What is...
Parameter Identification?

S. Röblitz

Zuse Institute Berlin (ZIB)

December 10, 2014

Outline

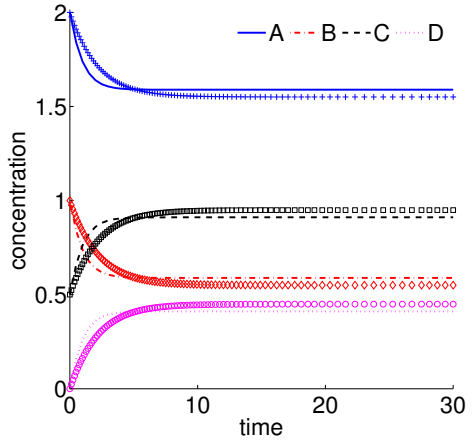
Outline

Introductory Example

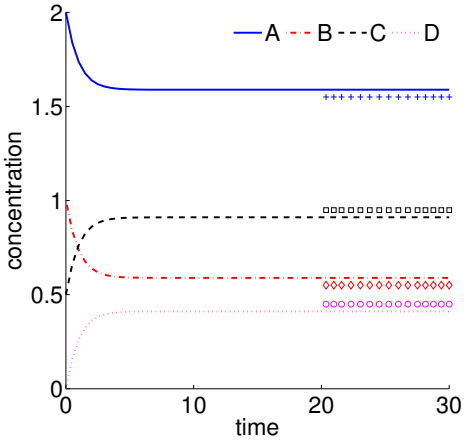
$$A + B \underset{k_2}{\overset{k_1}{\rightleftharpoons}} C + D$$

$$A' = B' = -k_1 AB + k_2 CD, \quad C' = D' = +k_1 AB - k_2 CD$$

$$A(0) = 2,\ B(0) = 1,\ C(0) = 0.5,\ D(0) = 0, \quad k_1^0 = 0.2,\ k_2^0 = 0.5$$



Case (T+E): measurement data cover both transient and equilibrium phase



Case (E): measurement data cover equilibrium phase only

Outline

Problem formulation

Model: $y(t,p) = (y_1(t,p), \ldots, y_d(t,p)) \in \mathbb{R}^d$

$$y' = f(y,p), \quad y(t_0) = y_0$$

Usually nonlinear in p!

Parameters: $p = (p_1, ..., p_q) \in \mathbb{R}^q$

Data: $z_{kl} \approx y_k(t_l, p),\ k = 1, \ldots, d,\ l = 1, \ldots, M_k$

mostly $m := \sum_k M_k \gg q \quad \rightsquigarrow$ data compression

Task: fit the model (ODE) to data, estimate parameters

Problem formulation

Questions:

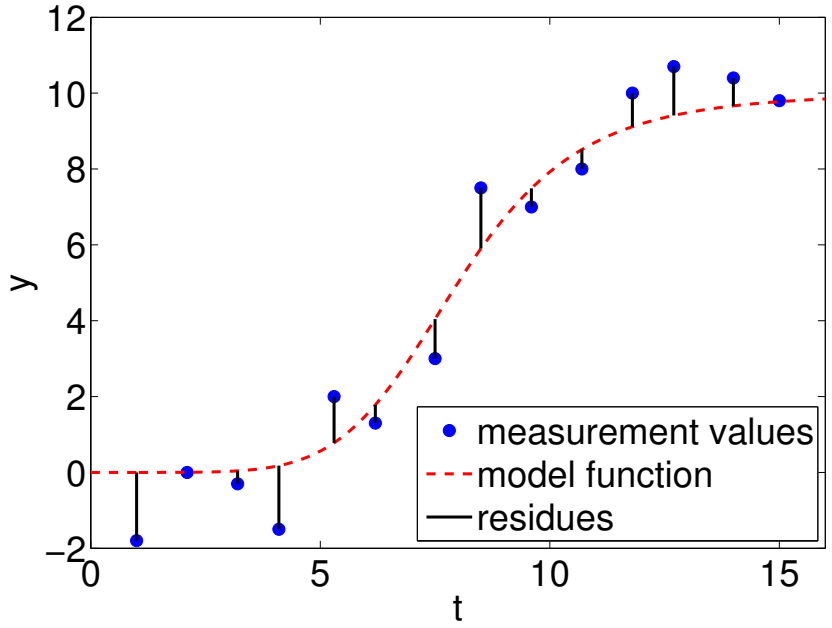
- What algorithm to use?
- How good is the fit?
- Is there a better model?
- Sensitivity?

Problem formulation

minimize least squares error

$$\|F(p)\|_2^2 = \sum_{k=1}^{d} \sum_{l=1}^{M_k} \frac{(z_{kl} - y_k(t_l, p))^2}{2\sigma_{kl}^2} \rightarrow \min$$

σ_{kl}: measurement error (standard deviation)



Approaches

- Frequentist
  - Assumes that there is an unknown but fixed parameter p
  - Estimates p with some confidence
  - Prediction by using the estimated parameter value
- Bayesian
  - Represents uncertainty about the unknown parameter
  - Uses probability to quantify this uncertainty; unknown parameters as random variables
  - Prediction follows rules of probability

Outline

Likelihood

assumption: normally distributed measurement noise

$$P(z|p) \propto \exp\left(-\sum_{k=1}^{d}\sum_{l=1}^{M_k}\frac{(z_{kl}-y_k(t_l,p))^2}{2\sigma_{kl}^2}\right)$$

maximum likelihood estimate (MLE):

$$p_{\text{MLE}} = \text{argmax}_p P(z|p)$$

Bayes theorem

Joint probability = Conditional Probability x Marginal Probability

$$P(p, z) = P(p|z)P(z) = P(z|p)P(p)$$

$$P(p|z) = \frac{P(z|p)P(p)}{P(z)}$$

$P(p)$: prior
$P(z|p)$: likelihood
$P(p|z)$: posterior (probability of parameter given the data)
$P(z) = \int P(z|p)P(p)dp$: evidence

maximum a posteriori estimate (MAP):

$$p_{\text{MAP}} = \text{argmax}_p P(p|z)$$

Evaluation of posterior

$$P(p|z) \propto P(z|p)P(p)$$

- evaluation of posterior by **Markov chain Monte Carlo (MCMC)** sampling (Metropolis-Hastings algorithm)
  $\rightarrow$ convergence?
- consider full high dimensional posterior PDF for statistical inference
- necessity of a prior!

