SIAM J. SCI. COMPUT.
Vol. 39, No. 2, pp. B221–B243

HYBRID AND ITERATIVELY REWEIGHTED REGULARIZATION BY UNBIASED PREDICTIVE RISK AND WEIGHTED GCV FOR PROJECTED SYSTEMS*

ROSEMARY A. RENAUT†, SAEED VATANKHAH‡, AND VAHID E. ARDESTANI‡

Abstract. Tikhonov regularization for projected solutions of large-scale ill-posed problems is considered. The Golub–Kahan iterative bidiagonalization is used to project the problem onto a subspace and regularization then applied to find a subspace approximation to the full problem. Determination of the regularization, parameter for the projected problem by unbiased predictive risk estimation, generalized cross validation, and discrepancy principle techniques is investigated. It is shown that the regularized parameter obtained by the unbiased predictive risk estimator can provide a good estimate which can be used for a full problem that is moderately to severely ill-posed. A similar analysis provides the weight parameter for the weighted generalized cross validation such that the approach is also useful in these cases, and also explains why the generalized cross validation without weighting is not always useful. All results are independent of whether systems are over- or underdetermined. Numerical simulations for standard one-dimensional test problems and two-dimensional data, for both image restoration and tomographic image reconstruction, support the analysis and validate the techniques. The size of the projected problem is found using an extension of a noise revealing function for the projected problem [I. Hnětynková, M. Plešinger, and Z. Strakoš, *BIT Numer. Math.*, 49 (2009), pp. 669–696]. Furthermore, an iteratively reweighted regularization approach for edge preserving regularization is extended for projected systems, providing stabilization of the solutions of the projected systems and reducing dependence on the determination of the size of the projected subspace.

Key words. large-scale inverse problems, Golub–Kahan bidiagonalization, regularization parameter estimation, unbiased predictive risk estimator, discrepancy principle, generalized cross validation, iteratively reweighted schemes

AMS subject classifications. 65F10, 65F22, 65R32

DOI. 10.1137/15M1037925

1. Introduction.

The solution of the numerically ill-posed linear system of equations

$$\mathbf{b} = A\mathbf{x}_{\text{ex}} + \boldsymbol{\eta}, \quad \mathbf{b} \in \mathcal{R}^m, \quad \mathbf{x} \in \mathcal{R}^n, \quad A \in \mathcal{R}^{m\times n} \tag{1.1}$$

for matrix A of large dimension with $m \geq n$ or $m < n$ is considered. Matrix A is ill-conditioned; the singular values of A decay exponentially to zero, or to the limits of the numerical precision. Noise in the data is represented by $\boldsymbol{\eta} \in \mathcal{R}^m$, i.e., $\mathbf{b} = \mathbf{b}_{\text{ex}} + \boldsymbol{\eta}$ for exact but unknown data $\mathbf{b}_{\text{ex}}$ that satisfies $\mathbf{b}_{\text{ex}} = A\mathbf{x}_{\text{ex}}$ for unknown exact model parameters $\mathbf{x}_{\text{ex}}$. Components η_i of $\boldsymbol{\eta}$ are assumed to be independently sampled from a Gaussian distribution with mean 0 and covariance s_i^2. Given A and $\mathbf{b}$, an estimate for $\mathbf{x}$ that predicts $\mathbf{b}_{\text{ex}}$ is desired.

Discrete ill-posed problems of the form (1.1) may be obtained by discretizing linear ill-posed problems, such as Fredholm integral equations of the first kind, and

*Submitted to the journal's Computational Methods in Science and Engineering section September 2, 2015; accepted for publication (in revised form) November 29, 2016; published electronically March 9, 2017.

http://www.siam.org/journals/sisc/39-2/M103792.html

Funding: This work was supported by NSF grants 1216559 and 1418377, and AFOSR grant 025717.

†School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287-1804 (renaut@asu.edu).

‡Institute of Geophysics, University of Tehran, Tehran, Iran (svatan@ut.ac.ir, ebrahimz@ut.ac.ir).

arise in many research areas including image deblurring, geophysics, etc. Due to the presence of the noise in the data and the ill-conditioning of A, regularization is needed in order to obtain an estimate for $\mathbf{x}$ approximating $\mathbf{x}_{\text{ex}}$. Standard Tikhonov regularization provides

$$\mathbf{x}(\alpha) = \arg\min_{\mathbf{x}} \{\|W_{\boldsymbol{\eta}}(A\mathbf{x} - \mathbf{b})\|_2^2 + \alpha^2 \|D(\mathbf{x} - \mathbf{x}_{\text{apr}})\|_2^2\} \tag{1.2}$$

for weighted data fidelity term $\|W_{\boldsymbol{\eta}}(A\mathbf{x}-\mathbf{b})\|_2^2$ and regularization term $\|D(\mathbf{x}-\mathbf{x}_{\text{apr}})\|_2^2$. D is a regularization matrix, assumed here to be invertible, and $\mathbf{x}_{\text{apr}}$ allows specification of a given reference vector of prior information for $\mathbf{x}$. The unknown regularization parameter α trades off between the data fidelity and the regularization terms. The noise in the measurements $\mathbf{b}$ is whitened when $W_{\boldsymbol{\eta}} = C_{\boldsymbol{\eta}}^{-1/2}$ for the covariance matrix $C_{\boldsymbol{\eta}} = \text{diag}(s_1^2, \ldots, s_m^2)$. Introducing $\tilde{\mathbf{b}} = W_{\boldsymbol{\eta}}\mathbf{b}$, $\tilde{A} = W_{\boldsymbol{\eta}}A$, shifting by the prior information through $\mathbf{y} = \mathbf{x} - \mathbf{x}_{\text{apr}}$, and assuming that the null spaces of $\tilde{A}$ and D do not intersect yields

$$\begin{aligned} \mathbf{y}(\alpha) &= \arg\min_{\mathbf{y}} \{\|\tilde{A}\mathbf{y} - \tilde{\mathbf{r}}\|_2^2 + \alpha^2\|D\mathbf{y}\|_2^2\}, \quad \tilde{\mathbf{r}} = (\tilde{\mathbf{b}} - \tilde{A}\mathbf{x}_{\text{apr}}) \\ &= (\tilde{A}^T\tilde{A} + \alpha^2 D^T D)^{-1}\tilde{A}^T\tilde{\mathbf{r}}. \end{aligned} \tag{1.3}$$

Analytically, when D is invertible, which is not always the case, we may write

$$(\tilde{A}^T\tilde{A} + \alpha^2 D^T D) = D^T((D^T)^{-1}\tilde{A}^T\tilde{A}D^{-1} + \alpha^2 I_n)D.$$

Thus, when it is feasible to calculate D^{-1}, or to solve systems of equations defined by invertible D, it is convenient to introduce the right preconditioned matrix $\tilde{\tilde{A}} = \tilde{A}D^{-1}$ and regularized inverse $\tilde{\tilde{A}}^\dagger(\alpha) = (\tilde{\tilde{A}}^T\tilde{\tilde{A}} + \alpha^2 I_n)^{-1}\tilde{\tilde{A}}^T$,[1] which provides

$$\mathbf{z}(\alpha) := \arg\min_{\mathbf{z}} \{\|\tilde{\tilde{A}}\mathbf{z} - \tilde{\mathbf{r}}\|_2^2 + \alpha^2\|\mathbf{z}\|_2^2\}, \quad \mathbf{z}(\alpha) = D\mathbf{y}(\alpha), \quad \text{and} \tag{1.4}$$

$$\mathbf{x}(\alpha) = \mathbf{x}_{\text{apr}} + \mathbf{y}(\alpha) = \mathbf{x}_{\text{apr}} + D^{-1}\tilde{\tilde{A}}^\dagger(\alpha)\tilde{\mathbf{r}}. \tag{1.5}$$

Although equivalent analytically, numerical techniques to solve (1.3) and (1.4) differ. For small-scale problems, for example, we may solve (1.3) using the generalized singular value decomposition (GSVD), e.g., [20], for the matrix pair $[\tilde{A}, D]$, but would use the singular value decomposition (SVD) of $\tilde{\tilde{A}}$ for (1.4), e.g., [6], as given in Appendix A, dependent on the feasibility of calculating D^{-1}. Still, the use of the SVD or GSVD is not viable computationally for large-scale problems unless the underlying operators possess a specific structure. For example, if the underlying system matrix and associated regularization matrix are expressible via Kronecker decompositions, e.g., [13], then the GSVD decomposition can be found via the GSVD for each dimension separately. Here we consider the general situation and the use of iterative Krylov methods to estimate $\mathbf{x}(\alpha)$.

1.1. Numerical solution by the Golub–Kahan bidiagonalization. In principle, iterative methods such as conjugate gradients or other Krylov methods can be employed to solve (1.2). Results presented in [12] demonstrate, however, that MINRES and GMRES should not be used as regularizing Krylov iterations due to the early

[1] Note that we use in general the notation $A^\dagger(\alpha)$ for the pseudoinverse of the augmented matrix $[A; \alpha I]$.

transfer of noise to the Krylov basis. Here we use the well-known Golub–Kahan bidiagonalization (GKB), implemented in the LSQR algorithm, which has been well-studied in the context of projected solutions of the least squares problem [22, 21]. Recently, there has also been some interest in the LSMR modification of LSQR [4], but due to our goal to investigate the regularization parameter α we focus on LSQR for which the noise regularizing properties of the iteration are better understood [12, 14]. Effectively, the GKB projects the solution of the inverse problem to a smaller subspace, say of size t.

Applying t steps of the GKB on matrix A with initial vector $\mathbf{b}$ of norm $\beta_1 = \|\mathbf{b}\|_2$, and defining $\mathbf{e}_1^{(t+1)}$ to be the unit vector of length $t+1$ with a 1 in the first entry, a lower bidiagonal matrix $B_t \in \mathcal{R}^{(t+1)\times t}$, and column orthonormal matrices $H_{t+1} \in \mathcal{R}^{m\times(t+1)}$, $G_t \in \mathcal{R}^{n\times t}$ are generated such that (see [10, 16])

$$AG_t = H_{t+1}B_t, \quad \beta_1 H_{t+1}\mathbf{e}_1^{(t+1)} = \mathbf{b}. \tag{1.6}$$

For $\mathbf{x}_t = G_t\mathbf{w}_t$, full, $\mathbf{r}_{\text{full}}(\mathbf{x}_t)$, and projected, $\mathbf{r}_{\text{proj}}(\mathbf{w}_t)$, residuals are related via

$$\begin{aligned}\mathbf{r}_{\text{full}}(\mathbf{x}_t) &= A\mathbf{x}_t - \mathbf{b} = AG_t\mathbf{w}_t - \beta_1 H_{t+1}\mathbf{e}_1^{(t+1)} \\ &= H_{t+1}B_t\mathbf{w}_t - \beta_1 H_{t+1}\mathbf{e}_1^{(t+1)} \\ &= H_{t+1}(B_t\mathbf{w}_t - \beta_1\mathbf{e}_1^{(t+1)}) = H_{t+1}\mathbf{r}_{\text{proj}}(\mathbf{w}_t)\end{aligned} \tag{1.7}$$

for which, by the column orthonormality of H_{t+1},

$$\|\mathbf{r}_{\text{full}}(\mathbf{x}_t)\|_2^2 = \|\mathbf{r}_{\text{proj}}(\mathbf{w}_t)\|_2^2. \tag{1.8}$$

Theoretically, therefore, an estimate for $\mathbf{x}$ with respect to a reduced subspace may be found by finding $\mathbf{w}_t$ and then projecting back to the full problem. Matrix B_t in most cases, however, inherits the ill-conditioning of the matrix A [22], and regularization of the projected problem is needed.

