Inverse Problems in Systems Biology

Heinz W. Engl
Johann Radon Institute for Computational and Applied Mathematics
Austrian Academy of Sciences
www.ricam.oeaw.ac.at

and

University of Vienna

see: H.W. Engl, C. Flamm, P. Kügler, J. Lu, S. Müller and P. Schuster,
Inverse problems in systems biology, Inverse Problems 25 (2009)
Systems biology is a relatively young biological discipline that claims to consider cells and organisms as entities in a holistic way. At the same time, it focuses on the interplay of components from the molecular to the systemic level. Quantitative measurements and recordings of biological processes are merged with advanced mathematical methods to yield predictive theoretical, mostly computational, models of biological systems.

High complexity:

Metabolism of a cell: several thousands of catalyzed biochemical reactions resulting in molecular concentrations of a large number of substrates, products and enzymes as functions of time. Data collection, data validation and model building are the great challenges of systems biology!

For example, the bacterial cell of the species *Escherichia coli* has 4,460 genes giving rise to roughly the same number of transcripts and translation products. The cell is robust against changes in the surrounding and collects information on the environment. The chemical reactions within the cell form a highly organized network controlled by a genetic regulatory system.
Human body: approximately 20,000 genes and 200 different cell types.

**Major goal of systems biology:**

To provide an understanding of properties and behavior of cells or organisms emerging as consequence of the interaction of large numbers of molecules, which organize themselves into highly intricate reaction networks that span various levels of cellular or organismal complexity. The number of nodes in metabolic networks amounts to several thousand molecules.

**Outlook:**

The efforts in systems biology to build accurate and predictive *in silico* models of biological systems and the rapid development of experimental techniques pave the ground for *synthetic biology* which seeks, on the one hand, to engineer existing biological systems and, on the other hand, to design *de novo* synthetic networks from well characterized biological parts or non-natural components exhibiting a desired qualitative dynamic behavior.
The goal of designing and constructing biological devices gives rise, from a mathematical point of view, to *multi-level inverse problems*:

(i) at the bottom level the individual biological entities must be characterized in terms of parameters,

(ii) at the next level of assembling the biological device, parameters influencing the coupling of the subsystems must be identified,

(iii) finally, putting the biological device into the cellular context will require parameter fine tuning to account for additional interactions between the device and the complex cellular matrix.
Nobel laureate Sidney Brenner in a talk at Scripps Institute entitled ‘Systems Biology and Inverse Problems: Much Ado about Nothing’:

“The program of systems biology is doomed”

Reasons given: lack of data, insensitivity of some phenomena to changes within the system (→ homeostasis ), within leads to ill-posedness of the inverse problem, lack of evolutionary perspective.

P. Sabatier: “well-posed questions for ill-posed problems”
Two levels of inverse problems:

- quantitative inference of structure/parameters of networks
- qualitative inverse problems that aim at reverse engineering of bifurcation patterns and other types of desired qualitative behavior.

Typical questions for qualitative inverse problems

Which are the parameters that maximize or minimize a stable closed orbit?

How can the bifurcation pattern of a given reaction network be influenced by changing certain parameters? A practical implication would be, for example, the determination of parameter sets that arrest the cell cycle of given cell types in a certain phase and prohibit malignant proliferation thereby.

How can one adjust the circadian rhythm such that it responds fast and smoothly to a pace maker reset in order to avoid negative consequences of jet lag?
Quantitative inverse problems:

Metabolic networks, which consist of enzyme catalyzed reactions that transform organic molecules for the purpose of acquiring energy and constructing complex compounds used in cellular functioning, are among the best studied networks in biology. Although the topology of these networks and the exact stoichiometry (i.e. the quantitative relationship between reactants and products) of the individual chemical reactions are well known, only little information is available on the functional form and the kinetic parameters of the enzymatic rate laws.
A typical feature of metabolic networks, which are among the best characterized networks in systems biology, is the existence of network hubs, i.e., a few nodes which are connected to a large number of other nodes in the network, while most of the nodes in the network possess only a few connections to other nodes. The existence of hubs is directly reflected in the sparsity of the corresponding stoichiometric matrix:

\[
\begin{align*}
\frac{d}{dt}x &= S \cdot v = \frac{d}{dt}x = \\
\begin{pmatrix}
    x_A \\
x_B \\
x_C \\
x_E \\
x_{EA} \\
x_{EB} \\
x_{EAB}
\end{pmatrix} &= \\
\begin{pmatrix}
    +1 & 0 & 0 & -1 & 0 & 0 & -1 & 0 \\
    0 & +1 & 0 & 0 & -1 & -1 & 0 & 0 \\
    0 & 0 & -1 & 0 & 0 & 0 & 0 & +1 \\
    0 & 0 & 0 & -1 & -1 & 0 & 0 & +1 \\
    0 & 0 & 0 & +1 & 0 & -1 & 0 & 0 \\
    0 & 0 & 0 & 0 & +1 & 0 & -1 & 0 \\
    0 & 0 & 0 & 0 & 0 & +1 & +1 & -1
\end{pmatrix} \cdot \\
\begin{pmatrix}
    v_1(x_A, q) \\
v_2(x_B, q) \\
v_3(x_C, q) \\
v_4(x_A, x_E, x_{EA}, q) \\
v_5(x_B, x_E, x_{EB}, q) \\
v_6(x_B, x_{EA}, x_{EAB}, q) \\
v_7(x_A, x_{EB}, x_{EAB}, q) \\
v_8(x_{EAB}, q)
\end{pmatrix}
\end{align*}
\]

Thus, the stoichiometric matrix \( S \) is a linear transformation between the reaction rate vector \( v \) and the vector of changes in species concentrations \( x \). In many cases entries in the reaction rate vector \( v \) cannot be measured directly and hence must be inferred indirectly from biological data.
First step: ODE models:

\[
\frac{dx}{dt} = Sv(x, q). \quad (1)
\]

with stoichiometric matrix \(S\) and reaction rates \(v\).

Specific forms of reaction rates based assumptions on underlying kinetics:

- mass action (derived from basic principles):

  rate of the \(j^{th}\) reaction is given by

  \[
v_j = k_j \prod_{i=1}^{n} x_{i}^{g_{ji}}, \quad j = 1, \ldots, l, \quad (2)
  \]

  where \(k_j \in \mathbb{R}^+\) is the rate constant and \(g_{ji} \in \mathbb{N}_0, i = 1, \ldots, n,\) are the kinetic orders which correspond to the numbers of molecules consumed by this reaction.
• composite laws combing (possibly many) elementary step. Most widely used:

**irreversible Michaelis-Menten** mechanism: the corresponding rate law amounts to

\[ v_j = V_j \frac{x_i}{K^M_j + x_i}, \]  \hspace{1cm} (3)

where \( V_j \) is the limiting rate, \( K^M_j \) the Michaelis constant, and \( x_i \) the concentration of the substrate converted by the reaction.

Also, gene regulatory networks are usually modelled via ODEs:

The fundamental processes of regulator binding, transcription, translation, and degradation can be modeled at various levels of detail. Traditionally, ODEs are formulated for mRNAs and proteins; other molecular species such as polymerases, nucleotides, ribosomes, and amino acids are supposed to abound, and hence their concentrations need not enter the models.
Assume there are \( n \) genes, and let the mRNA and protein concentrations be \( x_{mi} \) and \( x_{pi} \), \( i = 1, \ldots, n \). Then, the corresponding ODEs have the form

\[
\frac{dx_{mi}}{dt} = k_{ts}^{i} f_{i}(x_{1}^{p}, \ldots, x_{n}^{p}) - d_{mi}^{m} x_{mi}^{m}, \quad (4a)
\]
\[
\frac{dx_{pi}}{dt} = k_{tl}^{i} x_{mi}^{m} - d_{pi}^{p} x_{pi}, \quad (4b)
\]

where \( k_{ts}^{i} \), \( k_{tl}^{i} \) are transcription and translation constants, respectively, and \( d_{mi}^{m} \), \( d_{pi}^{p} \) are degradation constants.

