Technical Report PC-03-0808

Analysis of Emerson's MMI Estimation Algorithm

João P. Hespanha   Dale E. Seborg

University of California, Santa Barbara

August 8, 2003

Abstract

In this report we analyze Emerson's Multiple Model Interpolation (MMI) algorithm for parameter estimation and compare it with standard least-squares estimators. In certain cases these two algorithms provide the same estimate. We start be performing the analysis for the single-gain case and then expand the results for more general estimation problems.

1 Single-gain estimation

We consider the discrete-time, single-input, single-output, process model:

$$y(k) = K_p u(k-\tau) + n(k), \qquad \forall k \in \{1,2,3,\dots\} \tag{1}$$

where y denotes the output, u the control input, n measurement noise, τ a fixed known delay, and K_p the process steady-state gain. This model considers the effect of measurement noise, which was ignored in previous reports.

We assume that we have available a set of data

$$\Big\{ y(k), u(k-\tau) : k = 1,2,\dots,L \Big\} \tag{2}$$

collected over a time-window of length L. By stacking the inputs and outputs as vectors of length L

$$Y := \begin{bmatrix} y(1) \\ y(2) \\ \vdots \\ y(L) \end{bmatrix}, \qquad U := \begin{bmatrix} u(1-\tau) \\ u(2-\tau) \\ \vdots \\ u(L-\tau) \end{bmatrix}, \tag{3}$$

we can write the process model as

$$Y = K_p U + N, \tag{4}$$

where

$$N := \begin{bmatrix} n(1) \\ n(2) \\ \vdots \\ n(L) \end{bmatrix}, \tag{5}$$

is a vector of measurement noise. The value of N is not available to estimate K_p.

The data-set is processed multiple times by examining its fits with respect to a bank of M models that varies from iteration to iteration. The bank of models used in the ith iteration is given by

$$Y_m^j(i) = K_m^j(i) U, \qquad \forall j \in \{1,2,\dots,M\}, \tag{6}$$

where $Y_m^j(i)$ denotes the estimate of Y based on the jth model during the ith iteration. The corresponding prediction error is given by

$$E_m^j(i) := Y_m^j(i) - Y, \qquad \forall j \in \{1,2,\dots,M\}. \tag{7}$$

The *sum-of-squares error* $\mathrm{SSE}^j(i)$ for the jth model during the ith iteration is given by

$$\mathrm{SSE}^j(i) := \|E_m^j(i)\|^2 = (Y_m^j(i) - Y)'(Y_m^j(i) - Y), \tag{8}$$

and we define the corresponding *performance index* $J^j(i)$ by

$$J^j(i) := \frac{1}{\mathrm{SSE}^j(i)}. \tag{9}$$

Based on these definitions, we construct a *multiple model interpolation (MMI) estimator* by

$$\hat{K}_p(i) := \frac{\sum_{j=1}^M K_m^j(i) J^j(i)}{\sum_{j=1}^M J^j(i)}. \tag{10}$$

1.1 Moving multiple-models interpolation

The *moving multiple-models interpolation (MMMI) estimation algorithm* is defined as follows:

1. Set $i = 0$
2. Compute the MMI estimate $\hat{K}_p(i)$ based on the family of model defined by the candidate gains $K_m^j(i)$
3. Compute a new family of models by computing a new set of candidate gains $K_m^j(i+1)$ *centered* at $\hat{K}_p(i)$:

$$K_m^j(i+1) = \hat{K}_p(i) + \Delta_K\big(j - \frac{M+1}{2}\big) \tag{11}$$

4. Increment i and goto to 2 until there is no significant change in $\hat{K}_p(i)$.

We are assuming here that the number of models M is odd and there is a constant spacing Δ_K among the model gains.

