Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Fast Algorithm of Robust Kalman Filter via l_1 Regression by a Closed Form Solution

Yasuaki Kaneda *,** **Yasuharu Irizuki** ** **Masaki Yamakita** *

* *Tokyo Institute of Technology, Tokyo 152-8552, Japan (e-mail: {kaneda, yamakita}@ ac.ctrl.titech.ac.jp).*
** *Tokyo Metropolitan Industrial Technology Research Institute, Tokyo 135-0064, Japan*

Abstract: Robust Kalman filter (RKF) via l_1 regression is a linear filter for non-Gaussian measurement noise, and it can be formulated as a l_1 optimization problem. Generally, the optimization problem cannot be solved analytically, and some numerical iterative methods are needed. This paper proposes a closed form solution of RKF via l_1 regression by an approximation of its optimal solution and it gives a fast algorithm. The approximated solution can be calculated by upper and lower bounds of the optimal solution. Moreover, a bound of an estimation error of the approximated solution can be analyzed. Some numerical simulations demonstrate the effectiveness of the proposed algorithm.

Keywords: Kalman filters; Non-Gaussian processes; Robust estimation; Fast Kalman algorithms; Efficient algorithms.

1. INTRODUCTION

Kalman filter (KF) is a well known global optimal estimation for linear systems under some conditions. The condition is that an estimation error of an initial state and all noises are distributed by normal distributions. However, many real applications are contaminated by non-Gaussian measurement noise. Especially, outlier is one of the major non-Gaussian measurement noises and generated by heavier tailed distributions than the normal distribution. Hence, abnormal values, which are distant so much from mean values of distributions, are unusually occurred. In target tracking, an outlier is happened due to a reflection noise and it is called clutter [Bilik and Tabrikian, 2010]. In UAV using visual feedback [Guenard, 2008], temporary change of image contrast in background causes outliers in position data. Also in UGV using GPS [Kurashiki et al., 2010], radio disturbances due to some obstacles provide position data with outliers. For non-Gaussian measurement noise, KF is a unbiased minimum variance estimator, but not global optimal [Zanetti, 2012].

Many algorithms of KF for outliers have been proposed. A major KF for non-Gaussian measurement noise including outliers is a Gaussian sum filter [Sorenso and Alspach, 1971]. Gaussian sum filter can approximate arbitrary distributions by using a Gaussian mixture distribution, so it can provide global optimal estimates. However, a computational cost of the method is very high. In contrast, application of l_1 regression to KF gives a robust estimation under outliers and the method is called robust KF (RKF) via l_1 regression [Mattingley and Boyd, 2010, Kaneda et al., 2013]. The method evaluates the outliers by l_1 regression and the regression generates some thresholds to truncates the outliers Therefore, the method has little time delay to reduce effects of the outliers. However, RKF via l_1 regression needs to compute LMI and l_1 optimization problem, so the performance of the RKF depends on algorithms to compute the problems in practice.

Many useful tools to compute LMI are existing. For example, CVX [Mattingley and Boyd, 2010] and YALMIP [Löfberg, 2004] are description languages of convex optimization problems, and the languages can model LMI easily in MATLAB and the other softwares. SeDuMi [Sturm, 1999] and SDPT3 [Toh, et al., 1999] are solvers for optimization problems and implement some algorithms, e.g., an interior point method is the most famous algorithm. However, the algorithms need some numerical iterations in general, so a convergence rate and accuracy of the solutions depend on conditions of the iterations.

l_1 optimization problems can be formulated as QP optimization problems, and CVX can model also the optimization problems. Moreover, CVXGEN [Mattingley and Boyd, 2012] can generate custom C codes of the QP problems for online computations. Fast iterative shrinkage thresholding algorithm (FISTA) are proposed as effective computation methods for l_1 optimization problems [Beck and Marc, 2009]. However, similarly to computations of LMI, these methods demand some iterative methods. Homotopy method was also proposed and can compute a solution in a closed form [Garrigues and Ghaoui, 2008], but the method uses all past data.

In this paper, we derive upper and lower bounds of an optimal estimate of the RKF and compute an approximated estimate using the bounds. In addition, the approximated estimate is given by a closed form, so no iteration is needed and it gives a fast computation.

This paper is organized as follows: In section 2, RKF via l_1 regression is introduced. In section 3, a closed form solution of the RKF is proposed and the estimation error of the algorithm is analyzed. In section 4, some numerical simulations demonstrate effectiveness of the proposed algorithm. Conclusion is given in section 5.

Notation A vector is represented as a bold character and its element is as a subscript i. However, when variables and functions depend on time k, the time dependent functions are also expressed as a subscript k. For example, when a vector $\boldsymbol{x}$ depends on time k, the vector and its i-th element are expressed as $\boldsymbol{x}_k$ and $x_{k,i}$, respectively.

A derivative of an absolute function is defined as the following sub-gradient:

$$\frac{\partial |x_i|}{\partial x_i} :\in \begin{cases} \{1\} & x_i > 0 \\ [-1,\ 1] & x_i = 0 \\ \{-1\} & x_i < 0 \end{cases}.$$

It means that $\partial|x_i|/\partial x_i = 1$ for $x_i > 0$ and $\partial|x_i|/\partial x_i = -1$ for $x_i < 0$.

