# System Identification of Linear Parameter Varying State-space Models

Adrian Wills*and Brett Ninness

**Abstract**

This chapter examines the estimation of multivariable linear models for which the parameters vary in a time-varying manner that depends in an affine fashion on a known or otherwise measured signal. These locally linear models which depend on a measurable operating point are known as linear parameter varying (LPV) models. The contribution here relative to previous work on the topic is that in the Gaussian case, an expectation-maximisation (EM) algorithm-based solution is derived and profiled.

## 1 Introduction

This chapter considers the problem of estimating the parameters in linear parameter varying (LPV) model structures on the basis of observed input-output data records. LPV models are linear time varying structures wherein the time dependence is affinely related to a known "scheduling" signal. They have proven to be of importance in aerospace [18], engine control [2], and compressor control applications [8] amongst others [20].

Due to their relevance, the problem of estimating LPV models has attracted significant research attention. The contribution [16] was the first to not require state measurements, and work only with input-output measurements. It examined least squares prediction error (PE) estimation of multiple-input multiple-output (MIMO) state space structures with possible noise corruptions on the output measurements via a gradient-based search technique. This approach has been further examined and applied in [7, 23, 24].

*Both authors are with the School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW 2308, Australia. Email: {Adrian.Wills, Brett.Ninness}@newcastle.edu.au

The estimation of single-input single-output (SISO) transfer function LPV model structure has also been addressed. The work in [3] has considered estimation of a ARX-type LPV models via least mean square (LMS) and recursive least squares (RLS) algorithms. The recent paper [15] addresses more general Box–Jenkins type LPV models via an instrumental variable method.

A final key strand of research has been the development of subspace-based methods, again for the estimation of MIMO state-space LPV models [6, 21, 22, 25]. A main advantage of these methods is that they are non-iterative in nature, and their effectiveness has been demonstrated empirically.

This chapter addresses the estimation of MIMO state-space LPV systems that are affected by both state and measurement noise. Among that many possibilities, here a maximum-likelihood (ML) criteria is used. A key rationale for the consideration is the availability of underpinning theory assuring their general consistency, even in the presence of quite arbitrary stable closed loop control [17]. The particular importance of this for LPV systems has been emphasised and explained in [21].

There are several possible choices of algorithm for maximising the likelihood function. This chapter will develop and illustrate a new estimation approach using the expectation-maximisation (EM) algorithm. The performance of this technique for other linear time invariant, bilinear,frequency domain and other identification problems has been previously demonstrated [9, 10, 12–14, 26]. As will be illustrated here, it also delivers an effective and numerically robust solution for LPV estimation. Furthermore, it has reasonable computational requirements which scale modestly with increasing system state and input-ouput dimension.

Having developed the EM method for MIMO LPV systems, the chapter profiles it empirically via a simulation study.

## 2 Problem Formulation

This chapter considers the following linear time-varying state-space model structure

$$x_{t+1} = A_t x_t + B_t u_t + w_t, \tag{1a}$$

$$y_t = C_t x_t + D_t u_t + e_t. \tag{1b}$$

Here $x_t \in \mathbf{R}^n$ is the system state, $u_t \in \mathbf{R}^m$ is the system input, and $y_t \in \mathbf{R}^\ell$ is the system output. The noise terms $w_t \in \mathbf{R}^n$ and $e_t \in \mathbf{R}^\ell$ are assumed to be zero mean i.i.d. process for which

$$\operatorname{Cov}\left\{\begin{bmatrix} w_t \\ e_t \end{bmatrix}\right\} = \begin{bmatrix} Q & S \\ S^T & R \end{bmatrix} \triangleq \Pi \succ 0. \tag{1c}$$

The time variations in the system matrices in (1a), (1b) are assumed to be of the following linear forms

$$A_t = \sum_{i=0}^{n_\alpha} \alpha_t(i) A(i), \quad A(i) \in \mathbf{R}^{n \times n} \ \forall i, \tag{1d}$$

$$B_t = \sum_{i=0}^{n_\beta} \beta_t(i) B(i), \quad B(i) \in \mathbf{R}^{n \times m} \ \forall i \tag{1e}$$

$$C_t = \sum_{i=0}^{n_\alpha} \alpha_t(i) C(i), \quad C(i) \in \mathbf{R}^{\ell \times n} \ \forall i, \tag{1f}$$

$$D_t = \sum_{i=0}^{n_\beta} \beta_t(i) D(i), \quad D(i) \in \mathbf{R}^{\ell \times m} \ \forall i \tag{1g}$$

where the signals

$$\alpha_t = \begin{bmatrix} \alpha_t(0) \\ \vdots \\ \alpha_t(n_\alpha) \end{bmatrix} \in \mathbf{R}^{n_\alpha}, \quad \beta_t = \begin{bmatrix} \beta_t(0) \\ \vdots \\ \beta_t(n_\beta) \end{bmatrix} \in \mathbf{R}^{n_\beta} \tag{1h}$$

are termed "*scheduling parameters*" that are assumed known or measurable. A model of this form (1a)-(1h) is an example of a *linear parameter varying* (LPV) system, and in practice the values taken by the scheduling parameter specify a given system operating point.

In what follows, it will prove important to note the equivalent formulation

$$\begin{bmatrix} x_{t+1} \\ y_t \end{bmatrix} = \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix} \begin{bmatrix} (\alpha_t \otimes x_t) \\ (\beta_t \otimes u_t) \end{bmatrix} + \begin{bmatrix} w_t \\ e_t \end{bmatrix} \tag{2}$$

where

$$\mathcal{A} = \begin{bmatrix} A(0) & \dots & A(n_\alpha) \end{bmatrix}, \quad \mathcal{B} = \begin{bmatrix} B(0) & \dots & B(n_\beta) \end{bmatrix} \tag{3}$$

$$\mathcal{C} = \begin{bmatrix} C(0) & \dots & C(n_\alpha) \end{bmatrix}, \quad \mathcal{D} = \begin{bmatrix} D(0) & \dots & D(n_\beta) \end{bmatrix} \tag{4}$$

and $\otimes$ denotes the Kronecker tensor product [4].

Assume that the system matrices $\mathcal{A}, \mathcal{B}, \mathcal{C}, \mathcal{D}$, the disturbance covariance $\Pi$ and the initial state $x_1$ are appropriately parametrized by the elements of a vector $\theta \in \mathbf{R}^{n_\theta}$. Details on how this may be achieved are provided in what follows.

Then the problem considered in this chapter is the estimation of a value of $\theta$ that best explains (in a manner soon to be made precise) an observed length $N$ data record

$$U = [u_1, \cdots, u_N], \quad Y = [y_1, \cdots, y_N] \tag{5}$$

of the input-output response.

