Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

Parameter Identification in Systems Biology: Solving Ill-posed Inverse Problems using Regularization

S. Müller, J. Lu, P. Kuegler, H.W. Engl

RICAM-Report 2008-25

Parameter Identification in Systems Biology: Solving Ill-posed Inverse Problems using Regularization

S. Müller*,[1] J. Lu[1] P. Kügler[2] H.W. Engl[1,3]

Abstract

Biochemical reaction networks are commonly described by non-linear ODE systems. Model parameters such as rate and equilibrium constants may be unknown or inaccessible and have to be identified from time-series measurements of chemical species. However, parameter identification is an ill-posed inverse problem in the sense that its solution lacks certain stability properties. In particular, modeling errors and measurement noise can be amplified considerably. These problems can be overcome by the use of so-called regularization methods. More specifically, parameter identification can be formulated in a stable way as a minimization problem with a data mismatch and a regularization term. On a benchmark problem, we demonstrate the stabilizing effect of Tikhonov regularization, i.e. we are able to identify the parameters from noisy measurements in a stable and accurate manner. In the algorithmic realization, we use the adjoint technique to efficiently compute the gradient of the data mismatch. Furthermore, we have developed a software package which allows the identification of parameters in valid SBML models of biochemical reaction networks.

*stefan.mueller@oeaw.ac.at

[1]Radon Institute for Computational and Applied Mathematics, Austrian Academy of Sciences, Altenberger Straße 69, 4040 Linz, Austria

[2]Industrial Mathematics Institute, Johannes Kepler University Linz, Altenberger Straße 69, 4040 Linz, Austria

[3]Rector's Office, University of Vienna, Dr.-Karl-Lueger-Ring 1, 1010 Wien, Austria

August 8, 2008

1 Introduction

Via the recent progress in the technology of biological data acquisition (in particular genomics, proteomics, and functional genomics), a wealth of information has become available at the molecular level. In combination with the flowcharts of biochemical pathways involved in cellular metabolism that have been revealed by biochemists over the last few decades, it is now possible to develop comprehensive models of cellular dynamics. A goal of the discipline *Systems Biology* is to combine diverse knowledge from bioinformatics and biochemistry to model the genetic regulation of cellular dynamics and to understand how function and properties arise from the network of gene interactions.

In order to construct quantitative and predictive mathematical models of gene regulation, one requires not only the knowledge of network topology but also the reaction rates (as functions of the chemical species) involved and the values of rate and equilibrium constants. As opposed to mathematical models in fields such as physics and engineering where the various constants can often (but not always) be quantified from direct measurements, in biological systems the *in-vivo* parameters are usually inaccessible and need to be inferred from datasets. Thus, the *reverse engineering* of gene networks is a crucial step towards the underlying goal of biological discovery via the systems approach.

There exist many types of mathematical models for biological networks, ranging from boolean models, ordinary differential equation (ODE) models, stochastic models to partial differential equation (PDE) models [30]. Our paper is restricted to the deterministic setting of ODE models, where the inaccessible parameter values are to be identified from time-course data for the species that can be measured [29, 8]. Driven by the prevalence of such parameter identification problems in the ODE context, a number of methods and software tools have been developed in the systems biology community [7, 36, 14, 31, 35, 11]. Amongst the existing methods, the predominant strategy consists of finding parameter values that minimize the *data mismatch*, the discrepancy between the experimental observations and the simulated data. The underlying premise is that the optimal parameter set is one which gives rise to simulated data that match the experimental observations as much as

possible. Computationally, the minimization of the objective function may involve some combination of local and global minimization methods [28, 27].

Although quantitative data is becoming available in biology, due to the difficulty of *in-vivo* measurements and the complexity of cellular environments, data error is inevitable. Faced with the noise present in biological data, one naturally asks the question: how does the inaccuracy in the measurements propagate back to errors in the inferred parameters [16, 8, 28, 13]? Parameter identification problems are typically *ill-posed*, in particular unstable with respect to the data noise [3]. Therefore, it is important to consider methods that control the impact of data error on the identified parameters [19]. For models consisting of a large network of genes, this issue becomes especially important since the instability of parameter identification typically grows with the problem size. Thus, the development of strategies for identifying parameters in a reliable manner is an important mathematical problem with significant practical implications.

In this paper, we present mathematical methods from the field of *Inverse Problems* to identify unknown parameters in ODE models of cellular processes from *noisy* time-series measurements [3]. In particular, we examine a benchmark problem that has been widely studied in the noise-free case [23, 14, 27, 35]. By adding noise to the artificial data, we show that such a benchmark parameter identification problem is in fact highly unstable. Moreover, we demonstrate that *Tikhonov regularization* can be effectively used to deal with the presence of data noise.

1.1 Organization of the paper

The paper is intended for systems biologists with mathematical background as well as for mathematicians who work in systems biology but are not experts in the field of inverse problems. Therefore, the paper has a partly tutorial character. It is organized as follows: In Section 2, we define the forward problem of solving an initial value problem and the inverse problem of identifying parameters from time-series measurements. As usual, the inverse problem is formulated as a search for minimizers of the data mismatch. However, the solution of this minimization problem is unstable, which means that model errors and measurement noise can be amplified

arbitrarily. On a benchmark problem, we demonstrate the severe effect of measurement noise on the solution of the parameter identification problem. In Section 3, we suggest the use of regularization methods to overcome the instabilities in parameter identification. We consider stability and convergence of regularization methods and discuss strategies for choosing the regularization parameter. On the benchmark problem, we demonstrate the stabilizing effect of a specific regularization method, namely Tikhonov regularization. We are able to identify the parameters in a stable and accurate manner even in the case of noisy measurements. In order to illustrate the concepts of ill-posedness and regularization, we outline linear inverse problems in Section 4. In the linear (and finite-dimensional) setting, we can pinpoint the sources of non-uniqueness and noise amplification and show how Tikhonov regularization can be used to obtain a solution which is unique and stable with respect to data noise. Moreover, we can estimate the total error consisting of the propagated data error and the regularization error. After this tutorial section, we return to the non-linear problem of parameter identification. In Section 5, we briefly discuss local and global optimization methods and show how the gradient of the data mismatch (required by local methods) can be computed efficiently. In Section 6, we describe the implementation of our algorithm; the software package we have developed allows the identification of parameters (and initial conditions) in valid SBML models of biochemical reaction networks from noisy measurement data. In the final section, we draw our conclusions and give an outlook for future work on this subject.

2 Parameter identification

ODE models are widely used to describe the dynamics of biochemical reaction networks. In the following, we present mathematical methods from the field of inverse problems to identify unknown parameters in ODE models from (noisy) time-series measurements.

2.1 ODE models for biological systems

Before we deal with the inverse problem of parameter identification, we consider the *forward problem* of solving an initial value problem. More specifically, we consider an autonomous ODE system (containing n ODE variables and m parameters) together with the initial conditions for the ODE variables:

$$\begin{aligned} \dot{y}(t) &= f(y(t), x), \\ y(0) &= y_0. \end{aligned} \tag{1}$$

Let the vector field $f(y, x)$ be $\mathbf{C}^r$, $r \geq 1$, on some open set $U_y \times U_x$ with $U_y \subset \mathbb{R}^n$, $U_x \subset \mathbb{R}^m$ and let $(y_0, x) \in U_y \times U_x$. Then there is a time interval $[0, T]$ and a unique solution $y(t)$, $t \in [0, T]$, to the initial value problem (1), cf. e.g. [33]. Moreover, the solution $y(t) = y(t, y_0, x)$ is a $\mathbf{C}^r$ function of y_0 and x.

The forward problem consists in solving an initial value problem for given parameters. We define the *forward operator*,

$$\begin{aligned} F : U_x &\to C^r([0, T], U_y), \\ x &\mapsto y(t), \end{aligned} \tag{2a}$$

as the "solution operator" or "parameter-to-solution-map". Given any parameter vector $x \in U_x$, the operator F yields the solution $y(t)$ of the initial value problem (1). In practical situations, the values of the ODE variables are observed only at discrete time points $0 = t_0, \ldots, t_i, \ldots, t_N = T$, and the forward operator is to be

defined as

$$F : U_x \to (U_y)^{N+1}, \qquad (2b)$$
$$x \mapsto (y(t_0), \ldots, y(t_i), \ldots, y(t_N)).$$

2.2 Instability in parameter identification

Now we turn to the *inverse problem* of interest. Given time-series measurements of (a subset of) the ODE variables y, we want to identify (a subset of) the parameters x.

If only some parameters have to be identified, then the vector field $f(y, x) = \tilde{f}(y, x, \tilde{x})$ not only depends on the unknown parameters x, but also on the known parameters $\tilde{x}$, which for instance might represent experimental settings. However, for notational simplicity, the dependence of the vector field on the known parameters is not stated explicitly.

In many cases, only some of the ODE variables (or combinations thereof) are measured. Then, the relation between the n ODE variables y and the k measured variables z has to be specified. In other words, one has to utilize an observation operator $O : \mathbb{R}^n \to \mathbb{R}^k, y \mapsto z$. In our benchmark problem, all ODE variables can be measured, and hence the following formulation of the inverse problem does not involve an observation operator. However, our considerations can be easily generalized.

