Approximative inverse for linear and some nonlinear problems

A K Louis

Institute of Applied Mathematics, Universität des Saarlandes, D-66041 Saarbrücken

Abstract. In this paper we present a method for solving problems $Af = g$ by constructing an approximative inverse which maps the data g to a regularized solution of this equation of the first kind. No discretization for f is needed. The solution operator can be precomputed independently of the data. This works for linear problems and for nonlinear problems with a special structure. The regularization is achieved by computing mollified versions of the (minimum norm) solution. It is shown that this class of regularization operators contains as special cases the classical methods like Tikhonov – Phillips, iteration methods and also discretization methods. In case the operator has some invariance properties the storage needs are dramatically reduced.

Short title: Approximative inverse

July 31, 1995

1. Introduction

We study operator equations $Af = g$ for operators between Hilbert spaces X and Y. Both the cases of linear operators A and of nonlinear A with a special structure are treated. Approximate inverse means a solution operator which maps the data g to a stable approximation of the solution of the ill – posed problem $Af = g$. This inversion operator is precomputed without using the data g.
The method is based on two ideas. First, the computation of moments of the solution is stable; i.e., we compute instead of f the approximation $\langle f, e_\gamma \rangle$ with a suitable mollifier e_γ reducing in that way the high frequency components in the solution which are mostly affedted by the data noise. This can be reformulated as using a weaker topology in the space X, see [3], [5]. Examples for e_γ are given in Section 2, e_γ can be a basis function for projection methods, it can be chosen such that $\langle f, e_\gamma \rangle$ approximates a derivative of f; in wavelet language it can be a scaling function or a wavelet. Second, in the case of linear operators the computation of $\langle e_\gamma, f \rangle$ is then achieved by approximating e_γ in the range of the adjoint operator A^* by the reconstruction kernel v_γ : $A^* v_\gamma \simeq e_\gamma$. Then

$$\langle f, e_\gamma \rangle \simeq \langle f, A^* v_\gamma \rangle = \langle Af, v_\gamma \rangle = \langle g, v_\gamma \rangle \ .$$

This is the mollifier method presented in [7]. For nonlinear operators we combine these ideas with the results of Snieder [13] who generalized the Backus – Gilbert method to nonlinear problems. A detailed analysis of the method shows that the computational effort is much smaller than in [13].
In general the regularization of ill – posed problems is achieved by regularization methods like Tikhonov – Phillips, truncated singular value decomposition or iterative methods like Landweber or conjugate gradients. For references see for example [2], [4], [5], [10], [12]. These methods can easily be represented by the singular value decomposition of the compact operator A and a corresponding filter. On the other hand there are regularization methods like Backus – Gilbert or mollifier methods, see [1], [7], [12]. In Section 2 we derive the approximative inverse and we show that in the case of linear operators we compute in that way a mollified version of the minimum norm solution. We show in Section 3 that the above mentioned regularization methods are special cases of these mollifier methods. Also the relation to the Backus-Gilbert method is made precise. The next section presents the central difference quotient for approximating the first derivative as an approximative inverse where local averages of the solution are computed. Further on it shows an example where we determine the derivative of the solution of an integral equation of the first kind by properly chosing e_γ.
If the operator has some invariance properties the storage needs are reduced by applying group representations which intertwine with the operator. This is the content of Section 5 and the base for an efficient implementation. Section 6 contains the nonlinear case where for the sake of simplicity we start with a quadratic problem. The approach is then generalized to a larger class of problems. The last section contains numerical experiments for a nonlinear problem.

2. Approximative inverse for linear problems

In the following we assume A to be a linear, continuous operator between the Hilbert spaces X and Y. Especially we think of X as a space of functions and Y as a finite dimensional space of measurements. Hence, if necessary, we use $X = L_2(\Omega)$ for a suitable set $\Omega \subset I\!R^d$. Examples for mollifiers are

$$e_\gamma(x,y) = \frac{d}{\mathrm{vol}(S^{d-1})\gamma^d}\chi_\gamma(x-y)$$

where χ_γ is the characteristic function of the ball around 0 with radius γ and $\mathrm{vol}(S^{d-1})$ is the measure of the surface of the unit ball in $I\!R^d$. Here local averages of the solution are computed. With the band limiting filter

$$e_\gamma(x,y) = \left(\frac{\gamma}{\pi}\right)^d \mathrm{sinc}(\gamma(x-y))$$

the high – frequency components in the solution are eliminated. Fast decaying is the kernel of the heat equation

$$e_\gamma(x,y) = (2\pi)^{-d/2}\gamma^{-d}\exp(-|x-y|^2/(2\gamma^2)) \ .$$

In all cases the parameter γ acts as a regularization parameter. The mollifier e_γ is not necessarily a function with mean value 1. When the essential information we need are discontinuities in f we can use as e_γ a function such that $\langle f, e_\gamma(x,\cdot)\rangle$ approximates a derivative of $f(x)$. This means that e_γ can also be a wavelet, see e.g. [8].
For a motivation we consider the problem

$$Af(x) = \int_0^x f(t)\,dt = g(x) \quad , \quad 0 \le x \le 1$$

where the inverse operator A^{-1} is the differentiation of the first order. For a sufficiently smooth function ψ and for homogeneous or periodic boundary conditions we compute, see [5],

$$\langle \psi, f\rangle = \langle \psi, A^{-1}g\rangle = \langle (A^{-1})^*\psi, g\rangle = \langle \Lambda, g\rangle$$

where

$$\Lambda = (A^{-1})^*\psi = -\psi' \ .$$

The solution depends continuously on the data g as is seen by the simple calculation

$$|\langle \psi, f\rangle| = |\langle \psi, A^{-1}g\rangle| \le \|\Lambda\|\|g\| \ .$$