2. ROBUST KALMAN FILTER VIA L_1 REGRESSION

2.1 Formula

Let $\boldsymbol{x}_k \in \mathbb{R}^n$ and $\boldsymbol{y}_k \in \mathbb{R}^m$ be a state and measurement at time k, respectively. We consider the following linear time invariant system:

$$\boldsymbol{x}_k = A\boldsymbol{x}_{k-1} + \boldsymbol{w}_k, \qquad \boldsymbol{y}_k = C\boldsymbol{x}_k + \boldsymbol{v}_k + \boldsymbol{z}_k. \tag{1}$$

where $A \in \mathbb{R}^{n\times n}$ and $C \in \mathbb{R}^{m\times n}$ are known matrices of system and measurement, respectively. $\boldsymbol{w}_k \in \mathbb{R}^n$ is a system noise at time k, and $\boldsymbol{v}_k, \boldsymbol{z}_k \in \mathbb{R}^m$ are a Gaussian noise and outlier in a measurement at time k, respectively.

We assume that $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$ are mutually independent. Let $P \in \mathbb{R}^n$ be a covariance matrix of a state estimation error, and let $Q \in \mathbb{R}^{n\times n}$ and $R \in \mathbb{R}^{m\times m}$ denote covariance matrices of $\boldsymbol{w}_k$ and $\boldsymbol{v}_k$, respectively. Given Q, R, and an initial value of P, i.e., $P_{0|0}$, prediction and correction laws of RKF via l_1 regression are expressed as

$$\text{Predict:} \begin{cases} \hat{\boldsymbol{x}}_{k|k-1} = A\hat{\boldsymbol{x}}_{k-1|k-1}, \\ P_{k|k-1} = AP_{k-1|k-1}A^T + Q, \end{cases} \tag{2}$$

$$\text{Correct:} \begin{cases} L = P_{k|k-1}C^T(CP_{k|k-1}C^T + R)^{-1}, \\ \boldsymbol{e}_k = \boldsymbol{y}_k - C\hat{\boldsymbol{x}}_{k|k-1}, \\ \hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + L(\boldsymbol{e}_k - \boldsymbol{z}_k^*), \\ P_{k|k} = (I - LC)P_{k|k-1}, \end{cases} \tag{3}$$

where $\boldsymbol{z}_k^*$ is given by a solution of the following optimization problem with l_1 regression:

$$\boldsymbol{z}_k^* = \arg\min_{\boldsymbol{z}_k} \ (\boldsymbol{e}_k - \boldsymbol{z}_k)^T W (\boldsymbol{e}_k - \boldsymbol{z}_k) + \sum_{i=1}^m \lambda_i |z_{k,i}|, \tag{4}$$

where W is the following positive definite matrix:

$$\begin{aligned} W &= (I - CL)^T R^{-1}(I - CL) + L^T P_{k|k-1}^{-1} L \\ &= \left(CP_{k|k-1}C^T + R\right)^{-1}. \end{aligned} \tag{5}$$

$\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_m]^T \in \mathbb{R}^m$ is a regularization parameter, and it can be designed by the covariance matrices, Q and R, systematically [Kaneda et al., 2013].

Generally, the optimization problem (4) requires iterative computation methods like an interior point method. However, a convergence rate and accuracy of solutions depend on conditions of the iterative methods. For example, accuracy of a gradient method depends on its step size. And also, solutions rely on stop conditions in any algorithms. Parameters in iterative algorithms need to be tuned for each applications and the tuning is heuristic in general. In this paper, we propose a new algorithm without iterations by an approximation of the optimal solution, and we analyze a performance of the algorithm.

2.2 Condition to design regularization parameters of RKF

In Eq. (4), a first order necessary condition of an optimality condition derives the following inclusions:

$$\left(2W(\boldsymbol{e}_k - \boldsymbol{z}_k^*)\right)_i \in \begin{cases} \{\lambda_i\} & z_{k,i}^* > 0 \\ [-\lambda_i, \lambda_i] & z_{k,i}^* = 0 \\ \{-\lambda_i\} & z_{k,i}^* < 0 \end{cases}, \tag{6}$$

where $(\cdot)_i$ means i-th element of a vector. Eq. (4) is a convex optimization problem, so the condition is also sufficient. Note that Eq. (4) can be solved analytically if W is scalar or a diagonal matrix.

Assume that measurements include no outliers, i.e., $\boldsymbol{z}_k = \boldsymbol{0}$, Let $\boldsymbol{e}_k^* := \boldsymbol{e}_k|_{\boldsymbol{z}_k=0}$, Eq. (6) gives the following inclusion:

$$\left(2W\boldsymbol{e}_k^*\right)_i \in [-\lambda_i, \lambda_i]. \tag{7}$$

Let a sub-gradient of $|z_{k,i}|$ be $\eta_i \in [-1,\ 1]$. Eq. (7) can be rewritten as

$$[\,\lambda_1\eta_1 \ \cdots \ \lambda_m\eta_m\,]^T = 2W\boldsymbol{e}_k^*. \tag{8}$$

Since η_i can be randomly selected, it can be regarded as a stochastic variable without a loss of generality. Additionally, a selection of η_i is independent of other variables. Note that $\boldsymbol{e}_k^*$ is a prediction error considering only Gaussian noise as a measurement noise. A covariance matrix of $\boldsymbol{e}_k^*$, i.e., Σ_{e^*}, is given by

$$\Sigma_{e^*} = E[\boldsymbol{e}_k^* \boldsymbol{e}_k^{*T}] = CP_{k|k-1}C^T + R. \tag{9}$$

