2018 Annual American Control Conference (ACC)
June 27–29, 2018. Wisconsin Center, Milwaukee, USA

Performance Specified State Estimation With Minimum Risk

Elahe Aghapour Jay A. Farrell

Abstract— **Measurements that significantly deviate from those predicted by the model or from the normal pattern of sensed data are considered as outliers. Since outliers can degrade the performance of state estimation, outlier accommodation is critical. The traditional Neyman-Pearson Kalman filter approach is to ignore all residuals greater than a designer specified threshold. The criticism of such techniques is that they allow missed detections to pass through undetected thereby corrupting both the state estimate and covariance. This causes the state estimation gain and all subsequent outlier decisions to be based on an invalid model.**

In sensor rich applications, where large numbers of sensor measurements are available, all the sensor data may not be needed to achieve the system accuracy specification. Global navigation satellite systems (GNSS) is one such sensor rich application since various satellite systems are available, each of which supplies more than the minimal number of satellites needed to estimate the system state. Using more data than is required to meet the specification exposes the state estimate to unneeded risk of outlier inclusion. This paper formulates and solves the state estimation problem from the perspective of minimizing risk while achieving a performance specification.

I. INTRODUCTION

An outlier is an observation that deviates from other observations or the model prediction by enough to arouse suspicion that it was generated by a different mechanism [1]. The outlier detection problem is fundamental to data-driven applications including: system identification, state estimation [2], image processing [3], mapping [4], etc. In recent years, autonomous vehicles have found application in tasks such as navigation, search and rescue, and real-time monitoring with rich information sources [5], [6]. Due to advances in information technology, larger and larger amounts of data are collected in databases. Hampel [7] estimates that a routine data set may contain $1-10\%$ outliers. Achieving a performance specification and accurately predicting when it is or is not achieved are critical goals. Hence, outlier accommodation continues to attract much attention.

In applications such as navigation, the environment is signal rich: images contain many features and GNSS comprises many separate systems each of which oversupplies the number of satellites necessary for state estimation. To achieve a specified level of state estimation accuracy, the full set of measurements is typically not required. If the full set of measurements was used, then the state estimate will have been exposed to unnecessary risk, while predicting through the computed covariance that it is over-performing relative to the specification, all the while likely including outliers that make the state estimate and performance prediction incorrect. Various proposed approaches [8] try to detect and remove outlier measurements detected relative to a threshold fixed at design time. The choice of that threshold entails a tradeoff between missed detections and false alarms. When the threshold is fixed at design time, it is independent of the actual measurements available and the actual accuracy performance that is achieved.

In sensor-rich environments, the state estimation problem can be considered from fresh perspectives built around optimization, e.g., [4], [9]–[11], that do not explicitly characterize each measurement is an outlier or inlier. In fact, this distinction is generally not observable. In [4], [9], latent variables are introduce to deactivate outliers in pose graph problems. The authors of [10] search for a maximum subset of pose graph measurements that are consistent with the same system. The least soft-thresholded squares approach developed in [11] uses a quadratic cost for residuals up to a given magnitude with a linear penalty thereafter, thereby decreasing the effect of outliers.

This article focuses on estimating the state vector to achieve a specified level of performance while incurring minimum risk. The resulting approach shares ideas with and was motivated by [10]. Where [10] focuses on choosing the maximal set of measurements self-consistent with a model, our approach herein focuses on achieving a specification with minimal risk. Implementation of our approach uses the indicator variable method introduced in [10].

Selecting a subset of measurements is similar to the sensor selection problem [12]–[15], which chooses a given number of measurements to minimize a cost function, e.g.: entropy, or log volume of the confidence ellipsoid. Our problem which minimizes risk with a constraint on performance is distinct from the sensor selection problem that minimizes a cost with a constraint on the number of chosen sensors.

The paper is organized as follows. Section II, the problem notation is presented and the unobservability of outlier is discussed. Section III, we present the problem formulation as an constrained convex optimization to select a set of measurements to achieve a lower bound on the information. Section IV solves the optimization in two steps. Section V discusses an example simulation application.

II. PROBLEM STATEMENT

A. Notation

Let $x_k \in \mathbb{R}^n$ represent the state vector at discrete-time k. The equation for the time-evolution of the state is:

$$x_k = \Phi_{k-1} x_{k-1} + G_{k-1} u_{k-1} + \omega_{k-1} \tag{1}$$

Aghapour is a Ph.D. student and Farrell is a Professor at the Department of Electrical Engineering, University of California, Riverside, 92521. {eaghapour, farrell}@ee.ucr.edu.

