Optimization Methods for Calibration and Validation of Dynamic Models

Ekaterina A. Kostina
IWR, Heidelberg University

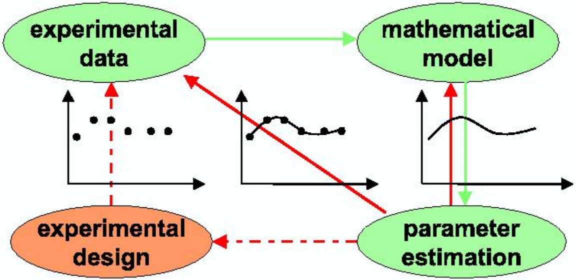
Chulalongkorn University-IWR Summer School on Applied Mathematics and Computational Science,

Chulalongkorn University, March 19-24, 2017

Outline

Outline

Model Validation:
Bringing Experiment and Modeling Together



We Need Models That Allow Simulation

From Merriam-Webster's Dictionary

Full Definition of *simulation*

1. *1* : the act or process of simulating
2. *2* : a sham object : counterfeit
3. *3a* : the imitative representation of the functioning of one system or process by means of the functioning of another *<a computer simulation of an industrial process>b* : examination of a problem often not subject to direct experimentation by means of a simulating device

Full Definition of *simulator*

1. : one that simulates; *especially* : a device that enables the operator to reproduce or represent under test conditions phenomena likely to occur in actual performance

Why Do We Need Simulation Models?

Why Do We Need Simulation Models?

- E.g. the concept of “digital twins” (GE, Siemens, ...)

GE Report Oct 4, 2015

Why Do We Need Simulation Models?

- E.g. the concept of “digital twins” (GE, Siemens, ...)

GE Report Oct 4, 2015

  - every new product or a physical economical process has a digital twin which includes a collection of models and algorithms;

Why Do We Need Simulation Models?

- E.g. the concept of “digital twins” (GE, Siemens, ...)

GE Report Oct 4, 2015

  - every new product or a physical economical process has a digital twin which includes a collection of models and algorithms;
  - which accompanies the process from the “origin” (modeling) through its life time, “growing” together with the process;

Why Do We Need Simulation Models?

- E.g. the concept of “digital twins” (GE, Siemens, ...)

GE Report Oct 4, 2015

  - every new product or a physical economical process has a digital twin which includes a collection of models and algorithms;
  - which accompanies the process from the “origin” (modeling) through its life time, “growing” together with the process;
  - is used among others to analyze data, predict malfunctioning and perform optimal operation.

Why Do We Need Simulation Models?

- E.g. the concept of “digital twins” (GE, Siemens, ...)

GE Report Oct 4, 2015

  - every new product or a physical economical process has a digital twin which includes a collection of models and algorithms;
  - which accompanies the process from the “origin” (modeling) through its life time, “growing” together with the process;
  - is used among others to analyze data, predict malfunctioning and perform optimal operation.
  - “The Wall Street Journal”, Jan 15, 2016: GE Digital CEO Says “Digital Twins” Will Optimize Machinery, Human Health

Why Do We Need Simulation Models?

- E.g. the concept of “digital twins” (GE, Siemens, ...)

GE Report Oct 4, 2015

  - every new product or a physical economical process has a digital twin which includes a collection of models and algorithms;
  - which accompanies the process from the “origin” (modeling) through its life time, “growing” together with the process;
  - is used among others to analyze data, predict malfunctioning and perform optimal operation.
  - “The Wall Street Journal”, Jan 15, 2016: GE Digital CEO Says “Digital Twins” Will Optimize Machinery, Human Health
- Another example: development and admission of drugs

Why Do We Need Simulation Models?

- E.g. the concept of “digital twins” (GE, Siemens, ...)

GE Report Oct 4, 2015

  - every new product or a physical economical process has a digital twin which includes a collection of models and algorithms;
  - which accompanies the process from the “origin” (modeling) through its life time, “growing” together with the process;
  - is used among others to analyze data, predict malfunctioning and perform optimal operation.
  - “The Wall Street Journal”, Jan 15, 2016: GE Digital CEO Says “Digital Twins” Will Optimize Machinery, Human Health
- Another example: development and admission of drugs
- Necessary precondition: validated models with reliable parameter estimates

Outline

Dynamic Process Models

- Ordinary Differential Equations (ODE)
- Boundary Conditions
- Measurement Functions
- Differential Algebraic Equations (DAE)
- Partial Differential Equations (PDE) and Method of Lines (MOL)
- Models with Switches

Ordinary Differential Equations (ODE)

System dynamics is influenced by controls/inputs and unknown parameters

$$\dot{x}(t) = f(t, x(t), p, u(t))$$

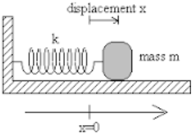
- simulation interval: $[t_0, t_{\text{end}}]$
- time $t \in [t_0, t_{\text{end}}]$
- state $x(t) \in \mathbb{R}^{n_x}$
- controls $u(t) \in \mathbb{R}^{n_u}$ $\longleftarrow$ inputs
- parameters $p \in \mathbb{R}^{n_p}$ $\longleftarrow$ unknown

ODE Example: Harmonic Oscillator

Mass m with spring constant k and unknown friction coefficient β:

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= -\frac{k}{m}(x_1(t) - u(t)) \quad - \beta x_2(t) \end{aligned}$$

- state — $x(t) \in \mathbb{R}^2$
- position of mass — $x_1(t)$
- velocity of mass — $x_2(t)$
- control: acceleration — $u(t) \in \mathbb{R}$ $\longleftarrow$ given input
- parameter: friction — $\beta \in \mathbb{R}$ $\longleftarrow$ unknown



Boundary Conditions

Constraints on initial or intermediate values are important part of dynamic model

Standard Form:

$$r(x(t_0), x(t_1), \dots, x(t_{\text{end}}), p) = 0, \quad r \in \mathbb{R}^{n_r}$$

E.g., fixed or parameter dependent initial value x_0:

$$x(t_0) - x_0(p) = 0 \qquad (n_r = n_x)$$

or periodicity:

$$x(t_0) - x(t_{\text{end}}) = 0 \qquad (n_r = n_x)$$

Note: Initial values $x(t_0)$ need not always be fixed!

Measurement Functions

We can measure functions of states and parameters:

$$M_i(x(t_i), p) \qquad i = 1, \ldots, N$$

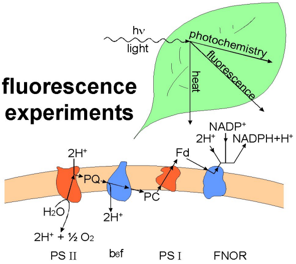
- $M_i \in \mathbb{R}^{m_i}$, often nonlinear
- altogether $\sum_{i=1}^{N} m_i$ measurements
- measurement times $t_1, \ldots, t_N \in [t_0, t_{\text{end}}]$

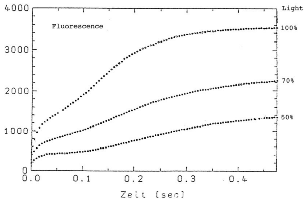
Example: The Light Reaction in Photosynthesis

Baake, Schlöder, 1992

three experiments with different light intensities

fluorescence experiments

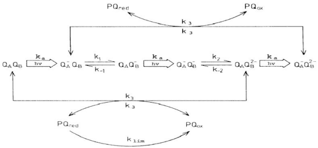




Laboratory Strasser, Stuttgart

Photosynthesis: ODE

electron transport chain
in photosynthesis:



- mathematical model:
  nonlinear ODE
  with 6 states and 4+2
  parameters

$$\begin{aligned}
\dot{x}_1 &= (k_a + k_3(p_{\text{tot}} - x_6))x_1 + k_3 x_5 x_6 \\
\dot{x}_2 &= k_a x_1 - (k_1 + k_3(p_{\text{tot}} - x_6))x_2 + k_{-1}x_3 + k_3 x_6 (1 - \textstyle\sum_{i=1}^{5} x_i) \\
\dot{x}_3 &= k_1 x_2 - (k_a + k_{-1})x_3 \\
\dot{x}_4 &= k_a x_3 - k_2 x_4 + k_{-2} x_5 \\
\dot{x}_5 &= k_3 x_1 (p_{\text{tot}} - x_6) + k_2 x_4 - (k_a + k_{-2} + k_3 x_6) x_5 \\
\dot{x}_6 &= -k_3 (1 - \textstyle\sum_{i=1}^{5} x_i) x_6 + k_3 (x_1 + x_2)(p_{\text{tot}} - x_6) + (p_{\text{tot}} - x_6) k_{\text{lim}}
\end{aligned}$$

with

$$k_a = \frac{I_2(1 - p_{2T})}{1 - p_{22} - p_{2T} + p_{22}p_{2T}(x_1 + x_3 + x_5)}$$

Photosynthesis: Boundary Conditions

- Initial values are given (partly depending on parameters):

$$r(x(0), p) = \begin{pmatrix} x_1(0) - c_1 \\ x_2(0) - c_2 \\ x_3(0) - c_3 \\ x_4(0) - c_4 \\ x_5(0) - c_5 \\ x_6(0) - p_{\text{tot}} \end{pmatrix} = 0$$

Photosynthesis: Measurement Function

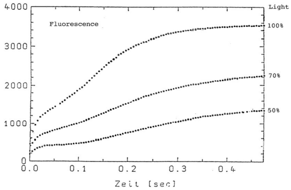
- Fluorescence is nonlinear function of states and parameters:

$$M_i(x(t_i), p) = \left\{ \frac{1 - p_{2T} - p_{22}}{p_{2T}} + \frac{1 - (x_1(t_i) + x_3(t_i) + x_5(t_i))}{1 + \frac{p_{22} p_{2T} (x_1(t_i) + x_3(t_i) + x_5(t_i))}{1 - p_{2T} - p_{22}}} \right\} \cdot S \cdot I_2$$

- Extra parameter (S) in measurement function (unknown gauge of apparatus)
- Fluorescence measured at 96 time points $t_1, \ldots, t_{96}$.
- Aim: Estimate model parameters from fluorescence measurements of living tobacco leaf

Photosynthesis: Multiple Experiment Structure

Data: 3 experiments with different light intensities (96 fluorescence measurements)



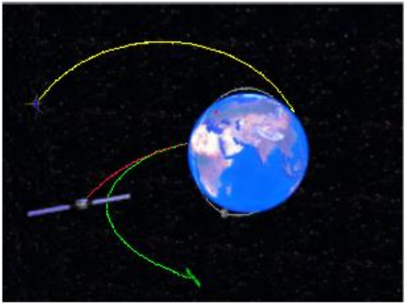
to be estimated:

4 system parameters $p_{\text{tot}}, p_{2T}, p_{22}, k_3$

\+ 1 measurement parameter S

\+ 3 x 2 parameters depending on experiment k_{lim}, I_2

Satellite Orbit Determination: False injection orbit

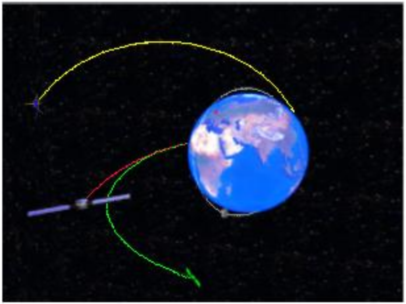


Actual injection orbit differs significantly due to launcher mal-function or underperformance

(SPACEFLIGHT NOW, 21.02.2002): ... the Ariane 5 launcher had propelled the Artemis satellite into a transfer orbit that was lower than expected, with the apogee at only 17 000 km rather than the nominal 36 000 km ...

Similar: Galileo satellite (August 2014)!

Satellite Orbit Determination: False injection orbit



Important: Fast and reliable determination of satellite orbit in order to

- predict future trajectory
- perform correction maneuvers

(SPACEFLIGHT NOW, 21.02.2002): ... the Ariane 5 launcher had propelled the Artemis satellite into a transfer orbit that was lower than expected, with the apogee at only 17 000 km rather than the nominal 36 000 km ...

Similar: Galileo satellite (August 2014)!

