since it penalizes strongly on the small residuals. In the extreme case where the error is not heavy tailed or there is no outliers, it is less effective than the sum-of-squares loss function. In practice, generally we expect the loss function can adaptively handle these cases. The Huber loss [Huber and Ronchetti, 2009] is a combination of the sum-of-squares loss and the LAD loss, which is quadratic on small errors but grows linearly for large values of errors. As a result, the Huber loss is not only more robust against outliers but also more adaptive for different types of data.

In this paper, besides adopting Huber loss, we propose to use the combination of the first order and the second order difference regularization of the trend component in trend filtering. The regularization based on the first order difference encourages to preserve abrupt trend change, but meanwhile it tends to introduce the staircases for the extracted trend component when slow trend change exists. One way to reducing the staircasing effect is the introduction of higher order difference in the regularizer. In fact, the second order difference can effectively eliminates the artifacts (e.g., staircasing) introduced by the first regularization without smoothing the

time series strongly. Also, it is proved recently that the introduction of the second order difference can achieve comparable results to higher order difference [Papafitsoros and Schönlieb, 2014]. Thus, we consider the combination of the first and second order difference for the best trade-off.

To sum up, in this paper we propose a novel trend filtering method based on the Huber loss and the combination of the first and second order dynamics of the trend component. The resulting optimization problem is solved via an efficient MM-ADMM algorithm. The complexity of the Huber loss makes it challenging to directly apply the ADMM framework [Boyd *et al.*, 2011]. Inspired by the majorization minimization [Sun *et al.*, 2017], we minimize an upper bound of the Huber loss in the ADMM framework. In the experiments, we compare ten different trend filtering algorithms and results on both synthetic and real-world data sets demonstrate the good performance of our trend filtering algorithm. To the best of our knowledge, our empirical comparisons is one of the most extensive and comprehensive studies which includes common popular trend filtering methods, including H-P trend filter, ℓ_1 trend filter, TV denoising filter [Chan *et al.*, 2001], mixed trend filter [Tibshirani, 2014], wavelet trend filter [Craigmile and Percival, 2006], repeated median filter [Siegel, 1982], robfilter [Fried, 2004], EMD trend filter [Moghtaderi *et al.*, 2011], and so on.

The rest of the paper is organized as follows. In Section 2, we briefly review the trend filtering and some related work. In Section 3, We introduce our robust trend filter and our efficient algorithm. We present the empirical studies of the proposed method in comparison with other state-of-the-art algorithms in Section 4, and conclude the paper in Section 5.

2 Related Work

There are various trend filtering methods proposed in the literature. As discussed in the previous section, the H-P trend filtering [Hodrick and Prescott, 1997] and ℓ_1 trend filtering [Kim *et al.*, 2009] extract the trend component by minimizing the weighted sum of the residual size and the smoothness of the trend component using the second order difference operator. The TV denoising filtering [Chan *et al.*, 2001] adopts the same loss function as the H-P and ℓ_1 trend filtering, but it utilizes the first order difference as the regularization to deal with abrupt trend changes. As a result, the TV denoising filtering can handle the abrupt trend change successfully, but it tends to introduce staircases for slow trend changes. In contrast, the H-P and ℓ_1 trend filtering generally prone to trend delay for abrupt trend changes. The mixed trend filtering [Tibshirani, 2014] adopts two difference orders as regularization. It is assumed that the observed time series were drawn from an underlying function possessing different orders of piecewise polynomial smoothness. However, it is difficult to test this assumption adaptively. Thus, the selection of the two order remains a challenging problem, and limits its usage in practice. Besides, due to the sum-of-squares loss, all the aforementioned trend filtering methods cannot handle heavy tailed errors and outliers.

In [Fried, 2004; Fried *et al.*, 2018], a robust trend filtering called robfilter is proposed. In robfilter, a robust regression

functional for local approximation of the trend is used in a moving window. To further improve the robustness of robfilter, outliers in the original time series is removed by trimming or winsorization based on robust scale estimators. Empirical mode decomposition (EMD) [Wu *et al.*, 2007] is a popular method for analyzing non-stationary time series. With iterative process, it finds local maxima and minima, and then interpolates them into upper and lower envelopes. The mean of the envelopes yields local trend, while the residual must comply two requirements to become an intrinsic mode function (IMF). This method assumes that the trend is “slowly varying” and the selection of “slowest” IMF sums up to be the trend. Since the local maxima and minima are sensitive to noise and outliers, the conventional EMD are not robust. Another variant, Ensemble EMD (EEMD), is a noise-assisted method. It is an ensemble of white noise-added data and treats mean of their output as the final true result. It resolves the mode mixing issue in original EMD but still suffer from outliers. The wavelet analysis [Afanasyev and Fedorova, 2016] is a transformation of the original time series into two types of coefficients, wavelet coefficients and scaling coefficients. Note that the wavelet coefficients are related to changes of average over specific scales, whereas scaling coefficients can be associated with averages on a specified scale. Since the scale associated with scaling coefficients is usually fairly large, the general idea behind the wavelet trend filter [Craigmile and Percival, 2006] is to associate the scaling coefficients with the trend, and the wavelet coefficients with the noise component. The benefits of trend filtering via wavelet transformation is that it can capture the abrupt change of trend, but how to choose orthogonal basis still remains a challenge.