978-1-5386-5428-6/$31.00 ©2018 AACC

where $\Phi_k \in \mathbb{R}^{n\times n}$ and $G_k \in \mathbb{R}^{n\times \ell}$. The variables $u, \omega_k \in \mathbb{R}^{\ell}$ are a vector of user-determined (known) inputs and white Gaussian process noise with $\omega_k \sim \mathcal{N}(0, Q_k)$, respectively. We assume the availability of a known Gaussian prior probability function (PDF) $\mathcal{N}(x^+_{k-1}, P^+_{k-1})$.

The measurement vector z_k is modeled as:

$$z_k = H_k x_k + \eta_k \tag{2}$$

where $H_k \in \mathbb{R}^{m\times n}$ is the measurement matrix and $\eta \sim \mathcal{N}(0, R_k)$ represents white Gaussian measurement noise. The covariance matrix R_k is assumed to be invertible and diagonal[1]. In addition to measurement noise, z_k may be affected by outliers at some time instants.

A canonical representation of $\mathcal{N}(\mu, P)$ is given by the information matrix $J = P^{-1}$ and the information vector $\zeta = P^{-1}\mu$. The information vector and matrix are propagated to time k from $k-1$ using eqn. (1) as [16]:

$$\begin{aligned} J^-_k &= (\Phi_{k-1}(J^+_{k-1})^{-1}\Phi^\top_{k-1} + Q_k)^{-1} \\ \zeta^-_k &= J^-_k\left[(\Phi_{k-1}(J^+_{k-1})^{-1}\zeta^+_{k-1}) + G_{k-1}u_{k-1}\right]. \end{aligned} \tag{3}$$

At time k a measurement z_k related to the state x_k becomes available which is assumed to be generated in accordance with eqn. (2). Assuming use of the optimal gain, the information vector and matrix are updated according to:

$$\begin{aligned} J^+_k &= H^\top_k R^{-1}_k H_k + J^-_k \\ \zeta^+_k &= H^\top_k R^{-1}_k z_k + \zeta^-_k. \end{aligned} \tag{4}$$

B. Hypotheses testing

Standard approaches to accommodating outliers include two stages: *a*) residual generation (RG), which can be accomplished by least squares [17], recursive least squares [18], parity space [19], etc.; and, *b*) decision making (DM) [20]. Standard decision making methods test the residual $r_k = z_k - H_k\hat{x}$ between each measurement and its expected value relative to a threshold confidence bound β:

$$||z - H\hat{x}|| \leq \beta. \tag{5}$$

When condition (5) fails then an outlier has been detected.

For outlier identification eqn. (2) is modified to

$$z_k = H_k x_k + s + n \tag{6}$$

where $s \in \mathbb{R}^m$ is the unknown outlier vector. Unfortunately, x_k and s cannot both be estimated from z_k because the number of constraints m is less than the number of unknowns $m+n$. Therefore, the outlier estimation problem is unobservable.

Hypotheses testing solves this issue by adding structure to s that decreases the number of unknowns. For example, to detect single outliers consider each $i = 1, \ldots, m$ separately. Defining $s = \mu_i e_i$ changes eqn. (6) to

$$z_k = H_k x_k + \mu_i e_i + n \tag{7}$$

[1]Note that there is no restriction attached to this assumption. The solution can be used for any invertible covariance matrix by using the transformation $z' = \Sigma_r z$ with $R^{-1} = \Sigma^\top_r \Sigma_r$, the measurement model for z' is:

$$z' = H'x + \eta' \text{ where } H' = \Sigma_r H \ , \ \eta' \sim \mathcal{N}(0, I)$$

where I is the identity matrix.

where $e_i \in \mathbb{R}^m$ is the i-th column of the identity matrix and $\mu_i \in \mathbb{R}$ is the outlier magnitude. With this model, for each i, there are only $(n+1)$ unknowns, so both x and μ_i can be estimated as long as $m \geq n+1$. In this case, if μ_i is sufficiently large, the i-th element of z_k is classified as the outlier.

