Arman Kheirati Roonizi

Kalman Filtering in Non-Gaussian Model Errors: A New Perspective

It is well known that the optimality of the Kalman filter relies on the Gaussian distribution of process and observation model errors, which in many situations is well justified [1]–[3]. However, this optimality is useless in applications where the distribution assumptions of the model errors do not hold in practice. Even minor deviations from the assumed (or nominal) distribution may cause the Kalman filter's performance to drastically degrade or completely break down. In particular, when dealing with perceptually important signals, such as speech, image, medical, campaign, and ocean engineering, measurements have confirmed the presence of non-Gaussian impulsive (heavy-tailed) and Laplace noises [4]. Therefore, the classical Kalman filter, which is derived under the nominal Gaussian probability model, is biased and even fails in such situations.

This article presents a simple modification to overcome this limitation of the Kalman filter. We show that in the smoothing Wiener filter, the estimated state is chosen to minimize the ℓ_2 norm of the variation of the system model error. The ℓ_2 norm, however, corresponds to Gaussian priors. It confirms the optimality of the Wiener and Kalman filters in linear Gaussian systems and explains why they suffer from sensitivity to Laplace and impulsive error statistics. To address this limitation, we propose a variation of the smoothing Wiener filter that substitutes a sum of absolute values (i.e., the ℓ_1 norm) for the sum of the squares used in the ℓ_2 smoothing Wiener filter to penalize variations in the system model error. The proposed ℓ_1 smoothing Wiener filter is suitable for analyzing systems with impulsive and Laplace model errors. The ℓ_1 smoothing Wiener filter is then used to recast and correct the system (or process) model equation. The modified model puts a Laplace or 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 the discrete-time steady state [5], the Kalman filter can be used to estimate the desired states through the modified model.

The optimality of the Kalman filter relies on the Gaussian distribution of process and observation model errors.

Introduction

This year, the Kalman filter turns 60. Introduced by Rudolf E. Kálmán, in his best-known contribution [6], it 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* and *Kalman filter* + *application* and discover thousands of other applications.

Before the Kalman filter, the Wiener filter, introduced by Nobert Wiener, was probably the first theory for merging dynamic systems and optimal estimation in the presence of noise [7]. It is a signal estimation algorithm that produces an optimal approximation of a random signal, using both the autocorrelation and cross correlation of a measurement signal with an original signal. The Wiener filter minimizes the MSE between the estimated random process and desired process. The MSE has long been regarded as the dominant quantitative performance metric for optimizing and assessing the Wiener filter and Kalman filter design. It is simple, memoryless, parameter free, and inexpensive to compute. In particular, 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, including Wiener and Kalman filters.

Digital Object Identifier 10.1109/MSP.2021.3134635
Date of current version: 27 April 2022

The MSE is defined by the ℓ_2 norm, which corresponds to Gaussian priors. This is the reason why 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 a Gaussian distribution). However, in many real applications, the MSE (or ℓ_2 regularization) exhibits weak performance and has been widely criticized for serious shortcomings, especially when the latent and observed variables are distributed in a non-Gaussian manner, for instance, when dealing with signals such as finance and ocean engineering, speech, and images, where a sparse distribution is appropriate to represent relevant statistics [4]. Impulse or sparse distribution also appears in various machine learning areas, including independent component analysis [9] and mobile and wireless communication [10]. The classical Wiener and Kalman filters are derived under Gaussian distribution assumptions and thus suffer from sensitivity to sparse and Laplace error statistics. The main purpose of this article is to provide a modified structure to optimize the Wiener and Kalman filters so that they do not lose their optimality in applications where system variables have an impulsive or Laplace distribution.

Recently, non-Gaussian robust Kalman filtering has been attracting more attention. A generalized maximum-likelihood Kalman filter (GMKF) was derived in [11] and extended in [12] through 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. The 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 algorithm. While most published algorithms (see [14] and the references therein) are advanced, in our method, a simple modification is applied to the system model, and the standard Kalman filter is employed for a given iteration. We neither modify the Kalman filter equations nor use other advanced techniques, but we show that through a simple change, the standard Kalman filter can be used to estimate system states even if the system model errors have non-Gaussian distribution.

Even minor deviations from the assumed (or nominal) distribution may cause the Kalman filter's performance to drastically degrade or completely break down.

Notation

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

Background

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

The adaptive Kalman filter is regarded as the optimal state estimator when process and observation noises are white Gaussian processes. It involves two stages: a prediction update,

$$\begin{cases}\hat{\boldsymbol{x}}_{k|k-1} = \boldsymbol{A}\hat{\boldsymbol{x}}_{k-1|k-1}\\ \boldsymbol{P}_{k|k-1} = \boldsymbol{A}\boldsymbol{P}_{k-1|k-1}\boldsymbol{A}^T + \boldsymbol{Q}_k\end{cases}, \tag{1}$$

and a measurement update,

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

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

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

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

where T_s is the sampling period and $\boldsymbol{x}_k = [p_k, v_k, a_k]^T$ contains the position, velocity, and acceleration. We assume that while the truck moves with a CA, the driver sometimes applies a sudden braking or acceleration input to the system, which produces instantaneous changes in the acceleration values. In this case, a discrete event phenomenon is observed in the presence of the CA signal, which is modeled as process noise; i.e., $\boldsymbol{w}_k = [T_s^2/2, T_s, 1]^T w_k$. 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 $\boldsymbol{y}_k = \boldsymbol{x}_k + \boldsymbol{\eta}_k$, where $\boldsymbol{y}_k = [y_{1,k}, y_{2,k}, y_{3,k}]^T$ 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 $\boldsymbol{y}_k$. The Kalman filter/smoother is then used to estimate the target states. The true and noisy states appear as red solid lines and gray dots in Figure 1(a)–(c), and the estimated target states provided by Kalman smoother are depicted as blue dashed lines.

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 the nominal Gaussian probability model (i.e., all latent and observed variables are assumed to have Gaussian distribution). In other words, the process noise in (3) is assumed to be white Gaussian noise by the Kalman filter.

