Physical and Black-Box Models in System Identification: the Case of Vibration Mechanics

Michèle Basseville, Albert Benveniste, Maurice Goursat, and Laurent Mével
IRISA, Rennes, France
To Lennart, who donated, to the system identification galaxy, one new algorithm for each drank bottle of coke
Black-Box Model

- Identification
- Data
- Black-box model
- Model validation
- Change detection
- Y/N
Black-Box and Physical Design Models

identification

data

model validation

change detection

Y/N

black-box model

model validation

model updating

diagnosis

failure

FE design model

data
Model updating in mechanical engineering

Model updating represents 90% of data based modeling effort in the design of mechanical structures subject to vibrations.

The classical way (an ill-posed problem since $\text{dim}($black-box$) \ll \text{dim}($FE$)$):

$\text{data} \rightarrow \text{black-box} \rightarrow \text{FE}$
Model updating in mechanical engineering

model updating is the most used path way to diagnosis of fatigues for structural health monitoring

the classical way (an ill-posed problem since $\text{dim(black-box)} \ll \text{dim(FE)}$):

```
data  \rightarrow  \text{black-box}  \rightarrow  \text{FE}
```
avoids ill-posedness

solves black-box and FE diagnosis

(could be used as a pre-processing for model updating)
Data/model gaps: how to design them?

Lennart to the rescue:

\[ \theta \in \Theta \text{ (model set)} \), \theta_\star \text{ (true system)} \), y_k \text{ (data)} \]

solving for \( \theta : \sum_{k=1}^{N} H(\theta, y_k) = 0 \) yields \( \hat{\theta}_N \)

ODE analysis: \( \theta_\star \) characterized by \( 0 = h(\theta) \triangleq E(H(\theta, y_k)) \)
Data/model gaps: how to design them?

Lennart to the rescue:

\[ \theta_0 \in \Theta \text{ (nominal model), } \theta_* \text{ (true system), } y_k \text{ (data)} \]

\[ \sum_{k=1}^{N} H(\theta_0, y_k) \text{ yields the desired gap} \]

analysis: \( \theta_0 = \theta_* \) iff \( 0 = h(\theta_0, \theta_*) \triangleq E_{\theta_*}(H(\theta_0, y_k)) \)
Data/model gaps: how to design them?

Lennart to the rescue:

\[ \theta_0 \in \Theta \text{ (nominal model)} , \theta_* \text{ (true system)} , y_k \text{ (data)} \]

how to compare \( \sum_{k=1}^{N} H(\theta_0, y_k) \) to 0?

analysis: \( \theta_0 = \theta_* \iff 0 = h(\theta_0, \theta_*) \triangleq E_{\theta_*}(H(\theta_0, y_k)) \)
In the sixties, Le Cam made the following observation: the longer the available data set is, the closer models we wish to discriminate can be.

Thus it makes sense assuming a deviation of order $1/\sqrt{N}$, where $N$ is the sample length,

$$\theta_0 - \theta_* = \frac{1}{\sqrt{N}} \delta,$$

and $\delta$ is a given fixed vector “gap” not depending on sample size.

Normalization (1) is known as the **statistical local approach**. Under (1), a number of simplifications occur that can be exploited for designing data/model gaps.
As a simple example, consider the case of scalar i.i.d. Gaussian random variables $y_k \sim \mathcal{N}(\theta, \sigma^2)$. Assume that $\theta_* = 0$ and compute

$$p_\theta(y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y-\theta)^2}{2\sigma^2}}, \quad -\log p_\theta(y) = \log(\sigma \sqrt{2\pi}) + \frac{(y - \theta)^2}{2\sigma^2}$$

$$\zeta_N(\theta) \triangleq \frac{1}{\sqrt{N}} \sum_{1}^{N} \frac{\partial}{\partial \theta} \log p_\theta(y_k) = \frac{1}{\sqrt{N}} \sum_{1}^{N} \frac{y_k - \theta}{\sigma^2}$$
Data/model gaps: Le Cam’s local approach

As a simple example, consider the case of scalar i.i.d. Gaussian random variables $y_k \sim \mathcal{N}(\theta, \sigma^2)$. Assume that $\theta_\star = 0$ and compute

$$p_{\theta}(y) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(y-\theta)^2}{2\sigma^2}}, \quad -\log p_{\theta}(y) = \log(\sigma \sqrt{2\pi}) + \frac{(y-\theta)^2}{2\sigma^2}$$

$$\zeta_N(\theta) \triangleq \frac{1}{\sqrt{N}} \sum_{1}^{N} \frac{\partial}{\partial \theta} \log p_{\theta}(y_k) = \frac{1}{\sqrt{N}} \sum_{1}^{N} \frac{y_k - \theta}{\sigma^2}$$

