Efficient Algorithms for the Regularization of dynamic inverse Problems – Part I: Theory

U Schmitt and AK Louis

Institute for Applied Mathematics, P.O. Box 151150, Saarland University, 66041 Saarbrücken, Germany.

E-mail: `schmitt@num.uni-sb.de`

Abstract. In this paper dynamic inverse problems are studied, where the investigated object is allowed to change during the measurement procedure. In order to achieve reasonable results, temporal *a priori* information will be considered. Here, "temporal smoothness" is used as a quite general, but for many applications sufficient, *a priori* information. This is justified in the case of slight movements during a x-ray scan in computerized tomography, or in the field of current density reconstruction, where one wants to conclude from electrical measurements on the heads surface to locations of brain activity.

First, the notion of a dynamic inverse problem is introduced, then we describe how temporal smoothness can be incorporated in the regularization of the problem, and finally an efficient solver and some regularization properties of this solver are presented.

This theory will be exploited in three practically relevant applications in a following paper.

Submitted to: *Inverse Problems*

AMS classification scheme numbers: 34A55, 49N45

1. Introduction

This paper is devoted to the study of dynamic inverse problems, where the object under consideration changes with time during the measuring process.

Starting point is a measuring procedure which needs a certain amount of time. During this time span, single measurements are taken at time steps t_i. Then, a dynamic problem is described by operators A_i, where i is a temporal index. That is, the linear operator A_i maps the properties of an investigated object to the measurements m_i at time step t_i.

Now two cases are considered:

- In the first case, the properties x of the examined object do not change during the measuring process. Thus, we have to solve

  $$A_i\, x = y_i \quad \text{for all } \ i.$$

  This is called a *static inverse problem*.

- In the second case, called *dynamic inverse problem*, the examined object is allowed to change during the measuring process, and we have to solve

$$A_i\, x_i = y_i \quad \text{for all } \; i. \tag{1}$$

Examples for dynamic inverse problems are current density reconstructions based on EEG/MEG measurements [1], dynamic electrical impedance tomography [2], process tomography [3, 4] or x-ray CT where slight patient movements can be detected.

Due to the degree of freedom in (1) and the instability of the problem, *a priori* information has to be considered to achieve reasonable and stable solutions of dynamic inverse problems. This kind of regularization is done by assuming *temporal smoothness* as *a priori* information and is considered by adding a penalty term

$$\sum_i \frac{\|x_{i+1} - x_i\|^2}{(t_{i+1} - t_i)^2}$$

in a suitable norm to the known Tikhonov-Phillips minimization task. An application of this smoothness measure in the context of inverse electrocardiography can be found in [5]. If we try to solve the corresponding minimization problem in a straight-forward manner as in [5], we get to a linear problem which is extremely large and thus too expensive to solve.

Our approach starts in the context of operators between Hilbert spaces and leads to a quite general formulation of our procedure. Discretization, which is needed to achieve implementable algorithms, is done as late as possible. Using this method in part two of this paper, we achieve a new type of temporal CT (computerized tomography) algorithm, which avoids direct discretization of the forward model. Statistical methods as Kalman-filters, which are discussed in part two of this paper, are discribed in terms of matrices. These methods are not able to deduce a procedure comparable to the temporal CT algorithm we will work out in part two.

In the following mathematical prerequisites will be supplied and two *efficient* procedures for the solution of dynamic inverse problems are developed. These procedures are formulated in terms of linear operator equations, which we will be discretized by suitable projection schemes. Finally it is observed that in the case of equidistant time steps these procedures are regularizations of the temporally uncoupled respectively of the static problem, depending whether the parameter in front of the penalty term $\sum_i \|x_{i+1} - x_i\|^2$ goes to zero or to infinity.

The temporal inverse problem described above could also be tackled in a statistical context, e.g. by using Kalman-Smoothers or Wiener Filters as proposed in [6]. Our approach introduced above is rather analytical and achieves superior results concerning efficiency: statistical procedures have to consider covariance matrices for each timestep, which are often expensive to compute and are "too large" which affects the efficiency of the procedure. A comparision of these two approaches based on a real world problem, namely temporal impedance tomography, can be found in the second part of this paper. There, we will notice a significant enhancement of speed.

2. The mathematical prerequisites

To set the stage, Hilbert spaces H and G_i and linear operators $A_i \in \mathcal{L}(H, G_i)$ are considered. The operator A_i maps the properties $x \in H$ of an examined object to measurements $y_i \in G_i$. The H and G_i are equipped with norms $\|\cdot\|_H$ and $\|\cdot\|_{G_i}$ and scalar products $\langle\cdot,\cdot\rangle_H$ and $\langle\cdot,\cdot\rangle_{G_i}$. These space related indices will be omitted in most cases if they are determined by their context. As the operators A_i all together are assumed to determine a static x sufficiently, the single A_i are in most cases *under-determined*. An operator A is called "under-determined" if

- A is a matrix over $\mathbb{K}$, which has more columns than rows,
- A maps from an infinite-dimensional Hilbert space to $\mathbb{K}^n$, or
- A maps from a function space over a manifold to a function space over a less-dimensional manifold. In the context of Sobolev spaces this can be stated as $A : H^s(M_1) \to H^t(M_2)$ and $\dim M_1 > \dim M_2$.

