The Dissertation Committee for Rahul Moghe
certifies that this is the approved version of the following dissertation:

Adaptive Algorithms for Identification of Symmetric and Positive Definite Matrices

Committee:

Maruthi R. Akella, Supervisor

Renato Zanetti, Co-Supervisor

Takashi Tanaka

Efsthatios Bakolas

Christopher D'Souza

Adaptive Algorithms for Identification of Symmetric and Positive Definite Matrices

by

Rahul Moghe

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT AUSTIN

May 2021

To my parents Medha and Milind, and my brother Rajeev.

Acknowledgments

I would like to thank my advisors Dr. Maruthi Akella and Dr. Renato Zanetti for the support and guidance that they have provided me. I consider myself fortunate to have had them as my advisors. They have given me the various opportunities that have set me up for success during and after my PhD.

I would like to thank my committee members Dr. Efsthatios Bakolas and Dr. Takashi Tanaka. Their courses have helped me immensely during the earlier years of my PhD. Thank you as well to Dr. Christopher D'Souza of NASA Johnson Space Center.

I would like to acknowledge the graduate students with whom I have shared most of my time. Much of my journey has overlapped with Dr. Soovadeep Bakshi of Rivian, Dr. Marcelino Almeida of Pensa Systems and Dr. Kirsten Tuggle of Draper Labs. I would like to thank Arjun, Scott, and the rest of the graduate students with whom I have had many conversations and experiences that have positively impacted my PhD experience.

I would also like to acknowledge the Aerospace and Mechanical Engineering faculty at UT Austin as well as the Systems and Control faculty at IIT Bombay for their teaching excellency and the motivation and encouragement they provided that helped me throughout this journey.

Finally, I would like to thank my dear family Aai, Baba, Rajeev and Priyanka

for their continued support during my journey. I owe my accomplishments to you all.

Adaptive Algorithms for Identification of Symmetric and Positive Definite Matrices

Publication No. ____________________

Rahul Moghe, Ph.D.
The University of Texas at Austin, 2021

Supervisor: Maruthi R. Akella
Co-Supervisor: Renato Zanetti

Adaptive estimation and identification algorithms involving unknown symmetric and positive definite (SPD) matrix-valued parameters are ubiquitous in engineering applications. The problem of estimating the noise covariance matrices in estimation algorithms is considered first. An adaptive Kalman filter to estimate the noise covariance matrix of the noises entering a linear time invariant system is introduced first. The convergence of the estimates as well as the states is guaranteed with mild assumptions on the system. Conditions of estimability of the noise covariance matrix are discussed. The generalization of the adaptive Kalman fitler to the linear time varying case is introduced next. To maintain positive definiteness of the noise covariance estimates a differential geometric approach is adopted. The geometry of the manifold of SPD matrices is used to develop a Riemannian optimization based adaptive Kalman filter that ensure positive definiteness of the

estimate. The convergence of the Riemannian optimization-based estimate and the adaptive Kalman filter is established under mild conditions of uniform observability and uniform controllability of the system. An adaptive control problem with an unknown SPD matrix is considered next. A novel projection scheme is introduced that ensures that the estimates of the unknown SPD matrix are SPD. Adaptive update laws for identifying the SPD matrix are also presented. The adaptive control laws are shown to globally stabilize systems in problems such as the adaptive angular velocity tracking, adaptive attitude control, and the adaptive trajectory tracking of robotic manipulators with parameter uncertainties within the generalized mass matrix. In general, such a method can be applied to estimation of symmetric matrices with eigenvalue constraints.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Motivation for SPD matrices

The set of symmetric and positive definite (SPD) matrices present a mathematically rich structure and are useful in various engineering applications. They appear as covariance matrices in multivariate statistics and estimation theory [73]. In the computer vision applications, the covariance descriptor which is a popular technique to depict the features in an image are also SPD matrices [148]. SPD matrices are used to represent principal diffusion directions in Diffusion Tensors Imaging [66] and to model the shape of objects in extended object tracking [83]. The inertial properties of a body in motion can also be represented at SPD matrices. The non rotational part of a deformation gradient tensor in continuum physics is modeled as a SPD matrix [93]. The compliance and the stiffness matrix of a robotic mechanism must be also positive definite [66]. Positive definite Lyapunov functions have also been used to estimate the largest basins of attraction of a dynamical system [144]. Estimating such Lyapunov functions formed using polynomials often involve identifying appropriate SPD matrices. All the above SPD matrices are critical to the behavior and properties of the system which they are a part of.

Estimation of SPD matrix parameters that drive physical systems is an important aspect of system identification. For example, estimating covariance matrices of a random variable from its samples is a well studied problem. Such an estimation problem is especially challenging when the number of samples is much less than the dimension of the covariance matrix [85, 61, 39]. The inertial properties of an body are estimated by observing its motion for improving the control performance. Such procedures involve extensive experimentation with the system. As more complex systems were designed, isolated experimentation became infeasible and *in situ* methods were developed. As a result, adaptive techniques for system identification were popularized which enable on-line identification of parameters. Since, the motion of the system depends heavily on the parameters to be estimated, adaptive techniques demanded stronger stability properties of the system under incorrect knowledge of the system parameters. Specifically for SPD matrix valued parameters, such guarantees are challenging because of the nonlinear nature of the eigenvalue constraint.

SPD matrices have been vital for various engineering application. In Kalman filtering, apart from estimating the uncertainty in state estimation, identifying the unknown noise covariance matrices that affect the dynamics and the observations has been well studied in the system identification literature [101]. Matrices representing the inertia properties of a body are often estimated to obtain an accurate model of the system. For example, in space applications, the measurements of the angular velocity of an unidentified tumbling rigid body is used to estimate its inertia, which in turn helps understanding its mass and density properties [2, 37]. Inertia properties of the robotic manipulator handling an object cannot always be modeled since the object being manipulated may have different shapes, sizes and mass. In such cases, the inertial properties of the combined manipulator object system has to be estimated in order to grasp and manipulate the object effectively [139, 133, 46]. In extended object tracking, the shape of an object is modeled as an ellipse represented by a SPD matrix [83]. The shape of the object is estimated along with its position to track multiple objects in the environment. In object detection and tracking, the covariance descriptor of the object is used as a measurement to estimate its position and velocity. In this scenario, a SPD matrix is processed as a measurement to track the object [148]. In all of the above examples, the estimate of the SPD parameter must also be SPD for it to be meaningful to the system. For example, having a negative eigenvalue in an inertia matrix or in the covariance of a Gaussian random variable defies the basic physics of the system. Hence, estimating SPD matrices need to be treated with special care.

1.2 Identification in Estimation

This dissertation addresses the problems of identifying SPD matrices while focusing on two applications. First, the problem of covariance estimation in filtering is considered, wherein the noise covariance matrices of the noises entering the system and the measurement are unknown. The Kalman filter is optimal in the mean squared sense for a linear system with additive white Gaussian process and measurement noises [79, 78]. However, it is no longer optimal when the noise covariance matrices are unknown or some of the elements are uncertain. Additionally, uncertain or unknown noise covariance matrices have been shown to cause filter divergence [128, 68, 94, 121, 127, 164]. These challenges associated with filter divergence motivate the development of adaptive filtering algorithms that simultaneously estimate the system states along with the covariance matrices. Although several adaptive formulations have found their application in many individual settings, rigorous theoretical foundations are usually absent and if present, are accompanied by restrictive assumptions such as steady state conditions, invertibility of the state transition and observation matrix or availability of the true state estimate.

An adaptive covariance matching technique for detectable linear systems based on correlation methods to recursively estimate the state vector and the noise covariance matrices are established in this dissertation. The salient features of the novel technique are given below.

- Guarantees on stability of the filter as well as the covariance matrix estimates,
- Convergence of the estimates to their true values, and
- Assurance that the covariance matrix estimates are physically meaningful, i.e., remain SPD.

1.3 Identification in Control

The second problem addressed in this dissertation is that of designing adaptive controllers in presence of an unknown SPD matrix parameters. Such problems appear in attitude control problems in applications such as space autonomy, robotic manipulators and underwater robots wherein the inertial properties of the object is unknown. Adaptive controllers have been designed to estimate the unknown matrices and guarantee stability of the controller through Lyapunov direct theorems [111]. A common assumption in these adaptive formulations is that the eigenvalues of the SPD matrix parameters are assumed to be bounded. In some cases, the bounds themselves are used to calculate the adaptive control gains. However, the update laws obtained from the Lyapunov analysis do not guarantee positive definiteness of the parameter estimates. Parameter constraints are often enforced by projecting the adaptive update law. Although several projection laws have been developed for vector parameters, they do not translate trivially to matrices with eigenvalue constraints. In such scenarios, it is critical to ensure positive definiteness of the estimated parameters either through new update laws or new projection schemes. A different projection scheme to enforce eigenvalue constraint would significantly improve the state of the art identification algorithms since it would be applicable to a larger class of problems. These issues motivate a new formulation that has the following features.

- Ensures that the estimate of the SPD matrix valued parameter are indeed SPD,
- Guarantees stability of the control law in presence of the new update law, and
- Ensures that the update law is domain independent and applicable to a wide variety of problems.

Most adaptive identification techniques treat the SPD matrix as a vector of its unique elements. In this way, symmetry of the estimate is readily ensured. However, ensuring positive definiteness or in a general manner of speaking, enforcing eigenvalue constraints are more challenging due to their nonlinear nature. Ensuring such constraints would not only provides a meaningful estimate, but also improve the performance of existing adaptive techniques.

1.4 Geometric structure of SPD matrices

SPD matrices follow a rich geometric structure. They form a cone in the vector space and also constitute a Riemannian manifold with a tangent identified with the set of symmetric matrices. Because of its geometric structure, geometric methods have been developed that respect the geometry of SPD matrices rather than ignoring it. Such methods have already shown promise for other manifolds over their Euclidean counterparts [30, 1]. Since ensuring positive definiteness is critical to real world applications, this dissertation leverages differential geometric methods to identify SPD matrix parameters for the problems mentioned above.

1.5 Contributions

In summary, the contributions of this dissertation are listed below.

- Derivation of an adaptive Kalman filter for a detectable linear system that estimates the process and measurement noise covariance matrices along with the states while ensuring stability and convergence is presented.
- A novel projection scheme is introduced that enforces eigenvalue constraints of matrix valued parameters in adaptive control formulation. The projection method is also shown to build stable controllers for a wide class of systems.

This applicability of the projection method is demonstrated for the attitude control problem and the trajectory tracking problem for robotic manipulators with unknown inertial properties.

- Riemannian optimization based identification methods are then formulated to obtain estimates for the problems of covariance estimation. Their global optimality and convergence is also discussed.

The rest of the dissertation is organized as follows. Chapter 2 performs a literature review of identification in estimation and control problems, as well as the differential geometry of SPD matrices. Chapter 3 presents the Adaptive Kalman filter to estimate the noise covariance matrices of a detectable linear time invariant system. The filter derivation and convergence results are proved in this chapter. The estimability of the noise covariance matrices is also discussed and certain limitations of this work is presented. In light of the limitations of the adaptive Kalman filter, a generalization to the linear time-varying case if presented in Chapter 5. A Riemannian optimization method to guarantee symmetry and positive definiteness of the noise covariance estimate is introduced. In particular, a least squares cost function of the noise covariance matrix elements is minimized using a Riemannian Trust-Region method. Global stability of the optimal solution is proved with assumptions of sufficient excitation of the system. The problem identification of SPD matrices in the context of adaptive control problems with an unknown SPD matrix valued parameter is presented in Chapter 4. Saliently, a projection scheme is introduced to bound the eigenvalues of the unknown SPD matrix parameter driving the system. Consequently, a positive definite function that is included in the Lyapunov analysis of systems is introduced and its boundedness properties are proved. Finally Chapter 6 summarizes the dissertation and discusses potential directions of future research.

Chapter 2

Technical Background

This chapter reviews past work on identification of SPD matrices in the context of covariance estimation and control problems. The discussion begins with the problem of adaptive Kalman filters.

2.1 Adaptive Covariance Identification

Covariance estimation from the point of view of statistical theory has been a topic of active research. In this field, the covariance matrix of a random variable is estimated from its samples. This problem is especially challenging when the dimension of the samples is much larger than the number of sample available and appear in the field of economics and finance [86, 57]. The focus of this dissertation is on adaptive techniques for identification wherein the covariance matrices are estimated on-line and the covariance estimate is used to estimate the states of the system.

2.1.1 Background on Adaptive Estimation

Kalman filtering is an optimal estimation algorithm for linear systems when driven by known inputs and white Gaussian noises. The optimality is lost when a complete knowledge of (a) the parameters that define the system model, (b) the output relations, and (c) the statistical description of uncertainties, is missing [99]. In many practical applications, the parameters of the model may be uncertain and estimation algorithms have to adjust to account for these uncertainties [20]. Various adaptive estimation algorithms for linear systems have hence been developed to that end. Even in the case when the system model is nonlinear, the algorithms for linear systems can be applied via linearization. Uncertainties in a linear system model can appear in various forms, namely, the state transition matrix and the observation matrix may be uncertain, the control inputs, if they exist, or their gains may be unknown, and finally the noise covariance matrices for both the process and measurement noise may be unknown.

An example of the case of unknown inputs is when an unknown object is to be tracked. In this scenario, an alternative to estimating the control input is to label the uncertainties in the input or its gain as an uncertainty in the system model. Consequently, this transpires as uncertainties in the process noise covariance and the covariance can then estimated or adjusted to account for the unknown control inputs. An adaptive estimation technique is to augment the state of the system with the unknown parameters. Such an approach requires an accurate characterization of the prior distribution of the parameter given that it is unknown.

Another adaptive technique is the multiple model estimation wherein the system is assumed to operate under finite number of modes [16, 17]. The parameter is assumed to belong to a set of finite values and each mode corresponds to a value from that set. A probability assigned to each model is then updated based on the measurements to choose the most probable value for estimation. Consequently, the parameter is assumed to belong to a discrete probability mass distribution which is updated using the system measurements. The unknown state transition and measurement matrices are estimated can be estimated using MLE techniques that do not require a priori information about the parameters [81, 98].

Finally, the uncertainties in the noise statistics are characterized by unknown or uncertain process and measurement noise covariance matrices. Various estimators have been derived [5, 23, 100, 101, 33, 122]. The methods to sequentially adjust the noise covariance matrices is arguably more robust due to the following reasons:

- The prior information for uncertain control inputs and system matrices may not be readily available
- All kinds of uncertainties in the models discussed above can be aggregated into uncertainties in the noise covariance matrices
- For nonlinear system, the above two issue are intensified by the presence of non-linearities

The study of adaptive estimation of the noise covariance matrices is arguably the most important and applicable form of adaptive filtering and hence is the topic of Chapter 3 of this dissertation.

An early study of adaptive algorithms to estimate noise covariance matrices is summarized in [101]. Broadly speaking, there are four approaches to adaptive filtering, namely, Bayesian methods, maximum likelihood estimation, correlation methods, and covariance matching techniques. For LTI systems, steady state convergence of an adaptive filter using the innovation sequence for estimation of the noise covariance matrix was originally reported in [100]. However, certain assumptions central to [100] such as steady-state conditions, observability, and invertibility of the state transition matrix makes the filter restrictive. Moreover, technical insufficiencies concerning the whiteness condition in [100] are clearly pointed out in [128, 4]. Specific differences between [100] and this work are described in further detail in the upcoming chapters. An approach to estimate only the process noise covariance using a backward shift operator is presented in [152]. This method uses a left co-prime factorization which is either not guaranteed to exist or can be difficult to analytically calculate. H_∞ filtering is used to estimate the statistics of the noises in [156]. This approach can be classified as offline since it uses the H_∞ filter estimate at future times to calculate the estimate of covariance matrices. The approach given in [56] uses linear regression to derive a robust Kalman filter. The adaptive filter presented in [167] does not provide rigorous convergence guarantees for their filter and their estimates.

A recent survey of covariance estimation results for LTI systems using correlation methods can be found in [55]. Standing out from among the various methods surveyed in [55] is the measurement averaging correlation method (MACM) which forms a stacked measurement model wherein the observability matrix is referred to as the observation matrix. The work in this dissertation shares some of the algebraic features of the MACM approach wherein a stacked measurement model is constructed. Another approach to estimate the process noise covariance matrix of LTI systems assuming a left-invertible observation matrix is presented in [58]. In this case, a linear stationary time series is formed by inverting the observation matrix and the covariance is estimated by squaring the sequence.

A covariance estimation result is presented in [53, 169, 90] for linear time-varying (LTV) systems using the linear time series construction. Specifically, a stacked measurement model is formed using an arbitrary N measurements to estimate the N-step predicted measurement and the measurement prediction error is shown to form a linear time series. The sample auto covariance of the linear time series is used to calculate estimates of the covariance matrices which is shown to be consistent assuming a full rank condition. Although the focus of this work is entirely on covariance estimation, no convergence guarantees are provided for the state estimation part. Adaptive filtering operations require stability guarantees on both the state and the covariance estimates which is missing from the literature.

2.1.2 The Covariance Matching Kalman Filter

Compared to previous approaches the developments in Chapter 3 takes a significant next step – that of simultaneously estimating the state of the system. A major contribution of this work is a proof of convergence for both the state as well as the covariance estimator. For purposes of calculating the covariance estimate using the covariance of the time series, full column rank of a certain coefficient matrix in assumed in [53]. The filter developed in Chapter 3 adopts concepts based on correlation methods, specifically in terms of forming a linear time series using the measurement sequence. Certain ideas from covariance matching techniques are also embraced by way of formally enforcing covariance consistency. Moreover, this result provides the explicit characterization of necessary conditions for estimability of the process and measurement noise covariance matrices.

2.2 Differential Geometry of SPD matrices

Ensuring symmetry and positive definiteness dates back to the time when the Kalman filter was introduced. In 1963, James Potter formulated the square-root filter to improve numerical properties of the update equation for the error covariance matrix of the extended Kalman filter used for the Apollo mission [22]. Consequently, various kinds of decompositions like the Cholesky decomposition [64, 151], the spectral decomposition [115], and the UDU factorization [27] were introduced for their numerical advantages. A byproduct of some of these decompositions was that symmetry and positive definiteness was automatically ensured.

A commonly adopted strategy to handle constrains such as SPD is to equip them with a differential geometric structure. Differential geometric tools help perform calculus on a constrained space as if it had a vector space-like structure and allow for developing optimal estimators similar to their vector space counterparts. For problems such as that of attitude estimation, the invariant Kalman filter utilizes the matrix lie group structure of rotation matrices [21]. The complimentary filter developed in [95] performs filtering for attitude estimation. In this method, the Lie Group structure of the $SO(3)$ rotation matrices was used to develop novel estimators that satisfy the constraint by virtue of the construction. Other methods for matrix lie groups have also been presented [168].

In this dissertation is the differential geometric structure of SPD matrices. Covariance and inertia, both SPD matrices, have a rich geometric and non-Euclidean structure [25]. Firstly, the set of $n \times n$ SPD matrices form a positive half cone in set of $n \times n$ matrices. The neighborhood of every SPD matrix can be viewed as being almost Euclidean. Mathematically, there exists a bijection between the neighborhood of every SPD matrix and the set of symmetric matrices. Sets containing such locally Euclidean structure are said to form manifolds embedded in vector spaces [29, 69]. Furthermore, the set of SPD matrices form a differentiable manifold with a Riemannian structure when equipped with an appropriate metric [25, Chapter 6].

The differential geometry was first introduced to characterize a geometric mean of a given set of SPD matrices [105]. Such a mean is the generalization of the geometric mean of positive numbers to SPD matrices and was called the Riemannian barycenter of a set of SPD matrices. As the arithmetic mean minimizes the sum of Euclidean distance between the mean and the sample set, the Riemannian barycenter minimizes a different distance function. The existence of a unique minimizer was shown by Karcher [80] for Riemannian manifolds with non-positive sectional curvature. Furthermore, the geometric mean was shown to satisfy additional properties relevant to the set of SPD matrices, such as invariance to inversion. The concept Riemannian barycenter and its uniqueness for SPD manifold was quickly adopted for applications such as averaging and classification of SPD tensors in elasticity theory [107, 106], diffusion tensor imaging [14], and computer vision problems [149]. In spite of an elegant formulation, till date, there is no closed-form solution for the Riemannian barycenter of a finite number of SPD matrices. Closed-form solutions have been found for specific case of calculating the mean of 2 SPD matrices. As a result, the all the applications of the geometric mean use iterative methods in their analysis.

In the context of estimation theory, the Riemannian framework was applied to covariance estimation problems, specifically, for estimation of covariance matrices from samples [162, 18, 116, 163]. Various definitions of probability distributions on the SPD manifold were introduced using several new metrics [123, 124]. Iterative algorithms for statistical inference on manifold valued data was then introduced [117, 118].

Kalman filtering algorithms for manifold valued states and observations were also introduced [102, 67, 31, 84, 34, 125, 52, 50, 51, 112, 10, 136, 137, 135]. Specifically for filtering on the SPD manifold, Kalman filter like algorithms have been introduced. The linear filter introduced in [150] provides a closed form update law based on the Affine-Invariant metric. A simplified linear observation model was used wherein, the full SPD state measurement was available. Such a method does not extend trivially to other linear models. The work in [41] provides a dynamic system model to track SPD matrices for appearance tracking. However, the problem of propagating and updating the estimates is approached with an optimization based framework. A tractable state-space model was introduced once again for full SPD state measurement and closed form update step was provided [165]. Closed form expressions for optimal filtering were also found for a specific class of models involving square root of the SPD state [157]. A Wishart prior and the Jensen-Bregman LogDet divergence was minimized to attain a closed form expression. Some other closed-form expressions have also been found for simplified models in [158, 126, 40]. However, no known closed-form results exist for estimating a linear model appearing in adaptive estimation that is found in Chapter 3.

For control problems, optimization based methods have utilized the differential geometric structure of the SPD inertia matrix [88, 89]. For discrete-time formulation of a continuous-time linear systems with a symmetric state transition matrix contains the exponential of a symmetric matrix, which is a SPD matrix. Such systems models are common in applications such as electrical network systems, multiagent network systems, and temperature dynamics. An optimization method was formulated for this problem using the Riemannian geometric structure of SPD matrices [130]. The physical consistency of the inertia matrix in robotic manipulators involves maintaining SPD properties of its estimate. An optimization approach on manifolds to solve the inertia identification problem was presented in various studies [147, 161, 161, 138]. Most recently, regularization methods have emerged to ensure the physical consistency of the system [28]. Optimization methods to solve SPD matrix valued equations are developed to obtain SPD estimates in Chapter 5 for the above identification problems of noise covariance estimation. The concept of natural gradient or Riemannian gradient for cost functions is used to obtain optimal solutions to the problem of identifying SPD matrices.

2.3 Adaptive Identification of SPD matrices

Identification of SPD matrices also appear in adaptive control problems. Various control theoretic applications such as rigid body control depends on a SPD matrix parameter. The autonomous control of a rigid body is extensively studied in the literature for applications pertaining to the control of spacecrafts, robotic manipulators, and underwater robots [160, 49, 166, 155, 91, 60]. Despite the non-linear nature of the system and the presence of uncertain parameters governing the system, linear feedback controllers with appropriate parameter adaptations were shown to provide global tracking of the states. Moreover, stability properties of the controller, proved through assumptions of bounded signals and parameters, made the controller highly applicable on physical systems. The inertial properties on the system, realized through the inertia matrix, is an important system parameter that drives the motion of the rigid body. Even though the inertia matrix is unknown, inertia-free adaptive controllers have been shown to exhibit global stability properties. Chapter 4 presents a novel adaptation technique, that is applicable to a fairly large set of problems is presented. Such a method, in the worst case, is shown to perform as well as the adaptive controller without projection.

For the attitude control problem, several inertia-free adaptive controllers have been developed in the literature. The work by Wen and Delgado [160] presented an adaptive controller with inertia matrix adaptation. An adaptive controller to realize linear closed loop dynamics in a 3-parameter representation with unknown inertia was presented in Ref. [76]. The work in [2] uses a similar PD+ controller with inertia adaptation with a quaternion representation and derives conditions for estimability of the inertia matrix. Adaptive controller for angular velocity tracking and inertia identification has been presented in [37]. In presence of disturbances an almost globally stable inertia-free adaptive controller has also been developed [129]. Certain non-rigid inertia models parametrized using constant unknown parameters were presented in [145]. An adaptive controller was developed to estimate the constant unknowns of the time varying model in this case. An adaptive controller design with attitude measurements has also been developed based on the certainty equivalence principle [103]. More recent work departs from the certainty equivalence assumption by unifying the convergence analysis for the observer and controller [47, 24]. Adaptive control of robotic manipulators also belongs to this class of problems. Some of the early results on adaptive control of robotic manipulators is given in [46, 133, 104]. Since then, various adaptive control laws for problems associated with robotics manipulators have been presented with uncertain inertial properties of the manipulators [153, 146, 62].

A common trait of the above control problems is that the unknown inertia matrix appears linearly in the dynamical equations of motion. In order to estimate the inertia matrix, a vector of unknown elements of the inertia matrix is formed. Consequently, adaptive update laws and control laws chosen through the Lyapunov analysis guarantee stability of the system. The choice of controller gains that make the derivative of the Lyapunov function negative depend on the bounds of the uncertain parameters. In the event that the unknown parameters are constrained, projection schemes are used to ensure that the estimates satisfy the bounds. Various projection methods have been developed in the literature to enforce norm constraints on the estimated elements. The adaptation presented by Bakker et. al. [19] for linear time-invariant systems bounds the scalar parameter estimate to lie between specified bounds. This adaptation was generalized to be applicable to a more general class of linear systems [9]. The work by Akella et. al. [3] extends the above parameter adaptation by providing a smooth parameter projection method for scalars. The projection method presented in [47] provides a differentiable projection operator for vectors to satisfy norm constraints. Such a projection was modified to a form that is sufficiently smooth [35]. For matrix valued parameters, however, the techniques mentioned above can not impose eigenvalue constraints. Even though norm constraints on the unique elements of a symmetric matrix indirectly enforce some upper bound on the maximum eigenvalue of the matrix, that bound can be explicitly chosen. Furthermore, lower bounding the eigenvalues is not possible via such methodology. For the adaptive control of robotic manipulators, projection methods have been introduced to maintain the uniform positive definiteness of the inertia matrix. For example, the projection method given in Refs. [47, 35] constrains the norm of the parameters.

A related problem in adaptive control of robotics manipulators is the problem of ensuring physical consistency of the inertial parameters. This problem reduces to maintaining positive definiteness of a $\mathbb{R}^{4\times 4}$ matrix containing the inertia matrix and mass parameters of the links of the manipulator [147]. The work in Ref. [154] projects the parameter update law such that the determinant of every primary sub matrix of the inertia matrix is constrained to be positive. Although a eigenvalue constraints could theoretically be imposed by including a diagonal matrix with eigenvalue bounds on the diagonal in the determinant function, this method is restrictive because an iterative scheme is necessary to perform this projection. A recent work adopted a differential geometric approach to maintain the physical consistency of the parameters [87]. A modified Bregman Divergence based function was used as a Lyapunov function to maintain physical consistency of the inertia matrix estimates. Although positive definiteness of the inertia matrix was maintained, such an algorithm can not be trivially extended to handle eigenvalue constraints. In light of these limitations, a novel projection method is presented in Chapter 4 that allows for explicit bounds on the eigenvalues of a symmetric matrix valued parameter. Specifically, the cases considered in this Chapter 4 when the eigenvalues of the symmetric matrix valued parameter are (i) lower bounded, (ii) upper bounded, and (iii) bounded above and below.

Constraining the eigenvalues of a symmetric matrix is performed using the eigen decomposition of the estimate of the matrix. The computational complexity of the eigen decomposition scales cubically with the dimension of the matrix. For applications involving a lower dimensional matrix, this may not be an issue. However, eigen decomposition may prove to be computationally expensive for applications such as the n-link manipulator problem in robotics. For this reason, a methodology to minimize the computation by directly updating the eigenvalues and the eigenvectors is provided in this section. This follows from the seminal work by Kato [82] on perturbation theory of linear operators. Given the update law for a matrix, the eigen decomposition of the matrix can be directly updated in most scenarios. Such an update leads to huge computational savings for real world applications.

Chapter 3

Adaptive Covariance Estimation in Kalman Filtering

This chapter presents an adaptive Kalman filter formulation for estimating the states as well as the noise covariance matrices online and guarantee convergence [1].

3.1 Problem Statement

Consider the LTI system of the form

$$\left.\begin{aligned} x_{k+1} &= \tilde{F} x_k + \tilde{w}_k \\ y_k &= \tilde{H} x_k + v_k \end{aligned}\right\} \tag{3.1}$$

wherein $\tilde{w}_k$ and v_k are the zero mean white Gaussian process and measurement noises. The definition for detectability of a general linear system is provided below [8].

Definition 3.1.1. *The pair $[F,H]$ is uniformly detectable if there exists integers $s,t \geq 0$ and constants $d \in [0,1)$ and $b \in (0,\infty)$ such that for all $\zeta \in \mathbb{R}^n$ and integers k that make*

$$\|F^k \zeta\| \geq d\|\zeta\|$$

[1]The research presented in this chapter is performed by the author and is previously published in a peer-reviewed journal:
Moghe, R., Zanetti, R., & Akella, M. R. (2019). Adaptive kalman filter for detectable linear time-invariant systems. Journal of Guidance, Control, and Dynamics, 42(10), 2197-2205.

true, the following is true as well.

$$\zeta^T M_s \zeta \geq b \zeta^T \zeta$$

wherein M_s is defined by the sum

$$M_s = \sum_{i=0}^{s} F^{T^s} H^T H F^s \tag{3.2}$$

In what follows we make the following assumption on the system in Eq. (3.1).