## 3 Maximum-Likelihood Estimation

One approach taken in this chapter to address this estimation problem is the employment of the classical Maximum-Likelihood (ML) technique. This requires the formulation of the joint density $p_\theta(Y)$ of the observation, which may be expressed using Bayes' rule as

$$p_\theta(Y) = p_\theta(y_0) \prod_{t=1}^{N} p_\theta(y_t \mid y_{t-1}, \cdots, y_1). \tag{6}$$

A maximum likelihood estimate $\widehat{\theta}_{\mathrm{ML}}$ of the parameter vector $\theta$ is then defined as any value $\widehat{\theta}_{\mathrm{ML}}$ satisfying

$$\widehat{\theta}_{\mathrm{ML}} \in \left\{ \theta \in \Theta : L(\theta) \leq L(\theta'), \, \forall \theta' \in \Theta \right\}, \tag{7}$$

where

$$L(\theta) \triangleq -\log p_\theta(Y) \tag{8}$$

and $\Theta \subseteq \mathbf{R}^{n_\theta}$ is a user chosen compact subset. In the above, the common notation $p_\theta$ is used to refer to a range of different densities, with the arguments to $p_\theta$ indicating what is intended. Furthermore, note that while the form of $p_\theta$ will depend on $U$, it is not formally an argument since it is not a random variable.

In the case where stochastic elements $w_t, e_t$ are Gaussian distributed

$$\begin{bmatrix} w_t \\ e_t \end{bmatrix} \sim \mathcal{N}(0, \Pi) \tag{9}$$

then the density $p_\theta(y_t \mid Y_{t-1})$ will also have a Gaussian density and lead to (ignoring constant terms that do not affect the minimisation (7)) [1]

$$L(\theta) = \sum_{t=1}^{N} \log \det\{\Lambda_t(\theta)\} + \varepsilon_t^T(\theta) \Lambda_t^{-1} \varepsilon_t(\theta) \tag{10}$$

where

$$\varepsilon_t(\theta) = y_t - \widehat{y}_{t|t-1}(\theta), \quad \Lambda_t(\theta) = \mathrm{Cov}\{\widehat{y}_{t|t-1}(\theta)\} \tag{11}$$

with $\widehat{y}_{t|t-1}(\theta)$ being the mean square optimal one step ahead predictor of $y_t$. This can be computed using a standard Kalman filter.

$$\widehat{y}_{t|t-1} = C_t\widehat{x}_{t|t-1} + D_t u_t, \tag{12}$$

$$\widehat{x}_{t+1|t} = A_t\widehat{x}_{t|t-1} + B_t u_k + K_t e_t, \tag{13}$$

$$K_t = \left(A_t P_{t|t-1} C_t^T + S\right) \Lambda_t^{-1}, \tag{14}$$

$$P_{t+1|t} = A_t P_{t|t-1} A_t^T + Q - K_t \Lambda_t K_t^T, \tag{15}$$

$$\Lambda_t = C_t P_{t|t-1} C_t^T + R. \tag{16}$$

where $\widehat{x}_{t|t-1}$ is the mean square optimal one step ahead state prediction which has associated covariance $P_{t|t-1}$.

## 4 The Expecation-Maximisation (EM) Algorithm for ML Estimation

Consider for a moment that in addition to $U$, $Y$ a record

$$X = [x_1, x_2, \cdots, x_{N+1}] \tag{17}$$

of the state sequence for a system modeled by (1a), (1b) is also available. Via the definition of conditional probability, this would permit the negative log-likelihood (8) to be expressed as

$$L(\theta) \triangleq \log p_\theta(Y) = \log p_\theta(X, Y) - \log p_\theta(X|Y). \tag{18}$$

The reason for considering this idea is that, as will be seen presently, under the LPV model (1a), (1b) the associated log-likelihood $p_\theta(X, Y)$ can be maximised in closed form. However, it is unreasonable to assume that a record of the state sequence $X$ is available. To address this, the log-likelihood $\log p_\theta(X, Y)$ can be approximated by its expected value

$$\mathcal{Q}(\theta, \widehat{\theta}_k) \triangleq \mathbf{E}_{\widehat{\theta}_k}\{\log p_\theta(X, Y) \mid Y\} \tag{19}$$

conditional on the data $Y$ observed, and an initial estimate $\widehat{\theta}_k$. Due to the expectation operator $\mathbf{E}_{\widehat{\theta}_k}\{\cdot \mid Y\}$ being linear, this approximation $\mathcal{Q}(\theta, \widehat{\theta}_k)$ can also be maximised in closed form.

Furthermore, via (18), this approximation is related to $L(\theta)$ according to

$$L(\theta) = \mathcal{Q}(\theta, \widehat{\theta}_k) - \mathcal{V}(\theta, \widehat{\theta}_k) \tag{20}$$

where

$$\mathcal{V}(\theta, \widehat{\theta}_k) \triangleq \mathbf{E}_{\widehat{\theta}_k}\{\log p_\theta(X \mid Y) \mid Y\}. \tag{21}$$

Therefore

$$\begin{aligned} L(\theta) - L(\widehat{\theta}_k) = &\left[\mathcal{Q}(\theta, \widehat{\theta}_k) - \mathcal{Q}(\widehat{\theta}_k, \widehat{\theta}_k)\right] + \\ &\left[\mathcal{V}(\widehat{\theta}_k, \widehat{\theta}_k) - \mathcal{V}(\theta, \widehat{\theta}_k)\right]. \end{aligned} \tag{22}$$

The second term

$$\mathcal{V}(\widehat{\theta}_k, \widehat{\theta}_k) - \mathcal{V}(\theta, \widehat{\theta}_k) = \int \log\left[\frac{p_{\widehat{\theta}_k}(X \mid Y)}{p_\theta(X \mid Y)}\right] p_{\widehat{\theta}_k}(X \mid Y)\,\mathrm{d}X \tag{23}$$

is the Kullback-Leibler divergence metric between $p_\theta(X|Y)$ and $p_{\widehat{\theta}_k}(X \mid Y)$, and hence is non-negative. As a result, any value of $\theta$ for which $\mathcal{Q}(\theta, \widehat{\theta}_k) > \mathcal{Q}(\widehat{\theta}_k, \widehat{\theta}_k)$ implies that $L(\theta) > L(\widehat{\theta}_k)$.

This suggests a strategy of maximising $\mathcal{Q}(\theta, \widehat{\theta}_k)$, which must increase $L(\theta)$ via (22), and then setting $\widehat{\theta}_{k+1}$ equal to this maximiser and repeating the process. This procedure is known as the expectation-maximisation (EM) algorithm, which can be stated in in abstract form as follows.

---

**Algorithm 4.1** ***EM Algorithm***

*1.* **E-Step**

$$\textit{Calculate:} \qquad \mathcal{Q}(\theta, \widehat{\theta}_k); \tag{24}$$

*2.* **M-Step**

$$\textit{Compute:} \qquad \widehat{\theta}_{k+1} = \arg\max_{\theta \in \Theta} Q(\theta, \widehat{\theta}_k). \tag{25}$$

---

# 5 EM for LPV Models

Application of the EM Algorithm to the LPV model (1a), (1b) requires development of how the E-Step and M-Step may be computed.

### 5.1 E-Step

The E-Step may be calculated by the following two lemmas.