$\mathbb{K}$ means either the real field or the complex field.

These operators have the property that equations of the form $AA^*u = y$ are easier or cheaper to solve (maybe numerically) than equations like $A^*Ax = A^*y$.

Definition 2.1 Given A_i, x, y_i as above, $x_i \in H$, we call

$$A_i\, x = y_i \quad \text{for all } \ i$$

a *static problem*, and

$$A_i\, x_i = y_i \quad \text{for all } \ i$$

a *dynamic problem*.

For the further steps we need the following definition.

Definition 2.2 The *Hilbert space sum* $H_1 \oplus \cdots \oplus H_n$ is the set $H_1 \times \cdots \times H_n$ equipped with the scalar product

$$\langle x, y \rangle := \sum_i \langle x_i, y_i \rangle_{H_i}.$$

The associated norm is defined accordingly.

Definition 2.3 An *operator matrix* is a collection of operators $A_{i,j} : G_j \to H_i, 1 \leq i \leq n, 1 \leq j \leq m$. These matrices can be multiplied by

$$(A \cdot B)_{i,j} = \sum_k A_{i,k} B_{k,j},$$

provided that the involved operators match. Addition of operator matrices is done entry by entry. To such a matrix we can assign an linear operator

$$[A] : G_1 \oplus \cdots \oplus G_m \to H_1 \oplus \cdots \oplus H_n$$

by

$$([A]x)_i = \sum_{j=1}^m A_{i,j} x_j.$$

The next theorem is important for the following calculations

Theorem 2.4 The map $A \mapsto [A]$ is an isomorphism between the set of operator matrices and the set of the linear operators $\mathcal{L}(H_1 \oplus \cdots \oplus H_n, G_1 \oplus \cdots \oplus G_m)$ according to $[A] \circ [B] = [A \cdot B]$ and $[A] + [B] = [A + B]$.

Proof: Most steps of the proof are easy. We only want to show that the considered map is surjective: for a given operator $B \in \mathcal{L}(H_1 \oplus \cdots \oplus H_n, G_1 \oplus \cdots \oplus G_m)$ we define the operator matrix

$$A_{i,j} = P_i B E_j$$

with canonical projections

$$P_i : H_1 \oplus \cdots \oplus H_n \to H_i$$

and canonical embeddings

$$E_j : H_j \to H_1 \oplus \cdots \oplus H_n.$$

Now it is easy to show $[A] = B$. ■

This theorem allows to calculate with operator matrices and operators which are assigned to operator matrices in the same way as with the well known matrices about fields. This includes that block matrices can be multiplied block wise, see [7], and that one needs not to distinguish between a matrix of operators and the assigned linear operator.

Furthermore the known Kronecker product of matrices [7] can be extended in the following way

Definition 2.5 Let M be a matrix in $\mathbb{K}^{n\times m}$ and B an linear operator $B : H \to G$. Then the *generalized Kronecker product* $M \otimes B$ is the operator

$$M \otimes B = \begin{bmatrix} m_{1,1}B & \cdots & m_{1,m}B \\ \vdots & & \vdots \\ m_{n,1}B & \cdots & m_{n,m}B \end{bmatrix}.$$

The following properties are easy to show:

$$\begin{aligned} (M \otimes A)(N \otimes B) &= (MN) \otimes (AB) \\ (M \otimes A)^{-1} &= M^{-1} \otimes A^{-1} \\ (M \otimes B)^{*} &= M^{*} \otimes B^{*} \end{aligned}$$

Furthermore $\otimes$ and $+$ are distributive, $\otimes$ is associative.

For simplifying the solution process of equations of the type

$$\sum_{i=1}^{T} (M_i \otimes A_i)\, x = y$$

in the special case that the A_i are matrices over a filed $\mathbb{K}$, the following theorem is quite useful. First we have to give a definition:

Definition 2.6 Suppose $x \in \mathbb{K}^{nm}$. Then define the rearrangement

$$\mathrm{Mat}_n(x) = X := \begin{pmatrix} x_1 & x_{1+n} & \ldots & x_{1+(m-1)n} \\ x_2 & x_{2+n} & \ldots & x_{2+(m-1)n} \\ \vdots & & & \vdots \\ x_n & x_{2n} & \ldots & x_{mn} \end{pmatrix}.$$

Now the following theorem can be stated. The proof is easy and therefore omitted.

Theorem 2.7 Given $M_i \in \mathbb{K}^{n\times n}, A_i \in \mathbb{K}^{m\times m}, 1 \leq i \leq T$ and $x, y \in \mathbb{K}^{nm}$, $X = \mathrm{Mat}_n(x)$ and $Y = \mathrm{Mat}_n(y)$, then the equation

$$\sum_{i=1}^{T} (M_i \otimes A_i)\, x = y$$

is equivalent to the matrix equation

$$\sum_{i=1}^{T} A_i X M_i^{\top} = Y.$$

Equations of the type $AXB^{\top} + CXD^{\top} = Y$ are called *generalized Sylvester equations* and can be solved much more efficiently than the equivalent Kronecker type equation $(A\otimes B + C\otimes D)\, x = y$; see [8].

