A general framework for nonlinear regularized Krylov-based image restoration

Serena Morigi[1], Lothar Reichel[2] and Fiorella Sgallari[1]

[1] Department of Mathematics, University of Bologna, Bologna, Italy
`{serena.morigi,fiorella.sgallari}@unibo.it`
[2] Department of Mathematical Sciences, Kent State University, Kent, USA
`reichel@math.kent.edu`

Abstract. This paper introduces a new approach to computing an approximate solution of Tikhonov-regularized large-scale ill-posed problems with a general nonlinear regularization operator. The iterative method applies a sequence of projections onto generalized Krylov subspaces using a semi-implicit approach to deal with the nonlinearity in the regularization term. A suitable value of the regularization parameter is determined by the discrepancy principle. Computed examples illustrate the performance of the method applied to the restoration of blurred and noisy images.

1 Introduction

We consider solution methods for minimization problems of the form

$$\min_{x \in \mathbb{R}^n} \left\{ \|Ax - b\|_2^2 + \mu \|\mathcal{L}(x)\|_2^2 \right\}, \tag{1}$$

where $A \in \mathbb{R}^{m \times n}$, $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, and $\mathcal{L} : \mathbb{R}^n \to \mathbb{R}^s$ is a linear or nonlinear operator. In particular, in this paper, we applied model (1) to the restoration of images, where the vector x represents the desired unknown image and the available noise- and possibly blur-contaminated image is represented by the vector b. Thus

$$b = \hat{b} + e, \tag{2}$$

where $\hat{b}$ is the unknown error-free vector associated with b, and e represents the error. The first term in (1) is commonly referred to as the *fidelity term* and the second term as the *regularization term*. The operator $\mathcal{L}$ is chosen to yield a computed solution with some known desired features. The scalar $\mu > 0$ is a regularization parameter. Its purpose is to balance the influence of the fidelity and regularization terms on the computed solution in a suitable manner.

In image restoration applications, A is a blurring operator which is generally severely ill-conditioned and may be singular. The purpose of the regularization term is to be able to determine a useful solution of (1) of moderate norm.

When $\mathcal{L}$ is a linear operator represented by a regularization matrix $L \in \mathbb{R}^{s \times n}$, the minimization problem (1) is the classical Tikhonov-regularized linear least-squares problem, and assuming that $\mathrm{rank} \begin{bmatrix} A \\ L \end{bmatrix} = n$, the problem (1) has the unique solution

$$x_\mu = \left(A^T A + \mu\, L^T L\right)^{-1} A^T b \tag{3}$$

for any $\mu > 0$. Here and below the superscript ^{T} denotes transposition. For large-scale problems and a fixed $\mu > 0$, an approximation of x_μ can be determined by applying an iterative method, such as LSQR, to (3). However, generally, a suitable value of the parameter μ is not known a priori and has to be determined during the solution process. Many methods for the selection of an appropriate regularization parameter μ require the normal equations for (3) to be solved repeatedly for many different values of the parameter μ.

A popular strategy for the selection of μ is based on the discrepancy principle. Assume that an estimate of the norm of the error e in the vector b in (2) is available,

$$\epsilon \approx \|e\|.$$

Then the discrepancy principle prescribes that the regularization parameter $\mu = \mu(\epsilon)$ be chosen so that the computed approximation $\widetilde{x}_\mu$ of the solution x_μ of (1) satisfies

$$\|A\widetilde{x}_\mu - b\| = \delta := \eta\epsilon, \tag{4}$$

where $\eta > 1$ is a user-specified constant, whose size depends on the accuracy of the estimate ϵ of $\|e\|$. A vector $\widetilde{x}_\mu$, such that (4) holds, is said to satisfy the discrepancy principle.

When L is the identity matrix, denoted by I, the Tikhonov regularization problem (1) is said to be in *standard form* and the solution can be computed efficiently by, e.g., partial Golub–Kahan bidiagonalization of A; see, e.g., [1–3]. The computed approximation $\widetilde{x}_\mu^k$ lives into a Krylov subspace

$$\mathcal{K}_k(A^T A, A^T b) = \mathrm{span}\{A^T b, (A^T A)A^T, \ldots, (A^T A)^{k-1}A^T b\} \tag{5}$$

for some $k \geq 1$.

Golub–Kahan bidiagonalization also can be applied when $L \neq I$, provided that the regularization problem can be transformed to standard form without too much effort by applying the *A-weighted pseudoinverse* of L; see [4] for details.