The nonlinearity is hidden in the regulation functions \( f_{i}(x_{1}^{p}, \ldots, x_{n}^{p}) \), which determine how one gene is activated or inhibited by the other genes.
The most common regulation functions are:

(i) customized functions,

(ii) sigmoid functions,

(iii) step functions.

In principle, regulator binding can be modeled using mass-action kinetics and the quasi-steady-state-assumption for the gene-regulator complexes. The knowledge of both binding kinetics and regulation logic allows the derivation of a customized function, which determines the activation or inhibition of one gene by the other genes.
However, many genes have several regulatory sites (such as in the case of cooperative binding) and the resulting regulation functions may become very complicated. In these cases, it is common to use sigmoid functions to model the typical switching behavior in gene regulation.

For example, let the $i^{\text{th}}$ gene be activated by the $j^{\text{th}}$ protein in cooperative manner.

Then, the regulation function is usually approximated by

$$ f_i(x_j^p) = \sigma^+(x_j^p, K_i, n_i) = \frac{(x_j^p)^{n_i}}{(x_j^p)^{n_i} + (K_i)^{n_i}}, \quad (5) $$

where $K_i$ is the concentration of half-activation (determining the threshold of the switch) and $n_i$ is the Hill coefficient (determining the steepness of the switch). This formula can be regarded as purely empirical, the Hill coefficient need not be integer.
protein product from gene expression binds to a regulatory region on the DNA and represses transcription

corresponding reaction rates with rate constants $x_1, x_2, x_3 \ldots$ are

- transcription: $\text{DNA} \rightarrow \text{DNA} + \text{mRNA} \Rightarrow \text{reaction rate } v = x_1 \cdot \text{DNA}$
- translation: $\text{mRNA} \rightarrow \text{mRNA} + \text{protein} \Rightarrow \text{reaction rate } v = x_2 \cdot \text{mRNA}$
- binding: $\text{DNA} + \text{protein} \rightarrow \text{DNA/protein} \Rightarrow \text{reaction rate } v = x_3 \cdot \text{DNA} \cdot \text{protein}$
- ...
A concrete example: a three step biochemical pathway (reproduced from Moles et al. (2003))

Solid arrows represent mass flow, and dashed arrows represent regulation, where → stands for activation and ⊣ stands for inhibition. Three genes are producing mRNAs G1, G2, G3 and enzymes E1, E2, E3 to regulate the transformation of substrate S into product P via the intermediate metabolites M1, M2.
Description as an ODE system in 8 variables:

\[
\begin{align*}
\frac{dG_1}{dt} &= \frac{V_1}{1+\left(\frac{P}{K_{i1}}\right)^{n_{i1}} + \left(\frac{K_{a1}}{S}\right)^{n_{a1}}} - k_1 \cdot G_1 \\
\frac{dG_2}{dt} &= \frac{V_2}{1+\left(\frac{P}{K_{i2}}\right)^{n_{i2}} + \left(\frac{K_{a2}}{M_1}\right)^{n_{a2}}} - k_2 \cdot G_2 \\
\frac{dG_3}{dt} &= \frac{V_3}{1+\left(\frac{P}{K_{i3}}\right)^{n_{i3}} + \left(\frac{K_{a3}}{M_2}\right)^{n_{a3}}} - k_3 \cdot G_3 \\
\frac{dE_1}{dt} &= \frac{V_4 \cdot G_1}{K_4 + G_1} - k_4 \cdot E_1 \\
\frac{dE_2}{dt} &= \frac{V_5 \cdot G_2}{K_5 + G_2} - k_5 \cdot E_2 \\
\frac{dE_3}{dt} &= \frac{V_6 \cdot G_3}{K_6 + G_3} - k_6 \cdot E_3 \\
\frac{dM_1}{dt} &= \frac{k_{cat1} \cdot E_1 \cdot \frac{1}{K_{m1}} \cdot (S-M_1)}{1+\frac{S}{K_{m1}} + \frac{M_1}{K_{m2}}} - \frac{k_{cat2} \cdot E_2 \cdot \frac{1}{K_{m3}} \cdot (M_1-M_2)}{1+\frac{M_1}{K_{m3}} + \frac{M_2}{K_{m4}}} \\
\frac{dM_2}{dt} &= \frac{k_{cat2} \cdot E_2 \cdot \frac{1}{K_{m3}} \cdot (M_1-M_2)}{1+\frac{M_1}{K_{m3}} + \frac{M_2}{K_{m4}}} - \frac{k_{cat3} \cdot E_3 \cdot \frac{1}{K_{m5}} \cdot (M_2-P)}{1+\frac{M_2}{K_{m5}} + \frac{P}{K_{m6}}} 
\end{align*}
\]

