A NUMERICAL METHOD FOR THE IDENTIFICATION OF LINEAR COMPARTMENTAL SYSTEMS*

SHAO XIU-MIN (邵秀民)
(*Institute of Mathematics, Academia Sinica, Beijing, China*)

Abstract

This paper deals with an inverse problem of a system of ordinary differential equations whose coefficient matrix is the so-called compartmental matrix. This problem arises in a variety of areas such as pharmacokinetics, biology, ecology, economics, and so on. A numerical method applicable to the cases of non-unique solutions is developed, by which the projection of the starting value of iteration onto the manifold of solution can be obtained. The convergence of the method is proved. A few examples are examined, which shows the effectiveness of the given method.

§ 1. Introduction

In recent years, a simple and effective mathematical model, called compartmental model, has been widely used in a variety of areas[1]. It originated from pharmacokinetics (which is still one of the major fields for use of the compartmental model up to now), and was then appllied gradually to other fields such as biology, ecology, medicine, chemistry, and even economics, to explore the quantitative law of transfer and exchange of mass or state. In biological and medical problems, it is used for explaining the processes of distribution, absorption, excretion or metabolism of medicinal, or physiological, or biochemical mass in organisms.

In this model, a system is considered to be made up of a finite number of parts. Each part is assumed to have a specified volume and a uniform distribution of mass (or a same state) at any time. The compartments interact with one another by exchanging the matter (or state). Exchange also occurs with the environment. To this exchange a certain law of conservation applies. A part satisfying the conditions mentioned above is called a compartment, and the whole system called compartmental system.

For instance, in pharmacological problems a part of human body which is homogeneous in drug density is taken as a compartment. (It is possible that different organs belong to one compartment, or one organ belongs to more than one compartment). The drug can diffuse between different compartments through biological membranes. Compartments are infused with drugs from environment by taking medicine or injection, and excrete drugs to environment through urine, excrement and sweat. The law of conservation of mass is valid.

In this paper the simplest case, the linear model, will be taken into consideration. The equations established according to the laws of conservation are usually of the following form:

* Received June 7, 1986.

$$\begin{cases} \dfrac{dx(t)}{dt} = Ax(t) + F(t), \\ x(0) = x_0, \end{cases} \tag{1.1}$$

where

$$\begin{aligned} & A = \{a_{ij}\}_{n\times n} \\ & a_{ij} = K_{ij}, \quad i \neq j \\ & a_{ii} = -K_{oi} - \sum_{\substack{j=1 \\ j\neq i}}^{n} K_{ji}, \ i, \quad j = 1, \cdots, n; \end{aligned} \tag{1.2}$$

K_{ij} is the rate coefficient from compartment j to compartment i, with subscript o denoting the environment; $K_{ij} \geqslant 0$. Therefore, the matrix A has the following properties:

1) Its off-diagonal entries are non-negative.
2) Its diagonal entries are negative.
3) It is diagonally dominant with respect to the columns,

$$|a_{ii}| \geqslant \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ji}.$$

We call this kind of matrix compartmental matrix.

In problem (1.2), $F(t)$, $x(0)$ and measurements of part or all of the components of $x(t)$

$$\hat{x}_i(t_j), \quad i = 1, \cdots, r, \ r \leqslant n; \ j = 1, \cdots, m$$

are given. Our goal is to find the coefficient matrix A according to known data, that is, to find the solution A of the following least squares problem:

$$\sum_{i=1}^{r} \sum_{j=1}^{m} |x_i(t_i, A) - \hat{x}_i(t_j)|^2 = \min_{A \in M_o}. \tag{1.3}$$

Here $x(t, A)$ is the solution of problem (1.2) with coefficient matrix A, and M_o is the set of compartmental matrices

$$M_o = \{A \mid a_{ii} < 0, \ a_{ij} \geqslant 0 \ (i \neq j), \ -a_{ii} \geqslant \sum_{\substack{j=1 \\ j\neq i}}^{n} a_{ji}, \ i, \ j = 1, \cdots, n\}. \tag{1.4}$$

This is a nonlinear least squares problem. The most common method for solving it is the Gauss-Newton method[4]. But, some difficulties arise:

The problem is ill-posed. Sometimes its solution is not unique. The examples in § 5 will show the following possibilities in the case that the solution exists:

1. The solution is unique.
2. There is a finite number of solutions
3. There is an infinite number of solutions, which often form a continuous manifold.

For the third case mentioned above, the general Gauss-Newton method can not be used.

Even if the solution is unique, the problem is ill-conditioned, which causes rather large computational error.

To mitigate these difficulties, a numerical method is developed in this paper. Suppose that a rough approximation of the solution is known. Take it as the starting point of iteration. Then an iterative technique based on Tikhonov's regularization[6] is given to approximate the projection of the starting value onto the manifold of the solution. The convergence of the method is proved. If the solution is unique, the projection is just the unique solution. In the other cases, it is reasonable to take it as an approximate value of the real solution. (In pharmacological problems, the starting approximate value mentioned above may be a so-called "animal solution", which can be obtained more easily than the "human body solution", because all its components can be measured from animal experiments. Or, if we want to derive the compartmental matrix A of somebody for a certain drug, and the compartmental matrix A' of another person for some other drug is already known, and these two problems have the same model, then the known matrix A' can be taken as an initial approximation for the matrix A.)

In Section 2, the method and its realization are described. The proof of the convergence of the iteration is given in Section 3. Section 4 is devoted to actual choice of the parameters in the method. Finally, in Section 5, a few examples are examined, whose numerical results show that the method is feasible for various cases.