Using the definition of the forward operator F, we can formulate the inverse problem of identifying the parameters x from time-series measurements y^δ of the ODE variables as a non-linear operator equation

$$F(x) = y^\delta \qquad (3)$$

for the unknown x. However, due to model errors and measurement noise, this problem may not even have a solution, i.e. no parameters x can be found such that the simulated data $y = F(x)$ and the measured data y^δ match. Hence, the problem

is redefined as a search for minimizers of the *data mismatch*,

$$\left\|F(x) - y^\delta\right\|^2 \to \min_x, \tag{4}$$

which has a solution on an appropriate set of admissible parameters. For given parameters x, the least squares term $\|F(x) - y^\delta\|^2$ quantifies the difference between the solution of the forward problem $y = F(x)$ and the time-series measurements y^δ. Depending on whether the data are continuous or discrete, the norm $\|.\|$ is chosen accordingly. For continuous data, we choose the L_2-norm,

$$\|y\|_{L_2} = \sqrt{\int_0^T y(t)^\top y(t)\,\mathrm{d}t}, \tag{5a}$$

but also other choices may be appropriate. For discrete data we choose the l_2-norm (or Euclidean norm),

$$\|y\|_{l_2} = \sqrt{\sum_{k=0}^{N} y(t_k)^\top y(t_k)}. \tag{5b}$$

In order to balance the contributions of the individual ODE variables, also weighted norms are frequently used (for both continuous and discrete data).

If more than one experimental data set is available, then the data mismatch involves a sum over all n_E experiments. Moreover, the individual experiments may have different settings, i.e. the vector field $f(y, x) = \tilde{f}(y, x, \tilde{x})$ may depend on parameters $\tilde{x}$, which are known but may have different values in the individual experiments. We write $f_{(i)}(y, x) = \tilde{f}(y, x, \tilde{x}_{(i)})$, $1 \le i \le n_E$, for the vector fields, $F_{(i)}$ for the resulting forward operators, and $y^\delta_{(i)}$ for the corresponding time-series measurements. Then, the data mismatch amounts to

$$\sum_{i=1}^{n_E} \left\|F_{(i)}(x) - y^\delta_{(i)}\right\|^2. \tag{6}$$

We denote the experiment index by (i), $1 \le i \le n_E$, in order to avoid confusion with the vector index.

As discussed in the introduction, the problem of parameter identification is typically ill-posed. According to Hadamard, a mathematical problem is called *well-posed* if (i) a solution exists (for all admissible data), (ii) the solution is unique (for all

admissible data), and (iii) the solution depends continuously on the data. If one of these requirements is not fulfilled, the problem is called *ill-posed*. On a benchmark problem, we demonstrate the instability in the identified parameters with respect to data noise; for a comprehensive mathematical treatment, see e.g. [3].

Benchmark problem

In Section 3, we study a metabolic pathway model which is described by an ODE system containing 8 variables, 36 parameters, and 2 experimental settings, cf. Table 1(a). Here, we state the ODE for just the first variable:

$$\frac{dG_1}{dt} = \frac{V_1}{1 + \left(\frac{P}{K_{i1}}\right)^{n_{i1}} + \left(\frac{K_{a1}}{S}\right)^{n_{a1}}} - k_1 \cdot G_1.$$

In the abstract notation, this corresponds to $\frac{dy_1}{dt} = f_1(y, x)$, where $y = (G_1, \ldots)$ is the 8-dimensional vector of ODE variables and $x = (V_1, K_{i1}, n_{i1}, K_{a1}, n_{a1}, k_1, \ldots)$ is the 36-dimensional vector of parameters. In order to account for the experimental settings we use the notation $f(y, x) = \tilde{f}(y, x, \tilde{x})$ with $\tilde{x} = (S, P)$.

The details of the benchmark problem (ODEs, initial values, parameters, experimental settings, data generation, minimization problem) will be given in Section 3. In Figure 1, we show some typical time-series measurements of two ODE variables for exact data and 5% data noise, and we display the corresponding relative errors in the identified parameters. While parameters are essentially identified correctly if exact data are used, using data with 5% relative noise leads to more than 100% relative error in some of the parameters. The severe effect of measurement noise demonstrates the ill-posedness of the parameter identification problem.

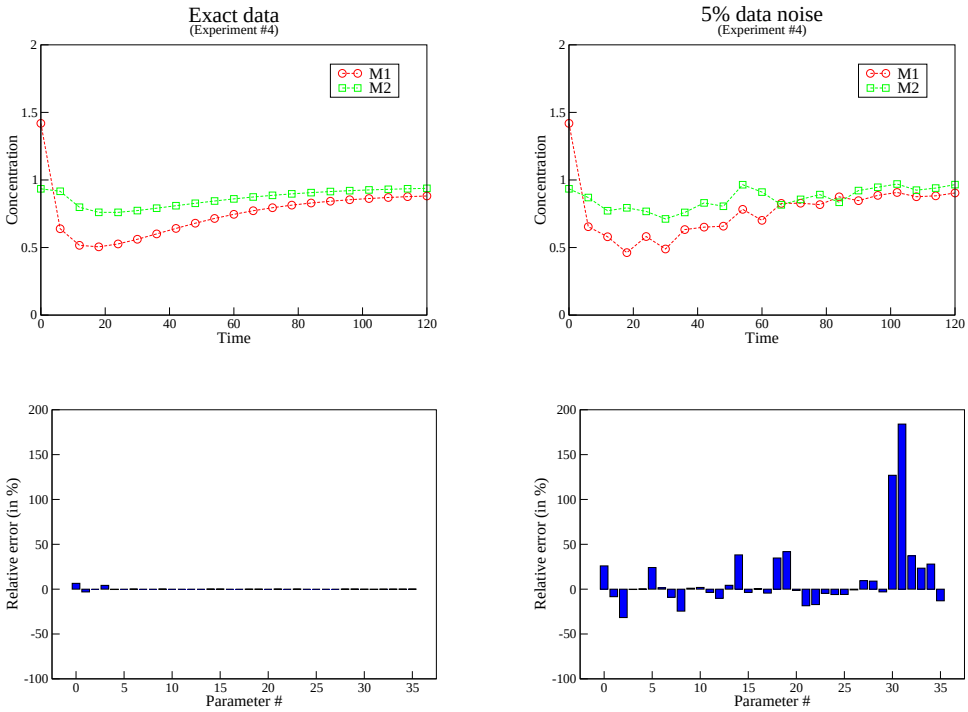

Figure 1: Parameter identification for a metabolic pathway model (see Section 3 for details). The corresponding ODE system contains 8 variables and 36 parameters. Upper row: typical time-series measurements of two ODE variables (for exact data and 5% data noise). Lower row: relative error in the identified parameters (for exact data and 5% data noise).

3 Regularization

In the previous section, we have defined the solution of the parameter identification problem as a minimizer of the data mismatch, cf. problem (4), and we have numerically demonstrated the instability of such a minimizer with respect to the data noise. In order to overcome the instability, regularization theory suggests to augment the data mismatch by a regularization term. In particular, we replace the minimization problem (4) by a family of nearby problems,

$$\left\|F(x) - y^\delta\right\|^2 + \alpha \left\|x - x^*\right\|^2 \to \min_x, \tag{7}$$

where α is the *regularization parameter* and x^* is an *a-priori estimate* for the solution. This approach is known as Tikhonov regularization. The minimization problem – and consequently its solutions – depends crucially on the regularization parameter. For $\alpha \to 0$ we get back the unregularized problem; for $\alpha \to \infty$ the problem has the solution $x = x^*$, thereby completely ignoring the measured data y^δ.

For general non-linear operators F, the data mismatch need not be convex and hence there may be global as well as local minimizers. In the following, we will consider only global minimizers as regularized solutions.

3.1 Theoretical background

Let us consider a scheme which assigns a regularized solution x_α^δ to any noisy data y^δ. In the case of linear inverse problems, there are several ways to construct an operator $R_\alpha : y^\delta \mapsto x_\alpha^\delta$ using the singular value decomposition of the forward operator. In the non-linear case, there are two main classes of methods for assigning solutions to data: variational methods, where the regularized solution is obtained as a minimizer of a functional (as in the case of Tikhonov regularization above); iterative methods, where the regularized solution is the output of an iterative algorithm and the iteration number plays the role of the regularization parameter [15].

Additionally, let $x^\dagger$ be an x^*-minimum norm solution for exact data y, i.e. $F(x^\dagger) = y$ and $\|x^\dagger - x^*\| \leq \|x - x^*\|$ for all x with $F(x) = y$, and let δ be an estimate for the

noise level, i.e. $\|y - y^\delta\| \leq \delta$.

Then, the requirements for a *regularization operator/scheme* are

(i) *stability*, i.e. continuous dependence of x_α^δ on y^δ, and

(ii) *convergence*, i.e. $x_\alpha^\delta \to x^\dagger$ as $\delta \to 0$.[1]

It can be shown [3] that, for convergence, the regularization parameter α has to depend on the noise level δ. The choice of the regularization parameter α in dependence on the noise level δ and the noisy data y^δ is called a *paramenter choice rule*. The rule is *a-priori* if $\alpha = \bar{\alpha}(\delta)$ and *a-posteriori* if $\alpha = \bar{\alpha}(\delta, y^\delta)$. For a paramenter choice rule, we require $\alpha \to 0$ as $\delta \to 0$.

