(Preprint) AAS 21-353

COVARIANCE ESTIMATION USING GEOMETRIC OPTIMIZATION ON SYMMETRIC POSITIVE DEFINITE MANIFOLDS

Rahul Moghe[*], Maruthi R. Akella[†], and Renato Zanetti[‡]

This paper provides a significant extension to our previous work on adaptive covariance estimation that estimates the noise covariance matrices of a discrete linear system with simultaneous guarantees for convergence of the noise covariance estimates and the state estimates. The specific advance established in this work is that the estimates of the covariance are calculated using a differential geometric optimization framework that ensures that the covariance estimates lie restricted to the symmetric positive definite (SPD) manifold. This property is desirable since the covariance estimates used in calculating the state estimates have to be SPD at all times for them to be meaningful. A cost function that is geodesically convex on the SPD manifold is chosen to provide a globally optimal solution.

INTRODUCTION

The Kalman filtering problem for linear system is well studied in the literature.[1,2] One of the important assumptions that allows for a reliable estimation algorithm is that the noise covariance matrices of the noises entering the system are completely known. This assumption is rarely true in practice because the noises in the system and the measurements are not readily available. In practice, the noise covariance matrices are either predetermined through experimental setups or are artificially inflated to adopt a conservative strategy. The case when inaccurate noise covariances are used is known to cause filter divergence.[3–8] These challenges motivate adaptive algorithms to estimate the noise covariance matrices.

There have been many notable contributions in the covariance matching techniques for estimating noise covariances.[9–17] Our recent work on adaptive Kalman filtering presented a significant advance in this field wherein for the first time, we derive the estimates of the unknown elements of the process and measurement noise covariance matrices along with theoretical guarantees for their convergence.[18] In this setting, although the estimates of the noise covariance matrices can be maintained to be symmetric at all times, there were no further assurances on their positive definiteness. In order to circumvent this issue, Ref. [18] adopts a convenient remedy, that is to revert back the covariance estimate to the most recent (prior) value when it was symmetric and positive definite (SPD). The overall convergence result still holds with this heuristic in place as the estimates are

*PhD Candidate, Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, email: rahul.moghe@utexas.edu.
†Ashley H. Priddy Centennial Professor, Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, AAS Fellow. email: makella@mail.utexas.edu.
‡Assistant Professor, Department of Aerospace Engineering and Engineering Mechanics, The University of Texas at Austin, AAS Fellow. email: renato@utexas.edu.

proved to be SPD in the limit with probability one. However, it motivates the question as to whether or not the estimator can be modified to guarantee the SPD property for the estimates at all times, throughout the estimation process.

In this paper, we provide a positive answer to the foregoing question by adopting a differential geometric approach that ensures SPD covariance estimates. 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.[19] This structure enables Geometric Optimization (GO) that enables us to solve optimization problems that are non-convex in the Euclidean sense, but are convex in the SPD manifold sense.[20] Instead of arbitrarily reverting to the most recent (past) estimate of the covariance when it was SPD, we estimate the covariance matrices using a differential GO technique that ensures SPD estimates at all times. Algorithms involving GO 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.[21,22] These methods capitalize on the concepts of geodesics and tangent spaces on the manifolds to ensure that the estimates through the iterations lie on the manifold. The advantages of GO techniques over traditional optimization methods with additional constraint requiring the estimates to lie on the manifold have been well documented in the literature.[21]

Geodesically convex (g-convex) cost functions were introduced to aid theoretical guarantees for optimality. Although g-convex functions are not convex in the traditional Euclidean sense, they provide guarantees of global optimality when viewed from the manifold perspective. The GO techniques have been used to estimate parameters of elliptically contoured distributions and covariance estimation.[23,24] In these studies, the covariance matrix of the distribution was estimated using samples drawn from that distribution. GO techniques have also been used subspace tracking adaptive signal processing.[25] In this paper, we use a particular g-convex cost function, namely, the S-divergence,[26] to provide SPD estimates of the noise covariance matrices of a detectable linear discrete time invariant system.

The paper is organized as follows. First, the adaptive Kalman filter to estimate the unknown elements of Q and R matrices is formulated. The expressions to adaptively estimate the noise covariance matrices are presented. This step closely mimics our original approach in Ref. [18]. A brief introduction to the geometry of the SPD manifold is presented. For brevity purposes, only relevant concepts required for the GO technique 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. g-convexity of the cost is then analyzed and conditions for g-convexity are derived. A few experimental results using existing GO tools to estimate the noise covariance matrices are then simulated.

ADAPTIVE FILTER FORMULATION

Problem Formulation

We first begin by presenting the system equations followed by the linear Kalman filter equations when the Q and the R matrices are known.

$$x_{k+1} = Fx_k + w_k \tag{1}$$

$$y_k = Hx_k + v_k \tag{2}$$

wherein $x_k \in \mathbb{R}^n$ denotes the state, $y_k \in \mathbb{R}^p$ denotes the measurements, $F \in \mathbb{R}^{n\times n}$ is the state transition matrix, $H \in \mathbb{R}^{p\times n}$ is the observation matrix, and $w_k \sim \mathcal{N}(\mathbf{0}_{n\times 1}, Q)$ and $v_k \sim \mathcal{N}(\mathbf{0}_{p\times 1}, R)$ are the white Gaussian process and measurement noise entering the system.

