PARAMETRIC IDENTIFICATION USING GLOBAL OPTIMIZATION METHODS

E. Balsa-Canto and J.R. Banga

IIM-CSIC (Spanish National Research Council)
Permanent staff

- J. R. Banga
- E. Balsa-Canto
- A. A. Alonso

Hired researchers

- L. Antelo
- J. A. Egea
- I. Otero-Muras
- J. O. H. Sendín
- M. Rodríguez-Fernández
- A. Franco
- M. Villafín

- Modelling
- Simulation
- Model analysis
- Parameter estimation
- Experimental design
- Dynamic optimization
- Global optimization
- Model identification
- Robust control in...
  (biochemical proc. & biological systems)
Outline

> Parameter estimation

- Problem definition
- Numerical techniques: single shooting vs. multiple shooting
- Typical difficulties: lack of practical identifiability, multimodality
- Global optimization methods
- Illustrative examples

> Related on-going work in our group

- Structural and practical identifiability analysis
- Global sensitivity analysis
- Optimal experimental design
The model calibration problem consists on finding the model unknowns to minimize the distance among the model predictions and the experimental data.

Mathematical model $M(p)$

The mathematical model will consist on two essential elements:

i) The set of **ordinary** differential equations describing the system behavior
   - $x \in X \subset \mathbb{R}^{n_x}$, vector of state variables
   - $u \in U \subset \mathbb{R}^{n_u}$, vector of control variables, external factors or stimuli
   - $p \in P \subset \mathbb{R}^{n_p}$, vector of parameters in the model
   - $t$, time

ii) The **observation function**, describing the relationship between the states in the model and the available measured quantities
   - $y^e \in Y \subset \mathbb{R}^{n_y}$, $n_o$ vectors of discrete time measurements per experiment
   - $n_{o,e}^\circ$, number of sampling times per observable per experiment
Experimental Scheme ε

The experimental scheme collects all information related to the experiment ε:

i) Observed / measured quantities
ii) Stimuli conditions
iii) Spatial domain of interest (if relevant)
iv) Locations (if spatial domain is relevant)
v) Experiment duration
vi) Sampling times
vii) Experimental noise: type/ quantity (if available)
Elements in the problem definition

**The model**

\[
\begin{align*}
\dot{x} &= f(x, u, p, t) \\
y^e(t_s, u, p) &= g(x(u, p, t_s), p, t) \\
s &= 1, \ldots, n_s^e
\end{align*}
\]

**Model unknowns**

\[
[\theta \; \theta_y^e]
\]

**Experimental data**

\[
y_m^e(t_s) \quad s = 1, \ldots, n_s^e
\]

**“Distance measure”**

Its definition depends on the prior information available:
- distribution in the parameters
- experimental noise

> **(Generalized) Least squares and modulus estimation:** almost no prior information is considered.

> **Maximum likelihood:** a probabilistic distribution in the noise is considered but without considering any uncertainty in the parameters.

> **Bayesian estimation:** introduces information about a probabilistic distribution in the parameters and noise.

(Not to be addressed here).
PARAMETER ESTIMATION

Problem Formulation

**Generalized least squares**

\[
J(\theta) = \sum_{\varepsilon=1}^{n_{\varepsilon}} \sum_{o=1}^{n_{o}} (y_{\varepsilon,o}^{}(\theta) - y_{m_{\varepsilon,o}}^{}(\theta))^T Q_{\varepsilon,o}^{}(y_{\varepsilon,o}^{}(\theta) - y_{m_{\varepsilon,o}}^{})
\]

> Quadratic cost functions are the most commonly used
> For linear models in the parameters (LP) the best estimate can be analytically obtained
> \( Q_{\varepsilon,o}^{} \) is a nonnegative definite symmetric weighting matrix. The weighting coefficients \( (\omega_{\varepsilon,o}^{}, s = 1, \ldots, n_{s}^{\varepsilon,o}) \) located in the diagonal of the matrix are positive or zero and fixed a priori.
> The solution \( \theta^* = \arg\min J(\theta) \) is regarded as the (weighted) L$_2$ estimator

