SIAM J. SCI. COMPUT.
Vol. 41, No. 5, pp. S149–S171

FLEXIBLE KRYLOV METHODS FOR ℓ_p REGULARIZATION*

JULIANNE CHUNG† AND SILVIA GAZZOLA‡

Abstract. In this paper we develop flexible Krylov methods for efficiently computing regularized solutions to large-scale linear inverse problems with an ℓ_2 fit-to-data term and an ℓ_p penalization term, for $p \geq 1$. First we approximate the p-norm penalization term as a sequence of 2-norm penalization terms using adaptive regularization matrices in an iterative reweighted norm fashion, and then we exploit flexible preconditioning techniques to efficiently incorporate the weight updates. To handle general (nonsquare) ℓ_p-regularized least-squares problems, we introduce a flexible Golub–Kahan approach and exploit it within a Krylov–Tikhonov hybrid framework. Furthermore, we show that both the flexible Golub–Kahan and the flexible Arnoldi approaches for $p = 1$ can be used to efficiently compute solutions that are sparse with respect to some transformations. The key benefits of our approach compared to existing optimization methods for ℓ_p regularization are that inner-outer iteration schemes are replaced by efficient projection methods on linear subspaces of increasing dimension and that expensive regularization parameter selection techniques can be avoided. Theoretical insights are provided, and numerical results from image deblurring and tomographic reconstruction illustrate the benefits of this approach, compared to well-established methods.

Key words. ℓ_p regularization, sparsity reconstruction, iterative reweighted norm, flexible Golub–Kahan, hybrid regularization, image deblurring

AMS subject classifications. 65F20, 65F22, 65F30

DOI. 10.1137/18M1194456

1. Introduction.

Inverse problems are prevalent in many important applications, ranging from biomedical to geophysical imaging, and solutions must be computed reliably and efficiently. In this work we consider discretized linear inverse problems of the form

$$\mathbf{b} = \mathbf{A}\mathbf{x}_{\text{true}} + \mathbf{e}\,, \tag{1}$$

where $\mathbf{b} \in \mathbb{R}^m$ is the observed data, $\mathbf{A} \in \mathbb{R}^{m\times n}$ is the ill-conditioned matrix that models the forward process, $\mathbf{x}_{\text{true}} \in \mathbb{R}^n$ is the desired solution, and $\mathbf{e} \in \mathbb{R}^m$ is the noise or perturbation affecting the observation. Due to the ill-posedness of the underlying problem, in order to recover a meaningful approximation of $\mathbf{x}_{\text{true}}$ in (1), some regularization is applied, i.e., problem (1) is replaced by a closely related one that is stable with respect to the corrupted data [18]. In this paper, we are interested in regularized problems of the form

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{\Psi}\mathbf{x}\|_p^p\,, \tag{2}$$

where $\|\cdot\|_p$ for $p \geq 1$ is the vectorial p-norm, $\lambda > 0$ is a regularization parameter, and $\mathbf{\Psi} \in \mathbb{R}^{n\times n}$ is a nonsingular matrix. For $p = 2$ and $\mathbf{\Psi} = \mathbf{I}$, (2) is the standard

*Received by the editors June 15, 2018; accepted for publication (in revised form) April 2, 2019; published electronically October 29, 2019.
https://doi.org/10.1137/18M1194456
Funding: This work was supported by EPSRC through grant numbers EP/K032208/1 and EP/R014604/1. The work of the first author was supported by NSF through grant numbers DMS 1654175 and DMS 1723005.

†Department of Mathematics, Computational Modeling and Data Analytics Division, Academy of Integrated Science, Virginia Tech, Blacksburg, VA (jmchung@vt.edu, http://www.math.vt.edu/people/jmchung/).
‡Department of Mathematical Sciences, University of Bath, United Kingdom (S.Gazzola@bath.ac.uk, http://people.bath.ac.uk/sg968/).

Tikhonov regularization problem, and many efficient techniques including hybrid iterative methods have been proposed (see, e.g., [5, 11, 22, 28]). However, optimization problems (2) for $p \neq 2$ can be significantly more challenging. For example, for $p = 1$, the so-called ℓ_1-regularized problem suffers from nondifferentiability at the origin; moreover, in some situations, one may wish to consider $0 < p < 1$, which results in a nonconvex optimization problem (see, e.g., [20, 24, 25]). In this paper, we will develop methods to compute an approximate solution to (2) for the case $p \geq 1$, where a unique solution exists. Henceforth, we will refer to problem (2) with $\mathbf{\Psi} = \mathbf{I}$ as the "ℓ_p-regularized" problem; problem (2) with $\mathbf{\Psi} \neq \mathbf{I}$ will be dubbed the "transformed ℓ_p-regularized" problem. Typically the transformed ℓ_p-regularized problem arises in cases where sparsity in some frequency domain (e.g., in a wavelet domain) is desired. Depending on the application, a sparsity transform may be included in both the fit-to-data and the regularization term. This was considered in [32], where the resulting minimization problem was solved with an inner-outer iteration scheme.

Most of the previously developed methods for ℓ_p minimization utilize nonlinear optimization schemes or iteratively reweighted optimization schemes, which can get very expensive due to inner-outer iterations [1, 16, 31, 32, 41]. Other popular approaches such as the split Bregman method [14], separable approximations [42], and accelerations of the iterative shrinkage thresholding algorithms [2] are fast alternatives, but a main disadvantage is that the regularization parameter must be selected a priori, which can be a difficult task. Krylov methods, on the other hand, have nice convergence and regularization properties, so there have been recent efforts to exploit Krylov methods to solve the ℓ_p-regularized problem, possibly without resorting to inner-outer iterations. For example, [20, 24] considered generalized Krylov methods for $\ell_p - \ell_q$ minimization, and Krylov methods based on the flexible Arnoldi algorithm were considered in [10, 35, 36]. Our proposed methods are mostly related to the latter approaches, which compute approximate solutions to the ℓ_p-regularized problem when $\mathbf{A}$ is square. Below we outline the main distinctions and contributions of our work.

In this paper, we propose new iterative hybrid methods based on a flexible Golub–Kahan decomposition to solve ℓ_p-regularized problems (2), where flexible preconditioning techniques are used to build appropriate solution subspaces. In particular, we describe two methods, namely, flexible LSQR and flexible LSMR, and show how Tikhonov regularization can be used to solve the projected problem, where the properties of the matrices associated with the flexible Golub–Kahan decomposition are exploited for efficient regularization parameter selection (in a hybrid fashion). We underline that methods based on the flexible Golub–Kahan algorithm can be implemented without explicitly constructing the matrix $\mathbf{A}$, i.e., by treating $\mathbf{A}$ and $\mathbf{A}^\top$ as linear operators acting on vectors. Furthermore, we describe a way to incorporate regularization terms expressed as the p-norm of the transformed solution within the flexible schemes (based on both the Arnoldi and the Golub–Kahan decompositions), i.e., to deal with the transformed ℓ_p-regularized problem.

One of the first major contributions, compared to [10], is that our methods can be used to solve problems with general (e.g., nonsquare) coefficient matrix $\mathbf{A}$. Second, we provide theoretical results that show optimality properties for the flexible approaches, and we prove that in exact arithmetic flexible LSMR iterates are the same as flexible GMRES iterates on the normal equations. Third, contrary to classical Krylov–Tikhonov methods [11], which can handle penalization terms evaluated in the 2-norm, the new methods can approximate penalization terms evaluated in the sparsity-inducing 1-norm and can include an invertible sparsity transformation,

which generalizes the flexible Arnoldi decomposition proposed in [10], as well as the flexible Golub–Kahan decomposition derived in this paper. Numerical comparisons to well-established ℓ_1 regularization methods reveal that the proposed strategies provide an easy-to-use approach for computing reconstructions with similar properties, but with two significant benefits: first, the regularization parameters can be selected automatically thanks to the hybrid framework; second, information from the current solution is incorporated via the regularization into the solution process as soon as it becomes available, with potentially great computational savings compared to methods involving inner-outer iterations.

The paper is organized as follows. In section 2 we review the ideas underlying the IRN approach for ℓ_p regularization and briefly review the flexible Arnoldi–Tikhonov approach. In section 3 we derive the flexible Golub–Kahan decomposition, leading to the introduction of the new flexible LSQR and flexible LSMR algorithms; hybrid approaches based on flexible LSQR and flexible LSMR are addressed, with a particular emphasis on the choice of regularization term and regularization parameter. Optimality properties for the new solvers and links to existing solvers are provided. In section 4 we describe how a sparsity transform can be handled within hybrid schemes based on the flexible Arnoldi and Golub–Kahan algorithms, and we investigate how the solution subspaces are modified by incorporating reweightings and sparsity transforms. Numerical results are presented in section 5, and conclusions and future work are provided in section 6.

2. Background on iteratively reweighted and flexible methods for ℓ_p regularization. A well-established strategy for solving the ℓ_p-regularized problem is the IRN algorithm [16, 32]. This approach requires solving a sequence of reweighted, penalized least-squares problems where the weights change at each iteration. When dealing with large systems, each least-squares problem is solved by an iterative method, so that an inner-outer iteration scheme is naturally established. In the following we use the acronym IRN to indicate a wide class of algorithms that leverage (outer) reweighing together with an (inner) iterative scheme. IRN methods are also closely related to iteratively reweighted least squares (IRLS) methods [4, Chapter 4]. Since IRN methods can get very costly, another common approach is to use iterative shrinkage thresholding algorithms [2], where an iterative two-step process is used.

Many of these methods assume that a good value of the regularization parameter is available a priori, but oftentimes this is not the case. And although there have been some recent works on selecting regularization parameters for ℓ_1 regularization, e.g., [13], these can still be quite costly for very large problems. Selecting regularization parameters for ℓ_p-regularized problems remains a tricky, yet crucial, task. For the special case where $p = 2$, significant works on hybrid methods have enabled successful simultaneous estimation of the regularization parameter and computation of large-scale reconstructions (see, e.g., [22, 31]). In these hybrid frameworks, the problem is projected onto Krylov subspaces of increasing size, and the task of choosing a suitable value of the regularization parameter is reduced to solving the smaller, projected problem. However, such approaches have not been fully investigated for general ℓ_p-regularized problems. The flexible hybrid framework for ℓ_p-regularized problems that we describe in section 3 incorporates simultaneous parameter selection and is based on the IRN reformulation.