Assumption 1. *The system given in Eq. (3.1) is uniformly detectable, i.e., the pair $(\tilde{F}, \tilde{H})$ is uniformly detectable.*

According to Assumption 1, there exists an invertible state transformation matrix W such that the system can be transformed into

$$\left.\begin{aligned} z_{k+1} &= \begin{bmatrix} F_{11} & \mathbf{0}_{l\times s} \\ F_{21} & F_{22} \end{bmatrix} \begin{bmatrix} z_k^O \\ z_k^{UO} \end{bmatrix} + W\tilde{w}_k \\ y_k &= \begin{bmatrix} H_1 & \mathbf{0}_{p\times s} \end{bmatrix} \begin{bmatrix} z_k^O \\ z_k^{UO} \end{bmatrix} + v_k \end{aligned}\right\} \tag{3.3}$$

wherein, the state and the transformed state at time t_k are respectively denoted by x_k and $z_k \in \mathbb{R}^n$. The integers l and s are the dimensions of the observable (^{O}) and unobservable (^{UO}) subspace of the state respectively such that $s + l = n$. Accordingly, the observable state $z_k^O \in \mathbb{R}^l$ and the unobservable state $z_k^{UO} \in \mathbb{R}^s$. The measurement at time t_k is given by $y_k \in \mathbb{R}^p$. The pair (F_{11}, H_1) is uniformly observable and F_{22} is stable (i.e. all eigenvalues within the unit sphere). The matrix W is such that $z_k = W x_k$ and define $w_k = W\tilde{w}_k = \begin{bmatrix} w_k^{O^T} & w_k^{UO^T} \end{bmatrix}^T$ to be the transformed process noise. The transformed state transition matrix is given by

$$F \triangleq W\tilde{F}W^{-1} = \begin{bmatrix} F_{11} & \mathbf{0}_{l\times s} \\ F_{21} & F_{22} \end{bmatrix} \tag{3.4}$$

and the observation matrix is defined as

$$H \triangleq \tilde{H}W^{-1} = \begin{bmatrix} H_1 & \mathbf{0}_{p\times s} \end{bmatrix}. \tag{3.5}$$

The process noise $w_k \in \mathbb{R}^n$ and the measurement noise $v_k \in \mathbb{R}^p$ are white Gaussian and uncorrelated with each other. The transformed state at time t_0 is denoted by z_0. Hence, $w_k \sim \mathcal{N}(\mathbf{0}_{n\times 1}, Q)$ and $v_k \sim \mathcal{N}(\mathbf{0}_{p\times 1}, R)$, wherein Q can be partitioned into observable and unobservable subspaces as $Q = \begin{bmatrix} Q^O & Q^{O,UO} \\ Q^{UO,O} & Q^{UO} \end{bmatrix}$.

The baseline Kalman filter equations for known Q and R matrices are given by [20]:

$$\left.\begin{aligned} \hat{z}^O_{k|k-1} &= F_{11}\hat{z}^O_{k-1|k-1} \\ \hat{z}^O_{k|k} &= \hat{z}^O_{k|k-1} + K_k(y_k - H_1\hat{z}^O_{k|k-1}) \\ P_{k|k-1} &= F_{11}P_{k-1|k-1}F_{11}^T + Q^O \\ K_k &= P_{k|k-1}H_1^T(H_1P_{k|k-1}H_1^T + R)^{-1} \\ P_{k|k} &= P_{k|k-1} - K_kH_1P_{k|k-1} \end{aligned}\right\} \tag{3.6}$$

wherein K_k is the Kalman gain, $P_{k|k-1}$ and $P_{k|k}$ are the state error covariance matrices after the prediction and update step respectively. The baseline Kalman filter equations contain the observable part of the Q matrix only. The unobservable state z_k^{UO} cannot be estimated as the system is not fully observable. The noise covariance matrices Q and R are constant for all time t_k. If the process and measurement noises are white Gaussian, the baseline Kalman filter in (3.6) is optimal in the mean squared error sense and the state error covariance $P_{k|k}$ converges to a steady-state value [7].

3.1.1 Problem Description

Given full knowledge of the system matrices (F_{11}, F_{21}, F_{22}, H_1, Q and R), the Kalman filter is the optimal estimator of the system given by (4.1). However, most practical applications approximate the values of Q and R. Given F_{11}, F_{21}, F_{22}, H_1, and measurements y_k, an adaptive algorithm to estimate both the state x_k and unknown elements of R and Q matrices is presented in this chapter. Certain restrictions on the number of elements that can be estimated as stated through the following assumptions.

3.1.2 Assumptions

The following assumptions are made.

Assumption 2. *The matrices forming the transformed state transition matrix, i.e., F_{11}, F_{21}, and F_{22}, and the observation matrix H_1 are assumed to be completely known. The measurement sequence y_k is also assumed to be accessible.*

Assumption 3. *It is assumed that the matrices F and H, the process noise covariance matrix $Q_k = Q \succ 0$, and the measurement covariance matrix $R_k = R \succ 0$ are constant with time.*

Assumption 4. *The pair $(\tilde{F}, \tilde{H})$ is assumed completely uniformly detectable, i.e., the pair (F_{11}, H_1) is uniformly observable and the matrix F_{22} has all its eigenvalues inside the unit circle in the complex plane. The pair $(F, Q^{\frac{1}{2}})$ is assumed to be stabilizable.*

Remark 3.1.1. *The preceding assumption ensures that the baseline Kalman filter given in (3.6) converges to a steady state [8].*

Assumption 5. *The pair $(F, Q^{\frac{1}{2}})$ has no unreachable nodes on the unit circle in the complex plane.*

Remark 3.1.2. *This assumption along with the others stated above ensure the existence of a stabilizing solution for the algebraic Riccati equation [128, 36].*

There is an additional assumption that is required for the proposed result to hold which is stated in the sequel.

3.2 Filter Derivation

The filter is derived in 3 subsections, namely, formulating modified stacked measurement model, forming a linear strictly stationary time series and finally, estimating the covariance to calculate the unknown elements.

3.2.1 Modified measurement model

For the n-dimensional discrete-time stochastic linear system given by (4.1), the observability matrix is defined by

$$\mathcal{O} = \begin{bmatrix} HF^{n-1} \\ HF^{n-2} \\ \vdots \\ HF \\ H \end{bmatrix} \tag{3.7}$$

Since the (F,H) pair is detectable, it follows from assumption 4 that $\mathcal{O}$ is column-rank deficient. However, since every system given by (4.1) can be transformed into one given by (3.3), the observability matrix for the transformed system is given by

$$\mathcal{O} = \begin{bmatrix} H_1 F_{11}^{n-1} & \mathbf{0}_{p \times s} \\ \vdots & \vdots \\ H_1 F_{11} & \mathbf{0}_{p \times s} \\ H_1 & \mathbf{0}_{p \times s} \end{bmatrix} = \begin{bmatrix} \mathcal{O}_1 & \mathbf{0}_{np \times s} \end{bmatrix} \tag{3.8}$$

where the $\mathcal{O}_1$ corresponds to the observability matrix of the pair (F_{11}, H_1). In this chapter, $\mathbf{0}_{a\times b}$ is matrix of zeros with size $a \times b$ for some positive integers a and b. Without loss of generality

$$\exists\, m \in \mathbb{N},\ 1 \leq m \leq n : M_o = \begin{bmatrix} H_1 F_{11}^{m-1} \\ \vdots \\ H_1 F_{11} \\ H_1 \end{bmatrix} \ :\ rank(M_o) = l \tag{3.9}$$

wherein $M_o \ \in \mathbb{R}^{mp\times l}$. This follows from the definition of detectability for LTI systems. The value of m corresponds to the buffer size, i.e., the number of past measurements that are stored in memory at each time step. In order to minimize the memory storage, a smallest such m which satisfies the rank condition. However, one may choose a larger value of m so that the matrix M_0 is well conditioned. Using this result, a stacked measurement model is formulated by stacking precisely m measurements,

$$z_{k+1} = F z_k + w_k \tag{3.10}$$

$$z_{k+2} = F^2 z_k + F w_k + w_{k+1} \tag{3.11}$$

$$z_{k+i} = F^i z_k + \sum_{j=0}^{j=i-1} F^j w_{k+i-1-j} \tag{3.12}$$

Now, the measurement equations for the corresponding times are given by

$$y_k = H z_k + v_k \tag{3.13}$$

$$y_{k+1} = HF z_k + H w_k + v_{k+1} \tag{3.14}$$

$$y_{k+2} = HF^2 z_k + HF w_k + H w_{k+1} + v_{k+2} \tag{3.15}$$

$$y_{k+i} = HF^i z_k + \sum_{j=0}^{j=i-1} HF^{k+i-1-j} w_{k+j} + v_{k+i} \tag{3.16}$$

wherein $i = 0, 1, \ldots, m-1$. Coagulating all the equations and the measurements for m time steps

$$\underbrace{\begin{bmatrix} y_{k+m-1} \\ y_{k+m-2} \\ \vdots \\ y_k \end{bmatrix}}_{\triangleq \mathcal{Y}_k} = \underbrace{\begin{bmatrix} HF^{m-1} \\ HF^{m-2} \\ \vdots \\ H \end{bmatrix}}_{\mathcal{O}} z_k + \begin{bmatrix} H & HF & HF^2 & \cdots & HF^{m-2} \\ \mathbf{0}_{p\times n} & H & HF & \cdots & HF^{m-3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & H \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} \end{bmatrix} \begin{bmatrix} w_{k+m-2} \\ w_{k+m-3} \\ \vdots \\ w_k \end{bmatrix} + \underbrace{\begin{bmatrix} v_{k+m-1} \\ v_{k+m-2} \\ \vdots \\ v_k \end{bmatrix}}_{\triangleq V_k} \tag{3.17}$$

Rewriting (3.17) using the detectability definition and eliminating the unobservable states

$$\mathcal{Y}_k = \underbrace{\begin{bmatrix} H_1F_{11}^{m-1} & \mathbf{0}_{p\times s} \\ H_1F_{11}^{m-2} & \mathbf{0}_{p\times s} \\ \vdots & \mathbf{0}_{p\times s} \\ H_1 & \mathbf{0}_{p\times s} \end{bmatrix}}_{\mathcal{O}} \begin{bmatrix} z_k^O \\ z_k^{UO} \end{bmatrix} +$$

$$+ \begin{bmatrix} \begin{bmatrix} H_1 & \mathbf{0}_{p\times s} \end{bmatrix} & \begin{bmatrix} H_1F_{11} & \mathbf{0}_{p\times s} \end{bmatrix} & \cdots & \begin{bmatrix} H_1F_{11}^{m-2} & \mathbf{0}_{p\times s} \end{bmatrix} \\ \mathbf{0}_{p\times n} & \begin{bmatrix} H_1 & \mathbf{0}_{p\times s} \end{bmatrix} & \cdots & \begin{bmatrix} H_1F_{11}^{m-3} & \mathbf{0}_{p\times s} \end{bmatrix} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \begin{bmatrix} H_1 & \mathbf{0}_{p\times s} \end{bmatrix} \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} \end{bmatrix} \begin{bmatrix} w_{k+m-2}^O \\ w_{k+m-2}^{UO} \\ w_{k+m-3}^O \\ \vdots \\ w_k^{UO} \end{bmatrix} + V_k$$

$$\mathcal{Y}_k = M_o z_k^O + \underbrace{\begin{bmatrix} H_1 & H_1F_{11} & \cdots & H_1F_{11}^{m-2} \\ \mathbf{0}_{p\times l} & H_1 & \cdots & H_1F_{11}^{m-3} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{p\times l} & \mathbf{0}_{p\times l} & \mathbf{0}_{p\times l} & H_1 \\ \mathbf{0}_{p\times l} & \mathbf{0}_{p\times l} & \mathbf{0}_{p\times l} & \mathbf{0}_{p\times l} \end{bmatrix}}_{\triangleq M_w} \underbrace{\begin{bmatrix} w_{k+m-2}^O \\ w_{k+m-3}^O \\ \vdots \\ w_k^O \end{bmatrix}}_{\triangleq W_k^O} + V_k. \tag{3.18}$$

Writing the modified measurement equation at the next time step

$$\mathcal{Y}_k = M_o z_k^O + M_w W_k^O + V_k \tag{3.19}$$

$$\mathcal{Y}_{k+1} = M_o z_{k+1}^O + M_w W_{k+1}^O + V_{k+1} \tag{3.20}$$

3.2.2 Formulating linear stationary time series

Since, the system is detectable, the matrix M_o is full column rank and hence its pseudo-inverse is unique and is defined by $M_o^\dagger = (M_o^T M_o)^{-1}(M)_o^T$. Projecting (3.19) and (3.20) onto the state space

$$M_o^\dagger \mathcal{Y}_k = z_k^O + M_o^\dagger M_w W_k^O + M_o^\dagger V_k \tag{3.21}$$

$$M_o^\dagger \mathcal{Y}_{k+1} = z_{k+1}^O + M_o^\dagger M_w W_{k+1}^O + M_o^\dagger V_{k+1} \tag{3.22}$$

Eliminating the observable state z_k^O by subtracting (3.21) from (3.22), the time series is formulated

$$\underbrace{M_o^\dagger \mathcal{Y}_{k+1} - F_{11} M_o^\dagger \mathcal{Y}_k}_{\triangleq \mathcal{Z}_k} = \underbrace{w_k^O + M_o^\dagger M_w W_{k+1}^O - F_{11} M_o^\dagger M_w W_k^O}_{\triangleq \mathcal{W}_k} + \underbrace{M_o^\dagger V_{k+1} - F_{11} M_o^\dagger V_k}_{\triangleq \mathcal{V}_k} \tag{3.23}$$

wherein $\mathcal{Z}_k$, $\mathcal{W}_k$ and $\mathcal{V}_k$ are concatenated of the measurement sequence, the process noise and the measurement noise respectively at different times. Although there is an abuse of notation in using the subscript k here, the actual time histories of the noise and measurement sequences is clearly specified in (3.17). The expressions of $\mathcal{W}_k$ and $\mathcal{V}_k$ as a co-efficient matrix multiplied by a concatenated vector of noise terms

$$\mathcal{W}_k = w_k^O + M_o^\dagger M_w W_{k+1}^O - F_{11} M_o^\dagger M_w W_k^O \tag{3.24}$$

$$\mathcal{W}_k = \underbrace{\Big[\begin{bmatrix} M_o^\dagger M_w & \boldsymbol{I}_{l\times l} \end{bmatrix} - \begin{bmatrix} \mathbf{0}_{l\times l} & F_{11} M_o^\dagger M_w \end{bmatrix} \Big]}_{\triangleq \mathcal{A}} \begin{bmatrix} w_{k+m-1}^O \\ w_{k+m-2}^O \\ \vdots \\ w_k^O \end{bmatrix} \tag{3.25}$$

$$\text{Let } \mathcal{A} = \begin{bmatrix} A_m & A_{m-1} & \cdots & A_1 \end{bmatrix}, \tag{3.26}$$

wherein, the matrix $\boldsymbol{I}_{l\times l}$ is the identity matrix of dimension l. The sequence $\mathcal{W}_k$ and its covariance can be written as follows.

$$\mathcal{W}_k = A_1 w_k + A_2 w_{k+1} + \cdots + A_m w_{k+m-1} \tag{3.27}$$

The sequence $\mathcal{V}_k$ and its covariance can be expressed in a similar way,

$$\mathcal{V}_k = M_o^{\dagger} V_{k+1} - F_{11} M_o^{\dagger} V_k \tag{3.28}$$

$$\mathcal{V}_k = \underbrace{\Big[\begin{bmatrix} M_o^{\dagger} & \mathbf{0}_{l\times p} \end{bmatrix} - \begin{bmatrix} \mathbf{0}_{l\times p} & F_{11} M_o^{\dagger} \end{bmatrix} \Big]}_{\triangleq \mathcal{B}\ \in \mathbb{R}^{l\times(m+1)p}} \begin{bmatrix} v_{k+m} \\ v_{k+m-1} \\ \vdots \\ v_k \end{bmatrix} \tag{3.29}$$

$$\text{Let } \mathcal{B} = \begin{bmatrix} B_m & B_{m-1} & \cdots & B_0 \end{bmatrix} \tag{3.30}$$

$$\mathcal{V}_k = B_0 v_k + B_1 v_{k+1} + \cdots + B_m v_{k+m} \tag{3.31}$$

3.2.3 Estimating the covariance matrices

Since the white Gaussian noise sequences w_k^O and v_k are both independent and identically distributed (i.i.d.), the sequence $\mathcal{Z}_k$, which is a function of the measurements, is a zero mean strictly stationary time series with these noise terms as inputs. Therefore, the covariance of $\mathcal{Z}_k$ is given by

$$Cov(\mathcal{Z}_k) = Cov(\mathcal{W}_k) + Cov(\mathcal{V}_k). \tag{3.32}$$

Writing down the expressions for the covariances in (3.32) using (3.27) and (3.31)

$$Cov(\mathcal{W}_k) = A_1 Q^O A_1^T + \cdots A_m Q^O A_m^T \tag{3.33}$$

$$Cov(\mathcal{V}_k) = B_0 R B_0^T + \cdots B_m R B_m^T \tag{3.34}$$

Note that the covariance matrices are constant for all time. Since $\mathcal{Z}_k$ is a function of the measurements, its covariance can be estimated using the following unbiased estimator Λ_k.

$$\Lambda_k = \frac{1}{k}\sum_{i=1}^{k} \mathcal{Z}_i \mathcal{Z}_i^T \tag{3.35}$$

In order to calculate the covariance recursively, the following recursive estimator is used with $\Lambda_0 = 0$,

$$\Lambda_k = \frac{k-1}{k}\Lambda_{k-1} + \frac{1}{k}\mathcal{Z}_k \mathcal{Z}_k^T \tag{3.36}$$

Note that first m measurements are used to get the first estimate of the covariance matrices. In order to estimate the unknown elements of the covariance matrices, $Cov(\mathcal{Z}_k)$ is replaced with its estimator Λ_k. The entire equation is then vectorized and the right hand side is split into a known part Θ_{known} and a product of matrix S and a vector of concatenated unknown elements to be estimated $\hat{\theta}_k$.

$$vec(\Lambda_k) = \Theta_{known} + S\hat{\theta}_k \tag{3.37}$$

The S matrix here is constructed using the matrices A_i and B_i from (3.27) and (3.31). The final assumption is stated below.

Assumption 6. *The matrix S used in (3.37) has full column rank.*

This assumption ensures the estimability of the unknown elements in the noise covariance matrices Q and R. This is well in line with a similar assumption was made about the number of unknown elements in the Q matrix by Mehra [100]. The implications of this assumptions are discussed in further detail in section 3.4. Hence, the unknown elements are calculated at each time using

$$\hat{\theta}_k = S^{\dagger}(vec(\Lambda_k) - \Theta_{known}) \tag{3.38}$$

wherein $S^\dagger = (S^T S)^{-1} S^T$ is the unique pseudo-inverse. This is a generalization of the case handled in [108]. It is clear that the estimability of the unknown elements depends on whether or not S is full column rank. However, since S is a constant matrix depending only on the state transition matrix F and the observation matrix H, it can be pre calculated and the bounds on the number of unknown elements or the conditions for estimability of the covariance matrices can be checked. This analysis is performed in section 3.4.

It is noteworthy that this algorithm can incorporate linear constraints between the elements of the unknown matrix. For example, if two of the elements of the unknown covariance matrix are known to be equal even though their value is unknown, the two unknowns can be denoted by the same variable while splitting the equation into known and unknown parts as in (3.37).

The special cases when only one of R or Q^O is unknown is given by

$$\Lambda_k - Cov(\mathcal{W}_k) = B_0 \hat{R}_k B_0^T + \cdots B_m \hat{R}_k B_m^T. \tag{3.39}$$

Vectorizing the above equation (denoted $vec(\cdot)$), let $vec(\Lambda_k - Cov(\mathcal{W}_k)) \triangleq U_k^r$

$$U_k^r = vec(B_0 \hat{R}_k B_0^T + \cdots B_m \hat{R}_k B_m^T) \tag{3.40}$$

$$U_k^r = \underbrace{(B_0 \otimes B_0 + \cdots B_m \otimes B_m)}_{\triangleq T_1} vec(\hat{R}_k) \tag{3.41}$$

$$vec(\hat{R}_k) = T_1^\dagger U_k^r \tag{3.42}$$

wherein, '$\otimes$' is the Kronecker product.

Assumption 7. *Matrix T_1 given in Eq. 3.42 has full column rank.*

For unknown Q^O case

$$\Lambda_k - Cov(\mathcal{V}_k) = A_1 \hat{Q}_k^O A_1^T + \cdots A_m \hat{Q}_k^O A_m^T. \tag{3.43}$$

Let $vec(\Lambda_k - Cov(\mathcal{V}_k)) \triangleq U_k^q$. Hence, vectorizing the above equation

$$U_k^q = vec(A_1 \hat{Q}_k^O A_1^T + \cdots A_m \hat{Q}_k^O A_m^T) \tag{3.44}$$

$$U_k^q = \underbrace{(A_1 \otimes A_1 + \cdots A_m \otimes A_m)}_{\triangleq T_2} vec(\hat{Q}_k^O) \tag{3.45}$$

Assuming that T_2 has full column rank, the estimate $\hat{Q}_k^O$ is given by

$$vec(\hat{Q}_k^O) = T^\dagger vec(U_k^q). \tag{3.46}$$

3.2.4 Algorithm outline

The pseudo-code of the algorithm is summarized in Algorithm 1. Note that since the covariance matrices estimated using the measurements use the $vec(\cdot)$ operation, they may not positive definite due to the randomness in the measurements. A simple condition check is embedded into the algorithm which ensures positive definiteness of the estimated covariance matrices.

3.3 Convergence Analysis for noise covariance matrices

This section analyzes the stability of the algorithm under the stated assumptions. The convergence of the covariance estimates to their true values are investigated first. Then, the stability of the state error covariance matrix sequences is evaluated and compared to the state error covariance of the baseline Kalman filter.

Algorithm 1 The Covariance Matching Kalman Filter

Input: $\hat{z}_0^O$, P_0, $Cov(\mathcal{Z})_0 = 0$, $\hat{Q}_0$, $\hat{R}_0$, $\{Q_{ij}, R_{ij}$ known$\}$, $y_i\big|_{i=1}^{k}$
Initialization: $(\hat{Q}_k)_1 = \hat{Q}_{k-1}$, $(\hat{R}_k)_1 = \hat{R}_{k-1}$
Output: $\hat{z}_k^O$, $\hat{Q}_k$, $\hat{R}_k$, $P_{k|k}$
for $k = 1 \to n$ **do**
    Using $\{y_k\}$ calculate $\mathcal{Y}_k$ ▷ Eq. (3.17)
    Using $\mathcal{Y}_k$ calculate $\mathcal{Z}_k$ ▷ Eq. (3.23)
    Using $\mathcal{Z}_k$ calculate Λ_k ▷ Eq. (3.36)
    Estimate $\hat{R}_k$ and $\hat{Q}_k^O$ using Λ_k ▷ Eq. (3.38)
    if $\hat{R}_k \preceq 0$ **then**
        $\hat{R}_k = \hat{R}_{k-1}$
    end if
    if $\hat{Q}_k^O \preceq 0$ **then**
        $\hat{Q}_k^O = \hat{Q}_{k-1}^O$
    end if
    $\hat{z}_{k|k-1}^O = F_{11}\hat{z}_{k-1|k-1}^O$
    $\hat{z}_{k|k}^O = \hat{z}_{k|k-1}^O + K_k(y_k - H_1\hat{z}_{k|k-1}^O)$
    $P_{k|k-1} = F_{11}P_{k-1|k-1}F_{11}^T + \hat{Q}_k^O$
    $K_k = P_{k|k-1}H_1^T(H_1P_{k|k-1}H_1^T + \hat{R}_k)^{-1}$
    $P_{k|k} = P_{k|k-1} - K_kH_1P_{k|k-1}$
end for

3.3.1 Convergence of noise covariance

Substituting the values of $\mathcal{W}_k$ and $\mathcal{V}_k$ using (3.27) and (3.31)

$$\mathcal{Z}_k = \sum_{i=1}^{m} A_i w^O_{k+i-1} + \sum_{i=0}^{m} B_i v_{k+i} \tag{3.47}$$

The above equation is a linear strictly stationary time series because of the zero mean, white Gaussian, and uncorrelated noise assumptions in place. Consider the autocovariance function of zero mean time series $\mathcal{Z}_k$ given by

$$C(k, k+\tau) := E[\mathcal{Z}_k \mathcal{Z}^T_{k+\tau}]. \tag{3.48}$$

If $|\tau| > m, C(k, k+\tau) = 0$. Hence, the autocovariance function $C(k_1, k_2)$ decays to 0 as k_1 and k_2 grow farther away from each other. Hence, the central limit theorem for linear stationary time series applies here which uses the weak law of large numbers [71]. This theorem ensures an element-wise convergence given by

$$\sqrt{k}\{\hat{C}(k,k) - C(k,k)\}_{ij} \xrightarrow{D} \mathcal{N}(\mathbf{0}, \Omega_{ij}) \tag{3.49}$$

wherein, $\hat{C}(k,k)$ is equal to the Λ_k calculated recursively in (3.36), the subscript ij denotes the element corresponding to the i^{th} row and j^{th} column of the matrix and the ^{D} signifies convergence in distribution. This result motivates the following convergence,

$$\forall \varepsilon > 0,\ Pr(|\{\hat{C}(k,k) - C(k,k)\}_{ij}| > \varepsilon) \xrightarrow{k\to\infty} 0 \tag{3.50}$$

The rate of convergence is directly proportional to $k^{-\frac{1}{2}}$ for all i and j. Using (3.38) and the existence of the pseudo-inverse for S matrix

$$\forall \varepsilon > 0,\ Pr(|\{\hat{\theta}_k - \theta\}_i| > \varepsilon) \xrightarrow{k\to\infty} 0 \tag{3.51}$$

wherein θ is the vector of true values of the unknown elements. Hence, the convergence of the covariance matrices is given by

$$\forall \varepsilon > 0,\ Pr(|\{\hat{Q}^O_k - Q^O\}_{ij}| > \varepsilon) \overset{k\to\infty}{\longrightarrow} 0, \tag{3.52}$$

$$\forall \varepsilon > 0,\ Pr(|\{\hat{R}_k - R\}_{ij}| > \varepsilon) \overset{k\to\infty}{\longrightarrow} 0 \tag{3.53}$$

where i and j correspond to the unknown elements of the covariance matrix. A similar convergence holds for autocovariance function with a lag $\tau \neq 0$,

$$\forall \varepsilon > 0,\ Pr(|\{\hat{C}(k,k+\tau) - C(k,k+\tau)\}_{ij} > \varepsilon) \overset{k\to\infty}{\longrightarrow} 0 \tag{3.54}$$

wherein, $\hat{C}(k,k+\tau) = \frac{1}{k}\sum_{i=1}^{k} \mathcal{Z}_i \mathcal{Z}_{i+\tau}^T$. Using additional autocovariance functions with non-zero τ augment and improve the ability to estimate covariance matrices. However, more measurements will have to be stored in the memory which might not be desirable. A trade off between the accuracy and memory may give the best performance. It is important to note that the positive definiteness checks in the algorithm do not affect the convergence of the covariance estimates. Retaining the previous estimate when positive definiteness is violated affects state estimation and the state error covariance matrix. However, since the autocovariance estimate $\hat{C}(k,k)$ is independent of the state estimate and is only dependent on the measurements, the convergence of $\hat{C}(k,k)$ to its true value is guaranteed. This causes the covariance estimate to converge to its true value. Hence, the results presented here hold regardless of the checks.