Using the local approach, renormalize $\theta_0 - \theta_\star = \frac{1}{\sqrt{N}} \delta$; this yields

$$\zeta_N(\theta_0) = -\frac{\delta}{\sigma^2} + \left(\frac{1}{\sqrt{N}} \sum_{1}^{N} \frac{y_k}{\sigma^2}\right) \Rightarrow \zeta_N(\theta_0) \sim -\frac{\delta}{\sigma^2} + \mathcal{N}\left(0, \frac{1}{\sigma^2}\right)$$

data-to-model gap $\zeta_N(\theta)$ is proportional to signal/noise ratio and subject to a $1/\sigma^2$ random perturbation. Let us generalize this to any $Y_k$ vector i.i.d. random sequence.
Consider a vector i.i.d. random sequence $Y_k \sim p_\theta(y), \theta \in \mathbb{R}^d$. Set

$$H(\theta, y) \triangleq \frac{\partial}{\partial \theta} \log p_\theta(y), \quad h(\theta_0, \theta_\star) \triangleq \mathbb{E}_{\theta_\star} H(\theta_0, Y_k)$$

We know: $\theta_0 = \theta_\star \iff h(\theta_0, \theta_\star) = 0$

Hence $h$ is a candidate nominal-to-true models gap. Estimate $h(\theta_0, \theta_\star)$ using the efficient score $\zeta_N(\theta_0)$:

$$h(\theta_0, \theta_\star) \approx \frac{1}{\sqrt{N}} \zeta_N(\theta_0), \text{ where } \zeta_N(\theta_0) \triangleq \frac{1}{\sqrt{N}} \sum_{k=1}^{N} H(\theta_0, Y_k)$$

With $I(\theta) = \mathbb{E}_\theta \left( H(\theta, Y_k) H^T(\theta, Y_k) \right)$ the Fisher Information Matrix, by Central Limit Theorem:

$$\theta_0 = \theta_\star \Rightarrow \zeta_N(\theta_0) \rightarrow \mathcal{N} \left( 0, I(\theta_\star) \right) \text{ when } N \rightarrow +\infty$$
Now, we follow Le Cam’s statistical local approach:

\[ \theta_0 - \theta_* = \frac{1}{\sqrt{N}} \delta, \quad (2) \]

where \( \delta \) is a given fixed “gap” not depending on sample size.

Under (2), taking a 1\textsuperscript{st}-order Taylor expansion of the efficient score around \( \theta_* \) yields:

\[
\zeta_N(\theta_0) - \zeta_N(\theta_*) \approx \left( \frac{1}{N} \sum_{k=1}^{N} \frac{\partial^2 \log p_{\theta}(Y_k)}{\partial \theta^2} \bigg|_{\theta=\theta_*} \right) \delta \approx -I(\theta_*) \delta,
\]

where we recall that \( I(\theta_*) \) is the \textit{Fisher Information Matrix}.

\[
\theta_0 - \theta_* = \frac{1}{\sqrt{N}} \delta \Rightarrow \zeta_N(\theta_0) \rightarrow N \left( -I(\theta_*) \delta, I(\theta_*) \right) \text{ when } N \rightarrow +\infty.
\]
Likelihood based data-to-model gap

Details:

\[ \zeta_N(\theta_0) - \zeta_N(\theta_*) = \frac{1}{\sqrt{N}} \sum_{k=1}^{N} (H(\theta_0, Y_k) - H(\theta_*, Y_k)) \]

(Taylor expansion) \approx \frac{1}{\sqrt{N}} \sum_{k=1}^{N} \left( \frac{\partial}{\partial \theta} H(\theta, Y_k) \bigg|_{\theta=\theta_*} \right) (\theta_0 - \theta_*)

(statistical local approach) \approx \left( \frac{1}{N} \sum_{k=1}^{N} \frac{\partial^2 \log p_\theta(Y_k)}{\partial \theta^2} \bigg|_{\theta=\theta_*} \right) \delta

(law of large numbers) \approx -\mathbf{I}(\theta_*) \delta,

\[ \theta_0 - \theta_* = \frac{1}{\sqrt{N}} \delta \Rightarrow \zeta_N(\theta_0) \rightarrow \mathcal{N} \left( -\mathbf{I}(\theta_*) \delta, \mathbf{I}(\theta_*) \right) \text{ when } N \rightarrow +\infty. \]
Likelihood based data-to-model gap

