Title
Outlier Accommodation in Sensor Rich Environments by Risk-Averse Performance-Specified State Estimation

Permalink
https://escholarship.org/uc/item/2rb1t446

Author
Aghapour, Elahe

Publication Date
2019

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA
RIVERSIDE

Outlier Accommodation in Sensor Rich Environments by
Risk-Averse Performance-Specified State Estimation

A Dissertation submitted in partial satisfaction
of the requirements for the degree of

Doctor of Philosophy

in

Electrical Engineering

by

Elahe Aghapour

June 2019

Dissertation Committee:

Prof. Jay A. Farrell, Chairperson
Prof. Amit K. Roy-Chowdhury
Prof. Kostas Karydis
The Dissertation of Elahe Aghapour is approved:

______________________________
Committee Chairperson

______________________________

University of California, Riverside
Acknowledgments

First and foremost, I would like to express my thanks to Prof. Jay A. Farrell for his utmost guidance, support, encouragement, and patience during my journey towards a Ph.D. He let me know how important it is to be rigorous as an engineer which is going to have a life-long influence on me.

I want to thank other members of my oral exam and defense committee, Prof. Amit K. Roy-Chowdhury, Prof. Anastasios Mourikis, Prof. Ertem Tuncel and Prof. Kostas Karydis. Without their help and advice, I cannot have been here. I want to thank all the professors and teaching assistants in Electrical and Computer Engineering Department who have taught me for their efforts and insights in teaching the graduate courses, from which I have learned a lot.

I want to thank all my friends especially Vahid Mardanlou and Taghi Farzad for being always supportive, Mostafa Khezri and Abtin Shahidi for bringing excitement to my daily routine, Mohammad Amin Baniasadi for all his poetical notes and Shaghayegh Gharghabi for her bright smile. I want to specially thank the support and encouragement from Zahra Barani, Fariborz Kargar, and Salar Rahili for their wonderful friendship during the past years and in future. During the past five years, they are always on my side and help me face every challenge.

I give my special thanks to my parents and my brother, for their endless and unconditional love, encouragement and support. I owe them too much. I dedicate this dissertation to them.
to the little prince, who tamed me at the age of three
ABSTRACT OF THE DISSERTATION

Outlier Accommodation in Sensor Rich Environments by
Risk-Averse Performance-Specified State Estimation

by

Elahe Aghapour

Doctor of Philosophy, Graduate Program in Electrical Engineering
University of California, Riverside, June 2019
Prof. Jay A. Farrell, Chairperson

Many applications require reliable, high precision state estimation while mitigating measurement outliers. This dissertation presents a novel state estimation approach to the challenge of preventing outlier measurements from affecting the accuracy and reliability of state estimation. Since outliers can degrade the performance of state estimation, outlier accommodation is critical. The most common method for outlier accommodation utilizes a Neyman-Pearson (NP) type threshold test in a (extended) Kalman filter (KF) to detect and remove residuals greater than a designer specified threshold. Such threshold based methods may use residuals arbitrarily close to the threshold, even when they are not needed to achieve an application’s performance specification. Outlier measurements that pass the residual test (i.e., missed detections) results in incorrect information being incorporated into the state and error covariance estimates. Once the state and covariance are incorrect, subsequent outlier decisions may be incorrect, possibly causing divergence.

The major contribution of this dissertation 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. Risk-averse performance-specified (RAPS) state estimation works within an optimization setting to choose a set of measurements that achieves a
performance specification with minimum risk of outlier inclusion. This dissertation derives and formulates the RAPS solution for outlier accommodation which applies to both linear and non-linear applications. The approach is also extended to moving horizon state estimation problem. Global Navigation Satellite Systems (GNSS) and inertial measurements for moving vehicle state estimation are used as an example to show the performance of the proposed approach.
# Contents

List of Figures x  
List of Tables xi  

## 1 Introduction  

## 2 Outlier-Free Trajectory and State Estimation  

2.1 System Model ........................................................................................................ 6  
2.1.1 Linear Model .................................................................................................. 6  
2.1.2 Nonlinear Model ............................................................................................. 7  
2.2 Outlier-Free Full Trajectory \( (L = K) \) MAP Estimation ........................................... 8  
2.2.1 Cholesky ........................................................................................................... 10  
2.2.2 SVD ............................................................................................................... 12  
2.2.3 QR .................................................................................................................. 13  
2.3 Outlier-Free Receding Horizon \( (L < k) \) Trajectory Estimation ......................... 14  
2.4 Outlier-Free Moving Horizon State Estimation ..................................................... 15  
2.5 Outlier-Free One-Time Epoch \( (L = 1) \) Estimation ................................................... 17  
2.6 Canonical Representation for Gaussian Distribution .......................................... 18  

## 3 RAPS State Estimation: Linear systems  

3.1 Problem Statement .................................................................................................. 19  
3.2 Problem Formulation .............................................................................................. 20  
3.3 Solution Methods ................................................................................................... 23  
3.3.1 Non-binary Solution ......................................................................................... 24  
3.3.2 Binary Solution ............................................................................................... 26  
3.3.3 Non-binary Solution Versus Binary Solution .................................................... 28  
3.4 Reducing Computational Load ............................................................................... 29  
3.4.1 Diagonal Elements Specification ..................................................................... 29  
3.4.2 Greedy Search ................................................................................................. 30  
3.5 Discussion ............................................................................................................... 31  
3.5.1 Specification for Reduced State ...................................................................... 31  
3.5.2 Choosing \( J_l \) ............................................................................................... 32  

## 4 RAPS State Estimation: Nonlinear systems  

4.1 Problem Statement .................................................................................................. 35  
4.2 Problem Formulation .............................................................................................. 35  
4.3 Solution Methods ................................................................................................... 38
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.3.1 Non-binary Solution</td>
<td>38</td>
</tr>
<tr>
<td>4.3.2 Binary Solution</td>
<td>41</td>
</tr>
<tr>
<td>4.4 Reducing Computational Load</td>
<td>43</td>
</tr>
<tr>
<td>4.4.1 Diagonal Elements Specification</td>
<td>43</td>
</tr>
<tr>
<td>4.4.2 Greedy Search</td>
<td>44</td>
</tr>
<tr>
<td>4.4.3 Measurement Selection Based on the Prior</td>
<td>45</td>
</tr>
<tr>
<td>5 Moving Horizon RAPS Trajectory Estimation: Linear systems</td>
<td>47</td>
</tr>
<tr>
<td>5.1 Problem Statement</td>
<td>47</td>
</tr>
<tr>
<td>5.2 Problem Formulation: Moving Horizon with Outliers</td>
<td>49</td>
</tr>
<tr>
<td>5.3 Solution Methods</td>
<td>51</td>
</tr>
<tr>
<td>5.4 Reducing Computational Load</td>
<td>53</td>
</tr>
<tr>
<td>5.5 Discussion</td>
<td>55</td>
</tr>
<tr>
<td>5.5.1 Discussion of the Selection Vector b</td>
<td>55</td>
</tr>
<tr>
<td>5.5.2 Most Recent State $x_k$ Accuracy Specification</td>
<td>56</td>
</tr>
<tr>
<td>5.5.3 Choosing $J_l$</td>
<td>57</td>
</tr>
<tr>
<td>5.5.1 Discussion of the Selection Vector b</td>
<td>55</td>
</tr>
<tr>
<td>5.5.2 Most Recent State $x_k$ Accuracy Specification</td>
<td>56</td>
</tr>
<tr>
<td>6 Moving Horizon RAPS Trajectory Estimation: Nonlinear systems</td>
<td>60</td>
</tr>
<tr>
<td>6.1 Problem Statement</td>
<td>60</td>
</tr>
<tr>
<td>6.2 Problem Formulation: Moving Horizon with Outliers</td>
<td>62</td>
</tr>
<tr>
<td>6.3 Solution Methods</td>
<td>64</td>
</tr>
<tr>
<td>6.4 Reducing Computational Load</td>
<td>66</td>
</tr>
<tr>
<td>7 Experimental Results</td>
<td>68</td>
</tr>
<tr>
<td>7.1 Background</td>
<td>68</td>
</tr>
<tr>
<td>7.1.1 GNSS Models</td>
<td>68</td>
</tr>
<tr>
<td>7.1.2 GNSS PVA Model</td>
<td>70</td>
</tr>
<tr>
<td>7.1.3 GNSS Aided INS Model</td>
<td>71</td>
</tr>
<tr>
<td>7.2 Hardware Setup and Data</td>
<td>72</td>
</tr>
<tr>
<td>7.2.1 Hardware Setup</td>
<td>72</td>
</tr>
<tr>
<td>7.2.2 Outlier Generation for GNSS Measurement</td>
<td>74</td>
</tr>
<tr>
<td>7.3 Binary RAPS Experimental Results: One-Time Epoch State Estimation</td>
<td>74</td>
</tr>
<tr>
<td>7.3.1 Algorithms</td>
<td>74</td>
</tr>
<tr>
<td>7.3.2 Performance Evaluation</td>
<td>75</td>
</tr>
<tr>
<td>7.3.3 Performance versus Outlier Mean</td>
<td>76</td>
</tr>
<tr>
<td>7.3.4 Experimental Results for Error, Risk, and GDOP</td>
<td>77</td>
</tr>
<tr>
<td>7.4 Non-binary RAPS Experimental Results: Moving Horizon State Estimation</td>
<td>79</td>
</tr>
<tr>
<td>7.4.1 Algorithms</td>
<td>79</td>
</tr>
<tr>
<td>7.4.2 Performance versus Outlier Mean</td>
<td>80</td>
</tr>
<tr>
<td>7.4.3 Experimental Results for Error, Risk, and GDOP: GNSS-PVA Approach</td>
<td>81</td>
</tr>
<tr>
<td>7.4.4 Experimental Results for Error, Risk, and GDOP: GNSS-INS Approach</td>
<td>83</td>
</tr>
<tr>
<td>8 Conclusion and Future Works</td>
<td>93</td>
</tr>
<tr>
<td>8.1 Conclusions</td>
<td>93</td>
</tr>
<tr>
<td>8.2 Publication list</td>
<td>94</td>
</tr>
<tr>
<td>8.3 Future Works</td>
<td>95</td>
</tr>
</tbody>
</table>

Bibliography
List of Figures

3.1 Find all feasible measurement subsets. .................................................. 27

7.1 Test trajectory. The bottom corner photos identify a variety of real-world environmental factors which adversely affect the performance of a GPS receiver, e.g. trees and tall buildings. ................................................................. 73

7.2 Mean horizontal position error and the percentage of selected measurements versus mean outlier magnitude $\mu \in [0, 20]$. The red curves display the result for the binary RAPS algorithm from Section 3.3.2. The yellow, green, blue and black curves show the results for the NP-(E)KF approach with $\gamma = 2, 3, 4, \text{ and } 5$, respectively. ................................................................. 86

7.3 Error, risk, and information diversity (i.e., GDOP) for the (nonlinear) GNSS-INS approach. The yellow, green, blue and black curves display the results for NP-EKF approach with $\gamma = 2, 3, 4, \text{ and } 5$, respectively. The red curve shows the performance of the RAPS approach. .................................................. 87

7.4 Performance comparison using GNSS data with the linear PVA model. The yellow, green, blue and black curves display the results for NP-KF approach $\gamma = 2, 3, 4, \text{ and } 5$, respectively. The red curve shows the RAPS performance. .................................................. 88

7.5 Mean horizontal position error and normalized mean $\|b\|_1$ versus mean outlier magnitude $\mu \in [0, 20]m$ using linear PVA model. The red curves display the results for the MH-RAPS algorithm from Section 5.5.2. The blue, green, yellow, and black curves show the results for the MH-NP approach with $\gamma = 2, 3, 4, \text{ and } 5$, respectively. .................................................. 89

7.6 Estimation error, risk, and information diversity (GDOP) versus time. The blue and green curves display the results for MH-NP approach with $\gamma = 2$ and 3, respectively. The red curve shows the performance of the MH-RAPS approach. .................................................. 90

7.7 Error, risk and information diversity (GDOP) versus time for $\mu = 2m$ and $L = 5$. The red curves display the result for the MH-RAPS algorithm from Section 6.3. The blue, green, yellow and black curves show the results for the MH-NP approach with $\gamma = 2, 3, 4, \text{ and } 5$, respectively. .................................................. 91

7.8 Error, risk and information diversity (GDOP) versus $L$ with $\mu = 2$. The red, blue, and green curves display the results for the MH-RAPS algorithm from Section 6.3 $L = 2, 4, \text{ and } 8$, respectively. .................................................. 92
List of Tables

2.1 Computational Complexity of Cholesky Solution .................. 11
2.2 Computational Complexity of SVD ................................. 13
3.1 Parameters for Choosing Different Specified Accuracy .......... 32
5.1 Covariance lower bound calculation for different $\alpha$ and $\beta$ ....... 59
7.1 GNSS-INS Horizontal Performance Statistics ...................... 85
7.2 GNSS-PVA Horizontal Performance Statistics ...................... 85
7.3 GNSS-PVA Horizontal Performance Statistics ...................... 89
7.4 GNSS-INS Horizontal Performance Statistics ...................... 91
7.5 MH-RAPS performance Statistics versus $L$ with $\mu = 2$ for GNSS-INS. . . . 92
Chapter 1

Introduction

An outlier can be defined as an observation that lies outside some overall pattern of
distribution and contains corrupted information about the variables to be estimated [30]. For
accurate and reliable estimation, it is important to avoid the effect of outliers. For many appli-
cations, state estimation is required while mitigating measurement outliers.

Given redundant measurements, outlier accommodation in most existing approaches
is accomplished in two stages: (a) residual generation (RG) [12, 26, 27] and (b) decision making
(DM) [8]. Such approaches evaluate measurement residuals against their expected values using
a fixed or an adaptive threshold test.

Choosing a threshold involves a trade-off between missed detections and false alarms
[42]. Moreover, there are two notable drawbacks with these methods: high computational costs
as the number of hypotheses increases and unobservability of outliers [13]. There are also
approaches such as least soft-thresholded squares [67] and median least squares [56, 59] that
avoid hypothesis testing. None of these approaches considers the question of whether additional
risky measurements are in fact needed to satisfy a performance specification.
Building on fundamental ideas [17, 25, 42, 44, 45, 69], GNSS Receiver Autonomous Integrity Monitoring (RAIM) techniques compute a parity vector from the measurement residual [10, 11, 60], assuming that there is enough measurement redundancy to discriminate the outlier source. The RAIM approach is well developed for multiple outlier detection [4, 11]. Extended RAIM (eRAIM) [31] incorporates an Inertial Measurement Unit (IMU) and Kalman filter based estimation into RAIM. RAIM methods are threshold based. They do not consider an accuracy specification in the measurement selection decision, so may incur unnecessary risk in sensor-rich situations. Outlier accommodation in GNSS aided inertial navigation applications has also been considered using the least-soft-squares approach in [52]. Data redundancy, quantified by the number of degrees-of-freedom (DOFs), is critical to successful outlier accommodation. Both RAIM and eRAIM are based on measurements from a single epoch, limiting data redundancy.

Redundancy can be enhanced both by adding additional sensors or by solving the estimation problem using all sensor data within a moving temporal window with time-horizon of length $L$. Such approaches have a long history and have been developed under various names: Moving Horizon Estimation (MHE) [2, 3, 29, 37, 41, 74]; Contemplative Real-time Estimation (CRT) [14, 15, 24, 51, 71, 72]; Simultaneous Localization and Mapping (SLAM) [18, 20–22, 34, 36, 57, 58, 63]; and others [32, 39]. In such approaches, 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 $L$ over which measurements are retained. To achieve a specified level of state estimation accuracy, the full set of measurements may not be required. In such cases, if the full set of measurements was used, then the state estimate would have been exposed to unnecessary outlier risk, while the computed error covariance would show that the estimator is over-performing relative to the specification.

When traditional outlier accommodation methods were developed for state estimation, the number of available sensors was limited. A main concern was maintaining observability of
the state vector. With rapidly decreasing sensor cost, some of today’s important applications are (or soon will be) sensor-rich – having many more measurements available than is required for observability. For example, several GNSS systems (e.g. GPS, GLONASS, Galileo) are available, with each individually supplying $6 - 10$ measurements per epoch, while only four (with suitable geometry) are required [38]. In the near future, the evolution of GNSS will bring the total number of GNSS satellites to over 100. Increasing the number of measurements used per epoch toward 100 would not significantly affect the achieved accuracy [65], but would significantly increase the probability of including outliers. Similarly, for feature-based navigation using camera data, only four suitably located features are required for observability [47], while tens or hundreds of features may be available per image. Such applications may also have a high number of measurements affected by outliers. We refer to such applications, where more measurements are available than are required either for observability or achieving a performance specification, as sensor-rich. In such application, the choice of the least risky subset of measurements that achieves the specification is an interesting question.

Motivated by the unobservability of outliers and applications with a high number of outliers, new outlier accommodation methods [13, 61, 62, 67] have been developed within an optimization framework. In [61, 62], latent variables are introduced to activate or deactivate measurements to achieve robust pose graph optimization. The authors of [13] search for a maximum subset of pose graph measurements that are consistent with the same system.

The RAPS approach is also optimization based, selecting a subset of measurements to satisfy a specified accuracy constraint while incurring minimum state estimation risk. The approach uses a sensor selection vector as in [13]. Our approach shares some similarity with sensor selection problems [16, 33, 54, 68], which select $k$ sensors from among $m$ sensor measurements to minimize a cost function (e.g., entropy or log volume of the confidence ellipsoid).
Our approach does not consider a constraint on the number of chosen measurements in subset selection. The existing approaches do not consider a performance specification as a constraint.

Considering the above challenges within outlier accommodation methods, changing the focus from explicitly characterizing each measurement as inlier/outlier to a new perspective developed within an optimization framework. Risk-averse performance-specified (RAPS) approach chooses a subset of measurements with minimum risk of outlier inclusion which can satisfy the specified accuracy. This dissertation also has the following other contributions to improve the accuracy, reliability and robustness of state estimation when measurements may contain outliers:

- RAPS is a novel state estimation approach to the challenge of preventing outlier measurements from affecting the accuracy, robustness, and reliability of state estimation. It is a general, but tractable solution for online state estimation.

- Proposed approach works within an optimization setting applicable to both linear and nonlinear applications. Binary and non-binary RAPS solutions are provided.

- Designed an extension RAPS solution to moving horizon (MH) state estimation for both linear and nonlinear applications which allows past outlier decisions to be reevaluated in the light of new data within the time window. MH-RAPS can improve the accuracy and robustness of state estimation in presence of outliers.

- Proposed approach is implemented and successfully solved the problem of estimating the state of a roving vehicle using experimental data. Both linear and nonlinear applications are presented. The linear application considered Global Navigation Satellite Systems (GNSS) data with a position, velocity, and acceleration model. The nonlinear application used the same GNSS data along with inertial measurement data that is combined through an aiding an Inertial Navigation System (INS).
The organization of this dissertation is as follows. Chapter 2 studies the problem of state estimation for outlier free measurements. Chapter 3 presents the RAPS state estimation optimization problem and solves the problem for linear systems. Chapter 4 extends the RAPS solutions for nonlinear applications. Chapter 5 and 6 presents the moving horizon RAPS state estimation solution for linear and nonlinear systems, respectively. Chapter 7 applies the RAPS solutions for vehicle state estimation using real data GNSS measurements (aiding an INS) for both linear and nonlinear applications. Chapter 8 concludes this dissertation and discusses potential future works.
Chapter 2

Outlier-Free Trajectory and State Estimation

Let $x_k \in \mathbb{R}^n$ represent the state vector at discrete-time $t_k = kT$ where the symbol $\mathbb{R}$ is used for the set of real numbers. This chapter addresses the state estimation problem for linear systems.

2.1 System Model

2.1.1 Linear Model

The discrete-time linear dynamical model is:

$$x_k = \Psi_{k-1}x_{k-1} + G_{k-1}u_{k-1} + \omega_{k-1}. \quad (2.1)$$

where $\Psi_k \in \mathbb{R}^{n\times n}$, 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 covariance $Q_k$.

The outlier free measurement vector $z_k$ at time $k$ is:

$$z_k = H_kx_k + \eta_k. \quad (2.2)$$
where $H \in \mathbb{R}^{n \times m}$ represents the measurement model and $\eta_k \sim \mathcal{N}(0, R_k)$ represents white Gaussian measurement noise. The covariance matrix $R_k$ is assumed to be invertible and diagonal\(^1\) which can be written as $R = \sum_{i=1}^{m} \sigma_i^2 e_i e_i^\top$ where $e_i \in \mathbb{R}^m$ is the $i$-th column of the identity matrix.

### 2.1.2 Nonlinear Model

The discrete-time dynamical model is:

$$x_k = f(x_{k-1}, u_{k-1}) + \omega_{k-1} \quad (2.3)$$

where $f : \mathbb{R}^n \times \mathbb{R}^\ell \rightarrow \mathbb{R}^n$. 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 covariance $Q_k$.

The outlier free measurement vector $z_k$ at time $k$ is:

$$z_k = h(x_k) + \eta_k \quad (2.4)$$

where $h : \mathbb{R}^n \rightarrow \mathbb{R}^m$ represents the measurement model and $\eta_k \sim \mathcal{N}(0, R_k)$ represents white Gaussian measurement noise. The covariance matrix $R_k$ is assumed to be invertible and diagonal which can be written as $R = \sum_{i=1}^{m} \sigma_i^2 e_i e_i^\top$ where $e_i \in \mathbb{R}^m$ is the $i$-th column of the identity matrix.

The linearization of (2.4) at $x_{k-1}^\bot$ using the first-order Taylor series yields the residual measurement model:

$$\delta z_k = H_k \delta x_k + \eta_k \quad (2.5)$$

where $\delta x_k = x_k - \hat{x}_k$, $z_k^\bot = h(x_k^\bot)$, $\delta z_k = z_k - z_k^\bot$ and $H_k = \nabla h(x)|_{x=x_k^\bot} \in \mathbb{R}^{m \times n}$.

