Available online at www.sciencedirect.com
SCIENCE DIRECT®

automatica

Automatica 42 (2006) 303–308

www.elsevier.com/locate/automatica

Brief paper

A new autocovariance least-squares method for estimating noise covariances☆

Brian J. Odelson[a], Murali R. Rajamani[b], James B. Rawlings[b,*]

[a] BP Research and Technology, 150 W. Warrenville Rd., Naperville, IL 60563, USA
[b] Department of Chemical and Biological Engineering, University of Wisconsin–Madison, 1415 Engineering Drive, Madison, WI 53706, USA

Received 25 August 2003; received in revised form 18 April 2005; accepted 7 September 2005
Available online 23 November 2005

Abstract

Industrial implementation of model-based control methods, such as model predictive control, is often complicated by the lack of knowledge about the disturbances entering the system. In this paper, we present a new method (constrained ALS) to estimate the variances of the disturbances entering the process using routine operating data. A variety of methods have been proposed to solve this problem. Of note, we compare ALS to the classic approach presented in Mehra [(1970). On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, *15*(12), 175–184]. This classic method, and those based on it, use a three-step procedure to compute the covariances. The method presented in this paper is a one-step procedure, which yields covariance estimates with lower variance on all examples tested. The formulation used in this paper provides necessary and sufficient conditions for uniqueness of the estimated covariances, previously not available in the literature. We show that the estimated covariances are unbiased and converge to the true values with increasing sample size. The proposed method also guarantees positive semidefinite covariance estimates by adding constraints to the ALS problem. The resulting convex program can be solved efficiently.

Keywords: Adaptive Kalman filter; Covariance estimation; Optimal estimation; Semidefinite programming; State estimation

1. Introduction

Model-based control methods, such as model predictive control (MPC), have become popular choices for solving difficult control problems. Higher performance, however, comes at a cost of greater required knowledge about the process being controlled. Expert knowledge is often required to properly commission and maintain the regulator, target calculator, and state estimator of MPC, for example. This paper addresses the required knowledge for the state estimator, and describes a technique with which ordinary closed-loop data may be used to remove some of the information burden from the user. Consider the usual linear, time-invariant, discrete-time model

$$x_{k+1} = Ax_k + Bu_k + Gw_k,$$
$$y_k = Cx_k + v_k$$

in which $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $G \in \mathbb{R}^{n\times g}$, $C \in \mathbb{R}^{p\times n}$, and $\{w_k\}_{k=0}^{N_d}$ and $\{v_k\}_{k=0}^{N_d}$ are uncorrelated zero-mean Gaussian noise sequences with covariances Q_w and R_v, respectively. The sequence $\{u_k\}_{k=0}^{N_d}$ is assumed to be a known input resulting from the actions of a controller. State estimates of the system are considered using a linear, time-invariant state estimator

$$\widehat{x}_{k+1|k} = A\widehat{x}_{k|k} + Bu_k,$$
$$\widehat{x}_{k|k} = \widehat{x}_{k|k-1} + L[y_k - C\widehat{x}_{k|k-1}]$$

in which L is the estimator gain, which is not necessarily the optimal gain. We denote the residuals of the output equations $(y_k - C\widehat{x}_{k|k-1})$ as the L-innovations when calculated using a state estimator with gain L. In order to use the optimal filter,

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Marco Campi under the direction of Editor I. Petersen.

* Corresponding author. Tel.: +1 608 263 5859; fax: +1 608 265 8794.
E-mail addresses: odelbj@bp.com (B.J. Odelson), rajamani@bevo.che.wisc.edu (M.R. Rajamani), rawlings@engr.wisc.edu (J.B. Rawlings).