As described in [32], the first step toward an IRN approach is to define a sequence of appropriate regularization operators to break the ℓ_p-regularized problem into a

sequence of 2-norm problems,

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{L}(\mathbf{x})\mathbf{x}\|_2^2 , \tag{3}$$

where

$$\mathbf{L}(\mathbf{x}) = \operatorname{diag}\left((|x_i|^{\frac{p-2}{2}})_{i=1,\dots,n} \right) . \tag{4}$$

Here x_i is the ith entry of vector $\mathbf{x}$. We remark that, when $p < 2$, care is needed when defining (4), because division by 0 may occur if $x_i = 0$ for some $i = 1, \dots, n$. To fix this potential issue, small thresholds $\tau_1, \tau_2 > 0$ are set, and the matrix in (4) is redefined as

$$\mathbf{L}(\mathbf{x}) = \operatorname{diag}((f_\tau(|x_i|)^{\frac{p-2}{2}})_{i=1,\dots,n}) , \text{ where } f_\tau(|x_i|) = \begin{cases} |x_i| & \text{if } |x_i| \geq \tau_1, \\ \tau_2 & \text{if } |x_i| < \tau_1. \end{cases} \tag{5}$$

Note that taking $\tau_2 < \tau_1$ enforces additional sparsity in $f_\tau(|x_i|)$. In the case $p = 1$, the IRN approach reduces the ℓ_1-regularized problem (2) to a sequence of least-squares problems involving a weighted 2-norm. That is,

$$\|\mathbf{x}\|_1 \approx \|\mathbf{L}(\mathbf{x})\mathbf{x}\|_2^2 ,$$

where $\mathbf{L}(\mathbf{x}) = \operatorname{diag}(1/\sqrt{f_\tau(|\mathbf{x}|)})$, $f_\tau(\cdot)$ is defined in (5), and the square root and absolute value operations are applied componentwise. We remark that problem (3) can be equivalently reformulated as

$$\min_{\widehat{\mathbf{x}}} \left\| \mathbf{A}\mathbf{L}(\mathbf{x})^{-1}\widehat{\mathbf{x}} - \mathbf{b} \right\|_2^2 + \lambda \|\widehat{\mathbf{x}}\|_2^2 , \tag{6}$$

where $\widehat{\mathbf{x}} = \mathbf{L}(\mathbf{x})\mathbf{x}$. This transformation into standard form is computationally convenient, as it only amounts to the inversion of a diagonal matrix.

For realistic scenarios, problems (3) and (6) are intrinsically nonlinear. In order to avoid nonlinearities, we follow the common practice of approximating the matrix $\mathbf{L}(\mathbf{x})$ by the matrix $\mathbf{L}_k = \mathbf{L}(\mathbf{x}_k)$, where $\mathbf{x}_k$ is an approximation of the solution obtained at the $(k-1)$st outer iteration. Then at the kth outer iteration, we solve the Tikhonov problem,

$$\min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{L}_k\mathbf{x}\|_2^2 . \tag{7}$$

The IRN method proposed in [32] prescribes to apply the conjugate gradient (CG) method to solve the normal equations corresponding to (7), i.e.,

$$(\mathbf{A}^\top\mathbf{A} + \lambda\mathbf{L}_k^\top\mathbf{L}_k)\mathbf{x} = \mathbf{A}^\top\mathbf{b} , \quad \mathbf{L}_k = \mathbf{L}(\mathbf{x}_k) . \tag{8}$$

Also preconditioned CG (PCG) can be applied at the kth outer iteration of IRN to solve the normal equations associated to a preconditioned version of (7), i.e.,

$$(\mathbf{L}_k^{-\top}\mathbf{A}^\top\mathbf{A}\mathbf{L}_k^{-1} + \lambda\mathbf{I})\widehat{\mathbf{x}} = \mathbf{L}_k^{-\top}\mathbf{A}^\top\mathbf{b} , \quad \mathbf{L}_k^{-1}\widehat{\mathbf{x}} = \mathbf{x}, \quad \mathbf{L}_k = \mathbf{L}(\mathbf{x}_k) . \tag{9}$$

We refer to this approach as the preconditioned IRN (PIRN) method, which is similar in essence to the inner-outer scheme proposed in [1] to handle total variation regularization. We emphasize that the term "preconditioned" is used in a somewhat

unconventional way: the "preconditioners" considered here are not aimed at accelerating the convergence of the iterative solvers but rather at enforcing some specific regularity into the associated solution subspace. Transformed ℓ_p-regularized problems can be suitably expressed in this framework too, as we will explain in section 4. We stress once more that, in the IRN framework, the matrix $\mathbf{L} = \mathbf{L}_k$ changes at each outer iteration, resulting in a sequence of least-squares problems to be solved. A more efficient alternative that is applied directly to problem (6) and that exploits flexible preconditioning to bypass inner-outer iterative schemes is summarized below.

Generalized Arnoldi–Tikhonov approaches. For completeness, we provide a brief overview of the generalized Arnoldi–Tikhonov (GAT) approach [10] to solve problem (6), or equivalently problem (3), for $\mathbf{A} \in \mathbb{R}^{n\times n}$ and for changing preconditioners $\mathbf{L}(\mathbf{x}_k) = \mathbf{L}_k$. Consider the flexibly preconditioned Arnoldi algorithm, in which, at the kth iteration, we have

$$\mathbf{A}\widehat{\mathbf{Z}}_k = \widehat{\mathbf{V}}_{k+1}\widehat{\mathbf{H}}_k\,, \tag{10}$$

where $\widehat{\mathbf{H}}_k \in \mathbb{R}^{(k+1)\times k}$ is upper Hessenberg, $\widehat{\mathbf{V}}_{k+1} = \begin{bmatrix}\widehat{\mathbf{v}}_1 & \dots & \widehat{\mathbf{v}}_{k+1}\end{bmatrix}$ contains orthonormal columns with $\widehat{\mathbf{v}}_1 = \mathbf{b}/\left\|\mathbf{b}\right\|_2$, and $\widehat{\mathbf{Z}}_k = \begin{bmatrix}\mathbf{L}_1^{-1}\widehat{\mathbf{v}}_1 & \dots & \mathbf{L}_k^{-1}\widehat{\mathbf{v}}_k\end{bmatrix} \in \mathbb{R}^{n\times k}$. Here and in the following, we assume an initial guess $\mathbf{x}_0 = \mathbf{0}$; extensions to include $\mathbf{x}_0 \neq \mathbf{0}$ are trivial and follow standard derivations. Also, throughout the paper, we assume that all of the algorithms are breakdown-free, i.e., the dimension of the kth solution subspace is k. We also note that, if the preconditioner is fixed for all iterations ($\mathbf{L}_i = \mathbf{L}$, $i = 1, \dots, k$), then $\widehat{\mathbf{Z}}_k = \mathbf{L}^{-1}\widehat{\mathbf{V}}_k$ and decomposition (10) reduces to the one associated with the standard right-preconditioned GMRES method. The GAT method computes approximate solutions of the form $\mathbf{x}_k = \widehat{\mathbf{Z}}_k\widehat{\mathbf{y}}_k$, where

$$\widehat{\mathbf{y}}_k = \arg\min_{\mathbf{y}} \left\|\widehat{\mathbf{H}}_k\mathbf{y} - \left\|\mathbf{b}\right\|_2\mathbf{e}_1\right\|_2^2 + \lambda\left\|\mathbf{y}\right\|_2^2\,, \tag{11}$$

where $\mathbf{e}_1 \in \mathbb{R}^{k+1}$ is the first column of the identity matrix of order $k+1$. For $\lambda = 0$, we have the flexible GMRES (FGMRES) method [34]. The main advantages of this approach are that *only one* solution subspace needs to be generated (versus multiple solves in IRN), one matrix-vector multiplication with $\mathbf{A}$ is required at each iteration (versus one with $\mathbf{A}$ and one with $\mathbf{A}^\top$ in CGLS), and a suitable value of the regularization parameter and an appropriate value of the threshold for the stopping criterion are determined automatically by exploiting the hybrid framework. In [10], the GAT method and its variants were used to efficiently compute approximate solutions to ℓ_1-regularized problems, but a limitation is that this method only works for square problems. A naïve extension of the GAT method to general least-squares problems by applying the flexible Arnoldi algorithm to the normal equations is not recommended, due to the squaring of the condition number of the coefficient matrix and the lack of a computationally convenient way to estimate the residual norm for the original problem (1). Although flexible versions of the so-called AB-GMRES and BA-GMRES methods [26] may be devised, in the following section we exploit a new computational tool from numerical linear algebra, namely, the flexible Golub–Kahan method. In this way, we avoid the normal equations and work directly with the residual from the original least-squares problem, which can be helpful in determining the regularization parameter and stopping criteria.

3. Flexible Golub–Kahan hybrid methods. In this section, we describe hybrid approaches based on the flexible Golub–Kahan process for computing an approximate solution to the Tikhonov problem (7), where $\mathbf{L}_k$ may change at each iteration. As

discussed in section 2, problem (7) approximates regularized problem (3). Similarly to the GAT method, the flexible Golub–Kahan hybrid methods follow an iterative two-step process. First we generate a basis for the solution by exploiting a flexible preconditioning framework to take into account a changing regularizer, and second, we compute an approximate solution to the inverse problem by solving an optimization problem in the projected subspace (where regularization can be done efficiently and with automatic regularization parameter selection for the projected problem). These iterative approaches are ideal for problems where $\mathbf{A}$ and $\mathbf{A}^\top$ can be accessed only by matrix-vector multiplication, where only a few basis vectors are required to obtain a good solution, and where a suitable value of the regularization parameter is not known a priori.

3.1. Incorporating weights: A flexible Golub–Kahan decomposition. In order to incorporate a changing preconditioner, we use a flexible variant of the Golub–Kahan bidiagonalization (GKB) algorithm to generate a basis for the solution. We call this the *flexible Golub–Kahan* (FGK) process and mention that it is closely related to the inexact Lanczos process [37, 40]. Given $\mathbf{A}, \mathbf{b}$ and changing preconditioners $\mathbf{L}_k$, the kth iteration of the FGK iterative process generates vectors $\mathbf{z}_k$, $\mathbf{v}_k$, and $\mathbf{u}_{k+1}$ such that

$$\mathbf{A}\mathbf{Z}_k = \mathbf{U}_{k+1}\mathbf{M}_k \quad \text{and} \quad \mathbf{A}^\top\mathbf{U}_{k+1} = \mathbf{V}_{k+1}\mathbf{T}_{k+1}, \tag{12}$$

where

- $\mathbf{Z}_k = \begin{bmatrix}\mathbf{z}_1 & \dots & \mathbf{z}_k\end{bmatrix} = \begin{bmatrix}\mathbf{L}_1^{-1}\mathbf{v}_1 & \dots & \mathbf{L}_k^{-1}\mathbf{v}_k\end{bmatrix} \in \mathbb{R}^{n\times k}$,
- $\mathbf{M}_k = [m_{i,j}]_{i=1,\dots,k+1;j=1,\dots,k} \in \mathbb{R}^{(k+1)\times k}$ is upper Hessenberg,
- $\mathbf{T}_{k+1} = [t_{i,j}]_{i,j=1,\dots,k+1} \in \mathbb{R}^{(k+1)\times(k+1)}$ is upper triangular,
- $\mathbf{U}_{k+1} = \begin{bmatrix}\mathbf{u}_1 & \dots & \mathbf{u}_{k+1}\end{bmatrix} \in \mathbb{R}^{m\times(k+1)}$ has orthonormal columns with $\mathbf{u}_1 = \mathbf{b}/\|\mathbf{b}\|_2$, and
- $\mathbf{V}_{k+1} = \begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_{k+1}\end{bmatrix} \in \mathbb{R}^{n\times(k+1)}$ has orthonormal columns.

