J. Phys. Chem. A **2009,** *113,* 2775–2785

Parameter Identification for Chemical Reaction Systems Using Sparsity Enforcing Regularization: A Case Study for the Chlorite−Iodide Reaction

Philipp Kügler,*,†,§ Erwin Gaubitzer,‡ and Stefan Müller§

Industrial Mathematics Institute, University of Linz, Altenbergerstrasse 69, 4040 Linz, Austria, Max F. Perutz Laboratories, University of Vienna, Dr. Bohr-Gasse 9, 1030 Vienna, Austria, and RICAM, Altenbergerstrasse 69, 4040 Linz, Austria

Received: October 5, 2008; Revised Manuscript Received: December 9, 2008

Complex chemical reactions are commonly described by systems of nonlinear ordinary differential equations. Rate and equilibrium constants of these models are usually not directly accessible and have to be indirectly inferred from experimental observations of the system. As a consequence, parameter identification problems have to be formulated and computationally solved. Because of a limited amount of information and uncertainties in the data, the solutions to such parameter identification problems typically lack uniqueness and stability properties and hence cannot be found in a reliable way by a pure minimization of the data mismatch (i.e., the discrepancy between experimental observations and simulated model output). To overcome these difficulties, so-called regularization methods have to be used. In this article, we suggest a sparsity promoting regularization approach that eliminates unidentifiable model parameters (i.e., parameters of low or no sensitivity to the given data). That way, the model is reduced to a core reaction mechanism with manageable interpretation while still remaining in accordance with the experimental observations. For the computational realization, we utilize the adjoint state technique for an efficient calculation of the gradient of the objective with respect to model parameters as well as uncertain initial and experimental conditions. Illustrations of our approach are given by means of the chlorite−iodide reaction for which reference parameter values are available.

Introduction

The field of nonlinear chemical kinetics has been investigated for about half a century, and yet only a few complex chemical reactions have been described by means of an experimentally backed system of ordinary differential equations. The most well-known is the chlorite−iodide reaction,[1] which involves only two inorganic ions and can be characterized by an established reaction mechanism (LLKE model[2]). Chemical reaction schemes are associated with systems of ordinary differential equations

$$\dot{c}(t) = f(c(t), q) \tag{1}$$

for the concentrations $c(t) = [c_1(t),..., c_m(t)]^T \in [\mathbb{R}_0^+]^m$ that describe the time course of the chemical species. The right-hand side f describes the coupling of the species c_i, which depends on a vector $q \in [\mathbb{R}_0^+]^n$ of n rate and equilibria constants. Given the vector field f, parameter values for q, and initial conditions on c, the time course of c can be numerically found by integration of eq 1, a problem one would refer to as a *direct* or forward problem.

A major challenge in chemical kinetics is that, if at all, only a few rate constants of complex reactions can be directly determined from measurements using the methods of initial rates and/or individual exponential fit of pseudo-first-order kinetic curves.[3] Hence, the unknown components of the parameter vector q have to be indirectly inferred from experimental observations of the system. Such parameter identification problems belong to the class of *inverse* problems, where in general terms one looks for the causes (e.g., parameters) for desired or observed effects (e.g., experimental observations). Inverse problems typically are inherently unstable (i.e., highly sensitive to data errors), such that regularization methods have to be used (see below).

Parameter identification for chemical reaction systems is based on a comparison between experimental data and numerical simulations obtained via eq 1. A common approach, also followed by the software packages ZITA, PARFIT, and PARKIN,[4−6] is to consider the least-squares minimization problem

$$\|\text{data} - \text{simulation}(q)\|^2 \rightarrow \min \tag{2}$$

over a set of admissible parameters q with some suitable norm $\|\cdot\|$ measuring the discrepancy between data and simulation. The underlying premise is that the optimal parameter set is one that gives rise to simulations that match experimental observations as much as possible.

Often, the possibilities for experimental observations of the system are limited, and only a few of the reactants are accessible to measurements. For instance, only iodine is exclusively detectable in the visible spectrum in the context of the chlorite−iodide reaction. As a consequence, the available experimental data often do not contain enough information to uniquely determine the model parameters. Then, solutions obtained by solving eq 2 may lead to inappropriate chemical conclusions, even in the case of simple kinetic systems.[3] An obvious approach to increase the information about the parameters contained in the data is to repeat the experiments under varying conditions. Though some relief can be expected, the problem of nonuniqueness of the solution to eq 2 may still remain.

---

* To whom correspondence should be addressed. E-mail: philipp.kuegler@jku.at.

† University of Linz.

‡ University of Vienna.

§ RICAM.

10.1021/jp808792u CCC: $40.75

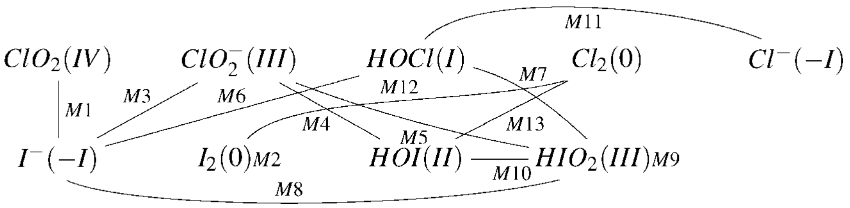


Figure 1. Network of RedOx reactions in the LLKE model.

Another difficulty of eq 2 arises from the fact that its solution, even if it is unique, may not depend continuously on the data. As a consequence of this inherent instability of parameter identification, even small errors in the data (e.g., due to practically inevitable measurement uncertainties) may be significantly amplified and parameter solutions, though able to reproduce the data, may become worthless. The issue of data noise amplification in parameter identification is widely discussed in the mathematical literature[7,8] but seems to be sparely addressed in the literature on chemical reactions. Especially, the pure least-squares approach in eq 2 is highly fault-prone in case of noisy experimental data. To counteract these instabilites, so-called regularization techniques have to be applied,[7] then allowing for the computation of stable approximations to the solution. In context of least-squares formulations, regularization theory suggests to replace the unstable problem in eq 2 by the neighboring minimization problem

$$\|\text{data} - \text{simulation}(q)\|^2 + \alpha \text{ penalty}(q) \rightarrow \min \qquad (3)$$

with strictly positive regularization parameter α, often referred to as variational regularization.[9] For appropriate choices of the penalty term, eq 3 can be shown to be stable such that its minimizer(s) $q(\alpha)$, as opposed to minimizers of eq 2, then depend continuously on the data. Clearly, the method parameter α acts as a balance between accuracy and stability of $q(\alpha)$. For very large α, the influence of the data in eq 3 is nearly negligible. Hence, the solution $q(\alpha)$ is rather insensitive to data perturbations but will lead to large deviations of the simulations from the data. A decrease of α will reduce the misfit between data and simulation but, since getting closer to eq 2, will also make $q(\alpha)$ more sensitive to data errors. Theory-based rules for optimal choices of α depend (besides of the actual penalty term) on available a priori information about the data errors and the data itself.[7]

In this article, we suggest to regularize the parameter identification problem in the presence of uncertain and limited data by using a penalty term in eq 3 that enforces sparse solutions (i.e., parameter vectors with only a few nonzero components). Regularization by sparsity is currently an active field of research with continuously growing results on both theoretical and algorithmic aspects.[10−14] In the context of parameter identification for chemical reaction networks, one advantage of the approach is that unidentifiable parameters are eliminated while the remaining ones can be recovered from the available data in a stable way, then still allowing to reproduce to data by means of a reduced reaction network. With the chlorite−iodide reaction, we have chosen one of the best described chemical reaction networks for the illustration of sparsity enforcing regularization and a validation by comparing our results with those accepted in the literature.

The article is organized as follows. First, we introduce the LLKE model[2] for the chlorite−iodide reaction utilizing the ODE formalism eq 1. Next, we describe the experimental data as well as the techniques and the varying closed system conditions under which they were obtained. The mathematical formulation of the parameter identification problem is the subject of the following section. Therein, we also specify the sparsity enforcing penalty terms that we used. Finally, results of our approach are presented and the interpretation of the sparsity pattern found is given. The numerical solution of the minimization problems in eq 3 by a combination of global and local search techniques is outlined in the Appendix. Especially, we give an explanation of the adjoint state technique for efficient gradient calculations of the objective functional.

