

# IEEE Signal Processing Magazine

## Accepted Article

---

---

---

# Kalman Filtering in Non-Gaussian Model Errors: A New Perspective<sup>a</sup>

ARMAN KHEIRATI ROONIZI

IT is well-known that the optimality of the Kalman filter relies on the Gaussian distribution of the process and observation model errors, which in many situations is well justified [1, 2, 3]. However, its optimality is useless in applications that the distribution assumptions of the model errors do not hold in practice. Even minor deviation from the assumed (or nominal) distribution may cause the Kalman filter's performance to drastically degrade or to completely break down. In particular, when dealing with perceptually important signals such as speech, images, medical, campaign and ocean engineering, the measurements have confirmed the presence of non-Gaussian impulsive (heavy tailed) or Laplace noises [4]. Therefore, the classical Kalman filter which is derived under nominal Gaussian probability model is biased or even breaks down in such situations. This article presents a simple modification approach to overcome this limitation of the Kalman filter. We show that in the smoothing Wiener filter, the estimated state is chosen to minimize the  $\ell_2$ -norm of the variation of the system model error. The  $\ell_2$ -norm, however, corresponds to Gaussian priors. It confirms the optimality of Wiener filter and Kalman filter in linear Gaussian systems and explains why they suffer from the sensitivity to Laplace or impulsive error statistics. To address this limitation, we propose a variation on smoothing Wiener filter which substitutes a sum of absolute values (i.e.,  $\ell_1$ -norm) for the sum of squares used in  $\ell_2$  smoothing Wiener filter to penalize variations in the system model error. The proposed  $\ell_1$  smoothing Wiener filter is suitable for analyzing systems with impulsive or Laplace model errors. The idea of  $\ell_1$  smoothing Wiener filter is then used to recast and correct the system (or process) model equation. The modified model puts a Laplace or a sparse distribution on the system model error to enforce the sparsity on the states of the system. Based on the fact that the formulations of the optimum filter by Wiener and Kalman are equivalent in discrete time steady state [5], the Kalman filter can be used to estimate the desired states using the modified model.

---

<sup>a</sup>This is the author's version of the manuscript accepted for publication in IEEE Signal Processing Magazine, Volume 38, Issue 2, March 2022. The paper can be found in its final format at: IEEE Signal Processing Magazine website doi: 10.1109/MSP.2021.3134635

20xx IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

---

## Introduction

This year Kalman filter celebrates 60 years old. The Kalman filter, introduced by Rudolf E. Kálmán in his most well-known contribution [6], is one of the most important and famous data fusion algorithms [an optimal estimator in

the minimum mean-square error (MMSE) sense] for estimating the hidden state of a linear dynamical system [2]. It has been widely applied in various fields such as aerospace engineering, control systems, signal and image processing, wireless communication, and many more. One can do an Internet search on the terms "state estimation + application" or

“Kalman filter + application” and discover thousands of other applications.

Before Kalman filter appearances, the Wiener filter introduced by Nobert Wiener was probably the first theory in merging the dynamic systems and optimal estimation in the presence of noise [7]. It is a signal estimation algorithm that produces an optimal estimate of a random signal using both the autocorrelation and crosscorrelation of the measurement signal with the original signal. The Wiener filter minimizes the MSE between the estimated random process and the desired process. The MSE criteria has long been regarded as the dominant quantitative performance metric to optimizing and assessing the Wiener filter and the Kalman filter design. The MSE is simple, memoryless, parameter free and inexpensive to compute. Specially, in the context of optimization and estimation theory it satisfies many important properties such as convexity, symmetry, and differentiability [8]. That is why the MSE is used to optimize a large variety of signal processing algorithms such as Wiener filter and Kalman filter. MSE is defined by  $\ell_2$ -norm which corresponds to Gaussian priors, the reason that both Wiener and Kalman filters are known as optimal estimators for linear systems subject to Gaussian priors (i.e., all the latent and observed variables have Gaussian distribution). However, in many real applications, the MSE (or  $\ell_2$  regularization) exhibits weak performance and has been widely criticized for serious shortcomings, especially when the latent or observed variables are distributed by non-Gaussian distribution, for instance, when dealing with signals such as finance and ocean engineering, speech and images where a sparse distribution is appropriate to represent their statistics [4]. Impulse or sparse distribution also appears in various machine learning areas such as independent component analysis [9], mobile and wireless communication [10]. The classical Wiener and Kalman filters are derived under Gaussian distribution assumptions and thus suffer from the sensitivity to sparse or Laplace error statistics. The main purpose of this article is to provide a modified structure to optimize the Wiener filter and the Kalman filter to not lose their optimality in applications where the system variables have impulsive or Laplace distribution.

Recently, the non-Gaussian robust Kalman filtering has been attracting more and more attention. A generalized maximum-likelihood Kalman filter (GMKF) has been derived in [11] and extended in [12] with an application to power grids. The key idea of the GMKF is to formulate a batch-mode regression after performing the Kalman filter prediction

step. Then a prewhitening method is performed to the batch-mode regression. Other main step is to apply the Schweppe-type Huber generalized maximum-likelihood estimator [13] to solve the regression via an iteratively reweighted least squares (IRLS) algorithm. While most published algorithms (see [14] and references therein) are advanced, in our method, a simple modification change is applied to the system model and then the standard Kalman filter is performed for a given iterations. We neither modify the Kalman filter equations nor employ other advanced methods, but we show that by a simple modification change to the system model, the standard Kalman filter can be used to estimate the states of the system even if the system model errors have non-Gaussian distribution.

## Prerequisites

This lecture note requires basic knowledge of system theory, state-space model, estimation theory and Kalman filter.

## Notation

Lowercase letters are used to denote scalars, e.g.,  $a$ ; boldface lowercase letters for vectors, e.g.,  $\mathbf{a}$ ; boldface uppercase letters for matrices, e.g.,  $\mathbf{A}$ . The subscript  $k$  stands for discrete time index while  $(\cdot)^T$  and  $(\cdot)^{-1}$  denote the matrix transpose and matrix inverse, respectively. Symbol  $\mathbb{E}\{\cdot\}$  refers to the expected value operation and  $*$  denotes the convolution. The  $\ell_p$ -norm of a vector  $\mathbf{u}$  is defined as  $\|\mathbf{u}\|_p = (\sum_i |u_i|^p)^{\frac{1}{p}}$ .

## Background

Model-based state estimation algorithms (e.g., Kalman filter) assume that the state of a system at time  $k$  evolved from the prior state at time  $k - 1$  according to the equation  $\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_k$ , where  $\mathbf{x}_k$  is the state vector containing the terms of interest for the system [e.g., target position ( $p_k$ ), velocity ( $v_k$ ) and acceleration ( $a_k$ )],  $\mathbf{A}$  is the state transition matrix of the process from the state at time  $k - 1$  to the state at time  $k$  and  $\mathbf{w}_k$  is the process noise vector with known covariance,  $\mathbf{Q}_k \triangleq \mathbb{E}\{\mathbf{w}_k^T \mathbf{w}_k\}$ . Measurements of the system are represented according to the equation  $\mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \boldsymbol{\eta}_k$ , where  $\mathbf{y}_k$  is the vector of observations,  $\mathbf{H}$  is the transformation matrix which relates the state variables to the measurements and

$\boldsymbol{\eta}_k$  is the vector containing the observation noise terms with known covariance,  $\mathbf{R}_k \triangleq \mathbb{E}\{\mathbf{v}_k^T \mathbf{v}_k\}$ . Adaptive Kalman filter is known as the optimal state estimator when both process and observation noises are white Gaussian process. It involves two stages: prediction and measurement updates:

*Prediction Update:*

$$\begin{cases} \hat{\mathbf{x}}_{k|k-1} = \mathbf{A}\hat{\mathbf{x}}_{k-1|k-1} \\ \mathbf{P}_{k|k-1} = \mathbf{A}\mathbf{P}_{k-1|k-1}\mathbf{A}^T + \mathbf{Q}_k \end{cases} \quad (1)$$