§ 2. Description of the Method

Throughout this paper we will not distinguish the matrix A and the vector corresponding to it

$$\{a_{11}, \cdots, a_{1n}, a_{21}, \cdots, a_{2n}, \cdots, a_{n1}, \cdots, a_{nn}\}^T.$$

(Generally, A should be considered as a vector rather than a matrix, except in some obvious cases.) Let $y(t, A) := D\times(t, A)\delta A$, where $D\times(t, A)$ represents the Jacobian matrix (with respect to A) of the vector $x(t, A)$. Then the formulas of the Gauss-Newton method are as follows:

$$A^{p+1} = A^p + \delta A, \quad p = 0, 1, \cdots,$$
$$\sum_{i=1}^{r}\sum_{j=1}^{m} |x_i(t_j, A^p) - \hat{x}_i(t_j) + y_i(t_j, A^p)|^2 = \min_{A^p+\delta A\in M_0}. \tag{2.1}$$

It is easy to prove:

Theorem 1. *Let $F(t)$ be continuous on $0\le t<\infty$, $A^p\in M_0$, A be in a neighborhood of A^p, and $x^p := x(t, A^p)$ be the solution of the initial value problem*

$$\begin{cases} \dfrac{dx}{dt} = A^p x + F, & 0<t<\infty \\ x(0, A^p) = x_0 \end{cases} \tag{2.2}$$

and $\delta A := A - A^p$. Then $y(t, A^p) := Dx(t, A^p)\delta A$ is the solution of the problem

$$\begin{cases} \dfrac{dy}{dt} = A^p y + \delta A\cdot x^p, \\ y(0, A^p) = 0. \end{cases} \tag{2.3}$$

This result makes it possible to get $y(t_j, A^p) := Dx(t_j, A^p)\delta A$ by using the numerical solution of (2.3). In fact, we can express the solution of (2.3) as follows:

$$y(t, A^p) = \sum_{k=1}^{n}\sum_{l=1}^{n} \delta a_{kl} y^{(k,l)}(t, A^p) \tag{2.4}$$

where $y^{(k,l)}(t, A^p)$ $(k, l = 1, \cdots, n)$ are the solutions of

$$\begin{cases} \frac{d}{dt} y^{(k,l)}(t, A^p) = A^p y^{(k,l)}(t, A^p) + x_l^p e_k, \\ y^{(k,l)}(0, A^p) = 0 \end{cases} \tag{2.5}$$

and e_k the k-th coordinate vector. Then the values $y^{(k,l)}(t_j, A^p)$ $(j=1, \cdots, m)$ can be obtained by solving (2.5) numerically, and, consequently, $y(t_j, A^p)$ $(j=1, \cdots, m)$ from (2.4). In such a way, the linear least squares problem (2.1) is reduced to the problem

$$\sum_{i=1}^{r}\sum_{j=1}^{m} |x_i(t_j, A^p) - \hat{x}_i(t_j) + \sum_{k=1}^{n}\sum_{l=1}^{n} \delta a_{kl} y_i^{(k,l)}(t_j, A^p)|^2 = \min_{\delta A \in M_c^p} \tag{2.6}$$

where $\delta A = \{\delta a_{kl}\}_{n\times n}$,

$$M_c^p = \{\delta A \mid A^p + \delta A \in M_c\}.$$

Introduce the notations

$$\begin{aligned} &j' = (k-1)n + l, \quad k, l = 1, \cdots, n, \\ &i' = (i-1)r + j, \quad i = 1, \cdots, r, \; j = 1, \cdots, m, \\ &n' = n^2, \quad m' = r \times m \end{aligned}$$

and the $m' \times n'$ matrix

$$C(A) = \{c_{i'j'}(A)\}, \quad c_{i'j'}(A) = y_i^{(k,l)}(t_j, A) \tag{2.7}$$

and the m'-dimensional vector

$$\begin{aligned} &\hat{x} = \{\hat{x}_{i'}\}, \quad \hat{x}_{i'} = \hat{x}_i(t_j), \\ &x^p = \{x_{i'}^p\}, \quad x_{i'}^p = x_i^p(t_j, A^p). \end{aligned} \tag{2.8}$$

Then (2.6) can be rewritten as

$$\|x^p - \hat{x} + C(A^p)\delta A\|_E^2 = \min_{\delta A \in M_c^p} \tag{2.6$'$}$$

where $\|\cdot\|_E$ is the Euclidean norm of a vector.

As mentioned in § 1, in general, problem (2.6)$'$ is ill-posed, or at least ill-conditioned. A certain kind of regularization techniques is necessary. Taking Tikhonov's idea for regularization, we choose the parameters α^p (the principle for choosing it will be given in § 3 and § 4) and turn to the following least squares problem

$$\|x^p - \hat{x} + C(A^p)\delta A\|_E^2 + \alpha^p \|A^{p+1} - A^0\|_E^2 = \min_{\delta A \in M_c^p}.$$

Then we have the iterative formula

$$A^p \in M_c, \quad p = 0, 1, \cdots,$$

$$\begin{cases} A^{p+1} = G(A^p, \alpha^p), \\ G(A^p, \alpha) = A^p + \{C^T(A^p)C(A^p) + \alpha I\}^{-1}\{C^T(A^p)(\hat{x} - x^p) + \alpha(A^0 - A^p)\}. \end{cases} \tag{2.9}$$

Now, we summarize the procedure of the algorithm mentioned above as follows:

1) Select a starting point $A^0 \in M_c$.

2) With $A^p (p=0, 1, \cdots)$ find numerically the solution of the initial value problem

$$\begin{cases} \dot{x} = A^p x + F, \\ x(0) = x_0 \end{cases}$$

and denote the obtained solution as $x^p(t_j)$ $(j=1, \cdots, m)$.