In this paper, we focus on trend filtering for time series without distinguishing periodic and trend components. When the seasonality/periodic and trend components need to be separated, the seasonal-trend decomposition method can be utilized, such as STL [Cleveland *et al.*, 1990], RobustSTL [Wen *et al.*, 2019], and TBATS [De Livera *et al.*, 2011].

3 Proposed Robust Trend Filter

3.1 Model Overview

We consider the time series of length N as $\mathbf{y} = [y_0, y_1, \dots, y_{N-1}]^T$, which can be decomposed into trend and residual components [Alexandrov *et al.*, 2012]

$$y_t = \tau_t + r_t \quad \text{or} \quad \mathbf{y} = \boldsymbol{\tau} + \mathbf{r}, \quad (1)$$

where $\boldsymbol{\tau} = [\tau_0, \tau_1, \dots, \tau_{N-1}]^T$ denotes the trend component, and $\mathbf{r} = [r_0, r_1, \dots, r_{N-1}]^T$ denotes the residual component. Note that in this paper we consider the time series faced in real-world applications. Therefore, the trend component may contains both slow trend change and abrupt trend change, while the residual component may contains both noise and outliers.

3.2 Preliminary

In the H-P trend filtering [Hodrick and Prescott, 1997], where the trend estimation τ_t is chosen to minimize the objective

function as follows

$$\frac{1}{2} \sum_{t=0}^{N-1} (y_t - \tau_t)^2 + \lambda \sum_{t=1}^{N-2} (\tau_{t-1} - 2\tau_t + \tau_{t+1})^2, \quad (2)$$

where the loss function (first term) measures the size of the residual signal after trend extraction while the regularization (second term) measures the smoothness of the extracted trend, and the regularization parameter $\lambda > 0$ controls the trade-off between them. The Eq. (2) can be equivalently formulated in the matrix form as

$$\frac{1}{2} \|\mathbf{y} - \boldsymbol{\tau}\|_2^2 + \lambda \|\mathbf{D}^{(2)} \boldsymbol{\tau}\|_2, \quad (3)$$

where $\mathbf{D}^{(2)}$ is the second-order difference matrix, and its definition is given in Eq. (6) below.

By replacing the ℓ_2 -norm with ℓ_1 -norm in the regularization term of H-P trend filtering, we obtain the ℓ_1 trend filtering [Kim *et al.*, 2009] as follows:

$$\frac{1}{2} \|\mathbf{y} - \boldsymbol{\tau}\|_2^2 + \lambda \|\mathbf{D}^{(2)} \boldsymbol{\tau}\|_1. \quad (4)$$

The ℓ_1 regularization in Eq. (4) promotes sparsity, which indicates $\|\mathbf{D}^{(2)} \boldsymbol{\tau}\|_1$ would have many zero elements. As a result, the ℓ_1 trend filter leads to a piecewise linear trend component.

As an extension of Eq. (4), piecewise polynomial trend signal can be obtained by using high-order difference operator in the regularization term [Kim *et al.*, 2009], i.e.,

$$\frac{1}{2} \|\mathbf{y} - \boldsymbol{\tau}\|_2^2 + \lambda \|\mathbf{D}^{(k+1)} \boldsymbol{\tau}\|_1, \quad (5)$$

where $\mathbf{D}^{(k+1)} \in \mathcal{R}^{(n-k-1) \times n}$ is the discrete difference operator of order $k + 1$, which can be defined recursively as $D^{(k+1)} = D^{(1)} \cdot D^{(k)}$ and the $D^{(1)}$ and $D^{(2)}$ are the first-order and second-order difference, i.e.,

$$\mathbf{D}^{(1)} = \begin{bmatrix} 1 & -1 & & & \\ & 1 & -1 & & \\ & & \ddots & \ddots & \\ & & & 1 & -1 \end{bmatrix}, \quad \mathbf{D}^{(2)} = \begin{bmatrix} 1 & -2 & 1 & & & \\ & 1 & -2 & 1 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 & -2 & 1 \end{bmatrix}. \quad (6)$$

When $k = 0$ in Eq. (5), it corresponds to the total variation (TV) filter [Chan *et al.*, 2001], which is suitable when the underlying trend signal is approximately piecewise constant.

3.3 Robust Trend Filtering with Sparse Model

All these trend filtering methods discussed in Section 3.2 adopt the sum-of-squares loss, which is appropriate when the residual part is near to Gaussian noise. However, in real-world applications, the residual often exhibits long-tailed distribution due to outliers. In this paper, we adopt the Huber loss from robust statistics [Huber and Ronchetti, 2009] to deal with possible long-tailed distributed residual. As discussed in Section 1, Huber loss combines the advantages of the sum-of-squares loss and the LAD loss adaptively.

