Identification of Nonlinear Systems

The difficulty with nonlinear systems is that a unified model structure is generally not possible. This means that besides requiring a priori information about the order of the system, a priori knowledge about the model structure is also required.

There may be special classes of nonlinear models where the model structure is linear in the unknown parameters. In this case, the identification may be done using linear parameter estimation methods even though the model is nonlinear in the system variables.
Consider the following example:

\[ X_{1k+1} = a_{11} X_{1k}^2 + a_{12} \sin(X_{2k}) + b_1 u_k \]
\[ X_{2k+1} = a_{21} \cos(X_{1k}) + a_{22} X_{2k} + b_2 u_k^2 \]

Assume that the above model structure is known and that the estimates of the parameters, \( a_{ij} \) and \( b_i \) are required. Because the outputs are linear in the parameters, the model may be written as

\[
X_{1k+1} = \begin{bmatrix} X_{1k}^2 & \sin(X_{2k}) & u_k \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ b_1 \end{bmatrix}
\]

\[
X_{2k+1} = \begin{bmatrix} \cos(X_{1k}) & X_{2k} & u_k^2 \end{bmatrix} \begin{bmatrix} a_{21} \\ a_{22} \\ b_2 \end{bmatrix}
\]
Now assuming that the measurements are

\[ Z_{1k} = \chi_{1k} + n_{1k} \]
\[ Z_{2k} = \chi_{2k} + n_{2k} \]

The linear parameter estimation for the first output is setup as follows.

\[
\begin{bmatrix}
Z_{1k+1} \\
Z_{1k+2} \\
\vdots \\
Z_{1k+r}
\end{bmatrix} =
\begin{bmatrix}
Z_{1k}^2 \\
Z_{1k+1}^2 \\
\vdots \\
Z_{1k+r-1}^2
\end{bmatrix}
\begin{bmatrix}
\sin(Z_{2k}) & u_k \\
\sin(Z_{2k+1}) & u_{k+1} \\
\vdots & \vdots \\
\sin(Z_{2k+r-1}) & u_{k+r-1}
\end{bmatrix}
\begin{bmatrix}
a_{11} \\
a_{12} \\
b_1
\end{bmatrix}
+ 
\begin{bmatrix}
n_{1k+1} \\
n_{1k+2} \\
\vdots \\
n_{1k+r-1}
\end{bmatrix}
\]

A similar estimation can be set up for the second set of parameters based on the measurements \( Z_{2k} \).
Provided that the states are directly measurable and that these measurements are a linear function of the model parameters, the linear estimation methods can be applied.

Note that the measurements of the states also appear in the measurement matrix $\mathcal{U}$, and therefore, the effective noise will be correlated through nonlinear filtering. This could lead to a difficult estimate bias problem.

If the original structure of the system model is not linear in the parameters, sometimes it can still be manipulated into a linear estimation problem.

$$\frac{a_2 \chi_{2k} + \chi_{1k}}{a_3 \chi_{1k} \chi_{2k}} \Rightarrow \left( \frac{a_2}{a_3} \right) \frac{1}{\chi_{1k}} + \left( \frac{1}{a_3} \right) \frac{1}{\chi_{2k}}$$
When the structure of the nonlinear system is not known other methods need to be considered. The general nonlinear state space representation can be written as

\[
X_{k+1} = f(X_k, u_k, b, k)\\
Z_k = g(X_k, u_k, b, k)
\]

Then a model structure and model parameters must be found to approximate the nonlinear functions \( f(X_k, u_k, b, k) \) and \( g(X_k, u_k, b, k) \).

One approach to this problem is to choose a power series approximation to each nonlinear function. For example, consider a two state nonlinear system using cubic polynomials to approximate the nonlinear functions.
The parameters of this chosen structure are then identified in the usual manner with a linear parameter estimator.

Other forms of polynomials such as Chebyseher polynomials may also be used.
The General Nonlinear Estimation Problem:

The functional block diagram below illustrates the structure of the general nonlinear parameter estimation problem.

\[ x_{k+1}^m = f(x_k^m, u_k, \beta, k) \]
\[ z_k^m = g(x_k^m, u_k, \beta, k) \]

Adjust \( \beta \) to minimize a scalar function of the error \( e_k = z_k - z_k^m \) for the assumed model structure \( f(\cdot), g(\cdot) \).
A common error function is the integral square error (ISE) or for discrete systems, the sum of the square error.

\[
J = \sum_{i=1}^{N} \varepsilon_i^2
\]

The problem can then be stated as the minimization of \( J \) w.r.t. \( \beta \).

\[
J^* = \min_{\beta} \sum_{i=1}^{N} (\varepsilon_i - Z_i^m)^2
\]

subject to the constraints

\[
\begin{align*}
    \mathbf{X}_{k+1}^m &= f \left( \mathbf{X}_k^m, \mathbf{u}_k, \mathbf{B}, \mathbf{k} \right) \\
    Z_k^m &= g \left( \mathbf{X}_k^m, \mathbf{u}_k, \mathbf{B}, \mathbf{k} \right)
\end{align*}
\]

It should be clear that the cost function, \( J \) is a nonlinear function of the parameter vector \( \beta \) through the variables \( Z^m \) and \( \mathbf{X}^m \).
This problem cannot generally be solved analytically, but can often be solved numerically. The minimum requirement to solve the problem numerically is a function to calculate the cost function $J$ for any given estimate of the parameters $\beta$. These calculations are performed as follows.

1. For a given input perturbation $u_k$, measure the process response $z_k$ for a given period of time. Store the sequences $u_k$ and $z_k$.

2. For an assumed model structure $f(\cdot)$ and $g(\cdot)$, and a parameter vector $\beta$, apply the input sequence $u_k$ from above and calculate the model output sequence $z_k^m$.