3) For all k, $l=1, \cdots, n$ such that $a^p_{kl}\neq 0$, solve numerically the initial value problems

$$\begin{cases} \dot{y}^{(k,l)}=A^p y^{(k,l)}+x^p_l e_k, \\ y^{(k,l)}(0)=0 \end{cases}$$

to get $y^{(k,l)}(t_j, A^p)$ $(j=1, \cdots, m)$, that is, get $C(A^p)$.

4) Choose regularization parameters $\alpha^p>0$, and solve the least squares normal equations

$$\{C^T(A^p)C(A^p)+\alpha^p I\}\delta A=C^T(A^p)(\hat{x}-x^p)+\alpha^p(A^0-A^p) \tag{2.10}$$

to produce δA.

5) A^{p+1} is obtained from $A^{p+1}=A^p+\delta A$.

6) It is required that the matrix A^{p+1} keeps the compartmental form. If not so, alter it in such a way:

1° If a certain off-diagonal entry is negative, let it be zero.

2° If a certain diagonal entry is non-negative, or although it is negative, its absolute value is smaller than the sum of the entries in the same column, then let it be negative sum of those entries.

7) Repeat 2)—6).

As will be seen in § 3, in the above procedure, 6) is necessary to ensure the convergence of the iterative process.

Of all steps mentioned above, two require a great deal of computational work:

1) The numerical solution of the systems (2.5) of ODEs: We select their computational method according to the following principles:

1° The method must be simple and fast, since the number of systems is quite large,

2° Appropriate accuracy of computation is necessary, since inverse problems are generally ill-conditioned,

3° But high accuracy methods are not necessary, because measurements are usually inaccurate.

Therefore, fast methods of moderate accuracy are suitable for our problem. In our calculation, the extrapolation based on the Euler method has been taken[5].

2) Solving the system (2.10) of algebraic equations: Considering the number of operations and storage, the QR-decomposition is preferable[2].

§ 3. The Convergence

First, notice the following important properties of the compartmental matrix:

Lemma 1. *If $A\in M_c$, then the eigenvalues λ_i of A satisfy*

$$\operatorname{Re}\lambda_i<0, \quad i=1, \cdots, n.$$

Proof. It is valid because a compartmental matrix is diagonally dominant, and its diagonal entries are negative.

Lemma 2. *If $A\in M_c$, then* $\exp(tA)$ *is bounded for all $t\geqslant 0$.*

Proof. Let $\lambda_1, \lambda_2, \cdots, \lambda_n$ be the eigenvalues of A, and J its Jordan normal form, and S a nonsingular matrix to reduce A to J, i.e.,

$$SAS^{-1}=J=\begin{bmatrix} J_1 & & \\ & \ddots & \\ & & J_r \end{bmatrix}.$$

Then

$$\|e^{tA}\|_2=\|Se^{tJ}S^{-1}\|_2\leqslant C\|e^{tJ}\|_2.$$

There exists a certain i_0, $1\leqslant i_0\leqslant r$, such that

$$\|e^{tJ}\|_2=\|e^{tJ_{i_0}}\|_2.$$

Let the eigenvalue in J_{i_0} be λ_{i_0}. Then

$$\|e^{tJ_{i_0}}\|_2\leqslant C\exp(t\operatorname{Re}\lambda_{i_0}).$$

From Lemma 1, we can obtain

$$\|e^{tJ_{i_0}}\|_2\leqslant C'.$$

Therefore,

$$\|e^{tA}\|_2\leqslant C'.$$

The boundedness of $\exp(tA)$ with other norms can be obtained from norm equivalence.

In the proof of our convergence theorem some results of the continuation method will be used. Let the homotopy defining the continuation process be given in the form

$$H(x,\ t)=x-G(x,\ t)=0,\quad t\in[0,\ 1]. \tag{3.1}$$

We have the following.

Theorem[4]. *Suppose that* G: $D\times[0,\ 1]\subset R^n\times R^1\to R^n$ *is Frechet-differentiable with respect to the first variable and that its Frechet-derivative* $\partial_x G(x,\ t)$ *is continuous on* $D\times[0,\ 1]$. *Assume further that* (3.1) *has a continuous solution* x: $[0,\ 1]\to\operatorname{int}(D)$ with known initial point $x^0=x(0)$, *and that* $\rho(\partial_x G(x(t),\ t))<1$ *for all* $t\in[0,1]$. *Then there exists a partition of* $[0,\ 1]$

$$0=t_0<t_1<t_2<\cdots<t_N=1$$

and integers $m_1,\cdots,m_{N-1}$ *such that the entire sequence* $\{x^{i,k}\}$ *defined by*

$$x^{i,k+1}=G(x^{i,k},\ t_i),\quad k=0,\ 1,\ \cdots,\ m_i-1,\ i=1,\ 2,\ \cdots,\ N-1,$$

$$x^{1,0}=x^0,\quad x^{i+1,0}=x^{i,m_i},$$

$$x^{N,k+1}=G(x^{N,k},\ 1),\quad k=0,\ 1,\ \cdots$$

remains in D *and*

$$\lim_{k\to\infty}x^{N,k}=x(1).$$

Definition. *A functional* $\Phi(z)$ *is called quasi-monotone in domain* S *if it has no local extrèmums in* S *except global minimum.*

The iterative process described in § 2 is not continuous because of the procedure 6). We can change it into a continuous process, if the constrained least squares problem is considered (requiring inequality constraints (1.4) to be satisfied). But in both cases, the proof of convergence is complicated. For the sake of simplicity, we will consider the unconstrained least squares problem and give proof under the stronger assumption that (continuous) iteration (2.9) keeps $A^p(p=0,\ 1,\ \cdots)$ compartmental matrices.

