Inverse matrix evaluation for linear systems

M Tadi[1] and Wei Cai[2]

[1] Department of Mechanical Engineering, University of Colorado at Denver, Campus Box 112,
PO Box 173364, Denver, CO 80217-3364, USA
[2] Department of Mathematics, University of North Carolina at Charlotte, 9201 University City
Boulevard, Charlotte, NC 28223, USA

Abstract
This paper is concerned with an iterative method to recover the system matrix
for a linear time-invariant dynamical system. It uses the states as given data and
generates a series of simulations, after which it is possible to directly invert for
the sought after matrix. It requires the solution to a nonlinear algebraic matrix
equation. Numerical results indicate that the algorithm requires only a few
iterations to converge to the true solution. The effect of noise on the accuracy
of the results is also investigated.

1. Introduction

In this paper we present an iterative method to recover the system matrix for a linear time-
invariant (LTI) dynamical system. Such systems arise in a variety of physical applications and,
as a result, they have been vigorously studied for various purposes. In this paper we concentrate
on monomolecular chemical reactions in which the concentration of the species is modelled by
an LTI dynamical system. For this case the parameters of interest are rate constants. The rate
constants determine the relative speed of each reaction and, ultimately, the equilibrium point
of the whole reaction mechanism the dynamics asymptotically approach [1–3]. It is possible
to measure the concentration of species (the states) as a function of time and therefore the
problem of interest is to recover the rate constants from the available laboratory data.

Motivated by numerous physical applications, inverse problems for various physical
systems have received considerable attention and the literature on this subject is vast. A
brief summary of results for systems described by ordinary differential equations can be found
in, for example, [4–6]. These approaches include the method of least squares, stochastic
least squares, maximum likelihood, Kalman filtering and a number of other variations. For
applications in chemical reactions, the kinetics is modelled by a system of nonlinear ordinary
differential equations that is often stiff. As a result, parameter identification efforts for such
systems have mostly relied on traditional methods such as least squares along with sensitivity
analysis techniques [7, 8]. Inverse eigenvalue problems [9] have also been studied in various
fields of applications. For instance, in mechanical vibrations, the interest is to recover the
system mass and stiffness matrices from the knowledge of the eigenvalues and eigenvectors.
A number of methods have been developed for these specific applications. These methods

include algebraic approaches [10] and the continuation methods [11]. In addition, a good review of the existing methods for inverse vibration problems can be found in [12].

In this paper we study an inverse problem for chemical reactions whose kinetics is modelled by a linear system of ordinary differential equations. The present method can be thought of as a direct approach in which the data are used directly to generate a matrix and, essentially, the inverse of the matrix is obtained using singular-value decomposition. In section 2 we present the formulation of the method and discuss the algorithm in detail. In section 3 we investigate the convergence of the iterations involved in the algorithm. In section 4 we use a number of numerical examples to show the applicability of the method and section 5 is devoted to conclusions.

2. Mathematical formulation

Consider an LTI system of equations given by

$$\dot{x} = Kx, \qquad x \in R^n, \qquad x(0) = x_0, \tag{1}$$

where the vector x contains the concentration of the species A_i, $i = 1, \ldots, n$ and the matrix K is made up of the unknown rate constants, k_{ij}, $i \neq j$. The rate constant, k_{ij}, is the speed of the reaction $A_i \rightarrow A_j$. The observation is the time history of the concentration of the species which can be measured in the laboratory. In chemical kinetics the individual rate constants, k_{ij}, need to be non-negative and the coefficient matrix, K, also needs to have a special structure, namely, that the diagonal elements are the negative of the sum of the rest of the elements in that column [13]. However, in the present method we do not enforce any structure onto the unknown matrix. We assume that all of the states can be measured and are available for inversion. The inverse problem for this system is then to recover the matrix of the rate constants, K, based on the vector of species concentrations x.

If the matrix Q is an initial guess for the rate constants, K, then the corresponding system response is given by

$$\dot{\hat{x}} = Q\hat{x} \qquad \hat{x} \in R^n, \qquad \hat{x}(0) = x_0. \tag{2}$$

The assumed matrix of rate constants, Q, is related to the actual unknown, K, according to $K = Q + H$, where now H is unknown. If we denote the error in the species concentrations by v, then the error dynamics is given by

$$\dot{v} = \dot{x} - \dot{\hat{x}} = Kx - Q\hat{x} = (Q + H)x - Q\hat{x} \tag{3}$$

$$\dot{v} = Qv + Hx. \tag{4}$$

Differentiating the above equation and employing equations (3) and (4) leads to

$$\ddot{v} = Q(Qv + Hx) + H(Q + H)x. \tag{5}$$

We next multiply the equation for the error dynamics by a positive scalar, α, and subtract it from equation (5), yielding

$$\ddot{v} - \alpha \dot{v} = Q(Q - \alpha I)v + \Phi x, \tag{6}$$

where

$$(Q - \alpha I)H + H(Q + H) = \Phi, \tag{7}$$

and I denotes the identity matrix. At this stage, the matrix Φ is an unknown and can be determined by integrating equation (6) on different time intervals. The unknown matrix H can be found from Φ by solving the Riccati equation (7) by iteration.