According to the MMMI algorithm we obtain

$$\hat{K}_p(i+1) = \frac{\sum_{j=1}^{M} K_m^j(i+1) J^j(i+1)}{\sum_{j=1}^{M} J^j(i+1)} \tag{12}$$

$$= \frac{\sum_{j=1}^{M} \Big(\hat{K}_p(i) + \Delta_K\big(j - \frac{M+1}{2}\big)\Big) J^j(i+1)}{\sum_{j=1}^{M} J^j(i+1)} \tag{13}$$

$$= \hat{K}_p(i) + \Delta_K \frac{\sum_{j=1}^{M} \big(j - \frac{M+1}{2}\big) J^j(i+1)}{\sum_{j=1}^{M} J^j(i+1)} \tag{14}$$

where the performance indexes $J^j(i+1)$ are given by (9). Using (4), (6), and (8) we can write the corresponding sum-of-squares errors for the jth model during the $(i+1)$th iteration as

$$\mathrm{SSE}^j(i+1) = (K_m^j(i+1)U' - K_pU' - N')(K_m^j(i+1)U - K_pU - N) \tag{15}$$

$$= (K_m^j(i+1) - K_p)^2\|U\|^2 - 2(K_m^j(i+1) - K_p)U'N + \|N\|^2 \tag{16}$$

$$= \big(\tilde{K}_p(i) + \Delta_K\big(j - \frac{M+1}{2}\big)\big)^2\|U\|^2 - 2(\tilde{K}_p(i) + \Delta_K\big(j - \frac{M+1}{2}\big))U'N + \|N\|^2 \tag{17}$$

where $\tilde{K}_p(i) := \hat{K}_p(i) - K_p$ denotes the estimation error at the ith iteration.

1.2 Equilibrium

Assume now that the MMMI algorithm converges to some value $\hat{K}_p$. Since $\hat{K}_p$ must be a fixed-point of (14), we conclude that at equilibrium

$$\Delta_K \frac{\sum_{j=1}^{M} \big(j - \frac{M+1}{2}\big) J^j}{\sum_{j=1}^{M} J^j} = 0 \quad \Leftrightarrow \quad \sum_{j=1}^{M} \big(j - \frac{M+1}{2}\big) J^j = 0, \tag{18}$$

where J^j denote the asymptotic value of $J^j(i)$ as $i \to \infty$. For the **3 moving models case** ($M = 3$) we simply have

$$\sum_{j=1}^{3} \big(j - 2\big) J^j = 0 \quad \Leftrightarrow \quad J^3 - J^1 = 0, \tag{19}$$

which, because of (9), is further equivalent to

$$\mathrm{SSE}^3 = \mathrm{SSE}^1 \tag{20}$$

where SSE^j denotes the asymptotic value of $\mathrm{SSE}^j(i)$ as $i \to \infty$. Using (17) we conclude that (20) is equivalent to

$$\big(\tilde{K}_p + \Delta_K\big)^2 \|U\|^2 - 2(\tilde{K}_p + \Delta_K)U'N + \|N\|^2 = \big(\tilde{K}_p - \Delta_K\big)^2 \|U\|^2 - 2(\tilde{K}_p - \Delta_K)U'N + \|N\|^2, \tag{21}$$

where $\tilde{K}_p$ denotes the asymptotic value of the estimation error $\tilde{K}_p(i)$ as $i \to \infty$. Equation (21) can further be simplified to

$$\tilde{K}_p \|U\|^2 - U'N = 0, \tag{22}$$

from which we conclude that

$$\tilde{K}_p = \frac{U'N}{\|U\|^2}, \tag{23}$$

as long as the input signal U is not identically zero. The equilibrium parameter estimate is therefore given by

$$\hat{K}_p = K_p + \frac{U'N}{\|U\|^2}. \tag{24}$$

It turns out that this is precisely the least-square estimate for the original data in (2). To verify this note that the sum-of-squares error for an arbitrary gain K is given by

$$\mathrm{SSE}(K) := (KU - Y)'(KU - Y) = (KU - K_pU - N)'(KU - K_pU - N). \tag{25}$$