The nominal Kalman filtering equations are as given below:[27]

$$\left.\begin{aligned} \hat{x}_{k|k-1} &= F\hat{x}_{k-1|k-1} \\ \hat{x}_{k|k} &= \hat{x}_{k|k-1} + K_k(y_k - H\hat{x}_{k|k-1}) \\ P_{k|k-1} &= FP_{k-1|k-1}F^T + Q \\ K_k &= P_{k|k-1}H^T(HP_{k|k-1}H^T + R)^{-1} \\ P_{k|k} &= (I - K_kH_k)P_{k|k-1}(I - K_kH_k)^T + K_kRK_k^T \end{aligned}\right\} \tag{3}$$

wherein $\hat{x}_{k|k-1}, \hat{x}_{k|k}$ are the predicted and updated state estimates, 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 noise covariance matrices Q and R are constant for all times. The system is assumed to be fully observable. Note that even though our previous result[18] holds for detectable systems with a state transformation, the non observable part of the system was separated and only the observable portions of the state and the process noise covariance matrix respectively could be estimated. Hence, without loss of generality, the proceeding discussion can be extended to detectable systems without any shortcomings. The adaptive Kalman filter equations can be specified as follows.

$$\left.\begin{aligned} \hat{x}^A_{k|k-1} &= F\hat{x}^A_{k-1|k-1} \\ \hat{x}^A_{k|k} &= \hat{x}^A_{k|k-1} + \hat{K}_k(y_k - H\hat{x}^A_{k|k-1}) \\ P^A_{k|k-1} &= FP^A_{k-1|k-1}F^T + \hat{Q}_k \\ \hat{K}_k &= P_{k|k-1}H^T(HP^A_{k|k-1}H^T + \hat{R}_k)^{-1} \\ P^A_{k|k} &= (I - K^A_kH_k)P_{k|k-1}(I - K^A_kH_k)^T + K^A_k\hat{R}_kK^{A^T}_k \end{aligned}\right\}. \tag{4}$$

Here $\hat{Q}_k$ and $\hat{R}_k$ are the estimates of the process and noise covariance matrices respectively. The superscripts ^{A} denote the quantities are drawn from the adaptive Kalman filter. If the process and measurement noises are white Gaussian, the baseline Kalman filter in (3) is optimal in the mean squared error sense and the state error covariance $P_{k|k}$ converges to a steady-state value.[28] Here, we assume that some elements of the Q and R matrices are unknown and we derive an adaptive filtering technique to estimate them.

Linear Time Series Formulation

The formulation of the linear time series below follows from our previous work.[18] Stacking m measurements in time, we get a modified observation model. The number of measurements to be stacked is calculated from the following equation.

$$\text{Define } O_i = \begin{bmatrix} HF^{i-1} \\ HF^{i-2} \\ \vdots \\ H \end{bmatrix} \tag{5}$$
$$m = \underset{i}{\operatorname{argmin}}\{\operatorname{rank}(O_i) = n\}$$

wherein, n is the dimension of the state space. The stacked measurement equation is given below.

$$\underbrace{\begin{bmatrix} y_k \\ y_{k-1} \\ \vdots \\ y_{k-m+1} \end{bmatrix}}_{\triangleq \mathcal{Y}_k} = \underbrace{\begin{bmatrix} HF^{m-1} \\ HF^{m-2} \\ \vdots \\ H \end{bmatrix}}_{O_m} x_{k-m+1} + \underbrace{\begin{bmatrix} H & HF & HF^2 & \cdots & HF^{m-2} \\ \mathbf{0}_{p\times n} & H & HF & \cdots & HF^{m-3} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & H \\ \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} & \mathbf{0}_{p\times n} \end{bmatrix}}_{\triangleq M_w} \underbrace{\begin{bmatrix} w_{k-1} \\ w_{k-2} \\ \vdots \\ w_{k-m+1} \end{bmatrix}}_{\triangleq W_k} + \underbrace{\begin{bmatrix} v_k \\ v_{k-1} \\ \vdots \\ v_{k-m+1} \end{bmatrix}}_{\triangleq V_k} \tag{6}$$

Eliminating the state from the above equation, we get the following linear time series.

$$\underbrace{O_m^\dagger \mathcal{Y}_k - FO_m^\dagger \mathcal{Y}_{k-1}}_{\triangleq \mathcal{Z}_k} = \underbrace{w_{k-m} + O_m^\dagger M_w W_k - FO_m^\dagger M_w W_{k-1}}_{\triangleq \mathcal{W}_k} + \underbrace{O_m^\dagger V_k - FO_m^\dagger V_{k-1}}_{\triangleq \mathcal{V}_k} \tag{7}$$

Here, † denotes the Moore-Penrose inverse of a matrix. Calculating the covariance of the linear time series gives the following linear matrix equation.

$$Cov(\mathcal{Z}_k) = A_1QA_1^T + \cdots A_mQA_m^T + B_0RB_0^T + \cdots B_mRB_m^T \tag{8}$$

The covariance is estimated using a Maximum a Posteriori (MAP) estimate of an Normal-inverse-Wishart (NIW) distribution. The mean and covariance are assumed to belong to the four parameter NIW distribution with a probability density function given below.[29]

$$\mathbf{NIW}(\mu, \Sigma|\nu, \lambda, \mu, \Phi) = |\Sigma|^{-(\frac{1}{2}(\nu+d)+1)}\exp\Big(-\frac{1}{2}\mathrm{tr}\big(\Phi\Sigma^{-1}\big) - \lambda(\mu-\mu_0)^T\Sigma^{-1}(\mu-\mu_0)\Big) \tag{9}$$

The posterior NIW distribution after observing r samples $x_i \sim \mathcal{N}(\mu, \Sigma),\ i = 1, \cdots, r$ is characterized by the following updated parameters.

