Multidirectional subspace expansion for single-parameter and multi-parameter Tikhonov regularization*

Ian N. Zwaan† Michiel E. Hochstenbach†

January 15, 2016

Abstract. Tikhonov regularization is a popular method to approximate solutions of linear discrete ill-posed problems when the observed or measured data is contaminated by noise. Multi-parameter Tikhonov regularization may improve the quality of the computed approximate solutions. We propose a new iterative method for large-scale multi-parameter Tikhonov regularization with general regularization operators based on a multidirectional subspace expansion. This expansion may be combined with subspace truncation to avoid excessive growth of the search space. Furthermore, we introduce a simple and effective parameter selection strategy based on the discrepancy principle and related to perturbation results.

Key words. Tikhonov, multi-parameter Tikhonov, generalized Krylov, multidirectional subspace expansion, subspace truncation, subspace method, linear discrete ill-posed problem, regularization, regularization parameter.

AMS subject classification. 15A29, 65F10, 65F22, 65F30, 65R30, 65R32

1. Introduction. We consider single-parameter and multi-parameter Tikhonov regularization problems of the form

$$\text{(1)}\qquad \operatorname*{argmin}_{\boldsymbol{x}} \|A\boldsymbol{x}-\boldsymbol{b}\|^2+\sum_{i=1}^{\ell}\mu^i\|L^i\boldsymbol{x}\|^2 \qquad (\ell\geq 1),$$

where $\|\cdot\|$ denotes the 2-norm and the superscript i is used as an index. We focus on large-scale discrete ill-posed problems such as the discretization of Fredholm integral equations of the first kind. More precisely, suppose A is an ill-conditioned or even singular $m\times n$ matrix with $m\geq n$ and L^i are $p^i\times n$ matrices such that the nullspaces of A and L^i intersect trivially. Let μ^i be nonnegative regularization parameters, and assume $\boldsymbol{b}$ is contaminated by an error $\boldsymbol{e}$ and satisfies $\boldsymbol{b}=A\boldsymbol{x}_\star+\boldsymbol{e}$, where $\boldsymbol{x}_\star$ is the exact solution. Finally, assume that a bound $\|\boldsymbol{e}\|\leq\varepsilon$ is available, so that the discrepancy principle (see, e.g., [8, Sect. 7.2]) can be used.

In single-parameter Tikhonov regularization ($\ell=1$), the choice of the regularization operator is typically significant, since frequencies in the nullspace of the operator remain unpenalized. Multi-parameter Tikhonov can be used when a satisfactory choice of the regularization operator is unknown in advance, or can be seen as an attempt to combine the strengths of different regularization operators. In some applications, using more than one regularization operator and parameter allows for more accurate solutions [1, 2, 12, 14].

Solving (1) for large-scale problems may be challenging. In case the μ^i are fixed *a priori*, methods such as LSQR [15] or LSMR [4] may be used. However, the problem becomes more complicated when the regularization parameters are not fixed in advance [9, 11, 12]. In this paper, we present a new subspace method consisting of three phases; a new expansion phase, an extraction phase, and a new truncation phase. To be more precise, let $\mathcal{X}_k\subset\mathbb{R}^n$ be a subspace of dimension $k\ll n$ with orthonormal basis X_k. Then we can compute matrix decompositions

$$\text{(2)}\qquad \begin{aligned} AX_k &= U_{k+1}\underline{H}_k \\ L^iX_k &= V_k^iK_k^i \qquad (i=1,2,\ldots,\ell),\end{aligned}$$

where U_{k+1} and V_k^i are orthonormal, $\beta\boldsymbol{u}_1=\boldsymbol{b}$, $\beta=\|\boldsymbol{b}\|$, $\underline{H}_k$ is a $(k+1)\times k$ Hessenberg matrix, and K_k^i is upper triangular. Denote $\boldsymbol{\mu}=(\mu^1,\ldots,\mu^\ell)$ for convenience. Now restrict the solution space to $\mathcal{X}_k$ so that $\boldsymbol{x}_k(\boldsymbol{\mu})=X_k\boldsymbol{c}_k(\boldsymbol{\mu})$, where

$$\text{(3)}\qquad \boldsymbol{c}_k(\boldsymbol{\mu})=\operatorname*{argmin}_{\boldsymbol{c}}\|AX_k\boldsymbol{c}-\boldsymbol{b}\|^2+\sum_{i=1}^{\ell}\mu^i\|L^iX_k\boldsymbol{c}\|^2=\operatorname*{argmin}_{\boldsymbol{c}}\|\underline{H}_k\boldsymbol{c}-\beta\boldsymbol{e}_1\|^2+\sum_{i=1}^{\ell}\mu^i\|K_k^i\boldsymbol{c}\|^2.$$

*This work is part of the research program "Innovative methods for large matrix problems", which is (partly) financed by the Netherlands Organization for Scientific Research (NWO).

†Department of Mathematics and Computer Science, TU Eindhoven, PO Box 513, NL-5600 MB Eindhoven, The Netherlands (www.ianzwaan.com, www.win.tue.nl/~hochsten).

The vector $\boldsymbol{e}_1$ is the first standard basis vector of appropriate length. Our paper has three contributions. First, a new expansion phase where we add multiple search directions to $\mathcal{X}_k$. Second, a new truncation phase which removes unwanted new search directions. Third, the selection of the regularization parameters μ_k^i in the extraction phase. The three phases work alongside each other: the intermediate solution obtained in the extraction phase is preserved in the truncation phase, whereas remaining perpendicular from the expansion phase are removed.

The paper is organized as follows. In Section 2 an existing nonlinear subspace method is discussed, whereafter we propose the new multidirectional subspace expansion of the expansion phase. Discussion of the truncation phase follows immediately. Sections 3 and 4 describe the extraction phase. In the former, a straightforward parameter selection strategy for multi-parameter regularization is given, in the latter, a justification using perturbation analysis. Numerical experiments are performed in Section 5 and demonstrate the competitiveness of our new method. We end with concluding remarks in Section 6.

2. Subspace expansion for multi-parameter Tikhonov. Let us first consider single-parameter Tikhonov regularization with a general regularization operator. Then $\ell = 1$ and we write $\mu = \mu^1$, $L = L^i$, and $K_k = K_k^1$, such that (1) simplifies to

$$\operatorname*{argmin}_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu\|L\boldsymbol{x}\|^2.$$

When $L = I$ we use the Golub–Kahan–Lanczos bidiagonalization procedure (see, e.g., [6, Sect. 9.3.3]) to generate the Krylov subspace

$$\mathcal{X}_k = \mathcal{K}_k(A^*A, A^*\boldsymbol{b}) = \operatorname{span}\{A^*\boldsymbol{b}, (A^*A)A^*\boldsymbol{b}, \ldots, (A^*A)^{k-1}A^*\boldsymbol{b}\}.$$

In this case $\underline{H}_k$ is lower bidiagonal and K_k is the identity and

$$\boldsymbol{x}_{k+1} = \frac{(I - X_k X_k^*)A^*\boldsymbol{u}_{k+1}}{\|(I - X_k X_k^*)A^*\boldsymbol{u}_{k+1}\|}$$