Note that $W^{-1} = CP_{k|k-1}C^T + R$, $E[\eta_i^2] \le 1$, and $E[\eta_i\eta_j] = 0$ $(i \ne j)$. Eq. (8) gives the following inequality:

$$\begin{aligned} \operatorname{diag}(\lambda_1^2, \cdots, \lambda_m^2) &\ge \operatorname{diag}(\lambda_1^2 E[\eta_1^2], \cdots, \lambda_m^2 E[\eta_m^2]) \\ &= 4WE\left[\boldsymbol{e}_k^* (\boldsymbol{e}_k^*)^T\right] W = 4W, \end{aligned} \tag{10}$$

where $\operatorname{diag}(\cdot)$ is a diagonal matrix. Therefore, Eq. (10) is a condition of the regularization parameter $\boldsymbol{\lambda}$.

A sparse solution, i.e., $z_{k,i}^* = 0$, can be often obtained if λ_i is large. $\boldsymbol{\lambda}$ should be determined in a small residual of both sides of Eq. (10). One of the solutions is given by the following semi-definite programming (SDP):

$$\begin{aligned} &\min_{\lambda_1^2, \cdots, \lambda_m^2} \ \lambda_1^2 + \cdots + \lambda_m^2 \\ &\text{s.t. } \operatorname{diag}(\lambda_1^2, \cdots, \lambda_m^2) \ge 4W. \end{aligned} \tag{11}$$

Claim 1. An assumption of $\boldsymbol{z}_k^* = \boldsymbol{0}$ derives Eq. (7). However, a dual problem of Eq. (4) can result in the same inclusion without the assumption [Kaneda et al., 2013].

Claim 2. In general, Eq. (4) cannot be solved analytically, and some iterative methods are needed. Also in Eq. (11), iterative methods are required to solve the LMI.

3. FAST ALGORITHM OF ROBUST KALMAN FILTER BY A CLOSED FORM SOLUTION

3.1 Derivation

Let $\hat{\boldsymbol{z}}_k^*$ be an approximated solution of $\boldsymbol{z}_k^*$. This section shows that use of upper and lower bounds of $\boldsymbol{z}_k^*$ provides the approximation of solution $\hat{\boldsymbol{z}}_k^*$, and $\hat{\boldsymbol{z}}_k^*$ can be written in a closed form.

It is assumed that the regularization parameter is given by the following inequality:

$$\Lambda \geq \sqrt{W}, \quad \Lambda = \mathrm{diag}(\lambda_1, \cdots, \lambda_m), \tag{12}$$

where $\sqrt{W}\left(\sqrt{W}\right)^T = W$. Suppose that $\sqrt{W}$ is computed by Cholesky decomposition, so $\sqrt{W}$ becomes a lower triangle matrix. The first order necessary condition of an optimality condition can be rewritten as

$$W(\boldsymbol{e}_k - \boldsymbol{z}_k^*) = \frac{1}{2}\Lambda \frac{\partial ||\boldsymbol{z}_k^*||_1}{\partial \boldsymbol{z}_k^*}.$$

$$\therefore \left(\frac{\partial ||\boldsymbol{z}_k^*||_1}{\partial \boldsymbol{z}_k^*}\right)^T W(\boldsymbol{e}_k - \boldsymbol{z}_k^*) = \frac{1}{2}\left(\frac{\partial ||\boldsymbol{z}_k^*||_1}{\partial \boldsymbol{z}_k^*}\right)^T \Lambda \frac{\partial ||\boldsymbol{z}_k^*||_1}{\partial \boldsymbol{z}_k^*}.$$

Therefore, the following inequality is satisfied:

$$\left(\left(\sqrt{W}\right)^T \frac{\partial ||\boldsymbol{z}_k^*||_1}{\partial \boldsymbol{z}_k^*}\right)^T \times \left(\left(\sqrt{W}\right)^T (\boldsymbol{e}_k - \boldsymbol{z}_k^*) - \frac{\partial ||\boldsymbol{z}_k^*||_1}{\partial \boldsymbol{z}_k^*}\right) \geq 0, \tag{13}$$

where $\left(\sqrt{W}\right)^T$ is an upper triangle matrix and represented by the following equation:

$$\left(\sqrt{W}\right)^T = \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1m} \\ 0 & w_{22} & \ddots & w_{2m} \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & w_{mm} \end{bmatrix}. \tag{14}$$

A sufficient condition for Eq. (13) is as follows:

$$\begin{cases} \left(\left(\sqrt{W}\right)^T \frac{\partial ||\boldsymbol{z}_{k,i}^*||_1}{\partial \boldsymbol{z}_{k,i}^*}\right)_i \geq 0, \\ \left(\left(\sqrt{W}\right)^T (\boldsymbol{e}_k - \boldsymbol{z}_k^*) - \frac{\partial ||\boldsymbol{z}_k^*||_1}{\partial \boldsymbol{z}_k^*}\right)_i \geq 0, \end{cases} \tag{15}$$

or

$$\begin{cases} \left(\left(\sqrt{W}\right)^T \frac{\partial ||\boldsymbol{z}_{k,i}^*||_1}{\partial \boldsymbol{z}_{k,i}^*}\right)_i \leq 0, \\ \left(\left(\sqrt{W}\right)^T (\boldsymbol{e}_k - \boldsymbol{z}_k^*) - \frac{\partial ||\boldsymbol{z}_k^*||_1}{\partial \boldsymbol{z}_k^*}\right)_i \leq 0. \end{cases} \tag{16}$$

