Preprints of the 19th World Congress
The International Federation of Automatic Control
Cape Town, South Africa. August 24-29, 2014

Noise Covariance Estimation for Time-varying and Nonlinear Systems

Ming Ge * **Eric C. Kerrigan** **

* *Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, UK (e-mail: ming.ge08@imperial.ac.uk).*
** *Department of Electrical and Electronic Engineering and Department of Aeronautics, Imperial College London, London SW7 2AZ, UK (e-mail: e.kerrigan@imperial.ac.uk).*

Abstract: Kalman-based state estimators assume a priori knowledge of the covariance matrices of the process and observation noise. However, in most practical situations, noise statistics are often unknown and need to be estimated from measurement data. This paper presents a new auto-covariance least squares method for noise covariance estimation of linear time-varying and nonlinear systems. The new algorithm is more general than existing methods, since we do not assume asymptotic stability of the open-loop system trajectories.

Keywords: Auto-covariance Least-squares, Noise Covariance Estimation, State Estimation, Linear Time-varying System, Nonlinear System, Kalman Filter, Extended Kalman Filter

1. INTRODUCTION

The performance of a Kalman filter relies on properly defined noise statistics. Failure to do so in the design of a Kalman filter could result in large estimation errors or even a divergence of state estimates. In the past four decades, many approaches have been taken for improving the accuracy of noise covariance estimation. The pioneering work of noise covariance estimation by Mehra (1972) introduced two correlation least-squares based algorithms, namely output and innovation correlation methods for obtaining noise covariance matrices in linear time-invariant systems. Mehra (1972) also stated conditions for finding a unique solution.

Odelson, Rajamani, and Rawlings (2006) presented a new algorithm for noise covariance estimation of linear time-invariant systems, which is a constrained auto-covariance least-squares (ALS) method inspired by the innovation correlation method of Mehra (1972). Odelson et al. (2006) estimates noise covariance matrices using least-squares semi-definite programming (SDP), which greatly reduces the variance of estimation compared to the innovation correlation method.

Rajamani and Rawlings (2009) introduced a new algorithm called the "one-column version", which is a simplified version of the ALS method. The computational complexity of constructing the ALS problem for high dimensional systems can be significantly reduced. Instead of using the identity matrix as the weight in the least-squares cost function, a method was proposed to calculate the optimal weighting for further minimizing the variance of the estimation error.

All the papers listed above are for linear time-invariant models only. Rajamani and Rawlings (2007) had successfully extended the standard ALS method to time-varying and nonlinear systems. However, due to the structure and approximations in their method, their algorithm may not work for time-varying neutrally stable or open-loop unstable systems. This paper provides a new algorithm for noise covariance estimation that can be applied to more general time-varying and nonlinear systems than those considered in Rajamani and Rawlings (2007).

2. NOMENCLATURE AND PRELIMINARIES

$\mathbb{E}[\cdot]$ and $\mathrm{cov}(\cdot)$ denote the expected value and covariance of a random variable, respectively. $\|.\|_F$ is the Frobenius norm of a matrix. The notation $P \succeq 0$ denotes that matrix P is positive-semidefinite and symmetric. The symbol $\otimes$ is the standard Kronecker product. The symbol $\oplus$ is the matrix direct sum:

$$\bigoplus_{k=1}^{N} G_k := \mathrm{diag}\left(G_1, \ \cdots, \ G_N\right) = \begin{bmatrix} G_1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & G_N \end{bmatrix}.$$

The symbols $\mathbf{1}_{n_r,n_c}$ and $\mathbf{0}_{n_r,n_c}$ represent $n_r \times n_c$ matrices with all entries equal to one or zero, respectively.

The symbol $(\cdot)_s$ denotes the vectorization of a matrix; recall also that

$$(ABC)_s = (C^\top \otimes A)(B)_s.$$

$\mathcal{I}_{N,p} \in \Re^{(pN)^2 \times p^2}$ denotes a permutation matrix that only contains zeros and ones (Odelson et al., 2006) so that

$$\left(I_N \otimes R\right)_s = \mathcal{I}_{N,p}(R)_s$$

The notation $(\cdot)_{ss}$ represents the column-wise stacked lower triangular elements of a symmetric matrix, hence it is possible to establish a relationship between $(\cdot)_s$ and $(\cdot)_{ss}$ as

$$(\hat{Q})_s = \mathscr{D}_r(\hat{Q})_{ss},$$

where $\mathscr{D}_r \in \Re^{r^2 \times \frac{r(r+1)}{2}}$ is a full column rank duplication matrix, which contains only zeros and ones (Magnus and Neudecker, 1999, p.57).

$\mathscr{M}_l^{r,c} \in \Re^{r \times c}$ denotes an auxiliary matrix containing only zeros and ones:

$$\mathscr{M}_l^{r,c} := \begin{bmatrix} \mathbf{0}_{r\times(l-1)} & I_r & \mathbf{0}_{r\times(c-r-l+1)} \end{bmatrix}.$$

$x \sim \mathcal{N}(\mu, P)$ denotes a random variable x with a normal distribution with mean μ and covariance matrix P.

$x \sim \mathcal{U}(a, b)$ denotes a uniformly distributed scalar random variable x within the interval $x \in [a, b]$.

3. LINEAR TIME-VARYING SYSTEMS

Consider the following discrete-time linear time-varying model:

$$\begin{aligned} x_{k+1} &:= A_k x_k + G_k w_k, \\ y_k &:= C_k x_k + v_k, \end{aligned}$$

where $A_k \in \Re^{n\times n}$ and $C_k \in \Re^{p\times n}$ are the dynamics and sensor matrices; $G_k \in \Re^{n\times r}$ is a time-varying noise matrix; $\{x_k\}_{k=1}^M$ is an unknown state sequence; $\{y_k\}_{k=1}^M$ are given output measurements; $\{w_k\}_{k=1}^M$ and $\{v_k\}_{k=1}^M$ are two unknown noise sequences, which affect the state and output, respectively.

