Moving Horizon Estimation (MHE)

James B. Rawlings

Department of Chemical and Biological Engineering
University of Wisconsin–Madison

Insitut für Systemtheorie und Regelungstechnik
Universität Stuttgart
Stuttgart, Germany
June 24, 2011
Outline

1. Full information estimation
2. Moving horizon estimation - zero prior weighting
3. Moving horizon estimation - nonzero prior weighting
4. Moving horizon estimation - constrained estimation
5. Moving horizon estimation - smoothing and filtering update
Now turn to the general problem of estimating the state of a noisy dynamic system given noisy measurements:

\[ x^+ = f(x, w) \]
\[ y = h(x) + v \]  \hspace{1cm} (1)

in which the process disturbance, \( w \), measurement disturbance, \( v \), and system initial state, \( x(0) \), are independent random variables with stationary probability densities.
Full information estimation will prove to have the best theoretical properties in terms of stability and optimality.

Unfortunately, it will also prove to be computationally intractable except for the simplest cases, such as a linear system model.

One method for practical estimator design therefore is to come as close as possible to the properties of full information estimation while maintaining a tractable online computation (MHE).
### Variables

Notation required to distinguish the system variables from the estimator variables:

<table>
<thead>
<tr>
<th></th>
<th>System variable</th>
<th>Decision variable</th>
<th>Optimal decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>state</td>
<td>$x$</td>
<td>$\chi$</td>
<td>$\hat{x}$</td>
</tr>
<tr>
<td>process disturbance</td>
<td>$w$</td>
<td>$\omega$</td>
<td>$\hat{w}$</td>
</tr>
<tr>
<td>measured output</td>
<td>$y$</td>
<td>$\eta$</td>
<td>$\hat{y}$</td>
</tr>
<tr>
<td>measurement disturbance</td>
<td>$\nu$</td>
<td>$\nu$</td>
<td>$\hat{\nu}$</td>
</tr>
</tbody>
</table>

The relationships between these variables:

\[
x^+ = f(x, w) \quad y = h(x) + \nu \\
\chi^+ = f(\chi, \omega) \quad y = h(\chi) + \nu \quad \eta = h(\chi) \\
\hat{x}^+ = f(\hat{x}, \hat{w}) \quad y = h(\hat{x}) + \hat{\nu} \quad \hat{y} = h(\hat{x})
\]
The full information objective function is

\[ V_T(\chi(0), \omega) = \ell_x(\chi(0) - \bar{x}_0) + \sum_{i=0}^{T-1} \ell_i(\omega(i), \nu(i)) \]  

subject to

\[ \chi^+ = f(\chi, \omega) \quad y = h(\chi) + \nu \]

in which \( T \) is the current time, \( y(i) \) is the measurement at time \( i \), and \( \bar{x}_0 \) is the prior information on the initial state.

The full information estimator is then defined as the solution to

\[ \min_{\chi(0), \omega} V_T(\chi(0), \omega) \]
What class of systems have a stable state estimator?

- Assume system observability?
  Too restrictive for even linear systems (recall the definition of detectability).
- We need a similar detectability definition for nonlinear systems – \textit{i-IOSS}:
What class of systems have a stable state estimator?

- Assume system observability?
  Too restrictive for even linear systems (recall the definition of detectability).
- We need a similar detectability definition for nonlinear systems – i-IOSS:

**Definition 1 (i-IOSS)**

The system $\dot{x} = f(x, w), \ y = h(x)$ is incrementally input/output-to-state stable (i-IOSS) if there exists some $\beta(\cdot) \in KL$ and $\gamma_1(\cdot), \gamma_2(\cdot) \in K$ such that for every two initial states $z_1$ and $z_2$, and any two disturbance sequences $w_1$ and $w_2$

\[
| x(k; z_1, w_1) - x(k; z_2, w_2) | \leq \beta(| z_1 - z_2 |, k) + \\
\gamma_1(\| w_1 - w_2 \|_{0:k-1}) + \gamma_2(\| y_{z_1,w_1} - y_{z_2,w_2} \|_{0:k})
\]
The notation \( x(k; x_0, w) \) denotes the solution to \( x^+ = f(x, w) \) satisfying initial condition \( x(0) = x_0 \) with disturbance sequence \( w = \{w(0), w(1), \ldots\} \).

The notation \( x(k; x_1, k_1, w) \) denotes the solution to \( x^+ = f(x, w) \) satisfying the condition \( x(k_1) = x_1 \) with disturbance sequence \( w = \{w(0), w(1), \ldots\} \).

One of the most important and useful implications of the i-IOSS property:

**Proposition 2 (Convergence of state under i-IOSS)**

If system \( x^+ = f(x, w), y = h(x) \) is i-IOSS, \( w_1(k) \to w_2(k) \) and \( y_1(k) \to y_2(k) \) as \( k \to \infty \), then

\[
\begin{align*}
x(k; z_1, w_1) & \to x(k; z_2, w_2) \quad \text{for all } z_1, z_2
\end{align*}
\]

(The proof of this proposition is discussed in Exercise 4.3.)
Convergent disturbances under i-IOSS

The class of disturbances \((w, v)\) affecting the system is defined as:

Assumption 3 (Convergent disturbances)

The sequence \((w(k), v(k))\) for \(k \in \mathbb{I}_{\geq 0}\) are bounded and converge to zero as \(k \to \infty\).