Satellite Dynamics

Kepler equations augmented by perturbations

$$\dot{r}(t) = v(t), \qquad \dot{v}(t) = -\frac{GM_\oplus}{\|r(t)\|^3} r(t) + \text{pert}(r(t), v(t), t)$$

due to external forces

- gravitational forces of sun and moon
- inhomogeneities of earth' gravitational field
- air drag
- solar radiation pressure
- dynamic solid tide
- relativistic effects
- gravitational forces from Venus and Jupiter

$\vdots$

Satellite Dynamics

Kepler equations augmented by perturbations

$\dot{r}(t) = v(t), \qquad \dot{v}(t) = -\frac{GM_\oplus}{\|r(t)\|^3} r(t) + \text{pert}(r(t), v(t), t)$

- Results in small but complex nonlinear differential equation system in six states with discontinuities in right-hand side
- Orbit uniquely defined if full state vector $\begin{pmatrix} r(t_0) \\ \dot{r}(t_0) \end{pmatrix}$ known at a time t_0

Data

- Observations of the satellite from different ground stations

Typical measurements:

- range
- range rate
- azimuth and elevation angles

Malindi Ground Station

Range Measurements

distance between ground station and satellite

$$M_1(x(t_M), t_M) \quad = \quad m(x(t_M), t_M) + M_1^{corr}(x(t_M), t_M)$$

where
$m(x(t_M), t_M) = 2\|r(t_M) - r_{stat}(t_M)\|_2$

$x(t_M) = \binom{r(t_M)}{v(t_M)}$ - position of the satellite at the moment t_M

$r_{stat}(t_M)$ - position of the ground station at the moment t_M

correction term $M_1^{corr}(x(t_M), t_M)$ takes into account:

- motion of satellite and station during signal travel time
- systematic errors of physical nature (atmosphere)
- systematic errors of technical nature (biases, delays, ...)

Range Rate Measurements

change in distance between ground station and satellite

$$M_2(x(t_M), t_M) \quad = \quad \frac{m(x(t_M + h), t_M + h) - m(x(t_M), t_M)}{h} + M_2^{corr}(x(t_M), t_M)$$

where

$m(x(t_M), t_M) = 2\|r(t_M) - r_{stat}(t_M)\|_2$

$x(t_M) = \begin{pmatrix} r(t_M) \\ v(t_M) \end{pmatrix}$ - position of the satellite at the moment t_M

$r_{stat}(t_M)$ - position of the ground station at the moment t_M

correction term: motion of station and satellite, biases, ...

Angle Measurements

observation direction (azimuth and elevation angle)

$$M_{3,4}(x(t_M), t_M) \quad = \quad W_{3,4}(x(t_M), t_M) + M_{3,4}^{corr}(x(t_M), t_M)$$

where

$W_3(x(t_M), t_M) = \arctan\left(\frac{s_E}{s_N}\right)$ for azimuth angle

$W_4(x(t_M), t_M) = \arctan\left(\frac{s_Z}{\sqrt{s_E^2 + s_N^2}}\right)$ for elevation angle

$$\begin{pmatrix} s_E \\ s_N \\ s_Z \end{pmatrix} = \begin{pmatrix} -sin\lambda & -sin\varphi cos\lambda & cos\varphi cos\lambda \\ cos\lambda & -sin\varphi sin\lambda & cos\varphi sin\lambda \\ 0 & cos\varphi & sin\varphi \end{pmatrix} \cdot (r - r_{stat})$$

λ and φ are the longitute and altitude of the ground station respectively

Satellite Orbit Determination

Parameters to be estimated:

Six orbit elements at a given time (epoch)

in coop with ESA

Dynamic Process Models

- Ordinary Differential Equations (ODE)
- Boundary Conditions
- Measurement Functions
- Differential-Algebraic Equations (DAE)
- Partial Differential Equations (PDE) and Method of Lines (MOL)
- Models with Switches

Differential-Algebraic Equations (DAE)

Augment ODE by **algebraic equations** g and **algebraic states** z

$$\begin{aligned} \dot{y}(t) &= f(t, y(t), z(t), u(t), p) \\ 0 &= g(t, y(t), z(t), u(t), p) \end{aligned}$$

- differential states $y(t) \in \mathbb{R}^{n_x}$
- algebraic states $z(t) \in \mathbb{R}^{n_z}$
- algebraic equations $g(\cdot) \in \mathbb{R}^{n_z}$

Example: Index 1 DAE

$$
\begin{aligned}
0 &= g(t, y, z) \\
0 &= \frac{d}{dt} g(t, y, z) = \frac{\partial g}{\partial t} + \frac{\partial g}{\partial y} \dot{y} + \frac{\partial g}{\partial z} \dot{z}
\end{aligned}
$$

If the matrix $\frac{\partial g}{\partial z} \in \mathbb{R}^{n_z \times n_z}$ is invertible then we can compute

$$
\dot{z} = -\left(\frac{\partial g}{\partial z}\right)^{-1} \left(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial y} \dot{y}\right)
$$

and obtain an ODE system for $x := \begin{pmatrix} y \\ z \end{pmatrix}$:

$$
\dot{x} = \begin{pmatrix} f(t, x) \\ -\left(\frac{\partial g}{\partial z}\right)^{-1} \left(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial y} \dot{y}\right) \end{pmatrix}
$$

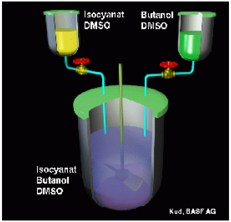
Example DAE: Urethane Reaction

S. Körkel, Cooperation with BASF SE

$$
\begin{aligned}
A + B &\rightarrow C \\
A + C &\rightleftharpoons D \\
3\,A &\rightarrow E
\end{aligned}
$$

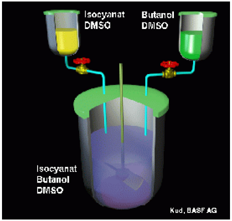
A: isocyanate B: butanol
C: urethane D: allophanate
E: isocyanurate L: solvent DMSO

- Prototype for polyurethane production
- Main product C, byproduct D
- Composition of the products determines physical properties of the polyurethane plastic material



Example DAE: Urethane Reaction

S. Körkel, Cooperation with BASF SE



- Reactor: ideally stirred tank
- Two controlled feeds:
  A in DMSO and B in DMSO
- Control of reactor temperature

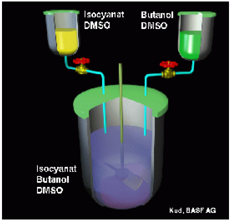
Example DAE: Urethane Reaction

S. Körkel, Cooperation with BASF SE

$$
\begin{aligned}
\dot{n_C} &= V \cdot (r_1 - r_2 + r_3) \\
\dot{n_D} &= V \cdot (r_2 - r_3) \\
\dot{n_E} &= V \cdot r_4 \\
0 &= n_A + n_C + 2n_D + 3n_E - n_{A0} - n_{Aea}(t) \\
0 &= n_B + n_C + n_D - n_{B0} - n_{Beb}(t) \\
0 &= n_L - n_{L0} - n_{Lea}(t) - n_{Leb}(t)
\end{aligned}
$$

$$n_C(t_0) = n_D(t_0) = n_E(t_0) = 0$$

$$
\begin{aligned}
r_1 &= k_1 \cdot \frac{n_A}{V} \cdot \frac{n_B}{V} \qquad r_3 = k_3 \cdot \frac{n_D}{V} \\
r_2 &= k_2 \cdot \frac{n_A}{V} \cdot \frac{n_C}{V} \qquad r_4 = k_4 \cdot (\frac{n_A}{V})^2 \\
k_{i=1,2,4} &= k_{ref\,i} \cdot \exp\left(-\frac{E_{ai}}{R} \cdot \left(\frac{1}{T(t)} - \frac{1}{T_{ref\,i}}\right)\right) \\
\frac{k_2}{k_3} &= k_{c2} \cdot \exp\left(-\frac{dh_2}{R} \cdot \left(\frac{1}{T(t)} - \frac{1}{T_{c2}}\right)\right)
\end{aligned}
$$



Urethane Reaction: Features

Model

- 6 state variables $n_A, n_B, n_C, n_D, n_E, n_L$, nonlinear Arrhenius kinetics
- 8 unknown parameters p : steric factors k_{ref_i}, activation energies E_{ai}, $i = 1, 2, 4$, equilibrium constant k_{c2}, reaction enthalpy dh_2
- 3 time dependent control functions $u(t)$: temperature $T(t)$, feed profiles $feed_1(t), feed_2(t)$
- 7 control variables q: initial molar numbers in the reactor n_{A0}, n_{B0}, n_{L0} and in the feeds $n_{A,e1,0}, n_{B,e2,0}, n_{L,e1,0}, n_{L,e2,0}$

Measurements

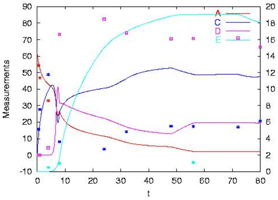
- 3 measurement methods (A, C/D, E) with different accuracies and different costs!

Urethane Example: Measurement Functions

Measurements: Mass percentage of A, C, D, E, e.g.

$$M_{n_C}(x,p) = 100 \cdot \frac{n_C M_C}{n_A M_A + \ldots + n_E M_E + n_L M_L}$$

Measurements and Simulated Model Response



Dynamical Process Models

- Ordinary Differential Equations (ODE)
- Boundary Conditions
- Measurement Functions
- Differential-Algebraic Equations (DAE)
- Partial Differential Equations (PDE) and Method of Lines (MOL)
- Models with Switches

Partial Differential Equations

- Instationary partial differential equations (PDE) arise, e.g., in transport processes, wave propagation, ...
- Also called “distributed parameter systems”
- Often PDE of subsystems are coupled with each other (e.g., flow connections)
- Method of Lines (MOL): discretize PDE in space to yield ODE or DAE system

Partial Differential Equations: Example

- Convection-Diffusion-Reaction Equation

$$\frac{\partial y_i}{\partial t}(t,x) = \frac{\partial}{\partial x}\left(D\frac{\partial y_i}{\partial x}(t,x)\right) - v\frac{\partial y_i}{\partial x} + \sum_{j=1}^{m} r_j(y(t,x),\alpha(t,x),T(t,x),p)\cdot\nu_{ij},\ \ i = 1,\dots,n$$

- Catalyst deactivation

$$\frac{\partial \alpha}{\partial t}(t,x) = -k(y(t,x),\alpha(t,x),T(t,x),p)$$

+ Initial and Boundary Conditions

- $y_i(t,x)$: concentration of species $i, i = 1,\dots,n$
- $\alpha(t,x)$: catalyst activity

Possible Solution Approach: Method of Lines

- Discretize the state variables $y_i(t) := y(t, x_i)$ on a grid $(x_i,\ i = 1, \ldots, N)$ with $\Delta x = x_{i+1} - x_i$
- Replace spatial derivatives by finite differences, e.g.

$$\frac{y(t, x_{i+1}) - y(t, x_{i-1})}{2\Delta x} = \frac{\partial y}{\partial x}(t, x_i) + \mathcal{O}(\Delta x^2)$$

$$\frac{y(t, x_{i+1}) - 2y(t, x_i) + y(t, x_{i-1})}{\Delta x^2} = \frac{\partial^2 y}{\partial x \partial x}(t, x_i) + \mathcal{O}(\Delta x^2)$$

- Substitute into PDE equation, obtain high-dimensional stiff sparse ODE/DAE system

Example: Nonlinear Exchange Rate Dynamics

with S. Jäger

- Exchange rate is a stochastic process $S_t, t \in [t_0, T]$ and satisfies stochastic differential equation

$$dS = \mu dt + \sigma dW, \ W \text{ is a Wiener process}$$

- The distribution $F_t(s) = P(S_t \leq s)$ is defined by a drift term μ and a variance σ^2

$$S_t = S_{t_0} + \int_{t_0}^{t} \mu dt + \int_{t_0}^{t} \sigma dW,$$

- The drift term μ and the variance σ^2 of are supposed to depend non-linearly on the real exchange rate s and a set of market fundamental variables Z (e.g. money amounts, real incomes, nominal interests)

$$\mu = \mu(s, Z), \quad \sigma^2 = \sigma^2(s, Z)$$

- initial values S_{t_0}, μ and σ^2 are unknown and need to be identified!