Compared to standard GKB [15], the key differences are that we now have an upper Hessenberg and an upper triangular matrix, instead of one bidiagonal matrix. Also, we must keep track of an additional set of vectors, namely, the basis vectors in $\mathbf{Z}_k$. Furthermore, since there is no bidiagonal structure to exploit, the additional computational requirement is orthogonalization with all previous vectors. However, these additional requirements are negligible if $k \ll \max\{m, n\}$. Moreover, as for standard GKB, the computational cost per iteration is dominated by a matrix-vector product with $\mathbf{A}$ and one with $\mathbf{A}^\top$. We remark that, if $\mathbf{L}_k = \mathbf{L}$, (12) reduces to the right-preconditioned GKB. The FGK process is summarized in Algorithm 3.1.

Algorithm 3.1 Flexible Golub–Kahan Process.

1: Initialize $\mathbf{u}_1 = \mathbf{b}/\beta_1$, where $\beta_1 = \|\mathbf{b}\|$
2: **for** $i = 1, \dots, k$ **do**
3: $\quad \mathbf{w} = \mathbf{A}^\top\mathbf{u}_i$, $t_{j,i} = \mathbf{w}^\top\mathbf{v}_j$ for $j = 1, \dots, i-1$
4: $\quad \mathbf{w} = \mathbf{w} - \sum_{j=1}^{i-1} t_{j,i}\mathbf{v}_j$, $t_{i,i} = \|\mathbf{w}\|$, $\mathbf{v}_i = \mathbf{w}/t_{i,i}$
5: $\quad \mathbf{z}_i = \mathbf{L}_i^{-1}\mathbf{v}_i$
6: $\quad \mathbf{w} = \mathbf{A}\mathbf{z}_i$, $m_{j,i} = \mathbf{w}^\top\mathbf{u}_j$ for $j = 1, \dots, i$
7: $\quad \mathbf{w} = \mathbf{w} - \sum_{j=1}^{i} m_{j,i}\mathbf{u}_j$, $m_{i+1,i} = \|\mathbf{w}\|$, $\mathbf{u}_{i+1} = \mathbf{w}/m_{i+1,i}$
8: **end for**

Notice that the column vectors of $\mathbf{Z}_k$ no longer span a Krylov subspace in a traditional sense, but they do provide a basis for the solution subspace. In section

5 we provide some qualitative observations regarding the basis vectors. For now, consider the fit-to-data term $\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2^2$ and consider solutions in the column space of $\mathbf{Z}_k$, denoted by $\mathcal{R}(\mathbf{Z}_k)$. Using the relationships in (12), the residual can be written as

$$\mathbf{A}\mathbf{Z}_k\mathbf{y}-\mathbf{b}=\mathbf{U}_{k+1}(\mathbf{M}_k\mathbf{y}-\beta_1\mathbf{e}_1).$$

We define the flexible LSQR (FLSQR) and flexible LSMR (FLSMR) iterates as $\mathbf{x}_k=\mathbf{Z}_k\mathbf{y}_k$, where

$$\mathbf{y}_k=\arg\min_{\mathbf{y}}\|\mathbf{M}_k\mathbf{y}-\beta_1\mathbf{e}_1\|_2^2 \tag{13}$$

and

$$\mathbf{y}_k=\arg\min_{\mathbf{y}}\|\mathbf{T}_{k+1}\mathbf{M}_k\mathbf{y}-\beta_1 t_{1,1}\mathbf{e}_1\|_2^2\,, \tag{14}$$

respectively. These definitions are analogous to the mathematical definitions of the LSQR and LSMR iterates in [8, 29, 30]. The FLSMR formulation exploits the following relationships:

$$\mathbf{A}^\top(\mathbf{A}\mathbf{Z}_k\mathbf{y}-\mathbf{b})=\mathbf{V}_{k+1}(\mathbf{T}_{k+1}\mathbf{M}_k\mathbf{y}-t_{1,1}\beta_1\mathbf{e}_1)\quad\text{and}\quad\mathbf{A}^\top\mathbf{b}=\mathbf{V}_{k+1}\beta_1 t_{1,1}\mathbf{e}_1\,.$$

We have the following optimality properties for FLSQR and FLSMR, which are analogous to the ones enjoyed by the standard counterparts of these methods and by FGMRES [34].

PROPOSITION 3.1. *The FLSQR iterate $\mathbf{x}_k$ obtained at the kth step minimizes the residual norm $\|\mathbf{A}\mathbf{x}_k-\mathbf{b}\|_2$ over $\mathcal{R}(\mathbf{Z}_k)$, and the FLSMR iterate $\mathbf{x}_k$ obtained at the kth step minimizes $\left\|\mathbf{A}^\top(\mathbf{A}\mathbf{x}_k-\mathbf{b})\right\|_2$ over $\mathcal{R}(\mathbf{Z}_k)$.*

We note that $\mathbf{T}_{k+1}\mathbf{M}_k$ is a $(k+1)\times k$ upper Hessenberg matrix and that the solution subspace generated by the FGK process is the same as the one generated by the flexible Arnoldi algorithm applied to the normal equations. More precisely, the following equivalence theorem holds.

THEOREM 3.2. *Let $\mathbf{A}\in\mathbb{R}^{m\times n}, m\geq n$ (full column rank), $\mathbf{b}\in\mathbb{R}^m$, $\mathbf{x}_0=\mathbf{0}$, and take the preconditioners $\mathbf{L}_i, i=1,2,\dots k$. Then, in exact arithmetic, the kth iterate of FLSMR applied to $\min_{\mathbf{x}}\|\mathbf{A}\mathbf{x}-\mathbf{b}\|_2$ is the same as the kth iterate of FGMRES applied to the normal equations*

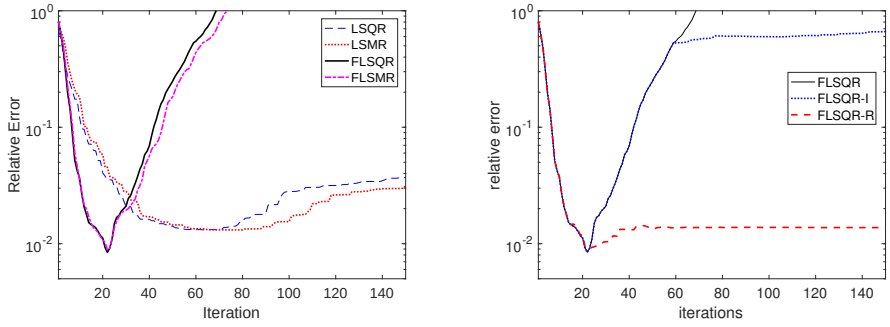
$$\mathbf{A}^\top\mathbf{A}\mathbf{x}=\mathbf{A}^\top\mathbf{b}. \tag{15}$$

Proof. After k iterations of FGMRES applied to (15), we have a matrix $\widehat{\mathbf{Z}}_k=\begin{bmatrix}\mathbf{L}_1^{-1}\widehat{\mathbf{v}}_1 & \dots & \mathbf{L}_k^{-1}\widehat{\mathbf{v}}_k\end{bmatrix}\in\mathbb{R}^{n\times k}$, an upper Hessenberg matrix $\widehat{\mathbf{H}}_k\in\mathbb{R}^{(k+1)\times k}$, and a matrix $\widehat{\mathbf{V}}_{k+1}\in\mathbb{R}^{n\times(k+1)}$ with orthonormal columns and $\widehat{\mathbf{V}}_{k+1}\mathbf{e}_1=\mathbf{A}^\top\mathbf{b}/\left\|\mathbf{A}^\top\mathbf{b}\right\|_2$, which satisfy the relationship

$$\mathbf{A}^\top\mathbf{A}\widehat{\mathbf{Z}}_k=\widehat{\mathbf{V}}_{k+1}\widehat{\mathbf{H}}_k\,. \tag{16}$$

The projected problem is given by

$$\min_{\mathbf{x}\in\mathcal{R}(\widehat{\mathbf{Z}}_k)}\left\|\mathbf{A}^\top\mathbf{A}\mathbf{x}-\mathbf{A}^\top\mathbf{b}\right\|_2^2=\min_{\mathbf{y}}\left\|\widehat{\mathbf{H}}_k\mathbf{y}-\left\|\mathbf{A}^\top\mathbf{b}\right\|_2\mathbf{e}_1\right\|_2^2\,, \tag{17}$$



(a) (b)

FIG. 1. `heat` *test problem from* [17]. (a) *Relative error norm,* $\|\mathbf{x}_k - \mathbf{x}_{\text{true}}\|_2 / \|\mathbf{x}_{\text{true}}\|_2$*, for LSQR, LSMR, FLSQR, and FLSMR. The semiconvergence behavior is evident for all of the methods.* (b) *Relative error norms for FLSQR-I and FLSQR-R with optimal regularization parameters, along with relative error norms for FLSQR provided for comparison.*

so the kth iterate of FGMRES is given by

$$\widehat{\mathbf{x}}_k = \widehat{\mathbf{Z}}_k \widehat{\mathbf{H}}_k^{\dagger} \left\| \mathbf{A}^\top \mathbf{b} \right\|_2 \mathbf{e}_1 \,,$$

where $\widehat{\mathbf{H}}_k^{\dagger} = (\widehat{\mathbf{H}}_k^\top \widehat{\mathbf{H}}_k)^{-1} \widehat{\mathbf{H}}_k^\top$ is the pseudoinverse. In exact arithmetic the solution subspaces generated by FGMRES and FGK in Algorithm 3.1 are the same and coincide with

$$\text{span}\{\mathbf{L}_1^{-1}\widehat{\mathbf{v}}_1, \mathbf{L}_2^{-1}\widehat{\mathbf{v}}_2, \ldots, \mathbf{L}_k^{-1}\widehat{\mathbf{v}}_k\} \,,$$

so that $\widehat{\mathbf{Z}}_k = \mathbf{Z}_k$. (This is immediate from factorizations (12) and (16).) The optimality condition for FGMRES (see Proposition 2.1 in [34]) and FLSMR (see Proposition 3.1) guarantee that the kth iterate of FLSMR and FGMRES both correspond to the solution of (17). ∎

3.2. Solving the regularized problem: Flexible hybrid algorithms. As explained in section 3.1, the FGK process can be used to build a solution subspace that can efficiently incorporate changing preconditioners, and one can solve the projected problems (13) and (14), which correspond to the FLSQR and FLSMR methods, respectively. However, it is well-known that, for inverse problems, iterative methods exhibit a semiconvergent behavior, where the relative reconstruction error norm $\|\mathbf{x}_k - \mathbf{x}_{\text{true}}\|_2 / \|\mathbf{x}_{\text{true}}\|_2$ decreases initially but at some point increases due to amplification of noise [18]. This phenomenon, which is common for most ill-posed inverse problems, occurs also for flexible methods, as can be seen in Figure 1(a).

Hybrid methods, where regularization is included on the projected problem, have been proposed to suppress the relative reconstruction error norms, i.e., to mitigate semiconvergence. The first hybrid approach that we propose is analogous to the GAT algorithm (see (11)), where we include a standard regularization term in (13), so that

$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \|\mathbf{M}_k \mathbf{y} - \beta_1 \mathbf{e}_1\|_2^2 + \lambda \|\mathbf{y}\|_2^2 \,. \tag{18}$$

Henceforth, we define FLSQR-I iterates as $\mathbf{x}_k = \mathbf{Z}_k \mathbf{y}_k$, where $\mathbf{y}_k$ is defined in (18).

