2018 European Control Conference (ECC)
June 12-15, 2018. Limassol, Cyprus

Performance-Specified Moving-Horizon State Estimation With Minimum Risk

Elahe Aghapour Jay A. Farrell

Abstract— **This paper is concerned with the estimation of the state of a linear dynamic system when the measurements may contain outliers. The most common method for outlier detection utilizes the traditional Neyman-Pearson (NP) Kalman filter approach which ignores all residuals greater than a designer specified threshold. When measurements with outliers are used (i.e., missed detections), the estimated state becomes incorrect and the computed state error covariance is too small, yielding an over confidence in the estimator in the incorrect state estimate. When valid measurements are ignored, information is lost, but this is only critical if it causes the performance specification to be violated. In signal rich applications, with a large number of sensor measurements, a smaller subset of measurements than is accepted by the NP approach, could be able to achieve the specified level of performance with lower risk of including an outlier in the set of utilized measurements.**

In the moving-horizon approach used herein, the number of measurements available for state estimation is affected by both the number of measurements per time step and the number of time steps over which measurements are retained. This moving horizon, performance-specified, risk-averse state estimation approach will be formulated in an optimization setting that selects measurements from within the window, to achieve a specified level of performance while minimizing the incurred risk. Simulation results are included, which demonstrate the application of the technique and its enhanced performance and robustness to outliers relative to traditional methods.

I. INTRODUCTION

For many applications, the optimal state estimation is required while mitigating sensor faults and measurement outliers. An outlier is generally defined as an observation that lies outside some overall pattern of distribution and contains no information about the system to be estimated [1]. Reliable process data are the key for system identification but measured data often contain noise and are also frequently contaminated by outliers which are produced by degradation, human error, or other unmeasured disturbances. Hence, the outlier detection problem is fundamental to data-driven applications including: state estimation [2], image processing [3], mapping [4], autonomous vehicles navigation [5], [6], and etc. Due to advances in information technology, larger and larger amounts of data are collected in databases while a routine data set may contain $1-10\%$ outliers [7].

Fault Detection and Removal includes two stages: *a*) residual generation (RG), which can be performed by least squares [8], recursive least squares [9], parity space [10], etc.; and, *b*) decision making (DM) [11]. However, two drawbacks attached to these methods. First, they suffer from high computational cost. Second, the outlier detection is unobservable.

Motivated by the fact that the distinction between inlier and outlier is generally not observable, [4], [12]–[14] considered the state estimation problem from fresh perspectives built around optimization that do not explicitly characterize each measurement is an outlier or inlier. The problem of outliers in pose graph optimization is surveyed in [4], [12] where latent variables are introduce to deactivate outliers in pose graph problems. In [15], a maximum subset of pose graph measurements is selected that are consistent with the same system.

This paper considers an alternative approach motivated by the ideas in [15] where choosing the maximal set of measurements self-consistent with a model in pose graph optimization is of interest. Similar to [16], the goal is to choose a subset of measurements in moving horizon which satisfies the accuracy specification and yields to minimum risk state estimate. We surveyed the problem for using single epoch of measurements at each time. There are two disadvantages in this method. first, incorrect linearization point can lead to select an outlier in chosen subset of measurements and consequently, an incorrect state estimate. Second, Data redundancy is critical to be able to choose a subset of them which satisfies the specified accuracy. Both drawbacks can be solved by estimating the state in a sliding window of time.

The paper is organized as follows. Section II, the problem notation for state estimation in a fixed lag moving horizon is presented. Section III, the problem is formulated as an constrained convex optimization which selects a set of measurements in a window to achieve a lower bound on the information. Section IV solves the optimization when *a*) the accuracy of the last state, and, *b*) choosing a maximal subset of measurements are of interest. Section V discusses an example simulation application and compares the results with method in [16].

II. PROBLEM STATEMENT

Let $x_k \in \mathbb{R}^n$ represent the state vector at time k. Assume a Gaussian prior probability function (PDF) $\mathcal{N}(x^+_{k-L}, P^+_{k-L})$ for the earliest state in the sliding window. The equation for the time-evolution of the state is:

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

where $\Phi_k \in \mathbb{R}^{n\times n}$ and $G_k \in \mathbb{R}^{n\times \ell}$. The variable $u_k \in \mathbb{R}^\ell$ is a vector of user-determined (known) inputs and $\omega_k \in \mathbb{R}^n$ is a white Gaussian process noise with $\omega_k \sim \mathcal{N}(0, Q_k)$.