By the column orthonormality of G_t, we have $\|\mathbf{x}_t\|_2^2 = \|G_t\mathbf{w}_t\|_2^2 = \|\mathbf{w}_t\|_2^2$. Thus, explicitly introducing regularization parameter ζ, distinct from α in order to emphasize regularization on the projected problem, yields the projected Tikhonov problem

$$\mathbf{w}_t(\zeta) = \arg\min_{\mathbf{w}\in\mathcal{R}^t}\{\|B_t\mathbf{w} - \beta_1\mathbf{e}_1^{(t+1)}\|_2^2 + \zeta^2\|\mathbf{w}\|_2^2\} \tag{1.9}$$

with solution

$$\begin{aligned}\mathbf{w}_t(\zeta) &= \beta_1(B_t^TB_t + \zeta^2 I_t)^{-1}B_t^T\mathbf{e}_1^{(t+1)} = \beta_1 B_t^\dagger(\zeta)\mathbf{e}_1^{(t+1)} \\ &= (G_t^TA^TAG_t + \zeta^2 I_t)^{-1}G_t^TA^T\mathbf{b} = (AG_t)^\dagger(\zeta)\mathbf{b}.\end{aligned} \tag{1.10}$$

In practice, one uses (1.10) to find $\mathbf{w}_t(\zeta)$ via the SVD for B_t, under the assumption that $t << m^* = \min(m,n)$, noting that an explicit solution for $\mathbf{w}_t$ is immediately available; see, e.g., Appendix A.

As already observed in [9, p. 302], the regularized LSQR algorithm now poses the problem of both detecting the appropriate number of steps t and finding the optimal parameter ζ_{opt}. One method of regularization is simply to avoid the introduction of the regularizer in (1.9) and find an optimal t at which to stop the iteration. Although it is known that the LSQR iteration is a regularizing iteration, it also exhibits a semiconvergence behavior so that eventually regularization is also needed.

This regularization may be achieved by either picking α in advance, namely, "regularize and project," or by the hybrid approach of regularizing the projected problem, e.g., [2, 16, 26]. The problem of first determining the appropriate size t for the projected space is discussed in, e.g., [14, 16] and more recently in [23] for large-scale geophysical inversion. Although the solutions obtained from "regularize then project" and "project then regularize" for a given t and $\alpha = \zeta$ are equivalent (see [16, Theorem 3.1] and [9, p. 301]), this does not immediately mean that ζ_{opt} for the subspace problem provides α_{opt} for the full problem [16].

Remark 1.1. Determining to which degree certain regularization techniques provide a good estimate for α_{opt} from the subspace problem estimate, and the conditions under which this will hold, is the topic of this work and is the reason we denote the regularization parameter on the subspace by ζ distinct from α.

1.2. Regularization parameter estimation. For the full problem, the question of determining an optimal parameter α_{opt} is well studied; see, e.g., [11, 30] for a discussion of methods including the Morozov discrepancy principle (MDP), the L-curve (LC), generalized cross validation (GCV), and unbiased predictive risk estimation (UPRE). The use of the MDP, LC, and GCV is also widely discussed for the projected problem, starting with the work of Kilmer and O'Leary [16] and continuing in [2]. Further, extensions for windowed regularization and hence multi-parameter regularization [1] are also applied for the projected problem [3]. Our attention is focused initially on the use of the UPRE. Effectively, the UPRE provides the correct estimate for α_{opt} in the context of a filtered truncated SVD (FTSVD) solution of (1.2) with t terms, provided that the LSQR factorization effectively captures the dominant right singular subspace of size t for matrix A. This observation does not immediately extend to the GCV. Applying a similar analysis as for the UPRE, however, provides a choice of the weighting parameter in the weighted GCV (WGCV) introduced in [2].

We stress that the approach assumes throughout, both numerically and theoretically, that the projected system is calculated with full reorthogonalization, a point not made explicit in many discussions, although it is apparent than many references implicitly make this assumption.

1.3. Overview. The paper is organized as follows. The regularization parameter estimation techniques of interest are presented in section 2. The discussion in section 2 is validated with one-dimensional simulations in section 3. Image restoration problems presented in section 4 illustrate the relevance for the two-dimensional case. In section 4.4 we extend the hybrid approach for use with an iteratively reweighted regularizer (IRR), which sharpens edges within the solution [24, 27, 28, 29, 32, 33], hence demonstrating that edge preserving regularization can be applied in the context of regularized LSQR solutions of the least squares problem on a projected subspace. Finally, in section 4.5 we also illustrate the algorithms in the context of sparse tomographic reconstruction of a walnut data set [7], demonstrating the more general use of the approach beyond deblurring of noisy data. Our conclusions are presented in section 5. It is of particular interest that our analysis applies for both over- and underdetermined systems of equations and is thus potentially of future use for other algorithms in which alternative regularizers are imposed and which also require repeated Tikhonov solves at each step. Further, this work extends our analysis of the UPRE in the context of underdetermined but small-scale problems in [28, 29], and demonstrates that IRR can be applied for projected algorithms.

2. Regularization parameter estimation. In order to use any specific regularization parameter estimation method for the projected problem, it is necessary to understand the derivation on the full problem. We thus provide a brief overview of the derivations as needed.

2.1. Unbiased predictive risk estimator. The predictive error, $\mathbf{p}_{\text{full}}(\mathbf{x}(\alpha))$, for the solution $\mathbf{x}(\alpha)$ is defined by

$$\mathbf{p}_{\text{full}}(\mathbf{x}(\alpha)) = A\mathbf{x}(\alpha) - \mathbf{b}_{\text{ex}} = AA^{\dagger}(\alpha)\mathbf{b} - \mathbf{b}_{\text{ex}} = (A(\alpha) - I_m)\mathbf{b}_{\text{ex}} + A(\alpha)\boldsymbol{\eta}, \tag{2.1}$$

where $A(\alpha) = AA^{\dagger}(\alpha)$ is the influence matrix. The residual may also be written in terms of the $A(\alpha)$ as

$$\mathbf{r}_{\text{full}}(\mathbf{x}(\alpha)) = (A(\alpha) - I_m)\mathbf{b} = (A(\alpha) - I_m)\mathbf{b}_{\text{ex}} + (A(\alpha) - I_m)\boldsymbol{\eta}. \tag{2.2}$$

In both equations the first term is deterministic, whereas the second is stochastic due to noise vector $\boldsymbol{\eta}$. To proceed, we need the trace lemma, e.g., [30, Lemma 7.2].

LEMMA 2.1. *For deterministic vector* $\mathbf{f}$*, random vector* $\boldsymbol{\eta}$ *with diagonal covariance matrix* $C_{\boldsymbol{\eta}}$*, matrix* F*, and expectation operator* E*,*

$$E(\|\mathbf{f} + F\boldsymbol{\eta}\|_2^2) = \|\mathbf{f}\|_2^2 + \operatorname{tr}(C_{\boldsymbol{\eta}}F^TF),$$

using $\operatorname{tr}(A)$ *to denote the trace of matrix* A*.*

Applying Lemma 2.1 to both (2.1) and (2.2) with the assumption that $C_{\boldsymbol{\eta}} = I_m$, due to whitening of noise $\boldsymbol{\eta}$, and using the symmetry of the influence matrix, we obtain

$$E(\|\mathbf{p}_{\text{full}}(\mathbf{x}(\alpha))\|_2^2) = \|(A(\alpha) - I_m)\mathbf{b}_{\text{ex}}\|_2^2 + \operatorname{tr}(A^T(\alpha)A(\alpha)), \tag{2.3}$$

$$E(\|\mathbf{r}_{\text{full}}(\mathbf{x}(\alpha))\|_2^2) = \|(A(\alpha) - I_m)\mathbf{b}_{\text{ex}}\|_2^2 + \operatorname{tr}((A(\alpha) - I_m)^T(A(\alpha) - I_m)). \tag{2.4}$$

Here $E(\|\mathbf{p}_{\text{full}}(\mathbf{x}(\alpha))\|_2^2)/m$ is the expected value of the risk of using the solution $\mathbf{x}(\alpha)$ to predict $\mathbf{b}_{\text{ex}}$. The first term on the right-hand side in each case cannot be obtained, but we may use $E(\|\mathbf{r}_{\text{full}}(\mathbf{x}(\alpha))\|_2^2) \approx \|\mathbf{r}_{\text{full}}(\mathbf{x}(\alpha))\|_2^2$ in (2.4). Thus, using linearity of the trace and eliminating the first term in the right-hand side of (2.3) gives the UPRE estimator to find α_{opt}:

$$\alpha_{\text{opt}} = \arg\min_{\alpha}\{U(\alpha) = \|(A(\alpha) - I_m)\mathbf{b}\|_2^2 + 2\operatorname{tr}(A(\alpha)) - m\}. \tag{2.5}$$

Typically, α_{opt} is found by evaluating (2.5) for a range of α, for example by the SVD (see, e.g., Appendix B), with the minimum found within that range of parameter values, as suggested in [10] for the GCV. See also, e.g., [29, equation (A.6)] for the formulae for calculating the function in terms of the SVD of matrix A.

2.1.1. Extending the UPRE for the projected problem. We observe that we may immediately write the predictive error and the residual in terms of the solution of the projected problem explicitly depending on the regularization parameter ζ. Specifically, defining the influence matrix $(AG_t)(\zeta) = AG_t(AG_t)^{\dagger}(\zeta)$ for the projected solution, we have

$$\mathbf{p}_{\text{full}}(\mathbf{x}_t(\zeta)) = AG_t\mathbf{w}_t(\zeta) - \mathbf{b}_{\text{ex}} = (AG_t)(\zeta)\mathbf{b} - \mathbf{b}_{\text{ex}}, \tag{2.6}$$

$$\mathbf{r}_{\text{full}}(\mathbf{x}_t(\zeta)) = AG_t\mathbf{w}_t(\zeta) - \mathbf{b} = ((AG_t)(\zeta) - I_m)\,\mathbf{b}. \tag{2.7}$$

By comparing (2.6) with (2.1) and (2.7) with (2.2), we obtain

$$U_{\text{full}}(\zeta) = \| \left((AG_t)(\zeta) - I_m\right) \mathbf{b}\|_2^2 + 2\,\text{tr}\left((AG_t)(\zeta)\right) - m.$$

Now by (1.8) it is immediate that the first term can be obtained without finding $\mathbf{x}_t(\zeta)$. For the second term we observe

$$\begin{aligned}(AG_t)(\zeta) &= AG_t((AG_t)^T AG_t + \zeta^2 I_t)^{-1}(AG_t)^T \\ &= H_{t+1}B_t((H_{t+1}B_t)^T(H_{t+1}B_t) + \zeta^2 I_t)^{-1}(H_{t+1}B_t)^T \\ &= H_{t+1}\left(B_t(B_t^T B_t + \zeta^2 I_t)^{-1}B_t^T\right)H_{t+1}^T = H_{t+1}B_t(\zeta)H_{t+1}^T,\end{aligned}$$

which yields

$$\text{tr}\left((AG_t)(\zeta)\right) = \text{tr}(H_{t+1}B_t(\zeta)H_{t+1}^T) = \text{tr}(B_t(\zeta)),$$

where the last equality follows from the cycle property of the trace operator for consistently sized matrices. Hence

$$U_{\text{full}}(\zeta) = \|\beta_1(B_t(\zeta) - I_{t+1})\mathbf{e}_1^{t+1}\|_2^2 + 2\,\text{tr}(B_t(\zeta)) - m \tag{2.8}$$

can be evaluated without reprojecting the solution for every ζ back to the full problem.