As an illustration, the distribution of the actual acceleration process model error is plotted in blue in Figure 2, where the distribution of the nominal acceleration process model error with different values of Q (the covariance of the process noise, w_k) is in yellow. It is evident 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

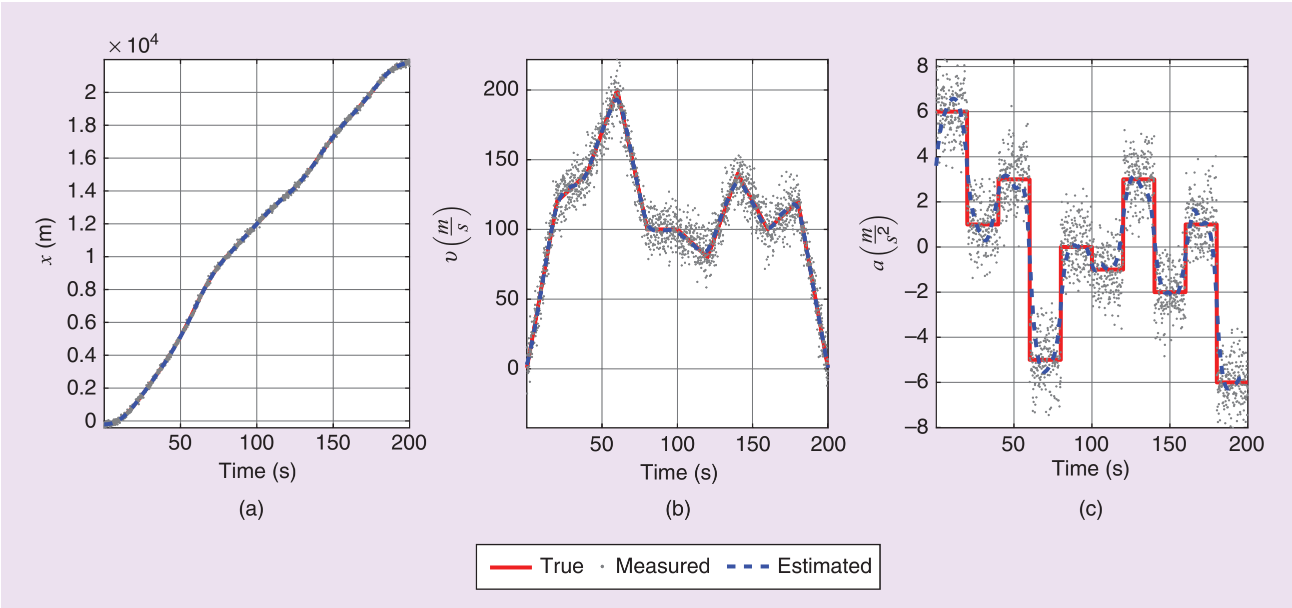

FIGURE 1. The results of adaptive Kalman smoother-based target state estimation. The (a) position, (b) velocity, and (c) acceleration.

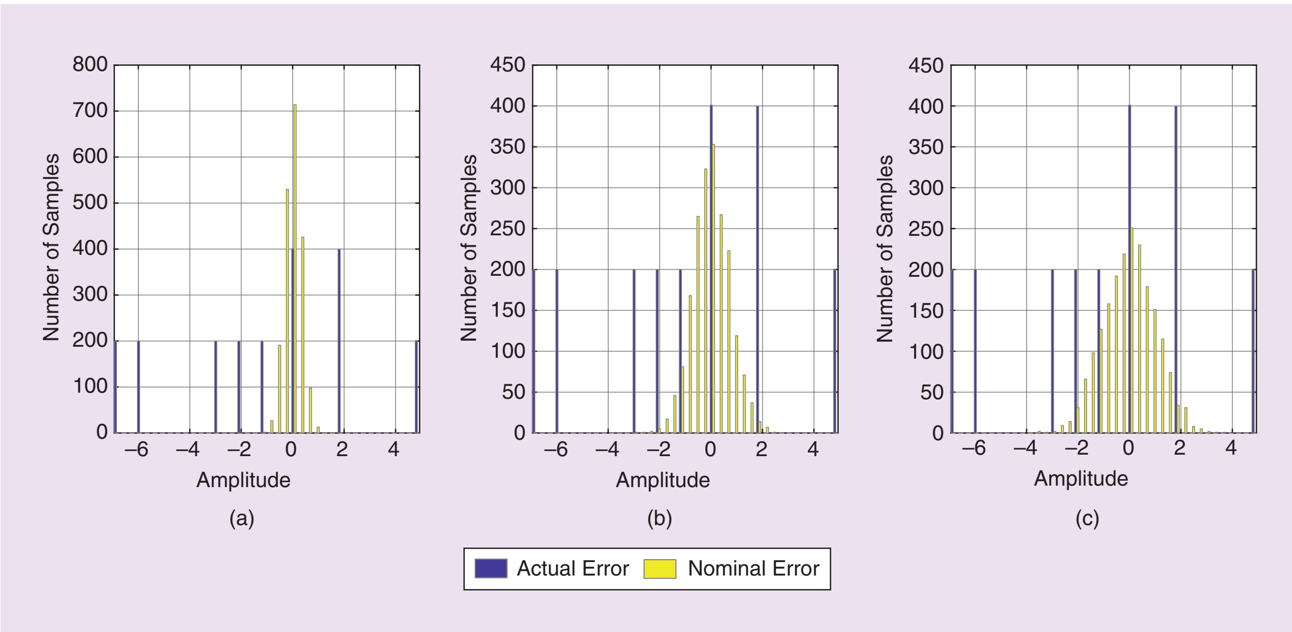

FIGURE 2. The acceleration model error. (a) $Q = 0.1$. (b) $Q = 0.5$. (c) $Q = 1$.

actual process model error, and a (sparse) fat tail distribution can provide a better representation to describe the model error than the normal distribution. Therefore, the optimality of the Kalman filter, which relies on the Gaussian distribution of the process and observation noises, is not supported in such situations.

Through a simple change, the standard Kalman filter can be used to estimate system states even if the system model errors have non-Gaussian distribution.

In this article, we analyze the preceding issue from a different perspective. To that end, we develop a joint perspective on Wiener and Kalman filtering. Note that the formulations of the optimum filter by Wiener and Kalman are equivalent in a steady state [5]. On the other hand, the smoothing Wiener filter can be viewed as an optimization problem with an ℓ_2 norm smoothing constraint on the variation of the process (system) model, as will become clear following the derivation outlined in the section "The ℓ_2 Smoothing Wiener Filter." The ℓ_2 norm corresponds to Gaussian priors and is sensitive to large values. Therefore, it exhibits weak performance when analyzing sparse data, such as the target acceleration in the previous example. That is why the estimated acceleration by the Kalman smoother does not accurately follow the true acceleration signal.

The remainder of the article describes a novel approach to overcome this limitation. We show that in the smoothing Wiener filter, the estimated state is chosen to minimize the ℓ_2 norm of the variation of the system model error, subject to the ℓ_2 norm of the data fidelity. Then, we propose a variation on the smoothing Wiener filter that substitutes a sum of absolute values (i.e., the ℓ_1 norm) for the sum of squares used in the ℓ_2 smoothing Wiener filter to penalize variations in the system model error. The ℓ_1 norm corresponds to Laplacian priors. Therefore, the proposed ℓ_1 smoothing Wiener filter is suitable for analyzing sparsely distributed data. The ℓ_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 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 a steady state, the Kalman filter can be employed to estimate the sparse states from the modified state-space model.