If (i) and (ii) are fulfilled,[2] then the pair consisting of the regularization operator/scheme and the parameter choice rule is called a *regularization method*.

It can be shown [5] that non-linear Tikhonov regularization has indeed a stabilizing effect in the sense of requirement (i). Moreover, with a suitable choice of α in dependence on (δ, y^δ), Tikhonov regularization is convergent in the sense of requirement (ii). More specifically, $\alpha = \bar{\alpha}(\delta, y^\delta)$ has to be chosen such that $\alpha \to 0$ and $\delta^2/\alpha \to 0$ as $\delta \to 0$. As a consequence, Tikhonov regularization together with a parameter choice rule satisfying the above conditions is a regularization method. A possible parameter choice rule is Morozov's discrepancy principle, which is discussed below.

The *total error* between x_α^δ (the regularized solution for noisy data) and $x^\dagger$ (the unregularized solution for exact data) can be estimated by the *propagated data error* and the *regularization error*:

$$\underbrace{\|x_\alpha^\delta - x^\dagger\|}_{\text{total error}} \quad \leq \quad \underbrace{\|x_\alpha^\delta - x_\alpha^0\|}_{\text{propagated data error}} \quad + \quad \underbrace{\|x_\alpha^0 - x^\dagger\|}_{\text{regularization error}} \quad . \tag{8}$$

The propagated data error measures the effect of the data noise on the regularized solution, whereas the regularization error measures the effect of the regularization on the solution for exact data. For $\alpha \to 0$, the propagated data error explodes (due

[1] In general, the *convergence rate* (of $x_\alpha^\delta \to x^\dagger$ as $\delta \to 0$) can be arbitrarily slow. Still, one can estimate convergence rates under additional assumptions on the solution $x^\dagger$ (for exact data) and the a-priori estimate x^*. Typically, $x^\dagger - x^*$ is required to be in the range of a certain operator. These assumptions are called *source conditions*, cf. [3].

[2] In the non-linear case, there may be several global minimizers of the functional. Hence, one has to define stability and convergence in an appropriate set-valued sense.

to the instability of the unregularized problem) and the regularization error tends to zero. For $\alpha \to \infty$, on the other hand, the propagated data error tends to zero (due to the damping effect of regularization) and the regularization error dominates. Cf. Figure 2.

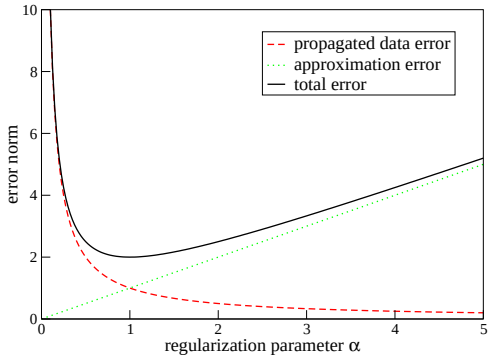

Figure 2: Dependence of the error terms on the regularization parameter. (Arbitrary units.)

Parameter choice rules aim at minimizing the total error, and there exists a variety of strategies for choosing the regularization parameter. A widely used (a-posteriori) rule is *Morozov's discrepancy principle* [3, p. 83]. It reflects the idea that one should not try to solve the inverse problem more accurately than the noise level. This means that one should avoid to match the noise in the data ("over-fitting"). More specifically, the discrepancy principle chooses the largest $\alpha = \bar{\alpha}(\delta, y^\delta)$ such that $\|F(x_\alpha^\delta) - y^\delta\| \leq \tau\,\delta$, where $\tau \geq 1$ is some fixed parameter.

In Section 4, we outline the basic theory of inverse problems in the simple linear and finite-dimensional setting. There, we illustrate the stabilizing effect of Tikhonov regularization, in particular, we give estimates for the propagated data error and the regularization error.

3.2 Tikhonov regularization for the benchmark problem: Numerical results

In the following, we study the metabolic pathway model depicted in Figure 3, which has been used as a benchmark problem in [23, 27]. The model describes the transformation of substrate S into product P via the intermediate metabolites M1, M2.

The enzymes E1, E2, E3 catalyzing the transformation (and the corresponding mRNAs G1, G2, G3) are produced by 3 genes, which in turn are regulated by the metabolites. The regulation involves activation as well as inhibition.

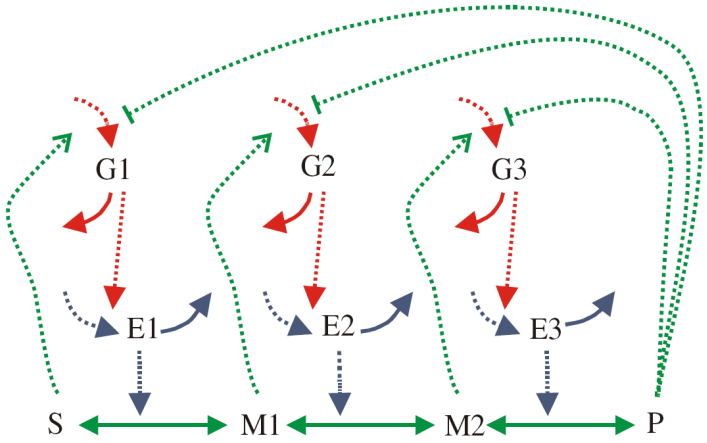

Figure 3: The three step biochemical pathway (reproduced from [23]). Solid arrows represent mass flow, and dashed arrows represent regulation, where $\rightarrow$ denotes activation and $\dashv$ denotes inhibition. Three genes are producing mRNAs G1, G2, G3 and enzymes E1, E2, E3 to regulate the transformation of substrate S into product P via the intermediate metabolites M1, M2.

The ODE system describing the dynamics of the model contains 8 variables, 36 parameters, and 2 experimental settings. It is given in Table 1(a). The 8 ODE variables are the concentrations of the mRNAs, enzymes, and metabolites. Using the notation of Section 2, we write $y = (G_1, G_2, G_3, E_1, E_2, E_3, M_1, M_2)$. The 36 parameters can be divided into the following classes: transcription/translation rates V, equilibrium constants K, Hill coefficients n, degradation rates k, and catalytic constants k_{cat}. We write $x = (V_1, K_{i1}, n_{i1}, K_{a1}, n_{a1}, k_1, \ldots, k_{cat3}, K_{m5}, K_{m6})$. The concentrations of substrate and product are constant in time, but they are used as experimental settings. In the notation of Section 2, we write $\tilde{x} = (S, P)$.

(a) ODE system

$$\frac{dG_1}{dt} = \frac{V_1}{1 + \left(\frac{P}{K_{i1}}\right)^{n_{i1}} + \left(\frac{K_{a1}}{S}\right)^{n_{a1}}} - k_1 \cdot G_1$$

$$\frac{dG_2}{dt} = \frac{V_2}{1 + \left(\frac{P}{K_{i2}}\right)^{n_{i2}} + \left(\frac{K_{a2}}{M_1}\right)^{n_{a2}}} - k_2 \cdot G_2$$

$$\frac{dG_3}{dt} = \frac{V_3}{1 + \left(\frac{P}{K_{i3}}\right)^{n_{i3}} + \left(\frac{K_{a3}}{M_2}\right)^{n_{a3}}} - k_3 \cdot G_3$$

$$\frac{dE_1}{dt} = \frac{V_4 \cdot G_1}{K_4 + G_1} - k_4 \cdot E_1$$

$$\frac{dE_2}{dt} = \frac{V_5 \cdot G_2}{K_5 + G_2} - k_5 \cdot E_2$$

$$\frac{dE_3}{dt} = \frac{V_6 \cdot G_3}{K_6 + G_3} - k_6 \cdot E_3$$

$$\frac{dM_1}{dt} = \frac{k_{cat1} \cdot E_1 \cdot \frac{1}{K_{m1}} \cdot (S - M_1)}{1 + \frac{S}{K_{m1}} + \frac{M_1}{K_{m2}}} - \frac{k_{cat2} \cdot E_2 \cdot \frac{1}{K_{m3}} \cdot (M_1 - M_2)}{1 + \frac{M_1}{K_{m3}} + \frac{M_2}{K_{m4}}}$$

$$\frac{dM_2}{dt} = \frac{k_{cat2} \cdot E_2 \cdot \frac{1}{K_{m3}} \cdot (M_1 - M_2)}{1 + \frac{M_1}{K_{m3}} + \frac{M_2}{K_{m4}}} - \frac{k_{cat3} \cdot E_3 \cdot \frac{1}{K_{m5}} \cdot (M_2 - P)}{1 + \frac{M_2}{K_{m5}} + \frac{P}{K_{m6}}}$$