**Generalized least modulus**

\[
J(\theta) = \frac{1}{n_{m}} \sum_{\varepsilon=1}^{n_{\varepsilon}} \sum_{o=1}^{n_{o}} \sum_{s=1}^{n_{s}^{\varepsilon,o}} \omega_{s}^{\varepsilon,o} \left| y_{s}^{\varepsilon,o}(\theta) - y_{m_{s}^{\varepsilon,o}} \right|
\]

\( n_{m} \): total number of experimental data
> Penalize large errors less than quadratic costs
> The solution is regarded as the (weighted) L$_1$ estimator

**Maximum log-likelihood**

\[
J(\theta) = \ln(\pi_{y}(y_{m}|\theta))
\]

> The maximum (log-)likelihood method looks for the value of the model unknowns that give the highest likelihood to the observed data.
> This approach allows to introduce in the design of the cost function the available information on the nature of the experimental noise.
> The type of probability density function will condition the type of function to be maximized: a Gaussian distribution is usually assumed in practice.
Problem Formulation

**Gaussian noise with known or constant (homocedastic) variance**

\[
\ln(\mathcal{P}_y (y^m|\theta)) = - \frac{1}{2} \sum_{s=1}^{n_s} \left( \frac{y_s(\theta) - y_m}{\sigma_s} \right)^2
\]

> Remark that in this case a L2 estimator is obtained by fixing the weights to the inverse of the variance of the associated noise.

> This estimator can be used even if the noise does not follow a Gaussian distribution, provided the values of the variances are known. It is called the *Gauss-Markow estimator*.

**Gaussian noise with unknown varying variance (heterocedastic)**

\[
\sigma_s^2 = a \left| y_s(\theta) \right|^b
\]

\[
J(\theta, b) = n_s \ln \left( \frac{1}{n_s} \sum_{s=1}^{n_s} \left( \frac{y_s(\theta) - y_m}{y_s(\theta)} \right)^b \right) + b \sum_{s=1}^{n_s} \ln \left| y_s(\theta) \right|^b
\]
Non-linear optimization problem

Find \( \begin{bmatrix} \theta \theta_y^\varepsilon \end{bmatrix} \) so as to minimize:

\[ J(\theta) \]

subject to the system dynamics:

\[ \dot{x} = f(x, u, p, t) \]
\[ y^\varepsilon(t_s, u, p) = g(x(u, p, t_s), p, t) \quad s = 1, \ldots \]

and bounds on the unknowns:

\[ \begin{bmatrix} \theta \theta_y^\varepsilon \end{bmatrix}_{\min} \leq \begin{bmatrix} \theta \theta_y^\varepsilon \end{bmatrix} \leq \begin{bmatrix} \theta \theta_y^\varepsilon \end{bmatrix}_{\max} \]
PARAMETER ESTIMATION

Typical difficulties

- Large number of parameters

4 ODEs, 9 parameters

14 non-linear ODEs, 30 parameters

More than 60 non-linear ODEs, more than 60 parameters

Number of parameters increases very rapidly with network (and model) complexity.
PARAMETER ESTIMATION

Typical difficulties

- Large number of parameters
- Limited Quantity and Quality of data
- Even the order of magnitude of the parameters is unknown

- Multimodal character of the resultant NLP
- Identifiability problems
PARAMETER ESTIMATION

Optimization problems

Multimodal NLP

Global maximum

Local maxima

Convex NLP

Global minimum

Local minimum
PARAMETER ESTIMATION

Optimization problems

Simple constrained problem

Unconstrained global maximum

Constrained global maximum

Constrained global minimum

Unconstrained global minimum
Iterative procedures

The way the iterates are computed determines the type of method:

Local / Global
Direct / Indirect
Adaptive / Population based
...

Optimization methods
PARAMETER ESTIMATION

**Optimization method**

**Single Shooting**

- Solves the system dynamics for every new iterate
- Intermediate solutions may be very far from the optimum
- Simulation may become unstable for some particular systems for some particular iterates
- The most widely used approach in combination with local and global NLP solvers
- Implementations with global methods overcome multimodality

**Multiple Shooting**

- The dynamics is partitioned in a number of shooting elements (less or equal that number of sampling times)
- The trajectory is discontinuous and the initial conditions in the intermediate steps become decision variables
- Constraints are necessary to enforce continuity in the optimum
- Intermediate solutions are usually closer to the experimental data
- Reduces multimodality but the cost to pay is a higher number of decision variables
- Current implementations combined with local constrained optimization methods

