Dynamic Systems Identification
Part 2 - Nonlinear systems

Reference:
Nonlinear dynamic systems

- Static models
  - Input/output data
  - Interpolation or extrapolation
- Dynamic model
  - Input/output data expanded with delayed input and output data
  - Learning for one-step-ahead prediction
  - Validation for multi-step-ahead prediction

\[ \hat{y}(k) = f(\hat{y}(k-1), \hat{y}(k-2), ..., u(k-1), ...) \]
RBF network – Example

\[ y(k + 1) = 0.2 \tanh(y(k)) + \sin(u(k)) \]

2207 I/O data, 20 basis functions distributed randomly in operating area

Original function

Approksimation with 20 basis functions
Experimental modelling of nonlinear systems

- 1990s: ANN = nonlinear systems identification
- Rule: Do not estimate what you already know!
- white box model, grey box model (physical modelling, semi-physical modelling), black box model
- Nonlinear black-box models: artificial neural networks, fuzzy models, wavelet models, etc.
Experimental modelling of nonlinear systems

- **Used terms (system theory vs. ANN/Machine learning):**
  - estimate, identify = train, learn
  - validate = generalize
  - model structure = network
  - estimation data, identification data = training set
  - validation data = generalization set
  - overfit = overtraining
Some practical concerns about nonlinear systems identification

- Identification procedure cannot/must not be fully automatised!
- Necessary: S/W, I/O data.
- We need experience on similar identification cases.
  Computer simulations of similar cases.
Experimental modelling of nonlinear systems

Nonlinear systems identification problem:
\[ y(k) = g(u(k-1), u(k-2), \ldots, y(k-1), y(k-2), \ldots) + v(k) \]
\[ \hat{y}(k | \theta) = g(\psi(k), \theta) \]

\[ \psi(t) = \text{vector of regressors} \]
\[ \theta = \text{vector of parameters} \]

- 2 subproblems
  - Selection of regressors \( \psi(k) \)
  - Selection of nonlinear mapping \( g(\varphi) \)
Regressors

- **linear systems**

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

- FIR (A=F=D=1, C=0)
- ARX (F=C=D=1)
- OE (A=C=D=1)
- ARMAX (F=D=1)
- BJ (A=1)
- State-space \( x(k) = Ax(k-1) + Bu(k-1) \)

\[ \hat{y}(k) = \theta^T \psi(k, \theta) \]
Regressors

- nonlinear systems

Regressors \( \psi \) determine different models:

- NFIR: \( u(k - i) \)
- NARX: \( u(k-i), y(k-i) \)
- NOE: \( u(k-i), \hat{y}(k-i) \)
- NARMAX: \( u(k-i), y(k-i), \varepsilon(k-i) = y(k-i) - \hat{y}(k-i) \)
- NBJ: \( u(k-i), y(k-i), \varepsilon(k-i), \varepsilon_n(k-i) = y(k-n) - \hat{y}_n(k-i) \)
- State-space

- other possible regressors
Nonlinear mappings

\[ g(\psi, \theta) = \sum \alpha_k g_k(\psi) \]

- Fourier series (scalar case)
- \( g_k(\psi) \) is a basis function
- “Known” structures:
  - wavelet functions
  - B splines
  - ARTIFICIAL NEURAL NETWORKS
    - Multilayer perceptron
    - radial basis function network
    - etc
  - FUZZY MODELS
Forms of known structures:

- Neural networks with sigmoid activation function
  \[ g_k(\psi) = \sigma(\beta_k \psi + \gamma_k) \]
  \( \gamma \) - position
  \( \beta \) - direction
  \( \gamma \) - scale

- Radial basis function networks
  \[ g_k(\psi) = r(\beta_k (\psi - \gamma_k)) \]

- Fuzzy models
  \[ g(\psi) = \sum_j y_j (\prod \mu_A(\psi)) \]
- Multilayer networks
- Recurrent networks

\[ \varphi(k) = g(\psi(k - i), \theta) \]

- Algorithms for estimation of parameters:
  - Gauss-Newton algorithms are very efficient,
  - Off-line identification, as well as on-line (recursive),
  - Gradient optimisation methods – time consuming.
Subproblems:

- Systematic selection of regressors
  - $u(k)$ - static nonlinearity
  - $u(k-i)$
  - $u(k-i), y(k-i)$
  - ...

- Selection of basis function
  - Most of them are universal “approximators”
  - There exist no exact criteria for basis function selection except subjective ones
  - “Radial” functions for low number of regressors (e.g. wavelet function for max. 3 regressors)
  - “ridge” functions for larger number of regressors (e.g. neural networks with sigmoid functions)
  - Fuzzy models, where heuristic knowledge exists.