TABLE 1: Mechanism of the Chlorine Dioxide/Chlorite−Iodide Reaction According to the LLKE Model[2]

M1	$ClO_2 + I^- \rightarrow ClO_2^- + {}^1/_2 I_2$
M2	$I_2 + H_2O \rightleftharpoons HOI + I^- + H^+$
M3	$HClO_2 + I^- + H^+ \rightarrow HOI + HOCl$
M4	$HClO_2 + HOI \rightarrow HIO_2 + HOCl$
M5	$HClO_2 + HIO_2 \rightarrow IO_3^- + HOCl + H^+$
M6	$HOCl + I^- \rightarrow HOI + Cl^-$
M7	$HOCl + HIO_2 \rightarrow IO_3^- + Cl^- + 2H^+$
M8	$HIO_2 + I^- + H^+ \rightleftharpoons 2HOI$
M9	$2HIO_2 \rightarrow IO_3^- + HOI + H^+$
M10	$HIO_2 + H_2OI^+ \rightarrow IO_3^- + I^- + 3H^+$
M11	$HOCl + Cl^- + H^+ \rightleftharpoons Cl_2 + H_2O$
M12	$Cl_2 + I_2 + 2H_2O \rightarrow 2HOI + 2Cl^- + 2H^+$
M13	$Cl_2 + HOI + H_2O \rightarrow HIO_2 + 2Cl^- + 2H^+$
M14	$HClO_2 \rightleftharpoons ClO_2^- + H^+$
M15	$H_2OI^+ \rightleftharpoons HOI + H^+$
M16	$I_2 + I^- \rightleftharpoons I_3^-$

LLKE Model of the Chlorite−Iodide Reaction

To illustrate the sparsity promoting regularization approach to parameter identification in chemical reaction systems, we chose the chlorite−iodide reaction network for which a detailed ODE model is available and commonly accepted (LLKE model[2]). Under certain conditions, the chlorite−iodide reaction can show unusual dynamic behavior. When the two inorganic ions chlorite ClO_2^- and iodide I^- are initially mixed, the color of the solution gets more and more yellow but then the solution suddenly becomes colorless again. This corresponds to a monotonic increase of the iodine concentration followed by a sudden drop of concentration. This chemical phenomenon is referred to as clock-type behavior, and the moment of concentration drop is called clock time. Examples of this iodine clock behavior are shown in Figure 2, where the differences in clock time and peak height are due to different initial mixings and pH values of the buffer solution.

The chemical reactions involved in the LLKE model are listed in Table 1; the network of RedOx reactions is illustrated in Figure 1. The edges of the network connect particular RedOx pairs and are labeled with the corresponding reaction number. Labels attached to nodes of the network (i.e., M2 and M9) represent disproportion reactions.

The ODE system resulting from the LLKE model[2] of the chlorine dioxide/chlorite−iodide reaction involves the concentrations $T_{I^-} = [I^-] + [I_3^-]$, $T_{I_2} = [I_2] + [I_3^-]$, $T_{HOI} = [HOI] + [H_2OI^+]$, $[HIO_2]$, $[ClO_2]$, $T_{ClO_2^-} = [ClO_2^-] + [HClO_2]$, $[HOCl]$, $[Cl_2]$, and $[Cl^-]$. The fast equilibrium assumption for reactions

M14, M15, and M16 allows to reduce the dimension of the original reaction system. More specifically, one can introduce the total concentrations T to eliminate the corresponding individual concentrations. Of course, the rate equations that involve the individual concentrations $[I^-]$, $[I_2]$, $[H_2OI^+]$, $[HOI]$, and $[HClO_2]$ have to be expressed by the total concentrations T_{I^-}, T_{I_2}, T_{HOI}, and $T_{ClO_2^-}$, see below.

Introducing the variables $q \in [\mathbb{R}_0{}^+]^{20}$, $c \in [\mathbb{R}_0{}^+]^9$, and $u \in \mathbb{R}^+$ (as listed in Table 2), the reaction mechanism of the LLKE model[2] (as listed in Table 1) involves the following reaction rates v_i

$$
\begin{aligned}
v_1 &= q_1 \cdot c_5 \cdot [I^-]\\
v_{2a} &= q_2 \cdot q_3 \cdot \frac{[I_2]}{u} - q_3 \cdot [HOI] \cdot [I^-]\\
v_{2b} &= \frac{q_2}{q_{15}} \cdot q_4 \cdot [I_2] - q_4 \cdot [H_2OI^+] \cdot [I^-]\\
v_3 &= q_5 \cdot [HClO_2] \cdot [I^-]\\
v_4 &= q_6 \cdot [HClO_2] \cdot [HOI]\\
v_5 &= q_7 \cdot [HClO_2] \cdot c_4\\
v_6 &= q_8 \cdot c_7 \cdot [I^-]\\
v_7 &= q_9 \cdot c_7 \cdot c_4\\
v_8 &= q_{10} \cdot q_{11} \cdot c_4 \cdot [I^-] \cdot u - q_{11} \cdot [HOI]^2\\
v_9 &= q_{12} \cdot c_4^2\\
v_{10} &= q_{13} \cdot c_4 \cdot [H_2OI^+]\\
v_{11} &= q_{17} \cdot q_{18} \cdot c_7 \cdot c_9 \cdot u - q_{18} \cdot c_8\\
v_{12} &= q_{19} \cdot c_8 \cdot [I_2]\\
v_{13} &= q_{20} \cdot c_8 \cdot [HOI]
\end{aligned}
\tag{4}
$$

with

$$
\begin{aligned}
[I^-] &= c_1 - \frac{1}{2} \cdot (q_{16} + c_1 + c_2) - \sqrt{\frac{1}{4} \cdot (q_{16} + c_1 + c_2)^2 - c_1 \cdot c_2}\\
[I_2] &= c_2 - \frac{1}{2} \cdot (q_{16} + c_1 + c_2) - \sqrt{\frac{1}{4} \cdot (q_{16} + c_1 + c_2)^2 - c_1 \cdot c_2}\\
[H_2OI^+] &= c_3 \cdot \frac{u}{q_{15} + u}\\
[HOI] &= c_3 \cdot \frac{q_{15}}{q_{15} + u}\\
[HClO_2] &= c_6 \cdot \frac{u}{q_{14} + u}
\end{aligned}
\tag{5}
$$

In particular, the reaction rates of the LLKE model[2] involve 20 parameters q_i, and their values and dimensions are shown for reference in Table 3. In case of closed reactor conditions, the change of the concentrations with respect to time t is described by the nonlinear system of ordinary differential equations

$$
\begin{aligned}
\dot{c}_1 &= -v_1 + v_{2a} + v_{2b} - v_3 - v_6 - v_8 + v_{10}\\
\dot{c}_2 &= \frac{1}{2} v_1 - v_{2a} - v_{2b} - v_{12}\\
\dot{c}_3 &= v_{2a} + v_{2b} + v_3 - v_4 + v_6 + 2v_8 + v_9 - v_{10} + 2v_{12} - v_{13}\\
\dot{c}_4 &= v_4 - v_5 - v_7 - v_8 - 2v_9 - v_{10} + v_{13}\\
\dot{c}_5 &= -v_1\\
\dot{c}_6 &= v_1 - v_3 - v_4 - v_5\\
\dot{c}_7 &= v_3 + v_4 + v_5 - v_6 - v_7 - v_{11}\\
\dot{c}_8 &= v_{11} - v_{12} - v_{13}\\
\dot{c}_9 &= v_6 + v_7 - v_{11} + 2v_{12} + 2v_{13}
\end{aligned}
\tag{6}
$$

to be complemented with initial values for the concentrations. In the following, this ODE system is compactly written as

$$
\begin{aligned}
\dot{c}(t) &= f(c(t), q, u)\\
c(0) &= c_0
\end{aligned}
\tag{7}
$$

with the vector $c_0 \in \mathbb{R}_0^{+9}$ of initial concentrations. Therein, the continuous function

$$f: \mathbb{R}_0^{+9} \times \mathbb{R}_0^{+20} \times \mathbb{R}^+ \rightarrow \mathbb{R}^9$$

represents the right-hand sides of eq 6. For example,

$$
\begin{aligned}
f_8(c, q, u) = {} & q_{17} \cdot q_{18} \cdot c_7 \cdot c_9 \cdot u - q_{18} \cdot c_8\\
& - q_{19} \cdot c_8 \cdot \left(c_2 - \frac{1}{2} \cdot (q_{16} + c_1 + c_2) \right.\\
& \left. - \sqrt{\frac{1}{4} \cdot (q_{16} + c_1 + c_2)^2 - c_1 \cdot c_2}\right)\\
& - q_{20} \cdot c_8 \cdot c_3 \cdot \frac{q_{15}}{q_{15} + u}
\end{aligned}
$$

With the concentrations c_i measured in M (= mol/L), the dimensions of of the model parameters in eq 7 are non uniform (see Table 3 for details). Since in addition the values of c and q may vary within wide ranges, it is advisable to make the model dimensionless and normalize all variables before doing any computations. To this end, all variables are scaled with typical values (e.g., $c_2 = c_2^{\text{typical}} c_2'$, $q_{13} = q_{13}^{\text{typical}} q_{13}'$) and eq 7 is transformed into a problem for and depending on dimensionless (primed) variables. However, for the sake of simplicity, we stay with the notation of eq 7 and consider it to be dimensionless whenever the focus of our discussion is on computational issues.

Experimental Setup and Data

In this section, we describe the iodine concentration data used to indirectly infer the parameters of the LLKE model. To this end, we considered different initial mixings of chlorite and iodide in solutions of various buffered pH values. The experiments were performed using only one buffer substance to limit the influence of different associations with it.[15] To provide a feature rich data set, we only selected those kinetic data curves showing clock reaction behavior.