Outline

1. Introductory Example
2. Problem Formulation
3. The Bayesian Approach
4. The Local Optimization Approach
   - Linear Least Squares Problems
   - Nonlinear Least Squares Problems
   - Specification for ODE Models

Linear least squares

$$y(t,p) = B(t)p, \quad B(t) \in \mathbb{R}^{d\times q}$$

Measurement time points: $t_1, \ldots, t_M \in \mathbb{R}$
Measurement values: $z_1, \ldots, z_M \in \mathbb{R}^d$
Set $A(t) = [B(t_1), \ldots, B(t_M)]^T, \quad z = [z_1, \ldots, z_M]^T$

Linear least-squares problem

For given $z \in \mathbb{R}^m$ and $A \in \mathbb{R}^{m\times q}$ with $m \geq q$ find $p \in \mathbb{R}^q$ such that

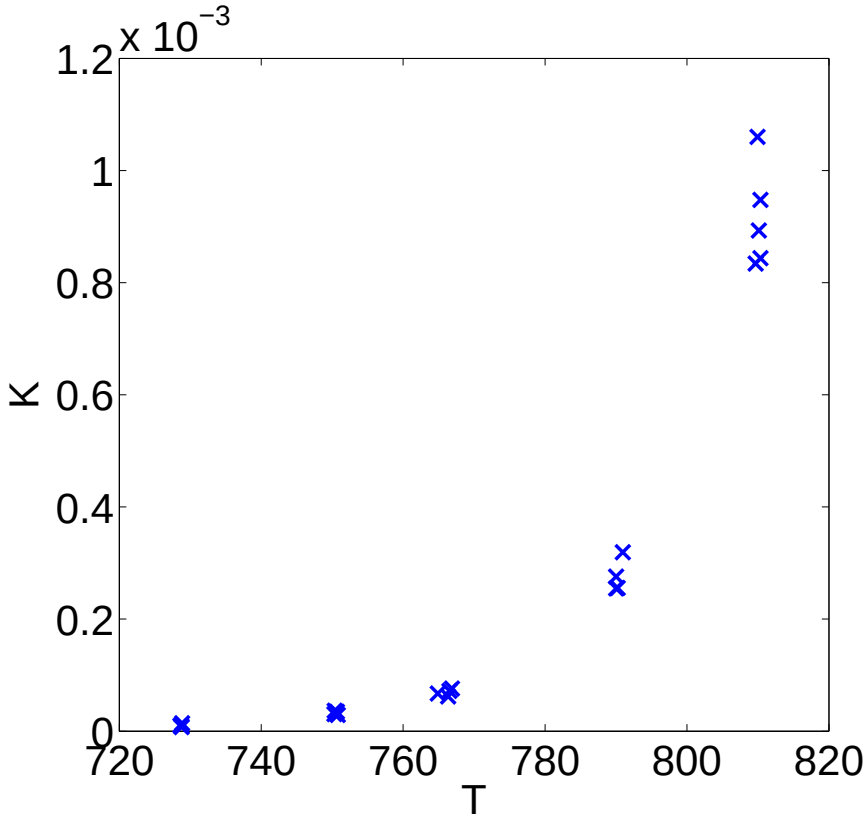
$$\|z - Ap\|_2 \to \min .$$

Example: Reaction kinetics

- Arrhenius law: dependence of rate constant K on temperature T

$$K(T) = C \cdot \exp\left(-\frac{E}{RT}\right)$$

- gas constant: $R = 8.3145 \frac{\text{J}}{\text{mol}\cdot\text{K}}$
- unknown parameters: pre-exponential factor C activation energy E
- measurements: $(T_i, K_i),\ i = 1, \ldots, m = 21$

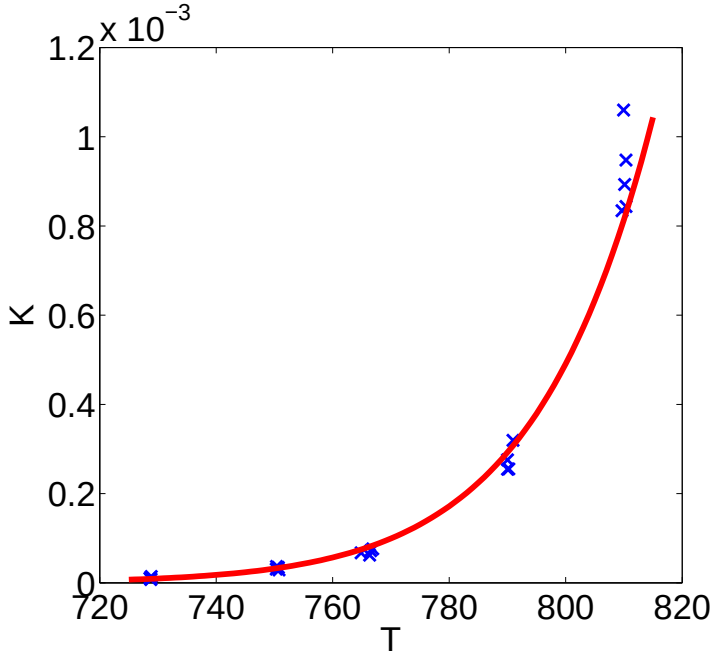


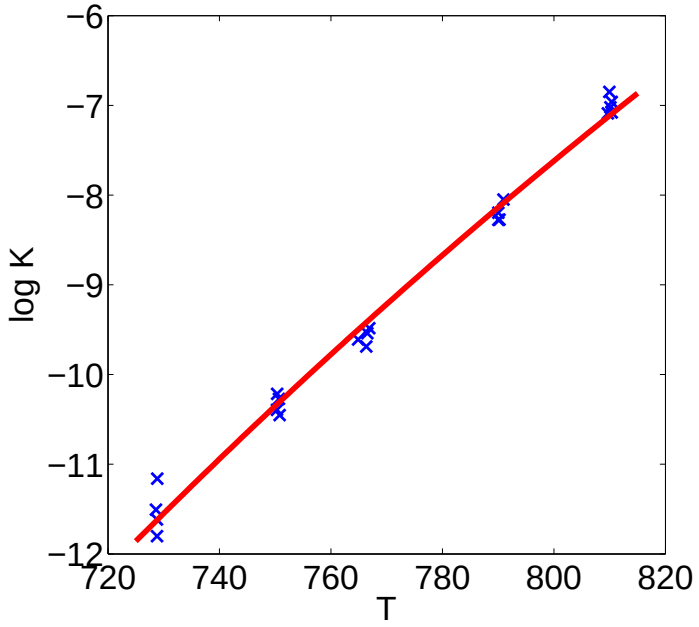
Example: Reaction kinetics

minimization problem:

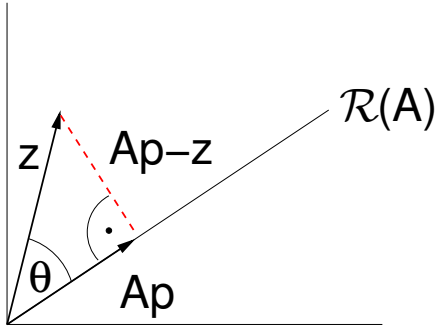
$$\log C - \frac{1}{RT_i} E = \log K_i \quad \Rightarrow \quad \| \log K - A \begin{pmatrix} \log C \\ E \end{pmatrix} \|_2 = \min$$

$$A_{i1} = 1,\ A_{i2} = -\frac{1}{RT_i},\ i = 1, \ldots, 21$$





The method of normal equations



$$\|z - Ap\|_2 = \min \Leftrightarrow z - Ap \in \mathcal{R}(A)^\perp$$
$$\mathcal{R}(A) = \{Ap : p \in \mathbb{R}^q\} \quad \text{(range space of } A\text{)}$$

$$\|z - Ap\|_2 = \min_{p' \in R^q} \|z - Ap'\|_2 \quad \Longleftrightarrow \quad A^T Ap = A^T z$$

uniquely solvable $\Leftrightarrow$ $A^T A$ invertible $\Leftrightarrow$ $\text{rank}(A) = q$

$\Rightarrow$ solution via Cholesky decomposition ($A^T A$ is spd)

if $\text{rank}(A) = q$: $\kappa_2(A)^2 \overset{\text{without proof}}{=} \frac{\lambda_{\max}(A^T A)}{\lambda_{\min}(A^T A)} = \kappa_2(A^T A)$

We need a more stable method for solving the normal eq.

Orthogonalization methods

Let $Q \in \mathbb{R}^{m \times m}$ be an orthogonal matrix, i.e. $Q^T Q = I$

$$\|z - Ap\|_2^2 = (z - Ap)^T Q^T Q (z - Ap) = \|Q(z - Ap)\|_2^2$$

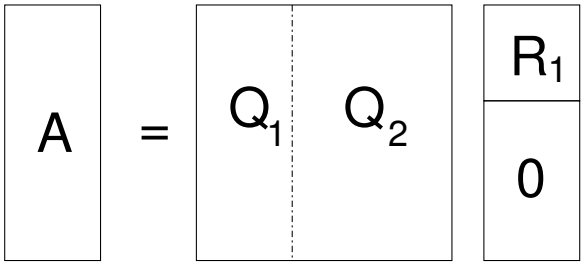
The Eucledian norm $\|\cdot\|_2$ is invariant under orthogonal transformation!

$$\|z - Ap\|_2 = \min \quad \Longleftrightarrow \quad \|Q(z - Ap)\|_2 = \min$$

Choice of the orthogonal transformation $Q \in \mathbb{R}^{m \times m}$?

QR factorization

QR factorization



$A = Q \begin{pmatrix} R_1 \\ 0 \end{pmatrix}$, $Q \in \mathbb{R}^{m \times m}$ orthogonal, $R_1 \in \mathbb{R}^{q \times q}$ upper triangular

Solution of least-squares problem: MATLAB: `[Q,R]=qr(A)`

$$\begin{aligned} \|z - Ap\|_2^2 &= \|Q^T(z - Ap)\|_2^2 = \left\| \begin{pmatrix} z_1 - R_1 p \\ z_2 \end{pmatrix} \right\|_2^2 \\ &= \|z_1 - R_1 p\|_2^2 + \|z_2\|^2 \geq \|z_2\|_2^2 \end{aligned}$$

$\text{rank}(A) = q \Rightarrow \text{rank}(R) = \text{rank}(A) = q \Rightarrow R_1$ is invertible

$$\|z - Ap\|_2 = \min \Leftrightarrow p = R_1^{-1} z_1 \Rightarrow \|r\|_2 = \|z_2\|_2$$

The QR-factorization is stable since $\kappa_2(A) = \kappa_2(R)$!

Householder QR

Columnwise application of Householder matrices to $A \in \mathbb{R}^{m \times q}$

$$Q^T A = (H_q H_{q-1} \cdots H_2 H_1) A = R$$

For $k = 1, \ldots, q$ we obtain: $\quad H_k = \mathrm{diag}(I_{k-1}, H_k') \in \mathbb{R}^{m \times m}$

$$A^{(k)} = H_k \begin{pmatrix} * & \cdots & \cdots & \cdots & \cdots & \cdots & * \\ & \ddots & & & & & \vdots \\ & & * & \cdots & \cdots & \cdots & * \\ & & & * & * & \cdots & * \\ & & & \vdots & \vdots & & \vdots \\ & & & \vdots & \vdots & & \vdots \\ & & & * & * & \cdots & * \end{pmatrix} = \begin{pmatrix} * & \cdots & \cdots & \cdots & \cdots & \cdots & * \\ & \ddots & & & & & \vdots \\ & & * & \cdots & \cdots & \cdots & * \\ & & & * & * & \cdots & * \\ & & & 0 & \vdots & & \vdots \\ & & & \vdots & \vdots & & \vdots \\ & & & 0 & * & \cdots & * \end{pmatrix}$$

Only the entries $*$ and $*$ are modified!

QR with column pivoting

Businger/Golub, Numer. Math. 7 (1965):

column permutation with the aim

$$|r_{11}| \geq |r_{22}| \geq \ldots \geq |r_{qq}|$$

step k:

$$A^{(k-1)} = \begin{pmatrix} R & S \\ 0 & T^{(k)} \end{pmatrix}$$

Denote the column j of $T^{(k)}$ by $T_j^{(k)}$.

Pivot strategy: push the column of $T^{(k)}$ with maximal 2-norm to the front so that after the change

$$|\alpha_k| = |r_{kk}| = \| T_1^{(k)} \|_2 = \max_j \| T_j^{(k)} \|_2 =: \| T^{(k)} \|_\square$$

Finally: $A\Pi = QR$ with permutation matrix Π

Rank decision

rank$(A) = n < q$:

$$Q^T A \Pi \overset{exact}{=} \begin{pmatrix} R & S \\ 0 & 0 \end{pmatrix} \overset{eps}{=} \begin{pmatrix} R & S \\ 0 & T^{(n+1)} \end{pmatrix}, \quad T^{(n+1)} \text{ “small”}$$

Note: The exact rank n cannot be computed since by rounding errors $\| T^{(n+1)} \|_\square = |r_{n+1,n+1}| \neq 0$

Idea: Stop the iteration as soon as

$$|r_{k+1,k+1}| < \delta |r_{11}|, \quad \delta : \text{ relative precision of } A$$

Definition: numerical rank n

$$|r_{n+1,n+1}| < \delta |r_{11}| \leq |r_{nn}|$$

choice of $\delta \quad \Longrightarrow \quad$ decision for n

Subcondition number

Deuflhard/Sautter, Lin. Alg. Appl. 29 (1980):