Wiener filter

Smoothing Wiener filter

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

$$\hat{\boldsymbol{x}}_k = \boldsymbol{\Psi}_k * \boldsymbol{y}_k \approx \boldsymbol{x}_k, \tag{4}$$

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

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

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

$$\begin{aligned}\boldsymbol{\Psi}(z) &= S_{\boldsymbol{xy}}(z) S_{\boldsymbol{yy}}^{-1}(z)\\ &= S_{\boldsymbol{xx}}(z)\boldsymbol{H}^T\\ &\quad \times [\boldsymbol{H} S_{\boldsymbol{xx}}(z)\boldsymbol{H}^T + S_{\boldsymbol{\eta\eta}}(z)]^{-1},\end{aligned} \tag{6}$$

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

$$\begin{aligned}\boldsymbol{\Psi}(z) &= [\boldsymbol{H}^T S_{\boldsymbol{\eta\eta}}^{-1}(z)\boldsymbol{H} + S_{\boldsymbol{xx}}^{-1}(z)]^{-1}\\ &\quad \times \boldsymbol{H}^T S_{\boldsymbol{\eta\eta}}^{-1}(z).\end{aligned} \tag{7}$$

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

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

In the following, the output response of the smoothing Wiener filter is computed using an optimization problem with an ℓ_2 smoothing constraint on the variation of the process (system) model.

The ℓ_2 smoothing Wiener filter

In this section, we show that in the smoothing Wiener filter, the estimated state is chosen to minimize the ℓ_2 norm of the process (or system) model error. For this purpose, we represent the output of (8) in the transform domain as $\hat{\boldsymbol{X}}(z) = \boldsymbol{\Psi}(z)\boldsymbol{Y}(z)$. According to (8), this can be written as

$$\begin{aligned}&[\boldsymbol{H}^T S_{\boldsymbol{\eta\eta}}^{-1}(z)\boldsymbol{H} + \boldsymbol{\Theta}^T(z^{-1}) S_{\boldsymbol{ww}}^{-1}(z)\boldsymbol{\Theta}(z)]\\ &\quad \times \hat{\boldsymbol{X}}(z) = \boldsymbol{H}^T S_{\boldsymbol{\eta\eta}}^{-1}(z)\boldsymbol{Y}(z).\end{aligned} \tag{9}$$

Equation (9) can be expressed as

$$\begin{aligned}&\boldsymbol{\Theta}^T(z^{-1}) S_{\boldsymbol{ww}}^{-1}(z)\boldsymbol{\Theta}(z)\hat{\boldsymbol{X}}(z) + \boldsymbol{H}^T S_{\boldsymbol{\eta\eta}}^{-1}(z)\\ &\quad \times \boldsymbol{H}\hat{\boldsymbol{X}}(z) - \boldsymbol{H}^T S_{\boldsymbol{\eta\eta}}^{-1}(z)\boldsymbol{Y}(z) = 0.\end{aligned} \tag{10}$$

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

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

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

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

where $\boldsymbol{\Theta}_k$ is the inverse z transform of $\boldsymbol{\Theta}(z) = \boldsymbol{I} - z^{-1}\boldsymbol{A}$; i.e., $\boldsymbol{\Theta}_k = \boldsymbol{\delta}_k - \boldsymbol{A}\boldsymbol{\delta}_{k-1}, \boldsymbol{\Theta}_k * \boldsymbol{x}_k = \boldsymbol{x}_k - \boldsymbol{A}\boldsymbol{x}_{k-1}$; $\boldsymbol{R}_k$ and $\boldsymbol{Q}_k$ were defined in the "Background" section. The first term in (12) describes a weighted ℓ_2 norm of the observation model error, while the second one describes a weighted ℓ_2 norm of the process model error. In the following, we propose a variation on the optimization problem to design a smoothing Wiener filter that can be used to overcome the issue discussed in the previous section.

Solution

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

$$\ell_{1,k}=\frac{1}{2}(\boldsymbol{y}_k-\boldsymbol{H}\boldsymbol{x}_k)^T\boldsymbol{R}_k^{-1}(\boldsymbol{y}_k-\boldsymbol{H}\boldsymbol{x}_k) + (\boldsymbol{\Theta}_k*\boldsymbol{x}_k)^T\boldsymbol{Q}_k^{-1}\operatorname{sgn}(\boldsymbol{\Theta}_k*\boldsymbol{x}_k), \tag{13}$$

where "sgn" is the sign (or signum) function:

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

For matrices, the sign function is defined elementwise. Equation (13) employs a sum of absolute values (i.e., the ℓ_1 norm) to penalize variations in the process model error, so we call it the ℓ_1 *smoothing Wiener filter*. The problem is that the second regularization term is nondifferentiable, which makes this a difficult minimization problem with no explicit solution. One way to deal with this is to replace the problem with a sequence of easier ones. This procedure is known as *majorization–minimization* (*MM*) [18]. Using the MM approach, (13) is converted to be simpler. For this, the MM technique proposes the following majorizer for $(\boldsymbol{\Theta}_k*\boldsymbol{x}_k)^T\boldsymbol{Q}_k^{-1}\operatorname{sgn}(\boldsymbol{\Theta}_k*\boldsymbol{x}_k)$ [18]:

$$(\boldsymbol{\Theta}_k*\boldsymbol{x}_k)^T\boldsymbol{Q}_k^{-1}\operatorname{sgn}(\boldsymbol{\Theta}_k*\boldsymbol{x}_k) \le \frac{1}{2}(\boldsymbol{\Theta}_k*\boldsymbol{x}_k)^T\boldsymbol{C}_{k,m}^{-1}\boldsymbol{Q}_k^{-1}(\boldsymbol{\Theta}_k*\boldsymbol{x}_k) + \frac{1}{2}\boldsymbol{C}_{k,m}, \tag{15}$$

where $\boldsymbol{C}_{k,m}=(\boldsymbol{\Theta}_k*\hat{\boldsymbol{x}}_k^{(m)})^T\operatorname{sgn}(\boldsymbol{\Theta}_k*\hat{\boldsymbol{x}}_k^{(m)})$ and $\hat{\boldsymbol{x}}_k^{(m)}$ denotes the estimated state after m iterations, with an initial condition.

For instance, one can initialize it using $\hat{\boldsymbol{x}}_k^{(0)}=(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{y}_k$, which is obtained by minimizing the trace of the observation error (i.e., $\boldsymbol{e}_o=\boldsymbol{y}-\boldsymbol{H}\boldsymbol{x}$) covariance matrix. Note that $\boldsymbol{C}_{k,m}$ is considered a constant value with respect to $\boldsymbol{x}_k$. Therefore, to solve (13), one can solve the following iterative optimization problem:

$$\ell_{1,k}=\frac{1}{2}(\boldsymbol{y}_k-\boldsymbol{H}\boldsymbol{x}_k)^T\boldsymbol{R}_k^{-1}(\boldsymbol{y}_k-\boldsymbol{H}\boldsymbol{x}_k) + \frac{1}{2}(\boldsymbol{\Theta}_k*\boldsymbol{x}_k)^T(\boldsymbol{Q}_k\boldsymbol{C}_{k,m})^{-1}(\boldsymbol{\Theta}_k + \frac{1}{2}\boldsymbol{C}_{k,m}. \tag{16}$$

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

$$\boldsymbol{\Psi}_{\ell_1}(z)=[\boldsymbol{H}^TS_{\eta\eta}^{-1}(z)\boldsymbol{H}+\boldsymbol{\Theta}^T(z^{-1})[S_{ww}(z)\boldsymbol{C}_m(z)]^{-1}\boldsymbol{\Theta}(z)]^{-1}\boldsymbol{H}^TS_{\eta\eta}^{-1}(z). \tag{17}$$

When compared, (17) and (8) differ regarding the term $S_{ww}(z)\boldsymbol{C}_m(z)$, which is simply $S_{ww}(z)$ in (8). Therefore, we conclude that in the ℓ_1 smoothing Wiener filter, the process noise is no longer a pure white Gaussian process but a weighted function of it. In other words, in the ℓ_1 smoothing Wiener filter, the state-space model is converted to

Algorithm 1: The proposed method.