$$\begin{aligned} \mu_r &= \frac{\lambda_0\mu_0 + r\bar{x}}{\lambda_0 + r} \\ \lambda_r &= \lambda_0 + r \\ \nu_r &= \nu_r + r \\ \Phi_r &= \Phi_0 + S + \frac{\lambda_0\, r}{\lambda_0 + r}(\bar{x}-\mu_0)^T(\bar{x}-\mu_0) \end{aligned} \tag{10}$$

wherein, $S = \sum_{i=1}^{r}(x_i - \bar{x})(x_i - \bar{x})^T$ is the sample covariance. The posterior estimates of μ and Σ are expressed as follows.

$$\begin{aligned} \hat{\mu}_r &= \mu_r \\ \hat{\Sigma}_r &= \frac{\Phi_r}{\nu_r - l} \end{aligned} \tag{11}$$

For further discussion about NIW distributions and a detailed calculation of the posterior, we refer the reader to Ref. [29]. Previously, Eq. (8) was solved through vectorization and separating out the known and unknown elements of the noise covariance matrices[18] as shown below.

$$\hat{\theta}_k = S^\dagger(vec(Cov(\mathcal{Z}_k)) - \Theta_{known}) \tag{12}$$

wherein, $\hat{\theta}_k$ is the estimate of the unknown elements of the noise covariance matrices, S is their coefficient matrix, $vec(\cdot)$ is the vectorization operation, and Θ_{known} is formed using the known elements from Eq. (8) through vectorization.

GEOMETRY OF THE SYMMETRIC POSITIVE DEFINITE MANIFOLD

Optimization algorithms require the characterization of the Gradient and the Hessian of the cost function over the set of SPD matrices. The set of SPD matrices represent a constraint on the eigenvalues of the matrix which is a nonlinear constraint on the matrix elements. Hence, the linear update steps used in traditional optimization algorithms do not ensure positive definiteness of a matrix. This is because there is a possibility of the updated matrix becoming indefinite or negative definite. The set of SPD matrices form a convex cone whose interior (strictly positive definite) forms a differentiable Riemannian manifold with non positive curvature. This means that the neighborhood of every SPD matrix resembles a Euclidean space of symmetric matrices locally.

We restate relevant definitions and properties of the SPD Riemannian Manifold that will motivate the Riemannian versions of the Gradient and the Hessian from.[19,22] Let the set of SPD matrices be denoted by $\mathbb{P}_n$. The tangent space to $\mathbb{P}_n$ at any point $X \in \mathbb{P}_n$, denoted by $T_X\mathbb{P}_n$, is identified with $\mathbb{S}_n$, the space of symmetric matrices.

Definition 1. *The Euclidean inner product on the $\mathbb{S}_n$ space is chosen to be the Frobenius inner product defined below.*

$$g_E(A, B) = \langle A, B\rangle_F = \mathrm{tr}(AB) \quad \forall\, A, B \in \mathbb{S}_n \tag{13}$$

wherein, $\langle\cdot,\cdot\rangle_F$ is the Frobenius inner product.

Definition 2. *The Riemannian inner product at the point X for $U, V \in T_X\mathbb{P}_n$ is defined as follows.*

$$g_X(U, V) = g_E(X^{-\frac{1}{2}}UX^{-\frac{1}{2}}, X^{-\frac{1}{2}}VX^{-\frac{1}{2}}) = \mathrm{tr}(X^{-1}UX^{-1}V) \tag{14}$$

Definition 3. *The Riemannian metric induced by the inner product on $T_X\mathbb{P}_n$ at X is given in a differential form as follows.*

$$ds = g_X(dX, dX)^{\frac{1}{2}} = \|X^{-\frac{1}{2}}dXX^{-\frac{1}{2}}\|_2 = \left[\mathrm{tr}(X^{-1}dX)^2\right]^{\frac{1}{2}} \tag{15}$$

wherein, $dX \in T_X\mathbb{P}_n$ denotes the differential tangent element.

Definition 4. *The unique geodesic (minimum length path) between two matrices $A, B \in \mathbb{P}_n$ is given by the following equation.*

$$\gamma(t) = A^{\frac{1}{2}}(A^{-\frac{1}{2}}BA^{-\frac{1}{2}})^t A^{\frac{1}{2}}, \quad t \in\ [0, 1] \tag{16}$$

Definition 5. *A retraction on $\mathbb{P}_n$ at X is defined as a smooth map $R_X : T_X\mathbb{P}_n \to \mathbb{P}_n$ given by the following expression.*

$$R_X(V) = X^{\frac{1}{2}}\mathrm{Exp}(X^{-\frac{1}{2}}VX^{-\frac{1}{2}})X^{\frac{1}{2}} \quad V \in T_X\mathbb{P}_n \tag{17}$$

wherein, Exp *denotes the matrix exponential operation. We use a shorthand notation, $R_X(V) = \mathrm{Exp}_X V$ in the sequel.*

The retraction at X in the direction V gives the $\mathbb{P}_n$ element obtained by moving a unit distance on the geodesic emanating from X in the V tangential direction.

Definition 6. *Consider a function $f : \mathbb{P}_n \to \mathbb{R}$. The directional derivative of f at $X \in \mathbb{P}_n$ in the direction $V \in T_X\mathbb{P}_n$ is defined as follows.*

$$Df(X)[V] = \frac{d}{dt} f(\gamma(t))\Big|_{t=0} \tag{18}$$

wherein, γ is a smooth curve on $\mathbb{P}_n$ passing through X at $t = 0$ with a $\gamma'(t) = V \in \mathbb{S}_n$ as its first derivative.