0005-1098/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2005.09.006

we need to know the covariances of the disturbances, Q_w, R_v from which we can calculate the optimal estimator's error covariance and the optimal Kalman filter gain. In most industrial process control applications, however, the covariances of the disturbances entering the process are not known. To address this requirement, estimation of the covariances from open-loop data has long been a subject in the field of adaptive filtering, and can be divided into four general categories: Bayesian (Alspach, 1974; Hilborn & Lainiotis, 1969), maximum likelihood (Bohlin, 1976; Kashyap, 1970), covariance matching (Myers & Tapley, 1976), and correlation techniques. Bayesian and maximum likelihood methods have fallen out of favor because of their sometimes excessive computation times. They may be well suited to a multi-model approach as in Averbuch, Itzikowitz, and Kapon (1991). Covariance matching is the computation of the covariances from the residuals of the state estimation problem. Covariance matching techniques have been shown to give biased estimates of the true covariances. The fourth category is correlation techniques, largely pioneered by Mehra (1970, 1972), Bélanger (1974) and Carew and Bélanger (1973), which we consider further in this paper.

2. Innovations based correlation techniques and the ALS estimator

With the standard linear state estimator, the state estimation error, $\varepsilon_k = x_k - \widehat{x}_{k|k-1}$, evolves according to

$$\varepsilon_{k+1} = \underbrace{(A - ALC)}_{\bar{A}}\varepsilon_k + \underbrace{[G, \quad -AL]}_{\bar{G}}\underbrace{\begin{bmatrix} w_k \\ v_k \end{bmatrix}}_{\bar{w}_k}. \tag{1}$$

We define the state-space model of the L-innovations as

$$\varepsilon_{k+1} = \bar{A}\varepsilon_k + \bar{G}\bar{w}_k,$$
$$\mathscr{Y}_k = C\varepsilon_k + v_k$$

in which

$$\mathscr{Y}_k = y_k - C\widehat{x}_{k|k-1}$$

and we require subsequently that the system is detectable and the chosen filter is stable.

Assumption 1. (A, C) is detectable.

Assumption 2. $\bar{A} = A - ALC$ is stable.

A stable filter gain L exists because of Assumption 1. In this formulation, the state and sensor noises are correlated:

$$E[\bar{w}_k(\bar{w}_k)^{\mathrm{T}}] \equiv \bar{Q}_w = \begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix},$$
$$E[\bar{w}_k v_k^{\mathrm{T}}] = \begin{bmatrix} 0 \\ R_v \end{bmatrix}.$$

Effect of initial condition: Assume the initial estimate error is distributed with mean m_0 and covariance P_0^-,

$$E(\varepsilon_0) = m_0, \quad \operatorname{cov}(\varepsilon_0) = P_0^-.$$

Propagating the estimate error through the state evolution equation gives an explicit formula for the mean

$$E(\varepsilon_k) = \bar{A}^k m_0$$

and the recursion for the covariance

$$\operatorname{cov}(\varepsilon_k) = P_k^-,$$
$$P_j^- = \bar{A}P_{j-1}^-\bar{A}^{\mathrm{T}} + \bar{G}\bar{Q}_w\bar{G}^{\mathrm{T}}, \quad j = 1, \ldots, k.$$

Because the filter is stable (Assumption 2), as k increases, the mean converges to zero and the covariance approaches a steady state given by the solution to the following Lyapunov equation

$$E(\varepsilon_k) \to 0,$$
$$\operatorname{cov}(\varepsilon_k) \to P^-,$$
$$P^- = \bar{A}P^-\bar{A}^{\mathrm{T}} + \bar{G}\bar{Q}_w\bar{G}^{\mathrm{T}}. \tag{2}$$

We therefore assume that we have chosen k sufficiently large so that the effects of the initial conditions can be neglected, or, equivalently, we choose the steady-state distribution as the initial condition.

Assumption 3. $E(\varepsilon_0) = 0$, $\operatorname{cov}(\varepsilon_0) = P^-$.