Let us first explain how to compute Φ. If we integrate equation (6) from $t = 0$ to τ and use $v|_0 = 0$, we obtain

$$\dot{v}|_0^\tau - \alpha v(\tau) = Q(Q - \alpha I)\int_0^\tau v\,dt + \Phi\eta(\tau), \qquad (8)$$

where $\eta(\tau) = \int_0^\tau x\,dt$. Next, we multiply equation (8) from the right by $\eta(\tau)^\top$ and obtain

$$[\dot{v}|_0^\tau - \alpha v(\tau)][\eta(\tau)]^\top = Q(Q - \alpha I)\left[\int_0^\tau v\,dt\right][\eta(\tau)]^\top + \Phi[\eta(\tau)][\eta(\tau)]^\top. \qquad (9)$$

Since the matrix $[\eta(\tau)][\eta(\tau)]^\top$ has only one nonzero eigenvalue and is therefore singular, the matrix Φ cannot be found using equation (9). However, if we perform the time integration for a series of time intervals, $[0, \tau_j]$, then we arrive at the equations for the outer products given by

$$[\dot{v}|_0^{\tau_j} - \alpha v(\tau_j)][\eta(\tau_j)]^\top = Q(Q - \alpha I)$$
$$\times\left[\int_0^{\tau_j} v\,dt\right][\eta(\tau_j)]^\top + \Phi[\eta(\tau_j)][\eta(\tau_j)]^\top, \qquad j = 1, \ldots, m \qquad (10)$$

where $m \geqslant n$ is the number of simulations. All of the above equations correspond to the same assumed rate constant matrix, Q, and we can add them to arrive at

$$A = Q(Q - \alpha I)B + \Phi C \qquad (11)$$

where

$$A = \sum_{j=1}^m [\dot{v}|_0^{\tau_j} - \alpha v(\tau_j)][\eta(\tau_j)]^\top, \qquad B = \sum_{j=1}^m \left[\int_0^{\tau_j} v\,dt\right][\eta(\tau_j)]^\top,$$
$$C = \sum_{j=1}^m [\eta(\tau_j)][\eta(\tau_j)]^\top. \qquad (12)$$

In the next section we will show that if the system matrix K has full rank then the matrix C has full rank for $m = n$, and can be inverted. For applications in chemical kinetics in which the concentration of species approach an equilibrium state, i.e. $Kx = 0$, the matrix K has $(n-1)$ nonzero eigenvalues. For this case, the matrix C becomes nearly singular due to numerical ill conditioning with only one eigenvalue close to zero and we use singular-value decomposition to obtain its psuedoinverse, i.e. C^+. Therefore, we need to add at least $m = n$ number of simulations. In practice, we use $m > n$ to improve the numerical conditioning of the matrix C. The set $\int_0^{\tau_j} x\,dt$, $j = 1, \ldots, m$, $m > n$ can at most span an n-dimensional space. The equation (10) can now be used to solve for the matrix Φ, which is given by

$$(A - Q(Q - \alpha I)B)C^+ = \Phi. \qquad (13)$$

The system dynamics is given by equation (1). For linear problems in chemical kinetics the concentration of species approaches a steady state condition, i.e. $Kx = 0$. Let the steady state condition be denoted by x_s. If the initial condition is chosen to be equal to the steady state solution, i.e. $x_0 = x_s$, then the states will be constant in time, and the vectors $\eta(\tau_j)$ will be linearly dependent for $j = 1, \ldots, m$. The matrix $C = \sum_{j=1}^m [\eta(\tau_j)][\eta(\tau_j)]^\top$ will still have only one nonzero eigenvalue, and equation (10) cannot be used to solve for the matrix Φ. However, in general, the initial condition is different from the steady state solution and the vectors $\eta(\tau_j)$ will not be linearly dependent. Once the matrix Φ is solved for, then equation (7) can be used to solve for the unknown matrix H. Equation (7) is a nonlinear algebraic Riccati equation which has been extensively investigated [14, 15]. One way to solve for the matrix H is an iterative method given by

$$(Q - \alpha I)H_{k+1} + H_{k+1}(Q + H_k) = \Phi, \qquad k = 1, \ldots, \ell. \qquad (14)$$

It is then necessary to separate the eigenvalue sets of the matrices $Q - \alpha I$ and $Q + H_k$. This is achieved by the introduction of α through equation (6). A necessary and sufficient condition for each iteration to have a unique solution is given by

$$\lambda_j + \mu_k \neq 0, \qquad j, k = 1, \ldots, n, \tag{15}$$

where λ_j are the eigenvalues of the matrix $(Q - \alpha I)$, and μ_k are the eigenvalues of the matrix $(Q + H_k)$, [6]. Choosing a large value for the parameter α ensures that the above condition is satisfied. It also makes the system diagonally dominant and, as a result, the above iterations converge to the solution. A good choice for α is a positive scalar which is larger than the absolute value of all of the eigenvalues of the matrix K. We choose the value of 100 for all the problems studied here. We can then recover the matrix of rate constants using the following algorithm.

Algorithm

(1) Select a series of time intervals $[0, \tau_j]$, $j = 1, m$, $\tau_1 < \tau_2 < \cdots < \tau_m \leqslant T$, where the data are given for the time period $[0, T]$. Also choose an initial guess for the matrix of rate constants Q.
(2) Use the given initial condition and the assumed rate constant matrix, Q, and solve the system of equations (2), forward until the time τ_m and, thereby, obtain the error $v(t)$.
(3) Compute the terms in equation (10) for all time intervals τ_j.
(4) Compute the matrices A, B, C using equations (12).
(5) Use singular-value decomposition to obtain the pseudoinverse of the matrix C, after which it can be used to obtain the matrix Φ using equation (13).
(6) Solve for the matrix H using the Riccati equation (7).
(7) Update the assumed value of the rate constant matrix according to $Q = Q + H$, and go to step (2). Repeat the process until the error is arbitrarily small and convergence is obtained.