Note that $\mathbb{P}_n$ can be identified as an embedded submanifold of the Euclidean space $E = \mathbb{R}^{n\times n}$. Let $\bar{f} : E \to \mathbb{R}$ be the extension of f to E, such that $\bar{f}(X) = f(X)\ \forall\ X \in \mathbb{P}_n$. The directional derivative can be calculated by extending the domain of f to the Euclidean space wherein $\mathbb{P}_n$ is embedded.[22]

$$Df(X)[V] = D\bar{f}(X)[V] = \lim_{t\to 0} \frac{\bar{f}(X + tV) - \bar{f}(X)}{t} \tag{19}$$

Definition 7. *The Riemannian gradient of f, denoted by* grad f*, is defined through the directional derivative as follows.*

$$Df(X)[V] = g_X(\text{grad } f, V) \quad \forall\ V \in T_X\mathbb{P}_n \tag{20}$$

Hence, for the $\mathbb{P}_n$ manifold, rearranging the expression for the directional derivative yields the Riemannian gradient at X from the Euclidean gradient as follows.

$$\text{grad } f = \frac{1}{2} X(\text{grad } \bar{f} + \text{grad } \bar{f}^T)X \tag{21}$$

Such a calculation ensures that $\text{grad } f \in T_X\mathbb{P}_n$, and subsequently, a first-order algorithm such as the steepest descent, that utilizes the retraction map ensures SPD estimates.

$$X_{k+1} = R_{X_k}(-\eta \text{ grad } f) \tag{22}$$

wherein, $X_k, X_{k+1} \in \mathbb{P}_n$ are the prior and updated estimates, η is the learning rate and $f : \mathbb{P}_n \to \mathbb{R}$ is the cost function.

Second-order methods involving the Hessian map provide a better control over the cost function for GO applications. To this effect, the second-order terms in the Taylor series expansion of the cost function needs to be calculated. This requires the calculation of Hessian maps and the derivative of vector fields on manifolds, known as connections or covariant derivatives. Without going into too much detail, a few general results from Ref. [22] are restated within the context of $\mathbb{P}_n$, our SPD manifold.

Definition 8. *The Riemannian Hessian of a map $f : \mathbb{P}_n \to \mathbb{R}$ at X is a linear operator* Hess f : $T_X\mathbb{P}_n \to T_X\mathbb{P}_n$ *defined as follows.*

$$\text{Hess } f(X)[U] = \nabla_U \text{ grad } f \tag{23}$$

wherein, ∇_U is the covariant derivative.

The covariant derivative, in simple terms, characterizes the derivative of a vector field along another vector field. For the $\mathbb{P}_n$ manifold, the Euclidean gradient of f represents a vector field on $\mathbb{P}_n$. The Hessian of f for $U \in T_X\mathbb{P}_n$ is expressed as a map below.

$$\text{Hess } f(X)[U] = X \text{ Hess } \bar{f}\ X + \frac{1}{2}(X \text{ grad } \bar{f}\ U + U \text{ grad } \bar{f}\ X) \tag{24}$$

The gradient and Hessian of the cost function enables us to optimize the cost function over the SPD manifold.

GEOMETRIC OPTIMIZATION

Although numerous cost functions may be constructed to ensure that Eq. (8) holds, optimality properties need to be rigorously established. Previous studies have introduced the concept of geodesic convexity that ensures a globally optimal solution to the GO problem. The log-det α-divergence and its variants have been known to provide efficient alternatives to the cost functions involving matrix logarithm and exponential calculations.[26,30,31] The expression for the S-divergence, which is used in this paper, is stated below.

$$J_S(A,B) = \det\Big(\frac{A+B}{2}\Big) - \frac{1}{2}\det A - \frac{1}{2}\det B \tag{25}$$

The above function was shown to be g-convex whenever A and B are SPD matrices.[26] This property ensures that the local and global minimizers of the cost function to coincide. In this paper, the optimization variable is the block diagonal version of the noise covariance matrices to be estimated.

$$X = \begin{bmatrix} Q & \mathbf{0}_{n\times p} \\ \mathbf{0}_{p\times n} & R \end{bmatrix} \in \mathbb{P}_{n+p} \tag{26}$$

The cost function can be broken down into three parts.

$$J(X) = J_{cov}(X) + J_{corr}(X) + J_{const}(X) \tag{27}$$

The first part of the cost function, namely $J_{cov}(X)$ enforces Eq. (8) using the S-divergence stated in Eq. (25).

$$J_{cov}(X) = \alpha J_S\left(C_k, \sum_{i=0}^{i=m} E_i\, X\, E_i^T\right). \tag{28}$$

Here, α is a positive constant, and $E_i = \big[A_i, B_i\big]$ for all $i = 0, \ldots, m$, $A_0 = \mathbf{0}_{n\times n}$. The rate of convergence of C_k, was shown to be $1/\sqrt{k}$ in our previous work.[18]

The second part of the cost function minimizes the cross-correlation between the process and measurement noises.

$$\begin{aligned} J_{corr}(X) &= \beta \mathrm{tr}\big(Z_1 X Z_2^T Z_2 X Z_1^T\big) \\ Z_1 &= \big[\mathbf{0}_{p\times n} \quad I_p\big] \\ Z_2 &= \big[I_p \quad \mathbf{0}_{p\times n}\big]. \end{aligned} \tag{29}$$

Here, β is a positive constant and the matrices Z_1 and Z_2 select the cross-correlation matrix from X and minimizes it. Although the cross-correlation can be estimated, we assume that it is zero.