Let $\mathbf{Z}_k = \mathbf{Q}_k \mathbf{R}_k$ be the thin QR factorization of $\mathbf{Z}_k$, where $\mathbf{R}_k \in \mathbb{R}^{k \times k}$ is upper triangular and $\mathbf{Q}_k \in \mathbb{R}^{n \times k}$ contains orthonormal columns. This is inexpensive to

compute if k is not too large. Then we also consider a hybrid method called FLSQR-R, in which iterates are constructed as $\mathbf{x}_k = \mathbf{Z}_k \mathbf{y}_k$, where

$$\mathbf{y}_k = \arg\min_{\mathbf{y}} \|\mathbf{M}_k \mathbf{y} - \beta_1 \mathbf{e}_1\|_2^2 + \lambda \|\mathbf{R}_k \mathbf{y}\|_2^2 . \tag{19}$$

The FLSQR-R method exhibits some desirable properties, especially for inverse problems. First, the FLSQR-R iterate can be interpreted as a best approximation in a subspace, in that $\mathbf{x}_k$ solves

$$\min_{\mathbf{x} \in \mathcal{R}(\mathbf{Z}_k)} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_2^2 . \tag{20}$$

Hence, the regularization parameter λ, which specifies the amount of regularization for the projected problem (19), corresponds to the amount of regularization for the constrained, full-dimensional problem (20). Second, using the following reformulation of the FLSQR-R subproblem (19),

$$\mathbf{w}_k = \arg\min_{\mathbf{w}} \left\|\mathbf{M}_k \mathbf{R}_k^{-1} \mathbf{w} - \beta_1 \mathbf{e}_1\right\|_2^2 + \lambda \|\mathbf{w}\|_2^2 , \quad \mathbf{y}_k = \mathbf{R}_k^{-1} \mathbf{w}_k ,$$

we can show that the singular values of the coefficient matrix $\mathbf{M}_k \mathbf{R}_k^{-1}$ provide good approximations to the singular values of $\mathbf{A}$. Indeed, we can see this by considering the following relations (where decomposition (12) and properties of the matrices appearing therein are extensively used):

$$\begin{aligned} \mathbf{R}_k^{-\top} \mathbf{M}_k^\top \mathbf{M}_k \mathbf{R}_k^{-1} &= \mathbf{R}_k^{-\top} \mathbf{M}_k^\top \mathbf{U}_{k+1}^\top \mathbf{U}_{k+1} \mathbf{M}_k \mathbf{R}_k^{-1} \\ &= \mathbf{R}_k^{-\top} \mathbf{Z}_k^\top \mathbf{A}^\top \mathbf{A} \mathbf{Z}_k \mathbf{R}_k^{-1} \\ &= \mathbf{Q}_k^\top \mathbf{A}^\top \mathbf{A} \mathbf{Q}_k . \end{aligned}$$

Since the eigenvalues are just the squares of the singular values, we see that as k increases, the singular values of $\mathbf{M}_k \mathbf{R}_k^{-1}$ provide better approximations to the singular values of $\mathbf{A}$.

Hybrid LSMR variants, namely, the FLSMR-I and FLSMR-R methods, can be defined analogously. Then, using Theorem 3.2, we can see that in exact arithmetic and for a fixed regularization parameter, the FLSMR-I iterates are the same as the GAT iterates applied to a Tikhonov problem with the fit-to-data term $\left\|\mathbf{A}^\top \mathbf{A}\mathbf{x} - \mathbf{A}^\top \mathbf{b}\right\|_2^2$. However, the benefit of the FGK approaches versus GAT on the normal equations is that FGK produces residual norms for the original problem, which can be important for tools such as the discrepancy principle for parameter selection and for stopping criteria.

Unless otherwise stated, the parameter choice methods considered here are based on the discrepancy principle: in particular, we either prescribe the discrepancy principle to be satisfied at each iteration, or we apply the "secant update" variant prescribing suitable updates of the regularization parameter at each iteration. More specifically, we determine an appropriate combination of regularization parameters (i.e., the number of performed iterations k and the Tikhonov parameter $\lambda > 0$) such that

$$\|\mathbf{b} - \mathbf{A}\mathbf{x}_k\|_2 \leq \eta \|\mathbf{e}\|_2 , \tag{21}$$

where $\mathbf{x}_k$ solves (20) and depends on both k and λ, and $\eta > 1$ is a safety factor. See [10] and [11] for a detailed description of these regularization parameter selection and stopping criteria strategies.

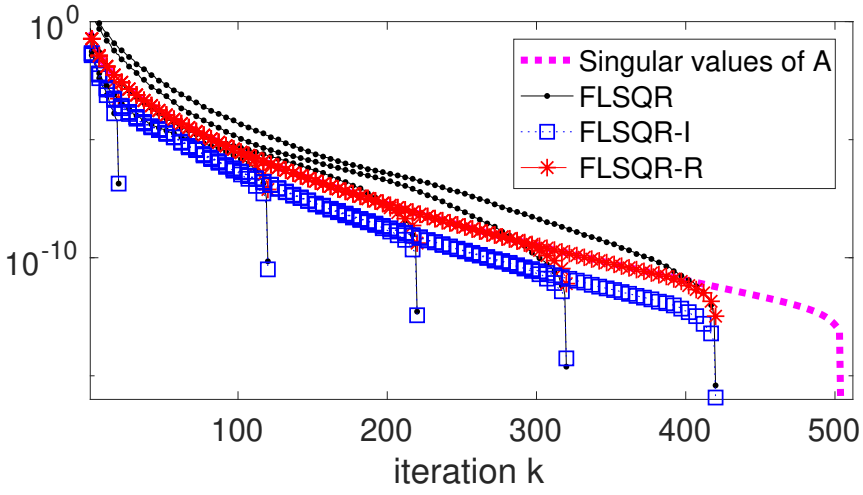


FIG. 2. `heat` *test problem from* [17]. *This plot compares the singular values of* $\mathbf{A}$ *to the singular values of* $\mathbf{M}_k$ *from FLSQR and FLSQR-I and to the singular values of* $\mathbf{M}_k\mathbf{R}_k^{-1}$ *from FLSQR-R for iterations* k *between* 20 *and* 420 *in increments of* 100.

An illustration. The goals of this illustration are (i) to demonstrate the higher quality of the solutions obtained by applying flexible methods (due to a better basis for the solution subspace), (ii) to motivate the need for a hybrid approach (by showing semiconvergence behavior of FLSQR and FLSMR), and (iii) to show that the singular values of the original problem can be approximated well by using FLSQR-R. More thorough numerical results and comparisons will be presented in section 5.

For this illustration, we use the `heat` example from Regularization Tools [17], where $\mathbf{A}$ has size 512×512, having a sparse true solution (50% of its entries are zero, so that $\mathbf{\Psi} = \mathbf{I}$ and $p = 1$ in (2)). White noise is added to the observed signal at noise level 10^{-4}, i.e., $\|\mathbf{e}\|_2/\|\mathbf{A}\mathbf{x}_{\text{true}}\|_2 = 10^{-4}$. In Figure 1(a), we provide relative reconstruction error norms per iteration for LSQR, LSMR, FLSQR, and FLSMR. The delayed semiconvergence of LSMR versus LSQR was noted in [5] and is also slightly visible for FLSMR versus FLSQR. The more pronounced feature that we see here is that the flexible variants converge faster but also exhibit stronger semiconvergence in that the relative error norms increase faster. Thus, there is a greater need for additional regularization. In Figure 1(b), we show that the hybrid methods FLSQR-I and FLSQR-R (here with the optimal regularization parameters, i.e., the ones that minimize the relative error norm at each iteration) can mitigate the semiconvergence behavior. Comparisons with different parameter selection methods can be found in section 5. We note that, for this particular test problem, flexible preconditioning speeds up the convergence of the iterative method. However, for our problems of interest (e.g., ℓ_p-regularized problems), flexible preconditioning is mainly used to improve the solution subspace. Thus, the particular choice of regularization for the projected problem is not so critical and is mostly required for supressing the errors.

Another important tool for the analysis of a regularization method is the approximation of the singular values of $\mathbf{A}$. For the standard GKB algorithm, it is well-known that the singular values of the bidiagonal matrix approximate the singular values of $\mathbf{A}$ [33]. However, these results do not directly extend to the FGK process. In Figure 2,

we display the singular values of $\mathbf{A}$ with a dashed line, which is partially covered by the FLSQR-R curve (continuous asterisked line). Then, for $k = 20$ to $k = 420$ in intervals of 100, we provide the singular values of upper Hessenberg matrix $\mathbf{M}_k$ for FLSQR (continuous circled line) and FLSQR-I (continuous squared line), and the singular values of $\mathbf{M}_k\mathbf{R}_k^{-1}$ for FLSQR-R. Note that, in the flexible methods, the previous iterate $\mathbf{x}_{k-1}$, which may include regularization, changes the preconditioner and hence the FGK matrices. It is evident that singular values of $\mathbf{M}_k\mathbf{R}_k^{-1}$ from FLSQR-R provide better approximations to the singular values of $\mathbf{A}$ than those of $\mathbf{M}_k$ from FLSQR and FLSQR-I. Furthermore, with more iterations, smaller singular values of $\mathbf{A}$ are being approximated, which motivates the need for regularization of the projected problem.

4. Flexible methods for the transformed problem. As mentioned in section 1, the goal in many applications is to compute solutions that are sparse with respect to some transformation (e.g., in a frequency domain). In this section, we focus on flexible Arnoldi and flexible Golub–Kahan hybrid methods for solving the transformed ℓ_p-regularized problem (2), where $\mathbf{\Psi} \neq \mathbf{I}$. Although any invertible transform matrix can be used here, we will focus on wavelet transforms mainly for two reasons. First, it is well-known that many images can be sparsely represented in the wavelet domain. Indeed, wavelet-based iterative methods have been widely considered for linear inverse problems (see, e.g., [6, 7, 23, 39]). Second, when taking orthonormal wavelet transforms, computations involving 2-norms of transformed quantities or inverse transforms can be easily performed. The specific strategy used to incorporate a wavelet transform into the flexible iterative solvers depends on the properties of the linear system at hand (which, eventually, depends on the properties of the inverse problem to be regularized) and, for all the methods, the regularization parameter can be automatically estimated.