Data: $\boldsymbol{y}_{1:T}, \boldsymbol{Q}_{1:T}, \boldsymbol{R}_{1:T}$
Result: $\hat{\boldsymbol{x}}_{k|k}$ for $k=1:T$
$\hat{\boldsymbol{x}}_{1:T}=(\boldsymbol{H}^T\boldsymbol{H})^{-1}\boldsymbol{H}^T\boldsymbol{y}_{1:T}$
for $m=1,2,3,\ldots$ **do**
  for $k=1:T$ **do**
    $\boldsymbol{\Lambda}_k=\hat{\boldsymbol{x}}_k-\boldsymbol{A}\hat{\boldsymbol{x}}_{k-1}$
    $\boldsymbol{C}_k=\boldsymbol{\Lambda}_k^T\operatorname{sgn}(\boldsymbol{\Lambda}_k)$
    $\hat{\boldsymbol{x}}_{k|k-1}=\boldsymbol{A}\hat{\boldsymbol{x}}_{k-1|k-1}$
    $\boldsymbol{P}_{k|k-1}=\boldsymbol{A}\boldsymbol{P}_{k-1|k-1}\boldsymbol{A}^T+\boldsymbol{C}_k^T\boldsymbol{Q}_k\boldsymbol{C}_k$
    $\hat{\boldsymbol{x}}_{k|k}=\hat{\boldsymbol{x}}_{k|k-1}+\boldsymbol{K}_k[\boldsymbol{y}_k-\boldsymbol{H}\hat{\boldsymbol{x}}_{k|k-1}]$
    $\boldsymbol{K}_k=\boldsymbol{P}_{k|k-1}\boldsymbol{H}^T(\boldsymbol{H}\boldsymbol{P}_{k|k-1}\boldsymbol{H}^T+\boldsymbol{R}_k)^{-1}$
    $\boldsymbol{P}_{k|k}=\boldsymbol{P}_{k|k-1}-\boldsymbol{K}_k\boldsymbol{H}\boldsymbol{P}_{k|k-1}$
    $\hat{\boldsymbol{x}}_k=\hat{\boldsymbol{x}}_{k|k}$
  end
end

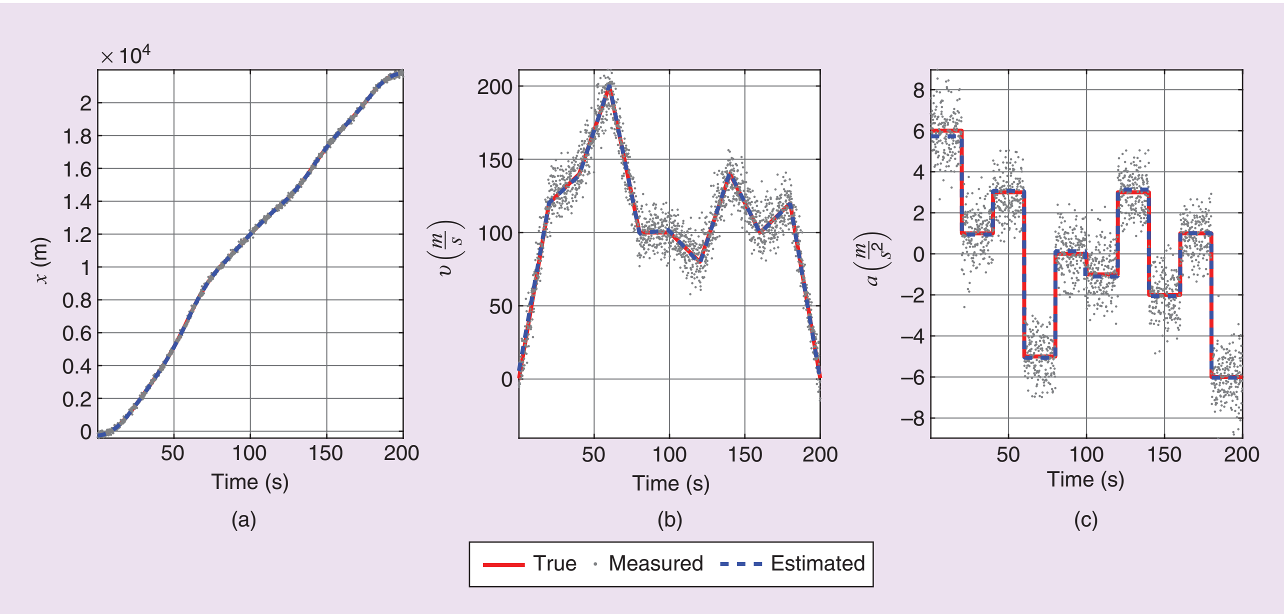

FIGURE 3. The results of the Kalman smoother-based state estimation of the modified model (20) after five iterations (m = 5). The (a) position, (b) velocity, and (c) acceleration.

$$\begin{cases} \boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{C}_{k,m}\boldsymbol{w}_k \\ \boldsymbol{y}_k = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{\eta}_k \end{cases}, \tag{18}$$

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

$$\begin{cases} \boldsymbol{x}_k = \boldsymbol{A}\boldsymbol{x}_{k-1} + \boldsymbol{w}_k \\ \boldsymbol{y}_k = \boldsymbol{H}\boldsymbol{x}_k + \boldsymbol{\eta}_k \end{cases}, \tag{19}$$

we suggest estimating the state of a dynamical system described by (18), instead. Rather than the Kalman filter implementation in (19), where we run the Kalman once across all k, we must run the Kalman for (18) across all k for a given iteration m; this must happen multiple times for $m = 1, 2, 3, \ldots$. We call the proposed method the ℓ_1 *Kalman filter/smoother*, as it is based on the ℓ_1 norm. It is summarized in Algorithm 1.

The classical Kalman filter/smoother is called the ℓ_2 *Kalman filter/smoother*, as it is based on the ℓ_2 norm. Note that if we set $\boldsymbol{C}_{k,m}$ in (18) to an identity matrix, then (19) and (18) are equivalent, which means the conventional ℓ_2 Kalman filter/smoother is a special case of the proposed ℓ_1 Kalman filter/smoother. The proposed technique is used to estimate the target states of the previous example.

