Regularization of Large–scale Ill–conditioned Least Squares Problems

Marielba Rojas L.

October 18, 1996

Abstract

Ill–conditioned problems arise in important areas like geophysics, medical imaging and signal processing. The fact that the ill–conditioning is an intrinsic feature of these problems makes it necessary to develop special numerical methods to treat them. Regularization methods belong to this class.

The lack of robust regularization methods for large–scale ill–conditioned problems motivated this project.

Our goal is to develop a regularization method for the least squares problem as a large–scale discrete ill–posed problem arising in seismic inversion.

Contents

1 Introduction

Ill–conditioned problems arise in important areas like geophysics, medical imaging and signal processing. Common problems in these areas are inverse problems which attempt to determine the structure of a system from the system's behavior.

Inverse problems are a natural source of (continuous) ill–posed problems. The numerical solution of these problems usually involves some kind of discretization which in turn originates a class of problems known as discrete ill–posed problems which are very ill–conditioned.

The fact that the ill–conditioning is an intrinsic feature of these problems makes it necessary to develop special numerical methods to treat them. Regularization methods belong to this class.

A regularization method computes an approximate solution, the *regularized solution*, to an ill–conditioned problem through a *regularization parameter*. A complete regularization method must address these two aspects.

A great variety of regularization methods are available for small to medium scale problems. The situation is different for the large–scale case. The lack of robust regularization methods for large–scale ill–conditioned problems motivated this project.

Our goal is to develop a regularization method for the least squares problem as a large–scale discrete ill–posed problem arising in seismic inversion.

In order to achieve that goal we define the following objectives:

1. Study the background on regularization of ill–conditioned problems that leads to a reasonable level of understanding of the problem.
2. Develop a strategy for computing the regularized solution.
3. Develop a strategy for computing the regularization parameter.
4. Implement and test the complete method.
5. Analyze the theoretical properties of the method.

The purpose of the present work is to address points 1 and 2. The organization of this work is the following. In section 2 we describe ill–conditioned problems and their features. Section 3 contains the description of the seismic application. In section 4 we review the least squares problem and analyze the ill–conditioned case in the presence of noisy data. Section 5 summarizes the existing regularization methods. In section 6 we discuss previous work for the large–scale case. In section 7 we describe the proposed approach. Final remarks are presented in section 8.

2 Ill–conditioned Problems

When ill–conditioned systems or least squares problems are encountered, the usual recommendation is not to trust any computed solution and to try to replace the coefficient matrix by a nearby well–conditioned one.

There are problems however, for which the ill–conditioning is an inherent feature and for which it is not possible to find a nearby problem with a well–conditioned coefficient matrix; this is the case for discrete ill–posed problems.

According to [26], we can distinguish two main classes of ill–conditioned problems based on the properties of their coefficient matrices:

1. **Rank–deficient Problems**. The properties of their coefficient matrices are:

   (a) There is a (usually) small cluster of small singular values.

   (b) There is a clear gap between large and small singular values, therefore the notion of numerical rank applies.

   (c) There is usually a reformulation that will eliminate the ill–conditioning.

   An example of the singular value distribution for a problem of this type is shown in figure 1.

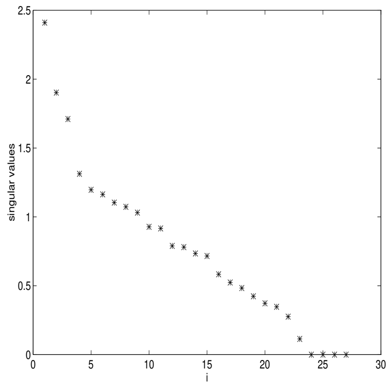

Figure 1: Singular Value distribution for a Rank–deficient Problem

2. **Discrete Ill–Posed Problems**. These problems come from the discretization of continuous ill–posed problems, and the properties of their coefficient matrices are:

   (a) There is a large cluster of small singular values (which increases with the dimension of the problem).

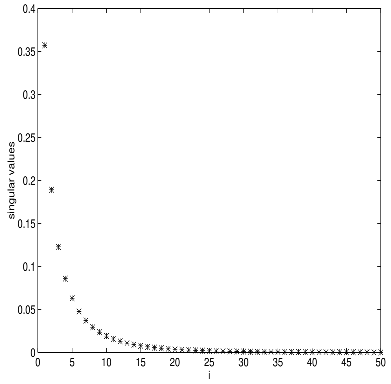

Figure 2: Singular Value distribution for a Discrete Ill–posed Problem

(b) The singular values decay gradually to zero with no gap in the singular spectrum, i.e. the notion of numerical rank does not apply.

(c) The components of b in the direction of the left singular vectors on average decay to zero faster than the singular values.

(d) There is usually no reformulation that can change these features.

The singular value distribution for a problem of this type (problem *heat* from [25]) can be observed in figure 2.

An extensive coverage of these two classes of ill–conditioned problems was presented in [26]. A detailed description of applications can be found in [20], [10] and the references therein. A specific application from geophysics will be described in this work.

Since standard methods fail to produce a meaningful solution for ill–conditioned problems and there are cases when it is not possible to eliminate the ill–conditioning, it is necessary to use other approaches to treat this kind of problems. One of such approaches is *Numerical Regularization*, which attempts to compute a more stable (or *regularized*) approximate solution by using additional information about the unknown exact solution.

The existing numerical regularization methods are designed for rank–deficient and discrete ill–posed problems. For other kinds of ill–conditioned problems, these methods are not suitable and we must use other techniques like iterative refinement, extended precision iterative refinement or preconditioning for the large–scale case.

In this project we focus on the regularization of discrete ill–posed problems.

3 Application

In 1923, Hadamard [18] introduced the concept of a well–posed and ill–posed problem. Ill–posed problems are those for which there is either no solution, or the solution is not unique, or the solution does not depend continuously on the data. By ill–posed he meant that the formulation was incorrect and the solution was therefore, meaningless. Since then, many industrial applications have appeared where solutions to ill–posed problems have well–defined physical meanings.

Ill–posed problems arise naturally from inverse problems, where one is interested in determining the structure of a system from its behavior. Oil exploration is one of the many fields where inverse problems are encountered.

In oil exploration, before the expensive drilling procedure takes place, experts need to determine the composition of the subsurface of the earth in places where geological data or other information indicates that oil or other valuable hydrocarbons are likely to be found.

The seismic inversion technique tries to determine the composition of the subsurface of the earth from the behavior of waves in that subsurface. To obtain data, seismic waves are generated (typically by an explosion) and their wave form (amplitude and direction) is measured by receivers located on the surface of the earth.

The goal of the experiment is to recover the slowness (inverse of wave propagation velocity) distribution of the subsurface. With this information (velocity with which waves travel in different media) experts can predict the composition of the subsurface.

The behavior of the waves travelling through the earth is modeled by means of the wave equation. After considerable simplifications, the problem to solve is

$$Find\ a\ function\ x\ such\ \ that\ \ \mathcal{F}x = b \tag{1}$$

where $\mathcal{F}$ is an ill–conditioned operator in most situations of practical interest. $\mathcal{F}$ being ill–conditioned means that some of its singular values (from the Singular Value Expansion (SVE) of the operator) are small. A discretization of $\mathcal{F}$ usually results in a matrix whose singular values are approximations of the singular values of the operator and whose singular vectors give information on the singular functions. We refer the reader to [23].

When b is replaced by real data $\bar{b}$, it will typically be the case that $\bar{b} \notin \mathcal{R}(\mathcal{F})$ and instead of (1), the following problem is considered

$$\min \ \| \mathcal{F}x - \bar{b} \|$$

where the minimization is on an appropriate function space and $\| \cdot \|$ is some suitable norm, usually the $\mathcal{L}_2$ norm.

For this choice of norm, the common approach to solving this minimization problem is known as *linearized inversion*. Two different techniques fall under this name. The first one, known as *asymptotic linearized inversion* (see [1]), finds an approximate inverse of the normal operator $\mathcal{F}^*\mathcal{F}$ (* denoting adjoint) analytically. If we denote by $\mathcal{P}$ that approximate inverse operator, the solution is computed as

$$x = \mathcal{P}\mathcal{F}^*\bar{b}$$

We are interested in the second approach known as *linearized least squares inversion* (see [50]). In this approach, $\mathcal{F}$ is discretized and $\| \cdot \|$ is chosen to be the l_2 norm in an appropriate vector space. The problem to solve is a discrete least squares problem.

Both techniques are used in practice. Asymptotic linearized inversion is efficient since there is no need to use an iterative process to compute the approximate solution. The disadvantage of the approach is that the accuracy of the solution is limited in view of the asymptotic nature of the approximate inverse operator used to compute it.

Because of the large size of the problems, linearized least squares inversion usually involves the use of an iterative process. While this increases the cost of the method, it also allows more control on the accuracy of the approximate solution. Moreover, linearized least squares inversion is the only option when it is not possible to compute the inverse of the normal operator analytically.

The problem to solve in least squares inversion can be stated as

Recover the solution x of

$$\min \ \| Ax - b \|_2 \tag{2}$$

from $\bar{x}$, the solution to

$$\min \ \| A\bar{x} - \bar{b} \|_2 \tag{3}$$

Where

1. $A \in \mathcal{R}^{m \times n}$ $(m \geq n)$ is the discretized operator.

2. m, n are large and A is unstructured.

3. A is not available explicitly, instead the action of A, A^T (and hence $A^T A$) on a vector are available.

4. The data vector $\bar{b} \in \mathcal{R}^m$ is a perturbation of the exact vector b caused by the noise in the experiment.

5. A inherits the ill–conditioning of $\mathcal{F}$ in the sense that it will have a similar singular value distribution, unless the discretization is so coarse that the singular value distribution of A does not resemble the singular value distribution of $\mathcal{F}$.

Since A comes from the discretization of an ill–posed operator, the resulting discrete least squares problem belongs to the class of discrete ill–posed problems introduced in section 1. In section 4 we study this kind of problem in more detail.

4 Least Squares Problems

The purpose of this section is to review some results on the least squares problem in general and analyze the special situation for the ill–conditioned case in the presence of noisy data.

The least squares problem has been thoroughly studied and is treated in most numerical linear algebra text books like [16] and [49], and also in specialized sources like the classic [31] and more recently in [2]. The problem is stated as follows

$$\min \parallel Ax - b \parallel_2$$

where $A \in \mathcal{R}^{m\times n}$, $m \geq n$ and $b \in \mathcal{R}^m$.

