EECE 574 - Adaptive Control
Recursive Identification in Closed-Loop and Adaptive Control

Guy Dumont

Department of Electrical and Computer Engineering
University of British Columbia

January 2012
All previous methods use the least-squares criterion

\[ V(t) = \frac{1}{t} \sum_{i=1}^{t} [y(i) - x^T(i)\hat{\theta}]^2 \]

and thus identify the average behaviour of the process.

For standard RLS, the estimation gain eventually converges to zero and adaptation stops.
When the parameters are time varying, it is desirable to base the identification on the most recent data rather than on the old one, not representative of the process anymore. This can be achieved by exponential discounting of old data, using the criterion

\[ V(t) = \frac{1}{t} \sum_{i=1}^{t} \lambda^{t-i} [y(i) - x^T(i) \hat{\theta}]^2 \]

where \(0 < \lambda \leq 1\) is called the forgetting factor.
Forgetting Factor

The new criterion can also be written

\[ V(t) = \lambda V(t-1) + [y(t) - x^T(t)\hat{\theta}]^2 \]

Then, it can be shown (Goodwin and Payne, 1977) that the RLS scheme becomes

**RLS with Forgetting**

\[
\hat{\theta}(t+1) = \hat{\theta}(t) + K(t+1)[y(t+1) - x^T(t+1)\hat{\theta}(t)] \\
K(t+1) = P(t)x(t+1)/[\lambda + x^T(t+1)P(t)x(t+1)] \\
P(t+1) = \begin{cases} 
  P(t) - \frac{P(t)x(t+1)x^T(t+1)P(t)}{[\lambda + x^T(t+1)P(t)x(t+1)]} & \text{if } P(t) > 0 \\
  0 & \text{otherwise}
\end{cases} \frac{1}{\lambda}
\]

In choosing \( \lambda \), one has to compromise between fast tracking and long term quality of the estimates. The use of the forgetting may give rise to problems.
The smaller $\lambda$ is, the faster the algorithm can track, but the more the estimates will vary, even the true parameters are time-invariant.

A small $\lambda$ may also cause blowup of the covariance matrix $P$, since in the absence of excitation, covariance matrix update equation essentially becomes

$$P(t + 1) = \frac{1}{\lambda} P(t)$$

in which case $P$ grows exponentially, leading to wild fluctuations in the parameter estimates.
Variable Forgetting Factor

- One way around this is to vary the forgetting factor according to the prediction error $\varepsilon$ as in
  \[ \lambda(t) = 1 - k\varepsilon^2(t) \]

- Then, in case of low excitation $\varepsilon$ will be small and $\lambda$ will be close to 1. In case of large prediction errors, $\lambda$ will decrease.
The following scheme\(^1\) is recommended:

**EFRA Algorithm**

\[
\begin{align*}
\varepsilon(t+1) &= y(t+1) - x^T(t+1)\hat{\theta}(t) \\
\hat{\theta}(t+1) &= \hat{\theta}^T(t) + \frac{\alpha P(t)x(t+1)}{\lambda + x^T(t+1)P(t)x(k+1)}\varepsilon(t) \\
P(t+1) &= \frac{1}{\lambda} \left[ P(t) - \frac{P(t)x(t+1)x^T(t+1)P(t)}{\lambda + x(t+1)^TP(t)x(t+1)} \right] \\
&\quad + \beta I - \gamma P(t)^2
\end{align*}
\]

where \(I\) is the identity matrix, and \(\alpha, \beta\) and \(\gamma\) are constants.