Assumption 1. The noise sequences $\{w_k\}_{k=1}^M$ and $\{v_k\}_{k=1}^M$ are i.i.d. random variables having Gaussian (or normal) distributions $\mathcal{N}(0, Q)$ and $\mathcal{N}(0, R)$, respectively, with zero mean and unknown covariance matrices Q and R.

Since the true noise covariance matrices Q and R are unknown, it is not possible to design an unbiased state estimator with a sequence of optimal filter gains $\{L_k\}_{k=1}^M$. Instead, we use a given sequence of sub-optimal filter gains $\{L_k^s\}_{k=1}^M$ and any appropriate given initial state guess $\hat{x}_{1|0}$ to obtain the estimated state sequence

$$\hat{x}_k := \hat{x}_{k|k-1} + L_k^s \left(y_k - \hat{y}_{k|k-1}\right), \quad k = 1, \ldots, M,$$

where $\hat{x}_{k+1|k}$ and $\hat{y}_{k|k-1}$ are the one-step ahead predicted state and output, respectively, given by

$$\begin{aligned} \hat{x}_{k+1|k} &:= A_k \hat{x}_k, \\ \hat{y}_{k|k-1} &:= C_k \hat{x}_{k|k-1}. \end{aligned}$$

The state error terms are defined as:

$$\varepsilon_k := x_k - \hat{x}_{k|k-1}, \qquad k = 1, \ldots, M,$$

hence

$$\begin{aligned} \hat{x}_{k+1|k} &= A_k(\hat{x}_{k|k-1} + L_k^s(y_k - \hat{y}_{k|k-1})) \\ &= A_k \hat{x}_{k|k-1} + A_k L_k^s (C_k x_k + v_k - C_k \hat{x}_{k|k-1}) \end{aligned}$$

where $\hat{x}_{k+1|k}$ and $\hat{y}_{k|k-1}$ are the one-step ahead predicted state and output, respectively, given by

$$\hat{x}_{k+1|k} := A_k \hat{x}_k, \qquad \hat{y}_{k|k-1} := C_k \hat{x}_{k|k-1},$$

and

$$\varepsilon_{k+1} = \underbrace{(A_k - A_k L_k^s C_k)}_{\bar{A}_k} \varepsilon_k + \underbrace{\begin{bmatrix} G_k & -A_k L_k^s \end{bmatrix}}_{\bar{G}_k} \underbrace{\begin{bmatrix} w_k \\ v_k \end{bmatrix}}_{\bar{w}_k}. \tag{1a}$$

Assumption 2. The initial state error term ε_1 has a Gaussian distribution: $\varepsilon_1 \sim \mathcal{N}(0, P_1)$ with zero mean and unknown covariance P_1.

We define the state space model of innovations as

$$z_k := y_k - \hat{y}_{k|k-1},$$

hence

$$z_k = C_k \varepsilon_k + v_k. \tag{1b}$$

A necessary and sufficient condition for the optimality of a Kalman filter is that the innovation sequence $\{z_k\}_{k=1}^M$ be white Gaussian noise (Mehra, 1970). However, for a sub-optimal filter, z_1, z_2, …, z_M are correlated with each other, thus we could produce an auto-covariance matrix of $\{z_k\}_{k=1}^M$ that represents the similarity between the original signal and some time lagged versions of itself.

For any $k \in \{1, \ldots, M\}$, the auto-covariance of vector z_k with j time-lags is defined as:

$$\begin{aligned} \mathscr{C}_j^{(1)}(z_k) &:= \mathbb{E}[(z_{k+j} - \mu_{k+j})(z_k - \mu_k)^\top] \\ &= \mathbb{E}[z_{k+j} z_k^\top] - \mu_{k+j} \mu_k^\top \end{aligned}$$

for $j = 0, 1, \ldots, N-1$, where N is the maximum number of time-lags and $\mu_{k+j} := \mathbb{E}[z_{k+j}]$.

Because the state error term ε_k is a function of ε_1 and $\{\bar{w}_k\}_{k=1}^M$, Assumption 2 ensures that

$$\forall\, j, k : \mu_{k+j} = 0 \implies \mathscr{C}_j^{(1)}(z_k) = \mathbb{E}[z_{k+j} z_k^\top].$$

In order to minimize the effect of the initial guess uncertainty, let us start from ε_{k_0-1}, $(k_0 > 1)$ and pick a fragment of innovations $\{z_k\}_{k=k_0}^{k_0+M_e-N}$. The auto-covariance with j time-lags is then given by:

$$\mathscr{C}_j\left(\{z_k\}_{k=k_0}^{k_0+M_e-N}\right) := \mathbb{E}\left[z_{k_0+j} z_{k_0}^\top \quad \cdots \quad z_{k_0+M_e-N+j} z_{k_0+M_e-N}^\top\right],$$

where M_e is the estimation data length such that $N \ll M_e$ and $M_e \leq M$.

Assumption 3. The state error term ε_{k_0-1} has a Gaussian distribution: $\varepsilon_{k_0-1} \sim \mathcal{N}(0, P_{k_0-1})$ with zero mean and unknown covariance P_{k_0-1}.