Well known facts about this problem are:

- There is always a solution.
- The solution with minimum norm is unique.
- Any solution satisfies the *Normal Equations*

$$A^T Ax = A^T b \tag{4}$$

An important tool for the analysis of the least squares problem is the **Singular Value Decomposition** (SVD) of the coefficient matrix. Let $A = U\Sigma V^T$ be such a decomposition, i.e.

$$\begin{aligned} &U \in \mathcal{R}^{m\times n}, \quad V \in \mathcal{R}^{n\times n} \\ &U^T U = I_n; \quad VV^T = V^T V = I_n \\ &\Sigma = diag(\sigma_1, \ldots, \sigma_n) \end{aligned}$$

where $\sigma_1 \geq \sigma_2 \geq \cdots \sigma_n \geq 0$ and σ_i's are the *singular values* of A. The columns of U are the *left singular vectors* and the columns of V are the *right singular vectors* of A.

To measure the sensitivity of the solutions to linear systems and least squares problems, we use the *condition number* of the coefficient matrix. We define this number with respect to the l_2 norm as

$$\kappa_2(A) = \frac{\sigma_1}{\sigma_n}$$

The sensitivity of the least squares solution is measured by either the condition number in the zero residual case or the *square* of the condition number in the nonzero residual case (see [16, Ch. 5]). This implies that when A is ill–conditioned, meaning that $\kappa_2(A)$ is large, the least squares solution will be unstable, i.e. very sensitive to perturbations in the data $A,\ b$.

Replacing A in (4) by its SVD yields

$$\begin{aligned} x &= V\Sigma^{\dagger}U^T b \\ &= \sum_{i=1}^{r} \frac{u_i^T b}{\sigma_i} v_i \end{aligned}$$

where † denotes the pseudoinverse as it is defined in [16] and r is the rank of A.

For the problems considered in this work, $\sigma_i \neq 0, \forall i$ (although some of them can be very small) and therefore $r = n$, which yields

$$x = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i \tag{5}$$

We note in passing that (5) is also the expression for the solution of the linear system $Ax = b$, when A is a square nonsingular matrix.

The norm of the least squares solution (5) is given by

$$\| x \|_2 = \left[\sum_{i=1}^{n} \left(\frac{u_i^T b}{\sigma_i}\right)^2 \right]^{\frac{1}{2}}$$

therefore, this norm will not be too large as long as $|u_i^T b| < \sigma_i$ for small σ_i. In view of this observation and in connection with discrete ill–posed problems, the following condition is usually assumed

Discrete Picard Condition (DPC)
$u_i^T b$ on average decay to zero faster than σ_i.

The need of this condition in order for most regularization methods to compute good approximate solutions was fully justified in [24].

Another fact usually observed in discrete ill–posed problems is that the singular vectors $u_i,\ v_i$ have more sign changes in their components as the singular values σ_i decrease, i.e. the high–frequency components correspond to small singular values.

As we explained in section 3, we are interested in recovering the solution x of

$$min \ \ \| Ax - b \|_2 \tag{6}$$

from $\bar{x}$, the solution to

$$min \ \ \| A\bar{x} - \bar{b} \|_2 \tag{7}$$

where $\bar{b} = b + \epsilon r$, with $\epsilon > 0$ and $r \in \mathcal{R}^m$ a random vector representing the noise.

The solution $\bar{x}$ of (7) is given by

$$\bar{x} = \sum_{i=1}^{n} \frac{u_i^T b}{\sigma_i} v_i + \sum_{i=1}^{n} \frac{\epsilon u_i^T r}{\sigma_i} v_i$$

Thus, $\bar{x}$ consists of two terms: one is the actual solution of the unperturbed problem (6) and the other is the contribution from the noise.

The difficulty in finding x from $\bar{x}$ comes from the contribution from the noise. Since it is not reasonable to expect that r (a vector of uncorrelated data) satisfies the DPC, it is possible that $u_i^T r$ increase or become constant for large i, causing the ratios $\frac{u_i^T r}{\sigma_i}$ to blow up as figure 3 illustrates. In figure 4 we show that in this case, $\bar{x}$ can differ considerably from x. We can also observe in figure 4 how $\bar{x}$ is dominated by high–frequency components.

The only type of errors considered in the previous analysis were measurement errors (noise) in b. We ignored other kind of errors in b because noise is typically larger than the rest. We also ignored errors in A (due to discretization or finite precision representation). If we take into account perturbations in both A and b, the difficulty of the problem increases considerably. The problem becomes a Total Least Squares problem (see [52]) and will not be treated in this work.

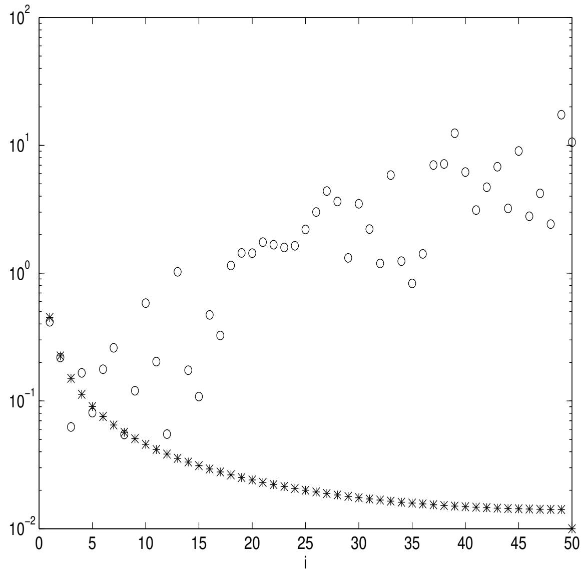

Figure 3: $|u_i^T b|/\sigma_i$ (in logarithmic scale) for exact (*) and noisy (o) data, problem *deriv*2 ($m = n = 50$) from the Regularization Tools package.

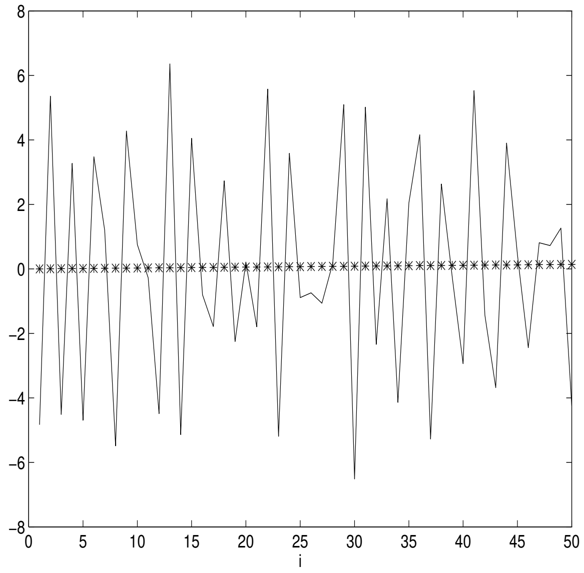

Figure 4: x (*) and $\bar{x}$ (-) for problem *deriv*2. The relative error in $\bar{x}$ is 43.78!

5 Regularization Concepts and Methods

Regularization is a technique that was originally devised for continuous ill–posed problems [51]. When the technique is extended to the discrete context it gives rise to *numerical regularization methods*. We limit our presentation to the latter case.

Numerical Regularization Techniques seek to approximate the exact (unknown) solution of an ill–conditioned problem by the solution of a related well–conditioned problem that includes information about the desired solution to the original problem.

The additional information is usually expressed as a constraint of the form

$$\| Lx \|_2 \leq \Delta \tag{8}$$

where L is typically the identity or a discretized derivative operator. In the latter case, the constraint is used to control the smoothness of the approximate solution. When $L = I$, the constraint controls the size of the solution.

To justify the use of (8), one must assume that the exact solution is smooth or that it has small norm. Regularization is also known as *smoothing* because it tries to damp nonsmooth components in the approximate solution. Note that when nonsmooth components correspond to small singular values, those components are magnified by the noise.

Regularization can also be regarded as a multi–objective optimization problem where one tries to balance the quality of the approximation and the effect of perturbations on the solution (this approach is taken in [29]). Associated with any regularization method there is a parameter, namely the *regularization parameter*, which controls how much perturbation effect is allowed.

Therefore, a complete regularization method has two aspects: the computation of the regularized solution and the computation of the regularization parameter.

In the rest of the section we summarize the main regularization techniques for both rank–deficient and discrete ill–posed problems. In section 6 we discussed methods for the large–scale case in detail. The methods presented here are suitable for ill–conditioned linear systems and least squares problems, the nonlinear least squares case has been studied in [54] and [11].

For a more detailed description of the methods, we refer the reader to [26]. Early surveys of regularization methods appeared in [3] and [53]. A common framework for the study of numerical regularization methods is proposed in [26] and [27].

5.1 Methods for Rank–deficient Problems

We recall from section 1 that the coefficient matrix A of a rank–deficient problem has a well–determined gap in the singular value spectrum, which makes possible to determine the *numerical rank* of the matrix.

The *numerical ϵ–rank* $r_\epsilon(A)$ of the matrix A is the number of columns that are numerically linearly dependent with respect to an error level ϵ, and it is defined as

$$r_\epsilon(A) = \min_{||E||_2 \leq \epsilon} rank(A + E)$$

In terms of the singular values of A

$$\sigma_{r_\epsilon(A)} > \epsilon \geq \sigma_{r_{\epsilon(A)}+1}$$

If $r_\epsilon(A)$ is ill–determined, i.e. if it is too sensitive to perturbations on ϵ and on the singular values, then it is better to use methods for problems with no gap in the singular spectrum.

The regularization strategy for rank–deficient problems consists of two steps

1. Replace A by a matrix of rank $k = r_\epsilon(A)$, i.e. ignore the small singular values of A. The usual choice for this rank–k matrix is

   $$A_k = \sum_{i=1}^{k} u_i \sigma_i v_i^T$$

   A_k is the closest rank–k matrix to A in the l_2 norm and in the Frobenius norm ([16, Ch. 2]).

2. Compute the approximate solution by

   $$x_k = \sum_{i=1}^{k} \frac{u_i^T b}{\sigma_i} v_i$$

   which is known as the *truncated SVD solution* and the regularization method is known as *truncated SVD*.

Since the SVD is expensive to compute, an alternative approach in practice is to compute rank–revealing decompositions like the QR decomposition with column pivoting ([16, Ch. 5]). In this case the matrix we use to replace A in step 1 above is still close to A, but not the closest one in any norm.

Although the truncated SVD method does not use additional information about the desired solution, the method is still regarded as a regularization method because it stabilizes (*regularizes*) the solution.