3. Convergence analysis

We first proceed to show that the iterative approach to solve the nonlinear matrix equation converges to a solution. Consider two such iterations, i.e.

$$(Q - \alpha I)H_{k+1} + H_{k+1}(Q + H_k) = \Phi \tag{16}$$
$$(Q - \alpha I)H_k + H_k(Q + H_{k-1}) = \Phi, \tag{17}$$

and let $\Delta_k = H_{k+1} - H_k$. Then, subtracting the above two equations leads to the equation for Δ_k given by

$$(Q - \alpha I)\Delta_k + \Delta_k(Q + H_k) = -H_k \Delta_{k-1}, \tag{18}$$

where we have added and subtracted the term $H_k H_k$. We can define two operators

$$\mathcal{L}_{\alpha,k} = (Q - \alpha I) \otimes I + I \otimes (Q + H_k)^\top = \mathcal{A} + I \otimes H_k^\top, \tag{19}$$
$$\mathcal{A} = (Q - \alpha I) \otimes I + I \otimes Q^\top \tag{20}$$

after which we can rewrite the above matrix equations in the vector form,

$$\mathcal{L}_{\alpha,k} H_{k+1} = \Phi, \qquad \mathcal{L}_{\alpha,k} \Delta_k = -H_{k+1} \Delta_{k-1}, \tag{21}$$

where $\otimes$ is the Kronecker product, and $\mathcal{A} \in R^{n^2 \times n^2}$. The following theorem is due to an unnamed referee.

Theorem 1. *Starting the iteration in equation (21) with H_0 and choosing the constant M such that $M \geqslant \|H_0\|$, the iteration converges to a unique solution H of equation (7) satisfying $\|H\| \leqslant M$ if*

$$\alpha > 2 \left(\|Q\| + M + \frac{\|\Phi\|}{M} \right). \tag{22}$$

Proof. Defining the operator $\mathcal{B}$ by $\mathcal{B}(F) = QF + FQ$, $\mathcal{A} = \mathcal{B} - \alpha I$ for $\alpha > 2\|Q\|$ and[3]

$$\|\mathcal{A}^{-1}\| \leqslant \sum_{m=0}^{\infty} \alpha^{-(m+1)} \|\mathcal{B}\|^m \leqslant \sum_{m=0}^{\infty} \alpha^{-(m+1)} (2\|Q\|)^m = \frac{1}{\alpha - 2\|Q\|}. \tag{23}$$

In particular, for such α we have

$$\|\mathcal{A}^{-1}\| \leqslant \frac{1}{\alpha - 2\|Q\|} \leqslant \frac{M}{2\|\Phi\|}, \tag{24}$$

implying that $\|\mathcal{A}^{-1}\| \|\Phi\| \leqslant \frac{1}{2} M$.

Let us now prove by induction that $\mathcal{L}_{\alpha,k}$ is invertible and $\|H_k\| \leqslant M (k = 0, 1, 2, \ldots)$. Indeed, assuming $\|H_k\| \leqslant M$ and using the identity $H_{k+1} = \mathcal{A}^{-1}(\Phi - H_{k+1}H_k)$, we estimate

$$\|H_{k+1}\| \leqslant \|\mathcal{A}^{-1}\|(\|\Phi\| + \|H_{k+1}\| \|H_k\|) \leqslant \frac{M}{2} + \frac{M\|H_{k+1}\|}{\alpha - 2\|Q\|} \tag{25}$$

which implies (because of $\alpha - 2\|Q\| \geqslant 2M$)

$$\|H_{k+1}\| \leqslant M, \tag{26}$$

as claimed. As a result,

$$\|\mathcal{A}^{-1}\mathcal{L}_{\alpha,k}(F) - F\| = \|\mathcal{A}^{-1}(FH_k)\| \leqslant M\|F\| \|\mathcal{A}^{-1}\| \leqslant \frac{M\|F\|}{\alpha - 2\|Q\|} \leqslant \frac{1}{2}\|F\|, \tag{27}$$

which implies the invertibility of $\mathcal{L}_{\alpha,k}$ and the norm estimate

$$\|\mathcal{L}^{-1}\| \leqslant \|\mathcal{A}^{-1}\| \sum_{m=0}^{\infty} \left(\frac{1}{2} \right)^m \leqslant \frac{2}{\alpha - 2\|Q\|}. \tag{28}$$

Next, we immediately see from the identity $\Delta_k = -\mathcal{L}_{\alpha,k}^{-1}(H_k \Delta_{k-1})$ that

$$\|\Delta_k\| \leqslant \frac{2M}{\alpha - 2\|Q\|} \|\Delta_{k-1}\|. \tag{29}$$

As a result, for $k > l$ we find that

$$\|H_k - H_l\| \leqslant \sum_{s=l}^{k-1} \|\Delta_s\| \leqslant \sum_{s=l}^{k-1} \left(\frac{2M}{\alpha - 2\|Q\|} \right)^s \|\Delta_0\| \tag{30}$$

$$\leqslant \left(\frac{2M}{\alpha - 2\|Q\|} \right)^l \frac{\|\Delta_0\|}{1 - \frac{2M}{\alpha - 2\|Q\|}}. \tag{31}$$

The latter estimate proves that $(H_k)_{k=0}^{\infty}$ is a Cauchy sequence and hence converges. Its limit H obviously satisfies $\|H\| \leqslant M$.

To prove uniqueness, we write the Riccati equation (7) in the form

$$H = \frac{1}{\alpha}(\mathcal{B}(H) + H^2 - \Phi). \tag{32}$$

---

[3] Here $\|\mathcal{B}\| = \sup_{\|F\|=1} \|\mathcal{B}(F)\| \leqslant 2\|Q\|$.

If H and $\tilde{H}$ are two solutions with norm $\leqslant M$, we easily see that

$$\|H - \tilde{H}\| \leqslant \frac{1}{\alpha}(\|\mathcal{B}\| + (\|H\| + \|\tilde{H}\|))\|H - \tilde{H}\| \leqslant \frac{2(\|Q\| + M)}{\alpha}\|H - \tilde{H}\|, \tag{33}$$

which implies that $H = \tilde{H}$, because $\alpha > 2(\|Q\| + M)$. This completes the proof. $\square$

Remark. If one initializes the iteration scheme by choosing H_0 with $\|H_0\| \leqslant \sqrt{\|\Phi\|}$, one takes $M = \sqrt{\|\Phi\|}$ and obtains a unique solution H for $\alpha > 2\|Q\| + 4\sqrt{\|\Phi\|}$ which satisfies $\|H\| \leqslant \sqrt{\|\Phi\|}$.