3.3.2 Convergence of the state error covariance

Three different error covariance matrix sequences are compared in this subsection. The matrix $\hat{P}_k$ is the one-step predictor error covariance of the filter which is calculated by propagating the initial covariance matrix P_0 using the estimated Kalman gain $\hat{K}_k$ and the estimated $\hat{Q}_k^O$ and $\hat{R}_k$ covariance matrices. The P_k matrix is the true error covariance of the filter which propagates the covariance using the estimated Kalman gain $\hat{K}_k$ and the true Q^O and R. Finally, $P_{k,opt}$ is the optimal error covariance matrix of the baseline Kalman filter given full knowledge of the noise statistics. Writing these matrix sequences down

$$\hat{P}_{k+1} = \hat{\bar{F}}_k \hat{P}_k \hat{\bar{F}}_k^T + \hat{K}_k \hat{R}_k \hat{K}_k^T + \hat{Q}_k^O \tag{3.55}$$

$$P_{k+1} = \hat{\bar{F}}_k P_k \hat{\bar{F}}_k^T + \hat{K}_k R \hat{K}_k^T + Q^O \tag{3.56}$$

$$P_{k+1,opt} = \bar{F}_k P_{k,opt} \bar{F}_k^T + K_k R K_k^T + Q^O \tag{3.57}$$

$$\text{wherein, } \hat{K}_k = F_{11} \hat{P}_k H_1^T (H_1 \hat{P}_k H_1^T + \hat{R}_k)^{-1}, \tag{3.58}$$

$$K_k = F_{11} P_{k,opt} H_1^T (H_1 P_{k,opt} H_1^T + R)^{-1}, \tag{3.59}$$

$$\hat{\bar{F}}_k = F_{11} - \hat{K}_k H_1, \textit{ and } \bar{F}_k = F_{11} - K_k H_1 \tag{3.60}$$

and the initial error covariances are equal, $\hat{P}_0 = P_0 = P_{0,opt}$. Subtracting (3.56) from (3.55), the asymptotic of the matrix sequence $\hat{P}_k - P_k$ can be analyzed by

$$\hat{P}_{k+1} - P_{k+1} = \hat{\bar{F}}_k (\hat{P}_k - P_k) \hat{\bar{F}}_k^T + \hat{K}_k (\hat{R}_k - R) \hat{K}_k^T + (\hat{Q}_k^O - Q^O) \tag{3.61}$$

Since the initial error covariance is the same, the matrix sequence can be expanded as

$$\hat{P}_{k+1} - P_{k+1} = \sum_{i=0}^{k} \{ \hat{\phi}_i \hat{K}_i (\hat{R}_i - R) \hat{K}_i^T \hat{\phi}_i^T + \hat{\phi}_i (\hat{Q}_i^O - Q^O) \hat{\phi}_i^T \} \tag{3.62}$$

where $\hat{\phi}_i = \hat{\hat{F}}_i \hat{\hat{F}}_{i-1} \cdots \hat{\hat{F}}_0$ is the state transition matrix corresponding to $\hat{\hat{F}}_k$ from initial time to the i^{th} time. Consider the partial sum $\Delta P_{m,n}$ from m to n defined as

$$\Delta P_{m,n} = (\hat{P}_{n+1} - P_{n+1}) - (\hat{P}_m - P_m) \tag{3.63}$$

$$\Delta P_{m,n} = \sum_{i=m}^{n} \{\hat{\phi}_i \hat{K}_i (\hat{R}_i - R) \hat{K}_i^T \hat{\phi}_i^T + \hat{\phi}_i (\hat{Q}_i^O - Q^O) \hat{\phi}_i^T\} \tag{3.64}$$

Each element of $\Delta P_{m,n}$ is a function of the elements of $\hat{R}_i - R$ and $\hat{Q}_i^O - Q^O$. Since, $\hat{\phi}_i$ and $\hat{K}_i$ are bounded from above, there exist ε_{ij}^r, ε_{ij}^q, δ_{ij}^r, δ_{ij}^q and corresponding N_{ij}^r and N_{ij}^q such that

$$\forall k > N_{ij}^q, Pr(|\{\hat{Q}_k^O - Q^O\}_{ij}| < \varepsilon_{ij}^q) > 1 - \delta_{ij}^q \tag{3.65}$$

$$\forall k > N_{ij}^r, Pr(|\{\hat{R}_k - R\}_{ij}| < \varepsilon_{ij}^r) > 1 - \delta_{ij}^r \tag{3.66}$$

$$\forall m, n > \max_{i,j}(N_{ij}^q, N_{ij}^r), \quad Pr(|\{\Delta P_{m,n}\}_{ij}| < \varepsilon_{ij}^p) > 1 - \delta_{ij}^p \tag{3.67}$$

Here each of ε_{ij}^p and δ_{ij}^p are functions of all the ε_{ij}^q and ε_{ij}^r, and δ_{ij}^q and δ_{ij}^r respectively for all i and j. Using (3.67)

$$\forall i, j\ \forall \varepsilon_{ij}^p > 0\ \lim_{m,n \to \infty} Pr(|\{\Delta P_{m,n}\}_{ij}| > \varepsilon_{ij}^p) \longrightarrow 0 \tag{3.68}$$

Hence, using the Cauchy criterion for random sequences given in Theorem 6.3.1 of [120]

$$\forall i, j\ \forall \varepsilon_{ij}^p > 0\ \lim_{k \to \infty} Pr(|\{\hat{P}_k - P_k\}_{ij}| > \varepsilon_{ij}^p) \longrightarrow 0 \tag{3.69}$$

$$\hat{P}_k - P_k \xrightarrow{P} \mathbf{0}_{l \times l} \tag{3.70}$$

Now consider the matrices defined in (3.58) and (3.60). The gain matrix $\hat{K}_k$ which is a continuous function of $\hat{P}_k$, $\hat{Q}_k^O$, and $\hat{R}_k$. Hence using the continuous mapping theorem given in Corollary 8.3.1 of [120]

$$\hat{K}_k - K_k \xrightarrow{P} \mathbf{0}_{l\times p} \tag{3.71}$$

$$\hat{K}_k - K_k \xrightarrow{P} \mathbf{0}_{l\times p} \implies \hat{\bar{F}}_k - \bar{F}_k \xrightarrow{P} \mathbf{0}_{l\times l} \tag{3.72}$$

Hence using the above results

$$P_k - P_{k,opt} \xrightarrow{P} \mathbf{0}_{l\times l} \tag{3.73}$$

$$\hat{P}_k - P_{k,opt} \xrightarrow{P} \mathbf{0}_{l\times l} \tag{3.74}$$

3.4 Estimability of noise covariance matrices

In (3.38), the existence of a pseudo-inverse of S matrix was assumed so that the unknown elements of Q^O and R matrix could be estimated. The S matrix can be pre calculated and the estimability of the unknown elements can be checked before the filter is deployed. However, a physical insight behind this assumption is uncovered via mathematical analysis in this section.

3.4.1 Estimability of *R* matrix

Say that the entire R is unknown and needs to be estimated while Q^O matrix is known. Consider the case when the system given by (4.1) has linearly dependent measurements. This is mathematically expressed as

$$\exists\ \xi \neq \mathbf{0}_{p\times 1}\ :\ \xi^T H_1 = \mathbf{0}_{1\times l} \tag{3.75}$$

$$\begin{bmatrix}\xi^T & \dots & \xi^T\end{bmatrix}_{1\times mp} M_o = \mathbf{0}_{1\times l} \implies M_o^{\dagger} \begin{bmatrix}\xi^T & \dots & \xi^T\end{bmatrix}^T_{mp\times 1} = \mathbf{0}_{l\times 1} \tag{3.76}$$

Hence, using the matrices defined in (3.30), the above condition translates to

$$B_i\xi = \mathbf{0}_{l\times 1},\ \forall\, i = 0,1,\ldots,m \implies \sum_{i=0}^{m}(B_i\xi\xi^T B_i^T) = \mathbf{0}_{l\times l} \tag{3.77}$$

$$\underbrace{(B_0\otimes B_0+\cdots+B_m\otimes B_m)}_{T_1} vec(\xi\xi^T) = \mathbf{0}_{l^2\times 1} \tag{3.78}$$

Hence, a common null space for all the coefficient matrices B_i for $i = 1,2,\ldots,m$ was found. Therefore, the matrix $vec(\xi\xi^T)$ belongs to the null space of the matrix T_1 defined in (3.41). In case of repeated or linearly dependent measurement, estimation of the R matrix is ambiguous. It is important to note that this unintentionally establishes a hard limit on the number of measurements available to the system. If the number of measurements is greater than the number of states, there always exists a ξ which satisfies the above condition and the R matrix cannot be unambiguously estimated. Intuitively, for example, say that the system has two identical sensors, This algorithm is unable to estimate the cross-covariance between the two noises of the identical sensors (with possibly different noise covariance matrices). The measurement equation can be modified by averaging out the linearly dependent measurements and estimating their aggregate covariance matrix. However, note that this is not a sufficient condition for estimability of R. There can be other cases which make $rank(T_1) = 0$.

3.4.2 Estimability of Q^O matrix

For estimability of the Q^O matrix, it has been established that the number of unknown elements that need be estimated cannot be more than $l\times p$ where p is the number of independent measurements [100].

Again let us assume that the entire Q^O matrix is to be estimated while R matrix is completely known. Let the matrix $X = (M_o^T M_o)^{-1}$ and $M_o^\dagger$ is stated as

$$M_o^\dagger = \begin{bmatrix} X(F_{11}^T)^{m-1}H_1^T & X(F_{11}^T)^{m-2}H_1^T & \dots & XH_1^T \end{bmatrix} \tag{3.79}$$

Note that X is an invertible matrix. Similar to the case for unknown R matrix, the idea is to find a null space common to all the matrix $A_i,\ i = 1,2,\dots,m$. First, the expression for $M_o^\dagger$ can be expressed as

$$M_o^\dagger = \begin{bmatrix} C_1 & C_2 & \dots & C_m \end{bmatrix} \tag{3.80}$$

$$C_i = X(F_{11}^T)^{m-i}H_1^T \quad i = 1,2,\dots,m \tag{3.81}$$

Using the above expression, the matrix $M_o^\dagger M_w$ is evaluated as

$$M_o^\dagger M_w = \begin{bmatrix} D_1 & D_2 & \dots & D_m \end{bmatrix} \tag{3.82}$$

$$D_1 = C_1 H_1 \tag{3.83}$$

$$D_2 = C_1 H_1 F_{11} + C_2 H_1 \tag{3.84}$$

$$\vdots$$

$$D_m = C_1 H_1 F_{11}^{m-2} + C_2 H_1 F_{11}^{m-3} + \dots + C_{m-1} H_1 \tag{3.85}$$

Evaluating $\mathcal{A}$ matrix using the expression above

$$\mathcal{A} = \begin{bmatrix} \begin{bmatrix} M_o^\dagger M_w & \boldsymbol{I} \end{bmatrix} - \begin{bmatrix} \mathbf{0}_{l\times l} & F_{11} M_o^\dagger M_w \end{bmatrix} \end{bmatrix} \tag{3.86}$$

$$A_m = D_1 = C_1 H_1 \tag{3.87}$$

$$\vdots$$

$$A_2 = D_m - F_{11} D_{m-1} \tag{3.88}$$

$$A_1 = \boldsymbol{I} - F_{11} D_m \tag{3.89}$$

Note that all matrices $A_i \in \mathbb{R}^{l \times l}$ for $i = 1, 2, \dots, m$. However, due to the matrix multiplication $H_1^T H_1$, their rank cannot be more than p with an exception of A_1. The estimability of Q^O thus directly depends on $rank(A_1 \otimes A_1 + \dots A_m \otimes A_m)$ and one of the cases in which this rank is 0 is if the null spaces of the matrices $A_i,\ i = 1, 2, \dots, m$ intersect. Let us say this is the case and there exists a non-zero vector $\kappa \in \mathbb{R}^{l \times 1}$ such that $A_i \kappa = \mathbf{0}_{l \times 1}$ for $i = 1, 2, \dots, m$. Trying to find an expression for κ

$$\forall i = 1, 2, \dots, m \quad A_i \kappa = \mathbf{0}_{l \times 1} \implies D_i \kappa = \mathbf{0}_{l \times 1} \tag{3.90}$$

However, note that in that case, $A_1 \kappa \neq \mathbf{0}_{l \times 1}$. Hence, a contradiction is reached and the matrices A_i do not share a part of their null space. However, since this is only a necessary condition, this analysis does not provide a condition for estimability of Q^O matrix.

3.5 Further comparison with prior literature

The problem of estimating the state and certain unknown elements of the process and measurement noise covariance matrices has received significant attention in prior literature, most notably in [100]. Several important facts that highlight crucial differences between [100] are stated below.

- In [100], the pair $(F, Q^{\frac{1}{2}})$ is assumed to be controllable and pair (F, H) is taken to be observable. The work respectively assumes stabilizability and detectability of the same matrix pairs as given in assumptions 4 which are clearly weaker technical hypotheses. The results presented in this chapter are applicable to a wider class of systems.

- Both the optimal (case of known covariance matrices) and suboptimal (case of unknown covariance matrices) are assumed to have reached steady-state conditions in [100]. This assumption is central to the developments therein since it is used to calculate the covariance matrix estimate. No such assumption as the steady-state properties are proven and convergence guaranteed.

- The state transition matrix F is assumed to be non-singular in [100] which is arguably a mild restriction. This assumption is subsequently used to calculate the estimate of the noise covariance matrices and hence is crucial to the formulation in [100]. On the other hand, the detectability assumption is sufficient for convergence and no additional non-singularity restrictions are placed upon the state transition matrix.

- The work in [128] analyzed discrete-time Kalman filtering with incorrect noise covariances. Corollary 3.3 in [128] states that for incorrect noise covariances obtained simply by multiplying with a positive scalar, the sequences are asymptotically white. This result, as the authors state, shows the insufficiency of the whiteness test used in [100] for estimating the noise covariance matrices that are required for constructing the steady-state filter. For related discussion, the readers are referred to [4]. This inadequacy of the whiteness condition should be noted to be a major limitation of the approach in [100] for estimating covariance matrices using the autocovariance function in case of sub-optimality of the filter. The method presented here is impervious to the above arguments involving the whiteness tests as the residual autocovariance is not used.

- In [100], convergence is proved for the asymptotic case, i.e., when the batch size (N) becomes infinitely large. However, such a filter can be potentially impractical (large memory buffer usage), and moreover, one has to always initiate the filter with a fixed batch size, the chosen batch size may not be large enough to guarantee convergence. In [100], it is assumed that the batch size N is much larger than n. All the results that follow prove convergence of the estimates for a large value of N. As given in (3.9), at most n measurements are stacked and the number of stacked measurements is pre-calculated using the system matrices (F and H).

3.6 Simulations

Consider a fictitious detectable system which satisfies the assumptions for this algorithm to converge.

$$x_k = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0.5 & 0 \\ 0.1 & 0.9 & 0.7 \end{bmatrix} x_{k-1} + w_{k-1} \tag{3.91}$$

$$y_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x_k + v_k \tag{3.92}$$

It is assumed that $w_k \sim \mathcal{N}(0,Q)$, and $v_k \sim \mathcal{N}(0,R)$ are both i.i.d white Gaussian noises. The values of the covariance matrices are

$$R = \begin{bmatrix} 5 & 0 \\ 0 & 4 \end{bmatrix}, \quad Q = \begin{bmatrix} 3 & 0.2 & 0 \\ 0.2 & 2 & 0 \\ 0 & 0 & 7.5 \end{bmatrix}$$

The elements R_{11}, Q_{11}, and Q_{22} are estimated and all other elements are known. The initial estimates for all these elements was chosen to be 10. From Monte Carlo simulation results, the estimate time history of $\hat{R}_k$ is given in Fig. 3.1. The time histories of estimates of the unknown elements of $\hat{Q}_k^O$ are shown to converge to its true value in Fig. (3.2). The predictor and true state error covariance converges to the optimal Kalman filter values Fig. (3.3) as shown in the simulations.

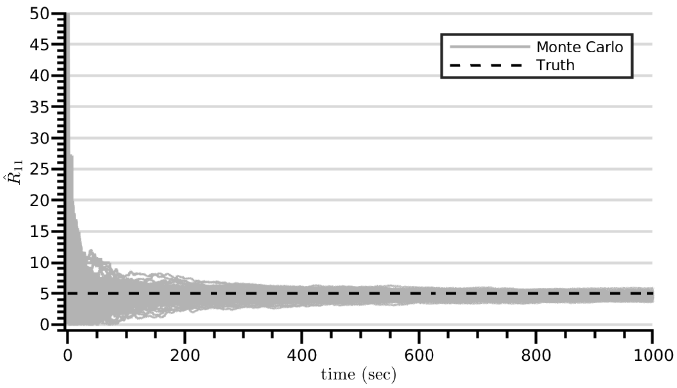

Figure 3.1: The estimate of the R_{11} element for 100 Monte Carlo simulations.

3.7 Conclusion

A novel algorithm to adaptively estimate the state and certain unknown elements of the process and measurement noise covariance matrices of a discrete linear time invariant stochastic system is formulated. The algorithm presented here is derived using a judicious combination of established adaptive filtering approaches such as correlation techniques and covariance matching techniques. The detectability property of the system is utilized for observing the state and formulating a time series containing measurement and process noise sequences independent of the state. A proof for the probabilistic convergence of the new algorithm is presented under additional assumptions of existence of a pseudo inverse used for unique-

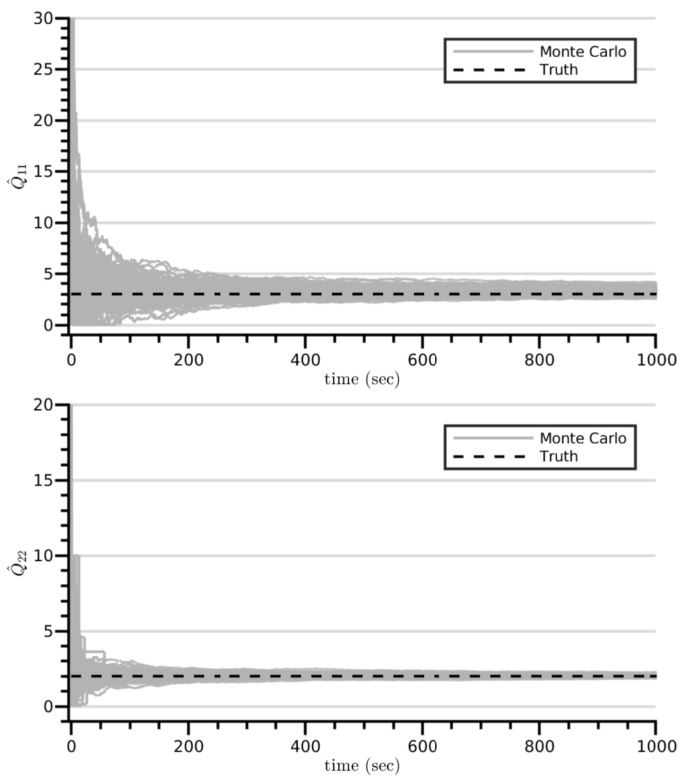

Figure 3.2: The estimates of the Q_{11} and Q_{12} elements for 100 Monte Carlo simulations.

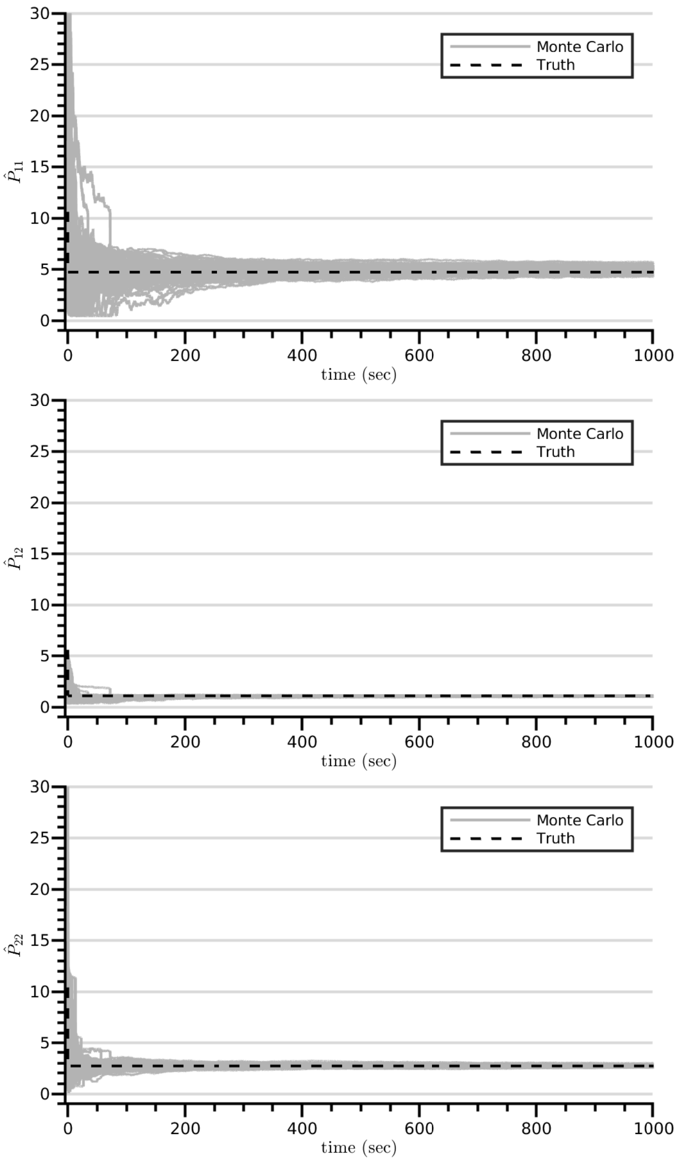

Figure 3.3: The elements of the estimated error covariances matrix from (3.55) along with the optimal Kalman filter truth from Eq. (3.57) for 100 Monte Carlo simulations.

ness of the estimates of the covariance matrices. This bears significant contrasts to approaches for adaptive covariance estimation algorithms reported in existing literature. Firstly, the technical assumptions of detectability rather than observability, no non-singularity requirements on the state transition and observation matrix like invertibility, and no requirement of reaching steady-state are less restrictive. Secondly, the proposed algorithm is independent of sub optimality tests for whiteness toward constructing the state-state filter. Lastly, the results do not require arbitrarily large batch sizes for ensuring convergence but rather, the memory usage and buffer size demands imposed by the algorithm can be a priori benchmarked in terms of the dimensionality of the state-space. Monte Carlo simulations demonstrate the effectiveness of the algorithm.

Chapter 4

Differential Geometric Identification Methods

4.1 Introduction

In the previous chapter an adaptive Kalman filter was introduced to estimate the noise covariance matrices for linear time invariant system. The positive definiteness of the noise covariance estimates is necessary for the stability of the resulting adaptive Kalman filter using the noise covariance matrix estimates. In the previous chapter, the most recent positive definite estimates were stored and a positive definiteness check on the estimates was required. Although the stability of the adaptive Kalman filter and the convergence of the noise covariance matrix estimates are immune to such an ad-hoc method, more effective identification methods can be developed. In this chapter, positive definite estimates of the unknown noise covariance matrices are obtained by adopting a differential geometric approach. Such methods enable minimization methods with a positive definite constraint to be transformed into unconstrained minimization on the manifold of symmetric and positive definite (SPD) matrices.

SPD matrices have a rich geometric structure and form a manifold as the topological space locally resembles the Euclidean space of symmetric matrices. The earliest use of differential geometry can be traced back to evaluating the geometric mean of a set of SPD matrices. Calculating the means are equivalent to minimizing the sum of squared norms. Minimizing the Frobenius norm, however, resulted in a swelling effect for SPD matrices [92]. This situation appears when the determinants of SPD matrices on a path between SPD matrices are artificially inflated. As a result, various other distance measures or metrics were used which gave the SPD matrices a Riemannian manifold structure.

The most common and geometrically rich metric is the affine-invariant (AI) metric [105, 118], also known as the GL-invariant metric or the Fischer metric from information geometry. The mean calculated by minimizing the metric is known as the Karcher mean or the Fréchet mean. The closed form solution to the Fréchet mean of n SPD matrices does not exist except for the case when n is two. Hence, optimization based methods have been derived to estimate the geometric mean [1, 30]. In spite of its geometrically rich nature of the AI metric, other metrics were popularized for their computational simplicity that is missing from the AI metric. Some other known metrics are the Log-Euclidean [14, 15], Cholesky decomposition based [159, 114], Log-Cholesky [92], Wasserstein [26, 96], and the Jensen-Bregman LogDet based metric [140, 42], and the Stein divergence [141].

In the context of adaptive identification of SPD matrices, several optimization based approaches have been developed. Differential geometric techniques have been applied to formulate various forms of Kalman filters to estimate manifold valued variables [150, 165, 102, 41]. Estimating the covariance matrix can be formulated as filtering on the SPD manifold. However, no closed form solution exists for the updates without making simplifying assumptions on the model. For example, the linear filter developed in [150] contains a closed form solution for a linear error model that is calculated with respect to a nominal trajectory known as the base model. However, the full manifold state observations are assumed to be available. Linearity of the model is arguably not a restrictive assumption since SPD matrices in most physical systems appear linearly. However, the affine model for covariance estimation given in Eq. (3.32), (3.33), and (3.34) does not fit into the framework given in [150]. The Stein center calculated using the square root of the LogDet divergence metric employed a closed form recursive solution that can be calculated using the square root of the SPD matrix [126]. However, such a recursive algorithm can only be used to calculate the geometric mean of SPD matrices using the Stein metric. Closed form solutions for a particular objective function were found in [158, 40] but do not readily generalize to the covariance estimation problem at hand.

In this chapter, the problem of estimating a symmetric positive definite matrix is posed as an Riemannian optimization problem on the SPD manifold. Consequently, SPD estimates of the noise covariance matrices are guaranteed. SPD matrices form a convex cone in the set of matrices of the same size. They also form a differentiable Riemannian manifold that is also a metric space with non positive curvature [25]. This structure enables one to solve constrained Euclidean optimization problems by posing them as unconstrained Riemannian optimization problems on the manifold [143]. Algorithms involving Riemannian optimization techniques have been used in the past for optimization on manifolds wherein manifold equivalents of first-order methods, second-order methods, line-search, Newton methods, and trust region methods have been developed [1, 30]. These methods utilize the concepts of geodesics, as well as, the gradients and possibly Hessians of objective functions defined on manifolds to ensure that the estimates are geometry aware. The advantages of such optimization methods over traditional optimization methods with additional constraint requiring the estimates to lie on the manifold have been well documented in the literature [1].

The chapter is organized as follows. First, the expression for the adaptive Kalman filter developed in Chapter 3 is restated here. A brief introduction to the geometry of the SPD manifold is presented. For brevity purposes, only relevant concepts required for the Riemannian optimization are stated. Next, the cost function which includes any additional element constraints on the noise covariance matrices is described. The Riemannian Gradient and Hessian of the cost function are then derived which are further used in a Riemannian Trust Region (RTR) optimization framework to obtain SPD estimates. Numerical simulations using the RTR based Adaptive Kalman filter to estimate the noise covariance matrices demonstrate the efficacy of the proposed approach.

4.2 Adaptive Filter Formulation

The basic structure of the adaptive filter formulated in this section follows from correlation based techniques [109, 53, 90].

4.2.1 Problem Formulation

A discrete linear time-varying (LTV) system is considered here with the system equations given by

$$\begin{aligned} x_{k+1} &= F_k x_k + G_k u_k + w_k \\ y_k &= H_k x_k + v_k \end{aligned} \tag{4.1}$$

wherein the process noise $w_k \sim \mathcal{N}(\mathbf{0}_{n\times 1}, Q)$ and the measurement noise $v_k \sim \mathcal{N}(\mathbf{0}_{p\times 1}, R)$ are uncorrelated white Gaussian noises with constant noise covariance matrices. Let $\phi_{i,k}$ be the associated state transition matrix such that $\phi_{k+1,k} = F_k$ and $\phi_{k,l} = \phi_{k,q}\phi_{q,l}$ for any $q \in Z$. No restrictions are made for the matrices F_k, G_k and H_k except for uniform observability of pair (F_k, H_k) and uniform controllability of the $(F_k, Q_k^{\frac{1}{2}})$ pair. To that end, the definition for uniform observability and uniform controllability is given by [73, Chapter 7.5]

Definition 4.2.1. *The pair (F_k, H_k) is uniformly observable if there exists an integer $s \geq 0$ and constants $0 < \alpha_1 < \alpha_2$ such that*

$$\alpha_1 \boldsymbol{I} \preceq M_{k+s,k} \preceq \alpha_2 \boldsymbol{I} \tag{4.2}$$

wherein,

$$M_{k+s,k} = \sum_{i=k}^{k+s} \phi_{i,k}^T H_i^T H_i \phi_{i,k} \tag{4.3}$$

and let $M_{k+l,k}$ for $l < s$ be the partial observability Gramian.

Definition 4.2.2. *The pair (F_k, E_k) is uniformly controllable if there exists an integer $s \geq 0$ and constants $0 < \beta_1 < \beta_2$ such that*

$$\beta_1 \boldsymbol{I} \preceq Y_{k+s,k} \preceq \beta_2 \boldsymbol{I} \tag{4.4}$$

wherein,

$$Y_{k+s,k} = \sum_{i=k}^{k+s} \phi_{k+s+1,i+1} E_i E_i^T \phi_{k+s+1,i+1}^T \tag{4.5}$$

Assumption 8. *Let the pair* (F_k, H_k) *be uniformly observable and the pair* $(F_k, Q_k^{\frac{1}{2}})$ *be uniformly controllable.*

The above assumption ensures the Kalman filter is exponentially stable [8, Theorem 5.3]. Subsequently, the following assumption is made on the Q and R noise covariance matrices.

Assumption 9. *The noise covariance matrices Q and R are both assumed to be constant and unknown.*

The aim of this paper is to estimate the unknown Q and R matrices while simultaneously estimating the states.

4.2.2 Measurement Difference Autocovariance approach

Since the pair (F_k, H_k) is uniformly controllable with constants $s \geq 0$ and $0 < \alpha_1 < \alpha_2$, consider $m \geq s$ measurements that are aggregated in time to form a linear time series. Such a development is described in the following result.

Proposition 1. *For the LTV system given by Eq.* (4.1) *with the Assumptions 9, and a $m \geq s$ from Definition 4.2.1, the measurements y_k of the system follow a linear time series given by*

$$\sum_{i=0}^{m} A_i^k y_{k-i} - \sum_{i=1}^{m} B_i^k G_{k-i} u_{k-i} = \sum_{i=1}^{m} B_i^k w_{k-i} + \sum_{i=0}^{m} A_i^k v_{k-i} \tag{4.6}$$

wherein, the coefficients A_i^k and B_i^k are completely determined from the system matrices F_k and H_k.

