Some Good Old But Sometimes Neglected Aspects in Numerical Parameter Estimation

P.W. Hemker

Amsterdam

Eindhoven, June 2009

90606:21h47

Contents:

- Some personal background
    - Problems from (bio-)chemistry
- A parameter estimation procedure
    - What is relevant information to deliver ?
    - Useful numerical tricks / insights
    - Example
- What to remember?

Background:

1965 - Data analysis in biochemistry lab
- 1968: Edinburgh: Course Computational Methods in Biochemistry
- Dundee: C.W. Gear introduces BDF for stiff ODEs

1970 - Numerical analysis

1972 - Parameter estimation in nonlinear differential equations
(1 paper, 2 reports, 1 ALGOL 60-code)
B.van Domselaar (MSc-thesis 1975)

1991 - Interest from chemical industry
1993-1997 - STW-project
W.J.H.Stortelder (PhD-thesis March 1998)

2007 - Renewed interest in data analysis for biochemistry

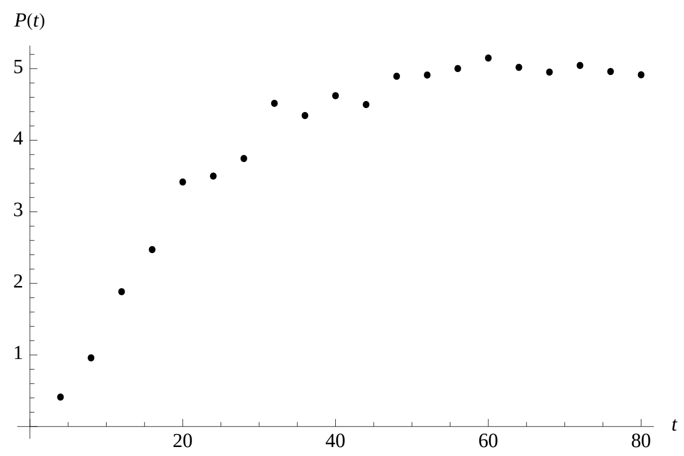
Simplest enzymatic reaction

$$S + E \underset{k_2}{\overset{k_1}{\rightleftarrows}} C \overset{k3}{\longrightarrow} E + P$$

Given:

$$S_0, \quad E_0, \quad C_0 = 0, \, P_0 = 0$$

Measurement data for $P(t)$:



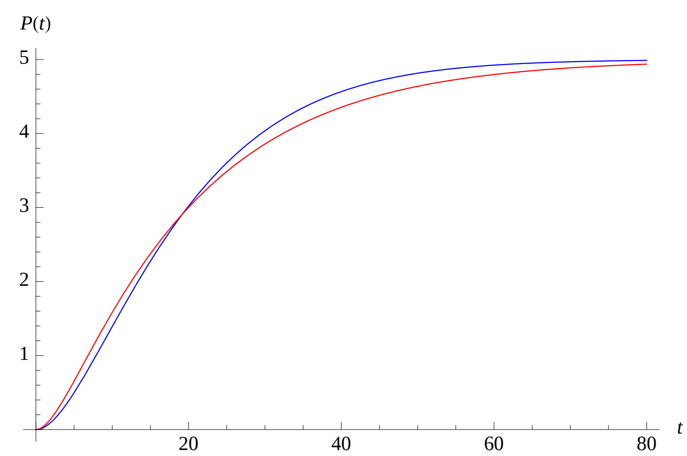
Determine: reaction constants: $k_1, \, k_2, \, k_3$

Simplest enzymatic reaction (2)

$$S + E \underset{k_2}{\overset{k_1}{\rightleftarrows}} C \overset{k3}{\longrightarrow} E + P$$

DAE system
$$\begin{cases} \frac{dc(t)}{dt} = +k_1 \cdot s(t) \cdot e(t) - (k_2 + k_3) \cdot c(t) \\ \frac{dp(t)}{dt} = +k_3 \cdot c(t) \\ E_0 = e(t) + c(t) \\ S_0 = s(t) + c(t) + p(t) \end{cases}$$

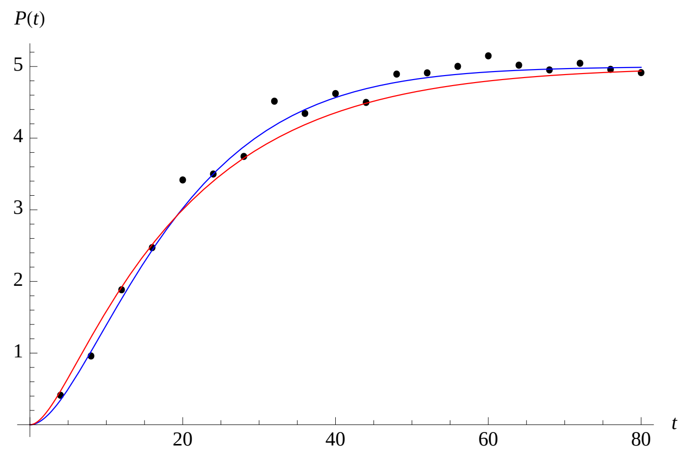
curve $p(t)$



depends on k_1, k_2, k_3

Simplest enzymatic reaction (3)

$$S + E \underset{k_2}{\overset{k_1}{\rightleftarrows}} C \overset{k3}{\longrightarrow} E + P$$



state: $\mathbf{y}(t) = \begin{pmatrix} s(t) \\ e(t) \\ c(t) \\ p(t) \end{pmatrix}$ parameters: $\mathbf{p} = \begin{pmatrix} k_1 \\ k_2 \\ k_3 \end{pmatrix}$

Error: normal distribution

The problem

The model (DAE system):

$$\boldsymbol{\alpha}\,\frac{d\mathbf{y}(t;\mathbf{p})}{dt} = f(\mathbf{y};\mathbf{p})\,, \qquad\qquad \mathbf{y}(0;\mathbf{p}) = \mathbf{y}_0(\mathbf{p})$$

The data: (arbitrarily spaced; usually a single component)

$$\{t_i,\ \widetilde{y}_{c_i,i}\}_{i=1,\dots,N}$$

Assumption

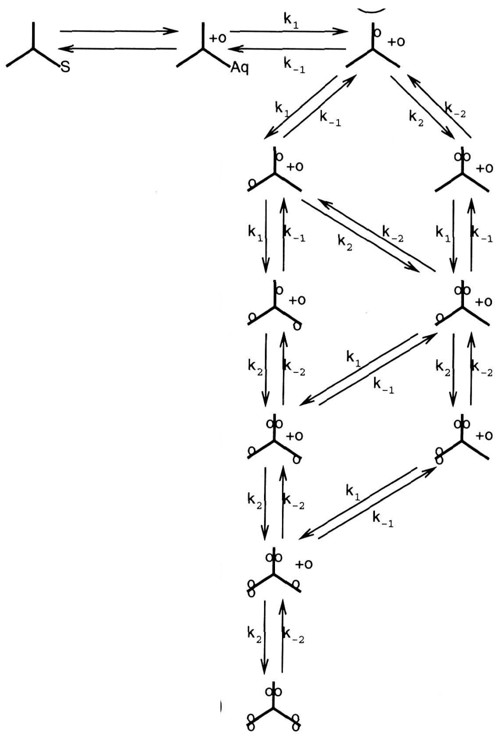
$$\widetilde{y}_{c_i,i} \approx \mathbf{y}_{c_i}(t_i;\mathbf{p}) \qquad \text{for some } \mathbf{p}$$

Measurement error:

- normal distribution
- mutually independent

First question: what $\mathbf{p}$?