For the further calculations we need another Lemma, it is about the solution of the so called Tikhonov-Phillips minimization problem:

Lemma 2.8 Given linear operators $A : H \to G_1, B : H \to G_2$ between Hilbert spaces H and G_1, G_2 respectively, such that B^*B is positive definitive. Furhermore let $x \in H, y \in G_1$ and $\lambda \in \mathbb{R}, \lambda \neq 0$. The unique solution of the minimization task

$$\|A\,x - y\|^2 + \lambda^2 \|B\,x\|^2 \to \min$$

can be determined by solving

$$(A^*A + \lambda^2 B^*B)\, x = A^*y. \tag{2}$$

If we have the relation

$$B^*BA = A^*E \tag{3}$$

for a positive definite $E : G_1 \to G_2$, equation (2) is equivalent to solving

$$(AA^* + \lambda^2 E)\, u = y \tag{4}$$

and setting $x = A^*u$.

Proof: Due to $\lambda \neq 0$ the functional $\Phi(x) = \|A\,x - y\|^2 + \lambda^2\|B\,x\|^2$ is strictly convex. So this functional has a unique solution, which can be achieved by solving $D\Phi(x) = 0$ where $D\Phi$ is the Frechet derivative of Φ. If we take into account that $DF(x)$ of $F(x) = \|A\,x - y\|^2$ fulfills We define $DF(x)\,h = 2\langle A^*(A\,x - y), h\rangle$, we get the linear equation stated above. The last statement is true because of

$$(A^*A + \lambda^2 B^*B)^{-1}A^* = A^*(AA^* + \lambda^2 E)^{-1}.$$

If A is under-determined, the last equation is easier respectivly cheaper to solve than the first one.

3. The procedure STR

In the following three procedures which can be used to solve dynamic inverse problems under the consideration of temporal smoothness as *a priori* information are presented. The first one, called normal equation approach, is not very efficient in the case of under-determined forward operators. Therefore a more efficient procedure called STR (=Spatio Temporal Regularizer) will be derived.

We start with linear operators $A_i, 1 \le i \le T$, which map the dynamic solutions x_i to measurements y_i. Further, it is supposed that the A_i either are compact operators between infinite-dimensional Hilbert spaces or are ill-conditioned matrices.

In order to solve the dynamic problem, we start with the minimization problem

$$\Phi(x) = \sum_{i=1}^{T} \|A_i x_i - y_i\|^2 + \lambda^2 \sum_{i=1}^{T} \|x_i\|^2 + \mu^2 \sum_{i} \frac{\|x_{i+1} - x_i\|^2}{(t_{i+1} - t_i)^2} \to \min . \tag{5}$$

Minimization of the first term forces compliance with the relation $A_i\, x_i = y_i$ for all i. The second term is of the type "spatial Tikhonov-Phillips-Regularization" and the third term measures the temporal smoothness of the x_i.

The following notations are introduced

$$\begin{aligned}
H^T &= H \oplus \cdots \oplus H \quad \text{(T times)}\\
A &= \operatorname{diag}(A_i) \in \mathcal{L}(H^T, G_1 \oplus \cdots \oplus G_T)\\
x &= (x_1, \ldots, x_T)^\top \in H^T\\
y &= (y_1, \ldots, y_T)^\top \in G_1 \oplus \cdots \oplus G_T\\
B &= D \otimes I_H \in \mathcal{L}(H^T, H^{T-1})\\
D &= \begin{pmatrix} \frac{1}{t_2 - t_1} & -\frac{1}{t_2 - t_1} & & & \\ & \frac{1}{t_3 - t_2} & -\frac{1}{t_3 - t_2} & & \\ & & \ddots & \ddots & \\ & & & \frac{1}{t_T - t_{T-1}} & -\frac{1}{t_T - t_{T-1}} \end{pmatrix} \in \mathbb{R}^{T \times (T-1)} .
\end{aligned}$$

One could use other forms of D. For example, if we assume equidistant timesteps $t_i = i$, $D = (-\delta_{i,j+1} + 2\delta_{i,j} - \delta_{i,j-1})_{i,j}$ leads to second order temporal smoothness of the x_i.

Now, the functional in (5) can be rewritten as

$$\Phi(x) = \|A\, x - y\|^2 + \lambda^2 \|x\|^2 + \mu^2 \|B\, x\|^2 \to \min .$$

As this functional is strict convex, a minimum exists, which is achieved by solving $D\Phi(x) = 0$. This derivative can be calculated according to the derivative occurring in the proof of Lemma 2.8, and we get the *normal equation*

$$(A^* A + \lambda^2 I + \mu^2 B^* B)\, x = A^*\, y. \tag{6}$$

As the A_i are under-determined this is a "quite large" problem. Unfortunately the technique used in the proof of Lemma 2.8 does not work here, a relation like (3) is not valid in this case. Thus, another technique must be used to achieve an efficient procedure, involving "small" operators $A_i A_j^*$.