The auto-covariance matrix (ACM) of $\{z_k\}_{k=k_0}^{k_0+M_e-N}$ can now be defined as

$$\mathcal{R} := \begin{bmatrix} \mathscr{C}_0\left(\{z_k\}_{k=k_0}^{k_0+M_e-N}\right) \\ \vdots \\ \mathscr{C}_{N-1}\left(\{z_k\}_{k=k_0}^{k_0+M_e-N}\right) \end{bmatrix}.$$

We also define matrix $\mathcal{R}_i$ as

$$\mathcal{R}_i := \mathbb{E}\begin{bmatrix} z_{k_0+i} z_{k_0}^\top \\ \vdots \\ z_{k_0+i+N-1} z_{k_0+i}^\top \end{bmatrix}, \quad i = 0, \ldots, M_e - N,$$

so that

$$\mathcal{R} = \begin{bmatrix} \mathcal{R}_0 & \mathcal{R}_1 & \cdots & \mathcal{R}_{M_e-N} \end{bmatrix}.$$

Note that the auto-covariance matrix $\mathcal{R}$ is a function of Q, R and P_{k_0-1}; an expression for $\mathcal{R}$ will be given by (4) in Section 4.

Next, we define the sample estimate of $\mathcal{R}_i$ as

$$\bar{\mathcal{R}}_i := \frac{1}{M_b - N + 1} \times \begin{bmatrix} z_{k_0+i} & \cdots & z_{k_0+i+M_b-N} \\ z_{k_0+i+1} & \cdots & z_{k_0+i+M_b-N+1} \\ \vdots & \ddots & \vdots \\ z_{k_0+i+N-1} & \cdots & z_{k_0+i+M_b-1} \end{bmatrix} \times \begin{bmatrix} z_{k_0+i}^\top \\ z_{k_0+i+1}^\top \\ \vdots \\ z_{k_0+i+M_b-N}^\top \end{bmatrix}$$

and

$$\bar{\mathcal{R}} := \begin{bmatrix} \bar{\mathcal{R}}_0 & \bar{\mathcal{R}}_1 & \cdots & \bar{\mathcal{R}}_{M_e-N} \end{bmatrix},$$

where M_b is the data length for estimating $\bar{\mathcal{R}}_i$ such that $M_e \leq M_b$. Building the matrix $\bar{\mathcal{R}}$ requires the innovation sequence $\{z_k\}_{k=k_0}^{k_0+M_e+M_b-N-1}$, hence, for fixed k_0, M_e, M_b and N, we require $M \geq k_0 + M_e + M_b - N - 1$.

We can now define a least squares optimization problem to estimate the true covariance:

$$(P^*_{k_0-1}, Q^*, R^*) := \arg \min_{\hat{P}_{k_0-1}, \hat{Q}, \hat{R}} \left\| \mathcal{R}(\hat{P}_{k_0-1}, \hat{Q}, \hat{R}) - \bar{\mathcal{R}} \right\|_F^2 \quad \text{s.t. } \hat{P}_{k_0-1},\ \hat{Q},\ \hat{R} \succeq 0. \tag{2}$$

Compared to time-invariant system models, the estimate error covariance $P_k := \mathbb{E}[(x_k - \hat{x}_{k|k-1})(x_k - \hat{x}_{k|k-1})^\top]$ in our model is the solution to a time-varying Riccati equation and does not reach a steady state value. Therefore, the state and measurement noise covariance Q and R cannot be estimated from a Lyapunov equation as in Rajamani and Rawlings (2007).

4. SOLVING THE OPTIMIZATION PROBLEM

Let us start from ε_{k_0-1} and consider the evolution of (1a) and (1b). The innovation sequence $\{z_k\}_{k=k_0}^{k_0+M_e-1}$ can be shown to be given by

$$z = \tilde{\mathcal{V}} \left(\tilde{\mathcal{E}} \varepsilon_{k_0-1} + \tilde{\mathcal{G}} \tilde{w} \right) + \tilde{v}, \tag{3}$$

where

$$\tilde{\mathcal{V}} := \tilde{\mathcal{C}} \tilde{\mathcal{A}}^{-1}, \quad \tilde{\mathcal{E}} := \begin{bmatrix} \bar{A}_{k_0-1} \\ \mathbf{0} \end{bmatrix}, \quad \tilde{\mathcal{G}} := \bigoplus_{k=k_0-1}^{k_0+M_e-2} \bar{G}_k,$$

$$\tilde{\mathcal{C}} := \bigoplus_{k=k_0}^{k_0+M_e-1} C_k, \quad \tilde{\mathcal{A}} := I_{nM_e} - \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \bigoplus_{k=k_0}^{k_0+M_e-2} \bar{A}_k & \mathbf{0} \end{bmatrix},$$

$$z := \begin{bmatrix} z_{k_0}^\top & z_{k_0+1}^\top & \cdots & z_{k_0+M_e-2}^\top & z_{k_0+M_e-1}^\top \end{bmatrix}^\top,$$
$$\tilde{w} := \begin{bmatrix} \bar{w}_{k_0-1}^\top & \bar{w}_{k_0}^\top & \cdots & \bar{w}_{k_0+M_e-3}^\top & \bar{w}_{k_0+M_e-2}^\top \end{bmatrix}^\top,$$
$$\tilde{v} := \begin{bmatrix} v_{k_0}^\top & v_{k_0+1}^\top & \cdots & v_{k_0+M_e-2}^\top & v_{k_0+M_e-1}^\top \end{bmatrix}^\top.$$

