Quantifying identifiability

(Unit 2 of 4)

Eric Walter

Eric.Walter@lss.supelec.fr

Laboratoire des Signaux et Systèmes

Parameter Estimation in Physiological Models, Euro Summer School, Lipari, September 2009.

Abstract:

While the answer to the test of a model for prior identifiability is qualitative (yes / no / sorry, I do not know), one would like to know *how much* the parameters of interest are identifiable.

This relates to *assessing parameter uncertainty*.

This unit concentrates on methods based on the *Fisher information matrix*. Their advantages and limitations are delineated, and alternative routes are briefly mentioned.

More on the material presented here can be found in

Eric Walter and Luc Pronzato, *Identification of Parametric Models from Experimental Data*, Springer, London, 1997.

For knowledge-based models, finding $\hat{\mathbf{p}}$ is not enough, in general.

Important to evaluate the uncertainty attached to $\hat{\mathbf{p}}$.

Several methods can be used (and possibly combined).

None is without drawbacks.

We shall concentrate on methods based on the Fisher information matrix (FIM), because

- they come at almost no cost,
- they nicely connect to (local) identifiability,
- they are simple enough to be usable (in Unit 3) for experiment design.

Alternative approaches will be briefly mentioned, though.

FIM-based methods

Let $\hat{\mathbf{p}}(.)$ be an *unbiased* estimator of $\mathbf{p}^*$, *i.e.*, such that

$$\underset{\mathbf{y}^S|\mathbf{p}^*}{E}\, \hat{\mathbf{p}}(\mathbf{y}^S) = \mathbf{p}^*.$$

[If it were possible to estimate $\hat{\mathbf{p}}$ an infinite number of times, then the mean of the estimates would coincide with the true value $\mathbf{p}^*$.]

The covariance matrix

$$\mathbf{P} = \underset{\mathbf{y}^S|\mathbf{p}^*}{E}\, \{[\hat{\mathbf{p}}(\mathbf{y}^S) - \mathbf{p}^*][\hat{\mathbf{p}}(\mathbf{y}^S) - \mathbf{p}^*]^T\}.$$

quantifies how the estimates $\hat{\mathbf{p}}$ are spread around $\mathbf{p}^*$.

One would like the estimates $\hat{\mathbf{p}}$ to be as concentrated as possible around $\mathbf{p}^*$, of course.

$\hat{\mathbf{p}}_1(.)$ with covariance matrix $\mathbf{P}_1$ is *more efficient* than $\hat{\mathbf{p}}_2(.)$ with covariance matrix $\mathbf{P}_2$ if $\mathbf{P}_1 < \mathbf{P}_2$.

$$\mathbf{P}_1 < \mathbf{P}_2 \Leftrightarrow \mathbf{P}_2 - \mathbf{P}_1 \text{ positive-definite} \Leftrightarrow \text{all eigenvalues of } \mathbf{P}_2 - \mathbf{P}_1 \text{ are strictly positive}.$$

A natural request is thus to make $\mathbf{P}$ as small as possible.

Cramér-Rao inequality provides a lower bound to what can be achieved.

Cramér-Rao inequality

Under technicalities, the covariance $\mathbf{P}$ of any unbiased estimator of $\mathbf{p}^*$ satisfies

$$\mathbf{P} \geq \mathbf{F}^{-1}(\mathbf{p}^*),$$

with $\mathbf{F}$ the FIM.

FIM can be computed in either of two ways:

$$\mathbf{F}(\mathbf{p}) = \underset{\mathbf{y}^s|\mathbf{p}}{\mathrm{E}} \{[\frac{\partial}{\partial \mathbf{p}} \ln \pi_y(\mathbf{y}^s|\mathbf{p})][\frac{\partial}{\partial \mathbf{p}} \ln \pi_y(\mathbf{y}^s|\mathbf{p})]^T\} = -\underset{\mathbf{y}^s|\mathbf{p}}{\mathrm{E}} \{\frac{\partial^2}{\partial \mathbf{p} \partial \mathbf{p}^T} \ln \pi_y(\mathbf{y}^s|\mathbf{p})\}.$$

- $\ln \pi_y(\mathbf{y}^s|\mathbf{p})$ is the log-likelihood of the data $\mathbf{y}^s$,
- $[\frac{\partial}{\partial \mathbf{p}} \ln \pi_y(\mathbf{y}^s|\mathbf{p})]$ is its gradient,
- $\frac{\partial^2}{\partial \mathbf{p} \partial \mathbf{p}^T} \ln \pi_y(\mathbf{y}^s|\mathbf{p})$ is its Hessian.

If $\mathbf{P} = \mathbf{F}^{-1}(\mathbf{p}^*)$ then the estimator is *efficient*.

FIM invertible $\Rightarrow$ all parameters (locally) identifiable.

Nice case: LP model + Gaussian noise

Assume

$$\mathbf{y}^{\mathrm{s}} = \mathbf{R}\mathbf{p}^{*} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon}$ is i.i.d. $\mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, with *known* covariance $\boldsymbol{\Sigma}$.