Typical example



$$
\begin{aligned}
\frac{\mathrm{d}[FM]}{\mathrm{d}t} &= -k_1[FM]\left(6[melAq]+4[mon]+2[di]+\right.\\
&\quad 4[NN]+2[NNN']+2[NNN'N']\left.\right)-\\
&\quad k_2[FM]\left([mon]+2[di]+3[tri]+[NNN']+2[tet]+[pen]\right)+\\
&\quad k_{-1}[H_2O]\left([mon]+2[di]+3[tri]+[NNN']+2[tet]+[pen]\right)+\\
&\quad k_{-2}[H_2O]\left(2[NN]+2[NNN']+2[tet]+\right.\\
&\quad 4[NNN'N']+4[pen]+6[hex]\left.\right)\ ,\\
\frac{\mathrm{d}[H_2O]}{\mathrm{d}t} &= -\frac{\mathrm{d}[FM]}{\mathrm{d}t}\ ,\\
\frac{\mathrm{d}[mon]}{\mathrm{d}t} &= 6k_1[FM][melAq]+2k_{-1}[H_2O][di]+2k_{-2}[H_2O][NN]-\\
&\quad 4k_1[FM][mon]-k_2[FM][mon]-k_{-1}[H_2O][mon]\ ,\\
\frac{\mathrm{d}[NN]}{\mathrm{d}t} &= k_2[FM][mon]+k_{-1}[H_2O][NNN']-\\
&\quad 4k_1[FM][NN]-2k_{-2}[H_2O][NN]\ ,\\
\frac{\mathrm{d}[di]}{\mathrm{d}t} &= 4k_1[FM][mon]+3k_{-1}[H_2O][tri]+2k_{-2}[H_2O][NNN']-\\
&\quad 2k_1[FM][di]-2k_2[FM][di]-2k_{-1}[H_2O][di]\ ,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}[NNN']}{\mathrm{d}t} &= 4k_1[FM][NN]+2k_2[FM][di]+4k_{-2}[H_2O][NNN'N']+\\
&\quad 2k_{-1}[H_2O][tet]-k_2[FM][NNN']-2k_1[FM][NNN']-\\
&\quad 2k_{-2}[H_2O][NNN']-k_{-1}[H_2O][NNN']\ ,\\
\frac{\mathrm{d}[tri]}{\mathrm{d}t} &= 2k_1[FM][di]+2k_{-2}[H_2O][tet]-k_2[FM][tri]-3k_{-1}[H_2O][tri]\ ,\\
\frac{\mathrm{d}[NNN'N']}{\mathrm{d}t} &= k_2[FM][NNN']+k_{-1}[H_2O][pen]-2k_1[FM][NNN'N']-\\
&\quad 4k_{-2}[H_2O][NNN'N']\ ,\\
\frac{\mathrm{d}[tet]}{\mathrm{d}t} &= 3k_2[FM][tri]+2k_1[FM][NNN']+4k_{-2}[H_2O][pen]-\\
&\quad 2k_2[FM][tet]-2k_{-2}[H_2O][tet]-2k_{-1}[H_2O][tet]\ ,\\
\frac{\mathrm{d}[pen]}{\mathrm{d}t} &= 2k_2[FM][tet]+2k_1[FM][NNN'N']-k_{-1}[H_2O][pen]-\\
&\quad 4k_{-2}[H_2O][pen]+6k_{-2}[H_2O][hex]-k_2[FM][pen]\ ,\\
\frac{\mathrm{d}[hex]}{\mathrm{d}t} &= k_2[FM][pen]-6k_{-2}[H_2O][hex]\ .
\end{aligned}
$$

scheme

11 differential equations

$$k_i(T) = \alpha_i \exp\left(\frac{-E_i}{RT}\right), \qquad i \in \{-2,-1,1,2\}.$$

16 parameters

Solution procedure (nonlinear least squares)

- choose $\mathbf{p}$ (initial estimate)
- solve $\mathbf{y}(t, \mathbf{p})$ from

$$\frac{d\mathbf{y}(t;\mathbf{p})}{dt} = f(\mathbf{y}, \mathbf{p})$$

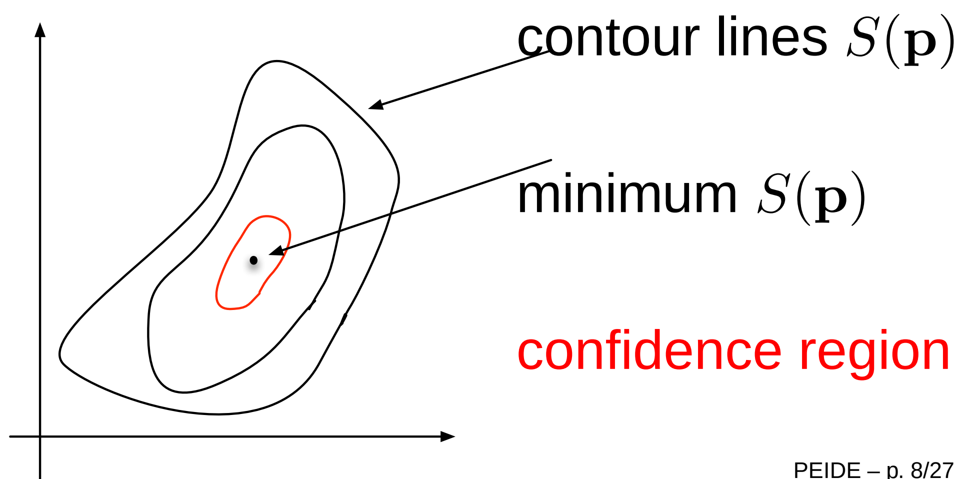
- compute discrepancy with the data

$$Y(\mathbf{p}) = \{\mathbf{y}_{c_i}(t_i, \mathbf{p}) - \widetilde{y}_{c_i,i}\}$$

$$S(\mathbf{p}) = \|Y(\mathbf{p})\|^2 = \sum_i w_i^2 \left(\widetilde{y}_{c_i,i} - \mathbf{y}_{c_i}(t_i, \mathbf{p})\right)^2$$