As an alternative, a scheme that projects L into the Krylov subspace (5) and determines an approximate solution of (1) in this subspace is described in [7]. We will refer to this Krylov subspace approach to the problem (1) as KS.

Methods based on reducing both A and L by generalized Arnoldi-type or Golub–Kahan-type methods are discussed in [8, 10].

An efficient iterative algorithm for the solution of large-scale Tikhonov regularization problems (1) with automatic selection of the regularization parameter μ is proposed in [9]. The method relies on successive orthogonal projections onto generalized Krylov subspaces of increasing and low dimension k. On the one

hand, the use of linear regularizing operators $\mathcal{L}$ different from the identity, has been proved to be very effective [8–10], on the other hand the application of nonlinear operators $\mathcal{L}$ is well-assessed in image restoration. This motivated the presented work that is an extension of the algorithm in [9] which can be applied to the solution of (1) for general nonlinear operators. By using a semi-implicit approach we replace (1) by

$$\min_{x \in \mathbb{R}^n} \left\{ \|Ax - b\|_2^2 + \mu \|\mathcal{L}(x, x_*)\|_2^2 \right\}, \tag{6}$$

where x_* is a known approximation, e.g. the result of a previous step in an iterative procedure. Therefore, the nonlinear operator $\mathcal{L}$ becomes a linear function of x. We refer to this method as the GKS method, which will be described in the following section.

2   The GKS method

The method proposed in [9] is designed for the solution of large-scale Tikhonov regularization problems (1) with a general linear regularization operator $\mathcal{L}$ and determines the regularization parameter $\mu > 0$ by the discrepancy principle [5]. The method determines a sequence of orthogonal projections of generalized Krylov subspaces onto subspaces of low dimension; the operator $\mathcal{L}$ is unchanged at each iteration.

In the proposed GKS extension, we assume $\mathcal{L}(x)$ to be a nonlinear regularization operator, such as those describe in subsection 2.1, which is updated at each iterative step and contributes to enlarge the generalized Krylov subspace. The GKS method is initialized with a user-chosen subspace $\mathcal{V}_0 \subset \mathbb{R}^n$ of low dimension $l \ll n$. In the numerical section we let $\mathcal{V}_0 = \mathcal{K}_l(A^T A, A^T b)$ for $l \leq 3$. Let the columns of the matrix $V_0 \in \mathbb{R}^{n \times k}$ form an orthonormal basis for the space $\mathcal{V}_0$.

Assume that an estimate $\delta > 0$ for the l_2-norm of the noise in b is available. The discrepancy principle (4) is then applied to determine a suitable value of $\mu = \mu_k$ for each iteration $k = 0, 1, \dots$ and compute the solution $x^{(k)}$ of the Tikhonov minimization problem (6) restricted to the subspace $\mathcal{V}_k$. This subspace is spanned by the orthonormal columns of the matrix V_k. Thus, $x^{(k)}$ is computed as follows:

$$y^{(k)} = \arg \min_{y \in \mathbb{R}^{l+k}} \left\{ \|AV_k y - b\|_2^2 + \mu_k \|L_{k-1} V_k y\|_2^2 \right\}, \quad x^{(k)} = V_k y^{(k)}, \tag{7}$$

where L_{k-1} denotes the discretization of the nonlinear operator $\mathcal{L}$ linearized at the approximate solution $x_\mu^{(k-1)}$, which was computed at the previous iteration step.

Define the $n \times n$ matrix

$$T(\mu_k) := A^T A + \mu_k L_{k-1}^T L_{k-1} . \tag{8}$$

The solution $y^{(k)}$ of the reduced minimization problem (7) can be determined by solving the corresponding normal equations,

$$V_k^T T(\mu_k) V_k\, y \;=\; V_k^T A^T b\,, \qquad (9)$$

for $y^{(k)}$. The associated approximate solution $x^{(k)}$ of the original unreduced problem (1) is given by $x^{(k)} = V_k y^{(k)}$.

Following the approach in the nonlinear Arnoldi method also used in [9], the subspace $\mathcal{V}_k$ is expanded to $\mathcal{V}_{k+1}$ by adding a new basis vector v_{new} to $\mathcal{V}_k$. This basis vector is determined by normalizing the residual $r^{(k)}$ of the unreduced problem (1). Thus,

$$v_{\mathrm{new}} = \frac{r^{(k)}}{\|r^{(k)}\|_2}\,, \qquad r^{(k)} = T(\mu_k) x^{(k)} - A^T b\,. \qquad (10)$$

Notice that the residual vector $r^{(k)}$ is parallel to the gradient of the functional minimized in the original unreduced problem (1) evaluated at $x^{(k)}$. In the absence of round-off errors, $r^{(k)}$ is orthogonal to the search space $\mathcal{V}_k$. To enforce orthogonality in the presence of round-off errors, the residual $r^{(k)}$ can be re-orthogonalized against $\mathcal{V}_k$ before normalization.