5.2 Methods for Discrete Ill–posed Problems

In the following we assume that $L = I$, whenever a constraint of the form (8) is used. If the problem is given in general form ($L \neq I$) it is possible to transform it to the standard form ($L = I$) by means of the algorithms given in [9] and [26].

I. Methods for computing the regularized solution

1. Direct Methods

   (a) Tikhonov Regularization.
   The method solves the following problem

   $$\min\{\| A\bar{x} - \bar{b} \|_2^2 + \lambda^2 \| \bar{x} \|_2^2\} \tag{9}$$

   where λ is the regularization parameter. Alternative formulations (for $\lambda \geq 0$) are

   $$(A^T A + \lambda^2 I)\bar{x} = A^T\bar{b}$$

   i.e. find a zero for the gradient of the objective function in (9), and

   $$\min \left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} \bar{x} - \begin{pmatrix} \bar{b} \\ 0 \end{pmatrix} \right\|_2$$

   which is a *damped least squares* problem.
   The solution of any of these problems is computed using either the SVD or the QR factorization of the matrix $\begin{pmatrix} A \\ \lambda I \end{pmatrix}$. The method is suitable for small to medium scale problems only, since one minimization problem has to be solved for each value of λ.

(b) Iterated Tikhonov Regularization.
This method is inspired in the iterative refinement technique for linear systems and consists of computing a sequence of approximations $\bar{x}^{(q+1)} = \bar{x}^{(q)} + \delta\bar{x}^{(q)}$, where $\delta\bar{x}_q$ solves

$$\left\| \begin{pmatrix} A \\ \lambda I \end{pmatrix} \delta\bar{x}^{(q)} - \begin{pmatrix} r^{(q)} \\ 0 \end{pmatrix} \right\|_2, \quad r^{(q)} = \bar{b} - A\bar{x}^{(q)}$$

(c) Quadratically constrained least squares.
In this approach, the following problems are considered

$$\begin{array}{ll} \min & \| A\bar{x} - \bar{b} \|_2 \\ s.t. & {\scriptstyle \|\bar{x}\|_2 \leq \Delta} \end{array} \tag{10}$$

$$\begin{array}{ll} \min & \| \bar{x} \|_2 \\ s.t. & {\scriptstyle \|A\bar{x}-\bar{b}\|_2 \leq \alpha} \end{array} \tag{11}$$

where Δ and α are the regularization parameters.

The theoretical aspects of problems (10) and (11) were discussed in [12].

Problem (10) is known as a trust region subproblem in the optimization literature. We will return to methods to solve this problem in section 7. Methods to solve (10) based on the SVD and the QR decomposition are presented in [16, Ch. 12].

A method for problem (10) in the large–scale case was presented in [17] and will be discussed in section 6.

In problem (11) the parameter α can be interpreted as the noise level in the data.

(d) Other methods. We mention mollifier methods where the regularized solution $\bar{x}_{reg}$ is computed as $\bar{x}_{reg} = A^{\#}\bar{b}$. $A^{\#}$ is a special matrix known as the *resolution matrix*. Other methods are discussed in [26].

2. Iterative Methods.

These methods are intended for large–scale problems for which factorizations are not affordable. For these problems, the coefficient matrix is usually not available explicitly. In this class of methods we have

(a) Landweber iteration.

$$\bar{x}_{k+1} = \bar{x}_k + \omega A^T(\bar{b} - A\bar{x}_k) \quad k = 0, 1, 2, \ldots$$

where ω is a parameter. This method in its original form is not used in practice since it is not very efficient. A modification has been proposed in [19] to accelerate convergence.

(b) Conjugate Gradient on the Normal Equations (CGLS).
This has been the method of choice for large–scale ill–conditioned least squares problems. We will discuss the method in section 6.

Other alternatives like the ν–methods referenced in [20] are still too restrictive to be used for practical problems.

II. Methods for computing the regularization parameter

Let us recall the multi–objective purpose of regularization, i.e. minimize the residual norm while minimizing the effect of perturbations in the data. Therefore, there are two kinds of errors in any regularized solution. These errors are known as *regularization error* and *perturbation error*. Regularization tries to balance both errors by means of the regularization parameter.

There are two kinds of methods for choosing the regularization parameter. Methods in the first class assume some knowledge on the noise level in the data. The only method in this class is the Discrepancy Principle. The second class of methods try to estimate the parameter assuming no knowledge on the noise level. To this latter class belong the Generalized Cross–Validation criterion and the L–curve criterion.

1. Discrepancy principle.
When we know a bound for the noise level α in the data, we can use this bound to implicitly compute a regularization parameter as the point where $\| A\bar{x} - \bar{b} \|_2 \leq \alpha$. The idea behind this criterion is that we cannot expect more accuracy in the approximate solution than the one present in the data. This method is attributed to Morozov [35].

2. Generalized Cross–Validation (GCV).
This method was presented in [15] and discussed also in [56]. The idea of using cross–validation to compute the regularization parameter is the following. If a data point $\bar{b}_i$ is excluded and a regularized solution $\bar{x}_{reg,i}$ based on the reduced data vector is computed, then if we use $\bar{x}_{reg,i}$ to compute an estimate of $\bar{b}_i$ we want this estimate to be good. While in

ordinary cross–validation the ordering of the data counts, generalized cross–validation is invariant to orthogonal transformations of the data vector.

The regularization parameter is chosen as the minimizer of the Generalized Cross–Validation (GCV) function

$$GCV(\lambda) = \frac{\| A\bar{x}_{reg} - \bar{b} \|_2^2}{[trace(I - AA^{\#}(\lambda))]^2}$$

where $A^{\#}(\lambda)$ is the matrix that maps the data vector $\bar{b}$ onto the regularized solution $\bar{x}_{reg}$, i.e. the resolution matrix mentioned before.

There are several difficulties associated with the GCV criterion. One is that sometimes the minimum of the function is difficult to compute numerically. The second one is that sometimes GCV cannot distinguish between correlated noise and signal. A third problems is that it may not be possible to compute $A^{\#}(\lambda)$, as it is the case for CGLS.

3. L–curve.
   This method is based on the trade–off curve between the two goals of regularization, i.e. minimizing the residual norm while keeping the solution norm from being too large. The L–curve is a logarithmic plot of the solution norm versus the residual norm for each value of the regularization parameter. The name comes from the fact that this curve is L–shaped.

   The optimal regularization parameter gives a solution that lies around the "corner" of the curve. Figure 5 shows the L–curve (and its corner) for problem *heat* from [25]. In this example, the curve is based on the values of the regularization parameter in Tikhonov regularization.

   The use of this curve to estimate the regularization parameter has been studied in [22] and [27]. The idea is to interpolate the curve in order to estimate the "corner". The L–curve method performs better than the GCV method when the noise in the data is correlated and comparably well for white noise. The advantage of the L–curve criterion over the GCV in the presence of correlated noise seems to come from the fact that the L–curve method uses information on both the residual norm and the solution norm, while the GCV method uses information on the residual norm only (see [27]).

   Both the L–curve and the GCV methods are *a posteriori* methods

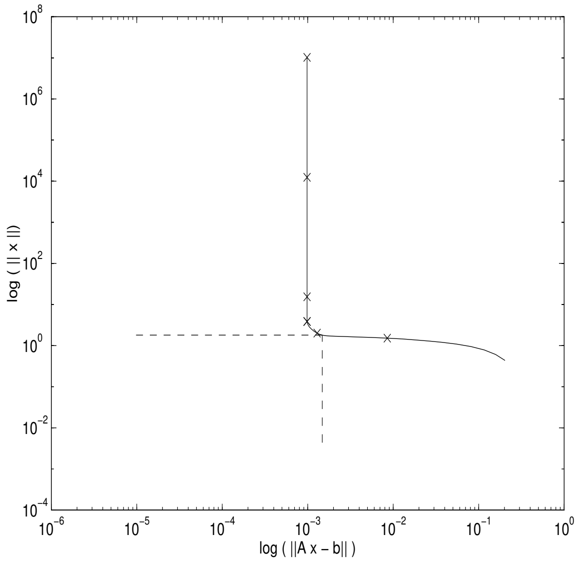

Figure 5: L–curve for problem *heat* and Tikhonov Regularization

since they need several approximate solutions in order to estimate the regularization parameter.

III. Hybrid Methods

These methods combine an iterative scheme to compute the regularized solution with a strategy to compute the regularization parameter, aiming to control the number of iterations in the iterative method. The method presented in [4] belong to this class. We will discuss that approach in section 6.

6 Regularization in the large–scale case

As we can see from section 5, there are very few methods for the regularization of large–scale ill–conditioned problems. The ones discussed here are the hybrid method presented in [4] (BGvD), the method for the quadratically constrained least squares problem presented in [17] (GvM) and the Conjugate Gradient method [28] applied to the normal equations (CGLS). The first method is the only complete regularization method in the sense that it computes both the regularization parameter and the regularized solution. The last two methods compute the regularized solution only.

6.1 BGvD Method

The problem is to recover the solution to min $\| Ax - b \|_2$ when b is replaced by noisy data.

The method uses an Implicitly Restarted Lanczos Bidiagonalization process to compute a sequence of approximations to the left and right singular vectors of A. The implicit restart technique from [46] is adapted to the Lanczos Bidiagonalization process [38] and zero shifts are used to filter out small singular values.

Reorthogonalization of the Lanczos vectors is carried out at every step of the Lanczos process. This makes possible the use of the Generalized Cross–Validation (GCV) function to estimate the regularization parameter k (in this case, the size of the factorization) when the noise level in the data is unknown.

The approximate solution computed by this method is a Truncated Singular Value Decomposition (TSVD) solution in a Krylov subspace of dimension k.

Advantages

- Uses A and A^T only in matrix–vector products.
- Effective when the minimum of the GCV function occurs at small values of k.

Drawbacks

- It is necessary to store the two matrices U_k and V_k for variable k.
- If the minimum of the GCV function occurs at a large k, as in the example in figure 6, then the size of the factorization also increases

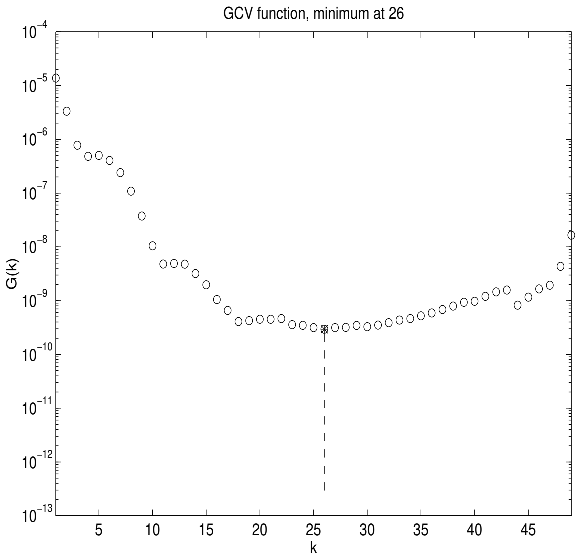

Figure 6: GCV function for problem *heat* from the Regularization Tools Package. The minimum occurs at $k = 26$ for a problem of dimension 50.

with the consequent increase in both storage and computational cost. For this method, the increase in computational cost is higher than usual due to the reorthogonalization step.