Remark 2.2. Although the UPRE function can be found from the projected solution alone, it is not clear whether (2.8) has any relevance with respect to the projected solution, i.e., does this appropriately regularize the projected solution? Otherwise it may not be appropriate to find ζ to minimize this function on the subspace.

The projected solution solves the problem with system matrix B_t and right-hand side vector $\beta_1\mathbf{e}_1^{t+1} = H_{t+1}^T\mathbf{b}$, which also consists of a deterministic and stochastic part, $H_{t+1}^T\mathbf{b}_{\text{ex}} + H_{t+1}^T\boldsymbol{\eta}$, where for white noise vector $\boldsymbol{\eta}$ and column orthogonal H_{t+1}, $H_{t+1}^T\boldsymbol{\eta}$ is a random vector of length $t+1$ with covariance matrix I_{t+1}. Thus the UPRE for the projected problem is

$$U_{\text{proj}}(\zeta) = \|\beta_1(B_t(\zeta) - I_{t+1})\mathbf{e}_1^{(t+1)}\|_2^2 + 2\,\text{tr}(B_t(\zeta)) - (t+1). \tag{2.9}$$

Comparing (2.8) with (2.9), it is immediate that minimizing (2.8) to minimize the risk for the projected solution also minimizes the risk for the full solution with respect to the given subspace.

It remains to determine whether there is any case in which finding ζ_{opt} also minimizes the predictive risk (2.5) for the full problem. Specifically, it is not immediate that $U_{\text{full}}(\alpha_{\text{opt}}) \approx U_{\text{full}}(\zeta_{\text{opt}})$ because α_{opt} is needed with respect to solutions in Range(V), not just restricted to Range$(G_t\tilde{V}_t)$. Here matrices V and $\tilde{V}_t$, the column orthogonal matrices arising in the SVDs of A and B_t, respectively, span the respective right singular subspaces. Although in exact arithmetic the large singular values of B_t provide a good approximation of the large singular values of A [6, section 9.3.3], the number of small singular values in the spectrum of B_t limits how well the full problem will be regularized by regularizing the projected problem. Adopting now the statement of *full regularization* of the LSQR as given in [15], namely that the LSQR iterate with t steps effectively captures the t-dimensional dominant right spectral space of A, suppose that t is such that the singular values of B_t approximate the t largest singular values of A with the natural order so that necessarily $\gamma_t > \sigma_{t^*+1}$ for $t \le t^*$. Equivalently, this requires that t^* is close to t and that the spectrum of B_t contains no singular value

approximating a very small spectral value of A. It is shown in [15, Theorem 2.3] that this requirement is more likely satisfied for severely and moderately ill-posed problems than for mildly ill-posed problems. Further, in such cases the LSQR solution on the space of size t approximates the truncated SVD (TSVD) solution of the full problem, namely, the solution of the full problem with filter factors $\phi_i(\alpha) = 0$ for $i > t$. Then,

$$\begin{aligned} \operatorname{tr}(A(\alpha)) &= \left(t - \alpha^2 \sum_{i=1}^{t} (\sigma_i^2 + \alpha^2)^{-1} \right) + \left((m^* - t) - \alpha^2 \sum_{i=t+1}^{m^*} (\sigma_i^2 + \alpha^2)^{-1} \right) \\ &= \left(t - \alpha^2 \sum_{i=1}^{t} (\sigma_i^2 + \alpha^2)^{-1} \right) \approx t - \alpha^2 \sum_{i=1}^{t} (\gamma_i^2 + \alpha^2)^{-1} = \operatorname{tr}(B_t(\alpha)). \end{aligned} \tag{2.10}$$

Thus, for the situation in which the LSQR iterate provides full regularization, determining ζ_{opt} to minimize (2.9) will yield α_{opt}, which is optimal for the filtered truncated SVD (FTSVD) solution of the full problem. This observation also follows Theorem 3.2 in [16], which connects the use of the TSVD of B_t for the solution with the solution obtained using the TSVD of A. We summarize below.

Remark 2.3 (UPRE). If t is such that $\phi_i(\alpha) \approx 0$ for $i > t$, and such that the LSQR iterate provides *full regularization* so that $\operatorname{tr}(A(\alpha)) \approx \operatorname{tr}(B_t(\alpha))$ and $\text{Range}(G_t \tilde{V}_t)$ approximates $\text{Range}(V_t)$, then $\zeta_{\text{opt}} \approx \alpha_{\text{opt}}$ when obtained using the UPRE. Further, the estimate is found without projecting the solution back to the full space, namely by minimizing (2.9).

When the LSQR does not provide *full regularization*, denoted as *partial regularization* in [15], the above result will not hold, and B_t captures the ill-conditioning of A through the inclusion of inaccurate small singular values in the spectrum of B_t.

2.2. Morozov discrepancy principle. Although it is well known that the MDP always leads to an overestimation of the regularization parameter, e.g., [16], it is still a widely used method for many applications and is thus an important baseline for comparison. The premise of the MDP [17] to find α is the assumption that the norm of the residual, $\|\mathbf{r}_{\text{full}}(\mathbf{x}(\alpha))\|_2^2$, follows a χ^2 distribution with δ degrees of freedom, $\|\mathbf{r}_{\text{full}}(\mathbf{x}(\alpha))\|_2^2 = \delta$. Heuristically, the rationale for this choice is seen by re-expressing (1.7),

$$\mathbf{r}_{\text{full}}(\mathbf{x}(\alpha)) = A\mathbf{x}(\alpha) - \mathbf{b} = A(\mathbf{x}(\alpha) - \mathbf{x}_{\text{ex}}) - \boldsymbol{\eta},$$

so that if $\mathbf{x}(\alpha)$ has been found as a good estimate for $\mathbf{x}_{\text{ex}}$, then the residual (1.7) should be dominated by the whitened error vector $\boldsymbol{\eta}$. For white noise, $\|\boldsymbol{\eta}\|^2$ is distributed as a χ^2 distribution with m degrees of freedom, for which $E(\|\boldsymbol{\eta}\|^2) = m$ and the variance is $2m$. Thus we seek a residual such that $\delta = \upsilon m$ using a Newton root-finding method (see Appendix B), where we take safety parameter $\upsilon > 1$ to handle the well-known oversmoothing of the MDP. Applying the same approach for the projected residual yields the noise term $H_{t+1}^T \boldsymbol{\eta}$ replacing $\boldsymbol{\eta}$. Thus the degrees of freedrom are reduced to $t+1$ and we seek a residual such that $\delta = \upsilon(t+1)$. A number of other suggestions for a projected discrepancy principle have been presented in the literature, e.g., [9, 16, 25], but these generally imply using $\delta \approx \upsilon \|\boldsymbol{\eta}\|_2^2 \approx \upsilon E(\|\boldsymbol{\eta}\|_2^2) \approx \upsilon m$ dependent on the noise level of the full problem with $\upsilon > 1$. It is reported in [9], however, that while the theory predicts choosing $\upsilon > 1$, numerical experiments support reducing υ. Alternatively, this may be seen as reducing the degrees of freedom, instead of reducing υ. We deduce that the interpretation for finding the regularization parameter based on the

statistical property of the projected residual in contrast to the full residual should be important in determining δ.

Remark 2.4 (MDP). For the MDP the degrees of freedom change from m to $t+1$ when the residual is calculated on the full space as compared to the projected space. Thus ζ_{opt} is not a good approximation for α_{opt} when obtained using δ_{proj} as a guide for the actual size of the projected residual. For the *full regularization* the degrees of freedom for the full problem are reduced and again $\zeta_{\text{opt}} \approx \alpha_{\text{opt}}$.

2.3. Generalized cross validation. Unlike the UPRE and MDP methods, the GCV method for finding the regularization parameter α does not require any information on the noise distribution for $\boldsymbol{\eta}$. The optimal parameter α is found as the minimizer of the function

$$G_{\text{full}}(\alpha) = \frac{\|\mathbf{r}_{\text{full}}(\mathbf{x}(\alpha))\|_2^2}{\left(\text{tr}(A(\alpha) - I_m)\right)^2}, \tag{2.11}$$

ignoring constant scaling of $G_{\text{full}}(\alpha)$ by n [5]. The obvious implementation of the GCV for the projected problem is the exact replacement in (2.11) using the projected system

$$G_{\text{proj}}(\zeta) = \frac{\|\mathbf{r}_{\text{proj}}(\mathbf{w}_t(\zeta))\|_2^2}{\left(\text{tr}(B_t(\zeta) - I_{t+1})\right)^2} \tag{2.12}$$

as indicated in [16]. It was recognized in [2, section 5.4], however, that this formulation tends to lead to solutions which are oversmoothed, and as an alternative the WGCV was introduced, dependent on parameter ω:

$$G_{\text{proj}}(\zeta, \omega) = \frac{\|\mathbf{r}_{\text{proj}}(\mathbf{w}_t(\zeta))\|_2^2}{\left(\text{tr}(\omega B_t(\zeta) - I_{t+1})\right)^2}.$$

Experiments illustrated that ω should be smaller for high noise cases, but in all cases $0 \le \omega \le 1$ is required to avoid the potential of a zero in the denominator. The choice for ω was argued heuristically and an adaptive algorithm to find ω was given.

Consider now the two denominators in (2.11) and (2.12). First of all, by (2.10) it is not difficult to show, for $0 < \omega < 1$, that

$$0 > \text{tr}(B_t(\alpha) - I_{t+1}) > \text{tr}(A(\alpha) - I_m),$$

so that $G_{\text{proj}}(\alpha) > G_{\text{full}}(\alpha)$ and α chosen to minimize the projected GCV will not minimize the full GCV term. For the weighted GCV, however,

$$\text{tr}(I_{t+1} - \omega B_t(\zeta)) = (1 + t - \omega t) + \omega\zeta^2 \sum_{i=1}^{t} \frac{1}{\gamma_i^2 + \zeta^2}, \tag{2.13}$$

and for the full regularization in which we approximate the full TSVD, $\phi_i(\alpha) \approx 0$ for $i > t$,

$$\text{tr}(I_m - A(\alpha)) = (m - m^*) + \alpha^2 \sum_{i=1}^{m^*} \frac{1}{\sigma_i^2 + \alpha^2} \approx (m - t) + \alpha^2 \sum_{i=1}^{t} \frac{1}{\sigma_i^2 + \alpha^2}. \tag{2.14}$$

Factoring for $m-t \neq 0$ and $(t+1-\omega t) \neq 0$ in (2.14) and (2.13), respectively, gives the scaled denominators

$$(m-t)\left(1+\frac{1}{m-t}\sum_{i=1}^{t}\frac{\alpha^2}{\sigma_i^2+\alpha^2}\right) \quad \text{and} \quad (1+t-\omega t)\left(1+\frac{\omega}{1+t-\omega t}\sum_{i=1}^{t}\frac{\zeta^2}{\gamma_i^2+\zeta^2}\right).$$

Ignoring constant scaling, the denominators are equilibrated by taking

$$\frac{1}{m-t}=\frac{\omega}{1+t-\omega t} \quad \text{yielding} \quad \omega=\frac{1+t}{m}<1.$$

This result suggests that we need $(t+1)/m \leq \omega \leq 1$ in order for ζ_{opt} to estimate α_{opt} found with respect to the projected space.

Remark 2.5 (GCV). Taking $\omega=(t+1)/m$ gives the regularization of the FTSVD of the full problem, in the case of the LSQR with full regularization, namely for moderately and severely ill-posed problems as defined in [15].