---

With the EFRA, the covariance matrix is bounded on both sides:

$$\sigma_{\text{min}}I \leq P(t) \leq \sigma_{\text{max}}I \quad \forall t$$

where

$$\sigma_{\text{min}} \approx \frac{\beta}{\alpha - \eta} \quad \sigma_{\text{max}} \approx \frac{\eta}{\gamma} + \frac{\beta}{\eta}$$

with

$$\eta = \frac{1 - \lambda}{\lambda}$$

With $\alpha = 0.5$, $\beta = \gamma = 0.005$ and $\lambda = 0.95$, $\sigma_{\text{min}} = 0.01$ and $\sigma_{\text{max}} = 10$. 
Identification in Closed Loop

The Identifiability Problem
Let the system be described by

\[ y(t) + a \cdot y(t - 1) = b \cdot u(t - 1) + e(t) \]

with

\[ u(t) = g \cdot y(t) \]

Let \( \hat{a} \) and \( \hat{b} \) be closed-loop estimates of \( a \) and \( b \).

Then, the closed-loop system can be written as:

\[ y(t) + (\hat{a} - \hat{b} \cdot g) y(t - 1) = e(t) \]
Hence any estimates $\hat{a}$ and $\hat{b}$ that satisfy

$$\hat{a} - \hat{b} \cdot g = a - b \cdot g$$

will give the same value for the identification criterion.

All estimates such that

$$\hat{a} = a + k \cdot g$$
$$\hat{b} = b + k$$

will give a good description of the process.
If the identification is performed using two feedback gains $g_1$ and $g_2$ or if the parameter $a$ is fixed, then the system becomes identifiable, because we have as many equations as unknowns.
Definitions

Let the discrete plant be described by

\[ y(t) = G_D(q^{-1})u(t) + G_N(q^{-1})e(t) \]

where \( G_D \) and \( G_N \) are linear rational transfer functions in the backward shift operator \( q^{-1} \) (i.e. \( q^{-1}y(t) = y(t - 1) \)) that can be parameterized by a vector \( \theta \), \( \{e(t)\} = N(0, \sigma) \). Let \( S \) denote this true system.

Let us assume that the identification is performed with the feedback controller \( R \) such that \( u(t) = R \cdot y(t) \).

The problem is then to find \( \hat{\theta} \), an estimate of \( \theta \) such that the model \( M(\hat{\theta}) \) given by

\[ y(t) = \hat{G}_D(q^{-1})u(t) + \hat{G}_N(q^{-1})e(t) \]

where \( \hat{G}_D(q^{-1}) \) and \( \hat{G}_N(q^{-1}) \) are parameterized by \( \hat{\theta} \), describes the system \( S \). Assume that an identification method denoted \( I \) is used to obtain the estimate \( \hat{\theta} \).
A loose and intuitive definition of identifiability is that $M(\hat{\theta})$ describes $S$ as the number of measurements $N$ tends to infinity. Let us define

$$D_T(S, M) = \{ \hat{\theta} | \hat{G}_D(q^{-1}) \equiv G_D(q^{-1}) \text{ and } \hat{G}_N(q^{-1}) = G_N(q^{-1}) \forall q \}$$

This is the set of desired estimates which corresponds to models $M(\hat{\theta})$ with the same plant and noise transfer functions as the actual system $S(\theta)$.

Note that this set does not depend on the regulator $R$ nor on the identification method $I$. Note also that the orders of the model transfer functions can be greater than those of the system, in which case there exist some pole-zero cancellations.

The actual estimates $\hat{\theta}$ depend on the number of measurements, the system and its model, the regulator and the identification method and can be written as $\hat{\theta}(N; S, M, I, R)$. 
Definitions

- **Definition 1:** The system $S$ is said to be **system identifiable** under $M, I, R$, i.e. $SI(M, I, R)$ if
  \[ \hat{\theta}(N; S, M, I, R) \rightarrow DT(S, M) \text{ w.p.1} \]
  as $N \rightarrow \infty$.

- **Definition 2:** The system $S$ is said to be **strongly system identifiable** under $I$ and $R$, i.e. $SSI(I, R)$ if it is $SI(M, I, R)$ for all $M$ s.t. $DT(S, M) \neq \emptyset$ where $\emptyset$ denotes the empty set.

- **Definition 3:** The system $S$ is said to be **parameter identifiable** under $M, I$ and $R$, i.e. $PI(M, I, R)$ if it is $SI(M, I, R)$ and $DT(S, M)$ consists of only one element.