Starting point for an efficient procedure involving operator matrices with entries $A_i A_j^*$ is the following minimization problem which is equivalent to (5): We introduce new variables d_i and solve

$$\Psi(x, d) = \sum_{i=1}^{T} \|A_i x_i - y_i\|^2 + \lambda^2 \sum_{i=1}^{T} \|x_i\|^2 + \mu^2 \sum_{i=1}^{T-1} \|d_i\|^2 \to \min \tag{7}$$

in connection with the constraints

$$d_i = (x_{i+1} - x_i)/(t_{i+1} - t_i). \tag{8}$$

In order to solve this constrained minimization problem, the constraints are coupled to the functional Φ by adding a penalty term. That is, we get solutions x_α and d_α of the unconstrained minimization task

$$\Psi_\alpha(x,d) = \sum_{i=1}^{T} \|A_i x_i - y_i\|^2 + \lambda^2 \sum_{i=1}^{T} \|x_i\|^2 + \mu^2 \sum_{i=1}^{T} \|d_i\|^2 \\ + \alpha^2 \sum_{i=1}^{T} \left\| d_i - \frac{x_{i+1} - x_i}{t_{i+1} - t_i} \right\|^2 \to \min \tag{9}$$

and achieve the solution x of (7), (8) by

$$x = \lim_{\alpha\to\infty} x_\alpha.$$

If d_i is scaled as $d_i = \lambda/\mu\, \delta_i$, and if the following notion is used

$$\delta = (\delta_1, \dots, \delta_{T-1})^\top \in H^{T-1},$$

the minimization problem (9) can be written as

$$\left\| \underbrace{\begin{bmatrix} A & 0 \\ \alpha B & \alpha \frac{\lambda}{\mu} I \end{bmatrix}}_{M_\alpha} \begin{pmatrix} x \\ \delta \end{pmatrix} - \begin{pmatrix} y \\ 0 \end{pmatrix} \right\|^2 + \lambda^2 \left\| \begin{pmatrix} x \\ \delta \end{pmatrix} \right\|^2 \to \min .$$

This is a Tikhonov-Phillips problem which can be solved as stated in Lemma 2.8:

$$\begin{pmatrix} x^\alpha \\ \delta^\alpha \end{pmatrix} = M_\alpha^* \left(M_\alpha M_\alpha^* + \lambda^2 I\right)^{-1} \begin{pmatrix} y \\ 0 \end{pmatrix}. \tag{10}$$

In order to determine x^α, first the following equation

$$\left(M_\alpha M_\alpha^* + \lambda^2 I\right) \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} y \\ 0 \end{pmatrix}.$$

is solved. We have

$$M_\alpha M_\alpha^* + \lambda^2 I = \begin{bmatrix} AA^* & \alpha AB^* \\ \alpha BA^* & \alpha^2(BB^* + \frac{\lambda^2}{\mu^2} I) \end{bmatrix} + \lambda^2 I.$$

Thus, one has to solve

$$AA^*\, u + \alpha AB^* v + \lambda^2\, u = y \tag{11}$$

$$\alpha BA^*\, u + \alpha^2(BB^* + \frac{\lambda^2}{\mu^2} I)\, v + \lambda^2\, v = 0. \tag{12}$$

v can be calculated from (12) as

$$v = -\frac{1}{\alpha} \left(BB^* + \left(\frac{\lambda^2}{\mu^2} + \frac{\lambda^2}{\alpha^2} \right) I \right)^{-1} BA^*\, u. \tag{13}$$

Substituted in (11), the following equation for u is achieved :

$$A\Bigg(I - \underbrace{B^* \left(BB^* + \left(\frac{\lambda^2}{\mu^2} + \frac{\lambda^2}{\alpha^2} \right) I \right)^{-1} B}_{N^\alpha}\Bigg) A^*\, u + \lambda^2\, u = y. \tag{14}$$

N^α can be simplified to

$$\begin{aligned} N^\alpha &= (D^\top \otimes I_H)\left[(D\otimes I_H)(D^\top I_H) + \left(\frac{\lambda^2}{\mu^2}+\frac{\lambda^2}{\alpha^2}\right)(I_{T-1}\otimes I_H)\right]^{-1}(D\otimes I_H) \\ &= (D^\top \otimes I_H)\Bigg[\underbrace{\left(DD^\top + \left(\frac{\lambda^2}{\mu^2}+\frac{\lambda^2}{\alpha^2}\right)I_{T-1}\right)^{-1}}_{Q^\alpha}\otimes I_H\Bigg](D\otimes I_H) \\ &= (D^\top Q^\alpha D)\otimes I_H. \end{aligned}$$

Then, (14) is equivalent to

$$A\left[I_T\otimes I_H - (D^\top Q^\alpha D)\otimes I_H\right]A^*\,u + \lambda^2\,u = y,$$

respectively

$$A\Big((\underbrace{I_T - D^\top Q^\alpha D}_{R^\alpha})\otimes I_H\Big)A^*\,u + \lambda^2\,u = y.$$