The 36 parameters can be divided into several classes: transcription/translation rates $V$, equilibrium constants $K$, Hill coefficients $n$, degradation rates $k$, and catalytic constants $k_{cat}$. 
“Naive” approach: identify parameters (from time-course measurements) via, e.g., a least-squares fit

\( p \): set of parameters
\( F \) maps parameters to solutions of ODE system, \( y := F(p) \)
\( \tilde{y} \): measured solutions

\[ \rightarrow \| F(p) - \tilde{y} \|^2 \rightarrow \min . \]

(maybe with some constraints),

minimization via some iterative algorithm (global-local):
\( \rightarrow (p_k) \)
Results for "exact" data

![Graph showing the objective function and average parameter error over iterations](image)

- **Obj. fun.**
- **Avg. par. err.**
Results with 5% data noise
10% data noise

![Graph showing objective function and average parameter error over iterations with 10% data noise.](image)
Observation: While $\|F(p_k) - \tilde{y}\|$ decreases, the $(p_k)$ seem not to converge to anything. Small “data misfit” does not imply anything about accuracy in parameters.

Abstract formulation:

$F$: operator mapping a **given** set of parameters to the data

evaluating $F$ involves solving the ODE-system

Inverse problem of determining the parameters:

solve

$$F(\text{parameters}) = (\text{noisy}) \text{ data}.$$ 

This is a nonlinear (ill-posed) equation:
Functional analytic theory of nonlinear ill–posed problems

\[ F(x) = y_0, \]

where \( F : D(F) \subset X \to Y \) is a nonlinear operator between Hilbert spaces \( X \) and \( Y \); assume that

\( F \) is continuous and \( F \) is weakly (sequentially) closed, i.e., for any sequence \( \{x_n\} \subset D(F) \), weak convergence of \( x_n \) to \( x \) in \( X \) and weak convergence of \( F(x_n) \) to \( y \) in \( Y \) imply \( x \in D(F) \) and \( F(x) = y \).

\( F \): forward operator for an inverse problem, e.g.

- parameter–to–solution map for a PDE (→ parameter identification)
- maps domain to the far field in a scattering problem (→ inverse scattering)
Notion of a “solution”: “$x^*\,$—minimum—norm—least—squares solution $x^{\dagger}$”:

$$\|F(x^{\dagger}) - y_0\| = \min\{\|F(x) - y_0\|/x \in D(F)\}$$

and

$$\|x^{\dagger} - x^*\| = \min\{\|x - x^*\|/\|F(x) - y_0\| = \|F(x^{\dagger}) - y_0\|\};$$

need not exist, if it does: need not be unique!

Choice of $x^*$ crucial: Available a–priori information has to enter into the selection criterion.

**Proposition:** $F$ compact, $F(x^{\dagger}) = y_0, x^{\dagger} \in \text{int } D(F), F(x) = y$ uniquely solvable in a neighborhood of $y_0$, dim $X = \infty$. Then $F^{-1}$ is discontinuous in $y_0$.