3. Simulations: One-dimensional problems. To illustrate the discussion in section 2 we examine the solution of ill-posed one-dimensional problems with known solutions. In all experiments we use MATLAB 2014b and test the problems `phillips` and `gravity`, which are discretizations of Fredholm integral equations of the first kind provided in the Regularization toolbox [10]. Problem `gravity` depends on a parameter d determining the conditioning of the problem; here we use $d=0.75$ yielding a severely ill-posed test problem. In contrast, problem `phillips` is moderately ill-posed and the Picard condition does not hold.[2] Simulations for over- and undersampled data are obtained by straightforward modification of the relevant functions in [10]. We discuss representative results obtained for the undersampled case with $m=152$ and $n=304$, for which the condition number of A is $4.05e+05$ and $3.38e+17$ for `phillips` and `gravity`, respectively. The function `bidiag_gk` associated with the software for [14] is used for finding the factorization (1.6) with full reorthogonalization for all of the basis vectors (the default).

For a given problem defined by (1.1) without noise, noisy data are obtained as

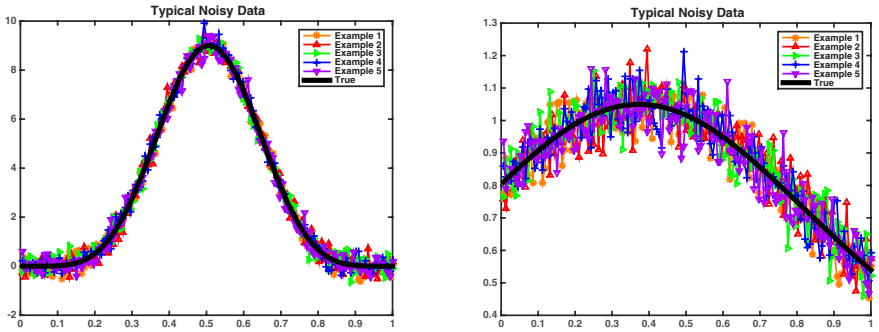
$$\mathbf{b}^c=\mathbf{b}_{\text{ex}}+\boldsymbol{\eta}^c=\mathbf{b}_{\text{ex}}+\eta\|\mathbf{b}_{\text{ex}}\|_2\boldsymbol{\epsilon}^c \tag{3.1}$$

for noise level η and with $\boldsymbol{\epsilon}^c$ the cth column of error matrix $\boldsymbol{\mathcal{E}}$ that has columns sampled from a random normal distribution using the MATLAB function `randn`(m,nc). The signal to noise ratio for the data given by

$$\text{BSNR}(\eta,m)=20\log10\left(\frac{\|\mathbf{b}_{\text{ex}}\|}{\|\mathbf{b}^c-\mathbf{b}_{\text{ex}}\|}\right)\approx -20\log10(\eta\sqrt{m}) \tag{3.2}$$

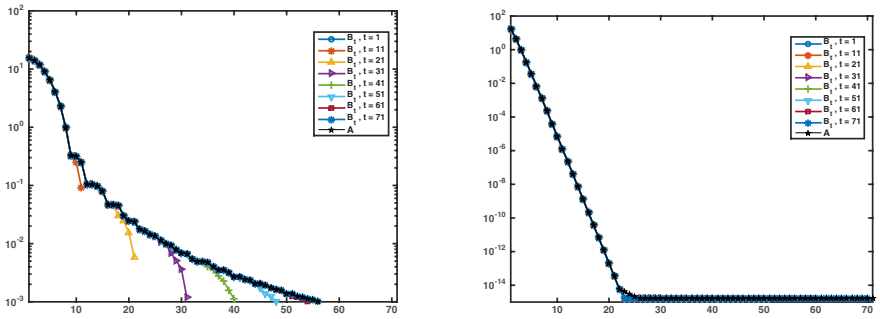
is independent of the test problem. In particular, $\text{BSNR}(.005,152)\approx 24.2$. Example simulation data are shown in Figure 1 for noise levels $\eta=.005$ for each test problem with $m=152$ and in each case for 5 samples of the noise, $\mathbf{b}^c$, $c=1\colon 5$. In all simulations the matrices and right-hand side data are weighted by the diagonal inverse square root of the covariance matrix, assuming colored noise.

[2]The continuous and discrete Picard conditions are well described in the literature, e.g., [11]. Basically, the Picard condition holds if the absolute values of the coefficients of the solution decay on average faster than the singular values.



(a) `phillips` (b) `gravity`, $d = .75$

FIG. 1. *Illustrative test data for noise level* $\eta = .005$ *for sample right-hand side data* $\mathbf{b}^c$, $c = 1\colon 5$, *with* $m = 152$ *and* $n = 304$. *Exact solutions are given by the solid lines in each plot.*

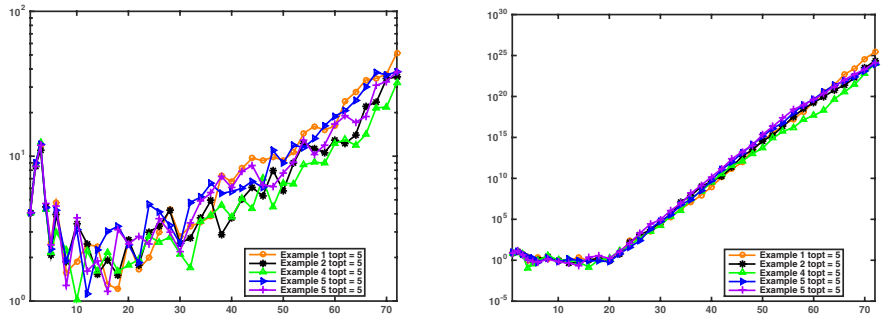


(a) `phillips` (b) `gravity` $d = .75$

FIG. 2. *Plots of singular values against the index of the singular value for matrix* B_t *for increasing* t, $t = 1 : 10 : 71$, *compared to the first* 71 *singular values of matrix* A *for underdetermined cases* $m = 152$ *and* $n = 304$.

3.1. Spectra of A and B_t. Figure 2 illustrates the spectra of the matrices A and B_t for each test problem for the pairs $(m, n) = (152, 304)$ for $t = 1\colon 10\colon 91$ and with noise level $\eta = .005$ used in the calculation of B_t. For `phillips` there are clear steps in the singular values at indices 9, 12, 16, 19, and 22, and the problem is only moderately ill-posed. In contrast, `gravity` is severely ill-posed: the singular values decay continuously and exponentially to machine precision. Because `gravity` is severely ill-posed, the LSQR iteration quickly captures the dominant right singular subspace for small t. On the other hand, for `phillips`, the slower decay of the spectrum and the generation of small Ritz values introduces inaccurate small singular values into the spectrum of B_t, as is seen by the departure of the spectrum of B_t from the spectrum of A. This may present difficulty for estimating the regularization parameter using the presented approaches, unless t is small.

3.2. Estimating t: The subproblem size. To determine the size of the subproblem we examine the approach suggested in [14, equation (3.9)] for determining the appearance of noise in the subspace. Denoting the diagonal components of B_t by θ_j, $j = 1\colon t$, and the subdiagonal entries by β_j, $j = 2\colon t+1$, the cumulative ratio, $\rho(t) = \prod_{j=1}^{t}(\theta_j/\beta_{j+1})$ shows the impact of β_{j+1} approaching the precision of the



(a) `phillips` (b) `gravity` $d = .75$

FIG. 3. *Noise revealing function* $\rho(t)$ *for the two test problems for underdetermined cases with noise level* $\eta = .005$ *as in Figure* 2.

algorithm as t increases. For large enough t and exact arithmetic, $\beta_{t+1} \ll \theta_t$ [14]. Noise is identified as entering at optimal iteration $t_{\text{opt}-\rho}$, given by

$$t_{\text{opt}-\rho} = \min\{\arg\max_{t>t_{\min}} (\rho(t))\} + 2. \tag{3.3}$$

Here $t_{\min}$ is problem dependent and chosen to ensure that noise has entered the problem and we take 2 steps beyond $t_{\min}$. There is little difference in the characteristic oscillation of $\rho(t)$ considered in our examples; noise enters quickly and $t_{\min} = 3$ is already sufficient. Figure 3 illustrates $\rho(t)$ for the test problems with 5 samples of the noisy data. $\rho(t)$ correctly identifies the point at which the singular values reach machine precision for problem `gravity`. We point out that the noise levels used in these examples, and the subsequent simulations, are substantially larger than the noise levels used in [14], for which the optimal choice of t is thus correspondingly larger. If we run our examples with less significant noise, we do obtain results consistent with [14]. It was already noted in [14] that (3.3) cannot be used when the discrete Picard condition does not hold, e.g., for `phillips`. An alternative method for identifying the subspace size is to minimize the GCV function for the TSVD [3, equation (3.12)]:

$$\mathcal{G}(t) = \frac{t_{\max}}{(t_{\max} - t)^2} \sum_{t+1}^{t_{\max}} |\mathbf{u}_i^T \mathbf{b}|^2. \tag{3.4}$$

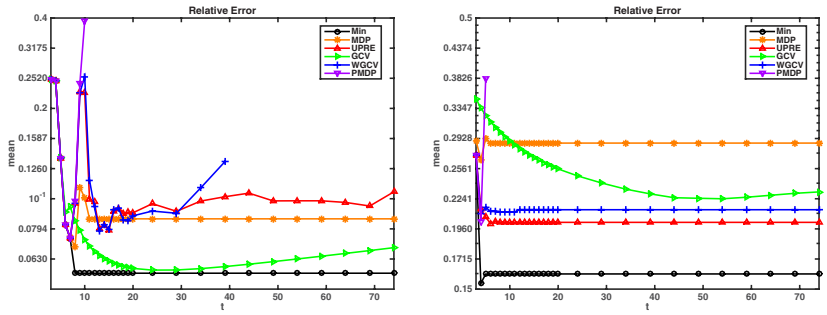
$\mathcal{G}(t)$ depends on the choice of $t_{\max}$, i.e., the size of the largest subspace considered; in contrast, $\rho(t)$ depends on the selection of $t_{\min}$ but is independent of $t_{\max}$. To assess the impact of the *correct* choice of t we also examine solutions obtained with larger subspaces.

3.3. Regularization parameter estimation. In implementing the parameter regularization we test the GCV with $\omega = 1$ as compared to WGCV with $\omega = (t + 1)/m$, and the MDP with $\delta = m$ as compared to projected MDP (PMDP) $\delta = (t + 1)$. Estimates using GCV, WGCV, MDP, PMDP, and UPRE are obtained by calculating the relevant functions for the same set of regularization parameters and then minimizing within the region of the optimum as used in the Regularization Tools for the GCV. For all tests the regularization parameter yielding a minimum error (denoted in results by MIN) for the projected space is also found by calculating

TABLE 1

Average RE over 50 *samples for problem size* $m = 152$ *and* $n = 304$. $t_{\min} = 3$ *with average* $t_{\text{opt}-\rho}$ *as given. Results are illustrated in Figure* 4.

	MIN	MDP	UPRE	GCV	WGCV	PMDP
	Average RE: Average $t_{\text{opt}-\rho} = 5$					
`phillips`	0.16	0.16	0.16	0.17	0.16	0.16
`gravity` $d = .75$	0.17	0.66	0.52	0.35	0.49	> 1
	Minimum RE (average t for the minimum)					
`phillips`	0.06 (9)	0.07 (8)	0.07 (7)	0.06 (24)	0.07 (7)	0.07 (7)
`gravity` $d = .75$	0.15 (4)	0.27 (4)	0.21 (4)	0.22 (54)	0.22 (4)	0.22 (4)



(a) `phillips` (b) `gravity`, $d = .75$

FIG. 4. *Average RE over all samples against* t *for the underdetermined case:* $m = 152$ *and* $n = 304$. *Noise level* $\eta = .005$.