*Measurement update:*

$$\begin{cases} \hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k [\mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}] \\ \mathbf{K}_k = \mathbf{P}_{k|k-1}\mathbf{H}^T (\mathbf{H}\mathbf{P}_{k|k-1}\mathbf{H}^T + \mathbf{R}_k)^{-1}, \quad (2) \\ \mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k\mathbf{H}\mathbf{P}_{k|k-1} \end{cases}$$

where  $\hat{\mathbf{x}}_{k|k-1} \triangleq \mathbb{E}\{\mathbf{x}_k | \mathbf{y}_{k-1}, \dots, \mathbf{y}_1\}$  is the *a priori* estimate of the state vector  $\mathbf{x}_k$  in the  $k$ -th stage using the observation  $\mathbf{y}_1$  to  $\mathbf{y}_{k-1}$ , and  $\hat{\mathbf{x}}_{k|k} \triangleq \mathbb{E}\{\mathbf{x}_k | \mathbf{y}_k, \dots, \mathbf{y}_1\}$  is the *a posteriori* estimate of the state vector after using the  $k$ -th observation  $\mathbf{y}_k$ . The matrices  $\mathbf{P}_{k|k-1} \triangleq \mathbb{E}\{(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k-1})^T\}$  and  $\mathbf{P}_{k|k} \triangleq \mathbb{E}\{(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})(\mathbf{x}_k - \hat{\mathbf{x}}_{k|k})^T\}$  are also defined as the *prior* and *posterior* state covariance matrices, while  $\mathbf{K}_k$  is the Kalman gain.

As an example of the application of Kalman filter/smoother, we consider a simple target tracking problem in which a truck moves along a straight line with a constant-acceleration (CA). In this case, the process model is defined by

$$\mathbf{x}_k = \begin{bmatrix} 1 & T_s & \frac{T_s^2}{2} \\ 0 & 1 & T_s \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{k-1} + \begin{bmatrix} \frac{T_s^2}{2} \\ T_s \\ 1 \end{bmatrix} w_k, \quad (3)$$

where  $T_s$  is the sampling period,  $\mathbf{x}_k = [p_k, v_k, a_k]^T$  contains the position, velocity and acceleration. We assume that while the truck moves with a constant acceleration, the driver sometimes applies a sudden braking input or sudden accelerating input to the system which results instantaneous changes in the acceleration values. In this case, a discrete event phenomenon is observed in the presence of constant-acceleration signal which is modelled as process noise, i.e.,  $w_k = [T_s^2/2, T_s, 1]^T w_k$ . We further assume that the truck is equipped with a GPS unit that can provide an estimate of the target states (position, velocity and acceleration) according to

the model  $\mathbf{y}_k = \mathbf{x}_k + \boldsymbol{\eta}_k$ , where  $\mathbf{y}_k = [y_{1,k}, y_{2,k}, y_{3,k}]^T$  is the vector of position, velocity and acceleration measurements and  $\boldsymbol{\eta}_k = [\eta_{1,k}, \eta_{2,k}, \eta_{3,k}]^T$  is the vector of observation noises. The observation noise,  $\boldsymbol{\eta}_k$ , is a white Gaussian process. The objective is to estimate the target states (position, velocity and acceleration) from the observation  $\mathbf{y}_k$ . The Kalman filter/smoother is then used to estimate the target states. The true states and the noisy states are shown via red solid line and gray dots in Figures 1(a)-1(c). The estimated target states provided by Kalman smoother are shown via blue dashed line in Figures 1(a)-1(c).

## Problem

We find that the estimated acceleration in Figure 1(c) does not have sharp discontinuities. The reason is that the Kalman filter/smoother is derived under nominal Gaussian probability model (i.e., all latent and observed variables are assumed to have Gaussian distribution). In other word, the process noise in (3) is assumed to be a white Gaussian noise by Kalman filter. As an illustration, the distribution of the actual acceleration process model error is plotted via blue color in Figure 2. The distribution of the nominal acceleration process model error with different values of  $Q$  (the covariance of the process noise,  $w_k$ ) is plotted via yellow color. It is seen that the nominal model error cannot adequately capture the actual model error even for large values of  $Q$ . A Gaussian distribution does not fit the actual process model error, but a (sparse) fat tail distribution can provide a better model to describe model error than the normal distribution. Therefore, the optimality of Kalman filter which relies on the Gaussian distribution of the process and observation noises, is not justified in such situations.

In this article, we analyze the above issue from a different perspective. To that end, we develop a joint perspective on Wiener filtering and Kalman filtering. Note that the formulations of the optimum filter by Wiener and Kalman are equivalent in steady state [5]. On the other hand, the smoothing Wiener filter can be viewed as an optimization problem with a  $\ell_2$ -norm smoothing constraint on the variation of process (system) model, as will become clear following the derivation outlined below in the “ $\ell_2$  smoothing Wiener filter” section.  $\ell_2$ -norm corresponds to Gaussian priors and is sensitive to the large values. Therefore, it exhibits weak performance in analyzing the sparse data such as the target acceleration

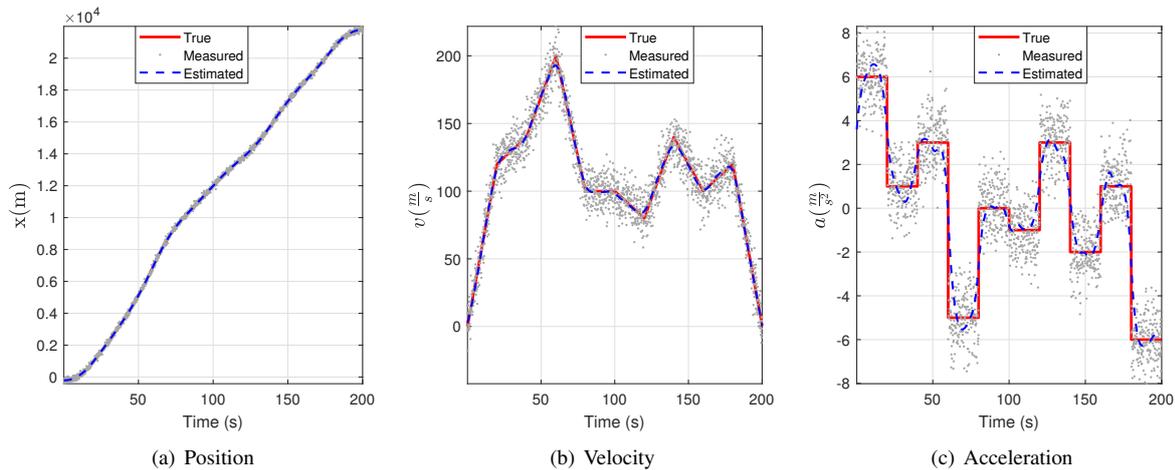


Figure 1: Parts (a), (b) and (c) show the results of adaptive Kalman smoother-based target state estimation.