Now consider the autocovariance, defined as the expectation of the data with some lagged version of itself (Jenkins & Watts, 1968),

$$\mathscr{C}_j = E[\mathscr{Y}_k\mathscr{Y}_{k+j}^{\mathrm{T}}]. \tag{3}$$

Using Eq. (1) and the steady-state initial condition (Assumption 3) gives for the autocovariance

$$E(\mathscr{Y}_k\mathscr{Y}_k^{\mathrm{T}}) = CP^-C^{\mathrm{T}} + R_v, \tag{4}$$
$$E(\mathscr{Y}_{k+j}\mathscr{Y}_k^{\mathrm{T}}) = C\bar{A}^jP^-C^{\mathrm{T}} - C\bar{A}^{j-1}ALR_v, \quad j \geqslant 1 \tag{5}$$

which are independent of k because of our assumption about the initial conditions. The autocovariance matrix (ACM) is then defined as

$$\mathscr{R}(N) = \begin{bmatrix} \mathscr{C}_0 & \cdots & \mathscr{C}_{N-1} \\ \vdots & \ddots & \vdots \\ \mathscr{C}_{N-1}^{\mathrm{T}} & \cdots & \mathscr{C}_0 \end{bmatrix}. \tag{6}$$

The number of lags used in the ACM is a user-defined parameter, N. The off-diagonal autocovariances are not assumed zero, because we do not process the data with the optimal filter, which is unknown. The ACM of the L-innovations can be written as follows:

$$\mathscr{R}(N) = \mathscr{O}P^-\mathscr{O}^{\mathrm{T}} + \Gamma\left[\bigoplus_{i=1}^{N}\bar{G}\bar{Q}_w\bar{G}^{\mathrm{T}}\right]\Gamma^{\mathrm{T}}$$
$$+ \Psi\left[\bigoplus_{i=1}^{N} R_v\right] + \left[\bigoplus_{i=1}^{N} R_v\right]\Psi^{\mathrm{T}} + \bigoplus_{i=1}^{N} R_v \tag{7}$$

in which

$$\mathscr{O} = \begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{N-1} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} 0 & 0 & 0 & 0 \\ C & 0 & 0 & 0 \\ \vdots & \ddots & & \vdots \\ C\bar{A}^{N-2} & \cdots & C & 0 \end{bmatrix},$$

$$\Psi = \Gamma \left[\bigoplus_{j=1}^{N} (-AL) \right].$$

In this result and those to follow, we employ the standard definitions of the Kronecker product, Kronecker sum and the direct sum (Brewer, 1978; Searle, 1982). In order to use the ACM relationship in a standard least-squares problem, we apply the "vec" operator, which is the columnwise stacking of a matrix into a vector (Brewer, 1978). If z_k is the kth column of an arbitrary Z matrix

$$\text{vec}(Z) = Z_s = [z_1^{\mathrm{T}} \ \cdots \ z_k^{\mathrm{T}}]^{\mathrm{T}}.$$

Throughout this paper, we use the s subscript to denote the outcome of applying the vec operator. Applying the vec operator to Eq. (7) and using the result of applying the vec operator on Eq. (2)

$$P_s^- = (\bar{A} \otimes \bar{A}) P_s^- + (\bar{G} \bar{Q}_w \bar{G}^{\mathrm{T}})_s \tag{8}$$

yields

$$\begin{aligned} [\mathscr{R}(N)]_s = {} & [(\mathscr{O} \otimes \mathscr{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1} \\ & + (\Gamma \otimes \Gamma)\mathscr{I}_{n,N}](G \otimes G)(Q_w)_s \\ & + \{[(\mathscr{O} \otimes \mathscr{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1} \\ & + (\Gamma \otimes \Gamma)\mathscr{I}_{n,N}](AL \otimes AL) \\ & + [\Psi \oplus \Psi + I_{p^2 N^2}]\mathscr{I}_{p,N}\}(R_v)_s \end{aligned} \tag{9}$$

in which $\mathscr{I}_{p,N}$ is a permutation matrix to convert the direct sum to a vector, i.e. $\mathscr{I}_{p,N}$ is the $(pN)^2 \times p^2$ matrix of zeros and ones satisfying

$$\left(\bigoplus_{i=1}^{N} R_v \right)_s = \mathscr{I}_{p,N} (R_v)_s.$$

Ideally, we would like to compute the autocovariance as the expectation of the product $\mathscr{Y}_k \mathscr{Y}_{k+j}^{\mathrm{T}}$. Practically, we approximate the expectation from the data using the time average, a valid procedure since the process is ergodic (Jenkins & Watts, 1968). The estimate of the autocovariance is computed as

