Optimal and Self-Tuning State Estimation for Singular Stochastic Systems: A Polynomial Equation Approach [1]

Huanshui Zhang   Lihua Xie   Yeng Chai Soh

BLK S2, School of Electrical and Electronic Engineering
Nanyang Technological University
Nanyang Avenue, Singapore 639798
Email: elhxie@ntu.edu.sg

Abstract

This paper is concerned with the optimal steady-state estimation for singular stochastic discrete-time systems using a polynomial equation approach. The key to the optimal estimation is to calculate an optimal estimator gain matrix. The main contribution of the paper is to present a simple method for computing the gain matrix. Our method involves solving one simple polynomial equation which is derived based on the uniqueness of the ARMA innovation model. The approach covers the prediction, filtering and smoothing problems. Further, when the noise statistics of model are not available, self-tuning estimation is performed by identifying one ARMA innovation model.

Keywords: Singular stochastic systems; optimal state estimation; innovation analysis; polynomial equation approach; self-tuning estimation.

1 Introduction

Singular systems can be used to describe non-causal phenomena which arise naturally in robotics, economic, electric and chemical systems [1-4]. The state estimation for linear singular systems has received much attention in the last few years ([4]). It is well known that the classic recursive filtering techniques such as the Kalman filtering and the Linear Least Squares method have been very useful for nonsingular systems. However, in singular systems the future dynamics can affect the present one, thus the problem of prediction and smoothing based on the classic filtering techniques is difficult and the calculation involved is complex ([4]).

Recently, the filtering, prediction and smoothing problems for linear singular systems have been addressed by employing a time-domain innovation method in [5, 7]. By transforming the singular system into an equivalent non-singular system with an output feedback, the main idea in [5] is to calculate the gain matrix directly by using projection formula. Note that the algorithm for computing the estimator in [5] is complicated as the estimator gain matrix is given in a complex way. In [7], the calculation for the gain matrix is simplified. Unfortunately, two matrices have to be found prior to the computation of an estimator, one is to find K_1 such that $det[M - (\Phi + K_1 H)q^{-1}] = \gamma q^{-n}$ for some scalar γ and the other is to select K_2 such that $M + K_2 H$ is nonsingular and $(M + K_2 H)^{-1}\Phi$ is stable. No general method has been known to compute the required K_1 and K_2, which may impose some difficulties when applying the algorithm.

In this paper, a new method for obtaining the gain matrix will be given which involves solving one simple polynomial equation. The polynomial equation is derived based on an innovation analysis method and the fact that a stable spectral factorization is unique. As compared to [5, 7], the present result greatly sim-

[1] This work has been supported by the Academic Research Funds of the Ministry of Education, Republic of Singapore.

plifies the calculation of estimator (filter, predictor or smoother). Further, when applied to normal systems, the result leads to a simple unified solution to the optimal filtering, smoothing and prediction.

It is also worth noting that the statistics of system and observation noises are assumed to be known in the existing work [4, 5, 7]. However, in many practical situations, the statistics of the noises are often unknown. In this paper we also consider a self-tuning (asymptotically optimal) filtering, smoothing and prediction for the singular systems.

The rest of the paper is organized as follows. The problem statement can be found in Section 2. The optimal and self-tuning estimators are derived in Section 3 and Section 4, respectively. An example is given in Section 5. Concluding remarks are made in Section 6.

2 Problem Statement

Consider a stochastic linear time-invariant system described by the following discrete-time model:

$$Mx(k+1) = \Phi x(k) + w(k) \quad (1)$$

$$y(k) = Hx(k) + v(k) \quad (2)$$

where $x(k) \in R^n, w(k) \in R^n, y(k) \in R^m, v(k) \in R^m$ represent the state, system stochastic noise, measurement output, and measurement noise, respectively. It is assumed that $w(k), v(k)$ are zero mean mutually uncorrelated white noises with $\mathcal{E}\left[w(k)w^T(j)\right] = Q_w\delta_{kj}$ and $\mathcal{E}\left[v(k)v^T(j)\right] = Q_v\delta_{kj}$, where $\mathcal{E}$ denotes the mathematical expectation, Q_w and Q_v are constant symmetric positive semi-definite matrices, δ_{kj} is the Kronecker delta and superscript T stands for the transpose. In this paper, the following assumptions will be made.

Assumption 2.1
The system (1) and (2) is observable, i.e.,

$$rank \begin{bmatrix} zM - \Phi \\ H \end{bmatrix} = n, \quad \forall z \in \mathcal{C}; \quad rank \begin{bmatrix} M \\ H \end{bmatrix} = n \quad (3)$$

where $\mathcal{C}$ is the set of complex numbers.

