# GENERALIZED ARNOLDI-TIKHONOV METHOD FOR SPARSE RECONSTRUCTION

SILVIA GAZZOLA* AND JAMES G. NAGY†

**Abstract.** This paper introduces two new algorithms, belonging to the class of Arnoldi-Tikhonov regularization methods, which are particularly appropriate for sparse reconstruction. The main idea is to consider suitable adaptively-defined regularization matrices that allow the usual 2-norm regularization term to approximate a more general regularization term expressed in the $p$-norm, $p \geq 1$. The regularization matrix can be updated both at each step and after some iterations have been performed, leading to two different approaches: the first one is based on the idea of the Iteratively Reweighted Least Squares method and can be obtained considering Flexible Krylov Subspaces; the second one is based on restarting the Arnoldi algorithm. Numerical examples are given in order to show the effectiveness of these new methods, and comparisons with some other already existing algorithms are made.

**Key words.** inverse problems, sparse reconstruction, regularization, total variation, Arnoldi method, Krylov subspace, preconditioning

**AMS Subject Classifications:** 65F20, 65F30

## 1. Introduction.

In this paper we consider large-scale linear ill-posed inverse problems of the form

$$b = Ax + e, \tag{1.1}$$

where the matrix $A \in \mathbb{R}^{N\times N}$ typically has singular values that quickly decay to, and cluster at zero; without loss of generality we assume $A$ to be scaled so that $\|A\|_2 = 1$. It is assumed that $b \in \mathbb{R}^N$ and $A$ are known, and the aim is to compute an approximation of $x \in \mathbb{R}^N$. The vector $e \in \mathbb{R}^N$ represents error, such as noise, that affects the data. Although $e$ is not known, in this paper we consider problems for which a good estimate of $\|e\|_2$ is known. Due to the ill-conditioning of $A$ and the presence of noise in the right-hand side, it is necessary to employ regularization in order to compute a meaningful approximation of $x$.

There are several techniques to regularize the linear inverse problem given by equation (1.1); see, for example, [11, 15, 26, 40]. One of the most well-known and well-established is *Tikhonov regularization*, which, in its general $\ell^2$ formulation, computes an approximation of $x$ by solving the following minimization problem

$$\min_{x\in\mathbb{R}^N} \left\{ \|Ax - b\|_2^2 + \lambda \|L(x - x_0)\|_2^2 \right\}, \tag{1.2}$$

where $\lambda > 0$ is called a regularization parameter, $L \in \mathbb{R}^{q\times N}$ is called a regularization matrix, and $x_0 \in \mathbb{R}^N$ is an initial guess for the solution (if no initial approximation is available, we simply take $x_0 = 0$). The choice of the regularization matrix and the value of the regularization parameter can be crucial to obtain a good approximation of $x$. For example, under-estimating $\lambda$ can lead to highly oscillatory (noisy) approximations of $x$, while over-estimating $\lambda$ can cause the solution to be overly smooth. Moreover, the regularization matrix enforces certain constraints on the solution; common choices for $L$ are the $N \times N$ identity matrix, $I$, or a scaled finite-difference

* Department of Mathematics, University of Padova, Italy. Email: gazzola@math.unipd.it.

†Department of Mathematics and Computer Science, Emory University, Atlanta, GA, USA. Email: nagy@mathcs.emory.edu. This work was supported by grant no. DMS-1115627 from the US National Science Foundation, and grant no. AF9550-12-1-0084 from the US Air Force Office of Scientific Research.

approximation of a derivative operator, such as the Laplacian. When $L = I$ and $x_0 = 0$, the problem is said to be in standard form.

Another popular way to solve (1.1) is by *iterative regularization*. In this setting an iterative method is applied directly to

$$\min_{x\in\mathbb{R}^N} \|Ax - b\|_2^2$$

and regularization is achieved by early termination, i.e. by stopping the iterations before the so-called inverted noise dominates the approximate solution. In this sense the number of iterations plays the role of a discrete regularization parameter.

A third approach, which is referred to as a *hybrid method*, combines a direct regularization scheme, such as Tikhonov, with an iterative Krylov subspace method. O'Leary and Simmons [30] originally introduced the idea of hybrid methods using Golub-Kahan (which is often called Lanczos) bidiagonalization. A hybrid approach based on the Arnoldi algorithm, which is referred to as the Arnoldi-Tikhonov method, was proposed in [7]. An advantage of using an Arnoldi based hybrid method is that each iteration requires only a matrix product with $A$, whereas in the Golub-Kahan based approaches, each iteration requires matrix vector multiplication with $A$ and $A^T$. Various authors have considered computation and implementation issues of hybrid methods, such as robust approaches to choose regularization parameters and stopping iterations; see for example, [3, 5, 8, 10, 12, 13, 18, 20, 21, 23, 32].

However, employing 2-norm filtering schemes, as is done in these previous works, is rather restrictive, and better approximations can be computed by considering more general optimizations of the form

$$\min_{x\in\mathbb{R}^N} \{\mathcal{J}(x) + \lambda\mathcal{R}(x)\} \tag{1.3}$$

where $\mathcal{J}(x)$ is a fit-to-data term, and $\mathcal{R}(x)$ is a regularization term. For example, it is known that solving

$$\min_{x\in\mathbb{R}^N} \left\{\|Ax - b\|_2^2 + \lambda\|x\|_p^p\right\} \tag{1.4}$$

reconstructs a sparse approximation of $x$ when $p = 1$. One can also consider changing the norm on the fit-to-data term, and solve the general optimization problem

$$\min_{x\in\mathbb{R}^N} \left\{\|Ax - b\|_q^q + \lambda\|x\|_p^p\right\}, \tag{1.5}$$

where we assume $1 \le p < 2$, $1 \le q < 2$. If the goal is to preserve jumps (i.e. edges) in $x$, then we may prefer to solve

$$\min_{x\in\mathbb{R}^N} \left\{\|Ax - b\|_2^2 + \lambda\mathrm{TV}(x)\right\}, \tag{1.6}$$

where $TV(x)$ is the *total variation* operator [35].

Nonlinear optimization methods are needed to solve problems (1.4)-(1.6), and they are therefore more computationally demanding than simply solving the linear least squares problem associated with the standard $\ell^2$ formulation of Tikhonov regularization (1.2). A lot of work has been done to propose suitable and efficient algorithms that can deal with the nonlinear problems; see, for example, [40] for problem (1.6) and [42] for problem (1.5), and the references therein.

In this paper, we describe a hybrid framework to solve these minimization problems. To simplify notation, we focus our discussion on problems (1.4) and (1.6), but the approach we describe is fairly general, and can be used to solve (1.3) for a variety of combinations of fit-to-data and regularization terms. Our approach is to approximate the regularization term by means of a quadratic functional, similarly to what is done in the iteratively reweighted least squares (IRLS) method (cf. [4], Ch.4). Specifically, we adaptively define a suitable regularization matrix (or, using the IRLS terminology, weighting matrix) $L$ that allows us to approximate

$$\|Lx\|_2^2 \simeq \mathcal{R}(x)\,,$$

where we focus on the cases $\mathcal{R}(x) = \|x\|_1$ and $\mathcal{R}(x) = \mathrm{TV}(x)$. The matrix $L$ is adaptively defined since it is automatically updated at each iteration or when a convenient number of iterations has been performed.

We note that the IRLS approach has been previously used for these problems. For example, in [41] the authors define a particular matrix $W$ that can be used to approximate $\|\cdot\|_1$ as well as the TV operator. A very similar approach is adopted in [31], where it is shown that the algorithm can be regarded as a majorization-minimization (MM) process [19]. However, to the best of our knowledge, the only published work that adopts the IRLS strategy to deal with problems of the form (1.5) is [34]. All of these approaches solve, at each iteration, a weighted least squares problem using the conjugate gradient method applied to the normal equations. In particular, each IRLS iteration generates *a new* Krylov subspace from scratch. Furthermore, to generate each Krylov subspace it is necessary to compute matrix-vector multiplications with both $A$ and $A^T$.

The approach described in this paper is designed to work in connection with the hybrid Arnoldi-Tikhonov method. Specifically, we show that our approach can be implemented very efficiently by interpreting the scheme as a flexibly preconditioned Krylov subspace method. This means that we generate *only one* Krylov subspace. Moreover, each iteration for our approach only requires one matrix-vector multiplication with $A$. One disadvantage of our approach is that storage and computational cost can become an issue in some cases, especially when the associated preconditioner is not easily inverted. For these cases, a restarting strategy is suggested.

This paper is organized as follows. In Section 2 we briefly describe a generalization of the Arnoldi Tikhonov method and we explain how to simultaneously choose the regularization parameter and stop the iterations. In Section 3 we describe how to solve the more general formulations (1.4)-(1.5) with an IRLS based approach. In Section 4 we describe an algorithm to solve these problems that is based on flexible Krylov subspaces, and in Section 5 we describe a second, storage efficient, algorithm that uses suitable restarts of the underlying iterative method. In Section 6 we present some numerical results and we make comparisons with already existing methods for sparse reconstruction. Concluding remarks are given in Section 7.