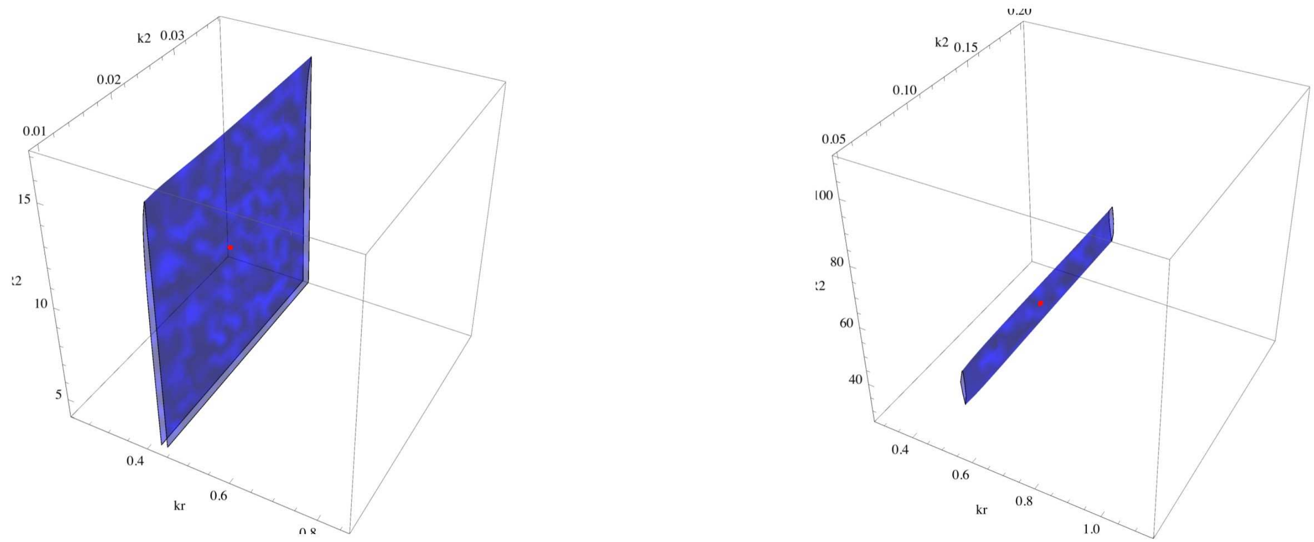
- find (local) minimum for $S(\mathbf{p})$

parameter space



What to compute ?

- $\mathbf{p}$ is not the most important information
- often there doesn't exists a unique $\mathbf{p}$
  - a unique $\mathbf{p}$ does exist in the mathematical sense
  - a unique $\mathbf{p}$ doesn't exist in the practical sense



- the location and the shape of the 'confidence region' is important ! in particular the "dimension"

Solution procedure (2)

- efficient and reliable minimisation $S(\mathbf{p}) = \|Y(\mathbf{p})\|^2$

- by iterative solution the non-linear least squares problem

$$\begin{aligned} J(\mathbf{p})\,\delta\mathbf{p} &= -Y(\mathbf{p}) \\ \mathbf{p}_{\text{new}} &= \mathbf{p}_{\text{old}} + \delta\mathbf{p} \end{aligned}$$

where

$\Rightarrow$ $Y(\mathbf{p})$
vector of discrepancies $\widetilde{y}_{j,i} - \mathbf{y}_{c_i}(t_i, \mathbf{p})$

$\Rightarrow$ $J(\mathbf{p}) = \frac{dY(\mathbf{p})}{d\mathbf{p}} = Y_p(p)$
sensitivity or Jacobian matrix $\left(\partial \mathbf{y}_{c_i}(t_i, \mathbf{p}) / \partial \mathbf{p}_j\right)$

are efficienly computed simultaneously !!!

Efficient solution of the DEAs and sensitivity equations

- original equations

$$\frac{d\mathbf{y}(t,\mathbf{p})}{dt} = f(\mathbf{y},\mathbf{p})$$

usually stiff equations $\rightarrow$ implicit integration rule needed
use BDF formula

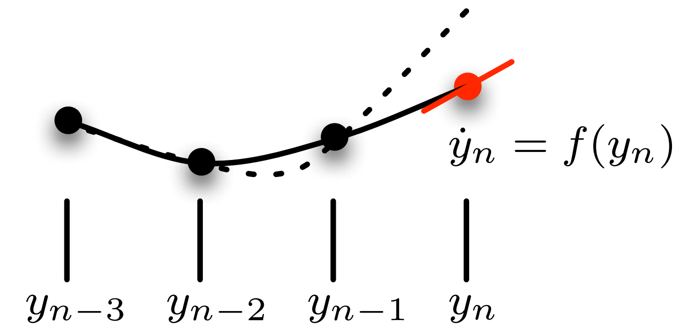
- sensitivity equations

$$\frac{d\mathbf{y_p}(t,\mathbf{p})}{dt} = f_{\mathbf{p}}(\mathbf{y},\mathbf{p}) + f_{\mathbf{y}}(\mathbf{y},\mathbf{p}) \,.\, \mathbf{y_p}(t,\mathbf{p})$$

- only one-sided dependence
- sensitivity equations are linear
- stiffness of sensitivity equations is identical with original equation
- stepsize and order control by original equations only
- given $f(\mathbf{y},\mathbf{p})$, the functions $f_{\mathbf{p}}(\mathbf{y},\mathbf{p})$ and $f_{\mathbf{y}}(\mathbf{y},\mathbf{p})$ are easily provided by a computer algebra package

$\Rightarrow$ the solution of the sensitivity equations is almost for free

The sensitivity / slave equations



BDF model equation

$$\dot{y} = f(y)$$

$$y_n = \beta h f(y_n) + \Phi(y_{n-1}, y_{n-2}, \ldots)$$

$$y_n^{(k+1)} = y_n^{(k)} - (I - \beta h f_y)^{-1} \left(y_n^{(k)} - h\beta f(y_n^{(k)} - \Phi \right)$$

BDF model + sensitivity equation

$$\begin{cases} \dot{y} &= f(y) \\ \dot{w} &= f_p(y) + f_y w \end{cases} \qquad w \equiv \partial \mathbf{y} / \partial \mathbf{p}$$

$$w_n = \beta h \ f_p(y_n) + \beta h \ f_y(y_n) \cdot w_n + \Psi_n$$

$$w_n = (I - \beta h f_y)^{-1} (\beta h \ f_p(y_n) + \Psi)$$

linear dependence — re-use of $(I - \beta h f_y)^{-1}$ — no iteration

Solution procedure (3)

- shape and location of the confidence region

  minimisation of $S(\mathbf{p}) = \|Y(\mathbf{p})\|^2$

– by integration of the DAEs we compute:

$$Y(\mathbf{p}) = \left(\mathbf{y}_{c_i}(t_i, \mathbf{p}) - \widetilde{y}_{c_i,i}\right)_{i=1\dots N}$$

$$J(\mathbf{p}) = \frac{dY(\mathbf{p})}{d\mathbf{p}} = \left(\frac{dy_{c_i}(t_i, \mathbf{p})}{d\mathbf{p}}\right)_{i=1\dots N,\ j=1\dots m}$$