Definition: subconditon $\mathrm{sc}(A) = |r_{11}|/|r_{qq}|$

Properties:

- $\mathrm{sc}(A) \geq 1$
- $\mathrm{sc}(\alpha A) = \mathrm{sc}(A)$
- $A \neq 0$ singular $\iff$ $\mathrm{sc}(A) = \infty$
- $\mathrm{sc}(A) \leq \kappa_2(A)$ (hence the name)

A is called *almost singular* if

$$\delta \mathrm{sc}(A) \geq 1 \quad \text{or equivalently} \quad \delta |r_{11}| \geq |r_{qq}|$$

Scaling

D_{row}, D_{col}: diagonal matrices

We consider the scaled least squares problem

$$\|D_{\text{row}}^{-1}(AD_{\text{col}}D_{\text{col}}^{-1}p - z)\|_2 = \|\bar{A}\bar{p} - \bar{z}\|_2 \,\text{min}$$

where $\bar{A} := D_{\text{row}}^{-1}AD_{\text{col}}, \quad \bar{p} := D_{\text{col}}^{-1}p, \quad \bar{z} := D_{\text{row}}^{-1}z$

- row (measurement) scaling: $D_{\text{row}} := \text{diag}(\delta z_1, \ldots, \delta z_m)$
  e.g. relative error concept: $\delta z_j = \varepsilon |z_j|$
- column (parameter) scaling: $D_{\text{col}} = \text{diag}(p_{1,\text{scal}}, \ldots, p_{q,\text{scal}})$
  (to account for different physical units)

in general: $\text{sc}(\bar{A}) \neq \text{sc}(A)$ and $\text{rank}(\bar{A}) \neq \text{rank}(A)$

Generalized Inverse

(1) $A \in \mathbb{R}^{q \times q}$, $\operatorname{rank}(A) = q$ (A is regular)

$\Rightarrow Ap = z$ has unique solution $p = A^{-1}z$ where $A^{-1}A = AA^{-1} = I_q$

(2) $A \in \mathbb{R}^{m \times q}$, $\operatorname{rank}(A) = n \leq q < m$

$\Rightarrow$ usually $Ap = z$ has no solution, but there is a minimizer

$$\|Ap - z\|_2 = \min$$

(2a) $n = q$: $\|Ap - z\|_2 = \min \Leftrightarrow A^T Ap = A^T z$ (normal eq.)
uniquely solvable since $A^T A$ is nonsingular:

$$p = \underbrace{(A^T A)^{-1} A^T}_{=A^+} z$$

(2b) $n < q$: write $p = A^+ z$ A^+: generalized/pseudo inverse

Penrose axioms

Let $A \in \mathbb{R}^{m \times q}$, $\text{rank}(A) = n \leq q \leq m$. Then A^+ is uniquely defined by

(i) $(A^+A)^T = A^+A$

(ii) $(AA^+)^T = AA^+$

(iii) $A^+AA^+ = A^+$

(iv) $AA^+A = A$

One can show: $p^* = A^+z$ is the shortest solution, i.e.

$$\|p^*\|_2 \leq \|p\|_2 \; \forall p \in L(z)$$

with

$$L(z) = \{p \in \mathbb{R}^q : \|z - Ap\|_2 = \min\} = p^* + \mathcal{N}(A)$$

Computation of shortest solution

$$Q^T A \Pi = \begin{pmatrix} R & S \\ 0 & 0 \end{pmatrix}, \quad A \in \mathbb{R}^{m \times q}$$

$\text{rank}(A) = n$, $R \in \mathbb{R}^{n \times n}$ upper triangular, $S \in \mathbb{R}^{n \times (q-n)}$

$$\Pi^T p = \begin{pmatrix} p_1 \\ p_2 \end{pmatrix}, \quad p_1 \in \mathbb{R}^n, \; p_2 \in \mathbb{R}^{q-n}$$

$$Q^T z = \begin{pmatrix} z_1 \\ z_2 \end{pmatrix}, \quad z_1 \in \mathbb{R}^n, \; z_2 \in \mathbb{R}^{m-n}$$

$$\Rightarrow \|z - Ap\|_2^2 = \|Q^T z - Q^T Ap\|_2^2 = \|z_1 - Rp_1 - Sp_2\|_2^2 + \|z_2\|_2^2$$

$$\|z - Ap\|_2^2 = \min \Leftrightarrow z_1 - Rp_1 - Sp_2 = 0 \Leftrightarrow p_1 = R^{-1}(z_1 - Sp_2)$$

Computation of shortest solution

$V := R^{-1}S,\ u := R^{-1}z_1$

$$\begin{aligned}\|p\|^2 = \|\Pi^T p\|^2 &= \|p_1\|^2 + \|p_2\|^2 = \|u - Vp_2\|^2 + \|p_2\|^2 \\ &= \|u\|^2 - 2\langle u, Vp_2\rangle + \langle Vp_2, Vp_2\rangle + \langle p_2, p_2\rangle \\ &= \|u\|^2 + \langle p_2, (I + V^TV)p_2 - 2V^Tu\rangle \\ &=: \phi(p_2)\end{aligned}$$

$$\phi(p_2) = \min \text{ if } \phi'(p_2) = 2(I + V^TV)p_2 - 2V^Tu = 0$$

$\Leftrightarrow (I + V^TV)p_2 = V^Tu$ (solution by Cholesky decomposition)

$$\phi''(p_2) = 2(I + V^TV) \text{ spd } \Rightarrow p_2 \text{ is minimum}$$

Note: if $n < q/2$, a faster algorithm exists

Outline

Problem formulation

Data:

$$(t_i, z_i), \quad t_i \in \mathbb{R},\ z_i \in \mathbb{R}^d, \quad i = 1, \ldots, M$$

Parameters: $p = (p_1, \ldots, p_q)^T$

Model:

$$y(t, p) : \mathbb{R} \times \mathbb{R}^q \mapsto \mathbb{R}^d \text{ nonlinear w.r.t. } p$$

Measurement tolerances:

$$\delta z_i \in \mathbb{R}^d, \quad D_i = \text{diag}(\delta z_i)$$

Define

$$F(p) = \begin{pmatrix} D_1^{-1}(z_1 - y(t_1, p)) \\ \vdots \\ D_M^{-1}(z_M - y(t_M, p)) \end{pmatrix} \in \mathbb{R}^m, \quad m \leq M \cdot d$$

Problem formulation

Nonlinear least-squares problem:

$$\sum_{i=1}^{M} \|D_i^{-1}(z_i - y(t_i, p))\|_2^2 = F(p)^T F(p) = \|F(p)\|_2^2 = \min$$