**Lemma 5.1** *With regard to the LPV model structure* (1a)-(1h) *and the following assumption on the initial state*

$$x_1 \sim \mathcal{N}(\mu, P_1) \tag{26}$$

*the joint log-likelihood approximation* $\mathcal{Q}(\theta, \widehat{\theta}_k)$ *defined via* (17), (19) *is given by*

$$\begin{aligned} -2\mathcal{Q}(\theta, \widehat{\theta}_k) = &\log\det P_1 + N \log\det \Pi + \\ &\operatorname{Tr}\left\{P_1^{-1} \mathbf{E}_{\widehat{\theta}_k}\{(x_1 - \mu)(x_1 - \mu)^T \mid Y\}\right\} + \\ &\operatorname{Tr}\left\{\Pi^{-1}\left[\Phi - \Psi\Gamma^T - \Gamma\Psi^T + \Gamma\Sigma\Gamma^T\right]\right\}, \end{aligned} \tag{27}$$

*where*

$$\Gamma \triangleq \begin{bmatrix} \mathcal{A} & \mathcal{B} \\ \mathcal{C} & \mathcal{D} \end{bmatrix} \tag{28}$$

$$\Phi \triangleq \sum_{t=1}^{N} \begin{bmatrix} \mathbf{E}_{\widehat{\theta}_k}\{x_{t+1}x_{t+1}^T \mid Y\} & \widehat{x}_{t+1|N}y_t^T \\ y_t\widehat{x}_{t+1|N}^T & y_t y_t^T \end{bmatrix} \tag{29}$$

$$\Psi \triangleq \sum_{t=1}^{N} \begin{bmatrix} \alpha_t^T \otimes \mathbf{E}_{\widehat{\theta}_k}\{x_{t+1}x_t^T \mid Y\} & \beta_t^T \otimes \widehat{x}_{t+1|N}u_t^T \\ \alpha_t^T \otimes y_t\widehat{x}_{t|N}^T & \beta_t^T \otimes y_t u_t^T \end{bmatrix} \tag{30}$$

$$\Sigma \triangleq \sum_{t=1}^{N} \begin{bmatrix} \alpha_t\alpha_t^T \otimes \mathbf{E}_{\widehat{\theta}_k}\{x_t x_t^T \mid Y\} & \alpha_t\beta_t^T \otimes \widehat{x}_{t|N}u_t^T \\ \beta_t\alpha_t^T \otimes u_t\widehat{x}_{t|N}^T & \beta_t\beta_t^T \otimes u_t u_t^T \end{bmatrix} \tag{31}$$

*with*

$$\widehat{x}_{t|N} \triangleq \mathbf{E}_{\widehat{\theta}_k}\{x_t \mid Y\}. \tag{32}$$

**Proof 1** *The Markov property of the LPV model structure* (2) *together with Bayes' rule yields*

$$p_\theta(Y, X_{N+1}) = p_\theta(x_1) \prod_{t=1}^{N} p_\theta(x_{t+1}, y_t \mid x_t). \tag{33}$$

*Furthermore, via* (1c),(2)

$$p_\theta\left( \left[ \begin{array}{c} x_{t+1} \\ y_t \end{array} \right] \middle| x_t \right) \sim \mathcal{N}(\Gamma z_t, \Pi), \tag{34}$$

*where*

$$\xi_t \triangleq \begin{bmatrix} x_{t+1} \\ y_t \end{bmatrix}, \quad z_t \triangleq \begin{bmatrix} \alpha_t \otimes x_t \\ \beta_t \otimes u_t \end{bmatrix} \tag{35}$$

*and by the assumption* (26) $p_\theta(x_1) = \mathcal{N}(\mu, P_1)$. *Therefore, using* (33) *and excluding constant terms*

$$\begin{aligned} -2\log p_\theta(Y, X_{N+1}) \quad = \quad & \log\det P_1 + N \log\det \Pi + \\ & (x_1 - \mu)^T P_1^{-1}(x_1 - \mu) + \\ & \sum_{t=1}^{N} (\xi_t - \Gamma z_t)^T \Pi^{-1} (\xi_t - \Gamma z_t). \end{aligned} \tag{36}$$

*Applying the conditional expectation operator* $\mathbf{E}_{\widehat{\theta}_k}\{\cdot \mid Y\}$ *to both sides of equation* (36) *yields* (27) *with*

$$\Phi = \mathbf{E}_{\widehat{\theta}_k}\{\xi_t \xi_t^T \mid Y\}, \quad \Psi = \mathbf{E}_{\widehat{\theta}_k}\{\xi_t z_t^T \mid Y\} \tag{37}$$

$$\Sigma = \mathbf{E}_{\widehat{\theta}_k}\{z_t z_t^T \mid Y\}. \tag{38}$$

*Using the definitions* (32),(35), *the fact that* $\alpha_t, \beta_t$ *are deterministic, and elementary properties of the Kronecker product then completes the proof.* □

This reduces much of the computation of $\mathcal{Q}(\theta, \widehat{\theta}_k)$ to that of calculating the Kalman smoothed state estimate $\widehat{x}_{t|N}$ together with its covariance $P_{t|N}$, which then delivers $\mathbf{E}_{\widehat{\theta}_k}\{x_t x_t^T \mid Y\} = P_{t|N} + \widehat{x}_{t|N}\widehat{x}_{t|N}^T$. What remains is the additional computation of $\mathbf{E}_{\widehat{\theta}_k}\{x_{t+1} x_t^T \mid Y\}$, which is not obtainable by a standard Kalman smoother.

The following lemma establishes how all these quantities my be obtained via numerically robust "square root" recursions.

**Lemma 5.2** *The components*

$$\begin{aligned}\mathbf{E}_{\widehat{\theta}_k}\{x_t^T \mid Y\} &\triangleq \widehat{x}_{t|N}^T, &(39)\\ \mathbf{E}_{\widehat{\theta}_k}\{x_t x_t^T \mid Y\} &= \widehat{x}_{t|N}\widehat{x}_{t|N}^T + P_{t|N}^{1/2}P_{t|N}^{T/2}, &(40)\\ \mathbf{E}_{\widehat{\theta}_k}\{x_t x_{t-1}^T \mid Y\} &= \widehat{x}_{t|N}\widehat{x}_{t-1|N}^T + M_{t|N} &(41)\end{aligned}$$

*required for the computation of* (27) *via (29)-* (32) *may be robustly computed as follows. The smoothed state estimate* $\{\widehat{x}_{t|N}\}$ *is calculated via the reverse-time recursion*

$$\begin{aligned}\widehat{x}_{t|N} &= \widehat{x}_{t|t} + J_t\left[\widehat{x}_{t+1|N} - \overline{A}_t\widehat{x}_{t|t} - \overline{B}_t u_t - SR^{-1}y_t\right], &(42)\\ J_t &\triangleq P_{t|t}\overline{A}_t^T P_{t+1|t}^{-1}, &(43)\end{aligned}$$

*where*

$$\overline{A}_t \triangleq A_t - SR^{-1}C_t, \qquad \overline{B}_t \triangleq B_t - SR^{-1}D_t. \tag{44}$$