According to those definitions, the system of Example 1 is neither $SI$ nor $PI$ for the class of models (2). Moreover, since for the model (2) $DT(S, M) \neq \emptyset$, the system is not $SSI$ either. However, when two regulators are used, the system becomes $PI$. It is also important to note that when a system is $SSI(I, R)$, the fact that the identification is performed under closed-loop is irrelevant and the identification can then be performed as if the system were operating in open-loop.
Consider the system

\[ A(q^{-1})y(t) = B(q^{-1})q^{-k}u(t) + C(q^{-1})e(t) \]

with the feedback

\[ F(q^{-1})u(t) = G(q^{-1})y(t) \]

The closed-loop system can then be described as

\[ (AF - q^{-k}BG)y(t) = CFe(t) \]

It is then obvious that any estimates \( \hat{A} \) and \( \hat{B} \) such that

\[ (\hat{A}F - q^{-k}\hat{B}G) = (AF - q^{-k}BG) \]

will describe the above system.

Hence, if \( L(q^{-1}) \) is an arbitrary polynomial, any \( \hat{A} \) and \( \hat{B} \) s.t.

\[
\begin{align*}
\hat{A} &= A + LG \\
\hat{B} &= B + q^kLF
\end{align*}
\]

(1)

are possible estimates.

This means that the true order of the system cannot be established from this type of closed-loop experiment and thus, has to be known à-priori.
Let the system $S$ be described by the following ARMAX process:

$$A(q^{-1})y(t) = q^{-k}B(q^{-1})u(t) + C(q^{-1})e(t)$$

where

\[
A(q^{-1}) = 1 + a_1 q^{-1} + \ldots + a_{n_a} q^{-n_a} \\
B(q^{-1}) = b_1 q^{-1} + \ldots + b_{n_b} q^{-n_b} \\
C(q^{-1}) = 1 + c_1 q^{-1} + \ldots + c_{n_c} q^{-n_c}
\]

$k \geq 0$ \hspace{0.5cm} $n_a \geq 0$ \hspace{0.5cm} $n_b \geq 1$ \hspace{0.5cm} $n_c \geq 0$

$A$, $B$, $C$ are assumed to co-prime, \textit{i.e.} have no common factors.
Identifiability Conditions

Let the regulator $R$ be given by:

$$F(q^{-1})u(t) = G(q^{-1})y(t)$$

where

$$F(q^{-1}) = 1 + f_1 q^{-1} + \ldots + f_n q^{-n_f}$$

$$G(q^{-1}) = g_0 + g_1 q^{-1} + \ldots + g_n q^{-n_g}$$

$F$ and $G$ are co-prime and $G$ may contain a delay, i.e. some of its leading coefficients may be zero.

Let the class of models $M(\hat{\theta})$ be

$$\hat{A}(q^{-1})y(t) = q^{-\hat{k}}\hat{B}(q^{-1})u(t) + \hat{C}(q^{-1})e(t)$$

with $\hat{n}_a \geq 0$, $\hat{n}_c \geq 0$, $\hat{k} \geq 0$, $\hat{n}_b \geq 1$
Identifiability Conditions

- Introduce

\[ n^* = \min \left[ (\hat{n}_a - n_a), (\hat{n}_b + \hat{k} - n_b - k), (\hat{n}_c - n_c) \right] \]

- The system with its feedback can be written as

\[ y = \frac{CF}{AF - q^{-k}BG} \]

- Let us define \( n_p \) be the number of common factors between \( C \) and \( (AF - q^{-k}BG) \) and let the order of the polynomial \( AF - q^{-k}BG \) be \( \max\left[ (n_a + n_f), (k + n_b + n_g) \right] - r \) with \( r \geq 0 \). Note that \( r > 0 \) means \( n_a + n_f = k + n_b + n_g \) and that some of highest order coefficients are zeros.