Materials and Spectrometry. All chemicals were obtained commercially, were of highest purity available, and were used without further purification. For the preparation of the experimental solutions purity for potassium iodide was assumed to be 100% and that of sodium chlorite 80% according to the specifications of the producer. Solutions were prepared fresh and used for two days at the most. Sodium chlorite solutions

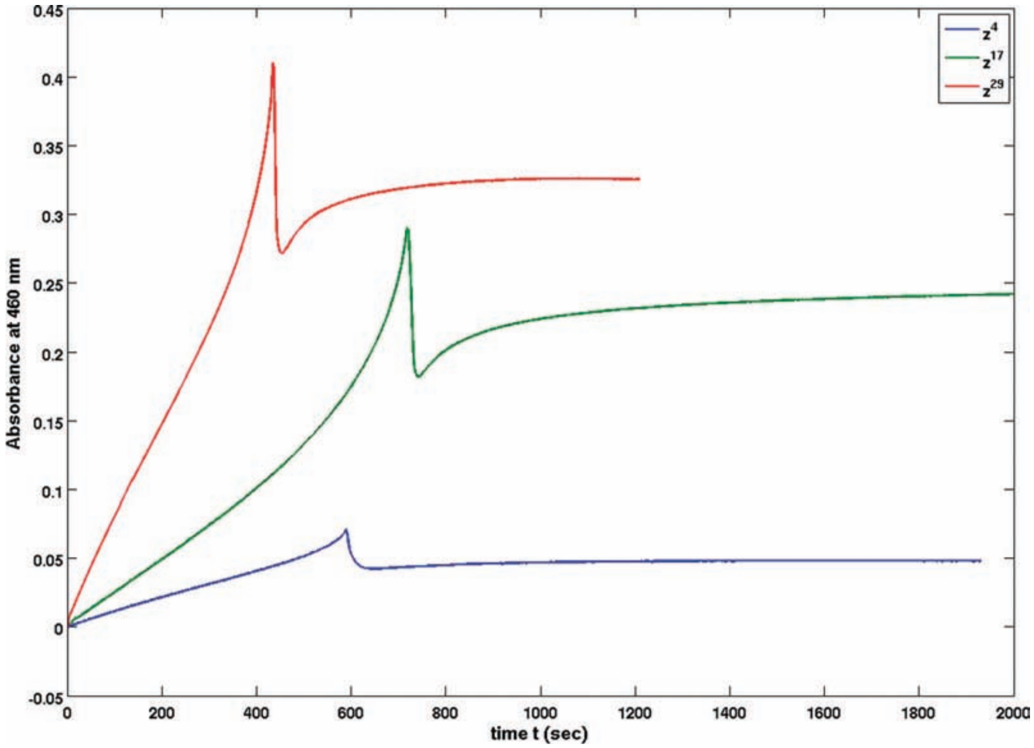


Figure 2. Observation of clock reaction under different conditions.

TABLE 2: Mapping of Model Parameters and Concentrations

$q_1 = k_1$	$q_{11} = k_{8r}$	$c_1 = T_{I^-}$
$q_2 = K_{2a}$	$q_{12} = k_9$	$c_2 = T_{I_2}$
$q_3 = k_{2ar}$	$q_{13} = k_{10}$	$c_3 = T_{HOI}$
$q_4 = k_{2br}$	$q_{14} = K_{14}$	$c_4 = [HIO_2]$
$q_5 = k_3$	$q_{15} = K_{15}$	$c_5 = [ClO_2]$
$q_6 = k_4$	$q_{16} = K_{16}$	$c_6 = T_{ClO_2^-}$
$q_7 = k_5$	$q_{17} = K_{11}$	$c_7 = [HOCl]$
$q_8 = k_6$	$q_{18} = k_{11r}$	$c_8 = [Cl_2]$
$q_9 = k_7$	$q_{19} = k_{12}$	$c_9 = [Cl^-]$
$q_{10} = K_8$	$q_{20} = k_{13}$	$u = [H^+]$

were prepared in 1 mM sodium hydroxide solution for stabilization. Potassium iodide was prepared in water. HSO_4^- was used as the buffer substance with an acidity constant of 1.96. Buffer and $NaClO_4$ solutions were prepared separately. $NaClO_4$ served to adjust the ionic strength to 0.3 M.

Precaution was taken to minimize the influence of incomplete mixing.[2] The solutions for the experiments were added in the following order: buffer solution, $NaClO_4$ solution, KI solution, $NaClO_2$ solution. Furthermore, the reactive compounds were prepared in such concentrations that the last mixing step involved equal volumes. That way, desired initial concentration values of chlorite and iodide could be achieved while preventing reactions caused by local excess concentrations.

The desktop absorption spectrometer Hitachi U-3300 was used to obtain absorbance values. The total iodine concentration was followed at the isosbestic point of iodine and triiodide at 460 nm. The absorbance coefficient was chosen[16] as 770 mol^{-1} cm^2. In terms of the notation introduced in Table 2, the iodine absorbance is represented as

$$770 \cdot c_2 \tag{8}$$

Experimental Data. We performed $N_e = 31$ experiments with variations in the initial concentrations of I^- and ClO_2^-, the H^+ concentration of the buffer solution, and the time interval $[0, T]$ of the observation. Labeling each experiment by the index j, we denote the absorbance data collected by

$$z^j, \ j = 1,\ldots, N_e$$

while the proper model for the j-th experiment becomes

$$\begin{aligned} \dot{c}(t) &= f(c, q, u^j) \text{ on } (0, T^j] \\ c(0) &= c_0^j \end{aligned} \tag{9}$$

with initial concentration vector

$$c_0^j = (c_{0,1}^j, 0, 0, 0, 0, c_{0,6}^j, 0, 0, 0)^T \tag{10}$$

compared with Table 2.

The chosen values of the experimental setup parameters

$$\bar{c}_{0,1}^j, \bar{c}_{0,6}^j, \bar{u}^j, j = 1,\ldots, N_e \tag{11}$$

are summarized in Table 5. For the initial concentrations, the error tolerances are ±5% (relative); for the pH values, linked to $[H^+]$ via $pH = -\log_{10}([H^+])$, we have ±0.1 (absolute). Hence, the actual values of $c_{0,1}^j$, $c_{0,6}^j$, and u^j are supposed to satisfy

$$\begin{aligned} c_{0,1}^j &\in [\bar{c}_{0,1}^j - 0.05\bar{c}_{0,1}^j, \bar{c}_{0,1}^j + 0.05\bar{c}_{0,1}^j] \\ c_{0,6}^j &\in [\bar{c}_{0,6}^j - 0.05\bar{c}_{0,6}^j, \bar{c}_{0,6}^j + 0.05\bar{c}_{0,6}^j] \\ u^j &\in [10^{-\bar{u}^j - 0.1}, 10^{-\bar{u}^j + 0.1}] \end{aligned} \tag{12}$$

With respect to the discussion of the inverse problem, we define the set

$$C = \{(c_{0,1}^1, c_{0,6}^1, u^1,\ldots, c_{0,1}^{N_e} c_{0,6}^{N_e}, u^{N_e}) | (12) \text{ is satisfied}\} \subset \mathbb{R}^{3N_e} \tag{13}$$

of possible setups.

Figure 2 shows the absorption data z^4, z^{17}, z^{29} corresponding to the experimental setups (eq 11) for $j = 4, 17, 29$ and observation times $T^4 = 1931$, $T^{17} = 2000$, $T^{29} = 1209$ (in seconds). Though the signal z^j is only recorded at discrete time steps between 0 and T^j, we will consider the data as continuous functions

$$z^j \in C([0, T^j]),\ j = 1,..., N_e \tag{14}$$

due to the high sampling rate. However, the following discussion could easily be repeated in case of time discrete data z^j.

Inverse Problem

Given a sequence of experimental data, the inverse problem of our interest is to identify the unknown model parameters for the underlying reaction mechanism. As outlined in our introduction, parameter identification is typically ill-posed and regularization methods have to be used to avoid misleading results due to instability and nonuniqueness of the solution. In this article, we suggest variational regularization using sparsity enforcing penalization and illustrate this approach by means of the chlorite−iodide reaction. The available experimental data have already been introduced in the previous section. In the following, we specifiy the simulation term in eq 3 and the measure for the misfit between data and simulation, followed by a discussion of the penalty term.

Simulation Operator and Data Mismatch. A numerical simulation of the LLKE model (eq 7) requires the assignment of values to all $n = 20$ components of the parameter vector q. In addition, values for the initial concentrations $c_{0,1}$, $c_{0,6}$, the H^+ concentration u, and the final integration time T have to be provided. In the following, we refer to the mapping that takes q, $c_{0,1}$, $c_{0,6}$, u, T onto the solution $c \in C([0, T], \mathbb{R}_0^{+9})$ of the ODE system (eq 7) as solution operator:

$$S: (q, c_{0,1}, c_{0,6}, u, T) \rightarrow c \tag{15}$$

$C([0, T], \mathbb{R}_0^{+9})$ denotes the space of all continuous functions $f:[0, T] \rightarrow \mathbb{R}_0^{+9}$; especially, $f(t) \in \mathbb{R}_0^{+9}$ for all $t \in [0, T]$. Note that S is nonlinear in its arguments and only implicitly defined via eq 7.

Having evaluated S, the iodine absorbance (eq 8) can be simulated by extraction of the second component of c and multiplication by the absorption constant 770. Defining the linear absorbance mapping

$$A: c \rightarrow 770 \cdot c_2$$

we finally can introduce the simulation operator

$$F = A \circ S: (q, c_{0,1}, c_{0,6}, u, T) \rightarrow 770 \cdot c_2 \tag{16}$$

as concatenation of S and A.