This is minimized by finding the value of K for which

$$\frac{\partial\, \mathrm{SSE}(K)}{\partial K} = 0 \quad \Leftrightarrow \quad 2U'(KU - K_pU - N) = 0, \tag{26}$$

giving the following sum-of-squares estimate

$$K = K_p + \frac{U'N}{\|U\|^2}. \tag{27}$$

Comparing (24) with (27) we concludes the following:

Lemma 1 (Equilibrium). *With **3 moving models** ($M = 3$) and a not identically zero input signal, the unique equilibrium point of the MMMI algorithm is the least-squares estimate of the gain parameter.*

1.3 Convergence

For **3 moving models** (14) can be written as

$$\begin{aligned}
\hat{K}_p(i+1) &= \hat{K}_p(i) + \Delta_K \frac{J^3(i+1) - J^1(i+1)}{\sum_{j=1}^3 J^j(i+1)} && (28)\\
&= \hat{K}_p(i) + \Delta_K \frac{\frac{1}{\mathrm{SSE}^3(i+1)} - \frac{1}{\mathrm{SSE}^1(i+1)}}{\sum_{j=1}^3 J^j(i+1)} && (29)\\
&= \hat{K}_p(i) + \Delta_K \gamma(i) \big(\mathrm{SSE}^1(i+1) - \mathrm{SSE}^3(i+1)\big) && (30)
\end{aligned}$$

where

$$\gamma(i) := \frac{1}{\mathrm{SSE}^1(i+1)\,\mathrm{SSE}^3(i+1)\sum_{j=1}^{3} J^j(i+1)}. \tag{31}$$

Subtracting K_p from both side of (30) we obtain the following recursion for the estimation error.

$$\tilde{K}_p(i+1) = \tilde{K}_p(i) + \Delta_K \gamma(i)\big(\,\mathrm{SSE}^1(i+1) - \mathrm{SSE}^3(i+1)\big) \tag{32}$$

On the other hand, from (17) we conclude that

$$\mathrm{SSE}^1(i+1) - \mathrm{SSE}^3(i+1) = -4\Delta_K\big(\|U\|^2\tilde{K}_p(i) - U'N\big) \tag{33}$$

Defining $v(i) := \|U\|^2\tilde{K}_p(i) - U'N$, we conclude from (32) and (33) that

$$\begin{aligned} v(i+1) &= \|U\|^2\tilde{K}_p(i+1) - U'N & (34)\\ &= \|U\|^2\tilde{K}_p(i) - 4\|U\|^2\Delta_K^2\gamma(i)v(i) - U'N & (35)\\ &= \big(1 - 4\|U\|^2\Delta_K^2\gamma(i)\big)\,v(i) & (36)\end{aligned}$$

This shows that $v(i)$ is monotone and bounded between $v(0)$ and 0 and therefore convergent. Assuming that the input U is not identically zero, this means that $\tilde{K}_p$ is also bounded as well as all remaining signals including the SSE^j and J^j. We thus conclude that $\gamma(i)$ will not converge to zero and therefore $1 - 4\|U\|^2\Delta_K^2\gamma(i)$ is bounded away from 1. From this it follows that $v(i)$ actually converges to zero and therefore

$$\hat{K}_p(i) \to K_p + \frac{U'N}{\|U\|^2}. \tag{37}$$

The convergence rate will be exponential. The following can be stated

Lemma 2 (Convergence). *With **3 moving models** ($M = 3$) and a not identically zero input signal, the MMMI algorithm converges exponentially fast to the least-squares estimate of the gain parameter.*

2 General case

We consider a general SISO ARX model for the process, which can be written as

$$y(k) = \varphi(k)'c(\theta_p) + n(k), \qquad \forall k \in \{1,2,3,\dots\} \tag{38}$$

where $\varphi(k)$ denotes the *regression vector* defined by

$$\varphi(k) := \begin{bmatrix} -y(k-1) & -y(k-2) & \cdots & -y(k-n_y) & u(k-1) & \cdots u(k-n_u)\end{bmatrix}' \in \mathbb{R}^{n_y+n_u}, \tag{39}$$

and $c(\theta_p)$ a column vector of coefficients that depends on some unknown parameter vector θ_p that belongs to a parameter set $\mathcal{P} \subset \mathbb{R}^n$.