First, consider a case of $i = m$. Assuming that diagonal elements of Eq. (14) are selected to be positive, conditions (15) and (16) result in the following inequalities:

$$z_{k,m}^* \leq e_{k,m} - \frac{1}{w_{mm}} \frac{\partial |z_{k,m}^*|}{z_{k,m}^*}, \quad \frac{\partial |z_{k,m}^*|}{z_{k,m}^*} \geq 0, \tag{17}$$

$$z_{k,m}^* \geq e_{k,m} - \frac{1}{w_{mm}} \frac{\partial |z_{k,m}^*|}{z_{k,m}^*}, \quad \frac{\partial |z_{k,m}^*|}{z_{k,m}^*} \leq 0. \tag{18}$$

Right hand sides of the inequalities can be interpreted as computations of upper and lower bounds of $z_{k,m}^*$. Let $\bar{z}_{k,m}^* \geq 0$ and $\underline{z}_{k,m}^* \leq 0$ be the upper and lower bounds of $z_{k,m}^*$, respectively. Assuming that signs of the upper and lower bounds are equal to one of the optimal solution, these bounds can be calculated by the following equations, respectively:

$$\bar{z}_{k,m}^* = \begin{cases} e_{k,m} - \frac{1}{w_{mm}} & e_{k,m} > \frac{1}{w_{mm}} \\ 0 & \text{otherwise} \end{cases}, \tag{19}$$

$$\underline{z}_{k,m}^* = \begin{cases} e_{k,m} + \frac{1}{w_{mm}} & e_{k,m} < -\frac{1}{w_{mm}} \\ 0 & \text{otherwise} \end{cases}. \tag{20}$$

Eq. (19) and (20) are 0 in a common domain. An estimate of $z_{k,m}^*$, i.e., $\hat{z}_{k,m}^*$, is defined as

$$\hat{z}_{k,m}^* = \bar{z}_{k,m}^* + \underline{z}_{k,m}^* = \begin{cases} e_{k,m} - \frac{1}{w_{mm}} & e_{k,m} > \frac{1}{w_{mm}} \\ 0 & \text{otherwise} \\ e_{k,m} + \frac{1}{w_{mm}} & e_{k,m} < -\frac{1}{w_{mm}} \end{cases}. \tag{21}$$

Assume that elements from $i+1$ to m are calculated. The condition (15) provides the following condition for i-th elements:

$$\begin{cases} \frac{\partial |z_{k,i}^*|}{z_{k,i}^*} \geq -\frac{1}{w_{ii}} \sum_{j=i+1}^m w_{ij} \frac{\partial |\hat{z}_{k,j}^*|}{\hat{z}_{k,j}^*}, \\ z_{k,i}^* \leq e'_{k,i} - \frac{1}{w_{ii}} \frac{\partial |z_{k,i}^*|}{z_{k,i}^*}, \\ e'_{k,i} = e_{k,i} + \frac{1}{w_{ii}} \sum_{j=i+1}^m w_{ij}(e_{k,j} - \hat{z}_{k,j}^*). \end{cases} \tag{22}$$

In the same way as $i = m$, Eq. (22) means a calculation of an upper bound of $z_{k,i}^*$, i.e., $\bar{z}_{k,i}^* \geq 0$. Note that $\bar{z}_{k,i}^* > 0$ gives $\frac{\partial |\bar{z}_{k,i}^*|}{\bar{z}_{k,i}^*} = 1$. Assuming that signs of $z_{k,i}^*$ and $\bar{z}_{k,i}^*$ are same, $\bar{z}_{k,i}^*$ can be computed by

$$\begin{aligned} &\text{if } -\frac{1}{w_{ii}} \sum_{j=i+1}^m w_{ij} \frac{\partial |\hat{z}_{k,j}^*|}{\hat{z}_{k,j}^*} \leq 1 \\ &\quad \text{then } \bar{z}_{k,i}^* = \begin{cases} e'_{k,i} - \frac{1}{w_{ii}} & e'_{k,i} > \frac{1}{w_{ii}} \\ 0 & \text{otherwise} \end{cases} \\ &\quad \text{else } \bar{z}_{k,i}^* = 0. \end{aligned} \tag{23}$$

where, for convenience, $\bar{z}_{k,i}^* = 0$ if the condition (15) is not satisfied. Similarly, note that $\frac{\partial |\underline{z}_{k,i}^*|}{\underline{z}_{k,i}^*} = -1$ for $\underline{z}_{k,i}^* < 0$, $\underline{z}_{k,i}^*$ is given by

$$\begin{aligned} &\text{if } -\frac{1}{w_{ii}} \sum_{j=i+1}^m w_{ij} \frac{\partial |\hat{z}_{k,j}^*|}{\hat{z}_{k,j}^*} \geq -1 \\ &\quad \text{then } \underline{z}_{k,i}^* = \begin{cases} e'_{k,i} + \frac{1}{w_{ii}} & e'_{k,i} < -\frac{1}{w_{ii}} \\ 0 & \text{otherwise} \end{cases} \\ &\quad \text{else } \underline{z}_{k,i}^* = 0. \end{aligned} \tag{24}$$

Also in the case, Eq. (23) and (24) become 0 in a common domain. Therefore, $\hat{z}_{k,i}^*$ is defined as

$$\hat{z}_{k,i}^* = \bar{z}_{k,i}^* + \underline{z}_{k,i}^*. \tag{25}$$

Algorithm 1 shows a fast algorithm of the RKF by a closed form computation.