Likelihood:

$$\pi_{\mathrm{y}}(\mathbf{y}^{\mathrm{s}}|\mathbf{p}) = [(2\pi)^{n_{\mathrm{t}}} \det \boldsymbol{\Sigma}]^{-1/2} \exp\left[-\frac{1}{2}(\mathbf{y}^{\mathrm{s}} - \mathbf{R}\mathbf{p})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}^{\mathrm{s}} - \mathbf{R}\mathbf{p})\right].$$

Gradient of log-likelihood:

$$\frac{\partial}{\partial \mathbf{p}} \ln \pi_{\mathrm{y}}(\mathbf{y}^{\mathrm{s}}|\mathbf{p}) = \mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{y}^{\mathrm{s}} - \mathbf{R}\mathbf{p}),$$

Hessian of log-likelihood:

$$\frac{\partial^2}{\partial \mathbf{p}\partial \mathbf{p}^{\mathrm{T}}} \ln \pi_{\mathrm{y}}(\mathbf{y}^{\mathrm{s}}|\mathbf{p}) = -\mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{R}.$$

FIM then given by:

$$\mathbf{F}(\mathbf{p}) = -\underset{\mathbf{y}^s|\mathbf{p}}{\mathrm{E}} \left\{\frac{\partial^2}{\partial\mathbf{p}\partial\mathbf{p}^T} \ln \pi_y(\mathbf{y}^s|\mathbf{p})\right\} = \underset{\mathbf{y}^s|\mathbf{p}}{\mathrm{E}} \{\mathbf{R}^T\boldsymbol{\Sigma}^{-1}\mathbf{R}\} = \mathbf{R}^T\boldsymbol{\Sigma}^{-1}\mathbf{R}.$$

$\mathbf{F}(\mathbf{p})$ therefore does not depend on the value of $\mathbf{p}$ (VERY convenient!)

Least-squares estimator, weighted by inverse of the noise covariance,

$$\hat{\mathbf{p}}_{ls} = (\mathbf{R}^T\boldsymbol{\Sigma}^{-1}\mathbf{R})^{-1}\mathbf{R}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}^s,$$

is unbiased and efficient. Its covariance matrix is

$$\mathbf{P}_{ls} = (\mathbf{R}^T\boldsymbol{\Sigma}^{-1}\mathbf{R})^{-1} = \mathbf{F}^{-1},$$

and $\mathbf{P}_{ls}$ can be obtain at little or no cost (hidden in the computation of $\hat{\mathbf{p}}_{ls}$).

95% confidence regions for the parameters then easy to derive.

EXAMPLE

Consider the (discrete-time) finite-impulse-response model (FIR)

$$y_{\mathrm{m}}(k, \mathbf{p}) = p_1 u(k-1) + p_2 u(k-2),$$

both LP and LI.

Assume

$$y(k) = y_{\mathrm{m}}(k, \mathbf{p}^*) + n(k), \qquad k = 1, \dots, 10,$$

with $n(k)$ i.i.d. $\mathcal{N}(0, \sigma^2)$.

Then

$$\mathbf{y}^{\mathrm{s}} = \begin{bmatrix} y(1) \\ \vdots \\ y(10) \end{bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} u(0) & u(-1) \\ \vdots & \vdots \\ u(9) & u(8) \end{bmatrix} \quad \text{and} \quad \boldsymbol{\Sigma} = \mathbf{I}\,\sigma^2.$$

So

$$\mathbf{F}(\mathbf{p}) = \mathbf{R}^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}\mathbf{R} = \frac{1}{\sigma^2}\begin{bmatrix} \sum_{i=0}^{9} u^2(i) & \sum_{i=0}^{9} u(i)u(i-1) \\ \sum_{i=0}^{9} u(i)u(i-1) & \sum_{i=0}^{9} u^2(i-1) \end{bmatrix}.$$

FIM depends on *u*, but not on **p**.

Unfortunately, most knowledge-based models are not LP!

Not so nice case: non-LP model + Gaussian noise

Maximum-likelihood estimator then (at best) *asymptotically* normal and *asymptotically* efficient.

$$\hat{\mathbf{p}}_{\mathrm{ml}} \xrightarrow{\text{law}} \mathcal{N}(\mathbf{p}^*, \mathbf{F}^{-1}(\mathbf{p}^*))$$

as number of data points tends to infinity.

Hence the idea of characterizing the uncertainty in $\hat{\mathbf{p}}_{\mathrm{ml}}$ by $\mathbf{F}^{-1}(\hat{\mathbf{p}}_{\mathrm{ml}})$.

Assume, for instance, that

$$y(t_i) = y_{\mathrm{m}}(t_i, \mathbf{p}^*) + \varepsilon(t_i), \quad i = 1, \ldots, n_{\mathrm{t}},$$

where the $\varepsilon(t_i)$'s are independent random variables distributed $\mathcal{N}(0, \sigma_{t_i}^2)$, with $\sigma_{t_i}^2$ known.