Nonlinear least-squares problem

Given a mapping $F : D \subseteq \mathbb{R}^q \to \mathbb{R}^m$ with $m > q$, find a solution $p^* \in D$ such that

$$\|F(p^*)\|_2^2 = \min_{p \in D} \|F(p)\|_2^2.$$

The nonlinear least-squares problem is called **compatible**, if

$$F(p^*) = 0.$$

Supplement: Multivariable calculus

Consider $p = (p_1, \ldots, p_q)^T \in \mathbb{R}^q$ and the scalar function $g(p) : \mathbb{R}^q \to \mathbb{R}$.
The derivative of g is the gradient (a column vector)

$$g'(p) := \nabla g(p) = \left(\frac{\partial g}{\partial p_1}, \cdots, \frac{\partial g}{\partial p_q} \right)^T$$

Consider the vector $F(p) = (F_1(p), \ldots, F_m(p))^T : \mathbb{R}^q \to \mathbb{R}^m$.
Its derivative is the Jacobian matrix

$$F'(p) = \begin{pmatrix} \frac{\partial}{\partial p_1} F_1(p) & \cdots & \frac{\partial}{\partial p_q} F_1(p) \\ \vdots & & \vdots \\ \frac{\partial}{\partial p_1} F_m(p) & \cdots & \frac{\partial}{\partial p_q} F_m(p) \end{pmatrix} \in \mathbb{R}^{m \times q}$$

Problem formulation

Minimization problem:

$$g(p) := \|F(p)\|_2^2 = \min$$

sufficient condition on p^* to be a local minimum:

$$g'(p^*) = 2F'(p^*)^T F(p^*) = 0 \quad \text{and} \quad g''(p^*) \text{ positive definite}$$

Find by Newton iteration a zero of the function $G : \mathbb{R}^q \to \mathbb{R}^q$

$$G(p) := F'(p)^T F(p)$$

(system of q nonlinear equations)

Newton iteration

Idea: approximation of G by a "simpler" function, namely its Taylor series expansion around a starting guess p^0

$$\begin{aligned} G(p) &= G(p^0) + G'(p^0)(p - p^0) + o(\|p - p^0\|) \quad \text{for} \quad p \to p^0 \\ &\doteq G(p^0) + G'(p^0)(p - p_0) =: \bar{G}(p) \end{aligned}$$

Zero p^1 of the linear substitute map $\bar{G}$:

$$G'(p^0)\Delta p^0 = -G(p^0), \quad p^1 = p^0 + \Delta p^0$$

Newton iteration:

$$G'(p^k)\Delta p^k = -G(p^k), \quad p^{k+1} = p^k + \Delta p^k$$

system of nonlinear equations $\Longrightarrow$ sequence of linear systems

Gauss-Newton iteration

$$G'(p) = F'(p)^T F'(p) + F''(p)^T F(p)$$

iteration in terms of $F(p)$:

$$\left(F'(p^k)^T F'(p^k) + F''(p^k)^T F(p^k)\right) \Delta p^k = -F'(p^k)^T F(p^k)$$

- compatible case:
  $F(p^*) = 0 \quad \Rightarrow \quad G'(p^*) = F'(p^*)^T F'(p^*)$
- almost compatible case: $F(p^*)$ is "small" $\Rightarrow$ save effort for evaluating $F''(p)$

$\Rightarrow$ Gauss-Newton iteration:

$$F'(p^k)^T F'(p^k) \Delta p^k = -F'(p^k)^T F(p^k), \quad p^{k+1} = p^k + \Delta p^k$$

Gauss-Newton iteration

$$F'(p^k)^T F'(p^k)\Delta p^k = -F'(p^k)^T F(p^k), \quad p^{k+1} = p^k + \Delta p^k$$

corresponds to normal equation for the linear least squares problem

$$\|F'(p^k)\Delta p^k + F(p^k)\|_2 = \min$$

formal solution

$$\Delta p^k = -F'(p^k)^+ F(p^k)$$

in case of convergence: $F'(p^*)^+ F(p^*) = 0$

if $\operatorname{rank}(F'(p)) = q \;\Rightarrow\; F'(p)^+ = \left(F'(p)^T F'(p)\right)^{-1} F'(p)^T$

Classification

- local GN methods: require sufficiently good starting guess p^0
- global GN methods: can handle bad guesses as well
- $m = q$: guaranteed local convergence
- $m > q$: convergence only for a small subclass of nonlinear LSQ problems, so-called adequate problems

Assumptions

We want to show that the Gauss-Newton method converges locally. This can only be expected under some assumptions on the function $F(p)$ and the starting guess p^0.

- $F : D \to \mathbb{R}^m$ continuously differentiable, $D \subseteq \mathbb{R}^q$ open and convex
- $F'(p)$ possibly rank-deficient Jacobian
- F' satisfies a particular Lipschitz condition (i.e. F' behaves "nicely") with Lipschitz constant $0 \leq \omega < \infty$

$$\|F'(\hat{p})^+(F'(\tilde{p}) - F'(p))(\tilde{p} - p)\|_2 \leq \omega \|\tilde{p} - p\|_2^2 \quad \text{(i)}$$

for all collinear $p, \tilde{p}, \hat{p} \in D$ (i.e. they lie on a single straight line) and $\tilde{p} - p \in \mathcal{R}(F'(p)^+)$

Assumptions

- compatibility condition (model and data must somehow fit each other)

$$\begin{aligned} \|F'(\tilde{p})^+(I - F'(p)F'(p)^+)F(p)\| &\leq \kappa(p)\|\tilde{p} - p\| \quad \forall p, \tilde{p} \in D \\ \kappa(p) &\leq \bar{\kappa} < 1 \quad \forall p \in D \quad (ii) \end{aligned}$$

- the initial update Δp^0 must be small enough and a ball $B_\rho(p^0)$ with radius ρ around p^0 is contained in D

$$\|\Delta p^0\| \leq \alpha \quad \text{and} \quad B_\rho(p^0) \subset D \qquad \text{(iii)}$$

$$h := \alpha\omega < 2(1 - \bar{\kappa}), \quad \rho := \alpha/(1 - \bar{\kappa} - h/2) \qquad \text{(iv)}$$

Convergence result

Under the above assumptions, the sequence $\{p^k\}$ of Gauss-Newton iterates is well-defined, remains in $B_\rho(p^0)$ and converges to some $p^* \in B_\rho(p^0)$ with

$$F'(p^*)^+ F(p^*) = 0.$$

The convergence rate can be estimated according to

$$\|p^{k+1} - p^k\|_2 \leq \frac{1}{2}\omega\|p^k - p^{k-1}\|_2^2 + \kappa(p^{k-1})\|p^k - p^{k-1}\|_2. \quad (*)$$

This result shows existence of a solution p^* in the sense that $F'(p^*)^+ F(p^*) = 0$ even in case of rank deficient Jacobian (rank may vary throughout D as long as $\omega < \infty$ and $\kappa < 1$).