**2. Generalized Arnoldi-Tikhonov (GAT) Method.** The Arnoldi-Tikhonov (AT) method was proposed in [7] with the basic aim of reducing problem (1.2) to a problem of much smaller dimension. In this section we describe a generalization of the approach [12, 18], which we call the generalized Arnoldi-Tikhonov (GAT) method. To describe this method, first consider the Krylov subspace

$$\mathcal{K}_m(A, r_0) = \mathrm{span}\{r_0, Ar_0, \ldots, A^{m-1}r_0\}, \quad r_0 = b - Ax_0, \quad m \ll N\,.$$

The Arnoldi algorithm [37] can be used to construct a basis for this Krylov subspace, which leads to the associated decomposition

$$AV_m = V_{m+1}\bar{H}_m, \tag{2.1}$$

where $V_{m+1} = [v_1, ..., v_{m+1}] \in \mathbb{R}^{N\times(m+1)}$ has orthonormal columns that span the Krylov subspace $\mathcal{K}_{m+1}(A, r_0)$, and $v_1$ is defined as $r_0/\|r_0\|_2$. The matrix $\bar{H}_m \in \mathbb{R}^{(m+1)\times m}$ is an upper Hessenberg matrix. Denoting by $h_{i,j}$ the entries of $\bar{H}_m$, in exact arithmetic the Arnoldi process terminates when $h_{m+1,m} = 0$, which means $\mathcal{K}_{m+1}(A, r_0) = \mathcal{K}_m(A, r_0)$.

The GAT method searches for approximations $x_m$ of the solution of problem (1.2) belonging to $x_0 + \mathcal{K}_m(A, r_0)$. Therefore, replacing $x = x_0 + V_m y$ ($y \in \mathbb{R}^m$ and $V_m$ is defined by (2.1)) into (1.2), yields the reduced minimization problem

$$y_m = \arg\min_{y\in\mathbb{R}^m} \left\{ \left\|\bar{H}_m y - V_{m+1}^T r_0\right\|_2^2 + \lambda \left\|LV_m y\right\|_2^2 \right\}, \tag{2.2}$$

where we have used that $V_{m+1}^T V_{m+1} = I$ and, since $v_1 = r_0/\|r_0\|_2$,

$$V_{m+1}^T r_0 = \|r_0\|_2 e_1, \quad \text{where} \quad e_1 = (1, 0, \ldots, 0)^T \in \mathbb{R}^{m+1}.$$

At each step of the Arnoldi algorithm, instead of solving (2.2) directly, we consider the following equivalent reduced-dimension least squares formulation

$$y_m = \arg\min_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} LV_m \end{pmatrix} y - \begin{pmatrix} c \\ 0 \end{pmatrix} \right\|_2^2, \tag{2.3}$$

where $c = \|r_0\|_2 e_1$. The coefficient matrix associated with (2.3) is typically tall, since $LV_m \in \mathbb{R}^{q\times m}$. Therefore, as suggested in [18], one could compute a QR factorization $LV_m = Q_m R_m$ ($Q_m \in \mathbb{R}^{q\times m}$ has orthonormal columns, $R_m \in \mathbb{R}^{m\times m}$ is upper triangular), and replace the matrix $LV_m$ in (2.3) by $R_m$. We also remark that, when dealing with standard form problems, the Arnoldi-Tikhonov formulation considerably simplifies thanks to the orthogonality of the columns of $V_m$ and, instead of (2.2), we can write

$$y_m = \arg\min_{y\in\mathbb{R}^m} \left\{ \left\|\bar{H}_m y - V_{m+1}^T\, b\right\|_2^2 + \lambda \left\|y\right\|_2^2 \right\}. \tag{2.4}$$

As a consequence, instead of (2.3) we can consider

$$y_m = \arg\min_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \bar{H}_m \\ \sqrt{\lambda} I \end{pmatrix} y - \begin{pmatrix} \|b\|_2\, e_1 \\ 0 \end{pmatrix} \right\|_2^2. \tag{2.5}$$

It should be noted that we consider Krylov subspaces $\mathcal{K}_m(A, r_0)$, and therefore the approach we describe in this paper cannot be directly used for rectangular matrices $A \in \mathbb{R}^{M\times N}$, $M \geq N$; however, it might be possible to adapt our approach to use, for example, the techniques proposed in [17] for least squares problem.

In the remaining part of this section we briefly review a discrepancy-principle based strategy originally introduced in [12]: this approach can successfully be employed to define $\lambda$ as well as a stopping criterion (i.e. the dimension $m$ of the Krylov subspace) when performing the GAT method. At each iteration we define the function

$$\phi_m(\lambda) = \|Ax_m - b\|_2 \tag{2.6}$$

$$= \|\bar{H}_m y_m - c\|_2, \tag{2.7}$$

where equality (2.7) is obtained exploiting relation (2.1) and the properties of its associated matrices. Assuming that a fairly accurate approximation of the quantity $\varepsilon = \|e\|_2$ is available, we say that the discrepancy principle is satisfied as soon as

$$\phi_m(\lambda) \leq \eta\varepsilon, \quad \text{where} \quad \eta \gtrapprox 1\,. \tag{2.8}$$

If a good estimate of the noise level $\widetilde{\varepsilon} = \|e\|_2/\|\bar{b}\|_2$ is known, the discrepancy principle reads $\phi_m(\lambda) = \eta\widetilde{\varepsilon}\|b\|$. By default we set $\eta = 1.01$.

At each iteration we approximate the discrepancy function (2.6) by the linear function

$$\phi_m(\lambda) \simeq \phi_m(0) + \lambda\beta_m, \tag{2.9}$$

where $\beta_m \in \mathbb{R}$ is defined by the ratio

$$\beta_m = \frac{\phi_m(\lambda_{m-1}) - \phi_m(0)}{\lambda_{m-1}}\,. \tag{2.10}$$

In the above expression, $\phi_m(\lambda_{m-1})$ is obtained by solving the $m$-dimensional problem (2.3) using the parameter $\lambda_{m-1}$, which is obtained at the previous step ($\lambda_0$ must be set to an initial value by the user); we also remark that $\phi_m(0)$ is the norm of the GMRES residual at the $m$-th iteration of the Arnoldi algorithm.

To select $\lambda_m$ for the next step of the generalized Arnoldi-Tikhonov algorithm we impose

$$\phi_m(\lambda_m) = \eta\varepsilon \tag{2.11}$$

and we force the approximation

$$\phi_m(\lambda_m) = \phi_m(0) + \lambda_m\beta_m\,. \tag{2.12}$$

Substituting in (2.12) the expression derived in (2.10), and using the condition (2.11), we obtain

$$\lambda_m = \left|\frac{\eta\varepsilon - \phi_m(0)}{\phi_m(\lambda_{m-1}) - \phi_m(0)}\right| \lambda_{m-1}\,, \tag{2.13}$$

where the absolute value has been considered in order to guarantee the positivity of $\lambda_m$; indeed, in early iterations, typically $\alpha_m \gg \eta\varepsilon$ and discarding the absolute value would produce a negative value for $\lambda_m$. Numerically, formula (2.13) is very stable, in the sense that after the discrepancy principle is satisfied, $\lambda_m$ is almost constant for growing values of $m$. We finally remark that this strategy is also very robust with respect to the initial choice of $\lambda$; we use the default value $\lambda_0 = 1$.

**3. Iteratively Reweighted Least Squares (IRLS) Method.** The Iterative Reweighted Least Squares (IRLS) method was originally introduced in a statistical framework to approximate the solution of

$$\min_x \|Ax - b\|_p^p, \ 1 \leq p < 2\,, \tag{3.1}$$

where an $\ell^p$ norm estimator is employed instead of the usual $\ell^2$ norm because it can lead to a more robust solution when recovering certain parameters in a linear model. The basic idea of IRLS is to reduce the minimization (3.1) to a sequence of least

squares problems involving a weighted $\ell^2$ norm. That is, a sequence of problems of the form

$$\min_x \|L(Ax-b)\|_2^2 \tag{3.2}$$

is solved, where $L$ is a diagonal *weighting* matrix that is updated at each step using the solution obtained at the previous iteration.

In the following we describe how the Arnoldi-Tikhonov method can be employed to approximate the solution of the sparse reconstruction problem (1.4) and the total variation reconstruction problem (1.6). Specifically, using the same idea underlying the IRLS method, we adaptively define a suitable regularization matrix by means of the information obtained at the previous steps.

**3.1. Sparse Reconstruction.** We first consider choosing preconditioners $L$ so that $\|Lx\|_2^2$ approximates $\|x\|_p^p$, $1 \le p < 2$. Of course, the best choice would be to take

$$L = \operatorname{diag}\left(\left(|[x]_i|^{\frac{p-2}{2}}\right)_{i=1,\dots,N}\right), \tag{3.3}$$

where we use the notation $[\cdot]_i$ to denote the $i$-th element of the vector inside the brackets[1]. But this is not possible in real problems, because it would require the exact solution $x$. However, it is natural to define, at iteration $m$, the matrix

$$L_m = \operatorname{diag}\left(\left(|[x_{m-1}]_i|^{\frac{p-2}{2}}\right)_{i=1,\dots,N}\right), \tag{3.4}$$

where $x_{m-1} = x_0 + V_{m-1}y_{m-1}$ is the solution obtained at the previous step of the Arnoldi-Tikhonov method. When the first iteration is performed (i.e. when $m=1$) we simply take $L_1 = I$. Typically, as the iterations proceed, the term $\|L_m x_m\|_2^2$ increasingly better approximates (3.3) and, therefore, it increasingly better approximates the quantity $\|x\|_p^p$. Since we are particularly interested in treating the case of $\|x\|_1$, we explicitly write the matrix (3.4) for $p=1$

$$L_m = \operatorname{diag}\left(\left(\frac{1}{\sqrt{|[x_{m-1}]_i|}}\right)_{i=1,\dots,N}\right). \tag{3.5}$$

We remark that, when $p<2$, care is needed when defining (3.4), because division by 0 may occur if $[x_{m-1}]_i = 0$ for some $i = 1,\dots,N$. Therefore it is safer to set a small threshold $\tau > 0$ and take

$$L_m = \operatorname{diag}\left((f_\tau([x_{m-1}]_i)_{i=1,\dots,N}\right), \tag{3.6}$$

where

$$f_\tau(\chi) = \begin{cases} |\chi|^{(p-2)/2} & \text{if } \chi > \tau \\ \tau^{(p-2)/2} & \text{if } \chi \le \tau \end{cases}. \tag{3.7}$$

[1]In this paper we generally use the notation $x_m$ to be the vector at the $m$-th iteration of the GAT method. To avoid confusion between iteration index and element of a vector, we therefore use the notation $[\cdot]_i$ to denote the $i$-th element of the vector inside the brackets. Thus, $[x]_i$ is the $i$-th element of the vector $x$, $x_m$ is the vector at iteration $m$, and $[x_m]_i$ is the $i$-th entry of the vector $x_m$. Analogous notation will be used when referring to a matrix entry; that is, $[W]_{ij}$ denotes the $(i,j)$ entry of the matrix $W$.