It is possible to use the above expressions to show that

$$\mathcal{R}(P_{k_0-1}, Q, R) = \Gamma \big(I_{M_e-N+1} \otimes P_{k_0-1} \big) \bar{\Gamma}^\top + \Omega \big(I_{N_d} \otimes Q \big) \bar{\Omega}^\top + \Phi \big(I_{N_d} \otimes R \big) \bar{\Phi}^\top + \Psi \big(I_{M_e-N+1} \otimes R \big), \tag{4}$$

where $N_d := \frac{(M_e-N+2)(M_e-N+1)}{2}$ and

$$\Gamma := \tilde{\mathcal{S}} \tilde{\mathcal{F}}, \quad \bar{\Gamma} := \tilde{\mathcal{S}}^d \tilde{\mathcal{F}}, \quad \Omega := \tilde{\mathcal{S}} \tilde{\mathcal{J}}, \quad \bar{\Omega} := \tilde{\mathcal{S}}^d \tilde{\mathcal{J}},$$
$$\tilde{\mathcal{F}}_s := \tilde{\mathcal{V}} \tilde{\mathcal{E}}, \qquad \tilde{\mathcal{F}} := I_{M_e-N+1} \otimes \tilde{\mathcal{F}}_s,$$
$$\Phi := \tilde{\mathcal{S}} \tilde{\mathcal{U}}, \quad \bar{\Phi} := \tilde{\mathcal{S}}^d \tilde{\mathcal{U}}, \quad \Psi := \begin{bmatrix} \mathbf{1}_{p \times p(M_e-N+1)} \\ \tilde{\mathcal{P}} \tilde{\mathcal{O}} \end{bmatrix},$$

$$\tilde{\mathcal{B}} := \tilde{\mathcal{V}} \bigoplus_{k=k_0-1}^{k_0+M_e-2} G_k, \quad \tilde{\mathcal{D}} := -\tilde{\mathcal{V}} \bigoplus_{k=k_0-1}^{k_0+M_e-2} A_k L_k^s,$$
$$\tilde{\mathcal{P}}_i := \mathscr{M}_{p(i+1)+1}^{p(N-1),pM_e}, \qquad \tilde{\mathcal{H}} := \mathscr{M}_1^{p,pN},$$
$$\tilde{\mathcal{J}}_i := \left(\mathscr{M}_1^{r(i+1),rM_e} \right)^\top, \quad \tilde{\mathcal{J}} := \bigoplus_{i=0}^{M_e-N} \tilde{\mathcal{B}} \tilde{\mathcal{J}}_i,$$
$$\tilde{\mathcal{U}}_i := \left(\mathscr{M}_1^{p(i+1),pM_e} \right)^\top, \quad \tilde{\mathcal{U}} := \bigoplus_{i=0}^{M_e-N} \tilde{\mathcal{D}} \tilde{\mathcal{U}}_i,$$
$$\tilde{\mathcal{O}}_i := \left(\mathscr{M}_{p+1}^{p,pM_e} \right)^\top, \qquad \tilde{\mathcal{O}} := \bigoplus_{i=0}^{M_e-N} \tilde{\mathcal{D}} \tilde{\mathcal{O}}_i,$$
$$\tilde{\mathcal{S}}_i := \mathscr{M}_{pi+1}^{pN,pM_e}, \quad \tilde{\mathcal{S}} := \begin{bmatrix} \tilde{\mathcal{S}}_0 & \tilde{\mathcal{S}}_1 & \cdots & \tilde{\mathcal{S}}_{M_e-N} \end{bmatrix},$$
$$\tilde{\mathcal{S}}^d := \bigoplus_{i=0}^{M_e-N} \tilde{\mathcal{H}} \tilde{\mathcal{S}}_i, \quad \tilde{\mathcal{P}} := \begin{bmatrix} \tilde{\mathcal{P}}_0 & \tilde{\mathcal{P}}_1 & \cdots & \tilde{\mathcal{P}}_{M_e-N} \end{bmatrix}.$$

In order to fit the problem into a standard linear least-squares formulation, we must vectorize the matrix $\mathcal{R}$, which is the column-wise stacking of a matrix into a vector. Hence, the vectorized matrix $(\mathcal{R})_s$ can be expressed as

$$(\mathcal{R})_s = (\bar{\Gamma} \otimes \Gamma) \mathcal{I}_{(M_e-N+1),n} (P_{k_0-1})_s + (\bar{\Omega} \otimes \Omega) \mathcal{I}_{N_d,r} (Q)_s + \left((\bar{\Phi} \otimes \Phi) \mathcal{I}_{N_d,p} + (I_{p(M_e-N+1)} \otimes \Psi) \mathcal{I}_{(M_e-N+1),p} \right) (R)_s.$$

Consider the dimension and dense structure of matrices $\bar{\Gamma}$, Γ, $\bar{\Omega}$, Ω, $\bar{\Phi}$, Φ and Ψ; calculating the Kronecker product of these matrices will be extremely slow and require significant amounts of computer memory. Alternatively, we could parallelize the computation of each vector $(\mathcal{R}_i)_s$ and combine them together to form the vector $(\mathcal{R})_s$. The vectorized matrix $(\mathcal{R}_i)_s$ can be expressed as

$$(\mathcal{R}_i)_s = (\bar{\Gamma}_i \otimes \Gamma_i) \mathcal{I}_{1,n} (P_{k_0-1})_s + (\bar{\Omega}_i \otimes \Omega_i) \mathcal{I}_{i+1,r} (Q)_s + ((\bar{\Phi}_i \otimes \Phi_i) \mathcal{I}_{i+1,p} + I_p \otimes \Psi_i)(R)_s,$$

where

$$\Gamma_i := \tilde{\mathcal{S}}_i \tilde{\mathcal{F}}_s, \quad \bar{\Gamma}_i := \tilde{\mathcal{H}} \Gamma_i, \quad \Omega_i := \tilde{\mathcal{S}}_i \tilde{\mathcal{B}} \tilde{\mathcal{J}}_i, \quad \bar{\Omega}_i := \tilde{\mathcal{H}} \Omega_i$$
$$\Phi_i := \tilde{\mathcal{S}}_i \tilde{\mathcal{D}} \tilde{\mathcal{U}}_i, \quad \bar{\Phi}_i := \tilde{\mathcal{H}} \Phi_i, \quad \Psi_i := \begin{bmatrix} I_p \\ \tilde{\mathcal{P}}_i \tilde{\mathcal{D}} \tilde{\mathcal{O}}_i \end{bmatrix}.$$

