



# Unbiased predictive risk estimation of the Tikhonov regularization parameter: convergence with increasing rank approximations of the singular value decomposition

Rosemary A. Renaut<sup>1</sup> · Anthony W. Helmstetter<sup>1</sup> · Saeed Vatankhah<sup>2</sup>

Received: 5 December 2018 / Accepted: 5 June 2019 / Published online: 14 June 2019  
© Springer Nature B.V. 2019

## Abstract

The truncated singular value decomposition may be used to find the solution of linear discrete ill-posed problems in conjunction with Tikhonov regularization and requires the estimation of a regularization parameter that balances between the sizes of the fit to data function and the regularization term. The unbiased predictive risk estimator is one suggested method for finding the regularization parameter when the noise in the measurements is normally distributed with known variance. In this paper we provide an algorithm using the unbiased predictive risk estimator that automatically finds both the regularization parameter and the number of terms to use from the singular value decomposition. Underlying the algorithm is a new result that proves that the regularization parameter converges with the number of terms from the singular value decomposition. For the analysis it is sufficient to assume that the discrete Picard condition is satisfied for exact data and that noise completely contaminates the measured data coefficients for a sufficiently large number of terms, dependent on both the noise level and the degree of ill-posedness of the system. A lower bound for the regularization parameter is provided leading to a computationally efficient algorithm. Supporting results are compared with those obtained using the method of generalized cross validation. Simulations for two-dimensional examples verify the theoretical analysis and the effectiveness of the algorithm for increasing noise levels, and demonstrate that the relative reconstruction errors obtained using the truncated singular value decomposition are less than those obtained using the singular value decomposition.

---

Communicated by Lothar Reichel.

---

Rosemary Renaut acknowledges the support of NSF Grant DMS 1418377: “Novel Regularization for Joint Inversion of Nonlinear Problems”.

---

✉ Rosemary A. Renaut  
renaut@asu.edu

Extended author information available on the last page of the article

**Keywords** Inverse problems · Tikhonov regularization · Unbiased predictive risk estimation · Regularization parameter

**Mathematics Subject Classification** 65F10

## 1 Introduction

We consider the solution of  $A\mathbf{x} \approx \mathbf{b}$ , or  $A\mathbf{x} \approx \mathbf{b}_{\text{true}} + \boldsymbol{\eta} = \mathbf{b}$  for noise (measurement error)  $\boldsymbol{\eta}$ , where  $A \in \mathcal{R}^{m \times n}$  is ill-conditioned, and the system of equations arises from the discretization of an ill-posed inverse problem that may be over or under determined. The general Tikhonov regularized linear least squares problem

$$\mathbf{x}^* = \arg \min_{\mathbf{x}} \left\{ \|A\mathbf{x} - \mathbf{b}\|_{W_b}^2 + \|D(\mathbf{x} - \mathbf{x}_0)\|_{W_x}^2 \right\}, \quad (1.1)$$

is a well-accepted approach for finding a smooth solution  $\mathbf{x}$ . Here  $\mathbf{x}_0$  is given prior information, possibly the mean of  $\mathbf{x}$ ,  $W_b$  and  $W_x$  are weighting matrices on the data fidelity and regularization terms, resp., and  $D$  is an optional regularization operator. Often  $D$  is imposed as a spatial differential operator, controlling the size of the derivative(s) of  $\mathbf{x}$ , but then (1.1) can be brought into standard form in which  $D$  is replaced by  $I$  [5,21]. Further, (1.1) can be rewritten in terms of a new variable  $\mathbf{y} = \mathbf{x} - \mathbf{x}_0$ . The weighted norm is defined by  $\|\mathbf{x}\|_W^2 := \mathbf{x}^T W \mathbf{x}$  and we use the notation  $\mathbf{m} \sim \mathcal{N}(\mathbf{m}_0, C_m)$  for random vector  $\mathbf{m}$  normally distributed with expected value  $E(\mathbf{m}) = \mathbf{m}_0$  and covariance matrix  $C_m$ ;  $E(\cdot)$  is used to denote expected value. When  $\boldsymbol{\eta} \sim \mathcal{N}(0, C_b)$ , then  $W_b = C_b^{-1}$  whitens the noise, i.e.  $W_b^{1/2} \boldsymbol{\eta} \sim \mathcal{N}(0, I)$ . Matrix  $W_x = C_x^{-1}$  can serve similarly as a prior on the inverse covariance of the noise in  $D\mathbf{y}$ . Using  $W_x = \alpha^2 I$ , as will be assumed here, corresponds to assuming the posterior distribution  $D\mathbf{y} \sim \mathcal{N}(0, \alpha^{-2})$ , see e.g. [27]. Here we discuss the solution of (1.1) with  $\mathbf{x}_0 = 0$ ,  $D = I$ ,  $W_x = \alpha^2 I$ ,  $W_b = I$  and explicitly assume common variance,  $\sigma^2$ , in the noise,  $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2 I)$ .

While solutions of (1.1) have been extensively studied, e.g. [17,21,22,41] there is still much discussion concerning the selection of  $W_x$  even for the single parameter case,  $W_x = \alpha^2 I$ . Suggested techniques include, among others, using the Morozov discrepancy principle (MDP) which assumes that the solution should be found within some prescribed  $\chi^2$  noise estimate [29], balance of the terms in (1.1) using the L-curve [21], the quasi-optimality condition [3,15,16] and minimization of the generalized cross validation (GCV) function [11] or of the statistically motivated Unbiased Predictive Risk Estimator (UPRE) [33,41]. Of these the MDP, GCV and UPRE approaches are all a posteriori estimators, the MDP on the  $\chi^2$  distribution of the predicted residual, the GCV through its derivation as a leave one out procedure to minimize the predictive error and the UPRE as an estimator of the minimum predictive risk of the solution. There is an extensive discussion of these methods in the standard literature e.g. [21,22,41] and many more are compared in [4]. We do not replicate that discussion here, rather we focus on the UPRE parameter choice method. The UPRE method has a firm theoretical foundation, is robust, and has been extensively applied in practical applications [1,18,25,27,35,37–40]. Our analysis extends the approach in [9] which

provided bounds on the regularization parameter for finding  $\alpha$  using the GCV; the analysis in [30] that examined convergence of the parameter with increasing resolution of the problem via the connection of the continuous and discrete singular value expansions for specific square integrable operators defining  $A$ ; and the discussion in [31] that demonstrated the relationship of the regularization parameter obtained when using the LSQR Krylov method for large scale problems. Moreover, our interest in the UPRE, instead of the MDP, arises because the UPRE depends only on the underlying knowledge of the noise distribution, whereas the MDP also introduces a secondary tolerance factor on the satisfaction of the  $\chi^2$  distribution, which is often needed to limit over smoothing of the solutions [2].

Throughout we use the Singular Value Decomposition (SVD)  $A = U \Sigma V^T$  [12], with columns  $\mathbf{u}_i$  and  $\mathbf{v}_i$  of orthonormal  $U$  and  $V$  respectively, and where the singular values  $\sigma_i$  of  $A$  are ordered on the principal diagonal of  $\Sigma$ , from largest to smallest. We assume that the matrix  $A$  has effective numerical rank  $r$ ;  $\sigma_r > 0$ , and  $\sigma_i, i > r$  is effectively zero as determined by the machine precision. In terms of the SVD components, the solution of (1.1) is given by

$$\mathbf{x}^* = \sum_{i=1}^r \frac{\sigma_i^2}{\sigma_i^2 + \alpha^2} \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i = \sum_{i=1}^r \gamma_i(\alpha) \frac{\mathbf{u}_i^T \mathbf{b}}{\sigma_i} \mathbf{v}_i, \quad \gamma_i(\alpha) = \frac{\sigma_i^2}{(\sigma_i^2 + \alpha^2)}. \quad (1.2)$$

The filter functions are  $\gamma_i(\alpha)$  and the given expansion applies, replacing  $r$  by  $k$ , when  $A$  is approximated by the TSVD,  $A_k = U_k \Sigma_k V_k^T$ . Throughout we use the subscript  $k$  to indicate variables associated with this rank  $k$  approximation, for example regularization parameter  $\alpha_k$  indicates the regularization parameter used for the  $k$ -term TSVD. Further, the use of the SVD for  $A$  provides useful insights on how the UPRE, and other methods, can be implemented when solving (1.1). Here we will show that the minimization of the underlying UPRE function is efficient and robust with respect to the  $k$ -term truncated singular value decomposition (TSVD). Moreover, there is a resurgence of interest in using a TSVD solution for the solution of ill-posed problems due to the increased feasibility of finding a good approximation of a dominant singular subspace even for large scale problems by using techniques from randomization, e.g. [7,8,13,26,28,32]. Thus the presented results will be more broadly relevant for efficient estimates of an approximate TSVD using these modern techniques applied for large scale problems, for which it is not feasible to find the full SVD expansion; necessarily  $k \ll r$ .

*Overview of main contributions.* An open source algorithm, Algorithm 1, for efficiently estimating optimal regularization parameters  $k_{\text{opt}}$  and  $\alpha_{k_{\text{opt}}}$ , defined to be the optimal number of terms to use from the TSVD, and the associated regularization parameter, resp., is presented. By `optimal` we mean that these parameters are optimal in the sense of minimizing the UPRE function. A MATLAB implementation of Algorithm 1 and a 2D test case using IR Tools [10] is available at [https://github.com/renautra/TSVD\\_UPRE\\_Parameter\\_Estimation](https://github.com/renautra/TSVD_UPRE_Parameter_Estimation). A Python 3.\* implementation using NumPy and SciPy is also available and relies on provision of the singular values and coefficients  $\mathbf{u}_i^T \mathbf{b}$ . In both cases an estimate for the noise variance in the data is required, as is standard for the UPRE method. The motivation for Algorithm 1 is based on the

theoretical results presented in Sect. 3. These results employ standard assumptions on the degree of ill-posedness of the underlying model and on the noise level in the data [23]. We briefly review how both the degree of ill-posedness and the noise level impact the choice of regularization parameter  $k$ , and demonstrate that the noise level is far more restrictive so that in general  $k \ll r$ . The convergence of  $\alpha_k$ , when found using both UPRE and GCV methods, is illustrated for examples from the Regularization toolbox [20]. The theory presented in Sect. 3 then leads to Theorems 3.1 and 3.2 which prove a lower bound for  $\alpha_k$  and that  $\alpha_k$  converges to  $\alpha_{k_{\text{opt}}}$ , under the assumption of a unique minimum of the UPRE function. Presented results for image deblurring verify the practicality of Algorithm 1 and demonstrate that the solutions obtained with  $k_{\text{opt}} < r$  yield smaller overall relative error than the solutions obtained without truncation of the SVD and  $\alpha_r$  found using the UPRE method.

The paper is organized as follows: In Sect. 2 we present background motivating results based on assumptions on the degree of ill-posedness of the problem in Sect. 2.1, a discussion of numerical rank in Sect. 2.2, how noise enters into the problem in Sect. 2.3 and the estimation of the regularization parameter in Sect. 2.4. The theoretical results providing our main contributions are presented in Sect. 3. A practical algorithm for estimating  $\alpha_{k_{\text{opt}}}$ , and hence also  $k_{\text{opt}}$ , is presented in Sect. 4 with simulations verifying the analysis and the algorithm for two dimensional cases. Conclusions and future extensions are provided in Sect. 5.