- Assume that a direct identification, using the sequences \( \{u(t)\} \) and \( \{y(t)\} \), is employed to estimates \( \hat{\theta} \) such that \( M(\hat{\theta}) \) describes \( S(\theta) \).
Identifiability Conditions

Theorem (Söderström, 1973)

* A necessary and sufficient condition for the set of desired estimates $D_T(S, M)$ to be non-empty is

$$n^* \geq 0 \text{ and } k^* \leq k$$

* From the definition of $n^*$, it is seen that the condition simply means that the true system is contained in the model set, *i.e.* that the model order is sufficient and that the dead-time is not overestimated.
Identification in Closed Loop

Identifiability Conditions

- A related well known sufficient identifiability condition is that the order of the feedback law be greater than or equal to the order of the forward path, see Goodwin and Payne (1977).

- The presence of dead-time $k > 0$ helps identifiability. This can be explained by the fact that the feedback signal then contains a component more or less independent of the current output $y(t)$.

- The previous condition unfortunately involve the true order of the system. If the true order is not known à-priori, then these conditions cannot be checked à-posteriori, since the order cannot be determined in closed-loop identification.

- These conditions are thus of little use in a practical application. The following theorem provides a way to guarantee $SSI$ without knowledge of the systems order.
Theorem (Söderström et al. (1975))

Consider the system $S$:

$$y(t) = GD(q^{-1})u(t) + GN(q^{-1})e(t)$$

where

$$\dim y = n_y, \dim u = n_u, \dim e = n_y, \ e = N(O, \Lambda)$$

and

$$GN(q^{-1}) \quad \text{has all zeros outside unit circle}$$

Let the control signal be given by

$$u(t) = Fi(q^{-1})y(t) + ki(q^{-1})v(t)$$

{$v$} is an external signal of $\dim v = n_v$ and $Fi$ and $ki$ are rational functions and $i = 1, \ldots r$. Assume, without loss of generality, that there is a dead-time either in the process or in the regulator, i.e. $GD(0)Fi(0) = 0$. Then, the system $S$ is identifiable if and only if

$$n_0\{ \begin{bmatrix} n_y \\ k_1(z) & \ldots & k_r(z) & F_i(z) & \ldots & F_r(z) \\ 0 & \ldots & 0 & I & \ldots & I \end{bmatrix} \} = n_y + n_u \ \forall z$$
A necessary condition for identifiability is

\[ r(n_v + n_y) \geq n_y + n_u \]

or

\[ r \geq (n_u + n_y) / (n_v + n_y) \]

If \( n_v = n_u \), then closed-loop identifiability always satisfied

Without external signal \( (n_v = 0) \),

\[ r \geq 1 + n_u / n_y \]

which when \( n_y = n_u \) becomes \( r \geq 2 \)
Under time-varying feedback, identifiability should be guaranteed.

Because in adaptive control, the controller parameters are constantly updated, identifiability should be guaranteed (at least during the transient phase).

Nonlinearities in the controller should also help identifiability.
Identification for Control

1. Experimental loop

2. Identified system

3. Design loop

4. Achieved loop

Optimal loop

(Gevers, 2005)
Prediction Error Identification

True system $\mathcal{S}$: $y_t = G_0(z)u_t + H_0(z)e_t$ \quad $e_t$ : white noise

Model set $\mathcal{M}$: $\{G(z, \theta), H(z, \theta), \theta \in D_\theta \subset \mathbb{R}^d\}$

Prediction error:
\[
\varepsilon_t(\theta) \triangleq y_t - \hat{y}_{t|t-1}(\theta) = H^{-1}(z, \theta)[(G_0(z) - G(z, \theta))u_t + v_t]
\]

Filtered prediction error: $\varepsilon^f_t(\theta) = D(z)\varepsilon_t(\theta)$

Prediction error criterion (Least Squares):
\[
V_N(\theta) = \frac{1}{N} \sum_{t=1}^{N} [\varepsilon^f_t(\theta)]^2 \quad \Rightarrow \quad \hat{\theta}_N = \arg \min_{\theta \in D_\theta} V_N(\theta)
\]
\[
\Rightarrow \hat{G} = G(z, \hat{\theta}_N), \quad \hat{H} = H(z, \hat{\theta}_N)
\]

(Gevers, 2005)
Properties of the Identified Model

Situation A: $S \in \mathcal{M}$

$\exists$ true $\theta_0$: $G(z, \theta_0) = G_0(z)$ and $H(z, \theta_0) = H_0(z)$.