For brevity, denote $C(A)=C$, $C(A^*)=C^*$. x^* is the vector replacing A^p in (2.8) with A^*, and $x(A)$ vector replacing A^p with A, and so on.

In the following discussion, we will only use the Euclidean norms of vector and matrix, and simply write them as $\|\cdot\|$ instead of $\|\cdot\|_E$.

Theorem 2 (Convergence theorem). *Let $F:[0, \infty)\in R^1 \to R^n$ in (1.1) be continuous on $[0, \infty)$, and A^* satisfy*

$$\hat{x}-x^*=0$$

(i.e., A^* is a point on the manifold of solution. We denoted this manifold by M_A.) Further, let A^* be a projection of a certain A^0 on M_A

$$\|A^*-A^0\|=\inf_{A\in M_A}\|A-A^0\|$$

and the projection is unique. Assume that in a neighborhood S of A^* the functional

$$\Phi(A, \alpha)=\|x(A)-\hat{x}\|^2+\alpha\|A-A^0\|^2$$

is quasi-monotone for every $\alpha>0$. And assume that the iteration (2.9) keeps A^p $(p=0, 1, \cdots)$ compartmental matrices. Then there exists a neighborhood $S(A^*, \delta)\subset S$ of A^* and a sequence of positive number $\{\alpha^p\}$, $\alpha^p\to 0\,(p\to\infty)$, such that for any $A^0\in S(A^*, \delta)$ the sequence $\{A^p\}$ defined by (2.9) remains in S and converges to A^*

$$\lim_{p\to\infty} A^p=A^*.$$

Proof. 1° The entries of matrix $C(A^p)$ are

$$e_i^T\int_0^{t_j} e^{A^p(t-\tau)}\left\{e_l^T\int_0^{\tau} e^{A^p(\tau-\xi)}F(\xi)e_k d\xi\right\}d\tau. \tag{3.2}$$

According to the hypothesis, F is continuous. From Lemma 2, $e^{A^p}(t-\tau)$ and $e^{A^p(\tau-\xi)}$ are bounded. So the integrals are well-defined. Again, in the expression of $G(A^p, \alpha)$, matrix $C^T(A^p)C(A^p)+\alpha I$ is positive definite when $\alpha>0$, so $\{C^T(A^p)C(A^p)+\alpha I\}^{-1}$ exists. Therefore, $G(A^p, \alpha)$ makes sense.

2° Now we give the expression of Frechet-derivative $\partial_A G(A, \alpha)$ of $G(A, \alpha)$ with respect to the first variable A.

$$\begin{aligned}\partial_A G(A, \alpha)&=I+D(C^TC+\alpha I)^{-1}\{C^T(\hat{x}-x)+\alpha(A^0-A)\}\\&\quad+(C^TC+\alpha I)^{-1}\{DC^T(\hat{x}-x)-C^TC-\alpha I\}\\&=D(C^TC+\alpha I)^{-1}\{C^T(\hat{x}-x)+\alpha(A^0-A)\}+(C^TC+\alpha I)^{-1}DC^T(\hat{x}-x),\end{aligned} \tag{3.3}$$

where D denotes the Frechet-derivative of the corresponding matrix with respect to the vector A, which is a tridimensional tensor. For example, write the vector A into $A=(a_1, a_2, \cdots, a_{n'})^T$; then

$$DC(A)=\left(\frac{\partial c_{ij}(A)}{\partial a_s}\right),\quad i=1, \cdots, m';\ j, s=1, \cdots, n'.$$

By the general formula of Frechet-derivative of the inverse of matrices, we have

$$\begin{aligned}D(C^TC+\alpha I)^{-1}&=-(C^TC+\alpha I)^{-1}D(C^TC+\alpha I)(C^TC+\alpha I)^{-1}\\&=-(C^TC+\alpha I)^{-1}(DC^T\cdot C+C^TDC)(C^TC+\alpha I)^{-1}.\end{aligned} \tag{3.4}$$

Substituting it into (3.3), we obtain the expression of $\partial_A G(A, \alpha)$. $\partial_A G(A, \alpha)$ exists and is continuous with respect to A and α for all $\alpha>0$. If the solution of the original problem is unique, then C^TC is nonsingular. In this case, $\partial_A G(A, \alpha)$ exists and is continuous also for $\alpha=0$.

3° Consider the equation

$$H(A, \alpha)=A-G(A, \alpha)=(C^TC+\alpha I)^{-1}\{C^T(\hat{x}-x)+\alpha(A^0-A)\}=0 \tag{3.5}$$

which is satisfied by the fixed point of the iteration $A^{p+1}=G(A^p, \alpha)$. When $\alpha\to\infty$, it has the solution $A=A^0$. From (3.3) and (3.4), $\partial_A G(A^0, \infty)=0$. Therefore,

$$\partial_A H(A^0, \infty)=I-\partial_A G(A^0, \infty)=I.$$

4° For the convenience of discussion, introduce the change of variable

$$\tau=\frac{1}{1+\alpha}, \quad \text{i.e.,} \quad \alpha(\tau)=\frac{1}{\tau}-1.$$