- As we discussed in section 5, the GCV criterion for computing the regularization parameter is not reliable in certain cases.

6.2 GvM Method

The problem in consideration is

$$\begin{array}{ll} \min & \| A\bar{x} - \bar{b} \|_2 \\ s.t. & \|\bar{x}\|_2 = \Delta \end{array}$$

where $\bar{b}$ is a vector of noisy data.

The method uses the Lanczos Bidiagonalization process [38] performed on A and the Cholesky factorization of small matrices, to compute a series of function pairs $(L_k(\lambda), U_k(\lambda))$ that bound the secular function

$$f(\lambda) = (A^T\bar{b})^T[(A^TA + \lambda I)^{\dagger}]^2(A^T\bar{b})$$

A zero–finding procedure is then applied to $L_k(\lambda)$ and $U_k(\lambda)$ to compute a new estimate for λ, the Lagrange multiplier associated to the minimization problem.

Advantages

- Uses A and A^T only in matrix–vector products.
- The approximate solution x can be computed in an inexpensive way after the optimal λ has been found, since a partial Bidiagonalization of A is available at that point.

Drawbacks

- It is necessary to store the two matrices U_k and V_k for variable k.
- The method works under the assumption that $\Delta <\| A^\dagger \bar{b} \|_2$. This assumption is not enough in the regularization context since it might still produce a solution with large norm. Instead, in order to use the method for regularization purposes Δ must be such that $\Delta \leq \| A^\dagger b \|_2 < \| A^\dagger \bar{b} \|_2$, where b is the exact data vector. This requires knowledge on the norm of the *exact* solution $A^\dagger b$ of the unconstrained problem.
- Computes the regularized solution only.

6.3 CGLS Method

One approach to solve the least squares problems is to apply the Conjugate Gradient method to the normal equations

$$A^T A \bar{x} = A^T \bar{b}$$

An implementation of such method should avoid forming the matrix $A^T A$ since doing so may introduce large rounding errors as it was pointed out in [16, example 5.3.2 on p.225]. The resulting method is known as Conjugate Gradient on the Normal Equations (CGNR, CGLS). The method has been used successfully to solve the least squares problems for noisy data due to an intrinsic regularization effect of the iteration.

The Conjugate Gradient method generates iterates in a Krylov subspace. In particular, for CGLS the iterates $\bar{x}_k \in \mathcal{K}_k(A^T A, A^T \bar{b})$ defined as

$$\mathcal{K}_k(A^T A, A^T \bar{b}) = \{A^T \bar{b}, (A^T A) A^T \bar{b}, \ldots, (A^T A)^{k-1} A^T \bar{b}\}$$

$\bar{x}_k$ is the solution to the following minimization problem

$$\begin{array}{ll} min & \| A\bar{x} - \bar{b} \|_2 \\ s.t. & {\scriptstyle \bar{x} \in \mathcal{K}_k(A^T A, A^T \bar{b})} \end{array}$$

The regularizing effect of CGLS comes from the fact that for some applications, $\mathcal{K}_k(A^T A, A^T \bar{b})$ approximates the subspace spanned by the right singular vectors v_i associated to the k largest singular values and therefore, $\bar{x}_k$ only has components in the direction of singular vectors associated to large singular values.

When the iteration starts approximating singular vectors associated to small singular values, contributions from the noise appear in the iterates $\bar{x}_k$, which start to diverge from that point. This behavior is known as *semiconvergence.*

It is necessary therefore to stop the iteration before the effect of the noise appears. The number of iterations k plays the role of the regularization parameter.

The semiconvergence behavior suggests the following stopping criterion: monitor the residual norm and stop the iteration when it starts to increase. The problem with this strategy is that in general, there is no warranty that *all* the singular vectors of interest have already converged when the iterates start to diverge. It is therefore necessary to estimate the regularization parameter k (i.e. where to stop the iteration) very accurately. Two options mentioned in [26] are the L–curve criterion and the Monte Carlo Cross–Validation procedure from [14].

For some problems arising in medical imaging (see [36] for example), all the singular vectors of interest converge first and therefore the stopping criterion suggested above can be used. The difficulty in those applications comes from the fact that the number of large singular values is large, requiring the use of efficient preconditioners to accelerate convergence.

Advantages

- Uses A and A^T only in matrix–vector products.
- The method requires very little storage: five vectors.
- Fast convergence.

Drawbacks

- The semiconvergence behavior makes it necessary to have an accurate estimate of the regularization parameter.
- Computes the regularized solution only.

6.3.1 Preconditioned CGLS in the ill–conditioned case

The issue of preconditioning in relation to ill–conditioned problems is very delicate. The usual goal of preconditioning is to improve the condition of the coefficient matrix by either clustering its eigenvalues or making them close to one. In the ill–conditioned case however, it is not desirable to change the whole spectrum of the matrix.

If we precondition all the eigenvalues of $A^T A$ (and therefore, the singular values of A) we would be computing iterates that contain contributions from the noise, since the preconditioner has probably mixed the small and large parts of the spectrum. Thus, in the ill–conditioned case we should precondition only the large part of the spectrum and leave the small part untouched. This fact has been observed before in [20], [26, Ch. 5], [21] and [36].

In general, it is not possible to distinguish a priori between the large and small parts of the spectrum of a matrix. This makes preconditioning for ill–conditioned problems a very difficult area and the object of current research. For problems where large and small parts of the spectrum can be identified, it has been possible to build efficient preconditioners as the ones reported in [7] and [36]. These problems are characterized by having a highly structured coefficient matrix (Toeplitz matrix) for which circulant preconditioners have proved to be successful. Multilevel preconditioners for more general problems have been proposed in [30].

7 Proposed Approach

In this section we present our approach for regularizing large–scale ill–conditioned least squares problems and propose a strategy for computing the regularized solution. We also present a brief discussion about the computation of the regularization parameter.

7.1 Computing the Regularized Solution

We compute the regularized solution of large–scale ill–conditioned least squares problems as the solution to the quadratically constrained least squares problem

$$\begin{array}{ll} \min & \frac{1}{2} \parallel A\bar{x} - \bar{b} \parallel_2^2 \\ s.t. & \|\bar{x}\|_2^2 \leq \Delta^2 \end{array} \tag{12}$$

where Δ is the regularization parameter and $\bar{b}$ is a vector of noisy data.

Problem (12) is a Trust Region Subproblem (TRS)

$$\begin{array}{ll} \min & \psi(x) = \frac{1}{2} x^T H x + g^T x + c \\ s.t. & \|x\|_2^2 \leq \Delta^2 \end{array} \tag{13}$$

where $H \in \mathcal{R}^{n \times n}$ symmetric, $g \in \mathcal{R}^n$ and $c \in \mathcal{R}$. In (12) $H = A^T A$ and $g = -A^T \bar{b}$.

Two important facts about the TRS are

- There is always a solution, since a continuous functional is being minimized on a compact set.
- There exists a characterization of the global solutions. This is a remarkeable result obtained independently by Gay [13] and Sorensen [45]. It is contained in lemma 7.1.

This problem has been treated extensively in the context of the trust–region globalization strategy for optimization methods.

In the small to medium scale setting, there are efficient methods to solve the problem. Such methods include Powell's dogleg method [40], Dennis and Mei's double dogleg method [8] and Moré and Sorensen's method [34]. The latter one being the method of choice in most applications.

A complete survey of methods was given by Moré in [33]. In that reference, Steihaug's method [48] is also mentioned. This method is suitable for the large–scale case and gives nearly optimal solutions when the Hessian of the quadratic is positive semidefinite. Methods for the general case were not available until recently.

It is worth noticing that Steihaug's approach could be used on problem (12). However, the fact that this method is based on the Preconditioned Conjugate Gradient method constitutes a major obstacle in the ill–conditioned case where finding efficient preconditioners is still an issue as we pointed out in section 6.

The known methods for the large–scale TRS were given by Sorensen in [47], Rendl and Wolkowicz in [41] and Santos and Sorensen in [43].

In [47] the TRS is recasted as a parameterized eigenvalue problem. The method relies on the Implicitly Restarted Lanczos Method (IRLM) ([46], [5]) to solve a sequence of large symmetric eigenvalues problems. The parameter is updated by an interpolation scheme and there is a separate treatment of a special case known as the *hard case*.

In [41] the same recasting as in [47] is used but the strategy for updating the parameter is a dual simplex method in the standard case and a primal simplex method in the hard case. The eigenvalue problems are solved with a block–Lanczos procedure.

In [43] the recasting of [47] is used again, but the interpolation scheme is different and there is a unified treatment of all cases. A small advantage of this method over the one in [41] is reported in [43].

We proposed to use the method for large–scale trust region subproblems (TRS method) presented in [43] to compute a solution for the constrained least squares problem (12).

The use of methods for the TRS in the regularization context has been suggested in [41] and [47]. In [32] a new trust region strategy is presented and applied to a regularization problem as an example. That particular strategy is not suitable for large–scale problems. [6] presents a method for the TRS for the special case of ill–conditioned least squares problems. The method works under the assumption that $\Delta < \| A^{\dagger}\bar{b} \|_2$.

Interestingly enough, in [44] a *nonlinear* least squares problem is solved by means of the Gauss–Newton method with a trust region globalization strategy, using Steihaug's method to solve the trust region subproblems. In that approach the TRS was not used as a way of regularizing the solution.

There is practically no reference to methods for the TRS in the regularization literature. This is maybe due to the fact that until now there was no method for the large–scale trust region subproblem and that there were efficient regularization strategies for smaller problems.

7.1.1 The method for the large–scale Trust Region Subproblem (TRS method)

The following definitions will be used

- $\delta_1, \delta_2, \ldots, \delta_n$ will be the eigenvalues of H in (13), with

$$\delta_1 \leq \delta_2 \leq \ldots \leq \delta_n$$

- The eigenspace associated with the smallest eigenvalue of H will be denoted by $\mathcal{S}_1$, i.e.

$$\mathcal{S}_1 = \{q | Hq = \delta_1 q\}$$

- The pseudoinverse will be denoted by † as in section 4.