The smoothness needed for the function ψ depends on the smoothing of the operator A. In the example considered here, the function g is once more differentiable than f and the function ψ has to be at least one times differentiable.
After this motivation we start with the procedure described in the introduction. First we assume the equation $A^*v_\gamma = e_\gamma$ to be solvable. Then we put

$$\langle f, e_\gamma\rangle = \langle f, A^*v_\gamma\rangle = \langle Af, v_\gamma\rangle = \langle g, v_\gamma\rangle =: S_\gamma g \ . \tag{1}$$

This is the technique to derive inversion formulas in x – ray computer tomography resulting in the so – called filtered backprojection methods, see e.g. [9]. If the equation $A^*v_\gamma = e_\gamma$ is not solvable we approximate v_γ by minimizing the defect $\|A^*v_\gamma - e_\gamma\|$ leading to the equation

$$AA^*v_\gamma = Ae_\gamma \ . \tag{2}$$

Then we get

$$\langle f, e_\gamma \rangle \simeq \langle f, A^* v_\gamma \rangle = \langle Af, v_\gamma \rangle = \langle g, v_\gamma \rangle =: S_\gamma g \ .$$

It is important to mention that no artificial discretization of f is needed as introduced by projection methods, see e.g. [5], [9], [10].

Definition 1. Let e_γ be a suitable function, and let v_γ be the solution of (2). Then $S_\gamma g := \langle g, v_\gamma \rangle$ is called the *approximative inverse* of the operator A and v_γ is the *reconstruction kernel.*

In the situation where $X = L_2(\Omega)$, $Y = \mathbb{R}^N$ this means

$$S_\gamma g(x) = \sum_{n=1}^{N} g_n v_{\gamma,n}(x) \ .$$

The minimum norm solution of the equation $Af = g$ is the solution with smallest norm. It is contained in the range of the adjoint operator A^* and therefore computed as

$$f_M = A^* u \quad \text{where} \quad AA^* u = g \ .$$

Hence

$$f_M = A^*(AA^*)^{-1} g \ . \tag{3}$$

The next theorem gives a simple relation between the approximative inverse and the minimum norm solution.

Theorem 2. The approximative inverse maps the right – hand side of $Af = g$ to the mollified version of the minimum norm solution f_M defined in (3); i.e.,

$$S_\gamma g = \langle f_M, e_\gamma \rangle \ . \tag{4}$$

Proof: The function v_γ can be written following (2) as

$$v_\gamma = (AA^*)^{-1} A e_\gamma \ ,$$

resulting in

$$S_\gamma g = \langle g, v_\gamma \rangle = \langle g, (AA^*)^{-1} A e_\gamma \rangle = (A^*(AA^*)^{-1} g, e_\gamma \rangle = \langle f_M, e_\gamma \rangle \ .$$

If $A : L_2(\Omega) \to \mathbb{R}^N$ is given as $(Af)_n = \langle f, k_n \rangle = \int_\Omega k_n(y) f(y) dy$ for $n = 1, \dots N$, then the adjoint operator $A^* : \mathbb{R}^N \to L_2(\Omega)$ is given as

$$A^* g(x) = \sum_{n=1}^{N} g_n k_n(x) \ .$$

The matrix representing AA^* is the Gram matrix of the system $\{k_n : 1 \le n \le N\}$ and computed to

$$(AA^*)_{mn} = \langle k_n, k_m \rangle = \int_\Omega k_m(y) k_n(y) dy$$

or

$$AA^* = \int_\Omega k(y)k^\top(y)\,dy$$

where $kk^\top$ denotes the dyadic product $(kk^\top)_{ij} = k_ik_j$ for $k = (k_1,\ldots,k_N)^\top \in I\!R^N$. The discrete equation inherits the instability of the integral operator, which means that the matrix is ill – conditioned. The v_γ are then computed by using Tikhonov – Phillips regularization with a rather small regularization parameter. The fine regularization is achieved by the parameter γ in e_γ, compare with [11].

3. Comparison with other methods

In this section we show that most methods used for stabilizing the solution are special cases of the approximative inverse. To this end we first present some of these methods. We then compare our result with the Backus - Gilbert method.

The most prominent regularization method for ill – posed problems is Tikhonov – Phillips regularization where the defect is penalized with an a priori information on the solution. For the sake of simplicity we use the L_2 – norm of the solution and get

$$J_\gamma(f) = \|Af-g\|^2 + \gamma\|f\|^2 \ .$$

The minimum of this functional is computed as solution of

$$(A^*A+\gamma I)f_\gamma = A^*g \ . \tag{5}$$

Now let A be a compact operator between the Hilbert spaces X and Y. Then it has a singular value decomposition

$$\{v_n, u_n; \sigma_n\}_n$$

where v_n, u_n are normalized and

$$Av_n = \sigma_n u_n \quad \text{and} \quad A^*u_n = \sigma_n v_n \ .$$

For non degenerate operators A; i.e., where the dimension of the range is not finite, the singular values σ_n decay to 0. The degree of the decay is a measure for the ill – posedness of the operator. With the singular value decomposition the solution of (5) can be written as

$$T_\gamma g = \sum_n F_\gamma(\sigma_n)\sigma_n^{-1}\langle g, u_n\rangle v_n \tag{6}$$

where the filter F_γ is given as

$$F_\gamma(\sigma) = \frac{\sigma^2}{\sigma^2+\gamma} \ .$$

Of course, also the truncated singular value decomposition can be written in the form (6) where

$$F_\gamma(\sigma) = \begin{cases} 1 & : \ \sigma \ge \gamma \\ 0 & : \ \sigma < \gamma \end{cases}$$