Assumption 2.2
The system (2.1) is regular, $det(zM - \Phi) \not\equiv 0$.

In this paper, we shall deal with the following problems:

Problem 1: Optimal state estimation

Find a steady-state estimator $\hat{x}(k+l \mid k)$ for $x(k+l)$ based on the observations $y(k), y(k-1), \cdots, y(0)$. Depending on the sign of l, the estimator will be a predictor ($l > 0$), a filter ($l = 0$), or a fixed lag smoother ($l < 0$).

Problem 2: Self-tuning estimation

When the noise statistics are unknown, i.e., Q_w and Q_v are unknown constant matrices, find an estimator which asymptotically converges with probability one to the optimal estimator. In this case, the estimator is called a *self-tuning estimator*.

3 Optimal State Estimation

3.1 Preliminaries

According to the standard decomposition of singular systems ([3]), under Assumption 2.1, there exist nonsingular matrices Q_1 and Q_2 such that

$$Q_1 M Q_2 = \begin{bmatrix} I_{n_1} & 0 \\ 0 & M_1 \end{bmatrix}, \quad Q_1 \Phi Q_2 = \begin{bmatrix} \Phi_1 & 0 \\ 0 & I_{n_2} \end{bmatrix} \quad (4)$$

where $n_1 + n_2 = n$, I_n is the $n \times n$ identity matrix, M_1 is a nilpotent matrix with index γ_0, i.e., $M_1^{\gamma_0} = 0$, $M_1^{\gamma_0 - 1} \neq 0$. By taking into consideration (4), $H(M - \Phi q^{-1})^{-1}$ can be expressed as

$$H(M - \Phi q^{-1})^{-1} = \frac{T(q^{-1})q^{\lambda_0}}{a(q^{-1})} \quad (5)$$

where

$$a(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{n_a} q^{-n_a} \quad (6)$$

$$T(q^{-1}) = T_0 + T_1 q^{-1} + \cdots + T_{n_t} q^{-n_t}, \; T_0 \neq 0 \quad (7)$$

and q^{-1} is the backward shift operator, i.e., $q^{-1}x(k) = x(k-1)$. Substituting (1) into (2) and noting (5) yields

$$a(q^{-1})y(k) = T(q^{-1})w(k + \lambda_0 - 1) + a(q^{-1})v(k) \quad (8)$$

It is obvious that $\gamma_0 \geq \lambda_0$.

Without loss of generality, $T(q^{-1})$ and $a(q^{-1})$ are assumed to have no common factors. Then the ARMA innovation model is obtained as

$$a(q^{-1})y(k) = D(q^{-1})\varepsilon(k) \quad (9)$$

and

$$D(q^{-1})\varepsilon(k) = T(q^{-1})w(k+\lambda_0-1) + a(q^{-1})v(k) \quad (10)$$

where the stable polynomial matrix

$$D(q^{-1}) = I_m + D_1q^{-1} + \cdots + D_{n_d}q^{-n_d} \quad (11)$$

with

$$n_d = max\{n_a, n_t\}$$

is a spectral factor and $\varepsilon(k)$ is a zero mean white noise sequence with covariance matrix

$$\mathcal{E}\left[\varepsilon(k)\varepsilon^T(j)\right] = Q_\varepsilon\delta_{kj}$$

satisfying

$$D(q^{-1})Q_\varepsilon D^T(q) = T(q^{-1})Q_wT^T(q) + a(q^{-1})Q_va(q) \quad (12)$$

By using the ARMA innovation model (9) and (10), the linear minimum mean square error variance estimators $\hat{w}(k+l \mid k)$, $\hat{v}(k+l \mid k)$ and $\hat{y}(k+l \mid k)$ of the white noises $w(k+l)$, $v(k+l)$ and the output $y(k+l)$ based on the measurements $\{y(k), y(k-1), \cdots, y(0)\}$ are given in the following Lemmas.