– by the iteration process $\mathbf{p} \rightarrow \mathbf{p} + \delta\mathbf{p}$
with $\delta\mathbf{p}$ from the normal equations

$$J^T J\ \delta\mathbf{p} = -J^T\ Y$$

better: Levenberg- Marquardt (to handle the nonlinearity)

$$\left(J^T J + \lambda\ I_m\right)\ \delta\mathbf{p} = -J^T\ Y$$

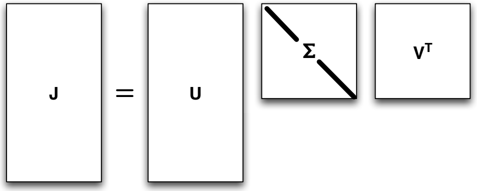
λ controls regularization & trust region

Minimization and confidence interval

Levenberg- Marquardt (λ for regularization and damping)

$$\left(J^T J + \lambda\, I_m\right)\ \delta\mathbf{p} = -J^T\, Y$$

Singular Value Decomposition !!! $J = U\,\Sigma\,V^T$



$$\delta\mathbf{p} = -V\ \left(\Sigma^T\Sigma + \lambda\, I_m\right)^{-1}\ \Sigma U^T Y$$

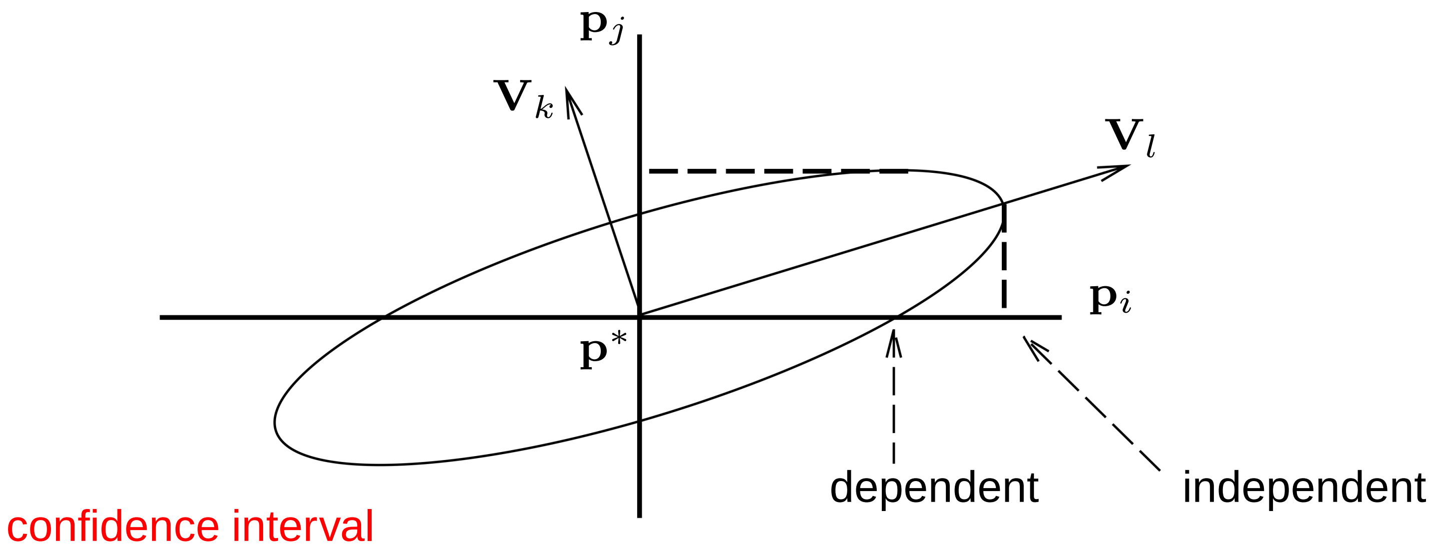
classical multivariate statistical analysis:

$$\Delta\mathbf{p}^T V \Sigma^2 V^T \Delta\mathbf{p} \ = \ \Delta\mathbf{p}^T J^T J\,\Delta\mathbf{p} \ \leq \ \frac{m}{N-m}\, S(\mathbf{p}^*)\,\mathcal{F}_\alpha(m, N-m)$$

$\mathcal{F}_\alpha(m, N-m)$: Fisher's F-distribution with m and $N-m$ d.o.f

Minimization and confidence interval (2)

$$\Delta\mathbf{p}^T\ V\ \Sigma^2\ V^T\ \Delta\mathbf{p} = \text{constant}$$



Linear approximation of reality

useful information

harder to visualise in more dimensions

Simple example (Barnes)

Scheme (Lotka-Volterra)

$$
\begin{aligned}
A &\xrightarrow{k_1} A + A \\
A + B &\xrightarrow{k_2} B + B \\
B &\xrightarrow{k_3} C
\end{aligned}
$$

Differential equations

$$
\frac{dx}{dt} = k_1\, x - k_2\, x\, y \; , \qquad \frac{dy}{dt} = k_2\, x\, y - k_3\, y
$$

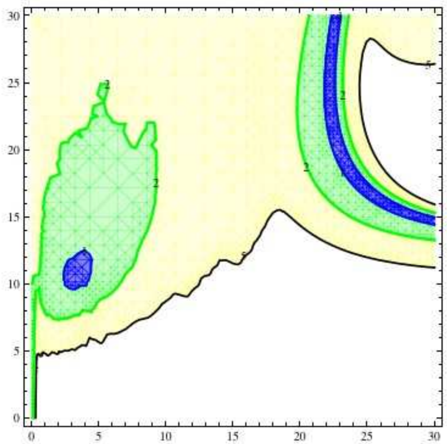
Data (Barnes, Edinburgh 1968)

t_i :	0.5	1.	1.5	2.	2.5	3.	3.5	4.	4.5	5.
x_i :	1.1	1.3	1.1	0.9	0.7	0.5	0.6	0.7	0.8	1.
y_i :	0.35	0.4	0.5	0.5	0.4	0.3	0.25	0.25	0.3	0.35

Contourplot residue (Barnes)