So we define

$$C^\alpha := A(R^\alpha\otimes I_H)A^* = \left[r^\alpha_{i,j}A_iA_j^*\right]_{i,j},$$

and

$$(C^\alpha + \lambda^2 I_{G_1\oplus\cdots\oplus G_T})\,u = y \tag{15}$$

has to be solved. If we consider (13), an analogous calculation yields

$$v = -\frac{1}{\alpha}\left[(Q^\alpha D)\otimes I_H\right]A^*u.$$

Because of (10) one gets

$$\begin{pmatrix} x^\alpha \\ r^\alpha \end{pmatrix} = M_\alpha^*\begin{pmatrix} u \\ v \end{pmatrix},$$

which supplies

$$\begin{aligned} x^\alpha &= A^*\,u - B^*\left[(Q^\alpha D)\otimes I_H\right]A^*\,u \\ &= \left[R^\alpha\otimes I_H\right]A^*\,u. \end{aligned}$$

Finally, one gets the following procedure for calculating x^α: First solve

$$(C^\alpha + \lambda^2 I_{G_1\oplus\cdots\oplus G_T})\,u = y,$$

then put

$$x^\alpha = \left[R^\alpha\otimes I_H\right]A^*\,u.$$

The last step to achieve a solution x of (7), (8) is to perform the process $\alpha\to\infty$: from (15) we get the equation

$$(C + \lambda^2 I_{G_1\oplus\cdots\oplus G_T})\,u = y$$

with

$$C = \lim_{\alpha\to\infty} C^\alpha = A\,(\lim_{\alpha\to\infty} R^\alpha)A^* = ARA^*,$$

whereas

$$\begin{aligned} R &= I_T - D^\top Q D \\ Q &= \left(DD^\top + \frac{\lambda^2}{\mu^2}I_{T-1}\right)^{-1}. \end{aligned} \tag{16}$$

So one gets the solution x of (7), (8) by

$$x = \lim_{\alpha\to\infty} x^{\alpha} = (R \otimes I_H) A^* u. \tag{17}$$

That is

$$x_i = \sum_j r_{i,j}\, A_j^* u_j.$$

If the calculations above are summarized, we get the following procedure STR for the solution of (5):

(i) **Input**: data y, spatial regularization parameter λ, temporal regularization parameter μ.

(ii) Calculate Q and R according to (16)

$$Q = \left(DD^\top + \frac{\lambda^2}{\mu^2} I_{T-1} \right)^{-1}$$
$$R = I_T - D^\top Q D = (r_{i,j})_{i,j}.$$

(iii) C is defined by

$$C = [r_{i,j} A_i A_j^*]_{i,j}.$$

(iv) Solve

$$(C + \lambda^2 I_{G_1 \oplus \cdots \oplus G_T})\, u = y.$$

(v) Finally, calculate the x_i by

$$x_i = \sum_j r_{i,j}\, A_j^* u_j.$$

In order to achieve efficiency, in a first step one calculates $A_j^* u_j$ for each j and afterwards x_i.

The procedure presented above involves ”small” operators $A_i A_j^*$ which leads to the announced efficiency compared to (6). Statistical procedures as the Kalman-filter which will be introduced in the second part of this paper, result in linear equations in terms of $A_i C_i A_i^*$. These are as ”small” as the operators in the procedure above but are in most cases expensive to compute due to the size of the appearing C_i.

As the linear operator C may be an operator between infinite dimensional Hilbert spaces, the procedure above *is not an algorithm*. In order to get an algorithm the operator equation in step (iv) must be solved numerically for instance by a projection scheme. In the next section it is explained how a projection scheme works, and then how such a scheme can be applied in order to approximate the solution of the equation in step (iv). In the end, an efficient algorithm for the numerical stable solution of dynamic inverse problems is achieved.

4. Solving operator matrix equations by projection schemes

Here we give a short description how projection schemes work and how we can apply them to the operator matrix equation emerging in procedure STR. More detailed information about projection schemes, especially convergence theorems can be found in [10].

Given are Banach spaces X, Y and a linear, continuous and injective operator $T : X \to Y$. The operator equation $Tx = y$ is considered. In order to calculate an approximative solution, one searches x_h in an finite dimensional subspace X_h of X such that

$$\Psi T x_h = \Psi y \qquad \text{for all} \quad \Psi \in Y_h^*.$$

Here Y_h^* is a finite dimensional subspace of Y^*. If it is assumed that $X_h = \mathrm{span}\{\phi_1, \ldots, \phi_n\}$ and $Y_h^* = \mathrm{span}\{\psi_1, \ldots, \psi_n\}$, one gets x_h as

$$x_h = \sum_{i=1}^{n} \alpha_i \phi_i,$$

where α fulfills

$$T_h \alpha = y_h.$$

Here $(T_h)_{i,j} = \psi_i T \phi_j$ and $(y_h)_i = \psi_i y$.