Then $\hat{\theta}_N \xrightarrow{N \to \infty} \theta_0$ and $(\hat{\theta}_N - \theta_0) \xrightarrow{N \to \infty} N(0, P_\theta)$

Cov $(\hat{\theta}_N) \sim \frac{1}{N} P_\theta$ : can be estimated from data.

The true $\theta_0$ belongs to an ellipsoid:

$$U_\theta = \{ \theta | (\theta - \hat{\theta}_N)^T P_\theta^{-1} (\theta - \hat{\theta}_N) < \chi \}$$

Situation B: $S \notin \mathcal{M}$

$\hat{\theta}_N \xrightarrow{N \to \infty} \theta^* \neq \theta_0$ : bias error

(Gevers, 2005)
Ljung (1985) was the first to formalize the situation when the true plant does not belong to the model set.

- Modelling error consists of two distinct components:
  - The **bias** error arises when the model structure is unable to represent the true system.
  - The **variance** error is caused by the noise and the finiteness of the data set.

- The variance error goes asymptotically to zero as the number of data goes to zero.

- If the model is not exact, then its quality should reflect its intended use.
Bias, variance and the key role of the input

True system: 6th order, Model: 2nd order

Left: white noise input  Right: low-pass input

(Gevers, 2005)
Bias and Variance Errors

- Introduce $\theta^*$ as the convergence point of the prediction error estimate:

$$
\theta^* = \lim_{N \to \infty} \hat{\theta}_N = \arg \min_{\theta \in \mathcal{D}_\theta} \bar{V}(\theta)
$$

- As defined by Ljung (1985), the bias and variance of the transfer function estimates are:

**Definition: Bias and Variance Errors**

$$
G_0(e^{j\omega}) - \hat{G}(e^{j\omega}, \hat{\theta}_N) = G_0(e^{j\omega}) - \hat{G}(e^{j\omega}, \theta^*) + \hat{G}(e^{j\omega}, \theta^*) - \hat{G}(e^{j\omega}, \hat{\theta}_N)
$$

- Bias

- Variance
Quantifying the Variance

- Define

$$\text{Var}(\hat{G}_N(e^{j\omega})) = E[|G(e^{j\omega}, \hat{\theta}_N) - E[G(e^{j\omega}, \hat{\theta}_N)]|^2]$$

- With $\Phi_u$ and $\Phi_v$ the power spectral densities of $u$ and $H_0e_0$, Ljung (1985) showed that

$$\lim_{n \to \infty} \lim_{N \to \infty} \text{Var}(\hat{G}_N(e^{j\omega})) = \frac{n \Phi_v(\omega)}{N \Phi_u(\omega)}$$

where $n$ is the model order

- Recently, exact expressions for finite-order models have been derived
Pseudo Random Binary Sequences (PRBS)


- A PRBS is a sequence of rectangular pulses of constant magnitude, modulated in width that approximates a white noise, and thus is rich in frequencies.
- A PRBS of length \(2^N - 1\) is generated by means of a shift register of \(N\) stages:

![Shift Register Diagram](image)

( summation modulo 2 )

Figure 5.7.1: *Generation of a PRBS of length \(2^5 - 1 = 31\) sampling periods*
Pseudo Random Binary Sequences (PRBS)

- Maximum duration of a PRBS rectangular pulse is $d_{\text{max}} = NT_s$
- For estimation of static gain, the duration of at least one pulse must be greater than the rise time of the process, $d_{\text{max}} = NT_s > t_R$

![Diagram of PRBS pulse duration with $NT_s > t_R$]