The final part of the cost function constraints certain elements of the Q and R matrices if they are known. Let the set of n_v known elements be denoted by the following set.

$$\mathcal{I}_{known} = \big\{k \,\big|\, 0 \le k \le n_v,\ X_{i_k, j_k} = \bar{x}_{i_k, j_k}\big\} \tag{30}$$

Known elements indexed using the above set can be constrained as follows.

$$J_{const}(X) = \gamma \sum_{l=0}^{n_v} \big(e_{i_l}^T X e_{j_l} - \bar{x}_{i_l, j_l}\big)^2 \tag{31}$$

wherein, $\gamma > 0$ is a constant, and e_i is a vector of zeros with 1 at the i^{th} place.

The Euclidean gradient of $J(X)$ is given below.

$$\begin{aligned} \text{grad } \bar{J} = {} & \alpha\big(\sum_{i=0}^{m} E_i^T (S(X)^{-1} - \frac{1}{2} P(X)^{-1}) E_i\big) \\ & + \beta\big(Z_1^T Z_1 X Z_2^T Z_2\big) \\ & + 2\gamma\big(\sum_{l=0}^{n_v} (e_{i_l}^T X e_{j_l} - \bar{x}_{i_l,j_l}) e_{j_l} e_{i_l}^T\big) \end{aligned} \tag{32}$$

wherein, $\bar{J}$ is the Euclidean extension of J, $P(X) = \sum_{i=0}^{m} E_i X E_i^T$, and $S(X) = \sum_{i=0}^{m} C_k + P(X)$. The Euclidean Hessian of the cost function is calculated below.

$$\begin{aligned} \text{Hess } \bar{J} = & -\alpha\big(\sum_{i=0}^{m} E_i^T (S(X)^{-1} P(U) S(X)^{-1} - \frac{1}{2} P(X)^{-1} P(U) P(X)^{-1}) E_i\big) \\ & + \beta\big(Z_1^T Z_1 U Z_2^T Z_2 + Z_2^T Z_2 U Z_1^T Z_1\big) \\ & + 2\gamma\big(\sum_{l=0}^{n_v} (e_{i_l}^T U e_{j_l} - \bar{x}_{i_l,j_l}) e_{j_l} e_{i_l}^T\big) \end{aligned} \tag{33}$$

Subsequently, the Riemannian Hessian can be calculated using the expression in Eq. (24) and the cost function minimized for the covariance estimates. The Riemannian Hessian of the cost function enables us to use Riemannian Trust Region method for optimization. This method is guaranteed to converge to a critical point of the cost function. For more details regarding the optimization algorithms, the reader is referred to Refs. [20–22].

GEODESIC CONVEXITY

The conditions for global optimality of the critical point found by minimizing the cost function in Eq. (27) are analyzed in this section. First we define a positive linear map. The following discussion is based on Ref. [20].

Definition 9. *A linear map Φ from one Hilbert Space $\mathcal{H}_1$ to another Hilbert space $\mathcal{H}_2$ is called positive if*

$$\mathbf{0} \preceq A \in \mathcal{H}_1 \implies \Phi(A) \succeq \mathbf{0}.$$

It is called strictly positive if $A \succ \mathbf{0} \implies \Phi(A) \succ \mathbf{0}$.

We immediately note that the map $\Phi(X) = C_k + \sum_{i=0}^{m} E_i X E_i^T$ is positive and linear. Additionally, Φ is strictly positive linear if the following condition is true.

$$\text{rank}\Big(\begin{bmatrix} E_0^T \\ E_1^T \\ \vdots \\ E_m^T \end{bmatrix}\Big) = n \tag{34}$$

Under this condition, it has been shown in Ref. [20, Collorary 2.11 and Remark 2.12] that the function $J_{cov}(\cdot)$ is g-convex on the $\mathbb{P}_n$ manifold. The other two parts of the cost function can similarly be proven to be g-convex. Such functions have the property that all the local minimizers of the function are also the global minimizers.

SUMMARY OF THE ALGORITHM

The pseudo code of the algorithm is described below.

Algorithm 1: Geometric Optimization based Covariance Estimation Algorithm

Input: $\hat{x}_0$, $\hat{Q}_0$, $\hat{R}_0$, $\hat{P}_0$, ν_0, λ_0, μ_0, Φ_0, $\{Q_{ij}, R_{ij}$ known$\}$
Data: $y_i\big|_{i=1}^{k}$
Calculate m from Eq. (5)
for $i = 1$ *to* k **do**
  Calculate the measurement $\mathcal{Z}_k$from LHS of Eq. (7)
  Update the distribution of $\mathcal{Z}_k$ to obtain ν_i, λ_i, μ_i, Φ_i given in Eq. (10)
  Calculate the MAP posterior estimate of $Cov(\mathcal{Z}_k)$ from Eq. (11)
  if $i > m$ **then**
    Calculate cost function from Eqs. (27), (28), (29), and (31)
    Calculate the Gradient of the cost function from Eqs. (32) and (21)
    Calculate the Hessian of the cost function from Eqs. (33) and (24)
    Solve the unconstrained GO problem to obtain $\hat{Q}_i$ and $\hat{R}_i$
    Update the state and state error covariance estimates using Eqs. (4)
  else
    Update the state and state error covariance estimates using $\hat{Q}_0$ and $\hat{R}_0$ in Eqs. (4)
  end
end
Result: $\hat{Q}_k$, $\hat{R}_k$, $\hat{P}_{k+1|k}$, $\hat{x}_k$, ν_k, λ_k, μ_k, Φ_k

SIMULATION

The numerical performance of our adaptive scheme is highlighted below. A detectable LTI system which satisfies the assumptions of the adaptive algorithms is simulated.