Example: Nonlinear Exchange Rate Dynamics

with S. Jäger

- Model: μ and σ^2 satisfy the Fokker-Planck or forward Kolmogorov equation for the density function $f(t,s) = \frac{dF_t(s)}{ds}$ of the stochastic process S_t

$$\frac{\partial f}{\partial t} = -\frac{\partial(\mu f)}{\partial s} + \frac{1}{2}\frac{\partial^2(\sigma^2 f)}{\partial s^2}$$

Example: Nonlinear Exchange Rate Dynamics

with S. Jäger

- The drift is modelled by the 3d order polynome (Creedy et al, 96; multiple equilibria, regime switching, multimodality)

$$\mu = \mu(t,s,Z) = a_0(Z)(s-a_1(Z))(s-a_2(Z))(s-a_3(Z)),$$
$$a_i(Z) = C_i \prod_{j=1}^{K} Z_j^{\alpha_{ij}}$$

and the variance is

$$\sigma^2 = {\gamma_0}^2 + {\gamma_1}^2 s$$

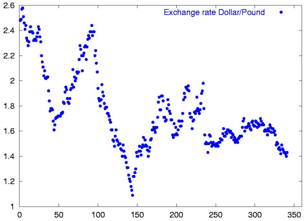
$Z_j, j = 1,...,K$ are the market fundamental variables

- Unknown parameters are $C_i,\ \alpha_{ij},\ j = 1,...,K,\ i = 0,1,2,3,\ \gamma_0,\ \gamma_1$

Example: Nonlinear Exchange Rate Dynamics

with S. Jäger

- Data η_j: monthly Dollar/Pound exchange rate



- Model response: expected value of the exchange rate

$$M(t_j) = \int_0^\infty f(t_j, s) s ds$$

Example: Nonlinear Exchange Rate Dynamics

with S. Jäger

- Boundary conditions:

$$\mu - \frac{1}{2}\frac{\partial(\sigma^2 f)}{\partial s}\bigg|_{s=s_{min}} = \mu - \frac{1}{2}\frac{\partial(\sigma^2 f)}{\partial s}\bigg|_{s=s_{max}} = 0,\ t \geq 0.$$

- Initial conditions: we assume that at the time moment $t = 0$ the stochastic process is stationary:

$$f(t,s)\mid_{t=0} = f^\star(s,p) = \exp\left(\int_0^s \frac{2\mu(\xi,p)}{\sigma^2(\xi,p)}d\xi - \ln\sigma^2(s,p) + \ln\sigma^2(0,p)\right)\eta^\star,$$

$\eta^\star$ is a normalizing constant

- alternative:

$$f(t,s)\mid_{t=0} = \exp\left(-(s-s_0)^2 a^2\right)\eta^\star,$$

with 2 additional parameters a and s_0 to estimate.

Method of Lines (MOL)

E.g. forward Kolmogorov equation for the density function $f(t,s) = \frac{dF_t(s)}{ds}$ of the stochastic process S_t

$$\frac{\partial f}{\partial t} = -\frac{\partial(\mu f)}{\partial s} + \frac{1}{2}\frac{\partial^2(\sigma^2 f)}{\partial s^2}$$

- introduce spatial grid points $s_0, \ldots, s_N$
- approximate spatial derivatives, e.g. by finite differences

$$\frac{\partial f(s_i)}{\partial s} \approx \frac{f(s_{i+1}) + f(s_i)}{h}, \quad \text{etc.}$$

- define state vector $y_{\text{col}} := (f(s_0), \ldots, f(s_N))^T$,
- obtain ODE

$$\dot{y}_{\text{col}}(t) = f_{\text{col}}(y_{\text{col}}(t), \mu(t), \sigma(t), p)$$

Transport and Degradation of Xenobiotics in Soil

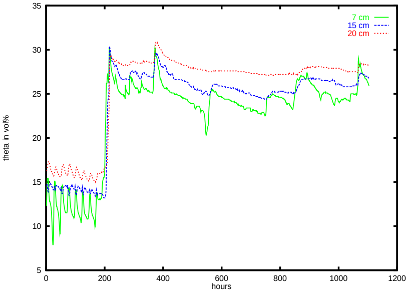
(..., A. Dieses, in coop. with O. Richter, TU Braunschweig)

Minilysimeter

- Investigation of fate of xenobiotics
- Expensive lysimeter experiments for registration
- To be replaced by computer experiments
- **Here: parameter estimation**
- **Later: Optimal lysimeter experiments**
  - + **optimal irrigation**
  - + **optimal solute application**
  - + **optimal sampling**

Field experiment: Water Transport (K. Aden)

- loamy sand without vegetation
- time-domain reflectrometry (TDR): hourly measurements of water content θ in 7, 15 and 20 cm
- period: Oct 28, 1997 - Dec 13, 1997



Model: Richards Equation

$$\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z}\left(D(\theta)\frac{\partial \theta}{\partial z} - K(\theta)\right)$$

$$K(\theta) = K_s \Theta^{1/2}\left[1 - \left(1 - \Theta^{n/(n-1)}\right)^{1-1/n}\right]^2, \qquad \Theta = \frac{\theta - \theta_r}{\theta_s - \theta_r}$$

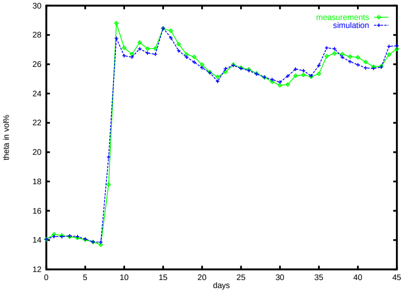
$$D(\theta) = K(\theta)\bar{C}(\theta)$$

$$\bar{C}(\theta) = \frac{1}{\alpha n m}\left(\Theta^{-1/m} - 1\right)^{-m}\Theta^{-1/m}\frac{1}{\theta - \theta_r}, \qquad m = 1 - \frac{1}{n}$$

- Initial condition: Linear interpolation of $\theta_{7cm}, \theta_{15cm}, \theta_{20cm}$ at begin of experiments (Oct 28, 1997)
- Upper boundary: Dirichlet condition (TDR data in 7 cm)
- Lower boundary: Dirichlet condition (TDR data in 20 cm)

Transport and Degradation of Xenobiotics in Soil

Result: Estimates for n, α and K_s
from TDR measurement data of water content in 15 cm depth



	guess	estimate
n	1.5	1.262 ± 0.0024
α	0.05	0.0324 ± 0.0024
K_s	35.0	20.92 ± 1.68

	α	K_s
n	0.14	-0.61
α	-	-0.94

Dynamical Process Models

- Ordinary Differential Equations (ODE)
- Boundary Conditions
- Measurement Functions
- Differential-Algebraic Equations (DAE)
- Partial Differential Equations (PDE) and Method of Lines (MOL)
- Models with Switches

Models with Switches

$$\begin{aligned} \dot{y}(t) = \quad & f(t, x(t), z(t), p, q, u(t), sign(\sigma(x(t), z(t), p)) \\ 0 = \quad & g(t, x(t), z(t), p, q, u(t), sign(\sigma(x(t), z(t), p)) \end{aligned} \quad t \in [t_0; t_f]$$

Discontinuous dynamics
Monitored by sign of ”switching functions” σ
E.g.

- modeling of phase transition in multi-phase reaction systems
- simplified modeling of fast transients in processes with varying time scales

Example: Growth of White Cabbage

Richter, Söndgerath 1990

- states: Cabbage's leaf biomass x_L, trunk biomass x_S and head biomass x_H
- 9 parameters a, r_L, μ_L, ρ, r_S, r_H, m_H, λ and t_H

$$\frac{dx_L}{dt} = r_L \frac{a+1}{a+\exp(\rho t)} x_L - \mu_L x_L$$

$$\frac{dx_S}{dt} = r_S x_L \left(\lambda - \frac{x_S}{x_L}\right)$$

$$\frac{dx_H}{dt} = \begin{cases} 0 & \text{for } t \le t_H \\ (r_H x_L - \mu_H x_H) & \text{for } t > t_H \end{cases}$$

- sign of "switching function" $\sigma := t - t_H$ determines which model is taken

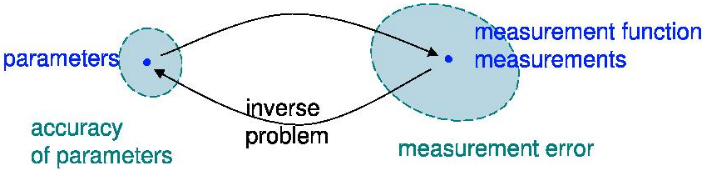
Outline

Parameter Estimation Problems

- Parameter Estimation: Problem Formulation
- Boundary Value Problem Approach
- Generalized Gauss Newton Methods
  - Optimization Criteria and Convergence
  - Practical Solution
- Sensitivity Analysis
- Examples:
  - Lotka-Volterra, Unstable Process
  - Orbit Determination Problem for Satellites
  - Photosynthesis
  - Bistable Belousov-Zhabotinskii Reaction
  - Enzyme Reaction Kinetics

What Is Parameter Estimation?

Determination of coefficients/parameters
in a model
from measured data
such that model matches data



Parameter Estimation Problem: Model

Model Equations? ("Forward Problem")

- Differential Algebraic Equations (DAE)

$$\begin{aligned} \dot{y} &= f(y, z, p, q, u) \\ 0 &= g(y, z, p, q, u) \end{aligned}$$

y: differential variables

z: algebraic variables, $x = (y, z)$

p: unknown parameters

q: design parameters

u: controls

stiff, nonlinear, with discontinuities, unstable modes, chaotic, ...

- Partial Differential Equations (PDE)

$$u_t - \nabla(K\nabla u) = f(u, p)$$

$\rightarrow$ semidiscretization in space

+ initial and boundary conditions!

Parameter Estimation Problem: Data

Experimental Data?

- Data from multiple experiments under varying experimental conditions
  - reactions with different initial composition of substances
  - different inputs (e.g., temperature, feedstreams, ...)
  - different experimental layout (e.g., a standing or sitting human being ...)

- Data for stationary or instationary states, bifurcations, oscillations, ...

  Each has specific model

- Outliers in the data

- Indirect observation functions ($\rightarrow$ and additional parameters), ...

Parameter Estimation Problem: Observation Model

$$\eta_{ij} \quad = \quad M_{ij}(x^{true}(t_j), p^{true}) + \varepsilon_{ij}$$

M: nonlinear function of states
(differential and algebraic)
and parameters

ε_{ij}: measurement error

- independent
- often normally distributed $\varepsilon_{ij} \in (\mathcal{N}(0, \sigma_{ij}^2))$

Parameter Estimation Problem: Match Model to the Data

Parameter Estimation Problem

$$\min_{x,p} \quad \sum_{i,j} \frac{(\eta_{ij} - M_{ij}(x(t_j), p))^2}{\sigma_{ij}^{\ 2}}$$

maximum likelihood

x, p satisfy

- DAE model
- additional constraints, e.g., boundary conditions, positivity, ...

$$r_2(x(t_1), \ldots, x(t_k), p) = 0 \text{ or } \geq 0$$

Parameter Estimation Problem: Match Model to the Data

Multiple Experiment Parameter Estimation Problem

$$\min_{x^l,p} \sum_{l=1}^{\#\,\text{Exp}} \sum_{i,j} \frac{(\eta_{ij}^l - M_{ij}^l(x^l(t_j^l),p))^2}{\sigma_{ij}^{l\,2}}$$

maximum likelihood

x^l, p satisfy

- DAE models, $l = 1, \ldots, \#\text{Exp}$
- additional constraints, e.g. boundary conditions, positivity, ...

$$r^l{}_2(x^l(t_1^l), ..., x^l(t_k^l), p) = 0 \text{ or } \geq 0 \,, \; l = 1, \ldots, \#\text{Exp}$$

Why Least-Squares Objective Function?

- Assumptions for distribution of measurement errors: $\varepsilon_{ij} \in \mathcal{N}(0, \sigma_{ij}^2)$
- Log-Likelihood function

$$
\begin{aligned}
\log \mathcal{L} &= \sum_{i,j} \log \left(\frac{1}{\sigma_{ij}\sqrt{2\pi}} \, e^{-\frac{(\eta_{ij} - M_{ij}(x,p))^2}{2\sigma_{ij}^2}} \right) \\
&= Const - \sum_{i,j} \frac{(\eta_{ij} - M_{ij}(x,p))^2}{2\sigma_{ij}^2} \\
\max_p \log \mathcal{L} &\equiv \min_p \sum_{i,j} \frac{(\eta_{ij} - M_{ij}(x,p))^2}{2\sigma_{ij}^2}
\end{aligned}
$$

- Then solution p^* of the parameter estimation problem is a maximum likelihood estimate for the parameters

Choice of Norm?