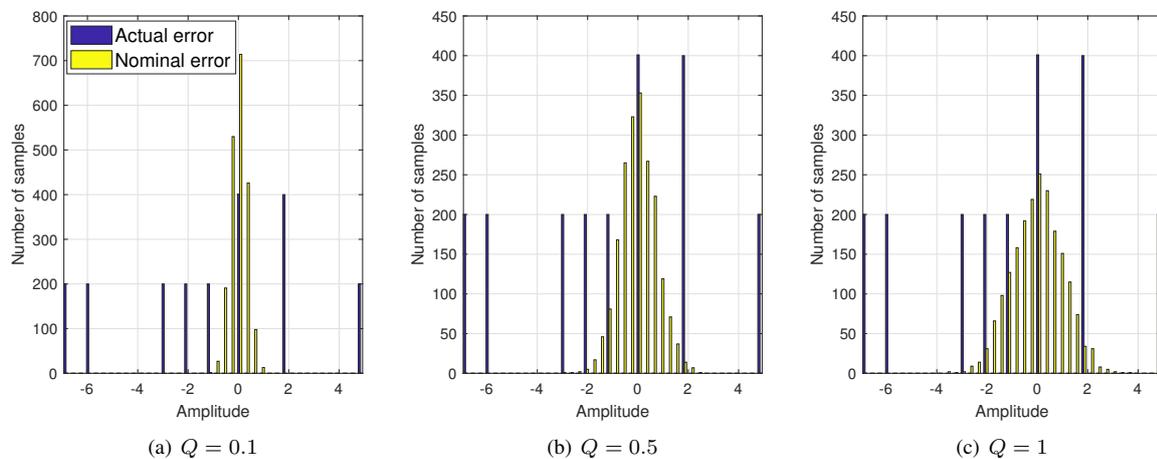


Figure 2: The histogram of the acceleration model error.

in the above example. That is why the estimated acceleration by Kalman smoother does not accurately follow the true acceleration signal. The remainder of this article describes a novel modification approach to overcome this limitation of the Kalman filter. We show that in the smoothing Wiener filter, the estimated state is chosen to minimize the  $\ell_2$ -norm of the variation of the system model error subject to the  $\ell_2$ -norm of the data fidelity. Then we propose a variation on smoothing Wiener filter which substitutes a sum of absolute values (i.e.,  $\ell_1$ -norm) for the sum of squares used in  $\ell_2$  smoothing Wiener filter to penalize variations in the system model error.  $\ell_1$ -norm corresponds to Laplacian priors. Therefore, the proposed  $\ell_1$  smoothing Wiener filter is suitable for analyzing the sparsely distributed data. The idea of  $\ell_1$  smoothing Wiener filter is then used to recast and correct the system (process)

model equation. The modified system model puts a Laplace distribution on the system model error to enforce the sparsity on the states of the system. Finally, using the fact that the formulations of the optimum filter by Wiener and Kalman are equivalent in steady state, the Kalman filter can be used to estimate the sparse states from the modified state space model.

## Wiener filter

In this section, we derive a KF model for  $\ell_2$  trend filtering and  $\ell_1$  trend filtering, respectively. First, we derive a dynamic model in state-space form to represent each approach and combine it with the Kalman filtering framework to estimate the desired signal.

## Smoothing Wiener filter

The aim of filter design is to find a filter such that when we apply the measurements to its input, it produces the MMSE estimate of the state of interest,  $\mathbf{x}_k$ :

$$\hat{\mathbf{x}}_k = \Psi_k * \mathbf{y}_k \approx \mathbf{x}_k \quad (4)$$

where  $\Psi_k$  is the filter impulse response and  $\hat{\mathbf{x}}_k$  is an estimate of the original state. To this aim, we define the error signal  $\mathbf{e}_k \triangleq \mathbf{x}_k - \hat{\mathbf{x}}_k$  and carry out the following minimization:

$$\min_{\Psi[\cdot]} (M = \mathbb{E}\{\mathbf{e}_k^T \mathbf{e}_k\}). \quad (5)$$

If  $\mathbf{x}_k$  and  $\boldsymbol{\eta}_k$  are uncorrelated, the linear system minimizing the MSE (the optimum smoothing Wiener filter) satisfies [7, 15, 16, 17]

$$\begin{aligned} \Psi(z) &= S_{\mathbf{x}\mathbf{y}}(z)S_{\mathbf{y}\mathbf{y}}^{-1}(z) \\ &= S_{\mathbf{x}\mathbf{x}}(z)\mathbf{H}^T [\mathbf{H}S_{\mathbf{x}\mathbf{x}}(z)\mathbf{H}^T + S_{\boldsymbol{\eta}\boldsymbol{\eta}}(z)]^{-1}, \end{aligned} \quad (6)$$

where  $S_{\mathbf{x}\mathbf{x}}(z)$  and  $S_{\boldsymbol{\eta}\boldsymbol{\eta}}(z)$  are the power spectral density (PSD) of  $\mathbf{x}_k$  and  $\boldsymbol{\eta}_k$ , respectively. (6) can also be expressed as

$$\Psi(z) = [\mathbf{H}^T S_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(z)\mathbf{H} + S_{\mathbf{x}\mathbf{x}}^{-1}(z)]^{-1} \mathbf{H}^T S_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(z) \quad (7)$$

According to the process model, the PSD of the hidden state and the PSD of the process noise are related as  $S_{\mathbf{x}\mathbf{x}}^{-1}(z) = \Theta^T(z^{-1})S_{\mathbf{w}\mathbf{w}}^{-1}(z)\Theta(z)$ , where  $S_{\mathbf{w}\mathbf{w}}(z)$  is the PSD of  $\mathbf{w}_k$  and  $\Theta(z) = \mathbf{I} - z^{-1}\mathbf{A}$ . Therefore, (7) is expressed as

$$\begin{aligned} \Psi(z) &= \left[ \mathbf{H}^T S_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(z)\mathbf{H} + \Theta^T(z^{-1}) \right. \\ &\quad \left. S_{\mathbf{w}\mathbf{w}}^{-1}(z)\Theta(z) \right]^{-1} \mathbf{H}^T S_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(z) \end{aligned} \quad (8)$$

In the following, the output response of smoothing Wiener filter is computed using an optimization problem with a  $\ell_2$  smoothing constraint on the variation of process (system) model.

### $\ell_2$ smoothing Wiener filter

In this section, we show that in the smoothing Wiener filter, the estimated state is chosen to minimize the  $\ell_2$ -norm of the process (or system) model error. To this purpose, we represent the output of (8) in the transform domain as  $\hat{\mathbf{X}}(z) =$

$\Psi(z)\mathbf{Y}(z)$ . According to (8), it can be written as

$$\begin{aligned} \left[ \mathbf{H}^T S_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(z)\mathbf{H} + \Theta^T(z^{-1}) \right. \\ \left. S_{\mathbf{w}\mathbf{w}}^{-1}(z)\Theta(z) \right] \hat{\mathbf{X}}(z) = \mathbf{H}^T S_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(z)\mathbf{Y}(z) \end{aligned} \quad (9)$$

(9) can be expressed as

$$\begin{aligned} \Theta^T(z^{-1})S_{\mathbf{w}\mathbf{w}}^{-1}(z)\Theta(z)\hat{\mathbf{X}}(z) \\ + \mathbf{H}^T S_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(z)\mathbf{H}\hat{\mathbf{X}}(z) - \mathbf{H}^T S_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(z)\mathbf{Y}(z) = 0 \end{aligned} \quad (10)$$

Taking the integral of (10) with respect to  $\mathbf{X}(z)$ , we obtain

$$\begin{aligned} \frac{1}{2} [\Theta(z)\hat{\mathbf{X}}(z)]^T S_{\mathbf{w}\mathbf{w}}^{-1}(z) [\Theta(z)\hat{\mathbf{X}}(z)] \\ + \frac{1}{2} [\mathbf{H}\hat{\mathbf{X}}(z) - \mathbf{Y}(z)]^T S_{\boldsymbol{\eta}\boldsymbol{\eta}}^{-1}(z) [\mathbf{H}\hat{\mathbf{X}}(z) - \mathbf{Y}(z)] = \boldsymbol{\kappa} \end{aligned} \quad (11)$$

where  $\boldsymbol{\kappa}$  stands for a constant function with respect to  $\mathbf{X}$ . It is straightforward to show that the inverse z-transform of (11), which is the the output response of the smoothing Wiener filter, is the solution of an optimization problem with smoothing constraint defined by the following loss function

$$\begin{aligned} \ell_{2,k} &= \frac{1}{2} (\mathbf{y}_k - \mathbf{H}\mathbf{x}_k)^T \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{H}\mathbf{x}_k) \\ &\quad + \frac{1}{2} (\Theta_k * \mathbf{x}_k)^T \mathbf{Q}_k^{-1} (\Theta_k * \mathbf{x}_k), \end{aligned} \quad (12)$$

where  $\Theta_k$  is the inverse z-transform of  $\Theta(z) = \mathbf{I} - z^{-1}\mathbf{A}$ , i.e.,  $\Theta_k = \boldsymbol{\delta}_k - \mathbf{A}\boldsymbol{\delta}_{k-1}$ ,  $\Theta_k * \mathbf{x}_k = \mathbf{x}_k - \mathbf{A}\mathbf{x}_{k-1}$  and  $\mathbf{R}_k$  and  $\mathbf{Q}_k$  were defined in Section . The first term in (12) describes a weighted  $\ell_2$ -norm of the observation model error while the second term describes a weighted  $\ell_2$ -norm of the process model error. In the following section, we propose a variation on the optimization problem to design the smoothing Wiener filter which can be used to overcome the problem discussed in previous section.