Meaning of this result:

the efficient score $\zeta_N(\theta_0) \triangleq \frac{1}{\sqrt{N}} \sum_{k=1}^{N} H(\theta_0, Y_k)$ depends only on known objects:

1/ the nominal model $\theta_0$
2/ the observed data set $Y_1, \ldots, Y_N$

its covariance matrix is the Fisher information matrix, i.e., reflects intrinsic uncertainty about the model

its mean is proportional to the nominal/true model gap, in the direction parallel to the Fisher information matrix: $\zeta_N(\theta_0) = \text{data-to-model gap}$.

$$\theta_0 - \theta_* = \frac{1}{\sqrt{N}} \delta \Rightarrow \zeta_N(\theta_0) \to \mathcal{N} \left( -\mathbf{I}(\theta_*) \delta, \mathbf{I}(\theta_*) \right) \text{ when } N \to +\infty.$$
Pseudo-score based data-to-model gap

\( \text{i.i.d. / likelihood} \)

\[ H(\theta, y) = \frac{\partial}{\partial \theta} \log p_\theta(y) \]

\[ h(\theta_0, \theta_*) = \mathbb{E}_{\theta_*} H(\theta_0, Y_k) \]

\[ h(\theta_0, \theta_*) = 0 \iff \theta_0 = \theta_* \]

score \( \zeta_N(\theta) \triangleq \frac{1}{\sqrt{N}} \sum_{1}^{N} H(\theta, Y_k) \)

\[ \theta_0 - \theta_* = \frac{\delta}{\sqrt{N}} \]

\[ \zeta_N(\theta_0) \sim \mathcal{N}\left(-\mathbf{I}(\theta_*) \delta, \mathbf{I}(\theta_*)\right) \]

\( \mathbf{I}(\theta_*): \) Fisher info matrix
### Pseudo-score based data-to-model gap

<table>
<thead>
<tr>
<th>i.i.d. / likelihood</th>
<th>general / pseudo-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>$H(\theta, y) = \frac{\partial}{\partial \theta} \log p_\theta(y)$</td>
<td>chose $H(\theta, y)$ such that $h(\theta_0, \theta_<em>) = E_{\theta_</em>} H(\theta_0, Y_k)$ yields $h(\theta_0, \theta_<em>) = 0 \iff \theta_0 = \theta_</em>$</td>
</tr>
<tr>
<td>$h(\theta_0, \theta_<em>) = E_{\theta_</em>} H(\theta_0, Y_k)$</td>
<td>pseudo-score $\zeta_N(\theta) \triangleq \frac{1}{\sqrt{N}} \sum_1^N H(\theta, Y_k)$</td>
</tr>
</tbody>
</table>
| $h(\theta_0, \theta_*) = 0 \iff \theta_0 = \theta_*$ | \begin{align*}
\theta_0 - \theta_* &= \frac{\delta}{\sqrt{N}} \\
\downarrow \\
\zeta_N(\theta_0) &\sim \mathcal{N} \left(-I(\theta_*) \delta, I(\theta_*)\right) \\
\downarrow \\
\zeta_N(\theta_0) &\sim \mathcal{N} \left(J(\theta_*) \delta, R(\theta_*)\right)
\end{align*} |
| score $\zeta_N(\theta) \triangleq \frac{1}{\sqrt{N}} \sum_1^N H(\theta, Y_k)$ | $\theta_0 - \theta_* = \frac{\delta}{\sqrt{N}}$ |
| $\theta_0 - \theta_* = \frac{\delta}{\sqrt{N}}$ | $\downarrow$ |
| $\zeta_N(\theta_0) \sim \mathcal{N} \left(-I(\theta_*) \delta, I(\theta_*)\right)$ | $\zeta_N(\theta_0) \sim \mathcal{N} \left(J(\theta_*) \delta, R(\theta_*)\right)$ |
| $\zeta_N(\theta_0) \sim \mathcal{N} \left(-I(\theta_*) \delta, I(\theta_*)\right)$ | $\downarrow$ |
| $\zeta_N(\theta_0) \sim \mathcal{N} \left(J(\theta_*) \delta, R(\theta_*)\right)$ | $\lim_{N \to \infty} \text{cov} \zeta_N(\theta_*)$ |

$I(\theta_*)$: Fisher info matrix

\[ J(\theta_*) = \left. \frac{\partial}{\partial \theta} h(\theta, \theta_*) \right|_{\theta = \theta_*} \]

\[ R(\theta_*) = \lim_{N \to \infty} \text{cov} \zeta_N(\theta_*) \]
Application: Vibrations Monitoring

Our group has been developing the above ideas in the late 80’s and early 90’s... without much success.