In order to simultaneously capture abrupt and slow trend changes, we consider two sparse regularization terms like the fused LASSO [Tibshirani *et al.*, 2005]. Specifically, we utilize first and second order difference operators in the regularization to deal with abrupt and slow trend changes, respectively. Since it is proved that the combination of the first and

second order difference regularization can achieve comparable results to higher order difference regularization [Papafitsoros and Schönlieb, 2014], in this paper we do not consider other higher order difference operators.

Based on the above discussion, we propose a Huber loss with a combined first and second order difference regularization as a robust trend filtering for time series (named RobustTrend in this paper), i.e.,

$$g_\gamma(\mathbf{y} - \boldsymbol{\tau}) + \lambda_1 \|\mathbf{D}^{(1)} \boldsymbol{\tau}\|_1 + \lambda_2 \|\mathbf{D}^{(2)} \boldsymbol{\tau}\|_1 \quad (7)$$

where $D^{(1)}$, $D^{(2)}$ are the first-order and second-order difference matrix, respectively. And $g_\gamma(\mathbf{x}) = \sum_i g_\gamma(x_i)$ is the summation of elementwise Huber loss function with each element as

$$g_\gamma(x_i) = \begin{cases} \frac{1}{2}x_i^2, & |x_i| \leq \gamma \\ \gamma|x_i| - \frac{1}{2}\gamma^2, & |x_i| > \gamma \end{cases} \quad (8)$$

and its derivative can be obtained as

$$g'_\gamma(x_i) = \begin{cases} x_i, & |x_i| \leq \gamma \\ \gamma \text{sgn}(x_i), & |x_i| > \gamma \end{cases} \quad (9)$$

where sgn is the sign function.

3.4 Efficient MM-ADMM Algorithm

In this section, we develop a specialized ADMM algorithm to solve the minimization problem of RobustTrend in Eq. (7). ADMM [Boyd *et al.*, 2011] is a powerful framework in optimization and also utilized in trend filtering with sum-of-square loss [Ramdas and Tibshirani, 2016]. However, the standard ADMM is not efficient in the RobustTrend filter since no exact solution exists in the updating step with Huber loss (will be shown later). One may transform Eq. (7) into a series of ordinary fused lasso problems as in [Polson and Scott, 2016]. But it is still inefficient since each ordinary fused lasso problem needs to be solved iteratively as well. To alleviate this problem, we utilize the majorization-minimization (MM) framework [Sun *et al.*, 2017] to approximate the Huber loss function in the ADMM updating step to obtain a closed form for efficient computation.

To simplify the ADMM formulation for Eq. (7), we rewrite it in an equivalent form as

$$\begin{aligned} \min \quad & g_\gamma(\mathbf{y} - \boldsymbol{\tau}) + \|\mathbf{z}\|_1 \\ \text{s.t.} \quad & \mathbf{D}\boldsymbol{\tau} = \mathbf{z} \end{aligned} \quad (10)$$

where

$$\mathbf{D} = \begin{bmatrix} \lambda_1 \mathbf{D}^{(1)} \\ \lambda_2 \mathbf{D}^{(2)} \end{bmatrix}. \quad (11)$$

Then we can obtain the augmented Lagrangian as

$$L_\rho(\boldsymbol{\tau}, \mathbf{z}, \mathbf{v}) = g_\gamma(\mathbf{y} - \boldsymbol{\tau}) + \|\mathbf{z}\|_1 + \mathbf{v}^T (\mathbf{D}\boldsymbol{\tau} - \mathbf{z}) + \frac{\rho}{2} \|\mathbf{D}\boldsymbol{\tau} - \mathbf{z}\|_2^2$$

where \mathbf{v} is the dual variable, and ρ is the penalty parameter. Following the ADMM procedure in [Boyd *et al.*, 2011], we can obtain the updating steps as

$$\boldsymbol{\tau}^{k+1} = \arg \min_{\boldsymbol{\tau}} \left(g_\gamma(\mathbf{y} - \boldsymbol{\tau}) + \frac{\rho}{2} \|\mathbf{D}\boldsymbol{\tau} - \mathbf{z}^k + \mathbf{u}^k\|_2^2 \right) \quad (12)$$

$$\mathbf{z}^{k+1} = \arg \min_{\mathbf{z}} \left(\|\mathbf{z}\|_1 + \frac{\rho}{2} \|\mathbf{D}\boldsymbol{\tau}^{k+1} - \mathbf{z} + \mathbf{u}^k\|_2^2 \right) \quad (13)$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{D}\boldsymbol{\tau}^{k+1} - \mathbf{z}^{k+1} \quad (14)$$

where $\mathbf{u} = (1/\rho)\mathbf{v}$ is the scaled dual variable to make the formulation more convenient. The above \mathbf{z} -minimization step can be efficiently solved by soft thresholding as

$$\mathbf{z}^{k+1} = S_\rho(\mathbf{D}\boldsymbol{\tau}^{k+1} + \mathbf{u}^k) \quad (15)$$

where the soft thresholding operator $S_\rho(a)$ is defined as

$$S_\rho(a) = \begin{cases} 0, & |a| \leq \rho \\ a - \rho \operatorname{sgn}(a), & |a| > \rho \end{cases} \quad (16)$$