However, if $L \neq I$ it may be more natural to consider a shift-independent generalized Krylov subspace of the form

$$\mathcal{X}_k = \mathcal{K}_k(A^*A, L^*L, A^*b),$$

spanned by the first k vectors in

$$\begin{array}{ll} \text{Group 0} & A^*\boldsymbol{b} \\ \text{Group 1} & (A^*A)A^*\boldsymbol{b}, (L^*L)A^*\boldsymbol{b} \\ \text{Group 2} & (A^*A)^2A^*\boldsymbol{b}, (A^*A)(L^*L)A^*\boldsymbol{b}, (L^*L)(A^*A)A^*\boldsymbol{b}, (L^*L)^2A^*\boldsymbol{b} \\ \quad \ldots & \end{array}$$

This has been studied studied by Li and Ye [13] and later by Reichel, Sgallari, and Ye [17]. An orthonormal basis can be created with a generalization of Golub–Kahan–Lanczos bidiagonalization [10]. However, while the search space grows linearly as a function of the number of matrix-vector products, the dimension of the generalized Krylov subspace grows exponentially as a function of the total degree of a bivariate matrix polynomial. As a result, if we take any vector $\boldsymbol{x} \in \mathcal{K}_k(A^*A, L^*L, A^*\boldsymbol{b})$ and write it as $p(A^*A, L^*L)A^*\boldsymbol{b}$, where p is a bivariate polynomial, then p has at most degree $\lfloor \log_2 k \rfloor$. This low degree may be undesirable especially for small regularization parameters μ.

An alternative approach is a greedy nonlinear method described by Lampe, Reichel, and Voss [12]. We briefly review their method and state a straightforward extension to multi-parameter Tikhonov regularization. Subsequently we present our new multidirectional approach.

Consider again the single-parameter case, then the low dimensional minimization (3) simplifies to

$$\boldsymbol{c}_k(\mu) = \operatorname*{argmin}_{\boldsymbol{c}} \|\underline{H}_k\boldsymbol{c} - \beta\boldsymbol{e}_1\|^2 + \mu\|K_k\boldsymbol{c}\|^2.$$

Next, choose a fixed value $\mu = \mu_k$ using, e.g., the discrepancy principle. It is easy to verify that

$$A^*\boldsymbol{b} - (A^*A + \mu_k L^*L)\boldsymbol{x}_k(\mu_k) = A^*U_{k+1}(\beta\boldsymbol{e}_1 - \underline{H}_k\boldsymbol{c}_k(\mu_k)) + \mu_k L^*V_k K_k\boldsymbol{c}_k(\mu_k)$$

is perpendicular to $\mathcal{X}_k$; this vector expands the search space. As usual, expansion and extraction repeat until suitable stopping criteria are met. Remark that the vector above is the gradient of the cost function

$$\boldsymbol{x} \mapsto \frac{1}{2}(\|A\boldsymbol{x}-\boldsymbol{b}\|^2+\mu\|L\boldsymbol{x}\|^2)$$

in the point $\boldsymbol{x}_k(\mu_k)$.

As previously stated, Lampe, Reichel, and Voss [12] consider only single-parameter Tikhonov regularization, however, their method readily extends to multi-parameter Tikhonov regularization. Again, the first step is to decide on regularization parameters $\boldsymbol{\mu}_k$. Next, use the residual of the normal equations

$$A^*\boldsymbol{b}-\Big(A^*A+\sum_{i=1}^{\ell}\mu_k^i L^{i^*}L^i\Big)\boldsymbol{x}_k(\boldsymbol{\mu}_k)=A^*U_{k+1}(\beta\boldsymbol{e}_1-\underline{H}_k\boldsymbol{c}_k(\boldsymbol{\mu}_k))-\sum_{i=1}^{\ell}\mu_k^i L^{i^*}V_k^iK_k^i\boldsymbol{c}_k(\boldsymbol{\mu}_k),$$

to expand the search space. Note that the residual is again orthogonal to $\mathcal{X}_k$ as well as the gradient of the cost function

$$\boldsymbol{x} \mapsto \frac{1}{2}(\|A\boldsymbol{x}-\boldsymbol{b}\|^2+\sum_{i=1}^{\ell}\mu^i\|L^i\boldsymbol{x}\|^2).$$

We summarize the method in Algorithm 1. In practice we initially use Golub–Kahan–Lanczos bidiagonalization and expand the search space with $A^*\boldsymbol{u}_k$ until a $\boldsymbol{\mu}_k$ can be found which satisfies the discrepancy principle.

Algorithm 1 (Generalized Krylov Subspace Tikhonov Regularization extension of [12])**.**
Input: Measurement matrix A, regularization operators $L^1,\dots,L^\ell$, and data $\boldsymbol{b}$.
Output: Approximate solution $\boldsymbol{x}_k \approx \boldsymbol{x}_\star$.

1. Initialize $\beta=\|\boldsymbol{b}\|$, $U_1=\boldsymbol{b}/\beta$, $X_0=[\,]$, $\boldsymbol{x}_0=\boldsymbol{0}$, and $\boldsymbol{\mu}_0=\boldsymbol{0}$.
   for $k=1,2,\dots$ **do**
2. Expand X_{k-1} with $A^*\boldsymbol{b}-(A^*A+\sum_{i=1}^{\ell}\mu_{k-1}^i L^{i^*}L^i)\boldsymbol{x}_{k-1}$.
3. Update $AX_k=U_{k+1}\underline{H}_k$ and $L^iX_k=V_k^iK_k^i$.
4. Select $\boldsymbol{\mu}_k$; see for example Section 3.
5. $\boldsymbol{c}_k=\operatorname{argmin}_{\boldsymbol{c}}\big\|\big[\underline{H}_k;\sqrt{\mu_k^1}K_k^1;\dots;\sqrt{\mu_k^\ell}K_k^\ell\big]\boldsymbol{c}-\beta\boldsymbol{e}_1\big\|$.
6. $\boldsymbol{x}_k=X_k\boldsymbol{c}_k$.

Suitable regularization operators often depend on the problem and its solution. Multi-parameter regularization may be used when a priori information is lacking. In this case, it is not obvious that the residual vector above is a "good" expansion vector. In particular if the intermediate regularization parameters $\boldsymbol{\mu}_k$ are not necessarily accurate. Inspired by the generalized Krylov subspace approach, we can remove the dependence on the parameters to some extend by expanding the search space with the vectors

$$A^*A\boldsymbol{x}_k(\boldsymbol{\mu}_k),\quad L^{1^*}L^1\boldsymbol{x}_k(\boldsymbol{\mu}_k),\quad\dots,\quad L^{\ell^*}L^\ell\boldsymbol{x}_k(\boldsymbol{\mu}_k),\tag{4}$$

separately. Here, we omit $A^*\boldsymbol{b}$ as it is contained in X_k. Since we expand the search space in multiple directions, we refer to this expansion as a "multidirectional" subspace expansion. Remark that the multidirectional expansion contains the previous residual expansion.