For the Landweber iteration,

$$f^{m+1} = f^m - \beta A^*(Af^m - g) \ ,$$

the filter is, when identifying the regularization parameter γ with $1/m$ where m is the iteration index,

$$F_m(\sigma) = 1 - (1 - \beta\sigma^2)^m \ .$$

The last example we want to mention is the method of conjugate gradients where the filter is given as

$$F_m(\sigma) = P_{m-1}(\sigma^2; g)\sigma^2$$

and the polynomials P_{m-1} describe the iterations, they depend on the data g, hence this is a nonlinear regularization method for a linear problem, and therefore this extremely well behaving method does not fit into the frame of approximative inverse we study here. For more details see e.g. [5].
The next result shows that the regularization methods as discussed above with the exception of the conjugate gradient method are special cases of the mollifier methods. To describe the result we make use of function spaces X, the mollifier is then a function of two variables and we consider $S_\gamma g(x) = \langle f, e_\gamma(x, \cdot)\rangle$ which means that the scalar product is taken with respect to the second variable of e_γ.

Theorem 3. Let the regularization method T_γ in (6) be given with a filter F_γ. Then this method can be written as an approximative inverse with mollifier

$$e_\gamma(x, y) = \sum_n F_\gamma(\sigma_n) v_n(x) v_n(y) \ . \tag{7}$$

Proof: The definition of v_γ as solution of $AA^* v_\gamma = Ae_\gamma$ in (2) leads with

$$v_\gamma(x) = \sum_n \sigma_n^{-1} \langle e_\gamma(x, \cdot), u_n\rangle v_n$$

to

$$v_\gamma(x) = \sum_n F_\gamma(\sigma_n) \sigma_n^{-1} u_n v_n(x) \ .$$

Then

$$S_\gamma g(x) = \langle g, v_\gamma(x)\rangle = T_\gamma g(x) \ .$$

Because of the versatility in the selection of the e_γ, which are functions of two variables, $e_\gamma(x, y)$, and which can differ for different x, it is not always possible to write the approximative inverse as a special regularization method. We can state the following result.

Theorem 4. Let e_γ be decomposed into

$$e_\gamma(x, y) = \sum_{mn} e_{\gamma,mn} v_n(x) v_m(y) \ .$$

Then the approximative inverse can be represented as regularization method based on a filter F_γ if and only if the filter satisfies

$$e_{\gamma,mn} = F_\gamma(\sigma_n)\delta_{mn} \ .$$

Proof: If we expand e_γ in terms of the v_n as

$$e_\gamma(x,y) = \sum_{mn} e_{\gamma,mn} v_n(x) v_m(y)$$

then

$$v_\gamma(x,\cdot) = \sum_n \sigma_n^{-1} u_n \sum_m e_{\gamma,mn} v_m(x) \ .$$

Hence

$$\begin{aligned} \langle v_\gamma(x,\cdot), g\rangle = & \sum_n \sigma_n^{-1} \langle u_n, g\rangle \sum_m e_{\gamma,mn} v_m(x) \\ = & \sum_n \sigma_n^{-1} F_\gamma(\sigma_n) \langle u_n, g\rangle v_n(x) \\ \Longleftrightarrow & \sum_m e_{\gamma,mn} v_m = F_\gamma(\sigma_n) v_n \ . \end{aligned}$$

With the orthonormality of the v_n follows

$$e_{\gamma,mn} = F_\gamma(\sigma_n)\delta_{mn} \ .$$

This leads to the above statement and shows that this is, at least from a theoretical point of view, the much more general approach. The regularization property can be interpreted in the following sense. For many integral operators the singular functions belonging to small singular values; i.e., which produce the instability, are highly oscillationg, see for example [5], [6], [14]. The local averaging in this method thus stabilizes the problem.

In the following we compare the method of approximative inverse with the Backus – Gilbert method for the situation $A : L_2(\Omega) \to \mathbb{R}^N$. In the classical Backus – Gilbert method an aproximation to the solution at point x is computed as

$$Sg(x) = \langle g, v(x)\rangle = \sum_{n=1}^N g_n v_n(x) \ ,$$

where $w_n(x)$ is determined by minimizing

$$\int_\Omega |x-y|^2 |\sum_{n=1}^N w_n(x) k_n(y)|^2 dy = \int_\Omega |x-y|^2 |A^* w(x)(y)|^2 dy$$

with the normalization

$$\sum_{n=1}^N w_n(x) \int_\Omega k_n(y) dy = \int_\Omega A^*(w(x))(y) = \langle A^* w(x), 1\rangle = 1 \ .$$

The first condition is interpreted as finding the $w(x)$ such that $A^* w(x)$ approximates the Delta distribution at the point x, the second condition gives a normalization because otherwise the solution to the minimization would be $w = 0$. The $w_n(x)$ are computed with Lagrangian mulitpliers; together with the multiplier μ they are the solution of

$$\begin{pmatrix} C(x) & h \\ h^\top & 0 \end{pmatrix} \begin{pmatrix} w(x) \\ \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

where the matrix elements of C are

$$C_{ij}(x) = \int_\Omega |x-y|^2 k_i(y) k_j(y) dy$$

and

$$h_i = \int_\Omega k_i(y) dy \ .$$

The matrix for determining the $v_n(x)$ depends here on the reconstruction point x, so it has to be decomposed for each x. Besides that the amount of smoothing, the spread of the point spread function, depends on the kernel $|x-y|^2$. If this should be changed then the kernel changes and the calculations have to start from the beginning. So the main difference to the approximative inverse is the attempt to approximate the Delta distributiuon instead of a mollifier and the extreme numerical effort.