Algorithm 1 Fast algorithm of measurement update of robust Kalman filter via l_1 regression at time k

$\boldsymbol{e}_k = \boldsymbol{y}_k - C\hat{\boldsymbol{x}}_{k|k-1}$

$W = (CP_{k|k-1}C^T + R)^{-1}$

compute $\left(\sqrt{W}\right)^T$ using Cholesky decomposition, where w_{ij} is an element of $\left(\sqrt{W}\right)^T$

$\hat{z}^*_{k,m} = \max(|e_{k,m}| - \frac{1}{w_{mm}}, 0)\ \mathrm{sign}(e_{k,m})$

for $i = m-1$ down to 1 **do**

  $e'_{k,i} = e_{k,i} - \frac{1}{w_{ii}} \sum_{j=i+1}^{m} \sigma_{ij} \frac{\partial |\hat{z}^*_{k,j}|}{\partial \hat{z}^*_{k,j}}$.

  if $-\frac{1}{w_{ii}} \sum_{j=i+1}^{m} w_{ij} \frac{\partial |\hat{z}^*_{k,j}|}{\hat{z}^*_{k,j}} > 1$ **then**

    $\bar{z}^*_{k,i} = 0$

  else

    $\bar{z}^*_{k,i} = \max\left(e'_{k,i} - \frac{1}{w_{ii}}, 0\right)$

  end if

  if $-\frac{1}{w_{ii}} \sum_{j=i+1}^{m} w_{ij} \frac{\partial |\hat{z}^*_{k,j}|}{\hat{z}^*_{k,j}} < -1$ **then**

    $\underline{z}^*_{k,i} = 0$

  else

    $\underline{z}^*_{k,i} = \min\left(e'_{k,i} + \frac{1}{w_{ii}}, 0\right)$

  end if

  $\hat{z}^*_{k,i} = \bar{z}^*_{k,i} + \underline{z}^*_{k,i}$.

end for

$L = P_{k|k-1}C^T W$

$\hat{\boldsymbol{x}}_{k|k} = \hat{\boldsymbol{x}}_{k|k-1} + L(\boldsymbol{e}_k - \hat{\boldsymbol{z}}^*_k)$

$P_{k|k} = (I - LC)P_{k|k-1}$

3.2 Analysis of an Estimation Error of Outliers and Innovation of RKF

In order to show a performance of the proposed algorithm, an estimation error of the solution is analyzed. Moreover, an innovation of the RKF using the algorithm is also analyzed.

Assume that the following condition is satisfied for estimated outliers $\hat{\boldsymbol{z}}^*_k$:

$$-1 \le -\frac{1}{w_{ii}} \sum_{j=i+1}^{m} w_{ij} \frac{\partial |\hat{z}^*_{k,j}|}{\hat{z}^*_{k,j}} \le 1, \quad \forall \hat{z}^*_i. \tag{26}$$

Eq. (25) can be simplified as

$$\hat{z}^*_{k,i} = \begin{cases} e'_{k,i} - \frac{1}{w_{ii}} & e'_{k,i} > \frac{1}{w_{ii}} \\ 0, & \text{otherwise} \\ e'_{k,i} + \frac{1}{w_{ii}} & e'_{k,i} < -\frac{1}{w_{ii}} \end{cases}. \tag{27}$$

Eq. (27) is equivalent to a solution of the following equation:

$$\left(\sqrt{W}\right)^T (\boldsymbol{e}_k - \hat{\boldsymbol{z}}^*_k) = \frac{\partial ||\hat{\boldsymbol{z}}^*_k||_1}{\partial \hat{\boldsymbol{z}}^*_k}. \tag{28}$$

For an estimation error of outliers, Eq. (1) and (28) yield the following equation:

$$\hat{\boldsymbol{z}}^*_k - \boldsymbol{z}_k = -\left(\sqrt{W}\right)^{-T} \frac{\partial ||\hat{\boldsymbol{z}}^*_k||_1}{\partial \hat{\boldsymbol{z}}^*_k} + C(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}) + \boldsymbol{v}_k.$$