It is unappealing for the search space to grow with $\ell+1$ basis vectors per iteration, because the cost of orthogonalization and the cost of solving the projected problems depend on the dimension of the search space. So, we wish to condense the best portions of the multiple directions in a single vector. We now give a more detailed description. Suppose we expand X_k with the vectors in (4) and obtain $\widetilde{X}_{k+\ell+1}$. Then we compute the decompositions

$$\begin{aligned}A\widetilde{X}_{k+\ell+1}&=\widetilde{U}_{k+\ell+2}\underline{\widetilde{H}}_{k+\ell+1}\\ L^i\widetilde{X}_{k+\ell+1}&=\widetilde{V}^i_{k+\ell+1}\widetilde{K}^i_{k+\ell+1}\qquad(i=1,2,\dots,\ell),\end{aligned}$$

analogous to (2) and determine parameters $\boldsymbol{\mu}_{k+1}$ and the approximate solution $\widetilde{\boldsymbol{c}}_{k+\ell+1}$. Next, we compute

$$\begin{aligned}A(\widetilde{X}_{k+\ell+1}Z^*)&=(\widetilde{U}_{k+\ell+2}P^*)(P\underline{\widetilde{H}}_{k+\ell+1}Z^*)\\ L^i(\widetilde{X}_{k+\ell+1}Z^*)&=(\widetilde{V}^i_{k+\ell+1}Q^{i*})(Q^i\widetilde{K}^i_{k+\ell+1}Z^*)\qquad(i=1,2,\dots,\ell),\end{aligned}\tag{5}$$

where Z, P, and Q^i orthonormal matrices of the form

$$Z=\begin{bmatrix}I_k & \\ & Z_{\ell+1}\end{bmatrix},\quad P=\begin{bmatrix}I_{k+1} & \\ & P_{\ell+1}\end{bmatrix},\quad Q^i=\begin{bmatrix}I_k & \\ & Q^i_{\ell+1}\end{bmatrix}.$$

Here I_k is the $k \times k$ identity matrix and $Z_{\ell+1}$ is an orthonormal matrix so that $Z_{\ell+1}\boldsymbol{c}_{k+1:k+\ell+1} = \gamma \boldsymbol{e}_1$. The matrices $P_{\ell+1}$ and $Q^i_{\ell+1}$ are computed to make $\underline{\widetilde{H}}_{k+\ell+1}Z^*$ and $\widetilde{K}^i_{k+\ell+1}Z^*$ respectively upper-Hessenberg and upper-triangular again. At this point we can truncate (5) to obtain

$$\begin{aligned} AX_{k+1} &= U_{k+2}\underline{H}_{k+1} \\ L^iX_{k+1} &= V^i_{k+1}K^i_{k+1} \qquad (i = 1, 2, \ldots, \ell), \end{aligned}$$

so that $\widetilde{X}_{k+\ell+1}\widetilde{\boldsymbol{c}}_{k+\ell+1} \subset \operatorname{span} X_{k+1} = \mathcal{X}_{k+1}$.

To illustrate our approach, let us consider an example where $\ell = 1$. Suppose we expand X_1 with vectors A^*AX_1 and L^*LX_1. Let $AX_{1+2} = U_{2+2}\underline{H}_{1+2}$ and $LX_{1+2} = V_{1+2}K_{1+2}$. Compute Z. Now $\underline{H}_{1+2}Z^*$, and $K_{1+2}Z^*$ are no longer upper-Hessenberg and upper-triangular respectively. Therefore we compute P and Q such that $P\underline{H}_{1+2}Z^*$ is again upper-Hessenberg and $QK_{1+2}Z^*$ is upper-triangular. Schematically we have

$$\xrightarrow{\underline{H}_{1+2}} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{\underline{H}_{1+2}Z^*} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{P\underline{H}_{1+2}Z^*} \begin{bmatrix} \times & \times & \times \\ \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix}$$

$$\xrightarrow{K_{1+2}} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{K_{1+2}Z^*} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix} \xrightarrow{QK_{1+2}Z^*} \begin{bmatrix} \times & \times & \times \\ 0 & \times & \times \\ 0 & 0 & \times \end{bmatrix}$$

accompanied by the decompositions

$$\begin{aligned} A(X_{1+2}Z^*) &= (U_{2+2}P^*)(P\underline{H}_{1+2}Z^*) \\ L(X_{1+2}Z^*) &= (V_{1+2}Q^*)(QK_{1+2}Z^*). \end{aligned}$$

At this point we can truncate the subspaces by removing the last columns from $(X_{1+2}Z^*)$, $(U_{2+2}P^*)$, $(P\underline{H}_{1+2}Z^*)$, $(X_{1+2}Z^*)$, $(V_{1+2}Q^*)$, and $(QK_{1+2}Z^*)$, as well as the bottom rows of $(P\underline{H}_{1+2}Z^*)$ and $(QK_{1+2}Z^*)$.

Below we summarize the steps of the new algorithm for solving problem (1). In our implementation we take care to use full reorthogonalization and avoid extending X_{k-1}, U_{k+1}, and V^i_k with numerically linearly dependent vectors. We omit these steps from the pseudocode for brevity. In addition, we initially expand the search space solely with $A^*\boldsymbol{u}_{k+1}$ until the discrepancy principle can be satisfied.

Algorithm 2 (Multidirectional Tikhonov regularization)**.**
Input: Measurement matrix A, regularization operators. $L^1, \ldots, L^\ell$, and data $\boldsymbol{b}$.
Output: Approximate solution $\boldsymbol{x}_k \approx \boldsymbol{x}_\star$.

1. Initialize $\beta = \|\boldsymbol{b}\|$, $U_1 = \boldsymbol{b}/\beta$, $X_0 = [\,]$, $\boldsymbol{x}_0 = \boldsymbol{0}$, and $\boldsymbol{\mu}_0 = \boldsymbol{0}$.
   for $k = 0, 1, \ldots,$ **do**
2. Expand X_k with $A^*A\boldsymbol{x}_k, L^{1*}L^1\boldsymbol{x}_k, \ldots, L^{\ell*}L^\ell\boldsymbol{x}_k$.
3. Update $AX_{k+\ell+1} = U_{k+\ell+2}\underline{H}_{k+\ell+1}$ and $L^iX_{k+\ell+1} = V^i_{k+\ell+1}K^i_{k+\ell+1}$.
4. Select $\boldsymbol{\mu}_k$; see for example Section 3.
5. $\boldsymbol{c}_{k+\ell+1} = \operatorname{argmin}_{\boldsymbol{c}} \left\| \left[\underline{H}_k; \sqrt{\mu^1_k}K^1_{k+\ell+1}; \ldots; \sqrt{\mu^\ell_k}K^\ell_{k+\ell+1}\right]\boldsymbol{c} - \beta\boldsymbol{e}_1 \right\|$.
6. Compute P, Q, and Z (see text).
7. Truncate $A(X_{k+\ell+1}Z^*) = (U_{k+\ell+2}P^*)(P\underline{H}_{k+\ell+1}Z^*)$ to $AX_{k+1} = U_{k+2}\underline{H}_{k+1}$.
   Truncate $L^i(X_{k+\ell+1}Z^*) = (V^i_{k+\ell+1}Q^{i*})(Q^iK^i_{k+\ell+1}Z^*)$ to $L^iX_{k+1} = V^i_{k+1}K^i_{k+1}$.
8. $\boldsymbol{x}_{k+1} = X_{k+1}\boldsymbol{c}_{k+1}$.

We have completed our discussion of the expansion and truncation phase of our algorithm. In the following section we discuss the extraction phase.