\(^1\)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 $R^{-1} = \Sigma_R^\top \Sigma_R$, the measurement model for $z'$ is:

$$z' = H' x + \eta' \quad \text{where } H' = \Sigma_R H , \quad \eta' \sim \mathcal{N}(0, diag(\sigma_1^2, \ldots, \sigma_m^2))$$
2.2 Outlier-Free Full Trajectory \((L = K)\) MAP Estimation

At each time instant \(k\), trajectory Maximum A Posteriori (MAP) approach would compute the entire trajectory of states \(X_M = [x_0^T, \ldots, x_k^T]^T \in \mathbb{R}^{(k+1)n}\) that maximize the posterior pdf [64]:

\[
\hat{X}_M = \arg\max_X P(X, U_{0:k-1}, Z_{1:k}),
\]

where \(U_{0:k-1} = \{u_i, \forall i = 0, \ldots, k-1\}\) and \(Z_{1:k} = \{z_i, \forall i = 1, \ldots, k\}\). Applying Bayes’ theorem to (2.6) yields:

\[
\hat{X}_M = \arg\max_X \left( p(x_0) \prod_{i=0}^{k-1} p(x_{i+1}|x_i, u_i) \prod_{j=1}^{k} p(z_j|x_j) \right),
\]

where \(p(x_{i+1}|x_i; u_i)\) and \(p(z_j|x_j)\) are the distribution of state evolution based on eqn. (2.1) and the measurement of the state based on eqn. (2.2), respectively and the prior probability density function (PDF) for the initial state \(x_0 \sim \mathcal{N}(\hat{x}_0, P_0)\) is known. Given the Gaussian assumptions for the prior, process, and measurement noise, the negative log-likelihood of (2.7) yields:

\[
\hat{X}_M = \arg\min_X \left( \|x_0 - \hat{x}_0\|_{P_0} + \sum_{i=0}^{k-1} \|x_{i+1} - (\Psi_i x_i + G_i u_i)\|^2_{Q_i} + \sum_{j=1}^{k} \|z_j - H x_j\|^2_{R_j} \right),
\]

where the squared Mahalanobis distance for vector \(v\) is denoted by \(\|v\|_W = v^T W^{-1} v = \|\Sigma_w v\|_2\) with \(W^{-1} = \Sigma_W^{-1} \Sigma_w\). Incorporating matrix \(W = \text{diag}(P_0, Q_0, \ldots, Q_{k-1}, R_1, \ldots, R_k)\) with eqn. (2.8) yields:

\[
\hat{X}_M = \arg\min_X \left( (r - D_M X)^T W^{-1} (r - D_M X) \right) \quad (2.9)
\]

\[
= \arg\min_X \|r - D_M X\|_W^2 \quad (2.10)
\]
where matrix $D_M \in \mathbb{R}^{(k+1)n \times km \times (k+1)n}$ and vector $r \in \mathbb{R}^{(k+1)n \times km}$ are

$$
D_M = \begin{bmatrix}
I & 0 & 0 & \cdots & 0 & 0 \\
\Psi_0 & -I & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \Psi_{k-1} & -I \\
0 & H_1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & H_k
\end{bmatrix},
$$

$$
r = \begin{bmatrix}
\hat{x}_0 \\
-G_0 u_0 \\
\vdots \\
-G_{k-1} u_{k-1} \\
z_1 \\
\vdots \\
z_k
\end{bmatrix}.
$$

The optimization (2.10) can be written in matrix form:

$$
\hat{X}_M = \arg\min_X \| \Sigma_W (r - D_M X) \|_2 = \arg\min_X \| (\bar{r} - \bar{D}_M X) \|_2.
$$

where Jacobian matrix $\bar{D}_M = \Sigma_w D$ and residual vector $\bar{r} = \Sigma_w r$ are defined by:

$$
\bar{D}_M = \begin{bmatrix}
\Sigma P_0 & 0 & 0 & \cdots & 0 & 0 \\
\Psi_0 & -\Sigma Q_0 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \Psi_{k-1} & -\Sigma Q_{k-1} \\
0 & H_1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & H_k
\end{bmatrix},
\bar{r} = \begin{bmatrix}
\Sigma P_0 \hat{x}_0 \\
-G_0 u_0 \\
\vdots \\
-G_{k-1} u_{k-1} \\
\Sigma R_1 z_1 \\
\vdots \\
\Sigma R_k z_k
\end{bmatrix}
$$

where $\bar{\Psi}_k = \Sigma Q_k \Phi_k$, $\bar{G}_k = \Sigma Q_k G_k$ and $\bar{H}_k = \Sigma R_i H_k$. The matrices $\bar{D}_M^0 \in \mathbb{R}^{n \times (k+1)n}$, $\bar{D}_M^1 \in \mathbb{R}^{kn \times (k+1)n}$, and $\bar{D}_M^2 \in \mathbb{R}^{km \times (k+1)n}$ represent the sub-matrices in $\bar{D}_M$ that correspond with the subvectors $\bar{r}^0 \in \mathbb{R}^n$, $\bar{r}^1 \in \mathbb{R}^{kn}$, and $\bar{r}^2 \in \mathbb{R}^{km}$ of $\bar{r}$, as indicated by the horizontal lines. The structure of submatrices $\bar{D}_M^0$ and $\bar{D}_M^1$ in eqn. (2.13) shows that the Jacobian matrix has full column rank. Eqn. (2.12) is a standard linear least-squares problem. The solution $X^*$ must satisfy the
equation

\((\bar{D}_M^\top D_M)X = \bar{D}_M^\top r\). \hspace{1cm} (2.14)

The theoretical solution, assuming that \(\bar{D}_M\) is full rank is

\(\hat{X}_M = (\bar{D}_M^\top D_M)^{-1}\bar{D}_M^\top r\); \hspace{1cm} (2.15)

however, eqn. (2.15) is not typically the best numerical solution approach for at least two reasons. First, computation of \((\bar{D}_M^\top D_M)\) squares the condition number of \(\bar{D}_M\). The condition number \(C\) of matrix \(\bar{D}_M\) is the ratio of the largest to smallest singular value of \(\bar{D}_M\). If the condition number of a matrix is very large, then the matrix is said to be ill-conditioned. The condition number quantifies the amplification of numeric errors and is greater than one. Therefore, squaring it is not beneficial to the solution. Second, it involves at matrix inversion. Alternative solution methods are discussed in the following subsections. Hereafter, having \(\bar{D}_M\) to be \(\bar{D}_M \in \mathbb{R}^{p \times q}\) where \(p = (m+n)k+n\), \(q = (k+1)n\), and \(p \geq q\). Matrix \(\bar{D}_M\) is also assumed to have full column rank (i.e., \(\text{Rank}(\bar{D}_M) = q\)). Standard methods to calculate the numeric complexities are considered. However, many different modifications on these methods can be found with improved computational cost which is out of our scope.

In the following, the sparsity of matrix \(\bar{D}_M\) is not considered. Sparse matrix decomposition methods can yield further computational savings [18, 22, 34].

### 2.2.1 Cholesky

Eqn. (2.14) is equivalent to \((\bar{D}_M^\top D_M)X - \bar{D}_M^\top r = 0\). The Cholesky decomposition provides an upper triangular matrix \(R \in \mathbb{R}^{q \times q}\) that is the square-root of a symmetric matrix:
\[ \bar{D}_M^\top \bar{D}_M = R^\top R \] This is useful for solving the normal equation:

\[ \hat{X}_M = \arg\min_X \| \bar{D}_M^\top r - \bar{D}_M^\top \bar{D}_M X \|_2^2 \]
\[ = \arg\min_X \| \bar{D}_M^\top r - R^\top RX \|_2^2 \]
\[ = \arg\min_X \| \bar{D}_M^\top r - R^\top \omega \|_2^2 \] (2.16)

where \( \omega = RX \). By applying backward-substitution to solve the lower triangular system in (2.16) for \( \omega \), then the upper triangular system \( \omega = RX \) can be solved by forward-substitution. The Cholesky decomposition method provides the fastest solution with complexity as summarized in Table 2.1.

<table>
<thead>
<tr>
<th>Steps</th>
<th>Computation Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Matrix multiplication ( \bar{D}_M^\top \bar{D}_M )</td>
<td>( O(q^2p) )</td>
</tr>
<tr>
<td>Cholesky decomposition</td>
<td>( O(4/3q^3) )</td>
</tr>
<tr>
<td>Backward substitution</td>
<td>( O(q^2) )</td>
</tr>
<tr>
<td>Forward substitution</td>
<td>( O(q^2) )</td>
</tr>
</tbody>
</table>

While the Cholesky decomposition avoids computation of a matrix inverse, it still computes \( \bar{D}_M^\top \bar{D}_M \) which since \( p \geq q \) is one of the more computationally expense portions of the algorithm. It also squares the condition number which can cause numeric instability when dealing with ill-conditioned matrices.
2.2.2 SVD

When the condition number is large (i.e., the problem is ill-conditioned), the factorization algorithm matters. This section considers solution by the Singular Value Decomposition (SVD).

Represent the SVD of \( \bar{D}_M \in \mathbb{R}^{p \times q} \) as \( \bar{D}_M = U\Sigma V^\top \). The matrices \( U \in \mathbb{R}^{p \times p} \) and \( V \in \mathbb{R}^{q \times q} \) are unitary and mutually orthogonal. The diagonal matrix \( \Sigma \in \mathbb{R}^{p \times q} \) is defined \( \Sigma = \text{diag}(\sigma_1, \ldots, \sigma_\rho, 0, \ldots, 0) \), where the singular values of \( \bar{D}_M \) are ordered such that \( \sigma_i \geq \sigma_{i+1} \). The rank of \( \bar{D}_M \) is \( \rho \). The conditional number of \( \bar{D}_M \) is \( C = \frac{\sigma_{\text{max}}}{\sigma_{\text{min}}} = \frac{\sigma_1}{\sigma_\rho} \).

Using the SVD decomposition method, the solution components associated with small singular values can be disregarded. For example, for a user-defined \( \varepsilon > 0 \), define \( r \) such that \( \sigma_\rho \geq \varepsilon \) and \( \sigma_{\rho+1} < \varepsilon \), then for all \( \ell > \rho \), set \( \sigma_\ell = 0 \).

Given the rank \( \rho \), the matrices \( V \) and \( U \) can be partitioned as

\[
V = [V_1, V_2], \quad U = [U_1, U_2]
\]

such that \( V_1 \in \mathbb{R}^{q \times \rho}, V_2 \in \mathbb{R}^{q \times (q-\rho)} \), \( \Sigma_1 = \text{diag}(\sigma_1, \ldots, \sigma_\rho) \), and \( U_1 \in \mathbb{R}^{p \times \rho}, U_2 \in \mathbb{R}^{p \times (p-\rho)} \).

With these definitions,

\[
\bar{D}_M = U\Sigma V^\top = U_1\Sigma_1 V_1^\top.
\]

Substituting the reduced SVD \( \bar{D}_M = U_1\Sigma_1 V_1^\top \) into either eqn. (2.14) or (2.12) yields the solution

\[
\hat{X}_M = V_1\Sigma_1^+ U_1^\top r,
\]

with \( \Sigma_1^+ = \text{diag}(\sigma_1^{-1}, \ldots, \sigma_\rho^{-1}) \) The quantity \( \bar{D}_M^+ = V_1\Sigma_1^+ U_1^\top \) is the pseudo-inverse of \( \bar{D}_M \).

The complexity of using the SVD decomposition is summarized in Table 2.2.

The SVD avoids an explicit matrix inversion and allows singular values that are too small to be discarded, along with their affects; however, the disadvantage of the SVD is its increased computational complexity relative to Cholesky (i.e., \( p^2q > pq^2 > q^3 \)).
Table 2.2: Computational Complexity of SVD

<table>
<thead>
<tr>
<th>Steps</th>
<th>Computation Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVD decomposition</td>
<td>$O(p^2q + pq^2 + q^3)$</td>
</tr>
<tr>
<td>Matrix multiplication</td>
<td>$O(p^2(p + q) + p^2)$</td>
</tr>
</tbody>
</table>

2.2.3 QR

The standard recommended solution approach uses the QR factorization $\bar{D}_M = Q[R]$, where $Q \in \mathbb{R}^{p \times p}$ is an orthogonal matrix and $R \in \mathbb{R}^{q \times q}$ is an upper triangular matrix. Substituting the QR factorization of $\bar{D}_M$ into eqn. (2.12):

$$\hat{X}_M = \arg\min_X \|r - D_MX\|^2$$

$$= \arg\min_X \|r - Q[R]X\|^2$$

$$= \arg\min_X \|Q^\top r - [R]X\|^2$$

$$= \arg\min_X \|[d] - [R]X\|^2$$

$$= \arg\min_X \|d - RX\|^2$$

where the following fact and definition are used: (1) $\|x\| = \|Qx\| = \|Q^\top x\|$; (2) $Q^\top r = [d]$. The final equation is solvable by back-substitution. The computation complexity is:

<table>
<thead>
<tr>
<th>Steps</th>
<th>Computation Complexity</th>
</tr>
</thead>
<tbody>
<tr>
<td>QR factorization</td>
<td>$O(2pq^2)$</td>
</tr>
<tr>
<td>Matrix multiplication, RX</td>
<td>$O(p^2)$</td>
</tr>
<tr>
<td>Forward substitution</td>
<td>$O(q^2)$</td>
</tr>
</tbody>
</table>

The QR factorization approach avoids the calculation of $\bar{D}_M^\top \bar{D}_M$. Instead, it works directly with $\bar{SVD}$. Therefore, the condition number is not squared, but remains unchanged,
yielding a more numerically stable algorithm. However, the computational cost is more than Cholesky factorization.

In this complete trajectory approach, the size of the trajectory vector \( X_M \) and the required computational cost increases linearly with the number of measurements \([18, 21, 22]\). In SLAM applications, where loop closures enhance MAP estimation accuracy, it is often a practice to maintain the entire trajectory. While direct solution approaches would have computational costs that increase as \( O((kn + n)^3) \), algorithms that \([18]\) capitalize on the sparsity of the problem achieve computational cost that is \( O((k + 1)n^3) \). In other applications, without loop closures, such as receding horizon estimation (RHE) and outlier accommodation, it may not be necessary to keep the entire trajectory and instead the problem is formulated over a sliding window to yield bounded computational cost, suitable for real time applications.

### 2.3 Outlier-Free Receding Horizon \((L < k)\) Trajectory Estimation

In the receding horizon approach, at time-step \( k \), the window will slide one epoch upon arrival of each new measurement vector \( z_k \). When the time-window slides, one old state vector \( \hat{x}_{k-L-1} \) will be marginalized out of \( X_{k-1,L} \) where the effect of old data \( z_{0:k-L} \) and \( u_{0:k-L-1} \) on state \( x_{k-L} \) are summarized in \( \Theta_{k-L}(x_{k-L}) \). Then, eqn. (2.8l) can be formulated for the trajectory \( X = [x_{k-L}^T, \cdots, x_k^T]^T \in \mathbb{R}^{(L+1)n} \) as \([2, 3, 29, 37, 41, 74]\):

\[
\hat{X}_{k,L} = \arg\min_X \left( \Theta_{k-L}(x_{k-L}) + \sum_{i=k-L}^{k-1} \omega_i^T Q_i^T \omega_i + \sum_{j=k-L+1}^{k} \eta_j^T R_j^T \eta_j \right)
\]

subject to: \( x_k = \Psi_{k-1} x_{k-1} + G_{k-1} u_{k-1} + \Gamma_{k-1} \omega_{k-1} \)

\[
z_k = H_k x_k + \eta_k
\]

\( x \in \mathcal{X}, \omega \in \mathcal{W}, \eta \in \mathcal{V} \)

where \( \mathcal{X}, \mathcal{W}, \) and \( \mathcal{V} \) represent the closed set constraints on variables \( x, \omega, \) and \( \eta \). The sets \( \mathcal{X}, \mathcal{W}, \) and \( \mathcal{V} \) correspond to physical characteristics of the system (e.g. in a chemical process
where states represent concentrations, states must be non-negative). Although formulating a probabilistic model for these variables is difficult, the designer will usually have knowledge about the range of these variables which can be included as constants.

The main focus in receding horizon approach is stability and accuracy. A main issue is, how to choose the function $\Theta_{k-L}$ that summarizes the effect of all previous data to the trajectory MAP solution. For the unconstrained problem, the choice [48]:

$$\Theta_{k-L}(x_{k-L}) = \|x_{k-L} - \hat{x}_{k-L}\|^2_{P_{k-L}}, \quad (2.19)$$

where $\hat{x}_{k-L}$ and $P_{k-L}$ are computed with $\hat{X}_{k-1,L}$, results in a stable solution where the RH solution is equivalent to that studied in Section 2.4. When the application has constraints, many forms of $\Theta_{k-L}(x_{k-L})$ may exist [19, 49, 70] of which eqn. (2.19) is one [48].

### 2.4 Outlier-Free Moving Horizon State Estimation

At time-step $k$, in a probabilistic framework, the process of sliding of the time window by one epoch upon the arrival of each new measurement vector is implemented by marginalizing out one old state vector $\hat{x}_{k-L-1}$ of the previous window to yield the Gaussian prior probability function (PDF) $\mathcal{N}(\hat{x}_{k-L}, P_{k-L})$ computed in $\hat{X}_{k-1,L}$ for the new window [22]. The reminder of this article is concerned with estimation of $X_{k,L}$. To simplify notation, the subscript will be dropped (i.e. $X$ will be used to denote $X_{k,L}$).

The Maximum A Posteriori (MAP) approach to estimate the trajectory $X = [x_{k-L}^T, \cdots, x_k^T]^T \in \mathbb{R}^{(L+1)n}$ by using Bayes’ theorem as:

$$\hat{X} = \arg\max_X P(X, U_{k-L:k-1}, Z_{k-L+1:k}) = \arg\max_X \left( p(x_{k-L}) \prod_{i=0}^{L-1} p(x_{k-i} | x_{k-1-i}, u_{k-L+i}) \prod_{j=0}^{L-1} p(z_{k-j} | x_{k-j}) \right). \quad (2.20)$$
Given the Gaussian assumptions for the prior, process, and measurement noise, the negative log-likelihood of (2.20) yields:

\[
\hat{X} = \underset{X}{\arg\min} \left( \| x_{k-L} - \hat{x}_{k-L} \|_{P_{k-L}}^2 + \sum_{i=k-L}^{k-1} \| x_{i+1} - (\Psi_i x_i + G_i u_i) \|_{Q_i}^2 + \sum_{j=k-L+1}^{k} \| z_j - H x_j \|_{R_j}^2 \right). 
\]

(2.21)

By defining \( W = \text{diag}(P_{k-L}, Q_{k-L}, \cdots, Q_{k-1}, R_{k-L+1}, \cdots, R_k) \), the optimization (2.21) can be transformed in to matrix form:

\[
\hat{X} = \underset{X}{\arg\min} \| r - DX \|_W^2 = \underset{X}{\arg\min} \| \Sigma W (r - DX) \|_2^2 = \underset{X}{\arg\min} \| \bar{r} - DX \|_2^2, 
\]

(2.22)

where \( W^{-1} = \Sigma_W \Sigma_W \) and the Jacobian matrix \( D \) and residual vector \( r \) are defined by:

\[
\bar{D} = \begin{bmatrix}
\Sigma_{P_{k-L}} & 0 & 0 & \cdots & 0 & 0 \\
\Psi_{k-L} & -\Sigma_{Q_{k-L}} & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & \Psi_{k-1} & -\Sigma_{Q_{k-1}} \\
0 & H_{k-L+1} & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & \cdots & 0 & H_k
\end{bmatrix},
\]

\[
\bar{\bar{D}}_0 = \begin{bmatrix}
\Sigma_{P_{k-L}} & \hat{x}_{k-L} \\
\Psi_{k-L} & -\hat{G}_{k-L} u_{k-L}
\end{bmatrix},
\]

\[
\bar{\bar{D}}_1 = \begin{bmatrix}
\Psi_{k-1} & -\hat{G}_{k-1} u_{k-1}
\end{bmatrix},
\]

\[
\bar{\bar{D}}_2 = \begin{bmatrix}
\Sigma_{R_{k-L+1}} z_{k-L+1} \\
\Sigma_{R_{k+1}} z_k
\end{bmatrix},
\]

\[
\bar{r} = \begin{bmatrix}
r_0 \\
r_1 \\
r_2
\end{bmatrix},
\]

(2.23)

where \( \Psi_i = \Sigma_Q \Psi_i \) and \( \hat{G}_i = \Sigma_Q G_i \) for \( i \in [k-L, k-1] \) and \( \tilde{H}_j = \Sigma_R H_j \) for \( j \in [k-L+1, k] \).

The sub-matrices \( \bar{\bar{D}}_0 \in \mathbb{R}^{n \times (L+1)n} \), \( \bar{\bar{D}}_1 \in \mathbb{R}^{Lx \times (L+1)n} \), and \( \bar{\bar{D}}_2 \in \mathbb{R}^{Lm \times (L+1)n} \) corresponds with the sub-vectors \( \bar{r}_0 \in \mathbb{R}^n \), \( \bar{r}_1 \in \mathbb{R}^{L_n} \), and \( \bar{r}_2 \in \mathbb{R}^{L_m} \), as indicated by the horizontal lines. Eqn. (2.22) is a linear least-squares problem which can be solved using QR decomposition, singular value decomposition (SVD), etc. The error covariance matrix for state trajectory \( X \) can be computed by the inverse of the tri-block diagonal information matrix \( J = \bar{\bar{D}}^TD \) with lower computational
cost than matrix inversion [35]. Articles [18, 22, 34] present efficient computational algorithms building on the sparsity structure of $D$.

Herein, each element of the vector $z_k$ may be affected by outliers at some time instants. Hence, additional reasoning is required. If $L = 1$, this is the Kalman filter (KF) with outliers. KF provides the optimal estimate conditioned on all past outlier decisions being correct. Using $L > 1$, (a) increases the length of the residual vector allowing more reliable outlier decisions; (b) allows various outlier hypotheses to be considered and compared; and (c) conditions $\hat{X}$ on outlier decisions made prior to $k - L$ which are unchangeable but allows past outlier decisions for $j = [k - L, \cdots, k]$ to be reevaluated in light of new data.