Log-likelihood:

$$\ln \pi_{\mathrm{y}}(\mathbf{y}^{\mathrm{s}}|\mathbf{p}) = (\text{term independent of } \mathbf{p}) - \frac{1}{2} \sum_{i=1}^{n_{\mathrm{t}}} \frac{[y(t_i) - y_{\mathrm{m}}(t_i, \mathbf{p})]^2}{\sigma_{t_i}^2} .$$

Gradient of log-likelihood:

$$\frac{\partial}{\partial \mathbf{p}} \ln \pi_{\mathrm{y}}(\mathbf{y}^{\mathrm{s}}|\mathbf{p}) = \sum_{i=1}^{n_{\mathrm{t}}} \frac{1}{\sigma_{t_i}^2} [y(t_i) - y_{\mathrm{m}}(t_i, \mathbf{p})] \frac{\partial}{\partial \mathbf{p}} y_{\mathrm{m}}(t_i, \mathbf{p}).$$

FIM:

$$\mathbf{F}(\mathbf{p}) = \underset{\mathbf{y}^{\mathrm{s}}|\mathbf{p}}{\mathrm{E}} \{[\frac{\partial}{\partial \mathbf{p}} \ln \pi_{\mathrm{y}}(\mathbf{y}^{\mathrm{s}}|\mathbf{p})][\frac{\partial}{\partial \mathbf{p}} \ln \pi_{\mathrm{y}}(\mathbf{y}^{\mathrm{s}}|\mathbf{p})]^{\mathrm{T}}\}$$

$$\mathbf{F}(\mathbf{p}) = \underset{\mathbf{y}^{\mathrm{s}}|\mathbf{p}}{\mathrm{E}} \{ \sum_{k=1}^{n_{\mathrm{t}}} \frac{1}{\sigma_{t_k}^2} \frac{\partial y_{\mathrm{m}}(t_k, \mathbf{p})}{\partial \mathbf{p}} [y(t_k) - y_{\mathrm{m}}(t_k, \mathbf{p})] \sum_{i=1}^{n_{\mathrm{t}}} [y(t_i) - y_{\mathrm{m}}(t_i, \mathbf{p})] \frac{1}{\sigma_{t_i}^2} \frac{\partial y_{\mathrm{m}}(t_i, \mathbf{p})}{\partial \mathbf{p}^{\mathrm{T}}} \}.$$

Since

$$\underset{\mathbf{y}^{\mathrm{S}}|\mathbf{p}}{\mathrm{E}} \{[y(t_k) - y_{\mathrm{m}}(t_k, \mathbf{p})][y(t_i) - y_{\mathrm{m}}(t_i, \mathbf{p})]\} = \sigma_{t_i}^2 \, \delta_{ik},$$

previous nightmare of a formula simplifies into

$$\mathbf{F}(\mathbf{p}) = \sum_{i=1}^{n_{\mathrm{t}}} \frac{1}{\sigma_{t_i}^2} \frac{\partial y_{\mathrm{m}}(t_i, \mathbf{p})}{\partial \mathbf{p}} \frac{\partial y_{\mathrm{m}}(t_i, \mathbf{p})}{\partial \mathbf{p}^{\mathrm{T}}} \quad ,$$

approximation of the Hessian used in the Gauss-Newton algorithm, probably already computed.

EXAMPLE

Consider the (non-LP) model

$$y_{\mathrm{m}}(t_k, \mathbf{p}) = p_1 \exp(-p_2 t_k) \quad \text{for } \; t_k \geq 0,$$

and assume that

$$y(t_k) = y_{\mathrm{m}}(t_k, \mathbf{p}^*) + n(k), k = 1, 2,$$

with $n(k)$ i.i.d. $\mathcal{N}(0, \sigma^2)$.

Then

$$\frac{\partial y_{\mathrm{m}}(t_k,\mathbf{p})}{\partial \mathbf{p}} = \begin{bmatrix} \dfrac{\partial y_{\mathrm{m}}(t_k,\mathbf{p})}{\partial p_1} \\ \dfrac{\partial y_{\mathrm{m}}(t_k,\mathbf{p})}{\partial p_2} \end{bmatrix} = \begin{bmatrix} \exp\,(-p_2 t_k) \\ -p_1 t_k \exp(-p_2 t_k) \end{bmatrix},$$

so

$$\mathbf{F}(\mathbf{p}) = \frac{1}{\sigma^2}\sum_{k=1}^{2}\begin{bmatrix}\exp(-p_2 t_k)\\ -p_1 t_k \exp(-p_2 t_k)\end{bmatrix}\cdot\begin{bmatrix}\exp(-p_2 t_k) & -p_1 t_k \exp(-p_2 t_k)\end{bmatrix}$$

$$= \frac{1}{\sigma^2}\sum_{k=1}^{2}\begin{bmatrix}\exp(-2p_2 t_k) & -p_1 t_k \exp(-2p_2 t_k)\\ -p_1 t_k \exp(-2p_2 t_k) & p_1^2 t_k^2 \exp(-2p_2 t_k)\end{bmatrix}.$$

FIM now depends on measurement times and **p**. ☹

Using $\mathbf{F}^{-1}(\hat{\mathbf{p}}_{ml})$ to characterize the uncertainty in $\mathbf{p}$ relies on the following daring approximations:

- $\mathbf{F}^{-1}(\mathbf{p}^*)$ is substituted for $\mathbf{P}_{ml}$, although number of observations finite, and sometimes quite small, so $\hat{\mathbf{p}}_{ml}$ is generally biased and Cramér-Rao does not apply.
- $\mathbf{F}^{-1}(\hat{\mathbf{p}}_{ml})$ is substituted for $\mathbf{F}^{-1}(\mathbf{p}^*)$.

Despites its shortcomings, this approach has the advantage of requiring far less computation than the alternative methods.

Results are more credible insofar as

- number of data points is large,
- nonlinearity of the model in $\mathbf{p}$ is mild,
- measurement errors are independently distributed and have small magnitudes.