**3.2. TV Reconstruction.** We now focus on approximating the TV regularization operator (1.6), which is often used in image restoration problems. If an image is represented in the discrete setting by a vector $x$, and $D_h$ and $D_v$ denote the finite difference approximations of the horizontal and vertical first derivative operators, respectively, then the TV functional is defined as

$$\mathrm{TV}(x) = \sum_{i=1}^{N} \sqrt{[D_h x]_i^2 + [D_v x]_i^2} = \|\sqrt{d}\|_1 \,, \tag{3.8}$$

where $d = (D_h x)^2 + (D_v x)^2$, and the squaring operation is done element wise. Before approximating the $\ell^1$ norm in (3.8), it is convenient to provide an alternative expression for the vector $d$. Consider the matrix

$$D = \begin{pmatrix} 1 & -1 & & \\ & \ddots & \ddots & \\ & & 1 & -1 \end{pmatrix} \in \mathbb{R}^{(n-1)\times n} \,,$$

which is a scaled finite difference approximation for the one-dimensional derivative. If $x$ is obtained from a discrete image by stacking its columns, then it is not difficult to show that [20]

$$\begin{array}{rcl} D_h x & = & (D \otimes I)x \in \mathbb{R}^{\bar{N}} \\ D_v x & = & (I \otimes D)x \in \mathbb{R}^{\bar{N}} \end{array} \,, \quad \text{where} \quad \bar{N} = (n-1)n = N - n \,.$$

Now, if we set

$$D_{hv} = \begin{bmatrix} D_h \\ D_v \end{bmatrix} \in \mathbb{R}^{2\bar{N}\times N} \tag{3.9}$$

then we obtain

$$\sum_{i=1}^{\bar{N}} [D_h x]_i^2 + [D_v x]_i^2 = \|D_{hv} x\|_2^2 \,.$$

Consider now the diagonal weighting matrix $W = \mathrm{diag}([\widetilde{W}, \widetilde{W}]) \in \mathbb{R}^{2\bar{N}\times 2\bar{N}}$, where

$$\widetilde{W} = \mathrm{diag}\left( \left( \left( [D_h x]_i^2 + [D_v x]_i^2 \right)^{-1/4} \right)_{i=1,\ldots,\bar{N}} \right) \in \mathbb{R}^{\bar{N}\times\bar{N}}.$$

We can verify that the optimal regularization matrix to choose in (1.2) in order to recover the TV regularization operator is $L = W D_{hv} \in \mathbb{R}^{2\bar{N}\times N}$ since

$$\begin{aligned} \|Lx\|_2^2 &= \sum_{i=1}^{\bar{N}} \left( [\widetilde{W}]_{ii}^2 \left( (D_h x)_i^2 + (D_v x)_i^2 \right) \right) \\ &= \sum_{i=1}^{\bar{N}} \left( \left( [D_h x]_i^2 + [D_v x]_i^2 \right)^{-1/2} \left( [D_h x]_i^2 + [D_v x]_i^2 \right) \right) \\ &= \sum_{i=1}^{\bar{N}} \left( [D_h x]_i^2 + [D_v x]_i^2 \right)^{1/2} = \mathrm{TV}(x) \,. \end{aligned}$$

We remark that, so far, all the computations have been carried out assuming that the exact $x$ is known. In real problems we do not know the exact $x$, so we have to exploit the iterative setting of the Arnoldi-Tikhonov method and consider, at step $m$,

$$L_m = W_m D_{hv}, \tag{3.10}$$

where

$$W_m = \mathrm{diag}([\widetilde{W}_m, \widetilde{W}_m])$$

and

$$\widetilde{W}_m = \mathrm{diag}\left(\left([D_h x_{m-1}]_i^2 + [D_v x_{m-1}]_i^2\right)^{-1/4}\right).$$

As before, we denote by $x_{m-1}$ the solution obtained at the end of the previous step of the Arnoldi-Tikhonov method and, to avoid division by zero when both $[D_h x_{m-1}]_i$ and $[D_v x_{m-1}]_i$ are null for some $i = 1, \ldots, \bar{N}$, we set a small threshold $\nu > 0$ and we instead consider

$$\widetilde{W}^{(k)} = \mathrm{diag}\left(g_\nu\left([D_h x_{m-1}]_i^2 + [D_v x_{m-1}]_i^2\right)\right),$$

where

$$g_\nu(\chi) = \begin{cases} \chi^{-1/4} & \text{if } \chi > \nu \\ \nu^{-1/4} & \text{if } \chi \le \nu \end{cases}. \tag{3.11}$$

**4. Flexible-AT Method.** In this section we describe the first strategy to practically implement the Generalized Arnoldi-Tikhonov method in connection with the regularization matrices just defined.

First of all, we recall that, at step $m$, the matrix associated with the least squares problem given by equation (2.3) is made of two blocks: the first one, $\bar{H}_m$, is a Hessenberg matrix of size $(m+1) \times m$ while the second one, $LV_m$, is typically a tall rectangular matrix of size $N \times m$. From a computational point of view this implies that, as long as $m \ll N$ is small, we can solve the regularized least squares problem (2.3) without much effort, for instance computing the so-called "skinny" QR factorization; however, as the number of iterations $m$ increases, solving directly the projected problem (2.3) becomes computationally demanding. On the other hand, if we are able to cheaply transform the original problem (1.2) into standard form, the matrix associated to the least squares problem (2.5) is altogether of dimension $(2m+1) \times m$ and therefore the computational effort to solve the direct problem is no longer very demanding, even when the number of iterations increases.

When we are dealing with the problem described in Section 3.1, the regularization matrices are square and nonsingular. In this case, transformation into standard form is formally rather simple, since it suffices to set $\hat{x} = Lx$, $\hat{x}_0 = Lx_0$, $\hat{A} = AL^{-1}$, solve the problem

$$\min_{\hat{x} \in \mathbb{R}^N} \left\{ \|\hat{A}\hat{x} - b\|_2^2 + \breve{\lambda} \|\hat{x} - \hat{x}_0\|_2^2 \right\}, \tag{4.1}$$

and come back to the solution of the original problem taking $x = L^{-1}\hat{x}$. Computing the inverse of the matrices (3.4) and (3.5) is not an issue because they are diagonal.

In general, when we want to apply the Arnoldi-Tikhonov method to the transformed system (4.1), we build the Krylov subspace $\mathcal{K}_m(AL^{-1}, b)$ and therefore the

inverse of the regularization matrix acts as a right-preconditioner for the original system. However, since we want to adopt the IRLS approach, the matrix $L$ must be updated at each step and, as a consequence, the preconditioner in $\mathcal{K}_m(AL^{-1}, b)$ changes at each iteration. For this reason we must consider particular Krylov subspaces that allow variable preconditioning: the so-called Flexible Krylov subspaces. We remark that, in this framework, the preconditioner is not used to accelerate convergence (as is typically done for well posed problems), but instead it is used to enforce some specific regularity conditions on the reconstructed solution. This interpretation of preconditioning for ill-posed problems is not new; see, for example, [9, 16].

Flexible Krylov subspaces [39] were introduced in various frameworks in order to incorporate an increasingly improved preconditioner into the original Krylov subspace: a typical situation is when the preconditioning matrix is itself obtained by iteratively solving a linear system, even employing a Krylov subspace method. In this case an inner-outer iterative scheme is established since not only the solution of the main linear system, but also the system defining the preconditioning, are updated at each iteration [38].

When we consider the Arnoldi-Tikhonov method with the regularization matrices (3.4), (3.5), (3.10) we are basically operating as described above, since at each iteration we update the regularization matrix exploiting the intermediate solutions computed at the previous steps and in this way we can approximate the optimal regularization matrix (3.3). In the following we explain how the Arnoldi-Tikhonov method can be used in connection with Flexible Krylov subspaces: we call this strategy Flexible-Arnoldi-Tikhonov (Flexi-AT) method. Formally this derivation is very similar to the one described in Section 2.