- if measurement errors are independent and normally distributed ($\varepsilon_{ij} \in N(0, \sigma_{ij}^2)$) then l_2 estimation is appropriate:

$$\min_{x,p} \sum_{i,j} \frac{(\eta_{ij} - M_{ij}(x(t_j), p))^2}{\sigma_{ij}^2} \quad \text{least squares}$$

$\rightarrow$ maximum likelihood: Legendre (1805), Gauss (1809)

Choice of Norm?

- if measurement errors are independent and normally distributed ($\varepsilon_{ij} \in N(0, \sigma_{ij}^2)$) then l_2 estimation is appropriate:

$$\min_{x,p} \sum_{i,j} \frac{(\eta_{ij} - M_{ij}(x(t_j), p))^2}{\sigma_{ij}^2}$$ least squares

$\rightarrow$ maximum likelihood: Legendre (1805), Gauss (1809)

- in case of Laplace distribution $\left(\frac{1}{2|\sigma_{ij}|} e^{-\frac{|t|}{|\sigma_{ij}|}} \right)$ l_1 estimation is appropriate:

$$\min_{x,p} \sum_{i,j} \frac{|\eta_{ij} - M_{ij}(x(t_j), p)|}{|\sigma_{ij}|}$$ least absolute deviation

$\rightarrow$ maximum likelihood: Boscovic (1758), Laplace (1812)

Properties of l_1-Parameter Estimation

under certain regularity assumptions

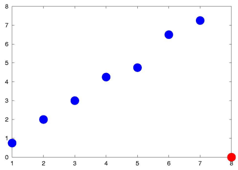
- l_1 optimal solution interpolates n "best" measurements
- consequently solution is less sensitive to outliers
- l_1 forms robust alternative to l_2 PE !

Properties of l_1-Parameter Estimation

under certain regularity assumptions

- l_1 optimal solution interpolates n “best” measurements
- consequently solution is less sensitive to outliers
- l_1 forms robust alternative to l_2 PE !

Data



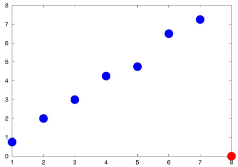
● Outlier!

Properties of l_1-Parameter Estimation

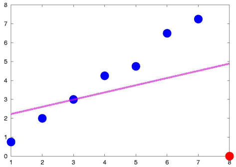
under certain regularity assumptions

- l_1 optimal solution interpolates n "best" measurements
- consequently solution is less sensitive to outliers
- l_1 forms robust alternative to l_2 PE !

Data



l_2



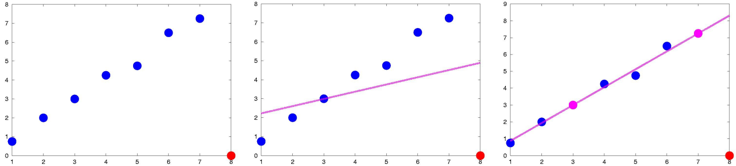
• Outlier!

Properties of l_1-Parameter Estimation

under certain regularity assumptions

- l_1 optimal solution interpolates n "best" measurements
- consequently solution is less sensitive to outliers
- l_1 forms robust alternative to l_2 PE !

Data $\quad$ l_2 $\quad$ l_1



● Outlier!

Robust Parameter Estimation

- robustness means “insensitivity to small deviations from the assumptions” (Huber 1981)
- even high-quality measurements are not exactly normally distributed, but typically longer-tailed
  (for scientific routine data 1–10% gross errors are the rule rather than the exception)
- gross errors often show up as outliers (although not all outliers are gross errors)
- a single outlier can completely spoil a least squares analysis

Robust PE: Choice of Cost Functional

- least squares norm of measurement errors (normally distributed measurement error)

$$\min \frac{1}{2} \sum_{i}^{\#Meas.} \left(\frac{\varepsilon_i}{\sigma_i} \right)^2$$

- robust against outliers l_1 norm of measurement errors (Laplace distributed measurement error)

$$\min \sum_{i}^{\#Meas.} \left| \frac{\varepsilon_i}{\sigma_i} \right|$$

- another robust estimator: hybrid “norm”, Huber-estimator

$$\min \frac{1}{2} \sum_{i:|\varepsilon_i/\sigma_i| \leq \gamma} \left(\frac{\varepsilon_i}{\sigma_i} \right)^2 + \sum_{i:|\varepsilon_i/\sigma_i| > \gamma} \left(\gamma \left| \frac{\varepsilon_i}{\sigma_i} \right| - \frac{1}{2} \gamma^2 \right)$$

Robust PE: Choice of Cost Functional

- least squares norm of measurement errors (normally distributed measurement error)

$$\min \frac{1}{2} \sum_{i}^{\#Meas.} \left(\frac{\varepsilon_i}{\sigma_i} \right)^2$$

- robust against outliers l_1 norm of measurement errors (Laplace distributed measurement error)

$$\min \sum_{i}^{\#Meas.} \left| \frac{\varepsilon_i}{\sigma_i} \right|$$

- another robust estimator: hybrid “norm”, Huber-estimator

$$\min \frac{1}{2} \sum_{i:|\varepsilon_i/\sigma_i| \leq \gamma} \left(\frac{\varepsilon_i}{\sigma_i} \right)^2 + \sum_{i:|\varepsilon_i/\sigma_i| > \gamma} \left(\gamma \left| \frac{\varepsilon_i}{\sigma_i} \right| - \frac{1}{2}\gamma^2 \right)$$

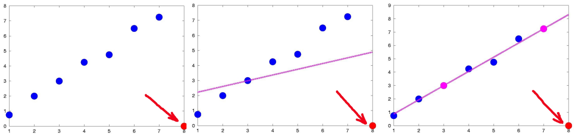
- partition constant γ can be determined by the ratio of “bad” data points in the measurement data for an assumed error probability $\epsilon = \frac{\varepsilon_i}{\sigma_i}$:

$\epsilon \to 0 \;\Rightarrow\; \gamma \to \infty$: converges to solution of least squares method,

$\epsilon \to 1 \;\Rightarrow\; \gamma \to 0$: converges to solution of l_1 approximation

Example: a Single Outlier

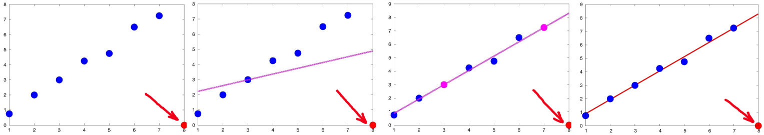
Data l_2 l_1



- Outlier!

Example: a Single Outlier

Data l_2 l_1 Huber ($\gamma = 0.02$)



- Outlier!

Parameter Estimation Problems

- Parameter Estimation: Problem Formulation
- Boundary Value Problem Approach
- Generalized Gauss Newton Methods
  - Optimization Criteria and Convergence
  - Aspects of Practical Solution
- Sensitivity Analysis
- Examples:
  - Lotka-Volterra, Unstable Process
  - Orbit Determination Problem for Satellites
  - Photosynthesis
  - Bistable Belousov-Zhabotinskii Reaction
  - Enzyme Reaction Kinetics

Direct "All-at-Once" Boundary Value Problem Methods

- the IVP approach: "single shooting"
  - integrate DAE over whole interval to yield $x(t; x_0, p)$
  - eliminate - infinite - state variables in favour of unknown parameters p, plug into suitable optimizer

Direct "All-at-Once" Boundary Value Problem Methods

- the IVP approach: "single shooting"
  - integrate DAE over whole interval to yield $x(t; x_0, p)$
  - eliminate - infinite - state variables in favour of unknown parameters p, plug into suitable optimizer

Direct "All-at-Once" Boundary Value Problem Methods

- the IVP approach: "single shooting"
  - integrate DAE over whole interval to yield $x(t; x_0, p)$
  - eliminate - infinite - state variables in favour of unknown parameters p, plug into suitable optimizer

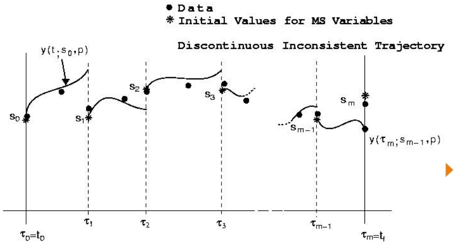
Bock and coworkers, 81, . . .

- the BVP approach: discretize DAE and solve simultaneously
  - optimization problem
  - discretized BVP as equality constraint
  - further constraints

  in one loop!

Flexible realization: multiple shooting
analogous for FD, collocation (e.g. Biegler)

The Multiple Shooting Method



- choose mesh $t_0 = \tau_0 < \tau_1 < \ldots < \tau_m = t_f$
- choose initial values $s_j = (y(\tau_j), z(\tau_j))$ as additional variables
- solve relaxed DAE IVP at each interval

$$\begin{aligned} \dot{y} &= f(y, z, p) \\ 0 &= g(y, z, p) \\ &\quad -\alpha(t) g(s_j, p) \end{aligned}$$

$\alpha(\tau_j) = 1$, $\alpha(t) \to 0$ for $t \to \infty$

- DAE discretization leads to additional matching conditions
  - for continuity: $s^y_{j+1} - y(\tau_{j+1}; s_j, p) = 0$
  - for consistency: $g(s_j, p) = 0$

After discretization: large scale nonlinear constrained approximation problem

$$\min_X \quad \frac{1}{2}||F_1(X)||_2^2$$

$$F_2(X) = 0 \qquad \text{(contains discretized BVP)} \quad \text{or } \geq 0$$

Difficulties

- nonlinear equality - and inequality - constrained optimization problem
- large number of variables from discretization

  e.g., in case multiple shooting: # of parameters +
  # of differential and algebraic variables
  × # of multiple shooting points
  × # of experiments!
- but special block structures

So Multiple Shooting Makes Things More Difficult?

FAQ: Why Multiple Shooting?

- key property: discretized states as add'l optimization variables
  - allows for *better initial guesses* using information about the process, helps to avoid ”far away” local minima
  - damps influence of poor parameter guesses
  - *reduces nonlinearity* and speeds up convergence (even up to one step convergence!)
  - method is *numerically stable* even for potentially instable, e.g. chaotic, differential equations
- efficient parallel implementation
- adaptive accuracy discretization strategies
- state-of-the-art solvers for DAE IVP applicable

An Unstable Test Problem

- state equations:

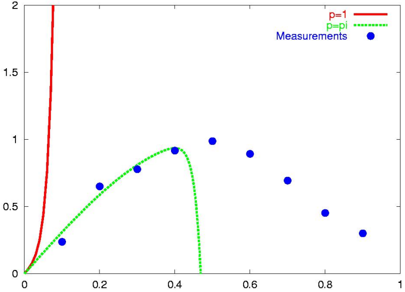
$$\dot{x}_1 = x_2 \qquad \dot{x}_2 = \mu^2 x_1 - (\mu^2 + p^2) \sin pt, t \in [0, 1]$$
$$x_1(0) = 0, \qquad x_2(0) = \pi$$

- special solution for “true” parameter value $p = \pi$:

$$x_1(t) = \sin \pi t, \qquad x_2(t) = \pi \cos \pi t.$$

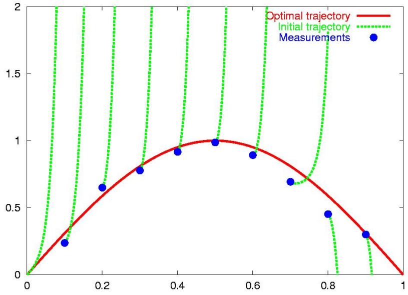
- $\mu = 60$, i.e. error propagation over $[0, 1]$ is $\exp \mu \approx 10^{27}$ - highly unstable
- pseudo random measurement noise, $\sigma = 0.05$

An Unstable Test Problem - Single Shooting - FAILS!



initial trajectory with $p = 1$ and with $p = \mathit{float}(\pi)$ in 64 bit

An Unstable Test Problem - Multiple Shooting - WORKS!



initial trajectory with $p = 1$ - convergence after 4 iterations in 64 bit

Multiple Shooting Reduces Nonlinearity: One Step Convergence

If:

- Dense data for all states are available
- Problem equations are linear in parameters
- Lengths of multiple shooting interval $h \to 0$

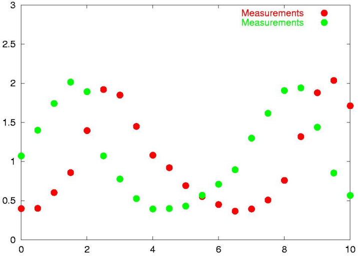
Then:

- One-step-convergence to true parameter values

$$p^1 = p^0 + \Delta p^0 = p^{true} + O(h^s)$$

Lotka-Volterra Problem: Model and Data

$$\begin{aligned} \dot{x}_1 &= -p_1 x_1 + p_2 x_1 x_2 \\ \dot{x}_2 &= +p_3 x_2 - p_4 x_1 x_2 \end{aligned}$$

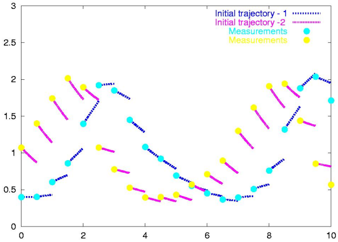


x_1: predators
x_2: preys
DE linear in parameters

Data: $\sigma = 5\%$

Lotka-Volterra Problem:
Solution with Multiple Shooting

Initial trajectory

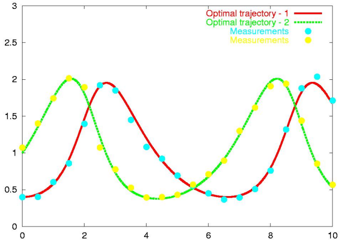


Initial guesses:

$p_1 = 0.5 \qquad p_2 = 0.5$

$p_3 = -0.5 \qquad p_4 = -0.2$

Solution trajectory



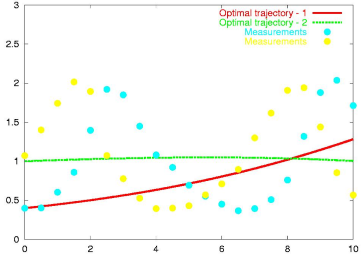
Solution:

$p_1 = 1.01 \pm 0.02 \qquad p_2 = 1.01 \pm 0.03$

$p_3 = 0.99 \pm 0.02 \qquad p_4 = 1.01 \pm 0.03$

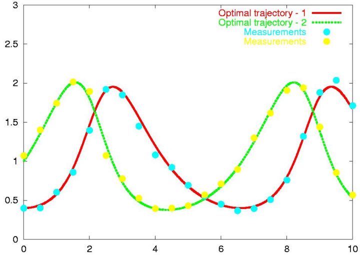
Comparison: Single vs. Multiple Shooting

Single Shooting



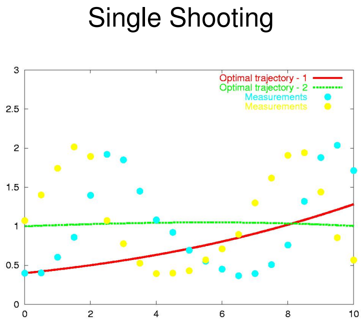
Convergence after 8 iterations

Multiple Shooting

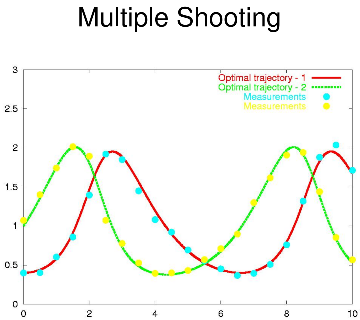


Convergence after 4 iterations

Comparison: Single vs. Multiple Shooting



Convergence after 8 iterations



Convergence after 4 iterations

- Multiple Shooting helps to avoid local minima

Generation of Initial States for Multiple Shooting Nodes

- important for problem solution $\rightarrow$ generation of good initial guesses for multiple shooting variables
- one possibility: solve *special nonlinear constrained least squares problem* at each multiple shooting node τ_j:

$$\min_{s} \quad \|s^j_{ref} - s\|_2^2$$
$$\text{s.t.} \quad \phi^j(s, \eta) = 0 \text{ or } \geq 0$$

s^j_{ref} is a reference value, e.g. the value of the computed trajectory at the end of the previous interval

constraints $\phi^j(s, \eta) = 0$ or ≥ 0 include, e.g., the requirement that initial values should satisfy the measurements at τ_j

Orbit Determination Problem

Minimize deviation of model response $M(y(t), p)$ from measurement values η

$$\min_{y,p} \sum_{j=1}^{l} \sum_{i=1}^{m_j} \left(\frac{\eta_{ij} - M_{ij}(y(t_j), p)}{\sigma_{ij}} \right)^2,$$

s.t. satellite dynamics is fulfilled

$$\begin{aligned} \dot{y}(t) &= f(t, y(t), p) \\ y(t_0) &= y_0(p) \end{aligned}$$

Orbit Determination Problem

Minimize deviation of model response $M(y(t), p)$ from measurement values η

$$\min_{y,p} \sum_{j=1}^{l} \sum_{i=1}^{m_j} \left(\frac{\eta_{ij} - M_{ij}(y(t_j), p)}{\sigma_{ij}} \right)^2,$$

s.t. satellite dynamics is fulfilled

$$\begin{aligned} \dot{y}(t) &= f(t, y(t), p) \\ y(t_0) &= y_0(p) \end{aligned}$$

Parameters to estimate are initial values for the states
$+$ callibration parameters in measurement functions
$+$ coefficients for air drag model and/or solar radiation pressure
$+$ ...

Orbit Determination Problem

Minimize deviation of model response $M(y(t), p)$ from measurement values η

$$\min_{y,p} \sum_{j=1}^{l} \sum_{i=1}^{m_j} \left(\frac{\eta_{ij} - M_{ij}(y(t_j), p)}{\sigma_{ij}} \right)^2,$$

s.t. satellite dynamics is fulfilled

$$\begin{aligned} \dot{y}(t) &= f(t, y(t), p) \\ y(t_0) &= y_0(p) \end{aligned}$$

Difficulties: No complete state observation
Outliers in the data

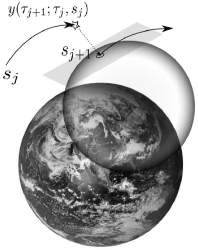
Generation of Initial States for Multiple Shooting Nodes: Orbit Determination for Satellites

S. Lenz in coop with ESA

Example:

Range Measurement $\eta_{r,j+1}$ at time τ_{j+1}

- Positions that fulfill the measurement are on a sphere around the station



1. Decompose solution of IVP in position and velocity

$$y(\tau_{j+1};\tau_j,s_j) \rightarrow \left((r_{j+1}^{int})^T, (\dot{r}_{j+1}^{int})^T \right)^T$$

2. Transform position vector into local tangent coordinate system: $r_{j+1}^{int} \rightarrow r_{j+1}^{lt}$
3. Scale position vector to measured range:

$$r_{j+1}^{lt,new} = \frac{\eta_{r,j+1}}{2} \frac{r_{j+1}^{lt}}{||r_{j+1}^{lt}||_2}$$

4. Transform new position vector back into inertial frame: $r_{j+1}^{lt,new} \rightarrow r_{j+1}^{new}$
5. Combine new position and unchanged velocity into a vector

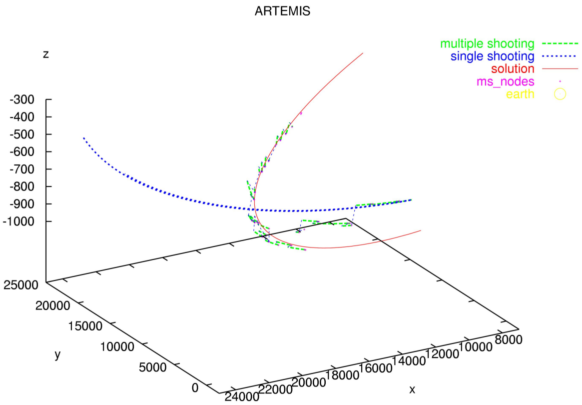
$$\left((r_{j+1}^{new})^T, (\dot{r}_{j+1}^{int})^T \right)^T \rightarrow s_{j+1}$$

6. Continue integration

Generation of Initial States for Multiple Shooting Nodes: Analytical Projection

Example: Orbit Determination for Satellites (ESA)

S. Lenz



Application: Orbit Determination Problems for Satellites (ARTEMIS-Launch)

initial multiple shooting trajectory (including projections)

(Loading ...)

Satellite Orbit Determination: Result

nominal orbit, actual orbit

(Loading ...)

Parameter Estimation Problems

- Parameter Estimation: Problem Formulation
- Boundary Value Problem Approach
- Generalized Gauss Newton Methods
  - Optimization Criteria and Convergence
  - Practical Solution
- Sensitivity Analysis
- Examples:
  - Lotka-Volterra, Unstable Process
  - Enzyme Reaction Kinetics
  - Photosynthesis
  - Bistable Belousov-Zhabotinskii Reaction

After discretization: large scale nonlinear constrained approximation problem

$$\min_X \quad \frac{1}{2}||F_1(X)||_2^2$$

$$F_2(X) = 0 \qquad \text{(contains discretized BVP)} \quad \text{or} \geq 0$$

numerical treatment with Newton type methods

unconstrained case (IVP methods)

constrained case (BVP methods)

examples for structure exploitation

Solution Methods - IVP Approach

- **Nonlinear unconstrained least squares problem**

$$\min \varphi(X) = \frac{1}{2} F_1^T(X) F_1(X), \quad X \in \mathbb{R}^n, \quad F_1 : \mathbb{R}^n \to \mathbb{R}^{m_1}$$

- optimal solution X^* solves the system of nonlinear equations

$$\nabla \varphi(X^*) = J_1^T(X^*) F_1(X^*) = 0$$

- Newton iteration: $X^+ = X^- + \Delta X$ for improving an approximate solution X of optimality equations, ΔX solves $\nabla^2 \varphi(X) \Delta X = -\nabla \varphi(X)$, or, equivalently,

$$\left(J_1^T J_1 + \sum_{i=1}^{m_1} F_{1,i} \nabla_{XX} F_{1,i} \right) \Delta X = -J_1^T F_1, \quad J_1 = \nabla F_1(X)$$

- variations of Newton's method involve the approximation of the term

$$S = \sum_{i=1}^{m_1} F_{1,i} \nabla_{XX} F_{1,i}(X).$$

- **Gauss-Newton:** $S = 0$

Solution Methods - BVP Approach

- **Nonlinear constrained least squares problem**

$$\min \varphi(X) = \frac{1}{2} F_1^T(X) F_1(X) \qquad s.t. \quad F_2(X) = 0$$

- Iteration: $X^{k+1} = X^k + \Delta X^k$
- The increment ΔX^k solves the quadratic problem:

$$\begin{array}{ll} \min\limits_{\Delta X \in \Omega^k} & \frac{1}{2} \Delta X^T A^k \Delta X + \nabla \varphi(X^k)^T \Delta X \\ \text{s.t.} & F_2(X^k) + J_2(X^k)^T \Delta X = 0 \end{array}$$

- A^k is an approximation of the Hessian of the Lagrangian:

$$\begin{array}{rcl} A^k & \approx & \nabla_{XX} L(X^k, \lambda^k), \quad L(X, \lambda) = \varphi(X) - \lambda^T F_2(X), \\ & \approx & J_1^T(X^k) J_1(X^k) + F_1(X^k)^T \nabla_{XX} F_1(X^k) - \lambda^T \nabla_{XX} F_2(X^k). \end{array}$$

Generalized Gauss-Newton

- **Generalized Gauss-Newton: ignore second order derivatives**

$$A^k = J_1(X^k)^T J_1(X^k)$$

- ΔX^k solves linear constrained problem

$$\min_{\Delta X \in \Omega^k} \quad \frac{1}{2}||F_1(X^k) + J_1(X^k)\Delta X||_2^2$$

$$\text{s.t.} \quad F_2(X^k) + J_2(X^k)\Delta X = 0$$

Optimality Criteria

- Constraint Qualification
  X is regular if Constraint Qualification (CQ) holds:

$$\text{rank}(J_2(X)) = \#Constraints$$

- Lagrange function:

$$L(X, \lambda) = \frac{1}{2}||F_1(X)||_2^2 - \lambda^T F_2(X)$$

Optimality Criteria

Necessary Conditions:

Let

- X^* be a regular solution of the nonlinear problem.