To this end, we changed the system model (3) to

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

where $\boldsymbol{C}_{k,m} = \boldsymbol{\Lambda}_{k,m}^T \operatorname{sgn}(\boldsymbol{\Lambda}_{k,m})$:

$$\boldsymbol{\Lambda}_{k,m} = \hat{\boldsymbol{x}}_k^{(m)} - \begin{bmatrix} 1 & T_s & \frac{T_s^2}{2} \\ 0 & 1 & T_s \\ 0 & 0 & 1 \end{bmatrix} \hat{\boldsymbol{x}}_{k-1}^{(m)},$$

$$\hat{\boldsymbol{x}}_k^{(0)} = \frac{1}{3}\boldsymbol{y}_k. \tag{21}$$

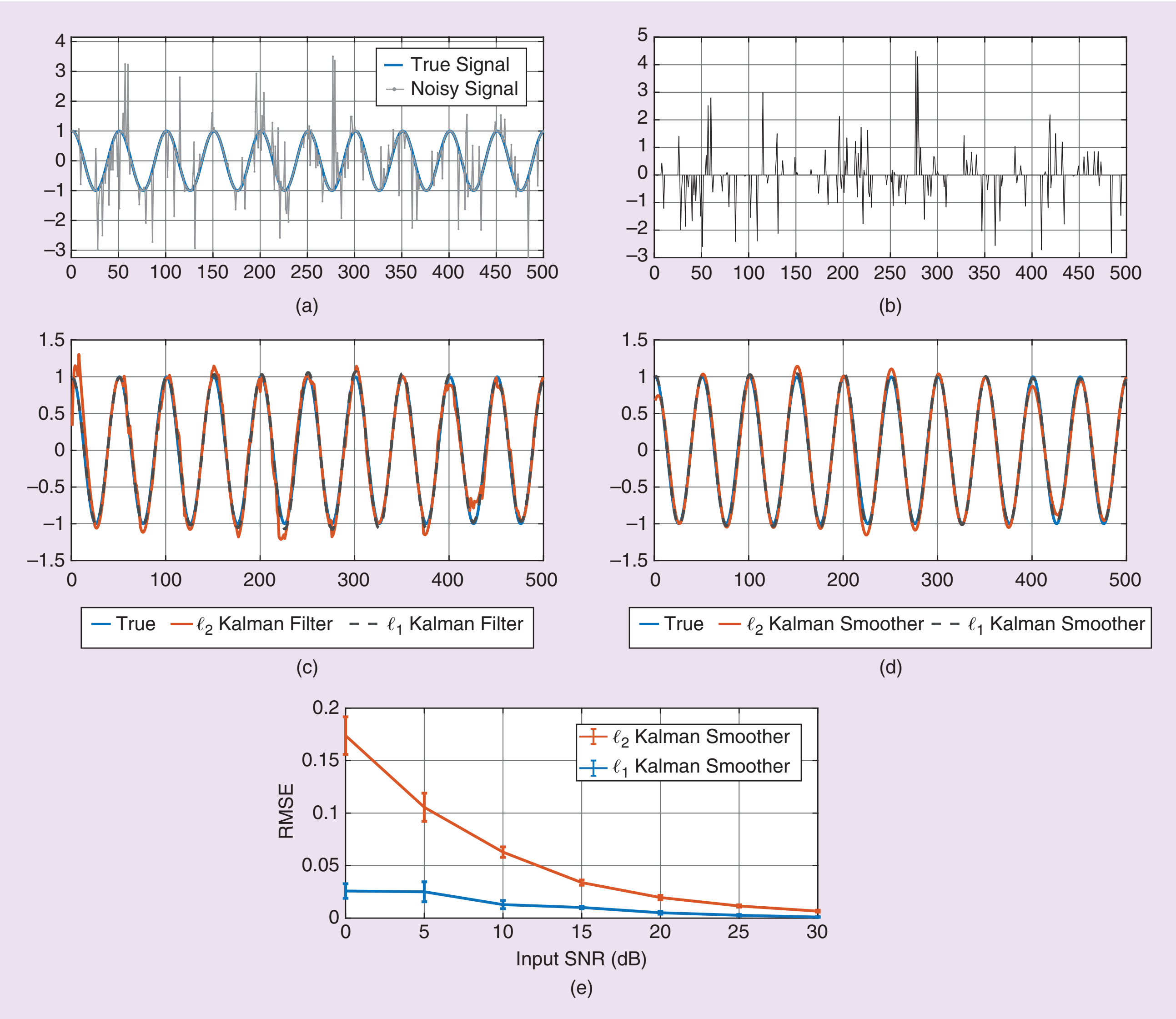

FIGURE 4. A sine wave with known frequency detection in Laplace noise, using a Kalman filter and smoother. (a) The true and noisy sine wave. (b) The impulsive noise. (c) The results obtained by the ℓ_2 and ℓ_1 Kalman filters. (d) The results obtained by the ℓ_2 and ℓ_1 Kalman smoothers. (e) The mean and standard deviation values of the root-mean-square error (RMSE) obtained by the ℓ_2 and ℓ_1 Kalman smoothers for different SNRs.

Here, $\boldsymbol{\Lambda}_{k,m}$ is the system model error after each iteration and $\boldsymbol{C}_{k,m}$ defines its absolute value.

Figure 3 illustrates the result of the Kalman smoother for tracking the target states in the previous example, using the state-space model (20) after five iterations ($m = 5$). The solid red curve and gray dots in Figure 3(a)–(c) indicate the theoretical state and its noisy data, respectively, which are the same as those in Figure 1. The estimated acceleration by the Kalman smoother using the modified model (20) is plotted in Figure 3(c) with a blue dashed line. The estimated acceleration has sharp discontinuities, which demonstrates that the proposed model, which is based on the ℓ_1 regularization and is less sensitive to large changes, is more accurate than the traditional Kalman filter/smoother. Note that we did not modify the Kalman filter equations, but we showed that via a simple modification to the system model, the Kalman filter can be used to optimally estimate the states of the system even if the latent and observed states have non-Gaussian (e.g., impulsive and Laplacian) distribution.

Examples

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

Detection of a sine wave with a 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 α, f, and ϕ are the amplitude, frequency, and phase, respectively] in the presence of impulsive or Laplace noise. The dynamical model of the system can be represented as

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

where $\boldsymbol{x}_k = [x_k \;\; x_{k-1}]^T$ and η_k is assumed to be impulsive or Laplace noise. Note that the dynamical model (22) no longer depends on the amplitude and phase. We generated 2,500 synthetic sine waves with different frequencies and contaminated them with impulsive and Laplace noise. For this purpose, we produced signals varying the power of η_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 sine wave estimation in impulsive noise, provided by a Kalman filter and smoother using (22), appears 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 the classical Kalman filter and smoother are illustrated in red in Figure 4(c) and (d), respectively. Again, we see that the classical Kalman filter and smoother (i.e., the ℓ_2 Kalman filter/smoother) suffer from sensitivity to impulsive error statistics.