The above approach can be extended to detect ℓ-tuples of outliers by including ℓ parameters and ℓ columns of the identity matrix. The resulting problem is solvable for $\ell \leq (m-n)$ as long as H_k augmented with the appropriate columns of the identity matrix is full rank; however, reliability of the test decreases as $(m-n-\ell)$ decreases. Also, the number of hypotheses grows rapidly with ℓ:

$$\sum_{\ell=1}^{m-n} \begin{pmatrix} m \\ \ell \end{pmatrix} = \sum_{\ell=1}^{m} \frac{m!}{(m-\ell)!\,\ell!}$$

Outlier identification executes consecutively for each alternative hypothesis. Depending on the number of hypotheses, this testing can have high computation cost. If the actual outlier is not in the hypotheses set, or even if it is, depending on the decision threshold, it may be missed.

C. Summary of Approach

This article considers an alternative approach motivated by the ideas in [10]. Instead of focusing on outlier rejection, the goal herein will be to estimate the state vector x_k with specified accuracy $J^+_k \geq J_\ell$ when measurements u_{k-1} and z_k are given, with least risk of the state estimate being corrupted by outliers.

III. Problem Formulation

From the Maximum A Posteriori (MAP) perspective the state estimate is [16]:

$$\begin{aligned} x^\star_k &= \underset{x_k}{\text{argmax}}\ p(x_k, x_{k-1}, u_{k-1}, z_k) \\ &= \underset{x_k}{\text{argmax}}\ p(x_{k-1})p(x_k|x_{k-1}, u_{k-1})p(z_k|x_k) \end{aligned}$$

Given the Gausian assumptions for the prior, process, and measurement noise, the negative log-likelihood can be employed to convert the solution into the minimization of an equivalent least-squares cost function:

$$\begin{aligned} x^\star_k = \underset{x_k}{\text{argmin}} \big(&\|(\Phi_{k-1}x_{k-1} + G_{k-1}u_{k-1}) - x_k\|^2_{Q_k} \\ &+ \|H_k x_k - z_k\|^2_{R_k} + \|x_{k-1} - \hat{x}_{k-1}\|_{P^+_{k-1}} \big) \end{aligned} \tag{8}$$

where the squared Mahalanobis norm of vector r with covariance Σ is denoted by $||r||^2_\Sigma = r^\top \Sigma^{-1} r$ which is equivalent to $||r||^2_\Sigma = ||\Sigma^{\frac{1}{2}} r||_2$. The notation $\hat{x}$ will be used to represent the expected value of x. At this point, we drop the subscript k in the optimization variable x_k for simplicity. Eqn. (8) is a least-squares problem that can be efficiently solved, to yield the standard Kalman filter in information form. That approach does not address the existence of outliers. Therefore, outliers will cause both the state estimate and the computed error covariance to become inaccurate.

In [10], the main new idea was to change the focus from detecting outliers to finding the largest subset of the

measurements that can be consistently produced by the assumed measurement model. Their approach introduced a binary vector $b = (b_1 \quad b_2 \quad \cdots \quad b_m)^\top$ with $b_i \in \{0,1\}$ that will be used to disable or enable a measurement. If $b_i = 1$, the i^{th} measurement is in the optimization process and if $b_i = 0$, the measurement is ignored. Therefore, eqn. (8) becomes

$$\begin{aligned} x_k^\star, b^\star = \underset{x,b}{\operatorname{argmin}} \big(& \|x_{k-1} - \hat{x}_{k-1}^+\|_{P_{k-1}^+} \\ & + \|(\Phi_{k-1}\hat{x}_{k-1}^+ + G_{k-1}u_{k-1}) - x\|_Q^2 \\ & + \|\Phi(b)\big(Hx - z_k\big)\|_R^2\big) \end{aligned} \tag{9}$$

where $\Phi(b) = diag(b)$ and the variables $\hat{x}_{k-1}$, P_{k-1}^+, u_{k-1} and z_k are known.

The model and prior are always trusted (i.e., outlier-free). Therefore, the first and second term of the optimization can be propagated using eqn. (3) yielding:

$$\begin{aligned} x_k^\star, b^\star = \underset{x,b}{\operatorname{argmin}} & \left[\|x - x_k^-\|_{P_k^-}^2 + \|\Phi(b)^\top (Hx - z_k)\|_R^2\right] \\ \text{subject to: } & b_i \in \{0,1\} \text{ for } i = 1,\ldots,m, \end{aligned} \tag{10}$$

where $x_k^- = \Phi_{k-1}\hat{x}_{k-1} + G_{k-1}u_{k-1}$ represents the prior state vector estimate and $P_k^- = \Phi_{k-1}P_{k-1}^+\Phi_{k-1}^\top + Q$ is the prior covariance.

