Automatica 45 (2009) 142–148

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

Estimation of the disturbance structure from data using semidefinite programming and optimal weighting☆

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

[a] BP Research and Technology, 150 W. Warrenville Road, Naperville, IL 60563, United States

[b] Department of Chemical and Biological Engineering, University of Wisconsin-Madison, 1415 Engineering Drive, Madison, WI 53706, United States

ARTICLE INFO

Article history:
Received 29 June 2007
Received in revised form
16 October 2007
Accepted 30 May 2008
Available online 10 December 2008

Keywords:
State estimation
Covariance estimation
Minimum independent disturbances
Minimum variance estimation
Semidefinite programming

ABSTRACT

Designing a state estimator for a linear state-space model requires knowledge of the characteristics of the disturbances entering the states and the measurements. In [Odelson, B. J., Rajamani, M. R., & Rawlings, J. B. (2006). A new autocovariance least squares method for estimating noise covariances. *Automatica*, *42*(2), 303–308], the correlations between the innovations data were used to form a least-squares problem to determine the covariances for the disturbances. In this paper we present new and simpler necessary and sufficient conditions for the uniqueness of the covariance estimates. We also formulate the optimal weighting to be used in the least-squares objective in the covariance estimation problem to ensure minimum variance in the estimates. A modification to the above technique is then presented to estimate the number of independent stochastic disturbances affecting the states. This minimum number of disturbances is usually unknown and must be determined from data. A semidefinite optimization problem is solved to estimate the number of independent disturbances entering the system and their covariances.

1. Introduction

We start with the linear time-invariant state-space model in discrete time:

$$x_{k+1} = Ax_k + Bu_k + Gw_k \tag{1a}$$

$$y_k = Cx_k + v_k \tag{1b}$$

in which $x_k \in \mathbb{R}^n$, $u_k \in \mathbb{R}^m$, $y_k \in \mathbb{R}^p$ are the state, input and measurement of the system at time t_k. The dimensions of the system matrices are $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $G \in \mathbb{R}^{n\times g}$ and $C \in \mathbb{R}^{p\times n}$. The noises corrupting the state and the measurement ($w_k \in \mathbb{R}^g$ and $v_k \in \mathbb{R}^p$) are modelled as zero-mean Gaussian noise sequences with covariances Q_w and R_v respectively. The noises w_k and v_k are assumed to be uncorrelated with each other for simplicity. The case where w_k and v_k are correlated can be handled as shown in Åkesson, Jørgensen, Poulsen, and Jørgensen (2008). The state estimation for the model given in Eqs. (1a) and (1b) when there are no constraints on the input and the state is given by the classical Kalman filter (Anderson & Moore, 1979).

If complete knowledge about the deterministic part of the model i.e. A, B, C is assumed, then the state estimator requires knowledge of the stochastic part of the model i.e. G, Q_w, R_v. In Odelson, Rajamani, and Rawlings (2006), a correlation based method for estimating the covariances Q_w, R_v was presented. The correlation based techniques were largely pioneered by Mehra (1970) and adapted by many others (Carew & Bélanger, 1973; Neethling & Young, 1974; Noriega & Pasupathy, 1997). All of these techniques assume that the G matrix is known. In the absence of any knowledge about G we use the model

$$x_{k+1} = Ax_k + Bu_k + \tilde{w}_k \tag{2a}$$

$$y_k = Cx_k + v_k \tag{2b}$$

in which $\tilde{w}$ is modelled as zero-mean Gaussian noise sequences with covariances Q. The two noise models are related by

$$Q = GQ_wG^{\mathrm{T}}.$$

For physical reasons, it is likely that the number of statistically independent process disturbances is less than n, however. In this paper we provide a method to estimate the number of independent process disturbances affecting the state. We also derive the formula for the minimum variance estimate of the covariances in addition to simple necessary and sufficient conditions for their uniqueness.

The rest of the paper is organised as follows: In Section 2 we give some mathematical preliminaries and present the single

☆ This paper was not presented at any IFAC meeting. This paper was recommended for publication in revised form by Associate Editor Tongwen Chen under the direction of Editor Ian R. Petersen.

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

0005-1098/$ – see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.automatica.2008.05.032

column formulation of the Autocovariance Least-Squares (ALS) technique simplified from Odelson et al. (2006). The main contributions of this paper are then presented in Sections 3–5. Simple mathematical conditions to check for uniqueness of the covariance estimates are proved in Section 3. In Section 4, we formulate the optimal weighting matrix to calculate the linear unbiased minimum variance covariance estimates. In Section 5 we present a modification to handle the case when G is unknown. (A version of the ALS code used in Sections 2 and 3 is available for free download at http://jbrwww.che.wisc.edu/software/als/.)

2. The one-column version of the Autocovariance Least-Squares (ALS) technique

We assume that the pair (A, C) is observable in the rest of this paper. We use the notation $\hat{x}_k$ to denote any estimate of the state x_k. If $L \in \mathbb{R}^{n\times p}$ is any arbitrary, stable filter gain, then the state estimates are calculated recursively as:

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + AL(y_k - C\hat{x}_k). \tag{3}$$

We can write the evolution of the state estimate error $\varepsilon_k = x_k - \hat{x}_k$ by subtracting Eq. (3) from (1a) and substituting (1b):

$$\begin{aligned} \varepsilon_{k+1} &= \underbrace{(A - ALC)}_{\bar{A}}\varepsilon_k + \underbrace{\begin{bmatrix} G & -AL \end{bmatrix}}_{\bar{G}} \begin{bmatrix} w_k \\ v_k \end{bmatrix} \\ \mathscr{Y}_k &= C\varepsilon_k + v_k \end{aligned} \tag{4}$$

in which $\mathscr{Y}_k$ are the L-innovations defined as $\mathscr{Y}_k \triangleq y_k - C\hat{x}_k$. Note that the L-innovations are uncorrelated in time if the initial state estimator L is optimal (i.e. $L = L_o$) (Anderson & Moore, 1979). We use the term L-innovations to distinguish them from the optimal innovations obtained by using the optimal state estimates.