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

The measurement vector z_k at time k is modeled as:

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

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

At time-step k, the trajectory of the state vectors $X = [x_{K-L}^\top, \cdots, x_k^\top]^\top \in \mathbb{R}^{(L+1)n}$ will be simultaneously estimated over a sliding time-window with length parameterized by L. The window will slide 1 epoch upon arrival of each new measurement vector z_k. At time k, when the time-window slides, one old state vector $\hat{x}_{k-L-1}$ will be marginalized out of the previous window to yield the Gaussian prior probability function (PDF) $\mathcal{N}(\hat{x}_{k-L}, P_{k-L})$ for the new window.

The Maximum A Posteriori (MAP) approach computes the trajectory of state vectors $X = [x_{K-L}^\top, \cdots, x_k^\top]^\top \in \mathbb{R}^{(L+1)n}$ that maximize the posterior pdf [13], [17], [18]:

$$\hat{X} = \underset{X}{\operatorname{argmax}}\, P(X, U, Z) \tag{3}$$

where $U = \{u_i,\ \forall\ i = k-L, \cdots, k-1\}$ and $Z = \{z_i,\ \forall\ i = k-L+1, \cdots, k\}$. Applying Bayes' theorem to (3) yields

$$\begin{aligned}\hat{X} &= \underset{X}{\operatorname{argmax}}\, P(X, U, Z) \\ &= \underset{X}{\operatorname{argmax}}\, p(x_{k-L}) \prod_{i=0}^{L-1} p(x_{k-L+1+i} | x_{k-L+i}; u_{k-L+i}) \\ &\qquad \prod_{j=1}^{L} p(z_{k-L+j} | x_{k-L+j})\end{aligned} \tag{4}$$

where $p(x_{i+1}|x_i; u_{i+1})$ is the distribution of state evolution based on eqn. (1) and $p(z_j|x_j)$ is the distribution of the measurement of state based upon eqn. (2). Maximizing the posterior pdf is equivalent to minimizing the negative log-likelihood of (4):

$$\begin{aligned}\hat{X} = \underset{X}{\operatorname{argmin}}\ & \|x_{k-L} - \hat{x}_{k-L}\|^2_{P_{k-L}} \\ &+ \sum_{i=k-L}^{k-1} \|x_{i+1} - (\Phi_i x_i + G_i u_i)\|^2_{Q_i} \\ &+ \sum_{j=k-L+1}^{k} \|z_j - H x_j\|^2_{R_j}.\end{aligned} \tag{5}$$

The Cholesky Decomposition to a matrix W provides Σ_W such that $W^{-1} = \Sigma_W^\top \Sigma_W$. With this notation, the squared Mahalanobis distance $\|v\|_W = \|\Sigma_W v\|_2$ where notation $\|x\|$ represents the 2-norm of vector x. The optimization (5) can be transformed to matrix form with

$$\hat{X} = \underset{X}{\operatorname{argmin}}\, \|\hat{r} - DX\|^2 \tag{6}$$

where the Jacobian matrix D and residual vector $\hat{r}$ are defined by:

$$D = \left[\begin{array}{cccccc} \Sigma_{P_{k-L}} & 0 & 0 & \cdots & 0 & 0 \\ \hline \bar{\Phi}_{k-L} & -\Sigma_{Q_{k-L}} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & & \\ 0 & 0 & 0 & \cdots & \bar{\Phi}_{k-1} & -\Sigma_{Q_{k-1}} \\ \hline 0 & \bar{H}_{k-L+1} & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & & \\ 0 & 0 & 0 & \cdots & 0 & \bar{H}_k \end{array}\right] = \left[\begin{array}{c} D_0 \\ \hline D_1 \\ \hline D_2 \end{array}\right]. \tag{7}$$

$$\hat{r} = \left[\begin{array}{c} \Sigma_{P_{k-L}} \hat{x}_{k-L} \\ \hline 0_{n\times 1} \\ \vdots \\ 0_{n\times 1} \\ \hline \Sigma_{R_{k-L+1}} z_{k-L+1} \\ \vdots \\ \Sigma_{R_k} z_k \end{array}\right] = \left[\begin{array}{c} \hat{r}_0 \\ \hline \hat{r}_1 \\ \hline \hat{r}_2 \end{array}\right].$$