Lemma 3.1 The white noise estimators $\hat{w}(k+l \mid k)$ and $\hat{v}(k+l \mid k)$ are given by

$$\hat{w}(k+l \mid k) = \mathcal{F}_l(q^{-1})\varepsilon(k), \quad \hat{v}(k+l \mid k) = \mathcal{E}_l(q^{-1})\varepsilon(k) \quad (13)$$

where

$$\begin{aligned} \mathcal{F}_l(q^{-1}) &= \bar{F}_0 + \bar{F}_1q^{-1} + \cdots + \bar{F}_{\lambda_0-l-1}q^{-\lambda_0+l+1}, \\ \bar{F}_i &= Q_wF^T_{\lambda_0-i-l-1}Q_\varepsilon^{-1} \end{aligned} \quad (14)$$

for $l < \lambda_0$ and $\mathcal{F}_l(q^{-1}) = 0$ for $l \geq \lambda_0$, and

$$\begin{aligned} \mathcal{E}_l(q^{-1}) &= \bar{E}_0 + \bar{E}_1q^{-1} + \cdots + \bar{E}_{-l}q^l, \\ \bar{E}_i &= Q_vE^T_{-i-l}Q_\varepsilon^{-1} \end{aligned} \quad (15)$$

while F_i and E_i are calculated recursively as

$$F_i = -\sum_{j=1}^{i} D_jF_{i-j} + T_i, \qquad E_i = -\sum_{j=1}^{i} D_jE_{i-j} + a_iI_m \quad (16)$$

with $F_i = T_0$, $E_0 = I_m$ and $T_i = 0$, $a_i = 0$ and $D_i = 0$ for $i > n_t$, $i > n_a$ and $i > n_d$, respectively.

Proof It follows directly from [5].

Lemma 3.2

The optimal output predictor $\hat{y}(k+l \mid k)(l > 0)$ is of the form

$$\hat{y}(k+l \mid k) = a^{-1}(q^{-1})\mathcal{S}_l(q^{-1})\varepsilon(k) \quad (17)$$

where

$$\mathcal{S}_l(q^{-1}) = S_0 + S_1q^{-1} + \cdots + S_{n_s}q^{-n_s} \quad (18)$$

with

$$n_s = max\{n_d - l, n_a - 1\}$$

and the coefficients S_i, $i = 0, \cdots, n_s$ are calculated using

$$S_i = -\sum_{j=1}^{l-1} G_ja_{l+i-j} + D_{l+i} \quad (19)$$

with

$$G_i = -\sum_{j=1}^{i} G_ja_{i-j} + D_i, \; G_0 = I_m, \; i = 0, 1, \cdots, l-1 \quad (20)$$

3.2 Main Result

In this subsection, we shall present our main result of optimal estimation.

First, denote

$$K_e(l) = M\mathcal{E}\left[x(k+l)\varepsilon^T(k)\right]Q_\varepsilon^{-1} \quad (21)$$

By applying (1)-(2) and the projection formula, we have

$$\begin{aligned} M\hat{x}(k+l \mid k) &= \Phi\hat{x}(k+l-1 \mid k-1) \\ &\quad + \hat{w}(k+l-1 \mid k-1) + K_e(l)\varepsilon(k) \end{aligned} \quad (22)$$

$$\hat{y}(k+l \mid k) = H\hat{x}(k+l \mid k) + \hat{v}(k+l \mid k) \quad (23)$$

In order to obtain the estimator $\hat{x}(k+l \mid k)$ based on the innovation $\varepsilon(k), \varepsilon(k-1), \cdots, \varepsilon(0)$ from (22), the key problem is to calculate the gain matrix $K_e(l)$. This problem has been considered in [5] but the computation is very involved. [7] has presented a simplified solution. However, the solution requires obtaining some matrices for which there is no general solution method. In the following we shall present a simple but efficient algorithm for computing the gain matrix which only involves solving one simple polynomial equation.

Theorem 3.1 Consider the system (1) and (2) satisfying Assumptions 2.1 and 2.2. Then, the steady-state gain matrix $K_e(l)$ for filtering, smoothing or prediction is the unique solution to the following equation:

$$T(q^{-1})K_e(l) = R(q^{-1}) \tag{24}$$

where for the filtering or smoothing,

$$\begin{aligned} R(q^{-1}) &= D(q^{-1})q^{-\lambda_0+l} - T(q^{-1})\mathcal{F}_l(q^{-1})q^{-1} \\ &\quad -a(q^{-1})\mathcal{E}_l(q^{-1})q^{-\lambda_0} \end{aligned} \tag{25}$$

and for the prediction,

$$R(q^{-1}) = \mathcal{S}_l(q^{-1})q^{-\lambda_0} - T(q^{-1})\mathcal{F}_l(q^{-1})q^{-1} \tag{26}$$