Assumption 1. The L-innovations data $\{\mathscr{Y}_1, \cdots \mathscr{Y}_{N_d}\}$ used in the techniques described in this paper are obtained after the system has reached steady state and any initial transience can be neglected when $\bar{A}$ is stable

Given a set of steady-state L-innovations data, we wish to form a weighted least-squares problem in the unknown disturbance covariances, Q_w, R_v, or Q, R_v depending on the choice of disturbance model. One of the motivations behind using a least-squares approach is to avoid a complicated nonlinear approach required for maximum likelihood estimation techniques Shumway and Stoffer (1982).

In the subspace ID literature (Gevers, 2006; Larimore, 1990; Van Overschee & De Moor, 1994), the identification procedures estimate the model and the disturbance covariances starting with the model in the innovations form, which is Eq. (3) rewritten as:

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + AL_o e_k \tag{5a}$$

$$y_k = C\hat{x}_k + e_k. \tag{5b}$$

Here e_k are the optimal innovations (as opposed to the L-innovations) and hence uncorrelated in time. The estimation of the system matrices $\hat{A}$, $\hat{B}$, $\hat{C}$ is carried out along with the Kalman filter gain $\hat{L}_o$.

Notice the difference between Eqs. (5a) and (5b) and Eqs. (1a) and (1b). If the subspace ID techniques are used to identify only the stochastic parameters from the above model, then the disturbance covariances are identified as $A\hat{L}_o H\hat{L}_o^{\mathrm{T}}A^{\mathrm{T}}$ instead of GQ_wG^{T} for the state noise and H instead of R_v for the measurements, where H is the covariance of e_k given by:

$$H = CP_oC^{\mathrm{T}} + R_v \tag{6}$$

and $P_o = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^{\mathrm{T}}]$ is calculated by solving the Riccati equation (Anderson & Moore, 1979).

Remark 2. As shown above, subspace ID techniques estimate a different set of covariances than G, Q_w, R_v. The aims of subspace ID are different and the estimates of the stochastic parameters are simply used to compute the optimal estimator gain. Finding the covariance parameters affecting the system (G, Q_w, R_v) on the other hand provides more flexibility in the choice of the state estimator. For example we may wish to employ a constrained, nonlinear moving horizon estimator (Rao, Rawlings, & Lee, 2001). In addition estimating G, Q_w, R_v gives a more informative handle in physical models to monitor the disturbances than monitoring changes in the optimal estimator gain.

Remark 3. A significant advantage of using the techniques presented in this paper over other identification techniques is the use of only steady state data in the calculations. This means that unlike other identification techniques there is no requirement for exciting inputs u_k to be applied to the system. The excitation is provided by the noises w_k, v_k, which are already present in the system.

Following the derivation along the lines of Odelson et al. (2006), we use Eq. (4) to write the following expectation of covariances:

$$E(\mathscr{Y}_k\mathscr{Y}_k^{\mathrm{T}}) = CPC^{\mathrm{T}} + R_v \tag{7}$$

$$E(\mathscr{Y}_{k+j}\mathscr{Y}_k^{\mathrm{T}}) = C\bar{A}^jPC^{\mathrm{T}} - C\bar{A}^{j-1}ALR_v \quad j \geq 1 \tag{8}$$

which are independent of k because of Assumption 1. Again using Eq. (4) we note that $P = E[(x_k - \hat{x}_k)(x_k - \hat{x}_k)^{\mathrm{T}}]$ satisfies the Lyapunov equation:

$$P = \bar{A}P\bar{A}^{\mathrm{T}} + \underbrace{\begin{bmatrix} G & -AL \end{bmatrix}}_{\bar{G}} \underbrace{\begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix}}_{\bar{Q}_w} \bar{G}^{\mathrm{T}}. \tag{9}$$

Using N to denote a user-defined number of lags and $\mathscr{R}_1(N)$ to denote the autocovariance matrix, Eqs. (7)–(9) give:

$$\begin{aligned} \mathscr{R}_1(N) &= E\begin{bmatrix} \mathscr{Y}_k\mathscr{Y}_k^{\mathrm{T}} \\ \vdots \\ \mathscr{Y}_{k+N-1}\mathscr{Y}_k^{\mathrm{T}} \end{bmatrix} \\ &= \underbrace{\begin{bmatrix} C \\ C\bar{A} \\ \vdots \\ C\bar{A}^{N-1} \end{bmatrix}}_{\mathscr{O}} PC^{\mathrm{T}} + \underbrace{\begin{bmatrix} I_p \\ -CAL \\ \vdots \\ -C\bar{A}^{N-2}AL \end{bmatrix}}_{\Gamma} R_v. \end{aligned} \tag{10}$$

In the results to follow, we employ the standard definitions and properties of the Kronecker product (Magnus & Neudecker, 1999). We use the 's' subscript to denote the column-wise stacking of a matrix into a vector.

Define $b \triangleq (\mathscr{R}_1(N))_s$. We then stack Eq. (10) and use the stacked form of Eq. (9) to substitute out P:

$$\begin{aligned} b = {} & [(C \otimes \mathscr{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}](GQ_wG^{\mathrm{T}})_s \\ & + [(C \otimes \mathscr{O})(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}(AL \otimes AL) + (I_p \otimes \Gamma)](R_v)_s. \end{aligned} \tag{11}$$

We next use the ergodic property of the L-innovations to estimate the autocovariance matrix $\mathscr{R}_1(N)$ from the given set of data. If $\{\mathscr{Y}_1, \mathscr{Y}_2, \cdots \mathscr{Y}_{N_d}\}$ are the set of L-innovations calculated from data as given by Eq. (4), and N is the window size used for the autocovariances then we define the matrix $\mathbb{Y}$ as follows:

$$\mathbb{Y} \triangleq \begin{bmatrix} \mathscr{Y}_1 & \mathscr{Y}_2 & \cdots & \mathscr{Y}_{N_d-N+1} \\ \mathscr{Y}_2 & \mathscr{Y}_3 & \cdots & \mathscr{Y}_{N_d-N+2} \\ \vdots & \vdots & \vdots & \vdots \\ \mathscr{Y}_N & \mathscr{Y}_{N+1} & \vdots & \mathscr{Y}_{N_d}. \end{bmatrix} \tag{12}$$