Proof. The proof follows from much of the past work on the measurement difference methods [53, 109]. We begin by accumulating m measurements by stacking them one on top of the other to form a modified measurement model given in Eq. (4.7). Defining $W_{k-1,k-m+1} = [w_{k-1}^T, \dots, w_{k-m+1}^T]^T$ and pre multiplying by

$$\underbrace{\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-m+1} \end{bmatrix}}_{\triangleq \mathcal{Y}_{k,k-m+1}} = \underbrace{\begin{bmatrix} H_k \phi_{k,k-m+1} \\ H_{k-1}\phi_{k-1,k-m+1} \\ \vdots \\ H_{k-m+1} \end{bmatrix}}_{\triangleq O_{k,k-m+1}} x_{k-m+1}$$
$$+ \underbrace{\begin{bmatrix} H_k & H_k F_{k-1} & H_k \phi_{k,k-2} & \cdots & H_k \phi_{k,k-m+2} \\ \mathbf{0}_{p\times n} & H_{k-1} & H_{k-1}F_{k-2} & \cdots & H_{k-1}\phi_{k-1,k-m+2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & H_{k-m+2} \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} \end{bmatrix}}_{\triangleq M^w_{k-1,k-m+1}} \tag{4.7}$$
$$\underbrace{\begin{bmatrix} w_{k-1} + G_{k-1}u_{k-1} \\ w_{k-2} + G_{k-2}u_{k-2} \\ \vdots \\ w_{k-m+1} + G_{k-m+1}u_{k-m+1} \end{bmatrix}}_{\triangleq U_{k-1,k-m+1}} + \underbrace{\begin{bmatrix} v_k \\ v_{k-1} \\ \vdots \\ v_{k-m+1} \end{bmatrix}}_{\triangleq V_{k,k-m+1}}$$

$O_{k,k-m+1}$, the invertible observability Gramian $M_{k,k-m+1}$ defined in Eq. (4.3) is recovered as shown below.

$$O^T_{k,k-m+1}\mathcal{Y}_{k,k-m+1} = M_{k,k-m+1}x_{k-m+1}+$$
$$O^T_{k,k-m+1}M^w_{k-1,k-m+1}U_{k-1,k-m+1} + O^T_{k,k-m+1}V_{k,k-m+1}$$

Inverting $M_{k,k-m+1}$ and using a one step predictor for the state x_{k-m+1}, a linear time series can be formed from the two equations given by

$$\begin{aligned} M^{-1}_{k,k-m+1}O^T_{k,k-m+1}\mathcal{Y}_{k,k-m+1} &= x_{k-m+1} + M^{-1}_{k,k-m+1}O^T_{k,k-m+1}M^w_{k-1,k-m+1}U_{k-1,k-m+1} \\ &\quad + M^{-1}_{k,k-m+1}O^T_{k,k-m+1}V_{k,k-m+1} \\ M^{-1}_{k-1,k-m}O^T_{k-1,k-m}\mathcal{Y}_{k-1,k-m} &= x_{k-m} + M^{-1}_{k-1,k-m}O^T_{k-1,k-m}M^w_{k-2,k-m}U_{k-2,k-m} \\ &\quad + M^{-1}_{k-1,k-m}O^T_{k-1,k-m}V_{k-1,k-m} \end{aligned}$$

Substituting $x_{k-m+1} = F_{k-m}x_{k-m} + G_{k-m}u_{k-m} + w_{k-m}$ and eliminating the state by subtraction, we get a linear time series given by

$$\mathcal{A}_k\mathcal{Y}_{k,k-m} = \mathcal{B}_kU_{k-1,k-m} + \mathcal{A}_kV_{k,k-m} \tag{4.8}$$

wherein,

$$\mathcal{A}_k = [M^{-1}_{k,k-m+1}O^T_{k,k-m+1}, \mathbf{0}_{n\times p}] - [\mathbf{0}_{n\times p}, F_{k-m}M^{-1}_{k-1,k-m}O^T_{k-1,k-m}] \tag{4.9}$$

and

$$\begin{aligned} \mathcal{B}_k = [M^{-1}_{k,k-m+1}O^T_{k,k-m+1}M^w_{k-1,k-m+1}, \boldsymbol{I}_{n\times n}] \\ - [\mathbf{0}_{n\times n}, F_{k-m}M^{-1}_{k-1,k-m}O^T_{k-1,k-m}M^w_{k-2,k-m}] \end{aligned} \tag{4.10}$$

Separating out individual components of the coefficients of y_k

$$\begin{aligned} A_0 &= M^{-1}_{k,k-m+1}\phi^T_{k,k-m+1}H^T_k \\ A_i &= M^{-1}_{k,k-m+1}\phi^T_{k-i,k-m+1}H^T_{k-i} - F_{k-m}M^{-1}_{k-1,k-m}F^T_{k-m}\phi^T_{k-i,k-m+1}H^T_{k-i} \\ A_m &= -F_{k-m}M^{-1}_{k-1,k-m}F^T_{k-m}H^T_{k-m} \end{aligned}$$

wherein, $i = 1, \ldots, m-1$ above and the coefficients of w_k is given by

$$B_1 = M_{k,k-m+1}^{-1} \phi_{k,k-m+1}^T M_{k,k}$$
$$B_i = M_{k,k-m+1}^{-1} \phi_{k-i+1,k-m+1}^T M_{k,k-i+1} - F_{k-m} M_{k-1,k-m}^{-1} \phi_{k-i+1,k-m}^T M_{k-1,k-i+1}$$
$$B_m = \boldsymbol{I}_{n\times n} - F_{k-m} M_{k-1,k-m}^{-1} F_{k-m}^T M_{k-1,k-m+1}$$

wherein, $i = 2, \ldots, m-1$ above. The statement of the proposition follows. □

Defining $\mathcal{Z}_k$ as the left hand side of Eq. (4.6), the autocovariance function of $\mathcal{Z}_k$ is given by

$$C_{k,k-p} = E[\mathcal{Z}_k \mathcal{Z}_{k-p}^T] = \sum_{i=p+1}^{m} B_i^k Q B_{i-p}^{k-p\,T} + \sum_{i=p}^{m} A_i^k R A_{i-p}^{k-p\,T} \tag{4.11}$$

wherein $p = 0, \ldots, m$. Notice that the autocovariance $C_{k,k-p} = \mathbf{0}_{n\times n}$ for $p > m$ vanishes. As long as the number of past measurements y_k stored at every time instant is greater than $m+1$, the autocovariance function can be estimated.

4.2.3 Covariance Matrix Estimation

The autocovariance is estimated using a single measurement as $\hat{C}_{k,k-p} = \mathcal{Z}_k \mathcal{Z}_{k-p}^T$. The elements of the autocovariance function can be rearranged using the vech$(\cdot)$ operation as follows.

$$\underbrace{\begin{bmatrix} \mathrm{vech}(\hat{C}_{k,k}) \\ \mathrm{vec}(\hat{C}_{k,k-1}) \\ \vdots \\ \mathrm{vec}(\hat{C}_{k,k-P}) \end{bmatrix}}_{\triangleq b_k} = \underbrace{\begin{bmatrix} \sum_{i=1}^{m} B_i^k \otimes_h B_i^k & \sum_{i=0}^{m} A_i^k \otimes_h A_i^k \\ \sum_{i=2}^{m} B_{i-1}^{k-1} \otimes_u B_i^k & \sum_{i=1}^{m} A_{i-1}^{k-1} \otimes_u A_i^k \\ \vdots & \vdots \\ \sum_{i=P+1}^{m} B_{i-P}^{k-P} \otimes_u B_i^k & \sum_{i=P}^{m} A_{i-P}^{k-P} \otimes_u A_i^k \end{bmatrix}}_{\triangleq D_k} \underbrace{\begin{bmatrix} \mathrm{vech}(Q) \\ \mathrm{vech}(R) \end{bmatrix}}_{\triangleq \theta} \tag{4.12}$$

A recursive least squares (RLS) estimation technique starting from an initial guess $(\hat{\theta}_0, \Psi_0)$ is given by

$$\begin{aligned}
\hat{\theta}_{k+1} &= \hat{\theta}_k + L_k(b_k - D_k\hat{\theta}_k) \\
\Psi_{k+1} &= (\boldsymbol{I} - L_kD_k)\Psi_k(\boldsymbol{I} - L_kD_k)^T + L_kR_WL_k^T \\
L_k &= \Psi_kD_k^T(R_W + D_k\Psi_kD_k^T)^{-1}
\end{aligned} \tag{4.13}$$

The convergence of the estimate has been established in Ref. [53, Theorem 8].

4.2.4 Adaptive Kalman Filter

Using the estimates $\hat{Q}_k$ and $\hat{R}_k$ of the noise covariance matrices, the following equations constitute the adaptive Kalman filter equations.

$$\begin{aligned}
\hat{x}_{k|k-1} &= F_{k-1}\hat{x}_{k-1} + G_{k-1}u_{k-1} \\
\hat{x}_{k|k} &= \hat{x}_{k|k-1} + \hat{K}_k(y_k - H_k\hat{x}_{k|k-1}) \\
\hat{P}_{k|k-1} &= F_{k-1}\hat{P}_{k-1|k-1}F_{k-1}^T + \hat{Q}_k \\
\hat{K}_k &= \hat{P}_{k|k-1}H_k^T(H_k\hat{P}_{k|k-1}H_k^T + \hat{R}_k)^{-1} \\
\hat{P}_{k|k} &= (\boldsymbol{I} - \hat{K}_kH_k)\hat{P}_{k|k-1}(\boldsymbol{I} - \hat{K}_kH_k)^T + \hat{K}_k\hat{R}_k\hat{K}_k^T
\end{aligned} \tag{4.14}$$

wherein, $\hat{P}_{k|k}$, $\hat{P}_{k|k-1}$, and $\hat{K}_k$ are the estimates of the quantities in the nominal Kalman filter [73, Chapter 7].

4.3 Riemannian Trust-Region method

Although the recursive least squares successfully estimates the elements of the noise covariance matrices, it does not guarantee SPD estimates of the covariance matrix. The convergence of the estimates to the true covariance matrices is guaranteed provided the matrix D_k is persistently excited. However, the transients are important when the covariance estimate is concurrently used to estimate the state vector. In this case, the filter may run into a problem of loss of observability or worse, provide negative information updates to the filter by virtue of a non positive definite noise covariance matrix estimate. As a result, having a SPD noise covariance matrix estimate is crucial to obtain a stable adaptive Kalman filter. To this end, a geometric optimization approach that respects the geometry of SPD matrices is introduced here. A brief summary of the geometry of SPD matrices is provided below (for a comprehensive review, see, e.g., [25] for SPD matrices and [1] for Riemannian optimization methods).

4.3.1 Geometry of Covariance Matrices

The space $\mathbb{S}^n_{++}$ forms a manifold with its tangent space at a point $X \in \mathbb{S}^n_{++}$ denoted by $T_X\mathbb{S}^n_{++}$ and identified with $\mathbb{S}^n$, the set of symmetric matrices. The affine invariant metric at $X \in \mathbb{S}^n_{++}$ defined by

$$\langle V_1, V_2 \rangle_X = \mathrm{Tr}\{X^{-1}V_1X^{-1}V_2\} \quad V_1, V_2 \in T_X\mathbb{S}^n_{++} \tag{4.15}$$

turns the manifold into a Riemannian manifold. The shortest path on the manifold between two points $X, Y \in \mathbb{S}^n_{++}$ is called the geodesic curve and is parameterized as

$$\gamma(s) = X^{\frac{1}{2}} \left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}} \right)^s X^{\frac{1}{2}} \quad s \in [0, 1] \tag{4.16}$$

wherein, $\gamma(0) = X$ and $\gamma(1) = Y$ denote the end points of the geodesic. A geodesic curve emanating from a point $X \in \mathbb{S}^n_{++}$ in the direction $V \in T_X\mathbb{S}^n_{++}$ is parameterized by

$$\gamma_{X,V}(s) = X^{\frac{1}{2}} \mathrm{Exp}\left(sX^{-\frac{1}{2}} V X^{-\frac{1}{2}}\right) X^{\frac{1}{2}} \tag{4.17}$$

and resides within $\mathbb{S}^n_{++}$ for any $s \in \mathbb{R}$. Given a smooth function $f : \mathbb{S}^n_{++} \to \mathbb{R}$, $\bar{f}$ as the extension of f to $\mathbb{R}^{n\times n}$, a smooth geodesic curve $\gamma : \mathbb{R} \to \mathbb{S}^n_{++}$ such that $\gamma(0) = X \in \mathbb{S}^n_{++}$ and $\dot{\gamma}(0) = V \in T_X\mathbb{S}^n_{++}$, the Euclidean gradient $\nabla\bar{f}$ defined using the directional derivative $\mathrm{D}\bar{f}(X)[V]$ of $\bar{f}$ at X in the direction V is given as

$$\mathrm{Tr}\{V\,\nabla\bar{f}(X)\} = \mathrm{D}\bar{f}(X)[V] \tag{4.18}$$

The Riemannian gradient of f at X, denoted by $\mathrm{grad}\, f(X) \in T_X\mathbb{S}^n_{++}$ is similarly defined as

$$\langle V, \mathrm{grad} f(X)\rangle_X = \frac{d}{dt} f(\gamma(t))\big|_{t=0} \tag{4.19}$$

Note that the Riemannian gradient is obtained from the Euclidean gradient by

$$\mathrm{grad} f(X) = X \mathrm{sym}(\nabla\bar{f}(X))X \tag{4.20}$$

From [74, Section 4.1.4], the Riemannian Hessian of f defined as a map $\mathrm{Hess} f(X) : T_X\mathbb{S}^n_{++} \to T_X\mathbb{S}^n_{++}$ is given by

$$\mathrm{Hess} f(X)[V] = \mathrm{D}(\mathrm{grad} f)(X)[V] - \mathrm{sym}(\mathrm{grad} f(X) X^{-1} V) \tag{4.21}$$

Using the above expressions for $\mathrm{grad} f$, the Hessian can be expressed in terms of the extension $\bar{f}$ as

$$\mathrm{Hess} f(X)[V] = X \mathrm{sym}(\mathrm{D}(\nabla\bar{f})(X)[V])X + \mathrm{sym}(V \mathrm{sym}(\nabla\bar{f})X) \tag{4.22}$$

4.3.2 Cost function, Gradient and Hessian

The cost function for the recursive least squares minimization from Eq. (4.13) is minimized with a Riemannian optimization framework. The recursive least squares cost function is given by

$$J_k(\theta) = \frac{1}{2}(D_k\theta - b_k)^T R_W^{-1}(D_k\theta - b_k) + \frac{1}{2}(\theta - \hat{\theta}_{k-1})^T \Psi_{k-1}^{-1}(\theta - \hat{\theta}_{k-1}) \quad (4.23)$$

Before evaluating the Riemannian gradient and the Riemannian Hessian, the cost function must be reformatted to explicitly depend on Q and R matrices. Such reformatting is possible via simple algebraic manipulation. The unique elements of a SPD matrix are given by

$$\mathrm{vech}(X) = \begin{bmatrix} \boldsymbol{I}_n^0 X e_i \\ \vdots \\ \boldsymbol{I}_n^{n-1} X e_n \end{bmatrix} = \underbrace{\begin{bmatrix} \boldsymbol{I}_n^0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_n^1 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{I}_n^{n-1} \end{bmatrix}}_{\triangleq \mathcal{I}_n \in \mathbb{R}^{n(n+1)/2 \times n^2}} (\boldsymbol{I}_n \otimes X)\mathrm{vec}(\boldsymbol{I}_n) \quad (4.24)$$

wherein, $X \in \mathbb{S}^n$, $e_i \in \mathbb{R}^n$ is the i^{th} canonical basis vector and $\boldsymbol{I}_n^i \in \mathbb{R}^{(n-i)\times n}$ is formed by deleting the first i rows of $\boldsymbol{I}_n$, the identity matrix. The following statement provides the expressions for the gradient of the least squares cost function.

Lemma 2. *Given the cost function in Eq.* (4.23)*, its Riemannian gradients at Q and R are given by*

$$\mathrm{grad}J_k(Q,R) = (Q\nabla_Q \bar{J}_k Q, R\nabla_R \bar{J}_k R) \quad (4.25)$$

wherein, $\bar{J}_k$ is the Euclidean extension of the cost J_k and the expressions for $\nabla_Q \bar{J}_k$

and $\nabla_Q \bar{J}_k$ *are given by*

$$\nabla_Q \bar{J}_k = \mathrm{BTr}_n \left\{ \mathrm{sym} \left(\mathrm{vec}(\boldsymbol{I}_n) \left(D_k^{Q^T} R_W^{-1} (D_k \theta - b_k) \right.\right.\right. \\ \left.\left.\left. + \ [\boldsymbol{I}_{m_q}, \boldsymbol{0}_{m_q \times m_r}] \Psi_{k-1}^{-1} (\theta - \hat{\theta}_{k-1}) \right)^T \mathfrak{I}_n \right) \right\} \tag{4.26}$$

$$\nabla_R \bar{J}_k = \mathrm{BTr}_p \left\{ \mathrm{sym} \left(\mathrm{vec}(\boldsymbol{I}_p) \left(D_k^{R^T} R_W^{-1} (D_k \theta - b_k) \right.\right.\right. \\ \left.\left.\left. + \ [\boldsymbol{0}_{m_r \times m_q}, \boldsymbol{I}_{m_r}] \Psi_{k-1}^{-1} (\theta - \hat{\theta}_{k-1}) \right)^T \mathfrak{I}_p \right) \right\} \tag{4.27}$$

wherein, $D_k = [D_k^Q, D_k^R]$*,* $\theta = [\mathrm{vech}(Q)^T, \mathrm{vech}(R)^T]^T$*,* $m_q = n(n+1)/2$ *and* $m_r = p(p+1)/2$*.*

Proof. Consider a geodesic $\gamma_{Q,V_Q}(t)$ as defined in Eq. (4.17). From the definition of the gradient in Eq. (4.18), the expression for $\nabla_Q \bar{J}_k$ is given by

$$\mathrm{D}\bar{J}_k(Q,R)[V_Q,V_R] = \frac{\partial \bar{J}_k(\gamma_{Q,V_Q}(s), \gamma_{R,V_R}(s))}{\partial s}\Big|_{s=0} = \mathrm{Tr}\{\nabla_Q \bar{J}_k V_Q\} + \mathrm{Tr}\{\nabla_R \bar{J}_k V_R\}$$

Separating the expressions into the parts containing Q and R, we get

$$\begin{aligned} \mathrm{Tr}\{\nabla_Q \bar{J}_k V_Q\} &= (D_k \theta - b_k)^T R_W^{-1} \left(D_k^Q \mathfrak{I}_n \boldsymbol{I}_n \otimes V_Q \mathrm{vec}(\boldsymbol{I}_n) \right) \\ &\quad + (\theta - \hat{\theta}_{k-1})^T \Psi_{k-1}^{-1} [\boldsymbol{I}_{m_q}, \boldsymbol{0}_{m_q \times m_r}]^T \mathfrak{I}_n \boldsymbol{I}_n \otimes V_Q \mathrm{vec}(\boldsymbol{I}_n) \\ &= \mathrm{Tr}\{\mathrm{vec}(\boldsymbol{I}_n) \left(D_k^{Q^T} R_W^{-1} (D_k \theta - b_k) + [\boldsymbol{I}_{m_q}, \boldsymbol{0}_{m_q \times m_r}] \Psi_{k-1}^{-1} (\theta - \hat{\theta}_{k-1}) \right)^T \mathfrak{I}_n \boldsymbol{I}_n \otimes V_Q\} \end{aligned}$$

Further simplification results in an expression given by

$$\mathrm{Tr}\{\nabla_Q \bar{J}_k V_Q\} = \mathrm{Tr} \left\{ \mathrm{BTr}_n \left\{ \mathrm{vec}(\boldsymbol{I}_n) \left(D_k^{Q^T} R_W^{-1} (D_k \theta - b_k) \right.\right.\right. \\ \left.\left.\left. + [\boldsymbol{I}_{m_q}, \boldsymbol{0}_{m_q \times m_r}] \Psi_{k-1}^{-1} (\theta - \hat{\theta}_{k-1}) \right)^T \mathfrak{I}_n \right\} V_Q \right\}$$

wherein, the expression $\mathrm{Tr}\{A(\boldsymbol{I}_n \otimes B)\} = \mathrm{Tr}\{\mathrm{BTr}_n\{A\}B\}$ is used. Comparing the expressions on both sides of the equations gives the result of the lemma. The expression for $\nabla_R \bar{J}_k$ results from a derivation similar to the one above and is omitted. □

The expression for the Riemannian Hessian of J_k can be derived from Eq. (4.22) and is given through the following statement.

Lemma 3. *The expression for the Riemannian Hessian of J_k from Eq.* (4.22) *is given by*

$$\mathrm{Hess}J_k(Q,R)[V_Q,V_R] = (\mathrm{Hess}_Q J_k(Q,R)[V_Q,V_R], \mathrm{Hess}_R J_k(Q,R)[V_Q,V_R]) \tag{4.28}$$

wherein

$$\mathrm{Hess}_Q J_k(Q,R)[V_Q,V_R] = Q\mathrm{sym}(\mathrm{D}(\nabla_Q \bar{J}_k)(Q,R)[V_Q,V_R])Q + \mathrm{sym}(V_Q \mathrm{sym}(\nabla_Q \bar{J}_k)Q)$$

$$\mathrm{Hess}_R J_k(Q,R)[V_Q,V_R] = R\mathrm{sym}(\mathrm{D}(\nabla_R \bar{J}_k)(Q,R)[V_Q,V_R])R + \mathrm{sym}(V_R \mathrm{sym}(\nabla_R \bar{J}_k)R)$$

The expressions for the directional derivatives of the Euclidean gradients are given by

$$\begin{aligned}\mathrm{D}(\nabla_Q \bar{J}_k)(Q,R)[V_Q,V_R] = \mathrm{BTr}_n \Big\{ \mathrm{sym}\Big(\mathrm{vec}(\boldsymbol{I}_n)\theta_v^T \Big(D_k^{Q^T} R_W^{-1} D_k \\ + [\boldsymbol{I}_{m_q}, \mathbf{0}_{m_q \times m_r}]\Psi_{k-1}^{-1} \Big)^T \mathcal{I}_n \Big) \Big\}\end{aligned} \tag{4.29}$$

and

$$\begin{aligned}\mathrm{D}(\nabla_R \bar{J}_k)(Q,R)[V_Q,V_R] = \mathrm{BTr}_p \Big\{ \mathrm{sym}\Big(\mathrm{vec}(\boldsymbol{I}_p)\theta_v^T \Big(D_k^{R^T} R_W^{-1} D_k \\ + [\mathbf{0}_{m_r \times m_q}, \boldsymbol{I}_{m_r}]\Psi_{k-1}^{-1} \Big)^T \mathcal{I}_p \Big) \Big\}\end{aligned} \tag{4.30}$$

wherein, $\theta_v = [\mathrm{vech}(V_Q)^T, \mathrm{vech}(V_R)^T]^T$.

Proof. The directional derivative of $\nabla_Q \bar{J}_k(Q,R)[V_Q,V_R]$ is obtained as

$$D(\nabla_Q \bar{J}_k)(Q,R)[V_Q,V_R] = \frac{\partial \nabla_Q \bar{J}_k(\gamma_{Q,V_Q}(s), \gamma_{R,V_R}(s))}{\partial s}\Big|_{s=0} \tag{4.31}$$

Since the gradient is affine in Q and R, the directional derivative is independent of the points Q and R where it is evaluated. The expression is obtained trivially by substituting V_Q and V_R in place of Q and R and removing the constant terms. □

4.3.3 Riemannian Trust-Region Method

The Riemannian trust-region (RTR) method is used to solve the quadratic least squares cost function in Eq. (4.23). At each step the RTR method performs an inner iteration that minimizes a quadratic approximation of a cost function at Q,R given by

$$\begin{aligned}\hat{m}_{(Q,R)}(V_Q,V_R) = J_k(Q,R) + \langle \mathrm{grad} J_k(Q,R), (V_Q,V_R)\rangle_{(Q,R)} \\ + \frac{1}{2}\langle \mathrm{Hess} J_k(Q,R)[V_Q,V_R], (V_Q,V_R)\rangle_{(Q,R)}\end{aligned} \tag{4.32}$$

The optimal $V_Q^* \in T_Q\mathbb{S}_{++}^n$ and $V_R^* \in T_R\mathbb{P}_p$ are obtained subject to a norm constraint on step size given by

$$\|(V_Q,V_R)\|_{(Q,R)} = \sqrt{\langle (V_Q,V_R),(V_Q,V_R)\rangle_{(Q,R)}} \leq \Delta \tag{4.33}$$

wherein, Δ is the trust-region radius. A truncated conjugate gradient (tCG) method [1, Algorithm 11] solves the inner iteration at each step. Then a verification step evaluates the decrease in the true and approximate cost function given by the ratio

$$\rho = \frac{J_k(Q,R) - J_k(\gamma_{Q,V_Q^*}(1), \gamma_{R,V_R^*}(1))}{\hat{m}_{(Q,R)}(\mathbf{0}_{n\times n}, \mathbf{0}_{p\times p}) - \hat{m}_{(Q,R)}(V_Q^*, V_R^*)} \tag{4.34}$$

and decide whether the optimal (V_Q^*, V_R^*) are accepted and whether the radius Δ should be decreased. Algorithm 2 describes the RTR algorithm. The constants used are taken from [1, Algorithm 10].

Remark 4.3.1. *The RTR algorithm ensures that the estimates are symmetric and positive definite. However, in practice, the SPD noise covariance estimates may be arbitrarily close to semidefiniteness. This may create numerical errors in the filter*

Algorithm 2 Riemannian Trust-Region Method

```
Input: Q̂_{k−1}, R̂_{k−1}, Ψ_{k−1}, D_k, b_k, R_W, Δ̄ > 0, Δ_1 ∈ (0, Δ̄), ρ ∈ [0, 1/4)
Initialization: (Q̂_k)_1 = Q̂_{k−1}, (R̂_k)_1 = R̂_{k−1}
Output: Q̂_k, R̂_k
for k = 1 → n do
    Minimize m̂_{(Q,R)}(V_Q, V_R)                                ▷ Eq. (4.32)
    subject to norm constraint with Δ_i                          ▷ Eq. (4.33)
    if ρ_i < 1/4 then
        Δ_{i+1} = (1/4) Δ_i
    else if ρ_i > 3/4 and ‖(V*_Q)_i, (V*_R)_i‖ = Δ_i then
        Δ_{i+1} = min(2Δ_i, Δ̄)
    else
        Δ_{i+1} = Δ_i
    end if
    if ρ_i > ρ_min then
        ((Q̂_k)_{i+1}, (R̂_k)_{i+1}) = (γ_{Q,V*_Q}(1), γ_{r,V*_R}(1))
    else
        ((Q̂_k)_{i+1}, (R̂_k)_{i+1}) = ((Q̂_k)_i, (R̂_k)_i)
    end if
end for
```

updates. To avoid this situation, the minimum eigenvalue of $\hat{Q}_k$ and $\hat{R}_k$ is lower bounded by a small positive constant $\varepsilon > 0$. The modified optimization variable is given by

$$\begin{aligned} \hat{Q}_k &= \varepsilon \boldsymbol{I}_n + \hat{Q}_k^{\varepsilon} \\ \hat{R}_k &= \varepsilon \boldsymbol{I}_p + \hat{R}_k^{\varepsilon} \end{aligned} \tag{4.35}$$

Such a modification ensures that the eigenvalues of the noise covariance estimates obtained by the RTR method are lower bounded by ε instead of zero. Such a modification merely results in a shift of the origin and does not affect the RLS solution.

The algorithm for the RTR-based AKF is summarized below.

Algorithm 3 Riemannian Trust-Region based Adaptive Kalman Filter (RTRAKF)

```
Input: x̂_0, Q̂_0, R̂_0, P̂_0, Ψ_0, m, P, y_i, i = 1,2,...
Output: Q̂_k, R̂_k, P̂_k, x̂_k
for k = 1 → n do
    if i > m + P then
        Calculate D_k and b_k                          ▷ Eq. (4.12)
        Calculate the Riemannian Gradient              ▷ Eq. (4.25)
        Calculate the Riemannian Hessian               ▷ Eq. (4.28)
        Use Algorithm 2 to obtain Q̂_i and R̂_i
        Update x̂_k and P̂_k                             ▷ Eq. (4.14)
    end if
end for
```

4.4 Stability Analysis

In this section, the main contributions of this paper, i.e., stability of the RTR-based covariance estimation scheme and the adaptive Kalman filter using the RTR-based covariance estimates is presented.

4.4.1 Convergence of the noise covariance estimates

The RTR method, by design, ensures that $\hat{Q}_k$ and $\hat{R}_k$ are SPD. Starting from SPD initial guesses the following results establish the convergence of the RTR-based noise covariance estimators by comparing them to the RLS solution.

Proposition 4. *Given that D_k is persistently excited,* $\Pr\{\hat{Q}_i^{RLS} \in \mathbb{S}_{++}^n, \hat{R}_k^{RLS} \in \mathbb{P}_p, \forall i > k\} \xrightarrow{k\to\infty} 1$.

Proof. The convergence of the batch least squares estimate $\hat{\theta}_k$ in the mean squared sense to the true value θ^* was established given that the combined coefficient matrix $D = [D_1^T, D_2^T, \ldots]^T$ is full column rank [53, Theorem 8]. Since, D_k is persistently excited, the full rank condition is automatically satisfied. Since, convergence in the mean squared sense implies convergence in probability, we know that $\Pr\{\|\hat{\theta}_k - \theta^*\| > 0\} \to 0$. Consequently, for any constant $\delta > 0$, $\Pr\{\|\hat{\theta}_k - \theta^*\| < \delta\} \xrightarrow{k\to\infty} 1$. We know that the true Q and R which are formed from the θ^* elements are SPD. Hence, there exists a δ such that $\forall\, \hat{\theta}_k : \|\hat{\theta}_k - \theta^*\| < \delta$, $\hat{\theta}_k$ is such that the matrices $\hat{Q}_k$ and $\hat{R}_k$ formed by its elements are SPD. Picking such a δ ensures that $\Pr\{\hat{Q}_k^{RLS} \in \mathbb{S}_{++}^n, \hat{R}_k^{RLS} \in \mathbb{P}_p\} \xrightarrow{k\to\infty} 1$ which in turn ensures the following statement.

$$\Pr\{\exists i > k, \hat{Q}_i^{RLS} \notin \mathbb{S}_{++}^n, \hat{R}_i^{RLS} \notin \mathbb{P}_p\} \xrightarrow{k\to\infty} 0.$$

The negation of the above statement proves the statement of the proposition. □

Proposition 5. *Given $\hat{Q}_k \in \mathbb{S}_{++}^n$, $\hat{R}_k \in \mathbb{P}_p$, $\Psi_k \in \mathbb{P}_{m_q+m_r}$, and the one step RLS and RTR solutions denoted by $(\hat{Q}_{k+1}^{RLS}, \hat{R}_{k+1}^{RLS})$ and $(\hat{Q}_{k+1}^{RTR}, \hat{R}_{k+1}^{RTR})$ respectively, if $\hat{Q}_{k+1}^{RLS} \in \mathbb{S}_{++}^n$ and $\hat{R}_{k+1}^{RLS} \in \mathbb{P}_p$ then $(\hat{Q}_{k+1}^{RLS}, \hat{R}_{k+1}^{RLS}) = (\hat{Q}_{k+1}^{RTR}, \hat{R}_{k+1}^{RTR})$.*

Proof. The cost function given in Eq. (4.23) is quadratic with a positive definite Euclidean Hessian and is hence convex in the argument θ. As a result, the recursive least squares minimizer produces unique solutions up to the error due to the stopping criterion. Similarly, the choice of constants in Algorithm 2 and the usage of exact Hessian ensures that $\lim_{k\to\infty} \operatorname{grad} J_k = 0$ for the RTR algorithm [1, Theorem 7.4.4]. Since, $\hat{Q}_k \in \mathbb{S}^n_{++}$ and $\hat{R}_k \in \mathbb{P}_p$, $\operatorname{grad} J_k = 0 \implies (\nabla_Q \bar{J}_k, \nabla_R \bar{J}_k) = (0,0)$. Hence, this solution exactly matches the solution from the RLS step up to the error induced due by the stopping criterion. The statement of the proposition follows. □

Theorem 6. *Given that D_k is persistently excited, the sequences $\hat{Q}_k^{RTR}$ and $\hat{R}_k^{RTR}$ found using the Algorithm 2 converge to their true values, Q and R respectively, in probability.*

Proof. From Proposition 4, we know that $\Pr\{\hat{Q}_i^{RLS} \in \mathbb{S}^n_{++}, \hat{R}_i^{RLS} \in \mathbb{P}_p, \forall i > k\} \xrightarrow{k\to\infty} 1$. Hence, from Proposition 5, we get that $\Pr\{(\hat{Q}_k^{RLS}, \hat{R}_k^{RLS}) \neq (\hat{Q}_k^{RTR}, \hat{R}_k^{RTR})\} \xrightarrow{k\to\infty} 0$. Since the RLS solution converges to the true value in the mean squares sense and the RTR solution matches the least squares solution with probability 1 as $k \to \infty$, the RTR solution converges in probability to (Q,R). □

4.4.2 Convergence of the state error covariance matrix

The stability properties of the adaptive Kalman filter using the RTR-based noise covariance estimates are established through the following statement.