The cost function in eqn. (10) can be written as

$$C(x,b) = \|r(x,b)\|^2 \tag{11}$$

where $r(x,b) = A_b x - c_b$ with

$$A_b = \begin{bmatrix} \Sigma_R \Phi(b)^\top H \\ \Sigma_{P_k^-} \end{bmatrix} \text{ and } c_b = \begin{bmatrix} \Sigma_R \Phi(b)^\top z_k \\ \Sigma_{P_k^-} x_k^- \end{bmatrix}$$

and $(P_k^-)^{-1} = (\Sigma_{P_k^-})^\top (\Sigma_{P_k^-})$ and $R^{-1} = \Sigma_R^\top \Sigma_R$. For any fixed b, the minimum cost $C(x,b)$ as a function of x quantifies the risk associated with using the measurements selected by those indices with $b_i = 1$. The i-th element of the binary vector b determines whether or not the i-th row of A_b and c_b is non-zero. When b is given, the optimization problem in (10) for x is a least-squares problem that can be converted to the Normal equation

$$A_b x = c_b.$$

As currently stated, the minimum risk associated with (10) is achieved for $b = 0 \in \mathbb{R}^{m\times 1}$ (i.e., discarding all the measurements). In [10], the authors proceeded by finding the largest subset of the given measurements for which there exists a configuration which explains those measurements.

Alternatively, herein, we consider an approach that finds the set of measurements that satisfy a performance specification with minimum risk. The performance constrained optimization problem is

$$\begin{aligned} x_k^\star, b^\star = \underset{x,b}{\operatorname{argmin}} & \, C(x,b) \\ \text{subject to: } & J_b^+ \geq J_l \\ & b_i \in \{0,1\} \text{ for } i = 1,\ldots,m, \end{aligned} \tag{12}$$

where J_l is an user-defined minimum accuracy specification and J_b^+ is the posterior information matrix corresponding to the using the measurements that have $b_i = 1$.

Note that the solution of this problem is not to simply select the smallest residuals that yield observability. One reason is the existence of a prior, which may make measurements in certain directions less useful. A second reason is that satisfaction of the performance constraint will require the selected rows of H to be sufficiently distinct.

The diagonal covariance matrix R can be written as

$$R = \sum_{i=1}^m \sigma_i^2 e_i e_i^\top. \tag{13}$$

Using eqn (13) and the definition of $\Phi(b)$ following eqn. (9), the Fisher information matrix J_b in the optimization (12) is

$$\begin{aligned} J_b^+ &= H^\top \Phi(b)^\top \Big(\sum_{i=1}^m \frac{1}{\sigma_i^2} e_i e_i^\top \Big) \Phi(b) H + J_k^- \\ &= H^\top \Big(\sum_{i=1}^m \frac{b_i^2}{\sigma_i^2} e_i e_i^\top \Big) H + J_k^- \\ &= \sum_{i=1}^m \frac{b_i^2}{\sigma_i^2} h_i^\top h_i + J_k^- \end{aligned} \tag{14}$$

where h_i is the i^{th} row of H.

Therefore, the optimization problem becomes:

$$\begin{aligned} P1 : \min_{x,b} & \left[\|x - x_k^-\|_{P_k^-}^2 + \|\sum_{i=1}^m \frac{b_i}{\sigma_i} e_i e_i^\top (Hx - z_k)\|^2\right] \\ \text{subject to: } & \big(\sum_{i=1}^m \frac{b_i^2}{\sigma_i^2} h_i^\top h_i + J_k^- \big) \geq J_l \\ & b_i \in \{0,1\} \text{ for } i = 1,\ldots,m. \end{aligned}$$

Problem $P1$ is not convex. By using the binary definition of b, $b_i^2 = b_i$, so that $P1$ is completely equivalent to

$$\begin{aligned} P2 : \min_{x,b} & \left[\|x - x_k^-\|_{P_k^-}^2 + \|\sum_{i=1}^m \frac{b_i}{\sigma_i} e_i e_i^\top (Hx - z_k)\|^2\right] \\ \text{subject to: } & \big(\sum_{i=1}^m \frac{b_i}{\sigma_i^2} h_i^\top h_i + J_k^- \big) \geq J_l \\ & b_i \in \{0,1\} \text{ for } i = 1,\ldots,m \end{aligned}$$