$$x_k = \begin{bmatrix} 0.8 & 0.2 & 0 \\ 0.3 & 0.5 & 0 \\ 0.1 & 0.9 & 0.7 \end{bmatrix} x_{k-1} + w_{k-1} \tag{35}$$

$$y_k = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} x_k + v_k \tag{36}$$

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. We assume

$$R = \begin{bmatrix} 5 & 0.7 \\ 0.7 & 4 \end{bmatrix}, \quad Q = \begin{bmatrix} 3 & 0.2 & 0 \\ 0.2 & 2 & 0 \\ 0 & 0 & 7.5 \end{bmatrix}$$

Note that since the above system is detectable, only the first two states are observable and hence, the covariance matrix corresponding to the first two states can be estimated. There is no such restriction on the measurement noise covariance matrix. In the first case, the elements R_{12}, R_{22}, and Q_{12} are assumed to be known and the R_{11}, Q_{11}, and Q_{22} are to be estimated. The initial estimates for all the unknown elements was chosen to be 10. The `Manopt` software for optimization on Riemannian manifolds is used for this example.[22] The estimated elements from our algorithm are compared to the results of the baseline adaptive scheme in Fig. (1). In the baseline adaptive scheme, the estimates were calculated using vectorization operation of Eq. (8). Whenever the baseline estimates were not SPD, the most recent SPD estimate was used. This is visualized with the flat regions in the plots. The predicted and the true state error covariance matrix elements are plotted in Figs. (2).

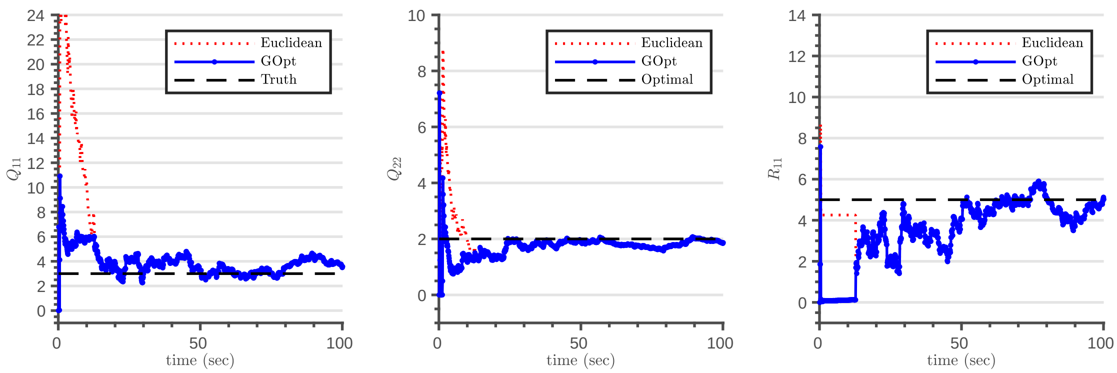

Figure 1. The unknown elements Q_{11}, Q_{12}, and R_{11} with the Euclidean, and the Geometric optimization scheme along with their true values.

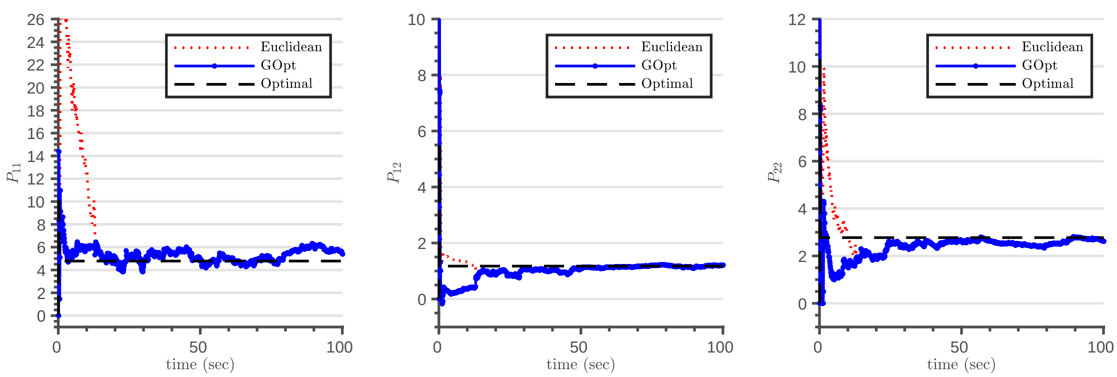

Figure 2. The state error covariance matrix elements for the Euclidean, Riemannian, and the optimal cases.

In the second case, an additional element R_{12} is assumed to be unknown. Note that this does not satisfy the assumptions of our baseline adaptive filter. This is simply because the coefficient matrix S in Eq. (12) of the unknown elements is not full rank. The GO method however proves to be robust and provide SPD estimates of noise covariance matrices. Although the noise covariance matrices are not guaranteed to converge to their true values, the state error covariances do converge and Riemannian adaptive scheme produces reasonable state and state error covariance estimates. In this case, the cost function, although g-convex, has multiple global minimizers which are no longer distinguishable. This was confirmed by checking that value of the cost function at the true values and the estimated values were equal. The estimates of the unknown elements of the noise covariance matrices and the state error covariance estimates are given in Figs. (3) and (4) respectively.