Let $\widetilde{\mathbf{\Psi}} \in \mathbb{R}^{m\times m}$ be an orthogonal matrix. Then, problem (2) is equivalent to

$$\min_{\mathbf{x}} \left\| \widetilde{\mathbf{\Psi}}\mathbf{A}\mathbf{\Psi}^{-1}\mathbf{\Psi}\mathbf{x} - \widetilde{\mathbf{\Psi}}\mathbf{b} \right\|_2^2 + \lambda \left\| \mathbf{\Psi}\mathbf{x} \right\|_p^p . \tag{22}$$

Moreover, after some variable transformations, (22) can be written as

$$\min_{\mathbf{s}} \|\mathbf{H}\mathbf{s} - \mathbf{d}\|_2^2 + \lambda \|\mathbf{s}\|_p^p , \quad \text{where} \quad \mathbf{H} = \widetilde{\mathbf{\Psi}}\mathbf{A}\mathbf{\Psi}^{-1},\ \mathbf{s} = \mathbf{\Psi}\mathbf{x},\ \mathbf{d} = \widetilde{\mathbf{\Psi}}\mathbf{b}, \tag{23}$$

which is an ℓ_p-regularized problem. The choice of $\widetilde{\mathbf{\Psi}}$ is problem-dependent and solver-dependent. For instance, when considering image deblurring problems where both $\mathbf{x}$ and $\mathbf{b}$ are images of the same size described by pixel values, it is natural to take $\widetilde{\mathbf{\Psi}} = \mathbf{\Psi}$ to be an orthogonal wavelet transform; this formulation was considered in [3]. If the GAT method is applied to problem (23) with $p = 1$ and variable preconditioner $\mathbf{L}(\mathbf{s}_k) = \mathbf{L}_k$, then the subspace $\mathcal{S}_k = \text{span}\{\mathbf{L}_1^{-1}\widehat{\mathbf{v}}_1, \mathbf{L}_2^{-1}\widehat{\mathbf{v}}_2, \ldots, \mathbf{L}_k^{-1}\widehat{\mathbf{v}}_k\}$ with $\widehat{\mathbf{v}}_1 = \mathbf{d}/\|\mathbf{d}\|_2$ is generated for the kth approximation of the transformed solution $\mathbf{s}$. This subspace enforces sparsity in the wavelet domain for the wavelet coefficients $\mathbf{s}$ of the original image $\mathbf{x}$. The solution subspace for the latter is given by $\mathbf{\Psi}^\top\mathcal{S}_k$, so that it is evident that first sparsity is enforced in the wavelet domain, and then the sparse wavelet coefficients are transformed back into the original pixel domain. However, for situations where one has no intuition regarding the sparsity properties of $\mathbf{b}$, one can simply take $\widetilde{\mathbf{\Psi}} = \mathbf{I}$. Analogously, if solvers based on the FGK process are applied to solve the same problem, then the solution subspace is given by

$$\mathbf{\Psi}^\top \text{span}\left\{\mathbf{L}_1^{-1}\mathbf{\Psi}\mathbf{v}_1, \ldots, \mathbf{L}_k^{-1}\mathbf{\Psi}\mathbf{v}_k\right\} \quad \text{with } \mathbf{v}_1 = \mathbf{A}^\top\mathbf{b}/\left\|\mathbf{A}^\top\mathbf{b}\right\|_2.$$

Notice that the choice of $\widetilde{\mathbf{\Psi}}$ is irrelevant for flexible methods based on FGK, since

$$\mathbf{H}^\top \mathbf{d} = \mathbf{\Psi}\mathbf{A}^\top \widetilde{\mathbf{\Psi}}^\top \widetilde{\mathbf{\Psi}}\mathbf{b} = \mathbf{\Psi}\mathbf{A}^\top \mathbf{b}, \quad \mathbf{H}^\top \mathbf{H} = \mathbf{\Psi}\mathbf{A}^\top \widetilde{\mathbf{\Psi}}^\top \widetilde{\mathbf{\Psi}}\mathbf{A}\mathbf{\Psi}^\top = \mathbf{\Psi}\mathbf{A}^\top \mathbf{A}\mathbf{\Psi}^\top.$$

An illustration. The goal of this illustration is to show that the solution space generated by the flexible Arnoldi algorithm applied to problem (23) is more suitable than the one generated by its standard counterpart. We consider a one-dimensional (1D) signal $\mathbf{x}$ with 64 entries, generated in such a way that only 8 of its 1-level Haar wavelet coefficients $\mathbf{s}$ are nonzero. The signal is corrupted by Gaussian blur with variance 2.25 and band 5, and white noise of level 10^{-2} is added. The exact and corrupted signals are displayed in Figure 3(a), and their wavelet coefficients are displayed in Figure 3(b). We choose $\lambda = 0$ in (23) so that the solution subspace does not depend on the specific parameter choice strategy that one may wish to consider. The threshold τ_1 in (5) is set to 0.2, while $\tau_2 = 10^{-14}$. Figure 3(c) displays the best reconstructions obtained by the FGMRES (11th iteration) and the GMRES (30th iteration) methods. One can clearly see that the FGMRES solution is of much higher quality than the GMRES one, and that the wavelet coefficients of the FGMRES solution are much sparser than the GMRES ones (see Figure 3(d)). The good performance of FGMRES for this example can be explained by looking at some of the basis vectors for the solution space, displayed in Figure 3(e)–(h). Indeed, the preconditioned basis vectors for the signal $\mathbf{x}$ have a piecewise-constant behavior, while the unpreconditioned ones display spurious oscillations; correspondingly, the preconditioned basis vectors for the wavelet coefficients $\mathbf{s}$ have a clear sparsity pattern, which is not reproduced by the unpreconditioned ones. Therefore, the FGMRES solution is better than the GMRES one as it is obtained by combining better basis vectors for the solution subspace. We remark that the basis vectors generated from the FGK process have similar properties, and thus are omitted. Also, a similar behavior of the preconditioned basis vectors can be observed in the more challenging experiments presented in section 5.

5. Numerical results. In this section, we provide three experiments to demonstrate the performance of the flexible Krylov hybrid methods on various test problems from image processing. The first two experiments are examples from image deblurring, where enforcing sparsity on the image and sparsity on the wavelet coefficients are investigated separately. The third experiment is concerned with tomographic reconstruction from undersampled data, where sparsity is imposed on the wavelet coefficients. All images are of size 256×256 pixels. For all of the experiments, the thresholds in (5) are $\tau_1 = 10^{-10}$, $\tau_2 = 10^{-16}$ (machine precision). All experiments are performed in MATLAB 2017a and use codes available in the Restore Tools [27] and AIR Tools II [19] software packages. MATLAB implementations of our methods, which should be used jointly with the IR Tools software package [9], are available at https://github.com/silviagazzola.

Experiment 1. In this experiment, we consider an image deblurring example from atmospheric imaging, with the true image, the point spread function (PSF), and the observed blurred image provided in Figure 4. For this problem, Gaussian white noise is added to the blurred image, such that the noise level is $5 \cdot 10^{-2}$.

For the reconstructions, we assume reflexive boundary conditions and solve the ℓ_1-regularized problem with $\mathbf{\Psi} = \mathbf{I}$, which is appropriate because the desired image is quite sparse. (Approximately 50% of its pixels are numerically zero.) First we provide a comparison of various Golub–Kahan-based methods. In Figure 5, we provide relative error norms per iteration for the flexible methods described in section 3, namely, FLSQR, FLSQR-I, and FLSQR-R with automatic regularization parameter

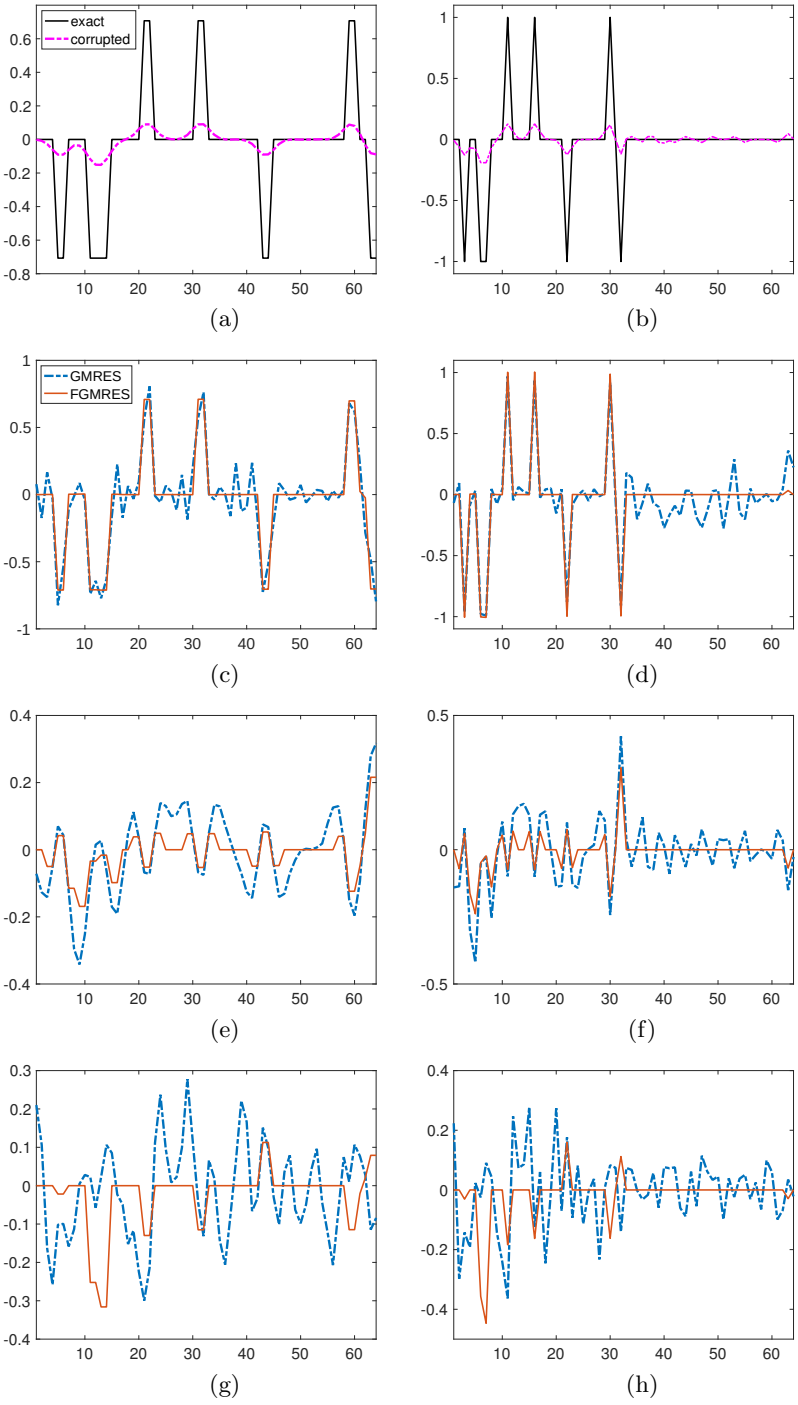


FIG. 3. *1D signal deblurring and denoising problem. The right column displays the 1D Haar wavelet coefficients of the signals displayed in the left columns. The first row shows the exact and corrupted signals. The second row shows the best reconstructions obtained by GMRES (dash-dot lines) and FGMRES (solid lines). The third and fourth row show the second and fourth basis vectors computed by GMRES (dash-dot lines) and FGMRES (solid lines), respectively.*

FIG. 4. *Experiment* 1: *Image deblurring example. Here we show the true image, the PSF, and the observed blurred and noisy image. The size of the images is* 256×256 *pixels.*

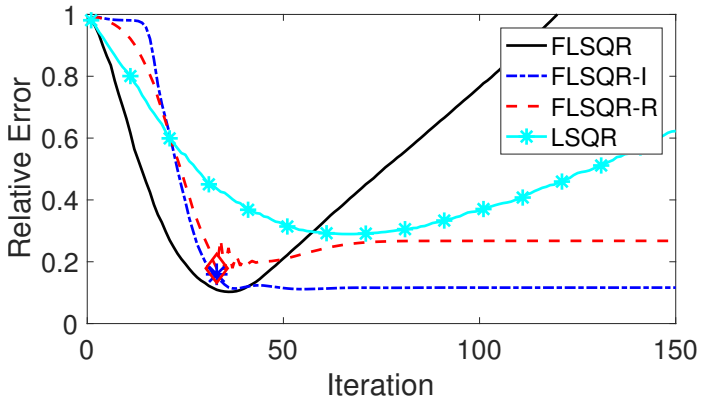


FIG. 5. *Experiment* 1: *Comparison of relative reconstruction error norms. The regularization parameter* λ *is selected automatically using the secant update method (discrepancy principle) for FLSQR-I and FLSQR-R;* $\lambda = 0$ *is set for FLSQR and LSQR. Automatically determined stopping iterations for the hybrid approaches are denoted by the diamond and star.*

selection using the secant update discrepancy principle (with safety factor $\eta = 1.01$ in (21)). In all experiments with the discrepancy principle, we use the true noise level but remark that estimates could be used [38]. Relative reconstruction error norms for LSQR are provided for comparison. Similarly to the observations made in section 3, the flexible methods exhibit faster convergence to more accurate solutions than the standard LSQR approach. Furthermore, we see that the flexible hybrid methods are able to stabilize the semiconvergent behavior by selecting an appropriate regularization parameter and stopping criterion (e.g., based on the secant update strategy or discrepancy principle).