Let us consider a sequence of problems of the kind (4.1), where $\hat{A} = AL_i^{-1}$ and $L_i$ is a regularization matrix that is updated at each step of the Arnoldi algorithm. At the $m$th step we want to find a solution of the form

$$x_m = x_0 + Z_m \breve{y}_m, \quad \text{where} \quad Z_m = [L_1^{-1}\breve{v}_1, \ldots, L_m^{-1}\breve{v}_m] \in \mathbb{R}^{N\times m}. \tag{4.2}$$

This is equivalent to saying that the vector $x_m - x_0$ is given by a linear combination of the columns of $Z_m$. Again, to compute the vector $\breve{v}_i$ and $z_i = L_i^{-1}\breve{v}_i$ we can consider a procedure similar to the Arnoldi algorithm (cf. [36]), which in matrix form can be written as

$$AZ_m = \breve{V}_{m+1}\breve{H}_m\,, \tag{4.3}$$

where now only the matrix $\breve{V}_{m+1} \in \mathbb{R}^{N\times(m+1)}$, whose first vector is $r_0/\|r_0\|_2$, has orthonormal columns; $\breve{H}_m \in \mathbb{R}^{(m+1)\times m}$ is still upper Hessenberg. We emphasize that the preconditioning is implicitly defined into the columns of $Z_m$. More details about the implementation of this procedure are given in Algorithm 1 at the end of this section. Relation (4.3) leads to the following regularized projected problem that must be solved at each iteration

$$\begin{aligned} \breve{y}_m &= \arg\min_{y\in\mathbb{R}^m} \left\{ \left\| \breve{H}_m y - \|r_0\|_2 e_1 \right\|_2^2 + \breve{\lambda}\,\|y\|_2^2 \right\} \\ &= \arg\min_{y\in\mathbb{R}^m} \left\| \begin{pmatrix} \breve{H}_m \\ \sqrt{\breve{\lambda}} I \end{pmatrix} y - \begin{pmatrix} \|r_0\|_2\, e_1 \\ 0 \end{pmatrix} \right\|_2^2 . \end{aligned} \tag{4.4}$$

To choose the regularization parameter $\breve{\lambda}$ at each step we can apply a strategy analogous to the one described in Section 2, since from the previous scheme we obtain an

expression for the projected discrepancy that is formally equivalent to (2.7). Indeed, thanks to relation (4.3), we can write the projected discrepancy as

$$\breve{\phi}_m(\breve{\lambda}) = \left\| \breve{H}_m \breve{y}_m - \|r_0\|_2 e_1 \right\|_2 \tag{4.5}$$

and consider the approximation

$$\breve{\phi}_m(\breve{\lambda}) = \breve{\phi}_m(0) + \breve{\lambda}\breve{\beta}_m,$$

where $\breve{\phi}_m(0)$ can now be regarded as the norm of the residual associated with the Flexible GMRES (FGMRES, [36]) method and, similarly to (2.10),

$$\breve{\beta}_m = \frac{(\breve{\phi}_m(\breve{\lambda}_{m-1}) - \breve{\phi}_m(0))}{\breve{\lambda}_{m-1}}.$$

Therefore we can update the regularization parameter at the $m$-th iteration of the Flexi-AT method by the following formula:

$$\breve{\lambda}_m = \left| \frac{\eta\varepsilon - \breve{\phi}_m(0)}{\breve{\phi}_m(\breve{\lambda}_{m-1}) - \breve{\phi}_m(0)} \right| \breve{\lambda}_{m-1}. \tag{4.6}$$

We remark that, since the only preconditioners involved when performing the Flexi-AT method are the inverse of the matrices (3.4), we do not have to worry anymore about division by zero and therefore it is not necessary to set a threshold as in (3.6). We summarize this approach in the following

---

ALGORITHM 1
*Flexible Arnoldi-Tikhonov (Flexi-AT) method*

---

Input: $A$, $b$, $x_0$, $\breve{\lambda}_0$, $\varepsilon$, $\eta$
Initialize: $r_0 = b - Ax_0$ and $\breve{v}_1 = r_0/\|r_0\|_2$.
For $i = 1, 2, \ldots$ until $\|b - Ax_i\| \leq \eta\varepsilon$
1. Compute $z_i = L_i^{-1}\breve{v}_i$ and $w = Az_i$.
   Compute $\breve{h}_{j,i} = w^T\breve{v}_j$ for $j = 1, \ldots, i$ and set $w = w - \sum_{j=1}^{i} \breve{h}_{j,i}\breve{v}_j$
2. Compute $\breve{h}_{i+1,i} = \|w\|_2$ and, if $\breve{h}_{i+1,i} \neq 0$, take $\breve{v}_{i+1} = w/\breve{h}_{i+1,i}$.
3. Compute the solution $\breve{y}_i$ of (4.4) with $\breve{\lambda} = \breve{\lambda}_{i-1}$.
4. Compute $\breve{\phi}_i(\breve{\lambda}_{i-1})$ by (4.5).
5. Update the value of $\breve{\lambda}$ by (4.6).

---

In the above algorithm, the first two points are meant to expand the Flexible Krylov subspace at each iteration, while the remaining points are devoted to perform regularization.

**5. The Restarting Strategy.** As said in the previous section, the approach based on the transformation of problem (1.2) into standard form is particularly convenient when the regularization matrix is cheaply invertible: this is not the case when, for instance, we want to approximate the total variation regularization (1.6) as described in Section 3.2. In this section we describe an approach, again based on the Generalized Arnoldi-Tikhonov method, that is alternative to Algorithm 1 and that, beside being still valid to solve problems like (1.4), can be also used in the case (1.6).

This strategy is essentially based on restarting the Arnoldi algorithm: at each restart the regularization matrix $L$, the initial guess $x_0$ and the initial parameter $\lambda_0$ are updated employing the last values computed when the discrepancy principle is satisfied; the first set of iterations is performed simply taking $L = I$, $x_0 = 0$ and $\lambda_0 = 1$. Waiting until the discrepancy principle is satisfied before restarting the Arnoldi algorithm guarantees that the solution employed to update $L$ and $x_0$ is quite accurate even if, especially during the first restarts, we are solving intermediate problems that roughly approximate the original (1.4) and (1.6). We remark that the quantities $L$, $x_0$, $\lambda_0$ depend on the restart we are performing; however, not to overload the notations, we omit the index indicating the number of the performed restarts.

About the number of iterations for each restart, it is well-known that the approximate solutions computed by the Arnoldi-Tikhonov method can very quickly fulfill the discrepancy principle and deliver a regularized solution belonging to a Krylov subspace of dimension $m \ll N$, even for problems of huge dimension [12], [32]. Therefore, considering both the standard form (2.4) and the general form (2.2) problems, the computational cost for each restart is kept low. Furthermore, as the number of restarts increases, the number of iterations required to satisfy the discrepancy principle decreases: this is due to the stable behavior of the GAT method, which after some restarts can compute solutions of similar quality. As a consequence, the regularization matrices $L$ and the Krylov subspaces generated by $A$ and $r_0$ tend to be the same and the discrepancy principle continues to be satisfied; eventually, the discrepancy principle continues to be satisfied after only one step of the Arnoldi-Tikhonov method has been performed, resulting in very cheap computations. In this situation, although the quality of the reconstruction may not be substantially improved, performing additional restarts could still be useful in order to keep updating the regularization matrix with slightly better approximations of the solution and, as a consequence, obtain slightly more accurate reconstructions (cf. Figure 6.7, upper frame). In the following we will denote by $m_D$ the number of iterations required to fulfill the discrepancy principle at every restart; $m_D$ is different at each restart, but not to overload the notations we omit the dependence on the number of restarts.

Determining when to exactly stop the restarts is not a crucial issue, mainly for two reasons: first of all, as the iterations proceed, the behavior of the solution is very stable and, secondly, because the cost of each restart is lower and lower. However we can employ some heuristic to set a stopping criterion: looking at the performed tests (cf. Section 6) it can be noted that, when the discrepancy is satisfied at the end of each restart, the values of the discrepancy function (2.6) keep decreasing and this can be regarded as a sign that we are computing a more accurate solution; therefore, we can decide to stop the iterations after reaching a pre-specified decrease of the discrepancy function. We can also choose to continue the iterations until a fixed maximum number of restarts has been carried out. In the following we will denote by $m_R$ the total number of restarts.

A variant of the approach just described is to restart the Arnoldi algorithm taking always $x_0 = 0$ and exclusively updating the regularization matrix. Although some improvements can be achieved with the restarts, the reconstructions are worse than the ones obtained when updating also $x_0$, and the behavior of the error is extremely non-monotone. Moreover, even if we do many restarts, the number of steps required to fulfill the discrepancy principle, after each restart, is almost constant. We can conclude that taking into account an initial guess both in the formulation (1.2) and in the definition of the Krylov subspaces are beneficial in order to improve the quality

of the solution.

The method described in this section is quite similar to the ones outlined in [31] and [41]; however, as mentioned in Section 1, in these papers the normal equations corresponding to (1.2) are considered and the resulting linear system, whose matrix is symmetric and positive definite, is solved by performing a limited number of Conjugate Gradient iterations.

We remark that, even if the matrix is easily invertible and the Flexible Krylov subspace strategy can be applied, one can also choose not to update the regularization matrix at each step and to rather employ the method described in this section. Provided that we apply the restarting strategy to the standard form problem (2.5) and that, at each restart, we update the matrix $L$ and we build the right-preconditioned Krylov subspaces $\mathcal{K}_m(AL^{-1}, r_0)$, the results are similar to the ones obtained applying the Flexi-AT method. We summarize the approach so far outlined in the following:

---

ALGORITHM 2
*Restarted GAT (ReSt-GAT) method*

---

Input: $A$, $b$, $\varepsilon$, $\eta$
Initialize: $\lambda_0 = 1$, $L = I$ and $x_0 = 0$;
For $k = 1, 2, \ldots, m_R$
1. For $m = 1, \ldots, m_D$
   Compute the decomposition (2.1) and solve (2.3) or (4.1) taking, as inputs, $L$, $\lambda_0$, $x_0$. Employ the relation (2.13) to choose the regularization parameter at each step.
2. Take $\lambda_0 = \lambda_{m_D}$, $x_0 = x_{m_D}$.
3. Define $L$ as (3.4) or (3.10), employing $x_0$.

---

In the above algorithm, the computations in the first step inside the iteration loop aim to expand the Krylov subspace and to solve the regularized least squares problem (2.3). The task of the remaining steps is to set the new quantities employed when restarting the Krylov subspace.