Now, we can describe how to achieve a projection scheme in case of an operator matrix $T = [C]$. We start from linear operators $C_{i,j} : H_j \to G_i$ with Banach spaces $G_i, H_j, 1 \le i, j \le n$. The spaces H_i are approximated by

$$H_i^h = \mathrm{span}\{\phi_{i,j} \,|\, 1 \le j \le m_i\} \subset H_i.$$

As functionals in G_i we have

$$(G_i^*)^h = \mathrm{span}\{\psi_{i,j} \,|\, 1 \le j \le n_i\} \subset G_i^*.$$

Now the operator equation

$$[C]f = g$$

with

$$[C] : \underbrace{H_1 \times \cdots \times H_m}_{H} \to \underbrace{G_1 \times \cdots \times G_n}_{G}$$

will be considered. Starting from the given subspaces H_i^h and $(G_i^*)^h$, the subspaces H^h and $(G^*)^h$ are constructed as follows. We define

$$p(i,j) = \sum_{l=1}^{i-1} n_l + j \quad \text{and} \quad q(i,j) = \sum_{l=1}^{i-1} m_l + j.$$

p maps $\{(i,j) \,|\, 1 \le i \le n, 1 \le j \le m_i\}$ bijectively to $\{1 \ldots \sum m_i\}$. q has an analogous property. Now one uses

$$\Phi_{q(i,j)} = (0, \ldots, \underbrace{\phi_{i,j}}_{i-\text{th place}}, \ldots, 0)$$

as a basis of a finite dimensional subspace H^h of H and

$$\Psi_{p(i,j)}(y_1, \ldots, y_n) = \psi_{i,j}(y_i)$$

as testing functionals in G^*. In other words:

$$H^h = \mathrm{span}\{\Phi_l \,|\, 1 \le l \le \textstyle\sum m_i\} \quad \text{and} \quad (G^*)^h = \mathrm{span}\{\Psi_l \,|\, 1 \le l \le \textstyle\sum n_i\}$$

are chosen.

Finally, the matrix D as a discretization of $[C]$ and $B_{i,j}$ as a discretization of $C_{i,j}$ are constructed. That is

$$D_{i,j} = \Psi_i [C] \Phi_j \quad \text{and} \quad (B_{i,j})_{k,l} = \psi_{i,k} C_{i,j} \phi_{j,l}.$$

Then we have

Lemma 4.1 The discretization D of $[C]$ can be constructed from the discretizations $B_{i,j}$ of $C_{i,j}$ by block wise compounding.

Proof: We have

$$\begin{aligned} D_{p(i,k),q(j,l)} &= \Psi_{p(i,k)}[C]\Phi_{q(j,l)} = \Psi_{p(i,k)}(C_{1,j}\phi_{j,l}, C_{2,j}\phi_{j,l}, \dots C_{n,j}\phi_{j,l}) \\ &= \psi_{i,k}C_{i,j}\phi_{j,l} = (B_{i,j})_{k,l}. \end{aligned}$$

■

Now we know how to solve step (iv) in the procedure STR numerically, and we get an efficient algorithm for the solution of dynamic inverse problems.

5. The procedure STR-C

In some applications, e.g. current density reconstruction, the operators A_i are not depending on i, that is $A_i = A_0$ for all i. In this case, the operator C is

$$C = [r_{i,j}A_0A_0^*] = R \otimes (A_0A_0^*).$$

If A_0 additionally is a matrix of size $n \times N$ we get according to Theorem 2.7, equation (17) and the relation $A^* = I_T \otimes A_0^*$ the following procedure STR-C ('C' means 'Constant operator'):

(i) **Input**: data y, spatial regularization parameter λ, temporal regularization parameter μ.

(ii) Calculate Q and R according to

$$\begin{aligned} Q &= \left(DD^\top + \frac{\lambda^2}{\mu^2}I_{T-1}\right)^{-1} \\ R &= I_T - D^\top QD. \end{aligned}$$

(iii) Solve the generalized Sylvester equation

$$(A_0A_0^*)UR + \lambda^2 U = \mathrm{Mat}_n(y) =: Y.$$

(iv) Calculate

$$X = A_0^*UR,$$

and get x_i as the i-th column of X.

The Sylvester type equation in (iii) can efficiently be solved by methods provided in [8].

6. Some remarks about efficiency

Now the costs of the three approaches "normal equation" (6), STR and STR-C will be compared. We start from matrices $A_i \in \mathbb{R}^{n\times N}$ and T time steps.

According to [7] the direct solution of a $n \times n$ system needs $2n^3/3$ FLOPS. According to [8] the costs for the solution of the Sylvester type equation in procedure

STR-C can be bounded by $25(n+T)^3$ FLOPS. The step (iv) in STR and STR-C needs $2Tn(T+N)$ FLOPS in each case.

If we assume $n = 64, N = 5000, T = 100$, we get

- a total cost of $2/3(NT)^3$ FLOPS, if we use the normal equation approach. Considering the given numbers, this is $8.3 \cdot 10^{16}$ FLOPS.
- a total cost of $2/3(nT)^3 + 2Tn(T+N)$ FLOPS in the case of procedure STR. Based on the numbers given above, this is an amount of $1.75 \cdot 10^{11}$ FLOPS.
- a total cost bounded by $25(n+T)^3 + 2Tn(T+N)$ FLOPS for the procedure STR-C, which makes $1.76 \cdot 10^8$ FLOPS.

So we see, that the procedures STR and STR-C are really efficient compared to the "naive" normal equation approach.