$\mathbb{Y} \in \mathbb{R}^{\tilde{p}\times\tilde{n}}$ where, $\tilde{n} \triangleq N_d - N + 1$ and $\tilde{p} \triangleq Np$. The estimate $\widehat{\mathscr{R}_1}(N)$ from data is then given by:

$$\widehat{\mathscr{R}_1}(N) = \frac{1}{N_d - N + 1}\mathbb{Y}\mathbb{Y}_1^{\mathrm{T}} \tag{13}$$

and $\hat{b} = (\widehat{\mathscr{R}_1}(N))_s$. Here, $\mathbb{Y}_1$ is the first row block of $\mathbb{Y}$ given by: $\mathbb{Y}_1 = \underbrace{[I_p\ 0\ \cdots\ 0]}_{\mathbb{E}}\mathbb{Y}$. Given the linear relation in Eq. (11) and the estimate $\hat{b}$ from Eq. (13), we can formulate the following positive definite constrained least-squares problem in the symmetric elements of the covariances Q, R_v. We define the least squares objective function

$$\Phi(Q, R_v) = \left\| \mathscr{A}\begin{bmatrix}(Q)_{ss}\\(R_v)_{ss}\end{bmatrix} - \hat{b}\right\|_W^2.$$

The following optimisation provides the estimates

$$\begin{aligned}&\min_{Q,R_v} \Phi(Q, R_v)\\ &\text{subject to,}\quad Q, R_v \geq 0.\end{aligned} \tag{14}$$

Here we introduce the notation of $(R_v)_{ss}$ to denote the column-wise stacking of only the symmetric $p(p+1)/2$ elements of the matrix R_v (eliminating the supra-diagonal elements). More explicitly there exists a unique matrix $\mathscr{D}_p \in \mathbb{R}^{p^2\times\frac{p(p+1)}{2}}$ called the *duplication matrix* (Magnus & Neudecker, 1999, p. 49) containing ones and zeros that gives the relation $(R_v)_s = \mathscr{D}_p(R_v)_{ss}$.

Using Eq. (11), we can then write $\mathscr{A}$ explicitly as:

$$\begin{aligned}\mathscr{A} &= [\mathscr{A}_1 \quad \mathscr{A}_2]\\ \mathscr{A}_1 &= [(C\otimes\mathscr{O})(I_{n^2} - \bar{A}\otimes\bar{A})^{-1}]\mathscr{D}_n\\ \mathscr{A}_2 &= [(C\otimes\mathscr{O})(I_{n^2} - \bar{A}\otimes\bar{A})^{-1}(AL\otimes AL) + (I_p\otimes\Gamma)]\mathscr{D}_p\end{aligned} \tag{15}$$

where, the *duplication matrices* $\mathscr{D}_n, \mathscr{D}_p$ are included to ensure symmetry in the covariance estimates. The estimation method in Eq. (14) is referred to as the Autocovariance Least-Squares (ALS) technique in the sequel.

3. Conditions for uniqueness

In this section, we assume that the $G \in \mathbb{R}^{n\times g}$ matrix is a known matrix. This case arises when the model (A, C, G) comes from a physical model. For physical models we often know G or know the locations of the zero elements in G based on the locations of process disturbances in the system. Without loss of generality we can also assume G to be full column rank. If G is not full column rank then a new matrix $\tilde{G}$ can be defined with its columns independent and such that $\tilde{G}\tilde{G}^{\mathrm{T}} = GG^{\mathrm{T}}$.

We next derive simple uniqueness conditions for the ALS covariance estimates with a known G. In the rest of this section we also assume that the weight is $W = I$. The objective function for the known G case is

$$\tilde{\Phi}(Q_w, R_v) = \left\| \tilde{\mathscr{A}}\begin{bmatrix}(Q_w)_{ss}\\(R_v)_{ss}\end{bmatrix} - \hat{b}\right\|^2$$

and the estimate is provided by solving

$$\begin{aligned}&\min_{Q_w,R_v} \tilde{\Phi}(Q_w, R_v)\\ &\text{subject to,}\qquad Q_w, R_v \geq 0\end{aligned} \tag{16}$$

where, $\tilde{\mathscr{A}} = \left[(C\otimes\mathscr{O})(I_{n^2} - \bar{A}\otimes\bar{A})^{-1}(G\otimes G)\mathscr{D}_g \quad \mathscr{A}_2\right]$.

Lemma 4. *The solution of the optimisation in Eq. (16) exists for all $\hat{b}$. The solution is unique if and only if $\tilde{\mathscr{A}}$ has full column rank.*

Proof. Since $\hat{b}$ is a finite value calculated from data, given the bound $B = \tilde{\Phi}(I_g, I_p)$ from Eq. (16), the inequality $\{Q, R|\tilde{\Phi}(Q, R) \leq B\}$ is a bounded and closed, hence compact, set (Haaser & Sullivan, 1991, Heine-Borel Theorem, p. 85). The set $\{Q, R|Q \geq 0, R \geq 0, Q = Q^{\mathrm{T}}, R = R^{\mathrm{T}}\}$ is the set of symmetric positive semidefinite matrices and hence is a closed set (Haaser, 2000, p. 69). The intersection of a compact set and a closed set is compact (Haaser & Sullivan, 1991, p. 80). The minimisation of a continuous function ($\tilde{\Phi}$) on a compact set achieves its minimum on the set (Weierstrass Theorem). Therefore, the solution to Eq. (16) exists for any set of finite measurements.

To prove uniqueness, we see that $\tilde{\mathscr{A}}$ having full column rank guarantees the objective function in Eq. (16) to be strictly convex. The positive semidefinite constraints in Eq. (16) are also convex (Vandenberghe & Boyd, 1996). The set $\{Q, R|\tilde{\Phi}(Q, R) \leq B\}$ defines a convex set, which when intersected with the semidefiniteness requirement still defines a convex set. Uniqueness then follows from the existence of a bounded solution for all $\hat{b}$ and the fact that a strictly convex objective function is subject to convex constraints (Boyd & Vandenberghe, 2004, p. 137). □

Definition 5. The noise covariances (Q_w, R_v) are defined to be ALS-estimatable if the solution to Eq. (16) is unique.

Assumption 6. We assume that the discrete-time state transition matrix A is non-singular. If the original A is singular, then a similarity transformation can be used to eliminate the states with zero eigenvalues and the noise covariances redefined.