Proposition 7. *Given the noise covariance estimates $\hat{Q}_k$ and $\hat{R}_k$ from the RTR Algorithm described in Algorithm 2, the adaptive Kalman filter from Eq.* (4.14) *is* *exponentially stable.*

Proof. The proof follows from [8, Theorem 5.3] and Definitions 4.2.1 and 4.2.2. The observability Gramian for the pair $(F_k, \hat{R}_k^{-\frac{1}{2}} H_k)$ corresponding to the adaptive Kalman filter is given by

$$\hat{M}_{k+m,k} = \sum_{i=k}^{k+m} \phi_{i,k}^T H_i^T \hat{R}_i^{-1} H_i \phi_{i,k}$$

Since, $\hat{R}_k \succ 0$ and the observability Gramian $M_{k+m,k}$ for the known case is SPD, the observability Gramian $\hat{M}_{k+m,k}$ for the adaptive Kalman filter using the RTR noise covariance matrix estimates is also SPD. Hence, the pair $(F_k, \hat{R}_k^{-\frac{1}{2}} H_k)$ is uniformly observable. Given that $\hat{Y}_{k+s,k} \succ \mathbf{0}$, the controllability Gramian for the pair $(F_k, \hat{Q}_k^{\frac{1}{2}})$ corresponding to the adaptive Kalman filter for the same $s > 0$ is given by

$$\hat{Y}_{k+s,k} = \sum_{i=k}^{k+s} \phi_{i,k} \hat{Q}_i \phi_{i,k}^T$$

Since $\hat{Q}_k \succ 0$, the $\hat{Y}_{k+s,k}$ is non-singular and the pair $(F_k, \hat{Q}_k^{\frac{1}{2}})$ is uniformly controllable. Hence from [8, Theorem 5.3], the adaptive Kalman filter is exponentially stable. □

The following statement establishes the convergence of the state error covariance matrix of the adaptive Kalman filter.

Theorem 8. *The state error covariance matrix sequence $\hat{P}_k$ of the adaptive Kalman filter converges to the state error covariance matrix ${}^oP_k$ of the optimal Kalman filter with probability* 1.

Proof. Consider three covariance sequences $\hat{P}_k$, P_k and ${}^oP_k$ given by [109]

$$\hat{P}_{k+1} = \hat{\bar{F}}_k \hat{P}_k \hat{\bar{F}}_k^T + \hat{K}_k \hat{R}_k \hat{K}_k^T + \hat{Q}_k \tag{4.36}$$

$$P_{k+1} = \hat{\bar{F}}_k P_k \hat{\bar{F}}_k^T + \hat{K}_k R \hat{K}_k^T + Q \tag{4.37}$$

$${}^oP_{k+1} = \bar{F}_k {}^oP_k \bar{F}_k^T + K_k R K_k^T + Q \tag{4.38}$$

wherein, $\bar{F}_k = F_k - \hat{K}_k H_k$, $\hat{\bar{F}}_k = F_k - \hat{K}_k H_k$,

$$\hat{K}_k = F_k \hat{P}_k H_k^T (H_k \hat{P}_k H_k^T + \hat{R}_k)^{-1}$$

$$K_k = F_k {}^oP_k H_k^T (H_k {}^oP_k H_k^T + R)^{-1}$$

Each of the three sequences denotes the one-step predictor state covariance matrix. The sequence $\hat{P}_k$ denotes the apparent state error covariance matrix of the adaptive Kalman filter and uses the noise covariance matrix estimates for its propagation. The sequence P_k denotes the actual state error covariance matrix of the adaptive Kalman filter and uses the Kalman gain from the apparent covariance sequence along with the true noise covariance matrices. The sequence ${}^oP_k$ denotes the optimal state error covariance which represents the case when Q and R are fully known. We will first prove the equivalence of $\hat{P}_k$ and P_k in the limit. Assuming the same error covariance at the initial time, the sequence formed by differencing $\hat{P}_k$ and P_k is given by

$$\hat{P}_{k+1} - P_{k+1} = (F_K - \hat{K}_k H_k)(\hat{P}_k - P_k)(F_K - \hat{K}_k H_k)^T + \hat{K}_k(\hat{R}_k - R)\hat{K}_k^T + (\hat{Q}_k - Q)$$

Since the RTR method ensures that $\hat{Q}_k$ and $\hat{R}_k$ are SPD, we conclude that $F_k - \hat{K}_k H_k$ is exponentially stable from Proposition 7. Additionally, from Theorem 6, both

$\hat{Q}_k$ and $\hat{R}_k$ converge in probability to Q and R respectively as $k \to \infty$. Hence, the exponentially stable matrix sequence converges to zero in probability, i.e., $\hat{P}_k \xrightarrow[\mathcal{P}]{k\to\infty} P_k$. From the expression for $\hat{K}_k$, since $\hat{P}_k \xrightarrow[\mathcal{P}]{k\to\infty} P_k$ and $\hat{R}_k \xrightarrow[\mathcal{P}]{k\to\infty} R$, we get

$$\hat{K}_k \xrightarrow[\mathcal{P}]{k\to\infty} F_k P_k H_k^T (H_k P_k H_k^T + R)^{-1}$$

as $k \to \infty$. Hence, the matrix sequence for P_k and ${}^oP_k$ is identical in the limit as $k \to \infty$ with probability 1. Invoking Proposition 7, the state transition matrix $F_k - K_k H_k$ is exponentially stable. The matrix sequence ${}^oP_k$ has a unique limit [73, Theorem 7.5]. Hence, $\hat{P}_k \xrightarrow[\mathcal{P}]{k\to\infty} {}^oP_k$ as $k \to \infty$. Finally, we conclude that $\hat{P}_k \xrightarrow[\mathcal{P}]{k\to\infty} {}^oP_k$ as $k \to \infty$.

□

4.5 Numerical Simulations

A LTV system is simulated in this section to demonstrate the RTR based adaptive Kalman filter. The dynamics of the system given in Eq. (4.1) with the system matrices given below [53].

$$F_k = \begin{bmatrix} 0 & 1 \\ -a_k b_k & -(a_k + b_k) \end{bmatrix} \quad G_k = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad H_k = \begin{bmatrix} 1 & d_k \end{bmatrix} \tag{4.39}$$

wherein, $\{a_k, b_k\} = c_k \pm i(0.4 + 0.2\sin(2\pi k/\tau)$, $c_k = -0.7 + 0.2\cos(2\pi k/\tau)$, $d_k = 2sin(10\pi k/\tau)$, and i is the imaginary unit. The measurements are assumed to be available every $1/\tau$ second with $\tau = 10^4$. The true noise covariance matrices are given by $Q = \begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$ and $R = 2$. For the purposes of the simulation, the control inputs u_k are assumed to be drawn from a unit normal distribution. The number of measurements stacked at every time step are $m = 3$. Fig. 4.1 shows the Frobenius norm of the error in estimating the Q matrix with the RTR and the RLS method. The estimates from both methods are shown to converge to zero. A similar trend is seen in the estimation error Frobenius norms in estimating R in Fig. 4.2 and the error between the state covariances of the adaptive Kalman filter and the optimal Kalman filter with known noise covariance matrices shown in Fig. 4.3. The difference between the two methods is seen when comparing the transient $\hat{Q}_k$ eigenvalues shown in Fig. 4.4. The RLS method sometimes leads to a negative eigenvalue while RTR method lower bounds the eigenvalue by a prescribed minimum value of 0.1 (Remark 4.3.1). Since the R matrix is a scalar, A similar trend is seen in Fig. 4.5 that shows the time history of its estimate $\hat{R}_k$. The result of negative eigenvalues of the noise covariance matrix estimates culminates as an inconsistent non positive definite error state covariance $\hat{P}_{k+1|k}$ shown in Fig. 4.6.

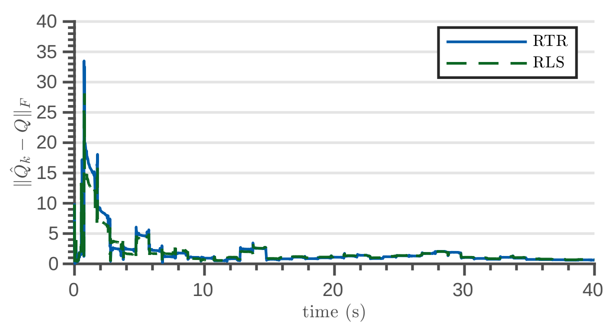

Figure 4.1: The Frobenius norm of the Q estimation error.

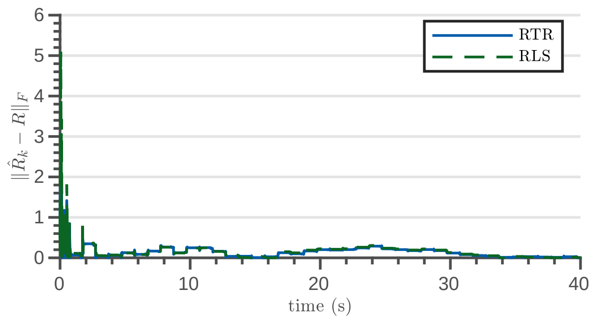

Figure 4.2: The Frobenius norm of the R estimation error.

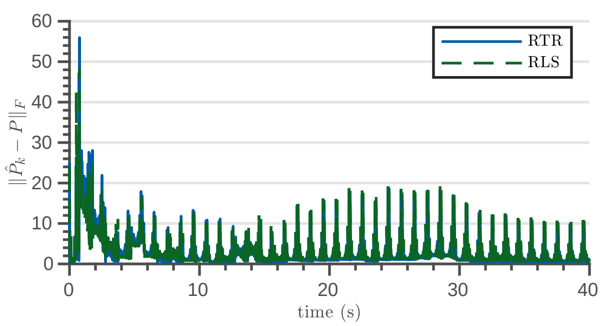

Figure 4.3: The Frobenius norm of the estimation error in the estimated state error covariance matrix $\hat{P}_{k|k=1}$ and the optimal ${}^{o}P_{k|k-1}$.

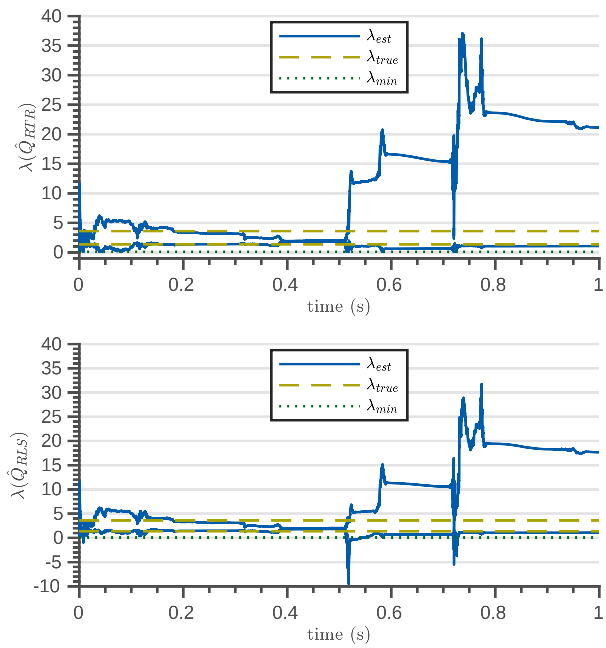

Figure 4.4: The transient eigenvalues of $\hat{Q}_k$, the true eigenvalues of Q, and $\lambda_{min} = 0.1$ from Remark 4.3.1.

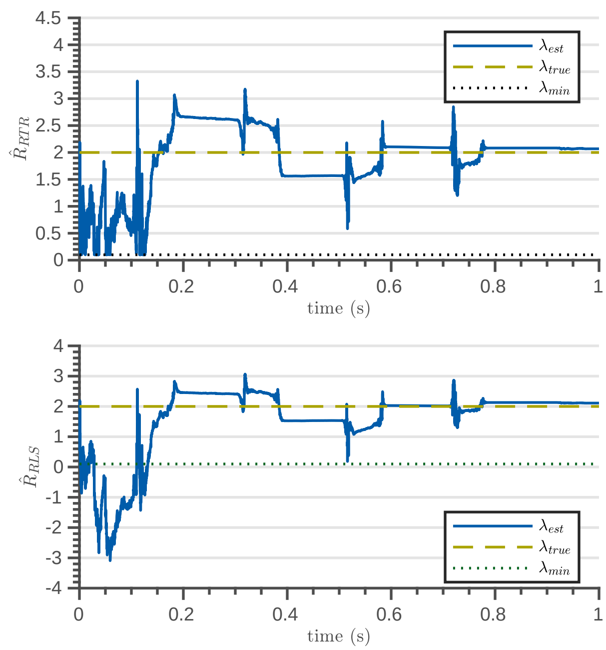

Figure 4.5: The transient values of $\hat{R}_k$, the true R, and $\lambda_{min} = 0.1$ from Remark 4.3.1.

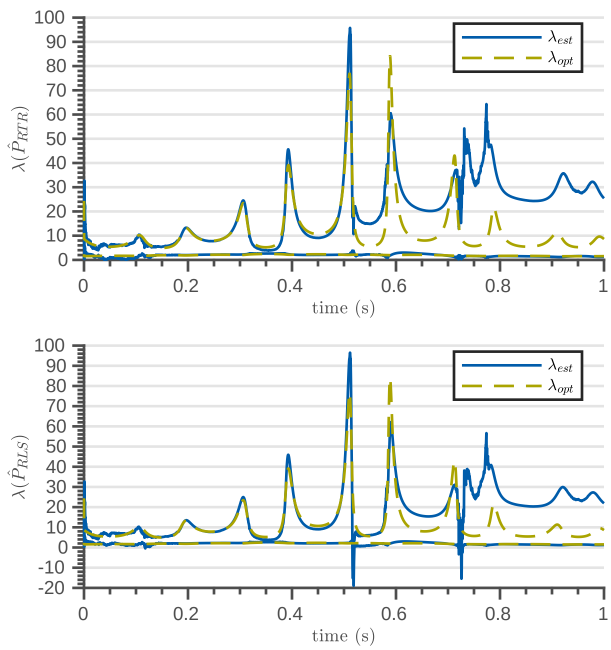

Figure 4.6: The transient eigenvalues of predicted state error covariance for the RTR-based adaptive Kalman filter.

4.6 Conclusion

A Riemannian Trust-Region (RTR) based Adaptive Kalman filter to estimate the states as well as the process and measurement noise covariance matrices is introduced in this chapter. A linear time series is constructed using a fixed buffer of past measurements in time. The autocovariance function of the time series is a linear function of the noise covariance matrices. A recursive least squares cost function is minimized using the RTR method to obtain symmetric and positive definite (SPD) noise covariance matrix estimates. The noise covariance matrix estimates are shown to converge to their true values provided there exists sufficient excitation of the system. Furthermore, the stability of the Adaptive Kalman filter is established by proving uniform observability and uniform controllability. The conditions of sufficient excitation are the counterparts to the full rank conditions of the S matrix in Eq. (3.37) for the LTI case.

Chapter 5

Adaptive Identification in Control

This chapter presents an identification technique for SPD matrix valued parameters in adaptive control applications. The developments of this chapter focus on continuous time formulations of identifying SPD matrices as opposed to the discrete time formulations from the previous chapters. Continuous time identification schemes for SPD matrix valued parameters frequently appear in control problems.

5.1 Notations and Preliminaries

Some notations and preliminaries specific to this chapter are presented here. The attitude of a rigid body can be expressed as a member of the special orthogonal group $SO(n)$. The associated Lie-algebra of this set is represented by the set of skew-symmetric matrices

$$\mathfrak{so}(n) = \{K \in \mathbb{R}^{n\times n} | K = -K^T\}$$

The skew-symmetric map $S : \mathbb{R}^n \to \mathfrak{so}(n)$ maps a vector to a member of the Lie-algebra. For $n = 3$ this map outputs a skew-symmetric matrix as follows.

$$S(v) = \begin{bmatrix} 0 & v_3 & -v_2 \\ -v_3 & 0 & v_1 \\ v_2 & -v_1 & 0 \end{bmatrix} \tag{5.1}$$

A matrix X can always be split into its symmetric and skew-symmetric parts,

$$X = \underbrace{\frac{1}{2}(X + X^T)}_{\triangleq (X)_S} + \underbrace{\frac{1}{2}(X - X^T)}_{\triangleq (X)_K} \tag{5.2}$$

The element-wise product or the Hadamard product of matrices $X, Y \in \mathbb{R}^{m\times n}$ is given by

$$(X \circ Y)_{ij} = (X)_{ij}(Y)_{ij} \tag{5.3}$$

The set of symmetric, symmetric and positive semidefinite, and symmetric and positive definite matrices are denoted by $\mathbb{S}^n$, $\mathbb{S}^n_+$, and $\mathbb{S}^n_{++}$ respectively. Let $\sigma(X) = \{\lambda \,|\, \det(X - \lambda \boldsymbol{I}) = 0\}$ be the spectrum of the matrix X. The map Diag : $\mathbb{R}^n \to \mathbb{S}^n$ which takes a vector and returns a diagonal matrix with the elements of the vector on the diagonal. Similarly, the map diag : $\mathbb{R}^{n\times n} \to \mathbb{R}^n$ return the diagonal elements of a matrix as a vector.

5.2 The Projection Scheme

The projection of symmetric matrices onto the cone positive semidefinite matrices is well known [32] and is known to be the closest point in the positive definite cone to the symmetric matrix with respect to the Frobenius norm.

$$\mathrm{P}_{\mathbb{S}^n_{++}}(X) = U\mathrm{Diag}(\max(0,\lambda_1),\ldots,\max(0,\lambda_n))U^T \tag{5.4}$$

wherein, $X = U\mathrm{Diag}(\lambda)U^T$ is an eigen decomposition. Consider two similarly defined projection within the cone of symmetric matrices, one with α lower bounding the minimum eigenvalue, and the other, with β upper bounding the maximum eigenvalue

$$
\begin{aligned}
\mathrm{P}^{min}_{\mathbb{S}^n}(X,\alpha) &= U\mathrm{Diag}(\max(\alpha,\lambda_1),\ldots,\max(\alpha,\lambda_n))U^T \\
\mathrm{P}^{max}_{\mathbb{S}^n}(X,\beta) &= U\mathrm{Diag}(\min(\beta,\lambda_1),\ldots,\min(\beta,\lambda_n))U^T
\end{aligned}
\tag{5.5}
$$

Note that $\mathrm{P}^{min}_{\mathbb{S}^n}(X,\alpha,\beta) = \mathrm{P}_{\mathbb{S}^n}(X-\alpha\boldsymbol{I})+\alpha\boldsymbol{I}$ and $\mathrm{P}^{max}_{\mathbb{S}^n}(X,\beta) = \beta\boldsymbol{I} - \mathrm{P}_{\mathbb{S}^n}(\beta\boldsymbol{I}-X)$ are alternative expressions for the projection. The projection for both lower and upper bounds is given by

$$
\mathrm{P}_{\mathbb{S}^n}(X,\alpha,\beta) = U\mathrm{Diag}(\lambda_1^{\alpha,\beta},\ldots,\lambda_n^{\alpha,\beta}))U^T \tag{5.6}
$$

wherein, $\lambda_i^{\alpha,\beta} = \min(\beta,\max(\lambda_i,\alpha))$ and $\mathrm{P}_{\mathbb{S}^n}(X,\alpha,\beta) = \mathrm{P}^{max}_{\mathbb{S}^n}(\mathrm{P}^{min}_{\mathbb{S}^n}(X,\alpha),\beta)$ is another expression. Consider a set of eigenvalues of X that are greater than a positive constant α indexed by the set given below.

$$
\begin{aligned}
\Omega(X,\alpha,\beta) &= \{k \mid \alpha \le \lambda_k(X) \le \beta\} \\
\Omega_-(X,\alpha) &= \{k \mid \lambda_k(X) < \alpha\} \\
\Omega_+(X,\beta) &= \{k \mid \lambda_k(X) > \beta\}
\end{aligned}
\tag{5.7}
$$

The notations Ω, Ω_-, and Ω_+ are used whenever the arguments are non ambiguous. The above sets partition the set $\{1,\ldots,n\} \in \mathbb{N}$. The above projection is differentiable at all points except for at $\lambda_i(X) \in \{\alpha,\beta\}$ for all i. It is in fact $\mathcal{C}^\infty$ at all points where it is differentiable [97, 131].

5.2.1 Stability with the projection mechanism

In various applications of adaptive control, the stability of the system is established using the Lyapunov direct method. For the case of unknown symmetric matrix, the Lyapunov candidate function for the states is augmented with the Lyapunov function for the error in the symmetric matrix parameters

$$V(t) = V_e(t) + V_J(t)$$

wherein, the $V_e(t)$ contains the state error term while $V_J(t)$ contains the error in the inertia matrix. In most cases, a squared term of the vector of unique elements of the symmetric matrix is added to the Lyapunov function.

In this section, a new Lyapunov term is proposed that can be augmented to the Lyapunov functions for a wide variety of applications. In the following let J be symmetric and positive definite matrix to be estimated. The matrix $\bar{J} \in \mathbb{S}^n$ as the matrix that tracks J and the estimate $\hat{J}$ as the projection of $\bar{J}$ given the eigenvalue bounds

$$\hat{J} = \mathrm{P}_{\mathbb{S}^n}(\bar{J}, \alpha, \beta) \quad \alpha, \beta > 0$$

Furthermore, the eigen decompositions for the matrices defined above are

$$\begin{aligned} J &= V \Sigma V^T \\ \bar{J} &= U \Lambda U^T \\ \hat{J} &= U \Lambda' U^T \end{aligned} \tag{5.8}$$

wherein, Λ' is the diagonal matrix with entries trimmed to lie within (α, β). A new formulation for $V_J(t)$ is presented below through a series of lemmas.

Lemma 9. *Let J be such that it satisfies Assumption 10 with parameter α and/or β and $\hat{J} = \mathrm{P}_{\mathbb{S}^n}(\bar{J}, \alpha, \beta)$ is the projected matrix. Let*

$$V_J(t) = \mathrm{Tr}\{(J - \bar{J})^2 - (\bar{J} - \hat{J})^2\} \tag{5.9}$$

Given the above definition,

$$V_J(t) \geq 0$$

with equality whenever $J = \bar{J}$.

Proof. Consider two cases for the eigenvalues of $\bar{J}$,

1. All the eigenvalues of $\bar{J}$ are within $[\alpha, \beta]$

2. Some of the eigenvalues are outside the interval $[\alpha, \beta]$

Case 1: Since $\Omega_-(\bar{J}, \alpha)$ and $\Omega_+(\bar{J}, \beta)$ are empty sets, $\hat{J}$ is equal to $\bar{J}$.

$$V_J(t) = \text{Tr}\{(J - \bar{J})(J - \bar{J})\} \geq 0 \tag{5.10}$$

Since, $J - \bar{J}$ is symmetric, the only way $V_J(J, \bar{J})$ vanishes is when J and $\bar{J}$ are equal.

Case 2: Consider the following algebraic manipulation.

$$\begin{aligned}
V_J(t) &= \text{Tr}\{(\bar{J} - J)^2 - (\bar{J} - \hat{J})^2\} \\
&= \text{Tr}\{(J^2 - 2J\bar{J} - \hat{J}^2 + 2\bar{J}\hat{J})\} \\
&= \text{Tr}\{(J^2 - 2J\hat{J} + \hat{J}^2 + 2J\hat{J} - \hat{J}^2 - 2J\bar{J} - \hat{J}^2 + 2\bar{J}\hat{J})\} \\
&= \text{Tr}\{\underbrace{(\hat{J} - J)^2}_{\succeq 0} + 2\underbrace{(J - \hat{J})(\hat{J} - \bar{J})}_{J_c}\}
\end{aligned}$$

The first term on the right hand side is the same as the first case and is non negative.

Using the eigen decomposition for the second term

$$\begin{aligned}
\text{Tr}\{J_c\} &= \text{Tr}\{(V\Sigma V^T - U\Lambda' U^T)(U\Lambda' U^T - U\Lambda U^T)\} \\
&= \text{Tr}\{(U^T V\Sigma V^T U - \Lambda')(\Lambda' - \Lambda)\} \\
&= \sum_{i \in \Omega_-} (d_{ii} - \alpha)(\alpha - \lambda_i) + \sum_{i \in \Omega_+} (d_{ii} - \beta)(\beta - \lambda_i)
\end{aligned}$$

wherein, all the terms in the index set $\Omega(\bar{J},\alpha,\beta)$ vanish, and d_{ii} are the diagonal terms of $U^T V\Sigma V^T U$. All the diagonal terms satisfy $d_{ii} \in [\alpha,\beta]$ by Assumption 10. Similarly, for the index set Ω_- and Ω_+, $\lambda_i < \alpha$ and $\lambda_i > \beta$ respectively. Hence, $\mathrm{Tr}\{J_c\}$ is non negative. Equality, in this case, occurs when J and $\hat{J}$ are equal. However, this condition is never true as the set union $\Omega_- \bigcap \Omega_+$ is non empty under Case two in general. Hence, the function $V_J(t) \geq 0$ with equality at $J = \bar{J}$. □

In applications involving Lyapunov functions, the derivative $\dot{V}(t)$ involves evaluating the derivative of $V_J(t)$ to chose an update law for $\bar{J}$. A general form of the update law for a system with states given by x commanded to follow a reference trajectory given by x_r and derivative of the reference trajectory given by $\dot{x}_r$

$$\dot{\bar{J}} = \Psi(x,x_r,\dot{x}_r) + \Gamma(\bar{J}-\hat{J}) + (\bar{J}-\hat{J})\Gamma \tag{5.11}$$

wherein, $\Psi(\cdot)$ is a symmetric matrix valued function and the projection $\hat{J}$, and Γ is a constant diagonal Hurwitz matrix. The particular definition of the function Ψ, states and their reference trajectories may vary for different applications. The derivative of the above Lyapunov function is bounded as shown in the following Lemma.

Lemma 10. *If $V_J(t)$ defined is as defined in Eq. (5.9), with an update law for $\bar{J}$ as given in Eq. (5.11) and the eigen decomposition defined in Eq. (5.8) then the following inequality is true almost everywhere.*

$$\dot{V}_J(t) \overset{a.e.}{\leq} 2\mathrm{Tr}\{(\hat{J}-J)\Psi(x,x_r,\dot{x}_r,\hat{J})\}$$

wherein, the $^{a.e.}$ stands for almost everywhere. This means that the derivative $\dot{V}_J(t)$ is defined at all points except on a set of Lebesgue measure zero.

Proof. Consider the expression for the derivative of $V_J(t)$ below with $\tilde{J}$ defined as the $\hat{J} - J$ error term.

$$
\begin{aligned}
\dot{V}_J(t) &= 2\mathrm{Tr}\{(\bar{J} - J)\dot{\bar{J}} - (\bar{J} - \hat{J})(\dot{\bar{J}} - \dot{\hat{J}})\} \\
&= 2\mathrm{Tr}\{\tilde{J}\Psi(x, x_r, \dot{x}_r, \hat{J})\} + 2\mathrm{Tr}\{2\Gamma\tilde{J}(\bar{J} - \hat{J}) + (\hat{J} - \bar{J})\dot{\hat{J}}\}
\end{aligned}
$$

The matrix $\hat{J}$ is formed by the new projection method. Note that the projection $\mathrm{P}_{\mathbb{S}^n}(X, \alpha, \beta)$ is differentiable at all points where α, β is does not belong to $\sigma(X)$. However, the set of points where the projection is not differentiable is of Lebesgue measure zero. A Clarke Generalized gradient has been used as a gradient at the points of non-differentiability [97]. The gradient of the projection map is given below [131].

$$
\nabla \mathrm{P}_{\mathbb{S}^n}(\bar{J}, \alpha, \beta)[H] = U(B(\mathrm{diag}(\Lambda)) \circ (U^T H U))U^T \tag{5.12}
$$

wherein, $H \in \mathbb{S}^n$ is the direction in which the gradient is evaluated, and the matrix valued function $B(\cdot)$ is defined below.

$$
\{B(x)\}_{ij} = \begin{cases} 1 & x_k \in (\alpha, \beta), k = i, j \\ 0 & x_k \notin (\alpha, \beta), k = i, j \\ \frac{x_i^{\alpha,\beta} - x_j^{\alpha,\beta}}{x_i - x_j} & \text{o.w.} \end{cases} \tag{5.13}
$$

wherein, $x_i^{\alpha,\beta} = \min(\beta, \max(x_i, \alpha))$. Since $\bar{J} \in \mathbb{S}^n_{++}$, $\dot{\bar{J}} \in \mathbb{S}^n$ and the derivative of $\hat{J}$ wherever it exists is defined below.

$$
\dot{\hat{J}} = \nabla \mathrm{P}_{\mathbb{S}^n}(\bar{J}, \alpha, \beta)[\dot{\bar{J}}] \tag{5.14}
$$

Similar to the treatment in Lemma 9, the cross terms evaluate to

$$
\mathrm{Tr}\{\Gamma\tilde{J}(\bar{J} - \hat{J})\} = \sum_{i \in \Omega_-} \Gamma_{ii}(d_{ii} - \alpha)(\alpha - \lambda_i) + \sum_{i \in \Omega_+} \Gamma_{ii}(d_{ii} - \beta)(\beta - \lambda_i)
$$

wherein, both the summations are negative due to the negative diagonal Γ_{ii} terms multiplying positive terms. The final term, denoted by $\text{Tr}\{J_x\} = \text{Tr}\{(\hat{J}-\bar{J})\dot{\hat{J}})\}$, evaluates to the following expression at all points except for those in a set of 0 measure.

$$\text{Tr}\{J_x\} \overset{\text{a.e.}}{=} \text{Tr}\{(\Lambda' - \Lambda)(B(\text{diag}(\Lambda)) \circ (U^T \dot{\bar{J}} U))\}$$

Let $b_{ij} = (B(\text{diag}(\Lambda)) \circ U^T \dot{\bar{J}} U)_{ij}$ denote the elements of the matrix.

$$\text{Tr}\{J_x\} \overset{\text{a.e.}}{=} \sum_{i \in \Omega_-} b_{ii}(\alpha - \lambda_i) + \sum_{i \in \Omega_+} b_{ii}(\lambda_i - \beta)$$

The Trace expression vanishes since $b_{ii} = 0$ for $i \in \Omega_- \bigcap \Omega_+$ almost everywhere and the statement of the lemma follows.