Remark 4 (Summable disturbances)

If the disturbances satisfy Assumption 26, then there exists a \(\mathcal{K}\)-function \(\gamma_w(\cdot)\) such that the disturbances are summable

\[
\sum_{i=0}^{\infty} \gamma_w \left( |(w(i), v(i))| \right) \text{ is bounded}
\]

See Sontag (1998, Proposition 7) for a statement and proof of this result.
Stage Cost

Given such class of disturbances, the estimator stage cost is chosen to satisfy the following property:

**Assumption 5 (Positive definite stage cost)**

The initial state cost and stage costs are continuous functions and satisfy the following inequalities for all $x \in \mathbb{R}^n$, $w \in \mathbb{R}^g$, and $v$ in $\mathbb{R}^p$

\[
\gamma_x(|x|) \leq \ell_x(x) \leq \gamma_x(|x|) \quad (4)
\]
\[
\gamma_w(||(w, v)||) \leq \ell_i(w, v) \leq \gamma_w(||(w, v)||) \quad i \geq 0 \quad (5)
\]

in which $\gamma_x, \gamma_w, \gamma_x, \gamma_w \in \mathcal{K}_\infty$ and $\gamma_w$ is defined in Remark 4.

Notice that if we change the class of disturbances affecting the system, we may also have to change the stage cost in the state estimator to satisfy $\ell_i(w, v) \leq \gamma_w(||(w, v)||)$ in (5).
Stability

Zero error system

First consider the zero estimate error solution for all $k \geq 0$ (initial state is equal to the estimator’s prior and zero disturbances). In this case, the optimal solution is:

$$\hat{x}(0|T) = x_0$$
$$\hat{w}(i|T) = 0 \text{ for all } 0 \leq i \leq T, T \geq 1$$
$$h(\hat{x}(i|T)) = y(i) \text{ for all } 0 \leq i \leq T, T \geq 1$$

The perturbation to this solution are: the system’s initial state (distance from $x_0$), and the process and measurement disturbances. We next define stability properties so that:

- **asymptotic stability** considers the case $x_0 \neq \overline{x}_0$ with zero disturbances.
- **robust stability** considers the case in which $(w(i), v(i)) \neq 0$. 
Global Asymptotic Stability

Definition 6 (Global asymptotic stability)

The estimate is based on the *noise-free* measurement $y = h(x(x_0, 0))$. The estimate is (nominally) globally asymptotically stable (GAS) if there exists a $\mathcal{KL}$-function $\beta(\cdot)$ such that for all $x_0, \bar{x}_0$ and $k \in \mathbb{I}_{\geq 0}$

$$|x(k; x_0, 0) - \hat{x}(k)| \leq \beta(|x_0 - \bar{x}_0|, k)$$

The standard definition of estimator stability for *linear systems* is consistent with Definition 6.
The estimate is based on the *noisy* measurement $y = h(x(x_0, w)) + v$. The estimate is robustly GAS if for all $x_0$ and $\bar{x}_0$, and $(w, v)$ satisfying Assumption 26, the following hold.

1. The estimate converges to the state; as $k \to \infty$

   $$\hat{x}(k) \to x(k; x_0, w)$$

2. For every $\varepsilon > 0$ there exists $\delta > 0$ such that

   $$\gamma_x(|x_0 - \bar{x}_0|) + \sum_{i=0}^{\infty} \gamma_w(||(w(i), v(i))||) \leq \delta$$

   (6)

   implies $|x(k; x_0, w) - \hat{x}(k)| \leq \varepsilon$ for all $k \in \mathbb{I}_{\geq 0}$. 

Stuttgart – June 2011
The first part of the definition ensures that converging disturbances lead to converging estimates.

The second part provides a bound on the transient estimate error given a bound on the disturbances.

Note that robust GAS implies GAS (see also Exercise 4.9).

With the pieces in place, we can state the main result of this section:

**Theorem 8 (Robust GAS of full information estimates)**

*Given an i-IOSS (detectable) system and measurement sequence generated by (1) with disturbances satisfying Assumption 26, then the full information estimate with stage cost satisfying Assumption 5 is robustly GAS.*
Robust GAS of full information estimates

Proof:
(a)
First we establish that the full information cost is bounded for all $T \geq 1$ including $T = \infty$. Consider a candidate set of decision variables

$$\chi(0) = x_0 \quad \omega(i) = w(i) \quad 0 \leq i \leq T - 1$$

The full information cost for this choice is

$$V_T(\chi(0), \omega) = \ell_x(x_0 - \bar{x}_0) + \sum_{i=0}^{T-1} \ell_i(w(i), v(i))$$

From Remark 4, the sum is bounded for all $T$ including the limit $T = \infty$. Therefore, let $V_\infty$ be an upper bound for the right-hand side. The optimal cost exists for all $T \geq 0$ because $V_T$ is a continuous function and goes to infinity as any of its arguments goes to infinity due to the lower bounds in Assumption 5.
Next we show that the optimal cost sequence converges:
Evaluate the cost at time $T - 1$ using the optimal solution from time $T$.
We have that

$$V_{T-1}(\hat{x}(0|T), \hat{w}_T) = V_T^0 - \ell_T(\hat{w}(T|T), \hat{v}(T|T))$$