Lemma 7. *If (A, C) is observable and A is non-singular, then the matrix $\tilde{\mathscr{A}}$ in Eq. (16) has a null space if and only if the matrix M defined by, $M = (C\otimes I_n)(I_{n^2} - \bar{A}\otimes\bar{A})^{-1}(G\otimes G)\mathscr{D}_g$ also has a null space, and the null space of $(C\otimes\mathscr{O})(I_{n^2} - \bar{A}\otimes\bar{A})^{-1}(G\otimes G)\mathscr{D}_g$ which multiplies $(Q_w)_{ss}$ in Eq. (16) is equal to the null space of M.*

The derivation is given in Appendix A.

Theorem 8. *If (A, C) is observable and A is non-singular, the covariances (Q_w, R_v) are ALS-estimatable if and only if M has full column rank, where:*

$$M = (C\otimes I_n)(I_{n^2} - \bar{A}\otimes\bar{A})^{-1}(G\otimes G)\mathscr{D}_g.$$

Proof. The proof follows from Definition 5 and Lemmas 4, 7. □

4. Minimum variance and optimal weighting

It is well known that for a linear model, the weighted least-squares estimator gives the minimum variance among the class of all linear unbiased estimators (Aitken, 1935). The weighted least-squares estimation of the covariances is given by the ALS technique as shown by Eq. (14). If $W = I$ is used as in Odelson et al. (2006), the minimum variance property for the estimates does not hold. If $\hat{b}$ is an unbiased estimator of b and $S \triangleq E[(\hat{b}-b)(\hat{b}-b)^{\mathrm{T}}] = \mathrm{cov}(\hat{b})$, then $W = S^{-1}$ is the weighting that gives minimum variance for the ALS problem. It is shown in Odelson et al. (2006) that $\hat{b}$ in Eq. (13) is an unbiased estimator. We derive in this section the formula for the minimum variance weighting matrix W to be used in the ALS objective.

Lemma 9. *Given the L-innovations from Eq. (4) and the definition of $\mathbb{Y}$ from Eq. (12), we have*

$$E[\mathbb{Y}] = 0 \quad \textit{and} \quad E[\mathbb{Y}\mathbb{Y}^{\mathrm{T}}] \triangleq E[\mathbb{Y}_s\mathbb{Y}_s^{\mathrm{T}}] = \Omega$$

with Ω as defined in Appendix B (Eq. (B.4)). The random matrix $\mathbb{Y}$ is distributed normally with $\mathbb{Y} \sim N(0, \Omega)$.

Proof of Lemma 9 is given in Appendix B.

Theorem 10. *The minimum variance weight to use in the ALS objective in Eq. (14) is given by $W = S^{-1}$, where,*

$$S = \frac{T(I_{\tilde{n}^2\tilde{p}^2} + K_{(\tilde{n}\tilde{p})(\tilde{n}\tilde{p})})((K_{\tilde{p}\tilde{n}}\Omega K_{\tilde{n}\tilde{p}}) \otimes (K_{\tilde{p}\tilde{n}}\Omega K_{\tilde{n}\tilde{p}}))T^{\mathrm{T}}}{(N_d - N + 1)^2} \tag{17}$$

and K_{ij} is the commutation matrix defined in Magnus and Neudecker (1999). T is defined as:

$$T = (\mathbb{E} \otimes I_p)(I_{\tilde{p}^2} \otimes (I_{\tilde{n}})_s)^{\mathrm{T}}(I_{\tilde{p}} \otimes K_{\tilde{p}\tilde{n}} \otimes I_{\tilde{n}})$$

and $\mathbb{E} = [I_p, 0 \cdots 0]$. $\tilde{n} = N_d - N + 1$ and $\tilde{p} = Np$.

Proof. Since $\mathbb{Y} \in \mathbb{R}^{\tilde{p}\times\tilde{n}}$ is a matrix as defined in Eq. (12) which is normally distributed with mean 0 and covariance Ω as defined in Lemma 9, the fourth moment of $\mathbb{Y}$ is given by: $\mathrm{cov}\,[\mathbb{Y}\mathbb{Y}^{\mathrm{T}}] \triangleq \mathrm{cov}((\mathbb{Y}\mathbb{Y}^{\mathrm{T}})_s)$.

The formula for the fourth moment of $\mathbb{Y}$ i.e $\mathrm{cov}(\mathbb{Y}\mathbb{Y}^{\mathrm{T}})$ for a normal distribution is given by:

$$\mathrm{cov}\,(\mathbb{Y}\mathbb{Y}^{\mathrm{T}}) = T_1(I_{\tilde{n}^2\tilde{p}^2} + K_{(\tilde{n}\tilde{p})(\tilde{n}\tilde{p})})((K_{\tilde{p}\tilde{n}}\Omega K_{\tilde{n}\tilde{p}}) \otimes (K_{\tilde{p}\tilde{n}}\Omega K_{\tilde{n}\tilde{p}}))T_1^{\mathrm{T}} \tag{18}$$

where, $T_1 = (I_{\tilde{p}^2} \otimes (I_{\tilde{n}})_s)^{\mathrm{T}}(I_{\tilde{p}} \otimes K_{\tilde{p}\tilde{n}} \otimes I_{\tilde{n}})$. The formula follows from the results in Ghazal and Neudecker (2000). See Ghazal and Neudecker (2000) for more details on the derivation.

We also have from Eq. (13):

$$\hat{b} = (\mathbb{Y}\mathbb{Y}^{\mathrm{T}}\mathbb{E}^{\mathrm{T}})_s = (\mathbb{E} \otimes I_p)(\mathbb{Y}\mathbb{Y}^{\mathrm{T}})_s.$$

From Eq. (18) we can then calculate S as:

$$S = \mathrm{cov}(\hat{b}) = \frac{(\mathbb{E} \otimes I_p)\mathrm{cov}(\mathbb{Y}\mathbb{Y}^{\mathrm{T}})(\mathbb{E}^{\mathrm{T}} \otimes I_p)}{(N_d - N + 1)^2}.$$

Thus we get Eq. (17) as the covariance of $\hat{b}$ where $T = (\mathbb{E} \otimes I_p)T_1$.

The optimal weight is then $W = S^{-1}$ following Aitken (1935). If S is singular, then without loss of generality we can take the Moore–Penrose pseudoinverse of S. □

The weight W is a complicated function depending on the values of the unknown covariances. A recursive calculation may be carried out for estimating W and the covariances (see Remark 12).