We assume that we have available a set of data

$$\Big\{y(k), \varphi(k) : k = 1,2,\dots,L\Big\} \tag{40}$$

collected over a time-window of length L. By stacking the outputs and regression vectors as follows

$$Y := \begin{bmatrix} y(1)\\ y(2)\\ \vdots\\ y(L)\end{bmatrix} \in \mathbb{R}^L, \qquad \Phi := \begin{bmatrix} \varphi(1)'\\ \varphi(2)'\\ \vdots\\ \varphi(L)'\end{bmatrix} \in \mathbb{R}^{L\times(n_y+n_u)}, \tag{41}$$

we can write the process model as

$$Y = \Phi c(\theta_p) + N, \tag{42}$$

where

$$N := \begin{bmatrix} n(1) \\ n(2) \\ \vdots \\ n(L) \end{bmatrix}, \tag{43}$$

is a vector with measurement noise. The value of N is not available to estimate p.

The data-set is processed multiple times by examining its fits with respect to a finite bank of models that varies from iteration to iteration. We denote by

$$\mathcal{M}(i) := \{\theta_m^1(i), \theta_m^2(i), \ldots, \theta_m^M(i)\} \subset \mathcal{P} \tag{44}$$

the values for the parameters for the bank of models used in the ith iteration. The estimate $Y_m^j(i)$ of Y based on the jth model during the ith iteration is defined by

$$Y_m^j(i) = \Phi\, c(\theta_m^j(i)), \qquad \forall j \in \{1, 2, \ldots, M\}. \tag{45}$$

and the corresponding prediction error is given by

$$E_m^j(i) := Y_m^j(i) - Y, \qquad \forall j \in \{1, 2, \ldots, M\}. \tag{46}$$

The *multiple model interpolation (MMI) estimator* is now given by

$$\hat{\theta}_p(i) := \frac{\sum_{j=1}^M J^j(i)\theta_m^j(i)}{\sum_{j=1}^M J^j(i)}. \tag{47}$$

where $J^j(i)$ denotes the *performance index* for the jth model during the ith iteration, defined by

$$J^j(i) := \frac{1}{\text{SSE}^j(i)}, \tag{48}$$

and $\text{SSE}^j(i)$ denotes the *sum-of-squares error* $\text{SSE}^j(i)$ given by

$$\text{SSE}^j(i) := \|E_m^j(i)\|^2 = (Y_m^j(i) - Y)'(Y_m^j(i) - Y). \tag{49}$$

Example 1. For a one-step delay system with unknown gain $\theta_p \in [1, 10]$, we have

$$\varphi(k) := u(k-1), \qquad c(\theta_p) := \theta_p, \qquad \mathcal{P} := [1, 10] \tag{50}$$

leading to a model similar to the one considered in Section 1 with $\tau = 1$:

$$y(k) = \theta_p u(k-1) + n(k), \qquad p \in [1, 10]. \tag{51}$$

Example 2. For a system with unknown gain $\theta_1 \in [1, 10]$ and unknown delay $\theta_2 \in \{1, 2, 3\}$, we would have

$$\varphi(k) := \begin{bmatrix} u(k-1) & u(k-2) & u(k-3) \end{bmatrix}', \quad c(\theta_1, \theta_2) := \begin{cases} \begin{bmatrix} \theta_1 & 0 & 0 \end{bmatrix}' & \theta_2 = 1 \\ \begin{bmatrix} 0 & \theta_1 & 0 \end{bmatrix}' & \theta_2 = 2 \\ \begin{bmatrix} 0 & 0 & \theta_1 \end{bmatrix}' & p_2 = 3 \end{cases}, \quad \mathcal{P} := [1, 10] \times \{1, 2, 3\} \tag{52}$$

leading to

$$y(k) = n(k) + \begin{cases} \theta_1 u(k-1) & \theta_2 = 1 \\ \theta_1 u(k-2) & \theta_2 = 2 \\ \theta_1 u(k-3) & \theta_2 = 3 \end{cases}, \qquad \theta_1 \in [1, 10]. \tag{53}$$

