EM Algorithm State Matrix Estimation for Navigation

Garry A. Einicke, Gianluca Falco, and John T. Malos

***Abstract*—The convergence of an expectation-maximization (EM) algorithm for state matrix estimation is investigated. It is shown for the expectation step that the design and observed error covariances are monotonically dependent on the residual error variances. For the maximization step, it is established that the residual error variances are monotonically dependent on the design and observed error covariances. The state matrix estimates are observed to be unbiased when the measurement noise is negligible. A navigation application is discussed in which the use of estimated parameters improves filtering performance.**

***Index Terms*—Kalman filtering, navigation, parameter estimation.**

I. Introduction

OPTIMAL minimum-variance filters and smoothers are widely used within bio-medical, communication, navigation and economic forecasting applications. These techniques minimize the variance of the estimation error provided that the state-space model parameters and noise statistics are known precisely. An iterative technique for estimating these unknowns is the expectation-maximization (EM) algorithm which is described in [1]–[6].

The EM algorithm for parameter estimation was first proposed by [1]. The algorithm consists of iterating a conditional expectation step and a likelihood maximization step. The expectation step of [1] involves least squares calculations on the incomplete observations using the current parameter iterations to estimate the underlying states. The maximization step involves re-estimating the parameters by maximizing a joint log-likelihood function using state estimates from the previous expectation step. In [2], a Kalman filter is used within the expectation step to recover the states. A multiparameter estimation problem is decoupled into separate maximum-likelihood estimations (MLEs) within the EM algorithm of [3].

General conditions for EM algorithm convergence are detailed in [4]. However, the likelihood functions for state-space parameter estimation do not exist in explicit closed form. This precludes straightforward calculation of the Hessians required in [4] to establish convergence. A recent study into the monotonicity of variance estimates is presented in [5]. In this sequel, similar monotonicity results for a sequence of state matrix estimates are presented.

Under simplified conditions, these EM algorithms are similar to subspace identification techniques [7]–[10], which typically consists of three steps. The first step involves applying pseudo-inverse or linear regression calculations on blocks of input and output matrices to obtain predicted states. It is explained in [8] that this step is equivalent to a bank of parallel Kalman filters. Indeed, Kalman filters are used for stochastic state estimation in [9] and the orthogonal decomposition technique of [10]. In step two, an information measure is used to determine the minimum model order [7]. Finally, the parameters are estimated from the states using least-squared techniques. It is assumed within [5] and herein that the underlying system is observable, which dispenses with the need for an order reduction step. In the case of state matrix estimation, the maximum-likelihood and least squares approaches yield the same formula. The EM algorithm described herein differs from the subspace identification techniques of [7]–[10] in the following respects. First, access to the input sequence is not essential. Second, it is applicable to on-line applications because the sums within the maximum-likelihood formulas may be updated over a receding horizon. Third, the state and parameter estimates can improve with successive iterations. However, the algorithm is confined to stable systems and the state matrix estimates are only unbiased when the measurement noise is negligible.

This correspondence is organized as follows. An EM algorithm for state matrix estimation is described in Section II. It is shown that the design and observed error covariances are monotonically dependent on the residual error variances. A navigation example is presented in Section III.

II. State Matrix Estimation

A. Problem Definition

Consider a linear time-invariant system G: $\mathbb{R}^n \rightarrow \mathbb{R}^p$ having the state-space realization

$$x_{k+1} = Ax_k + Bw_k \tag{1}$$

$$z_k = Cx_k + v_k \tag{2}$$

where $A \in \mathbb{R}^{n\times n}$, $B \in \mathbb{R}^{n\times m}$, $C \in \mathbb{R}^{p\times n}$ and $w_k \in \mathbb{R}^n$, $v_k \in \mathbb{R}^p$ are independent zero-mean stationary processes with $Q = E\{w_k w_k^T\}$ and $R = E\{v_k v_k^T\}$. EM algorithms which use filtered states to iteratively recalculate approximate maximum-likelihood estimates of uncertain Q and R matrices are reported in [3] and [5]. A similar procedure for calculating approximate maximum-likelihood estimates of an uncertain A is described below. This procedure uses a minimum-variance Kalman filter

Manuscript received October 25, 2009; revised February 04, 2010; accepted February 04, 2010. First published February 17, 2010; current version published March 17, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ricardo Merched.

