Introduction
Models for identification
Identification methods
Recursive identification
Tracking of time-varying parameters
Pseudo-random binary sequences (PRBS)
Identification in closed loop

Recommended textbook on identification:
System Identification

This is the field of building a dynamic model of a process from experimental data.

It consists of the following:

- Experimental planning
- Selection of model structure
- Parameter estimation
- Validation
Experiments are difficult and costly
Input signal should excite all modes
   Problematic in adaptive control with no active learning
Closed-loop identifiability helped by:
   Nonlinear, or time-varying feedback
   Setpoint changes
Models are really **approximations** of the real system

There is **NO** unique model to describe a process

Structure derived from prior knowledge and purpose of model

Natural to use general linear system, called a **black-box** model

- Impulse response model
- Step response model
- Transfer function model
- General stochastic model
- State-space model
- Laguerre model
A few words of caution:

*Do NOT* fall in love with your model...

*When map disagrees with Nature, TRUST Nature*

Swedish Army Manual
The Impulse Response Model

One of the simplest:

\[ y(k) = \sum_{j=0}^{\infty} h_j u(k - j - 1) \]

\( h_j \) is the \( j \)th element of the impulse response of the process. Assuming that the impulse response asymptotically goes to zero, it may be truncated after the \( n_h \)th element:

\[ y(k) = \sum_{j=0}^{n_h} h_j u(k - j - 1) \]

This is called a Finite Impulse Response (FIR) model and can only be used for stable processes. Used mainly in adaptive filtering.
The Step Response Model

Also one of the simplest: \( y(k) = \sum_{j=0}^{ns} s_j \Delta u(k - j - 1) \)

\( s_j \) is the \( j^{th} \) element of the step response of the process. \( \Delta \) is the differencing operator:

\[
\Delta u(k) = u(k) - u(k - 1) = (1 - q^{-1})u(k)
\]

here \( q \) is the forward-shift operator, i.e. \( qy(k) = y(k + 1) \). Conversely, \( q^{-1}y(k) = y(k - 1) \)

This is the Finite Step Response (FSR) model used in Dynamic Matrix Control (DMC). It can only be used for stable processes.
The Transfer Function Model

This is the classical discrete-time transfer function model:

\[ y(k) = \frac{B(q^{-1})}{A(q^{-1})} q^{-d} u(k - 1) \]

where \( d \) is the time delay of the process in sampling intervals (\( d \geq 0 \)) and the polynomials \( A \) and \( B \) are given by:

\[ A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_{n_A} q^{-n_A} \]
\[ B(q^{-1}) = b_0 + b_1 q^{-1} + \cdots + b_{n_B} q^{-n_B} \]

This can be used for both stable and unstable processes. Note that both FIR and FSR models can be seen as subsets of the transfer function model. This model requires less parameters than FIR or FSR, but assumptions about the orders \( n_A \) and \( n_B \) and the time delay \( d \) must be made.
The General Stochastic Model

This is better known as the Box-Jenkins model, and describes both the process and the disturbance:

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

where \( e(k) \) is a discrete white noise sequence with zero mean and variance \( \sigma_e \).

This general form is usually an overkill for practical applications.
The General Stochastic Model

The form most used in adaptive control is the so-called CARIMA, or ARIMAX model:

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

Putting \( \Delta \) in the numerator of the noise model means that the noise is nonstationary, \( (e(t)/\Delta \) is known as a random walk). This forces an integration in the controller.

This is the model used in Generalized Predictive Control (GPC).
The State-Space Model

Not used very often in adaptive control:

\[ x(k) = A x(k - 1) + B u(k - 1) \]

\[ y(k) = c^T x(k) \]

The ARMAX model can also be written in state-space form. One of the advantages of the state-space representation is that it simplifies the prediction. However, system identification is more complex for a state-space model than for a transfer function model.
The Laguerre Model

The use of this model in adaptive control will be developed in detail later on. In this representation, the process output is represented as a truncated weighted sum of Laguerre functions $l_i(k)$:

$$y(k) = \sum_{i=0}^{N} c_i l_i(k)$$

This representation is much more parsimonious than the FIR or FSR ones, but contrary to the transfer function model does not require assumptions about the order and time delay of the process. Although it is implemented in a state-space form, it is straightforward to use in system identification.
Developed by Karl Gauss in his study of the motion of celestial bodies “…the unknown parameters are chosen so that the sum of the squares of the differences between the observed and the computed values, multiplied by a number that measure the degree of precision is a minimum”

Applied to a variety of problems in mathematics, statistics, physics, economics, signal processing and control

Basic technique for parameter estimation
Example

Assume that data generated by $y = bx + n$ where $n$ is white measurement noise. The problem is to estimate $b$, from $k$ data pairs $(x, y)$.