In Figure 6, we provide the basis vectors (displayed as images) for the FLSQR-R and the LSQR solution subspaces for $k = 10, 20, 100$. Note that basis vectors for FLSQR-R correspond to the FGK vectors, while the LSQR ones correspond to the standard GKB vectors. It is evident that the basis images for the flexible method can better capture the flat regions of the image. Also, for large k, the FLSQR-R basis image is less affected by the noise amplification that is present in the LSQR basis image. Thus, we expect that, by constructing a better solution basis (i.e., one that is less affected by noise and that captures the sparsity properties of the image), the flexible methods can be successful for sparse image reconstruction. This behavior can be experimentally observed also at higher noise levels.

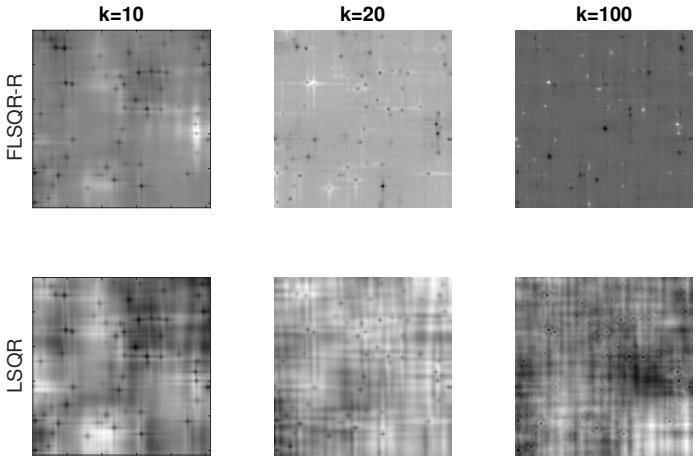


FIG. 6. *Experiment* 1: *Basis images for FLSQR-R and LSQR for* $k = 10, 20, 100$. *These are solution vectors (i.e.,* $\mathbf{z}_k$ *for FLSQR-R) that have been reshaped into images.*

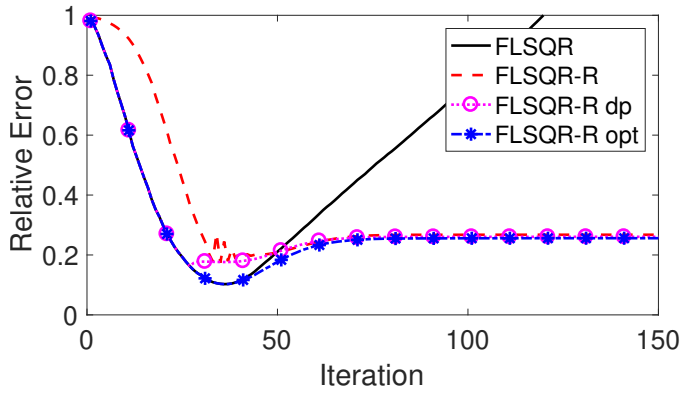


FIG. 7. *Experiment* 1: *Relative reconstruction error norms for different parameter choice methods. FLSQR-R and FLSQR-R dp use the secant update and the classical discrepancy principle, respectively, and thus require an estimate of the noise level. FLSQR-R opt corresponds to selecting the optimal regularization parameter at each iteration, which is not necessarily the overall best parameter because of flexibility.*

Next, we investigate some parameter choice methods. In Figure 7, we provide relative reconstruction error norms for FLSQR-R and FLSQR-R dp. Both methods use the discrepancy principle to obtain the regularization parameter, which requires prior knowledge of the noise level. More precisely, FLSQR-R utilizes the secant update parameter choice method described in [10], and FLSQR-R dp enforces the discrepancy principle to be satisfied at each iteration. Relative error norms for FLSQR-R opt correspond to selecting the regularization parameter at each iteration that minimizes the error norm of the current iterate minus the true solution. It is worth noting that, since the basis vectors are generated with respect to the current solution (because of flexibility), this approach does not necessarily produce the best overall regularization parameter for the problem.

Finally, we compare the FLSQR-R method to other methods for solving the ℓ_1-regularized problem. In Figure 8, we provide relative reconstruction error norms for

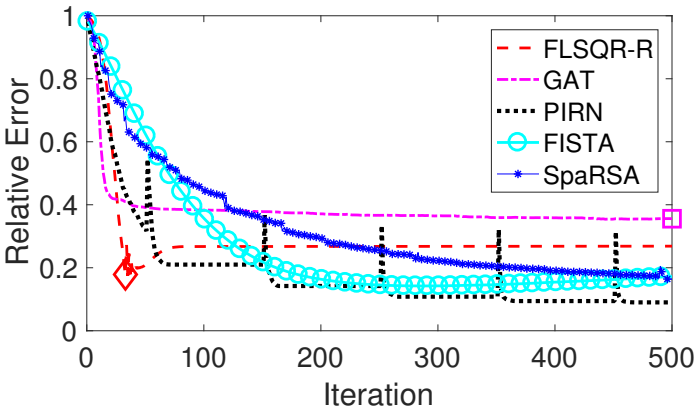


FIG. 8. *Experiment* 1: *Relative reconstruction error norms are provided to compare the FGK methods to some existing methods. It is important to note that the PIRN, FISTA, and SpaRSA regularization parameter is selected using our FLSQR-R approach.*

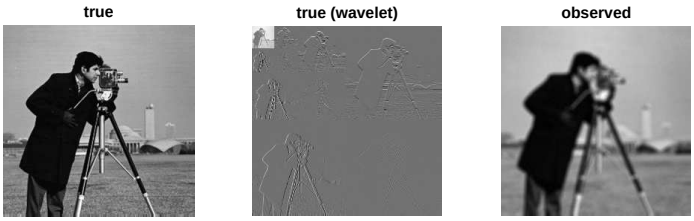


FIG. 9. *Experiment* 2: *Image deblurring example. Here we show the true image, the wavelet coefficients of the true image, and the observed image.*

GAT [10], PIRN, FISTA [2], and SpaRSA [42]. Since the regularization parameter for PIRN, FISTA, and SpaRSA must be selected prior to execution, we use the regularization parameter that is selected by FLSQR-R when the stopping criterion is satisfied (for this problem, $\lambda = 1.1 \cdot 10^{-5}$). We note that FISTA, SpaRSA, and PIRN compute reconstructions with similar or slightly better accuracy than FLSQR-R, but the two main advantages of the hybrid approaches are that the regularization parameter can be selected automatically, and the reconstruction can be obtained in fewer iterations. The main cost per iteration for all of these methods is one matrix-vector multiplication with $\mathbf{A}$ and one with $\mathbf{A}^\top$.

Experiment 2. In this experiment, we investigate the transformed ℓ_1-regularized problem for an image deblurring example. For this problem, we use the cameraman image shown in Figure 9, where out of focus blur (i.e., associated to a circular PSF of radius 4 pixels) and Gaussian white noise with noise level 0.01 are considered. Although a wide range of transformations $\mathbf{\Psi}$ can be employed, for simplicity we use a 2D Haar wavelet decomposition with 3 levels. For this example, the image itself is not sparse (only 27 pixels are numerically zero). However, slightly more that 10% of the pixels of the transformed true image (also provided in Figure 9) are numerically zero, and thus it is appropriate to consider the transformed ℓ_1-regularized problem.

First we investigate the Golub–Kahan-based methods. In Figure 10, we provide the relative reconstruction error norms for FLSQR, FLSQR-I, and FLSQR-R, where LSQR on the original problem is provided for comparison. Although the flexible methods take a few more iterations, they provide slightly smaller relative reconstruc-

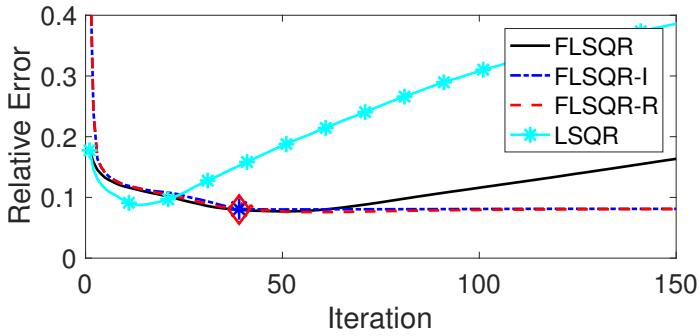


FIG. 10. *Experiment* 2: *Relative reconstruction error norms for Golub–Kahan-based approaches. The regularization parameter λ is selected automatically using the discrepancy principle for FLSQR-I and FLSQR-R; $\lambda = 0$ is set for FLSQR and LSQR.*

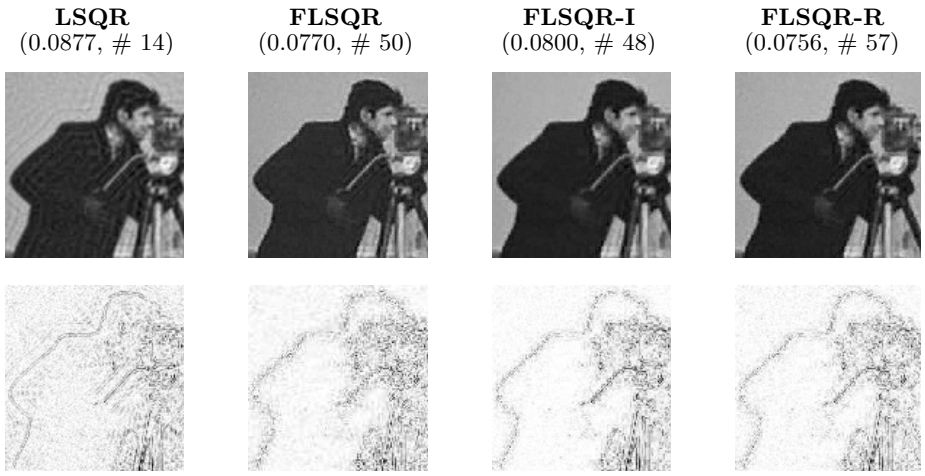


FIG. 11. *Experiment* 2: *Reconstructed subimages corresponding to the smallest relative reconstruction error norm for Golub–Kahan-based methods, along with absolute error subimages $|\mathbf{x}_k - \mathbf{x}_{\text{true}}|$ in inverted colormap (where white corresponds to small absolute error component). Relative reconstruction error norms and corresponding iteration numbers are reported in the titles.*

tion errors compared to the standard solvers and the reconstructions are improved, as evident in the displayed images. Subimages of the best reconstructions computed by Golub–Kahan-based methods are provided in Figure 11, along with the absolute error subimages $|\mathbf{x}_k - \mathbf{x}_{\text{true}}|$, for some values of k. The smallest relative error norm and the iteration number (preceded by #) are reported in brackets. We observe that although the relative error norms are comparable, the flexible methods better capture the flat regions of the image.