Monte-Carlo methods

Parameter estimation can be summarized as follows:

$$\text{System} \xrightarrow{\text{Experiment}} \text{Data } \mathbf{y}^s \xrightarrow{\text{Estimation}} \text{Estimate } \hat{\mathbf{p}}.$$

Identical experiments yield different results, due to perturbations and noise.

Before data collection, $\mathbf{y}^s$ is thus a random vector, and so is $\hat{\mathbf{p}}(\mathbf{y}^s)$:

$$\text{Random vector } \mathbf{y}^s \xrightarrow{\text{Estimation}} \text{Estimator } \hat{\mathbf{p}}(\mathbf{y}^s).$$

Actual measurements yield a particular realization of $\mathbf{y}^s$, with which a particular estimate $\hat{\mathbf{p}}(\mathbf{y}^s)$ is associated.

Monte-Carlo methods aim to determine statistical characteristics of the population of estimates for all possible realizations of $\mathbf{y}^s$ (*i.e.*, all possible experimental results).

Fictitious data vectors are generated for this purpose:

$$\text{Run of model } M(\hat{\mathbf{p}}) \xrightarrow{\text{Perturbations}} \text{Fictitious data } \mathbf{y}^{sf}.$$

Each $\mathbf{y}^{sf}$ gives a fictitious estimate $\hat{\mathbf{p}}^{\,f} = \hat{\mathbf{p}}(\mathbf{y}^{sf})$, computed as for real data.

$$\{\mathbf{y}^{sf}\} \xrightarrow{\text{Estimation}} \{\hat{\mathbf{p}}^{f}\},$$

Most often, $\hat{\mathbf{p}}^{f}$ is taken to be a normal random vector, and its distribution is simply characterized by the empirical mean and covariance matrix of the fictitious estimates.

This method thus requires a large number of *estimations* (and *a fortiori* of model runs).

The generation of credible fictitious data requires a realistic model for the perturbations, as provided by (successful) maximum-likelihood estimation.

Interesting as it is, this approach seems much too complicated to allow optimization of identifiability through experiment design (see Unit 3), and we shall not consider it further.

The approach to be considered next is one of the ingredients that will allow bypassing the notion of identifiability (see Unit 4).

Bounded-error set estimation

Two kinds of error cannot be avoided when estimating parameters from experimental data.

The first corresponds to measurement errors or state perturbations.

The second corresponds to structural errors, due to the fact that the model structure is at best an approximation of reality.

When

- the parameters to be estimated have a concrete meaning (knowledge-based models),
- or decisions have to be based on their numerical values (prediction, diagnosis, control…)

one should try to evaluate how these two types of error affect the estimates.

Usual statistical framework assumes the data corrupted by realizations of independent random variables:

- Estimator obtained, e.g., by maximum-likelihood.
- Quality of estimate characterized by taking advantage of properties of FIM.

To already mentioned limitations, one may add that this is badly adapted to

- *structural deterministic errors*, e.g., when a simple linear model is used to describe the behaviour of a complex nonlinear model, with perfectly repeatable deviations,
- errors for which the hypothesis of mutual independence is not tenable,
- errors for which only prior information is in the form of bounds.

Bounded-error estimation partly overcomes these limitations.

Bounded-error estimation aims to characterize the set of all values of **p** such that the associated errors lie between given prior bounds.

Consider, *e.g.*, an output error

$$e_{\mathrm{y}}(k, \mathbf{p}) = y(k) - y_{\mathrm{m}}(k, \mathbf{p}), \quad k = 1, \ldots, n_{\mathrm{t}},$$

where $y(k)$ is the kth scalar measurement, and $y_{\mathrm{m}}(k, \mathbf{p})$ is the corresponding model output.

Assume **p** to be feasible if and only if

$$e_y^{\mathrm{m}}(k) \leq e_{\mathrm{y}}(k, \mathbf{p}) \leq e_y^{\mathrm{M}}(k), \quad k = 1, \ldots, \mathrm{n_t},$$

where the bounds $e_y^{\mathrm{m}}(k)$ and $e_y^{\mathrm{M}}(k)$ are known *a priori* and differ.

These bounds may come from technical specifications, empirical knowledge, or intuition.

We aim to characterize the set of all feasible values of **p**.

Let $\mathbb{P}_{n_\mathrm{t}}$ be the set associated with the first n_t data.

Provided $e_y^\mathrm{m}(k) \neq e_y^\mathrm{M}(k)$, the inequalities associated with the kth measurement can be put in the *standard* form:

$$-1 \leq \bar{y}(k) - \overline{y_\mathrm{m}}(k, \mathbf{p}) \leq 1,$$

with

$$\bar{y}(k) = = \frac{2y(k) - e_y^\mathrm{M}(k) - e_y^\mathrm{m}(k)}{e_y^\mathrm{M}(k) - e_y^\mathrm{m}(k)} \quad \text{and} \quad \overline{y_\mathrm{m}}(k, \mathbf{p}) = \frac{2}{e_y^\mathrm{M}(k) - e_y^\mathrm{m}(k)}\, y_\mathrm{m}(k, \mathbf{p}).$$

We assume this transformation has been performed, and drop the upper bars.

