Advanced parameter estimation

Evaluating the quality of the estimates obtained by optimization

Katalin Hangos

University of Pannonia
Faculty of Information Technology
Department of Electrical Engineering and Information Systems

hangos.katalin@virt.uni-pannon.hu

March 2019

Lecture overview

1. Parameter estimation using optimization - a repetition
2. Analysing the residuals/prediction errors
   - Residual in the SISO case
   - Multiple output case
3. Analysing the covariances of the estimates
   - Basic LTI case
   - Nonlinear case

The prediction error

The prediction error series can be computed from the measured variables and the model output:

$$\varepsilon(k, p) = y(k) - \hat{y}(k|p) \qquad k = 1, \ldots, N$$

Principle of parameter estimation: A parameter estimation method generates an estimated parameter from the measured data :

$$D^N \rightarrow \hat{p}_N$$

The model is "good", i.e. the estimated parameters are "good" if the prediction errors are "small".

Magnitude of the prediction error
The "size" of the prediction error series $\varepsilon(k, p)$ is measured using an appropriate signal norm .

Minimizing the prediction error

Parameter estimation method: $D^N \to \hat{p}_N$

The general parameter estimation problem :

Given:

- **measured data**: $D[1, N] = D^N = \{(y(k), u(k)) \mid k = 1, \dots N\}$
- **predictive parametrized model** $\hat{y}(k|p) = g(k, D[1, k-1]; p)$
  prediction error series (discrete time signal):
  $\varepsilon(k, p) = y(k) - \hat{y}(k|p) \qquad k = 1, \dots, N$
- norm of the prediction error (**objective function**) - 2-norm, Least Squares: $V_{LS}(p, D^N) = \frac{1}{N} \sum_{k=1}^{N} w_k (\varepsilon(k, p))^2$
  w_k are positive scalar-valued weights, normally $w_k = 1$

Important

From the known D^N measurements and the p parameter vector we can compute the value of the $V_N(p, D^N)$ norm, that is minimized by the estimated $\hat{p}_N$ parameter vector.

Overview

1. Parameter estimation using optimization - a repetition
2. Analysing the residuals/prediction errors
   - Residual in the SISO case
   - Multiple output case
3. Analysing the covariances of the estimates

The residuals in the SISO case

Residual: realization of the prediction error series

$$\varepsilon(k,\theta) = y(k) - \hat{y}(k|\theta) \quad , \quad k = 1, \ldots, N$$

Basic case: ARX model (SISO LTI): the output noise is white

$$A^*(q^{-1}) \cdot y(k) = B^*(q^{-1}) \cdot u(k) + e(k)$$

Predictive form of the model:

$$\hat{y}(k|p) = -a_1 \cdot y(k-1) \ldots - a_n \cdot y(k-n) + b_0 \cdot u(k) + \ldots + b_m \cdot u(k-m)$$

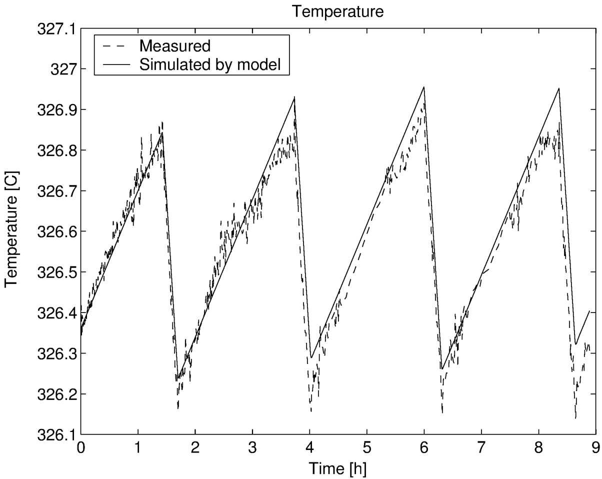
Parameter vector: $p = [-a_1 \;\; -a_2 \;\; \ldots \;\; -a_n \;\; b_0 \;\; b_1 \;\; \ldots \;\; b_m]^\top$
Prediction error (**white noise!**):

$$\varepsilon(k) = \hat{y}(k|p) - y(k) = e(k)$$

Important

For an unbiased estimation in the SISO LTI case, the residuals are uncorrelated and have 0 mean.

Example: quality of the estimate, prediction error



Measured and model computed (predicted) data

The residuals in the MO case

Residuals: defined component wise (individually) for $y_i, \; i = 1, ..., \ell$

$$\varepsilon_i(k, \theta) = y_i(k) - \hat{y}_i(k|\theta) \quad , \quad k = 1, \ldots, N$$

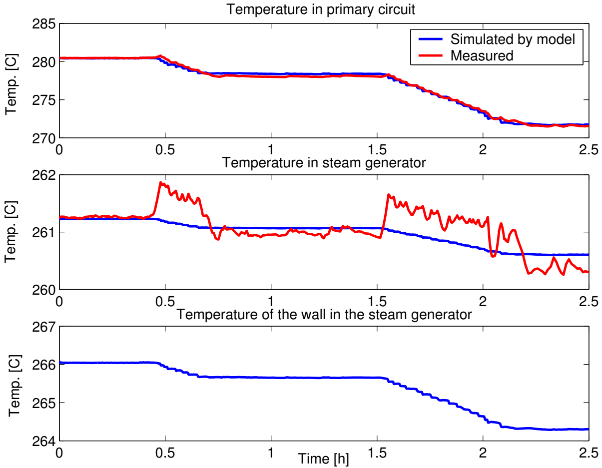
Basic case: component wise ARX (LTI) model: for an unbiased estimation the residuals are uncorrelated and have 0 mean component-wise