Let $\bar{b} := (\bar{\mathcal{R}})_s$. We can now rearrange our original optimization problem (2) into a least-squares problem with decision variables $\hat{P}_{k_0-1}, \hat{Q}$ and $\hat{R}$:

$$\min_\theta \left\| \underbrace{\begin{bmatrix} \mathscr{A}_0 \\ \vdots \\ \mathscr{A}_{M_e-N} \end{bmatrix}}_{\mathscr{A}} \underbrace{\begin{bmatrix} (\hat{P}_{k_0-1})_{ss} \\ (\hat{Q})_{ss} \\ (\hat{R})_{ss} \end{bmatrix}}_{\theta} - \underbrace{\begin{bmatrix} \bar{b}_0 \\ \vdots \\ \bar{b}_{M_e-N} \end{bmatrix}}_{\bar{b}} \right\|_2^2 \quad \text{s.t. } \hat{P}_{k_0-1},\ \hat{Q},\ \hat{R} \succeq 0, \tag{5}$$

where

$$\mathscr{A}_i := \Big[(\bar{\Gamma}_i \otimes \Gamma_i) \mathcal{I}_{1,n} \mathscr{D}_n \quad (\bar{\Omega}_i \otimes \Omega_i) \mathcal{I}_{i+1,r} \mathscr{D}_r \quad \big((\bar{\Phi}_i \otimes \Phi_i) \mathcal{I}_{i+1,p} + I_p \otimes \Psi_i \big) \mathscr{D}_p \Big],$$
$$\bar{b}_i := (\bar{\mathcal{R}}_i)_s.$$

5. PROPERTIES OF ALS ESTIMATE AND DISCUSSION

In this section, we analyze the properties of the ALS estimate, followed by a discussion and comparison with the approach introduced in Rajamani and Rawlings (2009).

Constructing the sequence of sub-optimal filter gains $\{L_k^s\}_{k=1}^M$ requires initial guesses of the noise covariance matrices, Q_g and R_g. These can either be chosen to be, for example

$$Q_g = I_r, \quad R_g = I_p,$$

or estimated using the output correlation method introduced by Mehra (1972), which does not require prior estimates of Q, R or L_k.

The correlation between z_k and z_{k+j} will keep decreasing and eventually become uncorrelated as the time-lag j increases. Hence, the maximum number of time-lags N can be determined by looking at the plot of the autocorrelation function of the innovation sequence $\{z_k\}_{k=k_0}^M$ against the time-lag variable, where for all $j > N$ the correlations between z_k and z_{k+j} are negligible.

Lemma 1. For a stochastic process $\{z_k\}_{k=k_0}^{k_0+M_b-N}$, the estimate of the auto-covariance

$$\widehat{\mathscr{C}_j}\left(\{z_k\}_{k=k_0}^{k_0+M_b-N}\right) := \frac{1}{M_b-N+1}\times \begin{bmatrix} \sum_{k=k_0}^{k_0+M_b-N} z_{k+j}z_k^\top & \cdots & \sum_{k=k_0+M_b-N}^{k_0+2M_b-2N} z_{k+j}z_k^\top \end{bmatrix}$$

is a consistent unbiased estimate [1] of $\mathscr{C}_j\left(\{z_k\}_{k=k_0}^{k_0+M_b-N}\right)$, such that

$$\mathbb{E}[\widehat{\mathscr{C}_j}] = \mathscr{C}_j \quad \forall\ M_b,\ j,$$
$$\operatorname{cov}(\widehat{\mathscr{C}_j}) \longrightarrow 0 \ \text{ as } \ M_b \longrightarrow \infty,$$

and

$$\widehat{\mathscr{C}_j} \longrightarrow \mathscr{C}_j \ \text{ as } \ M_b \longrightarrow \infty.$$

Proof. The proof of this result can be found in Jenkins and Watts (1968, p. 174-180). Another way of stating this fact is that the ensemble average $\mathscr{C}_j$ can be estimated by the time average $\widehat{\mathscr{C}_j}$; this is usually referred to as the ergodic property (Jenkins and Watts, 1968, p. 222). ■

Theorem 1. If the linear least squares problem (2) has a unique solution, then the ALS noise covariance estimates Q^* and R^* are unbiased for all sample sizes and converges asymptotically to the true covariance matrices Q and R as $M_b \to \infty$.

