Journal of Computational and Applied Mathematics 256 (2014) 180–195

Contents lists available at ScienceDirect

Journal of Computational and Applied Mathematics

journal homepage: www.elsevier.com/locate/cam

Automatic parameter setting for Arnoldi–Tikhonov methods

Silvia Gazzola *, Paolo Novati

Department of Mathematics, University of Padova, Italy

ARTICLE INFO

Article history:
Received 19 March 2012
Received in revised form 28 September 2012

Keywords:
Tikhonov regularization
Arnoldi algorithm
Discrepancy principle
Image restoration

ABSTRACT

In the framework of iterative regularization techniques for large-scale linear ill-posed problems, this paper introduces a novel algorithm for the choice of the regularization parameter when performing the Arnoldi–Tikhonov method. Assuming that we can apply the discrepancy principle, this new strategy can work without restrictions on the choice of the regularization matrix. Moreover, this method is also employed as a procedure to detect the noise level whenever it is just overestimated. Numerical experiments arising from the discretization of integral equations and image restoration are presented.

1. Introduction

In this paper we consider the solution of ill-conditioned linear systems of equations

$$Ax = b, \quad A \in \mathbb{R}^{N\times N},\ b \in \mathbb{R}^{N}, \tag{1}$$

in which the matrix A is assumed to have singular values that rapidly decay and cluster near zero. These kind of systems typically arise from the discretization of linear ill-posed problems, such as Fredholm integral equations of the first kind with a compact kernel; for this reason they are commonly referred to as linear discrete ill-posed problems (see [1], Chapter 1, for a background).

While working with this class of problems, one commonly assumes that the available right-hand side vector b is affected by noise, caused by measurement or discretization errors. Therefore, throughout the paper we suppose that

$$b = \overline{b} + e,$$

where $\overline{b}$ represents the unknown noise-free right-hand side, and we denote by $\overline{x}$ the solution of the error-free system $Ax = \overline{b}$. We also assume that a fairly accurate estimate of $\varepsilon = \|e\|$ is known, where $\|\cdot\|$ denotes the Euclidean norm.

Because of the ill-conditioning of A and the presence of noise in b, some sort of regularization is generally employed to find a meaningful approximation of $\overline{x}$. In this framework, a popular and well-established regularization technique is Tikhonov method, which consists in solving the minimization problem

$$\min_{x\in\mathbb{R}^n} \left\{ \|Ax - b\|^2 + \lambda \|Lx\|^2 \right\}, \tag{2}$$

where $\lambda > 0$ is the regularization parameter and $L \in \mathbb{R}^{(N-p)\times N}$ is the regularization matrix (see e.g. [2] and [1], Chapter 5, for a background). We denote the solution of (2) by x_λ. Common choices for L are the identity matrix I_N (in this case (2) is said to be in standard form) or scaled finite differences approximations of the first or the second order derivative (when $L \neq I_N$ (2) is said to be in general form). We remark that, especially when one has a good intuition of the behavior of the solution $\overline{x}$,

* Corresponding author. Tel.: +39 3403293160.
E-mail address: gazzola@math.unipd.it (S. Gazzola).

0377-0427/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.cam.2013.07.023

a regularization matrix different from the identity can considerably improve the quality of the approximation given by the solution of (2). The ideal situation is when the features of the exact solution that one wants to preserve belong to the null space of the matrix L, since L acts as a penalizing filter (see [3] and the references therein for a deeper discussion).

The choice of λ is also crucial, since it weights the penalizing term and so specifies the amount of regularization one wants to impose. Many techniques have been developed to determine a suitable value for the regularizing parameter, usually based on the amount of knowledge of the error on b (again we refer to [1], Chapter 7, for an exhaustive background; we also quote the recent paper [4] for the state of the art). When a fairly accurate approximation of ε is available (as in our case), a widely used method is the so-called discrepancy principle. It prescribes to take, as regularization parameter, the value of λ that solves the following equation

$$\|b - Ax_\lambda\| = \eta\varepsilon, \tag{3}$$

where $\eta > 1$ is a user-specified constant, typically very close to 1. The vector $b - Ax_\lambda$ is called discrepancy.