In many applications, for instance in image restoration problems, the solution is known to be nonnegative; however, methods based on Krylov subspaces are not guaranteed to compute nonnegative solutions. Finding a way to force nonnegativity can greatly improve the quality of the approximate solution [6], [29]. In the framework of the algorithm just described we can enforce nonnegativity at each restart, employing an approach very similar to one of those described in [6]. Referring to Algorithm 2, after the first step inside the iteration loop has been completed, and before updating the solution at step 2, we can project $x_{m_D}$ into the set

$$\mathbb{P} = \{x \in \mathbb{R}^N \, : \, [x]_i \geq 0 \, \forall \, i = 1, \ldots, N\} \tag{5.1}$$

of nonnegative vectors. In this way, at each restart we can consider an initial guess that is nonnegative even if, especially during the first restarts, this nonnegative vector is not guaranteed to satisfy the discrepancy principle. However, thanks to the stable behavior of the GAT method, after some restarts we can obtain a solution that is nonnegative *and* which also fulfills the discrepancy principle. We emphasize that our approach does not properly solve a constrained minimization problem whose constraint set is $\mathbb{P}$; it is a rather heuristic approach that forces nonnegative solutions by imposing proximity to a nonnegative initial guess at each iteration. The numerical

results reported in the next section clearly show the improvements obtained applying the strategy just described.

Using MATLAB-like notation, we describe the nonnegative version of Algorithm 2 in the following:

---

ALGORITHM 3
*Nonnegative Restarted GAT (NN-ReSt-GAT) method*

---

Input: $A$, $b$, $\varepsilon$, $\eta$
Initialize: $\lambda_0 = 1$, $L = I$ and $x_0 = 0$;
For $k = 1, 2, \ldots, m_R$
1. For $m = 1, \ldots, m_D$
   Compute the decomposition (2.1) and solve (2.3) or (4.1) taking, as inputs, $L$, $\lambda_0$, $x_0$. Employ the relation (2.13) to choose the regularization parameter at each step.
2. Project $x_{m_D}$ into the space $\mathbb{P}$ by taking $x_{m_D}(x_{m_D} < 0) = 0$.
3. Take $\lambda_0 = \lambda_{m_D}$, $x_0 = x_{m_D}$.
4. Define $L$ as (3.4) or (3.10), employing $x_0$.

---

**6. Numerical Experiments.** In this section we show the results of some numerical tests that contribute to validate the strategies described in this paper. Along with the reconstruction obtained using the new algorithms, we present some comparisons with other well-known methods to recover sparse solutions or to perform total variation regularization. All the tests were performed using MATLAB 7.10 (double precision) on a single processor, Intel Core i3-550, computer.

**Example 1.** For the first test we focus on sparse reconstruction and we take, as a test image, a synthetic astronomical image of size $256 \times 256$ pixels, characterized by a very sparse pattern: only the 0.7% of its elements corresponds to non-black pixels, i.e. can be considered different from zero (cf. Figure 6.3, upper left frame). We assume that the available image is corrupted by a spatially variant blur and it is divided into 25 different regions: the point spread function (PSF) is spatially invariant in each region. Gaussian white noise is added and we consider two successive noise levels: the first one is equal to $10^{-2}$ (cf. Figure 6.3, upper right frame), the second one is equal to $10^{-1}$. We refer to [27] for background on the solution of this kind of problem. Further information on this test problem, as well as the associated data, can be obtained from the MATLAB package *Restore Tools* [28].

In Figure 6.1 we plot the values of the relative error, the discrepancy function, and the regularization parameter (all displayed in logarithmic scale) versus the number of iterations. These results are obtained applying Algorithm 1; the noise level in the data is $\widetilde{\varepsilon} = 10^{-2}$. As said in Section 2, we choose as starting value for the regularization parameter $\lambda_0 = 1$ and, to define the discrepancy principle (2.8), we take the scalar $\eta = 1.01$. The value of the truncation parameter in (3.7) is set to $\tau = 10^{-8}$. In this case the stopping criterion determines an approximate solution that belongs to a Flexible Krylov subspace of dimension 23. We can note that, for this problem, all the quantities in Figure 6.1 exhibit a quite stable behavior after the discrepancy principle (and the stopping criterion) is satisfied: in particular we can observe that the relative error does not deteriorate as the iterations proceed. This feature is typical of the AT methods applied to ill-posed problems since, after the first iterations, the largest

singular values of the Hessenberg matrices $\bar{H}_m$ (2.1) and $\breve{H}_m$ (4.3) approximate the largest singular values of the original matrix $A$; by solving the projected problem by means of Tikhonov regularization (which can also be regarded as a spectral filtering method [14]) the computed quantities essentially depend on the decay of the singular values of the Hessenberg matrices, and therefore they tend to have a similar behavior after a certain number of iterations have been completed.

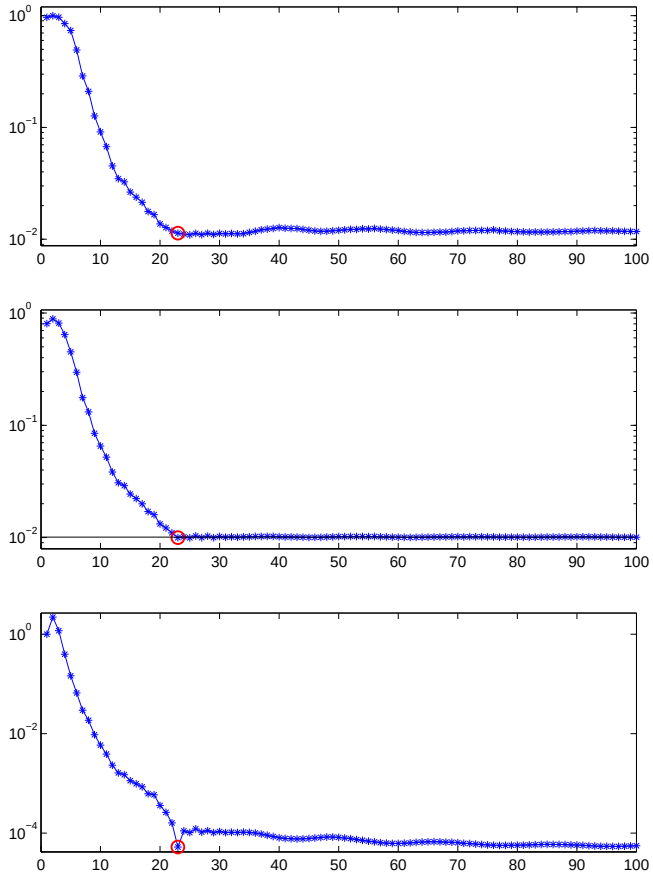



FIG. 6.1. *Values of the relative error (upper frame), the discrepancy function (middle frame) and the regularization parameter (lower frame) versus the number of iterations. These plots are obtained applying Algorithm 1 to the first test problem with* $\widetilde{\varepsilon} = 10^{-2}$. *The circle at the iteration 23 highlights the quantities obtained when the discrepancy principle is satisfied; we continue to iteration 100 to illustrate the stable behavior of our algorithms. The horizontal line in the second frame marks the threshold under which the discrepancy principle is satisfied.*

In Figure 6.2 we display the history of the relative errors obtained when considering different versions of Algorithm 2, including the nonnegative one described in Algorithm 3. The test problem and the parameters are as the ones above specified; in this example we consider 20 restarts. It is interesting to note that, when we restart the right-preconditioned Arnoldi scheme with an initial guess that has been projected into the space $\mathbb{P}$ (see equation (5.1)) we typically need slightly more iterations than when taking as initial guess the last computed intermediate solution. In particular, in this experiment, the number of inner iterations during the very first restart is dif-

ferent. Indeed, when we restart for the first time modifying the initial guess, the newly generated Krylov subspace can be pretty different to the old one and therefore we need some additional iterations to satisfy the discrepancy principle. We can also remark that, when restarts are performed taking simply $x_0 = 0$, the same number of iterations are needed at each restart and the quality of the computed solution does not significantly improve with each restart. In Figure 6.3 we show the images obtained applying the methods just considered.

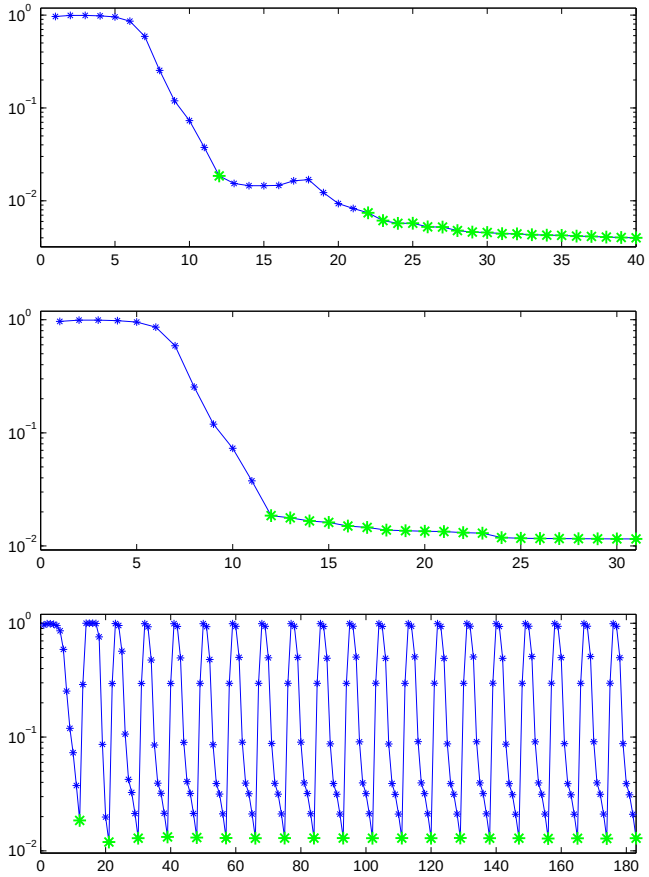



FIG. 6.2. *Behavior of the relative error versus the number of iterations obtained applying three variants of the restarting strategy. In the upper frame we force nonnegativity at each restarts (Algorithm 3), in the middle frame we take as initial guess the last solution computed at the end of the previous restart (Algorithm 2) and in the lower frame we take as initial guess* $x_0 = 0$ *at each restart. The bigger asterisks highlight the iteration at which a restart happens.*