Then τ varies from 0 to 1 as α varies from ∞ to 0. After that transformation (3.5) is reduced to

$$\overline{H}(A, \tau)=A-G(A, \alpha(\tau))=0. \tag{3.5$'$}$$

It can be proved that there is a continuous solution of (3.5)′ when τ moves on $[0, 1)$. In fact, $(A^0, 0)$ satisfies (3.5)′, and $\overline{H}(A, \tau)$ is continuous, and $\partial_A \overline{H}(A, \tau)$ exists on an open neighborhood of $(A^0, 0)$, and $\partial_A \overline{H}(A, \tau)$ is nonsingular and continuous at $(A^0, 0)$. By the implicit function theorem, there exist open neighborhoods S_1 and S_2 of A^0 and 0, respectively, such that for any $\tau\in S_2$, equation (3.5)′ has a unique solution $\widetilde{A}^\tau=h(\tau)\in S_1$, and the mapping h: $S_2\to R^{n'}$ is continuous. Again, taking a $\tau>0$ in S_2, neighboring $\tau=0$, we have $\widetilde{A}^\tau=h(\tau)$. From (3.3)

$$\partial_A\overline{H}(A^\tau, \tau)=I-\partial_A G(\widetilde{A}^\tau, \alpha(\tau)).$$

If we can prove (we will give this proof in 6°) that $\rho(\partial_A G(\widetilde{A}^\tau, \alpha(\tau)))<1$, then the matrix $\partial_A\overline{H}(\widetilde{A}^\tau, \tau)$ is nonsingular, and consequently, the previous procedure can be repeated. In this way, we will obtain a continuous solution curve

$$\widetilde{A}^\tau=h(\tau), \quad \tau\in[0, 1)$$

of equation (3.5)′ with initial point $A^0=h(0)$.

5° We prove that the end point of the continuous solution curve mentioned above is A^* when $\tau\to 1$. That is, for every sequence $\{\tau_p\}$, $\tau_p\to 1$ $(p\to\infty)$ (i.e., $\alpha_p=\alpha(\tau_p)\to 0$), the corresponding sequence of $\widetilde{A}^p$, $\widetilde{A}^p=h(\tau_p)$, has limit A^*

$$\lim_{p\to\infty}\widetilde{A}^p=A^*.$$

In fact, since $(\widetilde{A}^p, \tau)$ $(\tau\in[0, 1))$ is the solution of (3.5)′, it satisfies

$$f(\widetilde{A}^p, \tau)=C^T(\widetilde{A}^p)(\hat{x}-x(\widetilde{A}^p))+\alpha(\tau)(A^0-\widetilde{A}^p)=0. \tag{3.6}$$

So it is a critical point of the functional

$$\Phi(A, \tau)=\|x(A)-\hat{x}\|^2+\alpha(\tau)\|A-A^0\|^2.$$

Suppose that the radius of neighborhood S in the assumption of the theorem is R. Take A^0 such that $\|A^0-A^*\|\leqslant R/3$. Then $\widetilde{A}^\tau$ neighboring A^0 which was given in 4° must be in S. According to the assumption that $\Phi(A, \alpha)$ is quasi-monotone in S, $\widetilde{A}^\tau$ is a global minimizer of $\Phi(A, \tau)$. Thus

$$\begin{aligned}\alpha(\tau)\|\widetilde{A}^\tau-A^0\|^2&\leqslant\Phi(\widetilde{A}^\tau, \tau)\leqslant\Phi(A^*, \tau)\\&=\|x^*-x\|^2+\alpha(\tau)\|A^*-A^0\|^2=\alpha(\tau)\|A^*-A^0\|^2.\end{aligned}$$

So

$$\|\widetilde{A}^\tau-A^0\|\leqslant\|A^*-A^0\|.$$

This inequality and the selection of A^0 ensure that for all $\tau\in[0, 1)$, $\widetilde{A}^\tau$ remains in S, i.e., the sequence $\{\widetilde{A}^p\}$ is entirely in S. It has property

$$\|\widetilde{A}^p-A^0\|\leqslant\|A^*-A^0\|. \tag{3.7}$$

This means that $\{\tilde{A}^p\}$ is bounded, and therefore, has a convergent subsequence. Denoting the subsequence also as $\{\tilde{A}^p\}$, we obtain

$$\lim_{p\to\infty} \tilde{A}^p = B.$$

Taking limit in (3.7), we have

$$\|B-A^0\| \leqslant \|A^*-A^0\|. \tag{3.8}$$

Moreover, $\qquad \|x(\tilde{A}^p)-\hat{x}\|^2 \leqslant \Phi(\tilde{A}^p, \tau_p) \leqslant \Phi(A^*, \tau_p) = \alpha(\tau_p)\|A^*-A^0\|^2.$

Let $p\to\infty$; then $\alpha_p\to 0$,

$$\|x(B)-\hat{x}\|^2 = 0. \tag{3.9}$$

By the assumption of the theorem, A^* also satisfies (3.9), just like B, and it has the smallest distance to A^0, and such a solution is unique. So the equality in (3.8) is valid, and $B=A^*$.