The authors are with the Commonwealth Scientific and Industrial Research Organisation (CSIRO), Technology Court, Pullenvale, QLD 4069, Australia (e-mail: garry.einicke@csiro.au; Gianluca.Falco@csiro.au; John.Malos@csiro.au).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2010.2043151

to calculate predicted states $\hat{x}^{(m)}_{k/k-1}$ and corrected states $\hat{x}^{(m)}_{k/k}$ from the measurements z_k using an estimate $\hat{A}^{(m)}$ of A at iteration m. Let $\tilde{w}_k = w_k - \hat{w}_k$, where $\hat{w}_k$ denotes an estimate of w_k and define $\tilde{Q} = E\{\tilde{w}_k \tilde{w}_k^T\}$. The Kalman filter is given by

$$\hat{x}^{(m)}_{k/k} = \left(I - L^{(m)}C\right)\hat{x}^{(m)}_{k/k-1} + L^{(m)}z_k \quad (3)$$

$$\hat{x}^{(m)}_{k+1/k} = \left(\hat{A}^{(m)} - K^{(m)}C\right)\hat{x}^{(m)}_{k/k} + K^{(m)}z_k + B\hat{w}_k \quad (4)$$

where $L^{(m)} = P^{(m)}C^T(\Omega^{(m)})^{-1}$ and $K^{(m)} = \hat{A}^{(m)}L^{(m)}$ are the filter and predictor gains, respectively, $\Omega^{(m)} = CP^{(m)}C^T + R$, in which $P^{(m)} \in \mathbb{R}^{n\times n}$ is the solution of the algebraic Riccati equation (ARE)

$$P^{(m)} = \hat{A}^{(m)}P^{(m)}\left(\hat{A}^{(m)}\right)^T - K^{(m)}\Omega^{(m)}\left(K^{(m)}\right)^T + B\tilde{Q}B^T. \quad (5)$$

Suppose that there exists a residual error $s_k^{(m)} \in \mathbb{R}^n$ at iteration m such that

$$\hat{x}^{(m)}_{k+1/k+1} = \hat{A}^{(m)}\hat{x}^{(m)}_{k/k} + B\hat{w}_k + s_k^{(m)}. \quad (6)$$

Let $a_{i,j}$ and $\hat{a}^{(m)}_{i,j}$ denote the components of A and $\hat{A}^{(m)}$ at row i and column j, respectively. Then (6) may be written as $\hat{x}^{(m)}_{i,k+1/k+1} = \sum_{j=1}^{n} \hat{a}^{(m)}_{i,j}\hat{x}^{(m)}_{i,k/k} + B_i\hat{w}_{i,k} + s^{(m)}_{i,k}$, where $\hat{x}^{(m)}_{i,k}$, $s^{(m)}_{i,k}$, $\hat{w}_{i,k}$ and B_i denote the i^{th} element and i^{th} row of $\hat{x}^{(m)}_k$, $s^{(m)}_k$, $\hat{w}_k$ and B, respectively. Using the approaches of [3], [5], and [6, pp. 157–204], it is assumed that the observed states are normally distributed, i.e., $\hat{x}^{(m)}_{i,k+1/k+1} \sim N(\hat{a}^{(m)}_{i,j}\hat{x}^{(m)}_{i,k/k} + B_i\hat{w}_{i,k}, \sigma^2_{s_i^{(m)}})$, $k \in [0, N]$, an approximate maximum-likelihood estimate of $a_{i,j}$ can be calculated by

$$\hat{a}^{(m+1)}_{i,j} = \frac{\sum_{k=0}^{N-1}\left(\hat{x}^{(m)}_{i,k+1/k+1} - B_i\hat{w}_{i,k} - \sum_{j=1, j\neq i}^{n}\hat{a}^{(m)}_{i,j}\hat{x}^{(m)}_{j,k/k}\right)\hat{x}^{(m)}_{j,k/k}}{\sum_{k=0}^{N-1}\left(\hat{x}^{(m)}_{j,k/k}\right)^2}. \quad (7)$$

The advantage of applying the maximum-likelihood technique for normally distributed observations is that it can lead to unbiased estimates which achieve the Cramer–Rao-lower-bound, and have a gaussian probability density function; see of [6, pp. 157–204], [11]. In applications where the observations are Poisson, multinomial or exponential, the maximum-likelihood approaches described within [11] could be used instead.