k_1 vs k_2

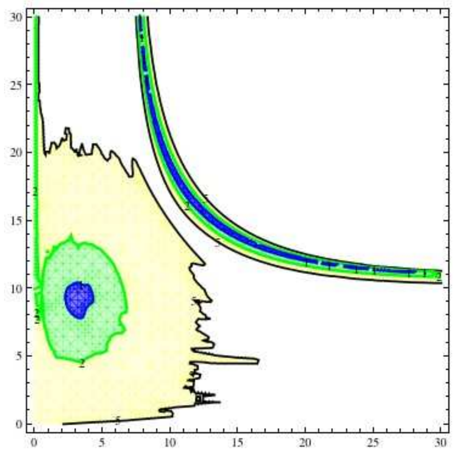
k_2 ↑



$k_3 = 10.0$ $\rightarrow k_1$

k_1 vs k_3

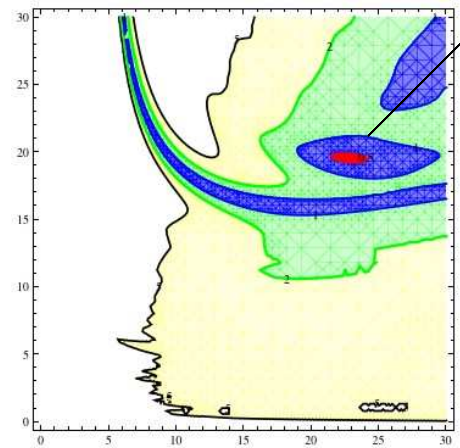
k_3 ↑



$k_2 = 10.0$ $\rightarrow k_1$

k_2 vs k_3

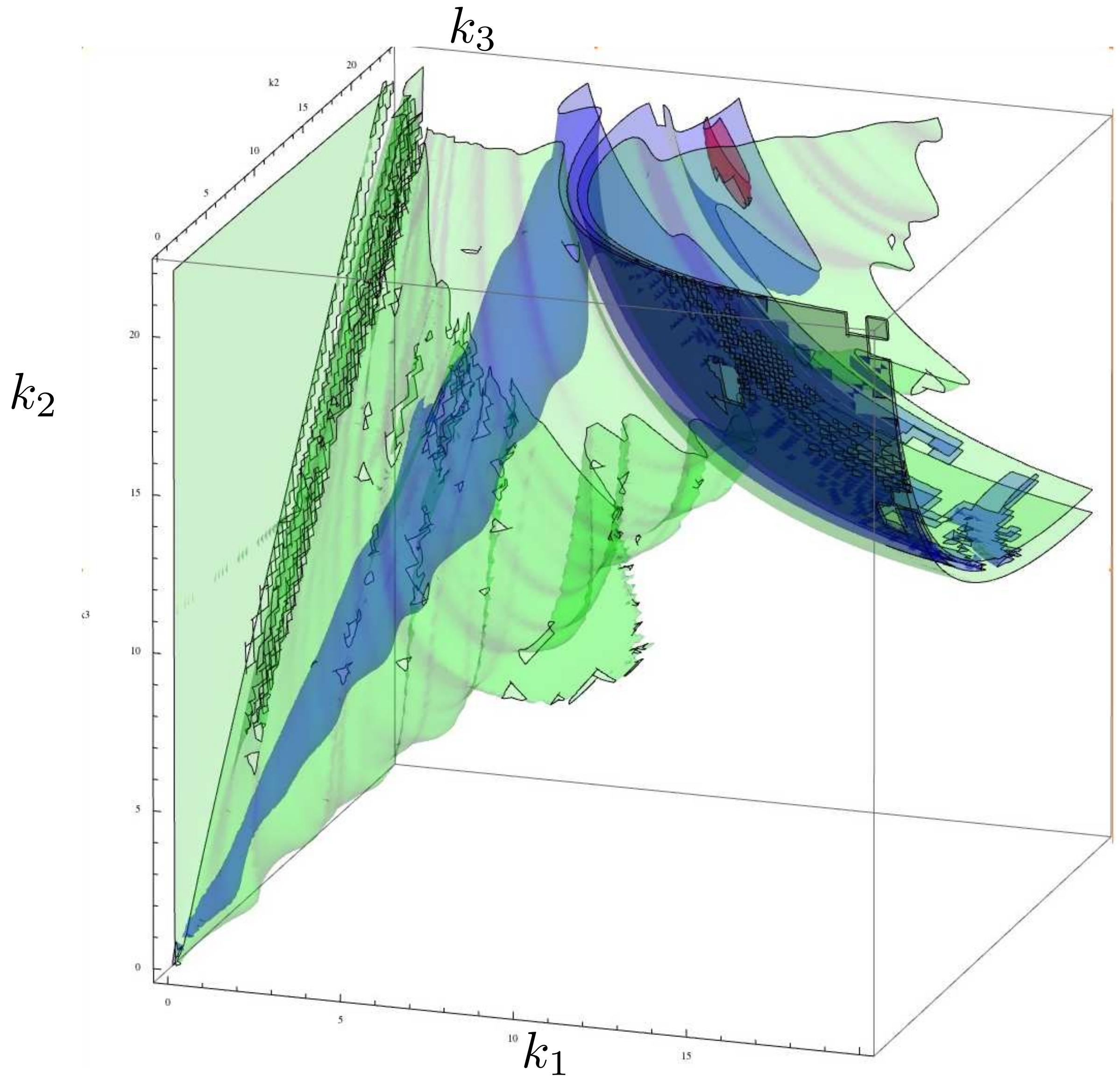
k_3 ↑



$k_2 = 10.0$ $\rightarrow k_1$

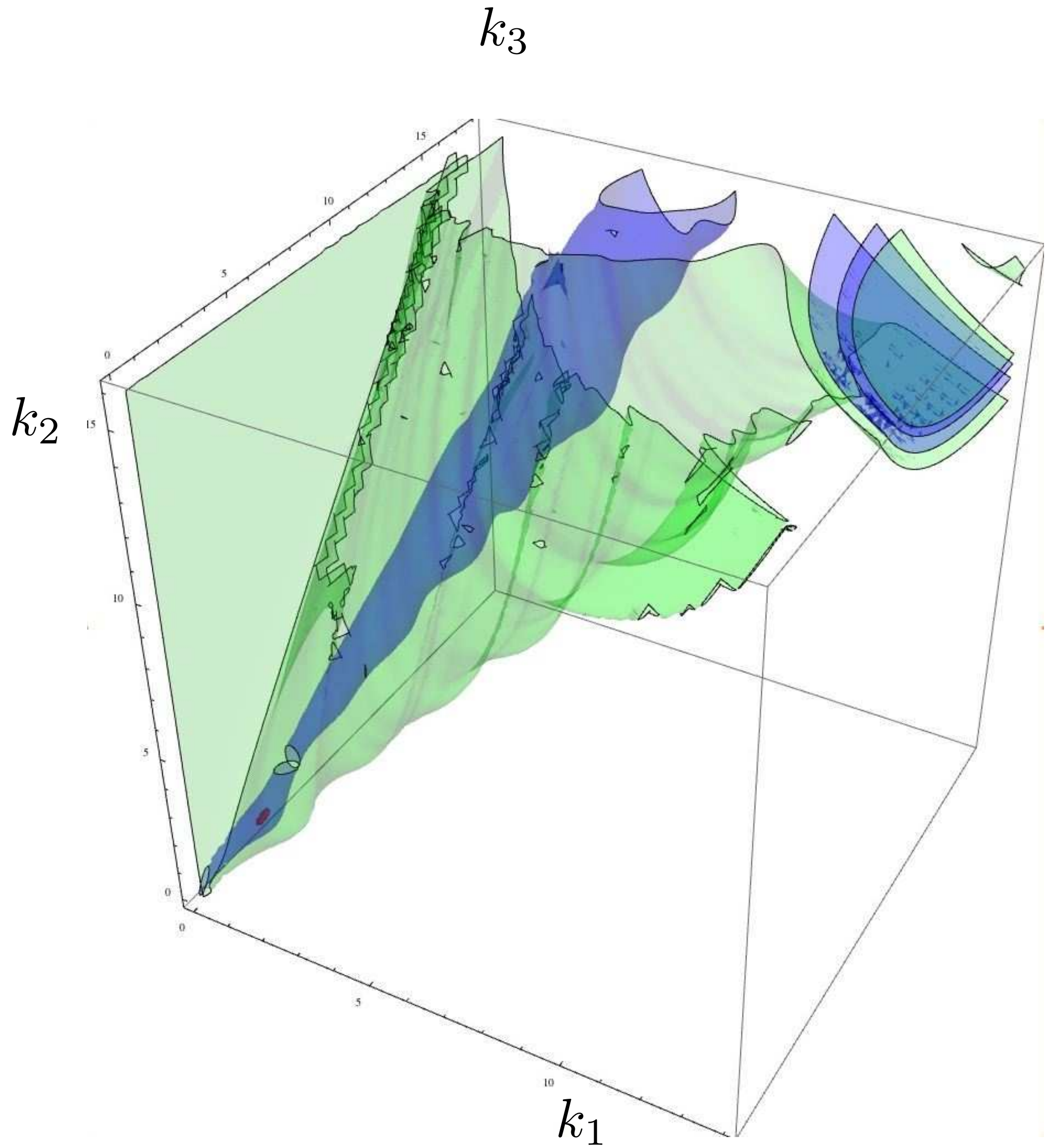
residu = $\|\widetilde{y}_{c_i,i} - \mathbf{y}_j(t_i)\|$ (sum of squares)