Uniqueness? $\Longrightarrow$ requires full rank Jacobian

Compatibility

- linear least squares problems, $F(p) = Ap - z$:

$$F'(p) = (Ap - z)' = A = (A\tilde{p} - z)' = F'(\tilde{p}) \overset{(i)}{\Longrightarrow} \omega = 0$$

$$F'(\tilde{p})^+(I - F'(p)F'(p)^+) = F'(p)^+(I - F'(p)F'(p)^+) = 0$$

$$\overset{(ii)}{\Longrightarrow} \kappa(p) = \bar{\kappa} = 0 \overset{(*)}{\Rightarrow} \|p^2 - p^1\| \leq 0 \Rightarrow \Delta p^2 = 0 \Rightarrow p^1 = p^*$$

convergence within one iteration

- systems of nonlinear equations ($m = q$):

$$\text{if } F'(p) \text{ nonsingular } \Rightarrow \quad F'(p)^+ = F'(p)^{-1}$$

$$\Rightarrow \; I - F'(p)F'(p)^+ = 0 \overset{(*)}{\Rightarrow} \kappa(p) = 0$$

quadratic convergence

- compatible problems: $F(p^*) = 0 \Rightarrow \kappa(p^*) = 0$

Incompatibility factor

General nonlinear least-squares problems ($m > q$):

$$\lim_{k\to\infty} \frac{\|p^{k+1} - p^k\|}{\|p^k - p^{k-1}\|} \overset{(*)}{\leq} \lim_{k\to\infty} \left(\frac{\omega}{2}\|p^k - p^{k-1}\| + \kappa(p^{k-1})\right) = \kappa(p^*)$$

$\kappa(p^*)$ is the asymptotic convergence rate.

Adequate nonlinear least-squares problems are characterized by the condition

$$\kappa(p^*) < 1.$$

Summary:

The Gauss-Newton method converges locally linearly for adequate and locally quadratically for compatible nonlinear least-squares problems.

Convergence monitor/Termination criteria

Define

$\overline{\Delta p}^{k+1} := -F'(p^k)^+ F(p^{k+1})$ (simplified Gauss-Newton corrections)

Stop iteration as soon as linear convergence dominates the iteration process:

$$\|\Delta p^{k+1}\| \approx \|\overline{\Delta p}^{k+1}\| \leq \text{XTOL}$$

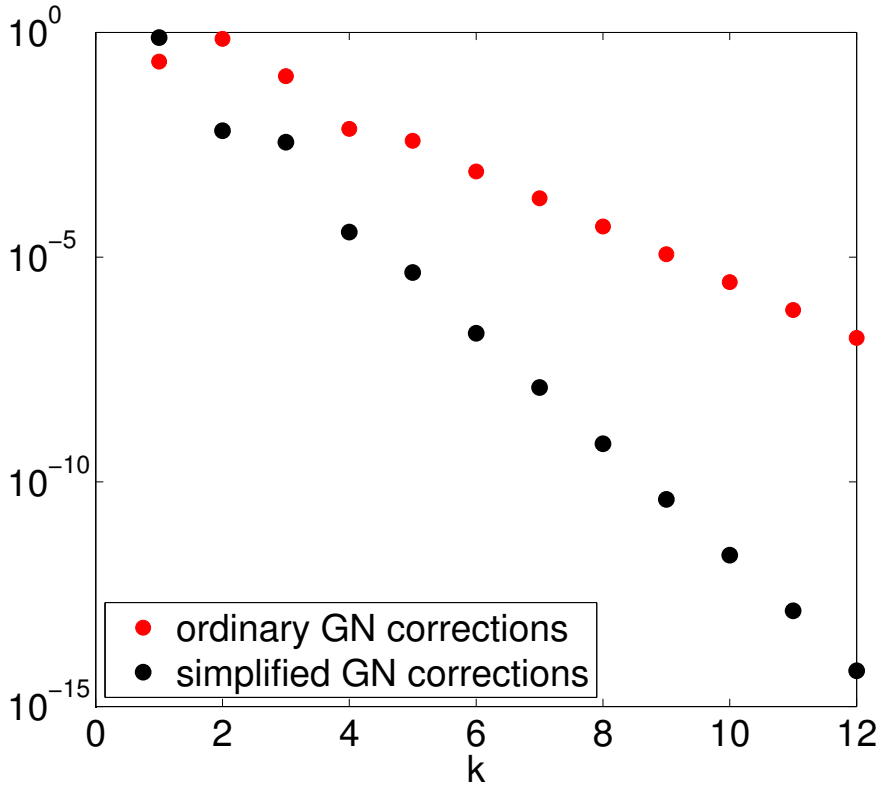
For incompatible problems in the convergent phase, it holds:

$$\|\Delta p^{k+1}\| \approx \kappa \|\Delta p^k\|, \quad \|\overline{\Delta p}^{k+1}\| \approx \kappa^2 \|\overline{\Delta p}^{k}\|$$

This provides an estimate for $\kappa(p^*)$:

$$\kappa(p^*) \approx \frac{\|\Delta p^{k+1}\|}{\|\Delta p^k\|}$$

Convergence monitor



Typical *scissor* between ordinary and simplified Gauss-Newton corrections in the linear convergence phase

Summary

Whenever the ordinary Gauss-Newton method with *full rank* Jacobian fails to converge, then one should rather improve the model (change equations or initial guess p^0) than just turn to a different iterative solver. In the convergent, but *rank-deficient* case, one should try to get more information by acquiring better data.

Identifiability

interpretation of result in case of **rank deficient** Jacobian:

Let p^* be the final iterate and $\text{rank}(F'(p^*)) = n < q$.