## 2 Motivating results

### 2.1 Degree of ill-posedness

As in [23, Definition 2.42], and subsequently adopted in [21], for the analysis we assume specific decay rates for the singular values dependent on whether the problem is mildly, moderately or severely ill-posed. Suppose that  $\zeta$  is an arbitrary constant, then the decay rates are given by

$$\sigma_i = \begin{cases} \zeta i^{-\tau} & \frac{1}{2} \leq \tau \leq 1 & \text{mild ill conditioning} \\ \zeta i^{-\tau} & \tau > 1 & \text{moderate ill conditioning,} \\ \zeta \tau^{-i} & \tau > 1 & \text{severe ill conditioning.} \end{cases} \quad (2.1)$$

Here  $\tau$  is a problem dependent parameter and it is assumed that the decay rates hold on average for sufficiently large  $i$ . Moreover, while defining  $\sigma_i$  in terms of index  $i$ , as is consistent with the literature, it will also be convenient to consider the definition in terms of the continuous variable  $i$ , so that  $\sigma_{i+\delta}$  is defined also for non-integer  $i + \delta$ . For ease, and without loss of generality, we pick the constant  $\zeta$  in (2.1) so that  $\sigma_1 = 1$  in all cases. Equivalently we use

$$\sigma_i = \begin{cases} i^{-\tau} & \frac{1}{2} \leq \tau \leq 1 & \text{mild ill conditioning,} \\ i^{-\tau} & \tau > 1 & \text{moderate ill conditioning,} \\ \tau^{1-i} & \tau > 1 & \text{severe ill conditioning,} \end{cases} \quad (2.2)$$

**Table 1** Number of significant singular values  $r$  for precision  $\varepsilon = 10^{-15}$  as a function of  $\tau$ . i.e.  $r$  is the numerical rank of the problem

|          | $\tau$ | 1.25  | 1.50  | 1.75 | 2.00 | 2.50 | 3.00 | 4.00 | 5.00 | 6.00 |
|----------|--------|-------|-------|------|------|------|------|------|------|------|
| Moderate | $r$    | 1e+12 | 1e+10 | 4e+8 | 3e+7 | 1e+6 | 1e+5 | 5623 | 1000 | 316  |
| Severe   | $r$    | 155   | 86    | 62   | 50   | 38   | 32   | 25   | 22   | 20   |

and note the recurrences

$$\sigma_{\ell+1} = \sigma_{\ell} \begin{cases} \left(\frac{\ell}{\ell+1}\right)^{\tau} & \text{mild or moderate ill conditioning,} \\ \tau^{-1} & \text{severe ill conditioning.} \end{cases}$$

## 2.2 Numerical rank

The precision of the calculations, as determined by the machine epsilon  $\varepsilon$ , is relevant in terms of the number of singular values that are significant in the calculation. This is dependent on the decay rate parameters of the singular values. We define the effective rank by  $r = \arg\max\{i : \sigma_i > \varepsilon\sigma_1\}$ .

**Proposition 2.1** *Assuming the normalization of the singular values as given by (2.2), the effective numerical rank  $r$  is bounded by*

$$r < \begin{cases} \varepsilon^{-1/\tau} & \text{mild/moderate decay,} \\ 1 - \frac{\log \varepsilon}{\log \tau} & \text{severe decay,} \end{cases} \quad (2.3)$$

where  $\varepsilon$  is the machine epsilon.

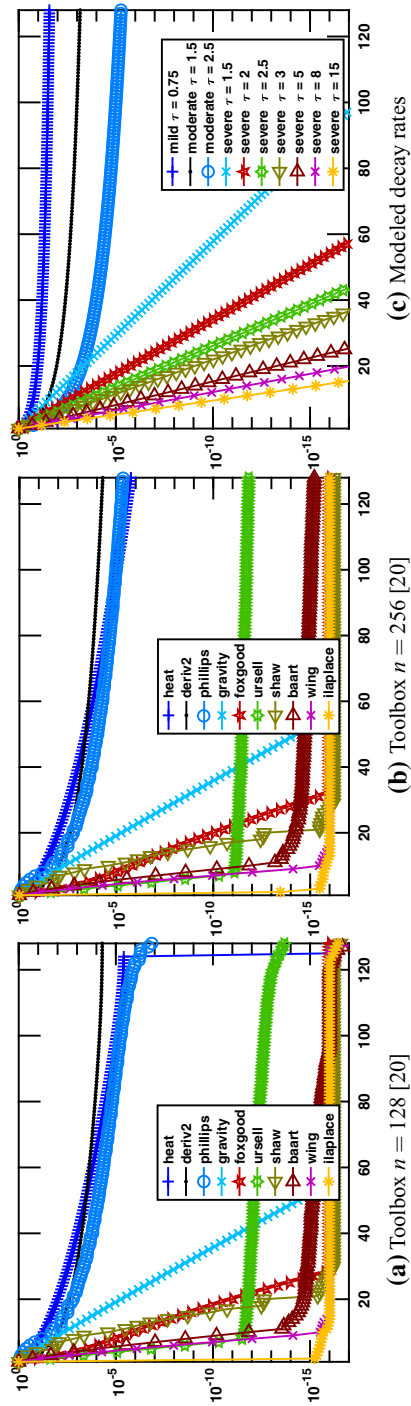
**Proof** Using (2.2) and normalization  $\sigma_1 = 1$ , it is immediate that we obtain (2.3) from

$$\begin{aligned} \text{mild/moderate: } r^{-\tau} > \varepsilon & \text{ implies } r < \varepsilon^{-1/\tau} \\ \text{severe : } \tau^{1-r} > \varepsilon & \text{ implies } r < 1 - \frac{\log \varepsilon}{\log \tau}. \end{aligned}$$

□

Estimates for numerical rank dependent on the decay rates, are given in Table 1 for moderate and severe decay. It is immediate that  $r$  is very small for cases of severe decay. Hence, for any problem exhibiting this severe decay and assuming that the discretization is sufficiently fine such that  $n \geq r$ , Table 1 suggests the maximum number of terms that one would use for the TSVD. Note that apart from the condition  $n \geq r$  the results in Table 1 are effectively independent of the discretization, thus the number of terms that can be used practically is largely independent of the discretization of the problem once  $n \geq r$ . Equivalently, with estimates of  $\tau$  and  $\varepsilon$  one may use (2.3) to determine first a minimum  $n$  and second the maximum number of terms for the TSVD, the maximum effective numerical rank of the problem.

These results are further illustrated in Fig. 1a in which we plot the singular values of test problems from the Regularization toolbox [20], with the normalization  $\sigma_1 = 1$  in



**Fig. 1** The singular value distribution for  $n = 128$  for noted examples from [20], normalized to  $\sigma_1 = 1$  in **a**. To show the independence of  $n$ , decay for  $n = 256$  for the same examples from [20] is also shown in **b**. For each of the toolbox examples it is possible to compare with a simulated case for a specific decay rate by illustrating (2.2) for severe, moderate or mild decay choices of  $\tau$ , as appropriate, given in **c**

each case. The plots show that these standard one dimensional test cases are primarily severely ill-posed, and thus, according to Table 1, guaranteed to have numerically very few accurate terms in the TSVD used for the solution (1.2). To show the relative independence of  $n$  we show in Fig. 1b the singular value distributions for the same cases and on the same scales as in Fig. 1a but using  $n = 256$ . This verifies that there is little to be gained by the use of problems with severe decay, as presented in [30], to validate convergence of techniques with increasing problem size. The dominant features are always represented by very few terms of the TSVD for cases with severe decay rates of the singular values.

As a comparison we also show in Fig. 1c the singular values generated using (2.3) with a selection of decay rates, as indicated in the legend. These show the dependence on  $\tau$  for mild, moderate and severe decays. Taken together the examples in Fig. 1 show that the results are not just an artificial artifact of the seemingly strong assumption in (2.2) that  $\zeta$  is fixed.

### 2.3 The discrete Picard condition and noise contamination

We now turn to the consideration of the noise in the coefficients  $s_i = \mathbf{u}_i^T \mathbf{b}$  and the impact of this noise on the potential resolution in the solution, as also discussed in [21, §4.8.1]. First, we assume that the singular values satisfy a decay rate condition (2.2). We also assume that the absolute values of the exact coefficients decay at least faster than the singular values

$$(s_i^2)_{\text{true}} \leq \sigma_i^{2(1+\nu)} \quad \text{for } 0 < \nu < 1. \quad (2.4)$$

Then the discrete Picard condition is satisfied [19,21, Theorem 4.5.1].

When the noise in the data has common variance  $\sigma^2$  we assume that there exists  $\ell$  such that  $E(s_i^2) = \sigma^2$ , for all  $i > \ell$ . Equivalently, we say that the coefficients are noise dominated for  $i > \ell$ . If this does not occur, then either the noise is insignificant,  $\sigma^2 < \sigma_i^2$  for all  $i$ , or totally dominates the solution  $\sigma^2 > 1$ , and these two cases are not of interest. Thus we can explicitly assume that there exists  $\ell$  such that  $\sigma_{\ell+1} < \sigma < \sigma_\ell$ , and more precisely that

$$\sigma_{\ell+1}^2 < \sigma_{\ell+1}^{1+\nu} < \sigma < \sigma_\ell^{1+\nu} < \sigma_\ell,$$

where we use definition (2.2) as a continuous function of  $i$  for a non integer index  $\ell + \delta$ .

**Proposition 2.2** *Let  $\sigma = \sigma_{\ell+\delta}^{1+\nu}$  for  $0 \leq \delta < 1$  and  $0 < \nu < 1$ , then  $E(s_i^2) = \sigma^2$  for  $i > \ell$  where*

$$\ell \approx \begin{cases} \sigma^{-1/(\tau(1+\nu))} - \delta & \text{mild/moderate decay,} \\ (1 - \delta) - \frac{\log \sigma}{(\nu+1) \log \tau} & \text{severe decay.} \end{cases} \quad (2.5)$$

**Proof** As in the proof of Proposition 2.1 we solve for  $\ell$  dependent on the decay rate with respect to the upper bound in (2.5). This gives

**Table 2** For different noise levels  $\sigma$  the size of  $\ell$  for given  $\tau$  and with  $\delta = \nu = 0.5$ . Entries calculated with rounding using (2.5)

| $\tau$   | 1.25           | 1.50 | 1.75 | 2.00 | 2.50 | 3.00 | 4.00 | 5.00 | 6.00 |
|----------|----------------|------|------|------|------|------|------|------|------|
| $\sigma$ | Moderate decay |      |      |      |      |      |      |      |      |
| 1e-1     | 3              | 2    | 2    | 2    | 1    | 1    | 1    | 1    | 1    |
| 1e-2     | 11             | 7    | 5    | 4    | 3    | 2    | 2    | 1    | 1    |
| 1e-4     | 135            | 59   | 33   | 21   | 11   | 7    | 4    | 3    | 2    |
| 1e-8     | 18478          | 3593 | 1115 | 464  | 135  | 59   | 21   | 11   | 7    |
| $\sigma$ | Severe decay   |      |      |      |      |      |      |      |      |
| 1e-1     | 7              | 4    | 3    | 3    | 2    | 2    | 2    | 1    | 1    |
| 1e-2     | 14             | 8    | 6    | 5    | 4    | 3    | 3    | 2    | 2    |
| 1e-4     | 28             | 16   | 11   | 9    | 7    | 6    | 5    | 4    | 4    |
| 1e-8     | 56             | 31   | 22   | 18   | 14   | 12   | 9    | 8    | 7    |

$$\begin{aligned} \text{mild/moderate: } (\ell + \delta)^{-\tau(1+\nu)} = \sigma &\text{ implies } \ell = \sigma^{-1/(\tau(1+\nu))} - \delta, \\ \text{severe: } \tau^{(1-(\ell+\delta))(1+\nu)} = \sigma &\text{ implies } \ell = (1 - \delta) - \frac{\log \sigma}{(\nu+1) \log \tau}. \end{aligned}$$

□

Estimates using  $\delta = \nu = 0.5$  are indicated in Table 2 showing that the number of terms is relatively small even for moderate decay of the singular values for acceptable noise estimates  $\sigma$ . Contrasting with Table 1 we see that the number of coefficients that can be distinguished from the noise is generally less than the numerical rank of the problem for relevant noise levels and machine precision. This limits the number of the terms of the TSVD to use. In particular, suppose that  $\alpha$  has to be found to filter the dominant noise terms with index  $i \geq \ell$ , then coefficients with  $i \gg \ell$  will be further damped because the filter factors given in (1.2) decrease as a function of  $i$ . These terms then become insignificant in terms of the expansion for the solution.

