Applied Mathematics and Computation 243 (2014) 258–268

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

# Generalized Cross-Validation applied to Conjugate Gradient for discrete ill-posed problems

P. Favati [a], G. Lotti [b], O. Menchi [c,*], F. Romani [c]

[a] IIT – CNR, Via G. Moruzzi 1, 56124 Pisa, Italy
[b] Dip. di Matematica, University of Parma, Parco Area delle Scienze 53/A, 43124 Parma, Italy
[c] Dip. di Informatica, University of Pisa, Largo Pontecorvo 3, 56127 Pisa, Italy

ARTICLE INFO

*Keywords:*
Regularization problems
Conjugate Gradient
Generalized Cross-Validation

ABSTRACT

In this paper we propose a new method to apply the Generalized Cross-Validation (GCV) as a stopping rule for the Conjugate Gradient (CG). In general, to apply GCV to an iterative method, one must estimate the trace of the so-called influence matrix which appears in the denominator of the GCV function. In the case of CG, unlike what happens with stationary iterative methods, the regularized solution has a nonlinear dependence on the noise which affects the data of the problem. This fact is often pointed out as a cause of poor performance of GCV. To overcome this drawback, our proposal linearizes the dependence by computing the derivatives through iterative formulas. We compare the proposed method with other methods suggested in the literature by an extensive numerical experimentation on both 1D and 2D test problems.



## 1. Introduction

Given a matrix $A \in \mathbf{R}^{n\times n}$ and a vector $\mathbf{b} \in \mathbf{R}^n$, we consider the system

$$A\mathbf{x} = \mathbf{b}. \quad (1)$$

We assume that $A$ is a large full rank matrix, having singular values which gradually decay to zero, so that it is difficult to determine its numerical rank. In many applications the available right-hand side of the system is contaminated by a noise $\boldsymbol{\eta}$ accounting for both the measurement errors and the process involved in the construction of the discrete model describing the underlying continuous phenomenon, i.e.

$$\mathbf{b} = \mathbf{b}^* + \boldsymbol{\eta}.$$

The vectors $\mathbf{b}^*$ and $\mathbf{x}^*$, such that $A\mathbf{x}^* = \mathbf{b}^*$, are considered the exact right-hand side and the exact solution of the system. Classical examples of this kind of problems arise from the discretization of Fredholm integral equations of the first kind, as for instance in the imaging deconvolution, where $A$ represents an imaging system, $\mathbf{x}^*$ an object, $\mathbf{b}^*$ the noise-free image of the object and $\mathbf{b}$ the noisy image.

Due to the ill-conditioning of the matrix and the presence of the noise, the solution $\widetilde{x} = A^{-1}\mathbf{b}$ is often a poor approximation of $\mathbf{x}^*$ even if the magnitude of $\boldsymbol{\eta}$ is small, and the problem of finding a good approximation of $\mathbf{x}^*$ turns out to be a discrete

* Corresponding author.
*E-mail addresses:* paola.favati@iit.cnr.it (P. Favati), grazia.lotti@unipr.it (G. Lotti), menchi@di.unipi.it (O. Menchi), romani@di.unipi.it (F. Romani).

http://dx.doi.org/10.1016/j.amc.2014.05.109
0096-3003/

ill-posed problem [7]. Special techniques called *regularization* methods are required to deal with this kind of problems. Both direct methods (as Tikhonov method) and iterative methods can be used to this aim. Iterative methods are suggested for large matrices $A$ without particular structure properties. The iterative method has to enjoy the semi-convergence property, i.e. in presence of the noise it reconstructs first the low-frequency components, which correspond to the largest singular values of $A$. The iteration should be stopped before the high-frequency components of the noise start to enter the computed solution. In this sense the iteration number plays the role of the regularization parameter.

Among the classical semi-convergent methods we consider here the *Conjugate Gradient* method (CG). The regularizing properties of CG are well known (see for example [13]). CG has in general a good convergence rate and finds quickly an optimal vector $\mathbf{x}_{opt}$ which minimizes the error with respect to $\mathbf{x}^*$. This behavior can be a disadvantage in the regularization context, because also the high-frequency components enter quickly the computed solution and the error increases sharply after the optimal number $k_{opt}$ of steps. As a matter of fact, the determination of $k_{opt}$ is very sensitive to the perturbation of the right-hand side [3]. As a consequence, the regularizing efficiency of CG depends heavily on the effectiveness of the stopping rule employed. Three widely used stopping techniques are the *Discrepancy Principle*, which is based on the idea that the residual norm should be related to an a priori knowledge of the noise level, the *Unbiased Predictive Risk Estimator* (UPRE) and the *Generalized Cross-Validation* rule (GCV) (see [7,15,16]). The last two methods are based on predictive error estimates. In this paper we focus our attention on GCV, which has the advantage over the other two techniques of not requiring information on the noise level.