### 2.5 Outlier-Free One-Time Epoch ($L = 1$) Estimation

One time epoch state estimation is a special case of moving horizon state estimation assuming that $L = 1$. Given a prior probability density function (PDF) $p(x_{k-1}) \sim \mathcal{N}(x_{k-1}^+, P_{k-1}^+)$, the problem of state estimation using all measurements at time $k$ from the Maximum A Posteriori (MAP) perspective is:

$$x_k^+ = \arg \max_{x} p(x, x_{k-1}, u_{k-1}, z_k)$$

$$= \arg \max_{x} p(x_{k-1})p(x|x_{k-1}, u_{k-1})p(z_k|x).$$  \hspace{1cm} (2.24)

Given the Gaussian assumptions for the prior, process, and measurement noise, the negative log-likelihood of the distribution (2.24) yields the nonlinear least-squares optimization:

$$x_k^+ = \arg \min_{x} \left(\|x_{k-1} - x_k^+\|_{P_{k-1}^+}^2 + \|H_k x_{k-1} + G_k u_{k-1} - x\|_{Q_k}^2\right)$$

$$+ \|H_k x - z_k\|_{R_k}^2.$$

Employing math manipulations to solve eqn. (2.26) yields to Kalman filter (KF) equations (see section 4.3 in [53]). Using eqn. (2.1), the state estimate $\hat{x}_k$ and its covariance $P_k$ are
propagated as:

\[ \hat{x}^-_k = \Psi^-_{k-1} \hat{x}^+_k + G_{k-1} u_{k-1} \]  
\[ P^-_k = \Psi^-_{k-1} P^+_1 \Psi^-_{k-1} + Q. \]  
(2.27)

With no outliers, the state estimate and its covariance corresponding to the optimal gain \( K_k \) are updated by:

\[ \hat{x}^+_k = \hat{x}^-_k + K_k (z_k - H_k \hat{x}^-_k) \]  
\[ P^+_k = (I - K_k H_k) P^-_k. \]  
(2.28)

Note that KF provides the optimal estimate conditioned on measurement \( z \) being outlier free. In following chapters, it is assumed that each element of the vector \( z_k \) in addition to measurement noise, may be affected by outliers at some time instants in which case eqn. (2.28) is no longer optimal.

### 2.6 Canonical Representation for Gaussian Distribution

While the Gaussian distribution is typically discussed in terms of its mean and covariance, \( \mathcal{N}(\mu, P) \), herein it will be more convenient to utilize the information form of representation \( \mathcal{N}^{-1}(\zeta, J) \). The information vector and information matrix are \( \zeta = P^{-1} \mu \) and \( J = P^{-1} \).

The information matrix is propagated through time and updated for measurements as [64]:

\[ J^-_k = (\Psi^-_{k-1} (J^+_k)^{-1} \Psi^-_{k-1} + Q_{k-1})^{-1} \]  
\[ J^+_k = H^-_k R^{-1}_k H_k + J^-_k. \]  
(2.29)
Chapter 3

RAPS State Estimation: Linear systems

3.1 Problem Statement

While the best estimation accuracy will be achieved by detecting and removing all outliers and then using all remaining measurements in the state estimation process, no detection approach will work perfectly. Missed detections defeat the reliability of all subsequent outlier decisions. Recently, new optimization based outlier accommodation approaches have been suggested [13, 61, 62, 67]. This article considers an alternative optimization based approach motivated by the ideas in [13]. Instead of focusing on outlier rejection, the goal herein will be to choose a subset of measurements to estimate the state vector $x_k$, where the subset is selected to have minimum risk of containing an outlier while achieving a specified accuracy. RAPS is an optimization based solution for state estimation problem which is looking for a subset of measurements with minimum risk of incurring outliers, satisfying a specified accuracy. In RAPS, a cost function corresponds to risk is minimized while the optimization is subject to an
accuracy constraint which is defined by information matrix. Analysis and discussion of RAPS optimization formulation and solutions is provided in following sections.

### 3.2 Problem Formulation

Let \( x_k \in \mathbb{R}^n \) represent the state vector at discrete-time \( t_k = kT \). Based on Section 2.5, given a prior probability density function (PDF) \( p(x_{k-1}) \sim \mathcal{N}(x_{k-1}^+, P_{k-1}^-) \) and Gaussian assumptions for the prior, process, and measurement noise, the problem of state estimation using all measurements at time \( k \) is:

\[
\begin{align*}
    x_k^+ &= \arg\max_x p(x, x_{k-1}, u_{k-1}, z_k) \\
    &= \arg\max_x p(x_{k-1}) p(x|x_{k-1}, u_{k-1}) p(z_k|x) \\
    &= \arg\min_x \| x_{k-1} - x_k^+ \|^2_{P_{k-1}^-} + \| (\Psi_{k-1} x_{k-1} + G_{k-1} u_{k-1}) - x \|^2_{Q_k} + \| H_k x - z_k \|^2_{R_k}
\end{align*}
\]  

(3.3)

where the squared Mahalanobis norm of vector \( r \) with covariance \( \Sigma \) is denoted by \( ||r||_\Sigma^2 = r^T \Sigma^{-1} r \) which is equivalent to \( ||r||_\Sigma^2 = \| \Sigma^{-\frac{1}{2}} r \|^2_2 \). The approach herein is concerned with sensor outliers; therefore, the model and prior are always trusted (i.e., outlier-free). With this assumption, the first and second terms of the optimization can be propagated as:

\[
\begin{align*}
    x_{k-1}^- &= \Psi_{k-1} x_k^+ + G_{k-1} u_{k-1} \\
    P_k^- &= \Psi_{k-1} P_{k-1}^+ \Psi_{k-1}^T + Q_{k-1}.
\end{align*}
\]  

(3.4)

Using eqn. (3.4) in optimization (3.3) yields:

\[
\begin{align*}
    x_k^+ &= \arg\min_x \left[ \| x - x_k^- \|^2_{P_k^-} + \| H_k x - z_k \|^2_{R_k} \right]
\end{align*}
\]  

(3.5)

which is a linear least squares problem and can be efficiently solved. This yields the standard Kalman filter in information form. So far, this approach does not address the existence of measurement outliers. Outliers included in the measurement update would cause both the
state estimate and the computed error covariance to become inaccurate, causing the prior to be incorrect for all future measurements.

Instead of detecting outliers, a main new idea introduced in [13] was to find the largest subset of the measurements that can be consistently produced by the assumed measurement model. This was implemented by introducing a binary vector \( b = [b_1, b_2, \cdots, b_m]^\top \) with \( b_i \in \{0, 1\} \) to disable or enable the \( i \)th measurement.

Incorporating this binary decision variable \( b \) in eqn. (3.1) yields:

\[
\begin{align*}
\dot{x}_k^+ &= \argmax_x p(x, x_{k-1}, u_{k-1}, z_k; b) \\
&= \argmax_x p(x_{k-1})p(x|x_{k-1}, u_{k-1})p(z|x_k; b).
\end{align*}
\]

This is the optimal estimate assuming that the measurements indicated by the pattern of ones in \( b \) are valid and that \( u \) and the system model are correct. For this problem formulation, eqn. (3.5) becomes:

\[
\begin{align*}
\dot{x}_k^+ \cdot b^* &= \argmin_{x, b} \left[ \|x - x_k^-\|^2_{P_k^-} + \|\Phi(b)^\top (H_k x - z_k)\|^2_{R_k} \right] \\
&\quad \text{subject to: } b_i \in \{0, 1\} \text{ for } i = 1, \ldots, m,
\end{align*}
\]

where \( \Phi(b) = \text{diag}(b) \) and the variables \( P_k^- \) and \( z_k \) are known. The cost function

\[
C(x, b) = \|x - x_k^-\|^2_{P_k^-} + \|\Phi(b)^\top (H_k x - z_k)\|^2_{R_k}.
\]

Incorporating the binary selection vector \( b \) into standard MAP state estimation yields to optimization (3.6) which quantifies the risk associated with each \( x \) and \( b \). 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 \( b_i = 1 \) (i.e. selecting an outlier by binary selection vector \( b_1 \) results incorrect state \( x_k^+ \) which is associated with bigger \( C(x_k^+, b_1) \) in compare with the case when non of the outliers are chosen by binary selection vector \( b \)). It is a summation of two positive terms. Its minimum value of zero is achieved for \( b = 0_m \in \mathbb{B}^{m \times 1} \) (i.e., discarding all the
measurements) where the symbol $\mathbb{R}$ is used for the set of Boolean numbers. In [13], the authors proceeded by finding the largest subset of the given measurements for which there exists a configuration which explains those measurements. Note that when a performance specification is stated, this largest subset may achieve greater performance than required at the expense of increased risk.

Alternatively, herein, we consider an approach that finds the set of measurements that satisfies the performance specification with minimum risk as quantified by $C(x, b)$. The performance specification can be defined by using either covariance or information matrices. Using Tchebycheff’s inequality for the covariance based specification $\text{cov}(x) = \sigma^2$ where $x \in \mathbb{R}$ implies:

$$\Pr(|x| > \gamma \sigma) \leq \frac{1}{\gamma^2}$$

where $\Pr$ represents the probability. Therefore, choosing $\sigma = \alpha \sqrt{\beta}$ in covariance based specification yields to performance specification $|x| \leq \alpha$ with probability $(1 - \beta)$. Similarly, the covariance based specification for $x \in \mathbb{R}^n$ is defined by:

$$P \leq P_l$$

where $P = \text{cov}(x)$ and $P_l = \text{diag}(\alpha_1 \sqrt{\beta}, \ldots, \alpha_n \sqrt{\beta})$. For matrices $A$, the notation $A \succeq 0$ means $A$ is a positive semidefinite matrix.

Due to the linear relation between selected measurements and information matrix in eqn. (2.29), the accuracy constraint is defined by using information matrix where $J = P^{-1}$ implies to have:

$$J \succeq J_l$$
where $J_l = P_l^{-1}$. Then, the performance constrained optimization problem is

$$x_k^*, b^* = \arg\min_{x, b} C(x, b)$$

subject to: $J_b^+ \geq J_l$

$$b_i \in \{0, 1\} \quad \text{for} \quad i = 1, \ldots, m,$$

where $J_b^+$ is the posterior information matrix corresponding to using the measurements that have $b_i = 1$ and $J_l \in \mathbb{R}^{n \times n}$ is an user-defined minimum accuracy specification. A discussion about choosing $J_l$ is provided in Section 3.5.2.

The Fisher information matrix $J_b^+$ in the optimization (3.7) for diagonal covariance $R$ is

$$J_b^+ = H_k^\top \Phi(b) \left( \sum_{i=1}^{m} \frac{1}{\sigma_i^2} e_i e_i^\top \right) \Phi(b) H_k + J_k^-$$

$$= H_k^\top \left( \sum_{i=1}^{m} \frac{b_i^2}{\sigma_i^2} e_i e_i^\top \right) H_k + J_k^-$$

$$= \sum_{i=1}^{m} \frac{b_i^2}{\sigma_i^2} h_i^\top h_i + J_k^-$$

where $h_i$ is the $i^{th}$ row of $H_k$. Therefore, the optimization problem (3.7) using the binary definition of $b$ (i.e. $b_i^2 = b_i$), becomes:

$$LP1 : \min_{x, b} \left[ \|x - x_k\|_P^2 + \sqrt{\sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} e_i^\top (H_k x - z_k)} \right]$$

subject to: $\left( \sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} h_i^\top h_i + J_k^- \right) \geq J_l$

$$b_i \in \{0, 1\} \quad \text{for} \quad i = 1, \ldots, m.$$

The solution of $LP1$ uses a subset of measurements selected by $b$ that minimizes the risk while satisfying the user defined accuracy specification $J_l$.

### 3.3 Solution Methods

The solution for RAPS optimization problem stated in (3.9) will be provided in this section.
3.3.1 Non-binary Solution

The optimization \( LP_1 \) is convex in \( x \) and Boolean in \( b \). The Boolean constraint on \( b \) can be relaxed to use the convex constraint \( b \in [0,1]^m \) as:

\[
\begin{align*}
\text{LP}_{1r}: \min_{x,b} & \left[ \|x - x_k^{-}\|_P^2 + \left\| \sum_{i=1}^m \frac{b_i}{\sigma_i} e_i^\top (H_k x - z_k) \right\|^2 \right] \\
\text{subject to:} & \left( \sum_{i=1}^m \frac{b_i}{\sigma_i} h_i^\top h_i + J - k \right) \geq J_l \\
& b_i \in [0,1] \text{ for } i = 1, \ldots, m.
\end{align*}
\]  

Problem \( LP_{1r} \) is convex separately in \( x \) and \( b \) with a convex feasible set for \( b \). Note that the feasible set for the relaxed optimization in \( LP_{1r} \) contains the feasible set for \( LP_1 \). Therefore, the objective value of \( LP_{1r} \) is a lower bound on the objective value of \( LP_1 \) [33].

Since \( LP_{1r} \) is a convex problem for either variable \( x \) or \( b \), separately, but not jointly, multi-convex programming can be employed to solve \( LP_{1r} \) by alternatively updating \( b \) and \( x \) using the modified algorithm described below based on results in [55]. Proximal terms are added in the cost function (see below) where the convergence of this algorithm is proved in [50]. This problem is solved iteratively. The iteration number will be indicated by a right superscript \( \ell \), starting at zero.

1. Selecting the measurements: In this step, the optimal \( b^{\ell+1} \) is found for fixed \( x_k^\ell \). Consequently, the first term \( \|x - x_k^{-}\|_P^2 \) in problem \( (LP_{1r}) \) will be dropped because it is independent of \( b \). Based on [7], the proximal term \( \lambda \|b - b^\ell\|^2 \) is required to penalize the change of \( b^{\ell+1} \) in comparison with \( b^{\ell} \), the optimization in standard form will be:

\[
\begin{align*}
\text{LP}_{2b}: \min_b & \left[ \left\| \sum_{i=1}^m \frac{b_i}{\sigma_i} e_i^\top (H_k x_k - z_k) \right\|^2 + \lambda \|b - b^{\ell}\|^2 \right] \\
\text{subject to:} & J_l - \left( \sum_{i=1}^m \frac{b_i}{\sigma_i} h_i^\top h_i + J_k^- \right) \leq 0 \\
& b_i \in [0,1] \text{ for } i = 1, \ldots, m.
\end{align*}
\]
where $\lambda > 0$ is a user-defined proximal parameter. This is a least squares problem constrained by a linear matrix inequality (LMI). Since, the constraint on the information matrix is a positive definite cone, optimization $LP_{2b}$ is a standard semidefinite programming (SDP) problem and can be 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:

$$LP_{2x} : \min_x \left[ \|x - x^\ell\|^2 + \| \sum_{i=1}^m b_i e_i^\top (H_k x - z_k) \|^2 + \beta \|x - x^\ell\|^2 \right]$$

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

$$\nabla_x C(x) = 2J_k^-(x - x_k^-) + 2H_k^\top \Phi(b)^2 R_k^{-1} (H_k x - z_k) + 2\beta (x - x^\ell).$$

(3.11)

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

$$x_k^+ = D^{-1} (J_k^- x_k^- + H_k^\top \Phi(b)^2 R_k^{-1} z_k + \beta x^\ell)$$

where $D = (J_k^- + H_k^\top \Phi(b)^2 R_k^{-1} H_k + \beta I)$.

Even for the linear problem $LP1$, the solution is iterative between $LP_{2b}$ and $LP_{2x}$. To initiate this iterative solution, the two steps of updating $b$ and $x$ are interchangeable. If an initial value of state $x^0$ is accurate (i.e. $J_k^-$ 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. $J_k^-$ is small), the optimization can start by finding $x^1$ in Step 2 assuming all the measurements are selected (i.e. $\Phi(b^0) = I$).
Note that the method above provides both an optimal state estimate and a non-binary selection vector \( b \in [0,1]^m \). The quantity \( \Phi(b)^{-2}R = diag \left( (\sigma_i/b_i)^2 \right) \) could be interpreted as a retuning of the measurement noise covariance \( R \) based on the assessment of risk indicated by \( b \).

Alternatively, at the conclusion of the algorithm, a threshold \( \tau \) could be specified such that

\[
b_i = \begin{cases} 
1 & b(i) \geq \tau \\
0 & \text{otherwise.}
\end{cases}
\]

When \( b_i = 0 \), the \( i^{th} \) row in \( H \) and \( R \) will be ignored. After thresholding \( b \), the state update step would need to be performed once more, without the proximal term, which is then a standard Kalman filter measurement update using the selected measurements. Using this threshold approach imposes two drawbacks: \( a \) The computed accuracy using binary \( b_i \) is different than the solution of the optimization (3.10) which may not satisfy the lower bound information constraint; and, \( b \) The designer needs to select an appropriate threshold.

### 3.3.2 Binary Solution

The optimization (3.9) solution can be divided into three steps:

1. **Definitions and Initialization.** Let \( \mathcal{M} \) be the set of those binary vectors \( b \) such that the performance constraint is satisfied. Each element of \( \mathcal{M} \) defines a subset of measurements that are feasible for \( LP1 \).

Let \( n_z(b) = (m - \|b\|_1) \) be the number of zero entries in \( b \). Let

\[
\mathcal{M}_s \subset \mathcal{M} \quad \text{and} \quad \mathcal{M}_s = \{ b | n_z(b) = s \text{ and } J_b^+ \geq J_l \}
\]

contain all feasible measurement subsets that have \( n_z(b) = s \) for \( 0 \leq s \leq m \).

Initialize the algorithm with \( s = 0 \) which has

\[
\mathcal{M}_0 = \{ [1,1,\cdots,1]^\top \} \quad \text{and} \quad \mathcal{M} = \mathcal{M}_0.
\]
Denote the cost function $LP1$ as
\[
C(x, b) = \|x - x_k^-\|_F^2 + \|\Phi(b)(H_k x - z_k)\|_R^2.
\]

Initialize the optimal selector vector $b^* = \{\}$ and minimum cost function value $c^* = C(\hat{x}^+, b^*) = +\infty$ where notation $\{\}$ represents an empty set.

2. Find all Feasible Measurement Subsets.

   (a) For each $b \in \mathcal{M}_s$, keep the zero elements of $b$ unchanged and deactivate exactly one of the active elements of $b$. Each $b \in \mathcal{M}_s$ will produce the $(m - s)$ permutations denoted as $b^\ell$ for $\ell = 1, \ldots, (m - s)$. Each of the resulting $b^\ell$ vectors will have:
   \[
   n_z(b^\ell) = (n_z(b) + 1) = (s + 1).
   \]

   (b) For $\ell = 1, \ldots, m - s$, check whether $b^\ell$ satisfies the performance constraint. If it does, then add $b^\ell$ to $\mathcal{M}$ and to $\mathcal{M}_{s+1}$.

   (c) If $\mathcal{M}_{s+1}$ is empty go to Step (3); otherwise set $s = s + 1$ and go to Step (2-(a)).

Fig. 3.1 illustrates how the algorithm works with $m = 4$. In the following, the candidate $b$ vectors will be denoted as $b_{\ell \sigma}$. The root has $b_{01} = [1111]^T$ which is determined to satisfy the constraint. Therefore $b_{01} \in \mathcal{M}_0 = \mathcal{M}$. In the second level, all four permutations of the active measurement are produced by deactivating one element of $b$. The performance
constraint is checked for all four. Nodes \( b_{12} = [1011]^\top \) and \( b_{13} = [1101]^\top \) satisfy the constraint. Therefore, for the example in Fig. 3.1,

\[
\mathcal{M}_1 = \{b_{12}, b_{13}\} \text{ and } \mathcal{M} = \mathcal{M} \cup \mathcal{M}_1.
\]

3. Risk Minimization. For each \( b^\ell \in \mathcal{M} \), solve the least square optimization:

\[
\hat{x}^\ell = \arg\min_x C(x, b^\ell).
\]  

(3.12)

Define \( c^\ell = C(\hat{x}^\ell, b^\ell) \). If \( c^\ell < c^* \), then set \( c^* = c^\ell \), \( \hat{x}^+ = \hat{x}^\ell \), and \( b^* = b^\ell \). Since we consider all \( b \)'s for which \( J_b^+ \geq J_\ell \) to find \( x \) with minimum cost, the final \( c^* \) is the global minimum value for the cost function and the final \( b^* \) and \( x^+ \) have the global minimum value of optimization \( P1 \) defined in (3.9).

3.3.3 Non-binary Solution Versus Binary Solution

In non-binary solution, interior point methods can be employed to solve the SDP problem in Selecting the measurements step. These methods typically require a few tens of iterations. Each iteration can be carried out with a complexity of \( O(m^3) \) operations. However, solving the RAPS optimization problem using a non-binary vector \( b \) has a few drawbacks:

a) The user-defined proximal parameters \( \lambda \) and \( \beta \) are challenging to select and affect the rate of convergence. b) The final solution converges to a local minimum, even when the function \( h(x) \) is convex.

In binary solution, the number of sensor combinations \( b \) to be checked is proportional to \( 2^s \) where \( s < m \). Solving the RAPS optimization problem using a binary vector \( b \) does find the global optimum \( c^* \) and does not involve any user-selected parameters. However, the combination of the high number of combinations \( b \) and the nonlinear optimization for each imposes a high computational cost and its computational cost becomes intractable as \( m \) increases. There-
fore, the non-binary solution is a tractable alternative when the number of measurements is large. Alternative binary solutions with lower computational cost are discussed in Section 3.4.

3.4 Reducing Computational Load

In the interest of reducing computational load while still finding a low risk combination of measurements that satisfies the specification, a few special topics are of interest.