As M is a singular matrix, the estimator $\hat{x}(k+l \mid k)$ cannot be calculated directly from (22) given the observations $y(k),\ y(k-1),\ \cdots, y(0)$. In the following, a simple method for computing $\hat{x}(k+l \mid k)$ is presented with the aid of output predictor. By combining (22) with (23) yields

$$\begin{aligned} \begin{bmatrix} M \\ H \end{bmatrix} \hat{x}(k+l \mid k) &= \begin{bmatrix} \Phi \\ 0 \end{bmatrix} \hat{x}(k+l-1 \mid k-1 \\ + \begin{bmatrix} I \\ 0 \end{bmatrix} \hat{w}(k+l-1 \mid k-1) &+ \begin{bmatrix} 0 \\ -I_m \end{bmatrix} \hat{v}(k+l \mid k) \\ + \begin{bmatrix} 0 \\ I_m \end{bmatrix} \hat{y}(k+l \mid k) &+ \begin{bmatrix} K_e(l) \\ 0 \end{bmatrix} \varepsilon(k)) \end{aligned} \tag{27}$$

Note that $\begin{bmatrix} M \\ H \end{bmatrix}$ is of full column rank, the steady state estimator is easily obtained from (27) as given in the theorem below.

Theorem 3.2 Consider the system (1) and (2) satisfying Assumptions 2.1 and 2.2. Then, the steady-state optimal estimator $\hat{x}(k+l \mid k)$ (filter, predictor and smoother) can be given recursively by

$$\hat{x}(k+l \mid k) = \tilde{M}\Phi\hat{x}(k+l-1 \mid k-1) + \mathcal{K}(q^{-1})\varepsilon(k) \tag{28}$$

where

$$\begin{aligned} \mathcal{K}(q^{-1}) &= \tilde{M}\mathcal{F}_l(q^{-1})q^{-1} - \tilde{H}\mathcal{E}_l(q^{-1}) + \tilde{M}K_e(l) + \\ &\quad \tilde{H}a^{-1}(q^{-1})\mathcal{S}_l(q^{-1}) \end{aligned} \tag{29}$$

with $\tilde{M} = (M^TM + H^TH)^{-1}M^T$ and $\tilde{H} = (M^TM + H^TH)^{-1}H^T$.

Proof The proof is straightforward from (27), Lemma 3.1 and Lemma 3.2.

Theorem 3.1 can be easily applied to normal systems, i.e., $M = I$. In this case, the nilpotent index $\gamma_0 = 0$ and the estimator for filtering, prediction and smoothing is readily obtained from (22) as stated in the following corollary.

Corollary 3.1 Consider the system (1)-(2) with $M = I$ and the pair (Φ, H) being observable. Then, the optimal estimator is given by

$$\begin{aligned} \hat{x}(k+l \mid k) &= \Phi\hat{x}(k+l-1 \mid k-1) + \\ &\quad [\mathcal{F}_l(q^{-1})q^{-1} + K_e(l)]\varepsilon(k) \end{aligned} \tag{30}$$

where the gain matrix $K_e(l)$ is the unique solution to the following equation:

$$T(q^{-1})K_e(l) = R(q^{-1}) \tag{31}$$

while for the filtering or smoothing

$$\begin{aligned} R(q^{-1}) &= D(q^{-1})q^l - a(q^{-1})\mathcal{E}_l(q^{-1}) \\ &\quad -T(q^{-1})\mathcal{F}_l(q^{-1})q^{-1} \end{aligned} \tag{32}$$

and for the prediction,

$$R(q^{-1}) = \mathcal{S}_l(q^{-1}) - T(q^{-1})\mathcal{F}_l(q^{-1})q^{-1} \tag{33}$$

4 Self-Tuning Estimation

We have seen in Section 3 that when the system model and the covariance matrices Q_w and Q_v are known, the optimal estimators can be obtained by Theorem 3.1.

In this section, our aim is to derive self-tuning estimators for the case when the covariance matrices Q_w and Q_v are unknown, while the other system matrices are known.