## 2.4 Regularization parameter estimation

We deduce from Tables 1 and 2 that the number of terms of the TSVD used for the solution of the regularized problem may strongly influence the choice for  $\alpha$ . Specifically the number of terms  $k$  of the TSVD to use should be less than the numerical rank,  $k < r$ , and is dependent on the noise level in the data. We are interested in investigating the choice of  $\alpha$  when obtained using the UPRE, but for comparison we also give the GCV function needed for the simulations, and note again that bounds on  $\alpha$  dependent on  $k$  have already been provided in [9]. The GCV and UPRE methods are derived without the use of the SVD [11,41], resp., but it is convenient for the analysis to express both methods in terms of the SVD. Ignoring constant terms in the UPRE that do not impact the location of the minimum, introducing  $\phi_i(\alpha) = 1 - \gamma_i(\alpha) = \alpha^2/(\sigma_i^2 + \alpha^2)$ , and noting  $\gamma_i(\alpha) = 0$ , for  $i > k$ , these are given by



$$U_k(\alpha) = \sum_{i=1}^k (1 - \gamma_i(\alpha))^2 (\mathbf{u}_i^T \mathbf{b})^2 + 2\sigma^2 \sum_{i=1}^k \gamma_i(\alpha) = \sum_{i=1}^k \phi_i^2(\alpha) s_i^2 + 2\sigma^2 \sum_{i=1}^k \gamma_i(\alpha), \quad (2.6)$$

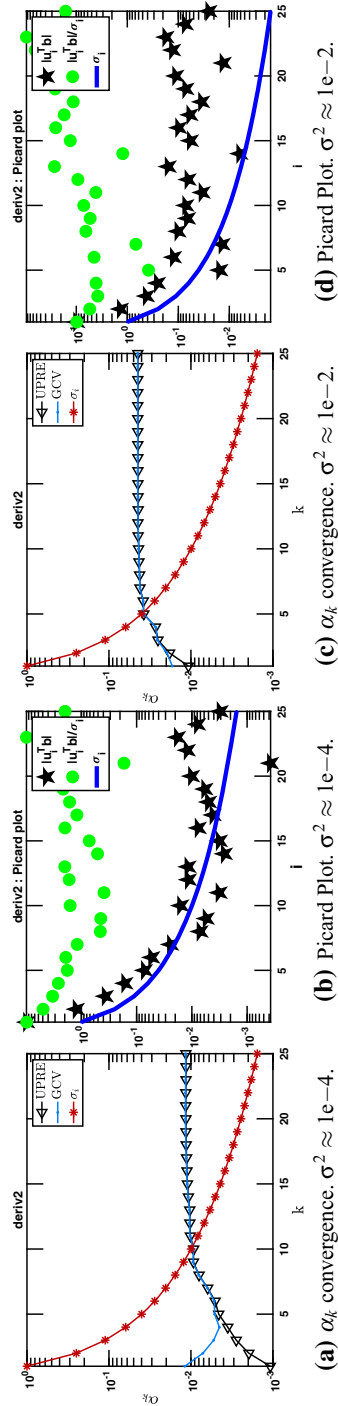
$$G_k(\alpha) = \frac{\sum_{i=1}^m (1 - \gamma_i(\alpha))^2 (\mathbf{u}_i^T \mathbf{b})^2}{\left(\sum_{i=1}^m (1 - \gamma_i(\alpha))\right)^2} = \frac{\sum_{i=1}^k \phi_i^2(\alpha) s_i^2 + \sum_{i=k+1}^m s_i^2}{\left((m - k) + \sum_{i=1}^k \phi_i(\alpha)\right)^2}. \quad (2.7)$$

Here the subscript  $k \leq r$  indicates that these are the expressions obtained using the TSVD, see e.g. [31, Appendix B] for derivations of the UPRE and GCV functions for arbitrary pairs  $(m, n)$ . Replacing  $k$  by  $r$  gives the standard functions for the full SVD. Further, we do not need all terms of the SVD to calculate the numerator in (2.7). Using  $\|\mathbf{b}\|_2^2 = \|U^T \mathbf{b}\|_2^2$  we can use

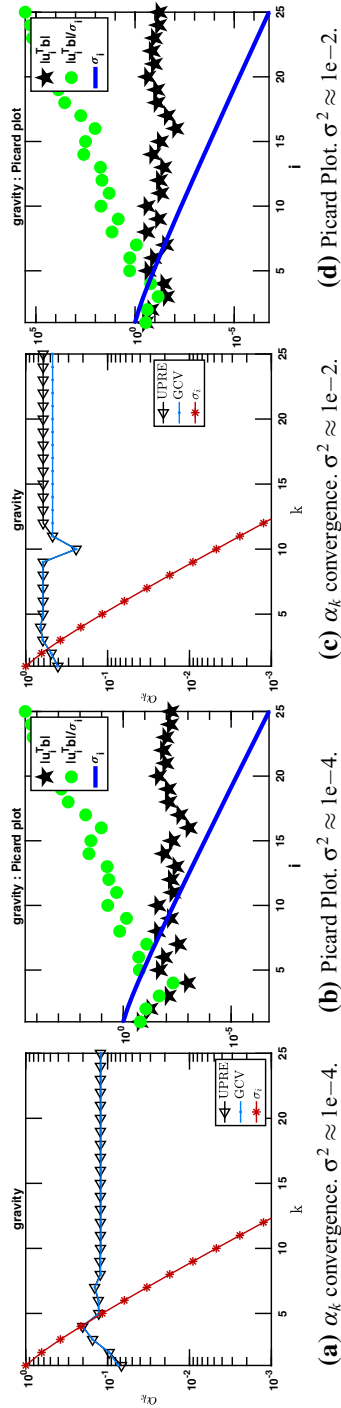
$$\sum_{i=k+1}^m (\mathbf{u}_i^T \mathbf{b})^2 = \|\mathbf{b}\|_2^2 - \sum_{i=1}^k (\mathbf{u}_i^T \mathbf{b})^2 = \|\mathbf{b}\|_2^2 - \sum_{i=1}^k s_i^2.$$

To illustrate how  $\alpha_k$  varies when found using these functions we illustrate an example of a problem that is only moderately ill-posed ( $\tau \approx 1.5$ ), showing the results of calculating the UPRE and GCV functions for data with noise variance  $\sigma^2 \approx 1\text{e-}4$  and  $\sigma^2 \approx 1\text{e-}2$  for the problem `deriv2`. The data and solution  $\mathbf{x}_{\text{true}}$  are initially normalized so that  $\|\mathbf{b}_{\text{true}}\|_2 = 1$ . Consistent with the decay rate assumptions the singular values are normalized by  $\sigma_1$ . This requires additional normalization of  $\mathbf{b}_{\text{true}}$  by  $\sigma_1$ , so that eventually  $\|\mathbf{b}_{\text{true}}\|_2 = \sigma_1^{-1}$ . Then noise contaminated data are generated as  $\mathbf{b} = \mathbf{b}_{\text{true}} + \boldsymbol{\eta}$  for  $\boldsymbol{\eta} \sim \mathcal{N}(0, \sigma^2 I)$ , for noise level  $\sigma$ . In these examples, the optimal value  $\alpha_k$  is obtained by first evaluating  $f(\alpha)$ ,  $f(\alpha) = U_k(\alpha)$  or  $f(\alpha) = G_k(\alpha)$  as specified in (2.6) or (2.7), resp., at  $\sigma_i$ ,  $1 \leq i \leq k$ . This provides  $\alpha_{\text{est}} = \arg \min_{1 \leq i \leq k} f(\sigma_i)$ . This estimate of the minimum is used as the initial value for minimizing  $f(\alpha)$  using Matlab `fminbnd` within the interval  $[.01\alpha_{\text{est}}, 100\alpha_{\text{est}}]$ . While this choice of lower and upper bounds on  $\alpha_k$  is somewhat arbitrary, it is similar to the approach used in [20] for minimizing the UPRE and GCV functions, and is chosen to assure that values for  $\alpha_k$  outside the interval  $[\sigma_k, \sigma_1]$  are possible when either  $\alpha_{\text{est}} = \sigma_k$  or  $\sigma_1$ . We pick this specific example in Fig. 2 to highlight the discussion as applied to a problem which is not severely ill-posed. We also give the same information in Fig. 3 for the severely ill-posed problem `gravity` ( $\tau \approx 1.5$ ), see Fig. 1a. To gain further insight the Picard plot, plots of  $\sigma_i$ ,  $|\mathbf{u}_i^T \mathbf{b}|$  and the ratio  $|\mathbf{u}_i^T \mathbf{b}|/\sigma_i$ , is given in each case in Fig. 2b, d for `deriv2` and in Fig. 3b, d for `gravity`. The solutions are contaminated by noise very quickly for small  $k$ , corresponding to fast convergence of  $\{\alpha_k\}$  with  $k$ .

Obtaining one-dimensional results, as shown in Figs. 2, 3, is trivial but motivates the theoretical study of convergence in Sect. 3, and then the application of that theory to standard two-dimensional problems in Sect. 4.



**Fig. 2** Example deriv2 from [20] showing the convergence of  $\{\alpha_k\}$  for UPRE and GCV functions for TSVD sizes of 1:25 as compared to the decay of the singular values, for the original problem of size 128 and the associated Picard plot for the data. **a-d** The noise variances are  $\sigma^2 \approx 1e-4$  and  $\sigma^2 \approx 1e-2$ , respectively. The converged relative errors in each case are .273 and .381 for the two noise variances  $\sigma^2 \approx 1e-4$  and  $\sigma^2 \approx 1e-2$ , respectively



**Fig. 3** Example gravity from [20] showing the convergence of  $\{\alpha_k\}$  for UPRE and GCV functions for TSVD sizes of 1:25 as compared to the decay of the singular values, for the original problem of size 128 and the associated Picard plot for the data. **a–d** The noise variances are  $\sigma^2 \approx 1e-4$  and  $\sigma^2 \approx 1e-2$ , respectively. The converged relative errors for noise variance  $\sigma^2 \approx 1e-4$  are .644 and .655 for UPRE and GCV respectively. For noise variance  $\sigma^2 \approx 1e-2$ , these errors are .787 and 1.143, respectively

### 3 Theoretical results

We aim to find effective practical bounds on the regularization parameter  $\alpha$  when found using the UPRE function. Observe first that we would not expect the regularization parameter to be larger than  $\sigma_1$ , otherwise all filter factors are less than  $1/2$ . Indeed imposing  $\alpha = \sigma_1$  would lead to over smoothed solutions, and all of the dominant singular value components (the components without noise contamination) would be represented in the solution with filtering e.g [22, Sections 4.4, 4.7]. In particular, the norm of the covariance matrix for the truncated filtered Tikhonov solution, the a posteriori covariance of the solution, is approximately bounded by  $\sigma^2/(4\alpha^2)$  which suggests smooth solutions for large  $\alpha$ . In contrast, the approximate bound for the a posteriori covariance when using the TSVD with  $k$  terms without filtering is given by  $\sigma^2/\sigma_k^2$  [22, Sections 4.4.2, 4.4]. Thus the filtered TSVD solution will be smoother than the TSVD solution when  $\alpha > \sigma_k$ : increasing  $\alpha$  reduces the covariance but provides more smoothing. Practically it is reasonable to impose the upper bound  $\alpha_{\max} \leq \sigma_1 = 1$  for  $\alpha$ . To limit the noise that can enter the solution it is also desirable to find the lower bound  $\alpha_{\min}$ . Solutions obtained for  $\alpha \in [\alpha_{\min}, \alpha_{\max}]$ , dependent on the spectrum of  $A$ , should be sufficiently filtered but retain relatively unfiltered dominant components of the solution. We proceed to determine  $\alpha_{\min}$  and to give a convergence analysis for  $\alpha_k$  as the number of terms in the TSVD is increased.

#### 3.1 Convergence of $\{\alpha_k\}$ calculated using UPRE

Denote the UPRE function (2.6) for the rank  $r$  problem by  $U(\alpha) = U_r(\alpha)$  and the optimal  $\alpha$  for the filtered TSVD solution with  $k$  components on the given interval as

$$\alpha_k = \operatorname{argmin}_{\alpha \in [\alpha_{\min}, \alpha_{\max}]} U_k(\alpha). \quad (3.1)$$

Ideally it would be helpful to find an interval  $[\alpha_{\min}, \alpha_{\max}]$  in which  $U_k(\alpha)$  is strongly convex, but we have not been able to show this in general. Instead, in the following we show that a useful estimate of  $\alpha_{\min}$  can be found.

For ease of notation within proofs we use  $\phi_i$  and  $\gamma_i$  to indicate  $\phi_i(\alpha)$  and  $\gamma_i(\alpha)$ , respectively, and denote differentiation of a function  $f(\alpha)$  with respect to  $\alpha$  as  $f'$ .