Given an estimate q for the unknown model parameters, the simulation of the data z^j of the j-th experiment is described by $F(q, c^j_{0,1}, c^j_{0,6}, u^j, T^j)$ or equivalently by $F(x^j, T^j)$ using the notation

$$x^j = (q, c^j_{0,1}, c^j_{0,6}, u^j) \in \mathbb{R}^{n+3} \tag{17}$$

While the final observation time T^j can be considered as precisely known, the values $c^j_{0,1}$, $c^j_{0,6}$, and u^j must be allowed to lie within a tolerance neighborhood of the measured values (eq 12). To measure the deviation between the data z^j and the simulation $F(x^j, T^j)$ of the j-th experiment, we use the square of the norm $||\cdot||_j$ on $L^2([0, T^j])$

$$||z^j - F(x^j, T^j)||_j^2 = \int_0^{T^j} (z^j(t) - 770 \cdot c^j_2(t))^2 \, dt \tag{18}$$

The space $L^2([0, T])$ contains all functions $f:[0, T] \rightarrow \mathbb{R}$ with $\int_0^T f(t)^2 \, dt < \infty$; all continuous functions in $C([0, T])$ belong to $L^2([0, T])$. The misfit (eq 18) of the j-th experiment can be influenced by altering the $n + 3$ variables $x^j = (q, c^j_{0,1}, c^j_{0,6}, u^j)$. To measure the total misfit over all N_e experiments, we use the weighted sum

$$\sum_{j=1}^{N_e} \omega_j \int_0^{T^j} (z^j(t) - 770 \cdot c^j_2(t))^2 \, dt \tag{19}$$

with scalar weights

$$\omega_j = \frac{1}{||z^j||_j^2} \tag{20}$$

Now, the total misfit (eq 19) can be influenced by alterations in the $N = n + 3N_e$ variables

$$x = (q, c^1_{0,1}, c^1_{0,6}, u^1, ..., c^{N_e}_{0,1}, c^{N_e}_{0,6}, u^{N_e}) \in \mathbb{R}^N \tag{21}$$

Our main interest is to identify the model parameters q belonging to a set of admissible parameters $Q \subset \mathbb{R}^{20}$ (Table 3). Still, we also have to take the uncertainties of the experimental setup parameters (eq 11) into account. Hence, if the total search space is defined as $X = Q \times C$ (eq 13), the unregularized parameter identification problem is given by

$$J_0(x) = \sum_{j=1}^{N_e} \omega^j ||z^j - F(x^j, T^j)||_j^2 \rightarrow \min_{x \in X \subset \mathbb{R}^N} \tag{22}$$

Sparsity Enforcing Penalty Term. As outlined in the Introduction, simply minimizing the pure least-squares functional (eq 19) is not advisable because of the ill-posedness of parameter identification. Instead, regularization techniques should be used to allow the computation of stable approximate solutions. The most popular approach for nonlinear problems is variational regularization (eq 3) where stability is enforced by expansion of the least-squares objective functional with an additive penalty term. Usually, the full argument, in our case x, is penalized. However, with narrow box constraints available for the experimental setup parameters and the equilibrium constants (eq 12 and Table 4), we only consider the penalization of the rate constants and to this end introduce the index set

$$I_R = \{1, 3, 4, 5, 6, 7, 8, 9, 11, 12, 13, 18, 19, 20\} \tag{23}$$

It contains those indices $i \in \{1,..., 20\}$ for which q_i represents a reaction rate (lower case k) but no equilibrium constant (upper case K) (compare with Table 2). A common choice for the penalty term in

$$J_\alpha(x) = J_0(x) + \alpha \text{ penalty } (q) \rightarrow \min_{x \in X \subset \mathbb{R}^N} \tag{24}$$

is the square of the Euclidean norm, which in our case is written as

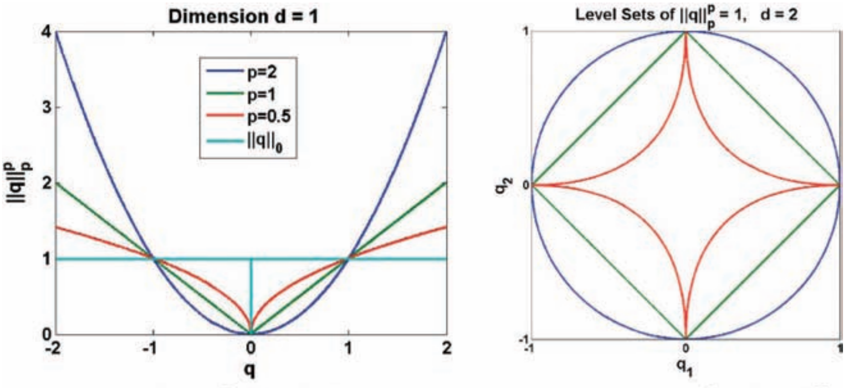
$$\text{penalty } (q) = \sum_{i \in I_R} q_j^2 \tag{25}$$

That way, eq 24 would amount to Tikhonov regularization.[7]

However, our goal is to enforce sparse solutions, that is, we look for model parameters such that

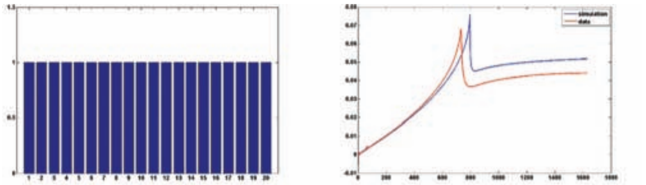
$$q_i = 0 \tag{26}$$

holds for as many indices $i \in I_R$ as possible while still allowing to explain the experimental data. The idea is to eliminate unidentifiable parameters while recovering the relevant ones in a stable manner. Note that any rate constant equal to zero leads to an elimination of the corresponding reaction given in eq 4. Hence, instead of a model containing all possible reaction pathways we strive for a reduced network still compatible with the experimental observations.



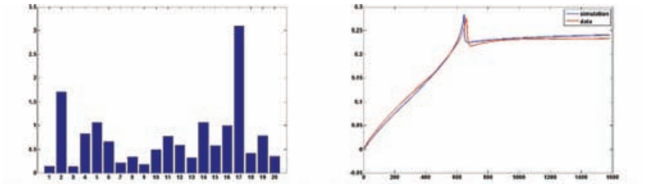
(a) $\|q\|_p^p$ vs. q in $\mathbb{R}$ (b) level set $\|q\|_p^p = 1$ in $\mathbb{R}^2$

Figure 3. Sparsity effect of the value of p in eq 28.



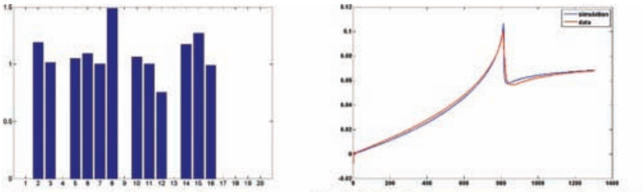
(a) literature values q^* normalized to 1 (b) data z^1 and corresponding simulation $F(q^*, \bar{c}^1_{0,1}, \bar{c}^1_{0,6}, \bar{u}^1, T^1)$

Figure 4. Simulation.



(a) solution q relative to literature values q^* (b) 3.5% deviation between data and simulation for $j = 12$

Figure 5. Result without regularization.



(a) sparse solution q relative to q^* (b) 4.1% deviation between data and simulation for $j = 27$

Figure 6. Result with sparsity enforcing regularization.

A first choice for a penalty term enforcing sparse solutions is

$$\|q\|_0 := \text{number of nonzero elements } q_i \text{ with } i \in I_R \qquad (27)$$

Though eq 27 does not satisfy the properties of a norm, this approach is called l_0-norm penalization. Since $\|\cdot\|_0$ is not even a continuous function, the computational costs for the minimization of objective functionals with l_0 penalty terms are high, in particular because of the use of combinatorial techniques. Hence, eq 27 is usually replaced by the p-th power of the l_p-norm

$$\|q\|_p := \left(\sum_{i \in I_R} |q_i|^p\right)^{1/p}$$

with some $p \in (0, 2]$ (the norm properties actually are only satisfied for $1 \leq p \leq 2$) such that the penalty term becomes

$$\text{penalty}(q) = \|q\|_p^p \qquad (28)$$

For $p = 2$, eq 28 coincides with eq 25, but then nonzero parameter values are only weakly penalized in the neighborhood of the origin. The smaller one chooses p the closer is the approximation of eq 27 by eq 28 and the stronger is the enforcement of sparsity. The effect of p is illustrated in Figure 3 for the one- and two-dimensional case. Note that (independent of the dimension of the parameter vector) convexity of the penalty functional is only given for $p \geq 1$. However, with respect to the numerical realization this does not pose additional difficulties since already the nonlinear least-squares functional J_0 typically is nonconvex.

In addition to the classical theory for Tikhonov regularization[7] with $p = 2$, a continuously increasing number of results on regularization properties for the case $p \in [1, 2]$ can be found in the literature (see the references given in the Introduction). Recently, it has been shown[17,18] that also $p < 1$ renders a regularization method. Hence, also for $p < 1$, the minimizers of

$$J_\alpha(x) = J_0(x) + \alpha\|q\|_p^p \rightarrow \min_{x \in X \subset \mathbb{R}^N} \qquad (29)$$

depend continuously on the data guaranteeing robustness of the approach despite the presence of data noise.