Remark. The uniqueness part of the proof shows that the nonlinear matrix Riccati equation (7) can also be solved by the iteration

$$H_{k+1} = \frac{1}{\alpha}(QH_k + H_kQ - H_k^2 - \Phi) \tag{34}$$

for any initialization matrix H_0 with $\|H_0\| \leqslant M$ if $\alpha > 2(\|Q\| + M)$. The convergence follows from the contraction mapping principle.

In all of the examples in this paper, the solution to the matrix equation is obtained with fewer than ten iterations. The method relies on the inversion of a matrix, C, which in turn depends on the property of the vector $\eta(t)$ as it changes in time. We next present a theorem involving outer products of vectors or dyads for completeness. In what follows we are mainly concerned with the existence results. Temporal data are given and we are free to choose an arbitrary set of time intervals $0 < \tau_1 < \tau_2 \cdots < \tau_m, m \geqslant n$, for the inversion. These times are chosen away from the steady state solution, i.e. $Kx(\tau_j) \neq 0, 1 < j \leqslant m$. First consider the case in which the matrix K is full rank.

Theorem 2. *Consider an LTI system $\dot{x} = Kx, x \in R^n, x(0) = x_0, \det(K) \neq 0$. The vectors $\int_0^{\tau_j} x(t)\,dt\ (j = 1, \ldots, m)$ with $m = n$ are linearly independent if:*

(a) the intervals $\tau_1, \ldots, \tau_m$ are distinct positive numbers such that none of the numbers $\lambda_k \tau_j$, where $\lambda_k\ (k = 1, \ldots, \nu$ with $\nu \leqslant n)$ ranges through the nonzero eigenvalues of K, is a multiple of $2\pi\mathrm{i}$ and
(b) $Kx_0 \neq 0$.

Proof. Integrating the above equation from zero to the time intervals, τ_j, leads to

$$x(\tau_1) - x_0 = K\int_0^{\tau_1} x\,dt, \ldots, x(\tau_m) - x_0 = K\int_0^{\tau_m} x\,dt. \tag{35}$$

If the vectors $\int_0^{\tau_j} x\,dt$ are linearly dependent then there must exist constants $c_1, \ldots, c_m$, not all equal to zero, such that

$$c_1\int_0^{\tau_1} x\,dt + \cdots + c_m\int_0^{\tau_m} x\,dt = 0. \tag{36}$$

Multiplying the equations in (35) by the c_j and adding them leads to

$$c_1(x(\tau_1) - x_0) + \cdots + c_m(x(\tau_m) - x_0) = K\left(c_1\int_0^{\tau_1} x\,dt + \cdots + c_m\int_0^{\tau_m} x\,dt\right) \tag{37}$$

which leads to

$$c_1(x(\tau_1) - x_0) + \cdots + c_m(x(\tau_m) - x_0) = 0. \tag{38}$$

We can rewrite the above equation in the form

$$\sum_{j=1}^{m} c_j (e^{K\tau_j} - I) x_0 = 0. \tag{39}$$

For a general $x_0 \neq 0$ the above equation can only be satisfied if the coefficient matrix is singular. The eigenvalues of the matrix $\sum_{j=1}^{m} c_j (e^{K\tau_j} - I)$ are the numbers $\sum_{j=1}^{m} c_j (e^{\lambda_k \tau_j} - 1)$. Writing the determinant of the coefficient matrix in terms of the product of its eigenvalues we arrive at

$$\prod_{k=1}^{n} \left(\sum_{j=1}^{m} c_j (e^{\lambda_k \tau_j} - 1) \right) = 0. \tag{40}$$

The above equation can only be satisfied if at least one of the factors in the product is equal to zero, i.e. for some $1 \leqslant \ell \leqslant n$

$$c^\top z_\ell = 0, \qquad z_\ell = [(e^{\lambda_\ell \tau_1} - 1), (e^{\lambda_\ell \tau_2} - 1), \dots, (e^{\lambda_\ell \tau_m} - 1)]^\top \tag{41}$$

where the vector c contains the coefficients c_j. Consider the first inner product $c^\top z_1$. Assume that the set τ_j is such that there exists a vector of coefficients $c \neq 0$ such that $c^\top z_1 = 0$. If we denote the last nonzero element of c by c_k then the inner product leads to

$$\sum_{j=1}^{k-1} c_j (e^{\lambda_1 \tau_j} - 1) + c_k (e^{\lambda_1 \tau_k} - 1) = 0, \qquad \text{or} \qquad \tau_k = \frac{1}{\lambda_1} \ln \left(1 - \frac{\sum_{j=1}^{k-1} c_j (e^{\lambda_1 \tau_j} - 1)}{c_k} \right). \tag{42}$$

It is sufficient to note that by changing τ_k so that $\tau_k \neq \frac{1}{\lambda_1} \ln(1 - \frac{\sum_{j=1}^{k-1} c_j (e^{\lambda_1 \tau_j} - 1)}{c_k})$ we can avoid the special cases. Following the same argument for z_ℓ, $\ell = 2, \dots, n$ we can obtain a set of τ_j such that the inner product can only be true for $c = 0$, which is a contradiction. This completes the proof. $\qquad\square$

In applications, we are free to choose a large number of simulations $m > n$ and the selection of the set τ_j can be at random. The following lemma from [16] is helpful to characterize the special cases.

Lemma 1. *Let y and z be any two real n-vectors and assume that $y \neq 0$. Then z satisfies $z^\top y = 0$ if and only if there exists a real, skew-symmetric, $n \times n$ matrix D, such that $z = Dy$.*

In chemical kinetics the coefficient matrix K has at most $(n-1)$ nonzero eigenvalues. Therefore we have the following lemma.