3.4.1 Diagonal Elements Specification

Often the diagonal elements of the information matrix are of primary interest. In this case, the information constraint can be manipulated as follows:

\[
\begin{align*}
\text{diag} \left( \sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} h_i^\top h_i + J_k^- \right) & \geq J_d \\
\sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} \text{diag}(h_i^\top h_i) + \text{diag}(J_k^-) & \geq J_d \\
\sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} \begin{bmatrix} h_{i1}^2 \\
\vdots \\
h_{im}^2 \end{bmatrix} + \text{diag}(J_k^-) & \geq J_d
\end{align*}
\]  

(3.13)

where \( J_d \) is the diagonal of the user-defined information matrix that encodes a minimum accuracy specification (See Section 3.5.2.). Then, the optimization problem (3.9) becomes:

\[
\begin{align*}
\text{LPD} : \min_{x, b} & \quad \left\| x - \tilde{x}_k \right\|_{P_k}^2 + \left\| \sum_{i=1}^{m} \frac{b_i}{\sigma_i} e_i e_i^\top (H_k x - z_k) \right\|_{P_k}^2 \\
\text{subject to:} & \quad \sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} \begin{bmatrix} h_{i1}^2 \\
\vdots \\
h_{im}^2 \end{bmatrix} + \text{diag}(J_k^-) \geq J_d
\end{align*}
\]

(3.14)

\[ b_i \in \{ 0, 1 \} \text{ for } i = 1, \ldots, m. \]

The binary solution in Section 3.3.2 can be employed to solve problem LPD. The binary solution starts with selection vector \( b \) when all the measurements are chosen (i.e. \( \| b \|_1 = \)
and then expands it to all feasible $b^\ell$ where the constraint in eqn. (3.13) is used to check the feasibility of each $b^\ell$ in Step 2. There are still on the order of $2^s$ values of $b$ to consider.

### 3.4.2 Greedy Search

Use of the diagonal accuracy constraint (3.13) allows the opportunity to employ the greedy search to find a feasible combination of sensors with low risk; although, the minimal risk measurement selection $b^*$ corresponds to global minimum risk $c^*$ may be missed. The greedy approach first finds a low cardinality, feasible combination of sensors $b_0$. Then it searches around $b_0$ to reduce risk.

The algorithm has three steps:

1. **Initialization.** Start with $\mathcal{M} = \{\}$, $b^* = \{\}$ and $c^* = C(\hat{x}^+, b^*) = +\infty$ where notation $\{\}$ represents an empty set.

2. **Find A Feasible $b_0$ with Greedy search.** The purpose of this step is to find a selection vector $b_0$ which can satisfy the information constraint (3.14) using a small number of measurements. An example greedy algorithm is given in Algorithm 1, where \( \setminus \) and \( \oplus \) represent the set difference and exclusive or (XOR) operators, respectively, and $h_{ij}$ represents the element of $H$ in the $i^{th}$ row and $j^{th}$ column.

3. **Permutation Generation.** Let $\ell = \|b_0\|_1$ which is the cardinality of $b_0$. Generate all $b$ with $\|b^\ell\|_1 = \{\ell, \ell + 1\}$. Then, if $b$ satisfies the accuracy constraint (3.13), $\mathcal{M} = \mathcal{M} \cup b$.

4. **Risk Minimization.** This step is the same as Step 3 in binary solution in Section 3.3.2.

Step 2 completes in less than or equal to $m$ steps. Steps 3 and 4 consider on the order of $O(\max\left(\binom{m}{\ell}, \binom{m}{\ell+1}\right))$ different $b$ vectors. This is approximately $O(m^j)$ for $j = \min(\ell, m - \ell)$, which compares well with the $O(2^s)$ approach in Section 3.3.2 ($s$ being a fraction of $m$).
Algorithm 1: Greedy Search for $b_0$

I Definitions.
Let $\ell$ be the number of selected measurements. The set $S$ contains the indices of deactivated bits in $b_0$ (i.e., $i \in S$ means the $i^{th}$ element of $b_0$ is 0).

II Initialization.
Initialize $\ell = 0$, $S = \{1, \cdots, m\}$, $b_0 = 0 \in \mathbb{R}^m$, and vector $J_p = diag(J_k^-) - J_d$.

III Choose the next measurement.
Let $[J_p]_j$ be the minimum element in $J_p$. Activate a non-active bit of $b_0$ to maximize the added information to $[J_p]_j$:

$$i = \arg\max_{i \in S} \left( \frac{h_{ij}}{\sigma_i} \right)^2.$$  \hspace{1cm} (3.15)

IV Update.
$$J_p = J_p + \frac{1}{\sigma_i} diag(h_i^\top h_i), S^{\ell+1} = S^\ell \setminus i, b_0 = b_0 \oplus e_i, \text{ and } \ell = \ell + 1.$$  
If $J_p < 0_n$, go to Step 1c. Otherwise stop.

3.5 Discussion

3.5.1 Specification for Reduced State

Sometimes constraining a submatrix of the posterior information matrix is of interest (i.e., certain states). In this case, consider $V^\top J_k^+ V \in \mathbb{R}^{s \times s}$ with $s \leq n$, where the matrix $V \in \mathbb{R}^{n \times s}$ has $V_{ij} \in \{0, 1\} \forall i, j$ and $VV^\top \in \mathbb{R}^{n \times n}$ is a diagonal matrix with binary values along the diagonal. The matrix of $V^\top V$ is equal to the identity matrix in $\mathbb{R}^{s \times s}$.

The performance constraint in (3.9) is then modified to:

$$V^\top J_k^+ V \geq J_i$$  \hspace{1cm} (3.16)
where \( J_l \in \mathbb{R}^{s \times s} \) is an user-defined lower bound for a subset of the posterior information matrix. All previously mentioned algorithms can be employed by checking the modified performance constraint in (3.16).

3.5.2 Choosing \( J_l \)

Let \( x \in \mathbb{R} \) is any random variable with Gaussian distribution \( x \sim \mathcal{N}(\mu, \sigma) \), then:

\[
Pr(|x - \mu| \leq \gamma \sigma) = erf\left(\frac{\gamma}{\sqrt{2}}\right)
\]

(3.17)

where \( erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \) is error function. Therefore, using eqn. (5.21), given parameters \( \alpha \) and \( \beta \), the performance specification \( |x - \mu| \leq \alpha \) is achieved with probability \( \beta \) if \( \sigma \leq \varepsilon \) where

\[
\varepsilon = \alpha \left( \sqrt{2} erf^{-1}(\beta) \right)^{-1}
\]

where \( erf^{-1} \) represents the inverse function of error function. Table 5.1 represents example values of \( \varepsilon \) in applications.

Similarly, for \( x \in \mathbb{R}^n \), we have:

\[
P \leq P_l
\]

where \( P = \text{cov}(x) \) and \( P_l = \text{diag}(\varepsilon_1^2, \cdots, \varepsilon_n^2) \). Then, the inverse relation between covariance and information matrices (i.e. \( J = P^{-1} \)) implies:

\[
J \geq J_l
\]

| \( \alpha \) (m) | 1 | 1 | 1 | 0.5 | 0.5 | 0.5 |
| \( \beta \) (%) | 99.7 | 95.4 | 68.2 | 99.7 | 95.4 | 68.2 |
| \( \varepsilon \) (m) | 0.336 | 0.505 | 0.995 | 0.168 | 0.252 | 0.497 |

Table 3.1: Parameters for Choosing Different Specified Accuracy
where $J_l \geq P_l^{-1}$.

The following example clarifies the simplicity of choosing the performance bound $J_l$. In highway vehicle applications horizontal position accuracy is of primary importance. In particular, the SAE specification [6] requires 1 m horizontal position accuracy at 95%. If the state vector in this application was $x = [p^T, v^T, a^T]^T \in \mathbb{R}^9$ where $p$, $v$ and $a \in \mathbb{R}^3$ represent the rover position, velocity and acceleration vectors. For this example the components of the position vector are north, east, and down. Then by defining

$$
V^T = \begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix},
$$

the constraint $V^T J_l V \geq \begin{bmatrix} \alpha_1 & 0 \\ 0 & \alpha_2 \end{bmatrix}$ where $\alpha_1, \alpha_2 > 0$ places an explicit constraint on the accuracy of the horizontal position. Based on Table 5.1, if $\alpha_1 = \alpha_2 = \frac{1}{(0.505)^2} \approx 3.9$ and the system is accurately modeled, then the SAE specification would be met. However, Table 5.1 represents the corresponding accuracy values for different parameters when states are uncorrelated. Due to correlated information for correlated states, $\alpha_1 = \alpha_2 = 3.9$ are required to be chosen bigger than 3.9.

Increasing $\alpha$ tightens the specification, which may require a larger number of measurements to be used, which increases the risk of outlier inclusion. 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 than the Fisher information matrix, then no feasible solution exists.

It is important to note that the states that are not included explicitly in the constraint may still be implicitly constrained. For example, position accuracy cannot be achieved without velocity and acceleration also being estimated. Also, in GNSS applications, due to all three
components of the line-of-sight vector being nonzero, explicitly constraining the horizontal position error will cause the vertical position error to also be observable.
Chapter 4

RAPS State Estimation: Nonlinear systems

4.1 Problem Statement

This chapter addresses the state estimation problem for nonlinear systems when, in addition to measurement noise, the measurement $z_k$ may also be affected by outliers. Risk-averse performance-specified (RAPS) state estimation works within an optimization setting to choose a set of measurements that achieves a performance specification with minimum risk of outlier inclusion. This Chapter derives and formulates the RAPS solution for nonlinear applications. The system model is presented in Section 2.1.2.

4.2 Problem Formulation

Let $x_k \in \mathbb{R}^n$ represent the state vector at discrete-time $t_k = kT$. Given a prior probability density function (PDF) $p(x_{k-1}) \sim \mathcal{N}(x_{k-1}^+, P_{k-1}^+)$, the problem of state estimation using all
measurements at time \( k \) from the Maximum A Posteriori (MAP) perspective is:

\[
x_k^+ = \arg\max_x p(x, x_{k-1}, u_{k-1}, z_k)
\]

\[
x_k^+ = \arg\max_x p(x_{k-1}) p(x|x_{k-1}, u_{k-1}) p(z_k|x).
\]  

(4.1)  

(4.2)

This problem will be studied when the measurement \( z_k \) may be affected by outliers.

Given the Gaussian assumptions for the prior, process, and measurement noise, the negative log-likelihood of the distribution (4.1) yields the nonlinear least-squares optimization:

\[
x_k^+ = \arg\min_x \left( \|x - x_{k-1}\|^2_{P_{k-1}} + \|f(x_{k-1}, u_{k-1}) - x\|^2_Q + \|h(x) - z_k\|^2_R \right).
\]  

(4.3)

The approach herein is concerned with sensor outliers; therefore, the model and prior are always trusted (i.e., outlier-free). With this assumption, the first and second terms of the optimization can be propagated as:

\[
x_{k-1}^- = f(x_{k-1}^-, u_{k-1})
\]  

(4.4)

By employing the first-order Taylor series approximation to (2.3), the model for propagation of the state error through time is:

\[
\delta x_k = \Psi_{k-1} \delta x_{k-1} + \omega_{k-1}
\]  

(4.5)

where \( \delta x_k = x_k - \hat{x}_k \), and \( \Psi_{k-1} = \nabla f(x_{k-1}, u_{k-1})|_{x=\hat{x}_{k-1}} \). The error covariance time propagation corresponding to (4.5) is:

\[
P_k^- = \Psi_{k-1} P_{k-1}^+ \Psi_{k-1}^\top + Q_{k-1}.
\]  

(4.6)

Using eqns. (4.4) and (4.6) in optimization (4.3) yields:

\[
x_k^+ = \arg\min_x \left[ \|x - x_{k-1}\|^2_{P_k^-} + \|h(x) - z_k\|^2_R \right].
\]  

(4.7)

which is a nonlinear least squares problem and can be efficiently solved. This yields the standard extended Kalman filter in information form. So far, this approach does not address the existence
of measurement outliers which would cause both the state estimate and the computed error covariance to become inaccurate, causing the prior to be incorrect for all future measurements.

Similar to section 3.2, incorporating this binary decision variable $b = [b_1, b_2, \ldots, b_m]^\top$ with $b_i \in \{0, 1\}$ in eqn. (4.1) yields:

$$x_k^+ = \arg\max_x p(x|x_{k-1}, u_{k-1}, z_k; b)$$

$$= \arg\max_x p(x_{k-1})p(x|x_{k-1}, u_{k-1})p(z|x_k; b).$$

This is the optimal estimate assuming that the measurements indicated by the pattern of ones in $b$ are valid. For this problem formulation, eqn. (4.8) becomes:

$$x_k^+, b^* = \arg\min_{x,b} \left[ \|x-x_k^-\|^2_{P_k^-} + \|\Phi(b)^\top (h(x) - z_k)\|^2_{R_k} \right]$$

subject to: $b_i \in \{0, 1\}$ for $i = 1, \ldots, m$,

where $\Phi(b) = \text{diag}(b)$ and the variables $P_k^-$ and $z_k$ are known. The cost function

$$C_{NL}(x, b) = \|x-x_k^-\|^2_{P_k^-} + \|\Phi(b)^\top (h(x) - z_k)\|^2_{R_k}$$

Similar to Section 3.2, we consider an approach that finds the set of measurements that satisfies the performance specification with minimum risk as quantified by $C_{NL}(x, b)$. Performance specification is defined by using Fisher information matrix. The Fisher information matrix $J_b^+$ for diagonal covariance $R$ is

$$J_b^+ = H_k^\top \Phi(b)^\top \left( \sum_{i=1}^m \frac{1}{\sigma_i^2} e_i e_i^\top \right) \Phi(b) H_k + J_k^-$$

$$= \sum_{i=1}^m \frac{b_i^2}{\sigma_i^2} h_i^\top h_i + J_k^-$$

where $h_i$ is the $i^{th}$ row of $H_k$ and $H_k = \nabla h(x)|_{x=x_k} \in \mathbb{R}^{m \times n}$. A discussion about choosing $J_l$ is provided in Section 3.5.2. Therefore, the RAPS optimization problem using the binary definition
of $b$ (i.e., $b_i^2 = b_i$), becomes:

$$NP1 : \min_{x, b} \left[ \|x - x_k^-\|_P^- + \| \sum_{i=1}^m \frac{b_i}{\sigma_i} e_i^\top (h(x) - z_k^-) \|_2 \right]$$

subject to:

$$\left( \sum_{i=1}^m \frac{b_i}{\sigma_i^2} h_i^\top h_i + J_k^- \right) \geq J_l$$

(4.11)

$$b_i \in \{0, 1\} \text{ for } i = 1, \ldots, m.$$  

The solution of $NP1$ uses a subset of measurements selected by $b$ that minimizes the risk while satisfying the user defined accuracy specification $J_l$. A discussion about choosing $J_l$ is provided in Section 3.5.2.

### 4.3 Solution Methods

This section provides solutions for the RAPS optimization problem stated in eqn. (4.11).

#### 4.3.1 Non-binary Solution

If the measurement function $h$ is convex, the optimization $NP1$ is convex in $x$ and Boolean in $b$.

The Boolean constraint on $b$ can be relaxed to use the convex constraint $b \in [0, 1]^m$ as:

$$NP1_r : \min_{x, b} \left[ \|x - x_k^-\|_P^- + \| \sum_{i=1}^m \frac{b_i}{\sigma_i} e_i^\top (h(x) - z_k^-) \|_2 \right]$$

subject to:

$$\left( \sum_{i=1}^m \frac{b_i}{\sigma_i^2} h_i^\top h_i + J_k^- \right) \geq J_l$$

(4.12)

$$b_i \in [0, 1] \text{ for } i = 1, \ldots, m.$$  

Problem $NP1_r$ is a convex problem for either variable $x$ or $b$, separately, but not jointly, multi-convex programming can be employed to solve $NP1_r$ by alternatively updating $b$ and $x$ [55]. Here, the solution described in Section 3.3.1 is modified for nonlinear systems. This problem
is solved iteratively. The iteration number will be indicated by a right superscript $\ell$, starting at zero.

1. **Selecting the measurements**: In this step, the optimal $b^{\ell+1}$ is found for fixed $x_k^\ell$. Consequently, the first term $\|x - x_k^\ell\|^2_{P_k^{-1}}$ in problem (NP1,) will be dropped because it is independent of $b$. Based on [7], the proximal term $\lambda \|b - b^\ell\|^2$ is added, the optimization in standard form will be:

$$NP_2 b : \min_b \| \sum_{i=1}^m b_i \sigma_i e_i^\top (h(x) - z_k) \|^2 + \lambda \|b - b^\ell\|^2$$

subject to: $J_{l} - (\sum_{i=1}^m b_i h_i^\top h_i + J_k^\ell) \leq 0$

$$b_i \in [0, 1] \text{ for } i = 1, \ldots, m.$$ 

where $\lambda > 0$ is a user-defined proximal parameter, $H_k = \nabla h(x)|_{x=x_k^\ell} \in \mathbb{R}^{m \times n}$, and $h_i$ is the $i$-th row of $H_k$. This is a least squares problem constrained by a linear matrix inequality (LMI) which is a standard semidefinite programming (SDP) problem and can be 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:

$$NP_2 x : \min_x \left[ \|x - x_k^\ell\|^2_{P_k^{-1}} + \|\Phi(b)^\top (h(x) - z_k)\|^2_{H_k} + \beta \|x - x^\ell\|^2 \right]$$

where $\beta > 0$ is a user-defined proximal parameter. This is a nonlinear least squares optimization problem which can be solved iteratively. Using counter $i$, starting from $x_k^0 = x^\ell$, the linearized problem at each iteration is:

$$NP_2 x_i : \min_{\delta x} \left[ \|\delta x - r_i^x\|^2_{P_k^{-1}} + \sum_{i=1}^m \frac{b_i}{\sigma_i} e_i^\top (H_i^\ell \delta x - r_i^x) \|^2 + \beta \|\delta x\|^2 \right]$$
where \( r'_k = x_k^i - x_k^i \), \( r'_z = z_k - h(x_k^i) \), \( H'_k = \nabla h(x)|_{x=x_k^i} \in \mathbb{R}^{m \times n} \), and \( x_{k+1}^i = x_k^i + \delta x_k^i \). The gradient of the cost function \( C(x) \) is:

\[
\nabla_x C(x) = 2J_r(x - r'_x) \\
+ 2H_k \Phi(b)^2 R_k^{-1}(H_k \delta x - r'_z) + 2\beta(\delta x).
\]

The optimal value for \( \delta x_k^i \) can be computed as the roots of the gradient:

\[
\delta x_k^i = D^{-1}(J_r - H_k \Phi(b)^2 R_k^{-1} x) \\
\text{where } D = (J_r + H_k \Phi(b)^2 R_k^{-1} H_k + \beta I).
\]

To initiate this iterative solution, the two steps of updating \( b \) and \( x \) are interchangeable. If an initial value of state \( x^0 \) is accurate (i.e. \( J_r \) 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. \( J_r \) is small), the optimization can start by finding \( x^1 \) in Step 2 assuming all the measurements are selected (i.e. \( \Phi(b^0) = I \)).

The method above provides both an optimal state estimate and a non-binary selection vector \( b \in [0,1]^m \). The quantity \( \Phi(b)^{-2}R = \text{diag} \left( (\sigma_i/b_i)^2 \right) \) could be interpreted as a retuning of the measurement noise covariance \( R \) based on the assessment of risk indicated by \( b \). Alternatively, a threshold \( \tau \) could be specified to have \( b \in \{0,1\}^Lm \) (see Section 3.3.1).

**Remark 1** If the measurement function \( h \) is non-convex, the nonlinear optimization \( NP_1 \), can be solved iteratively. For the iterative solution, using counter \( i \), starting from \( x_k^0 = x_k^i \), the linearized problem at each iteration is

\[
NP_{1i} : \min_{\delta x, b} \left[ \|\delta x - r'_x\|_{P_k}^2 + \|\sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} e_i e_i^\top (H_k \delta x - r'_z)\|_2^2 \right] \\
\text{subject to: } \left( \sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} h_i^\top h_i + J_r \right) \geq J_l \\
\text{where } b_i \in [0,1] \text{ for } i = 1, \ldots, m.
\]

The optimization \( NP_{1i} \) is convex separately in \( x \) and \( b \) with a convex feasible
set for \(b\). For problem \(NP1_i\), a two-step optimization solution can be employed. The overall solution may require multiple linearization iterations of step 1 and 2, until the convergence condition satisfied.

4.3.2 Binary Solution

In this Section, the binary solution, presented in Section 3.3.2, will be extended for nonlinear systems. The binary solution has three steps:

1. **Definitions and Initialization.** Let \(\mathcal{M}\) be the set of those binary vectors \(b\) and their corresponding state \(x\) such that the performance constraint is satisfied. Each element of \(\mathcal{M}\) defines a subset of measurements that are feasible for \(NP1\).

Let \(n_z(b) = (m - \|b\|_1)\) be the number of zero entries in \(b\). Let

\[
\mathcal{M}_s = \{b | n_z(b) = s \text{ and } J^+_b \geq J_l\}
\]

contain all feasible measurement subsets that have \(n_z(b) = s\) for \(0 \leq s \leq m\).

Denote the cost function \(NP1\) as

\[
C_{NL}(x, b) = \|x - x_k^-\|_{\mathcal{R}_s}^2 + \|\Phi(b)(h(x) - z_k)\|_{\mathcal{R}_k}^2.
\]

Initialize the algorithm with \(s = 0\) which has

\[
\mathcal{M}_0 = \{[1, 1, \cdots, 1]^\top\} \text{ and } \mathcal{M} = \{[1, 1, \cdots, 1]^\top; x\}
\]

where \(x = \arg\min_x C_{NL}(x, [1, 1, \cdots, 1]^\top)\).

Initialize the optimal selector vector \(b^* = \{\}\) and minimum cost function value \(c^* = C_{NL}(x^+, b^*) = +\infty\) where notation \(\{\}\) represents an empty set.
2. Find all Feasible Measurement Subsets.