$$\widehat{\mathscr{C}}_j = \frac{1}{N_d - j} \sum_{i=1}^{N_d - j} \mathscr{Y}_i \mathscr{Y}_{i+j}^{\mathrm{T}} \tag{10}$$

which is the so-called unbiased autocovariance estimator. The estimated ACM, $\widehat{\mathscr{R}}(N)$, is analogously defined using the computed $\widehat{\mathscr{C}}_j$. At this point we can define a least-squares problem to estimate Q_w, R_v. We summarize Eq. (9) as

$$\mathscr{A} x = b$$

in which

$$\mathscr{A} = \left[D(G \otimes G) \,\middle|\, \begin{matrix} D(AL \otimes AL) \\ +[\Psi \oplus \Psi + I_{p^2 N^2}]\mathscr{I}_{p,N} \end{matrix} \right], \tag{11}$$

$$D = [(\mathscr{O} \otimes \mathscr{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1} + (\Gamma \otimes \Gamma)\mathscr{I}_{n,N}],$$

$$x = [(Q_w)_s^{\mathrm{T}} \ (R_v)_s^{\mathrm{T}}]^{\mathrm{T}}, \quad b = \mathscr{R}(N)_s.$$

We define the ALS estimate as

$$\widehat{x} = \arg\min_x \|\mathscr{A} x - \widehat{b}\|_2^2 \tag{12}$$

in which

$$\widehat{x} = [(\widehat{Q}_w)_s^{\mathrm{T}} \ (\widehat{R}_v)_s^{\mathrm{T}}]^{\mathrm{T}}, \quad \widehat{b} = \widehat{\mathscr{R}}(N)_s.$$

The solution for the ALS estimate is the well known

$$\widehat{x} = \mathscr{A}^{\dagger} \widehat{b}, \quad \mathscr{A}^{\dagger} = (\mathscr{A}^{\mathrm{T}} \mathscr{A})^{-1} \mathscr{A}^{\mathrm{T}}.$$

The uniqueness of the estimate is a standard result of least-squares estimation (Lawson & Hanson, 1995). The estimated covariances are symmetric due to the structure of the least-squares problem.

Lemma 4. *The ALS estimate* (*Eq.* (12)) *exists and is unique if and only if $\mathscr{A}$ has full column rank.*

We note that the ACM could be written in the output form as well, in which the ACM is computed from the outputs instead of the L-innovations. There are number of reasons why the output covariance estimator is inadequate in practical applications. In the output-based formulation, the control law would have to be specified in order to consider closed-loop data. Additionally, output autocovariance methods are not suitable for estimating integrated white noise disturbances, which are used widely in industrial MPC implementations to remove steady offset.

Comments on the initial filter gain: In principle, any stable filter gain, L, may be used to calculate the L-innovations. This initial gain simply parameterizes the L-innovations. The covariances of the underlying noise sequences are contained in the *outputs* of the process. While the choice of the initial filter may impact the number of data points required to find reliable estimates of the covariances, we show in Section 3 that the initial choice of the filter is irrelevant for large data sets. The autocovariance matrix in Eq. (6) has nonzero off-diagonal elements for a suboptimal choice of L. Only when the true covariances (and optimal filter) are employed are the off-diagonal terms zero.

3. Properties of the ALS covariance estimates

In this section we evaluate the mean and variance of the ALS estimator. To this end, first we require the properties of the estimated autocovariance.

Lemma 5. *The expectation of the estimated autocovariance* ($\widehat{\mathscr{C}}_j$) *is equal to the autocovariance* ($\mathscr{C}_j$) *for all* j, *and the* *variance goes to zero inversely with sample size,* N_d

$$E[\widehat{\mathscr{C}}_j] = \mathscr{C}_j, \quad j = 0, \ldots, N,$$
$$\mathrm{cov}(\widehat{\mathscr{C}}_j) = O\left(\frac{1}{N_d - j}\right).$$

Proof. The expectation result follows from taking expectation of Eq. (10) and the definition of the autocovariance, Eq. (3). For brevity, the proof of the variance result is omitted. A derivation can be found in Bartlett (1946). □

Remark 1. The unbiased result in finite sample size is due to the strong assumption we have made on the initial conditions, Assumption 3. If we weaken this assumption and allow nonzero expectation of initial error or covariance of initial error not equal to P^-, then the bias is nonzero with finite sample size, but decreases exponentially to zero with increasing sample size.