3. A multi-parameter selection strategy. Parameter selection for single-parameter Tikhonov has been studied extensively. Hence, when $\ell = 1$, one may apply methods such as the discrepancy princple, the L-curve criterion, generalized cross validation, etc. to select the parameter μ^1_k. See, for example, Hansen [8, Sect. 7] for more information. Unfortunately, choosing satisfactory μ^i_k in multi-parameter Tikhonov regularization is more difficult than the corresponding single-parameter problem. See for example [1, 2, 5, 14].

In this section we focus on the discrepancy principle, which states that $\boldsymbol{\mu}_k$ must satisfy

$$\|A\boldsymbol{x}_k(\boldsymbol{\mu}_k) - \boldsymbol{b}\| = \eta\varepsilon, \tag{6}$$

where $\|\boldsymbol{e}\| \leq \varepsilon$ and $\eta > 1$ is a user supplied constant independent of ε. In single-parameter Tikhonov regularization where $\ell = 1$, root finding methods can be applied to the function $\varphi(\mu^1) = \|A\boldsymbol{x}_k(\mu^1) - \boldsymbol{b}\|^2 - \eta^2\varepsilon^2$ in order to find a nonnegative solution μ_k^1. A solution exists and is unique under mild conditions, see, for example, [3]. In multi-parameter Tikhonov regularization where $\ell > 1$, solutions are no longer unique and it is not obvious how to choose "good" parameters. Below we will discuss three different approaches.

Brezinski et al. [2] had some success with operators splitting. Substituting $\mu^i = \omega_k^i \widetilde{\mu}_k^i$ in (3) with nonnegative ω_k^i and $\sum_{i=1}^{\ell} \omega_k^i = 1$ leads to

$$\operatorname*{argmin}_{\boldsymbol{c}} \sum_{i=1}^{\ell} \omega_k^i (\|\underline{H}_k \boldsymbol{c} - \beta \boldsymbol{e}_1\|^2 + \widetilde{\mu}_k^i \|K_k^i \boldsymbol{c}\|^2).$$

This form of the minimization problem suggests the approximation of $X_k^* \boldsymbol{x}_\star$ by a linear combination of $\boldsymbol{c}_k^i(\widetilde{\mu}_k^i)$, where

$$\boldsymbol{c}_k^i(\mu) = \operatorname*{argmin}_{\boldsymbol{c}} \|\underline{H}_k \boldsymbol{c} - \beta \boldsymbol{e}_1\|^2 + \mu \|K_k \boldsymbol{c}\|^2 \qquad (i = 1, 2, \ldots, \ell), \tag{7}$$

and $\widetilde{\mu}_k^i$ is such that $\|\underline{H}_k \boldsymbol{c}_k^i(\widetilde{\mu}_k^i) - \beta \boldsymbol{e}_1\| = \eta\epsilon$. Alternatively, Brezinski et al. [2] consider solving

$$\boldsymbol{c}_k = \operatorname*{argmin}_{\boldsymbol{c}} \left\| \left[\underline{H}_k; \sqrt{\widetilde{\mu}_k^1} K_k^1; \ldots; \sqrt{\widetilde{\mu}_k^\ell} K_k^\ell \right] \boldsymbol{c} - \beta \boldsymbol{e}_1 \right\|,$$

where $\widetilde{\mu}^i$ are fixed and obtained from (7). The latter approach provides better results in exchange for an additional QR decomposition. In either case, operator splitting is a straightforward approach, but does not necessarily satisfy the discrepancy principle.

Lu et al. [14] rewrite the constrained minimization problem as a differential equation and approximate

$$F(\boldsymbol{\mu}) = \|\underline{H}_k \boldsymbol{c}_k(\boldsymbol{\mu}) - \beta \boldsymbol{e}_1\|^2 + \sum_{i=1}^{\ell} \mu^i \|K_k^i \boldsymbol{c}_k(\boldsymbol{\mu})\|^2$$

by a model function $m(\boldsymbol{\mu})$ which admits a straightforward solution to the constructed differential equation. However, it is unclear which $\boldsymbol{\mu}$ the method finds and its solution may depend on the initial guess. On the other hand, it is possible to keep all but one parameter fixed and compute a value for the free parameter such that the discrepancy principle is satisfied. This allows one to trace discrepancy hypersurfaces to some extent.

Gazzola and Novati [5] describe an interesting method. They start with a single-parameter problem and successively add parameters in a novel way until each parameter of the full multi-parameter problem has a value assigned. Especially in early iterations the discrepancy principle is not satisfied, but the parameters are updated in each iteration so that the norm of the residual is expected to approach $\eta\epsilon$. Unfortunately, we observed some issues in our implementation. For example, the quality of the result depends on initial values, as well as the order in which the considers the operators are added (that is, the indexing of the operators).

The methods discussed above compute parameters which approximately satisfy the discrepancy principle. However, the computed $\boldsymbol{\mu}$ are not unique, and which conditions are necessary to obtain a "good" set of parameters is an open question. We suggest a new method where we attempt to nudge the parameters in the right direction. We achieve this with new weights for the operator splitting approach.

Suppose we take $\mu^i = \mu_k \omega_k^i$, where ω_k^i are nonnegative, but do not necessarily sum to one, and μ_k is such that the discrepancy principle is satisfied. Then we obtain

$$\operatorname*{argmin}_{\boldsymbol{c}} \|\underline{H}_k \boldsymbol{c} - \beta \boldsymbol{e}_1\|^2 + \mu_k \sum_{i=1}^{\ell} \omega_k^i \|K_k^i \boldsymbol{c}\|^2. \tag{8}$$

Since the goal of regularization is to reduce sensitivity of the solution to noise, we use the weights

$$\omega_k^i = \frac{\|\boldsymbol{c}_k^i(\widetilde{\mu}_k^i)\|}{\|D_\mu \boldsymbol{c}_k^i(\widetilde{\mu}_k^i)\|}, \tag{9}$$

which bias the regularization parameters in the direction of lower sensitivity. If for some index $\|D_\mu \boldsymbol{c}_k^i(\widetilde{\mu}_k^i)\| = 0$, then we take $\boldsymbol{c}_k^i(\widetilde{\mu}_k^i)$ as the solution, or add a small positive constant. With this parameter choice, the solution is independent of the indexing of the operators, nor, up to a constant, on the scaling of A, $\boldsymbol{b}$, or any of the L^i. The former is easy to see; for the latter, consider the scaled problem

$$\operatorname*{argmin}_{\hat{\boldsymbol{x}}} \|\beta \boldsymbol{b} - \alpha A \hat{\boldsymbol{x}}\|^2 + \mu \sum_{i=1}^{\ell} \hat{\omega}^i \|\lambda^i L^i \hat{\boldsymbol{x}}\|^2$$

The noisy component of $\beta \boldsymbol{b}$ is $\beta \boldsymbol{e}$ and $\|\beta \boldsymbol{e}\| \leq \beta\epsilon$, hence the new discrepancy bound becomes

$$\|\alpha A\hat{\boldsymbol{x}} - \beta \boldsymbol{b}\| = \beta\eta\epsilon,$$

which is satisfied only for $\hat{\boldsymbol{x}} = \beta/\alpha \boldsymbol{x}$ when $\hat{\omega}^i = \alpha^2/(\lambda^i)^2\, \omega^i$ is non-zero for at least one index. In this case we obtain

$$\min_{\hat{\boldsymbol{x}}} \|\beta \boldsymbol{b} - \alpha A\hat{\boldsymbol{x}}\|^2 + \mu \sum_{i=1}^{\ell} \hat{\omega}^i \|\lambda^i L^i \hat{\boldsymbol{x}}\|^2 = \beta^2 \left(\min_{\boldsymbol{x}} \|A\boldsymbol{x} - \boldsymbol{b}\|^2 + \mu \sum_{i=1}^{\ell} \omega^i \|L^i \boldsymbol{x}\|^2 \right).$$