which is separately (not jointly) convex in x and Boolean for b. By replacing the nonconvex constraint $b \in \{0,1\}^m$ with the convex constraint $b \in [0,1]^m$, $P2$ is relaxed to:

$$\begin{aligned} P2_r : \min_{x,b} & \left[\|x - x_k^-\|_{P_k^-}^2 + \|\sum_{i=1}^m \frac{b_i}{\sigma_i} e_i e_i^\top (Hx - z_k)\|^2\right] \\ \text{subject to: } & \big(\sum_{i=1}^m \frac{b_i}{\sigma_i^2} h_i^\top h_i + J_k^- \big) \geq J_l \\ & b_i \in [0,1] \text{ for } i = 1,\ldots,m. \end{aligned} \tag{15}$$

which is a convex problem separately in x and b with convex feasible set for b.

Note that the feasible set for the relaxed optimization in $P2_r$ contains the feasible set for $P1$. Therefore, the objective value of $P2_r$ is a lower bound on the objective value of $P1$ [13].

IV. Solution Method

A. General Case

Problem $P2_r$ is a convex problem for either variable x or b, separately, but not jointly. Hence, problem $P2_r$ can be solved by using multi-convex programming by alternatively updating b and x using the modified algorithm described below [21]. The *proximal terms* are added in the cost function (see below) where the convergence of this algorithm is proved in [22]. This problem is solved iteratively. The iteration number will be indicated by a right superscript ℓ, starting at zero.

1) *Selecting the measurements:* In this step, the optimal $b^{\ell+1}$ is found for fixed x_k^{ℓ}. Consequently the first term $\|x-x_k^-\|^2_{P_k^-}$ in problem (P2) will be dropped because it is independent of b. Based on [23], the proximal term $\lambda\|b-b^{\ell}\|^2$ is required to penalize the change of $b^{\ell+1}$ in comparison with b^{ℓ}, the optimization in standard form will be:

$$P3: \min_{b} \|\sum_{i=1}^{m}\frac{b_i}{\sigma_i}e_i e_i^{\top}(Hx_k-z_k)\|^2+\lambda\|b-b^{\ell}\|^2$$
$$\text{subject to: } J_l-\big(\sum_{i=1}^{m}\frac{b_i}{\sigma_i^2}h_i^{\top}h_i+J_k^-\big)\le 0$$
$$b_i\in[0,1] \text{ for } i=1,\dots,m.$$

where $\lambda>0$ is the user-defined proximal parameter. This is a least squares problem constrained by a linear matrix inequality (LMI). Since, the constraint on information matrix is in positive definite cone (see Appendix), optimization $P3$ is a standard semidefinite programming (SDP) problem and can be efficiently solved by interior point methods.

2) *State update:* In this step, the variable $x^{\ell+1}$ is optimized with fixed $b^{\ell+1}$. The proximal term $\beta\|x-x^{\ell}\|^2$ penalizes the change of $x^{\ell+1}$ in comparison with the last iteration. The optimization is:

$$P4: \min_{x}\Big[\|x-x_k^-\|^2_{P_k^-}+\|\sum_{i=1}^{m}\frac{b_i}{\sigma_i}e_i e_i^{\top}(Hx-z_k)\|^2 + \beta\|x-x^{\ell}\|^2\Big]$$

where $\beta>0$ is the user-defined proximal parameter. which is an unconstrained least squares optimization problem. The gradient of the cost function $C(x)$ is:

$$\nabla_x C(x) = 2P_k^{-^{-1}}(x-x_k^-) + 2H^{\top}\Phi(b)^2R^{-1}(Hx-z_k)+2\beta(x-x^{\ell}). \tag{16}$$

The optimal value for x_k can be computed as the roots of the gradient which results:

$$x_k = D^{-1}(P_k^{-^{-1}}x_k^- + H^{\top}\Phi(b)^2R^{-1}z_k+\beta x^{\ell})$$

where $D=(P_k^{-^{-1}}+H^{\top}\Phi(b)^2R^{-1}H+\beta I)$.

To initiate the optimization solution, the two steps are interchangeable. If an initial value of x is accurate (J_k^- is large), the algorithm can start with Step 1 to optimize b for fixed x. When the initial value of x is not accurate (J_k^- is small), the optimization can be started with Step 2 to find x assuming all the measurements are selected i.e. $b_i=1$ for $\forall i=1,\cdots,m$ which is equivalent to applying Least Squares method to estimate x. In either case, subsequent optimization steps are employed iteratively until the convergence criteria is met.