Optimization at time $T - 1$ can only improve the cost giving

$$V_T^0 \geq V_{T-1}^0 + \ell_T(\hat{w}(T|T), \hat{v}(T|T))$$

and we see that the optimal sequence $\{V_T^0\}$ is nondecreasing and bounded
above by $V_\infty$. Therefore the sequence converges and the convergence
implies

$$\ell_T(\hat{w}(T|T), \hat{v}(T|T)) \to 0$$

as $T \to \infty$. The lower bound in (5) then gives that

$$\hat{v}(T) = y(T) - h(\hat{x}(T|T)) \to 0$$

and $\hat{w}(T|T) \to 0$ as $T \to \infty$. 
Robust GAS of full information estimates

Since the measurement satisfies $y = h(x) + v$, and $v(T)$ converges to zero, we have that

$$h(x(T)) - h(\hat{x}(T|T)) \to 0 \quad \hat{w}(T|T) \to 0 \quad T \to \infty$$

Because the system is i-IOSS, we have the following inequality for all $x_0$, $\hat{x}(0|k)$, $w$, $\hat{w}_k$, and $k \geq 0$,

$$|x(k; x_0, w) - x(k; \hat{x}(0|k), \hat{w}_k)| \leq \beta(|x_0 - \hat{x}(0|k)|, k) + \gamma_1(\|w - \hat{w}_k\|_{0:k-1}) + \gamma_2(\|h(x) - h(\hat{x}_k)\|_{0:k}) \quad (7)$$

Since $w(k)$ converges to zero, $w(k) - \hat{w}(k)$ converges to zero, and $h(x(k)) - h(\hat{x}(k))$ converges to zero.
From Proposition 2 we conclude that \[ |x(k; x_0, w) - x(k; \hat{x}(0|k), \hat{w}_k)| \]
converges to zero. Since the state estimate is \( \hat{x}(k) := x(k; \hat{x}(0|k), \hat{w}_k) \)
and the state is \( x(k) = x(k; x_0, w) \), we have that
\[ \hat{x}(k) \to x(k) \quad k \to \infty \]
and the estimate converges to the system state. This establishes part 1 of
the robust GAS definition.\(^1\)

\(^1\)It is not difficult to extend this argument to conclude \( \hat{x}(i|k) \to x(i; x_0, w) \) as \( k \to \infty \)
for \( k - N \leq i \leq k \) and any finite \( N \geq 0 \).
(b) Assume that (6) holds for some arbitrary $\delta > 0$. This gives immediately an upper bound on the optimal full information cost function for all $T$, $0 \leq T \leq \infty$, i.e., $V_\infty = \delta$. We then have the following bounds on the initial state estimate for all $k \geq 0$, and the initial state

$$\gamma_x(\|\hat{x}(0|k) - \bar{x}_0\|) \leq \delta \quad \gamma_x(\|x_0 - \bar{x}_0\|) \leq \delta$$

These two imply a bound on the initial estimate error,

$$|x_0 - \hat{x}(0|k)| \leq \gamma_x^{-1}(\delta) + \gamma_x^{-1}(\delta).$$

The process disturbance bounds are for all $k \geq 0$, $0 \leq i \leq k$

$$\gamma_w(\|\hat{w}(i|k)\|) \leq \delta \quad \gamma_w(\|w(k)\|) \leq \delta$$

and we have that

$$|w(i) - \hat{w}(i|k)| \leq \gamma_w^{-1}(\delta) + \gamma_w^{-1}(\delta).$$
Robust GAS of full information estimates

A similar argument gives for the measurement disturbance

\[ |\nu(i) - \hat{\nu}(i|k)| \leq \gamma_w^{-1}(\delta) + \gamma_w^{-1}(\delta). \]

Since

\[ -(\nu(i) - \hat{\nu}(i|k)) = h(x(i)) - h(\hat{x}(i|k)), \]

we have that

\[ |h(x(i)) - h(\hat{x}(i|k))| \leq \gamma_w^{-1}(\delta) + \gamma_w^{-1}(\delta) \]

We substitute these bounds in (7) and obtain for all \( k \geq 0 \)

\[ |x(k) - \hat{x}(k)| \leq \beta(\gamma_x^{-1}(\delta) + \gamma_x^{-1}(\delta)) + (\gamma_1 + \gamma_2)(\gamma_w^{-1}(\delta) + \gamma_w^{-1}(\delta)) \]

in which \( \beta(s) := \beta(s, 0) \) is a \( K \)-function. Finally we choose \( \delta \) such that the right-hand side is less than \( \varepsilon \), which is possible since the right-hand side defines a \( K \)-function, which goes to zero with \( \delta \). This gives for all \( k \geq 0 \)

\[ |x(k) - \hat{x}(k)| \leq \varepsilon \]

and part 2 of the robust GAS definition is established.
Consider again the estimation problem in the simplest possible setting with a linear time invariant model and Gaussian noise

\[ x^+ = Ax + Gw \quad w \sim N(0, Q) \]
\[ y = Cx + v \quad v \sim N(0, R) \]

and random initial state \( x(0) \sim N(\bar{x}(0), P^-(0)) \). In full information estimation, we define the objective function

\[ V_T(\chi(0), \omega) = \frac{1}{2} \left( |\chi(0) - \bar{x}(0)|^2_{(P^-(0))^{-1}} + \sum_{i=0}^{T-1} |\omega(i)|^2_{Q^{-1}} + |\nu(i)|^2_{R^{-1}} \right) \]

subject to \( \chi^+ = A\chi + G\omega, \ y = C\chi + \nu \).
Denote the solution to this optimization as

\[
(\hat{x}(0|T), \hat{w}_T) = \arg \min_{\chi(0), \omega} V_T(\chi(0), \omega)
\]

\[
\hat{x}(i+1|T) = A\hat{x}(i|T) + G\hat{w}(i|T)
\]

Because the system is *linear*, the estimator is stable if and only if it is stable with zero process and measurement disturbances.
Denote the solution to this optimization as

\[
(\hat{x}(0|T), \hat{w}_T) = \arg \min_{\chi(0),\omega} V_T(\chi(0), \omega)
\]

\[
\hat{x}(i+1|T) = A\hat{x}(i|T) + G\hat{w}(i|T)
\]

Because the system is \textit{linear}, the estimator is stable if and only if it is stable with zero process and measurement disturbances.