We now make some comparisons with other well-established and recently designed methods. In particular we focus on NESTA [1], SpaRSA [42], TwIST [2] and `l1_ls` [22]: these methods can efficiently handle a wide class of minimization problems whose objective function is the sum of a fit-to-data term and a regularization term. We also consider the method IRN-BPDN described in [34], which employs cyclically updated weighting matrices and therefore is quite close to the algorithms described in this paper. Most of the methods just mentioned basically require the user to set a suitable value for the regularization parameter: looking at the graph displayed in

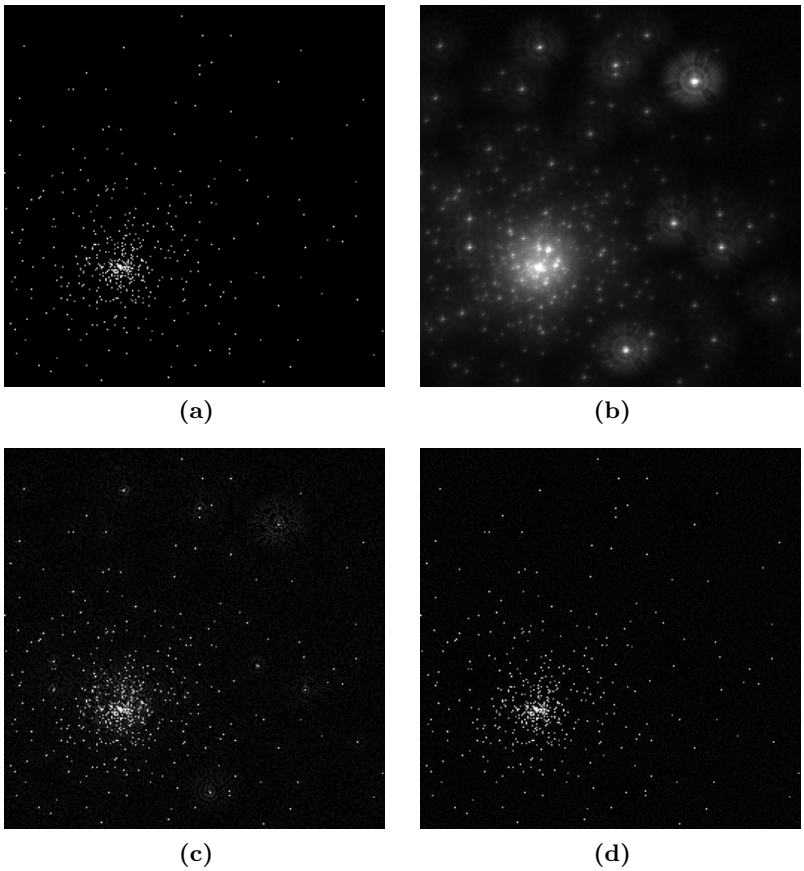



FIG. 6.3. *Images considered in the first example:* ***(a)*** *exact image;* ***(b)*** *blurred and noisy image, with* $\widetilde{\varepsilon} = 10^{-2}$*;* ***(c)*** *solution obtained at iteration 23 of Algorithm 1, i.e. when the discrepancy principle is satisfied;* ***(d)*** *solution obtained at the end of Algorithm 3, after 20 restarts have been performed.*

the lower frame of Figure 6.1 we can assume that a good value for this parameter relative to the problem at hand should be $\lambda = 10^{-4}$ because, going on with the number of iterations $m$, the $\lambda_m$ values stabilize around this point. In addition, we also consider the performance of the standard Arnoldi-Tikhonov method [12], and of its range-restricted version RR-AT [24]. In Figure 6.4 we show the behavior of the relative errors versus the number of iterations for many of the methods cited above; in Table 6.1 we report the value of the relative errors obtained when the stopping criterion of each method is fulfilled or when a maximum number of iterations has been performed, along with the total and average (per iteration) running time. In order to keep the comparisons fair, we decide to stop the iterations as soon as the relative change of the error drops below a certain threshold; moreover, we basically use the published version of each method along with the pre-specified parameters (except for $\lambda$): therefore, a more accurate tuning of all the parameters can possibly result in a better performance of some of the methods. Looking at the results displayed in Table 6.1 we can state that, for this example, the newly proposed algorithms exhibit excellent performance both in terms of quality of the results and computational time. The primary reason is that, adopting a Krylov subspace approach, at each iteration we

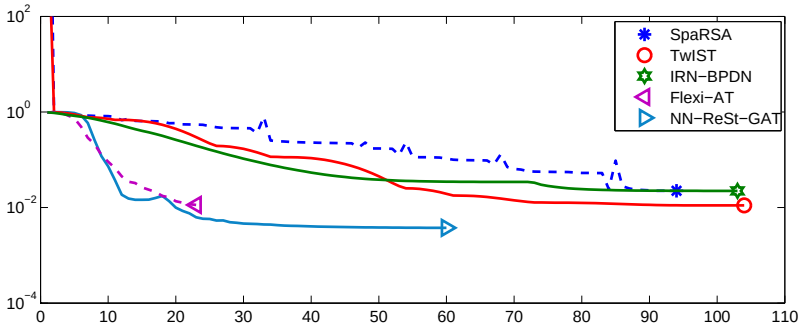



FIG. 6.4. *Behavior of the relative error versus the number of iterations for different optimization methods. The last computed value is highlighted with a different marker and corresponds to the one reported in Table 6.1..*

| Method | Relative Error | Iterations | Total Time | Average Time |
|---|---|---|---|---|
| SpaRSA | $2.2365 \cdot 10^{-2}$ | 94 | 24.76 | 0.26 |
| NESTA | $1.7800 \cdot 10^{-2}$ | 248 | 306.17 | 1.23 |
| TwIST | $1.1089 \cdot 10^{-2}$ | 104 | 28.02 | 0.27 |
| `l1_ls` | $2.2257 \cdot 10^{-2}$ | 298 | 683.55 | 2.29 |
| IRN-BPDN | $2.2294 \cdot 10^{-2}$ | 103 | 35.72 | 0.35 |
| AT | $1.8512 \cdot 10^{-2}$ | 12 | 0.91 | 0.08 |
| RR-AT | $1.9171 \cdot 10^{-2}$ | 18 | 3.77 | 0.21 |
| Flexi-AT | $1.1345 \cdot 10^{-2}$ | 23 | 2.44 | 0.11 |
| ReSt-GAT | $1.1033 \cdot 10^{-2}$ | 51 | 5.95 | 0.12 |
| NN-ReSt-GAT | $3.7530 \cdot 10^{-3}$ | 60 | 6.25 | 0.10 |

TABLE 6.1

*Comparisons of the performances of some algorithms developed to solve problem (1.4). The Relative Error reported is the one computed at the iteration displayed in the third column. The number of iterations is the minimum between the iterations required to fulfill the stopping criterion and the maximum number of allowed iterations. Both the Total Time and the Average Time are expressed in seconds.*

deal with projected quantities and therefore all the main computations are executed in reduced dimension. However some of the considered methods, such as NESTA and SpaRSA, can deal with much more general minimization problems - for instance involving a nonlinear fit-to-data functional: in this situation the algorithms described in the present paper cannot be straightforwardly applied.

Finally, in Figure 6.5 we show the same quantities displayed in Figure 6.1, but this time the noise level is $\widetilde{\varepsilon} = 10^{-1}$.

**Example 2.** We now consider another example regarding an image restoration problem and we apply the restarted GAT method in order to approximate the total variation regularization. As a test image we take a computer simulation of how a satellite can be detected by ground based telescopes; in this case the PSF is spatially invariant and models an atmospheric blur. This image is of size $256 \times 256$ pixels and we further corrupt the blurred image adding white noise, in such a way that the noise level is $\widetilde{\varepsilon} = 10^{-2}$; as in the previous example, the safety factor for the discrepancy principle is $\eta = 1.01$ and the threshold considered in (3.11) is $\nu = 10^{-8}$. Both the exact image and the PSF are available in the MATLAB package *Restore Tools* [28].

In Figure 6.6 we show the exact image, the blurred and noisy one and the reconstruction obtained applying Algorithm 3 to enforce nonnegativity: at each restart the