(b) Initial values

G_1	0.66667
G_2	0.57254
G_3	0.41758
E_1	0.4
E_2	0.36409
E_3	0.29457
M_1	1.419
M_2	0.93464

(c) Parameters

#	*name*	*tv*	*lb*	*ub*
1	V_1	1	10^{-1}	10^{+1}
2	K_{i1}	1	10^{-1}	10^{+1}
3	n_{i1}	2	10^{-1}	10^{+1}
4	K_{a1}	1	10^{-1}	10^{+1}
5	n_{a1}	2	10^{-1}	10^{+1}
6	k_1	1	10^{-1}	10^{+1}
7	V_2	1	10^{-1}	10^{+1}
8	K_{i2}	1	10^{-1}	10^{+1}
9	n_{i2}	2	10^{-1}	10^{+1}
10	K_{a2}	1	10^{-1}	10^{+1}
11	n_{a2}	2	10^{-1}	10^{+1}
12	k_2	1	10^{-1}	10^{+1}
13	V_3	1	10^{-1}	10^{+1}
14	K_{i3}	1	10^{-1}	10^{+1}
15	n_{i3}	2	10^{-1}	10^{+1}
16	K_{a3}	1	10^{-1}	10^{+1}
17	n_{a3}	2	10^{-1}	10^{+1}
18	k_3	1	10^{-1}	10^{+1}
19	V_4	0.1	10^{-2}	10^{+0}
20	K_4	1	10^{-1}	10^{+1}
21	k_4	0.1	10^{-2}	10^{+0}
22	V_5	0.1	10^{-2}	10^{+0}
23	K_5	1	10^{-1}	10^{+1}
24	k_5	0.1	10^{-2}	10^{+0}
25	V_6	0.1	10^{-2}	10^{+0}
26	K_6	1	10^{-1}	10^{+1}
27	k_6	0.1	10^{-2}	10^{+0}
28	k_{cat1}	1	10^{-1}	10^{+1}
29	K_{m1}	1	10^{-1}	10^{+1}
30	K_{m2}	1	10^{-1}	10^{+1}
31	k_{cat2}	1	10^{-1}	10^{+1}
32	K_{m3}	1	10^{-1}	10^{+1}
33	K_{m4}	1	10^{-1}	10^{+1}
34	k_{cat3}	1	10^{-1}	10^{+1}
35	K_{m5}	1	10^{-1}	10^{+1}
36	K_{m6}	1	10^{-1}	10^{+1}

(d) Experimental settings

S	0.1	0.46416	2.1544	10
P	0.05	0.13572	0.3684	1

Table 1: Three step biochemical pathway. (a) ODE system for the concentrations of mRNAs G_1, G_2, G_3, enzymes E_1, E_2, E_3, and metabolites M_1, M_2. (b) Initial values of the ODE variables. (c) List of parameters: true values, *tv*, for generating the experimental data; lower and upper bounds, *lb* and *ub*, for generating a random initial guess. (d) Experimental settings: 4 concentrations of substrate and product, S and P respectively, for generating $4 \times 4 = 16$ experiments.

Data generation

The "experimental" data are generated by numerical integration of the ODE system using the initial values given in Table 1(b), the "true" parameter values given in Table 1(c), and the experimental settings given in Table 1(d). As already mentioned, the concentrations of substrate and product serve as experimental settings. More specifically, we choose 4 values of S and P, respectively, and combine them to generate a total of 16 experimental data sets. Each time integration runs from 0 to 120 s with measurements every 6 s, which means that we obtain 20 data points for each ODE variable. The time-series are denoted by $y_{(i)}$, $1 \leq i \leq 16$. In order to generate noisy data $y^{\delta}_{(i)}$, we add Gaussian noise $n^{\delta}_{(i)}$ to the exact data $y_{(i)}$. We choose a relative noise level of 5%, i.e. $\|n^{\delta}_{(i)}\|/\|y_{(i)}\| = 5\%$. In Figure 4, we show exact and noisy data for a particular choice of the experimental settings.

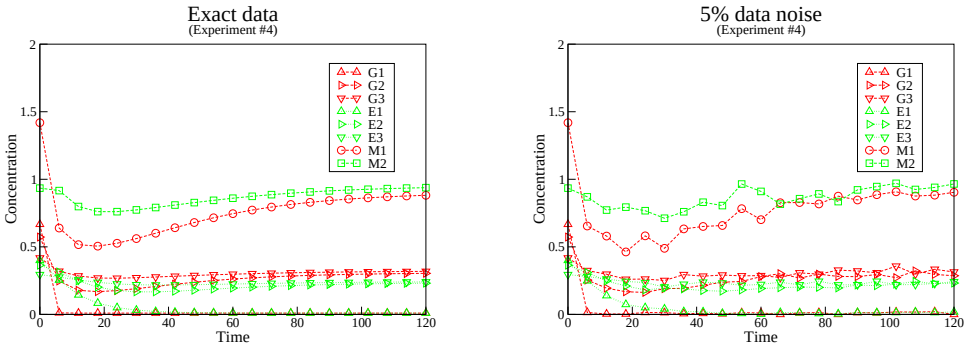

Figure 4: Experimental data generated by forward integration of the ODE system given in Table 1(a) using initial values given in Table 1(b), "true" parameter values given in Table 1(c), and experimental settings $S = 0.1$ and $P = 1$. Left: exact data. Right: data after adding 5% noise.

Minimization problem

The inverse problem consists of identifying all $m = 36$ parameters from time-series measurements of all $n = 8$ ODE variables. In the following, we use the first 8 of the 16 generated time-series as experimental data sets, i.e. $y^{\delta}_{(i)}$ with $1 \leq i \leq 8 = n_E$. As discussed in the theoretical part of this section, we identify the parameters x by minimizing the sum of the data mismatch and the regularization term,

$$\sum_{i=1}^{n_E} \left\| F_{(i)}(x) - y^{\delta}_{(i)} \right\|^2 + \alpha \left\| x - x^* \right\|^2 \rightarrow \min_x . \tag{9}$$

Here, the data mismatch involves a sum over the experiments. The use of the forward operators $F_{(i)}$, $1 \leq i \leq 8 = n_E$, reflects the different experimental settings being considered. The minimization is carried out for fixed choices of the regularization parameter α and the a-priori estimate x^*. In the following, we choose $x^* = 0$ thereby seeking solutions with minimal l_2-norm.

The minimization process starts by generating a random initial guess for the 36-dimensional parameter vector x, where each component of x has to be between the lower and upper bound specified in Table 1(c). This means that $x_i \in [lb_i, ub_i]$ for $i = 0, \ldots, m$. To be specific, we use the interval $[10^{-2}, 10^{+0}]$ for the parameters $V_4, k_4, V_5, k_5, V_6, k_6$ and $[10^{-1}, 10^{+1}]$ for all other parameters. Thus, each component of the initial guess for x may vary over 2 orders of magnitude.

As an illustration of our numerical results, we show the minimization history of the objective function and the average parameter error. In order to balance the contributions of the individual ODE variables (for each of the experiments), we modify the data mismatch term in (9). We use a weighted norm and define the *objective function* as

$$\sum_{i=1}^{n_E} \sum_{j=1}^{n} \frac{\|F_{(i)j}(x) - y^{\delta}_{(i)j}\|^2}{\|y^{\delta}_{(i)j}\|^2} + \alpha \|x\|^2 \tag{10}$$

where $F_{(i)j}$ and $y^{\delta}_{(i)j}$ refer to the j^{th} ODE variable (for the i^{th} experiment). The *average parameter error* between the current values x and the true values x^{tv} is defined as

$$\frac{1}{m} \sum_{i=1}^{m} \frac{|x_i - x_i^{tv}|}{|x_i^{tv}|}. \tag{11}$$

More precisely, this is the average over the absolute values of the relative parameter errors. In addition, we display the individual relative errors in the identified parameters.

Noise amplification

The following numerical results demonstrate the severe effect of measurement noise on the problem of parameter identification. The objective function to be minimized consists of the data mismatch only, i.e. we do not use any regularization ($\alpha = 0$).

In Figure 5, we show the minimization history of the data mismatch and the average parameter error for exact data and 5% data noise. Furthermore, we display the individual relative errors in the identified parameters. (The bar charts have already been displayed in Figure 1.) For exact data, the average relative error in the identified parameters is less then 1%, whereas for only 5% data noise, the relative error in some of the parameters is larger than 100%. These results motivate the use of regularization.