Proof. A similar proof for LTI systems can be found in Odelson et al. (2006). Recall the linear least-squares problem (5):

$$\theta^* := \arg\min_\theta \|\mathscr{A}\theta - \bar{b}\|_2^2.$$

A necessary and sufficient condition for a unique solution requires matrix $\mathscr{A}$ to be full rank (Lawson and Hanson, 1995, p. 36). Hence, the expected value of the estimate of $\hat{\theta}$ is

$$\begin{aligned}\mathbb{E}[\theta^*] &= \left(\mathscr{A}^\top\mathscr{A}\right)^{-1}\mathscr{A}^\top\mathbb{E}[\bar{b}] \\ &= \left(\mathscr{A}^\top\mathscr{A}\right)^{-1}\mathscr{A}^\top b \quad \text{(by Lemma 1, } \mathbb{E}[\bar{b}] = b) \\ &= \left(\mathscr{A}^\top\mathscr{A}\right)^{-1}\mathscr{A}^\top\mathscr{A}\theta = \theta.\end{aligned}$$

The covariance of the estimate is

$$\operatorname{cov}(\theta^*) = \left(\mathscr{A}^\top\mathscr{A}\right)^{-1}\mathscr{A}^\top\operatorname{cov}(\bar{b})\mathscr{A}\left(\mathscr{A}^\top\mathscr{A}\right)^{-1}.$$

According to Lemma 1, $\operatorname{cov}(\bar{b}) \to 0$ as $M_b \to \infty$, so that

$$\operatorname{cov}(\theta^*) \to 0 \quad \text{as} \quad M_b \to \infty.$$

Therefore, θ^* is unbiased for all sample sizes and converges asymptotically to the true θ as $M_b \to \infty$. ■

Rajamani and Rawlings (2007) provided a different approach for estimating the noise covariances for time-varying and nonlinear systems, which is based on the assumption that

$$\lim_{N\to\infty}\left(\prod_{k=k_0}^{N}\bar{A}_k\right)\varepsilon_{k_0} = 0. \tag{6}$$

The advantage of their algorithm is that the number of decision variables in the objective function is reduced from three to just two vectorized matrices: $(Q)_s$ and $(R)_s$. Hence, the computational effort of solving the auto-covariance least square is reduced.

However, for neutrally stable systems, the statement (6) does not hold. Even if the assumption is true, in order to remove the term $\hat{P}_{k_0-1}$ from the decision variables, the computational complexity to ensure

$$\left(\prod_{k=k_0}^{k_0+N}\bar{A}_k\right)\hat{P}_{k_0-1}\left(\prod_{k=k_0}^{k_0+N}\bar{A}_k\right)^\top \approx 0$$

mainly depends on the value of N, dimension and sparsity of matrices $\bar{A}_k$.

6. NONLINEAR SYSTEMS

Consider the following discrete-time nonlinear state space model:

$$x_{k+1} := f(x_k) + g(x_k)w_k$$
$$y_k := h(x_k) + v_k$$

where $x_k \in \Re^n$ is the state, $y_k \in \Re^p$ is the measurement, w_k and v_k satisfy Assumption 1. If we linearize the nonlinear functions $f(x_k)$, $g(x_k)$ and $h(x_k)$ around the current estimate $\hat{x}_k$ and predicted state $\hat{x}_{k|k-1}$, then we have

$$A_k := \left.\frac{\partial f(x_k)}{\partial x_k}\right|_{x_k=\hat{x}_k}, \quad G_k := g(\hat{x}_k), \tag{7a}$$

$$C_k := \left.\frac{\partial h(x_k)}{\partial x_k}\right|_{x_k=\hat{x}_{k|k-1}}. \tag{7b}$$

Assumption 3 requires the estimation errors $\{\varepsilon_k\}_{k=1}^M$ to be bounded and the expectation of ε_k to converge to zero as k increases. For nonlinear state estimation using the extended Kalman filter (EKF), Reif et al. (1999) states five conditions for improving the stability and convergence of

[1] An estimate is said to be unbiased if the expected value of the estimated variable is equal to the true value. If the covariance of an unbiased estimate tends to zero as the sample size M_b increases, the estimator is said to be consistent.

the EKF, including observability, small initial estimation error, small noise terms and no model error.

Given an appropriate guess of initial state $\hat{x}_{1|0}$ and noise covariance matrices Q_g and R_g, if Assumption 3 holds for the EKF, then the true noise covariance matrices Q and R can be estimated using the time-varying ALS method introduced in the previous section.

7. NUMERICAL EXAMPLES

7.1 Linear Time-varying System

Consider a neutrally stable linear time-varying system, for which the evolution of each state follows a random walk and the output matrix C_k varies with time:

$$x_{k+1} := \underbrace{\begin{bmatrix}1&0&0\\0&1&0\\0&0&1\end{bmatrix}}_{A_k} x_k + \underbrace{\begin{bmatrix}1\\1\\1\end{bmatrix}}_{G_k} w_k,$$
$$y_k := C_k x_k + v_k,$$

where $x_k \in \Re^{3\times 1}$, w_k and v_k satisfy Assumption 1 and the time-varying output matrix $C_k \in \Re^{1\times 3}$ consists of a pre-generated irregular sinusoidal wave [2] $\{\eta_k\}_{k=1}^{M+2}$, such that

$$C_k := [\eta_{k+2}\ \eta_{k+1}\ \eta_k], \qquad k = 1, \ldots, M.$$

We randomly pick an initial state $x_1 \sim \mathcal{N}(0, I_3)$, then generate output measurements $\{y_k\}_{k=1}^{5000}$ based on noise covariances [3] $Q = 4 \times 10^{-3}$ and $R = 5.5 \times 10^{-2}$. By choosing an initial state error covariance $P_{1|0} = I_3$ and the guessed initial state $\hat{x}_{1|0} = \mathbf{0}_{3,1}$. The sub-optimal filter gains $\{L_k^s\}_{k=1}^{M}$ and the state error covariance $\{P_k\}_{k=1}^{M}$ can be obtained from the Kalman filter equations (Humpherys et al., 2012):

$$P_{k|k-1} := A_k P_{k-1} A_k^\top + G_k Q_g G_k^\top, \tag{8a}$$
$$L_k^s := \left(P_{k|k-1} C_k^\top\right)\left(C_k P_{k|k-1} C_k^\top + R_g\right)^{-1}, \tag{8b}$$
$$P_k := \left(I - L_k^s C_k\right) P_{k|k-1}, \tag{8c}$$

with guessed noise covariances $Q_g = 1$ and $R_g = 1$.