3-dimensional situation (Barnes)



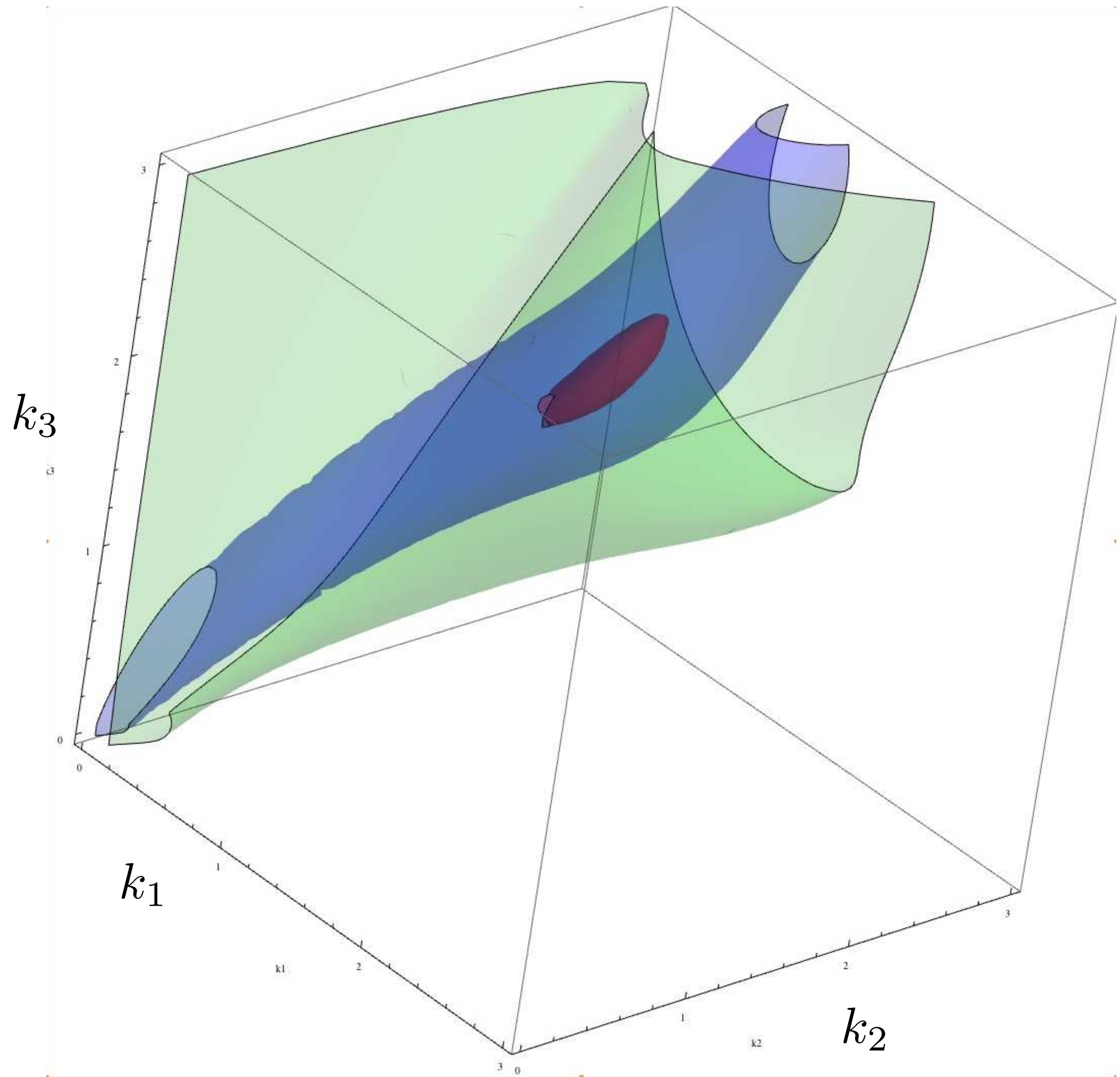
parameter space

3-dimensional situation (Barnes)