Next we compare FLSQR-R to the GAT method applied to the transformed problem, as well as to FISTA on the transformed problem. Relative reconstruction error norms are provided in Figure 12. Here, the automatically computed regularization parameter (i.e., the one selected by FLSQR-R upon fulfillment of the discrepancy principle) is $7.46 \cdot 10^{-5}$, but it seems too small for FISTA. Thus, we also provide in FISTA opt the results for FISTA with the optimal regularization parameter 0.1, which

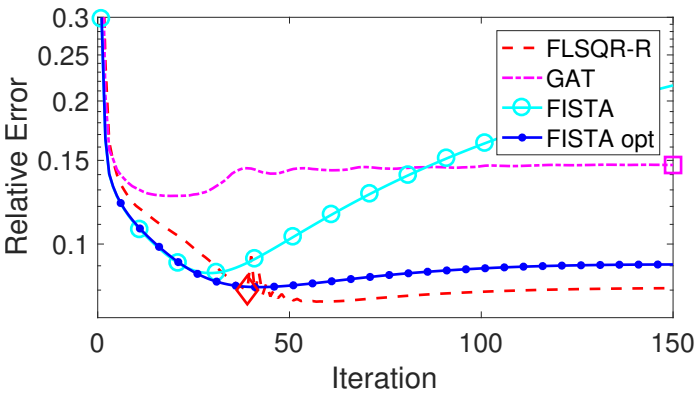


FIG. 12. *Experiment* 2: *Relative reconstruction error norms are provided to compare the FGK methods to some existing methods. FISTA uses the regularization parameter selected by FLSQR-R, and FISTA opt uses a regularization parameter that was found empirically using the true image.*

is determined by searching over 10 logarithmically equispaced values between 10^{-3} and 1, and selecting the one delivering the smallest final relative reconstruction error norm. We observe that for a good choice of the regularization parameter, FISTA reconstructions are similar to ours; however, for poor choices of the regularization parameter, FISTA reconstructions are either too blocky or contaminated with noise. The behavior of GAT is due to a poor automatically chosen regularization parameter.

Experiment 3. We consider a sparse X-ray tomographic reconstruction example with undersampled data. The goal of this experiment is to assess the performance of the new solvers based on the FGK decomposition for solving the transformed ℓ_1-regularized problem (2), where $\mathbf{A}$ is underdetermined and $\mathbf{\Psi}$ represents a 2D Haar wavelet transform with four levels. In [21] it is empirically shown that the compressive sensing theory applies when performing standard structured undersampling patterns and when solving either the ℓ_1 or the total variation regularized problems. The test problem considered here takes a vectorization of the well-known Shepp–Logan phantom as the exact solution $\mathbf{x}_{\text{true}}$; only roughly 40% of the pixels of the transformed exact solution $\mathbf{\Psi x}_{\text{true}}$ are numerically nonzero. A fairly underdetermined sparse matrix $\mathbf{A}$ of size 32580×65536 (i.e., roughly 50% undersampling) is generated using the `paralleltomo` function from AIR Tools II [19], which models a 2D equidistant parallel-beam scanning geometry, with the following parameters:

```
N = 256, theta = 0:2:179, p = round(sqrt(2)*N), d = sqrt(2)*N .
```

Here $\texttt{N}^2$ is the number of pixels of the phantom, `p` is the number of pixels of the detector, the source-detector pair is rotated at angles of projection `theta`, and `d` is the distance between the first and the last ray. Note that, with such undersampling and sparsity, and according to [21], recovery should be experimentally guaranteed. Gaussian white noise of level 10^{-2} is added to the exact data.

Figure 13 displays the history of the relative error norms associated with different purely iterative regularization methods (i.e., with $\lambda = 0$ in (23)): since we are dealing with a rectangular matrix, only LSQR and LSMR together with their flexible versions are considered. We can clearly see the benefits of introducing flexibility into the solution subspaces: indeed, a greater accuracy is achieved by the flexible methods (with a computational cost comparable to the standard solvers), together with a less pronounced semiconvergence. (This is particularly true for FLSMR, in accordance

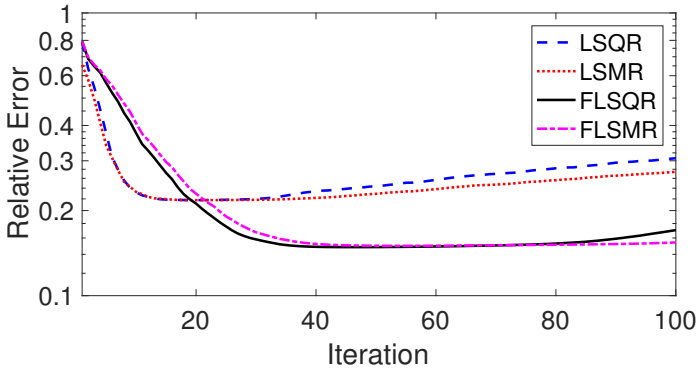


FIG. 13. *Experiment* 3: *History of the relative error norms, considering purely iterative Golub–Kahan-based methods.*

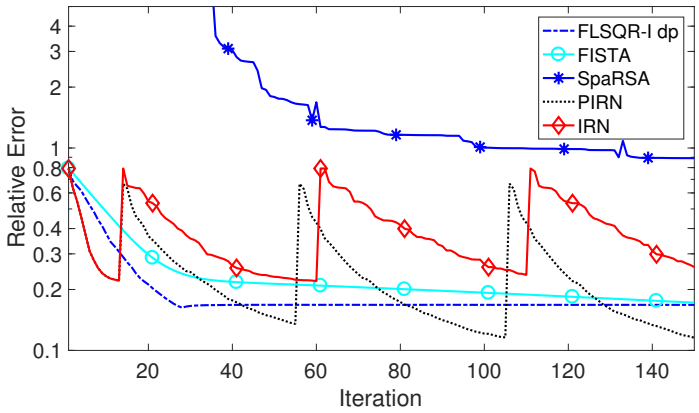


FIG. 14. *Experiment* 3: *History of the relative error norms, comparing the FLSQR-I method to FISTA, SpaRSA, IRN, and PIRN.*

with the observations in [5].) The only potential drawback is the doubling of the storage requirements for FGK compared to GKB, but this is not a serious concern if the required number of iterations k is relatively small (as it is for all of the presented test problems).

Figure 14 displays the history of the relative error norms when the FLSQR-I dp method is employed (with the regularization parameter chosen at each iteration by the discrepancy principle) and compares it to other solvers for (23). In particular, we compare with FISTA, SpaRSA, IRN, and PIRN. As already remarked, all of these well-established solvers require the regularization parameter λ to be set at the beginning of the iterative process: for this experiment we choose $\lambda = 3.6 \cdot 10^{-1}$, which is the value computed by the classical discrepancy principle at the end of the FLSQR-I dp iterations (when also some stabilization occurred in the iteration-dependent values of the regularization parameter). We can clearly see that SpaRSA does not perform well for this problem, and a more accurate tuning of the regularization parameter may improve its reconstruction. FISTA requires more iterations than FLSQR-I dp to compute reconstructions of similar quality. The reconstruction performances of both SpaRSA and FISTA depend heavily on the choice of the regularization param-

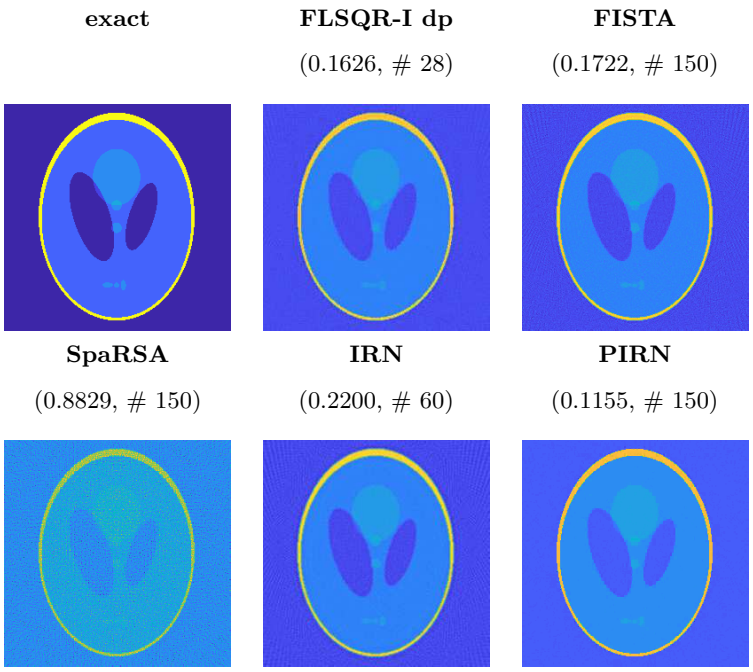


FIG. 15. *Experiment* 3: *Reconstructed subimages of best quality for various solvers. Smallest attained relative reconstruction error norms (up to* 150 *iterations) and corresponding iteration numbers (preceeded by #) are reported.*

eter. Of the considered methods, the PIRN method results in the smallest relative reconstruction error norms, but it requires more iterations than FLSQR-I dp to reach an optimal accuracy. The quality of the reconstruction does not significantly improve when additional PIRN or IRN iterations are performed. We do not show the behavior of the FLSQR-R, FLSMR-I, and FLSMR-R hybrid methods as they are very similar to the FLSQR-I method for this problem.

Figure 15 shows the best reconstructions computed by each method considered in Figure 14. The best relative error and the iteration number (preceded by #) are reported in brackets. Again, we remark that the computational cost for each iteration of these methods is dominated by a matrix-vector product with $\mathbf{A}$ and one with $\mathbf{A}^\top$.