*The associated state covariance matrices are computed from their square roots, for example,* $P_{t|t} = P_{t|t}^{1/2}P_{t|t}^{T/2}$*, via recursions involving the following* $\mathcal{Q}R$*-decompositions*

$$\begin{bmatrix} P_{t|t}^{T/2}\overline{A}_t^T & P_{t|t}^{T/2} \\ \overline{Q}^{T/2} & 0 \\ 0 & P_{t+1|N}^{T/2}J_t^T \end{bmatrix} = \mathcal{Q}^1 \begin{bmatrix} \mathcal{R}_{11}^1 & \mathcal{R}_{12}^1 \\ 0 & \mathcal{R}_{22}^1 \\ 0 & 0 \end{bmatrix}, \tag{45}$$

$$\begin{bmatrix} P_{t-1|t-1}^{T/2}\overline{A}_{t-1}^T \\ \overline{Q}^{T/2} \end{bmatrix} = \mathcal{Q}^2 \begin{bmatrix} \mathcal{R}_1^2 \\ 0 \end{bmatrix}, \tag{46}$$

$$\begin{bmatrix} R^{T/2} & 0 \\ P_{t|t-1}^{T/2}C_t^T & P_{t|t-1}^{T/2} \end{bmatrix} = \mathcal{Q}^3 \begin{bmatrix} \mathcal{R}_{11}^3 & \mathcal{R}_{12}^3 \\ 0 & \mathcal{R}_{22}^3 \end{bmatrix}, \tag{47}$$

*where* $\overline{Q} \triangleq Q - SR^{-1}S^T$ *and then setting*

$$P_{t|N}^{T/2} = \mathcal{R}_{22}^1, \qquad P_{t|t-1}^{T/2} = \mathcal{R}_1^2, \qquad P_{t|t}^{T/2} = \mathcal{R}_{22}^3. \tag{48}$$

*The matrices $M_{N|N}$ and $M_{N+1|N}$ are calculated via the initialisation*

$$M_{N|N} = (I - K_N C_N)\overline{A}_{N-1} P_{N-1|N-1}, \tag{49}$$

$$M_{N+1|N} = \overline{A}_N P_{N|N}, \tag{50}$$

*followed by the the backwards recursion $\{M_{t|N}\}_{t=2}^{N-1}$ given by*

$$M_{t|N} = P_{t|t} J_{t-1}^T + J_t(M_{t+1|N} - \overline{A}_t P_{t|t}) J_{t-1}^T. \tag{51}$$

*Finally, the reverse time recursion (42) is initialised by running to $t = N$ the (robust) Kalman filter recursions*

$$K_t = P_{t|t-1} C_t^T \left( C_t P_{t|t-1} C_t^T + R \right)^{-1} = \mathcal{R}_{12}^3, \tag{52}$$

$$\widehat{x}_{t|t-1} = \overline{A}_{t-1} \widehat{x}_{t-1|t-1} + \overline{B}_t u_{t-1} + S R^{-1} y_{t-1}, \tag{53}$$

$$\widehat{x}_{t|t} = \widehat{x}_{t|t-1} + K_t \left( y_t - C_t \widehat{x}_{t|t-1} - D_t u_t \right), \tag{54}$$

*for $t = 1, \ldots, N$.*

**Proof 2** *See Section 4.1 in [9].*

### 5.2 M-Step

With the completion of the E-Step delivering $\mathcal{Q}(\theta, \widehat{\theta}_k)$, attention turns to maximising it with respect to $\theta$. At this point, the details of how $\theta$ parametrizes the LPV model (1a)-(2) need to be established. For this purpose, we assume that $\theta$ is defined in a partitioned manner according to

$$\theta^T \triangleq [\beta^T, \eta^T]^T \tag{55}$$

where recalling (1c), (26)

$$\beta^T \triangleq \left[ \text{vec}\{\Gamma\}^T, \mu^T \right], \quad \eta^T \triangleq \left[ \text{vec}\{\Pi\}^T, \text{vec}\{P_1\}^T \right], \tag{56}$$

and the vec$\{\cdot\}$ operator creates a vector from a matrix by stacking its columns on top of one another. We likewise partition the set $\Theta$ of candidate parameter vectors as

$$\Theta = \Theta_1 \times \Theta_2, \quad \beta \in \Theta_1, \quad \eta \in \Theta_2 \tag{57}$$

where $\Theta_1 \subseteq \mathbf{R}^{n_\beta}$, $\Theta_2 \subseteq \mathbf{R}^{n_\eta}$ are both compact, and for which all $\eta \in \Theta_2$ imply symmetric positive definite $\Pi, P_1$.

With these definitions in place, local maximisers of $\mathcal{Q}(\theta, \widehat{\theta}_k)$ may be directly computed via the results of the following lemma.

**Lemma 5.3** *Let* $\Sigma$ *defined in* (31) *satisfy* $\Sigma > 0$ *and be used to define* $\widehat{\beta}$ *according to (*$\Psi$ *is defined in* (30)*)*

$$\widehat{\beta} \triangleq \left[\mathrm{vec}\left\{\widehat{\Gamma}\right\}^T, \widehat{\mu}^T\right]^T, \quad \widehat{\Gamma} = \Psi\Sigma^{-1}, \quad \widehat{\mu} \triangleq \widehat{x}_{1|N}. \tag{58}$$

*If* $\Theta$ *defined by* (57) *is such that* $\widehat{\beta}$ *lies within* $\Theta_1$*, then for any fixed* $\eta \in \Theta_2$*, the point (58) is the unique maximiser*

$$\widehat{\beta} = \arg\max_{\theta\in\Theta_1} \mathcal{Q}\left(\left[\begin{array}{c}\beta\\ \eta\end{array}\right], \widehat{\theta}_k\right). \tag{59}$$

*Furthermore,* $\widehat{\eta}$ *given by*

$$\widehat{\eta} \triangleq \left[\mathrm{vec}\left\{\widehat{\Pi}\right\}^T, \mathrm{vec}\left\{\widehat{P}_1\right\}^T\right]^T, \quad \widehat{\Pi} = \mathcal{R}_{22}^{T/2}\mathcal{R}_{22}^{1/2}, \tag{60}$$

$$\widehat{P}_1 \triangleq P_{1|N}^{1/2} P_{1|N}^{T/2}, \tag{61}$$

*forms a stationary point of* $\mathcal{Q}(\cdot, \widehat{\theta}_k)$ *with respect to* $\eta$*. Here,* $P_{1|N}^{1/2}$ *is defined by* (45) *and* $\mathcal{R}_{22}^{1/2}$ *is defined by the Cholesky factorisation (see Algorithm 4.2.4 of [11])*

$$\left[\begin{array}{cc}\Sigma & \Psi^T\\ \Psi & \Phi\end{array}\right] = \left[\begin{array}{cc}\mathcal{R}_{11} & \mathcal{R}_{12}\\ \varnothing & \mathcal{R}_{22}\end{array}\right]^T \left[\begin{array}{cc}\mathcal{R}_{11} & \mathcal{R}_{12}\\ \varnothing & \mathcal{R}_{22}\end{array}\right]. \tag{62}$$

*Note that the right hand side of the expression for* $\Pi$ *in (60) is*

$$\Pi = \Phi - \Psi\Sigma^{-1}\Psi^T \tag{63}$$

*realised in a numerically robust fashion that ensures essential properties of symmetry and non negative-definiteness of the result.*

**Proof 3** *This follows by an argument identical to that used to establish Lemma 4.3 of [10].*

### 5.3 A Summary of the Algorithm

The preceding derivations are now summarised in the interests of clearly defining the new algorithm proposed here.