4. Invariances Reduce the Storage Needs

Due to the fact that in the applications only a finite number of data is available we have $Y = \mathbb{R}^N$ or $Y = \mathbb{C}^N$. Then

$$S_\gamma g(z) = \sum_{n=1}^N g_n \ v_{\gamma,n}(z) \ ,$$

which means that we have to compute the reconstruction kernel $v_\gamma(x) \in \mathbb{R}^N$ for all reconstruction points x. These values are precomputed independently of the data. So computing time is not a big issue, but the storage needs can be prohibitively large.

In the following we use invariance properties of the operator A to reduce dramatically the storage needs for the method. To present the ideas we start with the following example.

Example 5. Let A be a convolution operator

$$Af(x) = \int_{\mathbb{R}^d} k(x-y) f(y) dy$$

and $X = Y = L_2(\mathbb{R}^d)$. Let T^z be the translation defined as

$$T^z f(x) = f(x-z) \ .$$

Then

$$T^z A = A T^z$$

and

$$T^z A A^* = A A^* T^z \ .$$

Proof: We easily see that

$$\begin{aligned} AT^z f(x) &= \int k(x-y) f(y-z) dy \\ &= \int k(x-z-u) f(u) du \\ &= Af(x-z) \\ &= T^z A f(x) \ . \end{aligned}$$

The operator AA^* is computed to

$$AA^* g(x) = \int_{I\!R^d} k^{(2)}(x,y) g(y) dy$$

where

$$k^{(2)}(x,y) = \int_{I\!R^d} k(x-u) k(y-u) du \ .$$

Then

$$\begin{aligned} AA^* T^z g(x) &= \int k^{(2)}(x,y) g(y-z) dy \\ &= \int k^{(2)}(x,u+z) g(u) du \\ &= \int k^{(2)}(x-z,u) g(u) du \\ &= T^z AA^* g(x) \end{aligned}$$

Let in the following E_γ be a function of one variable, we can think of $E_\gamma(y) = e_\gamma(0,y)$. Instead of a translation we consider arbitrary group representations.

Theorem 6. Let $A : X \to Y$ and let T_1^x be a group representation on X and T_2^x, T_3^x be group representations on Y such that

$$AT_1^x = T_2^x A \tag{8}$$

and

$$T_2^x AA^* = AA^* T_3^x \ . \tag{9}$$

Let w_γ be the solution of

$$AA^* w_\gamma = AE_\gamma \ . \tag{10}$$

Then the solution of

$$AA^* v_\gamma(x) = AT_1^x E_\gamma$$

is

$$v_\gamma(x) = T_3^x w_\gamma \ . \tag{11}$$

Proof : From the invariance properties follows

$$\begin{aligned} AT_1^x E_\gamma &= T_2^x AE_\gamma \\ &= AA^* T_3^x w_\gamma \end{aligned}$$

which completes the proof.

This means that only the solution w_γ has to be computed and stored, the kernels for other reconstruction points x are found by the action of T_1^x on E_γ and by T_3^x on w_γ.

For the above example of the convolution equation this can be applied in the following way. Let $T = T_1 = T_2 = T_3$ be the translation. The approximative inverse is then

$$\begin{aligned} S_\gamma g(x) &= \langle T^x w_\gamma, g\rangle \\ &= \int_{\mathbb{R}^d} w_\gamma(y-x) g(y) dy \end{aligned}$$

which again is of convolution type. Hence instead of a function of two variables $v_\gamma(x,y)$ only the function w_γ of one variable is needed.
In the case of a finite number of data where $(Af)_n = Af(x_n)$, $n = 1, \ldots, N$ for suitable points x_n the reconstruction kernel w_γ is a vector in $\mathbb{R}^N$ with $(w_\gamma)_n = w_\gamma(x_n)$. Then $w_\gamma(x_n - x)$ can be evaluated by linear interpolation between $(w_\gamma)_m$ and $(w_\gamma)_{m+1}$ with $x_m \leq x_n - x < x_{m+1}$.

For $Y = \mathbb{R}^N$ and M reconstruction points the storage needs originally was $M \times N$ real numbers, now it is only N real numbers!

5. Numerical examples

We study two examples in this section. First we consider $X = Y = L_2(0,1)$ and again the equation of the first kind

$$Af(x) = \int_0^x f(t)dt = g(x) \quad , \quad 0 \leq x \leq 1$$

with the solution $f = g'$. The adjoint operator is

$$A^* g(y) = \int_y^1 g(t)dt$$

and therefore the solution of

$$A^* v_\gamma(x,\cdot) = e_\gamma(x,\cdot)$$

is

$$v_\gamma(x,y) = -\frac{\partial}{\partial y} e_\gamma(x,y) \ .$$

With $e_\gamma(x,y) = \chi_{[-\gamma,\gamma]}(x-y)/(2\gamma)$ we get using distributional derivatives

$$v_\gamma(x,y) = \frac{1}{2\gamma}\left(\delta_{x+\gamma} - \delta_{x-\gamma}\right)$$

and hence the approximative inverse is the central difference quotient

$$S_\gamma g(x) = \frac{1}{2\gamma}\big(g(x+\gamma) - g(x-\gamma)\big)$$

which means that the central difference quotient with stepsize γ results in a local averaging of the searched – for solution with the characteristic function of the interval $[-\gamma,\gamma]$.
The bandlimiting filter $e_\gamma(x,y) = \gamma/\pi \mathrm{sinc}(\gamma(x-y))$ results in the bandlimiting differentiation formula with reconstruction kernel

$$v_\gamma(x,y) = \frac{\gamma(x-y)\cos\gamma(x-y) - \sin\gamma(x-y)}{\pi(x-y)^2} \ .$$

Both the mollifiers e_γ and the reconstruction kernels v_γ are of convolution type which minimizes both the storage and the computational effort, as discussed in the last section.