To improve the estimation, we employed the modified Kalman filter

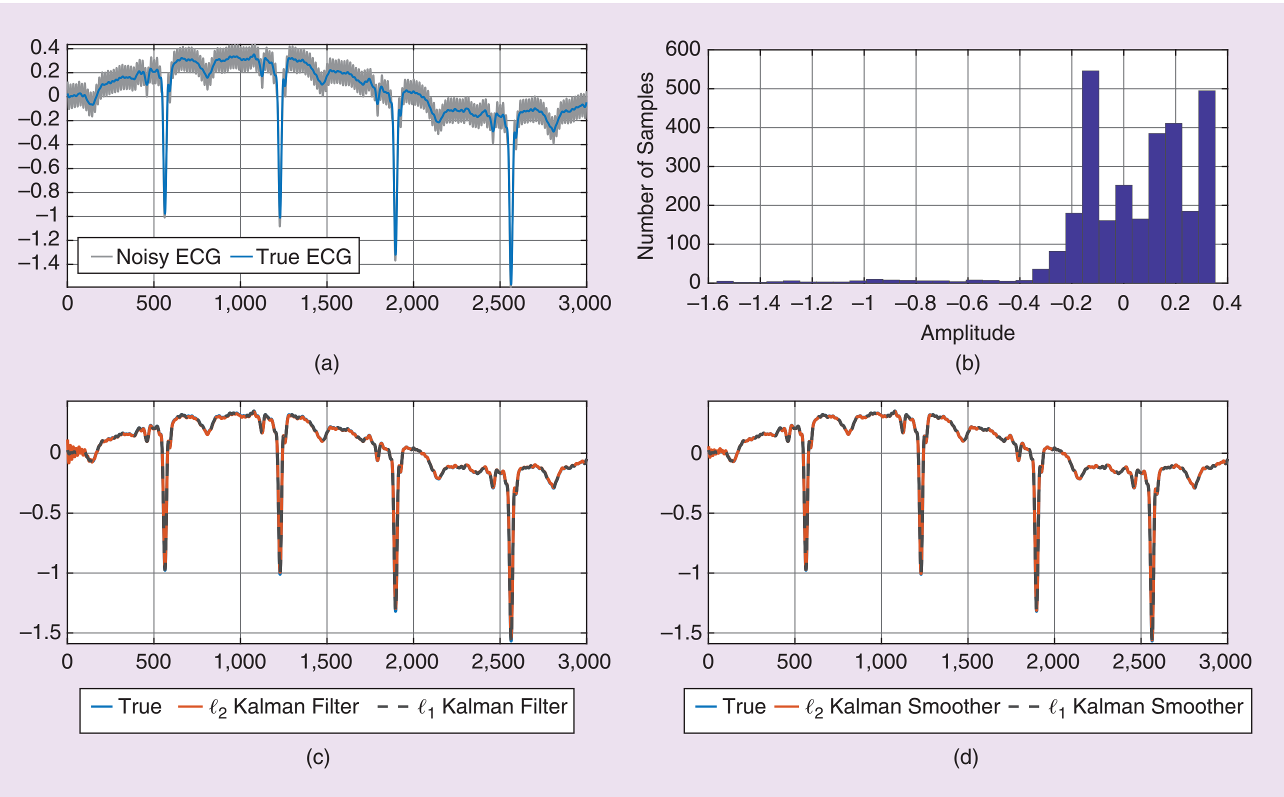

FIGURE 5. The ECG powerline noise cancellation using a Kalman filter and smoother. (a) The true and noisy ECG signal. (b) The distribution of the ECG signal. (c) The results obtained by the ℓ_2 and ℓ_1 Kalman filters. (d) The results obtained by the ℓ_2 and ℓ_1 Kalman smoothers.

and smoother (i.e., the ℓ_1 Kalman filter and smoother) to approximate the desired signal. The results of sine wave detection through the Kalman filter and smoother using the modified model are presented in Figure 4(c) and (d) as black dashed curves. The modified model improves the Kalman filter/smoother's performance for signal detection in impulsive noise. Finally, we employed the Kalman filter and smoother to estimate all generated signals with and without the modified model. The mean and standard deviation values of the root-mean-square error (RMSE) for signal reconstruction obtained by the ℓ_2 and ℓ_1 Kalman smoothers as a function of the SNR are in Figure 4(e). The results confirm that the modified structure can be used to optimize the Kalman filter and smoother so that they do not lose their optimality in applications where the system errors have impulsive or Laplace distribution. As a real application, (22) can be employed to model powerline noise and track and remove it form bioelectrical signals, e.g., electrocardiogram (ECG) signals. As an illustration, a real ECG signal contaminated by powerline noise is plotted in Figure 5(a). To remove the powerline noise, we first estimate it using the Kalman filter/smoother and then subtract it from the observation signal. The denoised ECG provided by the ℓ_2