B. Specification for Reduced State

Sometimes maximizing a subset of the posterior information matrix $J_s\in\mathbb{R}^{s\times s}$ with $s\le n$ is of interest. We can define a transfer matrix $\{V\in\mathbb{R}^{n\times s}|V_{ij}=0,1\forall i,j\}$ such that $VV^{\top}\in\mathbb{R}^{n\times n}$ is a block matrix. The entries in the block matrices of $VV^{\top}$ are equal to one if the corresponding block matrix in J_b is of interest otherwise they are equal to zero.

By using transfer matrix V, we will have:

$$P5: \min_{x,b}\Big[\|x-x_k^-\|^2_{P_k^-}+\|\sum_{i=1}^{m}\frac{b_i}{\sigma_i}e_i e_i^{\top}(Hx-z_k)\|^2$$
$$\text{subject to: } V^{\top}\big(\sum_{i=1}^{m}\frac{b_i}{\sigma_i^2}h_i^{\top}h_i+J_k^-\big)V\ge J_{l_s}$$
$$b_i\in[0,1] \text{ for } i=1,\dots,m.$$

where $J_{l_s}\in\mathbb{R}^{s\times s}$ is an user defined lower bound for a subset of the posterior information matrix. Equivalently, the optimization (P5) can be solved by an algorithm similar to the algorithm in Section (IV) and the optimal value of x and b will be computed by iteratively updating b and x in two steps optimization.

V. NUMERICAL RESULTS

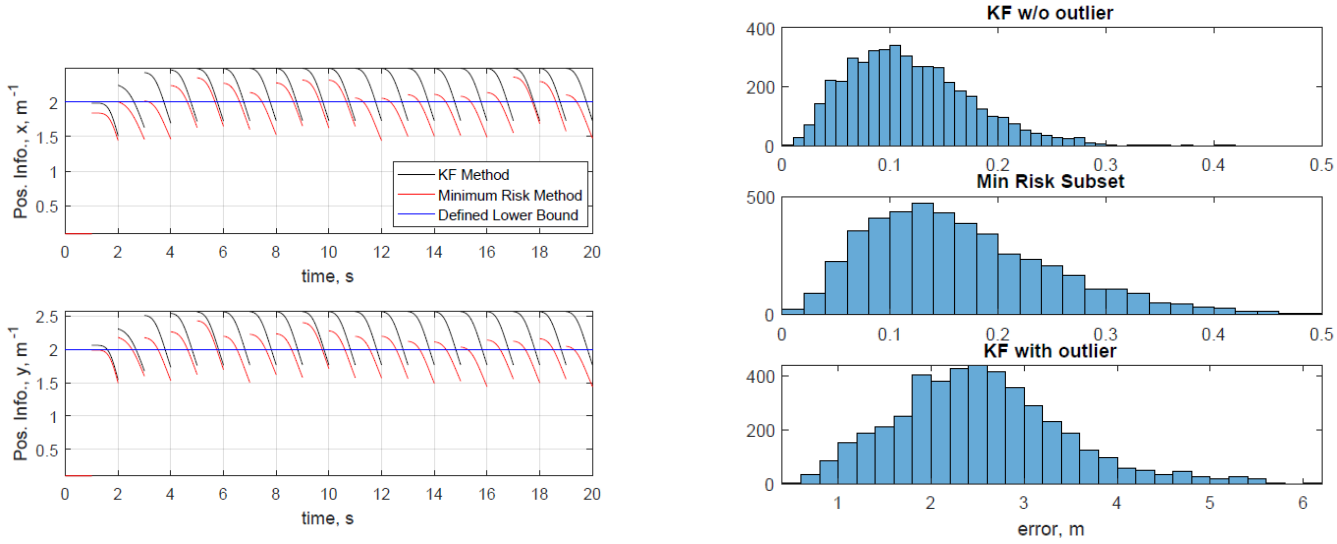
This section discusses a Matlab implementation to evaluate the performance of the approach proposed in Section IV-B. We will first consider the case without outliers, then evaluate our approach where outliers exist.