For the numerical experiment we consider $A : L_2(\Omega) \to I\!R^N$ and the integral equation

$$(Af)_n = \int_0^1 k(x_n, y) f(y) dy = g(x_n) \quad , \quad 1 \le n \le N \tag{12}$$

for equally distributed points $x_n = (n - 1/2)/N$ where the kernel is given as

$$k(x,y) = \begin{cases} x(1-y) & : \quad x \le y \\ y(1-x) & : \quad y \le x \end{cases}$$

Reconstructions with $e_\gamma(x,y) = \gamma \mathrm{sinc}(\gamma(x-y))/\pi$ are shown in [7]. Here we want to apply the approximative inverse with $e_\gamma(x,y) = (2\pi)^{1/2}\gamma^{-1}\exp((x-y)^2/(2\gamma^2))$ to determine both the solution and the derivative of the solution. For computing the derivative we start from

$$f'_\gamma(x) = \langle f', e_\gamma(x,\cdot)\rangle = -\langle f, e'_\gamma(x,\cdot)\rangle + f e_\gamma(x,\cdot)|_a^b \ ,$$

which means that we determine v_γ such that $\|A^* v_\gamma(x) + e'_\gamma(x,\cdot)\|$ is minimized. When $f(a)$ and $f(b)$ are known we can use directly this formula, otherwise we have to approximate these two values. The critical task of computing derivatives of the searched – for function is achieved here by differentiating the analytically given function e_γ instead of diferentiating the numerically determined function f. The prize to pay is that Ae'_γ has to be evaluated with high accuracy.

The right – hand side in the numerical example was chosen such that the exact solution is piecewise linear as shown in the following figures. Data errors occur also by the evaluation of Ae_γ. Fig. 4.2 shows the result with 5% data error.
The figures 4.3 and 4.4 show the computation of the derivative of the above function. Note that this function is not differentiable in the classical sense. The discontinuities and the behaviour of the derivative are clearly visible.

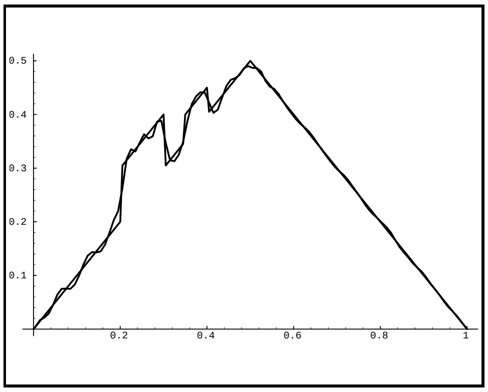

Fig. 4.1: Solution of (12) with exact data

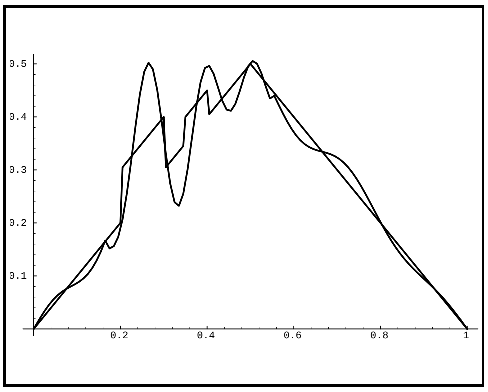

Fig. 4.2: Solution of (12) with noisy data, 5% noise

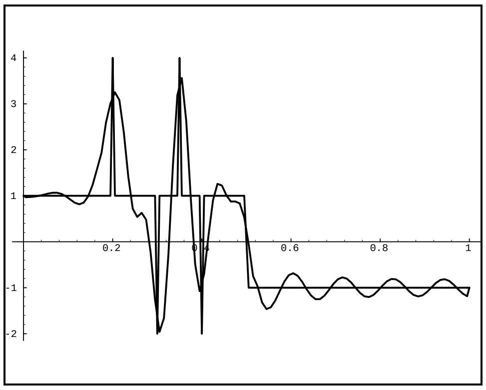

Fig. 4.3: Derivative of the solution of (12) with exact data

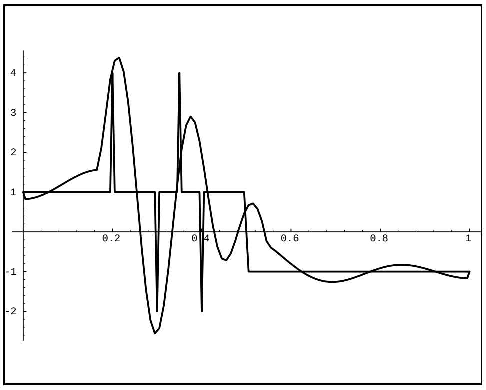

Fig. 4.4: Derivative of the solution of (12) with noisy data, 3% noise

6. Nonlinear problems

We start the presentation of the approach for the nonlinear problem

$$A : L_2(\Omega) \to I\!R^N$$

by considering for the sake of simplicity only a quadratic problem. Let the operator A be given as

$$Af = A_1 f + A_2 f$$

where A_1 is linear,

$$(A_1 f)_n = \int_\Omega k_n^1(y) f(y) dy$$

and A_2 is a quadratic operator defined as

$$(A_2 f)_n = \int_\Omega \int_\Omega k_n^2(y_1, y_2) f(y_1) f(y_2) dy_1 dy_2 \ .$$