(a) For each \( b \in \mathcal{M}_s \), keep the zero elements of \( b \) unchanged and deactivate exactly one of the active elements of \( b \). Each \( b \in \mathcal{M}_s \) will produce the \((m-s)\) permutations denoted as \( b^\ell \) for \( \ell = 1, \ldots, (m-s) \). Each of the resulting \( b^\ell \) vectors will have:

\[
n_z(b^\ell) = (n_z(b) + 1) = (s + 1).
\]

(b) For \( \ell = 1, \cdots, m-s \), check whether \( b^\ell \) satisfies the performance constraint. Due to the nonlinearity \( h(x) \), the measurement matrix \( H_k = \nabla h(x) \big|_{x = \hat{x}_k^\ell} \), and hence the feasibility, depend on the state estimate; therefore, for each \( b^\ell \), solve the nonlinear least squares optimization:

\[
\hat{x}_k^\ell = \arg\min_x C_{NL}(x, b^\ell).
\]

which can be iteratively solved in few iterations. Using counter \( i \), starting from \( x^0 = x_k^- \), the optimal value for \( \delta x_k^i \) can be computed by:

\[
\delta x_k^i = D^{-1} \left( J_k^- r_k^i + H_k^\top \Phi(b)^2 R_k^{-1} r_k^i \right)
\]

(4.16)

where \( D = (J_k^- + H_k^\top \Phi(b)^2 R_k^{-1} H_k) \) and \( x_k^{i+1} = x_k^i + \delta x_k^i \).

If \( b^\ell \) satisfies the performance constraint, then add \( b^\ell \) to \( \mathcal{M}_s+1 \) and \( \{b^\ell, \hat{x}_k^\ell\} \) to \( \mathcal{M} \).

(c) If \( \mathcal{M}_{s+1} \) is empty go to Step (3); otherwise set \( s = s + 1 \) and go to Step (2-(a)).

3. Risk Minimization. For each \( b^\ell; \hat{x}_k^\ell \in \mathcal{M} \), define \( c^\ell = C_{NL}(\hat{x}_k^\ell, b^\ell) \). If \( c^\ell < c^* \), then set \( c^* = c^\ell, x^+ = \hat{x}_k^\ell \), and \( b^* = b^\ell \).

Since we consider all \( b^\ell \)'s for which \( J_b^+ \geq J_b \) to find \( x \) with minimum cost, the final \( c^* \) is the global minimum value for the cost function and the final \( b^* \) and \( x^+ \) have the global minimum value of optimization \( NP1 \) defined in (4.11).
Remark 2 Non-binary Solution Versus Binary Solution

The computational cost for nonlinear solution depends on the convergence condition for \( x \). The computational cost comparisons for linear systems is provided in Section 3.3.3.

4.4 Reducing Computational Load

A few special topics to reduce the computational load while still finding a low risk combination of measurements that satisfies the specification, are of interest.

4.4.1 Diagonal Elements Specification

Often the diagonal elements of the information matrix are of primary interest. In this case, the information constraint can be manipulated as follows:

\[
\text{diag}\left( \sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} h_i^\top h_i + J_k^- \right) \geq J_d
\]

\[
\sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} \text{diag}(h_i^\top h_i) + \text{diag}(J_k^-) \geq J_d
\]

\[
\sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} \begin{bmatrix} h_{i1}^2 \\ \vdots \\ h_{in}^2 \end{bmatrix} + \text{diag}(J_k^-) \geq J_d
\]

where \( J_d \) is the diagonal of the user-defined information matrix that encodes a minimum accuracy specification. Then, the optimization problem (4.11) becomes:

\[
\text{NPD} : \min_b \left[ \| x - \bar{x}_k \|_{P_k}^2 + \| \sum_{i=1}^{m} \frac{b_i}{\sigma_i} e_i e_i^\top (h(x) - z_k) \|_{H_k}^2 \right]
\]

subject to:

\[
\sum_{i=1}^{m} \frac{b_i}{\sigma_i^2} \begin{bmatrix} h_{i1}^2 \\ \vdots \\ h_{in}^2 \end{bmatrix} + \text{diag}(J_k^-) \geq J_d
\]

\[
b_i \in \{0, 1\} \text{ for } i = 1, \ldots, m.
\]
The binary solution in Section 4.3.2 can be employed to solve problem NPD. The binary solution starts with selection vector $b$ when all the measurements are chosen (i.e. $\|b\|_1 = m$) and then expands it to all feasible $b^\ell$ where the constraint in eqn. (4.17) is used to check the feasibility of each $b^\ell$ in Step 2. There are still on the order of $2^s$ values of $b$ to consider.

### 4.4.2 Greedy Search

Use of the diagonal accuracy constraint (4.17) allows the opportunity to employ the greedy search to find a feasible combination of sensors with low risk; although, the optimal measurement selection $b^*$ with global minimum risk $c^*$ may be missed. The greedy approach first finds a low cardinality, feasible combination of sensors $b_0$. Then it searches around $b_0$ to reduce risk.

The algorithm has three steps:

1. **Initialization.** Initialize $\mathcal{M} = \{\}$, the optimal selector vector $b^\diamond = \{\}$ and minimum cost function value $c^\diamond = C(x^+, b^\diamond) = +\infty$.

2. **Find A Feasible $b_0$ with Greedy search.** The purpose of this step is to find a selection vector $b_0$ which can satisfy the information constraint (4.18) using a small number of measurements. An example greedy algorithm is given in Section 3.4.2 where $H = \nabla h(x)|_{x=x^-}$.

3. **Permutation Generation.** Let $n_1 = \|b_0\|_1$ which is the cardinality of $b_0$. Generate all $b$ with $\|b^\ell\|_1 = \{n_1, n_1 + 1\}$. Then, check whether generated $b^\ell$ satisfies the accuracy constraint (4.17). Due to the nonlinearity $h(x)$, the measurement matrix $H_k = \nabla h(x)|_{x=\hat{x}^\ell}$, and hence the feasibility, depend on the state estimate; therefore, an iterative solution can be employed to solve the nonlinear least squares optimization (4.15). Using counter $i$, starting from $x^0 = x^-\hat{x}$, the corresponding state $\hat{x}^\ell$ to selection vector $b^\ell$ can be found by using eqn. (4.16). If $b^\ell$ satisfies the accuracy constraint (4.17), $\mathcal{M} = \mathcal{M} \cup \{b^\ell, \hat{x}^\ell\}$. 

44
4. Risk Minimization. This step is the same as Step 3 in binary solution in Section 4.3.2 to find optimal solutions $b^*$ and $x_k^\dagger$.

Step 2 completes in less than or equal to $m$ steps. The computational cost depends on the algorithm which is being employed in step 3 and 4 to generate feasible selection vector and choosing one with minimum risk. The algorithm which is addressed in Steps 3 and 4, considers on the order of $O(\max((m_{n_1}), (m_{n_1+1})))$ different $b$ vectors which compares well with the generation of $O(2^s)$ different $b$ vectors in Section 4.3.2 ($s$ being a fraction of $m$).

4.4.3 Measurement Selection Based on the Prior

For nonlinear measurement model $h(x)$, the measurement matrix $H_k = \nabla h(x)|_{x=x_k}$, and hence the accuracy constraint, depend on the state estimate; iteratively solving the optimization problem (4.15) to find corresponding $x_k^\dagger$ to each $b^\ell$ increases the computational cost. To lower computational cost, the designer could select the vector $b$ based on the prior, calculating the risk and $H_k$ using fixed $x = x_k^\dagger$:

$$\min_b \left[ \| \sum_{i=1}^{m} \frac{b_i}{\sigma_i} e_i e_i^\top (h(x_k^\dagger) - z_k) \|^2 \right]$$

subject to: \( \left( \sum_{i=1}^{m} \frac{b_i}{\sigma_i} h_i^\dagger h_i + J_k^\dagger \right) \geq J_l \) \hspace{1cm} (4.19)

where $h_i$ is the $i^{th}$ row of $H = \nabla h(x)|_{x=x_k}$. There are still on the order of $2^s$ values of $b$ to consider; however, this approach to selecting $b$ is significantly cheaper as the residual $(h(x_k^\dagger) - z_k)$ is computed once and there is no iterative optimization.

After solving (4.19) to find the $b$ that minimizes risk based on the prior, then, only for this single value of $b$, will $x$ be updated by using eqn. (4.16).

This approach shares similarities with standard threshold-based outlier detection methods in that both choose which measurements to use (i.e., choice of $b$) based on the prior. This
saves computation, as only a single posterior state is computed. There are also differences since in RAPS the vector $b$ is chosen to ensure that the performance constraint is achieved, while traditional methods use a fixed threshold and do not consider performance relative to risk.

**Remark 3** Employing binary approach in Section 4.3.2 gives the optimal measurement selection $b^*$ with global minimum risk value $c^*$ while the optimal measurement selection $b^*$ may be missed by employing either greedy approach in Section 4.4.2 or measurement selection based on the prior in Section 4.4.3. Hence, the optimal minimum risk value $c^*$ could be greater than $c^*$ which may increase the state error.
Chapter 5

Moving Horizon RAPS Trajectory Estimation: Linear systems

5.1 Problem Statement

Let $x_k \in \mathbb{R}^n$ represent the state vector at time $k$. This article presents an algorithm for simultaneously estimating a trajectory of states $X_{k-1,L} = [x_{k-L}^T, \cdots, x_k^T]^T \in \mathbb{R}^{(L+1)n}$ when, in addition to measurement noise, the measurements $z_k$ may be affected by outliers. When the meaning is clear, we will use the symbol $X$ in place of $X_{k-1,L}$. The linear model is provided in Section 2.1.1. This section introduces necessary backgrounds. Moving horizon trajectory estimation is described in Section 2.4. At time-step $k$, the Maximum A Posteriori (MAP) approach to estimate the trajectory $X = [x_{k-L}^T, \cdots, x_k^T]^T \in \mathbb{R}^{(L+1)n}$ by using Bayes’ theorem as:

$$
\hat{X} = \arg\max_X P(X, U_{0:k-L-1}, Z_{1:k+1})
= \arg\max_X \left( \prod_{i=0}^{L-1} \frac{p(x_{k-L}|x_{k-i}^-u_{k-L-i}) \prod_{j=0}^{L-1} p(z_{k-j}|x_{k-j})}{p(x_{k-L}^-u_{k-L-i})} \right)
$$

(5.1)

where $U_{0:k-L-1} = \{u_i, \forall i = 0, \cdots, k-1\}$ and $Z_{1:k} = \{z_i, \forall i = 1, \cdots, k\}$. Given prior probability density function (PDF) for the state $x_{k-L} \sim \mathcal{N}(\hat{x}_{k-L}, P_{k-L})$ and the Gaussian assumptions for
the process, and measurement noise, the negative log-likelihood of (5.1) yields [18, 20–22, 34, 36, 57, 58, 63]:

$$\hat{X} = \arg\min_X \left( \|x_k - \hat{x}_k - L \|_{P_k - L} \|^2 + \sum_{i=k-L}^{k-1} \| x_{i+1} - (\Psi_i x_i + G_i u_i) \|^2 \Sigma_i + \sum_{j=k-L+1}^{k} \| z_j - H x_j \|^2 \Sigma_j \right).$$

(5.2)

By defining $W = \text{diag}(P_{k-L}, Q_{k-L}, \cdots, Q_{k-1}, R_{k-L+1}, \cdots, R_k)$, the optimization (5.2) can be transformed into matrix form:

$$\hat{X} = \arg\min_X \| r - DX \|^2 W = \arg\min_X \| \Sigma_W (r - DX) \|^2 = \arg\min_X \| \bar{r} - DX \|^2,$$

(5.3)

where $W^{-1} = \Sigma_W^{-1} \Sigma_W$. The matrix $D$ and vector $r$ are represented in eqn. (2.11). The Jacobian matrix $\bar{D}$ and residual vector $\bar{r}$ are defined by:

$$\bar{D} = \begin{bmatrix}
\Sigma_{P_0} & 0 & 0 & \cdots & 0 & 0 \\
\bar{\Psi}_0 & -\Sigma_{Q_0} & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & \bar{\Psi}_{k-1} & -\Sigma_{Q_{k-1}} \\
0 & \bar{H}_1 & 0 & \cdots & 0 & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & 0 & \bar{H}_k
\end{bmatrix}, \quad \bar{r} = \begin{bmatrix}
\Sigma_{P_0} \hat{x}_0 \\
-\bar{G}_0 u_0 \\
\vdots \\
-\bar{G}_{k-1} u_{k-1} \\
\Sigma_{R_1} z_1 \\
\Sigma_{R_k} z_k
\end{bmatrix}.$$

Eqn. (5.3) is a linear least-squares problem which can be solved using QR decomposition, singular value decomposition (SVD), etc. (see Section 2.2). The error covariance matrix for state trajectory $X$ can be computed by the inverse of the tri-block diagonal information matrix $J = \bar{D}^\top \bar{D}$ with lower computational cost than matrix inversion [35]. Articles [18, 22, 34] present efficient computational algorithms building on the sparsity structure of $\bar{D}$.

Herein, each element of the vector $z_k$ may be affected by outliers at some time instants. Hence, additional reasoning is required. If $L = 1$, this is the Kalman filter (KF) with outliers. KF provides the optimal estimate conditioned on all past outlier decisions being correct. Using
$L > 1$, (a) increases the length of the residual vector allowing more reliable outlier decisions; (b) allows various outlier hypotheses to be considered and compared; and (c) conditions $\hat{X}$ on outlier decisions made prior to $k - L$ which are unchangeable but allows past outlier decisions for $j = [k - L, \ldots, k]$ to be reevaluated in light of new data.

5.2 Problem Formulation: Moving Horizon with Outliers

To estimate the state trajectory $X$ over a time window with length $L$, the expression in eqn. (5.3) is minimized. However, if outliers are included in the set of measurements $Z$, they will cause the state estimate and error covariance to be incorrect. Herein, the main idea to overcome the problem of outlier measurements is to define a function to characterize the risk associated with any subset of the available measurements and then to find a subset with minimum risk that satisfies a constraint on the required performance.

To select a subset of measurements $[61, 62]$, a selection vector $b = [b_1, b_2, \ldots, b_L]^T \in \mathbb{B}^L m$ is introduced, where $b_i = [b_{i1}, b_{i2}, \ldots, b_{im}]^T \in \mathbb{B}^m$ is a binary indicator vector with $b_{ij} \in \{0, 1\} \forall i = 1, \ldots, L, j = 1, \ldots, m$. The binary selection vector disables the measurement that is the $j$th element of $z_{k-L+i}$ when $b_{ij} = 0$ or enables it when $b_{ij} = 1$.

Assuming that the prior and the propagation model are trusted, then the MAP estimation problem incorporating the decision vector $b$ is:

$$\hat{X} = \arg \max_{X, b} P(X, U, Z; b) = \arg \max_{X} P(Z|X, U; b) P(X, U)$$

$$= \arg \max_{X, b} p(x_{k-L}) \prod_{i=0}^{L-1} p(x_{k-i}|x_{k-1-i}; u_{k-1+i}) \prod_{j=1}^{L} p(z_{k-j+1}|x_{k-j+1}; b_j).$$

Similarly, employing the negative log-likelihood yields:

$$\hat{X} = \arg \min_{X, b} \left( \|x_{k-L} - \hat{x}_{k-L}\|^2_{P_{k-L}} + \sum_{i=k-L}^{k-1} \|x_{i+1} - (\Psi_i x_i + G_i u_i)\|^2_{Q_i} + \sum_{j=1}^{L} \|\Phi(b_j) (z_{k-j+1} - H_{k-j+1} x_{k-j+1})\|^2_{R_j} \right),$$

(5.4)
where \( \phi(b_j) = \text{diag}(b_j) \) for \( \forall j = 1, \cdots, L \). By defining \( I_n \) as the \( n \times n \) identity matrix and the block diagonal matrix \( \Phi(b) = \begin{bmatrix} I_n & 0 \\ 0 & \phi(b) \end{bmatrix} \), eqn. (5.4) can be converted to:

\[
X^*, b^* = \arg\min_{X,b} \| \Phi(b)(DX - \bar{r}) \|_2^2
\]
subject to: \( b_{ij} \in \{0,1\} \ \forall i = 1, \cdots, L, \ j = 1, \cdots, m \),

(5.5)

Note that eqns. (5.3) and (5.5) are equivalent if all measurements are selected (i.e., all elements \( b_{ij} = 1 \)).

For any \( X \) and \( b \), the cost function is:

\[
C(X,b) = \| \Phi(b)(DX - \bar{r}) \|_2^2,
\]

(5.6)

which is a function of \( X \) that quantifies the risk associated with using the measurements selected by \( b_{ij} = 1 \). The \( i, j \)-th element of the binary vector \( b \) determines whether or not the \([ (i - 1) m + j ]\)-th row of \( \bar{D}_2 \) and \( \bar{r}_2 \), respectively, are non-zero.

The minimum risk associated with the cost function is achieved when all the measurements are discarded (i.e. \( b = 0_{mL} \in \mathbb{R}^{mL} \)). This minimum risk solution yields the worst state accuracy, because no measurements are used. The trade-off of increased risk incurred to improve accuracy is fundamental; however, some measurements provide better risk versus accuracy than others.

The single epoch risk averse performance specified (RAPS) approach is proposed in Chapter 3 to choose a subset of measurements to estimate the state vector \( x_k \), where a subset is selected to have minimum risk of containing an outlier while achieving a specified accuracy. To extend the RAPS optimization to the problem of state estimation in the presence of outliers within a fixed-lag sliding window, the accuracy as quantified by the Fisher information matrix for the trajectory \( X \) corresponding to the selected measurements can be defined by:

\[
J_b = \bar{D}^\top \Phi(b)^\top \Phi(b) \bar{D}.
\]
Due to diagonal and symmetric selection matrix $\Phi(b)$ and binary assumption on $b$, $\Phi(b)^\top \Phi(b) = \Phi(b)^2 = \Phi(b)$. Therefore, the information constraint is:

$$
\tilde{D}^\top \Phi(b) \tilde{D} = \tilde{D}_0^\top \tilde{D}_0 + \tilde{D}_1^\top \tilde{D}_1 + \tilde{D}_2^\top \phi(b) \tilde{D}_2
$$

(5.7)

with $J_p = \tilde{D}_0^\top \tilde{D}_0 + \tilde{D}_1^\top \tilde{D}_1$. Hence, the RAPS optimization problem is:

$$
P_1 : X^*, b^* = \arg\min_{X, b} \left[ \|\tilde{D}_0^\top X - \tilde{r}_0\|^2 + \|\tilde{D}_1^\top X - \tilde{r}_1\|^2 + \|\Phi(b)(\tilde{D}_2^\top X - \tilde{r}_2)\|^2 \right]
$$

subject to: $J_p + \tilde{D}_2^\top \phi(b) \tilde{D}_2 \geq J_l$

$$
b_{ij} \in \{0, 1\}, \forall i = 1, \cdots, L, j = 1, \cdots, m.
$$

where $J_l$ is an user-defined lower bound specified for the accuracy. The choice of $J_l$ is discussed in Section 5.5.3.

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

$$
P_1^r : X^*, b^* = \arg\min_{X, b} \left[ \|\tilde{D}_0^\top X - \tilde{r}_0\|^2 + \|\tilde{D}_1^\top X - \tilde{r}_1\|^2 + \|\Phi(b)(\tilde{D}_2^\top X - \tilde{r}_2)\|^2 \right]
$$

subject to: $J_p + \tilde{D}_2^\top \phi(b) \tilde{D}_2 \geq J_l$

$$
b_{ij} \in [0, 1], \forall i = 1, \cdots, L, j = 1, \cdots, m,
$$

which is a convex problem in either variable $X$ or $b$, seperately, but not jointly. Note that the feasible set for the relaxed optimization $P_1^r$ (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 $P_1^r$ is a lower bound on the objective value of $P1$ [33].

### 5.3 Solution Methods

Problem $P1^r$ can be solved by applying multi-convex programming [55], which alternatively updates $b$ and $x$ using the modified algorithm described below. The following solution
is an extension of nonbinary solution, provided in Section 3.3.1 to moving horizon state estimation problem. This problem is solved iteratively until the convergence condition is met. The iteration number will be indicated by a right superscript \( \ell \), starting at zero.

1. Selecting the measurements: The optimal selection vector \( b^{\ell+1} \) is found for fixed \( X^\ell \). Hence, the first two terms in \( C(X, b) \) as defined in eqn. (5.8) can be dropped because they are independent of \( b \). A proximal term \( \lambda \|b - b^\ell\|^2 \) is required [7] to penalize the change of \( b^{\ell+1} \) in comparison with \( b^\ell \). Then, the optimization is:

\[
P_{2b} : b^{\ell+1} = \arg\min_b \| \phi(b)(\bar{D}_2X^\ell - \bar{r}_2) \|^2 + \lambda \|b - b^\ell\|^2
\]

subject to: \( J_l - J_p - \bar{D}_2^T \phi(b)\bar{D}_2 \leq 0 \)

\[
b_{ij} \in [0, 1], \; \forall i = 1, \ldots, L, \; j = 1, \ldots, m,
\]

where \( \lambda > 0 \) is the user-defined proximal parameter. This is a linear least squares problem constrained by a linear matrix inequality (LMI). The constraint on the information matrix is a positive definite cone which yields a positive semidefinite problem \( P_{2b} \). Hence, \( P_{2b} \) can be solved by interior point methods.

2. State update: This step optimizes the variable \( X^{\ell+1} \) for 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:

\[
P_{2x} : X^{\ell+1} = \arg\min_X C(X, b) + \beta \|X - X^\ell\|^2,
\]

where \( \beta > 0 \) is the user-defined proximal parameter. Problem \( P_{2x} \) is an unconstrained linear least squares optimization problem and can be transformed into matrix form:

\[
P_{2x} : \min_{X^{\ell+1}} \|Ax - c\|^2,
\]

where \( A^T = [\bar{D}_0^T, \bar{D}_1^T, \phi(b^{\ell+1})\bar{D}_2^T, \beta I_{(L+1)n}]^T \) and \( c^T = [\bar{r}_0^T, \bar{r}_1^T, \phi(b^{\ell+1})\bar{r}_2^T, X^\ell]^T \). Eqn. (5.9) could be minimized using Cholesky, singular value decomposition decomposition,
QR decomposition, etc (see Section 2.2). Efficient solutions based on the sparse structure of $A$ are discussed in [9, 43].