The set of **p** values that satisfy the constraints associated with $y(k)$ is then:

$$\mathbb{\Pi}_k = \{\mathbf{p} \in \mathbb{R}^{n_\mathrm{p}} \mid -1 \leq y(k) - y_\mathrm{m}(k, \mathbf{p}) \leq 1\},$$

$\mathbb{P}_{n_\mathrm{t}}$ is the intersection of the sets $\mathbb{\Pi}_k$ ($k = 1, \ldots, n_\mathrm{t}$).

$\mathbb{P}_{n_\mathrm{t}}$

- is a set estimator,
- may be empty if the hypotheses are wrong,
- is guaranteed to contain $\mathbf{p}^*$ (if it exists).

Size of $\mathbb{P}_{n_\mathrm{t}}$ is a measure of the uncertainty on $\mathbf{p}^*$.

The algorithms used to characterize $\mathbb{P}_{n_\mathrm{t}}$ depend on whether $M(\cdot)$ is LP (more generally on whether the error is affine in the parameters).

LP model structures

Consider an LP model structure, defined by

$$y_{\mathrm{m}}(k, \mathbf{p}) = \mathbf{r}^{\mathrm{T}}(k)\mathbf{p},$$

and an *output* error.

Since regressor vector $\mathbf{r}(k)$ is assumed known, $\mathbb{\Pi}_k$ is a strip in parameter space, bounded by parallel hyperplanes

$$\mathbb{H}^{+} = \{\mathbf{p} \in \mathbb{R}^{n_{\mathrm{p}}} \mid y(k) - \mathbf{r}^{\mathrm{T}}(k)\mathbf{p} = 1\}$$

and

$$\mathbb{H}^{-} = \{\mathbf{p} \in \mathbb{R}^{n_{\mathrm{p}}} \mid y(k) - \mathbf{r}^{\mathrm{T}}(k)\mathbf{p} = -1\}.$$

As the intersection of such strips, $\mathbb{P}_{n_{\mathrm{t}}}$ is a convex polyhedron.

Provided the $\mathbf{r}(k)$'s span $\mathbb{R}^{n_\mathrm{p}}$ (condition for identifiability), $\mathbb{P}_{n_\mathrm{t}}$ is bounded, and thus a polytope.

$\mathbb{P}_{n_\mathrm{t}}$ may become very complicated if n_t and especially dim $\mathbf{p}$ are large.

Hence the interest of constructing sets with simple shapes guaranteed to enclose $\mathbb{P}_{n_\mathrm{t}}$.

The most widely used are ellipsoids, orthotopes (boxes), parallelotopes and polyhedra with limited complexity (*e.g.*, simplexes).

Recursive determination of outer ellipsoids

Data taken into account one after the other to construct a succession of ellipsoids containing all values of **p** consistent with all previous measurements.

This is easily implementable on-line (possibly in real time).

After the first $k-1$ observations, $\mathbb{P}_{k-1}$ is characterized by the ellipsoid

$$\mathbb{E}(\hat{\mathbf{p}}^{k-1}, \mathbf{M}_{k-1}) = \{\mathbf{p} \mid (\mathbf{p} - \hat{\mathbf{p}}^{k-1})^{\mathrm{T}} \mathbf{M}_{k-1}^{-1} (\mathbf{p} - \hat{\mathbf{p}}^{k-1}) \leq 1\},$$

where $\hat{\mathbf{p}}^{k-1}$ is the centre, and $\mathbf{M}_{k-1}$ a positive-definite matrix that specifies size and orientation.

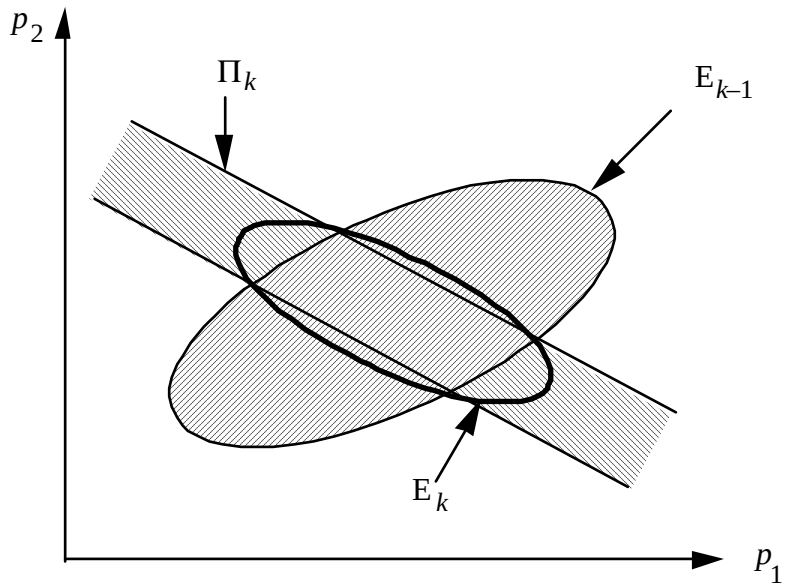
The volume of this ellipsoid is

$$\mathrm{vol}\, \mathbb{E}(\hat{\mathbf{p}}^{k-1}, \mathbf{M}_{k-1}) = \mathcal{V}(\dim \mathbf{p}) \sqrt{\det \mathbf{M}_{k-1}}\,,$$

with $\mathcal{V}(\dim \mathbf{p})$ the volume of the unit ball in parameter space.