With k^1 we denote the vector of the N components of k_n^1 and similarly with k^2 the vector of the N components of k_n^2. For the approximative inverse we follow Snieder [13] and make the following ansatz

$$f_\gamma(x) = \langle g, v_\gamma(x) \rangle + \langle g, V_\gamma(x) g \rangle$$

where $V_\gamma(x)$ is an $N \times N$ matrix. We replace g by Af and get

$$f_\gamma(x) \simeq \langle A_1 f, v_\gamma(x) \rangle + \langle A_2 f, v_\gamma(x) \rangle + \langle A_1 f, V_\gamma(x) A_1 f \rangle \tag{13}$$

where we omitted the higher order terms.
In the following we always use $x \in \Omega$ as the reconstruction point which is arbitrarily

fixed; and with y_1, y_2 and so on we denote the integration variables in the integral operators.
Of course we can attempt to approximate with the right – hand side an expression like $\langle f, e_\gamma(x,\cdot)\rangle$. But then the approximative inverse is not independent of the data, this does not lead to an approximative inverse as aimed for. We therefore follow [13] and consider the terms separately. We approximate the mollified solution with the linear term as well as possible which means

$$\langle A_1 f, v_\gamma(x)\rangle \simeq \langle f, e_\gamma(x,\cdot)\rangle \ ,$$

hence

$$A_1 A_1^* v_\gamma = A_1 e_\gamma \ .$$

The computation of v_γ thus follows exactly the lines as in the linear case. When this term presents the solution it only remains to make the rest as small as possible. Denoting the remainder by

$$R(x)(y_1,y_2) = v_\gamma(x)^\top k^2(y_1,y_2) + k^1(y_1)^\top V_\gamma(x) k^1(y_2) \tag{14}$$

where $a^\top b$ is the scalar product in the image space $\mathbb{R}^N$, we can write the rest in (13) as

$$\int_\Omega \int_\Omega R(x)(y_1,y_2) f(y_1) f(y_2) dy_1 dy_2 \ .$$

In order to make this term as small as possible we minimize with respect to $V_\gamma(x)$ for each reconstruction point x the norm

$$\|R(x)(\cdot,\cdot)\|^2_{L_2(\Omega\times\Omega)} \to \min \ .$$

We get

$$\begin{aligned}
\|R(x)\|^2 &= \int_{\Omega\times\Omega} \Big(v_\gamma^\top(x) k^2(y_1,y_2) + k^1(y_1)^\top V_\gamma(x) k^1(y_2) \Big)^2 dy_1 dy_2 \\
&= \int_{\Omega\times\Omega} \Big((v_\gamma(x)^\top k^2(y_1,y_2))^2 \\
&\quad + 2 v_\gamma(x)^\top k^2(y_1,y_2) k^1(y_1)^\top V_\gamma(x) k^1(y_2) \\
&\quad + (k^1(y_1)^\top V_\gamma(x) k^1(y_2))^2 \Big) dy_1 dy_2 \ .
\end{aligned}$$

Differentiating this expression with respect to the fixed matrix element $(V_\gamma)_{ij}(x)$ and equating the derivative to zero we find

$$\begin{aligned}
&\int_\Omega k_i^1(y_1) k^1(y_1)^\top dy_1 \ V_\gamma(x) \ \int_\Omega k_j^1(y_2) k^1(y_2)^\top dy_2 \\
&= - \int_{\Omega\times\Omega} v_\gamma^\top(x) k^2(y_1,y_2) k_i^1(y_1) k_j^1(y_2) dy_1 dy_2 \ , i,j = 1,\dots,N \ .
\end{aligned}$$

In matrix notion this is simply

$$(A_1 A_1^*) V_\gamma(x) (A_1 A_1^*) = - \sum_{n=1}^N v_{\gamma,n}(x) B_n \ , \tag{15}$$

where the matrices B_n on the right – hand side are

$$B_n = \int_{\Omega\times\Omega} k^1(y_1)k_n^2(y_1,y_2)k^1(y_2)^\top dy_1 dy_2 \ .$$

This means that with $C_n = (A_1A_1^*)^{-1}B_n(A_1A_1^*)^{-1}$ the function $k^{1^\top}(y_1)C_nk^1(y_2)$ is the orthogonal projection of the function $k_n^2(y_1,y_2)$ on the linear space span$\{k_i^1(y_1)k_j^1(y_2) : i,j = 1,\ldots,N\}$. Hence, if the k_n^1 are linearly independent, the matrix V_γ is simply

$$V_\gamma(x) = -\sum_{n=1}^N v_{\gamma,n}(x)C_n$$

and we get

$$f_\gamma(x) = \langle g, v_\gamma(x)\rangle - \sum_{n=1}^N v_{\gamma,n}(x)\langle g, C_n g\rangle \ .$$

For the approximation f_γ this means that

$$\begin{aligned} f_\gamma(x) &= \langle v_\gamma(x), g\rangle + \langle g, V_\gamma g\rangle \\ &= \langle v_\gamma, A_1 f\rangle + \sum_{n=1}^N v_{\gamma,n}(x)\Big((A_2f)_n - g^\top C_n g\Big) \ . \end{aligned}$$

If the kernels k_n^2 are already in the span of the product of the k_m^1 which means that $k_n^2(y_1,y_2) = k^1(y_1)^\top C_n k^1(y_2)$ and $(A_2f)_n = (A_1f)^\top C_n A_1 f$, then the contribution of the square term comes only from

$$\langle A_1f, V_\gamma(x)A_2f\rangle + \langle A_2f, V_\gamma(x)A_1f\rangle + \langle A_2f, V_\gamma(x)A_2f\rangle \ ,$$

see also the discussion in [13].