(1) Guess values for $\hat{Q}, \hat{R}_v$, where $Q = GQ_wG^{\mathrm{T}}$ and calculate Ω, $W = S^{-1}$ using Eqs. (17) and (B.4)
(2) Use the estimated weight in the ALS technique to estimate $\hat{Q}, \hat{R}_v$ using Eq. (14)
(3) Use estimates in previous step to recalculate W and iterate until convergence

Convergence of the above recursive scheme is not guaranteed, especially for large data sets.

Remark 11. If the initial estimator gain L was optimal, the L-innovations would be white and S (Eq. (17)) would be given as the second moment of the Wishart distribution (Anderson, 2003, p. 255). White innovations would also imply optimality of the filter and there would be no need to calculate the covariances. In the more practical situation when the L-innovations are not white, the assumption of 'whiteness' would lead to the incorrect weighting that is used in Dee, Cohn, Dalcher, and Ghil (1985).

Remark 12. The computation of S from Eq. (17) becomes prohibitively large even for a small dimensional problem with large data sets. This is a drawback for any practical application until efficient means for the computation are found. Although the weight may be estimated from data, a large data set is required before getting reliable estimates for the weight.

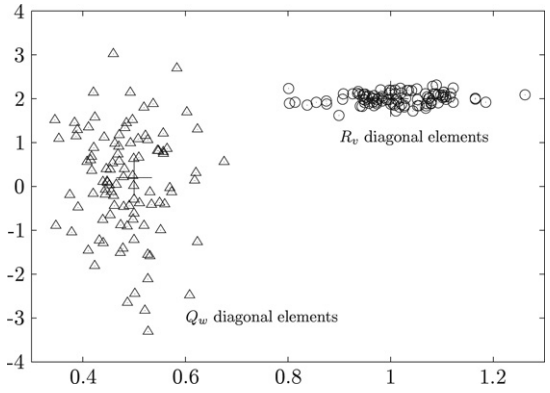


Fig. 1. Covariance estimates using $W = I$ in ALS.

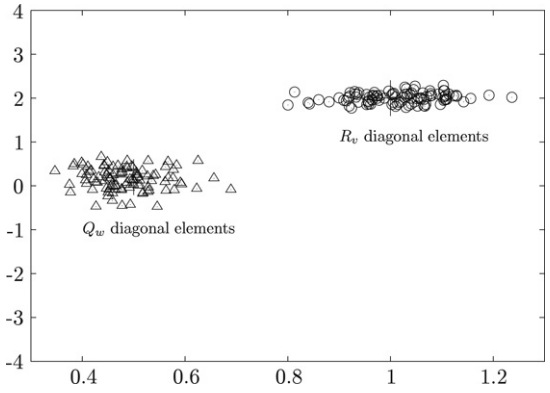


Fig. 2. Covariance estimates using the minimum variance weight in ALS.

Example of lower variance

Consider the following model for the system:

$$x_{k+1} = \begin{bmatrix} 0.732 & -0.086 \\ 0.172 & 0.990 \end{bmatrix} x_k + \begin{bmatrix} 1 & 0 \\ 0 & 0.2 \end{bmatrix} w_k$$

$$y_k = x_k + v_k.$$

Data is generated by drawing the noises from the following distributions:

$$w_k \sim N\,(0, \mathrm{diag}(0.5, 0.2))\,, \qquad v_k \sim N\,(0, \mathrm{diag}(1, 2))\,.$$

The ALS estimation of the covariances Q_w, R_v for a set of data of length 50 and simulated using $W = I$ and using the minimum variance weight (iterative scheme) from the above section is compared. The covariance estimation is repeated 100 times with a ALS window size of $N = 15$ and the results are plotted to check for the variance in the estimates. The diagonal elements of the estimated Q_w, R_v are plotted. As seen in Figs. 1 and 2 using the optimal weight gives estimates having much lower variance than using $W = I$.

5. ALS-SDP method

In this section the G matrix is assumed to be unknown and we use Eqs. (2a) and (2b) to describe the system. This case usually arises when the model comes from empirical identification methods based purely on (u, y) data. Our aim is to estimate the minimum rank Q and its decomposition $Q = FF^{\mathrm{T}}$, in which F has independent columns. The matrix F can be computed from the SVD of Q

$$Q = \begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} \Sigma_1 & \\ & 0 \end{bmatrix} \begin{bmatrix} U_1^{\mathrm{T}} \\ U_2^{\mathrm{T}} \end{bmatrix}$$

in which $U_1 \in \mathbb{R}^{n\times r}$, $\Sigma_1 \in \mathbb{R}^{r\times r}$ and r is the rank of Q. The matrix F is computed from

$$F = U_1\sqrt{\Sigma_1} \qquad FF^{\mathrm{T}} = U_1\Sigma_1 U_1^{\mathrm{T}} = Q.$$

This decomposition is unique within an orthogonal matrix multiplication. Thus by estimating F with fewest columns (Q with minimum rank), we obtain the number of statistically independent disturbances entering the state. We next present a method to estimate the number of independent process disturbances.

In the development of the remaining results in this section, we take the weight as $W = I$. If the rank is explicitly added to the objective in Eq. (14) using a tradeoff parameter ρ, the modified objective becomes

$$\Phi + \rho\mathrm{Rank}(Q). \tag{19}$$

Since the rank can take only integer values, minimisation of the above objective becomes a NP hard problem (Vandenberghe & Boyd, 1996). However, since the rank is equal to the number of nonzero eigenvalues of a matrix, a good heuristic substitute for the rank is the sum of its eigenvalues or the trace of the matrix. The trace of a matrix is also the largest convex envelope over the rank of the matrix (Fazel, 2002). The modified objective with the trace replacing the rank is

$$\Psi(Q, R_v) = \Phi(Q, R_v) + \rho\mathrm{Tr}\,(Q)$$

and the estimation problem is

$$\begin{aligned} &\min_{Q,R_v} \Psi(Q, R_v) \\ &\text{subject to,} \quad Q, R_v \geq 0. \end{aligned} \tag{20}$$

Eq. (20) can be easily rewritten in the form of a standard primal Semidefinite Programming (SDP) problem (Rajamani & Rawlings, 2007). We refer to Eq. (20) as the ALS-SDP (Autocovariance Least-Squares with Semidefinite Programming) in the sequel.