Note that $\frac{\partial ||\hat{\boldsymbol{z}}^*_k||_1}{\partial \hat{\boldsymbol{z}}^*_k}$ is a vector consisting of sub-gradients. This means that each element of the vector can be interpreted as mutually independent stochastic variables which are in $[-1, 1]$. From $E\left[\frac{\partial ||\hat{\boldsymbol{z}}^*_k||_1}{\partial \hat{\boldsymbol{z}}^*_k} \left(\frac{\partial ||\hat{\boldsymbol{z}}^*_k||_1}{\partial \hat{\boldsymbol{z}}^*_k}\right)^T\right] \le I$ and Eq. (5), a covariance matrix of an estimation error of outliers is given by

$$\begin{aligned} &E\left[(\hat{\boldsymbol{z}}^*_k - \boldsymbol{z}_k)(\hat{\boldsymbol{z}}^*_k - \boldsymbol{z}_k)^T\right] \\ &\quad \le \left(\sqrt{W}\right)^{-T} \left(\sqrt{W}\right)^{-1} + CP_{k|k-1}C^T + R \\ &\quad = W^{-1} + CP_{k|k-1}C^T + R \\ &\quad = 2\left(CP_{k|k-1}C^T + R\right). \end{aligned} \tag{29}$$

Moreover, for an innovation of the RKF, Eq. (28) yields

$$\boldsymbol{e}_k - \hat{\boldsymbol{z}}^*_k = \left(\sqrt{W}\right)^{-T} \frac{\partial ||\hat{\boldsymbol{z}}^*_k||_1}{\partial \hat{\boldsymbol{z}}^*_k}.$$

Therefore, its covariance matrix is given by

$$\begin{aligned} &E\left[(\boldsymbol{e}_k - \hat{\boldsymbol{z}}^*_k)(\boldsymbol{e}_k - \hat{\boldsymbol{z}}^*_k)^T\right] \\ &\quad \le \left(\sqrt{W}\right)^{-T} \left(\sqrt{W}\right)^{-1} = CP_{k|k-1}C^T + R. \end{aligned} \tag{30}$$

Claim 3. Note that Eq. (27) is an approximated solution, not optimal, again. However, a covariance matrix of an estimation error of $\hat{\boldsymbol{z}}^*_k$ is bounded by Eq. (29) under the condition (26).

Claim 4. The proposed algorithm assumes that $\sqrt{W}$ is a triangle matrix. If $\sqrt{W}$ is not a symmetric matrix, Eq. (12) cannot necessarily satisfy the condition of the regularization parameter (10). However, the proposed algorithm can provide the performance given by Eq. (30) if the condition (26) is satisfied. This means that the covariance matrix of an innovation of the RKF is bounded by that of standard KF without outliers. The fact shows that the performance of the RKF computed by the proposed algorithm becomes ideal one in some meanings.

Claim 5. The covariance matrices are satisfied only under the condition (26). In other words, the condition (26) can judge whether the proposed algorithm is good or not.

3.3 Analysis of a State Estimation Error

Eq. (1) and (3) yield the following equation:

$$\begin{aligned} \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k} &= \boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1} - L(\boldsymbol{e}_k - \hat{\boldsymbol{z}}^*_k) \\ &= (I - LC)(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k-1}) - L\boldsymbol{v}_k - L\left(\boldsymbol{z}_k - \hat{\boldsymbol{z}}^*_k\right), \end{aligned}$$

where $\hat{\boldsymbol{z}}^*_k$ is used in Eq. (3) instead of $\boldsymbol{z}^*_k$. Both in KF and the RKF, Eq. (3) updates a covariance matrix of a state estimation error. However, under outliers, an actual updated covariance matrix of a state estimation error is given by

$$\begin{aligned} P_{k|k} &= E\left[(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k})(\boldsymbol{x}_k - \hat{\boldsymbol{x}}_{k|k})^T\right] \\ &= (I - LC)P_{k|k-1} \\ &\quad + LE\left[(\boldsymbol{z}_k - \hat{\boldsymbol{z}}^*_k)(\boldsymbol{z}_k - \hat{\boldsymbol{z}}^*_k)^T\right] L^T. \end{aligned} \tag{31}$$

In the standard KF, $\hat{\boldsymbol{z}}^*_k = \boldsymbol{0}$, so the actual covariance matrix under outliers depends on a second moment of $\boldsymbol{z}_k$. For example, if $\boldsymbol{z}_k$ is distributed by a distribution whose second moment is infinite, like a Cauchy distribution[Idan and Speyer, 2010], the updated covariance matrix $P_{k|k}$ should be infinite in ideal, but it results in no update of a state. On the other hand, in the RKF using the proposed algorithm, Eq. (31) satisfies the following inequality and bounded:

$$\begin{aligned} P_{k|k} &\leq (I - LC)P_{k|k-1} + 2LCP_{k|k-1} \\ &= (I + LC)P_{k|k-1}. \end{aligned} \tag{32}$$

In the RKF using the proposed algorithm, the updated covariance matrix of a state estimation error can be selected among solutions satisfying Eq. (32). The update law (3) is one of the solutions.

4. SIMULATION

4.1 Problem Statement

We consider a state estimation problem of two-wheel vehicle shown in Fig. 1 under outliers [Kaneda et al., 2013]. Let m and J be a mass of the vehicle and moment of inertia about the center of gravity, respectively. Let f and τ_θ denote a driving force in the direction of motion and steering torque, respectively. Assuming $\boldsymbol{q} = [\, x \;\theta\; y \,]^T$ and $\boldsymbol{\tau} = [\, f \;\tau_\theta \,]^T$, a dynamic model of the vehicle is given by

$$\begin{bmatrix} m & 0 & 0 \\ 0 & J & 0 \\ 0 & 0 & m \end{bmatrix} \ddot{\boldsymbol{q}} = \begin{bmatrix} \cos\theta & 0 \\ 0 & 1 \\ \sin\theta & 0 \end{bmatrix} \boldsymbol{\tau}.$$