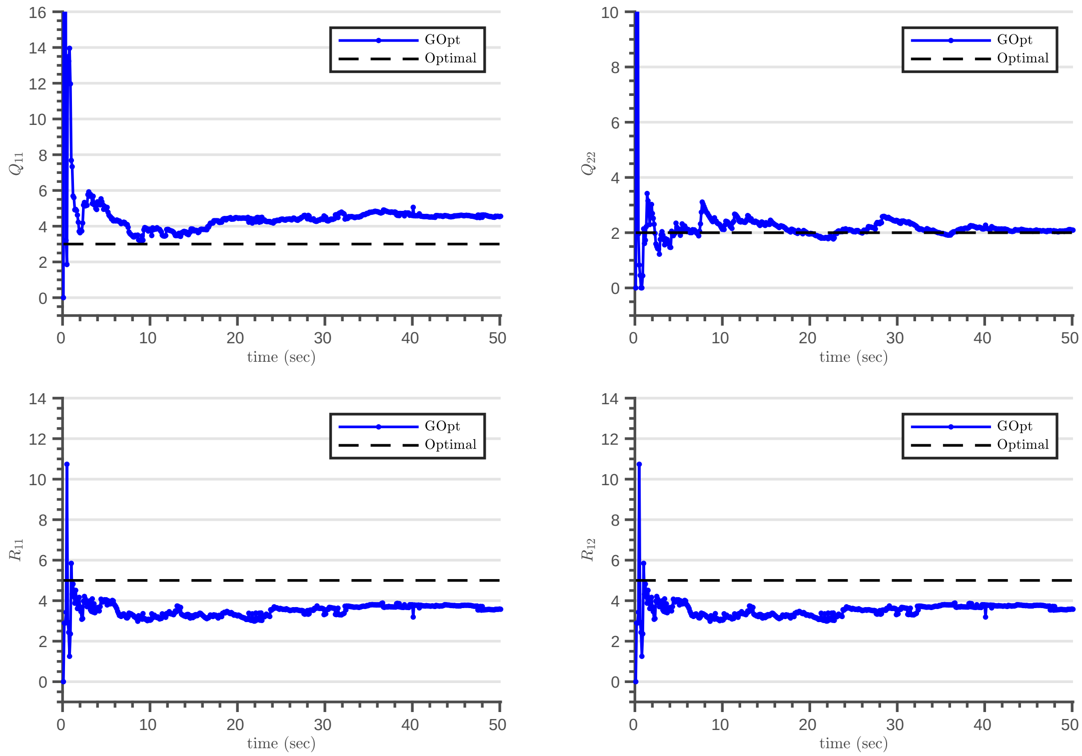

Figure 3. The unknown elements Q_{11}, Q_{12}, and R_{11} with the Euclidean, and the Geometric optimization scheme along with their true values.

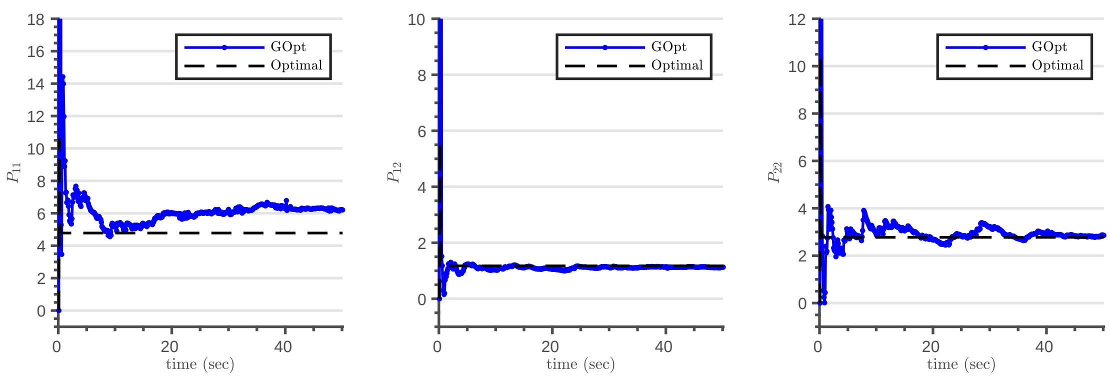

Figure 4. The state error covariance matrix elements for the Euclidean, Riemannian, and the optimal cases.

CONCLUSION

A geometric optimization (GO) based adaptive Kalman filter to estimate the states as well as the process and measurement noise covariance matrices is presented in this paper. A linear matrix equation in the noise covariance matrices was formulated by stacking the measurement equation in a fixed time window. Since the covariance matrices belong to the Riemannian manifold of symmetric positive definite (SPD) matrices, a geometric optimization problem on this manifold was formulated and solved whenever a new measurement is available. A log-det divergence based cost function was used for the optimization. The Riemannian versions of the Gradient and Hessians of the cost function are characterized which are used for second-order optimization techniques on the SPD manifold.

Conditions for the cost function to be geodesically convex were derived which, in turn, guarantee global optimality of all solutions. The previous covariance matching filter failed in cases where the S matrix in Eq. (12) was not full rank. The GO based algorithm had no such limitation and hence was found to be more applicable than its Euclidean counterpart. Even in such degenerate cases, the numerical experiments showed that the state error covariances and the noise covariance estimates converged to a steady state.

The case when the S matrix was rank deficient was found to be the case when the cost function was not strictly g-convex. This, in turn, leads to existence of multiple global minima. Although the Riemannian Trust Region optimization ensures convergence to a critical point, the inability to distinguish between multiple global minima is a limitation of this algorithm. This may occur in part because of the scalar nature of objective functions or simply because of lack of observability properties of the system. Characterizing the degenerate case is important. This analysis is a potential direction of future research. Optimizing over multiple geodesically convex cost functions may yield a different set of globally optimal solutions which may help reduce the degeneracy of such cases.

REFERENCES

[1] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME, Journal of Basic Engineering*, Vol. 82, No. 1, 1960, pp. 35–45.