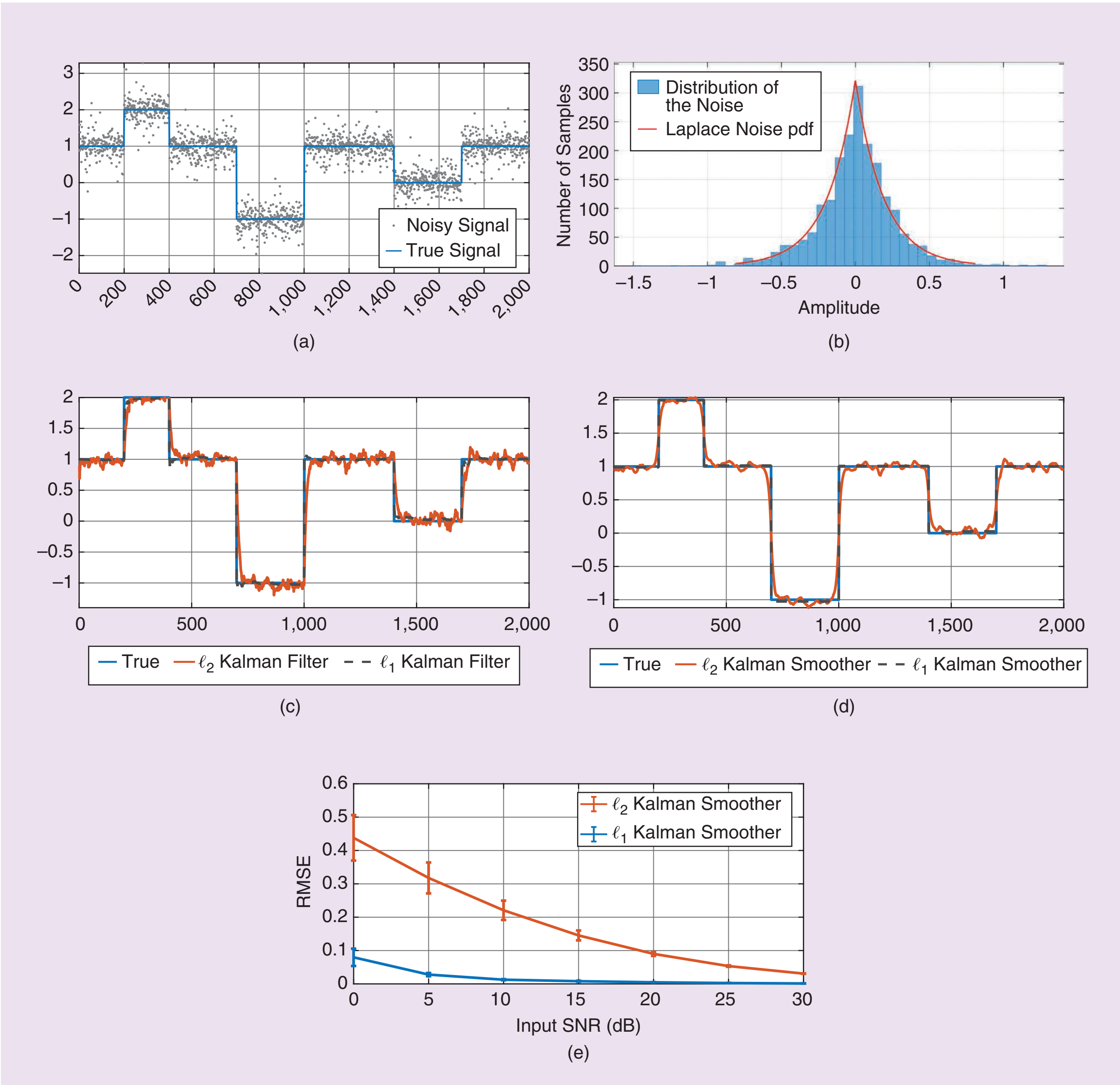

FIGURE 6. The piecewise constant signal detection in Laplace noise, using a Kalman filter and smoother. (a) The true and noisy constant signal. (b) The distribution and probability density function (pdf) of the noise. (c) The results obtained by the ℓ_2 and ℓ_1 Kalman filters. (d) The results obtained by the ℓ_2 and ℓ_1 Kalman smoothers. (e) The mean and standard deviation values of the RMSE obtained by the ℓ_2 and ℓ_1 Kalman smoothers for different SNRs.

and ℓ_1 Kalman filters and smoothers is in Figure 5(c) and (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

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

where w_k and η_k are non-Gaussian impulsive noises. The proposed technique was employed to distinguish the piecewise constant signal in Laplace noise. An example of piecewise constant signal estimation in Laplace noise via the Kalman filter and smoother is illustrated in Figure 6. Note that both the process and observation errors have non-Gaussian distribution. We repeated the experiments from the previous section for piecewise constant signal detection in Laplace noise. A comparison of these two approaches is provided in Figure 6(e), showing that the proposed technique improves the Kalman filter performance in non-Gaussian impulsive/Laplace model errors. Extending the proposed approach to piecewise polynomial signals (e.g., piecewise linear, piecewise quadratic, and so on) contaminated with Laplace noise is straightforward [20].

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

In this section, we consider the problem of estimating a sine wave corrupted by additive Cauchy noise, which frequently appears in engineering applications and is particularly interesting since it

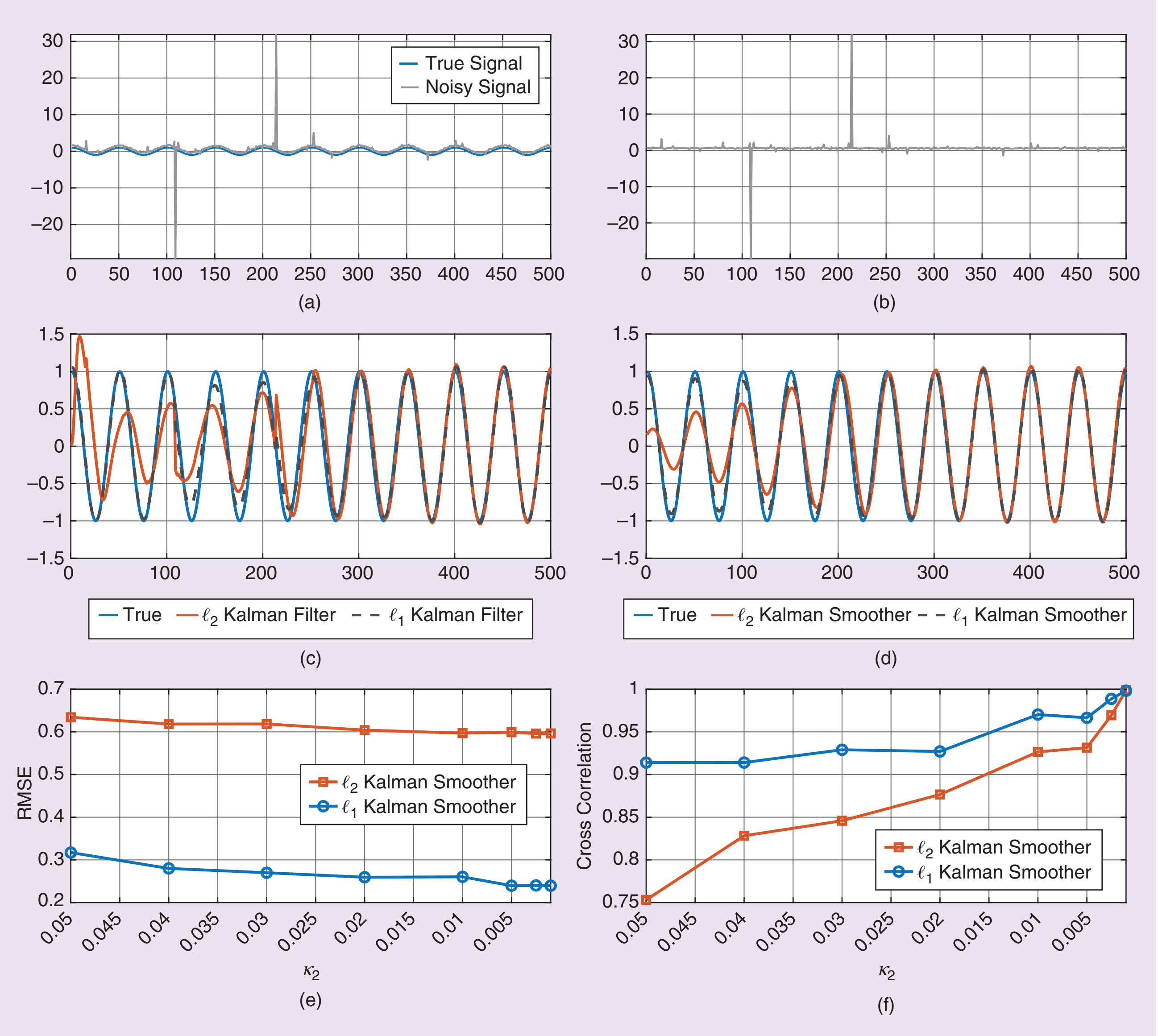

FIGURE 7. The sine wave detection in Cauchy noise, using a Kalman filter and smoother. (a) The true and noisy signal. (b) The Cauchy noise. (c) The results obtained by the ℓ_2 and ℓ_1 Kalman filters. (d) The results obtained by the ℓ_2 and ℓ_1 Kalman smoothers. The (e) mean values of the RMSE and (f) cross correlation between the original and estimated signal obtained by the ℓ_2 and ℓ_1 Kalman smoothers for different values of κ_2.