We remark that the space $\mathcal{V}_k$ is a Krylov subspace only in very special situations. Since the regularization parameter μ_k is updated during the iterations, so is the matrix $T(\mu_k)$. Therefore, the space $\mathcal{V}_k$, in general, is not a Krylov subspace when $L \neq I$. For this reason, the search space $\mathcal{V}_k$ is referred to as a generalized Krylov subspace.

2.1   Choice of the regularization operator $\mathcal{L}$

Nonlinear PDE-based methods have been applied successfully to noise-removal [12]. In the continuous setting, a popular general nonlinear diffusion model determines a denoised image w from a noise-contaminated image b by solving

$$\frac{\partial w}{\partial t} = \mathrm{div}(g(|\nabla w|)\nabla w), \quad w(x,0) = b. \qquad (11)$$

According to the definition of the diffusivity function g, that are, i.e.

$$g_{PM}(s) := 1/(1 + s^2/\rho^2), \quad g_{TV}(s) := 1/s, \quad g_{Lap}(s) := 1/s, \qquad (12)$$

we get the well-known Perona Malik (PM) [12], Total Variation (TV) [11], and Laplacian (Lap) restoration models.

The regularization matrix L_{k-1} at the iteration k of the iterative method GKS is obtained by the linearization of the right hand side of (11) that is, by evaluating g at $|\nabla w|$ using w from the previous time-step i, i.e.,

$$\mathcal{L}(w^{(i)}) \approx \mathrm{div}(g(|\nabla w^{(i-1)}|)\nabla w^{(i)}). \qquad (13)$$

Discretization of (13) gives the matrix $L_{k-1} \in \mathbb{R}^{n^2 \times n^2}$ with, generically, five nonvanishing entries in each row. Spatial partial derivatives are approximated by central finite differences; see [12] for details.

After a few time steps i (5 in the computed examples at most), the regularization matrix L_{k-1} is updated to obtain the new regularization matrix. In the numerical section, we compare results obtained when the matrix L is defined by the different choices in (12). We will denote these matrices by L_{PM}, L_{TV}, and L_{Lap}, respectively.

3 Numerical examples

This section illustrates the performance of the proposed GKS approach described in Section 2 when applied to the restoration of gray scale images `test`, `barbara`, `peppers`, and `cameraman`. These images are represented by arrays of 256×256 pixels stored column-wise in vectors in $\mathbb{R}^n$ with $n = 65536$.

The images are synthetically corrupted by blur and noise. Let the vector $x \in \mathbb{R}^n$ represent the original, unavailable blur- and noise-free image. A block Toeplitz with Toeplitz blocks blurring matrix $A \in \mathbb{R}^{n \times n}$ is generated with the function `blur` from [6]. This function has the parameters `band` and `sigma`, which determine the half-bandwidth of each Toeplitz block in A and the standard deviation of the underlying Gaussian point spread function, respectively.

We compare the quality of the restorations determined by the GKS and KS methods by measuring the Signal-to-Noise Ratio SNR defined by

$$\text{SNR}(x^*, x) := 10 \log_{10} \frac{\|x - E(x)\|_2^2}{\|x^* - x\|_2^2} \ (\text{dB}) \,, \tag{14}$$

where $E(x)$ denotes the mean gray-level of the uncontaminated image x and x^* is an accurate approximation of x; larger SNR-values generally indicate more accurate restorations.

Computational efficiency is measured in terms of the total number of iterations, Its in the Tables, required by the algorithms to satisfy the following stopping criterion. The number of iterations of the KS and GKS algorithms are terminated as soon as the relative error $\|x^{(k)} - x^{(k-1)}\| / \|x^{(k)}\|$ of the computed approximate solution $x^{(k)}$ drops below a user-specified threshold which for the reported experiments is $\tau = 5 \cdot 10^{-4}$.

The overall computational cost for Its iterations with the GKS algorithm is dominated by the count of arithmetic floating point operations required to evaluate $4Its$ matrix-vector products (MVP); each MVP with T is computed by evaluating one MVP with each one of the matrices L, L^T, A and A^T; cf. (8). Moreover, at each iteration, we computate a new regularization matrix L. This is not expensive due to the sparsity of L. The dominant computational cost for each iteration with the KS method is limited to 2 MVPs.