The stopping index is estimated through the minimum of the GCV function, whose denominator requires the computation of the trace of the CG influence matrix. GCV has been shown to be very effective when applied to iterative methods whose influence matrix does not depend on the noise, i.e. when the regularized solution depends linearly on the right-hand side of the system. However, this is not the case of CG, and some techniques have been proposed to overcome this drawback [4,6,7,14]. In order to approximate the denominator of the GCV function, we propose a new method which linearizes the dependence of the regularized solution on the noise. The novelty of our method consists in approximating the required derivatives by means of iterative formulas instead of using finite differences as suggested in [14]. Iterative formulas of this kind have been introduced in [1,12] for other regularization methods. The extensive numerical experimentation carried out on both 1D and 2D test problems shows the effectiveness of the method here described with respect to other methods proposed in the literature.

The outline of the paper is the following: preliminary definitions and the GCV function are given in Section 2. In Section 3 the CG code is recalled in order to derive the expressions used for computing the trace of the influence matrix. Unless the matrix $A$ has some special structure, the direct application of these expressions is impracticable for large dimensions, so a stochastic implementation based on the trace lemma is given. The special case of a circulant matrix $A$ is examined in Section 4. The numerical experimentation, described in Section 5, shows that the different approximations of the denominator of the GCV function are in general not very critical in detecting an acceptable stopping index. Anyway, a reasonable ranking of them can be obtained for the examples we consider.

Throughout the paper, $\boldsymbol{\eta}$ is assumed to be an uncorrelated Gaussian white noise, i.e. with distribution $\mathcal{N}(0,\sigma^2 I)$, and $\|\mathbf{v}\|$ denotes the Euclidean norm of a vector $\mathbf{v}$.

## 2. The regularized solution

Let $A = U\Sigma V^T$ be the singular value decomposition of $A$, where $U = [\mathbf{u}_1,\ldots,\mathbf{u}_n] \in \mathbf{R}^{n\times n}$ and $V = [\mathbf{v}_1,\ldots,\mathbf{v}_n] \in \mathbf{R}^{n\times n}$ have orthonormal columns, i.e. $U^TU = V^TV = I$, and $\Sigma = \mathrm{diag}(\sigma_1,\ldots,\sigma_n)$, where the $\sigma_i$ for $i = 1,\ldots,n$ are the singular values of $A$, gradually decaying toward zero. In practice, the last ones settle to values of the same magnitude of the machine precision.

The expansions of $\mathbf{b}^*$ and $\boldsymbol{\eta}$ in the basis $U$ are

$$\mathbf{b}^* = \sum_i b_i^* \mathbf{u}_i, \quad \boldsymbol{\eta} = \sum_i \eta_i \mathbf{u}_i, \quad \text{where} \quad b_i^* = \mathbf{u}_i^T\mathbf{b}^*, \quad \eta_i = \mathbf{u}_i^T\boldsymbol{\eta}.$$

Then

$$\mathbf{x}^* = A^{-1}\mathbf{b}^* = \sum_i x_i^* \mathbf{v}_i, \quad \text{where} \quad x_i^* = \frac{b_i^*}{\sigma_i}, \tag{2}$$

and

$$\widetilde{\mathbf{x}} = \mathbf{x}^* + A^{-1}\boldsymbol{\eta} = \mathbf{x}^* + \sum_i \frac{\eta_i}{\sigma_i}\mathbf{v}_i = \sum_i \widetilde{x}_i \mathbf{v}_i, \quad \text{where} \quad \widetilde{x}_i = x_i^* + \frac{\eta_i}{\sigma_i}. \tag{3}$$

The coefficients $\eta_i$ are typically of the same order for all $i$, with $|\eta_i| \sim \|\boldsymbol{\eta}\|/n$. If the last $\sigma_i$'s are much smaller than the corresponding $|\eta_i|$, the quantities $\eta_i/\sigma_i$ greatly increase with $i$. It follows that the low-frequency components of $\widetilde{\mathbf{x}}$ and $\mathbf{x}^*$ do not differ much, while the high-frequency components of $\widetilde{\mathbf{x}}$ are disastrously dominated by the high-frequency components of the noise and $\widetilde{\mathbf{x}}$ can be affected by a large error with respect to $\mathbf{x}^*$. The contribution of the high-frequency components of the noise should be damped in the regularized solution. Acceptable approximations can be obtained only if the $|b_i^*|$ decay to