It may be verified that the weights in (9) are indeed proportional to $\alpha^2/(\lambda^i)^2$. There are additional viable choices for ω^i, including two smoothed versions of the above:

$$\omega_k^i = \frac{\|\underline{H}_k \boldsymbol{c}_k^i(\widetilde{\mu}_k^i)\|}{\|\underline{H}_k D_\mu \boldsymbol{c}_k^i(\widetilde{\mu}_k^i)\|} \quad \text{and} \quad \omega_k^i = \frac{\|K_k \boldsymbol{c}_k^i(\widetilde{\mu}_k^i)\|}{\|K_k D_\mu \boldsymbol{c}_k^i(\widetilde{\mu}_k^i)\|}$$

We summarize the new parameter selection in Algorithm 3 below.

Algorithm 3 (Multi-parameter selection)**.**
Input: Projected matrices $\underline{H}_k, K_k^1, \ldots, K_k^\ell$, $\beta = \|\boldsymbol{b}\|$, noise estimate ϵ, uncertainty parameter η.
Output: Regularization parameters $\mu_k^1, \ldots, \mu_k^\ell$.

1. Use (7) to compute $\boldsymbol{c}^i$ and $\widetilde{\mu}_k^i$.
   if $\|D_\mu \boldsymbol{c}_k^i(\widetilde{\mu}_k^i)\| = 0$ for some i **then**
2. Replace $\|D_\mu \boldsymbol{c}_k^i(\widetilde{\mu}_k^i)\|$ by a small positive constant,
   or set $\mu_k^i = \widetilde{\mu}_k^i$ and $\mu_k^j = 0$ for $j \neq i$.
   else
3. Let $\omega_k^i = \|\boldsymbol{c}_k^i(\widetilde{\mu}_k^i)\|/\|D_\mu \boldsymbol{c}_k^i(\widetilde{\mu}_k^i)\|$.
4. Compute μ_k in (8) s.t. the discrepancy principle is satisfied.
5. Set $\mu_k^i = \mu_k \omega_k^i$.

In the next section we discuss perturbation theory and show how it relates to the weights ω_k^i in (9).

4. Perturbation analysis. The goal of regularization is to make reconstruction robust with respect to noise. By extension, a high sensitivity to the regularization parameters is undesirable. Consider a set of perturbed parameters $\boldsymbol{\mu}_\star + \Delta\boldsymbol{\mu}$; if $\|\Delta\boldsymbol{\mu}\|$ is sufficiently small

$$\begin{aligned} \boldsymbol{c}(\boldsymbol{\mu}_\star + \Delta\boldsymbol{\mu}) &= \boldsymbol{c}(\boldsymbol{\mu}_\star) + D\boldsymbol{c}(\boldsymbol{\mu}_\star)\Delta\boldsymbol{\mu} + \mathcal{O}(\|\Delta\boldsymbol{\mu}\|^2) \\ &= \boldsymbol{c}(\boldsymbol{\mu}_\star) - M^{-1}\Delta M \boldsymbol{c}(\boldsymbol{\mu}_\star) + \mathcal{O}(\|\Delta\boldsymbol{\mu}\|^2), \end{aligned}$$

where M and ΔM are defined as

$$M = \underline{H}_k^* \underline{H}_k + \sum_{i=1}^{\ell} \mu_\star^i K_k^{i\,*} K_k^i, \quad \Delta M = \sum_{i=1}^{\ell} \Delta\mu_k^i K_k^{i\,*} K_k^i. \tag{10}$$

Therefore, one might choose $\boldsymbol{\mu}_\star$ to minimize the sensitivity measure $\|M^{-1}\Delta M \boldsymbol{c}(\boldsymbol{\mu}_\star)\|$. We can solidify this statement by connecting it to the forward error in Proposition 1 and the backward error in Proposition 2.

Proposition 1. *Given regularization parameters $\mu_\star^i$ and perturbations $\mu_k^i = \mu_\star^i + \Delta\mu_k^i$, let $\boldsymbol{c}_\star = \boldsymbol{c}_k(\boldsymbol{\mu}_\star)$, $\boldsymbol{c}_k = \boldsymbol{c}_k(\boldsymbol{\mu}_k)$, $\boldsymbol{x}_\star = X_k \boldsymbol{c}_\star$, and $\boldsymbol{x}_k = X_k \boldsymbol{c}_k$. Assume $\underline{H}_k$ and all K_k^i are of full rank and define matrices M and ΔM as in* (10)*. If M and $M + \Delta M$ are nonsingular and the $\Delta\mu_k^i$ are sufficiently small so that $\|M^{-1}\Delta M\| < 1$, then*

$$\frac{\|\boldsymbol{x}_\star - \boldsymbol{x}_k\|}{\|\boldsymbol{x}_\star\|} \leq \frac{\|M^{-1}\Delta M\|}{1 - \|M^{-1}\Delta M\|}.$$

Proof. Observe that $\boldsymbol{c}_\star = M^{-1}\underline{H}_k^* \beta \boldsymbol{e}_1$ and $\boldsymbol{c}_k = (M + \Delta M)^{-1}\underline{H}_k^* \beta \boldsymbol{e}_1$. With a little manipulation we obtain

$$\boldsymbol{c}_k = (M + \Delta M)^{-1} M \boldsymbol{c}_\star = (I + M^{-1}\Delta M)^{-1} \boldsymbol{c}_\star = \sum_{j=0}^{\infty} (-M^{-1}\Delta M)^j \boldsymbol{c}_\star.$$

Now, using basic inequalities and the geometric series, we find

$$\frac{\|\boldsymbol{c}_k - \boldsymbol{c}_\star\|}{\|\boldsymbol{c}_\star\|} \leq \frac{\|M^{-1}\Delta M\|}{1 - \|M^{-1}\Delta M\|}.$$

Since X_k has orthonormal columns, the result of the proposition follows. □

Proposition 1 relates to the weights from (9) in the following way. Let $\boldsymbol{\mu}_\star = [\mathbf{0};\ \widetilde{\mu}_k^i;\ \mathbf{0}]$, and suppose we have a perturbation $\Delta\boldsymbol{\mu}_k$, such that all elements of $\boldsymbol{\mu}_\star + \Delta\boldsymbol{\mu}_k$ are nonnegative. Then, if $\|M^{-1}\Delta M\|$ is sufficiently small, the weight ω_k^i approximately satisfies the bound

$$\omega_k^i = \frac{\|\boldsymbol{c}_k(\boldsymbol{\mu}_\star)\|}{\|D_\mu \boldsymbol{c}_k(\boldsymbol{\mu}_\star\|} \geq \frac{\|\boldsymbol{c}_k(\boldsymbol{\mu}_\star)\|}{\|M^{-1}\Delta M \boldsymbol{c}_k(\boldsymbol{\mu}_\star)\|} = \frac{1}{\|M^{-1}\Delta M\|} \gtrsim \frac{\|\boldsymbol{c}_k(\boldsymbol{\mu}_\star)\|}{\|\boldsymbol{c}_k(\boldsymbol{\mu}_\star) - \boldsymbol{c}_k(\boldsymbol{\mu}_\star + \Delta\boldsymbol{\mu})\|}$$

Hence, the less sensitive $\boldsymbol{c}_k(\boldsymbol{\mu}_\star)$ is to perturbations in the parameters, the higher the weight ω_k^i.