## Solution

We propose the following variation on smoothing Wiener filter, in which the estimated state is chosen as the minimizer of

$$\begin{aligned} \ell_{1,k} &= \frac{1}{2} (\mathbf{y}_k - \mathbf{H}\mathbf{x}_k)^T \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{H}\mathbf{x}_k) \\ &\quad + (\Theta_k * \mathbf{x}_k)^T \mathbf{Q}_k^{-1} \text{sgn}(\Theta_k * \mathbf{x}_k), \end{aligned} \quad (13)$$

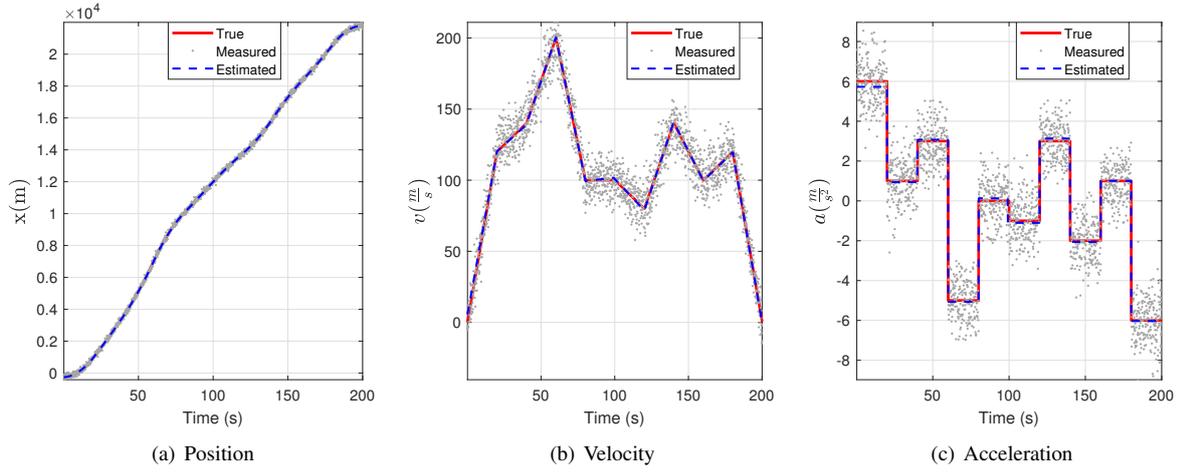


Figure 3: Results of the Kalman smoother-based state estimation of the modified model (20) after 5 iterations ( $m = 5$ ).

where  $\text{sgn}$  is the sign (or signum) function:

$$\text{sgn}(x) := \begin{cases} -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \\ 1 & \text{if } x > 0 \end{cases} \quad (14)$$

For matrices, the sign function is defined element wise. (13) employs a sum of absolute values (i.e.,  $\ell_1$ -norm) to penalize variations in the process model error. So, we call it  $\ell_1$  smoothing Wiener filter. The problem is that the second regularization term is non-differentiable which makes it a difficult minimization problem with no explicit solution. One way to deal with is to replace it with a sequence of simpler ones. This procedure is known as majorization-minimization (MM) method [18]. Using MM approach, (13) is converted to a simpler one. To this purpose, the MM approach proposes the following majorizer for the  $(\Theta_k * \mathbf{x}_k)^T \mathbf{Q}_k^{-1} \text{sgn}(\Theta_k * \mathbf{x}_k)$  [18]:

$$\begin{aligned} & (\Theta_k * \mathbf{x}_k)^T \mathbf{Q}_k^{-1} \text{sgn}(\Theta_k * \mathbf{x}_k) \\ & \leq \frac{1}{2} (\Theta_k * \mathbf{x}_k)^T \mathbf{C}_{k,m}^{-1} \mathbf{Q}_k^{-1} (\Theta_k * \mathbf{x}_k) + \frac{1}{2} \mathbf{C}_{k,m}, \end{aligned} \quad (15)$$

where  $\mathbf{C}_{k,m} = (\Theta_k * \hat{\mathbf{x}}_k^{(m)})^T \text{sgn}(\Theta_k * \hat{\mathbf{x}}_k^{(m)})$  and  $\hat{\mathbf{x}}_k^{(m)}$  denotes the estimated state after  $m$  iterations with an initial condition. For instance, one can initialize it using  $\hat{\mathbf{x}}_k^{(0)} = (\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{y}_k$  which is obtained by minimizing the trace of observation error (i.e.,  $\mathbf{e}_o = \mathbf{y} - \mathbf{H}\mathbf{x}$ ) covariance matrix. Note that  $\mathbf{C}_{k,m}$  is considered as a constant value with respect to  $\mathbf{x}_k$ . Therefore, in order to solve (13), one can solve the fol-

lowing iterative optimization problem:

$$\begin{aligned} \ell_{1,k} = & \frac{1}{2} (\mathbf{y}_k - \mathbf{H}\mathbf{x}_k)^T \mathbf{R}_k^{-1} (\mathbf{y}_k - \mathbf{H}\mathbf{x}_k) \\ & + \frac{1}{2} (\Theta_k * \mathbf{x}_k)^T (\mathbf{Q}_k \mathbf{C}_{k,m})^{-1} (\Theta_k * \mathbf{x}_k) + \frac{1}{2} \mathbf{C}_{k,m}. \end{aligned} \quad (16)$$

The second regularization term is now differentiable. By setting the derivative of (16) with respect to  $\mathbf{x}_k$  to zero (inspired by [19, Lemma 2]), and after some simplification, we find the following solution for the filter impulse response:

$$\begin{aligned} \Psi_{\ell_1}(z) = & \left[ \mathbf{H}^T S_{\eta\eta}^{-1}(z) \mathbf{H} + \Theta^T(z^{-1}) \right. \\ & \left. \left[ S_{ww}(z) \mathbf{C}_m(z) \right]^{-1} \Theta(z) \right]^{-1} \mathbf{H}^T S_{\eta\eta}^{-1}(z) \end{aligned} \quad (17)$$

Comparing (17) with (8), they differ with the term  $S_{ww}(z) \mathbf{C}_m(z)$  which is simply  $S_{ww}(z)$  in equation (8). Therefore, we conclude that in  $\ell_1$  smoothing Wiener filter, the process noise is no longer a pure white Gaussian process but it is a weighted function of it. In other word, in  $\ell_1$  smoothing Wiener filter, the state-space model is converted to

$$\begin{cases} \mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{C}_{k,m} \mathbf{w}_k \\ \mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \eta_k \end{cases}, \quad (18)$$

where  $\mathbf{C}_{k,m}$  is the absolute value of the model error ( $\hat{\mathbf{x}}_k - \mathbf{A}\hat{\mathbf{x}}_{k-1}$ ) after each iteration. Based on the above discussion, in order to estimate the sparse state of a dynamical system

described by

$$\begin{cases} \mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{w}_k \\ \mathbf{y}_k = \mathbf{H}\mathbf{x}_k + \eta_k \end{cases}, \quad (19)$$

the paper suggests to estimate the state of a dynamical system described by (18), instead. Against the Kalman filter implementation of (19) that we run Kalman only one time over all  $k$ , we must run Kalman for (18) over all  $k$ , for a given iteration  $m$ ; and then this has to run multiple times for  $m = 1, 2, 3, \dots$ . We call the proposed method  $\ell_1$  Kalman filter/smoothen as it is based on the  $\ell_1$ -norm method. It is summarized in Algorithm 1.