There is another motivation for the use of sparsity promoting penalization. Assume that model parameters $\tilde{q}$ are given from some trusted source (e.g., the values of the LLKE model[2] as listed in Table 3). However, because of the general imperfectness of mathematical models as images of real processes, it might still happen that experiments independently conducted cannot or only unsatisfactorily can be explained on the basis of $\tilde{q}$. Then, one might wish to only alter as few components of $\tilde{q}$ as possible while fitting the data. This can be achieved by enforcing sparsity of the difference between $\tilde{q}$ and that sought for solution. If we keep our focus on the index set I_R, the appropriate penalty term for enhancement of J_0 then becomes

$$\|q - \tilde{q}\|_p^p \qquad (30)$$

In the next section, we present our results obtained with both eqs 28 and 30, and the numerical realization of the approach (eq 29) is explained in the Appendix. We have run several tests with different choices of p and observed that $p < 1$ enforced sparser solutions than $p = 1$. In fact, a result according to which the l_1-norm solution actually is the sparsest solution is only available in the linear case and hence does not apply to the nonlinear problem at hand.[19]

The approach outlined in the article is suitable whenever internal noise due to random birth and death events of chemical species can be ignored and a deterministic ODE model (eq 1) is appropriate. Especially, the experimental system for the chlorite−iodide reaction studied in this article contained about 10^{16} reaction molecules. However, given a small-scale chemical reaction network where the effects of internal noise cannot be neglected, eq 1 is inappropriate and needs to be replaced by stochastic models. Similarly, parameter identification then needs to be treated as a statistical inverse problem (e.g., using so-called filtering techniques).[8,20]

Finally, we address the issue of experimental systems with drift. In our experiments, drift was avoided by the continuous calibration in the two-beam spectrophotometer. However, given a system with drift the resulting error can be counteracted if the drift can be reliably determined by accompanying measurements that allow to correspondingly adapt the observation operator A in eq 16.

Results

First, we ran a pure simulation of the $N_e = 31$ experiments using the reference parameter values q^* for the LLKE model[2] (Table 3) and the experimental conditions $\bar{c}_{0,1}^j$, $\bar{c}_{0,6}^j$, $\bar{u}^j$ listed in Table 5. Hence, no optimization was performed but the data z^j were only compared to the simulated outputs $F(q^*, \bar{c}_{0,1}^j, \bar{c}_{0,6}^j, \bar{u}^j, T^j)$. The corresponding value of the least-square objective J_0 is

$$\sum_{j=1}^{31} \omega^j \| z^j - F(q^*, \bar{c}_{0,1}^j, \bar{c}_{0,6}^j, \bar{u}^j, T^j) \|_j^2 = 0.4619 \qquad (31)$$

and the individual relative errors in the output

$$w^j \int_0^{T^j} (z^j(t) - c_2^j(t))^2 \, \mathrm{d}t \qquad (32)$$

reach from 6.14% for $j = 31$ to 20.19% for $j = 1$ with a mean of 11.64%. As will be shown below, the quality of the data fit based on Tables 3 and 5 is rather unsatisfactory. Figure 4 shows the experiment $j = 1$ with largest output error and, for later reference, the values of q^* normalized to 1.

In a first attempt to improve the quality of the data fit, we performed the minimization of J_0 only within the parameter search space $X = Q$ but kept the experimental conditions fixed with the values given in Table 5. Though the reference value in eq 31 could be slightly reduced that way, some large individual misfits remained. On the other hand, freezing the parameters to the values of Table 3 and minimizing J_0 only with respect to the setup variables $c_{0,1}^j$, $c_{0,6}^j$, u^j within $X = C$ led to acceptable fits of all data with objective value 0.0218. However, at the solution obtained many of the lower and upper bounds (eq 12) were active such that we rejected that approach and finally focused on the search space $X = Q \times C$.

The first computations over the $n + 3N_e = 113$-dimensional search space $X = Q \times C$ were concerned with the minimization problem as stated in eq 22 (i.e., only J_0 was minimized and regularization was not used). Not surprisingly, the larger degree of freedom allowed a much better fit of the data compared to the approaches mentioned above without hitting the boundaries of the search space. Figure 5 shows the experiment $j = 12$ with largest output error (and the identified model parameters). At the solution x to eq 22, the relative output errors (eq 32) range from 1.3 to 3.5% with an average value of 2.1%, and the objective functional assumes the value

$$J_0(x) = 0.0139 \qquad (33)$$

significantly below that given in eq 31. However, the model parameters identified (part of the solution x according to eq 21) strongly deviate from the literature values q^* (compare Figures 5a and 4a). Repeated independent minimizations of the unregularized objective (eq 22) led to solutions q with similar data fitting capabilities but large differences to q^* and also among themselves. These numerical examples underline the need for regularizing the parameter identification problem.

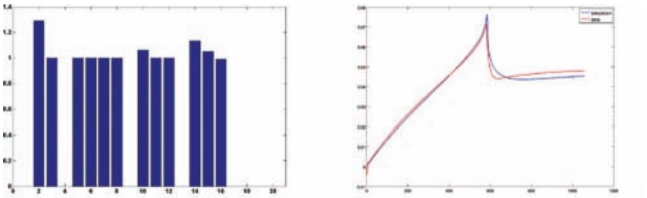

(a) sparse solution q relative to q^* (b) 5.1% deviation between data and simulation for $j = 5$

Figure 7. Result with sparsity penalization of $q - q^*$ using the pattern I_0.

To obtain more stable results, we then turned to the regularization approach (eq 29). For the actual minimization, we chose $\alpha = 0.1$ as well as the sparsity promoting penalty term (eq 28) with $p = 0.1$ and took advantage of a hierarchical strategy (see the Appendix). The corresponding result is illustrated in Figure 6. The relative output errors (eq 32) ranged from 1.4 to 4.1% with an average value 2.4% and J_0 value 0.0189, such that the quality of the data fit was only slightly worse compared to that of eq 33. However, because of the sparsity constraint the data now could be explained by a parameter vector $q \in \mathbb{R}^{20}$ with $q_i = 0$ for $i \in I_0$ with

$$I_0 = \{1, 4, 9, 13, 17, 18, 19, 20\} \qquad (34)$$

representing model parameters of no or low influence for reproducing the data. The nonzero values of q are much closer to the corresponding values of q^* than without regularization (Figure 5a). Independent reruns of the computations led to identical sparsity patterns and, as opposed to the unregularized situation, to significantly less fluctuation of the nonzero parameter values. This numerically demonstrates the stabilizing benefits of the penalty term and its capability to eliminate unidentifiable parameters. Therein, the sparsity structure implies a simplified reaction network that does not violate any chemical intuition but has a sound interpretation (see the Conclusions for details).

Finally, to examine the trustability of the reduced network structure suggested by eq 34, we set all q_i with $i \in I_0$ equal to zero and tested if our data could be explained by the remaining nonzero parameters q_i with $i \in I_X = \{2, 3, 5, 6, 7, 8, 10, 11, 12, 14, 15, 16\}$ in the vicinity of the corresponding reference values q_i^*, $i \in I_X$, of Table 3. Hence, we considered the minimization problem

$$J_\alpha(x) = J_0(x) + \alpha \| q - q^* \|_p^p \rightarrow \min_{x \in X}$$

with $\alpha = 0.1$, $p = 0.1$, and the penalty term only active over the index set I_X. The result, with relative output errors (eq 32) ranging from 1.6 to 5.1% with average value 2.6% and J_0 value 0.0245, is shown in Figure 7. As expected, a sparsity penalization of the difference between q and q^* forces the (nonzero) parameter values to stay close to the reference values. For the index set I_X, significant differences between q and q^* only exist at the indices $i = 2,10,14,15$ while still a satisfactory explanation of the experimental observations is possible.

Conclusions

In this article, we addressed the ill-posedness of parameter identification in nonlinear ODE systems describing chemical reaction networks and the need for regularization of the problem to computationally obtain stable solutions. For the common case of multiple solutions due to limited information content of available data, we suggested the use of variational regularization with sparsity enforcing penalty term. Among several solutions able to explain the experimental observations, a sparse solution represents a mathematical model that concentrates on the core reactions involved in the process. Parameters unidentifiable from the available data are eliminated, and reaction pathways of low or no impact on the model output are erased. This is also in accordance with Ockham's razor striving for simple models facilitating their manageability in the context of model-based applications beyond the scope of time-course simulations such as numerical bifurcation analysis, reaction control, or experimental design.

The sparsity pattern (eq 34) obtained for the LLKE model can be interpreted as follows (see also eq 4):

$\boldsymbol{q_1 = 0 \Rightarrow v_1 = 0}$. Reaction M1 represents the one-way conversion of chlorine dioxide to chlorite. Given an initial chlorine dioxide concentration equal to zero, that is, $c_5(0) = 0$ for the ODE model (eq 6), it is obvious that M1 actually does not take place, that is, $c_5(t) = 0$. The purpose of still including the reaction in the ODE model was to provide an indisputable test case for our sparsity approach. In all our computations with sparsity enforcing penalization the irrelevance of M1 was detected.