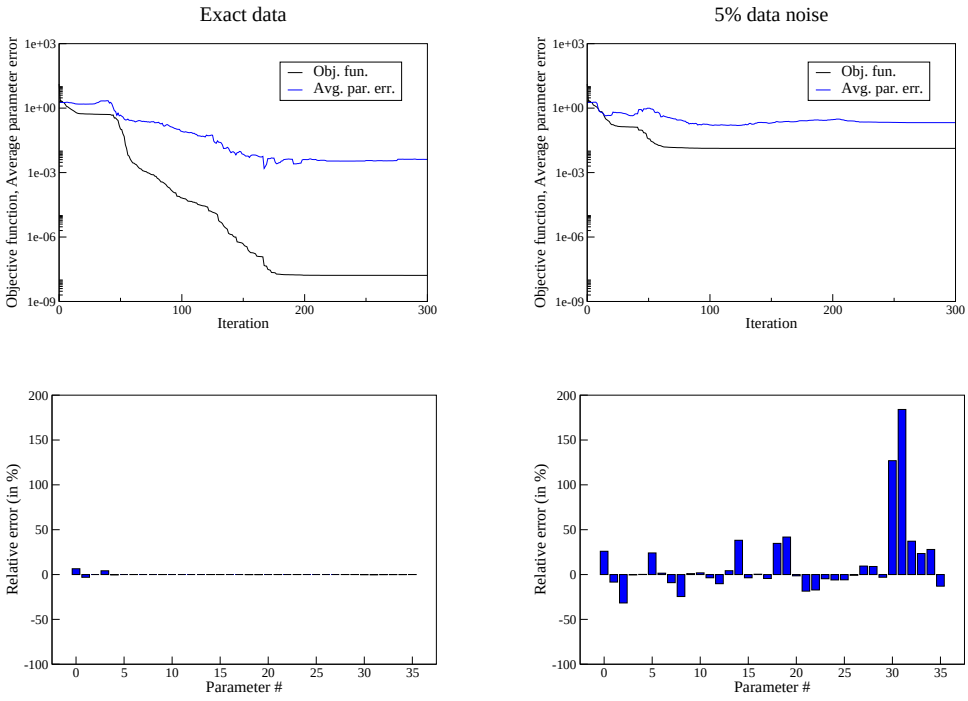

Figure 5: Parameter identification (without regularization) for exact and noisy data. Upper row: minimization history of the objective function and the average parameter error (for exact data and 5% data noise). Lower row: relative error in the identified parameters (for exact data and 5% data noise).

The stabilizing effect of regularization

Below, we present numerical results demonstrating the stabilizing effect of Tikhonov regularization on the problem of parameter identification. The objective function to be minimized consists of the data mismatch and the regularization term, cf. (10).

In Figure 6, we show the minimization history of the objective function and the parameter error for 5% data noise. On the left, we once again display the results using no regularization ($\alpha = 0$), but for the purpose of comparison, we use a different scale. As is typically observed, a decrease in the objective function need not be accompanied by a decrease in the parameter error. On the right, we display the results using Tikhonov regularization with $\alpha = 10^{-5}$. (For the choice of the regularization parameter, see below.) In the bar chart, we compare the individual relative errors in the identified parameters which have been obtained by using either no regularization or Tikhonov regularization. The noise amplification in individual parameters is damped considerably by using regularization.

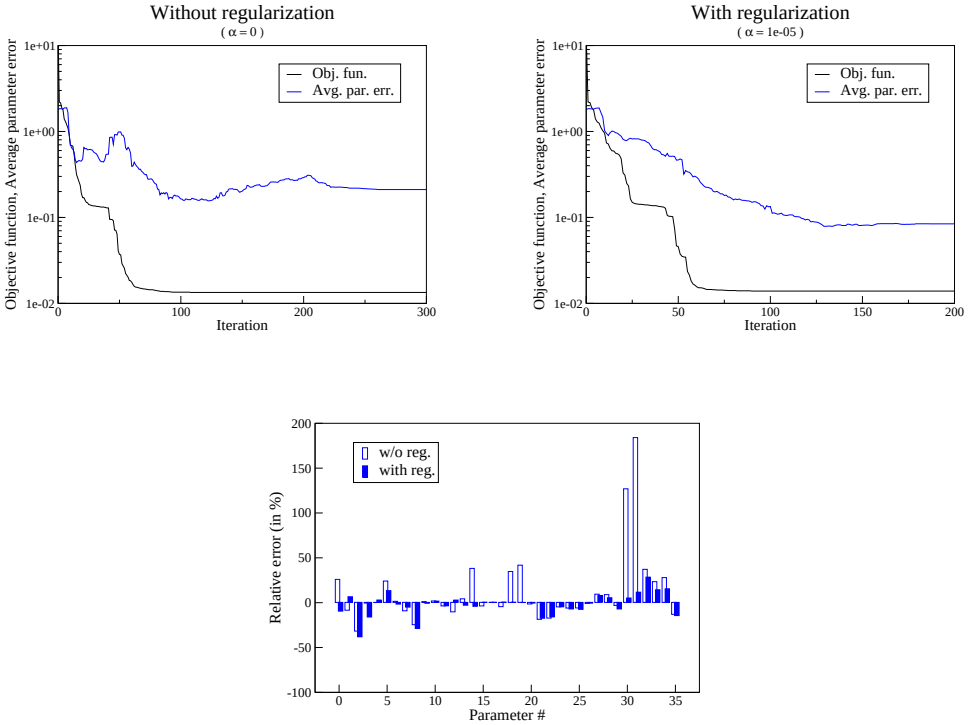

Figure 6: Parameter identification for 5% data noise. Upper row: minimization history of the objective function and the parameter error. Left: no regularization ($\alpha = 0$). Right: Tikhonov regularization with $\alpha = 10^{-5}$. Lower row: relative error in the identified parameters (without and with regularization).

The choice of the parameter α is a crucial issue for Tikhonov regularization. As we have seen in the theoretical part of this section, there exists a minimum of the total error as a function of the regularization parameter. In Figure 7, we show the corresponding numerical results, i.e. the average parameter error for several values of α. This figure numerically confirms Figure 2.

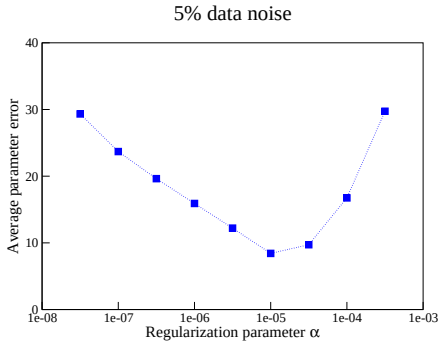

Figure 7: Parameter identification for 5% data noise using Tikhonov regularization. Average parameter error for several values of the regularization parameter.

In order to illustrate the effect of near optimal regularization, we have picked a parameter close to the best choice (as suggested by Figure 7). In practical problems (where the solution for exact data is not known), a plot as in Figure 7 is not available. Hence, parameter choice rules (such as Morozov's discrepancy principle [3, p. 83]) have to be used in order to determine the regularization parameter which minimizes the total error.

4 Illustration in the linear setting

Linear inverse problems can be analyzed more easily than general non-linear problems since one can use the singular value decomposition of the forward operator. In the finite-dimensional case, even simple methods from linear algebra suffice to pinpoint the sources of non-uniqueness and noise amplification and to illustrate the effect of Tikhonov regularization. Ill-posedness (in the sense of violating Hadamard's third criterion) is an infinite-dimensional effect, which can, however, cause numerical instabilities in finite-dimensional approximations.

In the following, we study inverse problems in the linear and finite-dimensional setting. In this case, the application of the forward operator $F(x)$ is just a matrix multiplication $A\,x$. As in the non-linear setting, one looks for a minimizer of the least squares error:

$$\|A\,x - y\|^2 \to \min_x, \tag{12}$$

where $x \in \mathbb{R}^m$, $y \in \mathbb{R}^n$, and $A \in \mathbb{R}^{n\times m}$. A necessary condition for a minimizer is a zero gradient. Taking the derivative of $\|A\,x - y\|^2$ with respect to x and setting it to zero yields the so-called normal equation

$$A^\top A\,x = A^\top y. \tag{13}$$

Since $\|A\,x - y\|^2$ is convex as a function of x, this condition is also sufficient.

4.1 Non-uniqueness and noise amplification

If the matrix $A^\top A$ is invertible, then the normal equation (13) – and equivalently the minimization problem (12) – has the unique solution $x = (A^\top A)^{-1}A^\top y$. If the matrix $A^\top A$ is not invertible, then there are infinitely many solutions of the normal equation. Among these one usually chooses the minimal norm solution $x^\dagger$, i.e. $\|x^\dagger\| \leq \|x\|$ for all solutions x of the normal equation.

But even if there is a unique solution, we may face the problem of noise amplification, since usually we have noisy data y^δ instead of exact data y. Let $A^\top A$ be invertible, x^δ be the unique solution for noisy data, and x be the unique solution for exact

data. Then, we can estimate

$$\|x^\delta - x\| \leq \frac{1}{\sigma_{\min}} \|y^\delta - y\|, \tag{14}$$

where $\sigma_{\min} = \sigma_{\min}(A)$ is the smallest singular value of the matrix A.

This and the following estimates can be easily derived using the normal equation (for exact or noisy data) and the Cauchy-Schwarz inequality. We note that the singular values σ_i of a matrix A are the square-roots of the positive eigenvalues λ_i of the matrix $A^\top A$, i.e. $\sigma_i(A) = \sqrt{\lambda_i(A^\top A)} > 0$.

The estimate (14) is sharp; the inequality becomes an equality if the noise occurs in a certain direction (more specifically, in the direction of the right singular vector belonging to the smallest singular value). This illustrates the problem of noise amplification in the presence of small singular values. For $\sigma_{\min} \ll 1$, even a small error $y^\delta - y$ in the data may lead to a significant error $x^\delta - x$ in the solution; the noise can be amplified by as much as $1/\sigma_{\min}$.