QR factorization with column pivoting and rank decision

$$Q^T F'(p^*)\Pi = R$$

$\Rightarrow$ reordering of parameters to $(p^*_{\pi_1}, p^*_{\pi_2}, \ldots, p^*_{\pi_n}, \ldots, p^*_{\pi_q})$ in such a way that

$$|r_{11}| \geq |r_{22}| \geq \cdots \geq |r_{nn}|$$

$\rightarrow$ parameters $(p^*_{\pi_1}, p^*_{\pi_2}, \ldots, p^*_{\pi_n})$ can actually be estimated from the current dataset

Parameter transformations

p: constrained parameters
u: unconstrained parameters
ϕ: transformation function

$$p = \phi(u), \quad F(p) = F(\phi(u)) =: \tilde{F}(u), \quad \tilde{F}'(u) = F'(p)\phi'(u)$$

constraint	transformation $\phi(u)$
$p > 0$	$p = \exp(u)$
$A \leq p \leq B$	$p = A + \frac{B-A}{2}(1 + \sin u)$
$p \leq C$	$p = C + \left(1 - \sqrt{1+u^2}\right)$
$p \geq C$	$p = C - \left(1 - \sqrt{1+u^2}\right)$

Levenberg-Marquardt method

Idea: local linearization only reasonable in a small neighborhood of p^k

$$\|F(p^k) + F'(p^k)\Delta p^k\|_2^2 + \rho\|\Delta p^k\|_2^2 = \min$$

$$(F'(p^k)^T F'(p^k) + \rho I_q)\Delta p^k = -F'(p^k)^T F(p^k)$$

- $\rho \to 0$: Levenberg-Marquardt $\to$ Gauss-Newton
- $\rho > 0$: $(F'(p^k)^T F'(p^k) + \rho I_q)$ always regular $\to$ masking of uniqueness of a given solution, incorrect interpretation of numerical results; solutions are often not minima of $g(p)$ but merely saddle points with $g'(p) = 0$

Scaling

- **row (measurement) scaling**

$$\tilde{F}(p) := D_{\text{row}}^{-1} F(p), \quad D_{\text{row}} = \text{diag}(\delta z_1, \ldots, \delta z_M)$$

$$\|\tilde{F}'(p^k)\Delta p^k + \tilde{F}(p^k)\| = \min \Leftrightarrow \|D_{\text{row}}^{-1}(F'(p^k)\Delta p^k + F(p^k))\| = \min$$

Different scaling leads to a different solution!

- **column (parameter) scaling**

$$p = D_{\text{col}}\tilde{p}, \quad H(\tilde{p}) = H(D_{\text{col}}^{-1} p) := F(p), \quad H'(\tilde{p}) = F'(p) D_{\text{col}}$$

assume $p^k = D_{\text{col}}\tilde{p}^k$:

$$\begin{aligned} \Delta\tilde{p}^k &= -H'(\tilde{p}^k)^+ H(\tilde{p}^k) = -(F'(p^k)D_{\text{col}})^+ F(p^k) \\ &\overset{F'(p^k) \text{ full rank}}{=} -D_{\text{col}}^{-1} F'(p^k)^+ F(p^k) = D_{\text{col}}^{-1}\Delta p^k \end{aligned}$$

No invariance in the rank deficient case!

Globalization

Global (or damped) Gauss-Newton method:

$$F'(p^k)\Delta p^k = -F(p^k), \quad p^{k+1} = p^k + \lambda_k \Delta p^k,\ 0 < \lambda_k \leq 1$$

How to choose λ_k?

Remember: we want to solve $F'(p)^+ F(p) = 0$

Idea: choose λ_k in such a way that $\|F'(p^k)^+ F(p^k)\|$ decays along the iterates (theory of level function, Gauss-Newton path)

damping factors λ_k are determined via a predictor-corrector strategy such that they satisfy the natural monotonicity test

$$\frac{\|\overline{\Delta p}^{k+1}(\lambda_k)\|}{\|\Delta p^k\|} < 1$$

divergence criterion: $\lambda_k < \lambda_{\min} \ll 1$

Codes

- NLSQ-ERR: **ERR**or-oriented Newton method for **N**onlinear **L**east-**SQ**uares
- NLSCON: solver for **N**onlinear **L**east-**S**quares with **CON**straints

Outline

Problem formulation

d species, q (unknown) parameters

model: $y' = f(y,p),\ y(0) = y_0$ with $y \in \mathbb{R}^d,\ p \in \mathbb{R}^q$

set of experimental data:
$(t_1, z_1, \delta z_1), \ldots, (t_M, z_M, \delta z_M),\ t_j \in \mathbb{R},\ z_j \in \mathbb{R}^d$

with measurement tolerances $\delta z_j \in \mathbb{R}^d,\ D_i = \mathrm{diag}(\delta z_i)$

$$F(p) := \begin{pmatrix} D_1^{-1}(y(t_1,p) - z_1) \\ \vdots \\ D_M^{-1}(y(t_M,p) - z_M) \end{pmatrix},\ F'(p) = \begin{pmatrix} D_1^{-1} y_p(t_1,p) \\ \vdots \\ D_M^{-1} y_p(t_M,p) \end{pmatrix}$$

Gauss-Newton method:

$$\Delta p^k = -F'(p^k)^+ F(p^k), \quad p^{k+1} = p^k + \lambda_k \Delta p^k$$

Computation of $F(p)$

requires solution of $y' = f(y, p)$ at time points $t_1, \ldots, t_M$

Problem: time points chosen by adaptive step-size control generally do not agree with $t_1, \ldots, t_M$

Solution: use integrator with dense output

Idea: construction of interpolation polynomial based on the available solution such that accuracy is not destroyed

Example: LIMEX has a dense output based on Hermite interpolation

Sensitivity analysis

- sensitivity matrix at time t:

$$S(t) := \frac{\partial y}{\partial p}(t) := y_p(t) \ \in \mathbb{R}^{d \times q}$$

$$F'(p) = \begin{pmatrix} D_1^{-1} S(t_1) \\ \vdots \\ D_M^{-1} S(t_M) \end{pmatrix}$$

- **scaled** sensitivity matrix (necessary whenever species or parameters cover a broad range of physical units):

$$\bar{S}_{ij}(t) := \frac{\partial y_i}{\partial p_j}(t) \cdot \frac{p_{\text{scal}}}{y_{\text{scal}}}$$

$p_{\text{scal}}, y_{\text{scal}}$: user specified scaling values

Computation of sensitivities

Differentiate $y' = f(y, p)$ with respect to p to obtain the variational equation

$$y'_p = f_y(y, p) \cdot y_p + f_p(y, p), \quad y_p(0) = 0$$

Compact notation:

$$S' = f_y \cdot S + f_p \,, \quad S = [s_1, \ldots, s_q], \; s_k \in \mathbb{R}^d$$

In practice: **solve combined system**

$$\begin{pmatrix} y' \\ s'_1 \\ \vdots \\ s'_q \end{pmatrix} = \begin{pmatrix} f \\ f_y \cdot s_1 + f_{p_1} \\ \vdots \\ f_y \cdot s_q + f_{p_q} \end{pmatrix} \in \mathbb{R}^{(q+1)d}$$

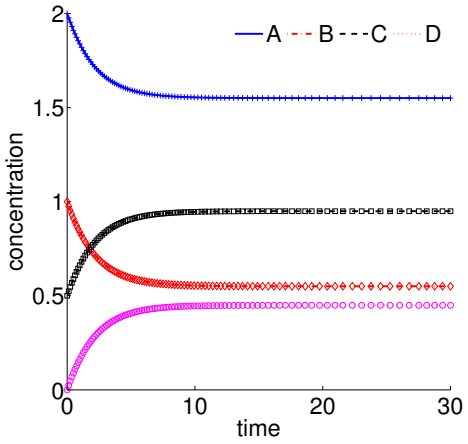
can be solved efficiently by LIMEX (inexact Jacobian, decoupling)