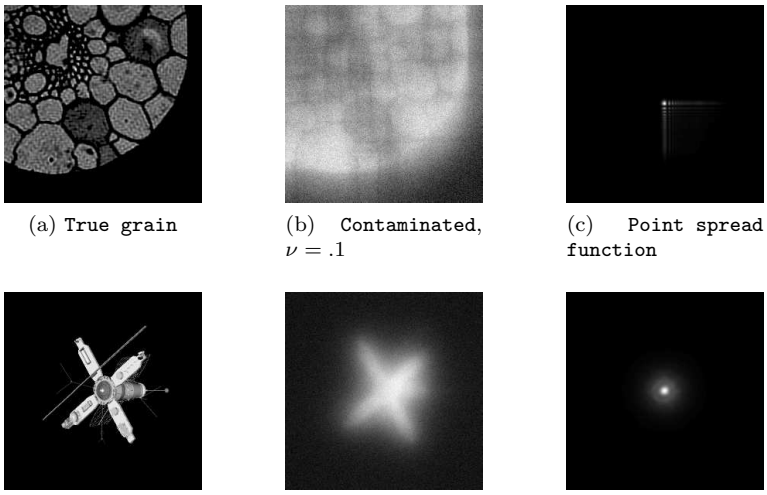
the regularization parameter at 1000 logarithmically sampled points between γ_1 and $\max(10^{-14}\gamma_1, \gamma_t)$.[3]

3.4. Evaluating the results. Contaminated data $\mathbf{b}^c$ are generated using (3.1) for $c = 1\colon 50$ yielding solutions $\mathbf{x}(t, c)$ for problem size t with $t = [3\colon 20, 24\colon 5\colon 74]$. The relative error (RE) of the solution with respect to the known true solution is given by

$$\text{RE}(t, c) = \|\mathbf{x}(t, c) - \mathbf{x}_{\text{ex}}\|_2 / \|\mathbf{x}_{\text{ex}}\|_2.$$

The results in Table 1 are the average RE over all 50 samples at the reported average $t_{\text{opt}-\rho}$ and the minimum RE over all samples and all t. These results summarize the graphs of the errors for the same cases in Figure 4. Immediately we observe that these one-dimensional problems generally need rather small subspaces to obtain optimal solutions with respect to the full problem. They also show that the estimators, other than the PMDP and WGCV, are quite robust to the choice of t, away from the optimal value. As suggested by the spectra of B_t and A shown in section 3.1, the WGCV is robust for `gravity`, for which the LSQR gives the full regularization, but not for `phillips`. Parameter dependent PMDP is not robust in either case. Also, the GCV has minimal error for a larger subspace, and thus estimating t_{opt} via $t_{\text{opt}-\rho}$ is not

[3] In practice, one would not take such a large selection of regularization parameters, but in tests we found that the minimization step in the UPRE may yield reduced error if the optimum is not found by sampling over a sufficiently fine distribution for the regularization parameter.



(a) `True grain` (b) `Contaminated`, $\nu = .1$ (c) `Point spread function`

(d) `True satellite` (e) `Contaminated`, $\nu = .1$ (f) `Point spread function`

FIG. 5. *Data for grain and satellite images with blur by the given point spread function and noise level* 10%, *corresponding to* $\nu = .1$.

effective. On the other hand, using (3.4), $t_{\text{opt}-\mathcal{G}}$ is larger (6 and 74, respectively, for the two problems), and a smaller error is achieved by GCV (.13 and .23, respectively).

The results confirm the expectation of the analysis on the performance of the techniques dependent on the degree of ill-posedness of the problem: `gravity` is more severely ill-conditioned and the WGCV and UPRE estimators give robust solutions independent of t, improving on the GCV and MDP solutions.

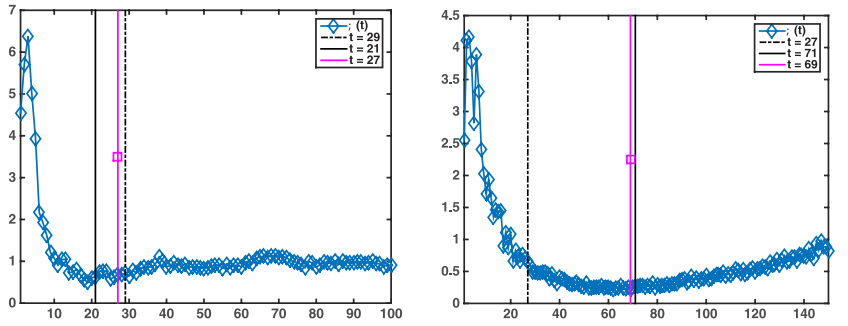
4. Two-dimensional simulations.

4.1. Image deblurring. We consider image deblurring problems, `grain` and `satellite`, of size 256×256 from RestoreTools [18]. Our main aim here is to first demonstrate the use of the regularization techniques PMDP, WGCV, and UPRE for increasing t, and then to examine a stabilizing technique using an IRR (section 4.4). Results without IRR are presented for completeness in section 4.3 and with IRR in section 4.4.2.

For contrast with the results presented in [18], we use noise levels $\eta = .00039$ and .00019 in (3.1), which correspond to noise levels $\nu = 10\%$ and 5%, respectively, in [18] with $\|\boldsymbol{\eta}^c\|_2 = \nu\|\mathbf{b}_{\text{ex}}\|_2$, yielding $\nu = \eta\sqrt{m}$. These correspond to BSNR 20 dB and 26 dB, as calculated by (3.2).[4] For immediate comparison with [18] we indicate the results using the noise level ν rather than η. In Figure 5 we give the true solution, blurred and noisy data, and the point spread function.

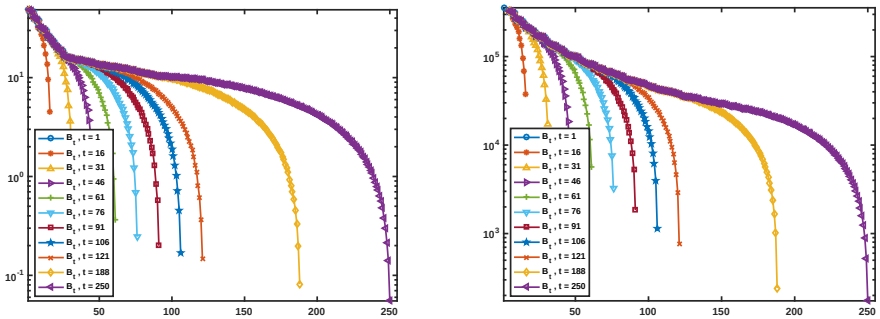
4.2. Algorithm details. In finding the restorations for the data indicated in Figures 5(b) and 5(e) we note first that the matrices A for the point spread functions indicated in Figures 5(c) and 5(f) do not satisfy the Picard condition. As illustrated

[4]For comparison with the results in [3], we note that there the BSNR is calculated using a definition in [19], and their results with 10 dB and 25 dB according to that definition yield 19 dB and 34 dB with (3.2), respectively, i.e., approximately 11% and 2% noise. Hence the results here for 10% may be approximately compared with the 10 dB results in [3].



(a) `grain` $\rho(t)$ (b) `satellite` $\rho(t)$

FIG. 6. *Noise revealing function* $\rho(t)$ *for the data illustrated in Figures* 5(b) *and* 5(e) *with* $t_{\min} = 25$. *Here the dot-dashed vertical line corresponds to the location of* $t_{\text{opt}-\rho}$, *the solid line with the symbol to* $t_{\text{opt}-\mathcal{G}}$, *and the solid line to* $t_{\text{opt}-\min}$.



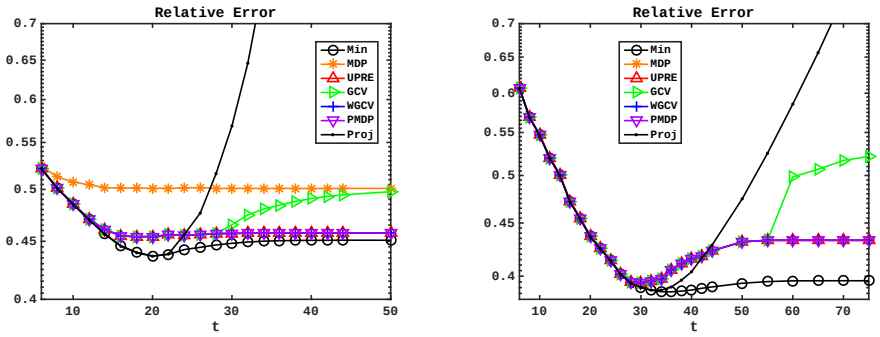
(a) `grain` (b) `satellite`

FIG. 7. *Plots of singular values against the index of the singular value for matrix* B_t *for increasing* t, $t = 1 : 15 : 121$, $t = 188$, *and* $t = 250$.

in Figure 6, $\rho(t)$ does not show the increase within the shown range of t, as is clear for $\rho(t)$ with large t obtained for the one-dimensional examples in Figure 3. The spectra for increasing t, shown in Figure 7, also demonstrate that these problems are only mildly to moderately ill-posed so that the LSQR iterate does not adequately capture the right singular subspace. Still, $\rho(t)$ attains a minimum within the shown range for t and then exhibits a gradual increase. This suggests that noise is entering the data after the minimum and that one may use

$$t_{\text{opt}-\min} = \arg\min(\rho(t)) + 2,$$

where again we advance 2 steps under the assumption that noise enters after the minimum. In Figure 6 the vertical lines indicate the positions of $t_{\text{opt}-\rho}$, $t_{\text{opt}-\mathcal{G}}$, and $t_{\text{opt}-\min}$. For `grain`, $t_{\text{opt}-\mathcal{G}}$ becomes quickly independent of the number of terms used, already stabilizing at $t_{\text{opt}-\mathcal{G}} = 27$ with just $t_{\max} = 50$ terms, with no change even out to a maximum size of $t_{\max} = 250$ in the calculation. For `satellite`, $t_{\text{opt}-\mathcal{G}}$ is less stable, only reaching 32 when $t_{\max} = 100$ terms are used in the estimation, but increasing to 96 if $t_{\max} = 250$ terms are imposed. Stability in the choice of $t_{\text{opt}-\rho}$ with respect to $t_{\min}$ also follows lack of stability in the choice of $t_{\text{opt}-\mathcal{G}}$, suggesting that it



(a) `grain` RE (b) `satellite` RE

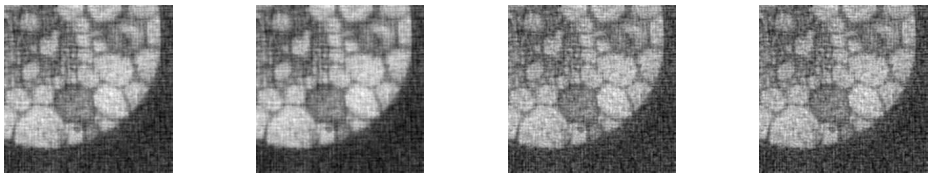
FIG. 8. *RE with increasing t with regularization parameter calculated by the different regularization techniques for examples illustrated in Figures* 5(b) *and* 5(e). *The solid line in each case is the solution with projection and without regularization.*

is preferable to use $t_{\text{opt}-\text{min}}$. In our experiments we have deduced that it is important to examine the characteristic shape of $\rho(t)$ in determining the optimal choice for the size of the subspace, and we will show results using $t_{\text{opt}-\text{min}}$, $t_{\text{opt}-\rho}$, and $t_{\text{opt}-\mathcal{G}}$.