6. Conclusions and future work. In this paper, we describe flexible hybrid iterative methods for computing approximate solutions to the (transformed) ℓ_p-regularized problem, for $p \geq 1$. To handle general (nonsquare) ℓ_p-regularized least-squares problems, we introduce a flexible Golub–Kahan approach and exploit it within a Krylov–Tikhonov hybrid framework. Theoretical results show that the iterates correspond to solutions of a full-dimensional Tikhonov problem that has been projected onto flexible Krylov subspaces of increasing dimensions. We describe various extensions for effectively computing solutions that are sparse with respect to some invertible transformation. Our proposed methods are *efficient* in that they can access $\mathbf{A}$ and $\mathbf{A}^\top$ as function evaluations and they avoid inner-outer schemes, and *automatic* in that parameters such as regularization parameters and stopping iterations can be naturally selected within a hybrid framework. Numerical results validate these observations.

Future work includes extensions to problems where $\mathbf{\Psi}$ is not invertible, and also to nonlinear regularization functionals (e.g., total variation) and nonconvex problems. Developing theoretical convergence results for flexible methods requires additional investigation and would also apply to other solvers based on flexible preconditioning, e.g., [10, 12]. Furthermore, by incorporating multilevel decompositions, these flexible hybrid methods can be exploited in a multiparameter regularization framework, where a different sparsity regularization parameter is incorporated for each level.

Acknowledgments. The authors are grateful to the anonymous referees for their detailed reading of the manuscript and for providing insightful remarks that helped to improve the paper. Also, the authors would like to thank the Isaac Newton Institute for Mathematical Sciences for support and hospitality during the program "Variational Methods and Effective Algorithms for Imaging and Vision" when work on this paper was undertaken.

REFERENCES

[1] S. Arridge, M. Betcke, and L. Harhanen, *Iterated preconditioned LSQR method for inverse problems on unstructured grids*, Inverse Problems, 30 (2014), 075009, https://doi.org/10.1088/0266-5611/30/7/075009.

[2] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2 (2009), pp. 183–202, https://doi.org/10.1137/080716542.

[3] M. Belge, M. E. Kilmer, and E. L. Miller, *Wavelet domain image restoration with adaptive edge-preserving regularization*, IEEE Trans. Image Process., 9 (2000), pp. 597–608, https://doi.org/10.1109/83.841937.

[4] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, Philadelphia, 1996, https://doi.org/10.1137/1.9781611971484.

[5] J. Chung and K. Palmer, *A hybrid LSMR algorithm for large-scale Tikhonov regularization*, SIAM J. Sci. Comput., 37 (2015), pp. S562–S580, https://doi.org/10.1137/140975024.

[6] I. Daubechies, M. Defrise, and C. De Mol, *An iterative thresholding algorithm for linear inverse problems with a sparsity constraint*, Comm. Pure Appl. Math., 57 (2004), pp. 1413–1457, https://doi.org/10.1002/cpa.20042.

[7] M. I. Espanol and M. E. Kilmer, *A wavelet-based multilevel approach for blind deconvolution problems*, SIAM J. Sci. Comput., 36 (2014), pp. A1432–A1450, https://doi.org/10.1137/130928716.

[8] D. C.-L. Fong and M. Saunders, *LSMR: An iterative algorithm for sparse least-squares problems*, SIAM J. Sci. Comput., 33 (2011), pp. 2950–2971, https://doi.org/10.1137/10079687X.

[9] S. Gazzola, P. C. Hansen, and J. G. Nagy, *IR Tools: A MATLAB package of iterative regularization methods and large-scale test problems*, Numer. Algorithms, 2018, https://doi.org/10.1007/s11075-018-0570-7.

[10] S. Gazzola and J. G. Nagy, *Generalized Arnoldi–Tikhonov method for sparse reconstruction*, SIAM J. Sci. Comput., 36 (2014), pp. B225–B247, https://doi.org/10.1137/130917673.

[11] S. Gazzola, P. Novati, and M. R. Russo, *On Krylov projection methods and Tikhonov regularization*, Electron. Trans. Numer. Anal., 44 (2015), pp. 83–123.

[12] S. Gazzola and Y. Wiaux, *Fast nonnegative least squares through flexible Krylov subspaces*, SIAM J. Sci. Comput., 39 (2017), pp. A655–A679, https://doi.org/10.1137/15M1048872.

[13] R. Giryes, M. Elad, and Y. C. Eldar, *The projected GSURE for automatic parameter tuning in iterative shrinkage methods*, Appl. Comput. Harmon. Anal., 30 (2011), pp. 407–422, https://doi.org/10.1016/j.acha.2010.11.005.

[14] T. Goldstein and S. Osher, *The split Bregman method for l1-regularized problems*, SIAM J. Imaging Sci., 2 (2009), pp. 323–343, https://doi.org/10.1137/080725891.

[15] G. Golub and W. Kahan, *Calculating the singular values and pseudo-inverse of a matrix*, SIAM Ser. B Numer. Anal., 2 (1965), pp. 205–224, https://doi.org/10.1137/0702016.

[16] I. F. Gorodnitsky and B. D. Rao, *A new iterative weighted norm minimization algorithm and its applications*, in Proceedings of the IEEE Sixth SP Workshop on Statistical Signal and Array Processing, 1992, pp. 412–415, https://doi.org/10.1109/SSAP.1992.246872.

[17] P. C. Hansen, *Regularization tools: A MATLAB package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 1–35, https://doi.org/10.1007/BF02149761.

[18] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*, SIAM, Philadelphia, 2010, https://doi.org/10.1137/1.9780898718836.

[19] P. C. Hansen and J. S. Jorgensen, *AIR Tools* II: *Algebraic iterative reconstruction methods, improved implementation*, Numer. Algorithms, 79 (2018), pp. 107–137, https://doi.org/10.1007/s11075-017-0430-x.

[20] G. Huang, A. Lanza, S. Morigi, L. Reichel, and F. Sgallari, *Majorization–minimization generalized Krylov subspace methods for $\ell_p - \ell_q$ optimization applied to image restoration*, BIT, 57 (2017), pp. 351–378, https://doi.org/10.1007/s10543-016-0643-8.

[21] J. S. Jorgensen and E. Y. Sidky, *How little data is enough? Phase-diagram analysis of sparsity-regularized X-ray computed tomography*, Philos. Trans. A, 373 (2015), 20140387, https://doi.org/10.1098/rsta.2014.0387.

[22] M. E. Kilmer and D. P. O'Leary, *Choosing regularization parameters in iterative methods for ill-posed problems*, SIAM J. Matrix Anal. Appl., 22 (2001), pp. 1204–1221, https://doi.org/10.1137/S0895479899345960.

[23] E. Klann, R. Ramlau, and L. Reichel, *Wavelet-based multilevel methods for linear ill-posed problems*, BIT, 51 (2011), pp. 669–694, https://doi.org/10.1007/s10543-011-0320-x.

[24] A. Lanza, S. Morigi, L. Reichel, and F. Sgallari, *A generalized Krylov subspace method for $\ell_p - \ell_q$ minimization*, SIAM J. Sci. Comput., 37 (2015), pp. S30–S50, https://doi.org/10.1137/140967982.

[25] A. Lanza, S. Morigi, I. Selesnick, and F. Sgallari, *Nonconvex nonsmooth optimization via convex—nonconvex majorization—minimization*, Numer. Math., 136 (2017), pp. 343–381, https://doi.org/10.1007/s00211-016-0842-x.

[26] K. Morikuni and K. Hayami, *Convergence of inner-iteration GMRES methods for rank-deficient least squares problems*, SIAM J. Matrix Anal. Appl., 36 (2015), pp. 225–250, https://doi.org/10.1137/130946009.

[27] J. G. Nagy, K. Palmer, and L. Perrone, *Iterative methods for image deblurring: A MATLAB object-oriented approach*, Numer. Algorithms, 36 (2004), pp. 73–93, https://doi.org/10.1023/B:NUMA.0000027762.08431.64.

[28] D. P. O'Leary and J. A. Simmons, *A bidiagonalization-regularization procedure for large scale discretizations of ill-posed problems*, SIAM J. Sci. and Stat. Comput., 2 (1981), pp. 474–489, https://doi.org/10.1137/0902037.

[29] C. C. Paige and M. A. Saunders, *Algorithm* 583*: LSQR: Sparse linear equations and least squares problems*, ACM Trans. Math. Software, 8 (1982), pp. 195–209, https://doi.org/10.1145/355993.356000.

[30] C. C. Paige and M. A. Saunders, *LSQR: An algorithm for sparse linear equations and sparse least squares*, ACM Trans. Math. Software, 8 (1982), pp. 43–71, https://doi.org/10.1145/355984.355989.

[31] R. A. Renaut, S. Vatankhah, and V. E. Ardestani, *Hybrid and iteratively reweighted regularization by unbiased predictive risk and weighted GCV for projected systems*, SIAM J. Sci. Comput., 39 (2017), pp. B221–B243, https://doi.org/10.1137/15M1037925.

[32] P. Rodriguez and B. Wohlberg, *An efficient algorithm for sparse representations with ℓ^p data fidelity term*, in Proceedings of 4th IEEE Andean Technical Conference (ANDESCON), 2008.

[33] Y. Saad, *On the rates of convergence of the Lanczos and the block-Lanczos methods*, SIAM J. Numer. Anal., 17 (1980), pp. 687–706, https://doi.org/10.1137/0717059.

[34] Y. Saad, *A flexible inner-outer preconditioned GMRES algorithm*, SIAM J. Sci. Comput., 14 (1993), pp. 461–469, https://doi.org/10.1137/0914028.

[35] A. K. Saibaba, T. Bakhos, and P. K. Kitanidis, *A flexible Krylov solver for shifted systems with application to oscillatory hydraulic tomography*, SIAM J. Sci. Comput., 35 (2013), pp. A3001–A3023, https://doi.org/10.1137/120902690.

[36] V. Simoncini and D. B. Szyld, *Flexible inner-outer Krylov subspace methods*, SIAM J. Numer. Anal., 40 (2002), pp. 2219–2239, https://doi.org/10.1137/s0036142902401074.

[37] V. Simoncini and D. B. Szyld, *Recent computational developments in Krylov subspace methods for linear systems*, Numer. Linear Algebra Appl., 14 (2007), pp. 1–59, https://doi.org/10.1002/nla.499.

[38] L. Tenorio, *An Introduction to Data Analysis and Uncertainty Quantification for Inverse Problems*, SIAM, Philadelphia, 2017, https://doi.org/10.1137/1.9781611974928.

[39] J. A. Tropp and S. J. Wright, *Computational methods for sparse solution of linear inverse problems*, Proc. IEEE, 98 (2010), pp. 948–958, https://doi.org/10.21236/ada633835.

[40] J. Van Den Eshof and G. L. Sleijpen, *Inexact Krylov subspace methods for linear systems*, SIAM J. Matrix Anal. Appl., 26 (2004), pp. 125–153, https://doi.org/10.1137/s0895479802403459.

[41] S. Vatankhah, R. A. Renaut, and V. E. Ardestani, 3-*D projected L*1 *inversion of gravity data using truncated unbiased predictive risk estimator for regularization parameter estimation*, Geophys. J. Internat., 210 (2017), pp. 1872–1887, https://doi.org/10.1093/gji/ggx274.

[42] S. J. Wright, R. D. Nowak, and M. A. Figueiredo, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Process., 57 (2009), pp. 2479–2493, https://doi.org/10.1109/icassp.2008.4518374.