Clearly, non-uniqueness is a violation of Hadamard's second criterion for well-posedness. The observed noise amplification, on the other hand, is not a violation of Hadamard's third criterion in the strict sense. A finite dimensional linear inverse problem is never ill-posed in the sense that the solution does not depend continuously on the data, but for very small singular values this situation is certainly approached.

4.2 Tikhonov regularization and error estimates

In our linear inverse problem, non-uniqueness is caused by a non-invertible matrix $A^\top A$ and noise amplification is caused by small singular values. The crucial idea behind regularization is to approximate the original (ill-posed) problem by a family of (well-posed) neighboring problems. Using Tikhonov regularization, the minimization problem is reformulated as

$$\|A\,x - y\|^2 + \alpha \|x\|^2 \to \min_x, \tag{15}$$

where $\alpha > 0$ is the regularization parameter. Solving the minimization problem (15) is equivalent to solving the regularized normal equation

$$(A^\top A + \alpha\, I)\, x_\alpha = A^\top y, \tag{16}$$

where we use the notation x_α to indicate the dependence of the solution on the regularization parameter.

The matrix $A^\top A + \alpha\, I$ has only positive eigenvalues and hence is invertible. (Since $\lambda_i(A^\top A + \alpha\, I) = \lambda_i(A^\top A) + \alpha > 0$.) Consequently, the regularized normal equation has the unique solution $x_\alpha = (A^\top A + \alpha\, I)^{-1} A^\top y$. At the same time, noise amplification is damped, which can be seen from the following estimates. We first deal with the *propagated data error* between the regularized solutions x_α^δ (for noisy data y^δ) and x_α (for exact data y):

$$\|x_\alpha^\delta - x_\alpha\| \leq \frac{1}{2\sqrt{\alpha}}\, \|y^\delta - y\|. \tag{17}$$

Next, we estimate the *regularization error* between the regularized solution x_α and an unregularized solution x (both for exact data y):

$$\|x_\alpha - x\| \leq \frac{\alpha}{(\sigma_{\min}^2 + \alpha)\, \sigma_{\min}}\, \|y\|. \tag{18}$$

Now, we assume that the data noise is bounded by

$$\|y^\delta - y\| \leq \delta, \tag{19}$$

and conclude that $\|y\| \leq \|y^\delta\| + \delta$. Then, by the triangle inequality, the *total error* can be estimated by

$$\begin{aligned} \|x_\alpha^\delta - x\| &\leq \|x_\alpha^\delta - x_\alpha\| + \|x_\alpha - x\| \\ &\leq \frac{1}{2\sqrt{\alpha}}\, \delta + \frac{\alpha}{(\sigma_{\min}^2 + \alpha)\, \sigma_{\min}}\, \left(\|y^\delta\| + \delta\right). \end{aligned} \tag{20}$$

For fixed noise level δ, the two error terms exhibit different qualitative behavior with respect to the regularization parameter α, cf. Figure 2. For increasing α, the propagated data error is decreasing, while the regularization error is increasing.

Hence, there exists an $\alpha = \bar{\alpha}(\delta, y^\delta)$, which minimizes the total error. As discussed in the theoretical part of Section 3, the choice of regularization parameter α in dependence on the noise level δ and the noisy data y^δ is called parameter choice rule. An a-priori choice, $\alpha = \bar{\alpha}(\delta)$, will not depend on the specific data, but on the noise level only. An a-posteriori choice, $\alpha = \bar{\alpha}(\delta, y^\delta)$, changes the regularization parameter for each specific data.

4.3 A numerical example

To give an illustration of the instability of linear inverse problems, we consider a 3-dimensional linear ODE system. The example is motivated by a biological system called the *repressilator* [2, 24]. The repressilator is a regulatory cycle of 3 genes where each gene represses its successor in the cycle. The model involves basal gene expression (b), repression (r), and degradation (d). The corresponding linear ODE system (a simplified and linearized version of the general non-linear system) is given by

$$\begin{aligned} \dot{y}(t) &= Jy(t) + k, \\ y(0) &= y_0, \end{aligned} \tag{21}$$

where

$$J = \begin{pmatrix} -d & 0 & -r \\ -r & -d & 0 \\ 0 & -r & -d \end{pmatrix}, \quad k = \begin{pmatrix} b \\ b \\ b \end{pmatrix}.$$

If we choose $d = 1$ and $r = 2$, then the eigenvalues of the Jacobian J can be computed to be $\{-3, -\sqrt{3}\,i, +\sqrt{3}\,i\}$. Thus, there are two imaginary eigenvalues leading to oscillations and one negative eigenvalue leading to exponential decay. We complete the problem specification by choosing $b = 3$ and $y_0 = (2, 1, 1)^\top$ and obtain the time-series depicted in Figure 8.

The initial value problem (21) can be solved analytically (using the matrix exponential of Jt):

$$y(t) = \underbrace{e^{Jt}}_{B}\, y_0 + \underbrace{(e^{Jt} - I)\, J^{-1}\, k}_{c}. \tag{22}$$

That is, the (affine) operator mapping the initial condition y_0 to the solution $y(t)$ involves the matrix $B = e^{Jt}$ and the constant vector c (arising from a non-zero vector k).

While the forward problem of solving the ODE system for given initial conditions is stable, the inverse problem of inferring the initial conditions y_0 from measurements at some later time, $y(\bar{t}\,)$, is highly unstable. At time $\bar{t} = 4$, the singular values of matrix $e^{J\bar{t}}$ are $\{6 \cdot 10^{-6}, 1, 1\}$. Due to the smallness of $\sigma_{\min} = 6 \cdot 10^{-6}$, we expect a tremendous amount of noise amplification.

Suppose we add measurement noise n^δ to the exact data $y(\bar{t}\,)$ in order to obtain noisy data $y^\delta(\bar{t}\,) = y(\bar{t}\,) + n^\delta$. The solution of the inverse problem can again be expressed analytically, which yields the following error between the solution y_0^δ for noisy data $y^\delta(\bar{t}\,)$ and the solution y_0 for exact data $y(\bar{t}\,)$:

$$y_0^\delta - y_0 = e^{-J\bar{t}} \left(y^\delta(\bar{t}\,) - y(\bar{t}\,) \right)$$

Let us assume a measurement noise of $n^\delta = \delta \cdot (1, 0, 0)^\top$ with a noise level of $\delta = 10^{-2}$. Now, we can plot the norm of the error as a function of the measurement time. In the left plot of Figure 8, we see an exponential growth in the error with respect to the measurement time.

As discussed in the previous section, the instability of this linear inverse problem can be treated with Tikhonov regularization (or other regularization methods, see e.g. [3]). We choose a regularization parameter of $\alpha = 10^{-2}$, and plot the norm of the error again. In the right plot of Figure 8, we see that the exponential growth in the error has been removed by the use of regularization. (Also note the difference in scales.) Since the noise amplification is damped, the initial conditions can be recovered in a stable manner.

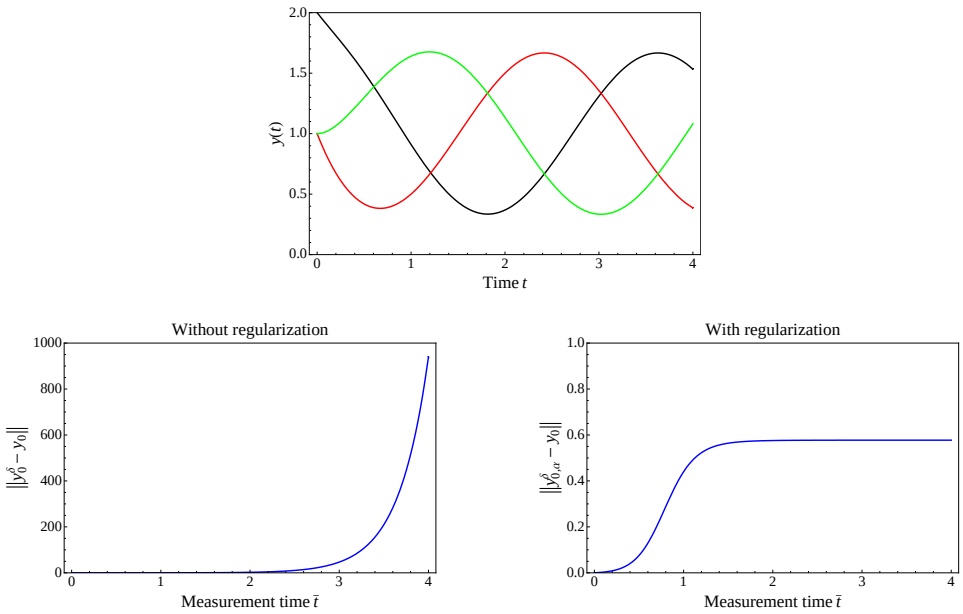

Figure 8: Identification of initial conditions from measurements at a later time for the *repressilator* (see text for details). The dynamics is modeled by a 3-dimensional linear ODE system. Upper row: typical time-series for the 3 ODE variables. Lower row: error in the identified initial conditions (for 1% measurement noise) as a function of the measurement time. Left: without regularization ($\alpha = 0$). Right: with regularization ($\alpha = 10^{-1}$).