The vehicle satisfies the following velocity constraint:

$$\dot{x}\sin\theta - \dot{y}\cos\theta = 0.$$

A model used in the estimation is the following constant acceleration model:

$$\boldsymbol{x}_k = A_a \boldsymbol{x}_{k-1} + \boldsymbol{w}_k,$$

where $\boldsymbol{x}_k = [x_k, \dot{x}_k, y_k, \dot{y}_k, \theta_k, \dot{\theta}_k]^T \in \mathbb{R}^6$ is a state of the vehicle at time k. A_a and $\boldsymbol{w}_k$ are given by

$$A_a = \mathrm{diag}(A,\ A,\ A), \quad \boldsymbol{w}_k = \left[\, G^T a_x \;\, G^T a_y \;\, G^T a_\theta \,\right]^T,$$

where $A = \begin{bmatrix} 1 & \Delta t \\ 0 & 1 \end{bmatrix}$ and $G = \left[\, \frac{\Delta t^2}{2} \;\; \Delta t \,\right]^T$. Δt is a sampling time, and a_x, a_y, and a_θ are accelerations of x_k, y_k, and θ_k, respectively. Assume that a_x, a_y, and a_θ are mutually independent. Let σ_{a_x}, σ_{a_y}, and σ_{a_θ} be standard deviations of a_x, a_y, and a_θ, respectively. A covariance matrix of $\boldsymbol{w}_k$, i.e., Q is given by

$$Q = E[\boldsymbol{w}_k \boldsymbol{w}_k^T] = \mathrm{diag}(\sigma_{a_x}^2 GG^T, \sigma_{a_y}^2 GG^T, \sigma_{a_\theta}^2 GG^T).$$

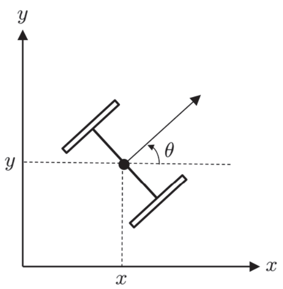

Fig. 1. Model of two-wheeled vehicle with non-holonomic constraint.

4.2 Noise Model

Two cases of distributions are considered as outliers, i.e., Cauchy and Gaussian mixture distributions. Cauchy distribution is often used to represent impulsive unexpected values of sensors [Idan and Speyer, 2010]. Gaussian mixture distribution is also used to express unusual outliers, e.g., clutter of radar tracking systems [Bilik and Tabrikian, 2010]. Cauchy and Gaussian mixture distributions, $p_c(x)$ and $p_g(x)$, are given by the following equations, respectively:

$$\begin{aligned} p_c(x) &= \frac{1}{\pi} \frac{\delta}{\delta^2 + (x - x_0)^2}, \\ p_g(x) &= (1 - \varepsilon)\mathcal{N}_x(0, \boldsymbol{\Sigma}_1^2) + \varepsilon \mathcal{N}_x(0, \boldsymbol{\Sigma}_2^2), \end{aligned}$$

where x_0 is a center and δ is a width of Cauchy distribution. ε is a random variable distributed by Bernoulli distribution whose probability is p. $\mathcal{N}_x(\mu, \boldsymbol{\Sigma}^2)$ is a normal distribution whose mean is μ and covariance matrix is $\boldsymbol{\Sigma}^2$.

4.3 Conditions

An initial value of the vehicle is $x_0 = [2.5 \;\; 1.0 \;\; 0.0]^T$, and parameters of the vehicle are $m = 1.0$ and $J = 0.1$. Torque inputs are determined appropriately.

A nominal measurement noise is Gaussian white noise whose mean is $\mathbf{0}$ and covariance matrix R is given by

$$R = \begin{bmatrix} 0.29 & 0.30 & 0.36 \\ 0.30 & 0.53 & 0.30 \\ 0.36 & 0.30 & 0.49 \end{bmatrix}.$$

In use of the Cauchy distribution, Cauchy noise is added to the nominal measurement noise. Its parameters are $x_0 = 0$ and $\delta = 5 \times 10^{-2}$. In use of the Gaussian mixture distribution, $p = 0.3$, $N_x(0, \Sigma_1^2)$ is a distribution of the nominal measurement noise, and $N_x(0, \Sigma_2^2)$ is a distribution whose standard deviation is 20 times more than that of the nominal measurement noise.

Parameters of the RKF are $P_{0|0} = I$, $\sigma_{a_x}^2 = \sigma_{a_y}^2 = 1.0 \times 10^4$, and $\sigma_{a_\theta}^2 = 5.0 \times 10$. A covariance matrix of the nominal measurement noise assumes to be known.

MATLAB is used to compute the simulation. A CPU of a computer used in the simulation is Xeon X5550 (2.66GHz) and memory is 3GB. The proposed algorithm is compared with four methods, i.e., CVX, CVXGEN, FISTA and use of only diagonal elements of W. Since the RKF can compute its solution analytically if W is a diagonal matrix, use of only diagonal elements results in a fast computation. CVXGEN generates C code of QP optimization problems in CVX, and the compiled code is used in MATLAB to accelerate a computation. A solution of Eq. (4) requires a computation of LMI (11). However, CVXGEN and FISTA cannot deal with SDP like LMI, so one regularization parameter calculated by CVX is fixed in CVXGEN and FISTA. Furthermore, an estimation procedure of outlier

Table 1. Sum of root mean squared errors.