The parameter η in (4) is 1.05.

Table 1 reports results for the restorations of the test images `test` and `cameraman` degraded by Gaussian blur with different parameters `band` and `sigma`, and a fixed Gaussian noise with $\sigma = 10$. Specifically, we considered the three regularization operators L_{Lap}, L_{PM}, and L_{TV} defined in Section 2.1.

The SNR values of the computed restorations are reported in the last column of the tables, and the SNR-values of the initial contaminated images are tabulated in the column labelled SNR_0.

Table 1 illustrates the good performance of the GKS approach in terms of both accuracy of the restorations and computational effort.

Figures 1 and 2 show restorations obtained by the GKS and KS algorithms when applied to the test images test and cameraman. The degraded images in Figures 1(b) and 2(b) were obtained from the original blur- and noise-free images in Figures 1(a) and 2(a), respectively, by applying Gaussian blur with parameters band = 5 and sigma = 1.5, for test and band = 5 and sigma = 2.0 for cameraman. The noise is Gaussian with $\sigma = 10$ and the regularization operator is L_{TV} for both images. Figures 1(c),2(c) and Figures 1(d),2(d) depict restorations obtained with the GKS and KS methods, respectively. For both test images, the GKS method yields a slightly more accurate restoration than the KS method both in terms of visual quality and SNR-value.

blur		Reg. Model	SNR_0	Its		SNR	
band	sigma	L		KS	GKS	KS	GKS
test							
		L_{Lap}		6	6	11.29	11.10
3	1	L_{PM}	2.61	3	7	11.61	11.81
		L_{TV}		6	3	11.30	11.44
		L_{Lap}		6	9	8.98	10.60
5	1.5	L_{PM}	3.80	12	8	10.25	10.42
		L_{TV}		10	3	9.92	10.63
cameraman							
		L_{Lap}		7	6	11.36	11.66
3	1	L_{PM}	3.77	8	8	11.76	12.00
		L_{TV}		11	2	11.42	11.76
		L_{Lap}		10	10	9.99	10.66
5	1.5	L_{PM}	4.80	12	9	10.65	10.80
		L_{TV}		13	2	10.07	10.34

Table 1. Comparison of the KS and GKS algorithms applied to the image restorations of test and cameraman images corrupted by different kinds of Gaussian blur and fixed Gaussian additive noise with $\sigma = 10$.

In Table 2, we consider the restoration of three different images corrupted by Gaussian blur with band = 5 and sigma = 2.0 (second column), and by additive zero-mean white Gaussian noise with different standard deviations σ. The benefits of the GKS algorithm using the L_{TV} regularization operator can

Image	noise	SNR_0	Its		SNR	
	σ		KS	GKS	KS	GKS
peppers	5	1.74	10	2	12.40	12.58
	10	1.60	13	3	10.84	11.26
	15	1.38	9	3	9.78	10.65
	20	1.08	8	3	8.87	8.81
cameraman	5	-0.33	4	2	10.54	10.80
	10	-0.39	16	3	9.20	9.81
	15	-0.49	36	2	8.37	9.30
	20	-0.63	36	3	7.71	7.90
barbara	5	-0.37	6	2	7.96	8.13
	10	-0.50	4	3	7.15	7.38
	15	-0.71	6	3	6.47	6.97
	20	-0.99	7	3	5.88	6.64

Table 2. Comparison of the GKS and KS algorithms using $L = L_{TV}$; the images have been contaminated by Gaussian blur with `band = 5` and `sigma = 2.0`, and additive zero-mean white Gaussian noise with standard deviation σ.

be observed both in terms of computational effort and quality of the restored images (SNR).

Summarizing the presented results, we see that GKS achieves higher accuracy than KS. Thus, using a sequence of matrices L to construct generalized Krylov subspaces is beneficial for image restorations.

4 Conclusions

The proposed scheme is an extension to the method in [9] and allows nonlinear regularization operators. This has been realized by a semi-implicit strategy based on updating the regularization matrix L at every iteration step. We have shown that this approach requires fewer iterations than application of a standard Krylov subspace method and produces restored images of higher quality.