5 Back to the non-linear problem: Algorithm

As we have seen in Section 3, the inverse problem of parameter identification from noisy measurements can be formulated in a stable way as a minimization problem with a data mismatch and a regularization term, cf. (7). In the following, we briefly discuss local and global optimization methods and show how the gradient of the data mismatch (required by local methods) can be computed efficiently.

5.1 Local and global optimization

In general, there are two main classes of methods to solve an optimization problem: *local* and *global* methods. Local search techniques [25] are often gradient-based, i.e., they use in some way the direction of steepest descent of the functional. For many local optimization algorithms convergence results are known; in particular, it can be proved that, provided certain conditions are satisfied, from a sufficiently close starting point the search converges to a local minimum. Global search techniques [12], on the other hand, strive for a global optimum. They comb through the search space following deterministic, stochastic, or combined strategies. Clearly, the size and dimension of the search space limit the efficiency of global optimization algorithms.

In order to obtain an algorithm which is efficient and converges to the global optimum, one often combines local and global methods [28, 27]. In this way, one joins the rapid convergence of local search techniques with the comprehensive nature of global search techniques. In an outer loop, the global algorithm locates a solution in the vicinity of the global optimum, whereas in the inner loop, the local algorithm refines a given solution up to a certain precision.

In this study, we focus on the instability of the unregularized problem of parameter identification and the stabilizing effect of Tikhonov regularization. In particular, we attempt to avoid the issue of non-uniqueness. For this purpose, we ensure that the random initial guess of the minimization problem is in the vicinity of the global minimum. In particular, we specify appropriate intervals for all parameters to be identified. Under these conditions, local optimization methods are sufficient to find the global minimum.

5.2 Efficient gradient computation

Local optimization algorithms require the gradient of the objective function. The gradient of the regularization term is obvious, hence below we consider the gradient of the data mismatch only. As an approximation, the gradient of

$$J(x) = \|F(x) - y^\delta\|^2 \tag{23}$$

could be easily computed by the finite difference method,

$$(\nabla J(x))_i \approx \frac{J(x + \epsilon e_i) - J(x)}{\epsilon}, \quad i = 1, \ldots, m, \tag{24}$$

with $\epsilon \ll 1$ and e_i denoting the i-th unit vector in $\mathbb{R}^m$. This computation involves $m + 1$ evaluations of $J(x)$ and hence as many evaluations of the forward operator $F(x)$ thereby requiring the solution of $m + 1$ initial value problems. This effort would be necessary in each step of the optimization process thereby making this approach quite inefficient.

The computational disadvantage of finite difference and related methods can be avoided if the gradient $\nabla J(x)$ is provided by means of the *adjoint technique*. We start by introducing the notation δx for small variations in the parameters. The resulting linearized variations in the solution $y = F(x)$ of the initial value problem (1) amount to

$$\delta y = F'(x)\, \delta x, \tag{25}$$

where $F'(x)$ is the derivative of the forward operator. More specifically, the definition of $F'(x)$ involves the differentiation of the initial value problem (1) with respect to the parameters x and thus requires the solution of m linear initial value problems. However, as we will see below, we need not evaluate the derivative of the forward operator.

In the following we assume continuous data and use the L_2-norm, i.e.

$$J(x) = \int_0^T (F(x) - y^\delta)^2 \, \mathrm{d}t. \tag{26}$$

Introducing a Lagrange multiplier ψ for the ODE constraint, we form the functional

$$L(y,x) = \int_0^T (y - y^\delta)^2 \,\mathrm{d}t + \int_0^T \psi^\top (\dot{y} - f(y,x)) \,\mathrm{d}t. \tag{27}$$

For the solution $y = F(x)$ of the initial value problem (1), the second term vanishes and we have $L(F(x),x) = J(x)$. We use integration by parts

$$\int_0^T \psi^\top \dot{y} \,\mathrm{d}t = \psi^\top y\Big|_0^T - \int_0^T \dot{\psi}^\top y \,\mathrm{d}t$$

and obtain

$$\begin{aligned} L(y,x) = &\int_0^T (y - y^\delta)^2 \,\mathrm{d}t + \psi(T)^\top y(T) - \psi(0)^\top y_0 - \int_0^T \dot{\psi}^\top y \,\mathrm{d}t \\ &- \int_0^T \psi^\top f(y,x) \,\mathrm{d}t. \end{aligned}$$

The variation δL due to variations δy, δx amounts to

$$\begin{aligned} \delta L = &\int_0^T 2\,(y - y^\delta)^\top \,\delta y \,\mathrm{d}t + \psi(T)^\top \delta y(T) - \int_0^T \dot{\psi}^\top \delta y \,\mathrm{d}t \\ &- \int_0^T \psi^\top (f_y(y,x)\,\delta y + f_x(y,x)\,\delta x) \,\mathrm{d}t. \end{aligned}$$

After collecting terms containing δy or δx, we have

$$\begin{aligned} \delta L = &\int_0^T \Big(2\,(y - y^\delta)^\top - \dot{\psi}^\top - \psi^\top f_y(y,x)\Big)\,\delta y \,\mathrm{d}t + \psi(T)^\top \delta y(T) \\ &- \int_0^T \psi^\top f_x(y,x) \,\mathrm{d}t\,\delta x. \end{aligned} \tag{28}$$

Now, let the Lagrange multiplier ψ satisfy the so-called *adjoint equations*:

$$\begin{aligned} \dot{\psi}(t) &= -\,f_y(y(t),x)^\top \psi(t) + 2\,(y(t) - y^\delta(t)), \\ \psi(T) &= 0. \end{aligned} \tag{29}$$

This is a linear ODE system with a terminal condition, i.e., a linear terminal value problem. Then, using (29) in (28), the variation δL does not depend on δy, i.e. $\delta L = L_x\,\delta x$. Additionally, we have $\delta L = \delta J$ (for $y = F(x)$) and $\delta J = \nabla J\,\delta x$ (by

definition). Hence, we can directly read off the gradient $\nabla J = L_x$ from (28):

$$\nabla J(x) = -\int_0^T \psi(t)^\top f_x(y(t), x)\,\mathrm{d}t. \tag{30}$$

This means that after solving the non-linear initial value problem (1) and the linear terminal value problem (29), we can calculate the gradient by a simple integration over time. As a consequence, gradient calculations become essentially independent of the number of parameters. From (24) we know that, using the finite difference method, one has to solve as many non-linear initial value problems as there are parameters to be identified. Using the adjoint technique, one can avoid this computational effort.

We conclude this section by summarizing our algorithm: (i) we formulate the problem of parameter identification as a *minimization problem* for the sum of the data mismatch and a regularization term (*Tikhonov regularization*); (ii) provided that the initial guess is in the vicinity of the global minimum, we apply *local (gradient-based) optimization* methods to find the global minimum; otherwise, we combine local and global optimization methods; (iii) in order to compute the gradient of the data mismatch efficiently, we use the *adjoint technique*. This algorithm has been applied to obtain the results presented in Section 3. In the next section, we describe the implementation of the algorithm.

6 Software

We have developed a software package, which allows the identification of parameters and initial conditions in ODE models of biochemical reaction networks. The models can be given either as explicit ODE systems or as reaction networks specified in the *Systems Biology Markup Language (SBML)* [6] in which case the resulting ODE systems are constructed automatically. Additionally, the problem specification includes a list of unknown parameters and initial conditions (with corresponding lower and upper bounds) and the measured data. The data can be measured either continuously or at discrete time points.

The problem of parameter identification is formulated as a minimization problem for

the sum of the data mismatch and a regularization term (Tikhonov regularization). In order to solve the minimization problem, we combine local and global optimization methods. In an outer loop, we have implemented a global method called Scatter Search [27] to locate a solution in the vicinity of the global optimum, whereas in the inner loop, our software uses the package *Interior Point Optimizer (Ipopt)* [32] to refine the solution by employing gradient-based methods. Any gradient-based optimizer requires the evaluation of the functional and the computation of its gradient. In our package, the values of the functional and its gradient are provided by the forward and adjoint solvers of the *SBML ODE Solver library (SOSlib)* [22, 20].

SOSlib is a *C/C++* programming library for the symbolic and numerical analysis of ODE systems derived from biochemical reaction networks encoded in SBML. It is written in *ANSI/ISO C* and distributed under the terms of the *GNU Lesser General Public License (LGPL)*. The package employs *libSBML*'s abstract syntax tree representation of mathematical formulas to construct the vector field of the ODE system from the reaction network. It also permits the symbolic differentiation of the vector field in order to obtain expressions for the Jacobian matrix and other derivatives. The numerical integration of the resulting ODE systems is performed by the *CVODES* solver from the *SUNDIALS* suite [10].

Recent efforts in the development of SOSlib have been focused on extensions that allow to address not only forward problems but also inverse problems associated with biochemical reaction networks [20]. In particular, we have extended SOSlib with capabilities to solve adjoint equations. This allows the efficient computation of the gradient of the functional, which in turn facilitates the computationally efficient identification of model parameters and initial conditions from experimental data.