Lemma 2. *Consider an LTI system $\dot{x} = Kx$, $x \in R^n$, $x(0) = x_0$, Rank$(K) = (n-1)$. The vectors $\int_0^{\tau_j} x(t) \, dt$ $(j = 1, \dots, m)$, $m = (n-1)$, are linearly independent if:*

(a) the intervals $\tau_1, \dots, \tau_m$ are distinct positive numbers such that none of the numbers $\lambda_k \tau_j$, where λ_k $(k = 1, \dots, \nu$ with $\nu \leqslant (n-1))$ ranges through the nonzero eigenvalues of K, is a multiple of $2\pi i$ and

(b) $Kx_0 \neq 0$.

Proof. Following the same steps as in theorem 2 we arrive at equation (39)

$$\sum_{j=1}^{m} c_j (e^{K\tau_j} - I) x_0 = 0 \qquad \text{for} \quad m = (n-1) \tag{43}$$

where $\sum_{j=1}^{m} c_j (e^{K\tau_j} - I)$ is an $n \times n$ coefficient matrix. If the vectors $\int_0^{\tau_j} x(t) \, dt$, $j = 1, \dots, m$ are linearly dependent then the rank of the coefficient matrix is less than m for nonzero

coefficients c_j. We can again express the rank deficiency in terms of the determinant of an $m \times m$ submatrix of the coefficient matrix, i.e.

$$\prod_{k=1}^{m} \left(\sum_{j=1}^{m} c_j (e^{\lambda_k \tau_j} - 1) \right) = 0, \tag{44}$$

where λ_k, $k = 1, \ldots, n-1$ are the nonzero eigenvalues of the K matrix. In a similar way we can obtain a set of distinct positive numbers τ_j such that the above relation can only be satisfied if $c_j = 0$ for $j = 1, \ldots, m$. $\qquad\square$

Remark. For n linearly independent vectors u_i, $u_i \in R^n$, $i = 1, \ldots, n$, the matrix $C = \sum_{i=1}^{n} u_i u_j^\top$, which is the sum of the outer products, has an inverse. The statement follows by noting that we can rewrite the outer product as

$$C = \sum_{i=1}^{n} u_i u_i^\top = [u_1 | u_2 | \cdots] \begin{bmatrix} u_1^\top \\ u_2^\top \\ \vdots \end{bmatrix} = U U^\top, \tag{45}$$

where it is sufficient to note that the square matrix U has linearly independent columns.

We next proceed to show that the method can recover the system matrix from a full knowledge of the states. The states can be fully measured, therefore both $v(t)$ and $x(t)$ are known in time. Rewriting equation (11), we have

$$A = Q(Q - \alpha I)B + [(Q - \alpha I)H + H(Q + H)]C. \tag{46}$$

Integrating the error equation (4) for times $[0 : \tau_j]$ simplifies the relations for A and B to

$$A = \sum_{j=1}^{m} ((Q - \alpha I)v(\tau_j) + H(x(\tau_j) - x_0))(\eta(\tau_j))^\top,$$
$$QB = \sum_{j=1}^{m} (v(\tau_j) - H\eta(\tau_j))(\eta(\tau_j))^\top. \tag{47}$$

After substituting for A, B and C in equation (46) and simplifying, this leads to

$$H \sum_{j=1}^{m} \left[(x(\tau_j) - x_0) - (Q + H) \int_0^{\tau_j} x \, dt \right] [\eta(\tau_j)]^\top = 0. \tag{48}$$

The arbitrary times τ_j are chosen such that $|\eta(\tau_j)| \neq 0$, and in general, $H \neq 0$; then

$$\sum_{j=1}^{m} \left[(x(\tau_j) - x_0) - (Q + H) \int_0^{\tau_j} x \, dt \right] = 0. \tag{49}$$

On the other hand, for the linear system $\dot{x} = Kx$, we have

$$(x(\tau_j) - x_0) - K \int_0^{\tau_j} x \, dt = 0, \qquad 0 < \tau_j < \tau_m \tag{50}$$

which, after adding the equations for $j = 1, \ldots, m$, leads to $K = Q + H$.

4. Implementations and numerical examples

We next use a number of numerical examples to show the applicability of the proposed algorithm. The algorithm requires an initial guess Q for the sought for unknown. We use $Q = -\beta I$, where β is a positive constant ($\beta = 1.5$) and I is the identity matrix. We only require that the initial guess leads to a stable system. Also, we need to have $m > n$ and we use $m = 22$.

Example 1. As a first example we consider the system studied in [13]. It is a monomolecular chemical kinetic system given by figure 1.

$$A_1 \rightleftharpoons A_2$$

$$\swarrow \nearrow \qquad k_{ij} : A_i \rightarrow A_j$$
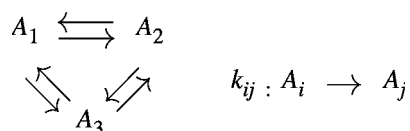
$$A_3$$

Figure 1. A chemical reaction involving three species.