Before presenting the TRS method of [43], let us first study the properties of problem (13).

The following lemma contains the characterization of the solutions of (13).

Lemma 7.1 *A feasible vector x_* is a solution of (13) if and only if x_* satisfies*

(i) $(H - \lambda I)x_* = -g$, *with* $H - \lambda I$ *positive semidefinite.*

(ii) $\lambda \leq 0$.

(iii) $\lambda(\| x_* \|_2 - \Delta) = 0$.

Proof: See [45] □

The solutions of (13) can occur on the boundary or in the interior of the trust region $\{x | \| x \|_2 \leq \Delta\}$. Interior solutions exist if and only if $\| x \|_2 < \Delta$ and H is positive definite (see [34]). In this case, the Lagrange multiplier λ is zero and the solution is given by $x = H^{-1}g$.

There is yet one more special situation that may arise only when g is orthogonal to $\mathcal{S}_1$ and H is not positive definite. In this case, depending on the value of Δ, it may not be possible to find $\lambda \leq 0$ that satisfies the conditions in lemma 7.1. This situation is known as the *hard case*. In the hard case, it is still possible to compute a solution on the boundary of the trust region. The precise statement is contained in the following lemma.

Lemma 7.2 *Assume that g is orthogonal to $\mathcal{S}_1$ and let $p = -(H - \delta_1 I)^\dagger g$. If $\delta_1 \leq 0$ and $\| p \|_2 < \Delta$, then the solutions of (13) consist of the set* $\{x | x = p + z, z \in \mathcal{S}_1, \| x \|_2 = \Delta\}$.

Proof: See [45] $\square$

Near hard case situations (when g is nearly orthogonal to $\mathcal{S}_1$) are very important in the context of regularization because they arise very frequently when the ill–conditioned TRS is treated. We will return to this in §7.1.2.

To describe the method in [43] we first explain how the TRS (13) is recasted as a parameterized eigenvalue problem, for this we define the following *bordered matrix*

$$B_\alpha = \begin{pmatrix} \alpha & g^T \\ g & H \end{pmatrix} \tag{14}$$

where α is a parameter. Note that for any value of α, the eigenvalues of B_α $\lambda_1(\alpha), \lambda_2(\alpha), \ldots, \lambda_{n+1}(\alpha)$ satisfy $\lambda_1(\alpha) \leq \delta_1 \leq \lambda_2(\alpha) \leq \delta_2 \leq \ldots \leq \delta_n \leq \lambda_{n+1}(\alpha)$ by the Cauchy interlacing theorem (see [39]).

If we now define $\varphi(x)$ as

$$\begin{aligned} \varphi(x) &= \frac{1}{2}(1 \quad x^T) B_\alpha (1 \quad x^T)^T \\ &= \frac{1}{2}\alpha - c + \psi(x) \end{aligned}$$

i.e. $\varphi(x)$ is a vertical translation of $\psi(x)$ in (13) and therefore, both functions will have the same minimizers.

Note that (12) is equivalent to

$$\begin{array}{ll} \min & \varphi(x) \\ s.t. & \|(1 \quad x^T)^T\|_2 \leq \sqrt{1+\Delta^2} \end{array} \tag{15}$$

Moreover, if B_α is indefinite, (15) has no interior solutions and the problem is equivalent to finding the smallest eigenpair of B_α with an eigenvector of the form $(1 \quad x^T)^T$ and norm $\sqrt{1+\Delta^2}$. The smallest eigenvalue is the minimum value of $\varphi(x)$ and x is the minimizer.

This observation suggests computing a (boundary) solution for (13) in terms of an eigenpair of the bordered matrix.

Suppose that we solve

$$B_\alpha = \begin{pmatrix} \alpha & g^T \\ g & H \end{pmatrix} \begin{pmatrix} 1 \\ x \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ x \end{pmatrix} \tag{16}$$

for the smallest eigenpair $\{\lambda_1(\alpha), (1 \;\; x^T)^T\}$ of B_α.

Observe that (16) is equivalent to

$$\begin{aligned} \alpha - \lambda &= -g^T x \\ (H - \lambda I)x &= -g \end{aligned}$$

moreover, $H - \lambda I$ is always positive semidefinite since, by the interlacing property, $\lambda_1(\alpha) \leq \delta_1$ for *any* value of α.

The idea is to adjust the parameter α in order to drive the x to the solution of (13). To accomplish this, the method works with the functions $\phi(\lambda)$ and its derivative defined as

$$\begin{aligned} \phi(\lambda) &= -g^T x \\ &= g^T (H - \lambda I)^\dagger g, \quad for \ (H - \lambda I)x = -g \\ &= \sum_{i=1}^{n} \frac{\gamma_i^2}{\delta_i - \lambda} \end{aligned}$$

$$\begin{aligned} \phi'(\lambda) &= x^T x \\ &= g^T [(H - \lambda I)^\dagger]^2 g, \quad for \ (H - \lambda I)x = -g \\ &= \sum_{i=1}^{n} \frac{\gamma_i^2}{(\delta_i - \lambda)^2} \end{aligned}$$

where γ_i is the component of g in the direction of the ith eigenvector of H, $i = 1, 2, \ldots, n$. Observe that the interlacing property of the eigenvalues of B_α and H also follows from the definition of $\phi(\lambda)$. Moreover, the definition of $\phi(\lambda)$ implies that $\lambda_1(\alpha) < \delta_1$, unless $\gamma_1 \neq 0$ (i.e. unless $g \perp \mathcal{S}_1$).

Since $\phi(\lambda)$ and $\phi'(\lambda)$ are expensive to compute, they are interpolated. A two point interpolating scheme is used in [43]. The interpolant $\hat{\phi}(\lambda)$ is built in such a way that convergence is guaranteed.

Figures 7 and 8 show $\phi(\lambda)$ and $\phi'(\lambda)$ respectively in the standard case for a problem of dimension three with eigenvalues $-2, \ -0.5, \ 2$.

To adjust α, $\hat{\lambda}$ is computed such that $\hat{\phi}'(\hat{\lambda}) = \Delta^2$ and then $\alpha = \hat{\lambda} + \hat{\phi}(\hat{\lambda})$. A new problem of type (16) is solved for this value of α to obtain a new pair $\{\lambda_1(\alpha), (1 \;\; x^T)^T\}$.

If the eigenvector associated to the smallest eigenvalue of B_α has a zero first component, then the required normalization cannot take place. This situation corresponds to a potential hard case situation. The following lemma establishes this.

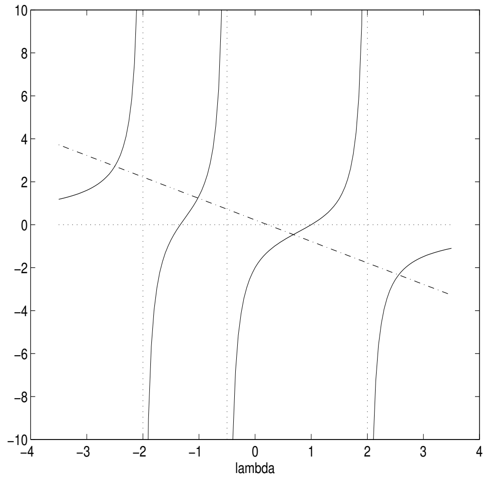

Figure 7: $\phi(\lambda)$ in the standard case and straight line $\alpha - \lambda$

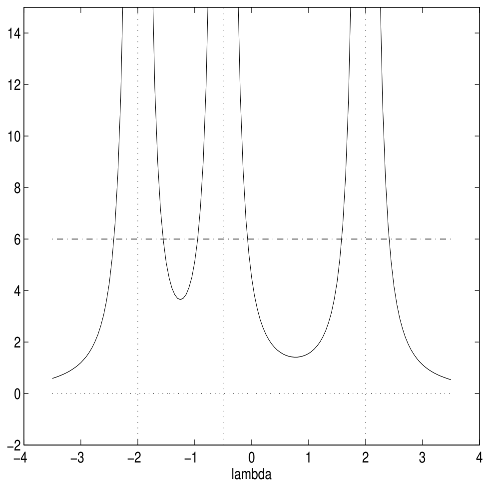

Figure 8: $\phi'(\lambda)$ in the standard case and straight line $\Delta^2 = 6$

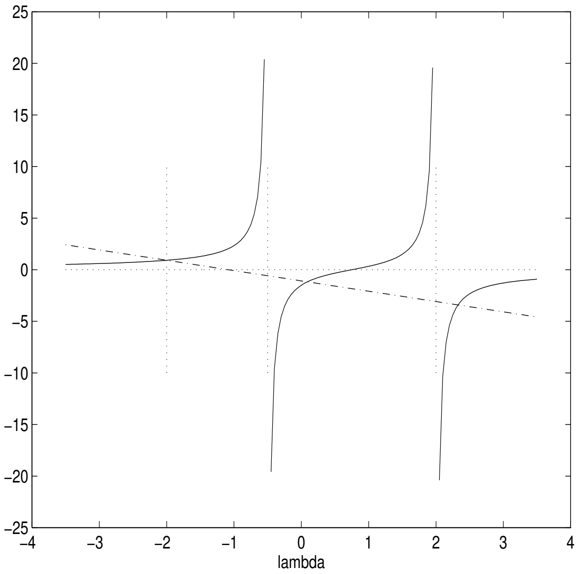

Figure 9: $\phi(\lambda)$ in the potential hard case and straight line $\alpha - \lambda$

Lemma 7.3 *For any $\alpha \in \mathcal{R}$ and $q \in \mathcal{S}_1$, $\{\delta_1, (0 \quad q^T)^T\}$ is an eigenpair of B_α if and only if g is orthogonal to $\mathcal{S}_1$.*

Proof: See [43] $\square$

A potential hard case situation is shown in figure 9. Note that in this case $g \perp \mathcal{S}_1$ and therefore, δ_1 is no longer a pole of $\phi(\lambda)$. As lemma 7.2 established, the way of computing the solutions in this case depends on the value of Δ.

We would like to point out that the assumption in the GvM method (see section 6) that $\Delta < \| A^\dagger \bar{b} \|_2$ rules out the possibility of the hard case, since for those values of Δ, we can always find $\lambda < 0$ such that the rest of the conditions of lemma 7.1 are satisfied.

The following lemma establishes that there exists a value of α for which it is always possible to find an eigenvector associated to the smallest eigenvalue of B_α that can be properly normalized when $g \perp \mathcal{S}_1$.