An EM algorithm for estimating an unknown A is proposed which involves iterating the following two-step procedure. Suppose that an initial estimate $\hat{A}^{(0)} > 0$ is available.

Expectation step: Use a Kalman filter designed with $\hat{A}^{(m)}$ within (3)–(4) is used to calculate filtered state estimates $\hat{x}^{(m)}_{k/k}$.

Maximization step: Using the $\hat{a}^{(m)}_{i,j}$ and the filtered states $\hat{x}^{(m)}_{i,k/k}$, calculate a candidate $\hat{a}^{(m+1)}_{i,j}$ from (7). Include the candidate $\hat{a}^{(m+1)}_{i,j}$ within $\hat{A}^{(m+1)}$ provided that $\hat{A}^{(m+1)} > 0$, to ensure stablity.

The above algorithm reuses past measurements and thus is suitable for retrospective or offline analyses. In online or real-time applications, the sums within (7) may be updated over a receding horizon using the most-recent state estimate. The problem addressed here is to determine conditions under which the residual error covariance $S^{(m)} = E\{s_k^{(m)}(s_k^{(m)})^T\}$ decreases monotonically. Monotonicity is established in two steps. For the expectation step, it is shown below that the design filter error covariance is monotonically dependent on the residual error covariance. Subsequently, it is shown for the maximization step that the residual error covariance is monotonically dependent on the design and observed error covariances.

B. Expectation Step Monotonicity

Suppose that there exists a $\overline{s}_k^{(m)} \in \mathbb{R}^n$ such that $\hat{A}^{(m)} = A$ and $\hat{x}^{(m)}_{k+1/k} = (A - K^{(m)}C)\hat{x}^{(m)}_{k/k} + K^{(m)}z_k + B\hat{w}_k + \overline{s}_k^{(m)}$. The state prediction error is then given by $\tilde{x}^{(m)}_{k+1/k} = x_k - \hat{x}^{(m)}_{k+1/k} = (A - K^{(m)}C)x^{(m)}_{k/k-1} - K^{(m)}v_k + B\tilde{w}_k - \overline{s}_k^{(m)}$. An alternative design error covariance follows by constructing $P^{(m)} = E\{\tilde{x}^{(m)}_{k+1/k}(\tilde{x}^{(m)}_{k+1/k})^T\}$, which yields

$$P^{(m)} = \left(A - K^{(m)}C\right)P^{(m)}\left(A - K^{(m)}C\right)^T + K^{(m)}R\left(K^{(m)}\right)^T + B\tilde{Q}B^T + \overline{S}^{(m)} \quad (8)$$

where $\overline{S}^{(m)} = E\{\overline{s}_k^{(m)}(\overline{s}_k^{(m)})^T\}$.

The Kalman filter design error covariance $P^{(m)}$, which is the solution of (5), or equivalently (8), assumes that the estimate $\hat{A}^{(m)}$ of the unknown A is correct. Since modeling error is present, the observed error covariance differs from $P^{(m)}$. The observed predicted state error is given by $\tilde{x}^{(m)}_{k+1/k} = (A - K^{(m)}C)x^{(m)}_{k/k-1} - K^{(m)}v_k + B\tilde{w}_k$. It follows that the observed predicted error covariance, namely $\Sigma^{(m)} = E\{\tilde{x}^{(m)}_{k+1/k}(\tilde{x}^{(m)}_{k+1/k})^T\}$ is given by

$$\Sigma^{(m)} = \left(A - K^{(m)}C\right)\Sigma^{(m)}\left(A - K^{(m)}C\right)^T + K^{(m)}R\left(K^{(m)}\right)^T + B\tilde{Q}B^T. \quad (9)$$

It is shown below that $P^{(m)}$ and $\Sigma^{(m)}$ are monotonically dependent on $\overline{S}^{(m)}$.

Lemma 3.1: Suppose the following:

i) A has its eigenvalues inside the unit circle and the pair (A, C) is observable;

ii) $P^{(1)}, P^{(2)}, \Sigma^{(1)}, \Sigma^{(2)} \in \mathbb{R}^{n\times n}$ are available which satisfy $P^{(2)} \le P^{(1)}$, $\Sigma^{(2)} \le \Sigma^{(1)}$;

iii) $\overline{S}^{(m+1)} \le \overline{S}^{(m)}$ for all $m \ge 1$.