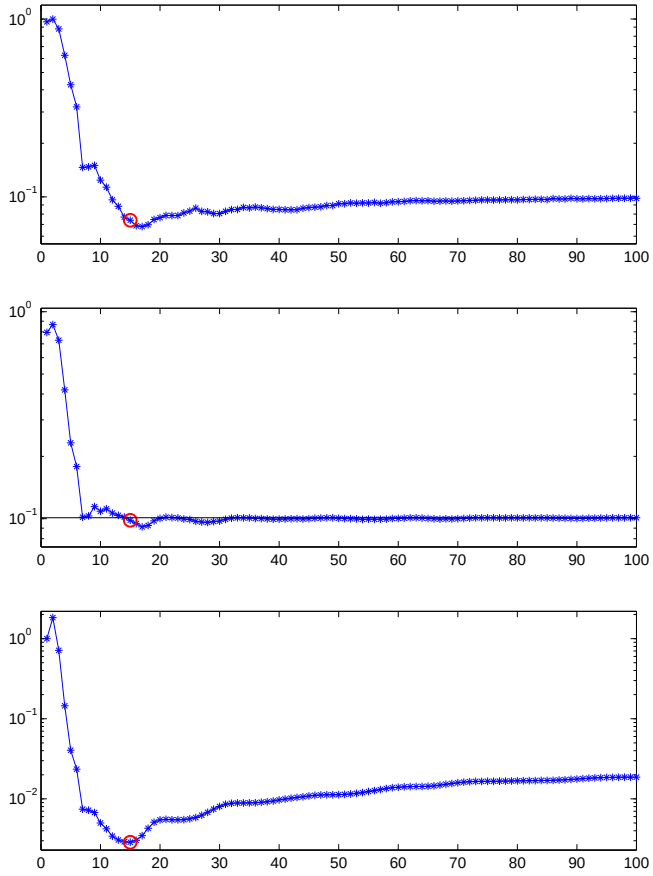
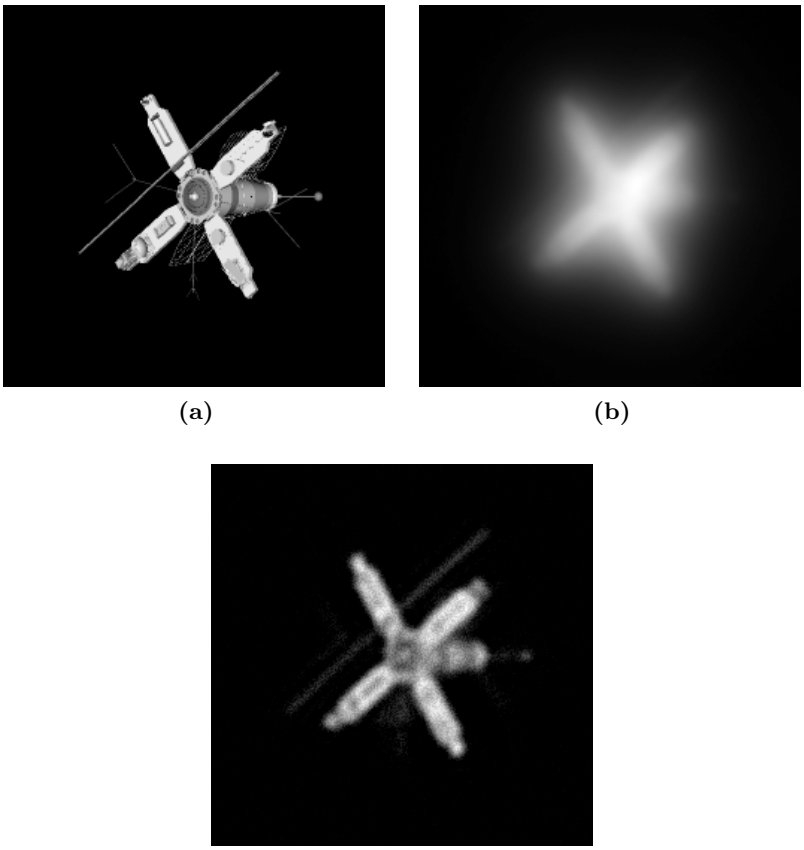



FIG. 6.5. *Values of the relative error, the discrepancy function and the regularization parameter versus the number of iterations. These plots are obtained applying Algorithm 1 to the first test problem with* $\widetilde{\varepsilon} = 10^{-1}$. *The circle at the 15th iteration highlights the quantities obtained when the discrepancy principle is satisfied.*

regularization matrix, defined by (3.10), is updated and the intermediate problems are solved using the Generalized Arnoldi Tikhonov method (2.3), since in this case the matrix (3.10) is not easily invertible. The fundamental difference between this test and the ones so far described is that in this case the solution belongs to the Krylov subspace $\mathcal{K}_m(A, b)$ defined taking into account exclusively the matrix $A$ while in the previous examples it belonged to the right-preconditioned Krylov subspace $\mathcal{K}_m(A(L)^{-1}, b)$, where the matrix $L$ was updated at each restart. We perform 200 restarts: this choice is supported by the fact that the discrepancy principle is satisfied after 9 iterations at the beginning, 3 iterations after the first restart and immediately, i.e. after just 1 iteration, in the following restarts. In this way the computational cost of each restart is very low. Moreover, evaluating the error, we see that it is always slightly decreasing.

Finally we examine the performance of the NN-ReSt-GAT Algorithm with respect to some other regularization methods. The first set of comparisons involves the standard Arnoldi-Tikhonov method and its generalized version applied with the



**(a)** **(b)**

**(c)**

FIG. 6.6. *Images considered in the second example:* ***(a)*** *exact image;* ***(b)*** *blurred and noisy image, with* $\widetilde{\varepsilon} = 10^{-2}$*;* ***(c)*** *reconstruction obtained applying Algorithm 3 to approximate total variation regularization, after 200 restarts are performed.*

fixed regularization matrix (3.9). In Figure 6.7 we display the behavior of the relative errors, of the discrepancy function and of the regularization parameter versus the number of iterations: looking at the upper frame we can clearly see that the approach based on total variation regularization can deliver better results than the other ones (this is not unexpected for image restoration problems, cf. [31]). It is also interesting to remark that, for this particular test problem, the standard Tikhonov method slightly outperforms the GAT method.

The second set of comparisons involves three methods that have been designed to iteratively deal with total variation regularization and that are closely related to the algorithms described in this paper, since they both adopt an IRLS procedure to linearly approximate the TV functional. The first one is the Adaptive Majorization-Minimization approach to total variation described in [31] (in the following we will refer to it as aMM-TV): this method is adaptive in the sense that a parameter selection strategy based on Bayesian considerations is derived. The second one is the algorithm IRN-TV derived in [33]: although in [25] the authors propose a strategy, based on statistical considerations, to automatically set the regularization parameter, we assign

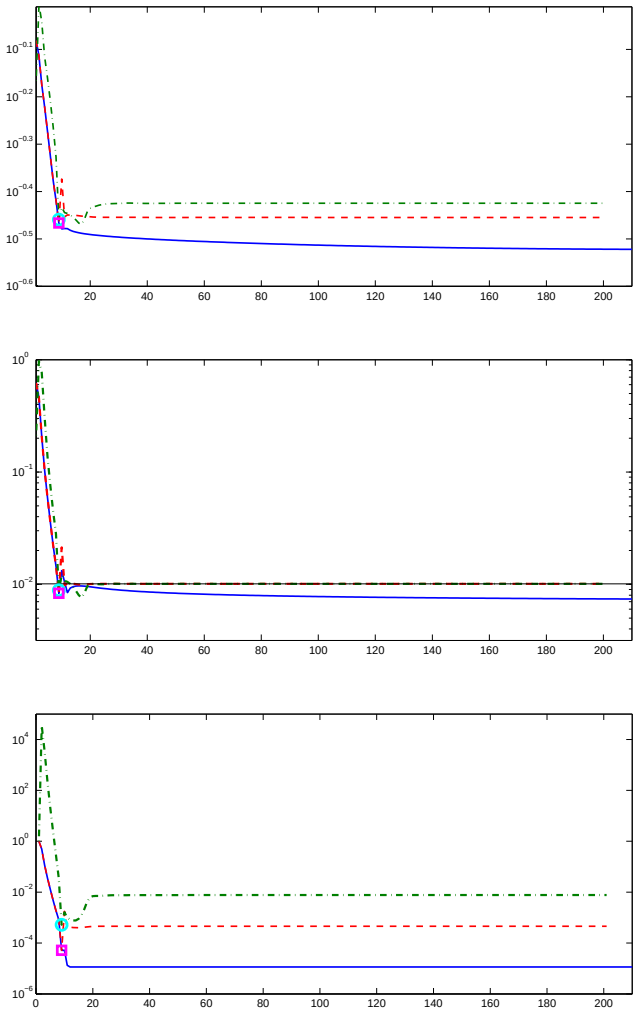


FIG. 6.7. *Comparison of the relative errors, the values of the discrepancy function and the values of the regularization parameter obtained when restoring the image of Example 2 with different methods. The solid line refers to Algorithm 2, the dashed line refers to the standard Arnoldi-Tikhonov method, and the dash-dot line refers to the Generalized Arnoldi-Tikhonov method with the regularization matrix $D_{hv}$ defined in (3.9). For both the AT and the GAT method the discrepancy principle is satisfied at the 9th iteration and in all the plots we mark it with a square and a circle, respectively.*

to it a fixed value: looking at the plot in the lower frame of Figure 6.7, we choose $\lambda = 10^{-5}$. Moreover, even if the authors do not seem to suggest it, we take as initial guess at each restart the last computed approximation. The third one is NESTA [1] and we still consider, as regularization parameter, $\lambda = 10^{-5}$.