The least-squares performance index is

$$J = \frac{1}{2} \sum_{i=1}^{k} [y(i) - x(i)b]^2$$
Defining the measurement vector:

\[ y = [y(1), \ldots, y(k)]^T \]

and the regressor:

\[ x = [x(1), \ldots, x(k)]^T \]

one can write

\[ J = \frac{1}{2} [y - bx]^T [y - bx] \]
The value that minimizes $J$ is then found as

$$\frac{dJ}{db} = -x^T[y - bx] = 0$$

The least-squares estimate $\hat{b}$ is then given by the solution to the *normal equation*:

$$-x^Ty + x^T\hat{b}x = 0$$

which, if $[x^Tx]$ is invertible, is

$$\hat{b} = [x^Tx]^{-1}x^Ty$$

- $x^Tx]^{-1}x^T$ is known as the (Moore-Penrose) pseudoinverse.
- This idea can easily be extended to dynamic systems.
Least-Squares Identification

Let a discretized dynamic system be described by

\[ y(t) = \sum_{i=1}^{n} -a_i y(t - i) + \sum_{i=1}^{n} b_i u(t - i) + w(t) \]

where \( u(t) \) and \( y(t) \) are respectively the input and output of the plant and \( w(t) \) is the process noise.

Defining

\[ \theta = [ a_1 \ldots a_n \ b_1 \ldots b_n]^T \]
\[ x = [ -y(t) \ldots -y(t - n) \ u(t - 1) \ldots u(t - n)]^T \]

the system above can be written as

\[ y(t) = x^T(t) \theta + w(t) \]
With $N$ observations, the data can be put in the following compact form:

$$Y = X\theta + W$$

where

$$Y^T = [ y(1) \cdots y(N) ]$$

$$W^T = [ w(1) \cdots w(N) ]$$

$$X = \begin{bmatrix} x^T(1) \\ \vdots \\ x^T(N) \end{bmatrix}$$
Defining the modelling error as
\[ \epsilon(t) = y(t) - x^T(t)\theta \]
the least-squares performance index to be minimized is:
\[ J = \frac{1}{2} \sum_{1}^{N} \epsilon^2(t) = \frac{1}{2} [Y - X\theta]^T[Y - X\theta] \]

Differentiate \( J \) with respect to \( \theta \) and equate to zero. \( \hat{\theta} \), the least-squares estimate of \( \theta \), which, if \([X^T X]\) is invertible, is: \(^{1}\)

\[ \hat{\theta} = [X^T X]^{-1} X^T Y \]

\(^{1}\)I am now tired of having to underline vectors, so I will stop doing so from now on
\[ \hat{\theta} = [X^T X]^{-1} X^T [X \theta + W] \]

or

\[ \hat{\theta} = \theta + [X^T X]^{-1} X^T W \]

Taking the expecting value gives

\[ E(\hat{\theta}) = \theta + E\{[X^T X]^{-1} X^T W\} \]

The second term in the above expression represents a bias in the estimate.
Note that the second term in the RHS is generally non-zero except when

1. \( \{w(t)\} \) is zero mean and uncorrelated, or
2. \( \{w(t)\} \) is independent of \( \{u(t)\} \) and \( y \) does not appear in the regressor
   - FIR and step models
   - Laguerre model
   - Output-error models