Lemma 7.4 *Suppose that g is orthogonal to $\mathcal{S}_1$ and let $p = -(H - \delta_1 I)^\dagger g$. If $\tilde{\alpha} = \delta_1 - g^T p$ then $\{\delta_1, (1 \quad p^T)^T\}$ is an eigenpair of $B_{\tilde{\alpha}}$. Moreover, $(1 \quad p^T)^T$ is orthogonal to $(0 \quad q^T)^T$, for every $q \in \mathcal{S}_1$.*

Proof: See [43] $\square$

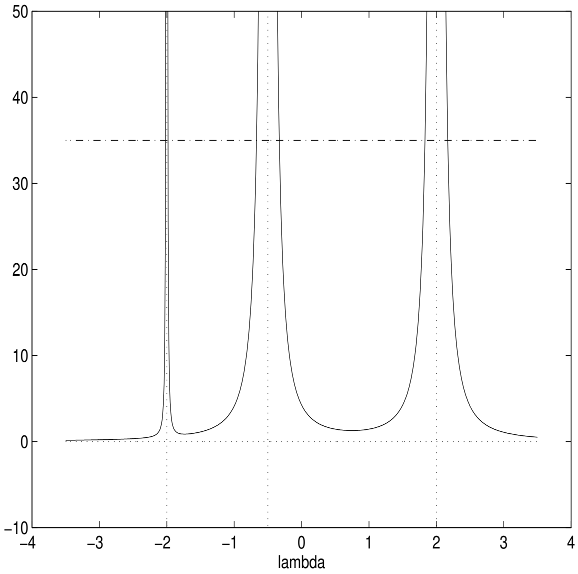

Figure 10: $\phi'(\lambda)$ in the near potential hard case and straight line $\Delta^2 = 35$

In practice, it may occur that the eigenvector of interest has a small first component which will produce large roundoff errors if the required normalization is carried out. This situation is called a near potential hard case and occurs when g is nearly orthogonal to $\mathcal{S}_1$.

As it can be observed in figure 10, computing $\hat{\lambda}$ such that $\hat{\phi}'(\hat{\lambda}) = \Delta^2$ can be a very ill–conditioned problem for large values of Δ in the near potential hard case situation. Depending on the shape of $\hat{\phi}'(\hat{\lambda})$, that may be the case even for small values of Δ (see §7.1.2.2).

The approach taken in [43] is to "ignore" the pole at δ_1 and use the *second* smallest eigenvalue of B_α as an interpolation point. In this way they try to approach the optimal value of α (the one that yields an optimal pair $\{\lambda_*, x_*\}$ for the TRS) from the right. It is shown in [43] that the use of the second smallest eigenpair is not needed eventually and that this strategy does not affect the q–superlinear convergence of the method.

Another important feature of the method in relation to convergence is the safeguarding of $\hat{\lambda}$ and α.

In the TRS method convergence is declared in any of the following cases:

- An interior solution is found.
- A boundary solution is found.
- The hard case is detected.
- A maximum number of iterations is reached.

With all these elements we are ready to present the complete method.

TRS method of [43]

Assume that initial guesses for $\lambda_{-1},\ \alpha$ are given.

1. **Initialization**
   - 1.1 Compute λ_0 from initial guess λ_{-1}
   - 1.2 Compute initial safeguards for $\lambda,\ \alpha$
   - 1.3 Safeguards $\lambda_0,\ \alpha$
   - 1.4 $k = 0$

2. **while** *(not convergence)* **do**
   - 2.1 Compute Interpolant $\hat{\phi}(\lambda)$ based on λ_{k-1}, λ_k
   - 2.2 Compute $\hat{\lambda}$ such that $\hat{\phi}'(\hat{\lambda}) = \Delta^2$
   - 2.3 Safeguard $\hat{\lambda}$ and update safeguards
   - 2.4 Adjust $\alpha = \hat{\lambda} + \hat{\phi}(\hat{\lambda})$
   - 2.5 Safeguard α and update safeguards
   - 2.6 Compute the *two* smallest eigenpairs of B_α
   - 2.7 **if** (*near hard case*) **then**

     Use *second smallest* eigenpair to update (λ_{k+1}, x_{k+1})

     else

     Use *smallest* eigenpair to update (λ_{k+1}, x_{k+1})

     end if
   - 2.8 $k = k + 1$

   end while

7.1.2 The TRS method on ill–conditioned problems

When the TRS method is applied to problem (12) and the matrix A is ill–conditioned, then $H = A^T A$ is nearly singular and it is easy to see that in this case, $g = -A^T\bar{b}$ is nearly orthogonal to $\mathcal{S}_1$.

Let $A = U\Sigma V^T$ be the SVD of A. Then

$$\begin{aligned} H &= V\Sigma^2 V^T \\ g &= -V\Sigma U^T\bar{b} \\ &= -V\begin{pmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_{n-1} & \\ & & & \sigma_n \end{pmatrix} U^T\bar{b} \end{aligned}$$

Problem	σ_1	$u_n^T b$	$\epsilon u_n^T r$	$g^T v_n$
heat	1.19×10^{-27}	4.59×10^{-17}	-0.0061	-7.38×10^{-30}
deriv2	3.33×10^{-5}	-3.33×10^{-7}	0.0051	-1.68×10^{-7}

Table 1: Nearly orthogonality of g with respect to $\mathcal{S}_1$

where $\sigma_i = \sqrt{\delta_{n-i+1}}$, since we order the eigenvalues and the singular values in opposite directions.

Suppose that $x \in \mathcal{S}_1$. Without loss of generality take $x = v_n$. Therefore,

$$g^T x \;=\; -(\sigma_1 u_1^T \bar{b}, \ldots, \sigma_{n-1} u_{n-1}^T \bar{b}, \sigma_n u_n^T \bar{b}) V^T x$$

and since $x \in \mathcal{S}_1$, $v_i^T x = 0, \; i = 1, \ldots, n-1$

$$g^T x \;=\; -\sigma_n u_n^T \bar{b} \tag{17}$$

In general, if σ_n is very small then g is nearly orthogonal to $\mathcal{S}_1$ and a near potential hard case situation occurs.

Expression (17) needs to be studied carefully in the ill–conditioned case. For noisy data $\bar{b} = b + \epsilon r$, (17) becomes

$$\begin{aligned} g^T x &= -\sigma_n u_n^T (b + \epsilon r) \\ &= -\sigma_n (u_n^T b + \epsilon u_n^T r) \end{aligned}$$

We distinguish two cases

1. When $\bar{b}$ contains no perturbations and if $u_n^T b$ is not too large (which is the case if the DPC holds), then g would be nearly orthogonal to $\mathcal{S}_1$.

2. In the presence of perturbations, if $u_n^T b$ is not too large, then $g^T x$ is dominated by $\sigma_n \epsilon u_n^T r$. In this case, g might not be nearly orthogonal to $\mathcal{S}_1$ since the coefficient $u_n^T r$ might be large enough to compensate for the small σ_n. In most severely ill–conditioned cases however, σ_n is so close to zero that even if $u_n^T r$ is large, g will still be nearly orthogonal to $\mathcal{S}_1$. An example of each case is shown in table 1.

As we saw in §7.1.1, a near potential hard case situation is handled by the TRS method through the second eigenpair of the bordered matrix. This approach may fail in the discrete ill–posed case due to the kind of eigenvalue distribution of H for these problems.

As we saw in section 2, H has a large cluster of small eigenvalues. Because of the interlacing property, the same clustered distribution can be expected for the eigenvalues of the bordered matrix, except perhaps for the smallest eigenvalue $\lambda_1(\alpha)$. Depending on the value of α (which depends on the value of $\hat{\lambda}$ computed from Δ and $\hat{\phi}'(\lambda)$), the smallest eigenvalue of B_α may or may not be separate from the rest of the spectrum.

For Δ large $\lambda_1(\alpha)$ will not be well separate from the rest of the spectrum. In the ill–conditioned least squares case $\lambda_1(\alpha)$ and $\lambda_2(\alpha)$ will actually be very close (and small) even for not too large values of Δ. In this case the corresponding eigenvectors will both have small first components and consequently, the normalization of the eigenvector associated to the second smallest eigenvalue will produce large errors. The situation is illustrated in §7.1.2.2.

7.1.2.1 An alternative for handling near potential hard case situations

In the near potential hard case situation we propose not to use the second smallest eigenpair if the second eigenvector also has a small first component. Instead of this, we suggest to use the following procedure to replace step 2.7 in the TRS method.

Assume that $\{\lambda_1(\alpha), (\nu_1 \;\; u_1^T)^T\}$ is the smallest eigenpair of B_α at iteration k of the TRS method.

> **if** (*near hard case*) **then**
>   **if** (*second smallest gives near hard case*) **then**
> $\delta = \frac{u_1^T H u_1}{u_1^T u_1}$
>   Solve $(H - \delta I)p = -g$
> $\alpha = \delta - g^T p$
>   **else**
>   Use *second smallest* eigenpair to update (λ_{k+1}, x_{k+1})
>   **end if**
> **else**
>   Use *smallest* eigenpair to update (λ_{k+1}, x_{k+1})
> **end if**

In other words, when the information on the second smallest eigenpair is not good enough, we try to compute an estimate for the special value $\tilde{\alpha}$, given by lemma 7.4, that guarantees that we can find an eigenvector with the desired structure. A good estimate for δ_1 is available in a near potential hard case situation, since in this case $\{\lambda_1(\alpha), u_1\}$ is a very good estimate for the smallest eigenpair of H and therefore $\delta = \frac{u_1^T H u_1}{u_1^T u_1}$ will be very close to δ_1.

To solve $(H-\delta I)p = -g$ we propose to compute l of the *largest* eigenvalues of H by means of the IRLM and take the solution in the corresponding Krylov subspace as an approximation to p.

The rationale behind this idea is that in the discrete ill–posed case, H usually has few significant eigenvalues and they are not clustered. Therefore, the solution in the Krylov subspace associated to the large eigenvalues will hopefully be a reasonably good approximation for p. Moreover, we expect that the IRLM will compute those eigenvalues efficiently since there is a small number of them and they are separate.

This idea is yet to be tried.

7.1.2.2 Examples

All the tests refer to problem *heat* from the Regularization Tools package [25]. The problem is an inverse heat equation and is very ill–conditioned. The dimensions are $m = n = 50$.

The right hand side was taken as $\bar{b} = b + \epsilon r$, where b is the exact data vector returned by routine **heat** from [25].