$\boldsymbol{q_4 = 0 \Rightarrow v_{2b} = 0}$ **and** $\boldsymbol{q_{13} = 0 \Rightarrow v_{10} = 0}$**.** Reactions M2b and M10 describe the generation and consumption of H_2OI^+, where M2b is an alternative to M2a. Our result $q_4 = 0$ indicates that the experimental data under consideration can still be explained if M2b is excluded from the reaction network and only M2a is used. It can be argued that the range of pH values was not wide enough to confirm the existence of M2b.

$\boldsymbol{q_9 = 0 \Rightarrow v_7 = 0}$. The rate constants in the LLKE model[2] are directly inferred from experiments. Our investigations show agreement for most of the needed core reactions. However, reaction M7 is found to be negligible by our sparsity approach. Given our set of experiments within a slightly smaller range of pH values compared to the LLKE model,[2] reaction M7 does not necessarily belong to the core reactions. In contrast to the chlorine-related differences (see below), this has no support by other recent publications.

$\boldsymbol{q_{18} = 0 \Rightarrow v_{11} = 0}$. Reaction M11 represents the generation of chlorine. In the LLKE model,[2] M11 is important only at $[H^+] > 0.01$ mol/L and $[HClO_2] > 10^{-3}$ mol/L. While some experiments were conducted with $[H^+] > 0.01$ mol/L, $[HClO_2]$ never exceeded its critical value because $[HClO_2] < [ClO_2]_0 < 10^{-3}$ mol/L (Table 5). The irrelevance of M11 under these conditions was detected by our sparsity approach, which led to $q_{18} = 0$.

$\boldsymbol{q_{17} = 0}$. The parameter q_{17} only occurs in the reaction rate v_{11} associated with reaction M11. Since q_{18} is zero, the value of q_{17} becomes irrelevant. Hence, the corresponding equilibrium constant cannot be identified on the basis of the data set considered and is set to zero.

$\boldsymbol{q_{19} = 0 \Rightarrow v_{12} = 0}$ **and** $\boldsymbol{q_{20} = 0 \Rightarrow v_{13} = 0}$. Reactions M12 and M13 consume chlorine and hence depend on the conversion rate v_{11} associated with M11. Since M11 was found to be unimportant, there is not sufficient educt for the reactions M12 and M13 such that their rates v_{12} and v_{13} become negligible.

Future work is devoted to investigate if the reduced model based on eq 34 is capable of reproducing the rich repertoire of dynamical features of the chlorine dioxide/chlorite−iodide reaction such as oscillations and bistability.[2,21]

Acknowledgment. This work was supported by the project MA05 of the Viennese Science and Technology Fund WWTF. We thank the members of the project team for fruitful discussions.

Appendix

Variational regularization techniques typically lead to minimization problems that may also be approached by gradient-based routines. The adjoint method for deriving an exact formula of the gradient, also with respect to varying experimental conditions, is exemplified by means of the LLKE model for the chlorite−iodide reaction. From the viewpoint of numerical realization, this approach avoids error-prone finite differences and unnecessary computational burden.

Evaluation of the Objective Functional

Clearly, any optimization technique for solving the minimization problem (eq 29) requires the evaluation of the objective functional J at the current iterate x. With a penalty term as in eq 28, the main effort in the evaluation of eq 29 is the computation of $J_0(x)$. Given the definition (eq 16) of the forward operator, the first step is to solve the ODE system (eq 9) N_e times with q and the experimental conditions $s^j = (c^j_{0,1}, c^j_{0,6}, u^j)$, $j = 1,..., N_e$, part of the current argument x according to eq 21. Second, the N_e time integrals of eq 19 have to be computed for which we used the MATLAB code *trapz*.

Gradient Calculation of the Objective

In our examples, the minimization problem (eq 29) was solved by a combined use of global and local optimization techniques. Global searches of the admissible set X were performed by the MATLAB routines *ga*, *patternsearch* (implementations of genetic and pattern search algorithms) to deliver suitable candidates for subsequent local searches. For the latter, the MATLAB code *fmincon*, a realization of the sequential quadratic programming method, was utilized. While *ga* and *patternsearch* only involve the evaluation of the objective J (as outlined above) for different arguments $x \in X$, *fmincon* in addition requires the calculation of the gradient of J.

Relaxation of the Penalty Term

With respect to gradient-based optimization, the nondifferentiability of eq 28 at 0 for $0 < p \leq 1$ (see also Figure 3) can be overcome by using a small relaxation parameter $\varepsilon > 0$ (chosen as 10^{-4} in our computations) and choosing the penalty

$$\|q\|^p_{p,\varepsilon} := \sum_{i \in I_R} (|q_i|^2 + \varepsilon)^{p/2} \tag{35}$$

instead of eq 28. Then, $\|q\|^p_{p,\varepsilon}$ is also differentiable at 0 (with zero slope, also see the hierarchical optimization approach

TABLE 3: Parameters of the LLKE Model,[2] Also Denoted by $\boldsymbol{q^*} \in \mathbb{R}^{20}$

$q_1 = k_1 = 6000\ M^{-1}\ s^{-1}$
$q_2 = K_{2a} = 5.4 \times 10^{-13}\ M^2$
$q_3 = k_{2ar} = 3.67 \times 10^9\ M^{-1}\ s^{-1}$
$q_4 = k_{2br} = 3.48 \times 10^9\ M^{-1}\ s^{-1}$
$q_5 = k_3 = 7.8\ M^{-1}\ s^{-1}$
$q_6 = k_4 = 6.9 \times 10^7\ M^{-1}\ s^{-1}$
$q_7 = k_5 = 10^6\ M^{-1}\ s^{-1}$
$q_8 = k_6 = 4.3 \times 10^8\ M^{-1}\ s^{-1}$
$q_9 = k_7 = 1500\ M^{-1}\ s^{-1}$
$q_{10} = K_8 = 4.55 \times 10^7\ M^{-1}$
$q_{11} = k_{8r} = 22\ M^{-1}\ s^{-1}$
$q_{12} = k_9 = 25\ M^{-1}\ s^{-1}$
$q_{13} = k_{10} = 110\ M^{-1}\ s^{-1}$
$q_{14} = K_{14} = 0.02\ M$
$q_{15} = K_{15} = 0.034\ M$
$q_{16} = K_{16} = 0.00135\ M$
$q_{17} = K_{11} = 1000\ M^{-2}$
$q_{18} = k_{11r} = 22\ s^{-1}$
$q_{19} = k_{12} = 1.5 \times 10^5\ M^{-1}\ s^{-1}$
$q_{20} = k_{13} = 10^6\ M^{-1}\ s^{-1}$

TABLE 4: Lower and Upper Bounds $\underline{q}_i$ and $\bar{q}_i$ Used for Confining the Parameter Search Space Q; Dimensions as in Table 3[a]

$\underline{q}_1$	$\underline{q}_2$	$\underline{q}_3$	$\underline{q}_4$	$\underline{q}_5$	$\underline{q}_6$	$\underline{q}_7$	$\underline{q}_8$	$\underline{q}_9$	$\underline{q}_{10}$
0	5.4×10^{-14}	0	0	0	0	0	0	0	2.3×10^{6}

$\underline{q}_{11}$	$\underline{q}_{12}$	$\underline{q}_{13}$	$\underline{q}_{14}$	$\underline{q}_{15}$	$\underline{q}_{16}$	$\underline{q}_{17}$	$\underline{q}_{18}$	$\underline{q}_{19}$	$\underline{q}_{20}$
0	0	0	0.016	0.01	0.0013	100	0	0	0

$\bar{q}_1$	$\bar{q}_2$	$\bar{q}_3$	$\bar{q}_4$	$\bar{q}_5$	$\bar{q}_6$	$\bar{q}_7$	$\bar{q}_8$	$\bar{q}_9$	$\bar{q}_{10}$
12000	5.4×10^{-12}	7.34×10^{9}	6.96×10^{9}	15.6	1.38×10^{8}	2×10^{6}	8.6×10^{8}	3000	4.55×10^{8}

$\bar{q}_{11}$	$\bar{q}_{12}$	$\bar{q}_{13}$	$\bar{q}_{14}$	$\bar{q}_{15}$	$\bar{q}_{16}$	$\bar{q}_{17}$	$\bar{q}_{18}$	$\bar{q}_{19}$	$\bar{q}_{20}$
44	50	220	0.024	0.34	0.014	1×10^{4}	44	3×10^{5}	2×10^{6}

[a] The selection of bounds for the optimizations were based on the published values available. Depending on the assessed accuracy of the measurements for the various rate and equilibrium constants, bounds for the optimizations were chosen appropriately. The measurements of equilibrium constants are generally easier to measure so its bounds were chosen as 1 magnitude higher and lower, respectively. For the kinetic constants the bounds were chosen asymmetrically. The upper bounds are normally easier to estimate so they were chosen as double of published values. For the lower bound no real limit can be assumed so the physical boundary of zero was reasonable.