The $\boldsymbol{\tau}$ -minimization step in Eq. (12) is nontrivial since no exact solution exists. One may use iterative algorithms, like standard gradient methods or conjugate gradient methods, to solve the $\boldsymbol{\tau}$ -minimization step since it is convex. However, this would increase the overall computation since the ADMM itself is an iterative algorithm. Inspired by the work of [Eckstein and Bertsekas, 1992] that the ADMM can still converge when the minimization steps are carried out approximately, we utilize one-iteration MM framework like [Pham and Venkatesh, 2013] to efficiently solve the $\boldsymbol{\tau}$ -minimization step in Eq. (12) with closed form solution. Specifically, we adopt the proved sharpest quadratic majorization for Huber loss as in [de Leeuw and Lange, 2009] to minimize the difference between the exact and approximate solution, i.e.,

$$\begin{aligned} g_\gamma(x_i) \leq \eta_\gamma(x_i|x_i^k) &= \frac{g'_\gamma(x_i^k)}{2x_i^k}(x_i^2 - (x_i^k)^2) + g_\gamma(x_i^k) \\ &= \frac{g'_\gamma(x_i^k)}{2x_i^k}x_i^2 + C, \end{aligned}$$

where $\eta_\gamma(x_i|x_i^k)$ denotes the majorization of element x_i 's Huber loss and x_i^k is the solution of x_i at ADMM's k th iteration. Then, the summation of Huber loss can be obtained as

$$g_\gamma(\mathbf{x}) \leq \eta_\gamma(\mathbf{x}|\mathbf{x}^k) = \sum_i \eta_\gamma(x_i|x_i^k) = \frac{1}{2}\mathbf{x}^T \mathbf{A} \mathbf{x} + C \quad (17)$$

where C represents a constant number and \mathbf{A} is a diagonal matrix as

$$\mathbf{A} = \operatorname{diag}(g'_\gamma(x_i^k)) \operatorname{diag}^{-1}(g_\gamma(x_i^k)). \quad (18)$$

Next, let define $\mathbf{x} = \mathbf{y} - \boldsymbol{\tau}$ and replace the Huber loss in Eq. (12) with the quadratic majorization in Eq. (17), we can obtain an approximate $\boldsymbol{\tau}$ -minimization step with one-iteration MM algorithm as

$$\mathbf{x}^{k+1} = \arg \min_{\mathbf{x}} \left(\frac{1}{2}\mathbf{x}^T \mathbf{A} \mathbf{x} + \frac{\rho}{2} \|\mathbf{D}\mathbf{x} - (\mathbf{u}^k - \mathbf{z}^k + \mathbf{D}\mathbf{y})\|_2^2 \right) \quad (19)$$

$$\boldsymbol{\tau}^{k+1} = \mathbf{y} - \mathbf{x}^{k+1} \quad (20)$$

To obtain the closed form of the above approximate $\boldsymbol{\tau}$ -minimization step, we take derivative of Eq. (19), set it to zero to obtain \mathbf{x}^{k+1} , and put \mathbf{x}^{k+1} into Eq. (20) to get the final $\boldsymbol{\tau}$ -minimization step as

$$\boldsymbol{\tau}^{k+1} = \mathbf{y} - \rho(\mathbf{A} + \rho\mathbf{D}^T\mathbf{D})^{-1}\mathbf{D}^T(\mathbf{u}^k - \mathbf{z}^k + \mathbf{D}\mathbf{y}). \quad (21)$$

Therefore, to obtain the extracted trend signal of the proposed RobustTrend filter, our designed MM-ADMM algorithm sequentially calculates Eq. (21), Eq. (15), and Eq. (14) in each ADMM iteration until termination.

For the termination criteria, besides the maximum iteration number, we also check the primal and dual residuals in each ADMM iteration until they are small enough. Specifically, the values of primal residual \mathbf{r}^k and dual residual \mathbf{s}^k at k th iteration are calculated by

$$\|\mathbf{r}^k\|_2 = \|\mathbf{D}\boldsymbol{\tau}^k - \mathbf{z}^k\|_2, \quad (22)$$

$$\|\mathbf{s}^k\|_2 = \rho\|\mathbf{D}^T(\mathbf{z}^k - \mathbf{z}^{k-1})\|_2. \quad (23)$$

Then, $\|\mathbf{r}^k\|_2$ and $\|\mathbf{s}^k\|_2$ are checked if they are smaller than the corresponding tolerance thresholds as [Boyd *et al.*, 2011]

$$\begin{aligned} \epsilon^{pri} &= \sqrt{2N-3}\epsilon^{abs} + \epsilon^{rel} \max\{\|\mathbf{D}\boldsymbol{\tau}^k\|_2, \|\mathbf{z}^{(k)}\|_2\}, \\ \epsilon^{dual} &= \sqrt{N}\epsilon^{abs} + \epsilon^{rel}\|\rho\mathbf{D}^T\mathbf{u}^{(k)}\|_2, \end{aligned}$$

where $\epsilon^{abs} > 0$ is an absolute tolerance and $\epsilon^{rel} > 0$ is a relative tolerance. A reasonable choice for the two tolerance is around 10^{-3} to 10^{-4} .