We re-generate output measurements using different initial state and noise sequences and repeat the simulation 200 times with $k_0 = 600$, $N = 30$, $M_e = 50$ and different values of M_b. Table 1 presents the mean and variance of the estimated results compared to their true values. Note that, as M_b increases, the variance of the 200 estimated results gets smaller and the mean of estimated results becomes more accurate, which verifies the statement in Theorem 1.

Figure 1 is a scatter plot of 200 estimations of the noise covariances Q^* and R^*, as well as the average of all 200 estimations. Figure 2 shows that, in this case, the eigenvalues of $\prod_{k=1}^{N} \bar{A}_k$ do not go to zero as N increases, therefore (6) does not hold.

7.2 Nonlinear System

Let us now consider tracking a sinusoidal wave whose amplitude, phase and frequency follow a random walk:

[2] The wave was generated based on the "nonlinear cyclical model" in Section 3.2.3 of Fusco (2009) with sample period $T_s = 0.1$ sec and an average amplitude of 1.5 m.

[3] The signal to noise ratio (SNR) is 5.03 dB.

Table 1. Noise Covariance Estimation Results

Q	Mean of Q^*	Variance of Q^*
True Value	4×10^{-3}	
$M_b = 500$	4.142×10^{-3}	3.148×10^{-6}
$M_b = 1000$	4.075×10^{-3}	1.439×10^{-6}
$M_b = 2000$	4.026×10^{-3}	6.201×10^{-7}
R	Mean of R^*	Variance of R^*
True Value	5.5×10^{-2}	
$M_b = 500$	5.813×10^{-2}	1.394×10^{-5}
$M_b = 1000$	5.646×10^{-2}	7.344×10^{-6}
$M_b = 2000$	5.531×10^{-2}	3.278×10^{-6}

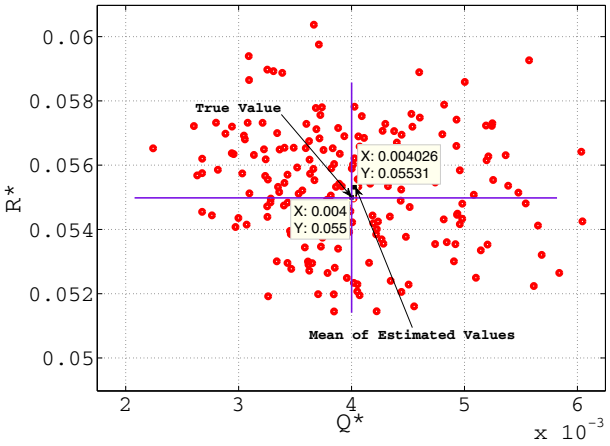

Fig. 1. Estimation of Noise Covariances Q^* and R^*

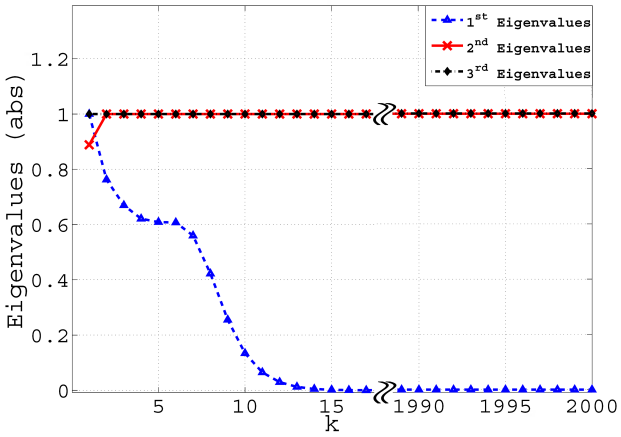

Fig. 2. Eigenvalues of Matrix $\prod_{k=1}^{N} \bar{A}_k$ as N increases

$$\underbrace{\begin{bmatrix}a_{k+1}\\b_{k+1}\\c_{k+1}\end{bmatrix}}_{x_{k+1}} := \underbrace{\begin{cases}+a_k \cos(c_k T_s) + b_k \sin(c_k T_s)\\ -a_k \sin(c_k T_s) + b_k \cos(c_k T_s)\\ +c_k\end{cases}}_{f(x_k)} + \underbrace{\begin{bmatrix}w_k^a\\w_k^b\\w_k^c\end{bmatrix}}_{w_k}$$
$$y_k := \underbrace{[0.8\ \ -0.5\ \ 0]\begin{bmatrix}a_k\\b_k\\c_k\end{bmatrix}}_{h(x_k)} + v_k$$

where $T_s = 0.1$ sec is the sampling time and w_k and v_k satisfy Assumption 1. We randomly pick an initial state from a uniform distribution, such that

Table 2. Noise Covariance Estimation Results

	True Value	Mean	Variance
$Q(1,1)$	3×10^{-4}	2.792×10^{-4}	8.361×10^{-9}
$Q(2,2)$	3×10^{-4}	2.735×10^{-4}	1.196×10^{-8}
$Q(3,3)$	2×10^{-4}	2.380×10^{-4}	3.771×10^{-9}
R	1×10^{-4}	1.039×10^{-4}	1.750×10^{-10}