An equivalent question

If noise-free data are provided to the estimator, \((w(i), v(i)) = 0\) for all \(i \geq 0\) in (8), is the estimate error asymptotically stable as \(T \to \infty\) for all \(x_0\)?
Then we make this statement precise:

- Define estimate error as \( \tilde{x}(i|T) = x(i) - \hat{x}(i|T) \) for \( 0 \leq i \leq T - 1 \), \( T \geq 1 \).

- The noise-free measurement satisfies
  \[
  y(i) - C\hat{x}(i|T) = C\tilde{x}(i|T), \quad 0 \leq i \leq T.
  \]

- The initial condition term can be written in estimate error as
  \[
  \hat{x}(0) - \bar{x}(0) = -(\tilde{x}(0) - a)\text{ in which } a = x(0) - \bar{x}(0).
  \]

The cost function of **noise-free measurement** is

\[
V_T(a, \tilde{x}(0), w) = \frac{1}{2} \left( |\tilde{x}(0) - a|^2_{(P(0))^{-1}} + \sum_{i=0}^{T-1} |C\tilde{x}(i)|^2_{R^{-1}} + |w(i)|^2_{Q^{-1}} \right)
\]

\[9\]
For estimation we solve

\[
\tilde{x}^0(0; a), w^0(a) = \min_{\tilde{x}(0), w} V_T(a, \tilde{x}(0), w)
\]  \hspace{1cm} (10)

subject to \( \tilde{x}^+ = A\tilde{x} + Gw \).

Now consider problem (10) as an optimal control problem using \( w \) as manipulated variable and minimizing an objective that measures size of estimate error \( \tilde{x} \) and control \( w \).

The stability analysis in estimation is to show that the origin for \( \tilde{x} \) is asymptotically stable, i.e., if there exists \( K\mathcal{L} \)-function \( \beta \) such that

\[
|\tilde{x}^0(T; a)| \leq \beta(|a|, T) \text{ for all } T \in \mathbb{II}_{\geq 0}.
\]
State Estimation as Optimal Control of Estimate Error

Differences between standard regulation and the estimation problem (10):

- (10) is slightly nonstandard because it contains an extra decision variable, the initial state, and an extra term in the cost function, (9).
- Convergence is a question about the terminal state in a sequence of different optimal control problems with increasing horizon length $T$. It is also not the standard regulator convergence question which asks how the state trajectory evolves using the optimal control law.
- In standard regulation, we inject the optimal first input and ask whether we are successfully moving the system to the origin as time increases.
- In estimation, we do not inject anything into the system; we are provided more information as time increases and ask whether our explanation of the data is improving (terminal estimate error is decreasing) as time increases.
Choose *forward* DP to seek the optimal terminal state $\tilde{x}^0(T; a)$ as a function of the parameter $a$ appearing in the cost function. (Check Exercise 4.12 to see how to solve (10).) There exists the following recursion for the optimal terminal state

$$\tilde{x}^0(k + 1; a) = (A - \tilde{L}(k)C) \tilde{x}^0(k; a) \quad (11)$$

for $k \geq 0$. The initial condition for the recursion is $\tilde{x}^0(0; a) = a$. The time-varying gains $\tilde{L}(k)$ and associated cost matrices $P^-(k)$ are

$$P^-(k + 1) = GQG' + AP^-(k)A'$$
$$- AP^-(k)C'(CP^-(k)C' + R)^{-1}CP^-(k)A \quad (12)$$

$$\tilde{L}(k) = AP^-(k)C'(CP^-(k)C' + R)^{-1} \quad (13)$$

in which $P^-(0)$ is specified in the problem.
Asymptotic stability of the estimate error can be established by showing that $V(k, \tilde{x}) := (1/2)\tilde{x}'P(k)^{-1}\tilde{x}$ is a Lyapunov function for (11) (Jazwinski, 1970, Theorem 7.4).

Although one can find Lyapunov functions valid for estimation, they do not have the same simple connection to optimal cost functions as in standard regulation problems, even in the linear, unconstrained case.

Stability arguments based instead on properties of $V_0^T(a)$ are simpler and more easily adapted to cover new situations arising in research problems.