Instead of studying the sensitivity of the approximate solution, one may wonder if it is possible to pick a vector $\boldsymbol{f}$ close to $\beta\boldsymbol{e}_1$ such that $\boldsymbol{c}_\star = (M+\Delta M)^{-1}\underline{H}_k^*\boldsymbol{f}$. Or in other words, given perturbed regularization parameters, is there a perturbation of $\beta\boldsymbol{e}_1$ such that the optimal approximation to the exact solution is obtained? The following proposition provides a positive answer.

Proposition 2. *Under the assumptions of Proposition 1, there exist vectors $\boldsymbol{f}$ and $\boldsymbol{g}$ such that $\boldsymbol{c}_\star = (M+\Delta M)^{-1}\underline{H}_k^*\boldsymbol{f}$ and $\boldsymbol{c}_k = M^{-1}\underline{H}_k^*\boldsymbol{g}$. Furthermore, $\boldsymbol{f}$ and $\boldsymbol{g}$ satisfy*

$$\frac{\|\beta\boldsymbol{e}_1 - \boldsymbol{g}\|}{\|\beta\boldsymbol{e}_1\|} \leq \kappa(\underline{H}_k)\|M^{-1}\Delta M\| \quad \textit{and} \quad \frac{\|\beta\boldsymbol{e}_1 - \boldsymbol{f}\|}{\|\beta\boldsymbol{e}_1\|} \leq \kappa(\underline{H}_k)\frac{\|M^{-1}\Delta M\|}{1-\|M^{-1}\Delta M\|},$$

where $\kappa(\underline{H}_k)$ is the condition number of $\underline{H}_k$.

Proof. The vector $\boldsymbol{f}$ is easy to derive using the *ansatz*

$$(M+\Delta M)^{-1}\underline{H}_k^*\boldsymbol{f} = M^{-1}\underline{H}_k^*\beta\boldsymbol{e}_1.$$

Let $\underline{H}_k = QR$ denote the QR-decomposition of $\underline{H}_k$, then

$$R^*Q^*\boldsymbol{f} = (M+\Delta M)M^{-1}R^*Q^*\beta\boldsymbol{e}_1,$$

and

$$\boldsymbol{f} = QR^{-*}(M+\Delta M)M^{-1}R^*Q^*\beta\boldsymbol{e}_1 + (I-QQ^*)\boldsymbol{v}$$

for arbitrary $\boldsymbol{v}$. Indeed, it is easy to verify that the above vector satisfies

$$\boldsymbol{c}_\star = (M+\Delta M)^{-1}\underline{H}_k^*\boldsymbol{f}.$$

If we choose $\boldsymbol{v} = \beta\boldsymbol{e}_1$, then

$$\boldsymbol{f} = QR^{-*}\Delta M M^{-1}R^*Q^*\beta\boldsymbol{e}_1 + \beta\boldsymbol{e}_1$$

so that

$$\frac{\|\beta\boldsymbol{e}_1 - \boldsymbol{f}\|}{\|\beta\boldsymbol{e}_1\|} = \|QR^{-*}\Delta M M^{-1}R^*Q^*\boldsymbol{e}_1\| \leq \|R^{-*}\|\,\|R^*\|\,\|\Delta M M^{-1}\|.$$

Here $\|R^{-*}\|\,\|R^*\|$ is the condition number $\kappa(\underline{H}_k)$ and $\|\Delta M M^{-1}\| = \|M^{-1}\Delta M\|$, since both M and ΔM are symmetric. This proves the first part of the proposition.

The second part is analogous. In particular, we use the *ansatz*

$$M^{-1}\underline{H}_k^*\boldsymbol{g} = (M+\Delta M)^{-1}\underline{H}_k^*\beta\boldsymbol{e}_1$$

and derive

$$\boldsymbol{g} = R^{-*}QM(M+\Delta M)^{-1}R^*Q^*\beta\boldsymbol{e}_1 + (I-QQ^*)\beta\boldsymbol{e}_1.$$

Again it is easy to verify that $\boldsymbol{c}_k = M^{-1}\underline{H}_k^*\boldsymbol{g}$. Observe that $\boldsymbol{g}$ can be rewritten as

$$\boldsymbol{g} = R^{-*}Q((I+\Delta M M^{-1})^{-1} - I)R^*Q^*\beta\boldsymbol{e}_1 + \beta\boldsymbol{e}_1$$

such that

$$\frac{\|\beta\boldsymbol{e}_1 - \boldsymbol{f}\|}{\|\beta\boldsymbol{e}_1\|} = \|R^{-*}((I+\Delta M M^{-1})^{-1} - I)R^*Q^*\boldsymbol{e}_1\| \leq \|R^{-*}\|\,\|R^*\|\,\|(I+\Delta M M^{-1})^{-1} - I\|.$$

Since $\|\Delta M M^{-1}\| = \|M^{-1}\Delta M\| < 1$, it follows that

$$\|(I+\Delta M M^{-1})^{-1} - I\| \leq \sum_{j=1}^{\infty}\|-\Delta M M^{-1}\|^j = \frac{\|M^{-1}\Delta M\|}{1-\|M^{-1}\Delta M\|},$$

which concludes the proof. □

We have discussed forward and backward error bounds which help motivate our parameter choice. Now that we have investigated each of the three phases of our method, we are ready to show numerical results.

5. Numerical experiments. We benchmark our algorithm with problems from Regularization Tools by Hansen [7]. Each problem provides an ill-conditioned $n \times n$ matrix A, a solution vector $\boldsymbol{x}_\star$ of length n and a corresponding measured vector $\boldsymbol{b}$. We let $n = 1024$ and add a noise vector $\boldsymbol{e}$ to $\boldsymbol{b}$. The entries of $\boldsymbol{e}$ are drawn independently from the standard normal distribution. The noise vector is then scaled such that $\varepsilon = \|\boldsymbol{e}\|$ equals $0.01\|\boldsymbol{b}\|$. We use $\eta = 1.01$ for the discrepancy bound in (6). We test the algorithms with one thousand different noise vectors for every triplet A, $\boldsymbol{x}_\star$, and $\boldsymbol{b}$ and report the median results.

The algorithms terminate when the relative difference between two subsequent approximations is less then 0.01, when $\boldsymbol{x}_{k+1}$ is (numerically) linear dependent in X_k, when both U_{k+1} and none of the V_k^i can be expanded, or when a maximum number of iterations is reached. For Algorithm 2 we use a maximum of 20 iterations and for Algorithm 1 a maximum of $(\ell + 1) \times 20$ iterations. For the sake of a fair comparison, the algorithms return the best obtained approximations and their iteration numbers.