The generalization of this approach to operators of arbitrary order in the following sense

$$Af = \sum_{\ell=1}^\infty A_\ell f \tag{16}$$

where

$$(A_\ell f)_n = \int_{\Omega^\ell} k_n^\ell(y_1,\ldots,y_\ell)f(y_1)\ldots f(y_\ell)dy_1\ldots dy_\ell \tag{17}$$

is now straight forward, but notationally difficult. Examples for operators of this type are given in [13]. With $b^\ell(g_1,\ldots,g_\ell)$ we denote a multilinear form; i.e., $b^\ell(g_1,\ldots,g_\ell)$ is linear in each argument. This form is generated by the tensor v^ℓ. For $\ell = 1$ we have

$$b^1(g) =< v_\gamma, g > \quad , \quad v^1 = v_\gamma \ ,$$

for $\ell = 2$

$$b^2(g_1,g_2) =< g_1, V_\gamma g_2 > \quad , \quad v^2 = V_\gamma \ .$$

If we use always the same argument we simply write $b^\ell(g)$. To be precise we have to indicate the dependence of b^ℓ on the reconstruction point x, hence $b^\ell(x;g)$. The ansatz for the approximative inverse is now

$$f_\gamma^L(x) = \sum_{\ell=1}^L b^\ell(x;g) \tag{18}$$

where we determine the tensors generating the multilinear form b^ℓ as described above for $L = 2$. We replace g by Af and collect the terms with same number of factors f :

$$\begin{aligned} f_\gamma(x) = \quad & b^1(x; A_1 f) \\ & + b^1(x; A_2 f) + b^2(x; A_1 f, A_1 f) \\ & + b^1(x; A_3 f) + b^2(x; A_2 f, A_1 f) + b^2(x; A_1 f, A_2 f) + \\ & \quad b^3(x; A_1 f, A_1 f, A_1 f) \\ & + \ldots \end{aligned}$$

where b^1 and b^2 are determined in the way just described and we continue recursively. The remainder corresponding to (14) is denoted by R_ℓ and we minimize

$$\|R_\ell(x; y_1, \ldots, y_\ell)\|_{L_2(\Omega^\ell)} \ .$$

We denote the components of the tensor v^ℓ by v^ℓ_α with the multiindex $\alpha \in I\!N^\ell$. Then

$$R_\ell(x; y_1, \ldots, y_\ell) = \sum_{\alpha_1=1}^{N} \cdots \sum_{\alpha_\ell=1}^{N} v^\ell_\alpha k^1_{\alpha_1}(y_1) \cdots k^1_{\alpha_\ell}(y_\ell) + r_\ell(x; y_1, \ldots, y_\ell)$$

where r_ℓ depends on terms determined before. The norm of the rest is

$$\begin{aligned} \|R_\ell(x)\|^2_{L_2(\Omega^\ell)} = & \int_{\Omega^\ell} \{ \Big(\sum_{\alpha_1=1}^{N} \cdots \sum_{\alpha_\ell=1}^{N} v^\ell_\alpha k^1_{\alpha_1}(y_1) \cdots k^1_{\alpha_\ell}(y_\ell) \Big)^2 \\ & + 2 \sum_{\alpha_1=1}^{N} \cdots \sum_{\alpha_\ell=1}^{N} v^\ell_\alpha k^1_{\alpha_1}(y_1) \cdots k^1_{\alpha_\ell}(y_\ell) \\ & \times r_\ell(x; y_1, \ldots, y_\ell) \} dy_1 \ldots dy_\ell \\ & + \|r_\ell\|^2 \ . \end{aligned}$$

Differentiating with respect to a fixed v^ℓ_α leads to the following system of equations for the v^ℓ_β

$$\begin{aligned} & \sum_{\beta_1=1}^{N} \cdots \sum_{\beta_\ell=1}^{N} v^\ell_\beta \int_\Omega k^1_{\alpha_1}(y_1) k^1_{\beta_1}(y_1) dy_1 \cdots \int_\Omega k^1_{\alpha_\ell}(y_\ell) k^1_{\beta_\ell}(y_\ell) dy_\ell \\ & = - \int_{\Omega^\ell} k^1_{\alpha_1}(y_1) \cdots k^1_{\alpha_\ell}(y_\ell) r_\ell(x; y_1, \ldots, y_\ell) dy_1 \cdots dy_\ell \end{aligned}$$

Note that $\int_\Omega k^1_{\alpha_m}(y_m) k^1_{\beta_m}(y_m) dy_m = (A_1 A_1^*)_{\alpha_m \beta_m}$. We can collect the results in the following theorem.

Theorem 7. The matrix for determining the tensor v^ℓ does not depend on x, it consists of factors of $A_1 A_1^*$. The tensor has the form

$$v^\ell(x) = \sum_{n=1}^{N} v^1_n(x) B^\ell_n \tag{19}$$

where B^ℓ_n is a tensor of order ℓ and it is independent of x.

Proof : The first assertion is obvious from the above derived system of equations, the second statement follows by induction. For $\ell = 2$ we have

$$r_2(x; y_1, y_2) = v_\gamma(x)^\top k^2(y_1, y_2) \ .$$

In the step from ℓ to $\ell + 1$ the sum starts with $v_1^\top(x) k^{\ell+1}(y_1, \dots, y_{\ell+1})$ plus terms which are contained already in r_ℓ.