Figure 5.7.2: Choice of a maximum duration of a pulse in a PRBS

- Typically, duration $L$ of test is taken s.t. $L = (2^N - 1)T_s$
- The condition $NT_s > t_R$ might result in too long an experiment
Pseudo Random Binary Sequences (PRBS)

- By choosing the PRBS clock frequency s.t.
  
  \[ f_{PRBS} = \frac{f_s}{p} \quad p = 1, 2, 3, \ldots \]
  
  the condition becomes \( d_{\text{max}} = pNT_s > t_R \)

- If \( p \) is the PRBS frequency divider, then

  \[ d_{\text{max}} = pNT_s \quad L' = pL \quad p = 1, 2, 3\ldots \]

- If \( N \) is increased by \( p - 1 \), then

  \[ d_{\text{max}} = (N + p - 1)T_s \quad L' = 2^{(p-1)}L \quad p = 1, 2, 3\ldots \]
Pseudo Random Binary Sequences (PRBS)

- Dividing the PRBS clock frequency reduces high frequencies while augmenting low frequencies
- Recommended to choose \( p < 4 \)
- For example for \( N = 8 \) and \( p = 1, 2, 3 \)

![Graphs showing spectral density of PRBS sequences with different values of p](Image)

Figure 5.7.3: Spectral density of a PRBS sequence \((f_s = 20 \text{ Hz})\), a) \(N=8, p=1\), b) \(N=8, p=2\), c) \(N=8, p=3\)
Optimal design requires knowledge of the true system!

This advocates iterative method...
- First use a PRBS
- Then, using estimated model, calculate an input optimal for that model
A signal $u$ is called **persistently exciting** of order $n$ if the matrix $C_n$ defined below is positive definite.

\[
C_n = \begin{pmatrix}
  c(0) & c(1) & \ldots & c(n-1) \\
  c(1) & c(0) & \ldots & c(n-2) \\
  \vdots & & \ddots & \vdots \\
  c(n-1) & c(n-2) & \ldots & c(0)
\end{pmatrix}
\]

\[
c(k) = \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{t} u(i)u(i-k)
\]
A signal $u$ is persistently exciting of order $n$ if and only if

$$U = \lim_{t \to \infty} \frac{1}{t} \sum_{k=1}^{t} (A(q)u(k))^2 > 0$$

for all nonzero polynomials $A$ of degree $n - 1$ or less.

- A step is PE of order 1
  $$(q - 1)u(t) = 0$$

- A sinusoid is PE of order 2
  $$q^2 - 2q \cos \omega h + 1)u(t) = 0$$

- To identify a transfer function of order $n$ one needs $n$ sinusoids
Example 1

Model

\[ y(t) + ay(t - 1) = bu(t - 1) + e(t) \]

Parameters

\[ a = -0.9\]
\[ b = 0.5\]
\[ \sigma = 0.5\]
\[ \hat{\theta}(0) = 0\]
\[ P(0) = 100\cdot \]

\[ \hat{\theta} = \begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix} \]

\[ \varphi(t - 1) = \begin{pmatrix} -y(t - 1) \\ u(t - 1) \end{pmatrix} \]
Example 1: Excitation

Input:

- Unit pulse at $t = 50$
- Square wave of unit amplitude and period 100
Example 1: Feedback

Case 1: $u(t) = -0.2y(t)$

Case 2: $u(t) = -0.32y(t - 1)$
Example 1: Forgetting Factor

Parameters: $\lambda = 1, \lambda = 0.999, \lambda = 0.99, \lambda = 0.95$
Example 2: Model Structure

- Consider the process

\[ y(t) - 0.8y(t - 1) = 0.5u(t - 1) + e(t) - 0.5e(t - 1) \]

- First use the estimation model with RLS

\[ y(t) + ay(t - 1) = bu(t - 1) + e(t) \]

- Then use the estimation model with RELS

\[ y(t) + ay(t - 1) = bu(t - 1) + e(t) + ce(t - 1) \]
Example 2: Model Structure

(a) $\hat{b}$

(b) $\hat{b}$

$\hat{a}$

$\hat{c}$