The perturbation ϵ was either 0 or 0.001, r is a random vector with components uniformly distributed in (0,1).

The experiments were carried out in MATLAB on a SUN SPARC station IPX, with IEEE standard double precision arithmetic. Machine precision is of order 10^{-16}.

For the IRLM we allowed a maximum of nine Lanczos vectors and started the method every time with the same random initial vector.

For the TRS method the following stopping criteria were used

- Interior Solution

  For a given tolerance ϵ_{INT}, an interior solution is declared if

$$(\| x \|_2 < \Delta) \;\; \& \;\; (\lambda > -\epsilon_{INT})$$

- Boundary Solution

  For a given tolerance ϵ_Δ, a boundary solution is declared if

$$\| x - \Delta \|_2 < \epsilon_\Delta * \Delta$$

- Maximum number of iterations. This limit was set to 50.

- Hard case. The hard case is verified by the criterion in lemma 5 from [43]. A tolerance ϵ_{HC} is needed for this test.
- KKT and no progress in α. KKT refers to the norm of the residual of the system in the Karush–Kuhn–Tucker (first order necessary conditions) for a minimizer of problem (13). When this condition is satisfied and α has the same value for many iterations, the current point x is returned as the solution.

  The criterion for the KKT condition is

$$\| g + (H - \lambda I)x \|_2 \leq \| g \|_2 \, \epsilon_{KKT};$$

In the following tests, the tolerances were chosen as

Tolerance	Value
ϵ_Δ	10^{-3}
ϵ_{INT}	10^{-4}
ϵ_{HC}	10^{-4}
ϵ_{KKT}	10^{-3}

The results of the experiments are summarized in table 2. The norm of the exact solution $\| x_* \|_2$ is 1.7308.

Test #	**Type of Data**	Δ	$\frac{\|\|x-x_*\|\|}{\|\|x_*\|\|}$	**Exit by**
1	exact	1.0	0.6460	Boundary Solution
2	exact	1.5	0.1819	KKT, no progress in α
3	exact	3	0.1752	Interior Solution
4	perturbed	1	0.6478	Boundary Solution
5	perturbed	1.5	0.2905	KKT, no progress in α
6	perturbed	3	0.2727	KKT, no progress in α

Table 2: Experiments

In test number 2 a value of $\alpha = 0.1842$ yielded $|\lambda_1(\alpha) - \lambda_2(\alpha)| \approx 10^{-17}$ which shows that the two smallest eigenvalues of B_α may be clustered. In this example, $\lambda_1(\alpha) \approx 10^{-27}$. The corresponding eigenvectors had very small first components.

A near potential hard case situation is declared if ν, the first component of the eigenvector associated to the smallest eigenvalue, satisfies

$$\frac{\| g \|_2 \, |\nu|}{\sqrt{1-\nu^2}} < \epsilon_\nu$$

In test numbers 2, 3, 5 and 6, the near potential hard case situation arises. Every time this situation was present, the first component of the eigenvector associated to the *second* smallest eigenvalue, also satisfied the previous criterion.

In figure 11 we show the exact and computed solutions for each test.

In figure 12 we show the function $\phi'(\lambda)$ for problem *heat* and perturbed data. The function is so steep close to δ_1, that even for small values of Δ, computing $\hat{\lambda}$ may be a very ill–conditioned problem.

7.1.2.3 Difficulties

The difficulties encountered when using the TRS method on discrete ill–posed problems are

1. In a near potential hard case situation, the second smallest eigenpair may fail to update α properly. A good strategy for handling the near hard case situation is needed, since as we saw in §7.1.2.2, this situation may arise even when a good estimate for Δ is available.

2. With respect to the solution of the eigenvalue problems, it is known that the Lanczos method might perform poorly when applied to matrices with clustered eigenvalues (see [16, Ch. 9]). We saw before that this kind of matrices is likely to arise in the ill–conditioned case.

7.1.2.4 Addressing the difficulties

We propose to investigate the following options to try to overcome the difficulties in §7.1.2.3.

1. To handle near potential hard case situations
   - Use the strategy described in 7.1.2.1.
   - Use special deflation techniques.

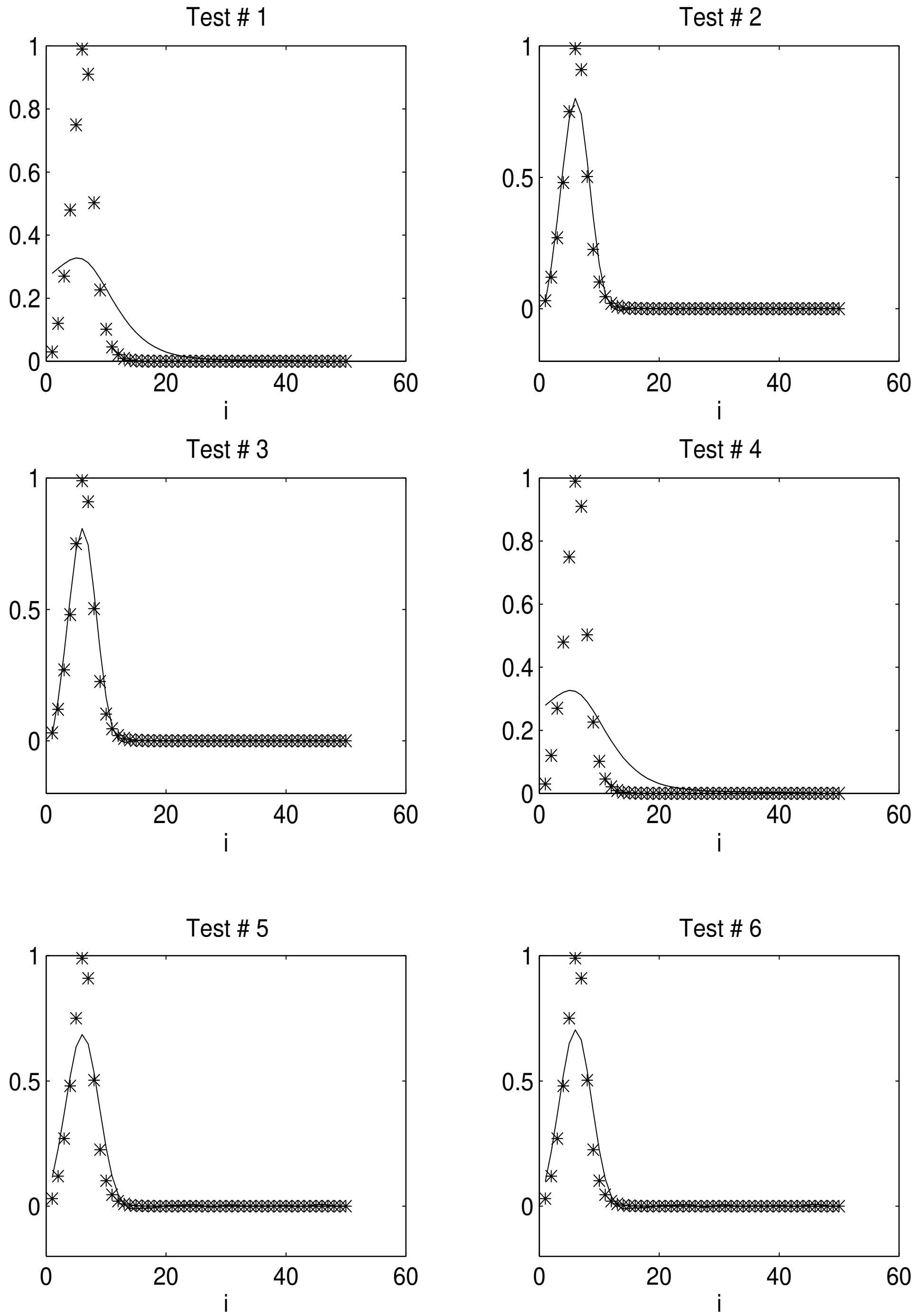

Figure 11: Exact solution (*) and approximate solution (-) for tests

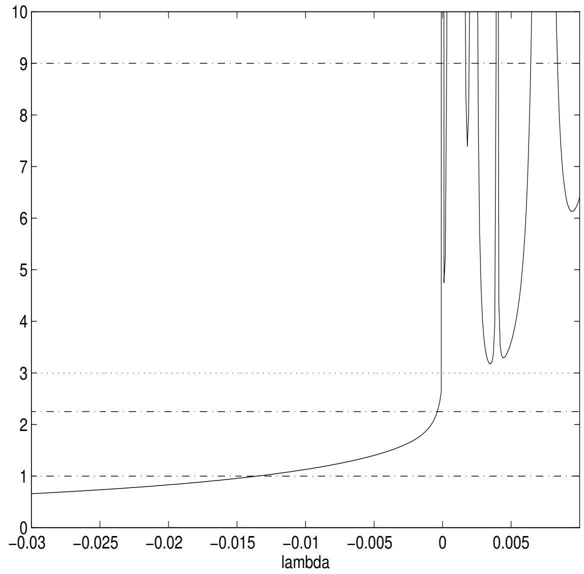

Figure 12: $\phi'(\lambda)$ for problem *heat* and (dash–dotted) straight lines at 1, $(1.5)^2$ and 9, corresponding to $\Delta = 1, 1.5$ and 3. The dotted line corresponds to $\| x_* \|_2 \approx 3$

2. To solve the eigenvalue problems more efficiently

   - Look at preconditioning strategies to accelerate the convergence of the IRLM when computing the smallest eigenpair of a matrix.
   - Use harmonic Ritz values instead of Ritz values as approximations to the eigenvalues. The harmonic Ritz values converge faster than the Ritz values to the smallest eigenvalues of a matrix.
   - Develop a hybrid eigenvalue method using the method proposed in [37] (Olsen's method) to compute the smallest eigenpair of the bordered matrix and the IRLM to compute an initial guess for the eigenvector. We expect that the use of Olsen's method will be a significant improvement since the method uses very little storage (two vectors) and inexpensive computations.
   - Use another eigenvalue method like the Jacobi–Davidson method.

3. To overcome the difficulties associated to the computation of the smallest eigenvalue from a cluster, it would be interesting to investigate the possibility of recasting the TRS in terms of the *largest* eigenvalues of H.

7.1.3 Advantages and Drawbacks

The use of the TRS method from [43] to compute a regularized solution for large–scale ill–conditioned problems has the following advantages and disadvantages.

Advantages

- Relies on matrix–vector products with either A, A^T or A^TA. This is an important efficiency factor in our real application (generated with the package DSO3.2 ([42])) in which applying A^TA directly is less expensive than applying A and A^T separately.
- The storage requirement is only V_k for k small and fixed.