Example

k	$\|F(p^k)\|$	$\|\Delta p^k\|$	λ_k	rank
0	4.81e-02	3.55e-01		2
1	1.11e-01	4.57e-01	1.000	2
	2.55e-02	1.75e-01	0.389	
2	1.04e-02	4.22e-02	1.000	2
3	9.16e-04	1.90e-03	1.000	2
4	8.00e-06	1.99e-05	1.000	2
5	23.21e-07	1.54e-09	1.000	2

$p^* = (k_1^*, k_2^*) = (0.10000001, 0.19999997) \quad \Rightarrow \quad k_{21}^* = 1.9999994$

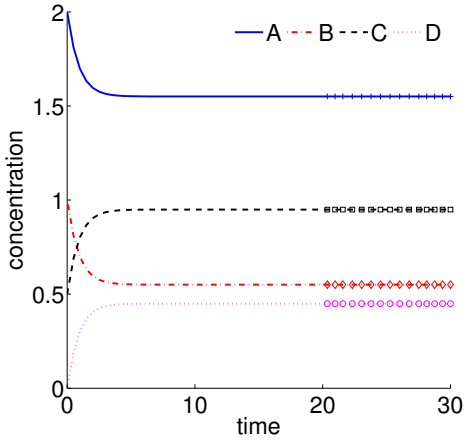
$\kappa = 0.008, \quad \text{sc}(F'(p^*)) = 6.3$



k	$\|F(p^k)\|$	$\|\Delta p^k\|$	λ_k	rank
0	3.85e-02	2.96e-02		1
1	2.40e-03	1.85e-03	1.000	1
2	1.11e-05	7.67e-06	1.000	1
3	6.67e-06	6.75e-11	1.000	1

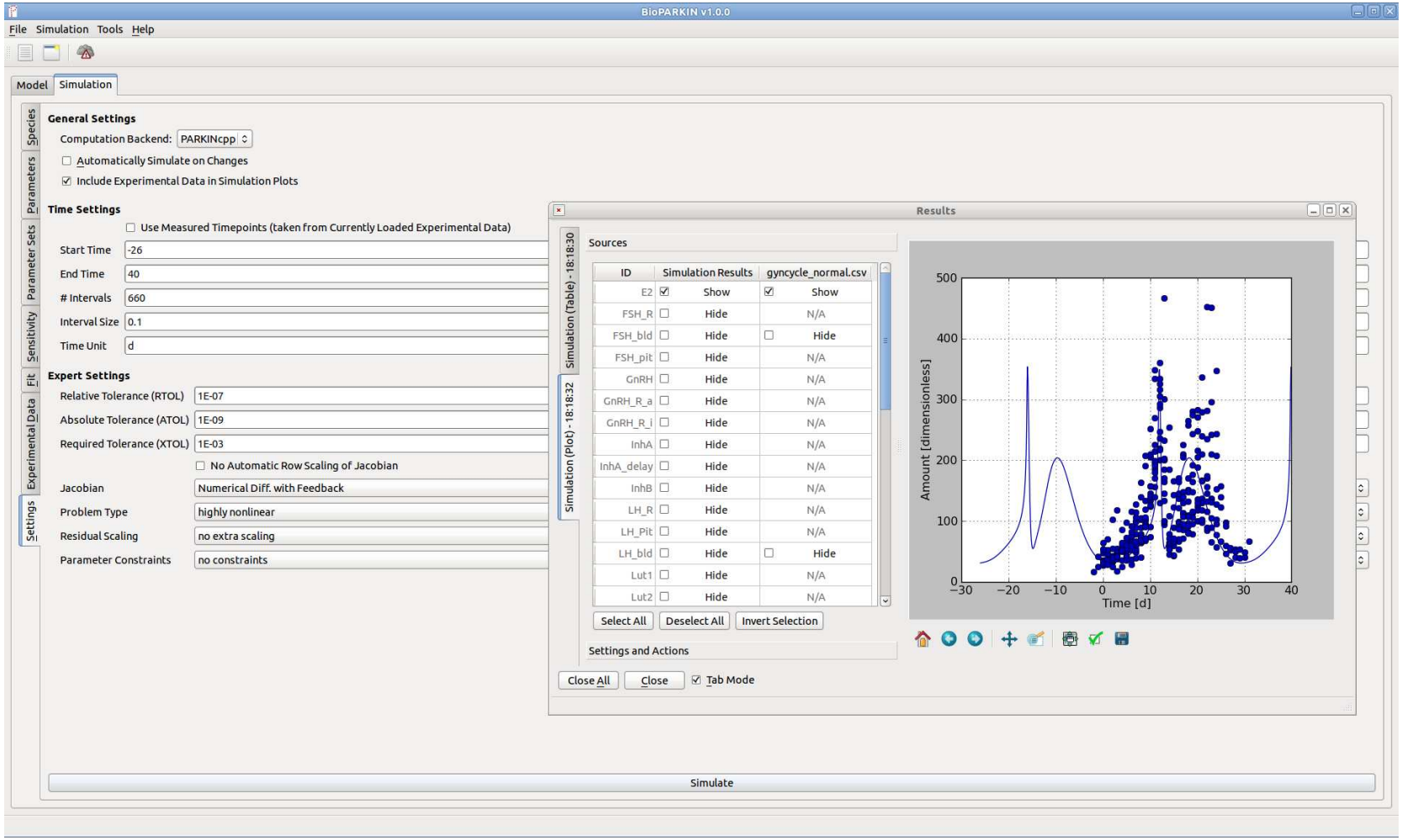
$p^{**} = (k_1^{**}, k_2^{**}) = (0.24155702, 0.48312906) \quad \Rightarrow \quad k_{21}^{**} = 2.0000622$

$\kappa = 0.004, \quad \text{sc}(F'(p^*)) = 4.5 \times 10^6$



BioPARKIN

integrated software environment for simulation and parameter identification, `http://bioparkin.zib.de`



BioPARKIN

- SBML import and export
- deterministic simulation methods (LIMEX; others will follow)
- sensitivity analysis based on variational equations
- parameter estimation via Gauss-Newton methods (tunable, possible scaling of species and parameters)
- output of information on identifiability of parameters
- time-shifting of data and concatination of IVPs for multi-experiment simulations

Thank you for your attention!

Contact:
Zuse Institute Berlin
Computational Systems Biology Group
http://www.zib.de/en/numerik/csb.html