□

5.3 Adaptive Control Applications

In this section three different applications of adaptive control with unknown symmetric matrices are considered. First, the adaptive angular velocity tracking problem is considered [37]. Next, the full adaptive attitude control problem with both orientation and angular velocity tracking is examined [160]. Lastly, the passivity-based adaptive control of robotic manipulator system is modified with the novel projection [133].

5.3.1 Adaptive Angular Velocity Tracking

The dynamical equations of motion are first described. Let $\omega(t)$ represents the angular velocity of the rigid body expressed in the body fixed frame, the unknown $J \in \mathbb{S}^3_{++}$ be its inertia matrix and $u(t) \in \mathbb{R}^3$ be the control input. The dynamics of the angular velocity of the rigid body is described in its body fixed frame as [37]

$$J\dot{\omega}(t) = -S(\omega)J\omega + u(t) \tag{5.15}$$

For the angular velocity tracking problem, the reference angular velocity $\omega_r(t)$ is assumed to be available in the body fixed frame. Since the angular velocity of the body is also expressed in the body fixed frame, the angular velocity error is expressed as

$$\omega_e(t) = \omega(t) - \omega_r(t) \tag{5.16}$$

and the error dynamics are given by

$$J\dot{\omega}_e = -S(\omega)J\omega + u(t) - J\dot{\omega}_r \tag{5.17}$$

Such a problem is relevant for system identification maneuvers in space applications.

Assumption 10. *The eigenvalues of the true inertia matrix J are bounded. Either the knowledge of either the lower bound is available in the form of a constant α,*

$$\lambda_{min}(J) \geq \alpha$$

or the knowledge of the upper bound is available in the form of a constant $\beta > \alpha$,

$$\lambda_{max}(J) \leq \beta$$

or both the bounds are available.

An adaptive controller to track the angular velocity in presence of an unknown or uncertain inertia matrix that satisfies Assumption 10 is formulated below. Consider the angular velocity tracking problem with the angular velocity error dynamics given in Eq. (5.17) with the adaptive control law given by

$$u(t) = -K\omega_e + S(\omega)\hat{J}\omega + \hat{J}\dot{\omega}_r \tag{5.18}$$

wherein, $K \in \mathbb{S}^n_{++}$ is a positive definite gain matrix and $\hat{J} = \mathrm{P}_{\mathbb{S}^n}(\bar{J}, \alpha, \beta)$ is the projection defined in Eq. (5.6). Consequently, the update law for $\bar{J}$ is designed as

$$\dot{\bar{J}} = -\gamma(\Xi(\omega, \omega_r, \dot{\omega}_r))_S + \Gamma(\bar{J} - \hat{J}) + (\bar{J} - \hat{J})\Gamma \tag{5.19}$$

wherein, $\Xi(\cdot) = \omega\omega_e^T S(\omega_r) + \dot{\omega}_r\omega_e^T$ denoted by Ξ for brevity below. Note that the update law for $\dot{\bar{J}}$ is in the form specified in Eq. (5.11). The stability properties of the closed loop system under the above update law are proved in the following theorem.

Theorem 11. *Consider a rigid body with error dynamics given by Eq. (5.17), the adaptive control law in Eqs. (5.18), and inertia update specified by Eq. (5.19). Then under Assumptions 10, the angular velocity error ω_e converges to zero asymptotically.*

Proof. Consider the following Lyapunov candidate function.

$$V(t) = \frac{1}{2}\omega_e^T J \omega_e + \frac{1}{2\gamma} V_J(t) \tag{5.20}$$

wherein, the expression for $V_J(t)$ is given in Eq. (5.9). From Lemma 9, $V(t)$ is known to be non negative and that $V(t)$ vanishes whenever ω_e and $\tilde{J} = \hat{J} - J$ vanish.

Since, the derivative of $V_J(t)$ does not exist on a set of measure 0, the derivative of $V(t)$ can be evaluated as follows almost everywhere.

$$\dot{V}(t) \overset{\text{a.e.}}{=} \omega_e^T J \dot{\omega}_e + \frac{1}{2\gamma} \dot{V}_J(t)$$

Substituting for the $\dot{V}_J(t)$ and the equations of motion, and using Lemma 10

$$\begin{aligned}
\dot{V}(t) &\overset{\text{a.e.}}{=} -\omega_e^T K \omega_e + \omega_e^T S(\omega) \tilde{J} \omega + \omega_e^T \tilde{J} \dot{\omega}_r + \frac{1}{2\gamma} \dot{V}_J(t) \\
&\overset{\text{a.e.}}{=} -\omega_e^T K \omega_e + \mathrm{Tr}\{\tilde{J} \omega \omega_e^T S(\omega_r) + \tilde{J} \dot{\omega}_r \omega_e^T\} + \frac{1}{2\gamma} \dot{V}_J(t) \\
&\overset{\text{a.e.}}{\leq} -\omega_e^T K \omega_e + \mathrm{Tr}\{\tilde{J}(\Xi)_K\}
\end{aligned}$$

Since, the trace of the product of a symmetric matrix and a skew-symmetric matrix is zero, the derivative of the Lyapunov candidate function evaluates to be non positive almost everywhere.

$$\dot{V}(t) \overset{\text{a.e.}}{\leq} -\omega_e^T K \omega_e \tag{5.21}$$

Since, the Lyapunov candidate function is not differentiable at all the points, the Lyapunov direct theorem for stability does not directly apply here. However, the non smooth Lyapunov function presented above satisfies the requirements of the corollaries of LaSalle-Yoshizawa wherein $\dot{V}(t)$ is upper bounded by a negative semidefinite function almost everywhere. Hence, the following convergence result is proved [59, 132, Collorary 1].

$$\omega_e \overset{t \to \infty}{\longrightarrow} 0$$

□

5.3.2 Adaptive Attitude Tracking

An adaptive controller for the full attitude control problem is formulated using the projection method described in Section 5.2. The dynamic equations are first stated. Although the quaternion representation is used for the attitude in this chapter, the convergence result with the projection can be similarly proved for formulations using other representations. Let q be the unit norm quaternion representing the attitude of the system in the inertial frame of reference. Let q_0 and $\boldsymbol{q}_v$ be the scalar and vector parts of the quaternion respectively. The Euler parameter kinematic differential equation is as given in [77].

$$\dot{\boldsymbol{q}}(t) = \frac{1}{2}E(q(t))\omega(t) = \frac{1}{2}\begin{bmatrix} -\boldsymbol{q}_v^T(t) \\ q_0\boldsymbol{I} + S(\boldsymbol{q}_v) \end{bmatrix}\omega(t) \tag{5.22}$$

Here, $\omega(t)$ represents the angular velocity of the rigid body expressed in the body fixed frame and $\boldsymbol{I}$ denotes the 3×3 identity matrix. The attitude dynamics of the rigid body are governed by Eq. 5.15 The rotation matrix corresponding to the attitude, also known as the Direction Cosine Matrix (DCM), is defined below.

$$C(\boldsymbol{q}) = (q_0^2 - \boldsymbol{q}_v^T\boldsymbol{q}_v)\boldsymbol{I} + 2\boldsymbol{q}_v\boldsymbol{q}_v^T - 2q_0S(\boldsymbol{q}_v) \tag{5.23}$$

Let q_r be the target reference quaternion expressed in the inertial frame. The DCM going from the reference frame to the body fixed frame is expressed as follows.

$$C(\boldsymbol{q}_e) = C(\boldsymbol{q})C(\boldsymbol{q}_r)^T \tag{5.24}$$

wherein, $\boldsymbol{q}_e = [q_{e_0}, \boldsymbol{q}_{e_v}^T]^T$ represents the error quaternion between the body fixed frame and the reference frame. The angular velocity tracking error is expressed in the reference frame as

$$\omega_e = \omega - C(\boldsymbol{q}_e)\omega_r \tag{5.25}$$

Consequently, the error quaternion and angular velocity dynamics are given by

$$\begin{aligned} \dot{\boldsymbol{q}}_e &= \frac{1}{2}E(\boldsymbol{q}_e)\omega_e \\ J\dot{\omega}_e &= -S(\omega)J\omega + u + J(S(\omega_e)C(\boldsymbol{q}_e)\omega_r - C(\boldsymbol{q}_e)\dot{\omega}_r) \end{aligned} \tag{5.26}$$

Let the adaptive control law be defined as

$$u(t) = -(K_2K_1 + \boldsymbol{I}_{3\times3})\boldsymbol{q}_{e_v} - K_2\omega_e + S(\omega)\hat{J}\omega - \hat{J}\Phi(\boldsymbol{q}_e, \omega_e, \omega_r, \dot{\omega}_r) \tag{5.27}$$

wherein,

$$\Phi(\cdot) = S(\omega_e)C(\boldsymbol{q}_e)\omega_r - C(\boldsymbol{q}_e)\dot{\omega}_r + \frac{1}{2}K_1(q_{e_0}\omega_e + S(\boldsymbol{q}_{e_v})\omega_e)$$

denoted by Φ for brevity, and $K_1, K_2 \in \mathbb{S}^3_{++}$ are the controller gains. The estimate $\hat{J}$ of the inertia matrix is defined using the projection as

$$\hat{J} = \mathrm{P}_{\mathbb{S}^3}(\bar{J}, \alpha, \beta)$$

wherein, $\bar{J} \in \mathbb{S}^n$ is the unbounded symmetric matrix used to define the estimate. The update law for $\bar{J}$ is chosen as

$$\dot{\bar{J}} = -\gamma(\mathfrak{S}(\boldsymbol{q}_e, \omega_e, \omega_r, \dot{\omega}_r))_S + \Gamma(\bar{J} - \hat{J}) + (\bar{J} - \hat{J})\Gamma \tag{5.28}$$

wherein,

$$\mathfrak{S}(\cdot) = \omega\omega_e^T S(C(\boldsymbol{q}_e)\omega_r) + K_1\boldsymbol{q}_{e_v}^T S(\omega) - \Phi\,(\omega_e + K_1\boldsymbol{q}_{e_v})^T$$

$\gamma > 0$ is a positive constant, and $\Gamma \prec 0$ is a diagonal Hurwitz matrix. Note that the update law for $\dot{\bar{J}}$ is once again in the form specified in Eq. (5.11). The stability of the adaptive controller is established through the following theorem.

Theorem 12. *Consider the rigid body with dynamics given in Eq. (5.26) with the adaptive control law in Eqs. (5.27) and inertia update specified by Eq. (5.28). Assume that the inertia matrix J satisfies Assumption 10 with for some known constants α or β or both. Then the states $\boldsymbol{q}_{e_v}$ and ω_e converge to zero asymptotically.*

Proof. Consider the Lyapunov candidate function given below.

$$V(t) = \frac{1}{2}(\omega_e + K_1\boldsymbol{q}_{e_v})^T J(\omega_e + K_1\boldsymbol{q}_{e_v}) + \boldsymbol{q}_{e_v}^T\boldsymbol{q}_{e_v} + (q_0 - 1)^2 + \frac{1}{2\gamma}V_J(t) \tag{5.29}$$

wherein, the expression for $V_J(t)$ is given in Eq. (5.9). From Lemma 9, $V_J(t)$ is non negative and it vanishes whenever $\hat{J} = J$. The rest of the Lyapunov candidate function is also positive definite with equality at $\omega_e = 0$ and $\boldsymbol{q}_e = [1,0,0,0]^T$ unit quaternion. The derivative of $V_J(t)$ is evaluated at all the points where it is differentiable. Since the set of points has Lebesgue measure 0, the equality below is qualified with an $^{\text{a.e.}}$ which stands for *almost everywhere*. The derivative of $V(t)$ is given by

$$\dot{V}(t) \overset{\text{a.e.}}{=} (\omega_e + K_1\boldsymbol{q}_{e_v})^T J(\dot{\omega}_e + K_1\dot{\boldsymbol{q}}_{e_v}) - 2\dot{q}_0 + \frac{1}{2\gamma}\dot{V}_J(t)$$

Expanding the above expression using the error dynamics and the control law

$$\dot{V}(t) \overset{\text{a.e.}}{=} -\boldsymbol{q}_{e_v}^T K_1\boldsymbol{q}_{e_v} - (\omega_e + K_1\boldsymbol{q}_{e_v})^T K_2(\omega_e + K_1\boldsymbol{q}_{e_v}) + \text{Tr}\{\tilde{J}\mathfrak{S}(\boldsymbol{q}_e, \omega_e, \omega_r, \dot{\omega}_r)\} + \frac{1}{2\gamma}\dot{V}_J(t)$$

The identity $\omega_e^T S(\omega_e)\tilde{J}\omega = 0$ was used for simplification. Substituting the inequality for $\dot{V}_J(t)$ from Lemma 10

$$\dot{V}(t) \overset{\text{a.e.}}{\leq} -\boldsymbol{q}_{e_v}^T K_1\boldsymbol{q}_{e_v} - (\omega_e + K_1\boldsymbol{q}_{e_v})^T K_2(\omega_e + K_1\boldsymbol{q}_{e_v}) + \text{Tr}\{\tilde{J}(\mathfrak{S}(\boldsymbol{q}_e, \omega_e, \omega_r, \dot{\omega}_r))_K\}$$

The trace of the product of a symmetric matrix and a skew-symmetric matrix is zero, the derivative of the Lyapunov candidate function evaluates to be non positive almost everywhere.

$$\dot{V}(t) \overset{\text{a.e.}}{\leq} -\boldsymbol{q}_{e_v}^T K_1 \boldsymbol{q}_{e_v} - (\omega_e + K_1 \boldsymbol{q}_{e_v})^T K_2 (\omega_e + K_1 \boldsymbol{q}_{e_v}) \tag{5.30}$$

The corollaries of LaSalle-Yoshizawa are again applied for the non smooth Lyapunov function, wherein $\dot{V}(t)$ is upper bounded by a negative semidefinite function almost everywhere. Such a theorem ensures the following convergence result [59, 132, Collorary 1]

$$\boldsymbol{q}_{e_v} \overset{t\to\infty}{\longrightarrow} \mathbf{0} \quad \text{and} \quad \omega_e \overset{t\to\infty}{\longrightarrow} \mathbf{0}$$

□

5.3.3 Adaptive control of robotic manipulator

Next, the problem of adaptive control of a n-link robotics manipulator system is considered. The dynamics of a n-link manipulator are given by [139]

$$M(q,\Theta)\ddot{q} + C(q,\dot{q},\Theta)\dot{q} + G(q,\Theta) = \tau, \tag{5.31}$$

wherein, $q \in \mathbb{R}^n$ is a vector of generalized coordinates, and $M(\cdot) \in \mathbb{S}^n_{++}$, $C(\cdot) \in \mathbb{R}^{n\times n}$, and $G(\cdot) \in \mathbb{R}^n$ denote the mass matrix, the Coriolis and the centrifugal terms, and the gravitational force vector respectively, $\tau \in \mathbb{R}^n$ is the motor torque input, and $\Theta = [\theta_1^T, \dots, \theta_n^T]^T \in \mathbb{R}^{10n}$ is a vector of unknown inertial parameters, wherein the unknown inertial parameters of each link consists of

$$\theta_i = [I_i^{xx}, I_i^{yy}, I_i^{zz}, I_i^{xy}, I_i^{yz}, I_i^{zx}, \boldsymbol{h}_i, m_i]^T \in \mathbb{R}^{10}$$

wherein, m_i, $\boldsymbol{h}_i \in \mathbb{R}^3$, and $I_i \in \mathbb{S}^3_{++}$ are the mass, first mass moment (mass multiplied by the location of the center of mass), and the second mass moment along any point in the body fixed frame. The inertia matrix and its inverse is assumed to uniformly positive definite

$$b_1\boldsymbol{I} \preceq M(q,\Theta) \preceq b_2\boldsymbol{I}$$

for some $b_1, b_2 > 0$ and C matrix is chosen such that $\dot{M} - 2C$ is skew-symmetric. The parameters defined above form the 4×4 pseudo-inertia matrix introduced in [161]

$$J_i = \begin{bmatrix} \frac{1}{2}\mathrm{Tr}\{I_i\} - I_i & \boldsymbol{h}_i \\ \boldsymbol{h}_i^T & m \end{bmatrix} \in \mathbb{S}^4_{++} \tag{5.32}$$

The physical consistency of the estimates of the inertial parameters was shown to be equivalent to the positive definiteness of J_i, the pseudo-inertia matrix [161]. The elements of the pseudo-inertia matrix are only dependent on the distribution of the link mass and independent of the configuration. Additionally, the pseudo-inertia is exhaustive in terms of parameters needed to define the inertial properties of the rigid link. In order to ensure physical consistency of the inertia estimates, eigenvalue bounds are imposed on the pseudo-inertia matrix for a link. Consequently, the following assumption can be made for the robotic manipulator.

Assumption 11. *Let there exist a set $\mathfrak{I} \subset \{1,2,\ldots,n\}$ wherein, n is the number of links of a manipulator, such that for all $i \in \mathfrak{I}$, the pseudo-inertia matrix for link i defined in Eq. (5.32) satisfies Assumption 10 with some constants α_i and β_i.*

Remark 5.3.1. *Every link of the robotic manipulator has an associated pseudo-inertia matrix. The adaptive control law formulated below does not require the* *knowledge of any eigenvalue bounds to be stable. However, given the bounds for some or all of the links, the adaptive control law can benefit from this information. At the very least, the pseudo-inertia matrix for every link can be constrained to have positive eigenvalues ($\alpha_i = 0$ and $\beta_i = \infty$).*

A passivity-based adaptive controller is designed to control the manipulator [133]

$$\begin{aligned} \tau &= M(q,\hat{\Theta})\ddot{q}_r + C(q,\dot{q},\hat{\Theta})\dot{q}_r + G(q,\hat{\Theta}) - K_D s \\ &= Y(q,\dot{q},\dot{q}_r,\ddot{q}_r)\hat{\Theta} - K_D s \end{aligned} \tag{5.33}$$

wherein, q_d is the joint reference trajectory,

$$\begin{aligned} \tilde{q} &= q - q_d \\ \dot{q}_r &= \dot{q}_d - \Upsilon\tilde{q} \\ s &= \dot{q} - \dot{q}_r = \dot{\tilde{q}} - \Upsilon\tilde{q} \end{aligned}$$

and the constant matrices $K_D \in \mathbb{S}^n_{++}$ and $\Upsilon \in \mathbb{S}^n_{++}$ is diagonal. The stability for the above controller with the new projection is proved through the following theorem.

Theorem 13. *Consider the system given in Eq. (5.31) with the passivity-based adaptive control law given in Eq. (5.33). Let the pseudo-inertia matrices of the links satisfy Assumption 11 with a set $\mathfrak{I}$ and the constants α_i, β_i where $i \in \mathfrak{I}$. Let $\bar{J}_i$ be designed to track the pseudo-inertia matrix J_i for all $i \in \mathfrak{I}$ and the corresponding projection operator be defined as $\hat{J}_i = \mathrm{P}_{\mathbb{S}^n}(\bar{J}_i, \alpha_i, \beta_i)$. The adaptive update law is given by*

$$\dot{\bar{J}}_i = -\gamma\mathfrak{B}_i(s,q,\dot{q},\dot{q}_r,\ddot{q}_r) + \Gamma(\bar{J}_i - \hat{J}_i) + (\bar{J}_i - \hat{J}_i)\Gamma \tag{5.34}$$

wherein $\mathfrak{B}_i(\cdot)$ is a symmetric matrix given by

$$\sum_{i=0}^{n} \mathrm{Tr}\{\tilde{J}_i \mathfrak{B}_i(s,q,\dot{q},\dot{q}_r,\ddot{q}_r)\} = \tilde{\Theta}^T Y^T(q,\dot{q},\dot{q}_r,\ddot{q}_r)s$$

for all links i under Assumption 11 with constants α_i and/or β_i. Under the above control law, $\tilde{q}$ converges to 0 as $t \to \infty$.

Proof. Consider the Lyapunov function given by

$$V(t) = \underbrace{s^T M(q,\Theta)s}_{\triangleq V_q(t)} + \frac{1}{2\gamma}\sum_{i=1}^{n} V_{J_i}(t) \tag{5.35}$$

It has been shown [161] that the derivate of the first part of the Lyapunov function is given by

$$\dot{V}_q(t) = -s^T K_D s + \tilde{\Theta}^T Y^T(q,\dot{q},\dot{q}_r,\ddot{q}_r)^T s \tag{5.36}$$

The derivative of the full Lyapunov function is given by

$$\dot{V}(t) \overset{\text{a.e.}}{=} -s^T K_D s + \sum_{i=1}^{n} \mathrm{Tr}\{\tilde{J}_i \mathfrak{B}_i\} + \frac{1}{2\gamma}\sum_{i=1}^{n} \dot{V}_{J_i}(t)$$

From Lemma 10, $\dot{V}_{J_i}(t) \overset{\text{a.e.}}{\leq} -2\gamma \mathrm{Tr}\{\tilde{J}_i \mathfrak{B}_i\}$ and hence

$$\dot{V}(t) \overset{\text{a.e.}}{\leq} -s^T K_D s \tag{5.37}$$

The LaSalle-Yoshizawa corollary [59, Collorary 1] for non smooth systems is then invoked to conclude that $s^T K_D s \to 0$ as $t \to \infty$. Consequently, as $s \to 0$, $\tilde{q}$ converges to a sliding stable surface given by

$$\dot{\tilde{q}} + \Upsilon \tilde{q} = \mathbf{0}$$

which in turn drives $\tilde{q} \to 0$. □

5.4 Computational Aspects of the Projection

Using the projection presented in this chapter requires computing the eigen decomposition of $\bar{J}$ at every time instant. The number of computations required for eigen decomposition of a matrix scales approximately with the cube of the dimension of the matrix. Such a computation may not be expensive for applications such as the three dimensional attitude control of spacecrafts. However, in robotics applications such as the attitude control of a n-link manipulator robot, evaluating the eigen decomposition may prove to be expensive. For this reason, a methodology to minimize the computation by directly updating the eigenvalues and the eigenvectors is provided in this section. The development of such a direct update follows from the perturbation theory of linear operators given by Kato [82].

Since, $\hat{J}$ was proven to be Lipschitz continuous [97] and assuming $\Psi(\cdot)$ is at least continuous, $\bar{J}$ is continuously differentiable. Even though $\hat{J}$ were to be a smooth projection, additional assumption would be needed on the reference trajectory ω_r and $\dot{\omega}_r$ in order to conclude smoothness of $\bar{J}$. The eigenvalues and eigenprojections of the time varying matrix $\bar{J}$ retain the continuous differentiability of $\bar{J}$ at all points except at exceptional points [82]. These are the points where the eigenvalues of the matrix split, cross or merge. Away from the exceptional points, a methodology to directly the eigenvalues and eigenprojections is presented below. Consequently, the following assumption is made.

Assumption 12. *Consider a time varying matrix $X(t) \in \mathbb{S}^n$ with the set of unique eigenvalues denoted by $\lambda_1, \ldots, \lambda_s$ wherein, $s \leq n$ at a given time t. In the neighborhood of t, $X(t)$ does not contain exceptional points, i.e., the multiplicities of the* *eigenvalues remain constant in the neighborhood of t. Equivalently, there is no splitting, merging or crossing of eigenvalues around t.*

Remark 5.4.1. *Note that the above assumption allows for repeated eigenvalues. However, the multiplicity of each repeated eigenvalues remains constant.*

In presence of repeated eigenvalues, the choice of eigenvectors is non unique as eigenvectors belong to a subspace of $\mathbb{R}^n$. However, the eigenprojections corresponding to the eigenvalues are unique. The eigenprojections are defined below.

Definition 5.4.1. *A projection matrix $P_i = P_i^2$ is known as the eigenprojection of the matrix X corresponding to the eigenvalue λ_i if and only if*

$$XP_i = \lambda_i P_i \quad i = 1, \ldots, s$$

For a real symmetric matrix $X \in \mathbb{S}^n$, the closed form expression for P_i is given by

$$P_i = \sum_{Xu_i = \lambda_i u_i} u_i u_i^T \tag{5.38}$$

wherein, u_i's are a orthonormal set of eigenvectors corresponding to the eigenvalue λ_i.

The following lemma enables direct eigenvalue and eigenprojection update.

Lemma 14. *Let $\bar{J}$, updated according to the update law given in Eq. (5.11), be differentiable at a given time t. Also, let the s unique eigenvalues of $\bar{J}$ and their corresponding eigenprojections be denoted by $\lambda_1, \ldots, \lambda_s$ and $P_1, \ldots, P_s$ respectively.*

If $\bar{J}$ satisfies Assumption 12 around t, then the first derivative of the unique eigenvalues and their eigenprojections are as given below.

$$\dot{\lambda}_i = \frac{\mathrm{Tr}\{P_i\dot{\bar{J}}\}}{\mathrm{Tr}\{P_i\}} \tag{5.39}$$

$$\dot{P}_i = P_i\dot{\bar{J}}S_i + S_i\dot{\bar{J}}P_i \tag{5.40}$$

wherein, $S_i = \sum_{j \neq i} \frac{P_j}{\lambda_i - \lambda_j}$.

Proof. Since, $\bar{J}$ is differentiable at t, so are the eigenvalues and the eigenprojections [82, Theorem 5.4]. The value of the eigenprojection in the neighborhood $\mathfrak{U}(t)$ is given by

$$P_i(t') = P_i + (t'-t)\dot{P}_i(t) + O((t'-t)^2)$$

wherein, $\dot{P}_i(t) = P_i\dot{\bar{J}}S_i + S_i\dot{\bar{J}}P_i$ and $S_i = \sum_{j \neq i} \frac{P_j}{\lambda_i - \lambda_j}$ is known as the value of the reduced resolvent of $\bar{J}$ at λ_i and $O(\cdot)$ is the big-O notation. The expression for the eigenvalue is similarly given by

$$\lambda_i(t') = \lambda_i + (t'-t)\dot{\lambda}_i(t) + O((t'-t)^2) \tag{5.41}$$

wherein, $\dot{\lambda}_i(t)$ is the repeated eigenvalue of $P_i\dot{\bar{J}}P_i$ defined as

$$P_i\dot{\bar{J}}P_iU_i = \dot{\lambda}_t(t)U_i$$

wherein, U_i contains an orthonormal set of eigenvectors of $P_i\dot{\bar{J}}P_i$ corresponding to the repeated eigenvalue $\dot{\lambda}_i$. Pre-multiplying by U_i^T and evaluating the trace

$$\mathrm{Tr}\{U_i^T P_i\dot{\bar{J}}P_iU_i\} = \dot{\lambda}_i n_i$$

wherein, n_i denotes the multiplicity of the $\dot{\lambda}_i$ eigenvalue. By definition, the trace of a eigenprojection is dimension of the eigen subspace. Hence, the expression for the first derivative simplifies to

$$\dot{\lambda}_i(t) = \frac{\mathrm{Tr}\{P_i\dot{\bar{J}}\}}{\mathrm{Tr}\{P_i\}}$$

wherein, the identity $P_i^2 = P_i$ was used for simplification. □

In general, since $\bar{J}$ is continuously differentiable, the number of exceptional points are at most countable in an infinite interval. Hence, the direct eigenvalue and eigenprojection update can be performed at all but countable points in any given interval. As the eigenvalues are directly available, the computational complexity of $O(n^3)$ eigen decomposition is eliminated.

The numerical simulations performed in the sequel use the above direct update to evaluate the eigenvalues and the eigenprojections at all the times regardless of the existence of exceptional points. Except for numerical errors accumulated over time, no difference was found in the eigenvalues calculated using the eigen decomposition and those calculated using the direct update law. In practical applications, exceptional points can be easily tracked by putting a threshold on the eigenvalues that approach each other. The eigen decompositions can also be evaluated at periodic intervals in time to minimize numerical errors.

5.5 Numerical Experiments

In this section, numerical simulation results showing the effectiveness of the novel projection scheme in adaptive control design are presented. The three problems presented in the previous sections are simulated here. The inertia matrix considered for the angular velocity tracking and the attitude tracking simulations is given below [2, 37].

$$J = \begin{bmatrix} 25 & 1.2 & 0.9 \\ 1.2 & 17 & 1.4 \\ 0.9 & 1.4 & 15 \end{bmatrix} \tag{5.42}$$

The initial angular velocity of the spacecraft is assumed to be $\boldsymbol{\omega} = [0.4, 0.2, -0.1]^T$ rad/s and its initial orientation with respect to the inertial frame (for attitude tracking only) is given by the quaternion $\boldsymbol{q} = [0.9837, -0.1037, 0.1037, -0.1037]^T$ and the initial value for both $\bar{J}$ and $\hat{J}$ is given by

$$\bar{J}(0) = \begin{bmatrix} 22 & 1.3 & 1.0 \\ 1.3 & 18 & 1.6 \\ 1.0 & 1.6 & 13 \end{bmatrix} \tag{5.43}$$

5.5.1 Adaptive Angular Velocity Tracking

The initial condition of $\omega_e(0) = [0, 0, 0]^T$ for the angular velocity was used. The reference angular velocity trajectory to be tracked is given by

$$\omega_r(t) = \begin{bmatrix} \sin(t) & \sin(2t) & \sin(3t) \end{bmatrix}^T \tag{5.44}$$

The parameters for the adaptive controller are given in Table 5.1. The angular velocity error norm and the inertia matrix estimation error norm are given in Figs. (5.2), and (5.3). The comparison of control histories was not provided since they were found to be similar. The eigenvalues of the inertia matrix estimate given in Fig. (5.4) show that the projection method ensures that the eigenvalues remain within the specified bounds.