**Algorithm 5.1 (EM Algorithm for Bilinear Systems)**

1. *Initialise the algorithm by choosing a parameter vector* $\widehat{\theta}_0$*.*

2. **(E-Step)** *Employ the square-root implementation of the modified Kalman smoother presented in Lemma 5.2 in conjunction with the parameter estimate* $\widehat{\theta}_k$ *to calculate the matrices* $\Phi, \Psi$ *and* $\Sigma$ *as shown in equations* (29) *through* (31)*.*

3. **(M-Step)** *In order to choose an updated parameter estimate* $\widehat{\theta}_{k+1}$*, select* $\Pi$*,* $\Gamma$*,* $\mu$ *and* $P_1$ *according to equations* (58)*,* (60) *and* (61)*.*

4. *If the algorithm has converged, terminate, otherwise return to step 2.*

Regarding step 4, obvious strategies for gauging convergence involve copying those developed for gradient based search [5, 19]. In particular, this chapter suggests a strategy of termination when relative likelihood increase on an iteration drops below a predetermined threshold. This approach is supported by empirical evidence. In the authors experience, once the rate of increase in the likelihood drops it rarely returns to higher levels.

# 6 Simulation Study

In this section we demonstrate the utility of the EM method detailed above for the estimation of LPV systems from input and output measurements.

## 6.1 SISO First Order Example

In order to gain some confidence that EM method is providing accurate estimates, we considered a first order SISO LPV system. More precisely, consider an LPV model in the form of (1) where $n = 1$, $m = 1$ and $p = 1$. Further $n_\alpha = n_\beta = 2$ delivering the LPV system

$$x_{t+1} = \alpha_t(1)A(1)x_t + \alpha_t(2)A(2)x_t + \beta_t(1)B(1)u_t + \beta_t(2)B(2)u_t + w_t \quad (64)$$

$$y_t = \alpha_t(1)C(1)x_t + \alpha_t(2)C(2)x_t + \beta_t(1)D(1)u_t + \beta_t(2)D(2)u_t + e_t \quad (65)$$

For the purposes of this simulation, $N = 1000$ samples of the output $y_t$ were generated using the following parameter values and signals. The input was generated as a square wave and is shown in the top plot of Figure 1. The scheduling parameters $\alpha_t = \beta_t$ were generated as sinusoidal waves that vary between $[0, 1]$ and so that $\alpha_t(1)$ and $\alpha_t(2)$ are $\pi$-radians out of phase,

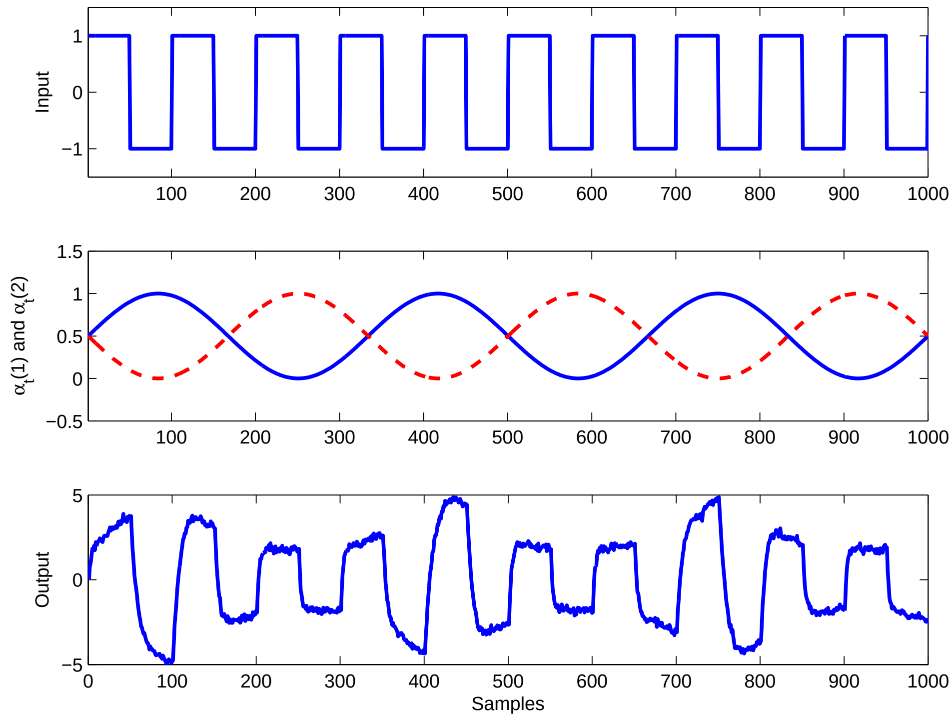


Figure 1: Top: input signal; Middle: $\alpha_t(1)$ (blue solid) and $\alpha_t(2)$ (red dashed) signals; Bottom: output signal.

which is shown in the middle plot in Figure 1. The noise signals $w_t,\ e_t$ were generated as an i.i.d. normally distributed signal via

$$\begin{bmatrix} w_t \\ e_t \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} Q & S \\ S & R \end{bmatrix}\right) \tag{66}$$

The "true" system parameters were chosen as

$$\begin{bmatrix} A(1) & A(2) & B(1) & B(2) \\ C(1) & C(2) & D(1) & D(2) \end{bmatrix} = \begin{bmatrix} 0.9 & 0.6 & 1 & 1 \\ 0.5 & 0.7 & 0 & 0 \end{bmatrix} \tag{67}$$

with

$$\begin{bmatrix} Q, & S, & R \end{bmatrix} = \begin{bmatrix} 0.01, & 0, & 0.01 \end{bmatrix} \tag{68}$$

The above can be arranged into a parameter vector $\theta$

$$\theta^T = [A(1),\ A(2),\ B(1),\ B(2),\ C(1),\ C(2),\ D(1),\ D(2),\ Q,\ S,\ R]$$

The output for one noise realization is shown in the bottom plot of Figure 1.

In order to examine the accuracy of the EM estimates $\widehat{\theta}$ of $\theta$ we performed a Monte-Carlo simulation based on $M = 1000$ runs. In each run only the noise realisation $w_t$ and $e_t$ were changed. Moreover, we were also interested in examining the robustness of the EM method to poor initial estimates of $\theta$. Therefore, the initial estimate $\widehat{\theta}_0$ was chosen as

$$\widehat{\theta}_0^T = \left[10^{-5}[0.9,\ 0.6,\ 1,\ 1,\ 0.5,\ 0.7],\ 1,\ 0,\ 1\right], \tag{69}$$

so that the system matrices are almost zero and the noise covariance is 100 times larger than the true values.