3. Calculate the cost function.

$$J = \sum_{i=1}^{N} (z_i - z_i^m)^2$$
The different methods of nonlinear parameter estimation (parameter optimization or nonlinear programming) are characterized by the way in which the parameter vector $\beta$ is updated to achieve a reduction in the cost function $J$.

Note: For every new estimate of the parameter vector $\beta$ in the iterative process, steps 2 and 3 from above must be performed. This can be computationally expensive.

There are two general categories of methods: direct methods, which only require the evaluation of the cost function $J$ on each iterative step; and indirect methods, which require some form of calculation of the expected change in the cost function $\Delta J$ for changes in the parameter vector $\Delta \beta$ as well as the calculation of the cost function $J$. 
Direct Search Methods:

i) Tabular methods: Calculate the cost function $J$ over an organized grid of the parameter space and pick out the minimum. Subdivide the space around the minimum to a finer grid and repeat the process until an accurate parameter estimate is found.
ii) **Creeping Random Search**: Calculate the cost function $J$ for a number of random guesses of $\beta$ in the neighborhood of an initial guess of $\beta$. Select the parameter guess that results in the lowest value of the cost function and make it the new starting point and repeat the process. The succession of these selected parameter vectors should statistically follow a path of decreasing cost functions until the local minimum is found.
iii) **The Simplex Method:** For a two parameter space, select **three points** that form an **equilateral triangle** and evaluate the cost function for each point. Then select as the **apex** the point with the highest cost function and flip the apex over the opposite side. Evaluate the cost function at the new point and repeat the process. When no further reduction of the cost function is achieved, reduce the size of the triangle and continue the process.

For a **three parameter vector**, a **pyramid** is used and for a larger number of parameters higher order equivalents of equidistant points are used.
Indirect Methods:

In these methods, the estimate of the change in the cost function for changes in the parameter vector \( \frac{\Delta J}{\Delta B} \bigg|_{B=B_i} \) is required on each iteration as well as the calculation of the cost function \( J \). These methods are generally called gradient methods.

Consider the Taylor series expansion of the cost function \( J \) about the estimated parameter vector on the \( i^{th} \) iteration \( \theta_i \).

\[
J(B) = J(\theta_i) + \left( \frac{\partial J}{\partial B} \right)^T_{B=\theta_i} (B - \theta_i) \\
+ \frac{1}{2} \left( \frac{\partial^2 J}{\partial B^2} \right)_{B=\theta_i} (B - \theta_i)^2 + \text{H.O.T.}
\]
Consider only the first-order terms.

\[ \Delta J = J(B) - J(B_i) = \left( \frac{dJ}{dB} \right)_{B = B_i}^{T} \Delta B \]

where \( \Delta B = B - B_i \)

To guarantee a negative change in the cost function \( \Delta J \) for a change in the parameter vector \( \Delta B \), choose

\[ \Delta B = B - B_i = -K \left( \frac{dJ}{dB} \right)_{B = B_i} \]

then

\[ B_{i+1} = B_i - K \left( \frac{dJ}{dB} \right)_{B = B_i} \]

This is the first-order gradient method.
First-Order Gradient Method
To maximize the negative cost function change \((-\Delta J)\) on each iteration consider retaining the second order term in the Taylor series expansion.

\[
\Delta J = \left( \frac{\partial J}{\partial \mathbf{B}} \right)^T \Delta \mathbf{B} + \frac{1}{2} \left( \frac{\partial^2 J}{\partial \mathbf{B}^2} \right) \Delta \mathbf{B}^2
\]

Then

\[
\frac{\partial \Delta J}{\partial \Delta \mathbf{B}} = \left( \frac{\partial J}{\partial \mathbf{B}} \right)_{\mathbf{B}=\mathbf{B}_i} + \left( \frac{\partial^2 J}{\partial \mathbf{B}^2} \right)_{\mathbf{B}=\mathbf{B}_i} \Delta \mathbf{B} = 0
\]

gradient vector Hessian Matrix

\[
(\mathbf{B} - \mathbf{B}_i) = - \left( \frac{\partial^2 J}{\partial \mathbf{B}^2} \right)_{\mathbf{B}=\mathbf{B}_i}^{-1} \left( \frac{\partial J}{\partial \mathbf{B}} \right)_{\mathbf{B}=\mathbf{B}_i}
\]

and

\[
\mathbf{B}_{i+1} = \mathbf{B}_i - \left( \frac{\partial^2 J}{\partial \mathbf{B}^2} \right)_{\mathbf{B}=\mathbf{B}_i}^{-1} \left( \frac{\partial J}{\partial \mathbf{B}} \right)_{\mathbf{B}=\mathbf{B}_i}
\]

This is the second-order gradient method.
The first-order gradient methods require the calculation of the gradient vector \( \frac{\partial J}{\partial \beta} \bigg|_{\beta = \beta_i} \) on each iteration. The selection of the gain \( K \) is experimental.

The second-order gradient method also requires the calculation of the Hessian matrix \( \frac{\partial^2 J}{\partial \beta^2} \bigg|_{\beta = \beta_i} \) which is computationally expensive, however, the convergence is much faster than the first-order method.

A compromise is the conjugate gradient method.

\[
\beta_{k+1} = \beta_k + K_k d_k
\]

where, \( K_k \) is chosen to minimize \( J(\beta_k + K_k d_k) \)

\[
d_k = -g_k + \gamma_k d_{k-1}
\]

\[
g_k = \left( \frac{\partial J}{\partial \beta} \right)_{\beta_k}, \quad \gamma_k = \frac{g_k^T g_k}{(g_{k-1}^T g_{k-1})}
\]
contours of constant cost

$J = 120$

$-d\k$

$-g\k$