Then

- X^* is feasible $F_2(X^*) = 0$
- there exists a unique vector λ^* such that

$$\nabla_X L(X^*, \lambda^*) = 0 \qquad \leftarrow \text{Stationarity}$$

- furthermore, second order necessary conditions hold:

$$d^T \nabla_{XX} L(X^*, \lambda^*) d \geq 0, \forall d \in \{w | J_2(X^*) w = 0\}$$

Optimality Criteria

Sufficient Condition:

Let

- (X^*, λ^*) satisfy first-order necessary conditions
- Positive Definiteness (PD) condition holds:

$$d^T \nabla_{XX} L(X^*, \lambda^*) d > 0, \forall d \in \{w \neq 0 | J_2(x^*) w = 0\}$$

Then

- X^* is a strict local minimum.

Optimality Criteria

for linear systems regularity conditions (CQ) and (PD) are equivalent to

$$\text{rank}(J_2(X)) = \#constraints$$

$$\text{rank}(J(X)) = \#variables, \quad J(X) = J = \begin{pmatrix} J_1 \\ J_2 \end{pmatrix}$$

Optimality Criteria

- Karush-Kuhn-Tucker Conditions: Feasibility + Stationarity

$$F_1^T(X^*)J_1(X^*) = 0$$
$$F_2(X^*) = 0$$

- (X^*, λ^*) is called a KKT-Point
- Under regularity conditions:
  (X^*, λ^*) is a KKT-Point of the nonlinear problem $\Leftrightarrow$ $(0, \lambda^*)$ is a KKT-Point of the linear problem

Solution of the Linear Least-Squares Problem

- At each GN iteration we need to solve:

$$\begin{array}{rl} \min & \frac{1}{2}||F_1(X^k) + J_1(X^k)\Delta X||_2^2, \\ \text{s.t.} & F_2(X^k) + J_2(X^k)\Delta X = 0 \\ \text{where} & J_i(X) = \nabla F_i(X) \end{array}$$

- KKT conditions:

$$\begin{pmatrix} J_1^T(X)J_1(X) & J_2^T(X) \\ J_2(X) & 0 \end{pmatrix} \begin{pmatrix} \Delta X \\ -\lambda \end{pmatrix} = \begin{pmatrix} -J_1^T(X)F_1(X) \\ F_2(X) \end{pmatrix}$$

- The linear system has unique solution, if (CQ) and (PD) are fulfilled

Solution of the Linear Least-Squares Problem

- ΔX^k can be formally written with the help of a solution operator J^+

$$\Delta X^k = -J^+(X^k)F(X^k)$$

J^+ is a generalized inverse: $J^+JJ^+ = J^+$

$$J = \begin{pmatrix} J_1 \\ J_2 \end{pmatrix}, \; F = \begin{pmatrix} F_1 \\ F_2 \end{pmatrix}$$

- The solution operator J^+ is explicitly given by

$$J^+(X) = \begin{pmatrix} \mathbb{I} & 0 \end{pmatrix} \begin{pmatrix} J_1^T(X)J_1(X) & J_2(X)^T \\ J_2(X) & 0 \end{pmatrix}^{-1} \begin{pmatrix} J_1(X)^T & 0 \\ 0 & \mathbb{I} \end{pmatrix}.$$

Local Contraction

Bock, 1987

Let (weighted) Lipschitz conditions be true for J und J^+:

$$(J) \qquad \frac{\|J^+(Y)[J(X+t(Y-X))-J(X)](Y-X)\|}{t\,\|Y-X\|^2} \leq \omega(X) \leq \omega < \infty \qquad \text{nonlinearity}$$

$$(J^+) \qquad \frac{\|[J^+(Z)-J^+(X)]R(X)\|}{\|Z-X\|} \leq \kappa(X) \leq \kappa < 1 \qquad \text{incompatibility}$$

$$\forall t \in [0,1], \quad X - Y = J^+(X)F(X), \quad R(X) := F(X) - J(X)J(X)^+F(X)$$

Then: For X^0 with $\|J(X^0)^+F(X^0)\|\omega/2 + \kappa < 1$

$X^{j+1} = X^j - J(X^j)^+F(X^j)$ is well defined

$X^k \rightarrow X^*$ stationary point with $J(X^*)^+F(X^*) = 0$

and $\|\Delta X^{j+1}\| \leq (\|\Delta X^j\|\,\omega/2 \,+\, \kappa)\,\|\Delta X^j\|$

linear convergence (convergence rate $\rightarrow \kappa$)

Local Contraction

Bock, 1987

Interpretation:
Nonlinearity ω

- ω is a measure for nonlinearity (weighted second derivative)
- ω^{-1} characterizes region of validity of the linear model

Incompatibility constant κ

- $\kappa < 1$: necessary for identifiability
- depends on compatibility of the model with measurements
- a stationary point with $\kappa < 1$ is statistically stable
- GGN method does not converge to large residual solutions (Advantage!)

Globalization Strategies

Line Search

- Iteration: $X^{k+1} = X^k + t^k \Delta X^k, \quad t^k \in]0,1]$, where t^k is a stepsize
- stepsize t^k is chosen such that the next iterate X^{k+1} is “better” than X^k: $T_1(X^{k+1}) < T_1(X^k)$
- exact penalty function as merit function

$$T_1(X) := \frac{1}{2}||F_1(X)||_2^2 + \sum_{i,Eq.} \alpha_i |F_{2i}(X)|$$

  with sufficiently large weights $\alpha_i > 0$
- t^k is (approximate) minimum of the merit function

$$t^k = \text{arg} \min_{0<t\leq 1} T_1(X^k + t\Delta X^k)$$

Globalization Strategies

Line Search

- Iteration: $X^{k+1} = X^k + t^k \Delta X^k, \quad t^k \in]0,1]$, where t^k is a stepsize
- stepsize t^k is chosen such that the next iterate X^{k+1} is "better" than X^k: $T_1(X^{k+1}) < T_1(X^k)$
- exact penalty function as merit function

$$T_1(X) := \frac{1}{2}||F_1(X)||_2^2 + \sum_{i,Eq.} \alpha_i |F_{2i}(X)|$$

  with sufficiently large weights $\alpha_i > 0$
- alternative: line search based on the natural level functions

$$T^k(X^k + t^k \Delta X^k) \quad = \quad ||J^+(X^k) F(X^k + t^k \Delta X^k)||_2^2$$

  $\rightarrow$ new effective "affine invariant" globalization strategy - guarantees full step in local convergence domain (Bock, K., Schlöder, 2000, K. 2004)

Numerical Solution of Linear Least-Squares Problems

Unconstrained case:

- QR factorization with column pivoting on the matrix J_1:

$$J_1 P = Q \begin{pmatrix} R \\ 0 \end{pmatrix} = (Q_1, \quad Q_2) \begin{pmatrix} R \\ 0 \end{pmatrix} = Q_1 R,$$

  where

  P is an $n \times n$ permutation matrix (orthogonal);
  Q is $m_1 \times m_1$ orthogonal;
  $Q_1 \in \mathbb{R}^{m_1 \times n}$, $Q_2 \in \mathbb{R}^{m_1 \times (m_1 - n)}$;
  R is $n \times n$ upper triangular.

- We get

$$||F_1 + J_1 \Delta X||_2^2 = ||Q_1^T F_1 + R P^T \Delta X||_2^2 + ||Q_2^T F_1||_2^2$$

- We minimize $||F_1 + J_1 \Delta X||_2^2$ by driving the first term in to zero:

$$\Delta X = -P R^{-1} Q_1^T F_1$$

Numerical Solution of Linear Least-Squares Problems

- alternative: use SVD of Jacobian J_1

$$J_1 = U \begin{pmatrix} S \\ 0 \end{pmatrix} V^T = (U_1, \quad U_2) \begin{pmatrix} S \\ 0 \end{pmatrix} V^T = U_1 S V^T,$$

  where

  U is an orthogonal $m_1 \times m_1$ matrix;
  $U_1 \in \mathbb{R}^{m_1 \times n}$, $U_2 \in \mathbb{R}^{m_1 \times (m_1 - n)}$;
  V is an orthogonal $n \times n$ matrix;
  S is a diagonal $n \times n$ matrix with elements $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$.

- the solution ΔX:

$$\Delta X = -VS^{-1} U_1^T F_1 = -\sum_i \frac{u_i^T F_1}{\sigma_i} v_i$$

Numerical Solution of Linear Least-Squares Problems

Constrained case:

- Orthogonal decomposition of J_2: $J_2 = LQ^T$ where
  - $L \in \mathbb{R}^{m_2 \times n}$,
  - $L = [L', \mathbf{0}]$, $L' \in \mathbb{R}^{m_2 \times m_2}$ is a lower triangular matrix,
  - $Q \in \mathbb{R}^{n \times n}$, $Q^T Q = I$.
- Linear problem can be rewritten:

$$
\begin{aligned}
\min \quad & \frac{1}{2}||F_1 + J_1 Q \Delta Y||_2^2, \\
\text{s.t.} \quad & F_2 + L \Delta Y = 0,
\end{aligned}
$$

with $\Delta Y = Q^T \Delta X$

Numerical Solution of Linear Least-Squares Problems

- The solution ΔY is:

$$\Delta Y = \begin{pmatrix} \Delta Y_2 \\ \Delta Y_1 \end{pmatrix},$$

where ΔY_2 is computed by

$$\Delta Y_2 = -(L')^{-1} F_2$$

and ΔY_1 solves unconstrained linear least squares problem:

$$\min_{\Delta Y_1} \quad \frac{1}{2} ||\tilde{F}_1 + \tilde{J}_1 \Delta Y_1||_2^2 = ||(F_1 - J_1 Q_1 (L')^{-1} F_2) + (J_1 Q_2) \Delta Y_1||_2^2.$$

- solution in original coordinates by back-transformation

$$\Delta X = Q \Delta Y$$

Treatment of Ill-Conditioned Problems

Regularization by a-priori information:

- Given a-priori information on values of variables $\tilde{X}_i$
- Given variances for this information σ_i^2
- i.e $X_i = \tilde{X}_i \pm \sigma_i$
- Then: Modify cost functional:

$$||F_1(X)||_2^2 \rightarrow ||F_1(X)||_2^2 + \sum_i^n \frac{(X_i - \tilde{X}_i)^2}{\sigma_i^2}$$

- $J_1^T J_1 \rightarrow J_1^T J_1 + \Sigma^{-2}$
- analogous: a-priori information with covariance matrix Cov

$\rightarrow$ important for moving horizon estimation in real-time

Treatment of Ill-Conditioned Problems

Regularisation by Rank Reduction

- in each iteration solve

$$\min_x \|Ax + b\|_2^2$$

- perform SVD, decompose $A = USV^T$

$$\min_y \|Sy + \hat{b}\|_2^2$$

where $y = V^T x, \quad \hat{b} = U^T b$ and
$S = \text{Diag}(s_i), \quad s_1 \geq s_2 \geq \dots \geq s_n \geq 0$

- condition $\quad cond(S) = \frac{s_1}{s_n} \qquad$ too large?
- rank reduction!
criterion: set rank to j^* where $j^* = max\{j | s_j \geq C\}$
choice of C ? $\qquad$ note: $\text{var}(y_j) = \frac{\beta^2}{s_j^2}$
for σ_{max}^2 maximal acceptable variance: choose $C \geq \frac{\beta}{\sigma_{max}}$

Treatment of Ill-Conditioned Problems

Regularisation by Rank Reduction

- alternative: QR-decomposition (with pivoting) $A = QR$
- estimate for the condition number

$$cond_{QR}: \quad \frac{|r_{11}|}{|r_{nn}|}$$

- rank citerion set rank to j^* where $j^* = max\{j||r_{jj}| \geq C\}$

How to Solve Linear Constrained l_1 Problem?

$$\min_{\Delta X \in \Omega^k} \quad ||F_1(X^k) + J_1(X^k)\Delta X||_1,$$

$$\text{s.t.} \quad F_2(X^k) + J_2(X^k)\Delta X = 0 \quad \text{or} \quad \geq 0$$

- cost function is piecewise-linear, very special structure that can be effectively epxloited $\rightarrow$ so called multiple pivoting or long steps (Osborne 76, Gabasov et al 79, K. et al 98, Osborne, K. 2006)

(Block-)Sparse Structures (Multiple Shooting)