Recall that the state-space model (1) is not uniquely parametrized. In general this presents a difficulty when comparing true system values, like those in (67), with estimated ones, since the estimates are likely to correspond to a different state coordinate system. However, for the first order example considered here, any similarity transformation of the state will cancel for the $A(1,2)$ terms and will also cancel when considering the product of $B(1)$ with $C(1)$ and $B(2)$ with $C(2)$.

Based on the above Figure 2 presents a histogram of the estimated $\widehat{A}(1,2)$, $\widehat{B}(1)\widehat{C}(1)$, $\widehat{B}(2)\widehat{C}(2)$ and $\widehat{D}(1,2)$ parameters for all the runs. From the figure, it appears that estimates are accurate, even though the EM method is initialised far from the true values.

### 6.2 10'th Order MIMO Example

Inspired by the good performance of the EM algorithm above we applied it to a more challenging situation where the system state was increased to $n = 10$ and the number of system inputs and outputs were chosen as $m = p = 2$. Again $n_\alpha = n_\beta = 2$, but this time $\alpha_t \neq \beta_t$.

In particular, the LPV system was formed by creating two random LTI-systems that were both ensured to be stable. This results in the matrices $A(1,2)$, $B(1,2)$, $C(1,2)$ and $D(1,2)$ to provide (recall from (28))

$$\Gamma = \begin{bmatrix} A(1) & A(2) & B(1) & B(2) \\ C(1) & C(2) & D(1) & D(2) \end{bmatrix} \tag{70}$$

The noise terms were chosen according to (recall from (1c))

$$\begin{bmatrix} w_t \\ e_t \end{bmatrix} \sim \mathcal{N}(0,\ \Pi) \tag{71}$$

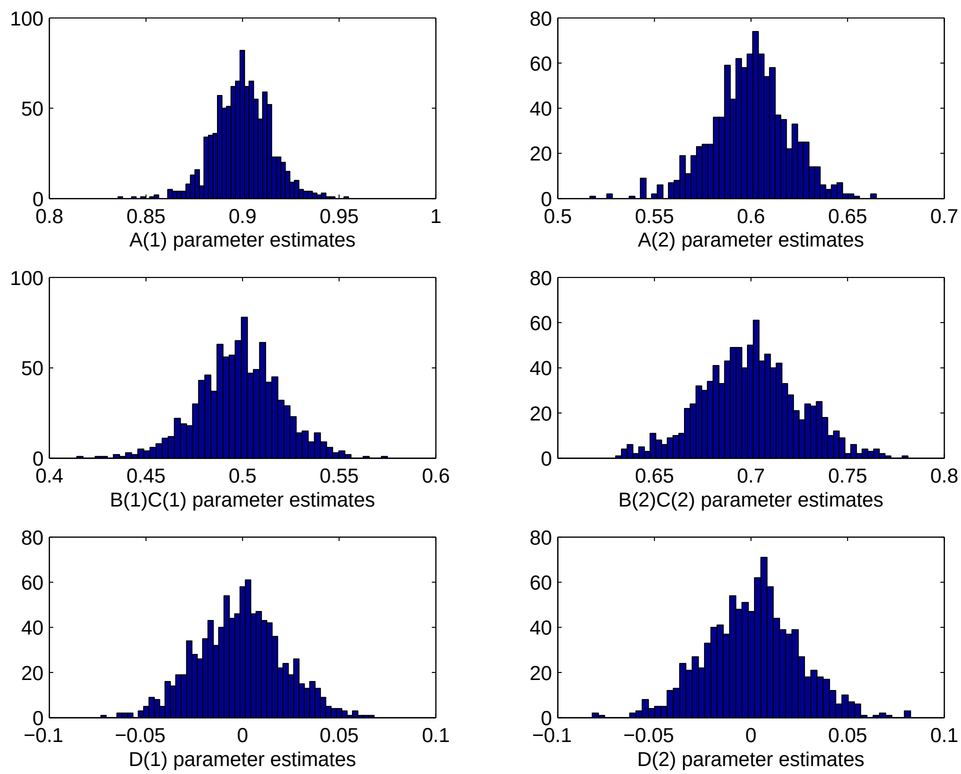


Figure 2: Histograms of estimated parameter values.

with

$$Q = 10^{-2} I_n, \qquad S = 0, \qquad R = 10^{-2} I_p. \tag{72}$$

The system was simulated using $N = 1000$ samples of the input (shown in the top plot of Figure 3) together with the scheduling sginals $\alpha_t(1,2)$ and $\beta_t(1,2)$ (shown in the middle and bottom plots of Figure 3, respectively).

This resulted in $N$ samples of the output $y_t$ according to (1). Based on the input and output measurements, the EM algorithm 5.1 was used to provide an estimate. The algorithm requires an initial estimate of the parameters and in an attempt to again demonstrate the robustness of the EM method to poor initial estimates, we chose

$$\widehat{\theta}_0^T = [\mathrm{vec}\left\{10^{-5}\Gamma\right\}^T, \ \mathrm{vec}\left\{100\Pi\right\}^T] \tag{73}$$

to ensure that the initial guess is far from the true values.

Unlike the previous example, it is more difficult to present a comparison of the true and estimated parameter values in the current case. Instead we have adopted the standard approach of providing a comparison of the true system output and the predicted one obtained from the EM estimate in Figure 4. This figure shows a close match between the two signals. Furthermore, we performed a standard whiteness test on the prediction errors as shown in Fifgure 5. At least according to the confidence bounds, the errors are uncorrelated. Additionally, and as a final test of the model we computed the cross-correlation between the prediction error and the input, shown in Figure 6. Again, according to the confidence intervals, the error is uncorrelated with the input signal.

# 7 Conclusion

In this chapter the expectation-maximisation (EM) algorithm has been presented and examined for the purpose of finding Maximum-Likelihood estimates of state-space linear parameter varying models.

Advantages of the EM method for LPV model estimation are that it is relatively straightforward to implement; it appears to be robust to bad initial parameter estimates; it scales linearly with the number of data points $N$; and, it straighforwardly handles possibly high-order MIMO LPV systems. A disadvantage is that the EM method does not straightforwardly handle structured LPV systems.