Note that for the types of problems to be solved with this method, we choose $N_d \gg N$, and therefore $\mathrm{cov}(\widehat{\mathscr{C}}_j) \to 0$ as $N_d \to \infty$, for all j. We can choose large N_d because we require only routine operating data, not identification testing data with input excitation. The properties of the autocovariance estimate then imply the ALS estimates of the covariances are unbiased for all sample sizes, and converge to the true values with increasing sample size.

Theorem 6. *Given* $\mathscr{A}$ *(Eq. (11)) has full column rank, the ALS noise covariance estimates* $(\widehat{Q}_w, \widehat{R}_v)$ *(Eq. (12)) are unbiased for all sample sizes and converge asymptotically to the true covariances* (Q_w, R_v) *as* $N_d \to \infty$.

Proof. For compactness, we use the notation of the least-squares problem of Eq. (12) in which $\mathscr{A}x = b$, and

$$\widehat{x} = \arg\min_x \|\mathscr{A}x - \widehat{b}\|_2^2.$$

The expected value of the estimate is

$$\begin{aligned} E[\widehat{x}] &= \mathscr{A}^\dagger E[\widehat{b}] \\ &= \mathscr{A}^\dagger b \quad \text{(by Lemma 5)} \\ &= \mathscr{A}^\dagger \mathscr{A} x = x. \end{aligned}$$

The covariance of the estimate is

$$\mathrm{cov}(\widehat{x}) = \mathscr{A}^\dagger \mathrm{cov}(\widehat{b})(\mathscr{A}^\dagger)^{\mathrm{T}}.$$

From Lemma 5, $\mathrm{cov}(\widehat{b}) \to 0$ as $N_d \to \infty$. Therefore

$$\mathrm{cov}(\widehat{x}) \to 0 \quad \text{as } N_d \to \infty. \qquad \square$$

Remark 2. Again, as in Lemma 5, the unbiased result in finite sample size is due to Assumption 3. If we remove this assumption, the bias is nonzero with finite sample size, but decreases exponentially to zero with increasing sample size.

4. Discussion and comparison to previous approaches

4.1. Comparison to correlation based methods

The pioneering work of Mehra (1970, 1972) and Carew and Bélanger (1973) has seen successful application using open-loop data and remains highly cited. Mehra employs a three-step procedure to estimate (Q_w, R_v): (i) Solve a least squares problem to estimate P^-C^{T} from the estimated autocovariances using Eqs. (4) and (5). (ii) Use Eq. (4) and the estimated P^-C^{T} to solve for R_v. (iii) Solve a least-squares problem to estimate Q_w from the estimated P^-C^{T} and R_v using Eq. (8). We offer two criticisms of the classic Mehra approach. Our first comment concerns the conditions for uniqueness of $(\widehat{Q}_w, \widehat{R}_v)$ in Mehra's approach. These conditions were stated (without proof) as

(1) (A, C) observable.
(2) A full rank.
(3) The number of unknown elements in the Q_w matrix, $g(g+1)/2$, is less than or equal to np.

These conditions were also cited by Bélanger (1974). As a counterexample, consider

$$A = \begin{bmatrix} 0.9 & 0 & 0 \\ 1 & 0.9 & 0 \\ 0 & 0 & 0.9 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad G = I.$$

The Mehra conditions predict that unique covariances exist, but the $\mathscr{A}$ matrix in Eq. (11) does not have full column rank for this case. Thus these conditions are not sufficient. The problem here is that although P^-C^{T} and R_v are uniquely estimatable from the data, Q_w is not. Examining the null space of the stacked version of the P^-C^{T} equation shows that any multiple of the following matrix can be added to an estimate of $\widehat{Q}_w$ without changing the fit to the autocovariance data

$$Q = \begin{bmatrix} 0.117 & -0.552 & 0 \\ -0.552 & -0.613 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