- Models order ($2n+1$, Takan’s theorem);
Measures of model quality

- Measure of model quality (example)
  \[
  \bar{V}(\theta) = E\|y(t) - g(\psi(t), \theta)\|^2 = \lambda + E\|g_0(\psi(t)) - g(\psi(t), \theta)\|^2
  \]

- Three sources of differences with true system:
  - noise \( e(t) \); variance \( \lambda = E(e^2(t)) \)
  - bias
    \[
    V = E\|g_0(\hat{\theta}_N, \psi(t)) - g^*(\psi(t))\|^2
    \]
  - Variance of estimation
    \[
    V = \lambda \frac{\text{dim } \theta}{N}
    \]
There is always a limitation to model fit.
Recommendations for practice:

- Look at the data
  (detection of nonlinearity, time constants).
- Try simple things first
  (try simple structures and model orders first, at first linear models and small number of estimation parameters).
- Look into the physics (ideas for regressors).
- Validation and estimation data.
- Center and scale the data.
- The bias-variance trade-off.
- The notion of efficient number of parameters (shrinking).
- Sampling time selection (same rule as for linear systems).
  Sampling time should be selected to grasp all interesting process dynamics.
- Input signal selection
  (look at magnitude distribution and input/output distribution of the data)
Model validation

- One-step-ahead prediction
  \[ \hat{y}(k) = g(y(k-1), y(k-2), \ldots u(k-1), u(k-2), \ldots) \]

- Simulation
  \[ \hat{y}(k) = g(\hat{y}(k-1), \hat{y}(k-2), \ldots u(k-1), u(k-2), \ldots) \]
Model validation

- Validation of residuals for one-step-ahead prediction

- Auto correlation function of prediction error

- Cross correlation fct of $u_1$ and prediction error

- Histogram of prediction errors
Model validation

- The consistency of input/output response
  - One-step-ahead prediction
    \[ \hat{y}(k) = g(y(k-1), y(k-2), \ldots, u(k-1), u(k-2), \ldots) \]
  - Simulation
    \[ \hat{y}(k) = g(\hat{y}(k-1), \hat{y}(k-2), \ldots, u(k-1), u(k-2), \ldots) \]
Model validation

- Model reduction (pruning)

- Model purposiveness or usefulness

Network after having pruned 59 weights
Example of identification of 1st order system

Mathematical model of the process with parameters

\[ y(k) = y(k-1) - 0.5 \tanh(y(k-1) + u(k-1)^3) \]

\[ u \text{ – input signal} \]

\[ y \text{ – output signal} \]
Nonlinearity of the system
Histograms of magnitudes and input/output data

- Estimation input signal histogram
- Validation input signal histogram

Magnitude

Amplitude

\( u(k) \)

\( y(k) \)

\( y(k+1) \)
Neural network, regressors, structure and parameters

Used software: NNSYSID Toolbox for Matlab

Regressors:
y(k-1), u(k-1)

Structure:
ARX (model error method)

Optimisation method: Levenberg-Marquardt

\[
\begin{bmatrix}
-0.5588 & -2.0621 & -1.9530 \\
0.5155 & 0.0499 & -0.8670 \\
-1.5149 & 0.3190 & 0.4768 \\
0.3366 & -1.2029 & 1.8379 \\
0.8411 & 1.3841 & 1.7123
\end{bmatrix}
\]

\[
\begin{bmatrix}
1.2054 & 1.7784 & 0.0810 & 1.1704 & 1.4048 & -0.0580
\end{bmatrix}
\]
Model validation (one-step-ahead prediction)
Validation of residuals (one-step-ahead prediction)
Model response and original system’s response on validation data (simulation, not one-step-ahead prediction)
Residuals of simulated validation data

Residuals autocorrelation

Cross-correlation between e and u

Systems modelling from data
Examination assignment

Select a discrete nonlinear system of the 1st order, identify it and make a written report. The report should contain the following items:

- mathematical model of the original system with all parameters;
- Simulation scheme or Matlab code that enables rerun of data acquisition (no masks);
- Show the nonlinearity of the system in 3D plot;
- Estimation signal and response, sampling time;
- Validation signal and response, sampling time;
• Histograms of magnitudes and input/output data pairs;
• Type of neural network and optimisation method;
• Figure of the final network structure;
• Values of parameters (weights);
• Figure of comparison between original system simulation response and model simulation response on validation signal.
• Simulation residuals validation on validation data (the figure of residuals, residuals histogram, $\phi_{ee}$, $\phi_{ue}$).

Eksperiments should be repeatable based on your report only. Reports should not be longer than 6 pages.