K	Γ	α	β	γ
$150\boldsymbol{I}_{3\times 3}$	$-2.5\boldsymbol{I}_{3\times 3}$	14	26	100

Table 5.1: Controller parameters for angular velocity tracking of the sinusoidal velocity profile.

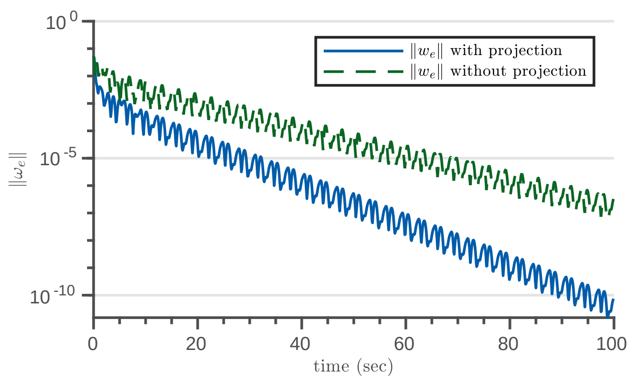

Figure 5.1: The norm of the angular velocity tracking error for angular velocity control problem with projection (top) and without projection (bottom).

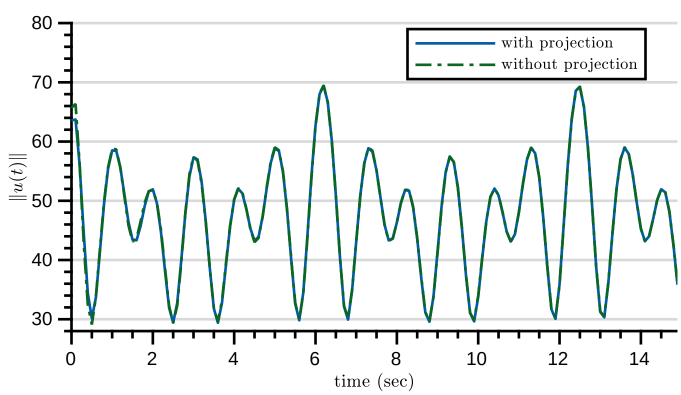

Figure 5.2: The norm of the control input for the angular velocity control problem with projection (top) and without projection (bottom).

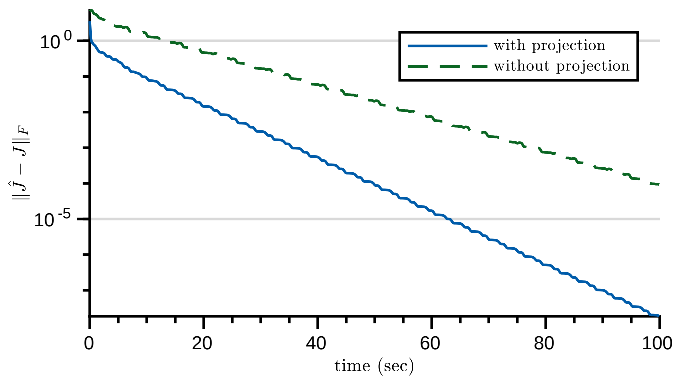

Figure 5.3: The norm of the inertia matrix estimation error for the adaptive angular velocity controller with (top) and without (bottom) projection.

K_1	K_2	Γ	α	β	γ
$10\boldsymbol{I}_{3\times3}$	$2.5\boldsymbol{I}_{3\times3}$	$-2.5\boldsymbol{I}_{3\times3}$	14	26	1

Table 5.2: Controller parameters for adaptive attitude tracking of a coning maneuver.

5.5.2 Adaptive Attitude Control

For the attitude tracking problem the spacecraft is commanded coning maneuver given in Ref. [2]. The motion of the desired reference frame is given by the 3-2-1 Euler angles trajectory given by

$$\psi(t) = 0.1t \text{ rad}$$
$$\theta(t) = -0.2222\pi \text{ rad}$$
$$\phi(t) = 0.5t \text{ rad}$$

and the reference trajectories are accordingly calculated. The parameters for the adaptive controller are given in Table 5.2 The initial estimate of $\bar{J}$ is as given in Eq. (5.43). The norm of the velocity part of the error quaternion and the angular

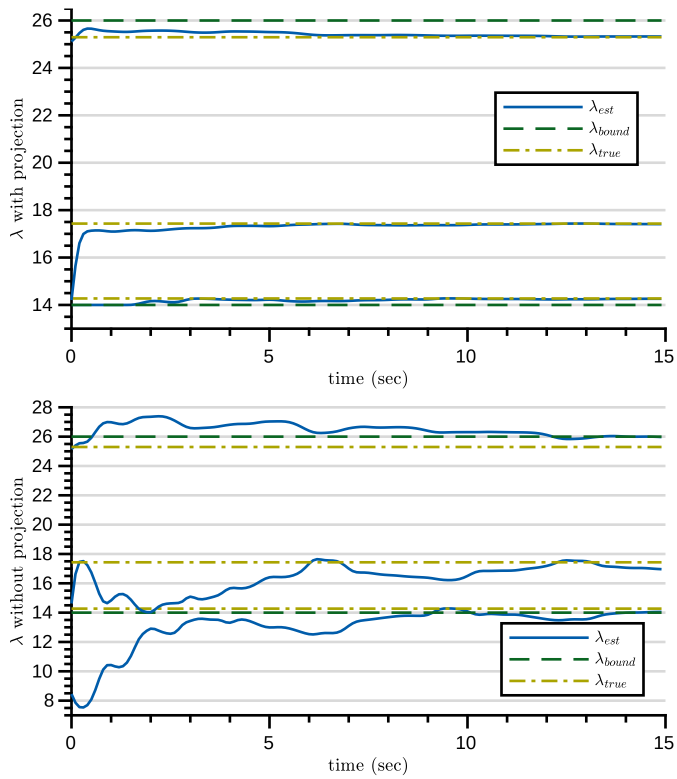

Figure 5.4: The eigenvalues of inertia matrix estimate for adaptive controller with (top) and without (bottom) projection. The bounds for the eigenvalues as well as the true eigenvalues are also provided.

velocity tracking error norms is given in Figs. (5.5) and (5.7). The inertia matrix estimation norm given in Fig. (5.8) shows a lower estimation error when using the new projection. The eigenvalues of the estimates given in Fig. (5.9) shows that the projection scheme ensures that the eigenvalues remain withing the specified bounds.

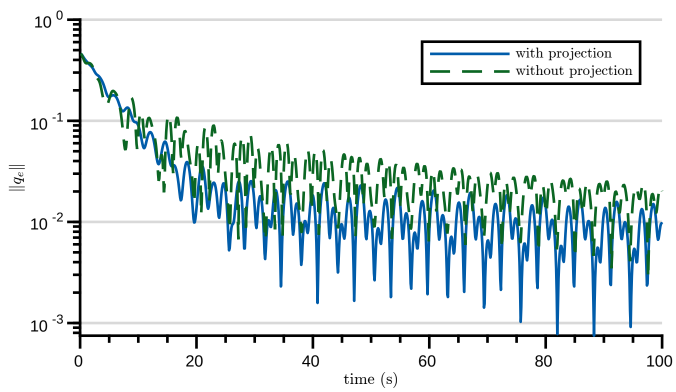

Figure 5.5: The norm of the error quaternion between the body fixed frame and the reference frame for the attitude control problem with (blue) and without (black) projection.

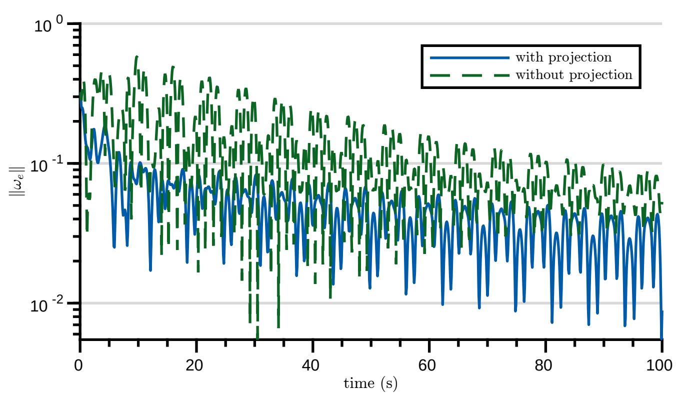

Figure 5.6: The norm of the angular velocity tracking error for attitude control problem with (blue) and without (black) projection.

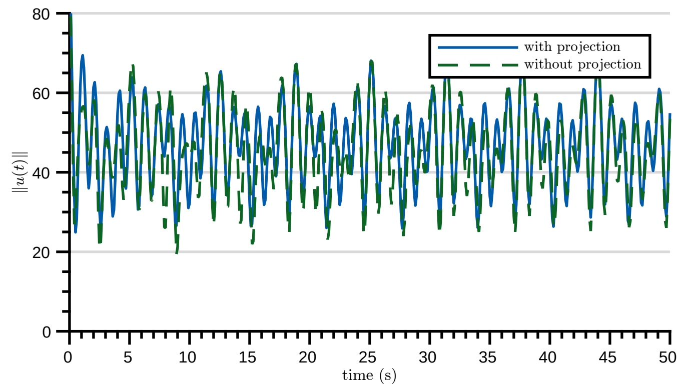

Figure 5.7: The norm of the control input for attitude control problem with (blue) and without (black) projection.

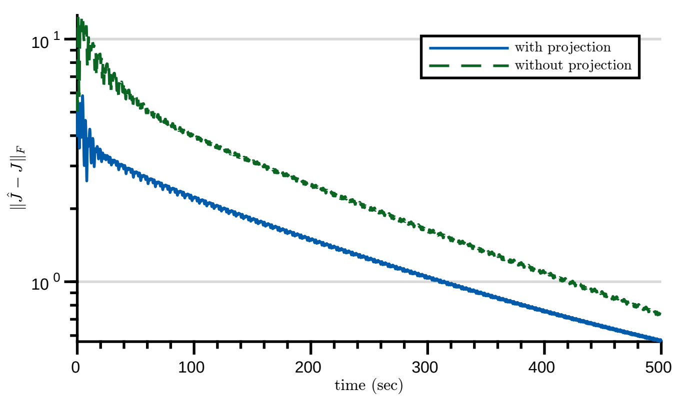

Figure 5.8: The Frobenius norm of inertia matrix estimate error for adaptive attitude controller with (blue) and without (black) projection.

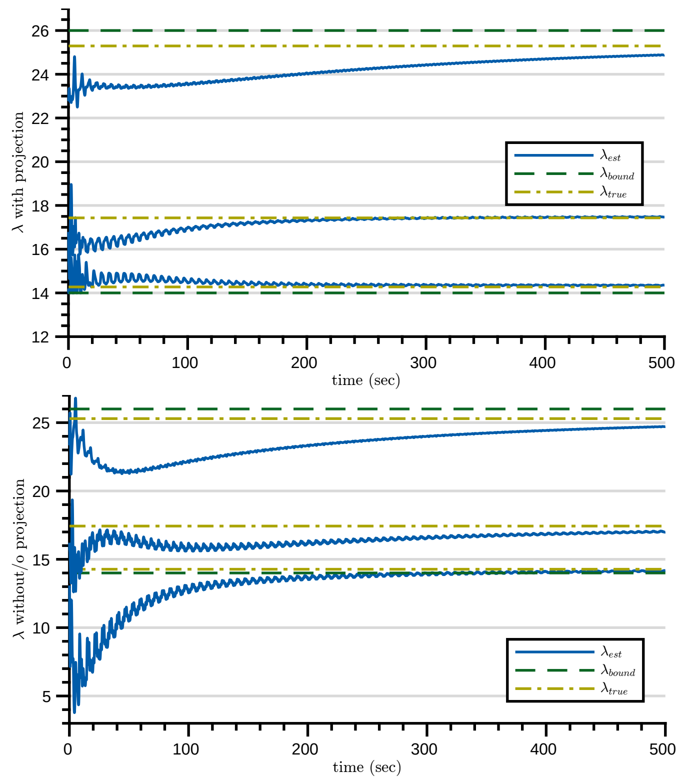

Figure 5.9: The eigenvalues of the inertia matrix estimate (blue), their true values (red) and the eigenvalue bounds (black) for adaptive attitude controller with (top) and without (bottom) projection.

K_D	Γ_i	α_i	β_i	γ_i
$50\boldsymbol{I}_{3\times3}$	$-2.5\boldsymbol{I}_{3\times3}$	0	1.5	1

Table 5.3: Parameters for the adaptive control of robotic manipulator with unknown inertial parameters. Note that the same parameters values are used for both the links $i = 1,2$.

5.5.3 Adaptive Control of Robotic Manipulator

A two link manipulator was commanded period reference trajectory for its joints specified by

$$
\begin{aligned}
q_{d_1}(t) &= 0.8\cos\left(\frac{2\pi t}{5}\right) \\
q_{d_2}(t) &= 0.4\cos\left(\frac{2\pi t}{3}\right)
\end{aligned} \tag{5.45}
$$

The parameters are specified in Table 5.3. The true values of the link pseudo-inertia matrices are given by

$$
J_1 = \begin{bmatrix} 0.59 & 0 & 0 & 0.4 \\ 0 & 0.25 & 0 & 0 \\ 0 & 0 & 1.25 & 0 \\ 0.4 & 0 & 0 & 1.0 \end{bmatrix} \tag{5.46}
$$

$$
J_2 = \begin{bmatrix} 0.39 & 0 & 0 & 0.6 \\ 0 & 0.25 & 0 & 0 \\ 0 & 0 & 1.25 & 0 \\ 0.6 & 0 & 0 & 1.0 \end{bmatrix} \tag{5.47}
$$

and the initial estimates were randomly generated with the true values as the mean and 0.05 as the variance. The norm of the joint angle error and the joint velocity error is given in Figs. (5.10) and (5.11). The errors in the pseudo-inertia matrix (Eq. (5.32)) for both the links is given Fig. (5.13). The comparison of the eigenvalues of the pseudo-inertia matrix estimates is shown in Figs. (5.14) and (5.15).

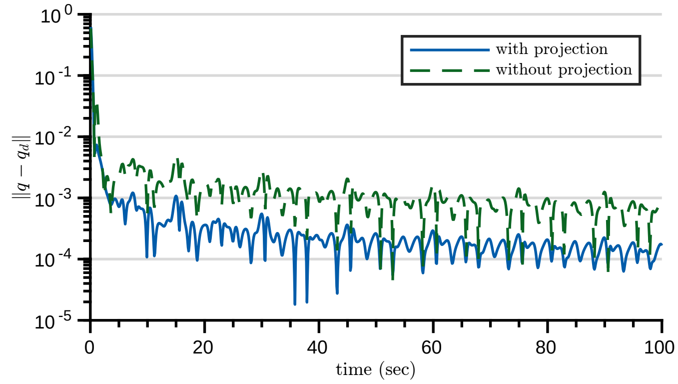

Figure 5.10: The comparison of the norm of the tracking error between the robot manipulator angles and the reference trajectory with and without projection.

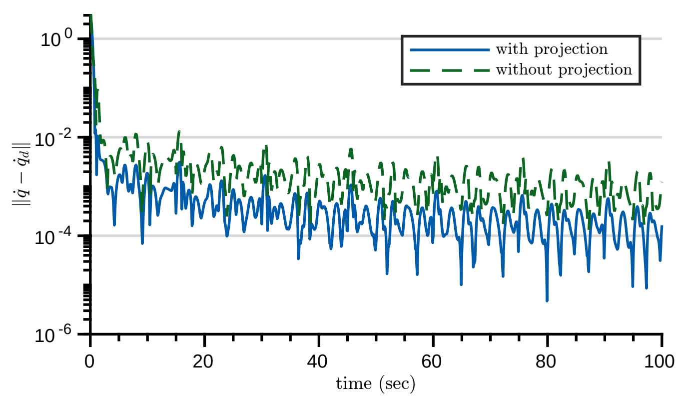

Figure 5.11: The comparison of the norm of the link velocity error between the robot manipulator angles and the reference trajectory with and without projection.

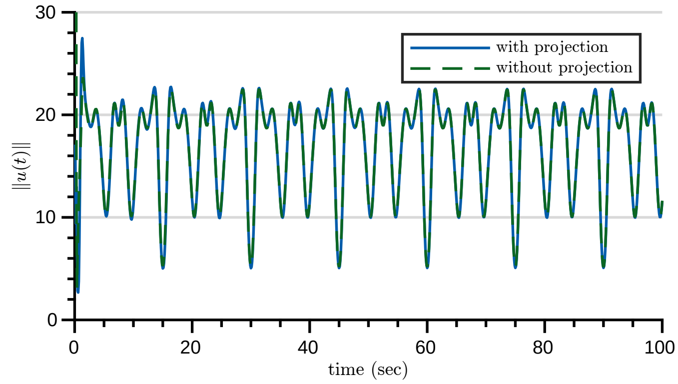

Figure 5.12: The comparison of the norm of the 2 joint torque input for the robot manipulator angles and the reference trajectory with and without projection.

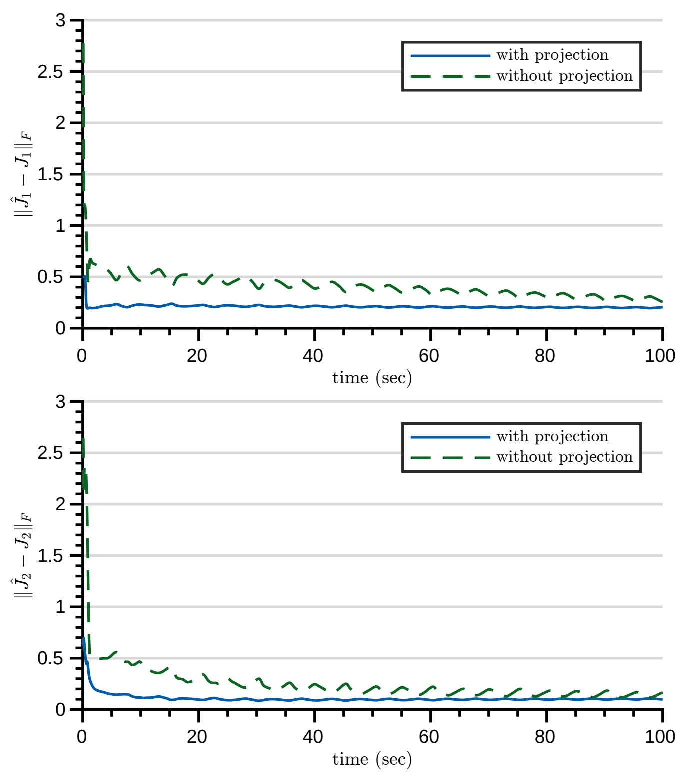

Figure 5.13: The Frobenius norm of pseudo-inertia matrix estimate error for the link 1 (top) and link 2 (bottom) link with (blue) and without (black) projection.

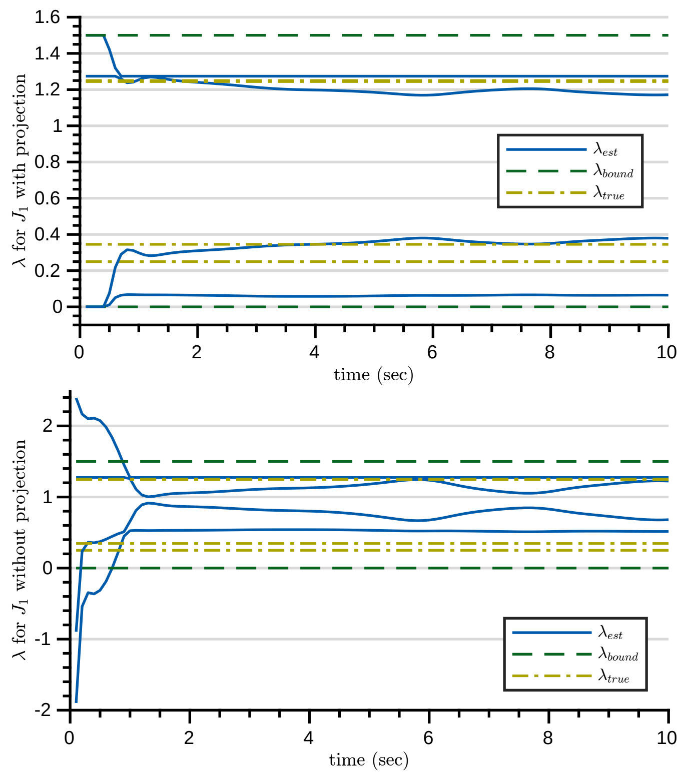

Figure 5.14: The eigenvalues of the link 1 inertia matrix estimate (blue), their true values (red) and the eigenvalue bounds (black) for adaptive attitude controller with (top) and without (bottom) projection.

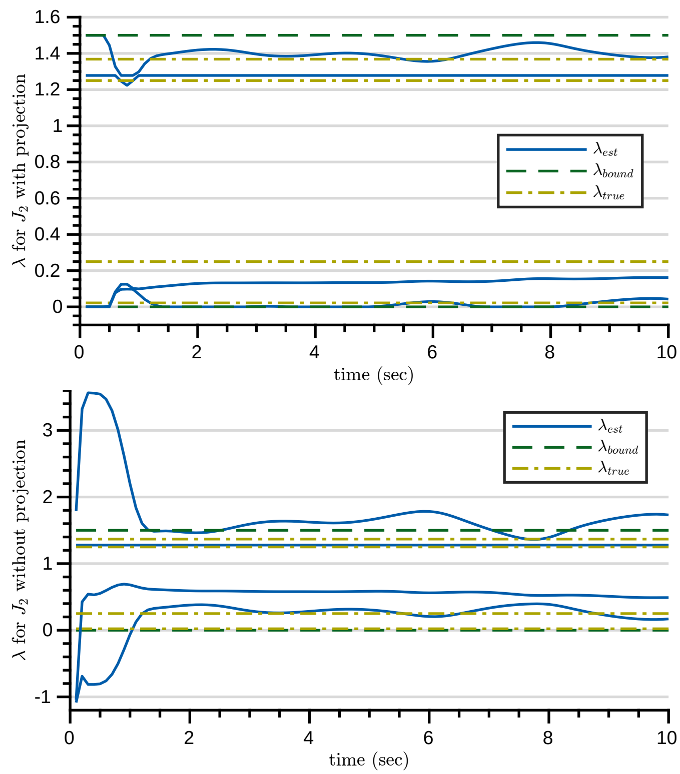

Figure 5.15: The eigenvalues of the link 2 inertia matrix estimate (blue), their true values (red) and the eigenvalue bounds (black) for adaptive attitude controller with (top) and without (bottom) projection.

5.6 Conclusion

This chapter addresses the problem of adaptive control in presence of unknown SPD matrix valued parameters. Such a problem is common in control applications such as attitude tracking wherein the inertia matrix is SPD and robotic manipulators wherein the inertial properties of every link, that can be expressed as an SPD matrix, are unknown. Previous chapters presented a methodology wherein information to identify SPD matrices were obtained in a discrete time setting. In this chapter, the tracking error of the control systems, that provides the information needed to update SPD matrix is obtained in a continuous manner.

The approach presented here is designed to handle a wide variety of control systems and not just systems with an SPD matrix parameter. In general, the projection method developed in this chapter can handle symmetric matrix valued parameters with eigenvalue constraints. Historic adaptive control formulations have not used the eigenvalue bounds to obtain better control performance. In order to address this gap in the literature, this chapter presented an globally convergent adaptive control formulations as well as ensure that the symmetric matrix parameter conforms with the eigenvalue bounds.

A novel projection scheme was presented to estimate a symmetric matrix valued parameter with bounded eigenvalues in adaptive control applications. A Lyapunov-like function was introduced that can be used in conjunction with a wide variety of adaptive control problems with symmetric matrix valued parameters. Globally convergent adaptive control laws for angular velocity tracking, attitude tracking, and the trajectory tracking of a robotic manipulator were derived. A direct update scheme to update the eigenvalues and the eigenprojection of a matrix without explicitly performing the eigen decomposition was derived. The update scheme was found to be stable even in the case of eigenvalue crossings. Numerical studies on the three problems described above produced a lower tracking error and lower estimation error as compared to existing adaptive control laws where the projection was not used. The increase in control effort due to high controller gains was also found to be less when the projection was used.

In spite of a non smooth projection, the control systems were shown to exhibit global stability properties. This was possible because the set of number of points of non-differentiability had Lebesgue measure 0. Smooth extensions of the above projection method present an attractive avenue of future work. The projection method presented here operates on the eigenvalues and eigenprojections of the symmetric matrix. Although the problem of performing eigen decomposition was addressed in Section 5.4, adaptive update laws that do not require eigen decompositions and enforce eigenvalue constraints could be useful to the community and are a topic of future research.

Chapter 6

Conclusion

Symmetric and positive definite matrices are often used to model engineering systems. This dissertation presents various adaptive techniques for identification of symmetric and positive definite (SPD) matrices. First, the problem of adaptive covariance estimation was introduced. An adaptive technique to estimate the noise covariance matrices in a linear model was shown to guarantee convergence. The rate of convergence of the estimates to the true value was shown to be proportional to $1/\sqrt{k}$ where k is the time variable. Although many adaptive Kalman filters have been developed in the literature, the algorithm presented here has two salient features. The algorithm presented here only assumes detectability of the system. The convergence of the state error covariance was also established in this dissertation. The combination of these two features makes the adaptive filter unique and widely applicable.

The adaptive covariance algorithm presented in this dissertation makes no prior assumptions on the covariance matrix. The algorithm presented here is more general than the Bayesian techniques that assume a prior distribution (Wishart or Inverse-Wishart) on the covariance matrix. The covariance estimation part in the formulation here is decoupled from the state estimation. This is because a linear time series is formed by eliminating the measurements. One way to make this algorithm more applicable is to have a continuous-time formulation. However, this requires an understanding of the derivative of noise and is a topic of future research. In terms of identifiability of the covariance matrices Mehra [100] provided an upper bound on the number of elements that could be estimated. This dissertation poses the condition of identifiability as rank condition on a matrix similar to the observability conditions for the state vector. However, whether or not this condition is stronger than the one presented by Mehra is important and is a direction of future research.

Next, the problem of adaptive identification of inertia matrices in control problems was addressed. The identification of inertia matrices in angular velocity tracking, attitude control, and trajectory tracking in robotic manipulator control was addressed. A general methodology to impose eigenvalue constraints on symmetric matrices was developed. This methodology can broadly be applied to symmetric matrices and not just SPD matrices. The projection method introduced projects the matrix onto a cone formed by the eigenvalue constraints. Since such a projection is non smooth, this leads to non smooth Lyapunov functions in stability analysis. The non smooth extensions to the Lyapunov theorems are applied to establish the stability under the new projection law. Although the projection is non smooth, the Clarke generalized gradient has been defined to characterize the gradient of the projection at the points of non differentiability [97, 131].

A geometric optimization approach for the estimation of covariance matrices was formulated in this dissertation. Such a formulation ensures that the covariance estimates are SPD and hence exponential stability of the adaptive Kalman filter [8]. Optimization procedures have been used to solve the estimation problems in the past. Although such procedures perform well on the problems in focus here, iterative procedures do not enjoy the computational efficiency of the other techniques developed in this dissertation. However, optimization procedures have to be used largely because of the lack of elegant closed form solutions for the data manipulation on the SPD manifold. Apart from the work in this dissertation, it is worth noting here that the non-iterative solutions found in Refs. [150, 126, 40, 158, 87] provide an attractive avenue for future work.

Bibliography

[1] P.-A. Absil, Robert Mahony, and Rodolphe Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, Princeton, NJ, USA, 2009.

[2] Jasim Ahmed, Vincent T. Coppola, and Dennis S. Bernstein. Adaptive asymptotic tracking of spacecraft attitude motion with inertia matrix identification. *Journal of Guidance, Control, and Dynamics*, 21(5):684–691, 1998.

[3] Maruthi R. Akella and Kamesh Subbarao. A novel parameter projection mechanism for smooth and stable adaptive control. *Systems & Control Letters*, 54(1):43–51, 2005.

[4] Danial L. Alspach. Comments on "On the identification of variances and adaptive Kalman filtering". *IEEE Transactions on Automatic Control*, 17(6):843–845, 1972.

[5] Danial L. Alspach. A parallel filtering algorithm for linear systems with unknown time varying noise statistics. *IEEE Transactions on Automatic Control*, 19(5):552–556, 1974.

[6] Daniel L. Alspach and Harold W. Sorenson. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic*

Control, 17(4):439–448, 1972.

[7] B. D. O. Anderson and John B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, NJ, 1979.

[8] B. D. O. Anderson and John B. Moore. Detectability and stabilizability of time-varying discrete-time linear systems. *SIAM Journal on Control and Optimization*, 19(1):20–32, 1981.

[9] Anuradha M. Annaswamy, Fredrik P. Skantze, and Ai-Poh Loh. Adaptive control of continuous time systems with convex/concave parametrization. *Automatica*, 34(1):33–49, 1998.

[10] Peter L. Antonelli, Robert J. Elliott, and Robert M. Seymour. Nonlinear filtering and Riemannian scalar curvature, $\mathcal{R}$. *Advances in Applied Mathematics*, 8(3):237–253, 1987.

[11] Ienkaran Arasaratnam and Simon Haykin. Cubature kalman filters. *IEEE Transactions on Automatic Control*, 54(6):1254–1269, 2009.

[12] Ienkaran Arasaratnam, Simon Haykin, and Robert J. Elliott. Discrete-time nonlinear filtering algorithms using gauss–hermite quadrature. *Proceedings of the IEEE*, 95(5):953–977, 2007.

[13] S. P. Arjun Ram and Maruthi R. Akella. Uniform exponential stability result for the rigid-body attitude tracking control problem. *Journal of Guidance, Control, and Dynamics*, 43(1):39–45, 2020.