where $\bar{\Phi}_k = \Sigma_{Q_k} \Phi_k$ and $\bar{H}_k = \Sigma_{R_k} H_k$. The matrices $D_0 \in \mathbb{R}^{n\times(L+1)n}$, $D_1 \in \mathbb{R}^{Ln\times(L+1)n}$, and $D_2 \in \mathbb{R}^{Lm\times(L+1)n}$ represent the sub-matrices in D that correspond with the subvectors $\hat{r}_0 \in \mathbb{R}^n$, $\hat{r}_1 \in \mathbb{R}^{Ln}$, and $\hat{r}_2 \in \mathbb{R}^{Lm}$ of $\hat{r}$, as indicated by the horizontal lines. Eqn. (6) is a linear least-squares problem which can be solved using the QR decomposition, the singular value decomposition (SVD), etc. However, measurement z_k may have outlier whose effect should be removed prior to trajectory estimation.

A. Summary of Approach

While the best strategy may be to detect and remove outlier measurements, no approach will work perfectly. This article considers an alternative approach motivated by the ideas in [15]. The authors of [15] search for a maximum subset of pose graph measurements that are consistent with the same system model, instead of focusing on outlier rejection. The method is implement through a measurement selection vector denoted by b.

In [16], the idea is proposed to choose a subset of measurements to estimate the state vector x_k, where subset is selected to have minimum risk of containing an outlier while achieving a specified accuracy. In the present article, we consider the problem of state estimation in the presence of outliers within a fixed-lag sliding window. Analysis and discussion of how selecting a subset of measurements can improve the state estimation is provided.

III. Problem Formulation

To estimate the state trajectory X over a time window with length L, the risk expression in eqn. (6) is minimized. However, outliers in the set of measurements Z can yield the incorrect state estimate and error covariance. Herein, the main idea to overcome this problem is changing the focus from detecting the outlier to finding a subset of the measurements which minimize the risk that satisfies the specified accuracy bound. To select a subset of measurements, a selection vector $b = (b_1 \quad b_2 \quad \cdots \quad b_L)^\top \in \mathbb{R}^{Lm}$ is introduced, where

$b_i = (b_{i1} \quad b_{i2} \quad \cdots \quad b_{im})^\top \in \mathbb{R}^m$ is a binary indicator vector with $b_{ij} \in \{0,1\}\ \forall i = 1,\cdots,L,\ j = 1,\cdots,m$. The binary selection vector disables the measurement represented by the j^{th} element of z_{k-L+i} when $b_{ij} = 0$ or enables it when $b_{ij} = 1$.

Assuming that the signal u and model (1) are trusted, then by defining the block diagonal matrix $\Psi(b) = \begin{bmatrix} I_{n(L+1)} & 0 \\ 0 & \phi(b) \end{bmatrix}$, eqn. (6) can be generalized as:

$$\begin{aligned} X^\star, b^\star = &\underset{x,b}{\operatorname{argmin}} \|\Psi(b)\big(DX - \hat{r}\big)\|^2 \\ &\text{subject to: } b_{ij} \in \{0,1\}\ \forall i = 1,\cdots,L,\ j = 1,\cdots,m, \end{aligned} \tag{8}$$

where $\phi(b) = diag(b)$ and I_n denotes the $n \times n$ identity matrix. Note that eqns. (8) and (6) are equivalent if the optimization over b is dropped and all measurements are selected (i.e., all elements of $b_{ij} = 1$).

For any fixed b, the cost $C(X,b) = \|\Psi(b)\big(DX - \hat{r}\big)\|^2$ as a function of X quantifies the risk associated with using the selected measurements with $b_{ij} = 1$. The ij-th element of the binary vector b determines whether or not the $[(i-1)m + j]$-th row of D_2 and $\hat{r}_2$, respectively, are non-zero. The optimization problem (8) for given b, is a linear least square problem.