The range for the regularization parameter is also important, as indicated through the windowing approach based on (3.4) [1]. From Figure 7 it is clear that LSQR iteration only provides a *partial regularization* for either problem and that B_t only captures a portion of the spectrum. Thus we use a single window defined by t^* and apply an FTSVD for the solution which is dominant for the first t^* terms, i.e., with filter factors $\phi_i(\zeta) \approx 1$ for $i < t^*$ and $\phi_i(\zeta) \to 0$ for $i > t^*$. With $\zeta = \tau\gamma_{t^*}$, $\phi_{t^*}(\zeta) = 1/(1+\tau^2) < 1$. In our results, $\tau = .1$, $t^* = \max(t_{\text{opt}-\rho}, t_{\text{opt}-\mathcal{G}})$, and we impose $\tau\gamma_{t^*} < \zeta < \gamma_1$ for the range of ζ. For the *minimal* (MIN) solutions, the range is adjusted to $10^{-1.5} \le \zeta \le 1$, consistent with the range for the regularization parameter used in [3]. This range is scaled by the mean of the standard deviation of the noise in the data, consistent with the inverse covariance scaling of the problems. In finding the MIN solution the range is sampled at 100 logarithmically sampled points.

4.3. Results. For quantitative measurement of a given solution as compared to the known solution we again use the RE. Other possibilities include using the signal to noise ratio, which is directly related to the RE, and the mean structural similarity index (MSSM) suggested in [31]. We found in our experiments that high MSSM corresponds to low RE, and thus providing these results delivers little in terms of further assessment of the algorithms for image deblurring. The REs using the regularization parameter estimators in contrast to MIN and Proj are illustrated in Figure 8 for restoration of the images in Figures 5(b) and 5(e). The results are consistent with the literature in terms of the semiconvergence behavior of the LSQR and also confirm the increase in error seen with GCV without the weighting parameter. Results obtained with WGCV, PMDP, and UPRE are consistent and verify the analysis in section 3.3, providing a stable solution for increasing t. Solutions found at the noted t_{opt} for UPRE, as compared to the *optimal* solution with minimum error, are illustrated in Figure 9 for problem `grain`. Results for the `satellite` image are similar. Overall the results demonstrate that the restorations are inadequate at this level of noise.

4.4. Iteratively reweighted regularization. Iteratively reweighted regularization (IRR) provides a cost effective approach for sharpening images, e.g., [32], and



(a) MIN $t_{\text{opt}} = 22$ (b) $t_{\text{opt}-\text{min}} = 21$ (c) $t_{\text{opt}-\mathcal{G}} = 27$ (d) $t_{\text{opt}-\rho} = 29$

FIG. 9. *Solutions for noise level* 10% *for* `grain` *using UPRE to find the regularization parameters and comparing the solutions obtained for* $t_{\text{opt}-\rho}$, $t_{\text{opt}-\text{min}}$, *and* $t_{\text{opt}-\mathcal{G}}$ *as compared to the solution with minimum error, MIN.*

has been introduced and applied for focusing geophysical inversion, denoted in this context as minimum support regularization [24, 27, 28, 29, 33]. Regularization operator D is replaced by a solution dependent operator $D^{(k)}$, initialized with $D^{(0)} = I$ and $\mathbf{x}^{(0)} = 0$, and yields iterative solution $\mathbf{x}^{(k+1)}(\alpha)$. For $k > 0$,

$$(D^{(k)})_{ii} = ((\mathbf{x}_i^{(k)} - \mathbf{x}_i^{(k-1)})^2 + \beta^2)^{-1/2},$$

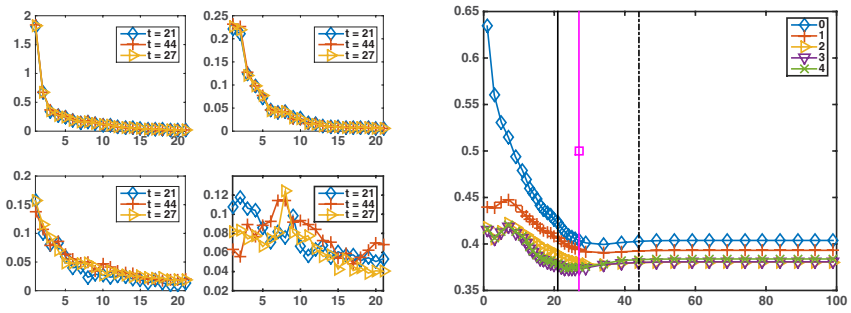
where $\beta > 0$ is a focusing parameter which assures that $D^{(k)}$ is invertible. Immediately,

$$(D^{(k)})_{ii}^{-1} = ((\mathbf{x}_i^{(k)} - \mathbf{x}_i^{(k-1)})^2 + \beta^2)^{1/2}.$$

Thus we can use (1.4) with system matrix $\tilde{\tilde{A}}^{(k)} = \tilde{A}(D^{(k)})^{-1}$ to obtain the iterative solution $\mathbf{x}^{(k+1)}$, $k > 0$. Furthermore, it is straightforward to modify the algorithm for calculation of the factorization (1.6) for the left and right preconditioned matrix $\tilde{\tilde{A}}^{(k)}$, also noting for the specific preconditioners that operations with the diagonal matrix are simple componentwise products.

4.4.1. Comments on the parameter β. Suppose that $\mathbf{x}_i^{(k)} = \mathbf{x}_i^{(k-1)}$ for $i \in \mathcal{I}$ and $\beta = 0$, then $(D^{-1})_{ii} = 0$ for $i \in \mathcal{I}$. Then, rather than solving $\tilde{A}D^{-1}\mathbf{z} \approx \mathbf{r}$, we solve the reduced system $\hat{A}\hat{\mathbf{z}} \approx \mathbf{r}$, where $\hat{A}$ is $\tilde{A}$ but with column i removed for $i \in \mathcal{I}$ and all other columns scaled by the relevant diagonal entries from $(D^{-1})_{ii}$. Matrix $\hat{A}$ is of size $m \times \hat{n}$ where $\hat{n} = n - |\mathcal{I}|$ and vector $\hat{\mathbf{z}}$ is vector $\mathbf{z}$ with entries $i \in \mathcal{I}$ removed. With regularization, $\hat{\mathbf{z}}$ is obtained as the solution of $(\hat{A}^T\hat{A} + \alpha^2 I_{\hat{n}})\hat{\mathbf{z}} = \hat{A}^T\mathbf{r}$. Thus $\hat{\mathbf{y}} = \hat{D}^{-1}\hat{\mathbf{z}}$, where $\hat{D}$ is obtained from D with the same diagonal entries $i \in \mathcal{I}$ removed. The update for $\mathbf{x}$ is therefore obtained using (1.5) with entries $\mathbf{x}_i^{(k)}(\alpha) = \mathbf{x}_i^{(k-1)} + \hat{\mathbf{y}}_i$ for $i \notin \mathcal{I}$ and $\mathbf{x}_i^{(k)}(\alpha) = \mathbf{x}_i^{(k-1)}$ for $i \in \mathcal{I}$. Forthwith we use $\beta = 0$ and factorize the reduced system with system matrix $\hat{A}$.

4.4.2. Algorithmic details for IRR. The approach for the iteration requires some explanation as to how the range of t is obtained at IRR iterations $k > 0$. Noise revealing function $\rho^{(k)}(t^{(k-1)}, t)$ depends on the subspace size $t^{(k-1)}$ from the prior step k and the current subspace size t. Further, $t_{\text{opt}-\rho}$, $t_{\text{opt}-\text{min}}$, and $t_{\text{opt}-\mathcal{G}}$ are all dependent on $t^{(k-1)}$ as well as $t_{\text{min}}^{(k)}(t^{(k-1)})$ and $t_{\text{max}}^{(k)}(t^{(k-1)})$, i.e., given a specific subspace size at $t^{(k-1)}$, the minimum and maximum sizes to be used at step k need to be specified. Because we anticipate that further noise will enter with increasing k, we expect that $t_{\text{min}}^{(k)} < t_{\text{min}}^{(k-1)}$ and that $t_{\text{max}}^{(k)}(t) < t_{\text{opt}}^{(k-1)}$. With these constraints, the cost



(a) $k > 0$ $\rho(t)$ (b) RE

FIG. 10. *Demonstration of determining the projected problem size with IRR iterations for problem* `grain` *with* 5% *noise.* (a) $\rho^{(k)}(t)$ *for increasing* k *for* $k = 1\colon 4$ *left to right and from top to bottom.* (b) *The dot-dashed vertical line corresponds to the location of* $t_{\text{opt}-\rho}$*, the solid line with the symbol to* $t_{\text{opt}-\mathcal{G}}$*, and the solid line to* $t_{\text{opt}-\text{min}}$.

of an IRR step will be less than the first step $k = 0$. Examination of $\rho^{(k)}(t^{(k-1)}, t)$ is useful in identifying the constraints on $t^{(k)}$. For each iteration the range for ζ is constrained using the current singular values, by $\tau\gamma_t^* \leq \zeta \leq \gamma_1$, where at step 0, $t^* = \max(t_{\text{opt}-\rho}, t_{\text{opt}-\mathcal{G}})$ and $t^* = t_{\text{opt}-\rho}$ for the IRR updates. We will examine the choices for the case with 5% noise.

4.4.3. Results with IRR for 5% noise. Here we only report results obtained using the UPRE as compared to the *optimal* solutions. These results are indicative of IRR implemented with the other methods for finding the regularization parameter. To examine the process carefully in one case, we focus on problem `grain` with 5% noise. Function $\rho(t)$ at the first step $k = 0$ does not differ significantly from the case with 10% noise, shown in Figure 6. We use a maximum subspace size with $t = 100$ for the calculation of $t_{\text{opt}-\mathcal{G}}$. Figure 10(a) shows $\rho^{(k)}(t^{(k-1)}, t)$ for the choices of t, $t_{\text{opt}-\rho}$, $t_{\text{opt}-\text{min}}$, and $t_{\text{opt}-\mathcal{G}}$. It is clear that $\rho^{(k)}(t^{(k-1)}, t)$ is almost independent of $t^{(k-1)}$ for the first steps, but that noise enters for $k = 4$. This is also reflected in the RE in Figure 10(b): the RE stabilizes for increasing t and decreases for the first three steps of IRR but increases at step 4. The impact of the IRR is similar for problem `satellite`.

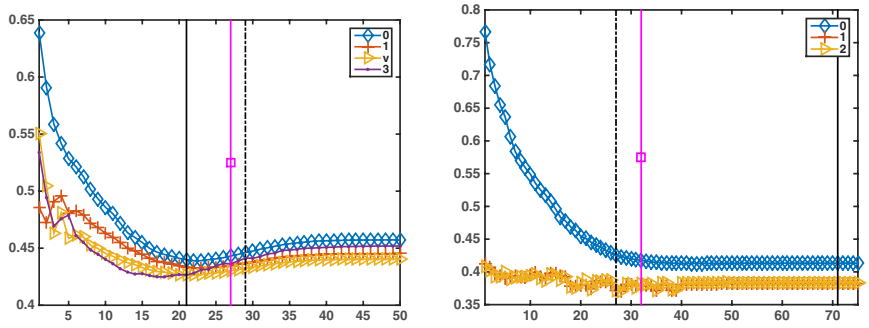
The RE for the two simulations, contrasted with MIN, are detailed in Table 2. The IRR stabilizes the solutions leading to results comparable to those of MIN. Moreover, it is also clear that one may not conclude that finding t_{opt} using $t_{\text{opt}-\text{min}}$ is preferable to using $t_{\text{opt}-\mathcal{G}}$ or $t_{\text{opt}-\rho}$. The results in Table 2 indicate that using $t_{\text{opt}-\rho}$ leads to the best solutions in one case, and $t_{\text{opt}-\mathcal{G}}$ in the other, although effectively the quality is comparable. Provided that the solutions are stabilized with the IRR, improvements in the solutions are obtained in a limited number of steps using IRR with relatively small subspaces for the iterative updates.