3.5 Online Extension for Data Streams

In order to support streaming data applications, the proposed RobustTrend filter is extended to online mode by applying sliding window scheme like robfilter and repeated median filter in [Fried, 2004; Fried *et al.*, 2018]. During each sliding window, the trend signal is extracted using our designed MM-ADMM algorithm described in Section 3.4. To speed up, the $\rho\mathbf{D}^T\mathbf{D}$ and $\mathbf{D}\mathbf{y}$ are computed once and cached for the $\boldsymbol{\tau}$ -minimization step in Eq. (21). Furthermore, warm start is adopted in each new sliding window, where the initial values for $\boldsymbol{\tau}$, \mathbf{z} , and \mathbf{u} are from previous window's final solution by removing the first element and copying last element once. Since the trend signal usually exhibits slow change, this warm start trick often provides an appropriate approximation to result in fewer ADMM iterations.

4 Experiments and Comparisons

In this section, we conduct experiments on both synthetic data and real-world data to demonstrate the effectiveness of the proposed RobustTrend filtering algorithm.

4.1 Baseline Algorithms

We tested nine state-of-the-art trend filters as baseline algorithms to perform a relative comprehensive evaluation, including H-P filter, ℓ_1 trend filter, TV denoising filter, Mixed trend filter [Tibshirani, 2014], Wavelet trend filter [Craigmile and Percival, 2006], Repeated median filter [Siegel, 1982], robfilter [Fried, 2004], EMD trend filter [Moghtaderi *et al.*, 2011] and its extension EEMD trend filter.

4.2 Dataset

For synthetic data, we first generate the trend signal with 1000 data points, which contains a sin wave, a triangle wave, and a square wave with 1.0 amplitude to represent smooth slow trend change as well as abrupt trend change with sharp discontinuity. Next, we add Gaussian noise with 0.2 standard

Metric Outlier ratio	MSE				MAE			
	1%	5%	10%	20%	1%	5%	10%	20%
H-P filter	0.0126	0.0209	0.0263	0.0502	0.0656	0.0985	0.1158	0.1730
ℓ_1 trend filter	0.0102	0.0170	0.0226	0.0469	0.0536	0.0860	0.1029	0.1684
TV denoising filter	0.0184	0.0223	0.0277	0.0473	0.0973	0.0983	0.1272	0.1827
Mixed trend filter	0.0070	0.0137	0.0183	0.0451	0.0562	0.0772	0.0963	0.1715
Wavelet trend filter	0.0167	0.0293	0.0348	0.0648	0.0857	0.1300	0.1400	0.1996
robfilter	0.0047	0.0109	0.0080	0.0110	0.0500	0.0569	0.0596	0.0586
Repeated median filter	0.0121	0.0117	0.0129	0.0145	0.0615	0.0599	0.0642	0.0724
EMD filter	0.0240	0.0366	0.0401	0.0902	0.0999	0.1410	0.1523	0.2293
EEMD filter	0.0200	0.0286	0.0393	0.0595	0.0917	0.1189	0.1479	0.1850
Proposed RobustTrend filter	0.0051	0.0054	0.0058	0.0079	0.0434	0.0442	0.0501	0.0638

Table 1: Performance of the proposed algorithm compared with other trend filter algorithms on synthetic data with different ratios of outliers.

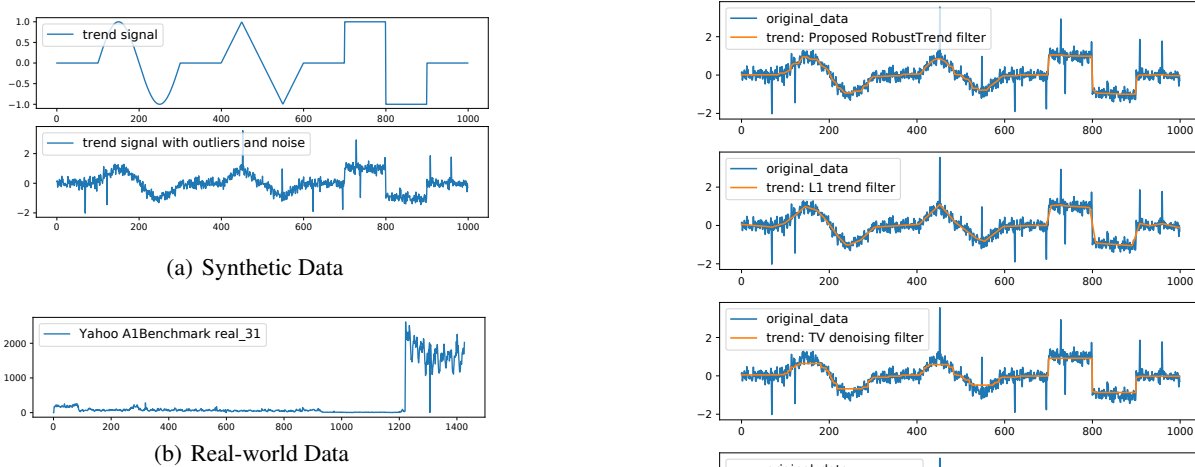


Figure 1: Synthetic and real-world data used for the experiment.