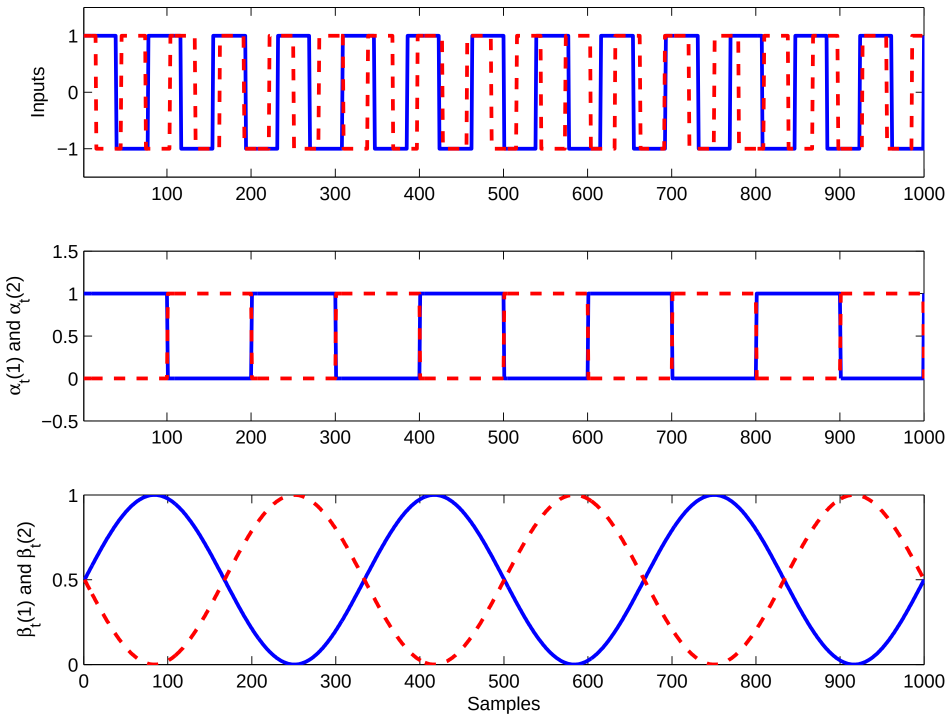


Figure 3: Top: input signals ; Middle: $\alpha_t(1)$ (blue solid) and $\alpha_t(2)$ (red dashed) signals; Bottom: $\beta_t(1)$ (blue solid) and $\beta_t(2)$ (red dashed) signals.

# References


[1] K.J. Åström. Maximum likelihood and prediction error methods. *Automatica*, 16:551–574, 1980.

[2] Gary J. Balas. Linear parameter-varying control and its application to a turbofan engine. *International Journal of Robust and Nonlinear Control*, 12(9):763–796, 2002.

[3] Bassam Bamieh and Laura Giarre. Identification of linear parameter varying models. *International Journal of Robust and Nonlinear Control*, (12):841–853, 2002.

[4] Dennis S. Bernstein. *Matrix Mathematics*. Princeton University Press, 2005.

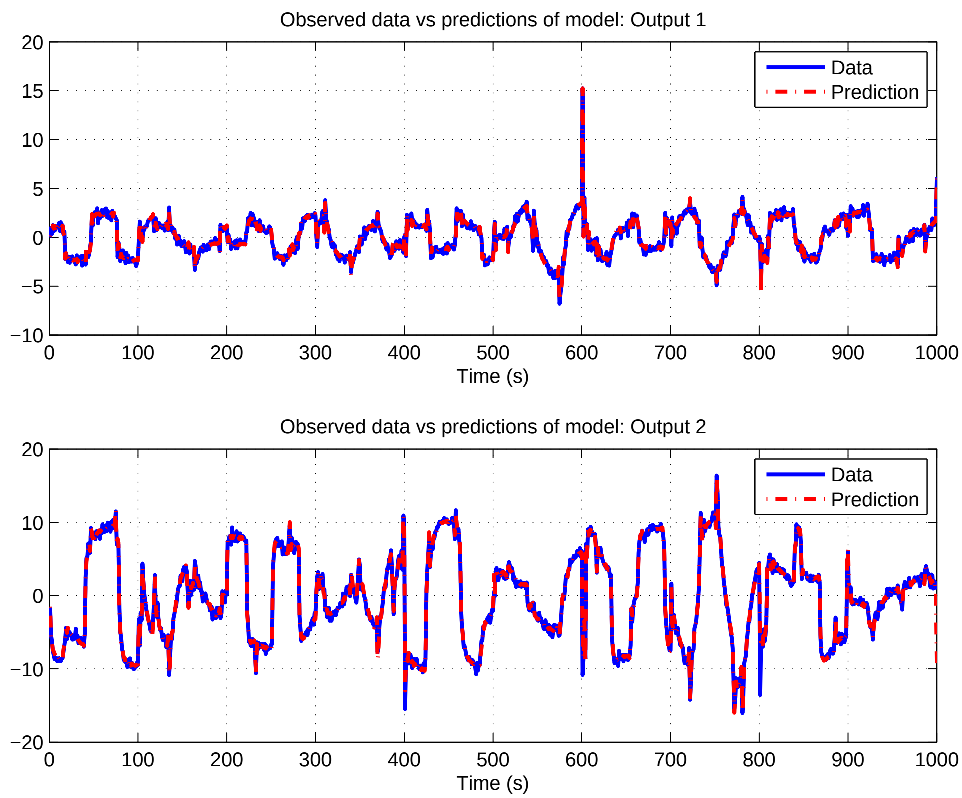


Figure 4: Predicted (red-dashed) and actual (blue-solid) output responses.

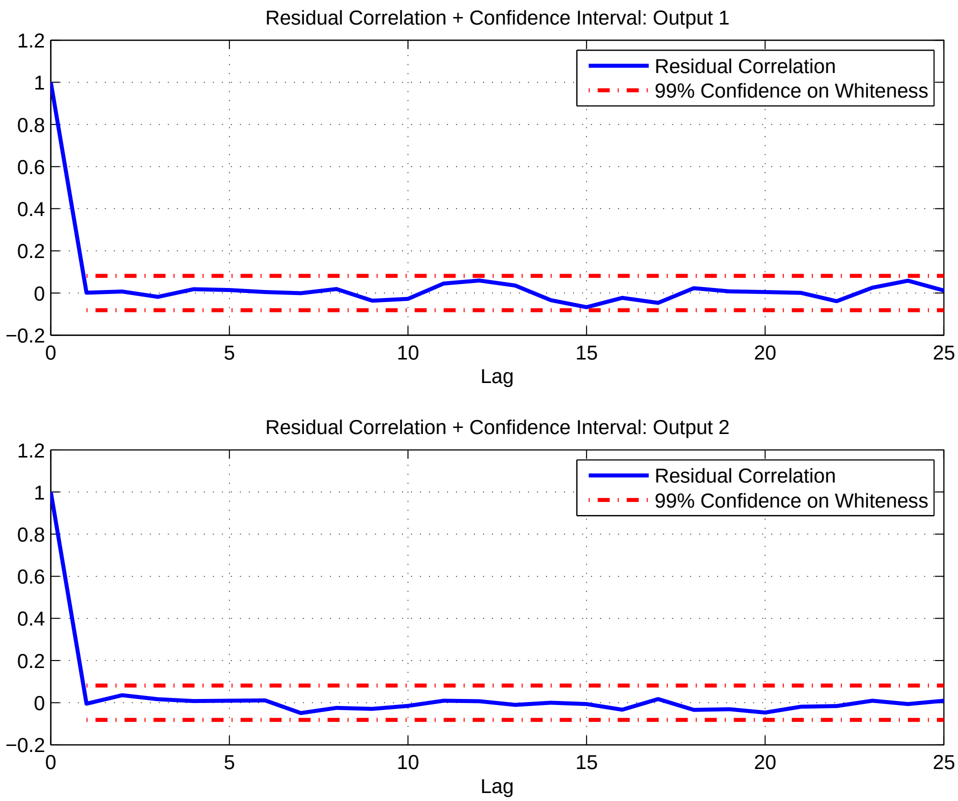


Figure 5: Autocorrelation of residual errors.