4.4.4. Terminating the IRR iteration. The graphs of $\rho(t)$ with increasing k in Figure 10(a) indicate that the properties of $\rho^{(k)}(t^{(k-1)}, t)$ can be used to determine effective termination of the IRR, based on the iteration k when noise enters into $\rho^{(k)}(t^{(k-1)}, t)$. Our experience has shown that the optimal solution in terms of image quality is achieved not at the step before noise enters in $\rho^{(k)}(t^{(k-1)}, t)$, but two steps before.

TABLE 2

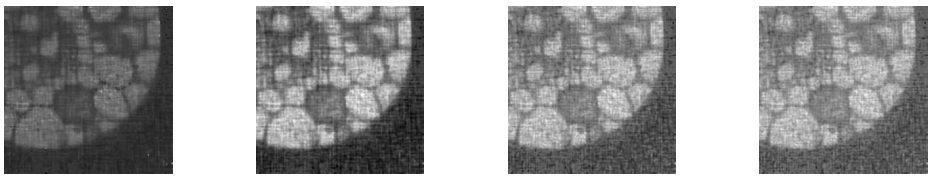
RE for problem `grain` *and problem* `satellite` *with* 5% *noise corresponding to* $\nu = .05$ *for solutions found using different selection of* t_{opt}*, as compared to the optimum found with UPRE and the overall optimum selected over the range for* ζ*.*

	`grain` with $t_{\text{opt}-\text{min}} = 21$ and $t_{\text{opt}-\mathcal{G}} = 27$			
Iteration	$t_{\text{opt}-\text{min}}$	$t_{\text{opt}-\mathcal{G}}$	MIN for UPRE	Overall MIN
1	0.4248	0.4074	0.4030	0.3993
2	0.4073	0.3962	0.3929	0.3903
3	0.3899	0.3811	0.3788	0.3776
4	0.3775	0.3731	0.3728	0.3731
	`satellite` with $t_{\text{opt}-\rho} = 42$ and $t_{\text{opt}-\mathcal{G}} = 69$			
Iteration	$t_{\text{opt}-\rho}$	$t_{\text{opt}-\mathcal{G}}$	MIN for UPRE	Overall MIN
1	0.3566	0.3520	0.3517	0.3863
2	0.3375	0.3430	0.3373	0.3382
3	0.3374	0.3431	0.3372	0.3352
4	0.3385	0.3485	0.3371	0.3352



(a) RE `grain` 10% noise (b) RE `satellite` 10% noise

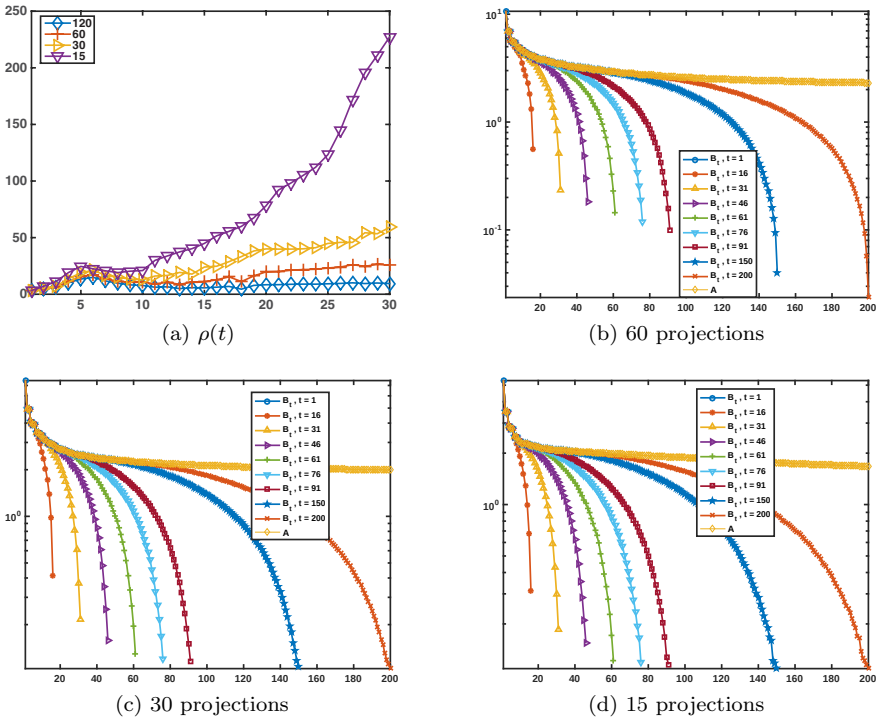
FIG. 11. *RE for problems* `grain` *and* `satellite` *with noise level* 10%*, in (a) and (b), respectively, with increasing iteration in each case for solutions calculated using UPRE. The solutions are stable out to* $k = 2$ *iterations of IRR. Here the dot-dashed vertical line corresponds to the location of* $t_{\text{opt}-\rho}$*, the solid line with the symbol to* $t_{\text{opt}-\mathcal{G}}$*, and the solid line to* $t_{\text{opt}-\text{min}}$*.*



(a) $k = 3$, MIN $t_{\text{opt}} = 19$ (b) $k = 2$, $t_{\text{opt}-\text{min}} = 21$ (c) $k = 2$, $t_{\text{opt}-\mathcal{G}} = 27$ (d) $k = 2$, $t_{\text{opt}-\rho} = 29$

FIG. 12. *Solutions for noise level* 10% *for* `grain` *using UPRE to find the regularization parameters and comparing the solutions obtained for different* t_{opt}*, as compared to the solution with minimum error, MIN, for IRR at the indicated step applied to the solutions in Figure* 9*.*

4.4.5. Results for 10% noise. In Figure 11 we illustrate how the RE changes with the iteration count. The vertical lines in Figure 11(a) demonstrate that using $t_{\text{opt}-\rho}$ or $t_{\text{opt}-\mathcal{G}}$ makes little difference to the quality of the solution when measured with respect to the RE. Example solutions for `grain` are given in Figure 12, for contrast with the solutions without IRR shown in Figure 9.



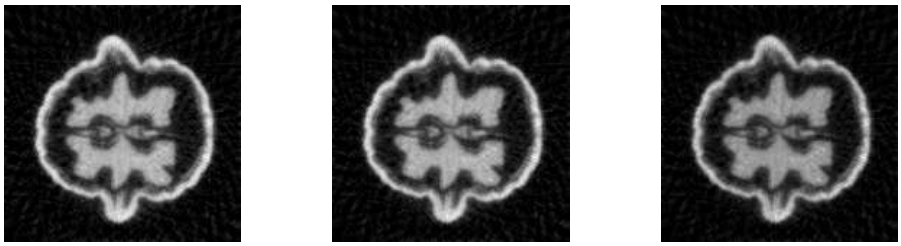
(a) $\rho(t)$ (b) 60 projections

(c) 30 projections (d) 15 projections

FIG. 13. $\rho(t)$ *for increasing sparsity for resolution* 164 *for the walnut data* (a) *and the spectra for increasing* t, $t = 1 : 15 : 91$, 150, 200 *with increasing sparsity* (b)–(d).

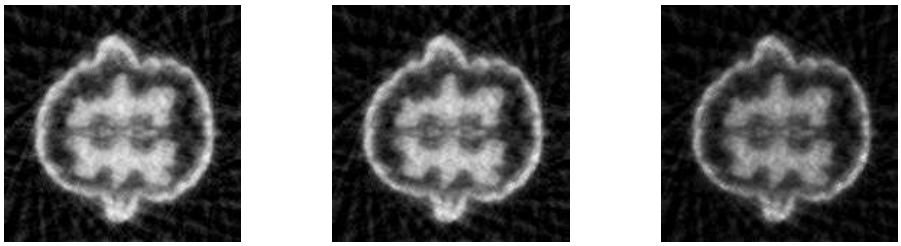
4.5. Sparse tomographic reconstruction of a walnut. To contrast the success of the regularization parameter estimation techniques in the context of a two-dimensional projection problem, we present results for the reconstruction of projection data obtained from tomographic x-ray data of a walnut, used for edge preserving reconstruction in [8] and with the description of the data described in [7]. Datasets `DataN` correspond to the resolution $N \times N$ in the image and use 120 projections, corresponding to 3° sampling. Data are provided with $N = 82$, 164, and 328. We use resolution 164 with 120 projections, and then downsampled to 60, 30, and 15 projections, i.e., angles 3°, 6°, 12°, and 24°. Results with resolutions 82 and 328 are comparable. Results are presented using the solution at $t_{\text{opt}-\rho}$ regularized using UPRE; see Figures 14 and 15. Results for 120 projections are almost perfect due to the apparent limited noise in the provided data, while the results with 15 projections clearly show the projection data. In all the presented results the parameters $t_{\text{opt}-\rho}$ are determined automatically, after manually picking $t_{\min} = 5$ from manual consideration of the plot for $\rho(t)$; see Figure 13. Further, from Figures 13(b)–13(d), it is immediate that in this case the LSQR iteration only offers a partial regularization independent of the sparsity and thus again it is necessary to identify the window for the usable spectrum. All parameters are then estimated in the same way as for the image restoration cases.

Figures 14–15 show results for one set of data with increasing sparsity; compare with [8, Figures 6.6 and 6.7], which give results with resolution for $N = 128$ and 256, respectively, and angle separation 2°, 4°, 6°, and 12°. The approach in [8] uses selected choices for the regularization parameter based on a sparsity argument



(a) UPRE: $k = 0$ (b) UPRE: $k = 1$ (c) UPRE: $k = 2$

FIG. 14. *Solutions at increasing iterations for a walnut with resolution* 164×164*,* $t_{\min} = 5$*,* $t_{\text{opt}-\rho} = 8$*, positivity constraint, and sampling at* 6° *intervals, for* 60 *projections.*



(a) UPRE: $k = 0$ (b) UPRE: $k = 1$ (c) UPRE: $k = 2$

FIG. 15. *Solutions at increasing iterations for a walnut with resolution* 164×164*,* $t_{\min} = 5$*,* $t_{\text{opt}-\rho} = 8$*, positivity constraint, and sampling at* 12° *intervals, for* 30 *projections.*

with prior information and seeks to support the use of the sparsity argument for reconstruction of sparse data sets. The images exhibit the rather standard total variation blocky structures when applied for truly sparsely sampled data. Our results show robust reconstructions with the automatically determined solutions, after first examining the plot for $\rho(t)$. IRR generates marginal improvements in qualitative solutions. To show the impact of the correct choice of t_{opt} on the solution, we show a set of results at iteration $k = 0$ using $t_{\min} = 18$ in Figure 16 with the positive constraint. The UPRE yields solutions qualitatively similar to the case with $t_{\min} = 5$. In the examples here, we do impose an additional positivity constraint on the solutions at each step, before calculating the iterative weighting matrix.

The results demonstrate that the projected problem with automatic determination of ζ_{opt} can be used to reconstruct sparsely sampled tomographic data, provided that an initial estimate for $t_{\min}$ is manually determined by consideration of the plot of $\rho(t)$. Further, IRR stabilizes the solutions. For the sparse data sets the solutions do not exhibit the characteristic blocky reconstructions of total variation image reconstructions as seen in [8].