H. Bock, 1981

**PARAMETER ESTIMATION**

**Local Optimization methods**

**LOCAL**

- **Direct**
  - Nelder-Mead

- **Indirect**
  - First order
    - Steepest Descent
  - Second order
    - Newton /qN
    - GC
    - L-BFGS
    - TN
    - SQP

  ✓ Proofs of convergence (to local extrema)
  ✓ Proofs of convergence (V. Torczon and col, 2002/2003)

- For least-squares minimization
  - Gauss - Newton
  - Levenberg - Marquardt

Local methods for parameter estimation:
Illustrative examples

Thermal isomerization of $\alpha$-pinene

$$\frac{dy_1}{dt} = -(k_1 + k_2)y_1$$
$$\frac{dy_2}{dt} = k_1 y_1$$
$$\frac{dy_3}{dt} = k_2 y_1 - (k_3 + k_4)y_3 + k_5y_5$$
$$\frac{dy_4}{dt} = k_3 y_3$$
$$\frac{dy_5}{dt} = k_4 y_3 - k_5 y_5$$

Nominal values for the parameters:

$k_1 = 5.93 \cdot 10^{-5}$; $k_2 = 2.96 \cdot 10^{-5}$;
$k_3 = 2.05 \cdot 10^{-5}$; $k_4 = 27.5 \cdot 10^{-5}$
$k_5 = 4.00 \cdot 10^{-5}$

Multistart of a local solver

<table>
<thead>
<tr>
<th>Box 1</th>
<th>[0.0, 10^{-3}]</th>
</tr>
</thead>
<tbody>
<tr>
<td>Box 2</td>
<td>[0.0, 10^{-2}]</td>
</tr>
<tr>
<td>Box 3</td>
<td>[0.0, 1.0]</td>
</tr>
</tbody>
</table>
PARAMETER ESTIMATION

Local Optimization methods

The Goodwin oscillator

\[
\begin{align*}
\dot{x}_1 &= \frac{a_1}{A_1 + z_1^{10}} - b_1 x_1 \\
\dot{y}_1 &= a_2 x_1 - \beta_1 y_1 \\
\dot{z}_1 &= \gamma_1 y_1 + \delta_1 z_1 ,
\end{align*}
\]

\[a = 3.4884, \ A = 2.1500,\]
\[b = 0.0969, \ \alpha = 0.0969,\]
\[\beta = 0.0581, \ \gamma = 0.0969, \ \sigma = 10,\]
\[\delta = 0.0775.\]

Comparison of single shooting with multiple shooting plus a local method
PARAMETER ESTIMATION

Local Optimization methods

Examples of local solutions

10% experimental noise.
Lsq~10

Absence of noise.
Lsq~1x10^{-2}
PARAMETER ESTIMATION

Global Optimization methods

GLOBAL

Deterministic

Stochastic

Hybrid

Deterministic global methods

- Take advantage of the problem’s structure, use gradient or Hessian information and may even guarantee convergence within a preselected level of accuracy.
- Although very promising and powerful, there are still limitations to their application, mainly due to the rapid increase of computational cost with the size of the considered system and the number of its parameters.

**Stochastic global methods**

- Do not require any assumptions about the problem's structure.
- Make use of pseudo-random sequences to determine search directions towards the global optimum.
- The main advantage of these methods is that they rapidly arrive to the proximity of the solution although further refinements may be too costly.
- Convergence to the global optimum cannot be, in general, guaranteed.
- The number of methods has rapidly increased in last decades. The most successful approaches lie in one (or more) of the following groups:
  - Pure random search and *adaptive sequential methods*
  - Clustering methods
  - *Population based methods*
  - *Nature inspired methods*
PARAMETER ESTIMATION

Global Optimization methods

Simulated Annealing:
- Adaptive stochastic
- Inspired on the process of slow and controlled cooling of melted metals.
- Allows for “uphill” moves
- Slow convergence as compared to other stochastic methods.