---

**Algorithm 1**  $\ell_1$  Kalman filter

---

**Require:**  $\mathbf{y}_{1:T}, \mathbf{Q}_{1:T}, \mathbf{R}_{1:T}$

**for**  $m = 1, 2, 3, \dots$  **do**

**for**  $k = 1 : T$  **do**

$$\mathbf{\Lambda}_k = \hat{\mathbf{x}}_k - \mathbf{A}\hat{\mathbf{x}}_{k-1}$$

$$\mathbf{C}_k = \mathbf{\Lambda}_k^T \text{sgn}(\mathbf{\Lambda}_k)$$

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{A}\hat{\mathbf{x}}_{k-1|k-1}$$

$$\mathbf{P}_{k|k-1} = \mathbf{A}\mathbf{P}_{k-1|k-1}\mathbf{A}^T + \mathbf{C}_k^T \mathbf{Q}_k \mathbf{C}_k$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k [\mathbf{y}_k - \mathbf{H}\hat{\mathbf{x}}_{k|k-1}]$$

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}^T (\mathbf{H}\mathbf{P}_{k|k-1} \mathbf{H}^T + \mathbf{R}_k)^{-1}$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{H} \mathbf{P}_{k|k-1}$$

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k}$$

**end for**

**end for**

**return**  $\hat{\mathbf{x}}_{k|k}$  for  $k = 1 : T$

---

The classical Kalman filter/smoothen is called  $\ell_2$  Kalman filter/smoothen as it is based on the  $\ell_2$ -norm method. Note that if we set  $\mathbf{C}_{k,m}$  in (18) to identity matrix, then (19) and (18) are equivalent which means the conventional  $\ell_2$  Kalman filter/smoothen is a special case of the proposed  $\ell_1$  Kalman filter/smoothen. The proposed technique is used to estimate the target states of the previous example. To this purpose, we changed the system model (3) to

$$\mathbf{x}_k = \begin{bmatrix} 1 & T_s & \frac{T_s^2}{2} \\ 0 & 1 & T_s \\ 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{k-1} + \mathbf{C}_{k,m} \mathbf{w}_k, \quad (20)$$

where  $\mathbf{C}_{k,m} = \mathbf{\Lambda}_{k,m}^T \text{sgn}(\mathbf{\Lambda}_{k,m})$ ,

$$\mathbf{\Lambda}_{k,m} = \hat{\mathbf{x}}_k^{(m)} - \begin{bmatrix} 1 & T_s & \frac{T_s^2}{2} \\ 0 & 1 & T_s \\ 0 & 0 & 1 \end{bmatrix} \hat{\mathbf{x}}_{k-1}^{(m)}, \quad \hat{\mathbf{x}}_k^{(0)} = \frac{1}{3} \mathbf{y}_k. \quad (21)$$

$\mathbf{\Lambda}_{k,m}$  is the system model error after each iteration and  $\mathbf{C}_{k,m}$  defines its absolute value. Figure 3 illustrates the result of the Kalman smoothen for tracking the target states in the previous example using the state space model (20) after 5 iterations ( $m = 5$ ). The solid red curve and dots gray in Figures 3(a)-3(c) denote the theoretical state and its noisy data, respectively, which are the same as those in Figure 1. The estimated acceleration by Kalman smoothen using the modified model (20) is plotted in Figure 3(c) with the blue dashed line. The estimated acceleration has sharp discontinuities which demonstrates that the proposed model, which is based on the  $\ell_1$  regularization and is less sensitive to large changes, is more accurate than the traditional Kalman filter/smoothen model. Note that we did not modify the Kalman filter equations, but we showed that by a simple modification to system model, Kalman filter can be used to optimally estimate the states of the system even if the the latent or observed states have non-Gaussian (e.g., impulsive, Laplacian) distribution.

## Examples

The modified structure can find various applications in state estimation in non-Gaussian error statistics. As proof of concept, we focus on three specific examples.

### Detection of a sine wave with known frequency in Impulsive/Laplace noise

In the following example, we seek to estimate a sine wave of known frequency [ $x_k = \alpha \cos(2\pi f k + \phi)$ , where  $\alpha$ ,  $f$  and  $\phi$  are amplitude, frequency and phase, respectively] in the presence of impulsive or Laplace noise. The dynamical model of the system can be represented as

$$\begin{cases} \mathbf{x}_{k+1} = \begin{bmatrix} 2 \cos(2\pi f) & -1 \\ 1 & 0 \end{bmatrix} \mathbf{x}_k + \begin{bmatrix} 1 \\ 0 \end{bmatrix} w_k \\ y_k = [1 \ 0] \mathbf{x}_k + \eta_k \end{cases}, \quad (22)$$

where  $\mathbf{x}_k = [x_k \ x_{k-1}]^T$  and  $\eta_k$  is assumed to be an impulsive or a Laplace noise. Note that the dynamical model