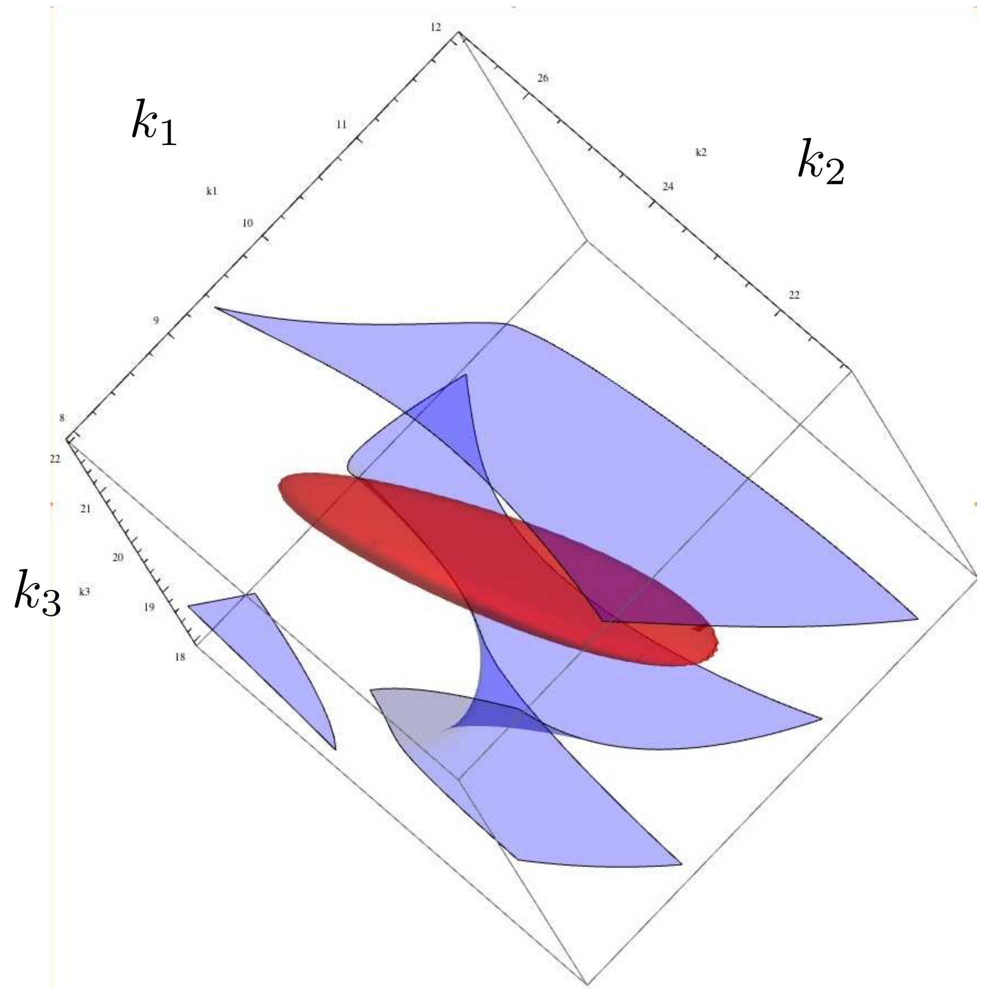
parameter space

Zoom in   bottom



parameter space

Zoom in top

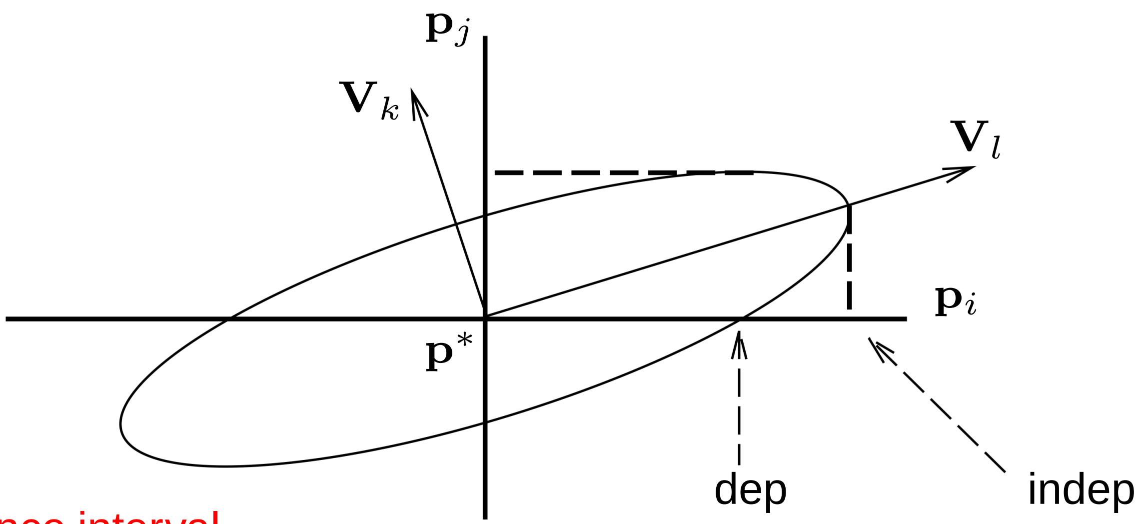


parameter space

Visualize singular value information

useful information for the user !!!

$$\Delta \mathbf{p}^T \; J^T J \; \Delta \mathbf{p} = \Delta \mathbf{p}^T \; V \; \Sigma^2 \; V^T \; \Delta \mathbf{p}$$



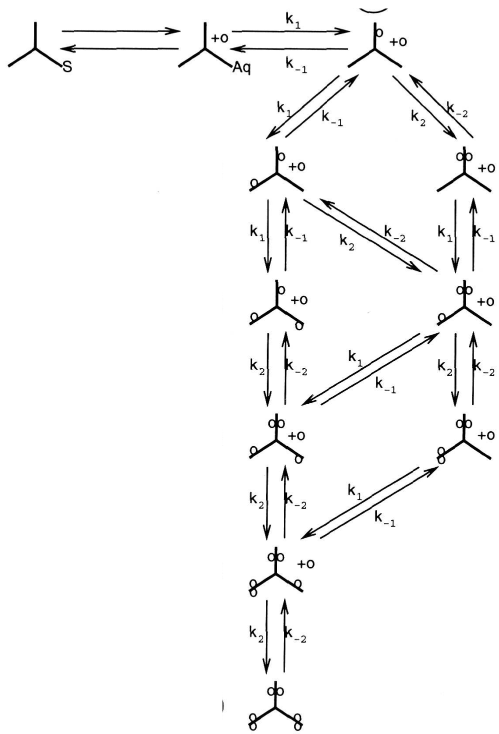
confidence interval

length of the axes: $\frac{1}{\sigma_k}, \frac{1}{\sigma_l}$

larger $\sigma \rightarrow$ higher accuracy

linear approximation of reality

Typical example



$$
\begin{aligned}
\frac{\mathrm{d}[FM]}{\mathrm{d}t} &= -k_1[FM]\,(6[melAq] + 4[mon] + 2[di] + \\
&\quad 4[NN] + 2[NNN'] + 2[NNN'N']) - \\
&\quad k_2[FM]\,([mon] + 2[di] + 3[tri] + [NNN'] + 2[tet] + [pen]) + \\
&\quad k_{-1}[H_2O]\,([mon] + 2[di] + 3[tri] + [NNN'] + 2[tet] + [pen]) + \\
&\quad k_{-2}[H_2O]\,(2[NN] + 2[NNN'] + 2[tet] + \\
&\quad 4[NNN'N'] + 4[pen] + 6[hex]) \ , \\
\frac{\mathrm{d}[H_2O]}{\mathrm{d}t} &= -\frac{\mathrm{d}[FM]}{\mathrm{d}t} \ , \\
\frac{\mathrm{d}[mon]}{\mathrm{d}t} &= 6k_1[FM][melAq] + 2k_{-1}[H_2O][di] + 2k_{-2}[H_2O][NN] - \\
&\quad 4k_1[FM][mon] - k_2[FM][mon] - k_{-1}[H_2O][mon] \ , \\
\frac{\mathrm{d}[NN]}{\mathrm{d}t} &= k_2[FM][mon] + k_{-1}[H_2O][NNN'] - \\
&\quad 4k_1[FM][NN] - 2k_{-2}[H_2O][NN] \ , \\
\frac{\mathrm{d}[di]}{\mathrm{d}t} &= 4k_1[FM][mon] + 3k_{-1}[H_2O][tri] + 2k_{-2}[H_2O][NNN'] - \\
&\quad 2k_1[FM][di] - 2k_2[FM][di] - 2k_{-1}[H_2O][di] \ ,
\end{aligned}
$$