References

1. Å. Björck, A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations, BIT Numer. Math., 28 (1988), pp. 659–670.
2. D. Calvetti and L. Reichel, Tikhonov regularization of large linear problems, BIT Numer. Math., 43 (2003), pp. 263–283.
3. D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari, Tikhonov regularization and the L-curve for large, discrete ill-posed problems, J. Comput. Appl. Math., 123 (2000), pp. 423–446.
4. L. Eldén, A weighted pseudoinverse, generalized singular values, and constrained least squares problems, BIT Numer. Math., 22 (1982), pp. 487–502.
5. H. W. Engl, M. Hanke and A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, The Netherlands, 1996.
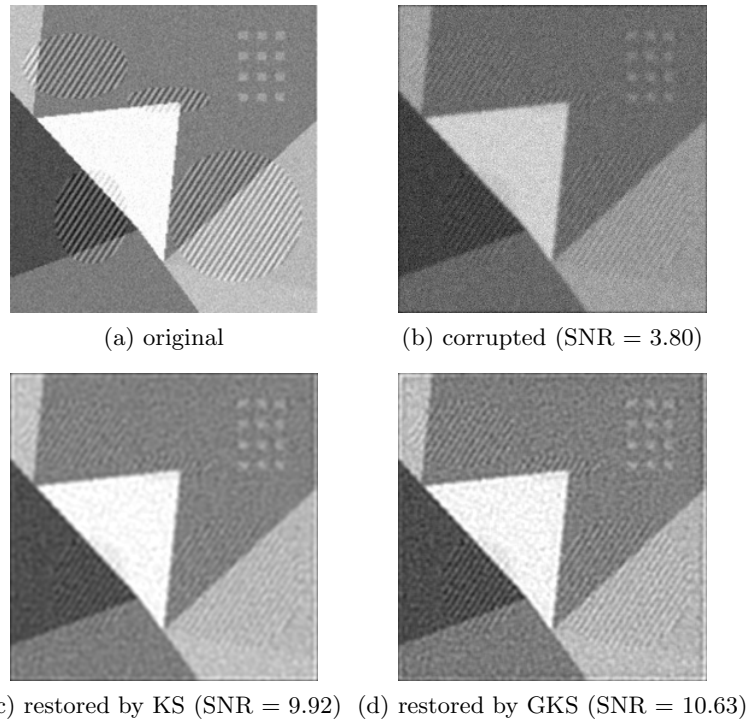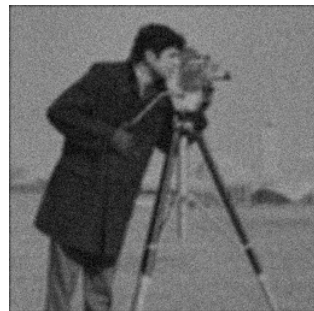
(a) original   (b) corrupted (SNR = 3.80)

(c) restored by KS (SNR = 9.92)   (d) restored by GKS (SNR = 10.63)

Fig. 1. Restoration results obtained by the KS and GKS algorithms applied to a `test` image that has been contaminated by Gaussian blur with `band` = 5 and `sigma` = 1.5, and Gaussian noise with standard deviation $\sigma = 10$.

6. P. C. Hansen, Regularization tools version 4.0 for Matlab 7.3, Numer. Algorithms, 46 (2007), pp. 189–194.
7. M. E. Hochstenbach and L. Reichel, An iterative method for Tikhonov regularization with a general linear regularization operator, J. Integral Equations Appl., 22 (2010), pp. 463–480.
8. M. E. Hochstenbach, L. Reichel, and X. Yu, A Golub–Kahan-type reduction method for matrix pairs, submitted for publication.
9. J. Lampe, L. Reichel, and H. Voss, Large-scale Tikhonov regularization via reduction by orthogonal projection, Linear Algebra Appl., 436 (2012), pp. 2845–2865.
10. L. Reichel, F. Sgallari, and Q. Ye, Tikhonov regularization based on generalized Krylov subspace methods, Appl. Numer. Math., 62 (2012), pp. 1215–1228.
11. L. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D, 60 (1992), pp. 259–268.
12. J. Weickert, B. M. H. Romeny, and M. A. Viergever, Efficient and reliable schemes for nonlinear diffusion filtering, IEEE Trans. Image Processing, 7 (1998), pp. 398–410.

(a) original   (b) corrupted (SNR_0 = -0.39)

(c) restored by KS (SNR = 9.30) (d) restored by GKS (SNR = 9.81)

Fig. 2. Restoration results obtained by the KS and GKS algorithms applied to a `cameraman` image that has been contaminated by Gaussian blur with `band` = 5 and `sigma` = 2.0, and Gaussian noise with standard deviation $\sigma = 10$.