4.1 Self-Tuning Predictor

In this subsection we consider the self-tuning predictor $\hat{x}(k+l \mid k)$ with $l > \lambda_0 - 1$. Note that when $l > \lambda_0 - 1$, $\mathcal{F}_l(q^{-1}) = 0$ and $\mathcal{E}_l(q^{-1}) = 0$. Then, it follows from (26) that $R(q^{-1}) = \mathcal{S}_l q^{-\lambda_0}$. The optimal predictor, using (28), is given as

$$\begin{aligned}\hat{x}(k+l \mid k) &= \tilde{M}\Phi\hat{x}(k+l-1 \mid k-1) + \tilde{M}K_e(l)\varepsilon(k) \\ &\quad + a^{-1}(q^{-1})\tilde{H}\mathcal{S}_l(q^{-1})\varepsilon(k) \qquad (34)\end{aligned}$$

where $K_e(l)$ is the unique solution to the following equation:

$$T(q^{-1})K_e(l) = \mathcal{S}_l(q^{-1})q^{-\lambda_0} \qquad (35)$$

The self-tuning algorithm for predictor with $l > \lambda_0 - 1$ is follows:

Step 1: From (9), estimate the coefficient matrices $D_i(i = 1, \cdots, n_d)$ of the polynomial $D(q^{-1})$ using the RELS or EKF method ([6]).

Step 2: Obtain the estimate $\hat{\mathcal{S}}(q^{-1})$ of $\mathcal{S}_l(q^{-1})$ using (19) and the estimate $\hat{D}(q^{-1})$.

Step 3: Solve $K_e(l)$ from the polynomial equation (35).

Step 4: Compute $\hat{x}(k+l \mid k)$ by substituting the estimates of $K_e(l)$ and $\mathcal{S}_l(q^{-1})$ into (34).

4.2 Self-Tuning Filtering, Smoothing and Prediction

In this subsection, we consider a general case for self-tuning estimation with an arbitrary l, including smoothing, filtering and prediction. The self-tuning algorithm is based on the estimation of statistics of innovation and noises.

Denote $z(k) = a(q^{-1})y(k)$.

Step 1: Estimate the coefficient matrices D_i of the polynomial $D(q^{-1})$ using the RELS. The innovation $\varepsilon(k)$ can be computed by

$$\hat{\varepsilon}(k) = z(k) - \hat{D}_1\hat{\varepsilon}(k-1) - \cdots - \hat{D}_{n_d}\hat{\varepsilon}(k-n_d) \qquad (36)$$

and the covariance matrix Q_ε can be computed recursively as

$$\hat{Q}_\varepsilon(k+1) = \hat{Q}_\varepsilon(k) + \frac{1}{k+1}[\varepsilon(k)\varepsilon^T(k) - \hat{Q}_\varepsilon(k)] \qquad (37)$$

Step 2: Computing the correlation function of the MA processes of both sides of (10) yields the matrix equations

$$\sum_{j=i}^{n_d} D_j Q_\varepsilon D_{j-i}^T = \sum_{j=i}^{n_t} T_j Q_w T_{j-i}^T + \sum_{j=i}^{n_a} a_j Q_v a_{j-i} \qquad (38)$$

where $i = 0, 1, \cdots n_d$. Substituting the estimates $\hat{D}_i(k)$ and $\hat{Q}_\varepsilon(k)$ into (38), we can obtain the estimates $\hat{Q}_w(k)$ and $\hat{Q}_v(k)$ (see example for details).

Step 3: With the use of $\hat{Q}_w(k)$ and $\hat{Q}_v(k)$, the estimates $\hat{\mathcal{F}}_l(q^{-1})$ and $\hat{\mathcal{E}}_l(q^{-1})$ are easily obtained from Lemma 3.1. The $K_e(l)$ is calculated uniquely from (24).

Step 4: Compute $\hat{x}(k+l \mid k)$ by substituting $\hat{K}_e(l)$, $\hat{\mathcal{F}}_l(q^{-1})$, $\hat{\mathcal{E}}_l(q^{-1})$ into (28).

5 Example

Consider the singular discrete-time system described by (1)-(2) with

$$M = \begin{bmatrix} 1 & 0 & 0 \\ -1 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \quad \Phi = \begin{bmatrix} 0.9 & 0 & 0 \\ -0.4 & 0.8 & 0 \\ -0.5 & 0 & 0.5 \end{bmatrix}$$

$$H = [0 \quad 1 \quad 0]$$

where $w(k) = [1 \quad -1 \quad 0]^T w_0(k)$, $w_0(k)$ and $v(k)$ are zero-mean white noises $Q_{w_0} = 1$, $Q_v = 1$, respectively. The variances of Q_{w_0}, Q_v will be assumed unknown in the simulation.