[2] R. E. Kalman and R. S. Bucy, "New results in linear filtering and prediction theory," *Transactions of the ASME, Journal of Basic Engineering*, Vol. 83, No. 1, 1961, pp. 95–108.

[3] S. Sangsuk-Iam and T. E. Bullock, "Analysis of discrete-time Kalman filtering under incorrect noise covariances," *IEEE Transactions on Automatic Control*, Vol. 35, No. 12, 1990, pp. 1304–1309.

[4] H. Heffes, "The effect of erroneous models on the Kalman filter response," *IEEE Transactions on Automatic Control*, Vol. 11, No. 3, 1966, pp. 541–543.

[5] N. Madjarov and L. Mihaylova, "Kalman filter sensitivity with respect to parametric noises uncertainty," *Kybernetika*, Vol. 32, No. 3, 1996, pp. 307–322.

[6] S. S. Saab, "Discrete-time Kalman filter under incorrect noise covariances," *American Control Conference, Proceedings of the 1995*, Vol. 2, IEEE, 1995, pp. 1152–1156.

[7] S. Sangsuk-Iam and T. E. Bullock, "Analysis of continuous-time Kalman filtering under incorrect noise covariances," *Automatica*, Vol. 24, No. 5, 1988, pp. 659–669.

[8] J. L. Willems and F. M. Callier, "Divergence of the stationary Kalman filter for correct and for incorrect noise variances," *IMA Journal of Mathematical Control and Information*, Vol. 9, No. 1, 1992, pp. 47–54.

[9] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Transactions on Automatic Control*, Vol. 15, No. 2, 1970, pp. 175–184.

[10] R. Mehra, "Approaches to adaptive filtering," *IEEE Transactions on Automatic Control*, Vol. 17, No. 5, 1972, pp. 693–698.

[11] H. Wang, Z. Deng, B. Feng, H. Ma, and Y. Xia, "An adaptive Kalman filter estimating process noise covariance," *Neurocomputing*, Vol. 223, 2017, pp. 12–17.

[12] X. Wang, M.-q. Liu, Z. Fan, and S.-l. Zhang, "A novel approach of noise statistics estimate using H∞ filter in target tracking," *Frontiers of Information Technology & Electronic Engineering*, Vol. 17, No. 5, 2016, pp. 449–457.

[13] Y. Yang and W. Gao, "An optimal adaptive Kalman filter," *Journal of Geodesy*, Vol. 80, No. 4, 2006, pp. 177–183.

[14] J. Duník, O. Straka, O. Kost, and J. 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*, Vol. 31, No. 11, 2017, pp. 1505–1543.

[15] J. Duník, O. Straka, and O. Kost, "Measurement difference autocovariance method for noise covariance matrices estimation," *Decision and Control (CDC), 2016 IEEE 55th Conference on*, IEEE, 2016, pp. 365–370.

[16] B. Feng, M. Fu, H. Ma, Y. Xia, and B. Wang, "Kalman filter with recursive covariance estimation-Sequentially estimating process noise covariance," *IEEE Transactions on Industrial Electronics*, Vol. 61, No. 11, 2014, pp. 6253–6263.

[17] J. Duník, O. Kost, and O. Straka, "Design of measurement difference autocovariance method for estimation of process and measurement noise covariances," *Automatica*, Vol. 90, 2018, pp. 16–24.

[18] R. Moghe, R. Zanetti, and M. R. Akella, "Adaptive Kalman Filter for Detectable Linear Time-Invariant Systems," *Journal of Guidance, Control, and Dynamics*, Vol. 42, No. 10, 2019, pp. 2197–2205, 10.2514/1.G004359.

[19] R. Bhatia, *Positive definite matrices*, Vol. 24. Princeton university press, 2009.

[20] S. Sra and R. Hosseini, "Conic Geometric Optimization on the Manifold of Positive Definite Matrices," *SIAM Journal on Optimization*, Vol. 25, No. 1, 2015, pp. 713–739.

[21] P.-A. Absil, R. Mahony, and R. Sepulchre, *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[22] N. Boumal, B. Mishra, P.-A. Absil, and R. Sepulchre, "Manopt, a Matlab Toolbox for Optimization on Manifolds," *Journal of Machine Learning Research*, Vol. 15, 2014, pp. 1455–1459.

[23] S. Sra and R. Hosseini, "Geometric optimisation on positive definite matrices for elliptically contoured distributions," *Advances in Neural Information Processing Systems*, Vol. 26, 2013, pp. 2562–2570.

[24] A. Wiesel, "Geodesic convexity and covariance estimation," *IEEE transactions on signal processing*, Vol. 60, No. 12, 2012, pp. 6182–6189.

[25] S. T. Smith, "Geometric optimization methods for adaptive filtering," *arXiv preprint arXiv:1305.1886*, 2013.

[26] S. Sra, "Positive definite matrices and the S-divergence," *Proceedings of the American Mathematical Society*, Vol. 144, No. 7, 2016, pp. 2787–2797.

[27] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with applications to tracking and navigation: theory algorithms and software*. John Wiley & Sons, 2004.

[28] B. Anderson and J. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[29] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian data analysis*. CRC press, 2013.

[30] Z. Chebbi and M. Moakher, "Means of Hermitian positive-definite matrices based on the log-determinant α-divergence function," *Linear Algebra and its Applications*, Vol. 436, No. 7, 2012, pp. 1872–1889.

[31] S. Sra, "A new metric on the manifold of kernel matrices with application to matrix geometric means," *Advances in neural information processing systems*, Vol. 25, 2012, pp. 144–152.