Lemma 13. *If $p < n$ (i.e. number of measurements is fewer than the number of states), then the following holds for $\mathscr{A}$ defined in Eq. (15):*

$$\dim[\mathrm{Null}(\mathscr{A})] \geq (n-p)(n-p+1)/2.$$

Proof. The dimension condition follows by substituting $G = I$ in Lemma 7, noting that $(I_{n^2} - \bar{A}\otimes\bar{A})$ is full rank and using the rank condition in Hua (1990). □

Theorem 14. *A solution $(\hat{Q}, \hat{R}_v)$ to the ALS-SDP in Eq. (20) is unique if dim $[Null(M)] = 0$ where,*

$$M = (C\otimes I_n)(I_{n^2} - \bar{A}\otimes\bar{A})^{-1}(\hat{F}\otimes\hat{F})\mathscr{D}_g.$$

Proof. The function Φ in Eq. (20) is identical to the objective in Eq. (14). Following Theorem 8 and Lemma 4, $\dim[\mathrm{Null}(M)] = 0$ then implies that Φ is strictly convex at the solution $\hat{Q}, \hat{R}_v$. Moreover, Φ is convex for all Q, R_v.

The second part of the objective in Eq. (20) i.e. $\mathrm{Tr}\,(Q)$ is linear in the variable Q and hence is also convex. The overall objective Ψ in Eq. (20) is then strictly convex at $\hat{Q}, \hat{R}_v$ and also convex on the convex set $Q, R_v \geq 0$. Using simple arguments of strict convexity and contradiction it can be proved that the solution $\hat{Q}, \hat{R}_v$ is also the unique minimiser of Ψ. □

The ALS-SDP optimisation is solved for each value of the tradeoff parameter ρ and the choice of ρ is made from a tradeoff plot of $\mathrm{Tr}\,(Q)$ versus Φ from Eq. (20). The value of ρ is chosen from the tradeoff plot such that the rank of Q is minimised without significant compromise on the original objective Φ. Efficient algorithms for solving the ALS-SDP can be found in Haaser (2000, chap. 10).

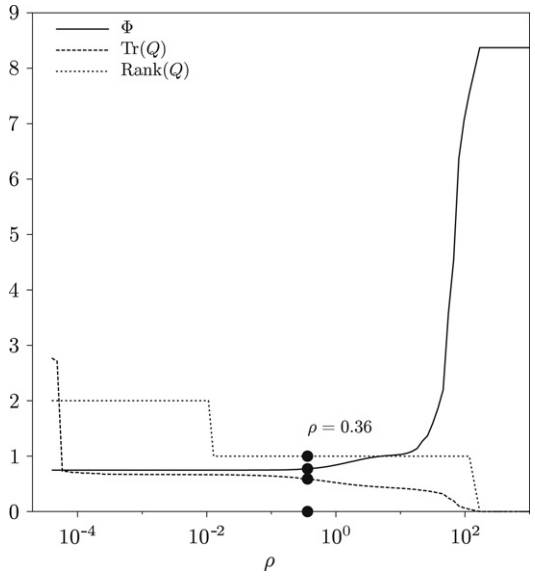


Fig. 3. Values of competing parts of the objective function in Eq. (20) and the rank of Q for different values of ρ.

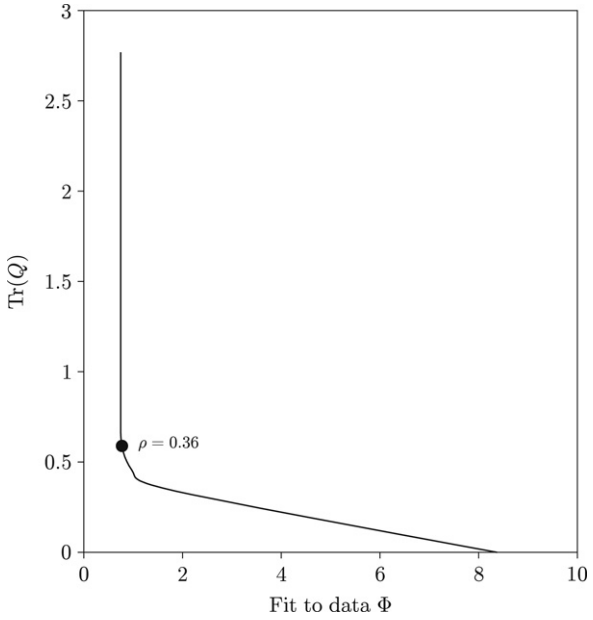


Fig. 4. Tradeoff plot between Φ and $Tr(Q)$ from Eq. (20) to choose the tradeoff parameter ρ.

Example of the ALS-SDP technique

Let the plant be simulated using the following state-space matrices.

$$A = \begin{bmatrix} 0.733 & -0.086 \\ 0.172 & 0.991 \end{bmatrix} \qquad C = \begin{bmatrix} 1 & 2 \end{bmatrix} \qquad G = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}$$

with noises drawn from the distributions: $w_k \sim N(0, 0.5)$ and $v_k \sim N(0, 1)$. Although the data is generated by a single column G matrix, we assume G is unknown and use the ALS-SDP procedure with $N = 15$ and $N_d = 5000$. The length of the data used for this example makes an iterative calculation of W from Section 4 infeasible. We use $W = I$ in this example. The results from the ALS-SDP are shown in Figs. 3 and 4. The plots show that choice of $\rho = 0.36$ is where the $\mathrm{Tr}\,(Q)$ is the minimum with no significant change in Φ. Also, the at $\rho = 0.36$ $\mathrm{rank}(Q) = 1$, which is the rank of the GQ_wG^{T} matrix in the simulated data. The estimated

covariances using $\rho = 0.36$ are:

$$\hat{Q} = \begin{bmatrix} 0.445 & 0.253 \\ 0.253 & 0.144 \end{bmatrix} \qquad \hat{R}_v = 1.06$$

in good agreement with

$$GQ_wG^{\mathrm{T}} = \begin{bmatrix} 0.5 & 0.25 \\ 0.25 & 0.125. \end{bmatrix} \qquad R = 1.$$

The decomposition of $\hat{Q} = \hat{F}\hat{F}^{\mathrm{T}}$ gives the one-column vector $\hat{F} = [0.67\ \ 0.38]^{\mathrm{T}}$, which is in good agreement with $G\sqrt{Q_w}$ used in the model.