The two steps of the optimization are interchangeable. Which step to begin with can be decided based on the accuracy of the prior $X^0$. 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. Otherwise, the optimization can start by finding $X^1$ in Step 2 assuming all the measurements are selected (i.e. $b^0_{ij} = 1$ for $\forall i = 1, \cdots, L, j = 1, \cdots, m$) which is equivalent to applying the Least Squares method to estimate $X$.

Solving the SDP in Step 1 is time consuming and the computational cost grows as the number of the measurements increases. Therefore, for real-time applications, the computational load versus performance trade offs related to $L$ must be carefully considered. In most applications, the accuracy specification is given as diagonal information matrix. Solution in this important special situation will be discussed in the next section.

5.4 Reducing Computational Load

When, the accuracy specification is given by a diagonal matrix, the information constraint in optimization $P1_r$ can be simplified:

$$
\begin{align*}
\text{diag}(J_p) + \text{diag}((\tilde{D}_2^T \phi(b) \tilde{D}_2) & \geq J_d \\
\mathbb{1}_p + \text{diag}(\sum_{i=1}^{L} \sum_{j=1}^{m} b_{ij} \bar{d}^T_{i-1} m + j \bar{d} (i-1) m + j) & \geq J_d \\
\mathbb{1}_p + \sum_{i=1}^{L} \sum_{j=1}^{m} b_{ij} \text{diag}((\tilde{d}^T_{i-1} m + j \tilde{d} (i-1) m + j) & \geq J_d \\
\mathbb{1}_p + \sum_{i=1}^{L} \sum_{j=1}^{m} b_{ij} \begin{bmatrix}
\tilde{d}^2_{(i-1) m + j, 1} \\
\vdots \\
\tilde{d}^2_{(i-1) m + j, (L+1) n}
\end{bmatrix} & \geq J_d,
\end{align*}
$$

(5.10)
where \( \mathbb{J}_p = \text{diag}(J_p) = \text{diag}(D_0^\top D_0 + D_1^\top D_1) \), \( d_i \) is the \( i^{th} \) row of \( D_2 \) and \( d_{(i-1)m+j} \) is the \( j^{th} \) element of \( d_{(i-1)m+j} \). Moreover, the variable \( J_d \) is an user-defined minimum accuracy specification for diagonal elements of the information matrix and \( d_{ij} \) selects the entry in \( i^{th} \) row and \( j^{th} \) column of \( D_2 \) (i.e. the \( j^{th} \) element in \( d_i \)). Define matrix \( D \in \mathbb{R}^{(L+1)n \times mL} \) as an element wise square of the matrix \( \bar{D}_2^\top \)

\[
D^\top = \begin{bmatrix}
    d_{1,1} & d_{1,2} & d_{1,3} & \cdots & d_{1,(L+1)n} \\
    \vdots & \vdots & \ddots & \vdots \\
    d_{mL,1} & d_{mL,2} & d_{mL,3} & \cdots & d_{mL,(L+1)n}
\end{bmatrix},
\]

then the constraint (5.10) can be converted to:

\[
\mathbb{J}_p + D b \geq J_d \tag{5.11}
\]

Consequently, the optimization is:

\[
P3 : X^*, b^* = \underset{X,b}{\text{argmin}} \left[ \|D_0 X - \bar{r}_0\|^2 + \|D_1 X - \bar{r}_1\|^2 + \|\phi(b)(\bar{D}_2 X - \bar{r}_2)\|^2 \right]
\]

subject to: \( \mathbb{J}_p + D b \geq J_d \) \tag{5.12}

\[
b_{ij} \in [0, 1], \ \forall i = 1, \cdots, L, \ j = 1, \cdots, m,
\]

which is a convex problem for either variables \( X \) or \( b \), separately. Similarly, the multi-convex programming can be applied to iteratively update \( b \) and \( X \). Step 2 will be the same as Section 5.3. Step 1 will be changed to solve the following optimization with fixed \( X \) as:

\[
P3_b : b^{f+1} = \underset{b}{\text{argmin}} \ |\phi(b)(\bar{D}_2 X^{f} - \bar{r}_2)|^2 + \lambda \|b - b^{f}\|^2
\]

subject to: \( J_d - (\mathbb{J}_p + D b) \leq 0 \)

\[
b_{ij} \in [0, 1], \ \forall i = 1, \cdots, L, \ j = 1, \cdots, m,
\]

which is a constrained linear least square problem which can be solved by Lagrangian method, interior point and etc. At each iteration, the constrained linear least square problem can be solved with a complexity of \( O((mL)) \) operations [73] while the general case solution discussed

54
in Section 5.3 employs interior point method in each iteration to solve the SDP problem. These methods typically require a few tens of iterations. Each iteration can be carried out with a complexity of $O((mL)^3)$ operations for dense matrices. However, the sparsity of $\bar{D}_2$ can reduce its computational cost.

### 5.5 Discussion

#### 5.5.1 Discussion of the Selection Vector $b$

The selection vector $b$ computed by the two methods discussed in Sections 5.3 and 5.4, is a non-binary vector (i.e. $b \in [0,1]^{mL}$). Partition the matrix $\Sigma_W$ into $\Sigma_W = \text{diag}(\Sigma_P, \Sigma_Q, \Sigma_R)$ with $\Sigma_P \in \mathbb{R}^{n \times n}$, $\Sigma_Q \in \mathbb{R}^{n(L+1) \times n(L+1)}$, and $\Sigma_R \in \mathbb{R}^{Lm \times Lm}$, then, the accuracy is:

$$D^\top \Psi(b) D = J_p + D_1^\top \phi(b) D_1$$

$$= J_p + D_1^\top \Sigma_R^\top \phi(b) \Sigma_R D_1 \tag{5.13}$$

Hence, the non-binary selection vector could be treated as a measurement noise covariance amplification factor that scales the measurement noise information matrix based on the risk-level of each measurement. Alternatively, the non-binary selection vector $b$ could be converted to a binary vector $b^b$ by using a threshold $\tau$ such that:

$$b^b_{ij} = \begin{cases} 1 & b_{ij} \geq \tau \\ 0 & \text{otherwise} \end{cases}$$

However, using the threshold approach imposes two drawbacks: (a) Using binary $b^b$, the computed accuracy $D^\top \Psi(b^b) D$ is different than the solution of the optimization (5.12) which may not satisfy the lower bound accuracy constraint; and, (b) A user-defined threshold must be selected.
5.5.2 Most Recent State $x_k$ Accuracy Specification

At each time $k$, the most important goal in control applications is to estimate the most recent state $\hat{x}_k$ and its accuracy $J_k$ which could also be achieved by the single epoch RAPS solution [1] (i.e. $L = 0$). However, using a single epoch of measurements may not provide sufficient redundancy to remove the effects of outliers resulting in an inaccurate state estimate and covariance (i.e. the covariance is too small and the state estimate is biased). Measurement redundancy can be increased by increasing the time window length $L$. Moreover, if an outlier is selected in RAPS at time instant $k$, there is an opportunity to correct the incorrect selection in $L - 1$ steps ahead. Hence, a trajectory of states $X = [x_{k-L}^T, \cdots, x_k^T] \in \mathbb{R}^{(L+1)n}$ is estimated by selecting measurements within the sliding window with length $L$ which has minimum risk and satisfies the accuracy constraint. However, instead of a specified accuracy on the trajectory of states $X$, a performance accuracy of the most recent state $x_k$ can be of interest.

If the Jacobian matrix is partitioned as:

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

with $J_{LL} \in \mathbb{R}^{nL \times nL}$ and $J_{kk} \in \mathbb{R}^{n \times n}$. Employing marginalization [22], the information matrix for the most recent state is

$$J_k = J_{kk} - J_{kL}J_{LL}^{-1}J_{Lk}. \quad (5.14)$$

To implement this approach, the accuracy constraint in problem $(P_{1r})$ should change to:

$$J_{kk} - J_{kL}J_{LL}^{-1}J_{Lk} \geq J_{ll}, \quad (5.15)$$

where $J_{ll}$ is an user-defined lower bound on the accuracy for the most recent state $x_k$. However, the information constraint in eqn. (5.15) is nonlinear in $b$.  

56
To avoid the nonlinear constraint, the designer might instead choose the linear constraint:

\[ J_{kk} \geq J_{lk} \]

which is equivalent to

\[ J_{kk} - J_{kl}J_{LL}^{-1}J_{lk} \geq J_{lk} - J_{kl}J_{LL}^{-1}J_{lk} \]

\[ J_k \geq J_{lk} - J_{kl}J_{LL}^{-1}J_{lk}. \]  

However, the constraint in eqn. (5.15) imposes:

\[ J_{kk} \geq J_{lk} + J_{kl}J_{LL}^{-1}J_{lk}. \]  

Obviously, satisfying the feasibility constraint of eqn. (5.16), which is equivalent to eqn. (5.18), does not guarantee the last state accuracy constraint in eqn. (5.19), because \( J_{LL} \geq 0 \) and \( J_{kl}J_{LL}^{-1}J_{lk} \) depends on \( b \).

### 5.5.3 Choosing \( J_l \)

In many applications, the user-defined specified accuracy \( J_l \) can be easily selected to satisfy the desired accuracy. For example, the experimental section of this article will discuss highway vehicle applications, where horizontal position accuracy is critically important; therefore, \( J_l \) is chosen such that the north and east components of the vehicle position are constrained.

For example, consider the typical situation where there is an accuracy constraint on the horizontal position accuracy of the most current state \( x_k \). The state vector is \( x = [p^T, v^T, a^T]^T \in \mathbb{R}^9 \) where \( p, v \) and \( a \) represent the rover 3D position, velocity and acceleration vectors, respectively. The vector \( p \) is comprised of north (\( N \)), east (\( E \)), and down (\( D \)) as \( p = [N, E, D]^T \). Then, based on Section 5.5.2, the last state accuracy constraint is:

\[ V(J_{kk} - J_{kl}J_{LL}^{-1}J_{lk})V^T \geq J_{lb}, \]  

\(^{57}\)
where $V = \begin{bmatrix} I_3 & 0_{3 \times 6} \\ 0_{6 \times 3} & 0_6 \end{bmatrix}$ and $J_{lb}$ is the user-defined specification for position accuracy. The constraint in eqn. (5.20) is nonlinear in $b$ (as discussed in Section 5.5.2). To avoid this nonlinear optimization, the experiment will instead use the constraint: $V J_{kk} V^\top \geq J_{lb}$.

Let $x \in \mathbb{R}$ is any random variable with Gaussian distribution $x \sim \mathcal{N}(\mu, \sigma)$, then:

$$\Pr(|x - \mu| \leq \gamma \sigma) = erf\left(\frac{\gamma}{\sqrt{2}}\right)$$  \hspace{1cm} (5.21)

where $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2)dt$ is error function. Therefore, using eqn. (5.21), for given $\alpha$ and $\beta$, the performance specification $|x - \mu| \leq \alpha$ is achieved with probability $\beta$ if $\sigma \leq \varepsilon$ where $\varepsilon$ is:

$$\varepsilon = \alpha \left(\sqrt{2} erf^{-1}(\beta)\right)^{-1}$$

where $erf^{-1}$ represents the inverse function of error function. Table 5.1 represents example values of $\varepsilon$ in applications.

Similarly, for $x \in \mathbb{R}^n$, we have:

$$P \leq P_l$$

where $P = \text{cov}(x)$ and $P_l = \text{diag}(\varepsilon^2_1, \cdots, \varepsilon^2_n)$. Then, the inverse relation between covariance and information matrices (i.e. $J = P^{-1}$) implies:

$$J \geq J_l$$

where $J_l \geq P_l^{-1}$.

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 require the use of more measurements, thereby increasing the risk of inclusion of outliers.

The largest possible lower bound $J_{lb}$ corresponds to the Fisher information matrix computed from using all measurements. If $J_{lb}$ is large, then the number of measurements may
not be enough to satisfy the information constraint (e.g. there is no feasible solution). Hence, all measurements will be chosen without considering their risk. For this case, or when a feasible solution exists but yields to high risk, it should activate a high risk alarm and the solution should only be used with caution. When the high risk alarm is activated, an alternative can be to decrease $J_I$ to select fewer measurements to achieve lower risk; however, this also yields lower expected accuracy.

Table 5.1: Covariance lower bound calculation for different $\alpha$ and $\beta$

<table>
<thead>
<tr>
<th>$\alpha(m)$</th>
<th>2</th>
<th>2</th>
<th>2</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>0.5</th>
<th>0.5</th>
<th>0.5</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta(%)$</td>
<td>0.997</td>
<td>0.954</td>
<td>0.682</td>
<td>0.997</td>
<td>0.954</td>
<td>0.682</td>
<td>0.997</td>
<td>0.954</td>
<td>0.682</td>
</tr>
<tr>
<td>$\varepsilon(m)$</td>
<td>0.673</td>
<td>1.010</td>
<td>1.991</td>
<td>0.336</td>
<td>0.505</td>
<td>0.995</td>
<td>0.168</td>
<td>0.252</td>
<td>0.497</td>
</tr>
</tbody>
</table>
Chapter 6

Moving Horizon RAPS Trajectory Estimation: Nonlinear systems

6.1 Problem Statement

Let $x_k \in \mathbb{R}^n$ represent the state vector at discrete-time $k$. This chapter addresses the problem of state estimation of the trajectory of states $X = [x_{k-L}^T, \cdots, x_k^T]^T \in \mathbb{R}^{(L+1)n}$ for nonlinear systems, over a window of duration $L$, when in addition to measurement noise each element of the measurement vector $z_k$ may be affected by outliers at some time instants. The nonlinear system equations are provided in Section 2.1.2. We assume a Gaussian prior probability function (PDF) $x_{k-L} \sim \mathcal{N}(\hat{x}_{k-L}, P_{k-L})$ is known. This section presents necessary backgrounds.

At time-step $k$, the moving horizon state trajectory (MH-ST)

$$X(k) = [x_{k-L}^T, \cdots, x_k^T]^T \in \mathbb{R}^{(L+1)n}$$

will be estimated. The Gaussian prior probability function $\mathcal{N}(\hat{x}_{k-L}, P_{k-L})$ is computed from the trajectory estimate at the prior time instant:

$$X(k-1) = [x_{k-L-1}^T, \cdots, x_{k-1}^T]^T$$
The Cholesky Decomposition of a matrix $\mathbf{U}$ where Taylor series to linearize the nonlinear functions. This nonlinear least squares optimization can be solved iteratively by employing the first order process noise, and measurement noise distributions are Gaussian, maximizing the posterior pdf is equivalent to minimizing the negative log-likelihood of (6.1):

$$\dot{X} = \arg\max_X \left( \|x_k - \hat{x}_k\|_{\mathbf{P}_{k-L}}^2 + \sum_{i=k-L}^{k-1} \|f(x_i, u_i) - x_{i+1}\|_{\mathbf{Q}_i}^2 + \sum_{j=k-L+1}^k \|h(x_j) - z_j\|_{\mathbf{R}_j}^2 \right).$$

This nonlinear least squares optimization can be solved iteratively by employing the first order Taylor series to linearize the nonlinear functions $h$ and $f$. Using counter $i$ with initial value $i = 0$, the trajectory $\hat{X}'$ is used for linearization at each iteration:

$$\delta X' = \arg\min_{\delta X} \left[ \|\delta x_{k-L} - r'_{k-L}\|_{\mathbf{P}_{k-L}}^2 + \sum_{j=k-L+1}^k \|H_j' \delta x_j - \delta z'_j\|_{\mathbf{R}_j}^2 \right] + \sum_{i=k-L}^{k-1} \|\Psi_i' \delta x_i - \delta x_{i+1} - r'_i\|_{\mathbf{Q}_i}^2$$

where $r'_{k-L} = \delta \mathbf{X}_L$, $r'_{k-L} = \mathbf{X}_L - \delta \mathbf{X}_L$, $r'_i = \dot{X}_{i+1} - f_i(\dot{X}_i, u_i)$, $\Psi_i' = \nabla_{x_i} f_i(x, u_i)|_{x=\dot{X}_i}$, and $\hat{X}' = \dot{X}' + \delta X'$. The Cholesky Decomposition of a matrix $W$ provides $\Sigma_W$ such that $W^{-1} = \Sigma_W^{-1} \Sigma_W$. With this notation, the squared Mahalanobis distance $\|v\|_W = \|\Sigma_W v\|$ where the notation $\|x\|$ represents the standard 2-norm of vector $x$.

The optimization (6.3) can be transformed to the matrix form

$$\delta X' = \arg\min_{\delta X} \|r' - D' \delta X\|^2.$$
The Jacobian matrix $\bar{D}$ and residual vector $\bar{r}$ are:

$$
\bar{D} = \begin{bmatrix}
\Sigma_{R_{k-1}} & 0 & 0 & \cdots & 0 & 0 \\
0 & \bar{F}_{k-L-1} & -\Sigma_{Q_{k-L}} & 0 & \cdots & 0 & 0 \\
0 & 0 & \bar{F}_{k-L} & -\Sigma_{Q_{k-L}} & \cdots & 0 & 0 \\
0 & 0 & 0 & \cdots & \bar{F}_{k} & -\Sigma_{Q_{k}} \\
0 & 0 & 0 & \cdots & 0 & \bar{H}_{k} \\
0 & 0 & 0 & \cdots & 0 & 0 & \bar{H}_{k} \\
\end{bmatrix}
$$

$$
\bar{r} = \begin{bmatrix}
\Sigma_{R_{k-1}}r'_{x_{k-L}} \\
\Sigma_{Q_{k-L}}r'_{x_{k-L}} \\
\Sigma_{Q_{k-1}}r'_{x_{k-L}} \\
\Sigma_{R_{k-1}}\delta z'_{k-L-1} \\
\Sigma_{R_{k}}\delta z'_{k} \\
\end{bmatrix}
$$

where $\bar{F}_{k} = \Sigma_{Q_{k}}F_{k}$ and $\bar{H}_{k} = \Sigma_{R_{k}}H_{k}$. The matrices $\bar{D}_1 \in \mathbb{R}^{(L+1)n \times (L+1)n}$ and $\bar{D}_2 \in \mathbb{R}^{Lm \times (L+1)n}$ represent the sub-matrices of $\bar{D}$ that correspond with the sub-vectors $\bar{r}_1 \in \mathbb{R}^{(L+1)n}$ and $\bar{r}_2 \in \mathbb{R}^{Lm}$ of $\bar{r}$, as indicated by the horizontal lines. Eqn. (6.4) is a linear least-squares problem which can be solved using the QR decomposition, the singular value decomposition (SVD), etc. (see Section 2.2).

### 6.2 Problem Formulation: Moving Horizon with Outliers

Outliers in the set of measurements $Z_{k-L-1:k}$ can yield the incorrect trajectory estimate and error covariance. Therefore, when the measurement $z_{k}$ may be effected by outliers, the effect of the outliers should be removed prior to trajectory estimation. Outliers directly affect the residual $\bar{r}_2$, but also indirectly affect the residual $\bar{r}_1$ through their effect on the estimated trajectory $\hat{X}'$. To address outliers, the cost function in eqn. (6.4) is used to quantify the risk of including outliers.
A binary selection vector \( b = [b_1, b_2, \cdots, b_L]^{\top} \in \mathbb{B}^L \) is introduced to allow selection of the measurements to minimize risk while satisfying a specified accuracy bound. In this vector, \( b_i = [b_{i1}, b_{i2}, \cdots, b_{im}]^{\top} \in \mathbb{B}^m \) is a binary 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 time-evolution defined in eqn. (2.3) are trusted (i.e., not affected by outliers), then by defining the block diagonal matrix \( \Phi(b) = \begin{bmatrix} I_{n(L+1)} & 0 \\ 0 & \phi(b) \end{bmatrix} \) where \( \phi(b) = \text{diag}(b) \), eqn. (6.4) can be generalized as:

\[
\begin{align*}
\delta X^*, b^* & = \arg\min_{\delta X, b} \| \Phi(b)(D\delta X - \bar{r}) \|^2 \\
\text{subject to: } b_{ij} & \in \{0, 1\} \\
& \forall i = 1, \cdots, L, \, j = 1, \cdots, m.
\end{align*}
\]

(6.5)

For any fixed \( b \), the cost \( C(\delta X, b) = \| \Phi(b)(D\delta X - \bar{r}) \|^2 \) is a function of \( \delta X \) that quantifies the risk associated with the measurements selected by \( b_{ij} = 1. \) Activating \( b_{ij} = 1 \) means that the \([(i-1)m+j]\)-th row of \( \bar{D}_2 \) and \( \bar{r}_2 \), respectively, are non-zero. The optimization problem (6.5) for given \( b \), is a linear least squares problem.

Each additional activated measurement increases the expected accuracy (predicted by the information matrix) and also the risk of outlier inclusion. Hence, the minimum risk achievable is zero, when none of the measurements are selected (i.e. \( b = 0_{mL} \in \mathbb{B}^{mL} \)). However, the expected accuracy of the set of measurements is also the worst possible.

For MH-RAPS state estimation, the accuracy is quantified by the Fisher information matrix corresponding to the selected measurements

\[
J_b = \bar{D}^\top \Phi(b)^\top \Phi(b) \bar{D}^\top.
\]

The selection matrix \( \Phi(b) \) is diagonal and symmetric (i.e. \( \Phi(b)^\top = \Phi(b) \)). In addition, the binary assumption on \( b \) yields \( \Phi(b)^2 = \Phi(b) \). Therefore, the information constraint based on
The optimization \( P_1 \) can be solved using multi-convex programming, alternatively updating \( b \) and \( x \) [55]. The following modifies the nonbinary MH-RAPS solution provided in Section 5.3 to nonlinear applications. The iteration counter \( \ell \) is used as a superscript \( \ell \), starting at zero. In the \( \ell^{th} \) iteration, \( X^{\ell+1} \) and \( b^{\ell+1} \) are computed in two steps:
1. **Select the measurements:** This step optimizes the selection vector $b^{\ell+1}$ for fixed $X^{\ell}$. The first term in $P_1$, can be dropped because it is independent of $b$. Based on [7], the proximal term $\lambda \| b - b^\ell \|^2$ is added to penalize the change of $b^{\ell+1}$ in comparison with $b^\ell$, the optimization is:

$$NP_{2b} : b^{\ell+1} = \arg\min_b \| \phi(b)(\bar{D}_2^T \delta X - \bar{r}_2^T) \|^2 + \lambda \| b - b^\ell \|^2$$
subject to: $J_p + \bar{D}_2^T \phi(b)\bar{D}_2 \geq J_0$

$$b_{ij} \in [0,1] \; \forall i = 1, \cdots, L, \; j = 1, \cdots, m.$$ 

where $\lambda > 0$ is the user-defined proximal parameter. The problem $NP_{2b}$ is a linear least squares problem constrained by a linear matrix inequality (LMI) which is a standard semidefinite programming problem and can be solved by employing an interior point method.