$$
\begin{aligned}
\frac{\mathrm{d}[NNN']}{\mathrm{d}t} &= 4k_1[FM][NN] + 2k_2[FM][di] + 4k_{-2}[H_2O][NNN'N'] + \\
&\quad 2k_{-1}[H_2O][tet] - k_2[FM][NNN'] - 2k_1[FM][NNN'] - \\
&\quad 2k_{-2}[H_2O][NNN'] - k_{-1}[H_2O][NNN'] \ , \\
\frac{\mathrm{d}[tri]}{\mathrm{d}t} &= 2k_1[FM][di] + 2k_{-2}[H_2O][tet] - k_2[FM][tri] - 3k_{-1}[H_2O][tri] \ , \\
\frac{\mathrm{d}[NNN'N']}{\mathrm{d}t} &= k_2[FM][NNN'] + k_{-1}[H_2O][pen] - 2k_1[FM][NNN'N'] - \\
&\quad 4k_{-2}[H_2O][NNN'N'] \ , \\
\frac{\mathrm{d}[tet]}{\mathrm{d}t} &= 3k_2[FM][tri] + 2k_1[FM][NNN'] + 4k_{-2}[H_2O][pen] - \\
&\quad 2k_2[FM][tet] - 2k_{-2}[H_2O][tet] - 2k_{-1}[H_2O][tet] \ , \\
\frac{\mathrm{d}[pen]}{\mathrm{d}t} &= 2k_2[FM][tet] + 2k_1[FM][NNN'N'] - k_{-1}[H_2O][pen] - \\
&\quad 4k_{-2}[H_2O][pen] + 6k_{-2}[H_2O][hex] - k_2[FM][pen] \ , \\
\frac{\mathrm{d}[hex]}{\mathrm{d}t} &= k_2[FM][pen] - 6k_{-2}[H_2O][hex] \ .
\end{aligned}
$$

scheme

11 differential equations

$$k_i(T) = \alpha_i \exp\left(\frac{-E_i}{RT}\right), \qquad i \in \{-2, -1, 1, 2\}.$$

16 parameters

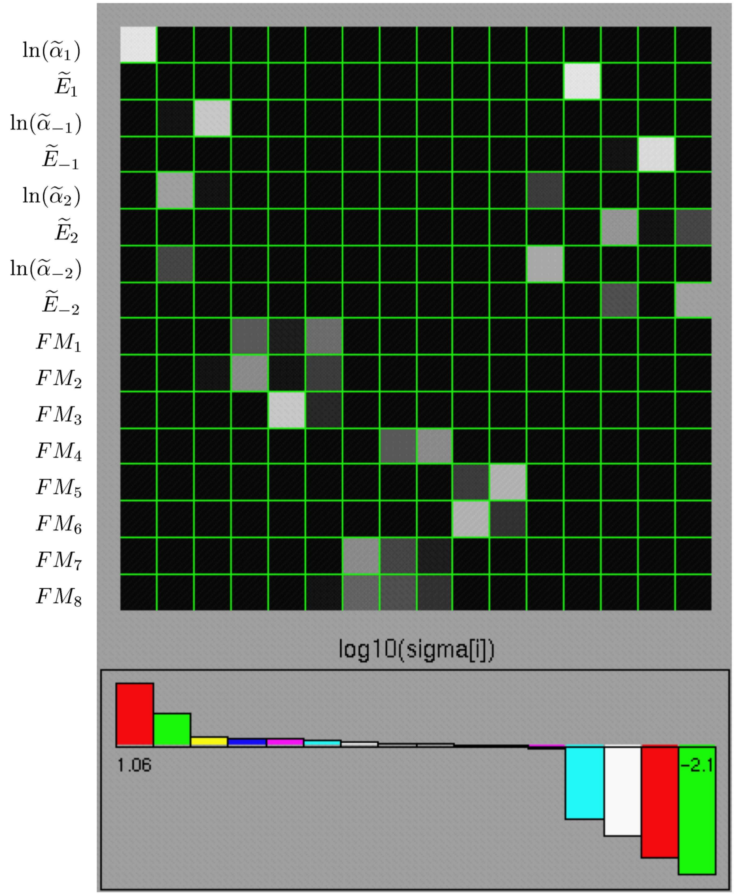
First results (16 parameter example)

	initial estimates (θ_{ini})	final estimates ($\widehat{\theta}$)	independent confidence regions ($\Delta^I \theta$)	dependent confidence regions ($\Delta^D \theta$)
$\ln(\widetilde{\alpha}_1)$	-2.74	-3.376	0.134	0.073
$\widetilde{E}_1$	98.00	65.33	14.0	7.38
$\ln(\widetilde{\alpha}_{-1})$	-4.68	-8.047	0.65	0.467
$\widetilde{E}_{-1}$	68.00	91.91	57.2	38.3
$\ln(\widetilde{\alpha}_2)$	-8.15	-4.181	0.621	0.261
$\widetilde{E}_2$	120.00	54.23	61.4	25.1
$\ln(\widetilde{\alpha}_{-2})$	-9.49	-7.986	0.893	0.405
$\widetilde{E}_{-2}$	90.00	53.03	88.9	38.2
FM_1	8.41	8.743	0.621	0.582
FM_2	7.61	8.534	0.608	0.578
FM_3	5.6	5.097	0.607	0.604
FM_4	5.58	6.097	0.712	0.702
FM_5	4.8	4.672	0.766	0.760
FM_6	4.81	4.723	0.768	0.752
FM_7	4.8	5.382	0.694	0.686
FM_8	5.58	6.065	0.703	0.683
$S(\theta)$	335.7	14.77		

Table 6.4: Initial and final estimates of θ, plus confidence regions (cf. (1.25) and (1.26) with $\alpha = 0.05$), after reparametrisation of the pre-exponential factor.

Singular value information

parameters
to be estimated



singular values in decreasing order

Conclusion

For the numerical solution of the DE: use BDF-formulas

- solves accurately stiff DEs and DAEs
- gradient information is cheap to obtain
- codes can be generated automatically

Information from the Jacobian's singular value decomposition

- is used for regularisation of the ill-posed problem.
- gives most invaluable information for the user

References

- Stortelder, W. J. H.
  "Parameter estimation in nonlinear dynamic systems"
  Series: CWI Tracts, Vol. 124, pp. 1 - 176,
  CWI, 1998, ISBN: 90-6196-482-2.

- L. Petzold, S. Li, Y. Cao and R. Serban
  "Sensitivity analysis of differential-algebraic equations and partial differential equations"
  Computers & Chemical Engineering, Volume 30, Issues 10-12, September 2006, pp. 1553 - 1559