6. Conclusions

Given a set of system matrices A, C, a known noise shaping matrix G and an initial arbitrary stable filter gain L, uniqueness of the estimates of Q_w and R_v using the ALS technique can be checked using the simple conditions in Theorem 8. The computational burden in checking these conditions is minimal even for large dimension systems. Estimates of the noise covariances from data are minimum variance only if the least-squares is weighted with the optimal weight. This weight was shown to depend on the fourth moment of data and a formula was derived (Theorem 10). An example was presented to show the reduced variance in the covariance estimates when using the minimum variance weight. The complicated nature of the formulae do not make them practical using current computational techniques. The weight however puts to rest the issue of the existence of the best linear unbiased estimator for the covariances.

One of the major uncertainties in the process industries is choice of the G matrix and the number of independent disturbances affecting the significant variables of the plant. For the case of an unknown G, we developed and applied a method for estimating the minimal rank of Q (number of columns of G) required to represent the data. An estimation procedure using SDP and a rank heuristic was shown to give a tradeoff between fit to the data and the minimisation of the rank. The 'knee' of the tradeoff curve was shown to give good estimates for the minimum number of disturbances and their covariances.

Acknowledgments

The authors thank Prof. S.P. Boyd for helpful discussions and acknowledge financial support through NSF grant #CNS–0540147 and PRF grant #43321–AC9.

Appendix A. Proof of Lemma 7

Let $[q_N, r_N]^{\mathrm{T}}$ be an element in the null space of $\tilde{\mathscr{A}}$ in Eq. (16), where the dimensions are $q_N \in \mathbb{R}^{\frac{g(g+1)}{2}}$ and $r_N \in \mathbb{R}^{\frac{p(p+1)}{2}}$. This implies:

$$\tilde{\mathscr{A}}\begin{bmatrix} q_N \\ r_N \end{bmatrix} = 0 \quad \text{or,} \quad \begin{bmatrix} (C \otimes \mathscr{O})A^{\dagger}(G \otimes G)\mathscr{D}_g & \mathscr{A}_2 \end{bmatrix}\begin{bmatrix} q_N \\ r_N \end{bmatrix} = 0$$

where, $\mathscr{A}_2$ is defined in Eq. (15) and $A^{\dagger} = (I_{n^2} - \bar{A} \otimes \bar{A})^{-1}$. We then have:

$$(C \otimes \mathscr{O})A^{\dagger}(G \otimes G)\mathscr{D}_g q_N \ + [(C \otimes \mathscr{O})A^{\dagger}(AL \otimes AL) + (I_p \otimes \Gamma)]\mathscr{D}_p r_N = 0. \tag{A.1}$$

We can rewrite $\mathscr{O}$ and Γ as:

$$\mathscr{O} = \begin{bmatrix} C \\ \mathscr{O}_1\bar{A} \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} I_p \\ \mathscr{O}_1(-AL) \end{bmatrix} \quad \text{where, } \mathscr{O}_1 = \begin{bmatrix} C \\ \vdots \\ C\bar{A}^{N-2} \end{bmatrix}.$$

Since (A, C) is observable, $\mathscr{O}_1$ has full column rank for $N \geq (n+1)$. Partitioning $\mathscr{O}$ and Γ as above, we can write Eq. (A.1) as the following Equations:

$$(C \otimes C)A^{\dagger}[(G \otimes G)\mathscr{D}_g q_N + (AL \otimes AL)\mathscr{D}_p r_N] + \mathscr{D}_p r_N = 0 \tag{A.2}$$

$$(C \otimes \mathscr{O}_1\bar{A})A^{\dagger}[(G \otimes G)\mathscr{D}_g q_N + (AL \otimes AL)\mathscr{D}_p r_N] + (I_p \otimes \mathscr{O}_1(-AL))\mathscr{D}_p r_N = 0. \tag{A.3}$$

By expanding $\bar{A} = A - ALC$ and using Eq. (A.2), Eq. (A.3) simplifies to:

$$(I_p \otimes \mathscr{O}_1 A)(C \otimes I_n)A^{\dagger}[(G \otimes G)\mathscr{D}_g q_N + (AL \otimes AL)\mathscr{D}_p r_N] = 0.$$

Since $\mathscr{O}_1$ is full column rank and A is non singular (Assumption 6), $(I_p \otimes \mathscr{O}_1 A)$ is also full column rank. This implies:

$$(C \otimes I_n)A^{\dagger}[(G \otimes G)\mathscr{D}_g q_N + (AL \otimes AL)\mathscr{D}_p r_N] = 0. \tag{A.4}$$

Substituting Eq. (A.4) in A.2 and noting that $(C \otimes C)$ can be written as $(I_n \otimes C)(C \otimes I_n)$, we get: $\mathscr{D}_p r_N = 0$ Eq. (A.4) then simplifies to:

$$(C \otimes I_n)A^{\dagger}(G \otimes G)\mathscr{D}_g q_N = 0.$$

Thus, q_N is an element in the null space of $M = (C \otimes I_n)(I_{n^2} - \bar{A} \otimes \bar{A})^{-1}(G \otimes G)\mathscr{D}_g$ and $r_N = 0$.

Proving the second part of the lemma is straightforward by starting with Eq. (A.4) and multiplying with $(I_n \otimes \mathscr{O})$, which is full column rank. □

Appendix B. Proof of Lemma 9

Let $Y_p = [\mathscr{Y}_1^{\mathrm{T}} \cdots \mathscr{Y}_{N_d}^{\mathrm{T}}]^{\mathrm{T}}$ and $\mathbb{E}_1$ be a permutation matrix containing 0's and 1's and giving the relation:

$$\mathbb{Y}_s = \mathbb{E}_1 Y_p. \tag{B.1}$$

From Eq. (4), we have

$$\varepsilon_k = \bar{A}^k\varepsilon_0 + \sum_{j=0}^{k-1}\bar{A}^{k-j-1}\bar{G}\begin{bmatrix} w_j \\ v_j \end{bmatrix}. \tag{B.2}$$