In Figure 6.8 we display the history of the relative errors for the aMM-TV, IRN-TV, NN-ReSt-GAT, ReSt-GAT and NESTA methods. Some comparisons regarding all the above listed methods used to solve the present test problem are summarized in Table 6.2, whose layout is identical to that of Table 6.1. Looking at Table 6.2 we

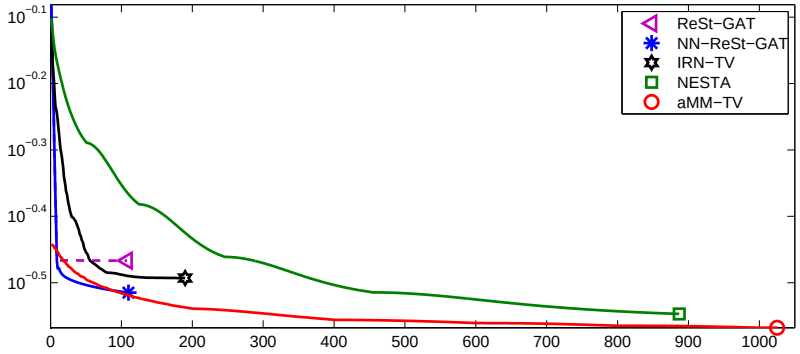



FIG. 6.8. *Behavior of the relative error versus the number of iterations for the NN-ReSt-GAT, ReSt-GAT, aMM-TV, NESTA and IRN-TV algorithms. The last computed value is highlighted with a different marker and corresponds to the one reported in Table 6.2.*

| Method | Relative Error | Iterations | Total Time | Average Time |
|---|---|---|---|---|
| aMM-TV | $2.7056 \cdot 10^{-1}$ | 1025 | 2159.35 | 2.10 |
| IRN-TV | $3.2141 \cdot 10^{-1}$ | 190 | 14.67 | 0.08 |
| NESTA | $2.8382 \cdot 10^{-1}$ | 887 | 69.57 | 0.08 |
| ReSt-GAT | $3.4138 \cdot 10^{-1}$ | 108 | 12.87 | 0.12 |
| NN-ReSt-TV | $3.0556 \cdot 10^{-1}$ | 110 | 13.37 | 0.12 |
| AT | $3.4176 \cdot 10^{-1}$ | 9 | 0.34 | 0.04 |
| GAT | $3.4809 \cdot 10^{-1}$ | 9 | 0.70 | 0.08 |
| RR-AT | $3.5321 \cdot 10^{-1}$ | 14 | 1.39 | 0.10 |

TABLE 6.2

*Comparisons of the performances of some algorithms developed to solve the problem (1.6). The Relative Error reported is the one computed at the iteration displayed in the third column. The number of iterations is that required to fulfill the stopping criterion or a fixed maximum number of allowed iterations. Both the Total Time and the Average Time are expressed in seconds.*

can state that the aMM-TV algorithm surely is the best one in terms of quality of the reconstruction, but it is also the most expensive one: indeed the parameter choice strategy proposed in [31] requires the method to perform a lot of iterations and a lot of restarts before determining a reasonable value for the regularization parameter, resulting in an overall slow convergence. We further remark that, among all the algorithms listed in the above table, NN-ReSt-GAT is the only one that produces nonnegative solutions.

**7. Concluding Remarks.** In this paper we have proposed two new methods to solve linear ill-posed problems by employing the Generalized Arnoldi-Tikhonov approach and approximating the $\ell^1$ and the total variation regularization terms by means of a weighted $\ell^2$ norm. Since these two strategies are based on Krylov subspace techniques, they are particularly efficient when dealing with large-scale problems. At each iteration the regularization parameter is chosen by adopting a simple scheme based on the discrepancy principle. A comparison with some well-established methods for $\ell^1$ and TV regularization shows that the new strategies are computationally much cheaper, especially for what concerns the parameter choice. Some experiments regarding image restoration problems attest that the quality of the reconstructions achieved applying the new approaches is comparable to other well-known methods. Therefore we believe that these strategies can be regarded as valid alternatives to deal

with sparse reconstruction problems.

## REFERENCES


[1] S. Becker, J. Bobin, and E. Candès. NESTA: A Fast and Accurate First-Order Method for Sparse Recovery. *SIAM J. Imaging Sciences*, 4(1):1–39, 2011.

[2] J. Bioucas-Dias and M. Figueiredo. A new TwIST: two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Proc.*, 16(12):2992–3004, 2007.

[3] Å. Björck. A bidiagonalization algorithm for solving large and sparse ill-posed systems of linear equations. *BIT*, 28(3):659–670, 1988.

[4] Å. Björck. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, PA, 1996.

[5] Å. Björck, E. Grimme, and P. van Dooren. An implicit shift bidiagonalization algorithm for ill-posed systems. *BIT*, 34(4):510–534, 1994.

[6] D. Calvetti, G. Landi, L. Reichel, and F. Sgallari. Non-negativity and iterative methods for ill-posed problems. *Inverse Problems*, 20:1747–1758, 2004.

[7] 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.*, 123:423–446, 2000.

[8] D. Calvetti and L. Reichel. Tikhonov regularization of large linear problems. *BIT*, 43(2):263–283, 2003.

[9] D. Calvetti, L. Reichel, and A. Shuibi. Invertible smoothing preconditioners for linear discrete ill-posed problems. *Appl. Numer. Math.*, 54:135–149, 2005.

[10] J. Chung, J. G. Nagy, and D. P. O'Leary. A weighted GCV method for Lanczos hybrid regularization. *Elec. Trans. Numer. Anal.*, 28:149–167, 2008.

[11] H. W. Engl, M. Hanke, and A. Neubauer. *Regularization of Inverse Problems*. Kluwer Academic Publishers, Dordrecht, 2000.

[12] S. Gazzola and P. Novati. Automatic parameter setting for Arnoldi-Tikhonov methods. *J. Comput. Appl. Math.*, 256:180–195, 2014.

[13] M. Hanke. On Lanczos based methods for the regularization of discrete ill-posed problems. *BIT*, 41(5):1008–1018, 2001.

[14] P. C. Hansen. *Rank-deficient and discrete ill-posed problems*. SIAM, Philadelphia, PA, 1997.

[15] P. C. Hansen. *Discrete Inverse Problems: Insight and Algorithms*. SIAM, Philadelphia, PA, 2010.

[16] P. C. Hansen and T. K. Jensen. Smoothing-norm preconditioning for regularizing minimum-residual methods. *SIAM J. Matrix Anal. Appl.*, 29:1–14, 2006.

[17] K. Hayami, J. F. Yin, and T. Ito. GMRES methods for least squares problems. *SIAM J. Matrix Anal. Appl.*, 31(5):2400–2430, 2010.

[18] M. Hochstenbach and L. Reichel. An iterative method for Tikhonov regularization with a general linear regularization operator. *J. Integral Equations Appl.*, 22:463–480, 2010.

[19] D. Hunter and K. Lange. A tutorial on MM algorithms. *The American Statistician*, 58:30–37, 2004.

[20] M. E. Kilmer, P. C. Hansen, and M. I. Español. A projection-based approach to general-form Tikhonov regularization. *SIAM J. Sci. Comput.*, 29(1):315–330, 2007.

[21] M. E. Kilmer and D.P. O'Leary. Choosing regularization parameters in iterative methods for ill-posed problems. *SIAM J. Matrix Anal. Appl.*, 22(4):1204–1221, 2001.

[22] S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinvesky. An interior-point method for large-scale $\ell_1$-regularized least squares. *IEEE J. Selected Topics in Image Processing*, 1:606–617, 2007.

[23] R. M. Larsen. *Lanczos Bidiagonalization with Partial Reorthogonalization*. PhD thesis, Dept. of Computer Science, University of Aarhus, Denmark, 1998.

[24] B. Lewis and L. Reichel. Arnoldi-tikhonov regularization methods. *J. Comput. Appl. Math.*, 226:92–102, 2009.

[25] Y. Lin, B. Wohlberg, and H. Guo. UPRE method for total variation parameter selection. *Signal Processing*, 90:2546–2551, 2010.

[26] J. L. Mueller and S. Siltanen. *Linear and Nonlinear Inverse Problems with Practical Applications*. SIAM, Philadelphia, PA, 2012.

[27] J. G. Nagy and D. P. O'Leary. Restoring images degraded by spatially-variant blur. *SIAM J. Sci. Comput.*, 19:1063–1082, 1998.

[28] J. G. Nagy, K. M. Palmer, and L. Perrone. Iterative methods for image deblurring: A Matlab object oriented approach. *Numerical Algorithms*, 36:73–93, 2004.
See also: `http://www.mathcs.emory.edu/~nagy/RestoreTools`.

[29] J. G. Nagy and Z. Strakoš. Enforcing nonnegativity in image reconstruction algorithms. In

D. C. Wilson, et. al., editor, *Mathematical Modeling, Estimation, and Imaging*, volume 3461, pages 182–190. SPIE, 2000.

[30] D. P. O'Leary and J. A. Simmons. A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems. *SIAM J. Sci. Stat. Comp.*, 2(4):474–489, 1981.

[31] J. P. Oliveira, J. M. Bioucas-Dias, and M. A. T. Figueiredo. Adaptive total vatiation image deblurring: A majorization-minimization approach. *Signal Processing*, 89:1683–1693, 2009.

[32] L. Reichel, F. Sgallari, and Q. Ye. Tikhonov regularization based on generalized Krylov subspace methods. *Appl. Numer. Math.*, 62:1215–1228, 2012.

[33] P. Rodríguez and B. Wohlberg. An iteratively reweighted norm algorithm for total variation regularization. In *Proceedings of the 40th Asilomar Conference on Signals, Systems and Computers (ACSSC)*, 2006.

[34] P. Rodríguez and B. Wohlberg. An efficient algorithm for sparse representations with $\ell^p$ data fidelity term. In *Proceedings of 4th IEEE Andean Technical Conference (ANDESCON)*, 2008.

[35] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[36] Y. Saad. A flexible inner-outer preconditioned GMRES algorithm. *SIAM J. Sci. Comput.*, 14(2):461–469, 1993.

[37] Y. Saad. *Iterative Methods for Sparse Linear Systems, 2nd Ed.* SIAM, Philadelphia, 2003.

[38] V. Simoncini and D. B. Szyld. Flexible inner-outer Krylov subspace methods. *SIAM J. Numer. Anal.*, 40(6):2219–2239, 2006.

[39] V. Simoncini and D. B. Szyld. Recent computational developments in Krylov subspace methods for linear systems. *Numer. Linear Algebra Appl.*, 14:1–59, 2007.

[40] C. R. Vogel. *Computational Methods for Inverse Problems.* SIAM, Philadelphia, PA, 2002.

[41] B. Wohlberg and P. Rodríguez. An iteratively reweighted norm algorithm for minimization of total variation functionals. *IEEE Signal Processing Letters*, 14:948–951, 2007.

[42] S. J. Wright, M. A. T. Figueiredo, and R. D. Nowak. Sparse Reconstruction by Separable Approximation. *IEEE Trans. Signal Processing*, 57(7):2479–2493, 2009.