Thus: Compactness and local injectivity $\Rightarrow$ ill–posedness (like in the linear case).
Tikhonov regularization

\[(\ast) \quad \|F(x) - y^\delta\|^2 + \alpha\|x - x^*\|^2 \to \min \text{ over } D(F)\]

\[\to x^\delta_\alpha \quad (\text{exists, need not be unique}).\]

– Stable for \(\alpha > 0\) (in a multi–valued sense)
– convergence to an \(x^*\)–minimum–norm solution if

\[\alpha = \alpha(\delta) \to 0 \quad \text{and} \quad \frac{\delta^2}{\alpha} \to 0\]

(Seidman-Vogel)

Convergence rates: see first talk
Identification of $q = q(u)$ in

$$-\text{div}(q(u)\text{grad } u) = f(x) \text{ in } \Omega \subseteq \mathbb{R}^n$$

$$u(x) = 0 \quad \text{on} \quad \partial \Omega$$

from (distributed or boundary) temperature measurements by Tikhonov regularization ("identification of a nonlinearity"):

the inverse problems of identifying $q = q(x)$ and $q = q(u)$ are markedly different;

– dimensionality of data vs. unknown

– main numerical problem: $q$ can be identified at best on the interval which is covered by $u = u(q)$, a–priori unknown.

Such "identifications of nonlinearities" appear in many fields, e.g.: chemical and biological networks (concentration–dependent reaction rates), steel production (temperature–dependent heat conductivities)

Convergence rate $O(\sqrt{\delta})$ under a source condition whose interpretation via the "Coarea Formula" requires (in addition to $q^\dagger - q^* \in H^4$) that $u(q^\dagger)$ has no isotherms of vanishing $(n - 1)$–dimensional Hausdorff measure.
Numerical results (P. Kögler):

The true parameter $q^\dagger$ and the noisy data $z_1$ at $t = 0.5$

$q_1$ identified from $z_1$ with $q^* = 4$ and $q^* = 3$
Many other “variational” regularization methods:

\[ \cdots + \alpha \| x - x^* \|^2 \]

is replaced by another (semi-)norm or a more general (usually convex) term. E.g. maximum entropy regularization:

\[ \| Tx - y^\delta \|^2 + \alpha E(x) \rightarrow \min . \]

where

\[ E(x) := \int_{\Omega} x(t) \log \frac{x(t)}{m(t)} dt. \]

Trick: use Nemitskii operator \( H \) such that \( E(Hv) = \| v \|^2 + c \),

\[ \rightarrow \| THv - y^\delta \|^2 + \alpha \| v \|^2 \]

apply theory of Tikhonov regularization for nonlinear problems (Engl-Landl). Alternative: use Bregman distance (Eggermont), recently also used e.g. for EM-method.
or TV-regularization:

$$\ldots + \alpha \int |\nabla x| dt \rightarrow \min.$$  

(Burger, Osher, Vese, ...)  
“sparsity enforcing” regularization

$$\ldots + \alpha \|x\|_p \quad (1 < p < 2)$$

(Ramlau, Teschke)

Different class of methods: “iterative regularization”

Penalty terms introduce _a-priori knowledge_ or _bias:_
Some reaction network “identified” (from the same data) using $L^2$ (left) and sparsity (right) penalty
Back to our initial systems biology problem:

results with an algorithm based on

– variational regularization with a choice of various penalty terms ($L^2$, sparsity ...)
– computing derivatives w.r.t. parameters (i.e., $F'(p)h$) via adjoint techniques
– combination of global and gradient-based local optimization

implemented in SBML-based framework SOSLIB (with S. Müller, J. Lu, C. Flamm, P. Schuster et al)
5% data noise
With regularization

Objective function, Average parameter error

- Obj. fun.
- Avg. par. err.
Tikhonov regularization:

\[ T x = y \]
\[ x_\alpha^\delta = (\alpha I + T^* T)^{-1} T^* y_\delta, \]

equivalent characterization:

\[ \| T x - y_\delta \|^2 + \alpha \| x \|^2 \rightarrow \text{min}. \]
Typical error behaviour as a function of $\alpha$:

$$\|x^\delta_\alpha - T^\dagger y\| \leq \|x_\alpha - T^\dagger y\| + \|x_\alpha - x^\delta_\alpha\|$$

regularization error $\rightarrow 0$ as $\alpha \rightarrow 0$ if $y \in D(T^\dagger)$, otherwise: $\rightarrow \infty$