The concentrations of the species are modelled by a system of first order rate equations given by

$$\dot{x}_1 = -(k_{21} + k_{31})x_1 + k_{12}x_2 + k_{13}x_3 \tag{51}$$

$$\dot{x}_2 = k_{21}x_1 - (k_{12} + k_{32})x_2 + k_{23}x_3 \tag{52}$$

$$\dot{x}_3 = k_{31}x_1 + k_{32}x_2 - (k_{13} + k_{23})x_3. \tag{53}$$

It can be written in the form $\dot{x} = Kx$, where the matrix of rate constants is given by

$$K = \begin{bmatrix} -(k_{21} + k_{31}) & k_{12} & k_{13} \\ k_{21} & -(k_{12} + k_{32}) & k_{23} \\ k_{31} & k_{32} & -(k_{13} + k_{23}) \end{bmatrix},$$

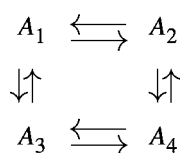$$K = \begin{bmatrix} -14.07 & 4.62 & 1.0 \\ 10.34 & -10.24 & 3.37 \\ 3.72 & 5.62 & -4.37 \end{bmatrix}. \tag{54}$$

The individual elements of the K matrix are not totally independent as is shown in the above equation. However, the proposed algorithm recovers the full matrix K. The data are the concentration of species as a function of time given in figure 3. We choose $m = 22$ time intervals with $\tau_1 = 0.04$, and $\tau_{j+1} = \tau_j + 0.03$, for $j = 1, 21$. If we consider the given data in figure 3, they suggest that the system reaches the steady state condition after about $t = 0.6$. If we included additional time intervals for $\tau_j > 0.6$, the vectors η_j generated from these additional simulations would be linearly dependent. These simulations would not add any new infomation to the matrix C through the outer product. For this case the singular values of the matrix C are given by $\sigma_1 = 1.31$, $\sigma_2 = 0.009\,84$, $\sigma_3 = 0.513 \times 10^{-5}$, and the matrix can readily be inverted. Table 1 shows the convergence of the elements of the matrix K. After only ten iterations the method is able to recover a very close approximation to the unknown rate constant matrix. At every iteration the error can be computed. It is given by

$$\text{error} = \sum_{k=1}^{r} v(k\Delta t)^{\top} v(k\Delta t), \tag{55}$$

where $T = r\Delta t$ is the time duration of the given data. The error is given in the second column.

Table 1. Convergence of the rate constant matrix for example 1. The reduction of the error is also presented.

Iter.	Error	K_{11}	K_{12}	K_{13}	K_{21}	K_{22}	K_{23}	K_{31}	K_{32}	K_{33}
1	0.789E+02	−13.336	4.285	1.003	1.201	−6.117	3.379	16.079	0.055	−4.385
2	0.105E+00	−14.026	4.597	1.002	8.633	−9.458	3.364	6.408	4.402	−4.369
3	0.421×10^{-2}	−14.075	4.622	1.000	10.021	−10.093	3.368	4.295	5.358	−4.369
4	0.207×10^{-3}	−14.073	4.621	1.000	10.280	−10.213	3.370	3.845	5.563	−4.370
5	0.105×10^{-4}	−14.071	4.620	1.000	10.329	−10.235	3.370	3.747	5.608	−4.370
6	0.558×10^{-6}	−14.070	4.620	1.000	10.338	−10.239	3.370	3.726	5.617	−4.370
7	0.302×10^{-7}	−14.070	4.620	1.000	10.340	−10.240	3.370	3.721	5.619	−4.370
8	0.167×10^{-8}	−14.070	4.620	1.000	10.340	−10.240	3.370	3.720	5.620	−4.370
9	0.940×10^{-10}	−14.070	4.620	1.000	10.340	−10.240	3.370	3.720	5.620	−4.370
10	0.538×10^{-11}	−14.070	4.620	1.000	10.340	−10.240	3.370	3.720	5.620	−4.370

$$A_1 \rightleftarrows A_2$$
$$\updownarrow \qquad\qquad \updownarrow$$
$$A_3 \rightleftarrows A_4$$

Figure 2. A chemical reaction involving four species.

Example 2. We next consider the reaction system given by figure 2.

The given data are the time history of the concentration of species which correspond to the actual rate constants, K, given by

$$K = \begin{bmatrix} -53 & 1 & 0 & 25 \\ 3 & -21 & 15 & 0 \\ 0.0 & 20 & -17 & 4 \\ 50 & 0 & 2 & -29 \end{bmatrix}. \tag{56}$$

Again we use $m = 22$ time intervals $\tau_1 = 0.04$, $\tau_{j+1} = \tau_j + 0.03$, for $j = 1, 21$. The singular values of the matrix C are $\sigma_1 = 0.951$, $\sigma_2 = 0.009\,81$, $\sigma_3 = 0.12 \times 10^{-5}$ and $\sigma_4 = 0.46 \times 10^{-9}$. We let the cut-off value be 0.1×10^{-9} and we can readily obtain the psuedoinverse matrix. After 14 iterations the algorithm converges to the matrix of rate constants given by

$$K = \begin{bmatrix} -52.999 & 0.999 & 0.001 & 25.000 \\ 2.998 & -20.997 & 14.997 & 0.001 \\ -0.003 & 20.004 & -17.003 & 4.002 \\ 50.001 & -0.002 & 2.001 & -29.001 \end{bmatrix}. \tag{57}$$

The results compare well to the actual rate constant matrix. This is quite an improvement over methods that are based on first-order necessary conditions obtained from the point of view of the minimization of a cost functional [17]. We considered this system in [17] and, based on the same observations, we needed 47 000 iterations to recover the K matrix.

Example 3. In this example we study the case in which the given data are noisy. Consider the reaction system in example 1 and assume that the given data are corrupted as shown in figure 4. If we simply use the noisy data and apply the scheme, then table 2 shows the convergence of the rate constants and the algorithm converges to a rate constant matrix given by

$$K = \begin{bmatrix} -11.091 & 3.582 & 0.856 \\ 12.147 & -10.896 & 3.192 \\ 4.464 & 5.358 & -4.482 \end{bmatrix}. \tag{58}$$
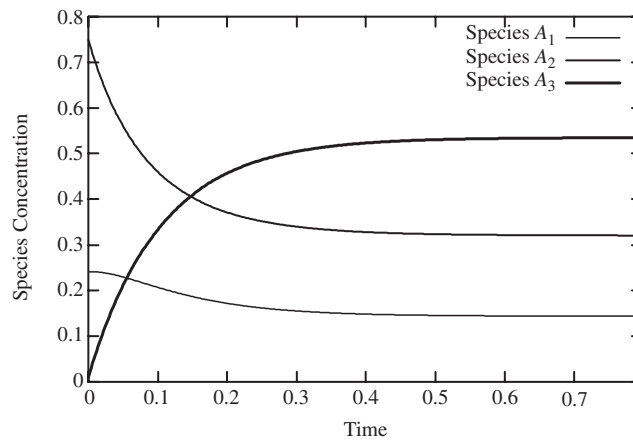
Figure 3. The concentration of the species for example 1 as a function of time. These data are provided as observation.