The performance constrained optimization problem is:

$$\begin{aligned} X^\star, b^\star = &\underset{x,b}{\operatorname{argmin}}\, C(X,b) \\ &\text{subject to: } J_b \geq J_l \\ &\qquad b_{ij} \in \{0,1\}\ \forall i = 1,\cdots,L,\ j = 1,\cdots,m, \end{aligned} \tag{9}$$

where J_l is an user-defined lower bound specified for the accuracy as quantified by the Fisher information matrix corresponding to the selected measurements

$$J_b = D^\top \Psi(b)^\top \Psi(b) D.$$

The selection matrix $\Psi(b)$ is diagonal and symmetric. In addition, the binary assumption on b yields $\Psi(b)^\top \Psi(b) = \Psi(b)^2 = \Psi(b)$. Hence, the optimization (9) is equivalent to:

$$\begin{aligned} P1: X^\star, b^\star = &\underset{x,b}{\operatorname{argmin}} \big[\|D_0 X - \hat{r}_0\|^2 + \|D_1 X - \hat{r}_1\|^2 \\ &+ \|\Phi(b)\big(D_2 X - \hat{r}_2\big)\|^2 \\ &\text{subject to: } D^\top \Psi(b) D \geq J_l \\ &\qquad b_{ij} \in \{0,1\}\ \forall i = 1,\cdots,L,\ j = 1,\cdots,m. \end{aligned}$$

Problem (P1) is an unconstrained convex optimization for x and constrained Boolean optimization for b. By replacing the nonconvex constraint $b \in \{0,1\}^{Lm}$ with the convex constraint $b \in [0,1]^{Lm}$, Problem (P1) is relaxed to:

$$\begin{aligned} P2: X^\star, b^\star = &\underset{x,b}{\operatorname{argmin}} \big[\|D_0 X - \hat{r}_0\|^2 + \|D_1 X - \hat{r}_1\|^2 \\ &+ \|\Phi(b)\big(D_2 X - \hat{r}_2\big)\|^2 \\ &\text{subject to: } D^\top \Psi(b) D \geq J_l \\ &\qquad b_{ij} \in [0,1]\ \forall i = 1,\cdots,L,\ j = 1,\cdots,m, \end{aligned} \tag{10}$$

which is a separately convex in X and b.

Note that the feasible set of $P2$ (i.e., convex constraint $b \in [0,1]^{Lm}$) contains the feasible set of $P1$ (i.e., Boolean nonconvex constraint $b \in \{0,1\}^{Lm}$). Therefore, the objective value of $P2$ is a lower bound on the objective value of $P1$ [19].

Problem P2 is distinct from standard techniques [8]–[10], [20], [21] that make decisions based on a user-defined threshold on the residual magnitude. The solution of P2 does not just select the smallest residuals; instead, it minimizes risk (as partially characterized by residual magnitude) while ensuring that a performance specification is satisfied. Satisfaction of the performance constraint will require the selected rows of D_2 to be sufficiently distinct. The existence of the prior (i.e., first two terms of eqn. (10)) may make the measurements in certain directions more or less useful.

IV. PROPOSED OPTIMIZATION SOLUTION

A. General Case

Optimization $P2$ is a problem in which the variables can be partitioned into sets (e.q. X and b) over which the problem is convex when the other variable is fixed. Hence, $P2$ can be solved by using multi-convex programming by alternatively updating b and x, as in [22], with the addition of *proximal terms* in the cost function (see below). This problem is solved iteratively. The iteration number will be indicated by a right superscript ℓ, starting at zero.

To initiate the iterative solution, the two steps of updating b and X are interchangeable. If an initial value of state X^0 is accurate (i.e. J^0 is large), the algorithm can start by finding b^1 for fixed X^0 in Step 1. When the initial value of X^0 is not accurate (i.e. J^0 is small), the optimization can start by finding X^1 in Step 2 assuming all the measurements are selected (i.e. $b_{ij}^0 = 1$ for $\forall i = 1,\cdots,L,\ j = 1,\cdots,m$) which is equivalent to applying the Least Squares method to estimate X. In the ℓ^{th} iteration, $X^{\ell+1}$ and $b^{\ell+1}$ are computed in two steps:

1) *Selecting the measurements:* Optimize the selection vector $b^{\ell+1}$ for fixed X^ℓ. The first two terms in $P2$ can be dropped because they are 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:

$$\begin{aligned} P3: &\min_b \|\Phi(b)\big(D_2 X^\ell - \hat{r}_2\big)\|^2 + \lambda\|b - b^\ell\|^2 \\ &\text{subject to: } J_l - D^\top \Psi(b) D \leq 0 \\ &\qquad b_{ij} \in [0,1]\ \forall i = 1,\cdots,L,\ j = 1,\cdots,m. \end{aligned}$$

where $\lambda > 0$ is the user-defined proximal parameter. This is a least squares problem constrained by a linear matrix inequality (LMI) that be solved using an interior point method, standard semidefinite programming (SDP), etc.