- for component wise independent models the residuals should be independent of each other, too
- for correlated individual models this is not true

Important

Most often the components of the output are not equally important and/or of different quality (e.g. measurement precision)

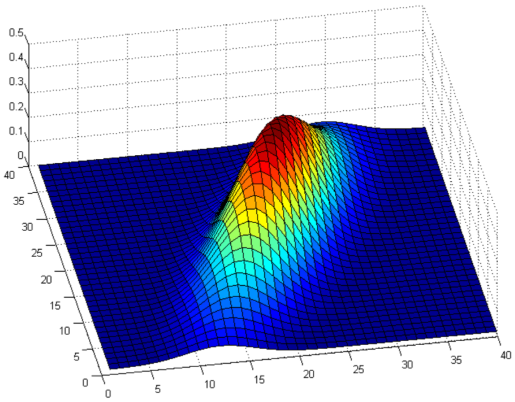
Example: quality of the estimate in the MO case



Measured and model computed (predicted) data

Overview

Multivariate Gaussian distribution – repetition



Level sets are ellipses – **confidence regions**

The parameter estimates in the LTI case

Basic case: ARX model (SISO): the output noise is white and Gaussian

$$A^*(q^{-1}) \cdot y(k) = B^*(q^{-1}) \cdot u(k) + e(k)$$

Parameter vector: $p = [-a_1 \;\; -a_2 \;\; \ldots \;\; -a_n \;\; b_0 \;\; b_1 \;\; \ldots \;\; b_m]^\top$

Important

For an unbiased estimate in the SISO LTI case, the estimate $\hat{p}_{LS}$ has a Gaussian distribution $\mathbb{N}(p, \Sigma)$, where the estimate of the covariance matrix Σ is

$$\hat{COV}\{\hat{p}_{LS}\} = \lambda_0 \cdot \left[\frac{1}{N} \sum_{k=1}^{N} \varphi(k) \cdot \varphi^\top(k) \right]^{-1}$$

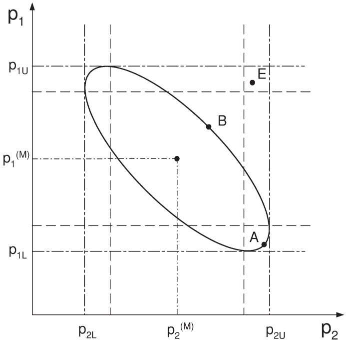
Quality of the estimate in the parameter space – LTI case

Analysis of the covariance matrix

Estimate of the covariance matrix:

$$R(N) = \frac{1}{N} \sum_{k=1}^{N} \varphi(k)\varphi^T(k)\Delta_\varepsilon$$

For a 'good' estimate, the parameter values are **uncorrelated** with **small variances**



Minimizing the prediction error by direct optimization

Method of parameter estimation: $D^N \to \hat{p}_N$

The general task of parameter estimation:
Given

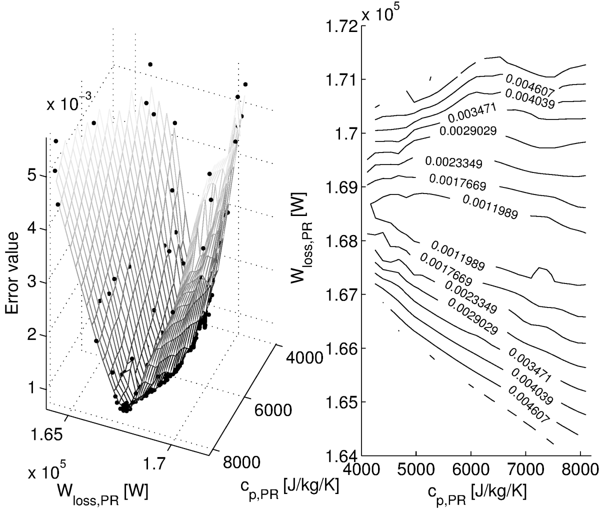
- measured values: $D[1, N] = D^N = \{(y(k), u(k)) \mid k = 1, \ldots, N\}$
- parametrized predictive model: $\hat{y}(k|p) = g(k, D[1, k-1]; p)$
  sequence of prediction errors (discrete-time signal): $\varepsilon(k, p) = y(k) - \hat{y}(k|p) \quad , \quad k = 1, \ldots, N$
- norm defined on the prediction error (2-norm, LS): $V_{LS}(\theta, D^N) = \frac{1}{N} \sum_{k=1}^{N} (\varepsilon(k, p))^2$

Compute:
The estimated parameter $\hat{p}_N$ is specified in the time instant $k = N$ so that

$$\hat{p}_{LS} = \hat{p}_{LS}(D^N) = \arg\min_{p} V_{LS}(p, D^N)$$

The optimum is determined by direct optimization , e.g. by using gradient method.

Example: estimated confidence regions



Level sets of the loss function