**Proposition 3.1** *The following equalities are required for the future discussion.*

$$\frac{\partial U_k}{\partial \alpha} = \frac{4}{\alpha} \left( \sum_{i=1}^k s_i^2 \phi_i^2(\alpha) \gamma_i(\alpha) - \sigma^2 \sum_{i=1}^k \phi_i(\alpha) \gamma_i(\alpha) \right), \quad (3.2)$$

$$\begin{aligned} \frac{\partial^2 U_k}{\partial \alpha^2} = & -\frac{1}{\alpha} \frac{\partial U_k}{\partial \alpha} + \frac{8}{\alpha^2} \left( \sum_{i=1}^k s_i^2 \phi_i^2(\alpha) \gamma_i(\alpha) (2\gamma_i(\alpha) - \phi_i(\alpha)) \right. \\ & \left. - \sigma^2 \sum_{i=1}^k \phi_i(\alpha) \gamma_i(\alpha) (\gamma_i(\alpha) - \phi_i(\alpha)) \right). \end{aligned} \quad (3.3)$$

**Proof** We use

$$\gamma'_i = -\frac{2\alpha\sigma_i^2}{(\sigma_i^2 + \alpha^2)^2} = -\frac{2}{\alpha}\phi_i\gamma_i = -\phi'_i < 0.$$

Directly differentiating  $U_k(\alpha)$  gives (3.2)

$$U'_k = \sum_{i=1}^k s_i^2 2\phi_i \phi'_i + 2\sigma^2 \sum_{i=1}^k \gamma'_i = \frac{4}{\alpha} \left( \sum_{i=1}^k s_i^2 \phi_i^2 \gamma_i - \sigma^2 \sum_{i=1}^k \gamma_i \phi_i \right).$$

Likewise for the second derivative

$$U''_k = -\frac{1}{\alpha} U'_k + \frac{4}{\alpha} \left( \sum_{i=1}^k s_i^2 (2\phi_i \phi'_i \gamma_i + \phi_i^2 \gamma'_i) - \sigma^2 \sum_{i=1}^k (\phi'_i \gamma_i + \phi_i \gamma'_i) \right), \quad (3.4)$$

giving (3.3) after substitution for the derivatives.  $\square$

**Proposition 3.2** Suppose that  $0 < \bar{\alpha} < \sigma_k/\sqrt{2}$  is a stationary point for  $U_k(\alpha)$ , for any  $1 \leq k \leq r$ . Then  $\bar{\alpha}$  is a unique minimum for  $U_k(\alpha)$  on the interval  $0 < \bar{\alpha} < \sigma_k/\sqrt{2}$ .

**Proof** Removing the first term from (3.4), identically zero at  $\alpha = \bar{\alpha}$  by assumption that  $\bar{\alpha}$  is a stationary point, gives

$$\begin{aligned} \frac{\partial^2 U_k}{\partial \alpha^2}(\bar{\alpha}) &= \frac{8}{\bar{\alpha}^2} \left( \sum_{i=1}^k s_i^2 \phi_i^2(\bar{\alpha}) \gamma_i(\bar{\alpha}) (2\gamma_i(\bar{\alpha}) - \phi_i(\bar{\alpha})) - \sigma^2 \sum_{i=1}^k \phi_i(\bar{\alpha}) \gamma_i(\bar{\alpha}) (\gamma_i(\bar{\alpha}) - \phi_i(\bar{\alpha})) \right) \\ &= \frac{8}{\bar{\alpha}^2} \left( \sum_{i=1}^k s_i^2 \phi_i^2(\bar{\alpha}) \gamma_i(\bar{\alpha}) (2 - 3\phi_i(\bar{\alpha})) - \sigma^2 \sum_{i=1}^k \phi_i(\bar{\alpha}) \gamma_i(\bar{\alpha}) (1 - 2\phi_i(\bar{\alpha})) \right). \end{aligned}$$

Now we substitute for  $\sum_{i=1}^k s_i^2 \phi_i^2(\bar{\alpha}) \gamma_i(\bar{\alpha}) = \sigma^2 \sum_{i=1}^k \gamma_i(\bar{\alpha}) \phi_i(\bar{\alpha})$  using (3.2) at  $\bar{\alpha}$  and note all terms are positive for  $1 - 3\phi_i(\bar{\alpha}) > 0$ ,  $i = 1 : k$ . But  $\phi_i$  is increasing with  $i$  due to the ordering of the  $\sigma_i$ . Thus  $1 - 3\phi_i(\bar{\alpha}) \geq 1 - 3\phi_k(\bar{\alpha}) > 0$  for  $\bar{\alpha} < \sigma_k/\sqrt{2}$  and  $U''_k(\bar{\alpha}) > 0$ . This result is true for any stationary point  $\bar{\alpha}$  on the interval. Hence  $U_k(\bar{\alpha})$  is a minimum for  $U_k(\alpha)$  and it is only possible to have a maximum at  $\alpha = 0$ , the end point of the given interval, but the end point is explicitly excluded from consideration. There are therefore no other stationary points within the interval and the minimum is unique.

**Remark 3.1** Although a minimum must exist in  $[0, \sigma_k/\sqrt{2}]$  because  $U_k(\alpha)$  is a continuous function on a compact set, this result does not show that a minimum exists in  $(0, \sigma_k/\sqrt{2})$ .

The next steps in the analysis rely on the following Assumptions 1–2 about the model and the data.

**Assumption 1** (Decay Rate [21,23]) The measured coefficients decay according to  $s_i^2 = \sigma_i^{2(1+\nu)} > \sigma^2$  for  $0 < \nu < 1$ ,  $1 \leq i \leq \ell$ , i.e. the dominant measured coefficients follow the decay rate of the exact coefficients.

**Assumption 2** (*Noise in Coefficients*) There exists  $\ell$  such that  $E(s_i^2) = \sigma^2$  for all  $i > \ell$ , i.e. that the coefficients  $s_i$  are noise dominated for  $i > \ell$ . Moreover, when  $i \leq \ell$  we assume that  $E(s_i^2) \approx s_i^2$ , so that the larger coefficients are effectively deterministic.

These assumptions have also been used in [21] for understanding how decay rates impact the convergence of iterative methods. We also recall that we use the non-restrictive normalization  $\sigma_1 = 1$  and use the notation  $E(a)$  for the expectation of scalar deterministic  $a$ .

For the remaining results we distinguish between the terms in the UPRE function that are, and are not, contaminated by noise.

**Proposition 3.3** *Suppose Assumption 2 holds, then for  $r > k + 1 > \ell$  there is an upper bound on  $E(U'_k)$  independent of  $\alpha$ :*

$$E\left(\frac{\partial U_r}{\partial \alpha}\right) < \cdots < E\left(\frac{\partial U_{k+1}}{\partial \alpha}\right) < E\left(\frac{\partial U_k}{\partial \alpha}\right) < \frac{\partial U_\ell}{\partial \alpha} \quad \forall \alpha. \quad (3.5)$$

The lower bound for  $E(U''_k)$  holds for a fixed lower bound on  $\alpha$

$$E\left(\frac{\partial^2 U_r}{\partial \alpha^2}\right) > \cdots > E\left(\frac{\partial^2 U_{k+1}}{\partial \alpha^2}\right) > E\left(\frac{\partial^2 U_k}{\partial \alpha^2}\right) > \frac{\partial^2 U_\ell}{\partial \alpha^2} \text{ if } \alpha > \frac{\sigma_{\ell+1}}{\sqrt{5}}, \quad (3.6)$$

whereas the upper bound depends also on an upper bound on  $\alpha$  that decreases with increasing  $k$

$$E\left(\frac{\partial^2 U_{k+1}}{\partial \alpha^2}\right) < E\left(\frac{\partial^2 U_k}{\partial \alpha^2}\right) \text{ if } \alpha < \frac{\sigma_{k+1}}{\sqrt{5}}. \quad (3.7)$$

**Proof** We note that the expectation operator is linear and when  $a$  is not a random variable  $E(a) = a$ . Applying these properties first to (3.2) yields

$$\begin{aligned} E(U'_k) &= E\left(U'_\ell + \frac{4}{\alpha} \sum_{i=\ell+1}^k \phi_i \gamma_i (s_i^2 \phi_i - \sigma^2)\right) \\ &\approx U'_\ell + \frac{4\sigma^2}{\alpha} \sum_{i=\ell+1}^k \phi_i \gamma_i (\phi_i - 1) < U'_\ell, \end{aligned}$$

where from line one to two we use linearity, and, by Assumption 2,  $E(U'_\ell) = U'_\ell$  and  $E(s_i^2) = \sigma^2$  for  $i > \ell$ . In particular, in expectation each term for  $i > \ell$  is negative and recursively both inequalities in (3.5) apply. Applying the expectation operator now to (3.3) gives

$$\begin{aligned}
E(U_k'') &\approx \frac{\partial^2 U_\ell''}{\partial \alpha^2} + \frac{4\sigma^2}{\alpha^2} \left( \sum_{i=\ell+1}^k \phi_i \gamma_i (1 - \phi_i) + 2 \left( \phi_i^2 \gamma_i (2\gamma_i - \phi_i) - \phi_i \gamma_i (\gamma_i - \phi_i) \right) \right) \\
&= U_\ell'' + \frac{4\sigma^2}{\alpha^2} \left( \sum_{i=\ell+1}^k \phi_i \gamma_i (1 - \phi_i + 2(\phi_i(2\gamma_i - \phi_i) - (\gamma_i - \phi_i))) \right) \\
&= U_\ell'' + \frac{4\sigma^2}{\alpha^2} \left( \sum_{i=\ell+1}^k \phi_i \gamma_i (1 - \phi_i + 2(\phi_i(2 - 3\phi_i) - (1 - 2\phi_i))) \right) \\
&= U_\ell'' + \frac{4\sigma^2}{\alpha^2} \left( \sum_{i=\ell+1}^k \phi_i \gamma_i (-6\phi_i^2 + 7\phi_i - 1) \right).
\end{aligned}$$

The sign of the second term depends on the sign of  $-6\phi_i^2 + 7\phi_i - 1$  which is increasing from  $-1$  as a function of  $\phi \leq 1$ . Hence

$$-6\phi_i^2 + 7\phi_i - 1 \begin{cases} \geq -6\phi_{\ell+1}^2 + 7\phi_{\ell+1} - 1 = \frac{\sigma_{\ell+1}^2(5\alpha^2 - \sigma_{\ell+1}^2)}{(\alpha^2 + \sigma_{\ell+1}^2)^2} > 0 & \text{if } \alpha > \frac{\sigma_{\ell+1}}{\sqrt{5}} \\ \leq -6\phi_k^2 + 7\phi_k - 1 = \frac{\sigma_k^2(5\alpha^2 - \sigma_k^2)}{(\alpha^2 + \sigma_k^2)^2} < 0 & \text{if } \alpha < \frac{\sigma_k}{\sqrt{5}}. \end{cases}$$

Again, in expectation, terms for  $i > \ell$  are all positive when  $\alpha \geq \sigma_{\ell+1}/\sqrt{5}$  and the nested inequalities in (3.6) apply. The requirement that the  $i^{\text{th}}$  term is necessarily positive becomes more severe as  $i$  increases, yielding the additional inequality with conditions on  $\alpha$  given in (3.7).

**Corollary 3.1** Suppose Assumption 2 holds, and that for  $\alpha_\ell > \sigma_{\ell+1}/\sqrt{5}$ ,  $U_\ell(\alpha_\ell)$  is a minimum for  $U_\ell(\alpha)$ . Then for  $\ell < k \leq r$ ,  $U_k(\alpha)$  is convex and decreasing at  $\alpha_\ell$ ,

$$E\left(\frac{\partial U_k(\alpha_\ell)}{\partial \alpha}\right) < 0 \quad \text{and} \quad E\left(\frac{\partial^2 U_k(\alpha_\ell)}{\partial \alpha^2}\right) > 0.$$

**Proof** If  $U_\ell(\alpha_\ell)$  is a minimum, then  $U_\ell'(\alpha_\ell) = 0$  and  $U_\ell''(\alpha_\ell) > 0$  and the inequalities follow immediately from (3.5) and (3.6).

**Corollary 3.2** Suppose Assumption 2 holds. If a stationary point  $\alpha_r < \sigma_r/\sqrt{5}$  exists there are no stationary points of  $U_k(\alpha)$  for  $\alpha \in (\sigma_r/\sqrt{5}, \sigma_k/\sqrt{2})$ .

