Approximate modeling: latency vs misfit

Ivan Markovsky, Jan C. Willems, Bart De Moor, and Sabine Van Huffel
Katholieke Universiteit Leuven, ESAT/SCD

Latency and misfit

Consider the linear static model $AX \approx B$, where A, B are given measurements.

Least squares approximation: $\min_{E,X} \|E\|_{\mathrm{F}}$ subject to $AX = B + E$

Interpretation: E—unobserved, latent variable, $\|E\|_{\mathrm{F}}$ — latency

Total least squares: $\min_{\Delta A, \Delta B, X} \big\|[\Delta A\ \Delta B]\big\|_{\mathrm{F}}$ s.t. $(A + \Delta A)X = B + \Delta B$.

Interpretation: ΔA, ΔB—data corrections, $\min_{\Delta A, \Delta B} \big\|[\Delta A\ \Delta B]\big\|_{\mathrm{F}}$ s.t. $(A + \Delta A)X = B + \Delta B$ — misfit

The latency approach corrects the model in order to make it match the data. The misfit approach corrects the data in order to make it match the model. Both approaches reduce the approximate modeling problem to exact modeling problems.

Note that: exact fit $\iff$ misfit = latency = 0.

Data fitting examples

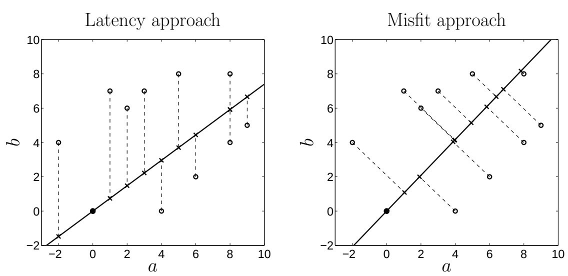
Consider a data set $\{ d_1, \dots, d_{10} \}$ consisting of 2 variables a and b, *i.e.*, $d_i = \left[\begin{smallmatrix} a_i \\ b_i \end{smallmatrix}\right]$.

Line fitting

If the data points were on a line, then they would satisfy a linear equation

$$a_i x = b_i, \quad \text{for } i = 1, \dots, 10 \quad \text{and} \quad \text{for some parameter } x.$$

Optimal fitting lines (—) and data corrections (- - -).



Geometrically, the latency is the sum of the squared vertical distances from the data points to the fitting line and the misfit is the sum of the squared orthogonal distances from the data points to the fitting line.

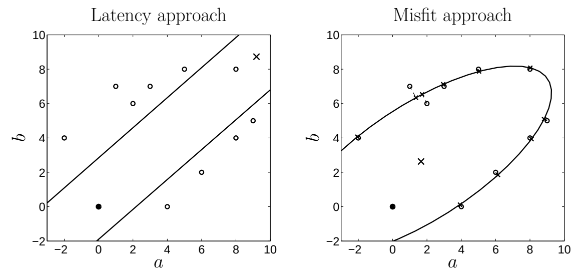
Ellipsoid fitting

If the data points were on an ellipse, then they would satisfy a quadratic equation

$$d_i^\top A d_i + \beta^\top d_i + c = 0, \quad \text{for } i = 1, \dots, 10 \quad \text{and} \quad \text{for some parameters } A, \beta, c.$$

The latency approach leads to what is called the algebraic fitting method and the misfit approach leads to what is called the geometric fitting method.

Optimal fitting ellipses (—) and data corrections (- - -) for the misfit approach.



LTI system identification

In this case the data is viewed as a vector time series. The considered model class consists of LTI systems with one input and one time lag

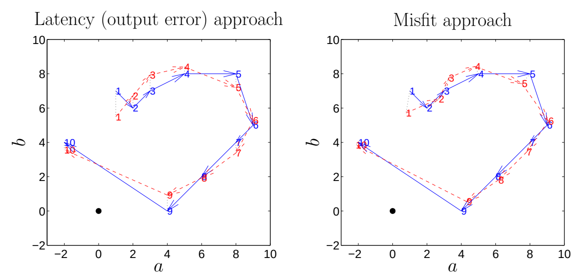
$$Q_0 a_i + Q_1 a_{i+1} = P_0 b_i + P_1 b_{i+1}.$$

The latency approach leads to ARMAX identification:

$$Q_0 a_i + Q_1 a_{i+1} - P_0 b_i - P_1 b_{i+1} = M_0 e_i + M_1 e_{i+1}.$$

The special case $M_0 = P_0$, $M_1 = P_1$ is called output error identification. Then the latent variable e acts like a correction on the output.

Data (—), optimal fitting trajectory (- - -), and data corrections ($\cdots$).



Output error identification can be viewed as the "dynamic least squares method". The misfit approach leads to the global total least squares problem and can be viewed as errors-in-variables system identification.

Classical vs behavioral and stochastic vs deterministic modeling

The data is a collection of points and a static model is a subset of the same space. In the dynamic case, the data set is viewed as an entity—a finite vector time series. A dynamical model is again a subset, however, consisting of time series. Classically the model is viewed as an equation, *e.g.*, $AX = B$ in the linear static.

One and the same modeling method can be derived and "justified" in deterministic as well as stochastic setting.

It is a discussion point what are the virtues and shortcomings of the two approaches.