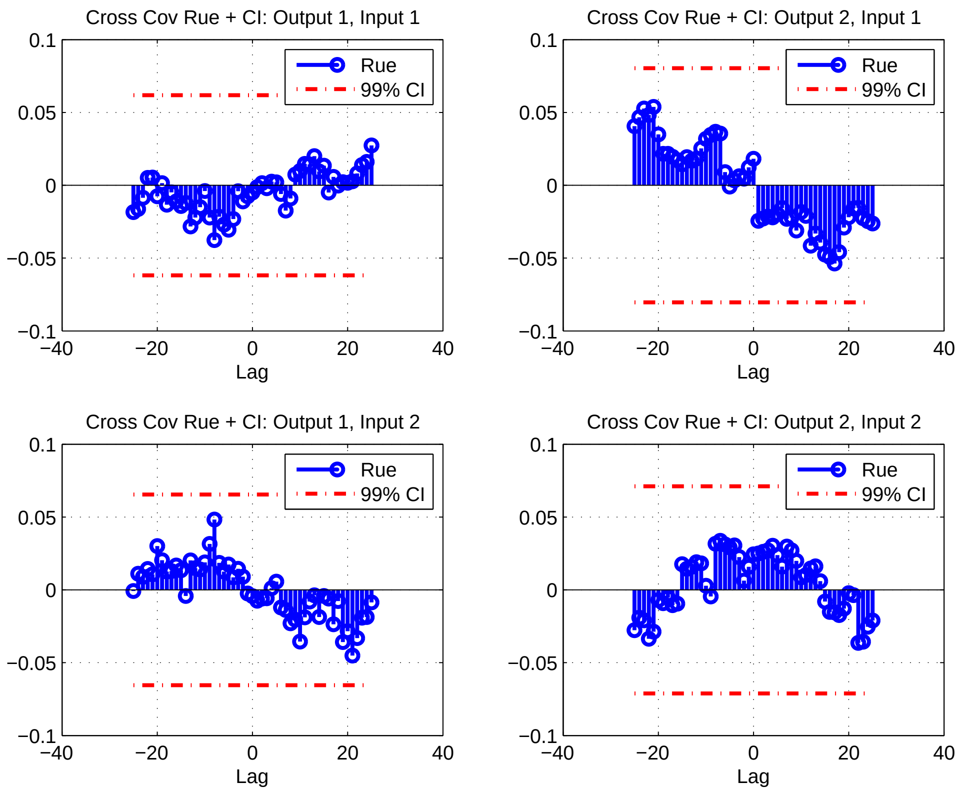


Figure 6: Cross-correlation of residual error with input.

[5] J.E. Dennis and Robert B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall, 1983.

[6] F. Felici, Jan-Willem van Wingerden, and Michel Verhaegen. Subspace identification of mimo lpv systems using a periodic scheduling sequence. *Automatica*, 43(10):1684–1697, 2007.

[7] A. Fujimori and Lennart Ljung. Model identification of linear parameter varying aircraft systems. *Proceedings of the Institute of Mechanical Engineers, Part G Journal of Aerospace Engineering*, 220(G4):337–346, 2006.

[8] L. Giarre, D. Baruso, P. Falugi, and B. Bamieh. Lpv model identification for gain scheduling control:An application to rotating stall and surge control problems. *Control Engineering Practice*, 14(4):351–361, 2006.

[9] Stuart Gibson and Brett Ninness. Robust maximum-likelihood estimation of multivariable dynamic systems. *Automatica*, 41(10):1667–1682, 2005.

[10] Stuart Gibson, Adrian Wills, and Brett Ninness. Maximum-likelihood parameter estimation of bilinear systems. *IEEE Transactions on Automatic Control*, 50(10):1581–1596, 2005.

[11] G. Golub and C. Van Loan. *Matrix Computations.* Johns Hopkins University Press, 1989.

[12] G.C. Goodwin and A. Feuer. Estimation with missing data. *Mathematical and Computer Modelling of Dynamical Systems*, 5(3):220–244, 1999.

[13] R. B. Gopaluni. A particle filter approach to identification of nonlinear processes under missing observations. *The Canadian Journal of Chemical Engineering*, 86(6):1081–1092, December 2008.

[14] Alf J. Isaksson. Identification of ARX-models subject to missing data. *IEEE Trans. Automat. Control*, 38(5):813–819, 1993.

[15] Vincent Laurain, Marion Gilson, Roland Tóth, and Hugues Garnier. Refined instrumental variable methods for identification of lpv box–jenkins models. *Automatica*, pages 959–967, 2010.

[16] L. H. Lee and K. Poolla. Identification of linear parameter-varying systems using nonlinear programming. *Journal of Dynamic Systems, Management, and Control*, 121:71–78, March 1999.

[17] L.Ljung. Convergence analysis of parametric identification methods. *IEEE Transactions on Automatic Control*, AC-23(5):770–783, 1978.

[18] A. Marcos and G.J. Balas. Development of linear parameter varying models for aircraft. *Journal of Guidance, Control and Dynamics*, 27(2):218–228, 2004.

[19] Jorge Nocedal and Stephen Wright. *Numerical Optimization*. Springer-Verlag, New York, 1999.

[20] W. Rugh and J. Shamma. Research on gain scheduling. *Automatica*, 36(10):1401–1425, 2000.

[21] Jan-Willem van Wingerden and Michel Verhaegen. Subspace identification of bilinear and LPV systems for open- and closed-lopp data. *Automatica*, 45:372–381, 2009.

[22] V. Verdult and M. Verhaegen. Subspace identification of multivariable linear parameter-varying systems. *Automatica*, 38(5):805–814, 2002.

[23] Vincent Verdult, Niek Bergboer, and Michel Verhaegen. Identification of fully parametrized linear and nonlinear state-space systems by projected gradient search. In *Proceedings of the 13th IFAC Symposium on System Identification, Rotterdam*, pages WeP07–3, 2003.

[24] Vincent Verdult, Marco Lovera, and Michel Verhaegen. Identification of linear parameter varying state space models with application to helicopter rotor dynamics. *International Journal of Control*, 77(13):1149–1159, 2004.

[25] Vincent Verdult and Michel Verhaegen. Kernel methods for subspace identification of multivariable lpv and bilinear systems. *Automatica*, 41(9):1557–1565, 2005.

[26] A.G. Wills, B. Ninness, and S.H. Gibson. Maximum likelihood estimation of state space models from frequency domain data. *IEEE Transactions on Automatic Control*, 54(1):19–33, 2009.