Drawbacks

- The difficulties pointed out in §7.1.2.3.
- Computes the regularized solution only.

7.2 Computing the Regularization Parameter

This issue still needs to be addressed. The use of the L–curve criterion seems to be a good option for computing the regularization parameter, as long as the DPC holds. It was showed in [55] that if that condition is not satisfied the L–curve criterion may fail to compute a good estimate for the regularization parameter.

8 Final Remarks

In this work we have shown that large–scale ill–conditioned least squares problems arise in important applications and that there is a lack of robust numerical methods to treat these problems.

We have proposed an approach to the regularization of ill–conditioned problems that allows us to use well known optimization results and new optimization techniques for the treatment of such problems. The application of those particular results to regularization problems has been attempted very few times in the past and never for the large–scale case.

At this stage of the project, we have accomplished the following:

- Cover the background on regularization of ill–conditioned problems.
- Propose a strategy for computing the regularized solution (use of the TRS method).
- Identify difficulties associated with the strategy and propose options for solving them.

In order to achieve our final goal, the following tasks need to be completed:

- Implement proposed improvements for the TRS method.
- Develop a strategy for computing the regularization parameter.
- Implement the complete method.
- Perform numerical experiments on test problems.
- Compare with other approaches.
- Apply the method to a real problem from seismology.
- Study the theoretical properties of the method.

We are currently addressing the first two tasks.

References

[1] G. Beylkin. Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform. *J. Math. Phys.*, 26:99–108, 1985.

[2] A. Björck. *Numerical Methods for Least Squares Problems.* SIAM, Philadelphia, 1996.

[3] A. Björck and L. Eldén. Methods in numerical algebra for ill–posed problems. Technical Report LiTH–MAT–R–33–1979, Department of Mathematics, Linkoping University, 1979.

[4] A. Björck, E. Grimme, and Paul van Dooren. An implicit shift bidiagonalization algorithm for ill–posed systems. *BIT*, 34:510–534, 1994.

[5] D. Calvetti, L. Reichel, and D.C. Sorensen. An implicitly restarted Lanczos method for large symmetric eigenvalue problems. *ETNA*, 2:1–21, 1994.

[6] T.F. Chan, J.A. Olkin, and D.W. Cooley. Solving quadratically constrained least squares using black box solvers. *BIT*, 32:481–495, 1992.

[7] N. Clinthorne, T. Pan, P. Chiao, W. Rogers, and J. Stamos. Preconditioning methods for improved convergence rates in iterative reconstruction. *IEEE Trans. Med. Imag.*, 12(1):78–83, 1993.

[8] J.E. Dennis and H.H. Mei. Two new unconstrained optimization algorithms which use function and gradient values. *J. Opt. Theory Appl.*, 28:453–482, 1979.

[9] L. Eldén. Algorithms for the regularization of ill–conditioned least squares problems. *BIT*, 17:134–145, 1977.

[10] H. W. Engl. Regularization methods for the stable solution of inverse problems. *Surveys Math. Ind.*, 3:71–143, 1993.

[11] J. Eriksson. *Optimization and regularization of nonlinear least squares problems.* Doctoral thesis, Umeå University, Umeå, Sweden, 1996.

[12] W. Gander. Least squares with a quadratic constraint. *Numer. Math.*, 36:291–307, 1981.

[13] D. Gay. Computing optimal locally constrained steps. *SIAM J. Sci. Stat. Comp.*, 2(2):186–197, 1981.

[14] D. A. Girard. A fast 'Monte–Carlo Cross–Validation' procedure for large least squares problems with noisy data. *Numer. Math.*, 56:1–23, 1989.

[15] G. Golub, M. Heath, and G. Wahba. Generalized cross–validation as a method for choosing a good Ridge parameter. *Technometrics*, 21(2):215–223, 1979.

[16] G. Golub and Ch. Van Loan. *Matrix Computations.* The John Hopkins University Press, Baltimore, 2nd edition, 1989.

[17] G. H. Golub and U. von Matt. Quadratically constrained least squares and quadratic problems. *Numer. Math.*, 59:561–580, 1991.

[18] J. Hadamard. *Lectures on Cauchy's problem in linear partial differential equations.* Yale University Press, New Haven, 1923.

[19] M. Hanke. Accelerated Landweber iterations for the solution of ill–posed equations. *Numer. Math.*, 60:341–373, 1991.

[20] M. Hanke and P. C. Hansen. Regularization methods for large–scale problems. *Surveys Math. Ind.*, 3:253–315, 1994.

[21] M. Hanke, J. Nagy, and R. Plemmons. Preconditioned iterative regularization for ill–posed problems. In L. Reichel, A. Ruttan, and R. Varga, editors, *Numerical Linear Algebra*, Berlin, 1992. Walter de Gruyter. Proceedings of the Conference in Numerical Linear Algebra and Scientific Computation, Kent (Ohio), USA.

[22] P. C. Hansen. Analysis of discrete ill–posed problems by means of the L–curve. *SIAM Review*, 34(4):561–580, 1992.

[23] P.C. Hansen. Computation of the Singular Value Expansion. *Computing*, 40:185–199, 1988.

[24] P.C. Hansen. The Discrete Picard Condition for discrete ill–posed problems. *BIT*, 30:658–672, 1990.

[25] P.C. Hansen. Regularization Tools: a MATLAB package for analysis and solution of discrete ill–posed problems. *Numer. Algo.*, 6:1–35, 1994. Available from **ftp.uni-c.dk** in the directory **uni-c/unipch/matlab/ReguTools**.

[26] P.C. Hansen. *Rank–Deficient and Discrete Ill–Posed Problems.* Doctoral thesis, UNI•C, Technical University of Denmark, Lyngby, Denmark, 1995.

[27] P.C. Hansen and D.P. O'Leary. The use of the L–curve in the regularization of discrete ill–posed problems. *SIAM J. Sci. Comp.*, 14(6):1487–1503, 1993.

[28] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Res. Nat. Bur. Standards.*, B49:409–436, 1952.

[29] L. Kaufman and A. Neumaier. Image reconstruction through regularization by envelope guided Conjugate Gradients. Technical Report 4–14, AT&T Bell Laboratories, 1994.

[30] J. T. King. Multilevel algorithms for ill–posed problems. *Numer. Math.*, 61:311–334, 1992.

[31] C. L. Lawson and R. J. Hanson. *Solving Least Squares Problems.* Classics in Applied Mathematics. SIAM, Philadelphia, 1995.

[32] J. M. Martínez and S. A. Santos. A trust region strategy for minimization on arbitrary domains. To appear in *Mathematical Programming*, 1995.

[33] J. J. Moré. Recent developments in algorithms and software for trust region methods. In A. Bachem, M. Grotschel, and B. Korte, editors, *Mathematical Programming Bonn 1982 – The state of the art*, pages 258–287. Mathematical Programming Society, Bonn, 1983.

[34] J. J. Moré and D. C. Sorensen. Computing a trust region step. *SIAM J. Sci. Stat. Comp.*, 4(3):553–572, 1983.

[35] V. A. Morozov. On the solution of functional equations by the method of regularization. *Sov. Math. Doklady*, 167(3):414–417, 1966.

[36] J. Nagy, V. Pauca, R. Plemmons, and T. Torgersen. Space–varying restoration of optical images. Technical report, Dept. of Mathematics and Computer Science, Wake Forest University, Winsdon–Salen, North Carolina, 1996.

[37] J. Olsen, P. Jorgensen, and J. Simons. Passing the one–billion limit in full configuration interaction (fci) calculations. *Chem. Phys. Letters*, 169(6):463–472, 1990.

[38] C. C. Paige and M. A. Saunders. LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Softw.*, 8(1):43–71, 1982.

[39] B.N. Parlett. *The Symmetric Eigenvalue Problem.* Prentice–Hall, Englewood Cliffs, NJ, 1980.

[40] M.J.D. Powell. A new algorithm for unconstrained optimization. In J.B. Rosen, O.L. Mangasarian, and K. Ritter, editors, *Nonlinear Programming.* Academic Press, 1970.

[41] F. Rendl and H. Wolkowicz. A semidefinite framework for trust region subproblems with applications to large scale minimization. Technical Report CORR Report 94–32, Department of Combinatorics and Optimization, University of Waterloo, December 1994.

[42] The Rice Inversion Project, Department of Computational and Applied Mathematics, Rice University. Houston, Texas. *DSO User manual and reference. Version 3.2*, January 1995.

[43] S. A. Santos and D. C. Sorensen. A new matrix–free algorithm for the large–scale trust–region subproblem. Technical Report 95–20, Department of Computational and Applied Mathematics, Rice University, July 1995.

[44] F. Santosa and W. W. Symes. *An Analysis of Least–Squares velocity inversion.* Geophysical monograph series, number 4. Society of exploration geophysicists, Tulsa, 1989.

[45] D. C. Sorensen. Newton's method with a model trust region modification. *SIAM J. Numer. Anal.*, 19(2):409–426, 1982.

[46] D. C. Sorensen. Implicit application of polynomial filters in a k–step Arnoldi method. *SIAM J. Mat. Anal. Appl.*, 13:357–385, 1992.

[47] D. C. Sorensen. Minimization of a large scale quadratic function subject to an ellipsoidal constraint. Technical Report 94–27, Department of Computational and Applied Mathematics, Rice University, August 1994.

[48] T. Steihaug. The conjugate gradient method and trust regions in large scale optimization. *SIAM J. Numer. Anal.*, 20(3):262–637, 1983.

[49] G. W. Stewart. *Introduction to Matrix Computations*. Academic Press, New York, 1973.

[50] A. Tarantola. *Inverse Problem Theory*. Elsevier, Amsterdam, 1987.

[51] A. N. Tikhonov and V. Y. Arsenin. *Solutions of ill–posed problems*. John Wiley & Sons, New York, 1977.

[52] S. Van Huffel and J. Vandewalle. *The Total Least Squares Problem: Computational Aspects and Analysis*, volume 9 of *Frontiers in Applied Mathematics*. SIAM, Philadelphia, 1991.

[53] J. Varah. A practical examination of some numerical methods for linear discrete ill–posed problems. *SIAM Review*, 21:100–111, 1979.

[54] C.R. Vogel. A constrained least squares regularization method for non-linear ill–posed problems. *SIAM J. Control Optim.*, 28(1):34–49, 1990.

[55] C.R. Vogel. Non–convergence of the L–curve regularization parameter selection method. *Inverse Problems*, 1996. to appear.

[56] G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.