A. Setup Details

Assume a sensor moves with unknown white Gaussian acceleration. The state vector $x(k)=[p^{\top},v^{\top},a^{\top}]^{\top}\in\mathbb{R}^9$ comprises the $3D$ position, velocity and acceleration. The task is to estimate the state of the sensor as a function of time using sensor position measurements $z(k)\in\mathbb{R}^m$ available at time instants $t_k=kT$ for $k=1,2,3,\cdots$. At each time step the first three columns of the H matrix are an $m\times 3$ random matrix. All other columns are zero. For observability, three linearly independent position measurements are required. To simulate a sensor-rich environment, this example uses $m=10$. For the results that include outlier measurements, the outlier rate will be 20%. The information lower bound for the position components is selected to be 4 m^2. The remaining diagonal elements of the lower bound are set small enough to have no impact (i.e., always feasible). At the conclusion of the optimization (15), the real values of b were mapped back to binary values using threshold of 0.1.

B. Results

The minimum risk approach proposed herein will be compared with two Kalman filters (KF) each using m measurements. One will be immune from outliers. The other will be subject to with outliers at the stated rate, without outlier



(a) Position Information Matrix Comparison

(b) Position Error Histogram

Fig. 1: Performance Comparison

rejection. The data includes time propagation at 125HZ and position measurement at 1 HZ.

Fig. 1a shows the temporal progression of the square-root of the horizontal position components p_1 and p_2 of the computed information matrix during one example run. (i.e. $[J(t)]_1$ and $[J(t)]_2$ as propagated by eqns. (3, 4) At each integer second a measurement update occurs which increases the information matrix. Between the measurement time instants the position information decreases due to the accumulating velocity and acceleration uncertainty. For the KF both with or without outliers, the progression of the computed information matrix is identical, because the KF is unaware of the outliers. For the KF that is immune from outliers, this information temporal progression is correct. For the KF that is affected by outliers, both the state estimate and the information matrix are wrong with the information matrix being too large (i.e., optimisitic). For the minimum risk (MR) approach, after $k = 1$ the information increase achieves the lower bound (blue), but is not as large as the Kalman filter, due to the cost function penalizing risk.

The position error vector is

$$E(k) = \|\hat{p}(k) - p(k)\| \tag{17}$$

where $\hat{p}$ and p are estimated and true position states, respectively. Simulations lasting 50 seconds were repeated for 100 experiments with different randomly chosen initial conditions, outlier profiles, and acceleration sequences. Fig. 1b compares the position error histograms of the three methods. While the KF that is unaffected by outliers slightly outperforms the MR approach, the MR approach does satisfy the 0.5 m position error specification. The KF that is affected by outliers has position errors up to 6 m.

VI. CONCLUSIONS

This paper presents a novel approach to improve the robustness of a state estimation when the measurement data may contain outliers. The core contribution of this article is changing the focus from outlier detection, to looking for a subset of measurements which have minimum risk while achieving a lower bounded information for state estimation. We present a solution for the case when accuracy specification for a subset of information is of interest. We proposed a general, but tractable framework, for real-time linear state estimation. Our final formulation (Problem $P2_r$) is a constrained separately convex linear least squares optimization and can be solved by interior point methods. The simulation results herein were not real-time due to the computational load of solving SDPs. Algorithms with faster computation are of research interest. Extensions to nonlinear and sliding-window application and real-time implementation are of interest for future work.

ACKNOWLEDGMENT

This work was partially supported by NSF grant IIS-1316934. We gratefully acknowledge this support. All opinions expressed in this article are those of the authors.

VII. APPENDIX

Positive Semidefinite Analysis For Information Constraint: The covariance matrix is positive definite then based on eqn. (3), posterior information matrix J_k^- which is the inverse of covariance $J_k^- = P_k^{-1}$ is also positive definite. The second term of J_b as $H^\top \Phi(b)^\top R^{-1} \Phi(b) H$ is positive definite.

Using the definition of positive definite matrices, assume a nonzero $a = [a_1, a_2, \cdots, a_m]^\top \in \mathbb{R}^n$ to have:

$$a^\top H^\top \Phi(b)^\top R^{-1} \Phi(b) H a = (\Phi(b) H a)^\top R^{-1} (\Phi(b) H a)$$

where $\big(\Phi(b)Ha\big)_j = \sum_{i=1}^{n} \mathfrak{H}_{ji} a_i$. Subscript j denotes the j^{th} entry of a vector and $\mathfrak{H}_{ji}$ represents the ij^{th} entry of matrix

$\Phi(b)H$. By using the diagonal representation of R, eqn. (VII) will be:

$$a^\top H^\top \Phi(b)^\top R^{-1} \Phi(b) H a = \sum_{j=1}^{m} \sum_{i=1}^{n} \frac{1}{\sigma_j^2} (\mathfrak{H}_{ji} a_i)^2$$

which is greater than zero $\forall a \neq 0$. Hence, $H^\top \Phi(b)^\top R^{-1} \Phi(b) H$ is positive definite. Consequently, the information constrained is positive definite if and only if the lower bound J_l is chosen to be positive definite matrix.