Simple algorithms are available for computing $\hat{\mathbf{p}}^{k}$ and $\mathbf{M}_k$ in such a way that

$$\mathbb{E}(\hat{\mathbf{p}}^{k}, \mathbf{M}_k) \supseteq \mathbb{E}(\hat{\mathbf{p}}^{k-1}, \mathbf{M}_{k-1}) \cap \mathbb{\Pi}_k \supseteq \mathbb{P}_{n\mathrm{t}}$$



One of them minimizes the volume of $\mathbb{E}(\hat{\mathbf{p}}^{k}, \mathbf{M}_k)$, which amounts to minimizing det $\mathbf{M}_k$.

Non-recursive determination of outer boxes

Determination of parameter uncertainty intervals for each components of **p** amounts to finding the smallest box with edges parallel to the axes enclosing $\mathbb{P}_{n_t}$.

Complexity of resulting description of $\mathbb{P}_{n_t}$ is moderate: $2n_p$ scalars for a box, versus $n_p(n_p + 3)/2$ scalars for an ellipsoid, where $n_p = \dim \mathbf{p}$.

The bounds of the ith parameter uncertainty interval are given by the minimal and maximal values of the cost $j(\mathbf{p}) = p_i$, when the feasible domain for **p** is $\mathbb{P}_{n_t}$, which is defined by linear inequalities.

The cost and constraints being linear, these bounds can be computed by linear programming.

Determination of the box thus requires solution of $2n_p$ linear-programming problems, with $2n_t$ linear constraints each.

Exact description

When the error is affine in $\mathbf{p}$, $\mathbb{P}_{n\mathrm{t}}$ can be written as

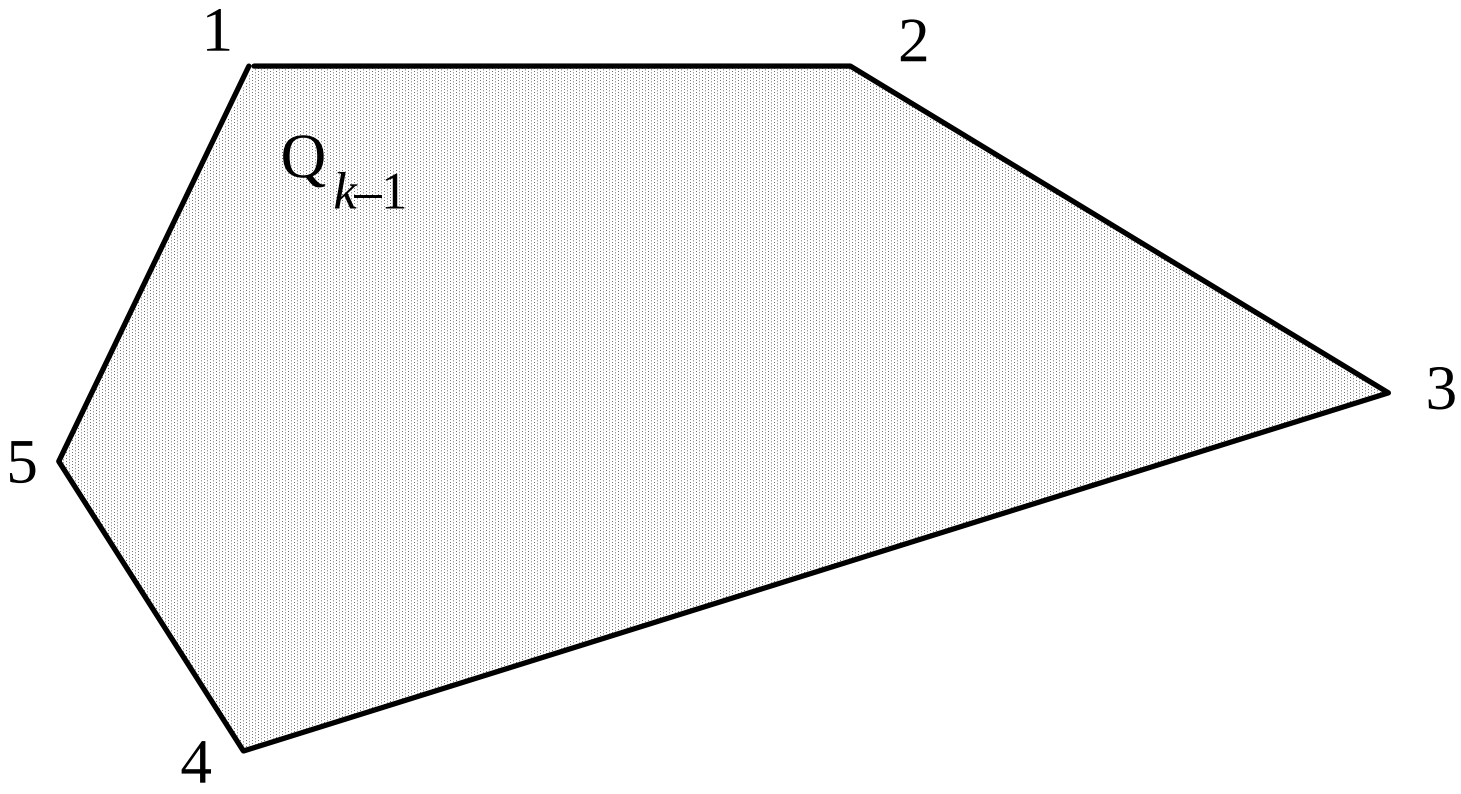
$$\mathbb{P}_{n\mathrm{t}} = \{\mathbf{p} \mid \mathbf{A}\mathbf{p} \geq \mathbf{b}\},$$

and computed recursively.