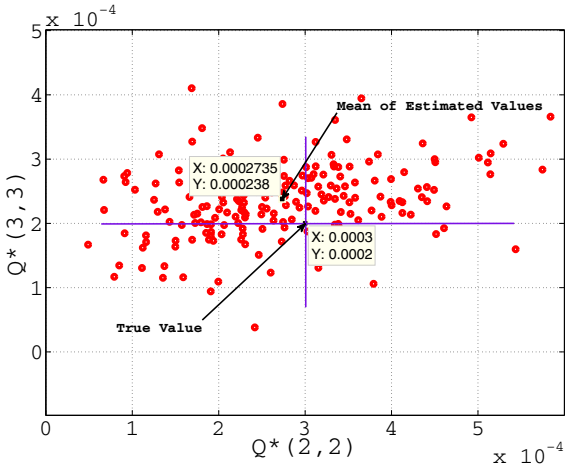

Fig. 3. Estimation of Noise Covariance $Q^*_{(2,2)}$ and $Q^*_{(3,3)}$

$$x_1 := \begin{bmatrix} x_1^1 \sim \mathcal{U}(-1,1) \\ x_1^2 \sim \mathcal{U}(-1,1) \\ x_1^3 \sim \mathcal{U}(0.3,\pi) \end{bmatrix},$$

then generate output measurements $\{y_k\}_{k=1}^{5000}$ based on noise covariance matrices [4]:

$$Q = \text{diag}\left(3 \times 10^{-4},\ 3 \times 10^{-4},\ 2 \times 10^{-4}\right),\ R = 1 \times 10^{-4}.$$

The initial state error covariance is $P_{1|0} = 0.1 \times I_3$ and the guessed initial state is $\hat{x}_{1|0} = \mathbf{0}_{3,1}$. The sub-optimal filter gains $\{L_k^s\}_{k=1}^M$ and the state error covariance $\{P_k\}_{k=1}^M$ are obtained from the EKF equations (7) and (8) with the guessed noise covariance matrices $Q_g = I_3$ and $R_g = 1$.

Similar to Example 7.1, we repeat the simulation 200 times with $k_0 = 600$, $N = 50$ and $M_e = 300$. Table 2 compares the mean and variance of the estimated results to their true values for $M_b = 2000$. Figures 3 and 4 are scatter plots of 200 estimates of some components of the noise covariance matrices Q^* and R^*. As in Example 7.1, we have also observed that the covariance of the estimates decrease as M_b increases.

8. CONCLUSIONS

In this paper, we have developed a noise covariance estimation algorithm for time-varying and nonlinear systems based on a constrained (positive semi-definite) autocovariance least-squares method. We used two examples to investigate the performance of the algorithm and both of them returned relatively good results. Compared with Rajamani and Rawlings (2007), our method added one more term into the decision variable during optimization, but in return can deal with more general models, such as

[4] The noise covariances are small, because we assume the irregular sinusoidal output $\{y_k\}_{k=1}^{5000}$ changed slowly over time with sampling frequency 10 Hz and the an average amplitude between 1 and 1.5 m. The signal to noise ratio (SNR) is 31.9 dB.

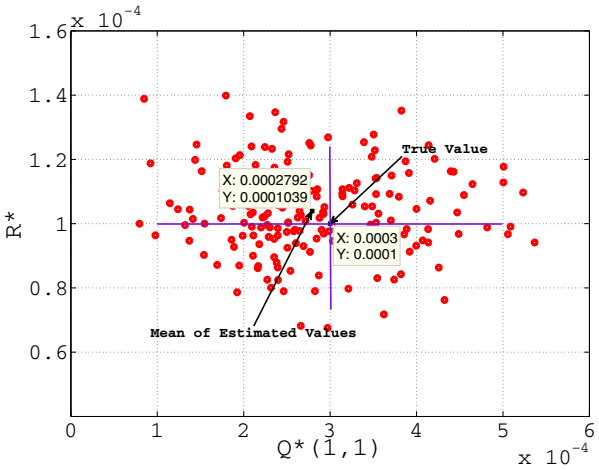

Fig. 4. Estimation of Noise Covariance $Q^*_{(1,1)}$ and R^*

neutrally stable systems. The overall computational time of constructing and solving the optimization problem can be significantly reduced by using parallel implementations and efficient SDP solvers. Future work could include applying this method in various fields for improving state estimation and prediction accuracy, such as chemical engineering, ocean wave prediction for wave energy converters and motion tracking using inertial measurement units.

REFERENCES

Fusco, F. (2009). Short-term wave forecasting as a univariate time series problem. Technical Report 2009-03, National University of Ireland.

Humpherys, J., Redd, P., and West, J. (2012). A fresh look at the Kalman filter. *SIAM Review*, 54(4), 801–823.

Jenkins, G.M. and Watts, D.G. (1968). *Spectral Analysis and its Applications*. Holden-Day, San Fransisco, CA.

Lawson, C.L. and Hanson, R.J. (1995). *Solving Least Squares Problems*. PA: SIAM, Philadelphia.

Magnus, J.R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, New York.

Mehra, R.K. (1970). On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, 15(2), 175–184.

Mehra, R.K. (1972). Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, 15(12), 903–908.

Odelson, B.J., Rajamani, M.R., and Rawlings, J.B. (2006). A new autocovariance least squares method for estimating noise covariance. *Automatica*, 42(2), 532–540.

Rajamani, M.R. and Rawlings, J.B. (2007). Application of a new data-based covariance estimation technique to a nonlinear industrial blending drum. Technical Report 2007-03, University of Wisconsin-Madison.

Rajamani, M.R. and Rawlings, J.B. (2009). Estimation of the disturbance structure from data using semidefinite programming and optimal weighting. *Automatica*, 45(1), 142–148.

Reif, K., Günther, S., Yaz, E., and Unbehauen, R. (1999). Stochastic stability of the discrete-time extended Kalman filter. *IEEE Transactions on Automatic Control*, 44(4), 714–728.