Evolutionary Search:
- Population based.
- Uses operators inspired in evolution: recombination, mutation, selection.
- Ex.: SRES


Differential Evolution:
- Population based
- Uses the following operators: differential mutation and crossover

PARAMETER ESTIMATION

Global Optimization methods

GLOBAL

Deterministic

Stochastic

Hybrid

Hybrid methods

> Combine global stochastic methods with local methods so as to enhance their strengths:
  - Global: fast approach to the vicinity of the solution
  - Local: Fast convergence to the closest solution

while compensating for their weaknesses:
  - Global: Slow refining of the solution
  - Local: convergence to local solutions when started far from the global

> There are two crucial aspects to take into account to develop a hybrid

- Design, i.e. the selection of the methods to be combined
- Tuning, i.e. the selection of the switching point
PARAMETER ESTIMATION

Hybrid methods

- **Sequential in two phases**
  - Different combinations of SRES and DE with local (first order) methods
  - “Manual” switching
  - Automatic switching


- **Sequential in several phases**
  - Use several switching points from different regions of the search space
  - Those switching points are determined following a specific heuristic
  - The recently developed *Scatter Search* method

[Image: Typical convergence curve]

http://www.iim.csic.es/~gingproc/software.html

New two-phases hybrid approach: an evolutionary method plus DIFITT (based on multiple shooting)

- Systematic switching strategy
- Using multiple shooting to reduce the multi-modality of the non-linear optimisation problem
- Reliable alternative for challenging parameter estimation problems in e.g. systems biology: ex. oscillations
- Is quite insensitive to the size of the search space
- Robust in the presence of experimental noise
- Large computational cost reductions as compared to the global methods

Back to the Godwin oscillator problem

PARAMETER ESTIMATION

Hybrid methods
Back to the Godwin oscillator problem

10% noise

Convergence curves with DE

Convergence curves with a Hybrid

objective function

CPU time [s]

box 5
box 10
box 100

Time
PARAMETER ESTIMATION

Hybrid methods

Scatter Search for Global Optimization in Matlab

---


Formulation of a parametric identification iterative procedure
**Identifiability analysis**

**Structural:** Will it be possible to uniquely estimate the model unknowns under given “ideal” (noise free and continuous) experimental conditions (observables + type of stimuli)?

Currently testing the possibilities of **Power Series based approaches**.

- Develop power series of the observables around a given time
- Use the uniqueness condition of the coefficients to formulate an algebraic system of equations on the parameters
- Try to solve the system of equations with the help of an “identifiability tableau”

**Practical:** Will it be possible to uniquely estimate the model unknowns under given “real” experimental conditions?

- To compute expected uncertainties, confidence regions and evaluate correlation among parameters
- Simplest approach use Fisher Information Matrix
- Robust Monte-Carlo based approach

Still much research needed to handle large scale models!!
**RELATED ON-GOING WORK**

**Ranking of parameters**

Based on the sensitivity analysis using a Latin Hypercube Sampling approach.
There are many design issues prior to performing experiments for modelling purposes:

- what to measure?
- which stimuli should be manipulated?
- and how?
- initial state of the system?
- when to measure (sampling times)?
- where to measure?
- number of experiments?
- etc.

The objective in OED is to find the **experimental scheme to maximize the richness of the experimental data as measured by the Fisher Information Matrix.**

\[
\mathcal{F} = \mathbb{E}_{\gamma_m | \mu} \left\{ \left[ \frac{\partial J(\theta)}{\partial \theta} \right] \left[ \frac{\partial J(\theta)}{\partial \theta} \right]^T \right\}
\]

**Optimal experimental design**
The problem can be formulated as general **Dynamic Optimization problem** and solved by, for example, the application of the **Control Vector Parameterization Approach**.

The problem is then transformed into a **non-linear programming problem** that may be solved by using global optimization methods.


Parameter estimation is formulated as a non-linear programming problem where the objective is to minimize a cost function related to the distance among model predictions and experimental data.

- Cost function allows to introduce information about experimental noise

- Typical difficulties are related to the presence of suboptimal solutions or to poor identifiability.

- Global optimization methods deal with multimodality.

- Hybrid methods have shown considerable efficiency

- Ranking of parameters and Optimal experimental design may help to improve identifiability properties.

- Often disregarded, Structural Identifiability is crucial to test for the possibility of estimating the parameters from available observables.

- Symbolic manipulation

- New methods are required to deal with large scale problems