$\alpha \rightarrow 0$ for fixed $\delta$
disadvantage of Tikhonov / variational regularization:

functional in general not convex, local minima

→ alternative: iterative regularization methods

**Iterative methods:**
Newton’s method for nonlinear well-posed problems:
fast local convergence
For ill-posed problems?

Linearization of $F(x) = y$ at a current iterate $x_k$:

$$F'(x_k)(x - x_k) = -\left(F(x_k) - y\right)$$  \hspace{1cm} (*)

$x_{k+1} \approx x$, where $x$ solves (*) should be a good approximation for a solution.

(*) is generally ill-posed

→ solve (*) by a regularization method for linear equations.
Tikhonov regularization leads to the
Levenberg-Marquardt method:
\[ x_{k+1} = x_k - (F'(x_k)^*F'(x_k) + \alpha_k I)^{-1}F'(x_k)^*(F(x_k) - y^\delta) \]
with \( \alpha_k \to 0 \) as \( k \to \infty \), \( \|y - y^\delta\| \leq \delta \).
Convergence for ill-posed problems: Hanke

Iteratively regularized Gauß–Newton method:
\[
x_{k+1} = x_k - (F'(x_k)^*F'(x_k) + \alpha_k I)^{-1}\left[ (F'(x_k)^*(F(x_k) - y^\delta) \\
+ \alpha_k(x_k - \zeta) \right]
\]
Convergence (rates): Bakushinskii, Hanke–Neubauer–Scherzer, Kaltenbacher

Landweber method:
\[ x_{k+1} = x_k + F'(x_k)^*(y^\delta - F(x_k)) \]
Convergence (rates): Hanke, Neubauer, Scherzer

Simple, but efficient?
Crucial: Choice of “stopping index” $n = n(\delta, y^\delta)$

Conditions needed:

- source conditions
- conditions restricting nonlinearity like:

$$||F(\tilde{x}) - F(x) - F'(x)(\tilde{x} - x)|| \leq \eta ||F(\tilde{x} - F(x)||$$

$$x, \tilde{x} \in B_\rho(x_0), \text{ with } \eta < 1/2$$

is hard to check

- Alternatives

1. Approach based on invariance properties (Deuflhard– Engl–Scherzer) leads to results about convergence rates, e.g.:

   Basic condition: “Newton–Mysovskii–condition”

   $$||(F'(x) - F'(x_\ast))F'(x_\ast)\sharp|| \leq C||x - x_\ast||,$$

where $F'(x_\ast)\sharp$ denotes a left inverse
Theorem: Newton–Mysovskii–condition, 
\[ \|F'(x)\| < C, \ x^\delta_k \text{ defined via Landweber. If} \]
\[ x^\dagger - x_0 = (F'(x)\dagger)^*F'(x)\dagger^*w \]
for some \( 0 < \nu \leq \frac{1}{2} \) and \( \|w\| \) sufficiently small, then
\[ \|x^\dagger - x^\delta_k\| \leq c_*\|w\|(k + 1)^{-\nu} \]
for \( 0 \leq k < k_* \). The stopping index \( k_* \) is defined via the discrepancy principle
\[ \|y^\delta - F(x_{k_*})\| \leq \tau\delta < \|y^\delta - F(x^\delta_k)\|, \ 0 \leq k \leq k_* \,, \]
with a suitable \( \tau > 2 \).
For \( \delta = 0 \) : rate holds for all \( k \). Otherwise:
\[ \|x^\delta_{k_*} - x^\dagger\| = O\left(\delta^{\frac{2\nu}{1+2\nu}}\right). \]
\( (\nu = \frac{1}{2} : x^\dagger - x_0 \in R(F'(x^\dagger)^*), \ \text{rate } O(\sqrt{\delta})) \).