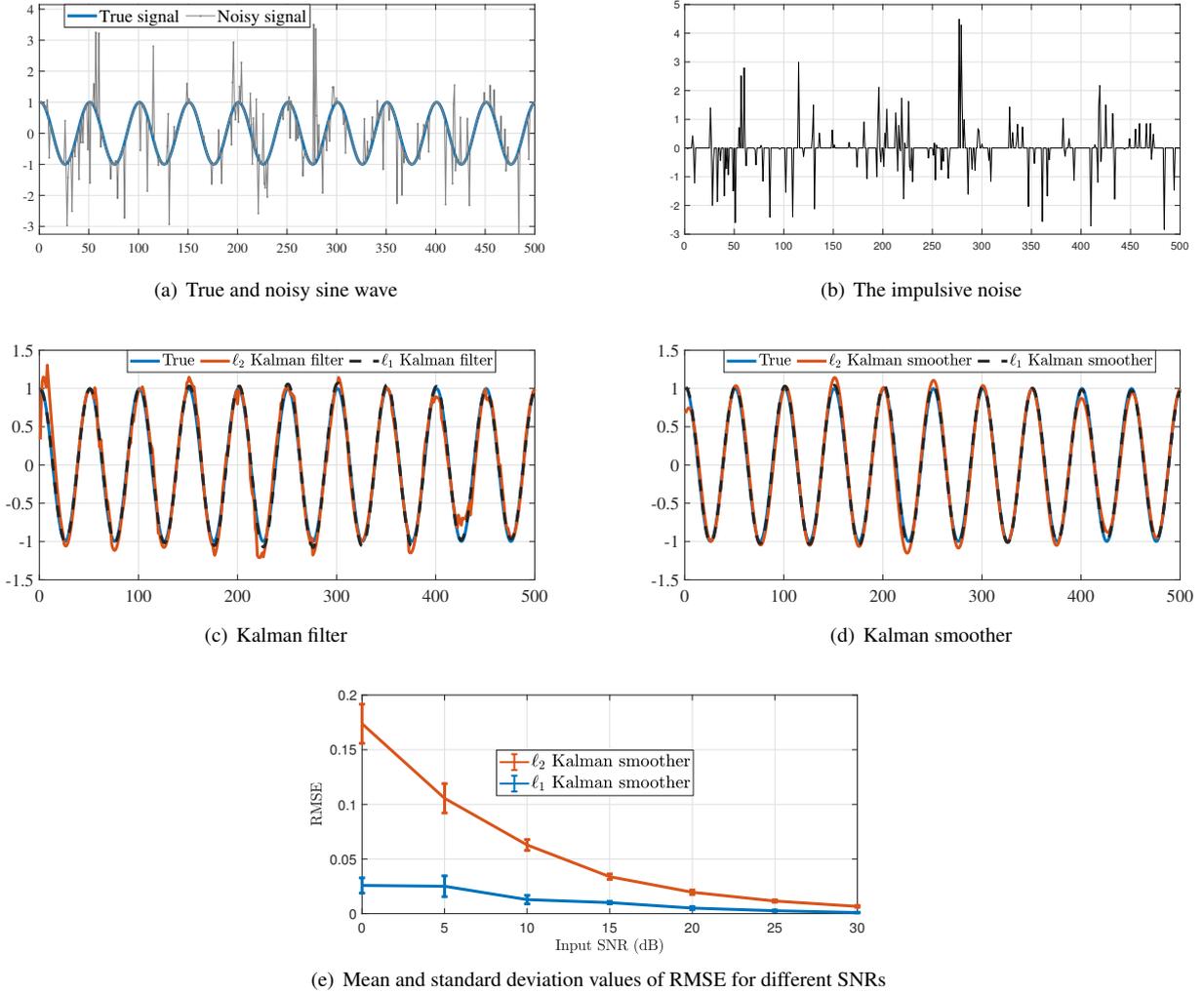


Figure 4: Sine wave with known frequency detection in Laplace noise using Kalman filter and smoother. a) true and noisy signal b) the result obtained by  $\ell_2$  and  $\ell_1$  Kalman filter c) the result obtained by  $\ell_2$  and  $\ell_1$  Kalman smoother d) the mean and standard deviation values of RMSE obtained by  $\ell_2$  and  $\ell_1$  Kalman smoother for different SNRs.

(22) no longer depends on the amplitude and phase. We generated 2500 synthetic sine waves with different frequencies and contaminated them with impulsive and Laplace noise. To this purpose, we produced signals varying the power of  $\eta_k$  in (22). The signal-to-noise ratio (SNR) was modulated from 0 to 30 dB. The Kalman filter and smoother were then used to estimate the desired signal  $x_k$ . An example of a sine wave estimation in impulsive noise provided by Kalman filter and smoother using (22) is illustrated in figure 4. Figure 4(a) shows the sine wave and its noisy measurements contaminated by impulsive noise. The noise is plotted in Figure 4(b). The estimated signals obtained by classical Kalman filter and smoother are illustrated via red color in Figure 4(c) and 4(d), respectively. Again, we see that the classical Kalman filter and smoother (i.e.,  $\ell_2$  Kalman filter/smoothing) suffer

from the sensitivity to impulsive error statistics. In order to improve the estimation, we employed the modified Kalman filter and smoother (i.e.,  $\ell_1$  Kalman filter and smoother) to estimate the desired signal. The results of sine wave detection provided by Kalman filter and smoother using the modified model are shown in Figure 4(c) and 4(d) via black dashed curves. The modified model improves the Kalman filter and smoother's performance for signal detection in impulsive noise. Finally, we employed Kalman filter and smoother to estimate all generated signals with and without modified model. The mean and standard deviation values of root mean squares error (RMSE) for signal reconstruction obtained by  $\ell_2$  and  $\ell_1$  Kalman smoother as a function of SNR is shown in Figure 4(e). The results confirm that the modified structure can be used to optimize the Kalman filter and smoother

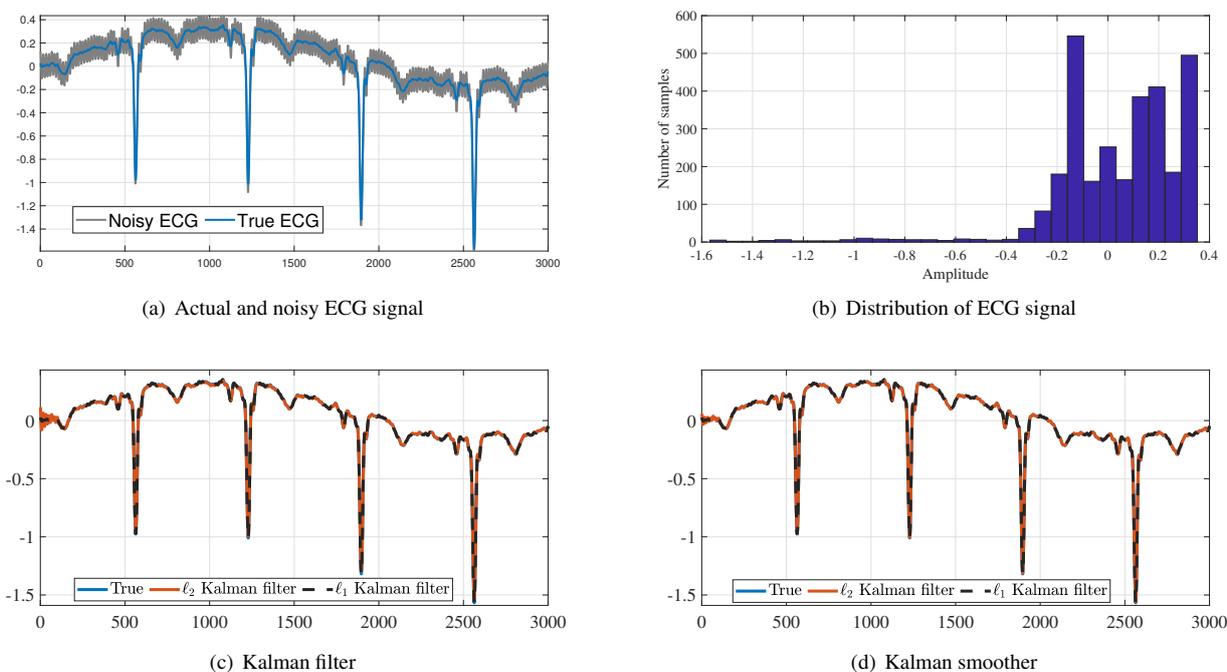


Figure 5: ECG powerline noise cancellation using Kalman filter and smoother. a) true and noisy ECG signal b) distribution of the ECG signal c) the result obtained by  $\ell_2$  and  $\ell_1$  Kalman filter e) the result obtained by  $\ell_2$  and  $\ell_1$  Kalman smoother.

to not lose their optimality in applications that the system errors have impulsive or Laplace distribution. As a real application, (22) can be used to model the powerline noise and to track and remove it from bioelectrical signals, e.g., electrocardiogram (ECG) signal. As an illustration, a real ECG signal contaminated by powerline noise is plotted in Figure 5(a). To remove the powerline, we first estimate it using Kalman filter/smoothing and then subtract it from the observation signal. The denoised ECG provided by  $\ell_2$  and  $\ell_1$  Kalman filter and smoother are plotted in Figure 5(c) and 5(d), respectively.

### Piecewise-polynomial signal detection in Laplace noise

In this example, we consider the problem of detecting a piecewise-constant signal  $x_k$  in additive Laplace noise. The dynamical model of the system can be represented as

$$\begin{cases} x_k = x_{k-1} + w_k \\ y_k = x_k + \eta_k \end{cases}, \quad (23)$$

where  $w_k$  and  $\eta_k$  are non-Gaussian impulsive noises. The proposed technique was employed to detect piecewise-constant signal in Laplace noise. An example of piecewise-constant signal estimation in Laplace noise using Kalman filter and

smoother is illustrated in figure 6. Note that both process and observation errors have non-Gaussian distribution. We repeated the same experiments as done in the previous section for piecewise-constant signal detection in Laplace noise. The results of comparison between these two approaches are summarized in Figure 6(e) which show that the proposed technique improves the Kalman filter performance in non-Gaussian impulsive/Laplace model errors. The extension of the proposed approach to piecewise-polynomial signals (e.g., piecewise-linear, piecewise-quadratic and etc.) contaminated with Laplace noise is straightforward.