7. Numerical Example

We present numerical results for a problem discussed in [13] where the remainder r^2 was displayed but where numerical reconstructions are not given.
The differential equation for the vibrating string is

$$u''(x) + \frac{\rho(x)\omega^2}{T} u(x) = 0$$
$$u(0) = u(L) = 0 \ .$$

We assume the function ρ to be

$$\rho(x) = \rho_0(1 + f(x)) \ .$$

For $f \equiv 0$ the eigenfunctions are

$$u_n^0(x) = \left(\frac{2}{L}\right)^{1/2} \sin \frac{n\pi x}{L}$$

and the eigenvalues

$$\omega_n^0 = \left(\frac{T}{\rho_0}\right)^{1/2} \frac{n\pi}{L} \ .$$

The data g_n in our problem are the relative frequency shifts given as

$$g_n = \frac{\omega_n^2 - (\omega_n^0)^2}{(\omega_n^0)^2} \ , n = 1, \dots, N \ ,$$

which means that we study $A : L_2(\Omega) \to I\!R^N$. For the integral operator we get the kernels

$$k_n^1(y) = -\frac{2}{L} \sin^2(\frac{n\pi y}{L})$$

and

$$k_n^2(y_1, y_2) = \frac{4}{L^2} \sin^2 \frac{n\pi y_1}{L} \sin^2 \frac{n\pi y_2}{L} + \sum_{m \neq n} \frac{4}{L^2} \frac{n^2}{n^2 - m^2} \sin \frac{n\pi y_1}{L} \sin \frac{m\pi y_1}{L} \sin \frac{n\pi y_2}{L} \sin \frac{m\pi y_2}{L} \ .$$

The matrix $A_1 A_1^* = \frac{1}{2} I + ee^\top$, where $e = (1, \dots, 1)^| top$. For the reconstruction we used the piecewise linear mollifier

$$e_\gamma(x - y) = \frac{1}{\gamma} \begin{cases} 1 - \frac{y-x}{\gamma} & : \quad x \le y \le x + \gamma \\ 1 + \frac{y-x}{\gamma} & : \quad x - \gamma \le y \le x \\ 0 & : \quad \text{otherwise} \end{cases} \ .$$

Here $A_1 e_\gamma(x, \cdot)$ can be computed exactly.
The reconstructions are presented here for $N = 15$ without noise, Fig. 6.1, and with 3% noise in Fig. 6.2. Exact data means that the first 15 eigenvalues are computed with high precision using hierarchical finite elements. The function to be reconstructed is piecewise linear, the dashed lines (- - -) are the results from the linear part of the problem, and the straight lines are the reconstructions including also the quadratic part. These examples show that the additional effort in computing the quadratic approximation improves the result considerably.

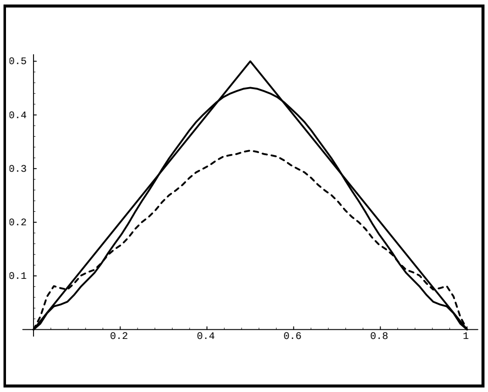

Fig. 6.1: Solution of inverse eigenvalue problem with exact data, linear and quadratic approximations of the exact solution

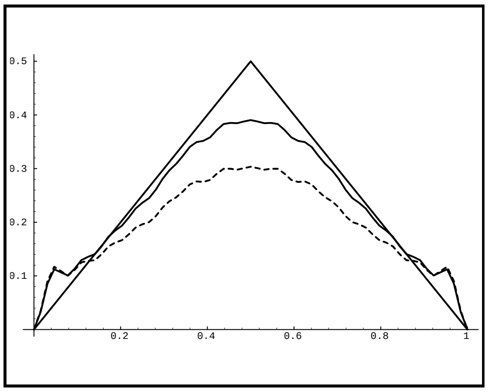

Fig. 6.2: Solution of inverse eigenvalue problem with noisy data (3%), linear and quadratic approximations of the exact solution

Acknowledgments

The work of the author was supported by the Deutsche Forschungsgemeinschaft in the Sonderforschungsbereich SFB 273 and by the Bundesministerium für Bildung, Wissenschaft, Forschung und Technologie (BMBF), grant No. 20M360 – 03LO7SAA-8.

References

[1] Backus G and Gilbert F 1967 *Geophys. J. R. Astron. Soc.* **13** 247-76
[2] Bertero M, de Mol C and Viano C 1980 in *Inverse scattering in optics* ed H P Baltes (Berlin:Springer) pp 161-214
[3] Eckhardt U 1976 *Computing* **17** 193-206
[4] Engl H W 1993 *Surveys on Mathematics for Industry* **3** 71-143
[5] Louis A K 1989 *Inverse und schlecht gestellte Probleme* (Stuttgart: Teubner)
[6] Louis A K 1992 *Inverse Problems* **8** 709-738
[7] Louis A K and Maaß P 1990 *Inverse Problems* **6** 427-440
[8] Louis A K and Maaß P, Rieder A (1994) *Wavelets* (Stuttgart:Teubner)

[9] Natterer F 1986 *The mathematics of computerized tomography* (Stuttgart: Teubner-Wiley)
[10] Parker R L 1994 *Geophysical Inverse Theory* (Princeton: U Press)
[11] Plato R and Vainikko G 1990 *Numerische Mathematik* **57** 63-79
[12] Sabatier P C 1987 in *Tomography and inverse problems* ed P C Sabatier (Bristol:Hilger) pp 471-667
[13] Snieder R 1991 *Inverse Problems* **7** 409-433
[14] Snieder R K, Beckers J and Neele F 1991 *J. Geophys. Res.* **96**501-515
[15] Strack T 1994 *Glättungsverfahren für nichtlineare inverse Probleme* (Saarbrücken: Master Thesis)