7 Conclusions and outlook

The identification of parameters from noisy data is a challenging problem, due to its ill-posedness. Via numerical examples and linear theory, we have illustrated the nature of instability in such an inverse problem; a small error in the data can result in a significant error in the parameters.

On a benchmark problem, we have shown that Tikhonov regularization allows one to remedy the ill-posedness of parameter identification. In particular, by choosing the regularization parameter appropriately based on the knowledge of a bound for the data noise, parameters can be identified in a stable and accurate manner.

There remain several topics for further research. In particular, we have so far only looked at identifying scalars in ODE systems; but in some situations, one would like to identify functions of ODE variables. For instance, in some biochemical applications one would like to determine the functional dependence of the reaction rates on the species concentrations instead of using finite-dimensional parameterizations such as Michaelis-Menten kinetics or Hill-functions. This leads to infinite-dimensional, possibly severely ill-posed inverse problems where choosing an appropriate regularization method is especially important [17, 18].

In this paper, we have only used the Euclidean norm of the parameters in the regularization term; but there are other regularization terms that may be more appropriate in other applications. For instance, one may want to identify the topology of a gene regulatory network. Starting from a completely connected network, one looks for the smallest network that is consistent with the data. Alternatively, one may have a-priori knowledge about the parameters values (from literature) and believes that only a few of them are actually inaccurate. In these situations, the use of sparsity-promoting regularization can be used to identify parameter sets of small cardinality [1, 26, 21, 34].

Aside from the class of variational regularization methods, iterative regularization methods could also be applied. In iterative methods, the regularized solution is the output of an iterative algorithm and the iteration number plays the role of the regularization parameter. One advantage of iterative methods is that the discrepancy principle (an a-posteriori rule for choosing the regularization parameter [3, p. 83])

can be implemented easily [9, 4, 15].

Acknowledgements

This work has been supported financially by the Vienna Science and Technology Fund "WWTF" (Project MA05). All support as well as fruitful discussions with Prof. Peter Schuster, Dr. Christoph Flamm, and MSc Lukas Endler, Rainer Machne, and Clemens Zarzer are gratefully acknowledged.

References

[1] Ingrid Daubechies, Michel Defrise, and Christine De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.*, 57(11):1413–1457, 2004.

[2] Michael B. Elowitz and Stanislas Leibler. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403:335–338, 2000.

[3] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[4] H. W. Engl and O. Scherzer. Convergence rates results for iterative methods for solving nonlinear ill-posed problems. In *Surveys on solution methods for inverse problems*, pages 7–34. Springer, Vienna, 2000.

[5] Heinz W. Engl, Karl Kunisch, and Andreas Neubauer. Convergence rates for Tikhonov regularisation of nonlinear ill-posed problems. *Inverse Problems*, 5(4):523–540, 1989.

[6] A. Finney and M. Hucka. Systems biology markup language: Level 2 and beyond. *Biochem Soc Trans*, 31(Pt 6):1472–1473, Dec 2003.

[7] K. G. Gadkar, R. Gunawan, and F. J. Doyle. Iterative approach to model identification of biological networks. *BMC Bioinformatics*, 6:155, 2005.

[8] R. Gunawan, K. G. Gadkar, and F. J. Doyle. Methods to identify cellular architecture and dynamics from experimental data. In Z. Szallasi, J. Stelling, and V. Periwal, editors, *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*, pages 179–183. MIT Press, Boston, 2006.

[9] Martin Hanke, Andreas Neubauer, and Otmar Scherzer. A convergence analysis of the Landweber iteration for nonlinear ill-posed problems. *Numer. Math.*, 72(1):21–37, 1995.

[10] Alan C. Hindmarsh, Peter N. Brown, Keith E. Grant, Steven L. Lee, Radu Serban, Dan E. Shumaker, and Carol S. Woodward. SUNDIALS: Suite of Nonlinear and Differential/Algebraic equation Solvers. *ACM Trans. Math. Softw.*, 31(3):363–396, 2005.

[11] S. Hoops, S. Sahle, R. Gauges, C. Lee, J. Pahle, N. Simus, M. Singhal, L. Xu, P. Mendes, and U. Kummer. COPASI–a COmplex PAthway SImulator. *Bioinformatics*, 22:3067–3074, Dec 2006.

[12] Reiner Horst, Panos M. Pardalos, and Nguyen V. Thoai. *Introduction to global optimization*, volume 48 of *Nonconvex Optimization and its Applications*. Kluwer Academic Publishers, Dordrecht, second edition, 2000.

[13] K. Jaqaman and G. Danuser. Linking data to models: data regression. In A. Heinrichs, E. Kritikou, B. Pulverer, and M. Raftopoulou, editors, *Systems biology: a user's Guide*, pages 27–33. Nature Publishing, New York, Nov 2006.

[14] X. Ji and Y. Xu. libSRES: a C library for stochastic ranking evolution strategy for parameter estimation. *Bioinformatics*, 22:124–126, Jan 2006.

[15] Barbara Kaltenbacher, Andreas Neubauer, and Otmar Scherzer. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, volume 6 of *Radon Series on Computational and Applied Mathematics*. de Gruyter, Berlin - New York, 2008.

[16] Andreas Kremling, Sophia Fischer, Kapil Gadkar, Francis J. Doyle, Thomas Sauter, Eric Bullinger, Frank Allgower, and Ernst D. Gilles. A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. *Genome Res*, 14(9):1773–85, 2004.

[17] P. Kügler. Identification of a temperature dependent heat conductivity by tikhonov regularization. *Journal of Inverse and Ill-posed Problems*, 10:67–90, 2002.

[18] P. Kügler. Identification of a temperature dependent heat conductivity from single boundary measurements. *SIAM J. on Numerical Analysis*, 41:1543–1564, 2004.

[19] F. Lei and S. B. Jorgensen. Estimation of kinetic parameters in a structured yeast model using regularisation. *J Biotechnol*, 88(3):223–37, 2001.

[20] J. Lu, S. Müller, R. Machne, and C. Flamm. SBML Ode Solver library: Extensions for inverse analysis. In *Proceedings of the Fifth International Workshop on Computational Systems Biology, WCSB 2008, Leipzig, Germany*, 2008.

[21] James Lu, Heinz W. Engl, Rainer Machne, and Peter Schuster. Inverse bifurcation analysis of a model for the mammalian G_1/S transition. In *Proceedings of Bioinformatics in Research and Development, Lecture Notes in Bioinformatics*. Springer, 2007.

[22] R. Machné, A. Finney, S. Müller, J. Lu, S. Widder, and C. Flamm. The SBML ODE Solver Library: a native API for symbolic and fast numerical analysis of reaction networks. *Bioinformatics*, 22(11):1406–7, 2006.

[23] Carmen G. Moles, Pedro Mendes, and Julio R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res*, 13(11):2467–74, 2003.

[24] Stefan Müller, Josef Hofbauer, Lukas Endler, Christoph Flamm, Stefanie Widder, and Peter Schuster. A generalized model of the repressilator. *J. Math. Biol.*, 53(6):905–937, 2006.

[25] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer Series in Operations Research and Financial Engineering. Springer, New York, second edition, 2006.

[26] Ronny Ramlau and Gerd Teschke. A Tikhonov-based projection iteration for nonlinear ill-posed problems with sparsity constraints. *Numer. Math.*, 104(2):177–203, 2006.

[27] Maria Rodriguez-Fernandez, Jose A. Egea, and Julio R. Banga. Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics*, 7:483, 2006.

[28] Maria Rodriguez-Fernandez, Pedro Mendes, and Julio R. Banga. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *Biosystems*, 83(2-3):248–65, 2006.

[29] B. Schoeberl, C. Eichler-Jonsson, E. D. Gilles, and G. Müller. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors. *Nat. Biotechnol.*, 20:370–375, Apr 2002.

[30] Z. Szallasi, J. Stelling, and V. Periwal, editors. *System Modeling in Cellular Biology: From Concepts to Nuts and Bolts*. MIT Press, Boston, 2006.

[31] N. A. van Riel and E. D. Sontag. Parameter estimation in models combining signal transduction and metabolic pathways: the dependent input approach. *Syst Biol (Stevenage)*, 153:263–274, Jul 2006.

[32] Andreas Waechter and Lorenz T. Biegler. On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Math. Program.*, 106(1):25–57, 2006.

[33] Stephen Wiggins. *Introduction to applied nonlinear dynamical systems and chaos*, volume 2 of *Texts in Applied Mathematics*. Springer-Verlag, New York, second edition, 2003.

[34] Clemens A. Zarzer. On Tikhonov regularization with non-convex sparsity constraints. submitted.

[35] Z. Zi and E. Klipp. SBML-PET: a Systems Biology Markup Language-based parameter estimation tool. *Bioinformatics*, 22:2704–2705, Nov 2006.

[36] Jason W. Zwolak, John J. Tyson, and Layne T. Watson. Parameter estimation for a mathematical model of the cell cycle in frog eggs. *J Comput Biol*, 12(1):48–63, 2005.