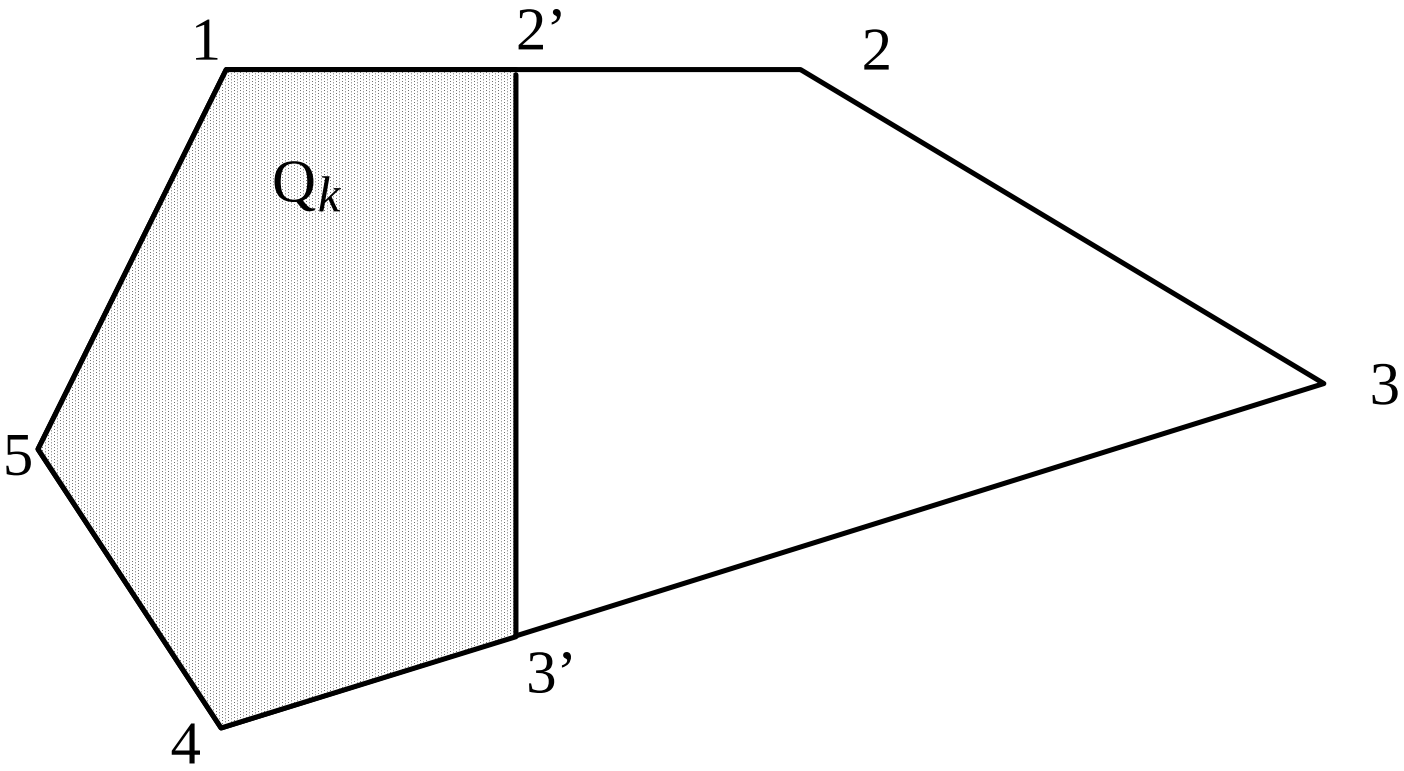
Each observation provides two inequalities, which are taken into account one after the other.

The basic idea of the algorithm is as follows:

Assume that the polyhedron $\mathbb{Q}_{k-1}$ formed by the first $(k-1)$ inequalities is



The only case deserving discussion is when the new inequality cuts into $\mathbb{Q}_{k-1}$:

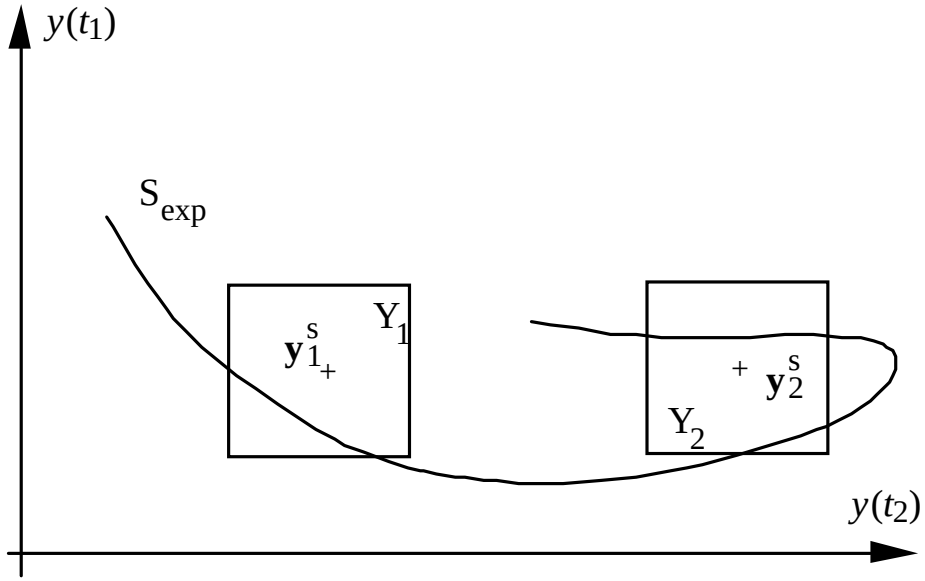


Important to keep track of adjacencies between vertices...

Non-LP model structures

The rule for knowledge-based models.

Posterior feasible set $\mathbb{P}_{n_t}$ no longer a polyhedron, may be non-connected, even if model s.g.i. Phenomenon best understood in the space of observations.



Set inversion

Yields *global results* even in the case of non-LP model structures. See Unit 4 for more details.

Assume posterior feasible set for **p** defined by

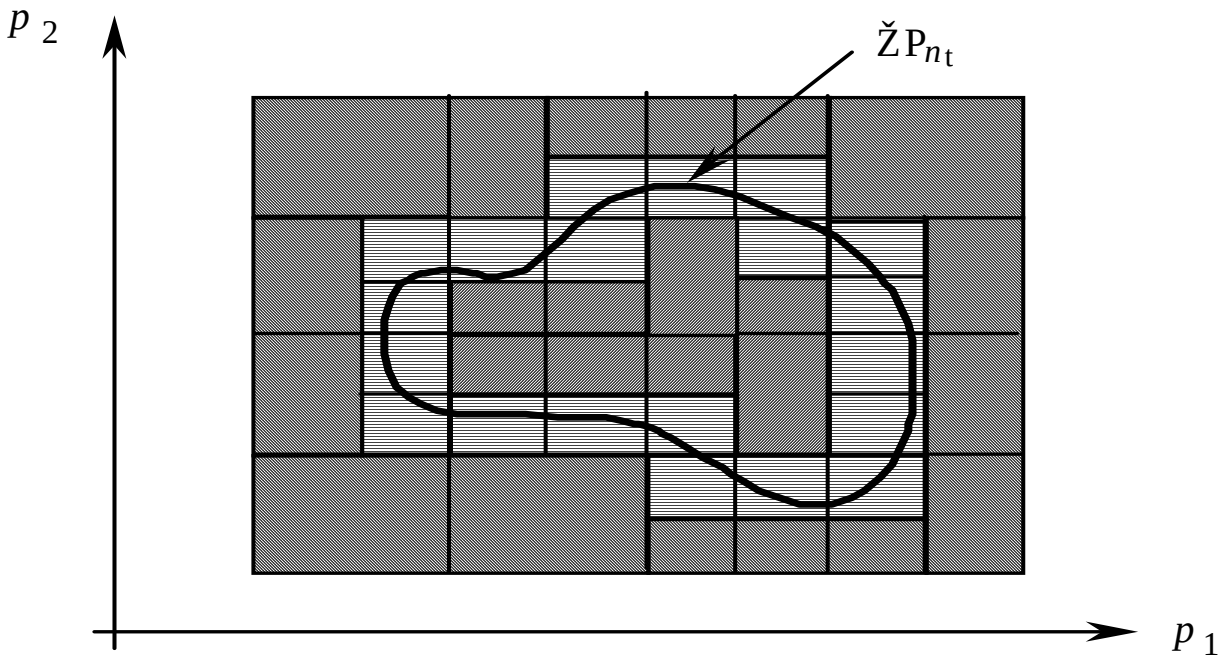
$$\mathbb{P}_{nt} = \{\mathbf{p} \in \mathbb{P}_0 \mid \mathbf{e}(\mathbf{p}) \in \mathbb{E}\},$$

where $\mathbb{P}_0$ is a prior box in parameter space, and $\mathbb{E}$ is a given box in error space. Then

$$\mathbb{P}_{nt} = \mathbf{e}^{-1}(\mathbb{E}) \cap \mathbb{P}_0,$$

where $\mathbf{e}^{-1}$ is the inverse function (in a set-theoretic sense) of the error function **e**.

Determining $\mathbb{P}_{nt}$ is thus a *set-inversion* problem, to be solved though approximate but guaranteed computations.



Subpaving K_{in} of all boxes that have been proved feasible

Subpaving K_i of all boxes still indeterminate

Subpaving K_{un} of all boxes that have been proved unfeasible

Conclusions

A variety of techniques available to characterize parameter uncertainty.

None can claim to be the best.

Method based on the inverse of the FIM is the simplest and most economical, but only justified asymptotically, except in important but very particular cases.

All methods rely on hypotheses that are not necessarily satisfied in practice, so one must be careful in interpreting their results.

The experiment performed (location of sensors and actuators, shape of the inputs applied, measurement times…) affects the uncertainty in the parameters.

Thus important to choose the experiment well, to maximize identifiability.

This is experiment design, to be considered in Unit 3.