2.1 Moving multiple-models interpolation

The *moving multiple-models interpolation (MMMI) estimation algorithm* is defined as follows:

1. Set $i = 0$ (iteration index)
2. Set $\ell = 1$ (parameter index)
3. Compute the MMI estimate $\hat{\theta}_p(i)$ based on the family of model defined by the candidate gains $\theta^j_m(i)$
4. Compute a new family of models by computing a new set of candidate gains $\theta^j_m(i+1)$ *centered at the* ℓ*th parameter* in $\hat{\theta}_p(i)$:

$$\theta^j_m(i+1) = \hat{\theta}_p(i) + \Delta_\ell\big(j - \frac{M+1}{2}\big)e_\ell, \tag{54}$$

   where e_ℓ denotes the ℓth element of the canonical basis of $\mathbb{R}^n$

5. Increment i and ℓ (modulo n) and go to 3 until there is no significant change in $\hat{\theta}_p(i)$.

We are assuming here that the number of models M is odd and there is a constant spacing Δ_ℓ among the model values for the ℓth parameter.

According to the MMMI algorithm we obtain

$$\hat{\theta}_p(i+1) = \frac{\sum_{j=1}^M \theta^j_m(i+1)J^j(i+1)}{\sum_{j=1}^M J^j(i+1)} \tag{55}$$

$$= \frac{\sum_{j=1}^M \Big(\hat{\theta}_p(i) + \Delta_\ell\big(j - \frac{M+1}{2}\big)e_\ell\Big)J^j(i+1)}{\sum_{j=1}^M J^j(i+1)} \tag{56}$$

$$= \hat{\theta}_p(i) + \Delta_\ell \frac{\sum_{j=1}^M \big(j - \frac{M+1}{2}\big)J^j(i+1)}{\sum_{j=1}^M J^j(i+1)} e_\ell \tag{57}$$

where the performance indexes $J^j(i+1)$ are given by (48). Using (42), (45), and (49) we can write the corresponding sum-of-squares errors for the jth model during the $(i+1)$th iteration as

$$\mathrm{SSE}^j(i+1) = (\Phi\, c(\theta^j_m(i+1)) - \Phi c(\theta_p) - N)'(\Phi\, c(\theta^j_m(i+1)) - \Phi c(\theta_p) - N) \tag{58}$$

$$= \tilde{c}^j_m(i+1)'\Phi'\Phi\tilde{c}^j_m(i+1) - 2\tilde{c}^j_m(i+1)'\Phi' N + \|N\|^2. \tag{59}$$

where

$$\tilde{c}^j_m(i+1) := c(\theta^j_m(i+1)) - c(\theta_p). \tag{60}$$

When the parametrization $c(\cdot)$ is **affine**[1], we actually have

$$\tilde{c}^j_m(i+1) = J(\theta^j_m(i+1) - \theta_p) = J\tilde{\theta}_p(i) + \Delta_\ell\big(j - \frac{M+1}{2}\big)Je_\ell, \tag{61}$$

where $\tilde{\theta}_p(i) := \hat{\theta}_p(i) - \theta_p$ denotes the estimation error at the ith iteration and J the (constant) Jacobian matrix of the map $c(\cdot)$, i.e., $c(\theta) = J\theta + c_0$.