If a Lyapunov function is required for further analysis, a converse theorem guarantees its existence.
Duality of Linear Estimation and Regulation

Estimator problem:

\[
\begin{align*}
  x(k + 1) &= Ax(k) + Gw(k) \\
  y(k) &= Cx(k) + v(k)
\end{align*}
\]

\( R > 0 \) \( Q > 0 \) \((A, C)\) detectable \((A, G)\) stabilizable

\[
\tilde{x}(k + 1) = \left(A - \tilde{L}C\right) \tilde{x}(k)
\]

Regulator problem:

\[
\begin{align*}
  x(k + 1) &= Ax(k) + Bu(k) \\
  y(k) &= Cx(k)
\end{align*}
\]

\( R > 0 \) \( Q > 0 \) \((A, B)\) stabilizable \((A, C)\) detectable

\[
x(k + 1) = (A + BK)x(k)
\]
Lemma 9 (Duality of controllability and observability)

\((A, B)\) is controllable (stabilizable) if and only if \((A', B')\) is observable (detectable).
In MHE we consider only the $N$ most recent measurements,

$$y_N(T) = \{y(T - N), y(T - N + 1), \ldots y(T - 1)\}.$$ 

For $T > N$, the MHE problem is defined to be:

$$\min_{\chi(T - N), \omega} \hat{V}_T(\chi(T - N), \omega) = \Gamma_{T - N}(\chi(T - N)) + \sum_{i=T-N}^{T-1} \ell_i(\omega(i), \nu(i))$$ 

subject to:

$$\chi^+ = f(\chi, \omega)$$
$$y = h(\chi) + \nu$$
$$\omega = \{\omega(T - N), \ldots, \omega(T - 1)\}$$
The designer chooses the prior weighting $\Gamma_k(\cdot)$ for $k > N$ until the data horizon is full.

For times $T \leq N$, we generally define the MHE problem to be the full information problem.

**Zero Prior Weighting**

When we choose $\Gamma_i(\cdot) = 0$ for all $i \geq N$:

$$\hat{V}_T(\chi(T - N), \omega) = \sum_{i=T-N}^{T-1} \ell_i(w(i), v(i))$$

Because it discounts the past data completely, this form of MHE must be able to asymptotically reconstruct the state using only the most recent $N$ measurements.

Establishing the convergence of the estimate error to zero for this form of MHE is straightforward.
Observability

To ensure solution existence, the system needs to be restricted further than i-IOSS:

**Definition 10 (Observability)**

The system $x^+ = f(x, w), y = h(x)$ is **observable** if there exist finite $N_o \in \mathbb{I}_{\geq 1}, \gamma_1(\cdot), \gamma_2(\cdot) \in \mathcal{K}$ such that for every two initial states $z_1$ and $z_2$, and any two disturbance sequences $w_1, w_2$, and all $k \geq N_o$

$$|z_1 - z_2| \leq \gamma_1(\|w_1 - w_2\|_{0:k-1}) + \gamma_2(\|y_{z_1, w_1} - y_{z_2, w_2}\|_{0:k})$$

At any time $T \geq N$, for the decision variables of $\chi(T - N) = x(T - N)$ and $\omega(i) = w(i)$ for $T - N \leq i \leq T - 1$, recall the cost function:

$$\hat{V}_T(\chi(T - N), \omega) = \sum_{i=T-N}^{T-1} \ell_i(w(i), v(i))$$

(14)

which is less than $V_\infty$ defined in the full information problem.
Observability ensures that for all $k \geq N \geq N_o$,

$$|x(k - N) - \hat{x}(k - N|k)| \leq \gamma_2(\|v\|_{k-N:k})$$

- Since $v(k)$ is bounded for all $k \geq 0$, observability has bounded the distance between the initial estimate in the horizon and the system state for all $k \geq N$.
- This fact along with continuity of $\hat{V}(\chi, \omega)$ ensures existence of the solution to the MHE problem by the Weierstrass theorem.
- But the solution does not have to be unique.
Definition 11 (Final-state observability)

The system $x^+ = f(x, w)$, $y = h(x)$ is \textit{final-state observable} (FSO) if there exist finite $N_0 \in \mathbb{I}_{\geq 1}$, $\gamma_1(\cdot)$, $\gamma_2(\cdot) \in \mathcal{K}$ such that for every two initial states $z_1$ and $z_2$, and any two disturbance sequences $w_1, w_2$, and all $k \geq N_0$

$$|x(k; z_1, w_1) - x(k; z_2, w_2)| \leq \gamma_1\left(\|w_1 - w_2\|_{0:k-1}\right) + \gamma_2\left(\|y_{z_1,w_1} - y_{z_2,w_2}\|_{0:k}\right)$$

- FSO is the natural system requirement for MHE with zero prior weighting to provide \textit{stability} and \textit{convergence}.
- FSO is weaker than observability and stronger than i-IOSS (detectability) as discussed in Exercise 4.11.
Convergence of MHE Cost Function

- Since \((w(i), v(i))\) converges to zero, (31) implies that \(\hat{V}_T\) converges to zero as \(T \to \infty\).
- The optimal cost at \(T\), \(\hat{V}_T^0\), is bounded above by \(\hat{V}_T\) so \(\hat{V}_T^0\) also converges to zero:

\[
\hat{V}_T^0 = \sum_{i=T-N}^{T-1} \ell_i(\hat{w}(i|T), y(i) - h(\hat{x}(i|T))) \to 0
\]

- Therefore \(y(i) - h(\hat{x}(i|T)) \to 0\) and \(\hat{w}(i|T) \to 0\).
- Since \(y = h(x) + v\) and \(v(i)\) converges to zero, and \(w(i)\) converges to zero, we also have

\[
h(x(i)) - h(\hat{x}(i|T)) \to 0 \quad w(i) - \hat{w}(i|T) \to 0
\]

(15)

for \(T - N \leq i \leq T - 1, \ T \geq N\).
Consider an observable system and measurement sequence generated by (1) with disturbances satisfying Assumption 26. The MHE estimate with zero prior weighting, $N \geq N_0$, and stage cost satisfying (5), is robustly GAS.
Nonzero Prior Weighting

There are two drawbacks of zero prior weighting:

- The system had to be assumed *observable* rather than detectable to ensure existence of the solution to the MHE problem.
- A large horizon $N$ may be required to obtain performance comparable to full information estimation.