Taking the expectation of the above expression and noting that $E[v_k] = E[w_k] = 0$, we get,

$$E[\varepsilon_k] = \bar{A}^k E[\varepsilon_0] = 0.$$

The equality follows from Assumption 1 since for k large enough, we have $\bar{A}^k = (A - ALC)^k \approx 0$. Taking the expectation of the L-innovations in Eq. (4), we get:

$$E[\mathscr{Y}_j] = CE[\varepsilon_j] + E[v_k] = 0, \quad \forall j \geq k.$$

Thus, we have $E[Y_p] = 0$. The covariance of Y_p is then also its second moment. Since $\Omega_p = E[Y_pY_p^{\mathrm{T}}]$, using Eqs. (7)–(9), we get:

$$\Omega_p = \underbrace{\begin{bmatrix} C \\ \vdots \\ C\bar{A}^{N_d-1} \end{bmatrix}}_{\mathscr{O}} P\mathscr{O}^T + \begin{bmatrix} R_v & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_v \end{bmatrix} + \Psi\begin{bmatrix} R_v & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_v \end{bmatrix} + \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 \\ C\bar{G} & 0 & 0 & 0 \\ \vdots & \ddots & & \vdots \\ C\bar{A}^{N_d-2}\bar{G} & \cdots & C\bar{G} & 0 \end{bmatrix}}_{\Gamma_f} \times \begin{bmatrix} \bar{Q}_w & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \bar{Q}_w \end{bmatrix}\Gamma_f^T + \begin{bmatrix} R_v & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & R_v \end{bmatrix}\Psi^T \tag{B.3}$$

where, $\Psi = \Gamma_f \begin{bmatrix} -AL & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & -AL \end{bmatrix}$, $\bar{Q}_w = \begin{bmatrix} Q_w & 0 \\ 0 & R_v \end{bmatrix}$.

Following Eq. (B.2), we see that ε_k is a linear combination of normally distributed noises given $\bar{A}^k \approx 0$ and hence is normal. This implies $\mathscr{Y}_k$ is also normally distributed. We then have: $Y_p \sim N(0, \Omega_p)$.

Since $\mathbb{Y}$ is a matrix, its mean and covariance are defined for the stacked version of the matrix i.e. $\mathbb{Y}_s$. Given the linear relationship between Y_p and $\mathbb{Y}_s$, we get, $\mathbb{Y}_s \sim N(0, \mathbb{E}_1 \Omega_p \mathbb{E}_1^{\mathrm{T}})$. Thus, the covariance of $\mathbb{Y}$ is:

$$\Omega = \mathbb{E}_1 \Omega_p \mathbb{E}_1^{\mathrm{T}} \tag{B.4}$$

where Ω_p is given by Eq. (B.3) and $\mathbb{E}_1$ satisfies Eq. (B.1). □

References

Aitken, A. C. (1935). On least squares and linear combinations of observations. *Proceedings of The Royal Society Edingburgh*, *55*, 42–48.

Åkesson, B. M., Jørgensen, J. B., Poulsen, N. K., & Jørgensen, S. B. (2008). A generalized autocovariance least-squares method for Kalman filter tuning. *Journal of Process Control*, *18*(7–8), 769–779.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis* (3rd ed.). New York: John Wiley & Sons.

Anderson, B. D. O., & Moore, J. B. (1979). *Optimal filtering*. Englewood Cliffs, NJ: Prentice-Hall.

Boyd, S., & Vandenberghe, L. (2004). *Convex optimization*. Cambridge University Press.

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

Dee, D., Cohn, S., Dalcher, A., & Ghil, M. (1985). An efficient algorithm for estimating noise covariances in distributed systems. *IEEE Transactions on Automatic Control*, *30*(11), 1057–1065.

Fazel, M. (2002). Matrix rank minimization with applications. *Ph.D. thesis*. Dept. of Elec. Eng., Stanford University.

Gevers, M. (2006). A personal view of the development of system identification. *IEEE Control Systems Magazine*, *26*(6), 93–105.

Ghazal, G. A., & Neudecker, H. (2000). On second-order and fourth-order moments of jointly distributed random matrices: A survey. *Linear Algebra and its Applications*, *321*, 61–93.

Haaser, N. B., & Sullivan, J. A. (1991). *Real analysis*. New York: Dover Publications, Inc.

Hua, D. (1990). On the symmetric solutions of linear matrix equations. *Linear Algebra and its Applications*, *131*, 1–7.

Larimore, W. (1990). Canonical variate analysis in identification, filtering, and adaptive control. In *Proceedings of the 29th conference on decision and control* (pp. 596–604).

Magnus, J. R., & Neudecker, H. (1999). *Matrix differential calculus with applications in statistics and econometrics*. New York: John Wiley.

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

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.

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.

Odelson, B. J., Rajamani, M. R., & Rawlings, J. B. (2006). A new autocovariance least-squares method for estimating noise covariances. *Automatica*, *42*(2), 303–308.

Rajamani, M. R., & Rawlings, J. B. (2007). Estimation of the disturbance structure from data using semidefinite programming and optimal weighting. *Technical Report 2007–02, TWMCC*. Department of Chemical and Biological Engineering, University of Wisconsin-Madison (Available at: http://jbrwww.che.wisc.edu/tech-reports.html).

Rao, C. V., Rawlings, J. B., & Lee, J. H. (2001). Constrained linear state estimation — a moving horizon approach. *Automatica*, *37*(10), 1619–1628.

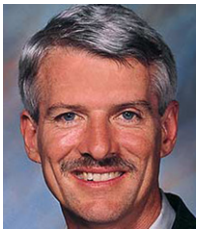
Shumway, R. H., & Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the em algorithm. *Journal of Time Series Analysis*, *3*, 253–264.

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

Van Overschee, P., & De Moor, B. (1994). N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, *30*(1), 75–93.

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

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 Mumbai University - Institute of Chemical Technology (formerly University Department of Chemical Technology, Mumbai University), India in 2002. In December 2007, he received his Ph.D. degree in Chemical Engineering at the University of Wisconsin-Madison, USA. He is currently working as a Research Engineer in the Advanced Control and Instrumentation group at BP Research and Technology in Naperville, Illinois. 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 the B.S. from the University of Texas in 1979 and the 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, molecular-scale chemical reaction engineering, monitoring and control, nonlinear model predictive control and moving horizon state estimation.