2) *State update:* Optimize the variable $X^{\ell+1}$ for fixed b^ℓ. The proximal term $\beta\|X - X^\ell\|^2$ penalizes the change of $X^{\ell+1}$ in comparison with the last iteration. The optimization is:

$$\begin{aligned} P4: \min_X &\big[\|D_0 X - \hat{r}_0\|^2 + \|D_1 X - \hat{r}_1\|^2 \\ &+ \|\Phi(b^\ell)\big(D_2 X - \hat{r}_2\big)\|^2 + \beta\|X - X^\ell\|^2 \end{aligned}$$

where $\beta > 0$ is the user-defined proximal parameter. Problem P4 is an unconstrained least squares optimization problem and can be transformed into:

$$P4: \min_X \|Ax - c\|^2, \tag{11}$$

where $A^\top = \left[D_0^\top, D_1^\top, \Phi(b^{\ell+1})D_2^\top, \beta I_{(L+1)n}\right]^\top$ and $c^\top = \left[\hat{r}_0^\top, \hat{r}_1^\top, \Phi(b^{\ell+1})\hat{r}_2^\top, X^{\ell\top}\right]^\top$. Hence, eqn. (11) can be minimized using QR decomposition, singular value decomposition (SVD), etc.

Note that problem $P2$ is solved by alternatively solving $P3$ and $P4$ until a convergence condition is met.

B. Discussion Of Trade-offs

The solution b to $P2$ defines a subset of measurements with minimum risk that satisfies the information constraint. There is a trade-off in the specified accuracy. Specifying a higher accuracy through a larger lower bound J_l for the Fisher information of the selected measurements would also increase the risk of inclusion of outliers.

The largest possible lower bound J_l corresponds to the Fisher information matrix computed from using all measurements. This is also the highest risk solution. If J_l is larger, then the performance specification yields no feasible solutions.

C. Last State x_k Accuracy Specification

At each time k, for real-time implementation, an important goal is to estimate the state $\hat{x}_k$ and its uncertainty P_k. However, using a single epoch of measurements may not provide sufficient redundancy to remove the effects of outliers. In presence of outliers this can yield to an inaccurate state estimate and covariance, with the covariance being too small and the state estimate being biased. Hence, we consider the problem to estimate a trajectory of state vectors $X = [x_{K-L}^\top, \cdots, x_k^\top]^\top \in \mathbb{R}^{(L+1)n}$ by selecting measurements from a sliding window of time with length L to have minimum risk and satisfy the accuracy specification. However, the estimate of the last state x_k and its covariance are most important. Hence, instead of a lower bounded on the information for the trajectory of states X, a performance accuracy of the last state x_k can be of primary interest.

The Jacobian matrix can be partitioned as

$$J = D^\top \Psi(b) D = \begin{bmatrix} J_{LL} & J_{Lk} \\ J_{kL} & J_{kk} \end{bmatrix}$$

with $J_{LL} \in \mathbb{R}^{nL \times nL}$ and $J_{kk} \in \mathbb{R}^{n \times n}$. Let $P_k = cov(\hat{x}_k)$, then $J_k = P_k^{-1}$ is the information matrix for the last state and $J_k = J_{kk} - J_{kL}J_{LL}^{-1}J_{Lk}$ [24]. When concerned only with the accuracy of $\hat{x}_k$, the problem (P2) would change to:

$$\begin{aligned} X^\star, b^\star = \underset{x,b}{\operatorname{argmin}} & \left[\|D_0 X - \hat{r}_0\|^2 + \|D_1 X - \hat{r}_1\|^2 \right. \\ & \left. + \|\Phi(b)\big(D_2 X - \hat{r}_2\big)\|^2 \right] \\ \text{subject to: } & J_{kk} - J_{kL}J_{LL}^{-1}J_{Lk} \geq J_{l_k} \\ & b_{ij} \in \big[0,1\big] \; \forall i = 1, \cdots, L, \; j = 1, \cdots, m, \end{aligned} \tag{12}$$

where J_{l_k} is a user defined lower bound for the last state x_k accuracy. The optimization (12) has a nonlinear constraint.