### Detection of a sine wave with known frequency in additive Cauchy noise

In this section, we consider the problem of estimating a sine wave corrupted by an additive Cauchy noise. Cauchy noise which frequently appears in engineering applications is particularly interesting as it does not have finite moments of order greater than or equal to one. It is a kind of impulsive non-Gaussian noise which its distribution looks similar to a normal distribution, but it has much heavier tails. In Figure 7, we give an example of sine wave detection in Cauchy noise using  $\ell_2$  and  $\ell_1$  Kalman filter and smoother. In this example, the parameters of the sine wave are chosen as  $\alpha = 1$ ,

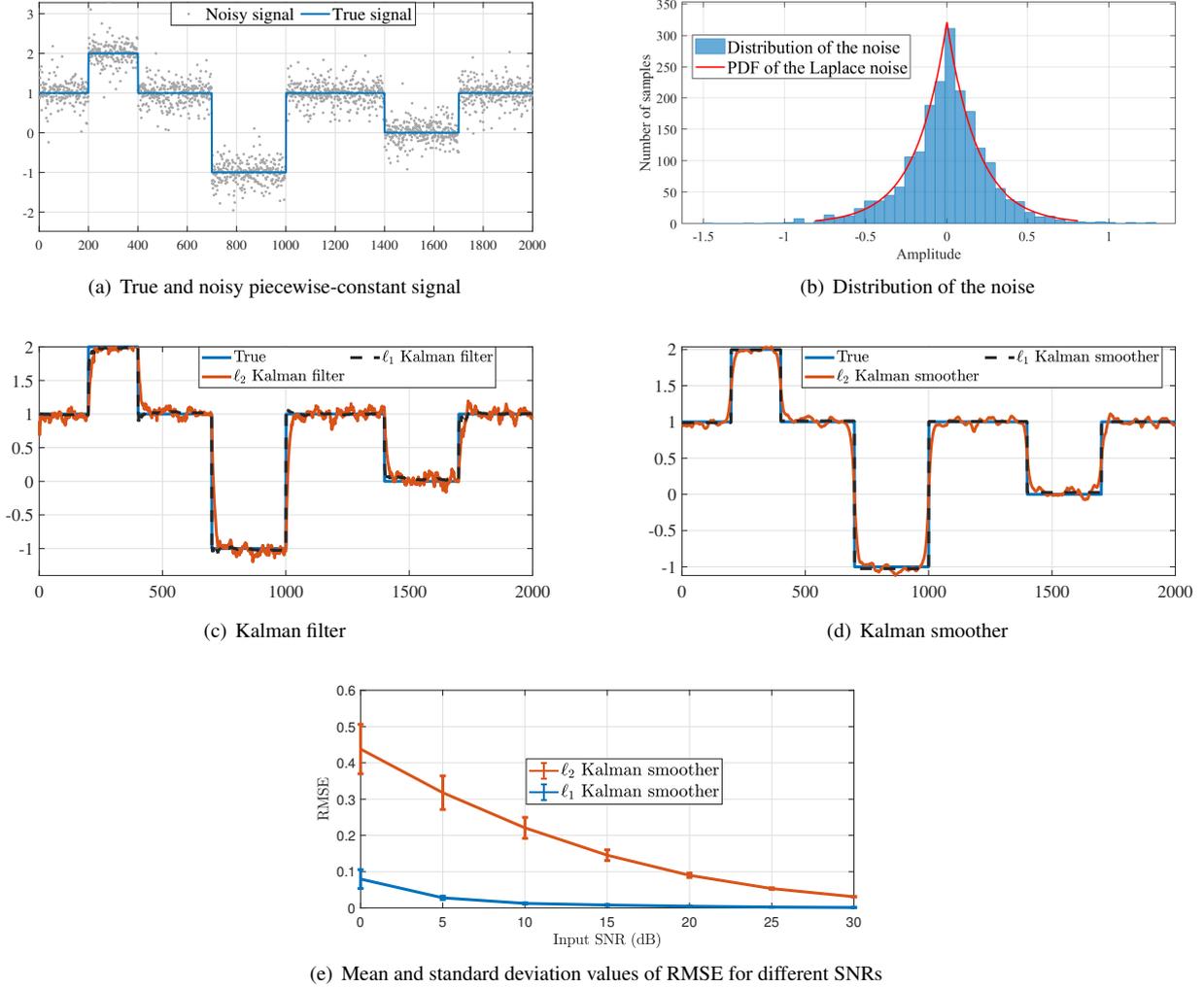


Figure 6: Piecewise-constant signal detection in Laplace noise using Kalman filter and smoother. a) true and noisy constant signal b) distribution and PDF of the noise c) the result obtained by  $\ell_2$  and  $\ell_1$  Kalman filter d) the result obtained by  $\ell_2$  and  $\ell_1$  Kalman smoother e) the mean and standard deviation values of RMSE obtained by  $\ell_2$  and  $\ell_1$  Kalman smoother for different SNRs.

$f = 0.02$  and  $\phi = 0$ . The Cauchy noise was generated from the Cauchy distribution,  $r = \kappa_1 + \kappa_2 \tan \pi(\text{rand} - 0.5)$ , where  $\kappa_1$  is the statistical median and  $\kappa_2$  is the half width at the half maximum density level. In this example, we set  $\kappa_1 = 0.6$  and  $\kappa_2 = 0.05$ . The estimated sine waves using  $\ell_2$  and  $\ell_1$  Kalman filter/smoothers for this specific example are shown in Figure 7(c) and 7(d). We also repeated the experiment for different sine waves with different frequencies and phases. To this purpose, we generated different Cauchy noises with different values of  $0.001 \leq \kappa_2 \leq 0.05$  and compared  $\ell_2$  and  $\ell_1$  Kalman filter/smoothers. We also computed the cross-correlation between the original and estimated sine wave. The mean values of RMSE and cross-correlation obtained by these two methods are reported in Figure 7(e) and

7(f), respectively. In the best case of signal reconstruction, the cross-correlation should be close to 1. The results show that  $\ell_1$  Kalman smoother outperforms  $\ell_2$  Kalman smoother.

## Summary

The classical Kalman filter is known as an optimal estimator for linear systems subject to Gaussian error statistics and thus suffers from the sensitivity to sparse or Laplace error statistics. In this article, we proposed a simple modification approach to system model that can be used to overcome this limitation of the Kalman filter. We showed that the smoothing Wiener filter can be viewed as a constrained optimization