Only in the above cases is the least-squares estimate unbiased. Furthermore, if it is the case as the number of observations increases to infinity, then we say that the method gives consistent estimates.
The estimate covariance is then given by

$$\text{cov} \hat{\theta} = E\{[X^T X]^{-1} X^T W W^T X[X^T X]^{-1}\}$$

which, if \{w(t)\} is white, zero-mean gaussian with covariance \(\sigma_w^2\), becomes

$$\text{cov} \hat{\theta} = \sigma_w^2 [X^T X]^{-1}$$

- The matrix \((X^T X)^{-1} X\) is called the pseudo-inverse of \(X\).
- Because of this, \([X^T X]^{-1}\) is sometimes called the covariance matrix.
- The condition that \(X^T X\) be invertible is called an excitation condition.
Alternatives to Least-Squares

- Need a method that gives consistent estimates in presence of coloured noise
  - Generalized Least-Squares
  - Instrumental Variable Method
  - Maximum Likelihood Identification
Generalized Least Squares

- Let the noise model be $w(t) = \frac{e(t)}{C(q^{-1})}$ where $\{e(t)\} = N(0, \sigma)$ and $C(q^{-1})$ is monic, of degree $n$

- System described as

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

- Defining filtered sequences

$$\tilde{y}(t) = C(q^{-1})y(t)$$
$$\tilde{u}(t) = C(q^{-1})u(t)$$

- System becomes

$$A(q^{-1})\tilde{y}(t) = B(q^{-1})\tilde{u}(t) + e(t)$$
If $C(q^{-1})$ is known, then least-squares gives consistent estimates of $A$ and $B$, given $\tilde{y}$ and $\tilde{u}$.

The problem however, is that in practice $C(q^{-1})$ is not known and $\{\tilde{y}\}$ and $\{\tilde{u}\}$ cannot be obtained.

An iterative method proposed by Clarke (1967) solves that problem.
**Generalized Least Squares**

1. Set $\hat{C}(q^{-1}) = 1$
2. Compute $\{\tilde{y}(t)\}$ and $\{\tilde{u}(t)\}$ for $t = 1, \ldots, N$
3. Use least-squares to estimate $\hat{A}$ and $\hat{B}$ from $\tilde{y}$ and $\tilde{u}$
4. Compute the residuals
   \[
   \hat{w}(t) = \hat{A}(q^{-1})y(t) - \hat{B}(q^{-1})u(t)
   \]
5. Use least-squares method to estimate $\hat{C}$ from
   \[
   \hat{C}(q^{-1})\hat{w}(t) = \varepsilon(t)
   \]
   i.e. $\hat{w}(t) = -c_1\hat{w}(t - 1) - c_2\hat{w}(t - 2) - \cdots + \varepsilon(t)$ where $\{\varepsilon(t)\}$ is white
6. If converged, then stop, otherwise repeat from step 2
Generalized Least Squares

- For convergence, the loss function and/or the whiteness of the residuals can be tested.
- An advantage of GLS is that it not only may give consistent estimates of the deterministic part of the system, but also gives a representation of the noise that the LS method does not give.
- The consistency of GLS depends on the signal to noise ratio, the probability of consistent estimation increasing with the S/N.
- There is, however, no guarantee of obtaining consistent estimates.

As seen previously, the LS estimate

\[
\hat{\theta} = (X^T X)^{-1} X^T Y
\]

is unbiased if \( W \) is independent of \( X \).

Assume that a matrix \( V \) is available, which is correlated with \( X \) but not with \( W \) and such that \( V^T X \) is positive definite, i.e.

\[
E[V^T X] \text{ is nonsingular}
\]

\[
E[V^T W] = 0
\]
Then, 

\[ V^T Y = V^T X \theta + V^T W \]

and \( \theta \) estimated by

\[ \hat{\theta} = [V^T X]^{-1} V^T Y \]

\( V \) is called the instrumental variable matrix.
Ideally $V$ is the noise-free process output and the IV estimate $\hat{\theta}$ is consistent.