We address these disadvantages by using **nonzero prior weighting**:

$$
\min_{\chi(T-N), \omega} \hat{V}_T(\chi(T-N), \omega) = \Gamma_{T-N}(\chi(T-N)) + \sum_{i=T-N}^{T-1} \ell_i(\omega(i), \nu(i))
$$
Definition 13 (Full information arrival cost)

The full information arrival cost is defined as

$$Z_T(p) = \min_{\chi(0), \omega} V_T(\chi(0), \omega)$$ \hspace{1cm} (16)

subject to

$$\chi^+ = f(\chi, \omega) \quad y = h(\chi) + \nu \quad \chi(T; \chi(0), \omega) = p$$

Here forward DP is used to decompose the full information problem exactly into the MHE problem (30) in which $\Gamma(\cdot)$ is chosen as arrival cost.

Lemma 14 (MHE and full information estimation)

The MHE problem (30) is equivalent to the full information problem (3) for the choice $\Gamma_k(\cdot) = Z_k(\cdot)$ for all $k > N$ and $N \geq 1$. 
Definition 15 (MHE arrival cost)

The MHE arrival cost $\hat{Z}(\cdot)$ is defined for $T > N$ as

$$\hat{Z}_T(p) = \min_{z, \omega} \hat{V}_T(z, \omega)$$

$$= \min_{z, \omega} \Gamma_{T-N}(z) + \sum_{i=T-N}^{T-1} \ell_i(\omega(i), \nu(i))$$ (17)

subject to

$$\chi^+ = f(\chi, \omega) \quad y = h(\chi) + \nu \quad \chi(T; z, T-N, \omega) = p$$

For $T \leq N$, usually define the MHE problem to be the full information problem, so $\hat{Z}_T(\cdot) = Z_T(\cdot)$ and $\hat{V}_T^0 = V_T^0$. 
Prior weighting

Choosing a prior weighting that underbounds the MHE arrival cost is the key sufficient condition for stability and convergence of MHE.

Assumption 16 (Prior weighting)

We assume that $\Gamma_k(\cdot)$ is continuous and satisfies the following inequalities for all $k > N$

1. Upper bound

$$\Gamma_k(p) \leq \hat{Z}_k(p) = \min_{z,\omega} \Gamma_{k-N}(z) + \sum_{i=k-N}^{k-1} \ell_i(\omega(i), \nu(i)) \quad (18)$$

subject to $\chi^+ = f(\chi, \omega), y = h(\chi) + \nu, \chi(k; z, k - N, \omega) = p$.

2. Lower bound

$$\Gamma_k(p) \geq \hat{V}_k^0 + \gamma_p (|p - \hat{x}(k)|) \quad (19)$$

in which $\gamma_p \in \mathcal{K}_\infty$. 
An upper bound for the MHE optimal cost is necessary to establish convergence of the MHE estimates.

A stronger result is that the MHE arrival cost is bounded above by the full information arrival cost.

**Proposition 17 (Arrival cost of full information greater than MHE)**

\[
\hat{Z}_T(\cdot) \leq Z_T(\cdot) \quad T \geq 1
\]  

Given (20) we also have the analogous inequality for the optimal costs of MHE and full information

\[
\hat{V}_T^0 \leq V_T^0 \quad T \geq 1
\]
Assumption 18 (MHE detectable system)

We say a system \( x^+ = f(x, w), \ y = h(x) \) is **MHE detectable** if the system augmented with an extra disturbance \( w_2 \)

\[
x^+ = f(x, w_1) + w_2 \quad y = h(x)
\]

is i-IOSS with respect to the augmented disturbance \((w_1, w_2)\).

- Note that MHE detectable is stronger than i-IOSS (detectable) but weaker than observable and FSO.
- See also Exercise 4.10.
Theorem 19 (Robust GAS of MHE)

Consider an MHE detectable system and measurement sequence generated by (1) with disturbances satisfying Assumption 26. The MHE estimate defined by (30) using the prior weighting function \( \Gamma_k(\cdot) \) satisfying Assumption 16 and stage cost satisfying Assumption 5 is robustly GAS.
Robust GAS of MHE

\[
\hat{x}(N+1|2N) \quad \hat{x}(2N|2N)
\]

\[
\hat{x}(N-1|2N) \quad \hat{x}(N|2N)
\]

\[
\hat{x}(N-1|N) \quad \hat{x}(N|N)
\]

\[
\hat{x}(0|N) \quad \hat{x}(1|N)
\]

\[
\hat{w}(0|N) \quad \hat{w}(N-1|N) \quad \hat{w}(N-1|2N) \quad \hat{w}(2N-1|2N)
\]

\[
\hat{x}(2N-1|2N) \quad \hat{x}(2N|2N)
\]

\[
\cdot \quad \cdot
\]

\[
N \quad 2N
\]

\[
0
\]

Stuttgart – June 2011
For estimation problem, some physically known facts should also be enforced such as:

- Concentrations of impurities must be nonnegative,
- Fluxes of mass must have the correct sign given concentration gradients.
- Fluxes of energy must have the correct sign given temperature gradients.
- ... 

However, unlike the regulator, the estimator has no way to enforce these constraints on the system.

It is important that any constraints imposed on the estimator are satisfied by the system generating the measurements.
Constrained MHE

It is straightforward to add constraints in MHE since it is posed as an optimization formulation.