Consider a second counterexample,

$$A = \begin{bmatrix} 0.1 & 0 \\ 0 & 0.2 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \quad C = [1 \;\; \delta].$$

When $\delta = 0$, this system is not observable, and thus does not meet Mehra's conditions. But $\mathscr{A}$ has full column rank for $\delta = 0$, the ALS method estimates unique covariances, and thus Mehra's conditions are also not necessary. In this example, one can use just state x_1 to uniquely distinguish the process disturbance from the output disturbance. It makes no difference whether or not the second state is observable.

Our second comment concerns the large variance associated with Mehra's method. This point was first made by Neethling and Young (1974), and seems to have been largely overlooked. First, step (ii) above is inappropriate because the zero-order lag autocorrelation estimate in Eq. (4) is not known perfectly. Second, breaking a single-stage estimation of Q_w and R_v into two stages by first finding P^-C^{T} and R_v and then using these

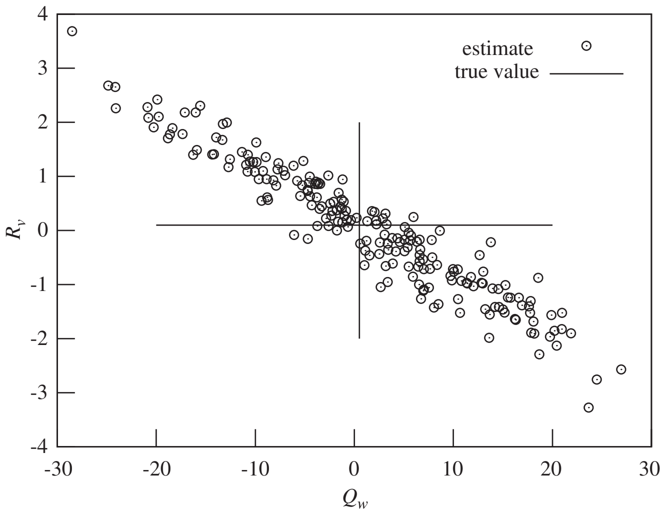

Fig. 1. Estimates of Q_w and R_v using Mehra's method.

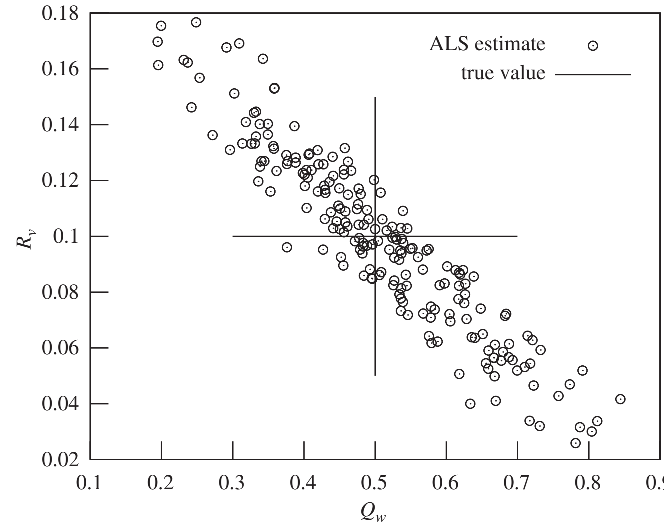

Fig. 2. Estimates of Q_w and R_v using proposed ALS method. Notice the axes have been greatly expanded compared to Fig. 1.

estimates to estimate Q_w in steps (i) and (iii) also increases the variance in the estimated Q_w.