To avoid the nonlinear constraint, a user might instead consider imposing the constraint $J_{kk} \geq J_{l_{kk}}$. However, this constraint is equivalent to

$$J_k + J_{kL}J_{LL}^{-1}J_{Lk} \geq J_{l_{kk}} \tag{13}$$

$$J_k \geq J_{l_{kk}} - J_{kL}J_{LL}^{-1}J_{Lk}. \tag{14}$$

Since $J_{LL} \geq 0$, this approach does not guarantee that $J_k \geq J_{l_{kk}}$.

V. NUMERICAL RESULTS

This section discusses a Matlab implementation to evaluate the performance of the approach proposed herein. We will compare the performance-specified minimum-risk (PSMR) approach with outliers to the standard approach in the ideal (outlier free) and outlier affected cases.

A. Setup Details

Let $x(k) = [p^\top(k), v^\top(k), a^\top(k)]^\top \in \mathbb{R}^9$ denote the state vector of a sensor which is comprised of $3D$ position, velocity and acceleration. Assume that the acceleration is unknown white Gaussian noise.

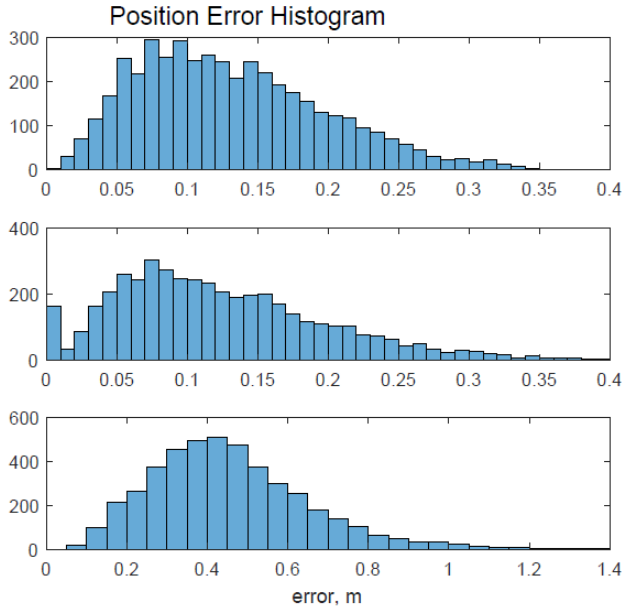
The task is to estimate the state of the sensor using measurements $z(k) \in \mathbb{R}^m$ that are a linear function of the position that are available at time instants $t_k = kT$ for $k = 1, 2, 3, \cdots$. At each time step, the first three columns of the matrix $H_k \in \mathbb{R}^{m \times 9}$ in eqn. (2) are a random matrix. The remaining six columns are zero. For observability, at least three linearly independent position measurements are required (i.e., $m \geq 3$). To simulate a sensor-rich environment, this example uses $m = 10$, which is reasonable in multi-constellation GNSS applications.

The MAP approach is implemented with a fixed duration sliding window of length $L = 4$. For the PSMR approach the information lower bound for the position components of the last state x_k is selected to be $4I\ \mathrm{m}^{-2}$ (inverse meters squared), while the remaining diagonal elements of the lower bound are set small enough to have no impact (i.e., always feasible). The lower bound of 4.0 corresponds to a position error standard deviation less than 0.5m. Two approaches are possible for accomodating the real b vector that results from the optimization. One appraoch is to use the real b. This approach effectively magnifies the noise variance from σ_i^2 to $\frac{\sigma_i^2}{b_i}$, see [25]. A second approach is to map the real b to binary b using a threshold. The results that follow use the second approach with a threshold of 0.1m.