39
2. Embedding into continuous methods, e.g,

- **continuous regularized Gauß-Newton-Method**

\[
\dot{x}(t) = (F'(x(t))^*F'(x(t)) + \alpha(t)I)^{-1}[F'(x(t))^*(y^\delta - F(x(t)) + \alpha(t)(x(t)) - \xi)]
\]

(Kaltenbacher, Neubauer, Ramm)

- **continuous Landweber**

\[
\dot{x}(t) = F'(x(t))^*(y^\delta - F(x(t)));
\]

(Tautenhahn)

and application of Lyapunov theory to the corresponding evolution equations for the error \(e(t) = x(t) - x^\dagger\) for proving

\[e(t) \to 0 \quad \text{as} \quad t \to \infty\]

connections to level set methods (Burger–Osher, Santosa, ...) and inverse scale space methods (Burger, Osher, Scherzer, ...) and to “online regularization” (Kügler)
Back to sparsity:

discrete: “as many 0’s as possible within range given by data with error bounds”

good approximation:

\[
\cdots + \alpha \|x\|_p \rightarrow \min.
\]

with

\[
0 < p < 1
\]

mathematical difficulties:

– non-convex

– non-differentiable

Method employed for analysis of maximum entropy regularization works for proving convergence, stability, convergence rates (Zarzer)
\( \ell_p - \text{Penalization} \)

- Penalty term

\[ \|x\|_p = \left( \sum_{i=1}^{\infty} |x_i|^p \right)^{1/p} \]

- the closer \( p \) gets to 0, the closer \( \|x\|_p \) gets to

\[ \|x\|_0^0 := \text{number of non-zero components of } x \]

<table>
<thead>
<tr>
<th>(1D) ( \ell_1 ) norm</th>
<th>(1D) ( \ell_p ) (quasi-)norm</th>
</tr>
</thead>
</table>

![Graphs showing \( \ell_p \) norms for different values of \( p \).]
Question: does the minimization of

\[ \| F(x) - y^\delta \|_2^2 + \alpha \| x \|_p^p \]

subject to \( x \in \ell_p \) with \( 0 < p \leq 1 \) render a regularization method?

- \( p = 1 \): yes, I. Daubechies, M. Defrise, C. Mol 04, (linear case), R. Ramlau, G. Teschke 05 (nonlinear case)

- \( p < 1 \): the penalty term \( \| \cdot \|_p^p \) is no norm and non-convex → ?

Idea (C.A. Zarzer): transformation of the original functional

\[ \| F(x) - y^\delta \|_2^2 + \alpha \| x \|_p^p, \quad 0 < p \leq 1 \]

to

\[ \| F(\mathcal{N}_{p,q}(\tilde{x})) - y^\delta \|_2^2 + \alpha \| \tilde{x} \|_q^q, \quad 1 \leq q \leq 2 \]

in order to obtain a convex penalty term satisfying norm-properties; similar to analysis of maximum entropy regularization
Qualitative inverse problems

invert map from parameters to bifurcation diagram, e.g., determine parameters such that

– distance to closest bifurcation point is maximized (→ robustness)
– width of hysteresis loop is maximized (→ switches)
– period of oscillations is kept constant over a wide range (→ circadian rhythm)

and do this in such a way that as few as possible parameters are changed (sparsity).
Inverse Bifurcation Analysis: Circadian Rhythm

- Autonomous oscillations with $\pm 24$ hour period


- Observed property: period of circadian rhythm insensitive to temperature variations
- Homeostasis: under environmental perturbations
  1. oscillatory behavior is conserved
  2. period of oscillation remains constant
Leloup-Goldbeter Circadian Rhythm Model