Until they found their way in vibrations monitoring and structural analysis. Reasons for this were:

- Small gaps are of interest
- Models (even black-box ones) are very large
  ⇒ heuristics for gaps perform poorly
- Good algorithms: (time and freq domain) subspace, ML
- Scilab toolbox COSMAD to support our methods
  http://www.irisa.fr/sisthem/cosmad/
Simulated bridge structure (courtesy of Etienne Balmes). The structure has 13668 nodes, 9642 finite elements, and 40976 degrees of freedom.
**Left:** stabilization diagram showing 3 frequencies for different model orders. **Right:** model/data gap when the 2 top frequencies are shifted below and above nominal value by 2.5%. 
Use for estimating confidence bounds

Confoundence bounds on a mode shape (real part). To make it readable, the size of the bound is multiplied by a factor of 10.

Error Cov ≈ J^T R^{-1} J

R = \lim_{N \to \infty} \text{cov} \zeta_N(\theta_*)

estimate R by bootstrapping over \zeta_N

compute J with model
Assume, for example, that one wishes to detect changes in a structure in case the ambient temperature also changed.

Problem: changes in temperature may result in bigger changes than a possible damage can do.

Hence there is a need to reject the temperature effect while keeping best sensitive to possible damages.
To this end, let \( H(\theta, \eta, Y_k) \) now depend both on the parameter \( \theta \) of interest (e.g., modes and mode shapes), and on the nuisance parameter \( \eta \) (e.g., the temperature). Let

\[
\zeta_N(\theta, \eta) \triangleq \frac{1}{\sqrt{N}} \sum_{1}^{N} H(\theta, \eta, Y_k)
\]

be the pseudo-score as before. Replace \( \zeta_N \) by its robust version

\[
\zeta_N(\theta/\eta) \triangleq K^T \zeta_N(\theta, \eta)
\]

where \( K \) is orthogonal, and of maximal rank such that

\[
K^T \left( \frac{\partial}{\partial \eta} h(\theta, \eta; \theta_*, \eta_*) \bigg|_{(\theta, \eta)=(\theta_*, \eta_*)} \right) = 0
\]

The robust pseudo-score \( \zeta_N(\theta/\eta) \) will not react to changes in \( \eta \) and will do its best at reacting to changes in \( \theta \).
Use for damage detection with nuisance (3)

Detailed FE simulation study of a bridge: temperature vs. damage. **Left:** effect on the 1\textsuperscript{st} mode of an increase of temperature on safe and damaged structure. **Right:** tests, with (solid) and without (dashed) rejection.

![Graph showing temperature vs. damage](image)
Use for damage detection with nuisance (4)

Detailed FE simulation study of a bridge: temperature vs. damage. **Left:** effect on the 1\textsuperscript{st} mode of an increase of temperature on safe and damaged structure. **Right:** tests, with deeper use of FE model
Use for damage FE diagnosis

\[ \theta_0 = \text{safe system; the system is re-parameterized with new coordinates } \mu, \text{ where changes can be expressed; ex.: } \mu = \text{FE parameters, } \theta = \text{modal parameters.} \]

test

\[ \mu = 0 \text{ (the system did not change at all)} \]

against

\[ \mu \neq 0 \text{ (the } \mu \text{ parameters changed)} \]

this reduces to testing whether \( \mathbb{E} \zeta_N(\theta_0) \neq 0 \) in:

\[
\theta_0 - \theta_* = \frac{1}{\sqrt{N}} J \mu \Rightarrow \zeta_N(\theta_0) \sim \mathcal{N} \left( J(\theta_*) J \mu, R(\theta_*) \right)
\]
$\mu = \text{FE parameters, } \theta = \text{modal parameters, } \dim(\mu) \gg \dim(\theta)$
Damage FE diagnosis: clustering

showing 4 classes
Damage FE diagnosis: simulated damage
Damage FE diagnosis: results
Concluding remarks

- LeCam + Ljung’s local approach allowed us to solve a variety of problems involving \{physical design model, black-box model, data\}

- is the local approach a panacea?
  - what does $\theta_0 - \theta_* = O(1/\sqrt{N})$ mean in practice?
  - so far we assumed that the true system belongs to the model set; still works in practice with model reduction, but theory missing
  - $J$ and $R$ matrices difficult estimating
AR
ARX
ARMA
ARMAX
NLARX
NLARMAX
FDARX
LLARX
LJUX
LJUNGEL