From the above discussion the limit of every convergent subsequence is A^*. Therefore, the sequence converges, and its limit is A^*,

$$\lim_{p\to\infty} \tilde{A}^p = A^*.$$

6° Now we prove that on the curve $A=h(\tau)$, $\tau\in[0, 1]$,

$$\rho(\partial_A G(A, \alpha(\tau)))<1.$$

From (3.3), at every point of this curve

$$\partial_A G(A, \alpha(\tau)) = [C^T(A)C(A)+\alpha(\tau)I]^{-1}DC^T(A)[\hat{x}-x(A)]. \tag{3.10}$$

Since $DC^T(A)$ and $x(A)$ are bounded, it is obvious that we only have to discuss the case that α is not very big. Now we estimate every column of the matrix $\partial_A G(A, \alpha(\tau))$. In the formula (3.10), $DC^T(A)[\hat{x}-x(A)]$ is a matrix. Take any one of its columns and denote it as $d(A)$. Suppose that the corresponding column of matrix $\partial_A G(A, \alpha(\tau))$ is z_α which is the solution of the following system of algebraic equations

$$[C^T(A)C(A)+\alpha(\tau)I]z_\alpha = d(A). \tag{3.11}$$

Consider systems

$$[C^T(A^*)C(A^*)+\alpha(\tau)I]z_\alpha^* = Qd(A) \tag{3.12}$$

and

$$C^T(A^*)C(A^*)z^* = Qd(A) \tag{3.13}$$

where $Qd(A)\in R(C^T(A^*)C(A^*))$. $R(C^T(A^*)C(A^*))$ is the range of values of the matrix $C^T(A^*)C(A^*)$, and Q is the orthogonal projection matrix on the subspace $R(C^T(A^*)C(A^*))$. We have

$$\|z_\alpha\| \leqslant \|z_\alpha - z_\alpha^*\| + \|z_\alpha^* - z^*\| + \|z^*\|. \tag{3.14}$$

Introduce the notations

$$\|C(A)-C(A^*)\| = h, \quad \|d(A)\| = \eta.$$

First, we make the estimation of $\|z^*\|$. Considering the normal solution (minimum-norm solution) of (3.13), we obtain

$$\|z^*\| \leqslant \|\{C^T(A^*)C(A^*)\}^+\| \|Q\| \|d\| = C_1\eta. \tag{3.15}$$

Here $C_1 = \|\{C^T(A^*)C(A^*)\}^+\|$.

Secondly, estimate $\|z_\alpha^* - z^*\|$. Let $\operatorname{rank}\{C^T(A^*)C(A^*)\} = k$. The positive eigenvalues of the matrix are $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_k > 0$, and the corresponding eigenvectors are $u_1, \cdots, u_k$,

$$C^T(A^*)C(A^*)u_j = \lambda_j u_j, \quad j=1, \cdots, k.$$

By (3.12), (3.13),

$$z_\alpha^* = \{C^T(A^*)C(A^*) + \alpha I\}^{-1} C^T(A^*)C(A^*)z^*.$$

Writing z^* in terms of the basis consisting of the above eigenvectors

$$z^* = \sum_{j=1}^{k} \beta_j u_j,$$

we get

$$z_\alpha^* = \sum_{j=1}^{k} \frac{\lambda_j \beta_j}{\lambda_j + \alpha} u_j.$$

Therefore,

$$z_\alpha^* - z^* = -\sum_{j=1}^{k} \frac{\alpha \beta_j}{\lambda_j + \alpha} u_j.$$

It follows that

$$\|z_\alpha^* - z^*\| \leqslant \frac{\alpha}{\lambda_k} \|z^*\| \leqslant \frac{C_1}{\lambda_k} \alpha\eta. \tag{3.16}$$

Finally, we estimate $\|z_\alpha - z_\alpha^*\|$. From (3.11), (3.12),

$$\begin{aligned}
\{C^T(A)C(A) + \alpha I\}(z_\alpha - z_\alpha^*) &= \{C^T(A)C(A) + \alpha I\}z_\alpha - \{C^T(A^*)C(A^*) + \alpha I\}z_\alpha^* \\
&\quad - \{C^T(A)C(A) - C^T(A^*)C(A^*)\}z_\alpha^* \\
&= (I-Q)d(A) - \{C^T(A)C(A) - C^T(A^*)C(A^*)\}z_\alpha^* \\
&= (I-Q)d(A) - \{C^T(A)[C(A) - C(A^*)] + [C^T(A) - C^T(A^*)]C(A^*)\}z_\alpha^*
\end{aligned}$$

which leads to

$$\|z_\alpha - z_\alpha^*\| \leqslant \frac{M_1\eta}{\alpha} + \frac{M_2 h}{\alpha} \|z_\alpha^*\|$$

considering the boundedness of the matrices $I-Q$, $C(A^*)$ and $C(A)$. Further, from (3.15), (3.16),

$$\begin{aligned}
\|z_\alpha - z_\alpha^*\| &\leqslant \frac{M_1\eta}{\alpha} + \frac{M_2 h}{\alpha}(\|z_\alpha^* - z^*\| + \|z^*\|) \\
&\leqslant \frac{M_1\eta}{\alpha} + \frac{M_2 h}{\alpha}\left(\frac{C_1}{\lambda_k}\alpha\eta + C_1\eta\right) \leqslant \frac{M_2 C_1}{\lambda_k} h\eta + (M_1 + M_2 C_1 h)\frac{\eta}{\alpha}.
\end{aligned}$$

Since $A = h(\tau)$ satisfies (3.6) and (3.7),

$$\alpha \geqslant \frac{\|C^T(A)[\hat{x} - x(A)]\|}{\|A^0 - A^*\|}.$$

Moreover,

$$\eta = \|d(A)\| \leqslant \|DC^T(A)[\hat{x} - x(A)]\|.$$

It is readily showed from Lemma 2 that there exists a constant C_2 such that

$$\frac{\|DC^T(A)[\hat{x} - x(A)]\|}{\|C^T(A)[\hat{x} - x(A)]\|} \leqslant C_2.$$

Therefore,

$$\|z_\alpha - z_\alpha^*\| \leqslant \frac{M_2 C_1}{\lambda_k} h\eta + (M_1 + M_2 C_1 h) C_2 \|A^0 - A^*\|. \tag{3.17}$$