Table 2. Convergence of the rate constant matrix for example 3 with noisy data.

Iter.	Error	K_{11}	K_{12}	K_{13}	K_{21}	K_{22}	K_{23}	K_{31}	K_{32}	K_{33}
1	0.401E+2	−10.514	3.301	0.866	4.460	−7.222	3.109	14.606	0.531	−4.375
2	0.888×10^{-1}	−11.057	3.559	0.861	10.907	−10.269	3.157	6.562	4.337	−4.449
3	0.144×10^{-1}	−11.090	3.580	0.857	11.932	−10.785	3.185	4.903	5.142	−4.474
4	0.918×10^{-2}	−11.091	3.582	0.856	12.110	−10.876	3.191	4.559	5.311	−4.481
5	0.839×10^{-2}	−11.091	3.582	0.856	12.141	−10.892	3.192	4.485	5.347	−4.482
6	0.823×10^{-2}	−11.091	3.582	0.856	12.146	−10.895	3.192	4.468	5.355	−4.482
7	0.819×10^{-2}	−11.091	3.582	0.856	12.147	−10.896	3.192	4.465	5.357	−4.482
8	0.818×10^{-2}	−11.091	3.582	0.856	12.147	−10.896	3.192	4.464	5.358	−4.482
9	0.817×10^{-2}	−11.091	3.582	0.856	12.147	−10.896	3.192	4.464	5.358	−4.482
10	0.817×10^{-2}	−11.091	3.582	0.856	12.147	−10.896	3.192	4.464	5.358	−4.482

The scheme was able to reduce the error by about four orders of magnitude; however, there exists a persistent error of 0.817×10^{-2}. The present scheme is based on using the data directly, and, as a result, the error in the data affects the accuracy of the result. If we use a filter to reduce the oscillations before it is used in the inversion algorithm, then the accuracy of the results can be further improved. We next use a local averaging to smooth out the given data. We use a five-point local averaging in which the data are regenerated using

$$\hat{y}(j) = \frac{y(j-2) + y(j-1) + y(j) + y(j+1) + y(j+2)}{5}. \tag{59}$$

Table 3 shows the convergence of the rate constant matrix for these data. The error has been reduced further and the accuracy of the results has improved. The recovered matrix of rate constants is given by

$$K = \begin{bmatrix} -14.293 & 4.836 & 1.009 \\ 8.506 & -9.406 & 3.340 \\ 3.021 & 5.932 & -4.413 \end{bmatrix}. \tag{60}$$
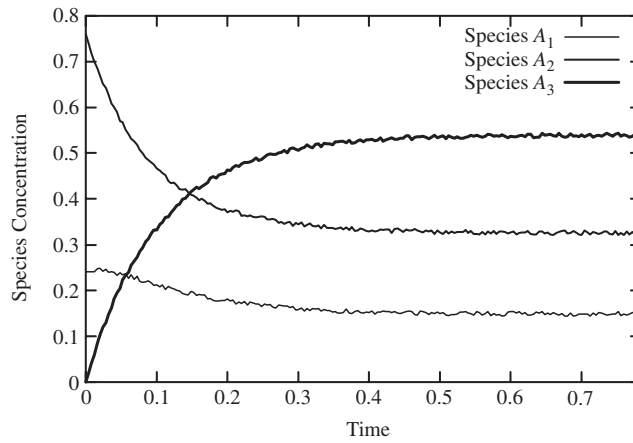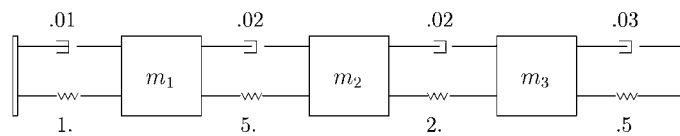
Figure 4. Noisy data that are provided for example 3.

Table 3. Convergence of the rate constant matrix for example 3 for the case in which the noisy data are filtered using a local averaging.

Iter.	Error	K_{11}	K_{12}	K_{13}	K_{21}	K_{22}	K_{23}	K_{31}	K_{32}	K_{33}
1	0.401E+2	−13.512	4.451	1.026	1.467	−6.033	3.258	13.027	1.161	−4.302
2	0.666×10^{-1}	−14.244	4.804	1.015	7.375	−8.833	3.308	5.001	4.962	−4.376
3	0.359×10^{-2}	−14.292	4.834	1.009	8.310	−9.305	3.334	3.425	5.732	−4.404
4	0.864×10^{-3}	−14.294	4.836	1.009	8.472	−9.388	3.339	3.106	5.890	−4.411
5	0.714×10^{-3}	−14.294	4.836	1.009	8.500	−9.403	3.340	3.039	5.923	−4.413
6	0.702×10^{-3}	−14.293	4.836	1.009	8.505	−9.405	3.340	3.025	5.930	−4.413
7	0.701×10^{-3}	−14.293	4.836	1.009	8.505	−9.406	3.340	3.022	5.932	−4.413
8	0.700×10^{-3}	−14.293	4.836	1.009	8.506	−9.406	3.340	3.021	5.932	−4.413
9	0.700×10^{-3}	−14.293	4.836	1.009	8.506	−9.406	3.340	3.021	5.932	−4.413
10	0.700×10^{-3}	−14.293	4.836	1.009	8.506	−9.406	3.340	3.021	5.932	−4.413