deviation. Then, we add 10 to 200 spikes and dips with 2.0 amplitude to represent 1% to 20% outlier ratio. As an example, the trend signal and trend with noise and 1% outlier are shown in Figure 1(a).

Note that we generate synthetic data with sin, triangle, and square wave as trend signal to simulate different types of trend changes in real-world data. If the trend signal only contains triangle/rectangle shape, the ℓ_1 /TV filter would be the optimal choice (under the case of white noise without outliers). However, these specific assumptions are of limited usage in practice. As the ground truth of trend is known, the mean squared error (MSE) and mean absolute error (MAE) are used for the performance evaluation later.

In addition to the synthetic data, we also adopt one real-world data from Yahoo’s Anomaly Detection Dataset¹, i.e., real_31 in A1Benchmark, which is shown in Figure 1(b). It can be seen that there is abrupt trend change around point 1220 and a outlier at point 1306.

4.3 Experiment Results on Synthetic Data

Figure 2 shows the trend filtering results from different algorithms when there is 1% outlier ratio. As can be seen in the figure, ℓ_1 trend filter, wavelet trend filter, and EEMD trend filter are heavily affected by the abrupt level change around the

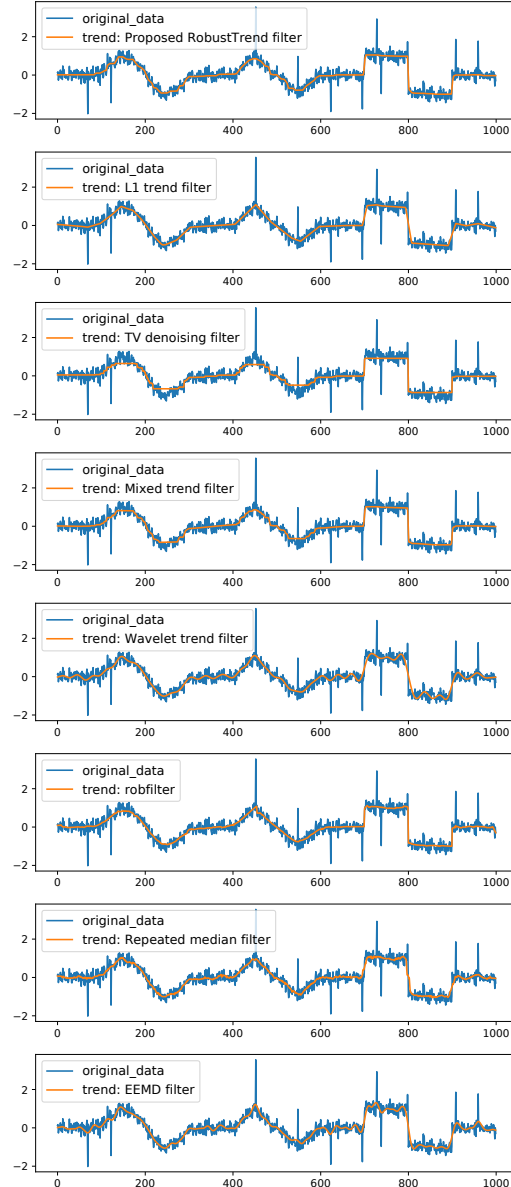


Figure 2: Trend filtering results on synthetic data.

¹<https://webscope.sandbox.yahoo.com/catalog.php?datatype=s>

square wave. TV denoising filter captures abrupt level change but exhibits staircasing around sine and triangle wave. Due to space limitation, we omit the results of H-P and EMD trend filters since their performances are worse than ℓ_1 and EEMD trend filters, respectively. To better evaluate the performance quantitatively, we calculate MSE and MAE to quantify the extraction accuracy. In addition to extracting trends on 1% of outliers, we also increase the ratio of outliers to 5%, 10%, and 20%. Table 1 summarizes the performance of our proposed algorithm along with other nine state-of-the-art trending filtering algorithms. The best results for each case are highlighted in bold fonts. Overall, it can be observed that our algorithm outperforms others.

The trend recovery near change points is important as it measures how prompt we can capture the trend change in time. Thus, we calculate MSE and MAE around 9 change points and their neighbors (one before and one after each change point) in the synthetic data (total 27 points) when outlier ratio is 5%. The results are summarized in Table 2 where the best results are highlighted in bold fonts. Clearly, our algorithm achieves significantly more accurate trend near change points.