**Proof** Suppose that  $\alpha_r \in [0, \sigma_r/\sqrt{5})$ . The existence of  $\alpha_r$  in this interval does not contradict Proposition 3.2 since  $\sigma_r/\sqrt{5} < \sigma_r/\sqrt{2}$ . By assumption,  $U_r'(\alpha_r) = 0$  and  $U_r''(\alpha_r) > 0$ . Thus by (3.5) and (3.7)  $U_k'(\alpha_r) > 0$  and  $U_k''(\alpha_r) > 0$ , and  $U_k(\alpha)$ ,  $\ell \leq k \leq r-1$  is convex and increasing at  $\alpha_r$ . Therefore, by continuity,  $U_k(\alpha)$  cannot reach a minimum for  $\alpha_r < \alpha_k < \sigma_k/\sqrt{2}$  without first passing through a stationary point which is a maximum. But by Proposition 3.2 there is no maximum of  $U_k(\alpha)$  to the left of  $\sigma_k/\sqrt{2}$  and thus there is also no minimum for  $\alpha_r < \alpha < \sigma_k/\sqrt{2}$ . In particular  $U_k(\alpha)$  has no stationary point for  $\sigma_r/\sqrt{5} \leq \alpha \leq \sigma_k/\sqrt{2}$ .

**Remark 3.2** We have shown through Corollary 3.2 that if  $U_r(\alpha_r)$  is a minimum for  $U_r(\alpha)$  and  $\alpha_r < \sigma_r/\sqrt{5}$  then  $U_k(\alpha_k)$  can only be a minimum for  $U_k(\alpha)$  if either  $\alpha_k \leq \alpha_r \leq \sigma_r/\sqrt{5}$  or  $\alpha_k > \sigma_k/\sqrt{2}$ , i.e. we may require  $\alpha_k > \sigma_k/\sqrt{2}$  under the assumption that we seek  $\alpha_r > \sigma_r$ . This applies for all  $k$  with  $1 \leq \ell \leq k \leq r-1$ .

Although this result does provide a refined lower bound for  $\alpha_k$ , it is dependent on  $k$  and decreasing with  $k$ , which is not helpful when  $k$  gets large, as needed for finding  $\alpha_r$ , i.e. this bound would suggest that  $\alpha_r$  needs to be found using the pessimistic lower bound  $\sigma_r/\sqrt{2}$ . We investigate now whether these lower bounds on  $\alpha$  are indeed realistic by looking for bounds on the UPRE functions  $U_k(\alpha)$ .

**Proposition 3.4** Suppose Assumptions 1 and 2 hold, then lower and upper bounds on  $U_k(\alpha)$  and its derivatives are given by  $\mathcal{L}_k(\alpha)$  and  $\mathcal{U}_k(\alpha)$  and their derivatives, respectively, where

$$0 < \mathcal{L}_k(\alpha) = G(\alpha) + F_k(\alpha) < E(U_k(\alpha)) < H(\alpha) + F_k(\alpha) = \mathcal{U}_k(\alpha) \quad (3.8)$$

$$\mathcal{L}'_k(\alpha) = G'(\alpha) + F'_k(\alpha) < E(U'_k(\alpha)) < H'(\alpha) + F'_k(\alpha) = \mathcal{U}'_k(\alpha), \quad (3.9)$$

$$\begin{aligned} \mathcal{L}''_k(\alpha) &= G''(\alpha) + F''_k(\alpha) < E(U''_k(\alpha)) < H''(\alpha) + F''_k(\alpha) = \mathcal{U}''_k(\alpha), \text{ for } \alpha \leq \sigma_\ell \text{ but} \\ \mathcal{U}''_k(\alpha) &= H''(\alpha) + F''_k(\alpha) < E(U''_k(\alpha)) < G''(\alpha) + F''_k(\alpha) = \mathcal{L}''_k(\alpha), \text{ for } \alpha > 1. \end{aligned} \quad (3.10)$$

Here  $G(\alpha)$  and  $H(\alpha)$  are independent of  $k$ , while  $F_k(\alpha)$  very clearly depends on the  $k$  terms in the sums as given by

$$G(\alpha) = \alpha^4 \sum_{i=1}^{\ell} \gamma_i^2, \quad H(\alpha) = \alpha^2 \sum_{i=1}^{\ell} \phi_i \gamma_i, \quad \text{and} \quad (3.11)$$

$$F_k(\alpha) = \sigma^2 \left( (k - \ell) + 2 \sum_{i=1}^{\ell} \gamma_i + \sum_{i=\ell+1}^k \gamma_i^2 \right). \quad (3.12)$$

**Proof** By (2.4) due to Assumption 1 for  $i \leq \ell$

$$\sigma_i^4 < \sigma_i^{2(1+\nu)} = s_i^2 < \sigma_i^2. \quad (3.13)$$

Thus

$$\alpha^4 \gamma_i^2 = \sigma_i^4 \phi_i^2 < \phi_i^2(\alpha) s_i^2 < \sigma_i^2 \phi_i^2 = \alpha^2 \phi_i \gamma_i. \quad (3.14)$$

Now from (2.6)

$$E(U_k(\alpha)) = \sum_{i=1}^{\ell} \phi_i^2 s_i^2 + \sigma^2 \left( 2 \sum_{i=1}^k \gamma_i(\alpha) + \sum_{i=\ell+1}^k \phi_i^2 \right) = \sum_{i=1}^{\ell} \phi_i^2 s_i^2 + F_k(\alpha),$$

may be bounded using (3.13). This yields immediately (3.8) with the noted definitions for  $G$ ,  $H$  and  $F_k$ , as given in (3.11)–(3.12).



To show (3.9) introduce  $D_i(\alpha) > 0$ ,  $i = 1, 2$ , given by

$$D_1(\alpha) = E(U_k(\alpha)) - (G(\alpha) + F_k(\alpha)) = \sum_{i=1}^{\ell} (\phi_i^2 s_i^2 - \alpha^4 \gamma_i^2)$$

$$D_2(\alpha) = (H(\alpha) + F_k(\alpha)) - E(U_k(\alpha)) = \sum_{i=1}^{\ell} (\alpha^2 \phi_i \gamma_i - \phi_i^2 s_i^2).$$

Then  $D_i$  are independent of  $k$  and

$$D'_1(\alpha) = \sum_{i=1}^{\ell} \left( \frac{2}{\alpha} (2\phi_i^2 \gamma_i s_i^2 + 2\alpha^4 \gamma_i^2 \phi_i) - 4\alpha^3 \gamma_i^2 \right) = \frac{4}{\alpha} \sum_{i=1}^{\ell} (\phi_i^2 \gamma_i s_i^2 - \alpha^4 \gamma_i^3) \text{ and}$$

$$D'_2(\alpha) = \sum_{i=1}^{\ell} (2\alpha \phi_i \gamma_i + \frac{2}{\alpha} (\alpha^2 \phi_i \gamma_i (1 - 2\phi_i) - 2\phi_i^2 \gamma_i s_i^2)) = \frac{4}{\alpha} \sum_{i=1}^{\ell} (\alpha^2 \phi_i \gamma_i^2 - s_i^2 \phi_i^2 \gamma_i).$$

But now again applying Assumption 1 we have

$$\alpha^4 \gamma_i^3 = \sigma_i^4 \phi_i^2 \gamma_i < s_i^2 \phi_i^2 \gamma_i < \sigma_i^2 \phi_i^2 \gamma_i = \alpha^2 \phi_i \gamma_i^2. \quad (3.15)$$

Therefore  $D'_i(\alpha) > 0$ ,  $i = 1, 2$  and we immediately obtain (3.9).

The second derivative result follows similarly using

$$D''_1(\alpha) = \frac{12}{\alpha^2} \sum_{i=1}^{\ell} \gamma_i (1 - 2\phi_i) (s_i^2 \phi_i^2 - \alpha^4 \gamma_i^2) > 0$$

$$D''_2(\alpha) = \frac{12}{\alpha^2} \sum_{i=1}^{\ell} \phi_i \gamma_i (1 - 2\phi_i) (\gamma_i \alpha^2 - s_i^2 \phi_i) > 0,$$

where in each case we apply (3.15) and note  $1 - 2\phi_i \geq 0$ , for  $1 \leq i \leq \ell$  and  $\alpha \leq \sigma_{\ell}$ . This then immediately gives the reverse inequalities for  $\alpha > 1$ .  $\square$

From (3.11)–(3.12) we see that we may write  $G$ ,  $H$  and  $F_k$  in terms of sums  $S_p(i_1, i_2) = \sum_{i=i_1}^{i_2} \gamma_i^p$  for  $p = 1$  and  $p = 2$  by writing  $\phi_i \gamma_i = \gamma_i - \gamma_i^2$ . Hence

$$G(\alpha) = \alpha^4 S_2(1, \ell), H(\alpha) = \alpha^2 (S_1(1, \ell) - S_2(1, \ell)) \text{ and}$$

$$F_k(\alpha) = \sigma^2 (k - \ell + 2S_1(1, \ell) + S_2(\ell + 1, k)).$$

Thus for  $U_{\ell}(\alpha)$  we have the bounding functions by Proposition 3.3

$$\mathcal{L}_{\ell}(\alpha) = G(\alpha) + F_{\ell}(\alpha) = \alpha^4 S_2 + 2\sigma^2 S_1$$

$$\mathcal{U}_{\ell}(\alpha) = H(\alpha) + F_{\ell}(\alpha) = \alpha^2 (S_1 - S_2) + 2\sigma^2 S_1,$$

where the sums all range from 1 to  $\ell$ . Moreover, also by Proposition 3.3,  $\mathcal{L}'_\ell(\alpha) < U_\ell(\alpha) < \mathcal{U}'_\ell(\alpha)$  where

$$\begin{aligned}\mathcal{L}'_\ell(\alpha) &= 4\alpha^3 S_2 + \alpha^4 S'_2 + 2\sigma^2 S'_1 = 4\alpha^3 (S_2 + S_3 - S_2) + \frac{4\sigma^2}{\alpha} (S_2 - S_1) \\ &= \frac{4}{\alpha} (\alpha^4 S_3 + \sigma^2 (S_2 - S_1)) \\ \mathcal{U}'_\ell(\alpha) &= 2\alpha (S_1 - S_2) + \alpha^2 (S'_1 - S'_2) + 2\sigma^2 S'_1 \\ &= 2\alpha (S_1 - S_2) + 2\alpha (S_2 - S_1 - 2(S_3 - S_2)) + \frac{4}{\alpha} \sigma^2 (S_2 - S_1) \\ &= \frac{4}{\alpha} (\alpha^2 (S_2 - S_3) + \sigma^2 (S_2 - S_1)),\end{aligned}$$

and we used  $\gamma'_i = -(2/\alpha)\gamma_i\phi_i = (2/\alpha)(\gamma_i^2 - \gamma_i)$  and  $(\gamma_i^2)' = -(4/\alpha)\gamma_i^2\phi_i = (4/\alpha)(\gamma_i^3 - \gamma_i^2)$ .

**Proposition 3.5** *Suppose Assumption 1 holds, then necessarily  $U'_\ell(\alpha) < 0$  for  $\alpha^2 < \sigma_{\ell+1}^2/(1 - \sigma_{\ell+1}^2)$ . Hence  $\alpha_\ell^2 > \sigma_{\ell+1}^2/(1 - \sigma_{\ell+1}^2)$ .*

**Proof** If the upper bound has a negative slope,  $\mathcal{U}'_\ell(\alpha) < 0$  for some  $\alpha$ , then  $U'_\ell(\alpha) < 0$  also. Immediately  $\mathcal{U}'_\ell(\alpha) < 0$  for  $\alpha^2(S_2 - S_3) + \sigma^2(S_2 - S_1) < 0$ , and for  $U'_\ell(\alpha) < 0$  it is sufficient that for  $1 \leq i \leq \ell$

$$0 > \alpha^2(\gamma_i^2 - \gamma_i^3) + \sigma^2(\gamma_i^2 - \gamma_i) = \gamma_i(\alpha^2\gamma_i(1 - \gamma_i) + \sigma^2(\gamma_i - 1)) = \gamma_i\phi_i(\alpha^2\gamma_i - \sigma^2),$$

and we need  $(\alpha^2\gamma_i - \sigma^2) < 0$ , or  $\alpha^2\sigma_i^2 - \sigma^2(\alpha^2 + \sigma_i^2) < 0$ . Now, for  $i \leq \ell$ ,  $\sigma_i^2 \geq \sigma_\ell^2 > \sigma^2$  and we obtain  $\alpha^2 < \min(\sigma^2\sigma_i^2/(\sigma_i^2 - \sigma^2))$  for all  $1 \leq i \leq \ell$ . But  $x^2/(x^2 - a^2)$  is decreasing with  $x$  for  $x^2 > a^2$ , hence we need  $\alpha^2 < \sigma^2/(1 - \sigma^2)$ . For  $\sigma_{\ell+1}^2 < \sigma^2 < \sigma_\ell^2$  and using  $x^2/(1 - x^2)$ , which is increasing with  $x \in (0, 1)$ , we obtain  $\alpha^2 < \sigma_{\ell+1}^2/(1 - \sigma_{\ell+1}^2)$ . Hence we must have  $\alpha_\ell^2 > \sigma_{\ell+1}^2/(1 - \sigma_{\ell+1}^2)$ .