7. Regularization properties of the procedure STR

As the solution x provided by the procedure STR depends on λ and μ, it can be written as $x = x_{ST}(\lambda, \mu)$. In the following, the processes $\mu \to 0$ and $\mu \to \infty$ will be considered. One additional assumption is $t_{i+1} = t_i + 1$ for all i, so that D has only entries 1 and -1. Other constant increments can be put into the parameter μ.

Now matrix $R^\infty = R^\infty(\lambda/\mu) = R^\infty(\rho)$ will be analyzed.

Before we start we need the following lemma. See, e.g. [9]:

Lemma 7.1 In the case of $t_{i+1} - t_i = 1$ the Matrix $DD^\top \in \mathbb{R}^{(T-1)\times(T-1)}$ has normed eigenvectors u_μ with

$$(u_\mu)_\nu = \sqrt{\frac{2}{T}} \sin\left(\frac{\nu\mu\pi}{T}\right) \qquad 1 \le \mu, \nu \le T-1$$

and corresponding eigenvalues

$$\lambda_\mu = 4\sin^2\left(\frac{\mu\pi}{2T}\right) \qquad 1 \le \mu \le T-1.$$

Theorem 7.2 We have

$$\lim_{\rho\to\infty} R^\infty(\rho) = (\delta_{i,j})_{i,j}$$

and

$$\lim_{\rho\to 0} R^\infty(\rho) = \left(\frac{1}{T}\right)_{i,j}.$$

Proof: It is $R^\infty(\rho) = I_T - D^\top(DD^\top + \rho^2 I_{T-1})^{-1}D$. Thus, the first statement follows from

$$(DD^\top + \rho^2 I_{T-1})^{-1} = \rho^{-2}(\rho^{-2}DD^\top + I_{T-1})^{-1}.$$

Due to Lemma 7.1 the eigenvalues of $DD^\top$ do not vanish, so the matrix $DD^\top$ is regular. Thus, we consider

$$M := \lim_{\rho\to 0} R^\infty(\rho) = I_T - D^\top(DD^\top)^{-1}D. \qquad (18)$$

The matrix M has the properties $MD^\top = DM = 0$, thus $m_{i,j} = m_{i,j+1}$ and $m_{i,j} = m_{i+1,j}$. So we can conclude that M is a matrix with constant entries. In order to calculate this entry, we analyze $(D^\top XD)_{1,1}$ with $X = (DD^\top)^{-1}$. Due to Lemma 7.1, X can be written as

$$X = \sum_{\mu=1}^{T-1} \lambda_\mu^{-1} u_\mu u_\mu^\top.$$

If we consider $(D^\top XD)_{1,1} = x_{1,1} - x_{1,2} - x_{2,1} + x_{2,2}$ together with Lemma 7.1 and the representation of X given above, we get

$$(D^\top XD)_{1,1} = \frac{1}{2T} \sum_{\mu=1}^{T-1} \left(\frac{\sin\left(\frac{\mu\pi}{T}\right)\sin\left(\frac{\mu\pi}{T}\right)}{\sin^2\left(\frac{\mu\pi}{2T}\right)} - 2\frac{\sin\left(\frac{\mu\pi}{T}\right)\sin\left(\frac{2\mu\pi}{T}\right)}{\sin^2\left(\frac{\mu\pi}{2T}\right)} + \frac{\sin\left(\frac{2\mu\pi}{T}\right)\sin\left(\frac{2\mu\pi}{T}\right)}{\sin^2\left(\frac{\mu\pi}{2T}\right)} \right).$$

If we write the sin function in terms of exponential functions, and if $x^2 - y^2 = (x+y)(x-y)$ is applied successively, we arrive at

$$(D^\top XD)_{1,1} = \frac{T-1}{T}.$$

Together with (18) one gets

$$(M)_{1,1} = 1 - \frac{T-1}{T} = \frac{1}{T},$$

which proves the second statement of the theorem.

■

For further interpretations of the limits of $x_{ST}(\lambda,\mu)$, two operators describing the temporal uncoupled problem and the static problem according to the linear operators A_i are introduced.

The operators are

$$A_{static} = \begin{bmatrix} A_1 \\ \vdots \\ A_T \end{bmatrix} \quad \text{and} \quad A_{uncoupled} = \begin{bmatrix} A_1 & & \\ & \ddots & \\ & & A_T \end{bmatrix}.$$

Then solving $A_{static}x = m$ provides the static solution $x \in H$ of $A_i\, x = m_i$ for all i , and solving $A_{uncoupled}x = m$ provides the temporal uncoupled solution $x = (x_i)_i \in H^T$ of $A_i\, x_i = y_i$ for all i.