To evaluate the different components of our RobustTrend filter, we also compare TV denoising filter with Huber loss (i.e., RobustTrend filter without second order difference regularization), ℓ_1 trend filter with Huber loss (i.e., RobustTrend filter without first order difference regularization), and RobustTrend filter with ℓ_2 -norm regularizations. The results are summarized in Table 3 with the outlier ratio setting to 5%, where the best results are highlighted in bold fonts. It can be seen that the Huber loss and the first order and second order difference ℓ_1 regularization terms bring significant performance improvements in trend filtering.

4.4 Experiment Results on Real-World Data

We perform experiment on one real-world dataset from Yahoo depicted in Figure 1(b). In this experiment, we apply the online mode of the trend filtering to investigate how it performs on streaming data. The results of top-3 performance filters (our RobustTrend, robfilter, and repeated median filter) and the popular ℓ_1 trend filter are summarized in Figure 3. Since the beginning of the data is almost constant, we only show the zoomed-in results after point 1200. It can be observed that our RobustTrend is able to follow the sudden increase of the trend and is not affected by the outlier. The ℓ_1 filter and robfilter show a delay in capturing the trend change. The ℓ_1 filter is also sensitive to the outliers in the original signal, leading to a dip in the trend. The repeated median algorithm overshoots the estimation of the trend and generates a higher trend than the actual one near the position where trend abruptly changes.

5 Conclusions

In this paper we propose a robust trend filtering method RobustTrend based on the Huber loss. The Huber loss can adaptively deal with residual outliers and different types of data. In order to impose the smoothness of the trend and meanwhile capture the abrupt trend change, we propose to use the

Algorithms	MSE	MAE
H-P filter	0.1463	0.3036
ℓ_1 trend filter	0.1404	0.2910
TV denoising filter	0.1100	0.2488
Mixed trend filter	0.1140	0.2526
Wavelet trend filter	0.1467	0.3052
robfilter	0.1900	0.2939
Repeated median filter	0.1378	0.2495
EMD filter	0.1618	0.3074
EEMD filter	0.1570	0.3004
Proposed RobustTrend filter	0.0862	0.1966

Table 2: Comparison of change points MSE and MAE on synthetic data of different trend filtering algorithms under 5% outlier ratio.

Algorithms	MSE	MAE
TV denoising filter	0.0223	0.0983
TV denoising filter with Huber loss	0.0077	0.0564
ℓ_1 trend filter	0.0170	0.0860
ℓ_1 trend filter with Huber loss	0.0097	0.0677
RobustTrend with L2 Reg	0.0093	0.0585
Proposed RobustTrend filter	0.0054	0.0442

Table 3: Comparison of MSE and MAE on synthetic data of different trend filtering algorithms under 5% outlier ratio.

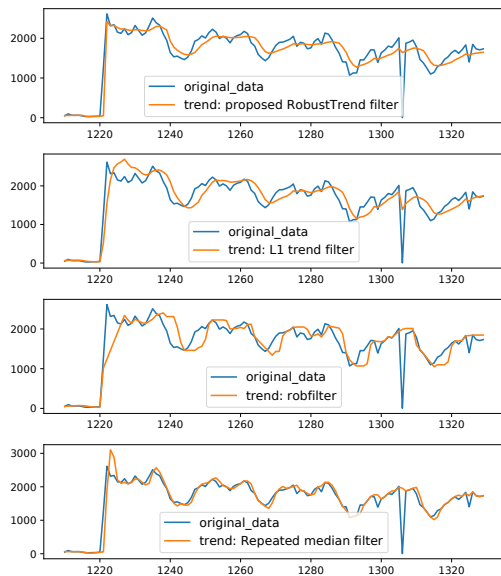


Figure 3: Zoomed-in version of the trend filtering results from different algorithms using online mode on real data.

combination of the first order and the second order difference of the trend component as regularization in trend filtering. To efficiently solve the resulting optimization problem, we design an MM-ADMM algorithm which applies majorization-minimization to facilitate the updating step for the Huber loss effectively.

References

[Afanasyev and Fedorova, 2016] Dmitriy O Afanasyev and Elena A Fedorova. The long-term trends on the electric-