To quantify the size of the variance inflation associated with Mehra's method, consider a third example, which has a *well-conditioned* observability matrix

$$A = \begin{bmatrix} 0.1 & 0 & 0.1 \\ 0 & 0.2 & 0 \\ 0 & 0 & 0.3 \end{bmatrix}, \quad G = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}, \quad C = [0.1 \;\; 0.2 \;\; 0].$$

Data are generated using noise sequences with covariance $Q_w = 0.5$, $R_v = 0.1$. The L-innovations are calculated with a filter gain corresponding to incorrect noise variances $Q_w = 0.2$ and $R_v = 0.4$. Mehra's method and the ALS method are run using $N_d = 1000$ data points, $N = 15$. The simulation is repeated 200 times to illustrate the mean and variances of the estimators. In Fig. 1, the estimates of (Q_w, R_v) using Mehra's method are plotted. The variance of the estimates is large, and many of the estimates are negative, which is unphysical. In Fig. 2, the ALS estimates of (Q_w, R_v) are plotted, on much tighter axes. The variance of the ALS estimates is much smaller than in Mehra's method, and none of the estimates are negative. Note that Neethling and Young (1974) discuss other examples with behavior similar to this one.

4.2. Enforcing semidefinite constraints

When dealing with a small sample of measurements or significant plant/model error, the ALS estimate of the covariances from Eq. (12) may not be positive semidefinite, even though the variance of the estimate may be smaller than the two-step procedure. Such estimates are physically meaningless. Most of the literature for estimating covariances does not address this issue. A recent ad hoc method of imposing positive semidefiniteness on the estimates of only R_v is given in Noriega and Pasupathy (1997).

Adding the semi-definite constraint directly to the estimation problem gives a constrained ALS estimation problem

$$\Phi = \min_{Q_w, R_v} \left\| \mathscr{A} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} - \widehat{b} \right\|_2^2 \quad \text{s.t.} \quad Q_w \geqslant 0, \quad R_v \geqslant 0. \tag{13}$$

The constraints in Eq. (13) are convex in Q_w, R_v and the optimization is in the form of a semidefinite programming (SDP) problem (Vandenberghe & Boyd, 1996). The matrix inequalities $Q_w \geqslant 0$, $R_v \geqslant 0$ can then be handled by adding a logarithmic barrier function to the objective. The optimization in Eq. (13) becomes:

$$\Phi = \min_{Q_w, R_v} \left\| \mathscr{A} \begin{bmatrix} (Q_w)_s \\ (R_v)_s \end{bmatrix} - \widehat{b} \right\|_2^2 - \mu \log \begin{vmatrix} Q_w & 0 \\ 0 & R_v \end{vmatrix} \tag{14}$$

in which, μ is the *barrier parameter* and $|\cdot|$ denotes the determinant of the matrix (Nocedal & Wright, 1999). The optimization in Eq. (14) is convex and the gradient can be evaluated analytically. A simple path following algorithm based on Newton steps provides a simple and efficient method to find the global optimum. The details of this type of algorithm can be found in Wolkowicz, Saigal, and Vandenberghe (2000, Chapter 10). The convexity of the optimization in Eq. (13) and Lemma 4 ensure uniqueness of the covariance estimate. The algorithm generalizes efficiently for large dimensional problems.

5. Conclusions

In this paper we have developed a new method (ALS) for using the autocovariance of the L-innovations to estimate the covariances of the system disturbances, which are required for optimal state estimation. We have shown that the ALS covariance estimates are unbiased and converge asymptotically to the true system covariances with increasing sample size. Using the standard properties of least-squares estimation, necessary and sufficient conditions for unique covariance estimates have been provided. Examples are provided to show that conditions previously stated in the literature are neither necessary nor sufficient for uniqueness of these estimates. Previously reported methods for this estimation have been shown to have unnecessarily large variance. The approach also guarantees semidefinite covariance estimates by solving a convex semidefinite programming problem.

Acknowledgments

Financial support came from the industrial members of the Texas-Wisconsin Modeling and Control Consortium and NSF through Grant #CTS-0105360. The authors thank the anonymous reviewers for many helpful comments, and Professors Stephen P. Boyd and John F. MacGregor for helpful discussions.

References

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

Averbuch, A., Itzikowitz, S., & Kapon, T. (1991). Radar target tracking—viterbi versus IMM. *IEEE Transactions on Aerospace and Electronic Systems*, *27*(3), 550–563.

Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society Supplement*, *8*, 27–41.