5. Conclusions. We have demonstrated that regularization parameter estimation by the method of UPRE can be effectively applied for regularizing the projected problem. Our results also motivate a choice for the weighting parameter in the WGCV. These results apply for the concept of *full regularization*, which was recently introduced in [15]. It was argued there, however, that in the case of full regularization, no additional regularization of the iterate is required, because the LSQR iterate approximates the TSVD solution of the full problem. It is known, however, that even

(a) UPRE: 60 (b) UPRE: 30 (c) UPRE: 15

FIG. 16. *Solutions for a walnut with resolution* 164×164 *without IRR, i.e., at* $k = 0$ *with* $t_{\min} = 18$ *yielding* $t_{\text{opt}-\rho} = 25$, 29, *and* 40, *automatically determined for sampling intervals* 6°, 12°, *and* 24°.

with truncation, additional filtering through the use of an FTSVD solution is required. Moreover, while the projected solution may not actually need regularization, without prior knowledge of the optimal subspace size t, it is difficult to determine the point at which semiconvergence contaminates the projected solution. Hence, effectively regularizing via the hybrid LSQR is still necessary. Our results demonstrate that the regularization estimators will find effective regularization of the FTSVD solution of the full problem. In the case of the *partial regularization*, i.e., in the cases in which a small Ritz value appears before the LSQR iterate has captured the dominant SVD components of A, as discussed in [15], regularization on the projected problem will not adequately regularize the full problem. Here we handled the situation in which LSQR does not sufficiently capture the dominant singular space by restricting the range for the regularization parameter dependent on the singular value for $\gamma_{t_{\text{opt}}}$, in order to effectively use an FTSVD solution of the projected problem. For future investigation, we suggest that it is important to identify the extent to which the LSQR iterate captures a right singular subspace for A, and then to potentially use more than one regularization parameter, one which is chosen to regularize the dominant terms of the spectrum, and one which handles the small singular values of B_t, extending the windowed regularization parameter techniques [1].

This work also demonstrates that edge preserving regularization, via iterative reweighting, can be applied to stabilize regularized solutions of the projected problem. Our results suggest that manual estimation of a minimal subspace size can then lead to useful estimates for an optimal projected space, with the use of the IRR leading to improvements in the solutions when t_{opt} is found by different methods, including the use of $t_{\text{opt}-\rho}$, $t_{\text{opt}-\min}$, and $t_{\text{opt}-\mathcal{G}}$, hence making the determination of this t_{opt} less crucial in providing an acceptable solution. Future work on this topic should include extending the use of more general iteratively reweighted regularizers, accounting for edges in more than one direction in conjunction with the projected solutions.

Appendix A. Expansion solutions. Suppose the SVD of matrix A, $A \in \mathcal{R}^{m\times n}$ is given by $A = U\Sigma V^T$, where the singular values are ordered $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{m^*} > 0$ and occur on the diagonal of $\Sigma \in \mathcal{R}^{m\times n}$ with $n - m$ zero columns (when $m < n$) or $m - n$ zero rows (when $m > n$), and $U \in \mathcal{R}^{m\times m}$ and $V \in \mathcal{R}^{n\times n}$ are orthogonal matrices [6]. Then,

$$\mathbf{x}(\alpha) = \sum_{i=1}^{m^*} \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i = \sum_{i=1}^{m^*} \phi_i(\alpha) \frac{\hat{b}_i}{\sigma_i} \mathbf{v}_i, \quad \hat{b}_i = \mathbf{u}_i^T \mathbf{b}. \tag{A.1}$$

For the projected case, $B_t \in \mathcal{R}^{(t+1)\times t}$, i.e., $m > n$, and the expression still applies with $\|\mathbf{b}\|_2\mathbf{e}_1^{(t+1)}$ replacing $\mathbf{b}$, ζ replacing α, γ_i replacing σ_i and $m^* = t$ in (A.1).

Appendix B. Regularization parameter estimation. All formulae apply using the SVD for B_t replacing that for matrix A.

B.1. Unbiased predictive risk estimator. The UPRE function is given by

$$U(\alpha) = \sum_{i=1}^{m^*} \left(\frac{1}{\sigma_i^2\alpha^{-2}+1} \right)^2 \hat{b}_i^2 + 2\left(\sum_{i=1}^{m^*} \phi_i \right) - m.$$

B.2. Morozov discrepancy principle. The MDP function is given by

$$\sum_{i=1}^{m^*} \left(\frac{1}{\sigma_i^2\alpha^{-2}+1} \right)^2 \hat{b}_i^2 + \sum_{i=n+1}^{m} \hat{b}_i^2 = \delta.$$

For the projected case, δ_{proj} replaces δ.

B.3. Generalized cross validation. Using the SVD for B_t, the WGCV function is given by

$$G(\zeta,\omega) = \frac{\sum_{i=1}^{t} \left(\frac{1}{\gamma_i^2\zeta^{-2}+1} \right)^2 \hat{b}_i^2 + \sum_{i=t+1}^{t+1} \hat{b}_i^2}{\left((1+t-\omega t) + \omega\zeta^2 \sum_{i=1}^{t} \frac{1}{\gamma_i^2+\zeta^2} \right)^2}.$$

With $\omega = 1$, this reduces to the expression for the projected GCV, (2.12).

REFERENCES

[1] J. Chung, G. Easley, and D. P. O'Leary, *Windowed spectral regularization of inverse problems*, SIAM J. Sci. Comput., 33 (2011), pp. 3175–3200, doi:10.1137/100809787.

[2] J. Chung, J. G. Nagy, and D. P. O'Leary, *A weighted GCV method for Lanczos hybrid regularization*, Electron. Trans. Numer. Anal., 28 (2008), pp. 149–167, http://etna.mcs.kent.edu/volumes/2001-2010/vol28/abstract.php?vol=28&pages=149-167.

[3] J. M. Chung, M. E. Kilmer, and D. P. O'Leary, *A framework for regularization via operator approximation*, SIAM J. Sci. Comput., 37 (2015), pp. B332–B359, doi:10.1137/130945363.

[4] 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, doi:10.1137/10079687X.

[5] G. H. Golub, M. Heath, and G. Wahba, *Generalized cross-validation as a method for choosing a good ridge parameter*, Technometrics, 21 (1979), pp. 215–223, doi:10.2307/1268518.

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

[7] K. Hämäläinen, L. Harhanen, A. Kallonen, A. Kujanpää, E. Niemi, and S. Siltanen, *Tomographic X-ray Data of a Walnut*, preprint, arXiv:1502.04064, 2015.

[8] K. Hämäläinen, A. Kallonen, V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen, *Sparse tomography*, SIAM J. Sci. Comput., 35 (2013), pp. B644–B665, doi:10.1137/120876277.

[9] M. Hanke and P. C. Hansen, *Regularization methods for large-scale problems*, Surv. Math. Ind., 3 (1993), pp. 253–315, http://numerik.mathematik.uni-mainz.de/~hanke/survey.pdf.

[10] P. C. Hansen, *REGULARIZATION TOOLS: A Matlab package for analysis and solution of discrete ill-posed problems*, Numer. Algorithms, 6 (1994), pp. 189–194, doi:10.1007/BF02149761.

[11] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 1998, doi:10.1137/1.9780898719697.

[12] P. C. Hansen and T. K. Jensen, *Noise propagation in regularizing iterations for image deblurring*, Electron. Trans. Numer. Anal., 31 (2008), pp. 204–220, http://etna.mcs.kent.edu/volumes/2001-2010/vol31/abstract.php?vol=31&pages=204-220.

[13] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra, and Filtering*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2006, doi:10.1137/1.9780898718874.

[14] I. Hnětynková, M. Plešinger, and Z. Strakoš, *The regularizing effect of the Golub-Kahan iterative bidiagonalization and revealing the noise level in the data*, BIT Numer. Math., 49 (2009), pp. 669–696, doi:10.1007/s10543-009-0239-7.

[15] Y. Huang and Z. Jia, *Some Results on the Regularization of LSQR for Large-Scale Discrete Ill-Posed Problems*, preprint, arXiv:1503.01864v3, 2015.

[16] 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, doi:10.1137/S0895479899345960.

[17] V. A. Morozov, *On the solution of functional equations by the method of regularization*, Sov. Math. Dokl., 7 (1966), pp. 414–417, http://www.citeulike.org/user/mpisarenco/article/11856055.

[18] 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, doi:10.1023/B:NUMA.0000027762.08431.64.

[19] R. Neelamani, H. Choi, and R. Baraniuk, *ForWaRD: Fourier-wavelet regularized deconvolution for ill-conditioned systems*, IEEE Trans. Signal Process., 52 (2004), pp. 418–433, doi:10.1109/TSP.2003.821103.

[20] C. C. Paige and M. A. Saunders, *Towards a generalized singular value decomposition*, SIAM J. Numer. Anal., 18 (1981), pp. 398–405, doi:10.1137/0718026.

[21] 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, doi:10.1145/355993.356000.

[22] 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, doi:10.1145/355984.355989.

[23] V. Paoletti, P. C. Hansen, M. F. Hansen, and M. Fedi, *A computationally efficient tool for assessing the depth resolution in large-scale potential-field inversion*, Geophysics, 79 (2014), pp. A33–A38, doi:10.1190/geo2014-0017.1.

[24] O. Portniaguine and M. S. Zhdanov, *Focusing geophysical inversion images*, Geophysics, 64 (1999), pp. 874–887, doi:10.1190/1.1444596.

[25] L. Reichel, F. Sgallari, and Q. Ye, *Tikhonov regularization based on generalized Krylov subspace methods*, Appl. Numer. Math., 62 (2012), pp. 1215–1228, doi:10.1016/j.apnum.2010.10.002.

[26] R. A. Renaut, I. Hnětynková, and J. Mead, *Regularization parameter estimation for large-scale Tikhonov regularization using a priori information*, Comput. Stat. Data Anal., 54 (2010), pp. 3430–3445, doi:10.1016/j.csda.2009.05.026.

[27] S. Vatankhah, V. E. Ardestani, and R. A. Renaut, *Automatic estimation of the regularization parameter in 2D focusing gravity inversion: Application of the method to the Safo manganese mine in the northwest of Iran*, J. Geophy. Eng., 11 (2014), 045001, doi:10.1088/1742-2132/11/4/045001.

[28] S. Vatankhah, V. E. Ardestani, and R. A. Renaut, *Application of the χ^2 principle and unbiased predictive risk estimator for determining the regularization parameter in 3-D focusing gravity inversion*, Geophys. J. Int., 200 (2015), pp. 265–277, doi:10.1093/gji/ggu397.

[29] S. Vatankhah, R. A. Renaut, and V. E. Ardestani, *Regularization parameter estimation for underdetermined problems by the χ^2 principle with application to 2D focusing gravity inversion*, Inverse Problems, 30 (2014), 085002, doi:10.1088/0266-5611/30/8/085002.

[30] C. Vogel, *Computational Methods for Inverse Problems*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2002, doi:10.1137/1.9780898717570.

[31] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, *Image quality assessment: From error visibility to structural similarity*, IEEE Trans. Image Process., 13 (2004), pp. 600–612, doi:10.1109/TIP.2003.819861.

[32] B. Wohlberg and P. Rodríguez, *An iteratively reweighted norm algorithm for minimization of total variation functionals*, IEEE Signal Process. Lett., 14 (2007), pp. 948–951, doi:10.1109/LSP.2007.906221.

[33] M. S. Zhdanov, *Geophysical Inverse Theory and Regularization Problems*, Methods in Geochemistry and Geophysics 36, Elsevier, Amsterdam, 2002.