Substituting (3.15), (3.16) and (3.17) into (3.14),

$$\|z_\alpha\| \leqslant C_1\eta + \frac{C_1}{\lambda_k}\alpha\eta + \frac{M_2 C_1}{\lambda_k} h\eta + (M_1 + M_2 C_1 h) C_2 \|A^0 - A^*\|. \tag{3.18}$$

So long as A^0 is close enough to A^*, h and η can be small enough, so that

$$\|z_a\|<\frac{1}{n'}.$$

It follows that

$$\|\partial_A G(A,\ \alpha(\tau))\|<1.$$

Then we have

$$\rho(\partial_A G(A,\ \alpha\ (\tau)))\leqslant\|\partial_A G(A,\ \alpha(\tau))\|<1.$$

7° Take a closed domain D, such that $D\subset S$ and $A=h(\tau)\subset D$ when $\tau\in[0,\ 1]$. On the domain D the hypothesises of Theorem[2] are all satisfied. (The continuity of the matrix $\partial_A G(A,\ \alpha(\tau))$ at $\tau=1$ is a consequence of 6°.) Applying this theorem, the desired conclusion can be obtained, i.e., there exists a sequence $\{\alpha^p\}$, $\alpha^p\to 0$ $(p\to\infty)$, and a corresponding sequence $\{A^p\}$, such that

$$\lim_{p\to\infty} A^p=A^*,$$

which completes the proof.

§4. Choice of $\{\alpha^p\}$

The questions about the prerequisites on the parameters $\{\alpha^p\}$ for ensuring the convergence and about the existence of such parameters have been discussed in Theorem 2. However, that result did not provide us with the actual values of these parameters. This problem will be settled in the present section.

It is easy to prove that the condition number of the system (2.10) of linear algebraic equations is

$$\mathrm{Cond}=\frac{\lambda_n+\alpha^p}{\lambda_1+\alpha^p},$$

where λ_1 is the smallest eigenvalue, and λ_n the largest eigenvalue of the matrix $C^T(A^p)C(A^p)$. It is a nonincreasing function of α^p. When $C^T(A^p)C(A^p)$ is singular, the condition number is infinite, if $\alpha^p=0$.

Therefore, from the consideration of stability, we can not get the "projection" solution A^*, which corresponds to $\alpha^p=0$. We can only get its approximation.

From the discussion in the preceding section and this section, the following way of determining the regularization parameters is proper:

1) α^0 should be rather large. The reasons are: (a) to ensure the convergence of the iteration (see Theorem 2); (b) to produce a faster iteration in approximating the needed solution, when the solutions are not unique; (c) to reduce the condition number;

2) Decrease α^p gradually.

3) When a given small value is attained by α^p, stop decreasing α^p.

In the numerical experiments with the examples shown in § 5, we changed the parameters as follows (of course, it is not a general regulation for the choice of parameters): Choose α^0 as 2 or 4; then reduce α successively at a rate $\alpha/4$, until it equals 0.0001. After that, keep α constant.

§ 5. Numerical Examples

In this section, a few artificial examples are constructed, by which the results obtained by using the method described above are compared with the theoretical solutions.

Example 1 (Unique solution). Two-compartmental model. The input $F(t)$ and the initial value $x(0)$ are

$$F(t)=\{-e^{-2t},\ e^{-t}\}^T,\quad x(0)=\{0,\ 2\}^T.$$

One of the components of the solution is known:

$$x_1(t)=te^{-t}.$$

Find the coefficient matrix A and another component $x_2(t)$.

The exact solution of this problem is

$$A=\begin{pmatrix}-1 & 1\\ 0 & -2\end{pmatrix},\quad x_2(t)=e^{-t}+e^{-2t}.$$

In computation, we take $m=24$, and

$$A^0=\begin{pmatrix}-0.95 & 1.2\\ 0 & -1.7\end{pmatrix}$$

and the times of iteration is 8. (From then on, the results can hardly be improved.) The comparison between the obtained results and the exact solution for A and $x_2(t)$, respectively, is presented in Table 1 and Table 2.

Table 1. A

	a_{11}	a_{12}	a_{21}	a_{22}
Exact sol.	−1.0000	1.0000	0.0000	−2.0000
Our results	−0.9956	1.0016	0.0000	−2.0035

Table 2. $x_2(t)$

t	0.1	0.3	0.5	0.7	0.9	1.2	1.6
Exact sol.	1.7236	1.2896	0.9744	0.7432	0.5719	0.3919	0.2427
Our results	1.7222	1.2871	0.9727	0.7392	0.5674	0.3878	0.2395
t	2.0	3.0	4.0	5.0	6.0	8.0	10.0
Exact sol.	0.1537	0.0523	0.0186	0.0067	0.0025	0.0003	0.0000
Our results	0.1514	0.0514	0.0184	0.0067	0.0024	0.0003	0.0000

Example 2 (Two solutions). Two-compartmental model. There is no output from the first compartment, so that the matrix A is of the form

$$A=\begin{bmatrix}-a_{21} & a_{12}\\ a_{21} & a_{22}\end{bmatrix}.$$

It is known that

$$F(t)=\begin{bmatrix}e^{-2t}\\ 2e^{-t}-te^{-t}-\dfrac{1}{2}e^{-2t}\end{bmatrix},\quad x(0)=\begin{bmatrix}0\\0\end{bmatrix},\quad x_1(t)=te^{-t}.$$

This problem has two solutions:

1. $$A=\begin{bmatrix}-3 & 2\\ 3 & -3\end{bmatrix},\quad x_2(t)=\frac{1}{2}(e^{-t}+2te^{-t}-e^{-2t});$$