We now extend the analysis to obtain a lower bound on  $\alpha_k$  for all  $k > \ell$ .

**Theorem 3.1** *Suppose Assumptions 1 and 2 hold, and that  $U_k(\alpha_k)$  is a minimum for  $U_k(\alpha)$ , then, for  $k \geq \ell$ ,  $\alpha_k > \alpha_\ell > \sigma_{\ell+1}/\sqrt{1 - \sigma_{\ell+1}^2} = \alpha_{\min}$ .*

**Proof** First suppose the contrary and that  $\alpha_k \leq \sigma_{\ell+1}/\sqrt{1 - \sigma_{\ell+1}^2}$ . Then  $U'_k(\alpha_k) = 0$  and by (3.5)  $U'_\ell(\alpha_k) > 0$ . But by Proposition 3.5  $U'_\ell(\alpha) < 0$  for  $\alpha \leq \sigma_{\ell+1}/\sqrt{1 - \sigma_{\ell+1}^2}$  and we have a contradiction yielding  $\alpha_k > \sigma_{\ell+1}/\sqrt{1 - \sigma_{\ell+1}^2} = \alpha_{\min}$ ,  $k \geq \ell$ . It remains to determine whether it is possible to have  $\sigma_{\ell+1}/\sqrt{1 - \sigma_{\ell+1}^2} < \alpha_k < \alpha_\ell$  where  $\alpha_\ell$  is the first minimum point of  $U_\ell(\alpha)$  to the right of  $\alpha_{\min}$ . Again we proceed by contradiction and suppose that  $\alpha_k \in [\alpha_{\min}, \alpha_\ell]$  exists. Then we have the following:

1. By (3.5)  $E(U'_k)(\alpha_\ell) < U'_\ell(\alpha_\ell) = 0$ , and by (3.6), noting  $\alpha > \sigma_{\ell+1}/\sqrt{5}$ ,  $E(U''_k)(\alpha_\ell) > U''_\ell(\alpha_\ell) > 0$ . Hence  $U_k(\alpha)$  is convex and decreasing at  $\alpha_\ell$ .
2. At the minimum critical point  $\alpha_k < \alpha_\ell$ ,  $U'_k(\alpha_k) = 0$ . Thus there must also be a second critical point which is a maximum for some  $\bar{\alpha}$  in the interval  $\alpha_k < \bar{\alpha} < \alpha_\ell$ , for which  $U'_k(\bar{\alpha}) = 0$  and  $U''_k(\bar{\alpha}) < 0$ .
3. At  $\bar{\alpha}$  we then have by (3.5) that  $U'_\ell(\bar{\alpha}) > 0$ . Hence  $U_\ell(\alpha)$  is increasing at  $\bar{\alpha} < \alpha_\ell$  but is decreasing at  $\alpha_{\min} < \bar{\alpha}$ , i. e.  $U'_\ell(\alpha)$  changes sign for some  $\alpha$  in the interval  $[\alpha_{\min}, \bar{\alpha}]$ . But by continuity then  $U_\ell(\alpha)$  has at least one minimum on this interval. By assumption, however,  $\alpha_\ell$  is the first minimum point of  $U_\ell(\alpha)$  to the right of  $\alpha_{\min}$  and we have arrived at a contradiction.

□

We have now obtained a tight lower bound on  $\alpha_k$

$$\alpha_{\min} = \frac{\sigma_{\ell+1}}{\sqrt{1 - \sigma_{\ell+1}^2}} < \alpha_k, \quad \ell \leq k \leq r. \quad (3.16)$$

It remains to discuss the convergence of  $\{\alpha_k\}$  to  $\alpha_{k_{\text{opt}}}$  with increasing  $k$ . We note that one approach would be to show that the  $U_k(\alpha)$  are convex for  $\alpha > \sigma_\ell$ , but the sign result in (3.10) only immediately applies for  $\alpha > 1$ , hence investigating the sign requires a more refined bound for each interval  $\alpha \in [\sigma_i, \sigma_{i-1}]$  for  $i \leq \ell$ . Instead we obtain the following result, which relies on the uniqueness of  $\alpha_k$ .

**Theorem 3.2** *Suppose Assumptions 1 and 2 hold and that  $\alpha_{k_{\text{opt}}}$  and each  $\alpha_k$ ,  $k > \ell$  are unique within the given interval  $\sigma_{\ell+1}/\sqrt{1 - \sigma_{\ell+1}^2} < \alpha < 1$ . Then, the sequence  $\{\alpha_k\}_{k>\ell}$  is on the average increasing with  $\lim_{k \rightarrow r} E(\alpha_k) = E(\alpha_{k_{\text{opt}}})$  and  $\{U_k(\alpha_k)\}$  is increasing.*

**Proof** It is immediate from (2.6) that  $U_k(\alpha) \geq U_\ell(\alpha)$  for any  $k > \ell$  and any  $\alpha$ , and that  $U_{k+1}(\alpha) \geq U_k(\alpha)$ . Thus the  $\{U_k(\alpha)\}$  is an increasing set of functions with  $k > \ell$ . By (3.5) of Proposition 3.3 we also have  $E(\frac{\partial U_{k+1}(\alpha)}{\partial \alpha}) < E(\frac{\partial U_k(\alpha)}{\partial \alpha}) < \frac{\partial U_\ell(\alpha)}{\partial \alpha}$ , and  $\{E(\frac{\partial U_k(\alpha)}{\partial \alpha})\}$  is a decreasing set of functions for  $k > \ell$ . In particular  $E(\frac{\partial U_{k+1}(\alpha_\ell)}{\partial \alpha}) < E(\frac{\partial U_k(\alpha_\ell)}{\partial \alpha}) < 0$ . Moreover, by Corollary 3.1 and (3.6) of Proposition 3.3, when  $\alpha_\ell > \sigma_{\ell+1}/\sqrt{5}$  the expected second derivatives at  $\alpha_\ell$  are positive and increasing with  $k$  so that the first derivative increases to 0 more quickly for larger  $k$ . Thus, not only do we have  $E(\alpha_k) > \alpha_\ell > \alpha_{\min}$  for all  $k$ , we also have that  $\{E(\alpha_k)\}$  converges from below to  $E(\alpha_{k_{\text{opt}}})$ .

**Corollary 3.3** (Faster Decay Rate of the Coefficients) *Suppose that the coefficients  $s_i$  decay at the rate  $s_i^2 = \sigma_i^{2(\rho+\nu)}$  for integer  $\rho > 1$ . Then the results of Theorems 3.1–3.2 still hold.*

**Proof** This holds by modifying the inequality (2.4) for the faster decay rate yielding

$$K_i \sigma_i^4 < \sigma_i^{2(\rho+\nu)} = s_i^2 < \sigma_i^2 K_i, \quad K_i = \sigma_i^{2(\rho-1)}.$$

Thus the coefficients are bounded as in (3.14) but with scale factor  $K_i$

$$\alpha^4 \gamma_i^2 K_i = \sigma_i^4 \phi_i^2 K_i < \phi_i^2(\alpha) s_i^2 < K_i \sigma_i^2 \phi_i^2 = K_i \alpha^2 \phi_i \gamma_i.$$

Using this relation all the results presented in Proposition 3.4 still hold with  $H(\alpha)$  and  $G(\alpha)$  replaced by

$$G_\rho(\alpha) = \alpha^4 \sum_{i=1}^{\ell} K_i \gamma_i^2, \quad \text{and} \quad H_\rho(\alpha) = \alpha^2 \sum_{i=1}^{\ell} K_i \phi_i \gamma_i.$$

Then again redefining the summations  $S_\rho$  to now depend on the coefficients with  $K_i$ , for  $H_\rho$  and  $G_\rho$ , following Proposition 3.5 yields the condition

$$\gamma_i \phi_i (\alpha^2 K_i \gamma_i - \sigma^2) < 0$$

for  $U'_\ell(\alpha) < 0$ . Continuing the argument as in the proof of Proposition 3.5 still yields the lower bound  $\alpha_\ell^2 > \sigma_{\ell+1}^2 / (1 - \sigma_{\ell+1}^2)$ . But this is all that is required for Theorems 3.1–3.2 and hence the results follow without modification.  $\square$

**Remark 3.3** This result shows that given a TSVD which sufficiently incorporates the dominant terms of the SVD expansion, including sufficient terms that are noise-contaminated,  $\alpha_k$  will be an increasingly good approximation for  $\alpha_{k_{\text{opt}}}$ . Moreover, including additional terms in the expansion will have limited impact on the solution, because  $\alpha_{k_{\text{opt}}} > \alpha_\ell$  and filter factor  $\gamma_i(\alpha_{k_{\text{opt}}})$  is decreasing with  $i$ . In particular, we are using  $\gamma_i(\alpha_{k_{\text{opt}}}) < \gamma_i(\alpha_\ell) < \gamma_{\ell+1}(\alpha_\ell) < \gamma_{\ell+1}(\sigma_{\ell+1}) = 1/2$ , for  $i > \ell + 1$  and  $\alpha_\ell > \sigma_{\ell+1}$ . These nested inequalities follow immediately because  $\gamma(x, y) = y^2(y^2 + x^2)^{-1}$  is decreasing as a function of  $x$  and increasing as a function of  $y$ .

**Remark 3.4** Although the main result of this paper effectively relies on an assumption that the UPRE functions have unique minima within the obtained bounds,  $\alpha_{\min} < \alpha_k < 1$ , proving that the minima are indeed unique seems to require using the discrete summations occurring in  $U_k(\alpha)$  as approximations to continuous integrals. This approach is very technical, not very general, being dependent on the decay rate parameter  $\tau$ , and serves only to tighten the lower bound for  $\alpha$ . We therefore chose not to present results along this direction, relying on the computational results that are supportive of the unique identification of a minimum within these realistic bounds.

**Remark 3.5** The results given depend on the assumption that summations with  $s_i^2$  for terms with  $i > \ell$  may be approximated in terms of the noise variance. For  $r - \ell$  small relative to  $r$ , this assumption breaks down. As  $r - \ell$  increases the assumptions become more reliable and less impacted by outlier data for  $s_i^2$ . Still the main convergence theorem holds only with respect to this analysis and we cannot expect that  $\{\alpha_k\}$  will always converge monotonically to  $\alpha_{k_{\text{opt}}}$  in practice. With sufficient safeguarding, as noted in the algorithm presented in Sect. 4, it is reasonable to expect that  $\alpha_{k_{\text{opt}}}$  is quickly and accurately identified.

**Remark 3.6** (*Posterior Covariance*) We have shown  $\{\alpha_k\}$  increases with  $k$ . Consequently, the approximate a posteriori covariance of the filtered TSVD solution  $\sigma^2/(4\alpha^2)$  decreases with  $k$ , to  $\sigma^2/(4\alpha_{k_{\text{opt}}}^2)$ . In trading-off the minimization of the risk by using the UPRE to find the optimal  $\alpha$ , the method naturally finds a solution which has increasing smoothness with increasing  $k$ . This limits the impact of the possibly non-smooth components of the solution corresponding to small singular values, most likely noise-contaminated, that would contaminate the unfiltered TSVD solution.

## 4 Practical application

The convergence theory for  $\{\alpha_k\} \rightarrow \alpha_{k_{\text{opt}}}$  as  $k \rightarrow k_{\text{opt}}$  presented in Sect. 3 motivates the construction of an algorithm to automatically determine the optimal index  $k_{\text{opt}}$ , defined as in Sect. 1 to be the optimal number of terms to use from the TSVD, and associated regularization parameter  $\alpha_{k_{\text{opt}}}$ . The algorithm is presented and discussed in Sect. 4.1 and tested for 2D test problems using IR Tools [10] in Sect. 4.2. These results also corroborate the convergence theory presented in Sect. 3.