REFERENCES

[1] D. Hawkins, *Identification of outliers*. Springer Science & Business Media, 2013.

[2] I. Kostavelis, E. Boukas, L. Nalpantidis, and A. Gasteratos, "Visual odometry for autonomous robot navigation through efficient outlier rejection," in *Imaging Systems and Techniques*. IEEE, 2013, pp. 45–50.

[3] S. Cho, J. Wang, and S. Lee, "Handling outliers in non-blind image deconvolution," in *Computer Vision*. IEEE, 2011, pp. 495–502.

[4] N. Sünderhauf and P. Protzel, "Switchable constraints vs. max-mixture models vs. RRR-a comparison of three approaches to robust pose graph SLAM," in *ICRA*. IEEE, 2013, pp. 5198–5203.

[5] A. Bahr, J. J. Leonard, and M. F. Fallon, "Cooperative localization for autonomous underwater vehicles," *The International Journal of Robotics Research*, vol. 28, pp. 714–728, 2009.

[6] J. Vaganay, J. J. Leonard, and J. G. Bellingham, "Outlier rejection for autonomous acoustic navigation," in *ICRA, 1996. Proc.*, vol. 3. IEEE, 1996, pp. 2174–2181.

[7] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust statistics: the approach based on influence functions*. John Wiley & Sons, 2011, vol. 114.

[8] S. Hewitson and J. Wang, "Extended receiver autonomous integrity monitoring (E RAIM) for GNSS/INS integration," *Journal of Surveying Engineering*, vol. 136, pp. 13–22, 2010.

[9] N. Sünderhauf and P. Protzel, "Towards a robust back-end for pose graph SLAM," in *Robotics and Automation (ICRA)*. IEEE, 2012, pp. 1254–1261.

[10] L. Carlone, A. Censi, and F. Dellaert, "Selecting good measurements via l1 relaxation: A convex approach for robust estimation over graphs," in *IROS*. IEEE, 2014, pp. 2667–2674.

[11] D. Wang, H. Lu, and M.-H. Yang, "Robust visual tracking via least soft-threshold squares," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, pp. 1709–1721, 2016.

[12] H. Wang, K. Yao, G. Pottie, and D. Estrin, "Entropy-based sensor selection heuristic for target localization," in *3rd Int. Information Processing in Sensor Networks*. ACM, 2004, pp. 36–45.

[13] S. Joshi and S. Boyd, "Sensor selection via convex optimization," *IEEE Transactions on Signal Processing*, vol. 57, pp. 451–462, 2009.

[14] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *IEEE Decision and Control Conference*, 2010, pp. 2572–2577.

[15] S. P. Chepuri and G. Leus, "Sparsity-promoting sensor selection for non-linear measurement models," *IEEE Transactions on Signal Processing*, vol. 63, pp. 684–698, 2015.

[16] S. Thrun, W. Burgard, and D. Fox, *Probabilistic robotics*. MIT press, 2005.

[17] P. M. Frank and X. Ding, "Survey of robust residual generation and evaluation methods in observer-based fault detection systems," *Journal of process control*, vol. 7, pp. 403–424, 1997.

[18] R. Brown and P. Hwang, "From RAIM to NIORAIM a new integrity approach to integrated multi-gnss systems," *Inside GNSS, May-June*, 2008.

[19] J. Gertler and D. Singer, "A new structural framework for parity equation-based failure detection and isolation," *Automatica*, vol. 26, pp. 381–388, 1990.

[20] R. V. Beard, "Failure accomodation in linear systems through self-reorganization." Ph.D. dissertation, Massachusetts Institute of Technology, 1971.

[21] X. Shen, S. Diamond, M. Udell, Y. Gu, and S. Boyd, "Disciplined multi-convex programming," in *Control And Decision Conf.* IEEE, 2017, pp. 895–900.

[22] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optim.*, vol. 23, no. 2, pp. 1126–1153, 2013.

[23] H. Attouch, J. Bolte, P. Redont, and A. Soubeyran, "Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Lojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010.