We used the following regularization operators to obtain the single-parameter Tikhonov regularization results. The first derivative operator L_1 with stencil $[1, -1]$ for Gravity-3, Heat-5, Heat, and Phillips. The second derivative operator L_2 with stencil $[1, -2, 1]$ for Deriv2-1, Deriv2-2, Foxgood, Gravity-1, and Gravity-2. The third derivative operator L_3 with stencil $[-1, 3 - 3, 1]$ for Baart. The fifth derivative operator L_5 with stencil $[-1, 5, -10, 10, -5, 1]$ and Deriv2-3. The derivative operators L_d are of size $(n - d) \times n$. In addition to a derivative operator L_d, we applied multi-parameter Tikhonov regularization with the identity operator I and the orthogonal projection $(I - N_d N_d^*)$. Here, the columns of N_d are an orthonormal basis for the nullspace $\mathcal{N}(L_d)$.

The results are listed in Table 1. For each test problem, the table lists the relative error obtained with Algorithm 1, abbreviated by E_{sd}, and Algorithm 2, abbreviated by E_{md}. Also listed are the ratio ρ_E of E_{md} to E_{sd} and the ratio ρ_{mv} of the number of matrix-vector products. That is,

$$\rho_E = \frac{E_{\text{md}}}{E_{\text{sd}}} \quad \text{and} \quad \rho_{\text{mv}} = \frac{\#\text{ MVs Algorithm 2}}{\#\text{ MVs Algorithm 1}}.$$

Only matrix-vector multiplications with A, A^*, L^i, and L^{i*} count towards the total number of MVs used by each algorithm. Finally, we use the multidirectional subspace expansion in conjunction with truncation throughout this section, since the results did not appear to suffer in quality from truncation.

Table 1: Benchmark results for problems from Regularization Tools.

	Single-parameter				**Multi-parameter**			
Problem	E_{sd}	E_{md}	ρ_E	ρ_{mv}	E_{sd}	E_{md}	ρ_E	ρ_{mv}
Baart	$1.73 \cdot 10^{-1}$	$1.11 \cdot 10^{-1}$	0.64	1.93	$1.72 \cdot 10^{-1}$	$5.39 \cdot 10^{-2}$	0.31	2.60
Deriv2-1	$2.44 \cdot 10^{-1}$	$2.44 \cdot 10^{-1}$	1.00	1.00	$2.27 \cdot 10^{-1}$	$5.82 \cdot 10^{-3}$	0.03	1.81
Deriv2-2	$2.35 \cdot 10^{-1}$	$2.35 \cdot 10^{-1}$	1.00	0.83	$2.29 \cdot 10^{-1}$	$2.03 \cdot 10^{-2}$	0.09	1.55
Deriv2-3	$4.35 \cdot 10^{-2}$	$4.35 \cdot 10^{-2}$	1.00	0.92	$4.35 \cdot 10^{-2}$	$4.32 \cdot 10^{-2}$	0.99	1.00
Foxgood	$3.31 \cdot 10^{-2}$	$3.30 \cdot 10^{-2}$	1.00	0.67	$3.29 \cdot 10^{-2}$	$1.10 \cdot 10^{-2}$	0.34	1.35
Gravity-1	$3.85 \cdot 10^{-2}$	$3.41 \cdot 10^{-2}$	0.88	1.08	$3.69 \cdot 10^{-2}$	$1.83 \cdot 10^{-2}$	0.50	1.18
Gravity-2	$5.53 \cdot 10^{-2}$	$5.26 \cdot 10^{-2}$	0.95	1.10	$5.52 \cdot 10^{-2}$	$3.97 \cdot 10^{-2}$	0.72	2.04
Gravity-3	$1.03 \cdot 10^{-1}$	$9.21 \cdot 10^{-2}$	0.90	1.08	$1.02 \cdot 10^{-1}$	$9.24 \cdot 10^{-2}$	0.91	1.89
Heat	$9.26 \cdot 10^{-2}$	$9.12 \cdot 10^{-2}$	0.99	1.05	$8.79 \cdot 10^{-2}$	$8.77 \cdot 10^{-2}$	1.00	1.19
Phillips	$2.50 \cdot 10^{-2}$	$2.50 \cdot 10^{-2}$	1.00	1.00	$2.49 \cdot 10^{-2}$	$2.47 \cdot 10^{-2}$	0.99	1.21

The single-parameter results in Table 1 show that multidirectional subspace expansion can obtain small improvements in the relative error at the cost of a small number of extra matrix-vector products. We stress that in these cases, Algorithm 1 was allowed to perform additional MVs, but converged with a higher relative error. If there is no improvement in the relative error, we see that multidirectional subspace expansion can improve convergence, for two of the Deriv2 problems as well as for Foxgood.

For the multi-parameter results in Table 1, we observe larger improvements in the relative error for multidirectional subspace expansion, but at the cost of a larger number MVs. We no longer see cases where multidirectional subspace expansion terminates with fewer MVs. In fact, the relative error is the same for Heat, even though more MVs are required. Finally, Figure 2 illustrates an example of the improved results which can be obtained by using multidirectional subspace expansion.

In the next tests we attempt to reconstruct an original image from a blurred and noisy observation. Consider an $n \times n$ grayscale image with pixel values in the interval $[0, 1]$. Then $\boldsymbol{x}$ is a vector of length n^2 obtained by

stacking the columns of the image below each other. The matrix A represents a Gaussian blurring operator, generated with `blur` from Regularization Tools. The matrix A is block-Toeplitz with half-bandwidth `band=11` and the amount of blurring is given by the variance `sigma=5`. The entries of the noise vector $\boldsymbol{e}$ are independently drawn from the standard normal distribution after which the vector is scaled such that $\epsilon = \mathbb{E}[\|\boldsymbol{e}\|] = 0.05\|\boldsymbol{b}\|$. We take η such that $\|\boldsymbol{e}\| \leq \eta\epsilon$ in 99.9% of the cases. That is,

$$\eta = 1 + \frac{3.090232}{\sqrt{2n^2}}.$$

For regularization we choose an approximation to the Perona–Malik [16] operator

$$\mathcal{L}(\boldsymbol{x}) = \operatorname{div}(g(|\nabla \boldsymbol{x}|^2)\nabla \boldsymbol{x}),$$

where $g(s) = \exp(-s/\rho)$ and ρ is a small positive constant. Because $\mathcal{L}$ is a nonlinear operator, we first perform a small number of iterations with a finite difference approximation $L_{\boldsymbol{b}}$ of $\mathcal{L}(\boldsymbol{b})$. The resulting intermediate solution $\tilde{\boldsymbol{x}}$ is used for a new approximation $L_{\tilde{\boldsymbol{x}}}$ of $\mathcal{L}(\tilde{\boldsymbol{x}})$. Finally, we run the algorithms a second time with $L_{\tilde{\boldsymbol{x}}}$ and more iterations; see Reichel, Sgallari, and Ye [17] for more information regarding the implementation of the Perona–Malik operator.