### 4.1 Algorithm

We propose an algorithm that works by iteratively minimizing (2.6) on the TSVD subspace of size  $k \leq r$  until a set of convergence criteria are met. These convergence criteria rest on the observation that *in general* for sufficiently large  $k$ , the relative change,  $c_k = |(\alpha_k - \alpha_{k+1})|/\alpha_k > 0$ , between successive parameter estimates,  $\alpha_k$  and  $\alpha_{k+1}$ , decreases as  $k$  increases towards  $r$ . If during the iterative procedure there exists a  $k$  such that it is reasonably believed that  $\alpha_k \approx \alpha_i$  for all  $i > k$ , the algorithm terminates, producing  $k_{\text{opt}}$  and  $\alpha_{k_{\text{opt}}}$ . A pseudo-code implementation is given as Algorithm 1.

---

#### Algorithm 1: Truncated UPRE Parameter Estimation

---

**Input:** SVD or TSVD; data  $\mathbf{b}$  and noise variance estimate  $\sigma^2$ ; initial index  $k_0$ ; maximum  $k$ ,  $k_{\text{max}}$ ; step size  $\Delta_k$ ; relative tolerance  $\delta$ ; window length  $w$ ; optional estimate for  $\ell$   
**Output:** Converged parameter  $\alpha_{k_{\text{opt}}}$ ; convergence index  $k_{\text{opt}}$ ; relative mean change  $\hat{c}_{iw}$ ;

```

1  $k \leftarrow k_0$ ;  $\hat{c}_{iw} \leftarrow \inf$ 
2 Initialize  $\alpha_{\min}$  according to (3.16) using  $\ell$  if provided, otherwise using  $k$ 
3  $\alpha(0) \leftarrow \arg \min_{\alpha} U_k(\alpha)$  over interval  $[\alpha_{\min}, 1]$ 
4 while ( $\hat{c}_{iw} > \delta$  and  $k < k_{\text{max}}$ ) or ( $\alpha(i) = \alpha_{\min}$ ) do
5    $i \leftarrow i + 1$ ;  $k \leftarrow k + \Delta_k$ 
6   If  $\ell$  not provided, update  $\alpha_{\min}$  according to (3.16) using  $k$ 
7    $\alpha(i) \leftarrow \arg \min_{\alpha} U_k(\alpha)$  over interval  $[\alpha_{\min}, 1]$ 
8    $c(i) = (|\alpha(i) - \alpha(i-1)|)/\alpha(i)$ 
9   if  $i \geq w$  then
10     $\hat{c}_{iw} \leftarrow \text{mean}(c(i), c(i-1), \dots, c(i-w+1))$ 
11  end
12 end
13 return  $k = k_{\text{opt}}$ ,  $\alpha(i) = \alpha_{k_{\text{opt}}}$ ,  $\hat{c}_{iw}$ 
```

---

Algorithm 1 takes as input a full or truncated SVD as well as a number of required and optional parameters which we now discuss. For large scale problems it is not necessary, and is even discouraged, to compute  $\alpha_k$  for all  $k \leq k_{\text{opt}}$ . For moderately or mildly ill-posed problems, and for problems with high signal to noise ratios in which the expected  $k_{\text{opt}}$  is likely to be large relative to the problem size, it is recommended to start the algorithm at some  $k_0 \neq 1$  and to increment  $k$  by some  $\Delta_k \neq 1$ , yielding the sequence  $\{k(i) : k_0, k_0 + \Delta_k, k_0 + 2\Delta_k, \dots, k_0 + i\Delta_k\}$ . The algorithm computes the sequence  $\{\alpha_{k_0}, \alpha_{k_0+\Delta_k}, \alpha_{k_0+2\Delta_k}, \alpha_{k_0+3\Delta_k}, \dots\}$ , each solving (3.1) for the given index, until either  $k_0 + i\Delta_k \geq k_{\text{max}}$  or until  $\alpha_k$  has converged, where  $k_0$ ,  $\Delta_k$ , and  $k_{\text{max}}$  are provided by the user. For each  $k_0 + i\Delta_k$  the relative change in  $\alpha$  is computed as  $c_i = |\alpha_{k_0+i\Delta_k} - \alpha_{k_0+(i-1)\Delta_k}| / \alpha_{k_0+i\Delta_k}$ . Noting again that  $c_i$  is only in general decreasing for sufficiently large  $i$ , it is unwise to determine stopping criteria by directly thresholding on  $c_i < \delta$ , for some user provided tolerance  $\delta$ . It is observed that higher confidence in convergence can be achieved by requiring  $\hat{c}_{iw} < \delta$  where  $\hat{c}_{iw}$  is the mean of multiple  $c_i$ 's calculated over the window of size  $w$ , i.e. over  $\{c_i, c_{i+1}, \dots, c_{i+w}\}$ . This protects against the possibility of stopping the parameter search too early and prior to the stabilization of  $\alpha_k$ . This occurs when  $c_i < \delta$ , while at the same time  $c_j \geq \delta$  for some  $j > i$ . Due to the impact of noise on calculating the parameter  $\alpha_k$ , if  $k$  is not yet sufficiently large so that  $\alpha_k$  has not stabilized then the relative changes between successive estimates of  $\alpha_k$  may be either extremely small or large. Comparing multiple values of  $c_i$  in the form of  $\hat{c}_{iw}$  to  $\delta$  enables a broader view of the convergence of  $\alpha_k$ , and the moving window average smooths out variation in  $c_i$ .

**Remark 4.1** (Parameter  $\Delta_k$ ) The choice of  $\Delta_k$  is influenced by the size of the problem and if known, an estimate for the expected number of terms to be used in the TSVD solution. While choosing  $\Delta_k$  large has computational advantages due to a larger step size in the search for  $k_{\text{opt}}$ , with  $\Delta_k$  too large one risks the possibility of Algorithm 1 producing a value of  $k_{\text{opt}}$  larger than necessary. Solutions with  $k_{\text{opt}}$  larger than necessary more closely resemble the full UPRE regularized solution. For the problem sizes considered here  $\Delta_k \in \{5, 10, 25\}$  all seemed to work well.

**Remark 4.2** (Parameter  $w$ ) The choice of  $w$  has a similar effect as  $\Delta_k$ . Choosing  $w$  large will delay the termination criteria. Parameters  $\Delta_k$  and  $w$  interact in the sense that they together determine the set  $\{c_i, c_{i+1}, \dots, c_{i+w}\}$  whose mean is compared to  $\delta$  in determining convergence. The choice of  $w$  determines how many values are being averaged, while  $w$  and  $\Delta_k$  determine the minimum and maximum  $k$  of the moving window over which  $\alpha_k$  is tested for convergence. Choosing  $w \in \{5, 10, 25, 50\}$  worked well for the problems considered here.

**Remark 4.3** (Parameter  $\delta$ ) Algorithm 1 is sensitive to  $\delta$  and we recommend choosing  $\delta \in [1e-5, 1e-3]$ . In our experiments  $\delta > 1e-3$  terminated the algorithm prior to convergence resulting in over smoothed solutions due to an underestimate of  $k_{\text{opt}}$ , while  $\delta < 1e-5$  produced  $k_{\text{opt}}$  far greater than necessary.

To summarize, the required input to the proposed algorithm is a full or truncated SVD, a starting index  $k_0$ , a step size between successive estimates  $\Delta_k$ , an upper-bound  $k_{\text{max}}$  dependent on the severity of the problem and the noise level, a tolerance  $\delta$ , and

a width  $w$  over which the moving average of relative changes in successive estimates of  $\alpha$  is computed.

The results of Theorem 3.1 are incorporated into Algorithm 1 with the inclusion of an optional parameter  $\ell$  specifying an estimate for the index at which noise dominates the coefficients. If a Picard plot is available  $\ell$  can be estimated visually, otherwise an approach relying on Picard parameter estimates similar to that used by [24,34] can be used. If an estimate for  $\ell$  is available,  $\alpha_{\min}$  is calculated according to (3.16), and  $\alpha_k$  is found using  $\alpha_{\min} = \sigma_{\ell+1}/\sqrt{1 - \sigma_{\ell+1}^2}$  and  $\alpha_{\max} = 1$  in (3.1). Otherwise, the bound  $\sigma_{k+1}/\sqrt{1 - \sigma_{k+1}^2}$  is used in (3.1). In either case if the lower bound is achieved then the theory indicates that noise has not yet dominated and the algorithm is allowed to continue. Thus, in the case where  $k < k_{\max}$ , necessary conditions for the termination of Algorithm 1 are  $\hat{c}_{iw} < \delta$  and  $\alpha_k$  should be greater than the specified  $\alpha_{\min}$ .

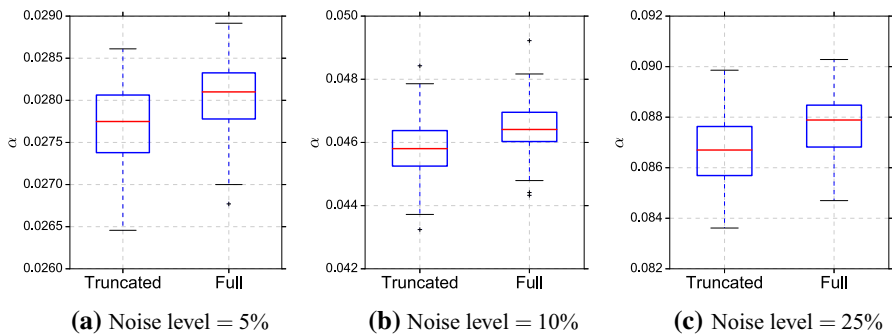
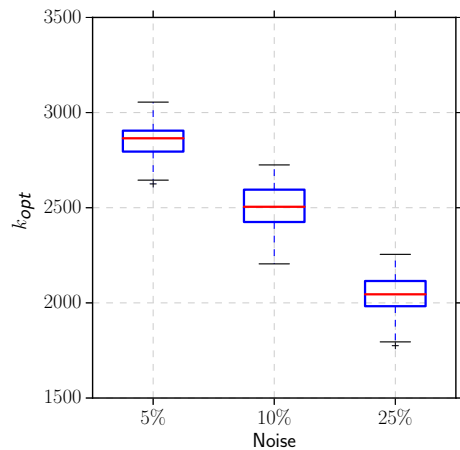
## 4.2 Verification of the algorithm and theory

We now present the evaluation of Algorithm 1 on a 2D test problem using the IR Tools package described in [10]. We report the results applying a Gaussian blur to test problem *Satellite* of size  $256 \times 256$  using *PRblur*, with medium blur. We considered noise levels of 5%, 10%, and 25%, with 100 noise instances generated for each noise level. The IR Tools function *PRnoise* was used to generate noise, where the noise level is defined as  $\|\mathbf{b}\|_2/\|\mathbf{b}\|$ . A moving window of size  $w = 5$  in computing  $\hat{c}_{iw}$  with relative tolerance of  $\delta = 1e-3$  was found to work well for each noise level, but may need to be adapted to the severity of the ill-posedness of the problem. Recorded in each run are the converged  $\alpha_{k_{\text{opt}}}$ , the size of the TSVD subspace  $k_{\text{opt}}$  to be used, and the relative reconstruction error (RRE). RRE is defined as  $\|x_{\text{true}} - x_{k_{\text{opt}}}\|_2/\|x_{\text{true}}\|_2$  where  $x_{k_{\text{opt}}}$  is the filtered,  $k_{\text{opt}}$ -truncated TSVD solution obtained by using  $\alpha_{k_{\text{opt}}}$  as the regularization parameter.