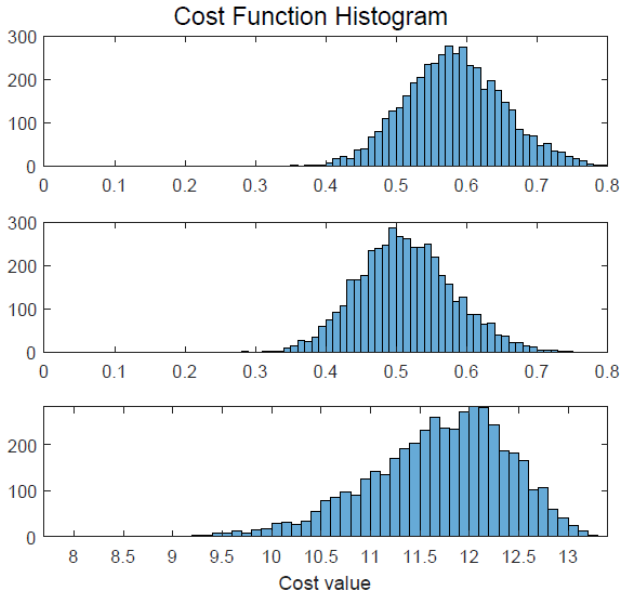
The PSMR approach will be compared with the minimum risk (MR) method described in [16] and two MAP implementations. One of the MAP implementations, the ideal case, will use outlier-free measurements. The other implementations, without outlier accommodation, will be subject to outliers with a 20% rate of occurrence.

Each simulation experiment is implemented for $k = 1, \cdots, 50$ with randomly chosen outliers. The number of experiments is 100.The position error vector is calculated as

$$E(k) = \|\hat{p}(k) - p(k)\|_2$$



(a) Position Error Histogram



(b) Cost Function Value Histogram

Fig. 1: Performance Comparison for time window $L = 4$: (Top) MAP with no outlier. (Middle) PSMR with outlier. (Bottom) MAP with outlier.

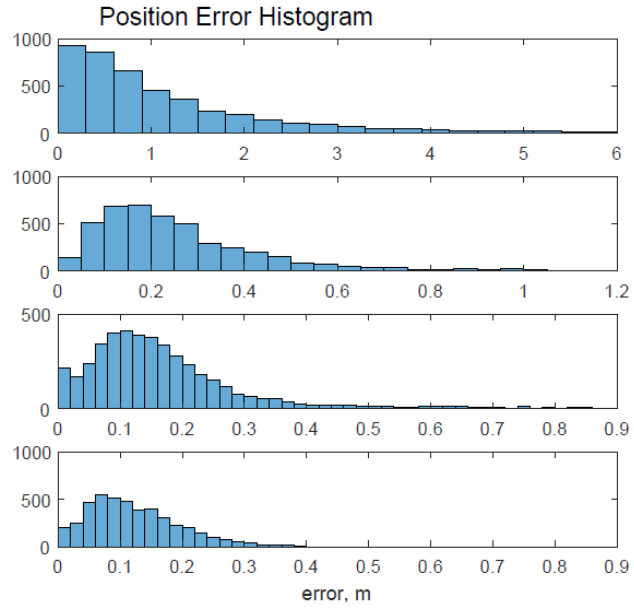


Fig. 2: Position Error Histogram: Labeled from top to bottom with a-d (a) KF with 3 Smallest residual. (b) KF with 6 Smallest residual. (c) KF with 8 Smallest residual. (d) PSMR at each time.

where $\hat{p}$ is the estimated position and p is the true position.

B. Results

Fig. 1a compares the position error histogram for the three state estimation methods described herein. Fig. 1b shows the cost (i.e., risk) histogram for each method in the corresponding rows. The top row displays the results when the sliding-window MAP approach uses all 10 measurements at each time and none of the measurements are affected by outliers. The maximum position error is 0.35m. The middle row displays the results for the PSMR approach for the stated performance specification and the same measurements, except to outliers affecting the measurements. Note that the performance is similar to the ideal case and the risk is slightly less (i.e., The histogram is shifted to the left.), even in the presence of outliers. The risk is decreased because the PSMR approach does not use all the measurements. Typically, over 60% of measurements in each time window are used. Using fewer measurements causes a slight increase in the maximum position errors to 0.5m; however, the performance specification is satisfied. The bottom row shows the results for the sliding-window MAP approach, without outlier accommodation using exactly the same outlier affected measurements as was used to generate the PSMR results. In this case, position errors are experienced up to 6 m (the figure axis is truncated) and the risk histogram shifts significantly to the left.