Bock 81, 87, Schlöder 83

- large block sparse super-structures from multiple experiments

$$
\begin{pmatrix}
\begin{matrix} E_{L1} \\ U_{L1} \end{matrix} & 0 & 0 & & 0 & \begin{matrix} E_{G1} \\ U_{G1} \end{matrix} \\
0 & \begin{matrix} E_{L2} \\ U_{L2} \end{matrix} & 0 & & 0 & \begin{matrix} E_{G2} \\ U_{G2} \end{matrix} \\
0 & 0 & \boxed{\begin{matrix} E_{L3} \\ U_{L3} \end{matrix}} & & 0 & \boxed{\begin{matrix} E_{G3} \\ U_{G3} \end{matrix}} \\
 & & & \ddots & & \vdots \\
0 & 0 & 0 & & \begin{matrix} E_{LN} \\ U_{LN} \end{matrix} & \begin{matrix} E_{GN} \\ U_{GN} \end{matrix}
\end{pmatrix}
\begin{matrix} \\ \\ \leftarrow \\ \\ \\ \end{matrix}
$$

Experiments N: $1 \sim 100$

(Block-)Sparse Structures (Multiple Shooting)

Bock 81, 87, Schlöder 83

- large block sparse super-structures from multiple experiments
- structures from parametrization in time, e.g. induced by multiple shooting $\rightarrow$ typical staircase structure

$$
\begin{array}{l}
U_{Lk} = \\ \\ E_{Lk} =
\end{array}
\left[\begin{array}{cccccc}
D_1^1 & D_1^2 & \cdot & \cdot & \cdot & D_1^q \\
D_2^1 & D_2^2 & \cdot & \cdot & \cdot & D_2^q \\
G_1^l & G_1^r & & & & G_1^q \\
& \boxed{G_2^l} & G_2^r & & & \boxed{G_2^q} \\
& & \ddots & \ddots & & \vdots \\
& & & G_{m-1}^l & G_{m-1}^r & G_{m-1}^q
\end{array}\right]
\qquad
\begin{array}{l}
U_{Gk} = \\ \\ E_{Gk} =
\end{array}
\left[\begin{array}{c}
D_1^p \\
D_2^p \\
G_1^p \\
\boxed{G_2^p} \\
\vdots \\
G_{m-1}^p
\end{array}\right]
$$

Meshpoints $m : 2 \rightarrow \geq 100$

Evaluation of Linear Systems
DAE Initial Value Problems and Derivatives

- BDF discretization for stiff systems
- adaptive integrators for ODE and relaxed DAE
- treatment of implicitly given discontinuities and jumps in dynamics
- fast and accurate computation of 1. and 2. order derivatives
  Combining
  > ”automatic differentiation” of model equations and
  > ”internal numerical differentiation” of adaptive discretization scheme
- in forward or reverse mode

⋮

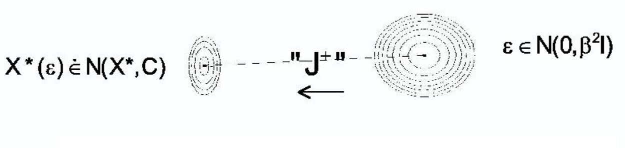
e.g. DAESOL, RKFSWT (Bauer et al ’98, Albersmeyer ’05, Kirches ’06)

Parameter Estimation Problems

- Parameter Estimation: Problem Formulation
- Boundary Value Problem Approach
- Generalized Gauss Newton Methods
  - Optimization Criteria and Convergence
  - Practical Solution
- Sensitivity Analysis
- Examples:
  - Lotka-Volterra, Unstable Process
  - Orbit Determination Problem for Satellites
  - Photosynthesis
  - Bistable Belousov-Zhabotinskii Reaction
  - Enzyme Reaction Kinetics

Assessment of Uncertainties in Parameter Estimates for Constrained Case (l_2)

- "good fit" is not sufficient - we need to know uncertainty of parameter estimate $X^\star(\varepsilon)$ depending on measurement errors, e.g. $\varepsilon \in N(0, \beta^2 \mathbb{I})$

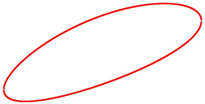


- first order expansion at $X^\star = X^\star(0)$:
  $$\Delta X = X^\star(\varepsilon) - X^\star \approx -J(X^\star)^+ \begin{pmatrix} \varepsilon \\ 0 \end{pmatrix}$$
- yields covariance-matrix approximation for states and parameters
  $$C := \mathcal{E}\left(\Delta X \Delta X^T\right) = \left(J(X^\star)^+ \begin{pmatrix} \varepsilon \\ 0 \end{pmatrix} \begin{pmatrix} \varepsilon \\ 0 \end{pmatrix}^T J(X^\star)^{+T} \right) =$$
  $$J(X^\star)^+ \begin{pmatrix} \beta^2 \mathbb{I} & 0 \\ 0 & 0 \end{pmatrix} J(X^\star)^{+T}$$

Assessment of Uncertainties in Parameter Estimates for Constrained Case (l_2)

- Nonlinear confidence region $G_N(\alpha)$ for the state and parameter estimates with error probability α is

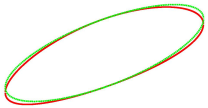
$$G_N(\alpha) := \{X |\, F_2(X) = 0,\ \|F_1(X)\|_2^2 - \|F_1(X^\star)\|_2^2 \leq \gamma^2(\alpha)\}$$



Assessment of Uncertainties in Parameter Estimates for Constrained Case (l_2)

- $G_N(\alpha)$ can be approximated through the linearized confidence region $G_L(\alpha)$

$$G_L(\alpha) := \{X \quad | \quad F_2(X^\star) + J_2(X^\star)(X - X^\star) = 0, \\ \|F_1(X^\star) + J_1(X^\star)(X - X^\star)\|_2^2 - \|F_1(X^\star)\|_2^2 \leq \gamma^2(\alpha)\}.$$



Assessment of Uncertainties in Parameter Estimates for Constrained Case (l_2)

- $G_L(\alpha)$ can be equivalently represented using generalized inverse J^+:

$$G_L(\alpha) = \{X | X = X^\star - J(X^\star)^+ \begin{pmatrix} \varepsilon \\ 0 \end{pmatrix}, ||\varepsilon||_2^2 \leq \gamma(\alpha)\}$$

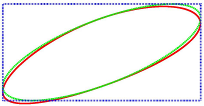
Assessment of Uncertainties in Parameter Estimates for Constrained Case (l_2)

- $G_L(\alpha)$ is contained in confidence box

$$G_L(\alpha) \quad \subset \quad \Pi_{i=1}^{n}[X_i^\star - \delta_i, X_i^\star + \delta_i], \quad \delta_i = C_{ii}^{1/2}\gamma(\alpha)^{1/2}$$

exactly, that is

$$\max_{X \in G_L(\alpha)} \qquad |X_i - X_i^\star| = \delta_i, \;\; i = 1, ..., n.$$



- $C_{ii}^{1/2}$ - standard deviations of parameters can be computed fast

Assessment of Uncertainties in Parameter Estimates for Constrained Case (l_2)

- Covariance matrix can be computed for all variables:

  $\rightarrow$ for parameters and state variables, $\rightarrow$ Prediction
- Diagonal elements of covariance matrix C_{ii} are variances of corresponding variables

Important: Variances for functions $g(x,p)$ of parameters and states can also easily be determined.
Procedure: Introduce new parameter p_{neu} and additional equality constraint $p_{new} = g(x,p)$.
Confidence interval for p_{new} describes quality of $g(x,p)$.

Characterization of Confidence Ellipsoids

Functions Φ_α of the covariance matrix $C(p, q, u, w)$ (unconstrained problem)

- A optimal: Average of the variances of the estimates

$$\Phi_1(C) = \frac{1}{n} trace C$$

- B optimal: maximal square root of diagonal elements of the covariance matrix (“maximal standard deviation”, proportional to maximal edge of enclosing box) (Bock, 1987)

$$\Phi_M(c) = \max_i (C_{ii}^{1/2})$$

- D optimal: Determinant of covariance matrix (“volume”)

$$\Phi_0(C) = det(C)$$

- E optimal: Maximum Eigenvalue of covariance matrix (“maximal semi axis”)

$$\Phi_\infty(c) = \lambda_{max}(C)$$

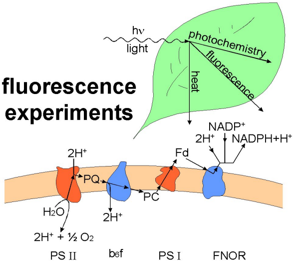
Parameter Estimation Problems

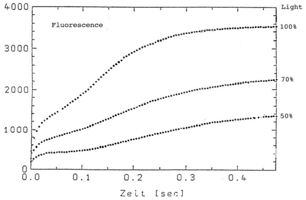
- Parameter Estimation: Problem Formulation
- Boundary Value Problem Approach
- Generalized Gauss Newton Methods
  - Optimization Criteria and Convergence
  - Practical Solution
- Sensitivity Analysis
- Examples:
  - Lotka-Volterra, Unstable Process
  - Orbit Determination Problem for Satellites
  - Photosynthesis
  - Bistable Belousov-Zhabotinskii Reaction
  - Enzyme Reaction Kinetics

Example: The Light Reaction in Photosynthesis

Baake, Schlöder, 1992

three experiments with different light intensities:

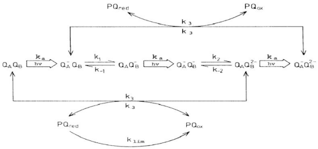




Laboratory Strasser, Stuttgart

Photosynthesis: ODE Model

electron transport chain in photosynthesis:



- mathematical model: nonlinear ODE with 6 states and 6 parameters

$$
\begin{aligned}
\dot{x}_1 &= (k_a + k_3(p_{tot} - x_6))x_1 + k_3 x_5 x_6\\
\dot{x}_2 &= k_a x_1 - (k_1 + k_3(p_{tot} - x_6))x_2 + k_{-1}x_3 + k_3 x_6 (1 - \textstyle\sum_{i=1}^{5} x_i)\\
\dot{x}_3 &= k_1 x_2 - (k_a + k_{-1})x_3\\
\dot{x}_4 &= k_a x_3 - k_2 x_4 + k_{-2} x_5\\
\dot{x}_5 &= k_3 x_1 (p_{tot} - x_6) + k_2 x_4 - (k_a + k_{-2} + k_3 x_6)x_5\\
\dot{x}_6 &= -k_3(1 - \textstyle\sum_{i=1}^{5} x_i)x_6 + k_3(x_1 + x_2)(p_{tot} - x_6) + (p_{tot} - x_6)k_{lim}
\end{aligned}
$$

with

$$
k_a = \frac{I_2(1 - p_{2T})}{1 - p_{22} - p_{2T} + p_{22}p_{2T}(x_1 + x_3 + x_5)}
$$

Photosynthesis: Measurement Function

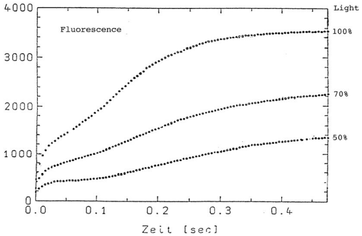
- Fluorescence is nonlinear function of states and parameters:

$$b_i(x(t_i), p) = \left\{ \frac{1 - p_{2T} - p_{22}}{p_{2T}} + \frac{1 - (x_1(t_i) + x_3(t_i) + x_5(t_i))}{1 + \frac{p_{22}p_{2T}(x_1(t_i)+x_3(t_i)+x_5(t_i))}{1-p_{2T}-p_{22}}} \right\} \cdot S \cdot I_2$$

- extra parameter (S) in measurement function (unknown gauge of apparatus)
- Fluorescence measured at 96 time points $t_1, \ldots, t_{96}$.
- Aim: Estimate model parameters from fluorescence measurements of living tobacco leaf

Photosynthesis: Multiple Experiment Structure

Data: 3 experiments with different light intensities (96 fluorescence measurements)



to be estimated:

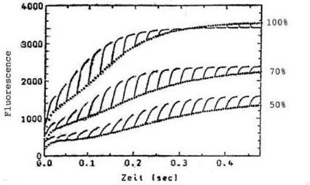
4 system parameter $p_{\text{tot}}, p_{2T}, p_{22}, k_3$

\+ 1 measurement parameter S

\+ 3 x 2 parameter depending on experiment k_{lim}, I_2

Initital Trajectories for Multiple Shooting Photosynthesis

Multiple shooting with 20 gridpoints

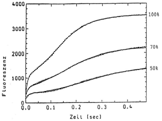


	k_3	p_{22}	p_{2T}	p_{tot}	S	I_2	k_{lim}
Exp 1	20	0.5	0.4	10	10	300	1
Exp 2						210	1
Exp 3						150	1

initial guesses

Photosynthesis: 3-Experiment Solution

Acc:10^{-3}, 12 Iterations, 3 damped



	k_3	p_{22}	p_{2T}	p_{tot}	S
solution	17.3	0.0710	0.841	11.5	18.1
standard error[1]	± 0.76	± 0.015	± 0.015	± 0.84	± 1.0