Bélanger, P. R. (1974). Estimation of noise covariance matrices for a linear time-varying stochastic process. *Automatica*, *10*, 267–275.

Bohlin, T. (1976). Four cases of identification of changing systems. In R. K. Mehra, & D. G. Lainiotis (Eds.), *System identification*: *Advances and case studies* (1st ed.). New York: Academic Press.

Brewer, J. (1978). Kronecker products and matrix calculus in system theory. *IEEE Transactions on Circuits and Systems*, *25*(9), 772–781.

Carew, B., & Bélanger, P. R. (1973). Identification of optimum filter steady-state gain for systems with unknown noise covariances. *IEEE Transactions on Automatic Control*, *18*(6), 582–587.

Hilborn, C. G., & Lainiotis, D. G. (1969). Optimal estimation in the presence of unknown parameters. *IEEE Transactions on Systems, Science, and Cybernetics*, *5*(1), 38–43.

Jenkins, G. M., & Watts, D. G. (1968). *Spectral analysis and its applications*. 500 Sansome Street, San Fransisco, CA: Holden-Day.

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

Lawson, C. L., & Hanson, R. J. (1995). *Solving least squares problems*. Philadelphia, PA: SIAM.

Mehra, R. K. (1970). On the identification of variances and adaptive Kalman filtering. *IEEE Transactions on Automatic Control*, *15*(12), 175–184.

Mehra, R. K. (1972). Approaches to adaptive filtering. *IEEE Transactions on Automatic Control*, *17*, 903–908.

Myers, K. A., & Tapley, B. D. (1976). Adaptive sequential estimation with unknown noise statistics. *IEEE Transactions on Automatic Control*, *21*, 520–523.

Neethling, C., & Young, P. (1974). Comments on "Identification of optimum filter steady-state gain for systems with unknown noise covariances". *IEEE Transactions on Automatic Control*, *19*(5), 623–625.

Nocedal, J., & Wright, S. J. (1999). *Numerical optimization*. New York: Springer.

Noriega, G., & Pasupathy, S. (1997). Adaptive estimation of noise covariance matrices in real-time preprocessing of geophysical data. *IEEE Transactions on Geoscience Remote Sensing*, *35*(5), 1146–1159.

Searle, S. R. (1982). *Matrix algebra useful for statistics*. New York: Wiley.

Vandenberghe, L., & Boyd, S. (1996). Semidefinite programming. *SIAM Review*, *38*(1), 49–95.

Wolkowicz, H., Saigal, R., & Vandenberghe, L. (Eds.). (2000). *Handbook of semidefinite programming*: *Theory*, *algorithms and applications*. MA: Kluwer Academic Publishers.

Brian J. Odelson received his B.S. from Purdue University in 1998 and his Ph.D. from the University of Wisconsin in 2003, both in Chemical Engineering. His Ph.D. research focused on improved state estimation in the presence of unknown noise, model mismatch, and various disturbance models. In 2003, he moved to Naperville, Illinois as a Senior Research Engineer for BP Research and Technology.

Murali R. Rajamani was born in Coimbatore, India, in 1980. He received his Bachelors in Chemical Engineering (B. Chem. Engg.) degree with a gold medal from University of Mumbai, Institute of Chemical Technology, Mumbai, India in 2002. Since then, he has been working towards a Ph.D. degree in the Department of Chemical and Biological Engineering, University of Wisconsin–Madison, USA. His research interests include model predictive control, state estimation, estimation of noise statistics, nonlinear control and applications of semidefinite optimization in control and estimation.

James B. Rawlings received his B.S. from the University of Texas in 1979 and his Ph.D. from the University of Wisconsin in 1985, both in Chemical Engineering. He spent one year at the University of Stuttgart as a NATO postdoctoral fellow and then joined the faculty at the University of Texas. He moved to the University of Wisconsin in 1995 and is currently the Paul A. Elfers Professor of Chemical and Biological Engineering and the co-director of the Texas-Wisconsin Modeling and Control Consortium (TWMCC).

His research interests are in the areas of chemical process modeling, monitoring and control, nonlinear model predictive control, moving horizon state estimation, particulate systems modeling, and crystal engineering.