- ity markets: Comparison of empirical mode and wavelet decompositions. *Energy Economics*, 56:432–442, 2016.
- [Alexandrov *et al.*, 2012] Theodore Alexandrov, Silvia Bianconcini, Estela Bee Dagum, Peter Maass, and Tucker S McElroy. A review of some modern approaches to the problem of trend extraction. *Econometric Reviews*, 31(6):593–624, 2012.
- [Boyd *et al.*, 2011] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends in Machine Learning*, 3(1):1–122, 2011.
- [Chan *et al.*, 2001] Tony F Chan, Stanley Osher, and Jianhong Shen. The digital TV filter and nonlinear denoising. *IEEE Transactions on Image Processing*, 10(2):231–241, 2001.
- [Cleveland *et al.*, 1990] Robert B Cleveland, William S Cleveland, Jean E McRae, and Irma Terpenning. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73, 1990.
- [Craigmile and Percival, 2006] Peter F Craigmile and Donald B Percival. Wavelet-based trend detection and estimation. *Encyclopedia of Environmetrics*, 6, 2006.
- [de Leeuw and Lange, 2009] Jan de Leeuw and Kenneth Lange. Sharp quadratic majorization in one dimension. *Computational Statistics & Data Analysis*, 53(7):2471–2484, 2009.
- [De Livera *et al.*, 2011] Alysha M De Livera, Rob J Hyndman, and Ralph D Snyder. Forecasting time series with complex seasonal patterns using exponential smoothing. *Journal of the American Statistical Association*, 106(496):1513–1527, 2011.
- [Eckstein and Bertsekas, 1992] Jonathan Eckstein and Dimitri P Bertsekas. On the Douglas—Rachford splitting method and the proximal point algorithm for maximal monotone operators. *Mathematical Programming*, 55(1-3):293–318, 1992.
- [Fried *et al.*, 2018] Roland Fried, Karen Schettlinger, and Matthias Borowski. robfilter: Robust time series filters. URL: <https://cran.r-project.org/web/packages/robfilter/index.html>. R package version 4.1.1. (visited on 11/11/2018), 2018.
- [Fried, 2004] Roland Fried. Robust filtering of time series with trends. *Journal of Nonparametric Statistics*, 16(3-4):313–328, 2004.
- [Hodrick and Prescott, 1997] Robert J Hodrick and Edward C Prescott. Postwar US business cycles: an empirical investigation. *Journal of Money, Credit, and Banking*, pages 1–16, 1997.
- [Huber and Ronchetti, 2009] Peter J. Huber and Elvezio M. Ronchetti. *Robust statistics*. John Wiley & Sons, New Jersey, 2nd edition, 2009.
- [Kim *et al.*, 2009] Seung-Jean Kim, Kwangmoo Koh, Stephen Boyd, and Dimitry Gorinevsky. ℓ_1 trend filtering. *SIAM Review*, 51(2):339–360, 2009.
- [Moghtaderi *et al.*, 2011] Azadeh Moghtaderi, Pierre Borgnat, and Patrick Flandrin. Trend filtering: empirical mode decompositions versus ℓ_1 and Hodrick-Prescott. *Advances in Adaptive Data Analysis*, 3(01n02):41–61, 2011.
- [Papafitsoros and Schönlieb, 2014] Konstantinos Papafitsoros and Carola-Bibiane Schönlieb. A combined first and second order variational approach for image reconstruction. *Journal of Mathematical Imaging and Vision*, 48(2):308–338, Feb 2014.
- [Pham and Venkatesh, 2013] Duc-Son Pham and Svetha Venkatesh. Efficient algorithms for robust recovery of images from compressed data. *IEEE transactions on image processing*, 22(12):4724–4737, 2013.
- [Polson and Scott, 2016] Nicholas G Polson and James G Scott. Mixtures, envelopes and hierarchical duality. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(4):701–727, 2016.
- [Ramdas and Tibshirani, 2016] Aaditya Ramdas and Ryan J Tibshirani. Fast and flexible ADMM algorithms for trend filtering. *Journal of Computational and Graphical Statistics*, 25(3):839–858, 2016.
- [Siegel, 1982] Andrew F Siegel. Robust regression using repeated medians. *Biometrika*, 69(1):242–244, 1982.
- [Sun *et al.*, 2017] Ying Sun, Prabhu Babu, and Daniel P Palomar. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Transactions on Signal Processing*, 65(3):794–816, Feb 2017.
- [Tibshirani *et al.*, 2005] Robert Tibshirani, Michael Saunders, Saharon Rosset, Ji Zhu, and Keith Knight. Sparsity and smoothness via the fused lasso. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(1):91–108, 2005.
- [Tibshirani, 2014] Ryan Tibshirani. Adaptive piecewise polynomial estimation via trend filtering. *The Annals of Statistics*, 42(1):285–323, 2014.
- [Wang *et al.*, 2007] Hansheng Wang, Guodong Li, and Guohua Jiang. Robust regression shrinkage and consistent variable selection through the LAD-Lasso. *Journal of Business & Economic Statistics*, 25(3):347–355, 2007.
- [Wen *et al.*, 2019] Qingsong Wen, Jingkun Gao, Xiaomin Song, Liang Sun, Huan Xu, and Shenghuo Zhu. Robust-STL: A robust seasonal-trend decomposition algorithm for long time series. In *AAAI*, pages 1501–1509, 2019.
- [Wu and Huang, 2009] Zhaohua Wu and Norden E Huang. Ensemble empirical mode decomposition: a noise-assisted data analysis method. *Advances in Adaptive Data Analysis*, 1(01):1–41, 2009.
- [Wu *et al.*, 2007] Zhaohua Wu, Norden E Huang, Steven R Long, and Chung-Kang Peng. On the trend, detrending, and variability of nonlinear and nonstationary time series. *Proceedings of the National Academy of Sciences*, 104(38):14889–14894, 2007.