Assumption 20 (Estimator constraint sets)

1. For all $k \in \mathbb{I}_{\geq 0}$, the sets $\mathbb{W}_k$, $\mathbb{X}_k$, and $\mathbb{V}_k$ are nonempty and closed, and $\mathbb{W}_k$ and $\mathbb{V}_k$ contain the origin.

2. For all $k \in \mathbb{I}_{\geq 0}$, the disturbances and state satisfy

$$x(k) \in \mathbb{X}_k \quad w(k) \in \mathbb{W}_k \quad v(k) \in \mathbb{V}_k$$

3. The prior satisfies $\bar{x}_0 \in \mathbb{X}_0$. 
**Constrained full information.** The constrained full information estimation objective function is:

\[
V_T(\chi(0), \omega) = \ell_x(\chi(0) - \bar{x}_0) + \sum_{i=0}^{T-1} \ell_i(\omega(i), \nu(i)) \tag{22}
\]

subject to

\[
\begin{align*}
\chi^+ &= f(\chi, \omega) \\
y &= h(\chi) + \nu \\
\chi(i) &\in \mathbb{X}_i \\
\omega(i) &\in \mathbb{W}_i \\
\nu(i) &\in \mathbb{V}_i \\
i &\in \mathbb{I}_0: T-1
\end{align*}
\]

The constrained full information problem is

\[
\min_{\chi(0), \omega} V_T(\chi(0), \omega) \tag{23}
\]
Theorem 21 (Robust GAS of constrained full information)

Consider an i-IOSS (detectable) system and measurement sequence generated by (1) with constrained, convergent disturbances satisfying Assumptions 26 and 20. The constrained full information estimator (23) with stage cost satisfying Assumption 5 is robustly GAS.
**Constrained MHE.** The constrained moving horizon estimation objective function is

\[
\hat{V}_T(\chi(T - N), \omega) = \Gamma_{T-N}(\chi(T - N)) + \sum_{i=T-N}^{T-1} \ell_i(\omega(i), \nu(i)) \tag{24}
\]

subject to

\[
\begin{align*}
\chi^+ &= f(\chi, \omega) & y &= h(\chi) + \nu \\
\chi(i) &\in \mathbb{X}_i & \omega(i) &\in \mathbb{W}_i & \nu(i) &\in \mathbb{V}_i & i \in I_{T-N:T-1}
\end{align*}
\]

The constrained MHE is given by the solution to the following problem

\[
\min_{\chi(T-N), \omega} \hat{V}_T(\chi(T - N), \omega) \tag{25}
\]
Theorem 22 (Robust GAS of constrained MHE)

Consider an MHE detectable system and measurement sequence generated by (1) with convergent, constrained disturbances satisfying Assumptions 26 and 20. The constrained MHE estimator (25) using the prior weighting function \( \Gamma_k(\cdot) \) satisfying Assumption 16 and stage cost satisfying Assumption 5 is robustly GAS.

- Because the system satisfies the state and disturbance constraints due to Assumption 20, both full information and MHE optimization problems are \textbf{feasible} at all times.

- The proofs of Theorems 21 and 22 closely follow the proofs of their respective unconstrained versions, Theorems 27 and 19, and are omitted.
Constrained Linear Systems

For the *constrained linear systems*:

\[
x^+ = Ax + Gw \quad y = Cx + v
\]  \hspace{1cm} (26)

- First, the i-IOSS assumption of full information estimation and the MHE detectability assumption both reduce to the assumption that \((A, C)\) *is detectable* in this case.
- We usually choose a constant quadratic function for the estimator stage cost for all \(i \in \mathbb{I}_{\geq 0}\)

\[
\ell_i(w, v) = (1/2)(|w|^2_{Q^{-1}} + |v|^2_{R^{-1}}) \quad Q, R > 0
\]  \hspace{1cm} (27)

- In the unconstrained linear problem, we can of course find the full information arrival cost exactly; it is:

\[
Z_k(z) = V_k^0 + (1/2)|z - \hat{x}(k)|{(P_{-(k)})^{-1}} \quad k \geq 0
\]
We use this unconstrained arrival cost to be the prior weighting in MHE:

**Assumption 23 (Prior weighting for linear system)**

\[ \Gamma_k(z) = \hat{V}_k^0 + (1/2)|z - \hat{x}(k)|(P_-(k))^{-1} \quad k > N \]  

(28)

in which \( \hat{V}_k^0 \) is the optimal MHE cost at time \( k \).

This choice implies robust GAS of the MHE estimator also for the *constrained case* as we next demonstrate.

To ensure the form of the estimation problem to be solved online is a quadratic program, we specialize the constraint sets to be polyhedral regions.

**Assumption 24 (Polyhedral constraint sets)**

For all \( k \in \mathbb{I}_{\geq 0} \), the sets \( \mathbb{W}_k, \mathbb{X}_k, \) and \( \mathbb{V}_k \) in Assumption 20 are nonempty, closed polyhedral regions containing the origin.
Corollary 25 (Robust GAS of constrained MHE)

Consider a detectable linear system and measurement sequence generated by (26) with convergent, constrained disturbances satisfying Assumptions 26 and 24. The constrained MHE estimator (25) using prior weighting function satisfying (28) and stage cost satisfying (27) is robustly GAS.

This corollary follows as a special case of Theorem 22.
The MHE approach uses the MHE estimate \( \hat{x}(T - N) \) and prior weighting function \( \Gamma_{T-N}(\cdot) \) derived from the unconstrained arrival cost as shown in (28).