First $\lim_{\mu\to\infty} x_{ST}(\lambda,\mu)$ will be analyzed. Due to Theorem 7.2 and step (iii) of STR we get u by solving

$$\left(\left[\frac{1}{T} A_i A_j^* \right] + \lambda^2 I \right) u = y,$$

which is the same as

$$\left(\frac{1}{T} A_{static} A_{static}^* + \lambda^2 I \right) u = y.$$

If we set $v = u/T$ this is equivalent to

$$\left(A_{static} A_{static}^* + \lambda^2 T I \right) v = y.$$

The last step in procedure STR delivers a vector x_{ST} with T constant entries

$$x_k = \sum_j \frac{1}{T} A_j^* u_j = A_{static}^* v \qquad 1 \le k \le T.$$

For stating the next theorem the Tikhonov-Phillips operator

$$\begin{aligned} T_{\lambda,mode}y &= \text{minarg}\left\{\|A_{mode}x - y\|^2 + \lambda^2\|x\|^2\right\} \qquad mode \in \{static, uncoupled\} \\ &= A^*_{mode}\left(A_{mode}A^*_{mode} + \lambda^2 I\right)^{-1} y \end{aligned}$$

and the projections

$$P_k(x_1, \dots, x_T) = x_k \qquad 1 \le k \le T$$

are introduced. The projections are needed, because x_{ST} is a vector of size T with constant entries.

Theorem 7.3

$$\lim_{\mu\to\infty} P_k(x_{ST}(\lambda,\mu)) = T_{\lambda\sqrt{T},static}\, y \qquad 1 \le k \le T$$

$$\lim_{\lambda\to 0}\lim_{\mu\to\infty} P_k(x_{ST}(\lambda,\mu)) = A^\dagger_{static}\, y \qquad 1 \le k \le T$$

Proof: The first statement was proven above. The second statement follows from regularization properties of the Tikhonov-Phillips operator, see [10, 11]. ■

Now $\lim_{\mu\to 0} x_{ST}(\lambda,\mu)$ is studied. Again, Theorem 7.2 shows that step (iii) in STR is equivalent to solving

$$(A_i A_i^* + \lambda^2 I)\, u_i = y_i \quad \text{for all } i$$

and x is calculated by

$$x_i = A_i^*\, u_i \quad \text{for all } i$$

These two steps can be written as

$$(A_{uncoupled}A^*_{uncoupled} + \lambda^2 I)\, u = y,$$

followed by

$$x = A^*_{uncoupled}u.$$

So we get

Theorem 7.4

$$\lim_{\mu\to 0} x_{ST}(\lambda,\mu) = T_{\lambda,uncoupled}\, y$$

$$\lim_{\lambda\to 0}\lim_{\mu\to 0} x_{ST}(\lambda,\mu) = A^\dagger_{uncoupled}\, y$$

Proof: Again, the first statement is proven above. The prove of the second statement is done in the same way as in the proof of Theorem 7.3. ■

So, one can say, that procedure STR delivers regularizations of the static as well as the uncoupled problem, depending on the limiting process of the temporal regularization parameter μ. STR produces a balance between the two extremes "static" and "uncoupled".

One last remark: The limiting processes in Theorems 7.3 and 7.4 can not be done in an arbitrary way. For example if we proceed $(\lambda,\mu) \to (0,0)$ by $(\lambda_i,\mu_i) = (\alpha/i, 1/i)$, we get the matrix $R^\infty(\lambda_i/\mu_i) = R^\infty(\alpha)$ and the result of the procedure STR does not converge to $A^\dagger_{static}\, y$ or $A^\dagger_{uncoupled}\, y$.

Acknowledgments

This work was supported by the German ministry of education and research (03-LO7SA2-2, 03-BU7AA1-4).

References

[1] U. Schmitt, A.K. Louis, F. Darvas, H. Buchner, and M. Fuchs. Numerical Aspects of spatio-temporal Current Density Reconstruction from EEG-/MEG-Data. *IEEE Trans Med Imag*, 20(4):314–324, 2001.

[2] M. Vauhkonen and P.A. Karjalainen. A Kalman Filter Approach to track fast Impedance Changes in electrical Impedance Tomography. *IEEE Trans. Biomed. Eng.*, 45(4):486–493, 1989.

[3] A. Seppänen, M. Vauhkonen, E. Somersalo, and J.P. Kaipio. State Space Models in Process Tomography - Approximation of state Noise Covariance. *Inv Probl Eng*, in press, 2000.

[4] R.A. Williams and M.S. Beck. *Process Tomography, Principles, Techniques and Applications*. Butterworth-Heinemann Ltd, Oxford, 1995.

[5] D.H. Brooks, G.F. Ahmad, and R.S. MacLeod an G.M Maratos. Inverse Electrocardiography by simultaneous Imposition of multiple Constraints. *IEEE Trans Biomed Eng*, 46(1):3–17, 1999.

[6] M.S. Grewal and A.P. Andrews, Kalman FIltering. Prentice Hall, 1993.

[7] G.H. Golub and C.F. van Loan. *Matrix Computations*. John Hopkins University Press, 3rd edition, 1996.

[8] J.D. Gardiner, A.J. Laub, J.J. Amato, and C.B. Moler. Solution of the Sylvester Matrix Equation $AXB^\top + CXD^\top = E$. *ACM Transactions on Mathematical Software*, 18(2):223–238, 1992.

[9] G.D. Smith. *Numerische Lösung von partiellen Differentialgleichungen*. Vieweg, 1970.

[10] A.K. Louis. *Inverse und schlecht gestellte Probleme*. Teubner, Stuttgart, 1989.

[11] D. Colton and R. Kress. *Inverse acoustic and electromagnetic scattering theory*. Springer, 1992.