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

$$NP_{2\delta X} : \delta X^{\ell+1} = \arg\min_{\delta X} \left[ \| \bar{D}_1^T \delta X - \bar{r}_1^T \|^2 \right. \left. + \| \phi(b)(\bar{D}_2^T \delta X - \bar{r}_2^T) \|^2 + \beta \| \delta X - \delta X^\ell \|^2 \right]$$

where $\beta > 0$ is the user-defined proximal parameter. Problem $NP_{2\delta X}$ is a linear least squares optimization problem which can be minimized using QR decomposition, singular value decomposition (SVD), etc. By finding $\delta X^{\ell+1}$, the state trajectory will be updated to $\hat{X}^{\ell+1} = \hat{X}^\ell + \delta X^\ell$. Then, the matrix $\bar{D}$ and vector $\bar{r}$ will be updated by linearizing around $\hat{X}^{\ell+1}$.

That problem $P_1$, is solved by alternatively solving $NP_{2b}$ and $NP_{2\delta X}$ until a convergence condition is met.
Note that the method above provides both an optimal state estimate and a non-binary selection vector \( b \in [0, 1]^m \). The quantities \( \phi(b) \bar{D}_2 \) and \( \phi(b) \bar{r}_2 \) could be interpreted as a retuning of the weighted measurements based on the assessment of risk indicated by \( b \).

**Remark 4** Note that all subjects discussed in Section 5.5 are applicable to nonlinear MH-RAPS solution described herein.

### 6.4 Reducing Computational Load

The RAPS problem is solved iteratively in two steps. Employing an interior point method to solve the SDP problem in the *select the measurements* step requires a few tens of iterations. Each iteration can be carried out with a complexity of \( O((mL)^3) \) operations for dense matrices. However, the matrix \( \bar{D}_2 \) is sparse which reduces the computational cost. The *state update* step can be solved with the computational cost of order \( O(L^3n^2m) \) for dense matrices. Efficient solutions based on the sparse structure of \( \bar{D} \) is studied in [9, 43]. However, solving the two steps iteratively and updating matrix \( \bar{D} \) and vector \( \bar{r} \) based on each new linearization points can increase the computational cost significantly. To decrease the computational cost, two alternatives could be considered as:

- In the first few iterations, the changes in \( \delta X^\ell \) compared with \( \delta X^{\ell-1} \) may be large which might yield to significant changes in matrix \( \bar{D} \) and vector \( \bar{r} \) based on the sensitivity of the functions \( f \) and \( h \) to \( x \). Then, as the difference from one iteration to the next decreases, the change in matrix \( \bar{D} \) and vector \( \bar{r} \) becomes negligible, converting the nonlinear problem to a linear problem focused on measurement selection. Each different choice of a set of measurements yields a different state estimate.

- The linearization points could be kept unchanged during the multi-convex iterations if the function \( f \) and \( h \) have low sensitivity to \( x \).
• The diagonal element accuracy described in Section 5.4 can be employed to decrease the computational cost of solving SDP problem in \textit{Select the measurements} step.
Chapter 7

Experimental Results

This chapter will apply the RAPS solutions for vehicle state estimation. Both linear and nonlinear applications will be considered. The linear application is referred to in the GNSS literature as a PVA model wherein the GNSS measurements are used to estimate the position, velocity, and acceleration of the GNSS antenna. In the PVA model there is no input \( u \). The nonlinear application is an inertial navigation system (INS) that integrates the inputs \( u \) from an inertial sensor. The error state of the INS is estimated using the same GNSS measurements. The following section briefly introduces essential GNSS and INS navigation system background and notation.

7.1 Background

7.1.1 GNSS Models

State estimation will be performed using double-differenced pseudorange and Doppler measurements [23, 38]. The first difference is performed between the measurements from the rover and corrections from a base station. This first difference removes the errors that are common between all receivers within a local area: ionosphere, troposphere, ephemeris, and satellite
clock. The second difference is between the single-differenced measurements of a pivot satellite and the corresponding single-differences of all the others satellites. This second difference eliminates the receiver clock. Herein, the pivot satellite at each epoch is the one with the highest elevation.

For satellite $s$, the single-differenced pseudorange model is

$$\Delta \rho_k^s = R(p_k, \hat{p}_s^k) + b_k + M_k^s + \eta_k^s$$

where $R(p_k, \hat{p}_s^k) = \|p_k - \hat{p}_s^k\|$ represents the range from the receiver location $p_k$ to the location of satellite $s$, $b_k$ is the receiver clock bias, $M_k^s$ is multipath, and $\eta_k^s \sim \mathcal{N}(0, R_\rho)$ is receiver pseudorange noise. The double-differenced pseudorange measurement model is:

$$\nabla \Delta \rho_k^s = R(p_k, \hat{p}_s^k) - R(p_k, \hat{p}_p^k) + \epsilon_k^s$$

where $\epsilon_k^s \sim \mathcal{N}(0, R_\rho)$ represents the combined effects of measurement noise and multipath in the double-difference pseudorange. Herein $\epsilon_k^s$ is assumed to be uncorrelated between satellites and white. Neither assumption is true in practice, due to the double-difference operation and multipath error.

After compensating for satellite velocity, the single-differenced Doppler $D_k^s$ measurement model is

$$\Delta D_k^s = h_k^s \cdot v_k + r_k + \zeta_k^s$$

where $v_k$ is the rover velocity vector, $r_k$ is the receiver clock drift rate, $\zeta_k^s$ represents Doppler measurement noise and $h_k^s = \left( \frac{p - \hat{p}_s^k}{\|p - \hat{p}_s^k\|} \right)_p$ is the unit vector pointing from the satellite to the estimated receiver location. The double-differenced Doppler measurement model is:

$$\nabla \Delta D_k^s = (h_k^s - h_p^s) \cdot v_k + \epsilon_{D_k}^s$$

where $\epsilon_{D_k}^s \sim \mathcal{N}(0, R_D)$ is assumed to be white and uncorrelated between satellites.

---

1 The equations to compute $\hat{p}_s^k$ from ephemeris data are in [23, 38].
Per satellite, there are two measurements modeled by:

\[
\begin{align*}
    z_k^s &= h^s(x_k) + y_k^s, \\
    y_k^s &= h^s(x_k) - R(p, \hat{p}_k^s) - (h^s_k - h^s_p) \cdot v.
\end{align*}
\]  

where \( z_k^s = [\nabla \Delta \rho_k^s, \nabla \Delta D_k^s] \) and \( h^s = [R(p, \hat{p}_k^s) - R(p, \hat{p}_k^p), (h^s_k - h^s_p) \cdot v] \). The noise vector \( d_k^s \sim \mathcal{N}(0, R) \) with \( R = 
\begin{bmatrix}
    R_p + R_p^p & 0 \\
    0 & R_d + R_d^p
\end{bmatrix}
\). The vector of measurements at epoch \( k \) corresponding to the model in eqn. (2.5) concatenates the measurements per satellite: \( z_k = [z_1^s, \ldots, z_m^s] \).  

7.1.2 GNSS PVA Model

The rover state is \( x = [p^\top, v^\top, a^\top] \in \mathbb{R}^9 \) where \( p, v \) and \( a \in \mathbb{R}^3 \) represent the rover position, velocity and acceleration vectors, respectively. The continuous-time PVA vehicle model is

\[
    \dot{x}(t) = \begin{bmatrix}
        0 & I & 0 \\
        0 & 0 & I \\
        0 & 0 & -\lambda_a I
    \end{bmatrix} x(t) + \begin{bmatrix}
        0 \\
        0 \\
        1
    \end{bmatrix} \omega_a(t),
\]

where \( \omega_a(t) \) is modeled as Gaussian white noise with power spectral density \( Q_a = \sigma_a^2 \).

The corresponding discrete-time PVA vehicle time propagation model is given by

\[
    x_k = \Psi_{k-1} x_{k-1} + \Gamma_{k-1} \omega_{k-1}
\]

with

\[
    \Psi_k = 
    \begin{bmatrix}
        I & T I & a_3 I \\
        0 & I & a_2 I \\
        0 & 0 & a_1 I
    \end{bmatrix},
\]

\[
    \Gamma_k \approx
    \begin{bmatrix}
        T^{5/2}/\sqrt{20} I \\
        T^{3/2}/\sqrt{3} I \\
        \sqrt{T} I
    \end{bmatrix},
\]

and \( Q_d = \begin{bmatrix}
    0 \\
    0 \\
    \sigma_d^2 I
\end{bmatrix} \)

where all submatrices are three by three with \( a_1 = e^{-\lambda_a T} \),

\[
    a_2 = (1 - e^{-\lambda_a T})/\lambda_a, \quad a_3 = (\lambda_a T - 1 + e^{-\lambda_a T})/\lambda_a^2
\]

70
and $\omega_k \sim \mathcal{N}(0, Q_d)$ where $Q_d$ is a covariance matrix. The approximation indicated in $\Gamma_k$ yields
the correct diagonal of the discrete-time noise covariance matrix, but $\Gamma_k Q_d \Gamma_k^T$ approximates the
off-diagonal terms relative to the exact calculation. The details of the model and its parameters
are in [46].

The measurement residual is $r_k^i = z_k^i - h^i(\hat{x}_k)$. The linearized measurement model is:

$$r_k^i = H^i_{dk} \delta x_k + d \gamma_{rk}^i \quad (7.2)$$

where the measurement matrix is

$$H^i_{dk} = \begin{bmatrix} (h_k^i - h^i_{pk}) & 0 & 0 \\ 0 & (h_k^i - h^i_{pk}) & 0 \end{bmatrix} \quad (7.3)$$

Concatenating all satellite residuals $r_k^i$ into the vector $r_k$ yields:

$$r_k = H_d \delta x_k + d \gamma_k \quad (7.4)$$

### 7.1.3 GNSS Aided INS Model

The rover state vector $x = [p^T, v^T, q^T, b_a^T, b_g^T]^T \in \mathbb{R}^{n_s}$ containing the 3D position ($p$), velocity ($v$), accelerometer bias ($b_a$) and gyro bias ($b_g$), and the attitude quaternion ($q \in \mathbb{R}^4$).

Thus $n_s = 16$. The kinematic equation for the rover state is:

$$\dot{x}(t) = f(x(t), u(t)) \quad (7.5)$$

where $f : \mathbb{R}^{n_s} \times \mathbb{R}^6 \rightarrow \mathbb{R}^{n_s}$ represents the kinematics and $u \in \mathbb{R}^6$ is the specific force and angular
rate vector. The form of the function $f$ can be found in many references, e.g., [23, 40].

Let $\tau_i$ denote the time instants at which IMU measurements are available. Assume a
distribution for the initial state $x(t_{k-1}) \sim \mathcal{N}(x_{k-1}^+, P_{k-1})$ and the IMU measurements $\tilde{u}(\tau_i)$ are
given. The INS propagates the estimate of the rover state between aiding measurement time
instants as the integral of

$$\dot{\hat{x}}(t) = f(\hat{x}(t), \tilde{u}(t)) \quad (7.6)$$
where \( \hat{u}(t) = \tilde{u}(t) - \hat{b}(t) \)

The solution of (7.6) over the time interval \( t \in [\tau_{i-1}, \tau_i] \) from initial condition \( x_{i-1} \) is represented as the operator:

\[
\psi(x_{i-1}, u_{i-1}) = x_{i-1} + \int_{\tau_{i-1}}^{\tau_i} f(x(\tau), u(\tau)) d\tau.
\]

This integral operator can be iterated for all IMU measurements \( U_k = \{ \tilde{u}(\tau) \text{ for } \tau \in [t_k-1, t_k] \} \) to propagate the state from \( t_{k-1} \) to \( t_k \). The iterated integral operation is denoted as:

\[
x_k^- = \psi_k(x_{k-1}^+, U_{k-1})
\]

(7.7)

where \( x_k^- \) is the propagated state and \( t_k \) denotes the time instants at which GNSS measurements are valid. The error propagation model is:

\[
\delta x_k^- = \Psi_k \delta x_{k-1}^+ + \omega_k
\]

(7.8)

where \( \omega_k \) is process noise with covariance \( Q_d \) and \( \Gamma_k = I \) and \( \Psi_k \) is the state transition matrix.

The derivations are contained in [23, 40] with formulas for \( \Psi_k \) and \( Q_d \). The error covariance matrix will be propagated through time using eqn. (4.6).

The state estimation is updated using double-differenced GNSS pseudorange and Doppler measurements using the models in Section 7.1.1.

### 7.2 Hardware Setup and Data

#### 7.2.1 Hardware Setup

This experimental hardware includes two GNSS receivers and one IMU. A single-frequency u-blox M8T GNSS receiver provides data for the state estimation measurement updates. For the GNSS-PVA results, it is used alone, in a linear estimation application. For the GNSS-INS results, it is used with IMU data in a nonlinear estimator. The IMU sensor is the
NV-IMU 1000. Real-time differential corrections were obtained from the nearby ESRI base-station via an NTRIP caster using the RTCM standard [5] through a cellular connection. A two-frequency NOVATEL OEMV2 GNSS receiver was also on board to provide more accurate data for computation of a ground-truth trajectory that will be used for state estimation accuracy analysis. Both receivers are connected to the same antenna (Antcomm ANN-MS-0-005). The antenna was mounted on a vehicle’s rooftop.

The two-frequency, integer-resolved, carrier phase GNSS data from the OEMV2 and NV-IMU data are used with the smoothing algorithm in [66] to generate a ground truth trajectory at centimeter accuracy. Ground truth trajectory estimation is performed in post-processing.

The car traveled around a testbed outside the University of California Riverside (UCR) campus. The test environment contains several one-storied buildings and some palm trees. This caused very few instances of GNSS signal obstruction, delay, multipath, or loss. As a result, the dataset itself can be considered as having a clear sky, with few outliers. The test setup is shown in Fig. 7.1.

Figure 7.1: Test trajectory. The bottom corner photos identify a variety of real-world environmental factors which adversely affect the performance of a GPS receiver, e.g. trees and tall buildings.
7.2.2 Outlier Generation for GNSS Measurement

Given the clear sky dataset described in Section 7.2.1, we are able to study the effects of outliers on the algorithms of interest by artificially adding outliers at each epoch. The average number of GNSS measurements per epoch over the entire trajectory was 7. At each epoch, outliers will be added to the pseudorange measurements of 2 randomly selected satellites.

During each experiment, the magnitude of each outlier is selected from the uniform distribution $U(\mu - \alpha, \mu + \alpha)$, with $\alpha = 1.5m$. In different experiments, the outlier mean $\mu$ will be varied from $0 - 20m$ to study performance versus outlier magnitude. For $\mu < \alpha$, the outlier generation distribution was $U(0, \mu + \alpha)$.

7.3 Binary RAPS Experimental Results: One-Time Epoch State Estimation

This section presents experimental results that evaluate the performance of the RAPS approach proposed in Sections 3.3.2 and 4.3.2 relative to conventional approaches. For the linear GNSS-PVA application described in Section 7.1.2, RAPS will be compared with the Neyman-Pearson Kalman filter (NP-KF). For the nonlinear GNSS-INS application described in Section 7.1.3, binary RAPS solved iteratively for $x$ and will be compared with the Neyman-Pearson extended Kalman filter (NP-EKF).

7.3.1 Algorithms

Results are compared for two algorithms:

- **NP-(E)KF** has all the measurements available to it, but artificially generated outliers have been added, as described in Section 7.2.2. The NP-(E)KF ignores all measurements for which the absolute value of their residual is greater than the threshold $s_{ii} =$
\( \gamma \sqrt{R_{ii} + h_i P^{-1} h_i^T} \) where \( \gamma \) is a positive constant. Herein, results are shown for \( \gamma = 2, 3, 4, \) and 5. Note that \( \gamma = 3 \) is typical for outlier accommodation.

- **RAPS** has all the measurements available to it, but outliers have been added. The added outliers are exactly the same as those used for NP-(E)KF. The RAPS binary selection vector determines whether or not measurements are used.

### 7.3.2 Performance Evaluation

The main performance criteria will by horizontal position error, because it is of critical interest in highway applications. Let \( p_r(k) \) denote the ground truth rover position at time \( t_k \). Let \( \hat{p}(k) \) denote the real-time estimate of the rover position at time \( t_k \). The horizontal position error at time \( k \) is computed as

\[
E_p(k) = \| \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} (p_r(k) - \hat{p}(k)) \|,
\]

(7.9)

Performance analysis includes the mean and standard deviation of the norm of this position error, the percentage of examples with sub-meter accuracy, and the maximum position error.

The state estimation accuracy can be directly improved by more measurements and increased measurement information diversity. The enhanced accuracy that might be achieved by including each additional measurement should be balanced by the increased risk associated with that measurement. For performance comparison, it is interesting to simultaneously consider: horizontal position error, risk, and measurement information diversity.

The risk \( R_k \) is defined by:

\[
R_k = \frac{1}{m_k} \sum_{i=1}^{m_k} b_i \frac{|r_i^T|}{\sigma_i}
\]

(7.10)
with \( r_k^i = (z_k^i - h(\hat{x}_k^i)) \), \(|a|\) denotes the absolute value of \( a \), and \( m_k \) is the number of residuals available at time \( k \). In NP-(E)KF, if the \( i^{th} \) measurement passes the threshold test, then \( b_i = 1 \), otherwise \( b_i = 0 \).

Measurement information diversity can be evaluated by Geometric Dilution of Precision (GDOP). The GDOP definition from the GNSS literature is adapted to account for measurement selection as:

\[
GDOP = \sqrt{\text{trace}(H_k^T \Phi(b_k)H_k)^{-1}}
\]  

(7.11)

where observation matrix \( H_k \) is defined following eqn. (7.3). GDOP is a scalar metric that quantifies the diversity of user-to-satellite direction vectors for the selected satellite measurements at time instant \( k \). GDOP has an inverse relationship with the information matrix for the selected measurements. Consequently, the better diversity yields to lower GDOP.

The performance of RAPS versus NP-(E)KF can be compared using their corresponding error, risk, and GDOP figures.

### 7.3.3 Performance versus Outlier Mean

In a sequence of Monte Carlo tests, artificially generated outliers are added to measurements as described in Section 7.2.2. Figs. (7.2a) and (7.2b) show NP-(E)KF and RAPS performance using the linear PVA (left) and nonlinear INS (right) model, respectively. For each value of \( \mu \in [0,20] \) meter, each point on each curve is produced as the average of horizontal position error over 300 seconds and 10 Monte Carlo experiments. For each Monte Carlo run, the same outlier corrupted data is used for both NP-(E)KF and RAPS. Each Monte Carlo experiment generates a new set of outliers for new randomly selected satellites to be used both for NP-(E)KF and RAPS.

For Figs. 7.2a and 7.2b, the y-axis in the top sub-figure is the mean of the horizontal position error. The y-axis in the bottom sub-figure is the percentage of measurements used.
by the state estimation algorithm. At each time instant, the measurement set (pseudorange and Doppler) contains 9% generated outliers. For both PVA and INS applications, the NP-(E)KF mean horizontal position error initially rises and later falls as the magnitude of the outlier increases. The initial rise is due to small outliers being likely to pass through the threshold test. As the magnitude of the outlier (i.e., $\mu$) increases, each NP threshold test removes an increasingly higher percentage of the outliers, until it is correctly removing all the measurements which are affected by outliers. Correct rejection of all outliers occurs for smaller values of $\mu$ as the NP threshold $\gamma$ decreases.

The INS has both more information (i.e., IMU data) and a more accurate (though nonlinear) model. The result is that the GNSS-INS approach yields measurement residuals and error covariance matrix $P_k^{-}$ that are smaller than the GNSS-PVA approach. Therefore, the NP-EKF tests are able to detect and remove outliers at significantly smaller values (i.e., for $\gamma = 5$, $\mu$ equal to 5.4 for GNSS-INS versus 14.4 for GNSS-PVA).

For both the GNSS-PVA and GNSS-INS implementations, the RAPS mean horizontal error is always less than that for NP-(E)KF, regardless of the value of the threshold $\gamma$. This shows that the RAPS approach is robust to the magnitude of the outlier, without the designer having to pick a value for a test parameter (e.g., the NP threshold $\gamma$).

### 7.3.4 Experimental Results for Error, Risk, and GDOP

Figs. 7.3 and 7.4 plot the error, risk, and GDOP metrics for portions of a single experiment using the GNSS-INS data and GNSS-PVA data, respectively. The intent of this figures is to enable a discussion of the interplay between GDOP and risk on the horizontal position error that is achieved. All subfigures include one curve for each of the five algorithms under consideration. Some curves may not be visible at times where multiple curves have identical values.
both figures 7.3 and 7.4 allow performance comparison between the five algorithms for two different values of the outlier mean magnitude $\mu$. In Fig. 7.3, the experiment for the left figure generated outliers using $\mu = 2$ (i.e. outliers magnitude distributed in $U[0.5,3.5]$). The experiment for the right figure generated outliers using $\mu = 7$ (i.e. outliers magnitude distributed in $U[5.5,8.5]$). Fig. 7.4a presents data for $\mu = 6$ (i.e. outliers magnitude distributed in $U[4.5,7.5]$). Fig. 7.4b presents data for $\mu = 17$ (i.e. outliers magnitude distributed in $U[15.5,18.5]$). The artificially generated outliers are different for the two values of $\mu$. The generated outliers affect two randomly chosen satellites at each time instant. At some epochs, the set of satellites unaffected by outliers can have a high GDOP.