– Model: 10 variables, 38 parameters
– Consider variation in the maximum mRNA degradation rate, $v_{mT}$
\[
\begin{align*}
\frac{dC_C}{dt} &= -k_{dC} \cdot C_C - k_1 \cdot C_C + k_2 \cdot C_n - k_4 \cdot C_C + k_3 \cdot P_2 \cdot T_2 \\
\frac{dC_n}{dt} &= -k_{dN} \cdot C_n + k_1 \cdot C_C - k_2 \cdot C_n \\
\frac{dM_p}{dt} &= \frac{K_{IP}^n \cdot vsP(K_{IP}^n + C_n^n) - (k_d \cdot M_p + V_{mP} \cdot M_p)}{(K_{mP} + M_p)} \\
\frac{dM_t}{dt} &= \frac{K_{IT}^n \cdot V_{sT}}{(K_{IT}^n + C_n^n)} - \frac{(k_d \cdot M_t + V_{mT} \cdot M_t)}{(K_{mT} + M_t)} \\
\frac{dP_0}{dt} &= \frac{k_{sP} \cdot M_p - k_d \cdot P_0 - V_{1P} \cdot P_0}{(K_{1P} + P_0)} + \frac{V_{2P} \cdot P_1}{(K_{2P} + P_1)} \\
\frac{dP_1}{dt} &= \frac{V_{1P} \cdot P_0}{(K_{1P} + P_0)} - \frac{k_d \cdot P_1 - V_{2P} \cdot P_1}{(K_{2P} + P_1)} - \frac{V_{3P} \cdot P_1}{(K_{3P} + P_1)} + \frac{(V_{4P} \cdot P_2)}{(K_{4P} + P_2)} \\
\frac{dP_2}{dt} &= \frac{V_{3P} \cdot P_1}{(K_{3P} + P_1)} - \frac{V_{4P} \cdot P_2}{(K_{4P} + P_2)} - \frac{k_d \cdot P_2}{(K_{dP} + P_2) + k_4 \cdot C_C - k_3 \cdot P_2 \cdot T_2} \\
\frac{dT_0}{dt} &= \frac{k_{sT} \cdot M_t - k_d \cdot T_0 - V_{1T} \cdot T_0}{(K_{1T} + T_0)} + \frac{V_{2T} \cdot T_1}{(K_{2T} + T_1)} \\
\frac{dT_1}{dt} &= \frac{V_{1T} \cdot T_0}{(K_{1T} + T_0) - k_d \cdot T_1} - \frac{V_{2T} \cdot T_1}{(K_{2T} + T_1)} - \frac{V_{3T} \cdot T_1}{(K_{3T} + T_1)} + \frac{V_{4T} \cdot T_2}{(K_{4T} + T_2)} \\
\frac{dT_2}{dt} &= \frac{V_{3T} \cdot T_1}{(K_{3T} + T_1)} - \frac{V_{4T} \cdot T_2}{(K_{4T} + T_2)} + \frac{k_4 \cdot C_C - k_3 \cdot P_2 \cdot T_2 - k_d \cdot T_2 - V_{dT} \cdot T_2}{(K_{dT} + T_2)} 
\end{align*}
\]
- Drosophila model with parameter set $q^0$ exhibits stable limit cycles within a certain range of the parameter $v_{mT}$
- However, the period of oscillation varies with variations of $v_{mT}$
Inverse Bifurcation Problem

– Given upper and lower bounds for the parameter $v_{mT}$, find a parameter set $q$ such that the period of oscillations becomes exactly 24 hours for $v_{mT} \in [v, \bar{v}]$

– Furthermore, in finding a solution $q$, change as few components of the original parameter set $q^0$ as possible

– Use of $l_p$ -functional, $p \leq 1$, for identifying sparse parameter set

– Constrained minimization problem:

$$\min_{q} J(q) = TV(\text{period} - 24\text{hr}) + \alpha\|q - q^0\|^p_p$$

s.t. $LPC_{\text{left}} \leq v,$

$LPC_{\text{right}} \geq \bar{v}$

where $[LPC_{\text{left}}, LPC_{\text{right}}]$ denotes the region which allows for stable limit cycles (depending on $q$)
Test Case: Result

Optimized bifurcation diagram shows strictly constant period within given range of parameter $v_{mT}$.
Effectiveness of $l_p$ Functional for Sparsity

Smoothed $l_{0.1}$ penalty functional

$I_2$ penalty functional
Identification of parameters

Rates associated with the inferred parameters

May be relevant for drug development against

– Familian Advanced Sleep Phase Syndrom
– Delayed Sleep Phase Syndrom
Needs from systems biology:

– much larger ODE-systems

– multi-level aspects

– spatial aspects (PDE models)

– stochasticity