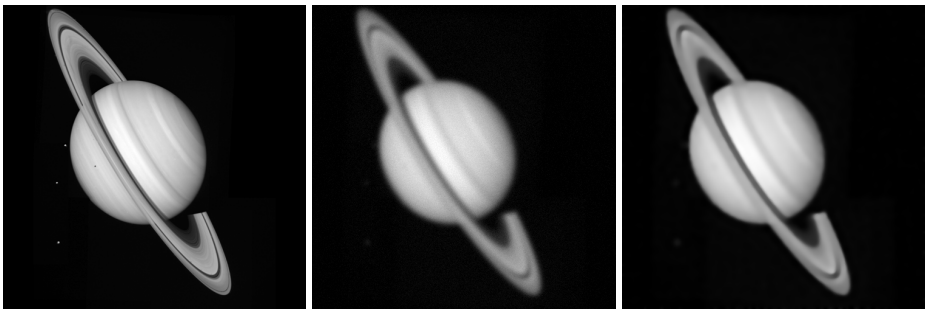

Figure 1: Deblurring results for `Saturn`. The original (left), observed (middle), and reconstructed images (right).

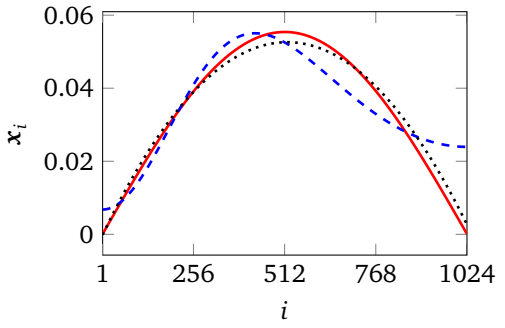

Figure 2: `baart` test problem.

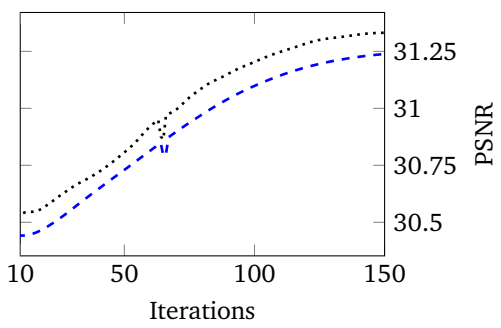

Figure 3: Convergence history for `Saturn`.

The solid line is the exact solution. The dashed line represents the results obtained with multi-parameter regularization and the residual subspace expansion (Algorithm 1). The dotted line represents the results obtained with multi-parameter regularization and multidirectional subspace expansion (Algorithm 2).

For our test we use an image of Saturn, see Figure 1, with $\rho = 0.03$, 25 iterations for the first run, and 150 iterations for the second run. In both cases we stop the iterations around the point where convergence flattens out, which can be seen for the second run from the convergence history in Figure 3. The figure uses the peak signal-to-noise ratio (PSNR) given by $-20\log_{10}(\|\boldsymbol{x}_\star - \boldsymbol{x}_k\|/n)$ versus the iteration number k. A higher PSNR implies a lower error.

We observe that multidirectional subspace expansion may allow convergence to a more accurate solution. Because multidirectional subspace expansion requires extra matrix-vector products, we investigate the performance in Table 2 and when Algorithm 2 achieves parity with Algorithm 1. There is only a small difference in the total number of matrix-vector products when parity is achieved, but a large improvement in wall clock time. This improvement is in large part due to the block operations which can only be used Algorithm 2. For reference, the runtimes were obtained on an Intel Core i7-3770 and with MATLAB R2015b on 64-bit Linux 4.2.5.

Table 2: The number of matrix-vector products and wall clock time used by the different methods. The results in the upper rows are for Lizards and the results in the lower rows are for Saturn.

Method	Total	A	A^*	L	L^*	Time (s)
Alg 1	599	150	150	150	149	82.3
Alg 2	889	295	150	295	149	98.4
Parity	637	211	108	211	107	62.3

6. Conclusions. We have presented a new method for large-scale Tikhonov regularization problems. The method combines a new multidirectional subspace expansion with an optional truncation to produce a higher quality search space. The multidirectional expansion generates a richer search space, whereas the truncation ensures moderate growth. Truncating the search space did not affect the quality of the results in our tests. Furthermore, numerical results illustrate that our method can yield more accurate results or faster convergence. Additionally, we have introduced a straightforward parameter selection for multi-parameter Tikhonov regularization. The parameters are selected with scale invariant relative weights consisting of the norm of single-parameter solutions and the norm of efficiently computable derivatives. We connect the weights to error bounds with perturbation theory.

REFERENCES

[1] M. Belge, M. E. Kilmer, and E. L. Miller. *Efficient determination of multiple regularization parameters in a generalized L-curve framework*. Inverse Problems 18.4 (2002), pp. 1161–1183.

[2] C. Brezinski, M. Redivo-Zaglia, G. Rodriguez, and S. Seatzu. *Multi-parameter regularization techniques for ill-conditioned linear systems*. Numer. Math. 94.2 (2003), pp. 203–228.

[3] D. Calvetti and L. Reichel. *Tikhonov regularization of large linear problems*. BIT 43.2 (2003), pp. 263–283.

[4] D. Fong and M. A. Saunders. *LSMR: an iterative algorithm for sparse least-squares problems*. SIAM J. Sci. Comput. 33.5 (2011), pp. 2950–2971.

[5] S. Gazzola and P. Novati. *Multi-parameter Arnoldi–Tikhonov methods*. Electron. Trans. Numer. Anal. 40 (2013), pp. 452–475.

[6] G. H. Golub and C. F. Van Loan. *Matrix Computations*. 3rd ed. Johns Hopkins University Press, 1996.

[7] P. C. Hansen. *Regularization Tools: A Matlab package for analysis and solution of discrete ill-posed problems*. Numer. Algorithms 6 (1994), pp. 1–35.

[8] P. C. Hansen. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. Society for Industrial and Applied Mathematics, Philadelpia, PA, 1998.

[9] M. E. Hochstenbach and L. Reichel. *An iterative method for Tikhonov regularization with a general linear regularization operator*. J. Integral Equations Appl. 22.3 (2010), pp. 465–482.

[10] M. E. Hochstenbach, L. Reichel, and X. Yu. *A Golub–Kahan-type reduction method for matrix pairs*. J. Sci. Comput. (2015), pp. 1–23.

[11] M. E. Kilmer, P. Hansen, and M. Español. *A projection-based approach to general-form Tikhonov regularization*. SIAM J. Sci. Comput. 29.1 (2007), pp. 315–330.

[12] J. Lampe, L. Reichel, and H. Voss. *Large-scale Tikhonov regularization via reduction by orthogonal projection*. Linear Algebra Appl. 436.8 (2012), pp. 2845–2865.

[13] R.-C. Li and Q. Ye. *A Krylov subspace method for quadratic matrix polynomials with applications to constrained least squares problems*. SIAM J. Matrix Anal. Appl. 25.2 (2003), pp. 405–528.

[14] S. Lu, S. V. Pereverzev, Y. Shao, and U. Tautenhahn. *Discrepancy curves for multi-parameter regularization*. J. Inverse Ill-Posed Probl. 18.6 (2010), pp. 655–676.

[15] C. C. Paige and M. A. Saunders. *LSQR: An algorithm for sparse linear equations and sparse least squares*. ACM Trans. Math. Softw. 8.1 (1982), pp. 43–71.

[16] P. Perona and J. Malik. *Scale-space and edge detection using anisotropic diffusion*. IEEE Trans. Pattern Anal. Mach. Intell. 12.7 (1990), pp. 629–639.

[17] L. Reichel, F. Sgallari, and Q. Ye. *Tikhonov regularization based on generalized Krylov subspace methods*. Appl. Numer. Math. 62.9 (2012), pp. 1215–1228.