Then $P^{(m+1)} \le P^{(m)}$ and $\Sigma^{(m+1)} \le \Sigma^{(m)}$ for all $m \ge 1$.

Proof: Condition i) implies that the solutions of the AREs (5) and (8) exist and are unique. Condition ii) is the initialization step of an induction argument. For the induction step, it can be verified that $P^{(m+1)} \leq P^{(m)}$ and $\Sigma^{(m+1)} \leq \Sigma^{(m)}$ assuming Condition ii) within the proof of [Lemma 2.2, 5]. ■

C. Maximization Step Monotonicity

Combining (6) and (7) yields

$$\hat{a}_{i,j}^{(m+1)} = \hat{a}_{i,j}^{(m)} + \left(\sum_{k=0}^{N-1} s_{i,k}^{(m)} \hat{x}_{j,k/k}^{(m)}\right)\left(\sum_{k=0}^{N-1}\left(\hat{x}_{j,k/k}^{(m)}\right)^2\right)^{-1} \tag{10}$$

It can be seen from (10) that the monotonicity of $\hat{A}^{(m)}$ and $S^{(m)}$ are co-dependent. Conditions for the monotonicity of $S^{(m)}$ are established below.

Lemma 3.2: Under condition i) of Lemma 3.1, suppose the following:

i) $S^{(1)}, \overline{S}^{(1)}, S^{(2)}, \overline{S}^{(2)} \in \mathbb{R}^{n\times n}$ are available which satisfy $S^{(2)} \leq S^{(1)}, \overline{S}^{(2)} \leq \overline{S}^{(1)}$; and

ii) $P^{(m+1)} \leq P^{(m)}$ and $\Sigma^{(m+1)} \leq \Sigma^{(m)}$ for all $m \geq 1$.

Then $S^{(m+1)} \leq S^{(m)}$ and $\overline{S}^{(m+1)} \leq \overline{S}^{(m)}$ for all $m \geq 1$.

Proof: Substituting (3) and (4) into (6) yields $s_k^{(m)} = L^{(m)}(C\tilde{x}_{k+1/k}^{(m)} + v_k)$, which implies

$$S_k^{(m)} = L^{(m)}\left(C\Sigma^{(m)}C^T + R\right)\left(L^{(m)}\right)^T. \tag{11}$$

Condition i) is the initialization step of an induction argument. For the induction step, $P^{(m+1)} \leq P^{(m)}$ implies $L^{(m+1)}(L^{(m+1)})^T \leq L^{(m)}(L^{(m)})^T$, which together with $\Sigma^{(m+1)} \leq \Sigma^{(m)}$ and (11) yields $S^{(m+1)} \leq S^{(m)}$. From the approach of [5], $P^{(m+1)} \leq P^{(m)}$ is equivalent to

$$\begin{bmatrix} B\tilde{Q}B^T + \overline{S}^{(m+1)} & A^T \\ A & -C^TR^{-1}C \end{bmatrix} \leq \begin{bmatrix} B\tilde{Q}B^T + \overline{S}^{(m)} & A^T \\ A & -C^TR^{-1}C \end{bmatrix}. \tag{12}$$

The result $\overline{S}^{(m+1)} \leq \overline{S}^{(m)}$ follows from (12), since $A, B, C, \tilde{Q}$ and R are time-invariant. ■

D. Low-Measurement Noise Asymptote

It is observed that if $\tilde{w}_k$ is independent of w_{k-1} and the measurement noise is negligible, the estimates (7) are unbiased.

Lemma 3.3: Suppose that C is full rank. If $E\{\tilde{w}_k\tilde{w}_k^T\} = 0$ then $\lim_{R\to 0} \hat{a}_{i,j}^{(m+1)} = a_{i,j}$.

Proof: Premultiplying both sides of (3) by C and using $\lim_{R\to 0} CL_i = I$ results in $\hat{x}_{k/k}^{(m)} = x_k$. It follows from (6) that $s_k^{(m)} = (A - \hat{A}^{(m)})x_k + B\tilde{w}_k$, which together with (7) yields

$$\hat{a}_{i,j}^{(m+1)} = a_{i,j} + \left(\sum_{k=0}^{N-1} B_i\tilde{w}_{i,k}x_{j,k}\right)\left(\sum_{k=0}^{N-1}(x_{j,k})^2\right)^{-1}. \tag{13}$$

The result follows from $\tilde{w}_k = w_k - \hat{w}_k$ and (13), since $x_k = Ax_{k-1} + Bw_{k-1}$ is independent of w_k and $\hat{w}_k$. ■

III. Examples

It is demonstrated below that if the model (1)–(2) is known but the state matrix elements are uncertain, the sequence of residual error variances will be monotonically nonincreasing.