Example 4. We next study an example from mechanical vibrations. Consider a second-order mass–spring system given by

$$M\ddot{x} + D\dot{x} + Kx = f(t), \qquad M = I, \qquad f(t) = 0$$

The mass matrix is assumed to be known and, here, it is taken to be identity. The problem is then to recover the stiffness and damping matrices from the knowledge of the system response. We can rewrite the equation in a first-order form given by

$$\begin{bmatrix} \dot{x} \\ \ddot{x} \end{bmatrix} = \begin{bmatrix} 0 & I \\ -K & -D \end{bmatrix} \begin{bmatrix} x \\ \dot{x} \end{bmatrix}$$

$$K = \begin{bmatrix} k_1 + k_2 & -k_2 & 0 \\ -k_2 & k_2 + k_3 & -k_3 \\ 0 & -k_3 & k_3 + k_4 \end{bmatrix}, \qquad D = \begin{bmatrix} d_1 + d_2 & -d_2 & 0 \\ -d_2 & d_2 + d_3 & -c_3 \\ 0 & -d_3 & d_3 + d_4 \end{bmatrix}.$$

We assume that the response of the system to a nonzero initial condition is known and we use the algorithm to recover the system matrices. We use the same number of time intervals and, for this case, the eigenvalues of the C matrix are given by

$$\sigma_i = \underline{34.5}, \qquad \underline{3.73}, \qquad \underline{1.41}, \qquad \underline{0.727}, \qquad \underline{0.0748}, \qquad \underline{0.002\,87}.$$

Note that the matrix has a full rank and we can simply invert it. After two iterations we have

$$\begin{bmatrix}
0.008 & 0.021 & 0.006 & 1.016 & 0.010 & -0.010 \\
-0.009 & -0.024 & -0.006 & -0.019 & 0.988 & 0.012 \\
0.002 & 0.005 & 0.001 & 0.004 & 0.002 & 0.997 \\
-6.019 & 5.018 & 0.083 & 0.018 & 0.069 & -0.006 \\
5.022 & -7.021 & 1.905 & -0.035 & -0.096 & 0.027 \\
-0.005 & 2.004 & -2.480 & 0.011 & 0.032 & -0.052
\end{bmatrix}$$

and after only six iterations we have

$$\begin{bmatrix}
0.0 & 0.0 & 0.0 & 1.0 & 0.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\
0.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \\
-6.0 & 5.0 & 0.0 & -0.03 & 0.02 & 0.0 \\
5.0 & -7.0 & 2.0 & 0.02 & -0.04 & 0.02 \\
0.0 & 2.0 & -2.5 & 0.00 & 0.02 & -0.05
\end{bmatrix}.$$

Note that we do not enforce any structure on the matrix. The number of iterations needed for the recovery is quite attractive. Examples 2 and 4 show that the number of iterations needed does not significantly increase as the order of the system increases.

5. Conclusion

In this paper we have presented an iterative method to recover the matrix of rate constants for a monomolecular chemical reaction. The algorithm uses a series of simulations after which it is possible to directly invert for the sought for unknown matrix. We have used a number of numerical examples to describe the applicability of the proposed method. The algorithm requires only a few iterations. It can also recover a close approximation to the unknown rate constants for which the given data are noisy.

Acknowledgments

References

[1] Deuflhard P and Bock H G 1981 Modelling of chemical reaction systems *Numerical Treatment of Inverse Problems in Chemical Reaction Kinet.* ed K H Ebert (New York: Springer)
[2] Milstein J 1981 Modeling of chemical reaction sytems *The Inverse Problem: Estimation of Kinetic Parameters* ed K H Ebert and P Deuflhard (New York: Springer)
[3] Hanson R K and Salimian S 1984 *Combustion Chemistry* ed W C Gardiner (New York: Springer)
[4] Gelb A, Kasper J F, Nash R A, Price C F and Sutherland A A 1974 *Applied Optimal Estimation* (Cambridge, MA: MIT Press)
[5] Ross G J S 1990 *Nonlinear Estimation* (New York: Springer)
[6] Sage A P and White C C 1977 *Optimum Systems Control* (Englewood Cliffs, NJ: Prentice-Hall)

[7]   Allen M T, Yetter R A and Dryer F L 1995 The decomposition of nitrous oxide at 1.5–10.5 atm and 1130–1173 K
        Int. J. Chem. Kinet. **27** 883–909
[8]   Vakhtin A B 1996 The rate constant for the recombination of trifluoromethyl radicals at $T = 296$ K *Int. J. Chem.
        Kinet.* **28** 443–52
[9]   Ji Xingzhi 1998 On matrix inverse eigenvalue problems *Inverse Problems* **14** 275–85
[10] Starek L and Inman D J 1991 On the inverse vibration problem with rigid-body modes *ASME J. Appl. Mech.* **58**
        1101
[11] Chu M T 1990 Solving additive inverse eigenvalue problems for symmetric matrices by the homotopy method
        IMA J. Numer. Anal. **9** 331–42
[12] Gladwell G M L 1996 Inverse problems in vibration—II *Appl. Mech. Rev.* **49** S25–34
[13] James Wei and Prater C D 1962 The structure and analysis of complex reaction systems *Advances in Catalysis*
        13 ed D D Eley, P W Selwood and P B Weisz (New York: Academic) pp 203–392
[14] Potter J E 1966 Matrix quadratic solutions *SIAM J. Appl. Math.* **14** 496–501
[15] Laub A J 1979 A Schur method for solving algebraic Riccati equations *IEEE Trans. Automat. Control* **24** 913–21
[16] Liu Ruey-Wen and Leake R J 1966 Exhaustive equivalence classes of optimal systems with separable controls
        SIAM J. Control **4** 678–85
[17] Tadi M 1998 An inverse problem for linear chemical reactions *Inverse Problems Eng.* **6** 15–31