It is important to note that the PSMR approach is not simply using the smallest residuals. Selecting the smallest residual might not satisfy the performance constraint, because the corresponding rows of the H matrix may not be sufficiently diverse. Fig. 2 is included to illustrate this fact.

Fig. 2 (a-c) show the position estimation results for a Kalman filter using only the 3, 6, or 8 smallest residuals, respectively (i.e., the performance constraint is ignored). As the number of measurements increases, the expected information from those measurement increases, but so does the risk of inclusion of outliers. For the KF using only the 3 smallest residuals per time step, the position error is sometimes high, due to the condition of the observability matrix being low (i.e, insufficient diversity in the rows of H). The performance specification is not achieved. Increasing the number of included measurements shifts the position error histogram to the left, but outliers still result in some

high position errors that violate the specification. The bottom graph shows the results of the PSMR approach ($L = 1$ for direct comparison with the KF). In this approach, the optimization process ensures that the selected measurements result in sufficient diversity in the rows of H to satisfy the position error specification, while minimizing the risk of including outliers. The PSMR may ignore some of the smaller residuals, as they would add risk without supplying a sufficient amount of new information. In the PSMR approach the number of measurements used at each time varies. In this example with $m = 10$ and an outlier rate of 20%, the number of selected measurements varied in the range of six to eight, with the median number of selected measurements being 7. Either approach has some risk of incorporating outliers. For the KF approach using the eight smallest residuals, outliers affected 12% of the measurements that were used. For the PSMR approach, outliers only affected 1% of the measurements that were selected.

VI. CONCLUSIONS

This paper presents a novel approach to improve the robustness of a state estimation when the measurement data may contain outliers. Instead of trying to detect and remove outliers, as in [15], [16], this article tries to select valid measurements within an optimization framework. The approach herein chooses a subset of measurements in a moving-horizon sliding-window which have minimum risk while achieving a lower bounded information for state estimation. We proposed a general but tractable framework for real-time linear state estimation. We consider the solution for the case when accuracy specification for the last state is of interest and analyze the trade off between increasing specified accuracy and risk.

Our final formulation (Problem $P2$) is a constrained separately convex linear least squares optimization and can be solved using multi-convex programming. Simulation evaluation is included. Topics of interest for future work include faster solution algorithms without user defined parameters, extensions to nonlinear systems, and real-time implementations.

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.

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.

[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.

[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.

[6] J. Vaganay, J. J. Leonard, and J. G. Bellingham, "Outlier rejection for autonomous acoustic navigation," in *ICRA*, 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] 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.

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

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

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

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

[13] M. C. Graham, J. P. How, and D. E. Gustafson, "Robust incremental SLAM with consistency-checking," in *IROS*. IEEE, 2015, pp. 117–124.

[14] 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.

[15] L. Carlone, A. Censi, and F. Dellaert, "Selecting good measurements via l 1 relaxation: A convex approach for robust estimation over graphs," in *(IROS)*. IEEE, 2014, pp. 2667–2674.

[16] E. Aghapour and J. A. Farrell, "Performance specified state estimation with minimum risk," *Submitted to American Control Conference ACC*, 2018.

[17] K. R. Muske, J. B. Rawlings, and J. H. Lee, "Receding horizon recursive state estimation," in *ACC*. IEEE, 1993.

[18] D. Q. Mayne and H. Michalska, "Adaptive receding horizon control for constrained nonlinear systems," in *Decision and Control Conference*. IEEE, 1993, pp. 1286–1291.

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

[20] D. Simpson, "Introduction to Rousseeuw (1984) Least Median of Squares Regression," in *Breakthroughs in Statistics*. Springer, 1997, pp. 433–461.

[21] A. J. Stromberg, "Computing the exact least median of squares estimate and stability diagnostics in multiple linear regression," pp. 1289–1299, 1993.

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

[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-łojasiewicz inequality," *Mathematics of Operations Research*, vol. 35, no. 2, pp. 438–457, 2010.

[24] R. M. Eustice, H. Singh, and J. J. Leonard, "Exactly sparse delayed-state filters for view-based SLAM," vol. 22, no. 6, pp. 1100–1114, 2006.

[25] E. Aghapour, F. Rahman, and J. A. Farrell, "Risk-averse performance-specified state estimation," *Proc. of IEEE/ION PLANS, in press*, 2018.