[14] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Log-euclidean metrics for fast and simple calculus on diffusion tensors. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 56(2):411–421, 2006.

[15] Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, 2007.

[16] Michael Athans, David Castanon, K-P. Dunn, C. Greene, Wing Lee, N. Sandell, and A. Willsky. The stochastic control of the f-8c aircraft using a multiple model adaptive control (mmac) method–part i: Equilibrium flight. *IEEE Transactions on Automatic Control*, 22(5):768–780, 1977.

[17] Michael Athans and Chaw-Bing Chang. Adaptive estimation and parameter identification using multiple model estimation algorithm. Technical Report 28, Massachusetts Institute of Technology Lexington Lincoln Lab, 1976.

[18] Augusto Aubry, Antonio De Maio, Luca Pallotta, and Alfonso Farina. Maximum likelihood estimation of a structured covariance matrix with a condition number constraint. *IEEE Transactions on Signal Processing*, 60(6):3004–3021, 2012.

[19] R. Bakker and Anuradha M. Annaswamy. Stability and robustness proper-

ties of a simple adaptive controller. *IEEE Transactions on Automatic Control*, 41(9):1352–1358, 1996.

[20] Yaakov Bar-Shalom, X. Rong Li, and Thiagalingam Kirubarajan. *Estimation with applications to tracking and navigation: theory algorithms and software*. Wiley, New York, NY, USA, 2004.

[21] Axel Barrau and Silvere Bonnabel. Invariant kalman filtering. *Annual Review of Control, Robotics, and Autonomous Systems*, 1:237–257, 2018.

[22] Richard H. Battin. *Astronautical guidance*. McGraw-Hill, New York, NY, USA, 1964.

[23] Pierre R. Belanger. Estimation of noise covariance matrices for a linear time-varying stochastic process. *Automatica*, 10(3):267–275, 1974.

[24] Abdelaziz Benallegue, Yacine Chitour, and Abdelhamid Tayebi. Adaptive attitude tracking control of rigid body systems with unknown inertia and gyro-bias. *IEEE Transactions on Automatic Control*, 63(11):3986–3993, 2018.

[25] Rajendra Bhatia. *Positive Definite Matrices*, volume 24. Princeton University Press, Princeton, NJ, USA, 2009.

[26] Rajendra Bhatia, Tanvi Jain, and Yongdo Lim. On the Bures–Wasserstein distance between positive definite matrices. *Expositiones Mathematicae*, 37(2):165–191, 2018.

[27] Gerald J. Bierman. *Factorization Methods for Discrete Sequential Estimation*. Academic Press, New York, NY, USA, 1977.

[28] Nicholas M. Boffi and Jean-Jacques E. Slotine. Implicit regularization and momentum algorithms in nonlinearly parameterized adaptive control and prediction. *Neural Computation*, 33(3):590–673, 2021.

[29] William M. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. Academic Press, New York, NY, USA, 2 edition, 1986.

[30] Nicolas Boumal, Bamdev Mishra, P.-A. Absil, and Rodolphe Sepulchre. Manopt, a Matlab toolbox for optimization on manifolds. *Journal of Machine Learning Research*, 15:1455–1459, 2014.

[31] Guillaume Bourmaud, Rémi Mégret, Marc Arnaudon, and Audrey Giremus. Continuous-discrete extended Kalman filter on matrix Lie groups using concentrated Gaussian distributions. *Journal of Mathematical Imaging and Vision*, 51(1):209–228, 2015.

[32] Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, London, UK, 2004.

[33] H. Whitin Brewer. Identification of the noise characteristics in a Kalman filter. In *Control and dynamic systems*, volume 12, pages 491–581. Elsevier, 1976.

[34] Martin Brossard, Silvere Bonnabel, and Jean-Philippe Condomines. Unscented Kalman filtering on Lie groups. In *2017 IEEE/RSJ International*

Conference on Intelligent Robots and Systems (IROS), pages 2485–2491. IEEE, 2017.

[35] Zhijun Cai, Marcio S. de Queiroz, and Darren M. Dawson. A sufficiently smooth projection operator. *IEEE Transactions on Automatic Control*, 51(1):135–139, 2006.

[36] Siew Chan, Graham C. Goodwin, and Kwai Sin. Convergence properties of the riccati difference equation in optimal filtering of nonstabilizable systems. *IEEE Transactions on Automatic Control*, 29(2):110–118, 1984.

[37] Nalin A. Chaturvedi, Dennis S. Bernstein, Jasim Ahmed, Fabio Bacconi, and N. Harris McClamroch. Globally convergent adaptive tracking of angular velocity and inertia identification for a 3-dof rigid body. *IEEE Transactions on Control Systems Technology*, 14(5):841–853, 2006.

[38] Zeineb Chebbi and Maher Moakher. Means of Hermitian positive-definite matrices based on the log-determinant α-divergence function. *Linear Algebra and its Applications*, 436(7):1872–1889, 2012.

[39] Yilun Chen, Ami Wiesel, Yonina C. Eldar, and Alfred O. Hero. Shrinkage algorithms for mmse covariance estimation. *IEEE Transactions on Signal Processing*, 58(10):5016–5029, 2010.

[40] Yixin Chen and John E. McInroy. Estimation of symmetric positive-definite matrices from imperfect measurements. volume 47, pages 1721–1725. IEEE, 2002.

[41] Guang Cheng and Baba C. Vemuri. A novel dynamic system in the space of SPD matrices with applications to appearance tracking. *SIAM Journal on Imaging Sciences*, 6(1):592–615, 2013.

[42] Anoop Cherian, Suvrit Sra, Arindam Banerjee, and Nikolaos Papanikolopoulos. Jensen-bregman logdet divergence with application to efficient similarity search for cvariance mtrices. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(9):2161–2174, 2012.

[43] Eduardo F Costa and Alessandro Astolfi. A necessary and sufficient condition for semi-stability of the recursive Kalman filter. In *American Control Conference, 2008*, pages 1280–1285. IEEE, 2008.

[44] Eduardo F Costa and Alessandro Astolfi. On the stability of the recursive Kalman filter for linear time-invariant systems. In *American Control Conference, 2008*, pages 1286–1291. IEEE, 2008.

[45] Eduardo F. Costa and Alessandro Astolfi. Bounds related to the riccati difference equation for linear time varying systems. In *Decision and Control, 2009 held jointly with the 2009 28th Chinese Control Conference. CDC/CCC 2009. Proceedings of the 48th IEEE Conference on*, pages 2321–2326. IEEE, 2009.

[46] John J. Craig, Ping Hsu, and S. Shankar Sastry. Adaptive control of mechanical manipulators. *The International Journal of Robotics Research*, 6(2):16–28, 1987.

[47] Anton H. J. de Ruiter. Observer-based adaptive spacecraft attitude control with guaranteed performance bounds. *IEEE Transactions on Automatic Control*, 61(10):3146–3151, 2015.

[48] Arnaud Doucet, Simon Godsill, and Christophe Andrieu. On sequential monte carlo sampling methods for bayesian filtering. *Statistics and computing*, 10(3):197–208, 2000.

[49] Haibo Du, Shihua Li, and Chunjiang Qian. Finite-time attitude tracking control of spacecraft with application to attitude synchronization. *IEEE Transactions on Automatic Control*, 56(11):2711–2717, 2011.

[50] TE Duncan. Estimation for jump processes in the tangent bundle of a Riemann manifold. *Applied Mathematics and Optimization*, 4(1):265–274, 1977.

[51] TE Duncan. Stochastic filtering in manifolds. *IFAC Proceedings Volumes*, 14(2):553–556, 1981.

[52] Tyrone E Duncan. Some filtering results in Riemann manifolds. *Information and Control*, 35(3):182–195, 1977.

[53] Jindřich Duník, Oliver Kost, and Ondřej Straka. Design of measurement difference autocovariance method for estimation of process and measurement noise covariances. *Automatica*, 90:16–24, 2018.

[54] Jindřich Duník, Ondřej Straka, and Oliver Kost. Measurement difference autocovariance method for noise covariance matrices estimation. In *Deci-*

sion and Control (CDC), 2016 IEEE 55th Conference on, pages 365–370. IEEE, 2016.

[55] Jindřich Duník, Ondřej Straka, Oliver Kost, and Jindřich Havlík. Noise covariance matrices in state-space models: A survey and comparison of estimation methods-part i. *International Journal of Adaptive Control and Signal Processing*, 31(11):1505–1543, 2017.

[56] Z. M. Durovic and Branko D. Kovacevic. Robust estimation with unknown noise statistics. *IEEE Transactions on Automatic Control*, 44(6):1292–1296, 1999.

[57] Jianqing Fan, Yuan Liao, and Han Liu. An overview of the estimation of large covariance and precision matrices. *The Econometrics Journal*, 19(1):C1–C32, 2016.

[58] Bo Feng, Mengyin Fu, Hongbin Ma, Yuanqing Xia, and Bo Wang. Kalman filter with recursive covariance estimation-sequentially estimating process noise covariance. *IEEE Transactions on Industrial Electronics*, 61(11):6253–6263, 2014.

[59] Nicholas Fischer, Rushikesh Kamalapurkar, and Warren E. Dixon. Lasalle-yoshizawa corollaries for nonsmooth systems. *IEEE Transactions on Automatic Control*, 58(9):2333–2338, 2013.

[60] O. -E. Fjellstad and Thor I. Fossen. Position and attitude tracking of AUV's: a quaternion feedback approach. *IEEE Journal of Oceanic Engineering*,

19(4):512–518, 1994.

[61] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. Sparse inverse covariance estimation with the graphical lasso. *Biostatistics*, 9(3):432–441, 2008.

[62] Miroslaw Galicki. An adaptive regulator of robotic manipulators in the task space. *IEEE Transactions on Automatic Control*, 53(4):1058–1061, 2008.

[63] Andrew Gelman, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. *Bayesian Data Analysis*. Chapman & Hall, Boca Raton, Fl, USA, 2013.

[64] Gene H. Golub and Charles F. Van Loan. *Matrix Computations*, volume 3. Johns Hopkins University Press, Baltimore, MD, USA, 2012.

[65] Neil J. Gordon, David J. Salmond, and Adrian FM Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings F (Radar and Signal Processing)*, 140:107–113(6), April 1993.

[66] Clement Gosselin. Stiffness mapping for parallel manipulators. *IEEE Transactions on Robotics and Automation*, 6(3):377–382, 1990.

[67] Søren Hauberg, François Lauze, and Kim Steenstrup Pedersen. Unscented Kalman filtering on Riemannian manifolds. *Journal of Mathematical Imaging and Vision*, 46(1):103–120, 2013.

[68] H. Heffes. The effect of erroneous models on the Kalman filter response. *IEEE Transactions on Automatic Control*, 11(3):541–543, 1966.

[69] Sigurdur Helgason. *Differential Geometry, Lie Groups, and Symmetric Spaces*. Academic Press, Orlando, FL, USA, 1979.

[70] Christopher Hide, Terry Moore, and Martin Smith. Adaptive Kalman filtering for low-cost ins/gps. *The Journal of Navigation*, 56(1):143–152, 2003.

[71] Yuzo Hosoya and Masanobu Taniguchi. A central limit theorem for stationary processes and the parameter estimation of linear processes. *The Annals of Statistics*, pages 132–153, 1982.

[72] Peter J. Huber. *Robust Statistics*, volume 523. Wiley, New York, NY, USA, 2004.

[73] Andrew H. Jazwinski. *Stochastic processes and filtering theory*. Courier Corporation, 2007.

[74] Ben Jeuris, Raf Vandebril, and Bart Vandereycken. A survey and comparison of contemporary algorithms for computing the matrix geometric mean. *Electronic Transactions on Numerical Analysis*, 39(1):379–402, 2012.

[75] Simon Julier, Jeffrey Uhlmann, and Hugh F. Durrant-Whyte. A new method for the nonlinear transformation of means and covariances in filters and estimators. *IEEE Transactions on Automatic Control*, 45(3):477–482, 2000.

[76] John L. Junkins, Maruthi R. Akella, and Rush D. Robinett. Nonlinear adaptive control of spacecraft maneuvers. *Journal of Guidance, Control, and Dynamics*, 20(6):1104–1110, 1997.

[77] John L. Junkins and Hanspeter Schaub. *Analytical Mechanics of Space Systems*. American Institute of Aeronautics and Astronautics, Reston, VA, USA, 2009.

[78] Rudolph E. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME, Journal of Basic Engineering*, 82(1):35–45, 1960.

[79] Rudolph E. Kalman and Richard S. Bucy. New results in linear filtering and prediction theory. *Transactions of the ASME, Journal of Basic Engineering*, 83(1):95–108, 1961.

[80] Hermann Karcher. Riemannian center of mass and mollifier smoothing. *Communications on pure and applied mathematics*, 30(5):509–541, 1977.

[81] Rangasami L. Kashyap. Maximum likelihood identification of stochastic linear systems. *IEEE Transactions on Automatic Control*, 15(1):25–34, 1970.

[82] Tosio Kato. *Perturbation Theory for Linear Operators*, volume 132. Springer Science & Business Media, Berlin, Germany, 2013.

[83] Johann Wolfgang Koch. Bayesian approach to extended object and cluster tracking using random matrices. *IEEE Transactions on Aerospace and Electronic Systems*, 44(3):1042–1059, 2008.

[84] Junghyun Kwon, Hee Seok Lee, Frank C Park, and Kyoung Mu Lee. A geometric particle filter for template-based visual tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(4):625–643, 2013.

[85] Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

[86] Olivier Ledoit and Michael Wolf. The power of (non-) linear shrinking: a review and guide to covariance matrix estimation. *Journal of Financial Econometrics*, 2019.

[87] Taeyoon Lee, Jaewoon Kwon, and Frank C. Park. A natural adaptive control law for robot manipulators. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1–9. IEEE, 2018.

[88] Taeyoon Lee and Frank C Park. A geometric algorithm for robust multibody inertial parameter identification. *IEEE Robotics and Automation Letters*, 3(3):2455–2462, 2018.

[89] Taeyoon Lee, Patrick M Wensing, and Frank C. Park. Geometric robot dynamic identification: A convex programming approach. *IEEE Transactions on Robotics*, 36(2):348–365, 2019.

[90] Tony T. Lee. A direct approach to identify the noise covariances of Kalman filtering. *IEEE Transactions on Automatic Control*, 25(4):841–842, 1980.

[91] Joshua G. Leonard, Naomi E. Graver. Model-based feedback control of autonomous underwater gliders. *IEEE Journal of Oceanic Engineering*, 26(4):633–645, 2001.

[92] Zhenhua Lin. Riemannian geometry of symmetric positive definite matrices via Cholesky decomposition. *SIAM Journal on Matrix Analysis and Applications*, 40(4):1353–1370, 2019.

[93] Jacob Lubliner. *Plasticity theory*. Dover Publications, New York, NY, USA, 2008.

[94] Nikola Madjarov and Ludmila Mihaylova. Kalman filter sensitivity with respect to parametric noises uncertainty. *Kybernetika*, 32(3):307–322, 1996.

[95] Robert Mahony, Tarek Hamel, and Jean-Michel Pflimlin. Nonlinear complementary filters on the special orthogonal group. *IEEE Transactions on Automatic Control*, 53(5):1203–1218, 2008.

[96] Luigi Malagò, Luigi Montrucchio, and Giovanni Pistone. Wasserstein Riemannian geometry of Gaussian densities. *Information Geometry*, 1(2):137–179, 2018.

[97] Jérôme Malick and Hristo S. Sendov. Clarke generalized jacobian of the projection onto the cone of positive semidefinite matrices. *Set-Valued Analysis*, 14(3):273–293, 2006.

[98] Peter S. Maybeck. *Combined Estimation of States and Parameters for On-Line Applications*. PhD thesis, Massachusetts Institute of Technology, 1972.

[99] Peter S. Maybeck. *Stochastic models, estimation, and control*, volume 141 of *Mathematics in Science and Engineering*. 1979.

[100] Raman Mehra. On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, 15(2):175–184, 1970.

[101] Raman Mehra. Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, 17(5):693–698, 1972.

[102] Henrique M. T. Menegaz, João Y. Ishihara, and Hugo T. M. Kussaba. Unscented Kalman filters for Riemannian state-space systems. *IEEE Transactions on Automatic Control*, 64(4):1487–1502, 2018.

[103] Travis H. Mercker and Maruthi R. Akella. Rigid-body attitude tracking with vector measurements and unknown gyro bias. *Journal of Guidance, Control, and Dynamics*, 34(5):1474–1484, 2011.

[104] Richard H. Middleton and Graham C. Goodwin. Adaptive computed torque control for rigid link manipulations. *Systems & Control Letters*, 10(1):9–16, 1988.

[105] Maher Moakher. A differential geometric approach to the geometric mean of symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 26(3):735–747, 2005.

[106] Maher Moakher. On the averaging of symmetric positive-definite tensors. *Journal of Elasticity*, 82(3):273–296, 2006.

[107] Maher Moakher and Andrew N Norris. The closest elastic tensor of arbitrary symmetry to an elasticity tensor of lower symmetry. *Journal of Elasticity*, 85(3):215–263, 2006.

[108] Rahul Moghe, Renato Zanetti, and Maruthi R. Akella. Covariance matching Kalman filter for observable LTI systems. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 6372–6377. IEEE, 2018.

[109] Rahul Moghe, Renato Zanetti, and Maruthi R. Akella. Adaptive Kalman filter for detectable linear time-invariant systems. *Journal of Guidance, Control, and Dynamics*, 42(10):2197–2205, 2019.

[110] Kenneth Myers and Byron D. Tapley. Adaptive sequential estimation with unknown noise statistics. *IEEE Transactions on Automatic Control*, 21(4):520–523, 1976.

[111] Kumpati S. Narendra and Anuradha M. Annaswamy. *Stable Adaptive Systems*. Prentice-Hall, Englewood Cliffs, NJ, USA, 1989.

[112] S. K. Ng and Peter E. Caines. Nonlinear filtering in Riemannian manifolds. *IFAC Proceedings Volumes*, 17(2):817–821, 1984. 9th IFAC World Congress: A Bridge Between Control Science and Technology, Budapest, Hungary, 2-6 July 1984.

[113] Romeo Ortega and Mark W. Spong. Adaptive motion control of rigid robots: A tutorial. *Automatica*, 25(6):877–888, 1989.

[114] Daniel Osborne, Vic Patrangenaru, Leif Ellingson, David Groisser, and Armin Schwartzman. Nonparametric two-sample tests on homogeneous Riemannian manifolds, Cholesky decompositions and diffusion tensor image analysis. *Journal of Multivariate Analysis*, 119:163–175, 2013.

[115] Yaakov Oshman and Itzhack Y. Bar-Itzhack. Square root filtering via covariance and information eigenfactors. *Automatica*, 22(5):599–604, 1986.

[116] Frédéric Pascal, Lionel Bombrun, Jean-Yves Tourneret, and Yannick Berthoumieu. Parameter estimation for multivariate generalized Gaussian distributions. *IEEE Transactions on Signal Processing*, 61(23):5960–5971, 2013.

[117] Xavier Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *Journal of Mathematical Imaging and Vision*, 25(1):127–154, 2006.

[118] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. A Riemannian framework for tensor computing. *International Journal of computer vision*, 66(1):41–66, 2006.

[119] T Rapcsák. Geodesic convexity in nonlinear optimization. *Journal of Optimization Theory and Applications*, 69(1):169–183, 1991.

[120] Sidney I. Resnick. chapter 6, 8, pages 167–201. Birkhäuser Boston, Boston, MA, 2014.

[121] Samer S. Saab. Discrete-time Kalman filter under incorrect noise covariances. In *American Control Conference, Proceedings of the 1995*, volume 2, pages 1152–1156. IEEE, 1995.

[122] Andrew P. Sage and Gary W. Husa. Algorithms for sequential adaptive estimation of prior statistics. In *1969 IEEE Symposium on Adaptive Processes (8th) Decision and Control*, pages 61–61. IEEE, 1969.

[123] Salem Said, Lionel Bombrun, Yannick Berthoumieu, and Jonathan H Manton. Riemannian Gaussian distributions on the space of symmetric positive definite matrices. *IEEE Transactions on Information Theory*, 63(4):2153–2170, 2017.

[124] Salem Said, Hatem Hajri, Lionel Bombrun, and Baba C. Vemuri. Gaussian distributions on Riemannian symmetric spaces: statistical learning with structured covariance matrices. *IEEE Transactions on Information Theory*, 64(2):752–772, 2017.

[125] Salem Said and Jonathan H Manton. On filtering with observation in a manifold: Reduction to a classical filtering problem. *SIAM Journal on Control and Optimization*, 51(1):767–783, 2013.

[126] Hesamoddin Salehian, Guang Cheng, Baba C. Vemuri, and Jeffrey Ho. Recursive estimation of the stein center of SPD matrices and its applications. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1793–1800, 2013.

[127] Suwanchai Sangsuk-Iam and Thomas E. Bullock. Analysis of continuous-time Kalman filtering under incorrect noise covariances. *Automatica*, 24(5):659–669, 1988.

[128] Suwanchai Sangsuk-Iam and Thomas E. Bullock. Analysis of discrete-time Kalman filtering under incorrect noise covariances. *IEEE Transactions on Automatic Control*, 35(12):1304–1309, 1990.

[129] Amit Sanyal, Adam Fosbury, Nalin Chaturvedi, and Dennis S. Bernstein. Inertia-free spacecraft attitude tracking with disturbance rejection and almost global stabilization. *Journal of Guidance, Control, and Dynamics*, 32(4):1167–1178, 2009.

[130] Kazuhiro Sato, Hiroyuki Sato, and Tobias Damm. Riemannian optimal identification method for linear systems with symmetric positive-definite matrix. *IEEE Transactions on Automatic Control*, 65(11):4493–4508, 2019.

[131] Hristo S. Sendov. Generalized hadamard product and the derivatives of spectral functions. *SIAM Journal on Matrix Analysis and Applications*, 28(3):667–681, 2006.

[132] Daniel Shevitz and Brad Paden. Lyapunov stability theory of nonsmooth systems. *IEEE Transactions on Automatic control*, 39(9):1910–1914, 1994.

[133] Jean-Jacques E. Slotine and Weiping Li. On the adaptive control of robot manipulators. *The International Journal of Robotics Research*, 6(3):49–59, 1987.

[134] Steven T. Smith. *Geometric Optimization Methods for Adaptive Filtering*. PhD thesis, Harvard University, 1993.

[135] Hichem Snoussi. Particle filtering on Riemannian manifolds. Application to covariance matrices tracking. In *Matrix Information Geometry*, pages 427–449. Springer, 2013.

[136] Hichem Snoussi and Ali Mohammad-Djafari. Particle filtering on Riemannian manifolds. In *AIP Conference Proceedings*, volume 872, pages 219–226. American Institute of Physics, 2006.

[137] Hichem Snoussi and Cédric Richard. Monte carlo tracking on the Riemannian manifold of multivariate normal distributions. In *2009 IEEE 13th Digital Signal Processing Workshop and 5th IEEE Signal Processing Education Workshop*, pages 280–285. IEEE, 2009.

[138] Cristóvão D Sousa and Rui Cortesao. Physical feasibility of robot base inertial parameter identification: A linear matrix inequality approach. *The International Journal of Robotics Research*, 33(6):931–944, 2014.

[139] Mark W. Spong, Seth Hutchinson, and Mathukumalli Vidyasagar. *Robot Modeling and Control*, volume 3. Wiley, New York, NY, USA, 2006.

[140] Suvrit Sra. A new metric on the manifold of kernel matrices with application to matrix geometric means. *Advances in Neural Information Processing Systems*, 25:144–152, 2012.

[141] Suvrit Sra. Positive definite matrices and the S-divergence. *Proceedings of the American Mathematical Society*, 144(7):2787–2797, 2016.

[142] Suvrit Sra and Reshad Hosseini. Geometric optimisation on positive definite matrices for elliptically contoured distributions. *Advances in Neural Information Processing Systems*, 26:2562–2570, 2013.

[143] Suvrit Sra and Reshad Hosseini. Conic geometric optimization on the manifold of positive definite matrices. *SIAM Journal on Optimization*, 25(1):713–739, 2015.

[144] Weehong Tan and Andrew Packard. Stability region analysis using polynomial and composite polynomial lyapunov functions and sum-of-squares programming. *IEEE Transactions on Automatic Control*, 53(2):565–571, 2008.

[145] Divya Thakur, Sukumar Srikant, and Maruthi R. Akella. Adaptive attitude-tracking control of spacecraft with uncertain time-varying inertia parameters. *Journal of Guidance, Control, and Dynamics*, 38(1):41–52, 2015.

[146] Patrizio Tomei. Robust adaptive friction compensation for tracking control of robot manipulators. *IEEE Transactions on Automatic Control*, 45(11):2164–2169, 2000.

[147] S. Traversaro, S. Brossette, A. Escande, and F. Nori. Identification of fully physical consistent inertial parameters using optimization on manifolds. In

2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pages 5446–5451, 2016.

[148] Oncel Tuzel, Fatih Porikli, and Peter Meer. Region covariance: A fast descriptor for detection and classification. In *European conference on computer vision*, pages 589–600. Springer, 2006.

[149] Oncel Tuzel, Fatih Porikli, and Peter Meer. Pedestrian detection via classification on Riemannian manifolds. *IEEE transactions on pattern analysis and machine intelligence*, 30(10):1713–1727, 2008.

[150] Ambrish Tyagi and James W. Davis. A recursive filter for linear systems on Riemannian manifolds. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, 2008.

[151] Michel Verhaegen and Paul Van Dooren. Numerical aspects of different Kalman filter implementations. *IEEE Transactions on Automatic Control*, 31(10):907–917, 1986.

[152] Hairong Wang, Zhihong Deng, Bo Feng, Hongbin Ma, and Yuanqing Xia. An adaptive Kalman filter estimating process noise covariance. *Neurocomputing*, 223:12–17, 2017.

[153] Hanlei Wang. Adaptive control of robot manipulators with uncertain kinematics and dynamics. *IEEE Transactions on Automatic Control*, 62(2):948–954, 2016.

[154] Hanlei Wang and Yongchun Xie. On the uniform positive definiteness of the estimated inertia for robot manipulators. *IFAC Proceedings Volumes*, 44(1):4089–4094, 2011.

[155] Xiangke Wang, Changbin Yu, and Zhiyun Lin. A dual quaternion solution to attitude and position control for rigid-body coordination. *IEEE Transactions on Robotics*, 28(5):1162–1170, 2012.

[156] Xie Wang, Mei-qin Liu, Zhen Fan, and Sen-lin Zhang. A novel approach of noise statistics estimate using h∞ filter in target tracking. *Frontiers of Information Technology & Electronic Engineering*, 17(5):449–457, 2016.

[157] Yin Wang, Octavia Camps, Mario Sznaier, and Biel Roig Solvas. Jensen bregman logdet divergence optimal filtering in the manifold of positive definite matrices. In *European Conference on Computer Vision*, pages 221–235. Springer, 2016.

[158] Zhizhou Wang and Baba C. Vemuri. DTI segmentation using an information theoretic tensor dissimilarity measure. *IEEE Transactions on Medical Imaging*, 24(10):1267–1277, 2005.

[159] Zhizhou Wang, Baba C. Vemuri, Yunmei Chen, and Thomas H. Mareci. A constrained variational principle for direct estimation and smoothing of the diffusion tensor field from complex dwi. *IEEE transactions on Medical Imaging*, 23(8):930–939, 2004.

[160] J. T. -Y. Wen and K. Kreutz-Delgado. The attitude control problem. *IEEE Transactions on Automatic Control*, 36(10):1148–1162, 1991.

[161] Patrick M. Wensing, Sangbae Kim, and Jean-Jacques E. Slotine. Linear matrix inequalities for physically consistent inertial parameter identification: A statistical perspective on the mass distribution. *IEEE Robotics and Automation Letters*, 3(1):60–67, 2017.

[162] Ami Wiesel. Unified framework to regularized covariance estimation in scaled Gaussian models. *IEEE Transactions on Signal Processing*, 60(1):29–38, 2011.

[163] Ami Wiesel. Geodesic convexity and covariance estimation. *IEEE transactions on signal processing*, 60(12):6182–6189, 2012.

[164] Jacques L. Willems and Frank M. Callier. Divergence of the stationary Kalman filter for correct and for incorrect noise variances. *IMA Journal of Mathematical Control and Information*, 9(1):47–54, 1992.

[165] Jesse Windle and Carlos M. Carvalho. A tractable state-space model for symmetric positive-definite matrices. *Bayesian Analysis*, 9(4):759–792, 2014.

[166] Bin Xian, Marcio S. de Queiroz, Darren M. Dawson, and Ian D. Walker. Task-space tracking control of robot manipulators via quaternion feedback. *IEEE Transactions on Robotics and Automation*, 20(1):160–167, 2004.

[167] Yuanxi Yang and Weiguang Gao. An optimal adaptive Kalman filter. *Journal of Geodesy*, 80(4):177–183, 2006.

[168] Chi Zhang, Amirhossein Taghvaei, and Prashant G. Mehta. Feedback particle filter on Riemannian manifolds and matrix Lie groups. *IEEE Transactions on Automatic Control*, 63(8):2465–2480, 2018.

[169] J Zhou and RH Luecke. Estimation of the co variances of the process noise and measurement noise for a linear discrete dynamic system. *Computers & Chemical Engineering*, 19(2):187–195, 1995.

Vita

Rahul Moghe was born and raised in Mumbai, India. He graduated from the Indian Institute of Technology Bombay in 2015. He was then enrolled in a Master of Science degree at the University of Texas at Austin in the Mechanical Engineering Department at the ReNeu Robotics Lab. After completing his Masters in 2017, Rahul went on to pursue his doctoral degree at the University of Texas at Austin in the Aerospace Engineering Department under the guidance of Dr. Maruthi Akella and Dr. Renato Zanetti. His research interests include geometric methods in control and estimation theory and machine learning for space applications.

Permanent address: rahul.moghe@utexas.com

This dissertation was typeset with LaTeX[†] by the author.

[†]LaTeX is a document preparation system developed by Leslie Lamport as a special version of Donald Knuth's TeX Program.