RAPS performance (red curves in Figs. 7.3 and 7.4 and bottom row of Tables 7.1 and 7.2) is almost the same for both outlier scenarios. At all times, RAPS generated a feasible solution where its risk was the minimum achieved across all the algorithms. Since it was choosing the minimum risk feasible combination of measurements, RAPS is willing to discard unneeded satellites; therefore, its GDOP is typically greater than or equal to that of the NP-(E)KF implementations. Nonetheless, its horizontal error performance is better.

The difference in performance is more noticeable in Figs. 7.3a and 7.4a, as the smaller outliers are more likely to be missed by the NP-(E)KF threshold test.

NP-(E)KF often uses more satellites - all that pass the threshold test - yielding a smaller GDOP. Some have larger residuals, giving it higher risk. While NP-(E)KF has better (i.e., lower) GDOP, which indicates improved accuracy in outlier-free situations, due to the higher risk of outliers that it incurs, the NP-(E)KF actual performance (i.e., achieved accuracy) is frequently worse than RAPS, especially when risk is high. Statistical horizontal error performance data for these same two experiments is summarized in Tables 7.1 and 7.2.
7.4 Non-binary RAPS Experimental Results: Moving Horizon State Estimation

This section analyzes experimental results to evaluate the performance of the MH-RAPS approach (proposed in Sections 5.4 and 6.3) relative to the MH-NP approach (described below).

7.4.1 Algorithms

Results are compared for two algorithms. The algorithms have the potential to use all the available GNSS data, which includes artificially generated outliers. The added outliers are identical for both algorithms.

- **MH-NP** at each time instant, applies a threshold test (see section 7.3.1) to all measurements within time window \([k - L + 1, k]\). Then, measurements which pass the threshold test, will be employed in MH state estimation. Results will be shown for \(\gamma = 2, 3, 4, \text{ and } 5\). Note that \(\gamma = 3\) is typical for outlier accommodation.

- **MH-RAPS** solves the optimization in Sections 5.4 and 6.3, which provides a non-binary selection vector \(b \in [0, 1]^m\) and a trajectory of states \(X\).

Here, the length of sliding window is \(L = 5\) in both algorithms.

The horizontal position error is the main performance criteria since it is of critical interest in highway applications. The norm of the horizontal position error is defined using the rover position at time \(k\) which yields to eqn. (7.9). The mean and standard deviation of the norm of the horizontal position error, the percentage of examples with sub-meter accuracy, and the maximum position error are studied to compare the performance of each method.
Any outlier accommodation method removes (or lessens the effect of) certain measurements. In addition to evaluating each method based on its achieved accuracy, it is interesting to consider the risk that they incur and the information diversity they utilize. Since in online control related applications, the last state accuracy is of interest, the last state error, risk and information diversity (GDOP) are studied. GDOP for the last state \( x_k \) is defined by using eqn. (7.11). At time \( k \), the normalized risk \( R_k \) for the last state \( x_k \) is quantified by:

\[
R_k = \frac{1}{m_k} \sum_{j=1}^{m_k} b_{kj} \frac{|r_{kj}|}{\sigma_{r_{kj}}}
\]

(7.12)

where \( r_{kj} \) is the \( j^{th} \) element of the residual vector \( r_k = (z_k - H \hat{x}_k) \), \( |a| \) denotes the absolute value of \( a \), and \( m_k \) is the number of residuals available at time \( k \).

### 7.4.2 Performance versus Outlier Mean

Fig. 7.5 shows the performance of the MH-NP and MH-RAPS approaches, averaged over \( m = 15 \) Monte Carlo simulations. Artificially generated outliers are added to randomly chosen measurements as described in Section 7.2.2 in a sequence of Monte Carlo tests. Each point on each curve of the top sub-figure, is produced as the average of horizontal position error over 300 seconds and \( m \) Monte Carlo experiments for each corresponding value of outlier mean \( \mu \in [0, 20] \) meter. Similarly, at each time instant, the normalize average \( \|b\|_1 = \frac{1}{mt} \sum_{i=1}^{L} \sum_{j=1}^{m} b_{ij} \) is computed for both MH-NP and MH-RAPS. In MH-NP, if the \( j^{th} \) measurement at lag \( i \) passes the threshold test, then \( b_{ij} = 1 \), otherwise \( b_{ij} = 0 \). For each value of \( \mu \in [0, 20] \) meter, each point on each curve of the bottom sub-figure, is the average of \( \|b\|_1 \) over 300 seconds and \( m \) Monte Carlo experiments. Each Monte Carlo experiment generates a new set of outliers added to new randomly chosen satellites to be used both for MH-NP and MH-RAPS. Both MH-NP and MH-RAPS use the same outlier corrupted data in each Monte Carlo test.
In Fig. 7.5, the $y$-axis in the top sub-figure is the mean horizontal position error. The $y$-axis in the bottom sub-figure is the normalized average $||b||_1$. The $x$-axis in both figures is the mean of outlier $\mu$. At each time instant, outliers are added to 2 randomly selected code measurements. The measurement set (pseudorange and Doppler) on average contains 10% generated outliers. For small $\mu$, the mean MH-NP horizontal position error increases with $\mu$. Small outliers are likely to pass through the NP threshold and their effect increases with their magnitude. For larger $\mu$, the mean MH-NP horizontal position error eventually decreases as the value of $\mu$ (i.e. mean outlier magnitude) increases. This is because, as the magnitude of the outliers increases, an increasing percentage of the outliers will be removed by the NP threshold test. For $\mu$ large enough, each NP threshold test successfully removes all the measurements that are affected by outliers. Smaller values of $\gamma$ correctly remove all outliers for smaller values of $\mu$. Hence, for smaller value of $\gamma$, the MH-NP horizontal position error mean drops at smaller value of $\mu$.

Fig. 7.5 shows that MH-RAPS performance is robust to outlier magnitude. Also, its mean horizontal position error is always less than or equal to that for MH-NP, regardless of the value of the threshold $\gamma$. Note also that MH-RAPS uses less than 60% of available measurements, as its objective is to satisfy the specification with minimum risk. Additional measurement incurs additional unneeded risk.

7.4.3 Experimental Results for Error, Risk, and GDOP: GNSS-PVA Approach

Fig. 7.6 shows the horizontal position error, risk, and GDOP for portions of one experiment with added outliers using the GNSS-PVA data. By using more measurements, the expected accuracy (i.e. GDOP) might be improved; however, there is a balanced between the enhanced expected accuracy that might be achieved by including each additional measurement and the increased risk associated with the corresponding measurement, as these two combine.
to yield an actual experimental accuracy. Fig. 7.6 shows the interplay between risk, expected accuracy, and the horizontal position error that is achieved for two different values of \( \mu \). All subfigures contain three curves. The blue and green curves display the results for MH-NP with \( \gamma = 2, 3 \), respectively. The error and risk for MH-NP with \( \gamma = 4, 5 \) are higher than \( \gamma = 2, 3 \) and have been discarded to improve the visibility of the other curves. The red curve shows the performance of the RAPS approach.

The left figure presents the results by using measurements affected by artificially generated outliers with \( \mu = 3 \). The right figure shows the results by using measurements affected by artificially generated outliers with \( \mu = 7 \). At each time instant, the generated outliers are added to two randomly selected satellite.

The performance of MH-NP with \( \gamma = 2, 3 \) is different for \( \mu = 3 \) and \( \mu = 7 \). In Fig. (7.6a), the outlier magnitude is small; therefore more likely to pass the NP threshold test. This corresponds to increased risk of higher error due to measurements that are affected by outliers being missed by the NP threshold test. In Fig. (7.6b), the outlier magnitude is larger; hence, MH-NP with \( \gamma = 2 \) can remove a higher percentage of the measurements affected by outliers, while some of the outliers are missed by NP-MH with \( \gamma = 3 \). Therefore, the error and risk corresponds to MH-NP with \( \gamma = 2 \) is lower than for MH-NP with \( \gamma = 3 \). Its GDOP is higher than MH-NP with \( \gamma = 3 \) since NP-MH with \( \gamma = 2 \) uses less measurements.

Table 7.3 summarizes positioning performance statistics using the following criteria: (a) mean, (b) standard deviation, (c) the percentage of samples with sub-meter accuracy, and (d) maximum horizontal position error. The table is divided into two sections, each section contains five rows. The left section contains MH-NP and MH-RAPS for the experiment with \( \mu = 3 \). The right section contains results for MH-NP and MH-RAPS for the experiment with \( \mu = 7 \).

MH-RAPS performance (red curves in Fig. 7.6 and bottom row of Table 7.3) is almost the same for both outlier scenarios (i.e. robust to different values of \( \mu \)). In the MH-RAPS ap-
proach, when feasible, the optimization process ensures that sufficiently diverse measurements are chosen to satisfy the position accuracy specification, while minimizing the corresponding risk. As the number of used measurements increases, the expected information from those measurements increases, but so also does the risk correspond to each added measurement. MH-RAPS may discard some of the measurements with large or small residuals, when they would add risk without adding a sufficient amount of new information. Therefore, the MH-RAPS GDOP is greater than the MH-NP implementations while its risk is lower. MH-RAPS horizontal error performance is also better.

7.4.4 Experimental Results for Error, Risk, and GDOP: GNSS-INS Approach

At each time instant, RAPS is choosing the minimum risk set of measurements that satisfies the accuracy specification. Note that neither the number nor magnitude of the acceptable residuals is explicitly constrained. The RAPS approach may adjust the number and diversity of satellites used to avoid unnecessary risk, while achieving the performance specification. The MH-NP approach uses all residuals that pass its threshold test, possibly achieving a great expected accuracy (lower GDOP) at higher risk, which results in a worse actual accuracy.

Fig. 7.7 shows the horizontal position error, risk and GDOP for a portion of the experiment. The length of the moving horizon sliding window is \( L = 5 \) in both algorithms. This gives each algorithm the ability to use all measurements within a 4 second window. For this figure, artificial outliers are randomly generated as described above with \( \alpha = \mu = 2 \) (i.e. from the Uniform distribution \( U(0,4) \)).

Fig. 7.7 shows that at all times, the RAPS approach (red curve) achieves the lowest risk (as expected). Also, at most times, RAPS has a worse expected accuracy (GDOP), as RAPS is using fewer measurements to minimize risk; nonetheless, the actual achieved accuracy of the
RAPS approach is more consistent and typically better than the MH-NP approach regardless of the value of $\gamma$.

Table 7.4 summarizes positioning accuracy performance statistics including the mean and standard deviation of the horizontal position error in eqn. (7.9), the percentage of examples with sub-meter accuracy, and maximum horizontal position error.

Fig. 7.8 and Table 7.5 present data related to the performance versus the MH window length $L$. Because RAPS is attempting to minimize risk while satisfying an expected performance criteria, as $L$ increases RAPS is able to select among more choices, therefore the risk decreases while the expected performance as quantified by GDOP remains relatively constant (i.e., the constraint is achieved). The actual performance improves slightly as $L$ increases, because the expected performance is achieved with lower risk of outlier inclusion. Table 7.5 presents statistics related to the horizontal position error and the selection vector $b$. In the last column shows the two-norm of the selection vector $b_k$ for the most recent measurement $z_k$ relative to the two-norm of the selection vector $b$ for the entire window. As this ratio decreases, RAPS is choosing to use more old measurements relative to current measurements, showing the value of increasing the window length $L$. 
Table 7.1: GNSS-INS Horizontal Performance Statistics

<table>
<thead>
<tr>
<th>Methods</th>
<th>$\mu = 2$</th>
<th>$\mu = 7$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean of</td>
<td>Std. of</td>
</tr>
<tr>
<td></td>
<td>error</td>
<td>error</td>
</tr>
<tr>
<td>NP-EKF $\gamma = 5$</td>
<td>0.63</td>
<td>0.35</td>
</tr>
<tr>
<td>NP-EKF $\gamma = 4$</td>
<td>0.58</td>
<td>0.26</td>
</tr>
<tr>
<td>NP-EKF $\gamma = 3$</td>
<td>0.45</td>
<td>0.23</td>
</tr>
<tr>
<td>NP-EKF $\gamma = 2$</td>
<td>0.28</td>
<td>0.15</td>
</tr>
<tr>
<td>RAPS</td>
<td>0.24</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Table 7.2: GNSS-PVA Horizontal Performance Statistics

<table>
<thead>
<tr>
<th>Methods</th>
<th>$\mu = 6$</th>
<th>$\mu = 17$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean of</td>
<td>Std. of</td>
</tr>
<tr>
<td></td>
<td>error</td>
<td>error</td>
</tr>
<tr>
<td>NP-EKF $\gamma = 5$</td>
<td>0.72</td>
<td>0.52</td>
</tr>
<tr>
<td>NP-EKF $\gamma = 4$</td>
<td>0.66</td>
<td>0.42</td>
</tr>
<tr>
<td>NP-EKF $\gamma = 3$</td>
<td>0.64</td>
<td>0.41</td>
</tr>
<tr>
<td>NP-EKF $\gamma = 2$</td>
<td>0.37</td>
<td>0.33</td>
</tr>
<tr>
<td>RAPS</td>
<td>0.35</td>
<td>0.31</td>
</tr>
</tbody>
</table>
Figure 7.2: Mean horizontal position error and the percentage of selected measurements versus mean outlier magnitude $\mu \in [0, 20]$. The red curves display the result for the binary RAPS algorithm from Section 3.3.2. The yellow, green, blue and black curves show the results for the NP-(E)KF approach with $\gamma = 2, 3, 4,$ and $5$, respectively.
Figure 7.3: Error, risk, and information diversity (i.e., GDOP) for the (nonlinear) GNSS-INS approach. The yellow, green, blue and black curves display the results for NP-EKF approach with $\gamma = 2, 3, 4,$ and $5$, respectively. The red curve shows the performance of the RAPS approach.
Figure 7.4: Performance comparison using GNSS data with the linear PVA model. The yellow, green, blue and black curves display the results for NP-KF approach $\gamma = 2, 3, 4, \text{and} 5$, respectively. The red curve shows the RAPS performance.
Figure 7.5: Mean horizontal position error and normalized mean $\|b\|_1$ versus mean outlier magnitude $\mu \in [0, 20]$m using linear PVA model. The red curves display the results for the MH-RAPS algorithm from Section 5.5.2. The blue, green, yellow, and black curves show the results for the MH-NP approach with $\gamma = 2, 3, 4, \text{ and } 5$, respectively.

Table 7.3: GNSS-PVA Horizontal Performance Statistics

<table>
<thead>
<tr>
<th>Methods</th>
<th>$\mu = 3$</th>
<th>$\mu = 7$</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean of error</td>
<td>Std. of error</td>
<td>Submeter accuracy</td>
<td>Max error</td>
<td>Mean of error</td>
<td>Std. of error</td>
</tr>
<tr>
<td>MH-NP $\gamma = 5$</td>
<td>0.61 (m)</td>
<td>0.46 (m)</td>
<td>78 (%)</td>
<td>1.90 (m)</td>
<td>0.81 (m)</td>
<td>0.65 (m)</td>
</tr>
<tr>
<td>MH-NP $\gamma = 4$</td>
<td>0.60 (m)</td>
<td>0.46 (m)</td>
<td>78 (%)</td>
<td>1.85 (m)</td>
<td>0.78 (m)</td>
<td>0.53 (m)</td>
</tr>
<tr>
<td>MH-NP $\gamma = 3$</td>
<td>0.58 (m)</td>
<td>0.45 (m)</td>
<td>80 (%)</td>
<td>1.85 (m)</td>
<td>0.69 (m)</td>
<td>0.53 (m)</td>
</tr>
<tr>
<td>MH-NP $\gamma = 2$</td>
<td>0.50 (m)</td>
<td>0.33 (m)</td>
<td>88 (%)</td>
<td>1.60 (m)</td>
<td>0.33 (m)</td>
<td>0.25 (m)</td>
</tr>
<tr>
<td>MH-RAPS</td>
<td>0.21 (m)</td>
<td>0.19 (m)</td>
<td>100 (%)</td>
<td>0.96 (m)</td>
<td>0.29 (m)</td>
<td>0.20 (m)</td>
</tr>
</tbody>
</table>
Figure 7.6: Estimation error, risk, and information diversity (GDOP) versus time. The blue and green curves display the results for MH-NP approach with $\gamma=2$ and 3, respectively. The red curve shows the performance of the MH-RAPS approach.
Figure 7.7: Error, risk and information diversity (GDOP) versus time for $\mu = 2$ and $L = 5$. The red curves display the result for the MH-RAPS algorithm from Section 6.3. The blue, green, yellow and black curves show the results for the MH-NP approach with $\gamma = 2, 3, 4,$ and $5$, respectively.

Table 7.4: GNSS-INS Horizontal Performance Statistics

<table>
<thead>
<tr>
<th>Methods</th>
<th>$\mu = 2$</th>
<th>Mean of error(m)</th>
<th>Std. of error(m)</th>
<th>Submeter accuracy(%)</th>
<th>Max error(m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MH-NP $\gamma = 5$</td>
<td></td>
<td>0.61</td>
<td>0.40</td>
<td>84</td>
<td>1.98</td>
</tr>
<tr>
<td>MH-NP $\gamma = 4$</td>
<td></td>
<td>0.59</td>
<td>0.38</td>
<td>87</td>
<td>1.96</td>
</tr>
<tr>
<td>MH-NP $\gamma = 3$</td>
<td></td>
<td>0.53</td>
<td>0.28</td>
<td>92</td>
<td>1.50</td>
</tr>
<tr>
<td>MH-NP $\gamma = 2$</td>
<td></td>
<td>0.48</td>
<td>0.24</td>
<td>96</td>
<td>1.24</td>
</tr>
<tr>
<td>MH-RAPS</td>
<td></td>
<td>0.31</td>
<td>0.11</td>
<td>100</td>
<td>0.61</td>
</tr>
</tbody>
</table>
Figure 7.8: Error, risk and information diversity (GDOP) versus $L$ with $\mu = 2$. The red, blue, and green curves display the results for the MH-RAPS algorithm from Section 6.3 $L=2$, 4, and 8, respectively.

Table 7.5: MH-RAPS performance Statistics versus $L$ with $\mu = 2$ for GNSS-INS.

<table>
<thead>
<tr>
<th>Methods</th>
<th>$\mu = 2$</th>
<th>Mean of error (m)</th>
<th>Std. of error (m)</th>
<th>Sub-meter accuracy (%)</th>
<th>Maximum $|b_k|/|p|$</th>
<th>Error (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$L = 2$</td>
<td></td>
<td>0.33</td>
<td>0.12</td>
<td>100</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>$L = 4$</td>
<td></td>
<td>0.31</td>
<td>0.11</td>
<td>100</td>
<td>0.61</td>
<td>0.49</td>
</tr>
<tr>
<td>$L = 8$</td>
<td></td>
<td>0.26</td>
<td>0.09</td>
<td>100</td>
<td>0.57</td>
<td>0.37</td>
</tr>
</tbody>
</table>
Chapter 8

Conclusion and Future Works

This chapter concludes this dissertation and points out potential future works.

8.1 Conclusions

This dissertation considers the challenge of preventing outlier measurements from affecting the accuracy and reliability of state estimation. The major contribution of this dissertation 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. This dissertation also provided the following contributions to robust state estimation in presence of outliers:

- Instead of the traditional outlier accommodation methods which are threshold based, an optimization based state estimation method is proposed which is general, but tractable solution for online applications.

- Proposed RAPS optimization solutions for both linear and nonlinear systems are provided. The nonbinary RAPS solution provides measurement weights that trade-off risk
versus information. The binary solution determines a set of selected measurements which satisfies the accuracy condition while minimizing risk.

- RAPS solution is extended to moving horizon state estimation for both linear and nonlinear applications which gives an opportunity to reevaluate past outlier decisions within time window in light of new measurement.

- RAPS is applied to vehicle state estimation using GNSS experimental data from a moving platform using a PVA model for linear applications and GNSS data along with inertial measurement data for nonlinear applications. Outliers where artificially generated and added to randomly chosen satellites at each time instant to enable direct comparison between all approaches. RAPS solved the GPS vehicle state estimation problem while demonstrating robustness to outliers and significantly improving positioning performance in compare with Neyman-Pearson (extended) Kalman filter approach.

### 8.2 Publication list


8.3 Future Works

The following three directions may be interesting for further investigations:

• The nonbinary RAPS solution presented herein, uses multi-convex programming in which user-defined proximal values are required. Solution algorithms with lower computational cost and without user-defined parameters are open for future work.

• This dissertation has not considered the case where no feasible solution exists or the case where the best feasible solutions have very high risk. Adjusting the length of time window \( L \) in RAPS solution when for given window length either no feasible solution exists or the best feasible solution has high risk, are of interest.

• The ideas we presented herein can be beneficially extended in various directions where outliers have to be suspected, like SLAM problems, image processing, machine learning and etc. Extensions to real-time implementation in other applications and comparison with alternative methods proposed by other authors are of interest for future work.
Bibliography


[13] Luca Carlone, Andrea Censi, and Frank Dellaert. Selecting good measurements via 1
relaxation: A convex approach for robust estimation over graphs. IEEE IROS, pages 2667–
2674, 2014.

ambiguity resolution in multiepoch GPS/INS: a common-position-shift approach. IEEE

navigation with long baseline transceivers: A near-real-time approach. IEEE Transactions

[16] Sundeep Prabhakar Chepuri and Geert Leus. Sparsity-promoting sensor selection for non-

[17] EYEY Chow and AS Willsky. Analytical redundancy and the design of robust failure

[18] Frank Dellaert and Michael Kaess. Square Root SAM: Simultaneous localization and
mapping via square root information smoothing. The International Journal of Robotics

for nonlinear MPC and moving horizon estimation. pages 391–417. Springer Nonlinear
Model Predictive Control, 2009.

[20] M.W.M. Gamini Dissanayake, Paul Newman, Steve Clark, Hugh F Durrant-Whyte, and
Michael Csorba. A solution to the simultaneous localization and map building (SLAM)


[26] Paul M Frank and Xianchun Ding. Survey of robust residual generation and evaluation
methods in observer-based fault detection systems. J. of Process Control, 7:403–424,
1997.


[28] Matthew C Graham, Jonathan P How, and Donald E Gustafson. Robust incremental SLAM