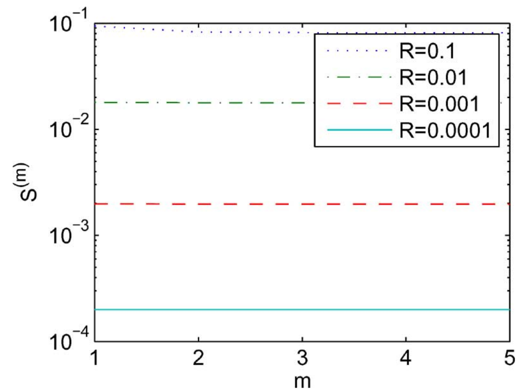

Fig. 1. Mean residual error variances versus iteration number for 5 EM algorithm iterations.

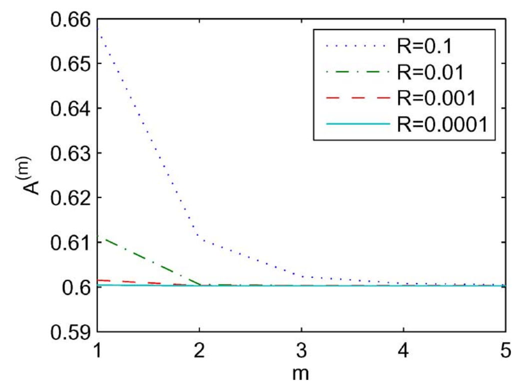

Fig. 2. Mean state matrix estimates versus iteration number for 5 EM algorithm iterations.

Example 1: In respect of the model (1), (2), suppose that $A = 0.6$, $B = C = 1$, $\sigma_w^2 = 0.2$ and $\sigma_v^2 = 0.1$. The design error variance was initialized by solving the ARE (5) with $\hat{A}^{(0)} = 0.9999$. Simulations were conducted with 100 realizations of Gaussian process noise and measurement noise with $N = 500\,000$. The mean residual error variances for 5 EM algorithm iterations are shown in Fig. 1. The figure shows that the variance sequence is monotonically nonincreasing which is consistent with Lemmata 3.1 and 3.2. In this example, all variance sequences are found to be monotonically decreasing, however, this becomes imperceptable at low R due to the limited resolution of the plot. The mean state matrix estimates are shown in Fig. 2. It can be seen that estimates asymptotically approach the true value of $A = 0.6$ when the measurement noise becomes negligible, which illustrates Lemma 3.3.

An example is presented below to demonstrate EM algorithm state matrix estimation even though the underlying model is unknown.

Example 2: A GPS receiver possessing a Sirf 3 chip set was driven along a surveyed 236-m-long track for approximately 11 minutes at CSIRO's Pullenvale site. The position, velocity and time were calculated using pseudorange and ephemeris data according to [12, ch.2, pp. 147–171] and [13, ch. 8, pp. 139–142]. It is desired to filter the noisy position estimates. In respect of the model (1)–(2), let $A = \mathrm{diag}(a_{1,1}, a_{2,2}, a_{3,3})$, $B = \mathrm{diag}(1,1,1)$, $C = \mathrm{diag}(1,1,1)$, $Q = \mathrm{diag}(\sigma_{w_1}^2, \sigma_{w_2}^2, \sigma_{w_3}^2)$, $R = \mathrm{diag}(100, 100, 250)$, where

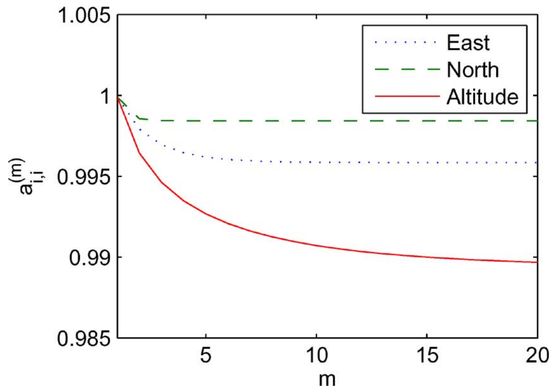

Fig. 3. Sequence of state parameter estimates versus iteration number.