Type of methods	Cauchy noise	Gaussian mixture noise
KF without outliers	1.31	
KF with outliers	50.1	12.2
RKF using CVX	1.66	4.35
RKF using CVXGEN with fixed regularization parameter	1.54	4.94
RKF using FISTA with fixed regularization parameter	2.18	4.35
RKF using only diagonal elements	2.04	2.17
proposed method	1.52	1.80
proposed method (Compiled ver.)	1.52	1.78

Table 2. Average of computation time at one time step.

Type of methods	Time [ms]
KF	0.06
RKF using CVX	239
RKF using CVXGEN with fixed regularization parameter	0.11
RKF using FISTA with fixed regularization parameter	0.30
RKF using only diagonal elements	0.09
proposed method	0.10
proposed method (Compiled ver.)	0.08

in the proposed algorithm is implemented in C code, and the compiled code is also compared with them.

4.4 Results

Table 1 shows summations of root mean square errors (RMSEs) of each state, and Table 2 shows averaged computation times of each algorithm at one time step. These values are averages of 10 times simulations. Results of standard KF with and without outliers are also shown in the table for comparison. In the simulation, results using the proposed algorithm satisfy the condition (26).

Table 1 shows that the RKF can reduce effects of both Cauchy and Gaussian mixture noises. However, accuracy of the solutions depends on the algorithms, and it can be seen that the proposed algorithm gives smaller RMSE than the other algorithms.

Table 2 shows that, using CVXGEN, FISTA, use of only diagonal elements, and the proposed algorithm, computation times are about 1/1000 times less than one using CVX. Moreover, the compiled version of the proposed algorithm is more accelerated, and a computation time of the compiled version comes close to that of KF.

5. CONCLUSION

In this paper, we proposed a fast algorithm of RKF via l_1 regression, which consists of a l_1 optimization problems. The proposed algorithm approximates the optimal solution by using its upper and lower bounds, and the approximated solution is given by a closed form. Moreover, it was shown that the proposed algorithm has an almost same performance as KF without outliers. Effectiveness was demonstrated by some numerical simulations.

In larger scale optimization problems, or in some conditions of a covariance matrix of Gaussian noise, the condition (26) was not satisfied at times, and performances of the RKF were deteriorated. A proposition of an efficient algorithm for the situations is one of our future works.

ACKNOWLEDGEMENTS

This work was partially supported by Grant-in-Aid for Scientific Research (B) 24360166.

REFERENCES

A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems, *SIAM Journal on Imaging Sciences*, Vol. 2, No. 1, pp. 183–202 (2009)

H. W. Sorenso and D. L. Alspach. Recursive bayesian estimation using gaussian sums, *Automatica*, Vol. 7, No. 4, pp. 465–479 (1971)

I. Bilik and J. Tabrikian. Maneuvering target tracking in the presence of glint using the nonlinear gaussian mixture kalman filter, *IEEE Trans. on Aero. Elec. Sys.*, Vol. 46, No. 1, pp. 246–262 (2010)

J. F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones, *Optimization Methods and Software*, Vol. 11, No. 1, pp. 625-653 (1999)

J. Lofberg. YALMIP: a toolbox for modeling and optimization in MATLAB, *Proc. 2004 IEEE International Symposium on Computer Aided Control Systems Design*, pp. 284–289 (2004)

J. Mattingley and S. Boyd. Real-time convex optimization in signal processing – Recent advances that make it easier to design and implement algorithms –, *IEEE Signal processing magazine* (2010)

J. Mattingley and S. Boyd. CVXGEN: A code generator for embedded convex optimization, Optimization and Engineering, Vol. 13, No. 1, pp. 1–27 (2012)

K. C. Tho, M. J. Todd, and R. H. Tütüncü. SDPT3 – a MATLAB software package for semidefinite programming, version 1.3. *Optimization Methods and Software*, Vol. 11, No. 1, pp. 545-581 (1999)

K. Kurashiki, T. Fukao, K. Ishiyama, T. Kamiya, and N. Murakami. Orchard traveling UGV using particle filter based localization and inverse optimal control, *Proc. IEEE/SICE Int. Symp. on Sys. Integration*, pp 31-36 (2010)

M. Idan and J. L. Speyer. Cauchy estimation for linear scalar system, *IEEE Trans. Automat. Contr.*, Vol. 55, No. 6, pp. 1329–1342 (2010)

N. Guenard, T. Hamel, and R. Mahony. A practical visual servo control for an unmanned aerial vehicle, *IEEE Trans. on Robotics*, Vol. 24, No. 2, pp. 331–340 (2008)

P. J. Garrigues and L. El Ghaoui. An homotopy algorithm for the lasso with online observations, *Proc. Neural Information Processing Systems* (2008)

R. Zanetti. Recursive update filtering for nonlinear estimation, *IEEE Trans. Automat. Contr.*, Vol. 57, No. 6, pp. 1481–1490 (2013)

Y. Kaneda, Y. Irizuki, and M. Yamakita. Design method of robust Kalman filter for multi output systems based on statistics, *Proc. the 2013 American Control Conference*, pp. 1346–1351 (2013)