TABLE 5: Initial Reactant and H$^+$ Concentrations (in M) Used in the N_e = 31 Experiments

j	$\bar{c}^j_{0,1}$	$\bar{c}^j_{0,6}$	$\bar{u}^j$	j	$\bar{c}^j_{0,1}$	$\bar{c}^j_{0,6}$	$\bar{u}^j$
1	9.9638×10^{-5}	1.9991×10^{-4}	0.0050	17	4.4903×10^{-4}	8.9991×10^{-4}	0.0012
2	9.9638×10^{-5}	1.9991×10^{-4}	0.0079	18	5.9871×10^{-4}	1.1998×10^{-3}	0.0012
3	9.9638×10^{-5}	1.9991×10^{-4}	0.0126	19	1.4967×10^{-4}	2.9997×10^{-4}	0.0021
4	9.9638×10^{-5}	1.9991×10^{-4}	0.0200	20	4.4903×10^{-4}	8.9991×10^{-4}	0.0021
5	9.9638×10^{-5}	1.9991×10^{-4}	0.0316	21	5.9871×10^{-4}	1.1998×10^{-3}	0.0021
6	1.9927×10^{-4}	3.9983×10^{-4}	0.0032	22	1.4967×10^{-4}	2.9997×10^{-4}	0.0041
7	1.9927×10^{-4}	3.9983×10^{-4}	0.0050	23	4.4903×10^{-4}	8.9991×10^{-4}	0.0041
8	1.9927×10^{-4}	3.9983×10^{-4}	0.0079	24	5.9871×10^{-4}	1.1998×10^{-3}	0.0041
9	1.9927×10^{-4}	3.9983×10^{-4}	0.0126	25	4.4903×10^{-4}	8.9991×10^{-4}	0.0015
10	1.9927×10^{-4}	3.9983×10^{-4}	0.0200	26	5.9871×10^{-4}	1.1998×10^{-3}	0.0015
11	1.9927×10^{-4}	3.9983×10^{-4}	0.0316	27	1.4967×10^{-4}	2.9997×10^{-4}	0.0028
12	2.9891×10^{-4}	7.9966×10^{-4}	0.0050	28	4.4903×10^{-4}	8.9991×10^{-4}	0.0028
13	2.9891×10^{-4}	7.9966×10^{-4}	0.0079	29	5.9871×10^{-4}	1.1998×10^{-3}	0.0028
14	2.9891×10^{-4}	7.9966×10^{-4}	0.0126	30	1.4967×10^{-4}	2.9997×10^{-4}	0.0052
15	2.9891×10^{-4}	7.9966×10^{-4}	0.0200	31	5.9871×10^{-4}	1.1998×10^{-3}	0.0052
16	2.9891×10^{-4}	7.9966×10^{-4}	0.0316				

below) and its derivative with respect to q_i, $i \in I_R$, is given by

$$p\frac{q_i}{(|q_i|^2+\varepsilon)^{1-p/2}} \tag{36}$$

Using eq 35 as penalty term yields the minimization problem

$$J_\alpha^\varepsilon(x) = J_0(x) + \alpha\,\|q\|_{p,\varepsilon}^p \rightarrow \min_{x\in X} \tag{37}$$

Finite Difference Approximation

Turning toward J_0, its gradient in principle could be computed by the finite difference method as, for example, used in the ZITA package.[4] With e_i denoting the i-th unit vector in $\mathbb{R}^N$ and some $\varepsilon << 1$,

$$\frac{\partial J_0(x)}{\partial x_i} \approx \frac{J_0(x+\varepsilon e_i) - J_0(x)}{\varepsilon}, i = 1, ..., N \tag{38}$$

serves as an approximation of the i-th component of $\nabla J_0(x) \in \mathbb{R}^N$. Numerical differentiation, as a prime example for an ill-posed problem, has to be handled with much care since otherwise errors in the evaluation of J_0 (e.g., due to errors in the data z^j) can lead to faulty calculations of the derivatives. Furthermore, the finite difference approach (eq 38) requires solving $N_e + (n + 3)\cdot N_e$ nonlinear ODE systems (eq 9), an effort necessary in each iteration step of the local optimization procedure *fmincon*.

The computational disadvantages of finite differences can be avoided if the gradient of J_0 is provided by means of the adjoint technique. A first step toward its introduction is the derivation of a formula for the gradient of J_0 based on a set of linearizations of the nonlinear model (eq 9).

Exact Gradient Formula Using a Set of Linearized Equations

The gradient of J_0 at argument x is denoted by $\nabla J_0(x) \in \mathbb{R}^N$. For any $h \in \mathbb{R}^N$ it satisfies

$$\lim_{\lambda\to 0} \frac{|J_0(x+\lambda h) - J_0(x) - \lambda(\nabla J_0, h)_N|}{\lambda} = 0 \tag{39}$$

where $(\cdot,\cdot)_N$ denotes the scalar product in $\mathbb{R}^N$. Actually, $(\nabla J_0(x), h)_N$ is the directional derivative of J_0 at x in direction h, often also denoted

$$J_0'(x)h \tag{40}$$

The N components of the vector $\nabla J_0(x)$ are given by $(\nabla J_0(x), e_i)_N$, $i = 1,..., N$, or eq 40 with directions $h = e_i$, $i = 1,..., N$. Hence, eq 40 is the key to any routine for calculating the gradient of J_0.

For the following discussion, it is advantageous to consider a general direction $h \in \mathbb{R}^N$ (where $h = e_i$ then is a special case) and to split the variables of h similar to eq 21 into

$$h = (s, d_{0,1}^{1}, d_{0,6}^{1}, v^{1}, ..., d_{0,1}^{N_e}, d_{0,6}^{N_e}, v^{N_e})$$

Therein, $s \in \mathbb{R}^n$ and $d_{0,1}^j$, $d_{0,6}^j$, $v^j \in \mathbb{R}$ denote possible additive perturbations of q, $c_{0,1}^j$, $c_{0,6}^j$, and u^j. Furthermore, for given h we introduce the auxiliary vectors

$$h^j = (s, d_{0,1}^j, d_{0,6}^j, v^j) \in \mathbb{R}^{n+3}$$

for $j = 1,..., N_e$ similar to those in eq 17.

Because of the chain rule, the directional derivative (eq 40) of eq 22 involves directional derivatives $F'(x^j, T^j)h^j$ of the forward operator (eq 16) at x^j in directions h^j. That is, we have

$$J_0'(x)h = -2\sum_{j=1}^{N} \omega^j (z^j - F(x^j, T^j), F'(x^j, T^j)h^j)_j \qquad (41)$$

where $(\cdot,\cdot)_j$ denotes the inner product in $L^2([0, T^j])$. Since the absorption operator A is linear, the derivative of F can easily be built once that of the nonlinear solution operator S is characterized. To this end, we consider two arguments x^j and $x^j + \lambda h^j$ with some scalar λ. Then, similar to eq 39, the derivative of S at x^j in direction h^j is defined as the linear operator L which satisfies

$$\lim_{\lambda \to 0} \frac{\|S(x^j + \lambda h^j, T_j) - S(x^j, T^j) - \lambda L h^j\|_{L^2([0,T^j],\mathbb{R}^9)}}{\lambda} = 0$$

in the following denoted by $L = S'(x^j, T^j)$. It can be shown that $S'(x^j, T^j)$ maps h^j onto the solution of a linearization of the ODE system (eq 9), and more particular, we have

$$S'(x^j, T^j): h^j \rightarrow w(h^j)^j$$

where $w(h^j)^j \in C([0, T^j], \mathbb{R}^9)$ denotes the solution of the linear ODE system

$$\dot{w}(t) = f_c(c^j, q, u^j)w + f_q(c^j, q, u^j)s + f_u(c^j, q, u^j)v^j, \quad \text{on } (0, T^j] \qquad (42)$$

$$w(0) = (d_{0,1}^j, 0, 0, 0, 0, d_{0,6}^j, 0, 0, 0)^T \qquad (43)$$

Therein,

$$f_c(c, q, u) = \frac{\partial f}{\partial c}(c, q, u)$$

$$= \begin{pmatrix} \frac{\partial f_1}{\partial c_1}(c, q, u) & ... & \frac{\partial f_1}{\partial c_9}(c, q, u) \\ \vdots & ... & \vdots \\ \frac{\partial f_9}{\partial c_1}(c, q, u) & ... & \frac{\partial f_9}{\partial c_9}(c, q, u) \end{pmatrix} \in \mathbb{R}^{9\times 9} \qquad (44)$$

is the Jacobian matrix of f with respect to c, for instance, the matrix entry at position (4, 6) is

$$f_c(c, q, u)_{4,6} = q_6 \cdot \frac{u}{q_{14} + u} \cdot c_3 \cdot \frac{q_{15}}{q_{15} + u} - q_7 \cdot \frac{u}{q_{14} + u} \cdot c_4$$

Furthermore, $f_q(c, q, u) \in \mathbb{R}^{9\times n}$ is the Jacobian matrix of f with respect to q and $f_u(c, q, u) \in \mathbb{R}^9$ is the derivative of f with respect to u.

Using the definitions of the operators involved, we can write eq 41 as

$$J_0'(x)h = -2\sum_{j=1}^{N_e} \omega^j \int_0^{T^j} (z^j(t) - 770 \cdot c_2^j(t)) \cdot 770 \cdot w(h^j)_2^j(t)\, \mathrm{d}t \qquad (45)$$

In principle, this formula could already be used for building the gradient of J_0 at x without the need for finite differences by evaluation of eq 45 for all $h = e_i$, $i = 1,..., N$. However, this would require solving N_e nonlinear ODE systems (eq 9) for c^j, $j = 1,..., N$, as well as $(n + 3)\cdot N_e$ linear ODE systems (eq 42) for the various $w(h^j)^j$. A significant reduction of these computational costs can be achieved if, instead of using eq 45, the gradient of J_0 is built by means of the adjoint technique.