We call this approach a “filtering update” because the prior weight at time \( T \) is derived from the solution of the MHE “filtering problem” at time \( T - N \), i.e., the estimate of \( \hat{x}(T - N) := \hat{x}(T - N | T - N) \) given measurements up to time \( T - N - 1 \).

For implementation, this choice requires storage of a window of \( N \) prior filtering estimates to be used in the prior weighting functions as time progresses.
In the **smoothing update** we wish to use $\hat{x}(T - N|T - 1)$ (instead of $\hat{x}(T - N|T - N)$) for the prior and wish to find an appropriate prior weighting based on this choice.

When constraints are added to the problem, the smoothing update provides a different MHE than the filtering update.

The smoothing prior weighting maintains the stability and GAS robustness properties of MHE with the filtering update.

Recall the unconstrained full information arrival cost is given by

$$Z_{T-N}(z) = V^0_{T-N} + (1/2)|z - \hat{x}(T - N)|^2_{(P-(T-N))^{-1}} \quad T > N \quad (29)$$

Now consider the proper weight when using $\hat{x}(T - N|T - 2)$ in place of $\hat{x}(T - N) := \hat{x}(T - N|T - 1)$ ...
Should it be the smoothed covariance $P(T - N|T - 2)$ instead of $P^{-}(T - N) := P(T - N|T - 1)$?
Should it be the smoothed covariance $P(T - N|T - 2)$ instead of $P^-(T - N) := P(T - N|T - 1)$?

Correct, but not complete.
Smoothing Update

Should it be the smoothed covariance $P(T - N|T - 2)$ instead of $P^-(T - N) := P(T - N|T - 1)$?

Correct, but not complete.

- The smoothed prior $\hat{x}(T - N|T - 2)$ is influenced by the measurements $y_{0:T-2}$.
- But the sum of stage costs in the MHE problem at time $T$ depends on measurements $y_{T-N:T-1}$.
- Therefore we have to adjust the prior weighting so we do not double count the data $y_{T-N:T-2}$.
For any square matrix $R$ and integer $k \geq 1$, define $\text{diag}_k(R)$ to be the following:

$$\text{diag}_k(R) := \begin{bmatrix} R & \cdots & R \\ & & \\ & & \end{bmatrix}$$

$k$ times

$$W_k = \text{diag}_k(R) + \mathcal{O}_k(\text{diag}_k(Q))\mathcal{O}'_k$$
Smoothing Update

From the following recursion (Rauch, Tung, and Striebel, 1965; Bryson and Ho, 1975)

\[
P(k|T) = P(k) + P(k)A'(P^-(k + 1))^{-1}\left( P(k + 1|T) - P^-(k + 1) \right) (P^-(k + 1))^{-1} AP(k)
\]

We iterate this equation backwards \(N - 1\) times starting from the known value \(P(T - 1|T - 2) := P^-(T - 1)\) to obtain \(P(T - N|T - 2)\).
The smoothing arrival cost is then given by

\[
\tilde{Z}_{T-N}(z) = \hat{V}_{T-1}^0 + (1/2) |z - \hat{x}(T - N| T - 2)|^2_{(P(T-N|T-2))^{-1}} \\
- (1/2) |y_{T-N:T-2} - O_{N-1}z|^2_{(W_{N-1})^{-1}} \quad T > N
\]

- The second term accounts for the use of the smoothed covariance and the smoothed estimate.
- The third term subtracts the effect of the measurements that have been double counted in the MHE objective as well as the smoothed prior estimate.

\[2\text{See Rao, Rawlings, and Lee (2001) and Rao (2000, pp.80–93) for a derivation that shows } \tilde{Z}_T(\cdot) = Z_T(\cdot) \text{ for } T > N.\]
Smoothing Update

\[ y_{T-N:1}^{T-2} \]

Filtering update \( y_{T-2N:T-N-1} \)

Smoothing update \( y_{T-N:T-1} \)

MHE problem at \( T \)
Recommended exercises

- Observability, detectability, i-IOSS. Exercises 4.1, 4.2, 4.3, 4.4, 4.5, 4.7, 4.8, 4.10, 4.11.\(^3\)

---

\(^3\)Rawlings and Mayne (2009, Chapter 4). Downloadable from www.che.wisc.edu/~jbraw/mpc.
Open research question

Weaken the restriction on the disturbances from convergent to bounded.

Assumption 26 (Bounded disturbances)

The sequence \((w(k), v(k))\) for \(k \in \mathbb{I}_{\geq 0}\) are bounded, i.e., \(\|w\|\) and \(\|v\|\) are finite.
Then establish the following kind of result.

**Theorem 27 (Robust GAS of full information estimate)**

*Given an i-IOSS (detectable) system and measurement sequence generated by (1) with disturbances satisfying Assumption 26, then the full information estimate with stage cost satisfying Assumption 5 is robustly GAS.*

\[
|x(k; x_0, w) - \hat{x}(k; \hat{x}_0, \hat{w})| \leq \beta(|x_0 - \bar{x}_0|, k) + \gamma_1(\|w\|_{[0:k-1]}) + \gamma_2(\|v\|_{[0:k-1]})
\]

for all \( k \in \mathbb{II}_{\geq 0} \).
Further Reading I


Further Reading II


Acknowledgments

The author is indebted to Luo Ji of the University of Wisconsin and Ganzhou Wang of the Universität Stuttgart for their help in organizing the material and preparing the overheads.