In this paper we solve (2) using an iterative scheme called Arnoldi–Tikhonov (AT) method, first proposed in [5]. This method has proved to be particularly efficient when dealing with large scale problems, as for instance the ones arising from image restoration. Indeed, it is based on the projection of the original problem (2) onto Krylov subspaces of smaller dimensions computed by the Arnoldi algorithm. However, for reasons closely related to the parameter choice strategy, this method has been experimented mostly when (2) is in standard form [6]; only recently an extension which employs generalized Krylov subspaces and that therefore can deal with general form problems has been introduced in [3].

Here we mainly focus our attention on general form problems, but we adopt a different approach from the one derived in [3], since we work with the usual Krylov subspaces $\mathcal{K}_m(A, b) = \text{span}\{b, Ab, \dots, A^{m-1}b\}$ (or, if an approximate solution x_0 is available, with $\mathcal{K}_m(A, b - Ax_0)$). We call this method Generalized Arnoldi–Tikhonov (GAT) to avoid confusion with the standard implementation of the AT method. The parameter choice strategy presented in this paper is extremely simple and does not require the problem (2) to be in standard form. Moreover, this new algorithm can handle rectangular matrices L, which is an evident advantage since in many applications this option is the most natural one. Our basic idea is to use a linear approximation of the discrepancy

$$\|b - Ax_m\| \approx \alpha_m + \lambda\beta_m,$$

where x_m is the mth approximation of the GAT method, and to solve with respect to λ the corresponding equation

$$\alpha_m + \lambda\beta_m = \eta\varepsilon.$$

As we shall see, the value of α_m in the above equation will be just the GMRES residual, whereas β_m will be defined using the discrepancy of the previous step. In this way, starting from an initial guess λ_0, we will actually construct a sequence of parameters $\lambda_1, \lambda_2, \dots$, such that λ_{m-1} will be used to compute x_m until the discrepancy principle (3) is satisfied. We will be able to demonstrate that the above technique is in fact a secant zero finder.

As we shall see, the procedure is extremely simple and does not require any hypothesis on the regularization matrix L. For this reason, in the paper we also consider the possibility of using the GAT method to approximate the noise level ε whenever it is just overestimated by a quantity $\overline{\varepsilon} > \varepsilon$. In a situation like this the discrepancy principle generally yields poor results if the approximation of ε is coarse. Anyway, our idea consists in restarting the GAT method, and to use the observed discrepancy to improve the approximation of ε step by step. The examples so far considered have demonstrated that this approach is really effective, and the additional expense due to the restarts of the GAT method does not heavily affect the total amount of work. This is due to the fact that the GAT method is extremely fast whenever an initial approximation x_0 is available.

The paper is organized as follows. In Section 2 we review the AT method and we describe its generalized version, the GAT method. In Section 3 we introduce the new technique for the choice of λ. In Section 4 we display the results obtained performing common test problems, as well as some examples of image restoration. In Section 5 we suggest an extension of the previous method that allows us to work even when the quantity ε is overestimated. Finally, in Section 6, we propose some concluding remarks.

2. The Arnoldi–Tikhonov method

The Arnoldi–Tikhonov (AT) method has been introduced in [5] with the basic aim of reducing the problem

$$\min_{x \in \mathbb{R}^N} \left\{ \|Ax - b\|^2 + \lambda\|Lx\|^2 \right\}, \tag{4}$$

in the case of $L = I_N$, to a problem of much smaller dimension. The idea is to project the matrix A onto the Krylov subspaces generated by A and the vector b, i.e., $\mathcal{K}_m(A, b) = \text{span}\{b, Ab, \dots, A^{m-1}b\}$, with $m \ll N$. The method was even introduced to avoid the matrix–vector multiplication with A^T required by Lanczos type schemes (see e.g [7,5,8,9]). For the construction of the Krylov subspaces the AT method uses the Arnoldi algorithm (see [10], Section 6.3, for an exhaustive background), which yields the decomposition

$$AV_m = V_{m+1}H_{m+1}, \tag{5}$$