[1]This implicitly assumes that $\mathcal{P}$ is the whole $\mathbb{R}^n$

2.2 Equilibrium

Assume now that the MMMI algorithm converges to some value $\hat{\theta}_p$. Since $\hat{\theta}_p$ must be a fixed-point of (57) for every ℓ, we conclude that at equilibrium

$$\Delta_\ell \frac{\sum_{j=1}^{M} \big(j - \frac{M+1}{2}\big) J_\ell^j}{\sum_{j=1}^{M} J_\ell^j} e_\ell \quad \Leftrightarrow \quad \sum_{j=1}^{M} \big(j - \frac{M+1}{2}\big) J_\ell^j = 0, \qquad \forall \ell \tag{62}$$

where J_ℓ^j denote the asymptotic value of $J^j(i)$ for the parameter index ℓ as $i \to \infty$. For the **3 moving models case** ($M = 3$) we simply have

$$\sum_{j=1}^{3} \big(j - 2\big) J_\ell^j = 0 \quad \Leftrightarrow \quad J_\ell^3 - J_\ell^1 = 0, \qquad \forall \ell \tag{63}$$

which, because of (48), is further equivalent to

$$\mathrm{SSE}_\ell^3 = \mathrm{SSE}_\ell^1, \qquad \forall \ell \tag{64}$$

where SSE^j denotes the asymptotic value of $\mathrm{SSE}^j(i)$ for the parameter index ℓ as $i \to \infty$. Using (59) and (61) we conclude that for the **affine** $c(\cdot)$ case, (64) is equivalent to

$$(\tilde{\theta}_p + \Delta_\ell e_\ell)' J' \Phi' \Phi J (\tilde{\theta}_p + \Delta_\ell e_\ell) - 2(\tilde{\theta}_p + \Delta_\ell e_\ell)' J' \Phi' N + \|N\|^2 = \\ (\tilde{\theta}_p - \Delta_\ell e_\ell)' J \Phi' \Phi J (\tilde{\theta}_p - \Delta_\ell e_\ell) - 2(\tilde{\theta}_p - \Delta_\ell e_\ell)' J' \Phi' N + \|N\|^2, \qquad \forall \ell \tag{65}$$

where $\tilde{\theta}_p$ denotes the asymptotic value of the parameter estimation error $\tilde{\theta}_p(i)$ as $i \to \infty$. Equation (65) can further be simplified to

$$e_\ell'(J' \Phi' \Phi J \tilde{\theta}_p - J' \Phi' N) = 0, \quad \forall \ell \quad \Leftrightarrow \quad J' \Phi' \Phi J \tilde{\theta}_p - J' \Phi' N = 0. \tag{66}$$

from which we conclude that

$$\tilde{\theta}_p = (J' \Phi' \Phi J)^{-1} J' \Phi' N, \tag{67}$$

as long as $J' \Phi' \Phi J$ is nonsingular. The equilibrium parameter estimate is therefore given by

$$\hat{\theta}_p = \theta_p + (J' \Phi' \Phi J)^{-1} J' \Phi' N. \tag{68}$$

It turns out that this is precisely the least-square estimate. To verify this note that the sum-of-squares error for an arbitrary value θ of the parameter is given by

$$\mathrm{SSE}(\theta) = (\Phi J(\theta - \theta_p) - N)'(\Phi J(\theta - \theta_p) - N). \tag{69}$$

This is minimized by finding the value of θ for which

$$\frac{\partial\, \mathrm{SSE}(\theta)}{\partial \theta} = 0 \quad \Leftrightarrow \quad J' \Phi' (\Phi J(\theta - \theta_p) - N) = 0 \tag{70}$$

giving the following sum-of-squares estimate

$$\theta = \theta_p + (J' \Phi' \Phi J)^{-1} J' \Phi' N \tag{71}$$

Comparing (68) with (71) we concludes the following:

Lemma 3 (Equilibrium). *With **3 moving models** ($M = 3$), $c(\cdot)$ **affine**, and $J' \Phi' \Phi J$ is nonsingular, the unique equilibrium point of the MMMI algorithm is the least-squares estimate of the parameter θ.*