$a_{1,1}$, $a_{2,2}$, $a_{3,3}$, $\sigma_{w_1}^2$, $\sigma_{w_2}^2$ and $\sigma_{w_3}^2$ are unknowns. The state matrix was initialized with $\hat{a}_{i,i}^{(0)} = 0.9999$, $i = 1\ldots 3$. It follows from $x_{i,k+1} = a_{i,i}x_{i,k} + w_{1,k}$ that $\sigma_{w_i}^2 = (1 - a_{i,i}^2)\sigma_{x_i}^2$, where $\sigma_{x_i}^2$ is the sample variance of $x_i = \{x_{i,1}, \ldots, x_{i,N}\}$. Therefore, the process noise variances were estimated using $\hat{\sigma}_{w_i}^2 = (1 - (a_{i,i}^{(0)})^2)\sigma_{z_i}^2$, where $\sigma_{z_i}^2$ is the sample variance of the measurements $z_i = \{z_{i,1}, \ldots, z_{i,N}\}$. The sequence of state matrix estimates, calculated from the EM algorithm described in Section II-A, is shown in Fig. 3. It can be seen that the sequences of $\hat{a}_{i,i}^{(m)}$ are monotonic nonincreasing, which is consistent with Lemmata 3.1 and 3.2. The root-mean-square errors of the north, east and altitude GPS measurements were 6.84, 3.13 m, and 12.7 m, respectively. The filtered north, east, and altitude root-mean-square errors after 20 EM algorithm iterations were 3.85 m, 2.10 m, and 5.98 m, respectively. Since a performance improvement occurred, it is suggested that the state matrix estimates are reasonable.

IV. Conclusion

An EM algorithm for joint estimation of the states and state matrix parameters from noisy measurements is described. The following is established: a) the sequence of design and observed error covariances is monotonically dependent on the residual error variances; b) the residual error variances are monotonically dependent on the design and observed error covariances; and c) when the measurement noise becomes negligible, the MLEs of the state matrix elements asymptotically approach the actual values.

It is demonstrated that filtering noisy GPS receiver measurements can yield improved mean-square-error performance when an EM algorithm is used to estimate the unknown parameters.

References

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood estimation from incomplete data via the EM algorithm," *J. Royal Statist. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.

[2] R. H. Shumway and D. S. Stoffer, "An approach to time series smoothing and forecasting using the EM algorithm," *J. Time Series Anal.*, vol. 3, no. 4, pp. 253–264, 1982.

[3] Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Signal Process.*, vol. 36, no. 4, pp. 477–489, Apr. 1988.

[4] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Statist.*, vol. 11, no. 1, pp. 95–103, Mar. 1983.

[5] G. A. Einicke, J. T. Malos, D. C. Reid, and D. W. Hainsworth, "Riccati equation and EM algorithm convergence for inertial navigation alignment," *IEEE Trans. Signal Process.*, vol. 57, no. 1, Jan. 2009.

[6] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1993, ch. 7.

[7] W. E. Larimore, "Maximum likelihood subspace identification for linear, nonlinear and closed-loop systems," in *Proc. Amer. Control Conf.*, Jun. 2005, pp. 2305–2319.

[8] P. Van Overschee and B. De Moor, "A unifying theorem for three subspace system identification algorithms," *Automatica*, vol. 21, no. 2, pp. 1853–1864, 1995.

[9] F. Ding, L. Qiu, and T. Chen, "Reconstruction of continuous-time systems from their non-uniformly sampled discrete-time systems," *Automatica*, vol. 45, no. 2, pp. 324–332, 2009.

[10] T. Katayama, *Subspace Methods for System Identification*. London, U.K.: Springer-Verlag London, 2005.

[11] A. Van Den Bos, *Parameter Estimation for Scientists and Engineers*. Hoboken, NJ: Wiley, 2007.

[12] E. D. Kaplan and C. J. Hegarty, *Understanding GPS: Principles and Applications*, 2nd ed. Boston, MA: Artech House, 2006.

[13] K. Borre and D. M. Akos, *A Software-Defined GPS and Galileo Receiver*. Boston, MA: Birkhauser, 2007.