Efficient Gradient Calculation by Means of the Adjoint Technique

As an alternative to eq 45, a formula for the derivative (eq 40) can be derived that involves solutions to so-called adjoint state equations and, as a consequence, can be computed much cheaper. We start by introducing the vector valued function $r^j \in C([0, T^j], \mathbb{R}^9)$ defined as

$$r_2^j(t) = 770 \cdot (z^j(t) - 770 \cdot c_2^j(t)) \text{ and } r_j^i(t) \equiv 0 \text{ for } i \neq 2 \qquad (46)$$

Then, eq 45 is equivalent to

$$J_0'(x)h = -2\sum_{j=1}^{N_e} \omega^j \int_0^{T^j} \langle r^j(t), w(h^j)^j(t)\rangle\, \mathrm{d}t \qquad (47)$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product in $\mathbb{R}^9$. The residual function r is furthermore used in the linear ODE system

$$\dot{\chi}(t) = -f_c(c^j, q, u^j)^T \chi + r^j, \text{ on } [0, T^j) \qquad (48)$$

$$\chi(T^j) = 0 \qquad (49)$$

referred to as the adjoint equation. Its solution, the adjoint state, is denoted by $\chi^j \in C([0, T^j], \mathbb{R}^9)$.

Next, if we use in eq 47 the representation for r^j induced by eq 48, partially integrate with respect to time, consider eqs 43 and 49, and take advantage of eq 42, we obtain the alternative formula

$$\begin{aligned} J_0'(x)h = & -2\sum_{j=1}^{N_e} \omega^j \int_0^{T^j} \langle f_c(c^j, q, u^j)^T \chi^j + \dot{\chi}^j, w(h^j)^j\rangle\, \mathrm{d}t \\ = & -2\sum_{j=1}^{N_e} \omega^j \Big\{ \int_0^{T^j} \langle \chi^j, f_c(c^j, q, u^j)w(h^j)^j - \dot{w}(h^j)^j\rangle \mathrm{d}t \\ & - \Big\{\langle \chi^j(0), w(h^j)^j(0)\rangle\Big\} \\ = & -2\sum_{j=1}^{N_e} \omega^j \Big\{ \int_0^{T^j} \langle \chi^j, f_q(c^j, q, u^j)s + f_u(c^j, q, u^j)v^j\rangle \mathrm{d}t \\ & + \Big\{\chi_1^{\ j}(0) \cdot d_{0,1}^j + \chi_6^j(0) \cdot d_{0,1}^j\Big\} \end{aligned} \qquad (50)$$

for eq 40.

The advantage of eq 50 over eq 45 comes from the fact that, as opposed to the solutions $w(h^j)^j$ of eq 42, the solutions χ^j, $j = 1,..., N_e$, of eq 48 do not depend on the variables h^j. Hence, the adjoint solutions can be computed independently of the search directions h or h^j, respectively. Overall, the computation of the gradient $\nabla J(x_0)$ according to eq 50 only requires to solve (in addition to the N_e nonlinear systems for c^j) N_e linear ODE systems (eq 48) for χ^j, $j = 1,..., N_e$, compared to $(n + 3)\cdot N_e$

linear systems (eq 42) in case of the approach (eq 45). This clearly demonstrates the benefits of the adjoint technique (which of course again avoids finite differences).

Note that the adjoint problem (eq 48) has to be solved backward in time because of the terminal condition (eq 49). By means of a change of variables $\tau = T^j - t$ and

$$\hat{\chi}(\tau) = \chi(T^j - \tau),\ r^{\wedge j}(\tau) = r^j(T^j - \tau),\ c^{\wedge j}(\tau) = c^j(T^j - \tau)$$

eq 48 can be transformed to

$$\begin{aligned} \hat{\chi}_\tau(\tau) &= f_c(c^{\wedge j}, q, u^j)^T \hat{\chi} - r^{\wedge j},\ \text{on } (0, T^j] \\ \hat{\chi}(0) &= 0 \end{aligned} \tag{51}$$

i.e., a problem forward in time. Clearly, the original solution χ^j needed in eq 50 is then obtained as $\chi^j(t) = \hat{\chi}^j(T^j - t)$.

Summarizing, the gradient calculation of J_0 at x as in eq 21 via the adjoint method involves the following steps:

1. Compute the solutions c^j of eq 9 for all x^j (see eq 17), $j = 1,..., N_e$. Typically, this task is already performed by a preceding evaluation of $J_0(x)$.
2. Build the residuals r^j as in eq 46 for $j = 1,..., N_e$.
3. Solve the adjoint state equations (eq 51) for $\hat{\chi}^j$, $j = 1,..., N_e$.
4. For $h = e_i$, $i = 1,..., N$, compute the inner products, time integrals, and sums of eq 50 for $J_0(x)'e_i$ to build $\nabla J_0(x) \in \mathbb{R}^N$.

The gradient of J_α^ε in eq 37 is then easily build using eqs 50 and 36.

Hierarchical Optimization Approach

Parameter components q_i, $i \in I_R$, to which the output $F(x^j, T^j)$ is not or only weakly sensitive, are detected by our sparsity enforcing approach. Still, they are possibly not exactly driven to zero but only to some small positive values during the optimization procedure (e.g., due to the zero slope of the relaxed penalty term (eq 35) at zero). Nevertheless, small values are forced by eq 35 to become very small such that strictly sparse solutions can be obtained by the following hierarchical strategy for solving the parameter identification problem:

1. Solve the optimization problem (eq 37) with rate index set I_R as in eq 23.
2. Set components q_i, $i \in I_R$, below the threshold ε equal to zero, i.e.,

$$q_i \le \varepsilon \text{ with } i \in I_R \rightarrow q_i \equiv 0$$

3. Eliminate indices $i \in I_R$ with $q_i = 0$ from I_R such that the set I_R is reduced.
4. Repeat steps 1−3 until sparse components can no longer be found and I_R can no longer be reduced.

References and Notes

(1) Kepper, P. D.; Boissonade, J.; Epstein, I. R. *J. Phys. Chem.* **1990**, *94*, 6525–6536.

(2) Lengyel, I.; Li, J.; Kustin, K.; Epstein, I. R. *J. Am. Chem. Soc.* **1996**, *118*, 3708–3719.

(3) Kormányos, B.; Horváth, A. K.; Peintler, G.; Nagypál, I. *J. Phys. Chem. A* **2007**, *111*, 8104–8109.

(4) Peintler, G. *Zita, version 4.1: A Comprehensive Program Package for Fitting Parameters of Chemical Reaction Mechanisms*; Attila Jozsef University: Szeged, Hungary, 1997.

(5) Bock, H.-G. *Randwertproblemmethoden zur Parameteridentifizierung in Systemen Nichtlinearer Differentialgleichungen*; Bonner Mathematische Schriften 183; Rheinische Friedrich-Wilhelms-Universität: Bonn, Germany, 1987.

(6) Nowak, U.; Deuflhard, P. *Appl. Numer. Math.* **1985**, *1*, 59–75.

(7) Engl, H. W.; Hanke, M.; Neubauer, A. *Regularization of Inverse Problems*; Kluwer Academic Publishers: Boston, 1996.

(8) Ljung, L. *System Identification: Theory for the User*; Prentice Hall: Upper Saddle River, NJ, 1999.

(9) Burger, M.; Osher, S. *Inverse Probl.* **2004**, *20*, 1411–1420.

(10) Daubechies, I.; Defrise, M.; Mol, C. D. *Commun. Pure Appl. Math.* **2004**, *57*, 1413–1457.

(11) Grasmair, M.; Haltmeier, M.; Scherzer, O. Inverse Probl. [Online] 2008, *24*, 055020. http://www.iop.org/EJ/abstract/0266-5611/24/5/055020/.

(12) Ramlau, R. *Electron. Trans. Numer. Anal.* **2008**, *30*, 54–74.

(13) Ramlau, R.; Teschke, G. *Numer. Math.* **2006**, *104*, 177–203.

(14) Griesse, R.; Lorenz, D. A. Inverse Probl. [Online] 2008, *24*, 035007. http://www.iop.org/EJ/abstract/0266-5611/24/3/035007/.

(15) Schmitz, G. *Int. J. Chem. Kinet.* **2004**, *36*, 480–493.

(16) Dateo, C. E.; Orban, M.; Kepper, P. D.; Epstein, I. R. *J. Am. Chem. Soc.* **1982**, *104*, 504–509.

(17) Grasmair, M. Well-Posedness and Convergence Rates for Sparse Regularization with Sublinear l^q Penalty Term; Technical Report 74, FSP 092: Joint Research Program of Industrial Geometry; University of Innsbruck: Innsbruck, Austria, August 2008.

(18) Zarzer, C. *Inverse Probl.* Submitted for publication, **2008**.

(19) Donoho, D. L. *Commun. Pure Appl. Math.* **2006**, *59*, 797–829.

(20) Kaipio, J.; Somersalo, E. *Statistical and Computational Inverse Problems*; Springer: New York, 2004.

(21) Lengyel, I.; Li, J.; Epstein, I. R. *J. Phys. Chem.* **1992**, *96*, 7032–7037.

JP808792U