There are many possible ways to construct the instrumental variable. For instance, it may be built using an initial least-squares estimates:

$$\hat{A}_1 \hat{y}(t) = \hat{B}_1 u(t)$$

and the $k^{th}$ row of $V$ is given by

$$v_k^T = [-\hat{y}(k-1), \ldots, -\hat{y}(k-n), u(k-1), \ldots, u(k-n)]$$
Consistent estimation cannot be guaranteed in general

A two-tier method in its off-line version, the IV method is more useful in its recursive form

Use of instrumental variable in closed-loop

- Often the instrumental variable is constructed from the input sequence. This cannot be done in closed-loop, as the input is formed from old outputs, and hence is correlated with the noise $w$, unless $w$ is white. In that situation, the following choices are available.
Instrumental variables in closed-loop

- Delayed inputs and outputs. If the noise $w$ is assumed to be a moving average of order $n$, then choosing $v(t) = x(t - d)$ with $d > n$ gives an instrument uncorrelated with the noise. This, however, will only work with a time-varying regulator.

- Reference signals. Building the instruments from the setpoint will satisfy the noise independence condition. However, the setpoint must be a sufficiently rich signal for the estimates to converge.

- External signal. This is effect relates to the closed-loop identifiability condition (covered a bit later...). A typical external signal is a white noise independent of $w(t)$. 
The maximum-likelihood method considers the ARMAX model below where $u$ is the input, $y$ the output and $e$ is zero-mean white noise with standard deviation $\sigma$:

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

where

$$A(q^{-1}) = 1 + a_1 q^{-1} + \cdots + a_n q^{-n}$$

$$B(q^{-1}) = b_1 q^{-1} + \cdots + b_n q^{-n}$$

$$C(q^{-1}) = 1 + c_1 q^{-1} + \cdots + c_n q^{-n}$$

The parameters of $A$, $B$, $C$ as well as $\sigma$ are unknown.
Defining
\[ \theta^T = [a_1 \cdots a_n \ b_1 \cdots b_n \ c_1 \cdots c_n] \]
\[ x^T = [-y(t) \cdots u(t-k) \cdots e(t-1) \cdots ] \]
the ARMAX model can be written as
\[ y(t) = x^T(t)\theta + e(t) \]

Unfortunately, one cannot use the least-squares method on this model since the sequence \( e(t) \) is unknown.

In case of known parameters, the past values of \( e(t) \) can be reconstructed exactly from the sequence:
\[ \epsilon(t) = [A(q^{-1})y(t) - B(q^{-1})]/C(q^{-1}) \]
Maximum-Likelihood Identification

\[ V = \frac{1}{2} \sum_{t=1}^{N} \epsilon^2(t) \]

- Minimize \( V \) with respect to \( \hat{\theta} \), using for instance a Newton-Raphson algorithm. Note that \( \epsilon \) is linear in the parameters of \( A \) and \( B \) but not in those of \( C \). We then have to use some iterative procedure

\[ \hat{\theta}_{i+1} = \hat{\theta}_i - \alpha_i (V''(\hat{\theta}_i))^{-1} V'(\hat{\theta}_i) \]

- Initial estimate \( \hat{\theta}_0 \) is usually obtained from a least-squares estimate

- Estimate the noise variance as

\[ \hat{\sigma}^2 = \frac{2}{N} V(\hat{\theta}) \]
If the model order is sufficient, the MLE is consistent, i.e. $\hat{\theta} \to \theta$ as $N \to \infty$.

The MLE is asymptotically normal with mean $\theta$ and standard deviation $\sigma_\theta$.

The MLE is asymptotically efficient, i.e. there is no other unbiased estimator giving a smaller $\sigma_\theta$. 
The Cramer-Rao inequality says that there is a lower limit on the precision of an unbiased estimate, given by

$$\text{cov} \, \hat{\theta} \geq M_{\theta}^{-1}$$

where $M_{\theta}$ is the Fisher Information Matrix

$$M_{\theta} = -E[(\log L)_{\theta \theta}]$$

For the MLE

$$\sigma_{\theta}^2 = M_{\theta}^{-1} = \sigma^2 V_{\theta \theta}^{-1}$$

if $\sigma$ is estimated, then

$$\sigma_{\theta}^2 = \frac{2V}{N} V_{\theta \theta}^{-1}$$
1. Specify a model structure
2. Compute the best model in this structure
3. Evaluate the properties of this model
4. Test a new structure, go to step 1
5. Stop when satisfactory model is obtained
• Most used package
• Graphical User Interface, automates all the steps, easy to use
• Familiarize yourself with it by running the examples