	100%		70%		50%	
	I_2	k_{lim}	I_2	k_{lim}	I_2	k_{lim}
solution	195.	1.07	143.	1.92	101.	1.67
standard error[1]	± 9.0	± 0.36	± 6.9	± 0.26	± 5.3	± 0.17

[1] estimated; multiplication by 4.5 yields 95% confidence intervals

Belousov Zhabotinskii Reaction

Nonlinear ODE

9 chem. species 5 control parameters

14 unknown parameters

The stochiometric system

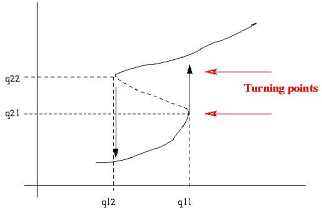
NOYES-FIELD-THOMPSON-MECHANISMUS

$$
\begin{aligned}
BrO_3^- + Br^- + 2H^+ &\underset{p_{-1}}{\overset{p_1}{\rightleftharpoons}} HBrO_2 + HOBr \\
HBrO_2 + Br^- + H^+ &\underset{p_{-2}}{\overset{p_2}{\rightleftharpoons}} 2HOBr \\
HOBr + Br^- + H^+ &\underset{p_{-3}}{\overset{p_3}{\rightleftharpoons}} Br_2 + H_2O \\
BrO_3^- + HBrO_2 + H^+ &\underset{p_{-4}}{\overset{p_4}{\rightleftharpoons}} 2BrO_2^- + H_2O \\
Ce^{3+} + BrO_2^- + H^+ &\underset{p_{-5}}{\overset{p_5}{\rightleftharpoons}} Ce^{4+} + HBrO_2 \\
Ce^{4+} + BrO_2^- + H_2O &\underset{p_{-6}}{\overset{p_6}{\rightleftharpoons}} Ce^{3+} + BrO_3^- + 2H^+ \\
2HBrO_2 &\underset{p_{-7}}{\overset{p_7}{\rightleftharpoons}} BrO_3^- + HOBr + H^+
\end{aligned}
$$

$$
\begin{aligned}
[HBrO_2] \quad \dot{A} &= p_1 BCH^2 - p_{-1}AD - p_2 ACH + p_{-2}D^2 \\
&\quad - p_4 ABH + p_{-4}F^2 + p_5 FGH - p_{-5}AI \\
&\quad - 2p_7 A^2 + 2p_{-7}BDH - k_E A \\
[BrO_3^-] \quad \dot{B} &= -p_1 BCH^2 + p_{-1}AD - p_4 ABH + p_{-4}F^2 \\
&\quad + p_6 FI - p_{-6}BGH^2 + p_7 A^2 - p_{-7}BDH - k_E(B - B_E) \\
[Br^-] \quad \dot{C} &= -p_{-1}BCH^2 + p_{-1}AD - p_2 ACH + p_{-2}D^2 \\
&\quad - p_3 CDH + p_{-3}E - k_E(C - C_E) \\
[HOBr] \quad \dot{D} &= p_1 BCH^2 - p_{-1}AD + 2p_2 ACH - 2p_{-2}D^2 \\
&\quad - p_3 CDH + p_{-3}E + p_7 A^2 - p_{-7}BDH - k_E D \\
[Br_2] \quad \dot{E} &= p_3 CDH - p_{-3}E - k_E E \\
[BrO_2] \quad \dot{F} &= 2p_4 ABH - 2p_{-4}F^2 - p_5 FGH + p_{-5}AI \\
&\quad - p_6 FI + p_{-6}BGH^2 - k_E F \\
[Ce^{3+}] \quad \dot{G} &= -p_5 FGH + p_{-5}AI + p_6 FI - p_{-6}BGH^2 \\
&\quad - k_E(G - G_E) \\
[H^+] \quad \dot{H} &= -2p_1 BCH^2 + 2p_{-1}AD - p_2 ACH + p_{-2}D^2 \\
&\quad - p_3 CDH + p_{-3}E - p_4 ABH + p_{-4}F^2 - p_5 FGH \\
&\quad + p_{-5}AI + 2p_6 FI \\
&\quad - 2p_{-6}BGH^2 + p_7 A^2 - p_{-7}BDH - k_E(H - H_E) \\
[Ce^{4+}] \quad \dot{I} &= p_5 FGH - p_{-5}AI - p_6 FI + p_{-6}BGH^2 - k_E I
\end{aligned}
$$

Belousov Zhabotinskii Reaction – PE Problem

Description of experiments

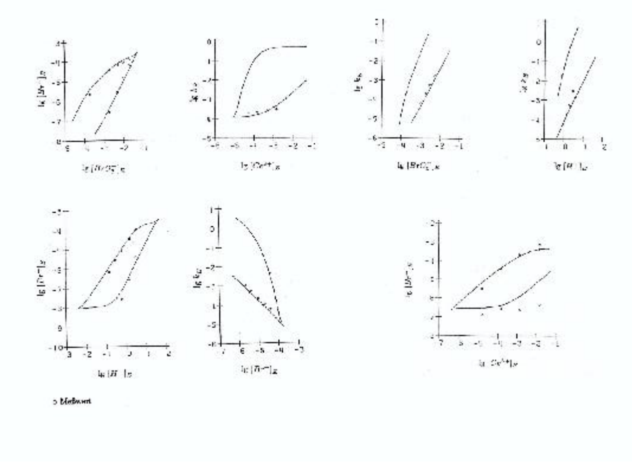


$$\left.\begin{array}{rcl} f(y, q, p) & = & 0 \\ f_y(y, q, p) \cdot h & = & 0 \\ h^T \cdot h & = & 1 \end{array}\right\} \otimes$$

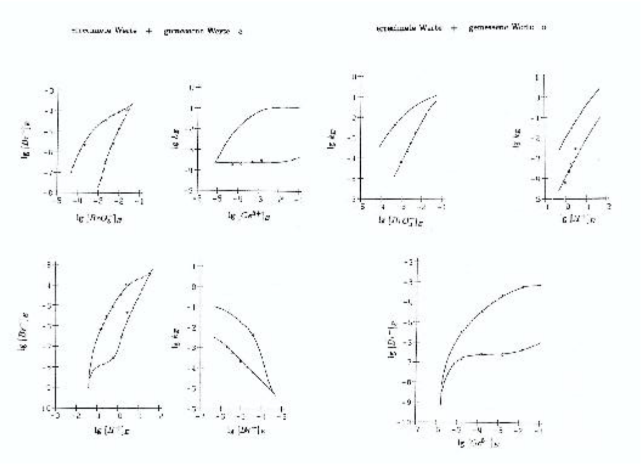
- Measurement points: 44 points on 4-dim manifold of turning points (Geiseler, Bar-Eli '81)
- Manifold depends on 14 unknown parameters (rate constants)
- Problem: fit 4-dim manifold on 44 5-dim points, which are implicitly given by $\otimes$

$\rightarrow$ 894 variables; 838 nonlinear equations 88 least squares terms 498 positivity constraints

Belousov Zhabotinskii Reaction: Simulation for Initial Guesses

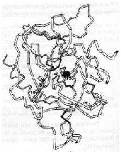


Belousov Zhabotinskii Reaction – Solution



Enzyme Reaction Kinetics

K. et al 2001



- enzymes = biocatalysts, highly active
- demand from industry since they accelerate biochemical reactions
- but: great expenses for the evaluation of the long-term behaviour
- practice: very many expensive experiments are carried out

Enzyme Reaction Kinetics

K. et al 2001

$$
\begin{array}{ccc}
\text{N} & \overset{k_{NU}}{\underset{k_{UN}}{\rightleftarrows}} & \text{U} \\
k_{ND} \searrow & & \swarrow k_{UD} \\
 & \text{D} &
\end{array}
$$

N: native enzyme, measurable
U: unfolded enzyme, not measurable
D: deactivated enzyme, not measurable

Enzyme Reaction Kinetics

K. et al 2001

$$\frac{d\ C_D}{d\ t} = \left(k_d^0 \exp\left(\frac{-\Delta h_u^*}{RT}\right) K_U + k_N^0 \exp\left(\frac{-\Delta h_N^*}{RT}\right)\right) \frac{C_{E_0} - C_D}{1 + K_U},$$

$$\frac{d\ C_S}{d\ t} = \frac{\dot{V}}{V}\left(C_S^0 - C_S\right) - r_{max}\frac{C_S}{k_m + C_S},$$

$$C_D(0) = 0, \qquad C_S(0) = C_S^0,$$

$$K_U = \exp\left(\frac{-\Delta h_u^0}{RT}\right) \exp\left(\frac{\Delta S_u^0}{R}\right),$$

$$r_{max} = A \exp\left(\frac{-\Delta h_E^*}{RT}\right) \frac{C_{E_0} - C_D}{1 + K_U}.$$

Application: Enzyme Reaction Kinetics

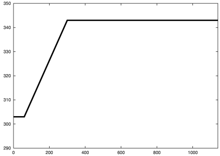
Cooperation with degussa.

- nonlinear Arrhenius kinetics
- 8 unknown parameters p
- 1 time dependent control function $u(t)$: temperature
- 1 indirect measurement: consumption of base necessary to neutralize the acidic reaction product (side-reaction!)
- quantities describing stability (total turn-over number and half-life) are of interest!
- problem is too ill-conditioned, impossible to identify parameters from 1 experiment!

Experiments with *Candida antarctica* on ionic resin ("Novozym")

l_2 parameter estimation from the standard experiment: estimated values of parameters $\pm$ standard deviation after parameter estimation

	initial profile
p_1	27.86 $\pm$ 4.42
p_2	48.98 $\pm$ 10.92
p_3	1.73 $\pm$ 2.39 $\times 10^5$
p_4	634.20 $\pm$ 806.00 $\times 10^6$
p_5	-1.43 $\pm$ 1.50 $\times 10^7$
p_6	-7.50 $\pm$ 4.16 $\times 10^7$
p_7	-4.15 $\pm$ 0.091
p_8	-8.63 $\pm$ 2.00



Can we find better experiments?

Question: Can We Determine Better Experimental Conditions?

Aim:

- choose experimental conditions $\xi = (u, q, w)$,
  - control functions: temperature profiles, feed streams,
  - control parameters: volume, initial conditions,
  - sampling design: measurement devices and times
- aim: “to maximize information gain”, here: “to minimize uncertainty of resulting parameter estimate”
- subject to state, control and parameter constraints
  - safety, domain of model validity, costs, feasibility of experiments

Outline

Summary

- numerical methods and applications
  - parameter estimation for DAE
- based on multiple shooting
- Generalized Gauss-Newton
  for l_2 and l_1 problems
- sensitivity analysis as basis for optimum experimental design
- complex nonlinear problems can be treated
- next: methods for nonlinear optimum experimental design

THANK YOU VERY MUCH FOR YOUR ATTENTION!

References

Baake, E. and Schlöder, J. P.: Modelling the fast fluorescence rise of photosynthesis. Bull. Math. Biol., 6(54), 999-1021 (1992).

Bauer, I., Bock, H. G., Körkel, S., and Schlöder, J. P.: Numerical methods for optimum experimental design in DAE systems. Journal of Computational and Applied Mathematics, 120, 1-25 (2000).

Bock, H. G.: Randwertproblemmethoden zur Parameteridentifizierung in Systemen nichtlinearer Differentialgleichungen. Bonner Mathematische Schriften, 183, Bonn (1987).

Bock H. G., Kostina E., and Schlöder, J. P.: Numerical Methods for Parameter Estimation in Nonlinear Differential Algebraic Equations GAMM-Mitt. 30, No. 2, 376-408 (2007)

Dieses, A.: Numerical methods for optimization problems in water flow and reactive solute transport processes of xenobiotics in soils. Dissertation, SFB Preprint 2001-07, University of Heidelberg (2001).

Gabasov, R., Kirillova, F. M., and Kostina, E. A.: An adaptive method of solving l1 extremal problems. Zhurnal vychislitelnoy matematiki i matematicheskoy fiziki, 38, 9, 1461- 1472 (1998) (in Russian, trans. to English: Computational Mathematics and Mathematical Physics, 38, 9, 1400-1411 (1998).