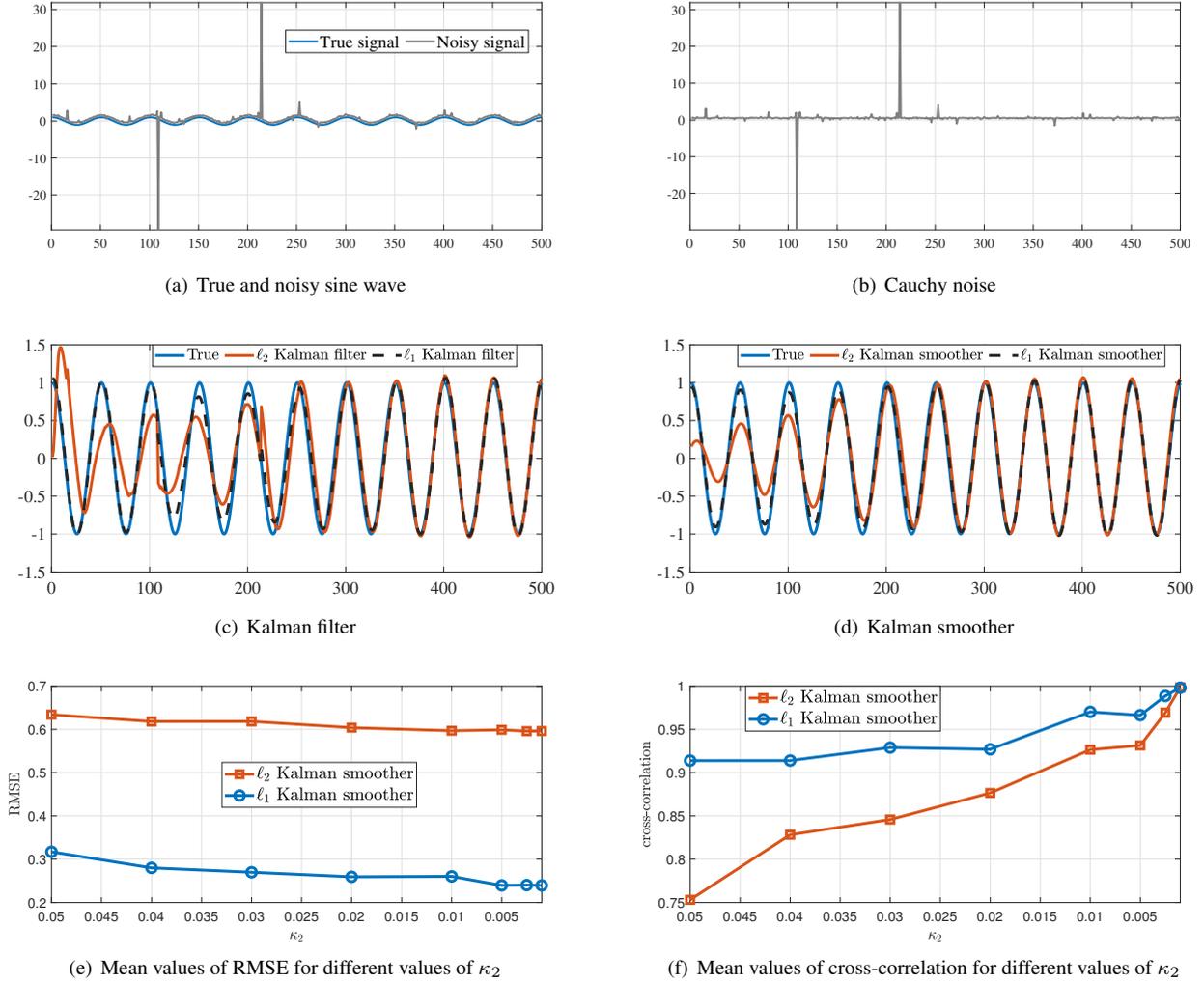


Figure 7: Sine wave detection in Cauchy noise using Kalman filter and smoother. a) true and noisy signal b) the Cauchy noise c) the result obtained by  $\ell_2$  and  $\ell_1$  Kalman filter d) the result obtained by  $\ell_2$  and  $\ell_1$  Kalman smoother. e) the mean values of RMSE and cross-correlation between the original and estimated signal obtained by  $\ell_2$  and  $\ell_1$  Kalman smoother for different values of  $\kappa_2$ .

problem that minimizes the  $\ell_2$ -norm of the system model error subject to the  $\ell_2$ -norm of the data fidelity. Then a variation on smoothing Wiener filter was proposed that substitutes a sum of absolute values (i.e.,  $\ell_1$ -norm) for the sum of squares used in  $\ell_2$  smoothing Wiener filter to penalize variations in the model error. The idea of  $\ell_1$  smoothing Wiener filter was then used to correct the dynamic model. The Kalman filter/smoothing can be used to estimate the states using the modified dynamic model. Note that we did not modify the Kalman filter equations, but we have shown that by recasting the state space model, the Kalman filter can be used to estimate the states of the system even if the process or observation errors have non-Gaussian (e.g., impulsive/Laplacian) distribution.

Although the proposed  $\ell_1$  Wiener filter or Kalman filter is suitable for systems with Laplace or sparse distribution, there are some improvements that can be done in the future. The extension of the proposed method to  $\ell_p$  Wiener filter or Kalman filter can be considered as one of the future works. Specially, for the lower value of  $p$ , i.e.,  $0 < p < 1$ , the proposed approach can be used for estimating the states of a system with very impulsive model error.

## Author

Arman Kheirati Roonizi (ebad.kheirati.roonizi@gmail.com) received his Ph.D. degree in computer science from the University of Milan, Italy, in 2017. During 2017-2018, he held

a Postdoctoral Research position with GIPSA-lab, Grenoble, France. He is currently an assistant Professor in the Department of Computer Science, Fasa University, Iran. His re-

search interests include signal modeling, processing, filtering and separation, multimodal signal processing, and functional data analysis.

## References

- [1] BDO Anderson and JB Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ, 1979.
- [2] D. Simon. *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. Wiley, 2006.
- [3] Ramsey Faragher. Understanding the Basis of the Kalman Filter Via a Simple and Intuitive Derivation. *IEEE Signal Process. Mag.*, 50:128–132, 2012.
- [4] S. Kotz, T. Kozubowski, and K. Podgorski. *The Laplace Distribution and Generalizations: A Revisit with Applications to Communications, Economics, Engineering, and Finance*. Progress in Mathematics. Birkhäuser Boston, 2001.
- [5] Jerry M Mendel. *Lessons in Digital Estimation Theory*. Prentice-Hall, Inc., USA, 1986.
- [6] R. E. Kalman. A New Approach to Linear Filtering and Prediction Problems. *J. Basic Eng.*, 82:35–45, 1960.
- [7] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series : with engineering applications*. Technology Press of the Massachusetts Institute of Technology, London, 1949.
- [8] Zhou Wang and Alan C. Bovik. Mean squared error: Love it or leave it? A new look at signal fidelity measures. *IEEE Signal Process. Mag.*, 26(1):98–117, 2009.
- [9] Hyun Park and Te Lee. Modeling Nonlinear Dependencies in Natural Images using Mixture of Laplacian Distribution. In L. Saul, Y. Weiss, and L. Bottou, editors, *Advances in Neural Information Processing Systems*, volume 17. MIT Press, 2005.
- [10] D. Middleton. Non-Gaussian noise models in signal processing for telecommunications: new methods an results for class A and class B noise models. *IEEE Trans. Inf. Theory*, 45(4):1129–1149, 1999.
- [11] Mital A. Gandhi and Lamine Mili. Robust Kalman Filter Based on a Generalized Maximum-Likelihood-Type Estimator. *IEEE Trans. Signal Process*, 58(5):2509–2520, 2010.
- [12] Junbo Zhao, Marcos Netto, and Lamine Mili. A Robust Iterated Extended Kalman Filter for Power System Dynamic State Estimation. *IEEE Trans. Power Syst.*, 32(4):3205–3216, 2017.
- [13] Peter J. Huber. *Robust Statistics*, pages 1248–1251. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [14] Junbo Zhao and Lamine Mili. A Robust Generalized-Maximum Likelihood Unscented Kalman Filter for Power System Dynamic State Estimation. *IEEE J Sel Top Signal Process*, 12(4):578–592, 2018.
- [15] W.H. Press, B.P. Flannery, S. Teukolsky, and W.T. Vetterling. *Numerical recipes, The art of scientific computing*. Cambridge University Press, New York, NY, Jan 1986.
- [16] T. Kailath. *Lectures on Wiener and Kalman Filtering*. CISM International Centre for Mechanical Sciences Series. Springer, 1981.
- [17] H.L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory*. Wiley, 2004.

- [18] Y. Sun, P. Babu, and D. P. Palomar. Majorization-minimization algorithms in signal processing, communications, and machine learning. *IEEE Trans. Signal Process.*, 65(3):794–816, 2017.
- [19] Arman Kheirati Roonizi and Christian Jutten. Forward-backward filtering and penalized least-squares optimization: A unified framework. *Signal Processing*, 178:107796, 2021.