First, $a(q^{-1})$, $T(q^{-1})$ and λ_0 in (8) are calculated as $a(q^{-1}) = 1 - 0.9q^{-1}$, $\lambda_0 = 2$ and

$$T(q^{-1}) =$$
$$[0.5q^{-1} \quad -1.25q^{-1}(1-0.9q^{-1}) \quad 2.5(-1+0.9q^{-1})],$$

By (12), the spectral factors $D(q^{-1})$ and Q_ε are given by

$$D(q^{-1}) = 1 + dq^{-1} = 1 - 0.6166q^{-1}, Q_\varepsilon = 2.7264$$

The predictor and filter are given by

$$\hat{x}(k+l \mid k) = \tilde{M}\Phi\hat{x}(k \mid k-1) + \mathcal{K}(q^{-1})\varepsilon(k) \qquad (39)$$

where

$$\tilde{M} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}$$

For the one-step ahead predictor, $l = 1$ and

$$\begin{aligned} \mathcal{K}(q^{-1}) &= 2(0.9+d)\begin{bmatrix} 1 & 0 & 1 \end{bmatrix}^T + \\ &\quad (0.9+d)\begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T (1-0.9q^{-1})^{-1} \end{aligned}$$

For the filter, $l = 0$ and

$$\begin{aligned} \mathcal{K}(q^{-1}) &= 1.25\begin{bmatrix} 1 & 0 & 0 \end{bmatrix}^T Q_{w_0}Q_\varepsilon^{-1}q^{-1} - \\ &\quad \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T Q_vQ_\varepsilon^{-1} + \\ &\quad 2\begin{bmatrix} 1 & 0 & 1 \end{bmatrix}(1 - 1.5625Q_{w_0}Q_\varepsilon^{-1} - Q_vQ_\varepsilon^{-1}) \\ &\quad + \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}^T (1-0.9q^{-1})^{-1}(1+dq^{-1}) \end{aligned}$$

As for the self-tuning filter, $Q_{w_0}Q_\varepsilon^{-1}$ and $Q_vQ_\varepsilon^{-1}$ are solved from the above equations:

$$\begin{aligned} Q_{w_0}Q_\varepsilon^{-1} &= 2.6177 + 5.2644d + 2.6177d^2 \\ Q_vQ_\varepsilon^{-1} &= -2.2721 - 5.6806 - 2.2721d^2 \end{aligned}$$

Note that in the self-tuning scheme, we estimate the parameter d rather than the variances Q_{w_0} and Q_v directly.

6 Conclusion

In this paper, we have presented a new and simple solution to the optimal filtering, prediction and smoothing problems for singular systems. The gain matrices for filter, predictor and smoother, which is the key to the optimal estimation, are computed directly from one simple polynomial equitation. As compared with the existing results, the calculation has been significantly simplified. The result has also been applied to normal systems, which gives a much simpler solution than the Kalman filtering formulation, especially for the case of smoothing as it does not require augmenting the system state. In the case when the statistics of white noises are unknown, the self-tuning (asymptotically optimal) state estimation has been performed by identifying one ARMA innovation model.

References

[1] S.L. Campell, *Singular systems of differential equations.* Pitman, London, 1980.

[2] F.L. Lewis, "A survey of linear singular systems," *Circuits, Syst. and Signal Processing,* vol. 5, pp. 3-36, 1986.

[3] J.D. Cobb, "Controllability, observability and duality in singular systems," *IEEE Trans. Automat. Control,* vol. 29, pp. 1076-1082, 1984.

[4] R. Nikoukhah, A.S. Willsky and B.C. Levy, "Kalman filtering and Riccati equations for descriptor systems," *IEEE Trans. Automat. Control*, vol. 37, pp. 1325-1342, 1992.

[5] H.S. Zhang, T.Y. Chai and X.J. Liu. "A unified approach to optimal estimation for discrete-time stochastic singular linear systems," *Automatica*, vol. 34, pp. 777-781, 1998.

[6] L. Ljung., *System identification: Theory for the user.* Prentice Hall, Englewood Cliffs, NJ, 1987.

[7] H.S. Zhang, L. H. Xie and Y. C. Soh, "Optimal recursive filtering, prediction and smoothing for singular stochastic discrete-time systems," *IEEE Trans. Atomat. Contr.*, vol. 44, no. 11, pp. 2154-2158, 1999.

[8] L. Ljung, "Convergence analysis of parametric identification methods," *IEEE. Trans. Automat. Contr.,* 23, 770-783, 1978.

[9] V. Kucera, "New results in state estimation and regulation", *Automatica*, vol. 17, no. 5, pp. 745-748, 1981.

[10] B. D. O. Anderson and J. B. Moore, *Optimal filtering.* Prentice-Hall, Englewood Cliffs, N.J, 1979.