2. $$A=\begin{bmatrix}-1 & \frac{2}{3}\\ 1 & -\frac{7}{3}\end{bmatrix},\quad x_2(t)=\frac{3}{2}(e^{-t}-e^{-2t}).$$

The following A^0 were chosen, respectively, in numerical experiments:

$$\begin{bmatrix}-2.9 & 2.2\\ 2.8 & -3.2\end{bmatrix},\quad \begin{bmatrix}-0.95 & 0.8\\ 0.9 & -2.2\end{bmatrix}$$

The results for the first solution are (iterative times=7)

Table 3. A

	a_{11}	a_{12}	a_{21}	a_{22}
First sol.	−3.0000	2.0000	3.0000	−3.0000
Our results	−3.0072	1.9798	3.0072	−3.0224

Table 4. $x_2(t)$

t	0.1	0.3	0.5	0.7	0.9	1.2	1.6
First sol.	0.1335	0.3182	0.4226	0.4726	0.4865	0.4667	0.4036
Our results	0.1337	0.3181	0.4217	0.4705	0.4832	0.4613	0.3961
t	2.0	3.0	4.0	6.0	8.0	10.0	12.0
First sol.	0.3292	0.1730	0.0823	0.0161	0.0028	0.0005	0.0000
Our results	0.3201	0.1627	0.0740	0.0123	0.0014	0.0000	0.0000

The results for the second solution are (iterative times=6):

Table 5. A

	a_{11}	a_{12}	a_{21}	a_{22}
Second sol.	−1.0000	0.6667	1.0000	−2.3333
Our results	−1.0026	0.6528	0.9225	−2.2167

Table 6. $x_2(t)$

t	0.2	0.4	0.6	0.8	1.0	1.2	1.6
Second sol.	0.2226	0.3315	0.3714	0.3711	0.3438	0.3157	0.2417
Our results	0.2245	0.3359	0.3778	0.3784	0.3560	0.3222	0.2458
t	2.0	3.0	4.0	5.0	6.0	8.0	10.0
Second sol.	0.1755	0.0710	0.0270	0.0100	0.0037	0.0005	0.0000
Our results	0.1772	0.0690	0.0246	0.0082	0.0025	0.0000	0.0000

Example 3 (Infinitely many solutions). Two-compartmental model. It is known that

$$F(t)=\begin{bmatrix}4te^{-t}\\ 4(2-t)e^{-t}\end{bmatrix},\quad x(0)=\begin{bmatrix}0\\ 2\end{bmatrix},\quad x_1(t)=4te^{-t}.$$

This problem has an infinite number of solutions. In fact, the matrices whose entries satisfy the following relations

$$\begin{cases}a_{12}=2,\\ a_{11}+a_{22}=-7,\\ a_{11}a_{22}-2a_{21}+a_{22}=3\end{cases}$$

are all solutions. One of them is

$$A=\begin{bmatrix}-2 & 2\\ 1 & -5\end{bmatrix}, \quad x_2(t)=2e^{-t}.$$

Taking $m=32$, and the initial point of iteration

$$A^0=\begin{bmatrix}-1.8 & 1.8\\ 0.8 & -4.6\end{bmatrix}$$

we obtained (iterative times $=9$)

Table 7. A

	a_{11}	a_{12}	a_{21}	a_{22}
One of sol.	−2.0000	2.0000	1.0000	−5.0000
Our results	−2.0138	1.9901	1.0004	−4.9314

Table 8. $x_2(t)$

t	0.2	0.4	0.6	0.8	1.0	1.2	1.6
One of sol.	1.6375	1.3406	1.0976	0.8987	0.7358	0.6024	0.4038
Our results	1.6530	1.3594	1.1155	0.9136	0.7479	0.6124	0.4104
t	2.0	3.0	4.0	5.0	7.0	9.0	11.0
One of sol.	0.2707	0.0996	0.0366	0.0135	0.0018	0.0002	0.0000
Our results	0.2745	0.1033	0.0373	0.0134	0.0017	0.0002	0.0000

Conclusion. This work is an attempt to compute inverse problems which have not unique solution. Obtained numerical results exhibit that the given method has a fair accuracy and hence is applicable in some cases. But its requirement for initial point seems to be slightly stricter. Widening the domain of convergence and heightening the accuracy will be the subjects of further work.

Acknowledgement. This work has been supported by the Deutsche Forschungsgemeinschaft, Sonderforschungsbereich 123. The following programs, worked out by the Numerical Analysis Group of Institute of Applied Mathematics, Heidelberg University, were used: EULEX, DECCON, SOLCON, SPLIKO, SPLIVA.

The author wishes to thank Prof. P. Deuflhard for the helpful discussions in the course of the work, and for reading the manuscript of this paper and giving important opinions.

The author thanks J. Zugck for his program, which was utilized in this work.

References

[1] Anderson, David H.: Compartmental Modeling and Tracer Kinetics, Springer-Verlag, 1983.

[2] Deuflhard, P.: Recent Progress in Extrapolation Methods for Ordinary Differential Equations, Invited talk presented at the SIAM 30th Anniversary Meeting, Stanford/Calif., July 19—23, 1982.

[3] Deuflhard, P.; Sautter, W.: On rank-deficient pseudoinverses, *Linear Algebra and Its Applications*, **29** (1980), 91—111.

[4] Ortega, J. M.; Rheinboldt, W. C.: Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York and London, 1970.

[5] Stoer, J.; Bulirsch, R.: Introduction to Numerical Analysis, Springer-Verlag, 1980.

[6] Tikhonov, A. N.; Arsenin, V. Y.: Solutions of Ill-posed Problems, Halsted-Press, 1977.