does not have finite moments of order greater than or equal to one. It is a kind of impulsive non-Gaussian noise with a distribution that looks similar to a normal one, but it has much heavier tails. In Figure 7, we give an example of sine wave detection in Cauchy noise by using the ℓ_2 and ℓ_1 Kalman filters and smoothers. In this example, the parameters of the sine wave are chosen as $\alpha = 1$, $f = 0.02$, and $\phi = 0$. The Cauchy noise was generated from the Cauchy distribution, $r = \kappa_1 + \kappa_2 \tan \pi(rand - 0.5)$, where κ_1 is the statistical median and κ_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 the ℓ_2 and ℓ_1 Kalman filter/smoothers for this example are in Figure 7(c) and (d). We repeated the experiment for different sine waves with various frequencies and phases. To this end, we generated different Cauchy noises with varying values of $0.001 \le \kappa_2 \le 0.05$ and compared the ℓ_2 and ℓ_1 Kalman filter/smoothers. We also computed the cross correlation between the original and estimated sine waves. The mean values of the RMSE and cross correlation obtained by these two methods are reported in Figure 7(e) and (f), respectively. In the best case of signal reconstruction, the cross correlation should be close to one. The results show that the ℓ_1 Kalman smoother outperforms the ℓ_2 Kalman smoother.

The aim of filter design is to find a filter such that when we apply measurements to its input, it produces the MMSE estimate of the state of interest.

Summary

The classical Kalman filter is known as an optimal estimator for linear systems subject to Gaussian error statistics and thus suffers from sensitivity to sparse or Laplace error statistics. In this article, we proposed a simple modification to the system model that can be used to overcome this limitation. We showed that the smoothing Wiener filter can be viewed as a constrained optimization problem that minimizes the ℓ_2 norm of the system model error, subject to the ℓ_2 norm of the data fidelity. Then a variation on the smoothing Wiener filter was proposed that substitutes a sum of absolute values (i.e., the ℓ_1 norm) for the sum of squares used in the ℓ_2 smoothing Wiener filter to penalize variations in the model error. The ℓ_1 smoothing Wiener filter was then used to correct the dynamic model.

The Kalman filter/smoother can be employed to estimate states through 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 and observation errors have non-Gaussian (e.g., impulsive/Laplacian) distribution. Although the proposed ℓ_1 Wiener and Kalman filters are suitable for systems with Laplace or sparse distribution, there are some improvements that can be made. Extending the proposed method to the ℓ_p Wiener and Kalman filters can be considered future work. In particular, for a lower value of p, i.e., $0 < p < 1$, the approach can be used to estimate 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. He is currently an assistant professor in the Department of Computer Science, Fasa University, Fasa, 74617-81189, Iran. From 2017 to 2018, he held a postdoctoral research position with the Grenoble Images, Speech, Signal, and Control lab, France. His research interests include signal modeling, processing, filtering and separation, multimodal signal processing, and functional data analysis.

References

[1] B. Anderson and J. Moore, *Optimal Filtering.* Englewood Cliffs, NJ, USA: Prentice Hall, 1979.

[2] D. Simon, *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches.* Hoboken, NJ, USA: Wiley, 2006.

[3] R. Faragher, "Understanding the basis of the Kalman filter via a simple and intuitive derivation," *IEEE Signal Process. Mag.*, vol. 29, no. 5, pp. 128–132, 2012, doi: 10.1109/MSP.2012.2203621.

[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 Series). Boston, MA, USA: Birkhäuser, 2001. [Online]. Available: https://books.google.com/books?id=cb8B07hwULUC

[5] J. M. Mendel, *Lessons in Digital Estimation Theory.* Englewood Cliffs, NJ, USA: Prentice Hall, 1986.

[6] R. E. Kalman, "A new approach to linear filtering and prediction problems," *J. Basic Eng.*, vol. 82, no. 1, pp. 35–45, 1960, doi: 10.1115/1.3662552.

[7] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series: With Engineering Applications.* London: Technology Press of the Massachusetts Institute of Technology, 1949.

[8] Z. Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at signal fidelity measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, 2009, doi: 10.1109/MSP.2008.930649.

[9] H. Park and T. Lee, "Modeling nonlinear dependencies in natural images using mixture of Laplacian distribution," in *Proc. Adv. Neural Inf. Process. Syst.*, L. Saul, Y. Weiss, and L. Bottou, Eds., vol. 17. Cambridge, MA, USA: MIT Press, 2005, pp. 1041–1048.

[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*, vol. 45, no. 4, pp. 1129–1149, 1999, doi: 10.1109/18.761256.

[11] M. A. Gandhi and L. Mili, "Robust Kalman filter based on a generalized maximum-likelihood-type estimator," *IEEE Trans. Signal Process.*, vol. 58, no. 5, pp. 2509–2520, 2010, doi: 10.1109/TSP.2009.2039731.

[12] J. Zhao, M. Netto, and L. Mili, "A robust iterated extended Kalman filter for power system dynamic state estimation," *IEEE Trans. Power Syst.*, vol. 32, no. 4, pp. 3205–3216, 2017, doi: 10.1109/TPWRS.2016.2628344.

[13] P. J. Huber, *Robust Statistics.* Berlin: Springer-Verlag, 2011, pp. 1248–1251.

[14] J. Zhao and L. Mili, "A robust generalized-maximum likelihood unscented Kalman filter for power system dynamic state estimation," *IEEE J. Sel. Topics Signal Process.*, vol. 12, no. 4, pp. 578–592, 2018, doi: 10.1109/JSTSP.2018.2827261.

[15] W. Press, B. Flannery, S. Teukolsky, and W. Vetterling, *Numerical Recipes, The Art of Scientific Computing.* New York, NY, USA: Cambridge Univ. Press, Jan. 1986.

[16] T. Kailath, *Lectures on Wiener and Kalman Filtering.* (CISM International Centre for Mechanical Sciences Series). New York, NY, USA: Springer-Verlag, 1981, pp. 1–143.

[17] H. Van Trees, *Detection, Estimation, and Modulation Theory, Part I: Detection, Estimation, and Linear Modulation Theory.* Hoboken, NJ, USA: 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.*, vol. 65, no. 3, pp. 794–816, 2017, doi: 10.1109/TSP.2016.2601299.

[19] A. Kheirati Roonizi and C. Jutten, "Forward-backward filtering and penalized least-squares optimization: A unified framework," *Signal Process.*, vol. 178, p. 107,796, Jan. 2021, doi: 10.1016/j.sigpro.2020.107796.

[20] A. Kheirati Roonizi, "ℓ_2 and ℓ_1 trend filtering: A Kalman filter approach [Lecture Notes]," *IEEE Signal Process. Mag.*, vol. 38, no. 6, pp. 137–145, 2021, doi: 10.1109/MSP.2021.3102900.