Figure 4 is a box plot<sup>1</sup> showing the spread of  $k_{\text{opt}}$  values for the 100 noise instances run for each noise level, where in each case  $k_{\text{opt}} \ll r = 65,536$ . Figure 5 is a box plot comparing the  $\alpha_{k_{\text{opt}}}$  returned by the algorithm, and  $\alpha_r$  obtained by minimizing the UPRE on the full space. These figures together reaffirm that the optimal regularization parameter found by UPRE is largely determined by a relatively small number of terms in the TSVD, and less impacted by the tail of the coefficients dominated by noise. The estimation of  $\{\alpha_k\}$  with increasing number of terms in the TSVD is depicted in Fig. 6 for the first 10 runs of each noise level, where the point of convergence ( $k_{\text{opt}}, \alpha_{k_{\text{opt}}}$ ) is represented as a cyan triangle. It should be noted that the estimated lower bound  $\alpha_{\min}$  was not used, and  $\alpha$  was minimized over the interval  $(0, 1)$  using *fminbnd* (*fminbound* is used for the Python implementation). A tolerance of  $\delta = 1e-3$  was found to produce a value for  $\alpha_{k_{\text{opt}}}$  just prior to the point where  $\{\alpha_k\}$  began to stabilize. A smaller  $\delta$  will necessarily increase  $k_{\text{opt}}$ , but with negligible changes in  $\alpha_{k_{\text{opt}}}$ . In these simulations averaged over all 100 runs,  $\alpha_{k_{\text{opt}}}$  was within 1.22%, 1.47%, and 1.17% of

<sup>1</sup> A box plot is a visual representation of summary statistics for a given sample. Horizontal lines of each plotted box represent the 75%, 50% (median), and 25% quantiles, with outliers plotted as individual crosses or points.

**Fig. 4** Box plots showing the index  $k_{\text{opt}}$  produced by Algorithm 1 for problem *Satellite* computed from 100 runs for noise levels 5%, 10%, and 25%. The number of terms  $k$  in the TSVD that provide useful information decreases as the noise level increases

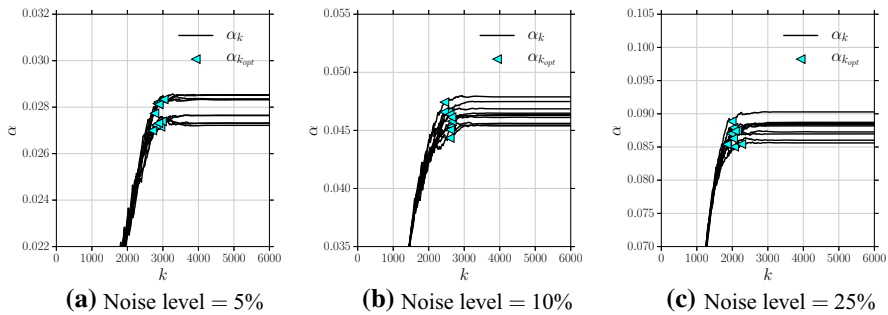


**Fig. 5** Box plots comparing parameter estimates  $\alpha_{k_{\text{opt}}}$  with  $\alpha_r$  for problem *Satellite* computed from 100 runs for noise levels 5%, 10%, and 25%. For each noise level, the estimate  $\alpha_{k_{\text{opt}}}$  produced by Algorithm 1 is generally less than  $\alpha_r$ , demonstrating that by including more terms in the TSVD,  $k > k_{\text{opt}}$ , greater regularization is required. Note that the limits on the y-axes vary across subplots to better visualize the parameter distributions across noise levels

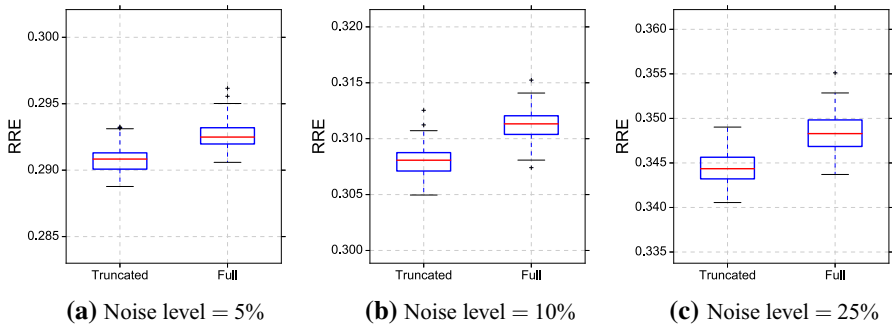
$\alpha_r$  for noise levels 5%, 10%, and 25% respectively using fewer than 5% of the SVD components.

In terms of RRE, the solution obtained using the truncated UPRE and a subspace of size  $k_{\text{opt}}$  with parameter  $\alpha_{k_{\text{opt}}}$  generated by Algorithm 1 generally provided a better solution than obtained using the full UPRE for each noise level. Figures 7 and 8 show box plots and histograms respectively of the RRE comparing the regularized TSVD and the full UPRE solution. Over all noise levels, the median and mean reconstruction error of 100 noise instances is lower in the regularized TSVD solution. Similar to the Picard parameter approaches of [34], Algorithm 1 identifies an index  $k_{\text{opt}}$  for which coefficients  $s_k$  are dominated by noise for  $k > k_{\text{opt}}$ . Our approach, however, does not rely on performing statistical tests on the coefficients, but instead examines the stabilization of  $\alpha_k$  as  $k$  increases. Once  $\alpha_k$  has stabilized, adding additional noise dominated terms in the solution delivers no benefit. Furthermore, if a TSVD with  $k_{\text{max}}$  terms has been calculated, then either  $\alpha_k$  converges for  $k < k_{\text{max}}$  or we know that





**Fig. 6** Line plots showing the calculated estimates for  $\{\alpha_k\}$  with increasing number of terms  $k$  in the TSVD. The results are given for problem *Satellite* for noise levels 5%, 10%, and 25%, for 10 random noise instances at the specified noise level. The resulting point  $(k_{\text{opt}}, \alpha_{k_{\text{opt}}})$  produced by Algorithm 1 is displayed as a cyan triangle. Note that the limits on the y-axes vary across subplots to better visualize the convergence across noise levels

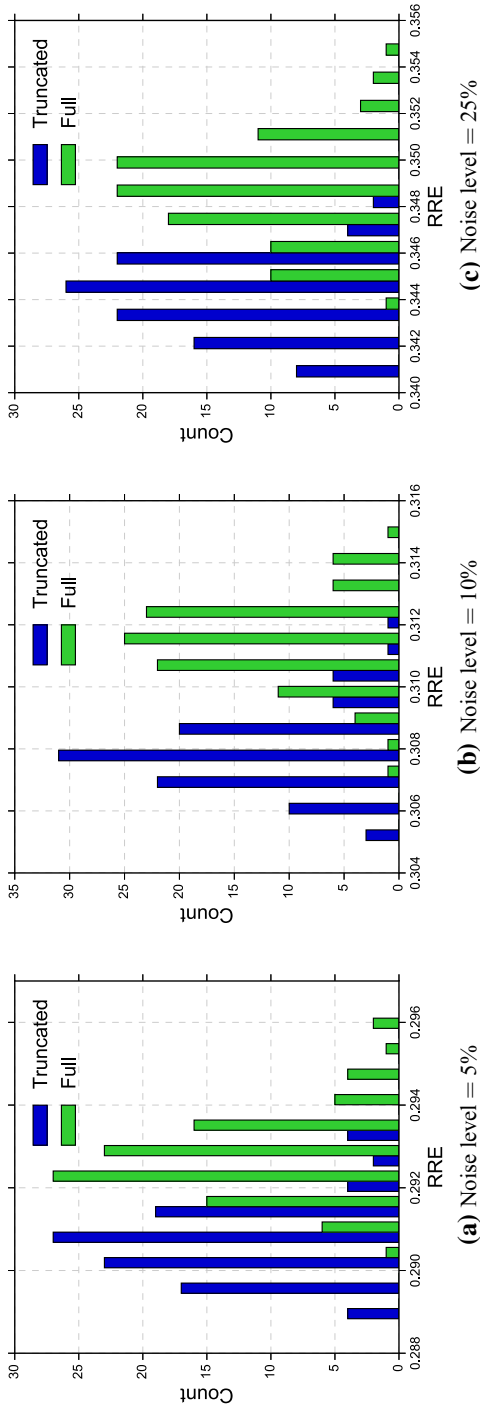


**Fig. 7** Box plots of RRE comparing solutions using truncated UPRE with parameter  $\alpha_{k_{\text{opt}}}$  and solutions using full UPRE with parameter  $\alpha_r$  for problem *Satellite* computed from 100 runs for noise levels 5%, 10%, and 25%. Regularization parameter  $\alpha_{k_{\text{opt}}}$  obtained by UPRE on a TSVD generally has lower error, as evident from Truncated UPRE plots being vertically shifted downwards relative to full UPRE boxplots. Note that the limits on the y-axes vary across subplots to better visualize the spread of the distributions across noise levels

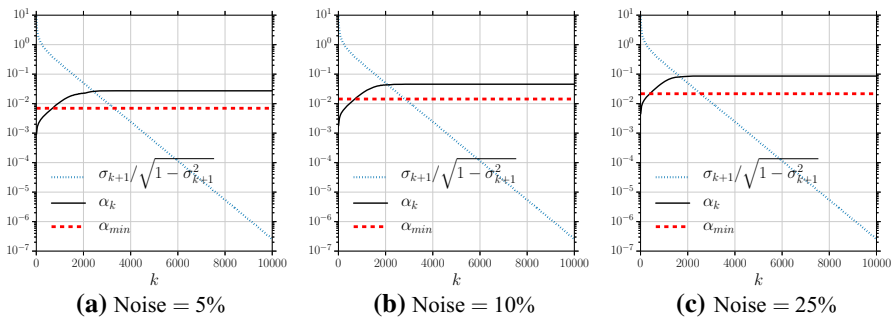
the optimal choice  $k_{\text{opt}}$  is greater than  $k_{\text{max}}$ , and that  $\hat{c}_{iw}$  provides some estimate for whether  $k_{\text{opt}} \gg k_{\text{max}}$  or whether the given TSVD can be assumed to be sufficient in providing a good estimate for the solution  $\mathbf{x}$ .

In these simulations  $\ell$  is not known precisely but was estimated by visual inspection of the Picard coefficients, as well as by comparing the distributions of the noise contaminated and noise free coefficients. This approach for estimating  $\ell$  is not possible in general as the noise free coefficients are unknown in practice, but this method of estimating  $\ell$  was employed for the purpose of validating the results of Theorem 3.1. An estimate for the lower bound  $\alpha_{\min}$  obtained from (3.16) is depicted as the red dashed curve in Fig. 9, with  $\{\alpha_k\}$  the solid black line. It can be seen that  $\alpha_{\min}$  serves as a tight lower bound for the converged parameter  $\alpha_{k_{\text{opt}}}$ , and the lower bound  $\sigma_{k+1}/\sqrt{1 - \sigma_{k+1}^2}$  can be used effectively in cases where an estimate of  $\ell$  is not available.

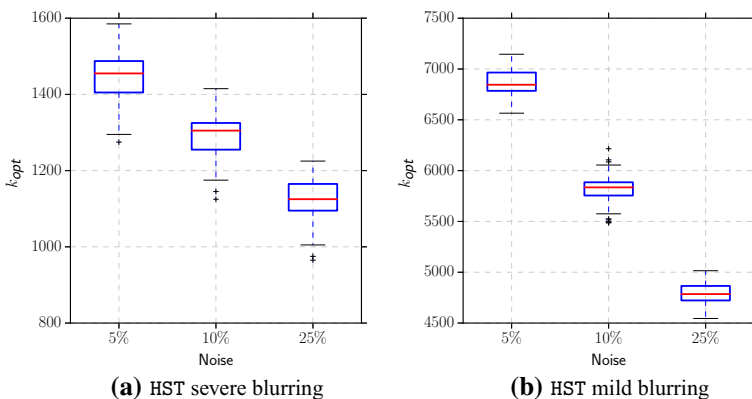
In addition to test image *Satellite* with a medium Gaussian blur applied, we also applied Algorithm 1 with the same parameters to test image *HST* with both mild



**Fig. 8** Histograms of RRE comparing solutions using  $\alpha_{k_{\text{opt}}}$  and solutions using  $\alpha_r$  for problem Satellite computed from 100 runs for each noise level 5%, 10%, and 25%. Regularization parameter  $\alpha_{k_{\text{opt}}}$  obtained by UPRE on a TSVD generally has lower error, as evident from the truncated histograms having peaks shifted to the left relative to the full UPRE



**Fig. 9** Line plots showing the convergence of  $\{\alpha_k\}$  for problem Satellite for noise levels 5%, 10%, and 25%. In each subplot,  $\alpha_r$  is plotted as a solid black line for 10 random noise instances at the specified noise level. The dotted blue curve represents the lower bound in (3.16) as a function of  $k$ , with the red dashed line representing the lower bound according to Theorem 3.1 and dependent on  $\ell$  for a single run (colour figure online)



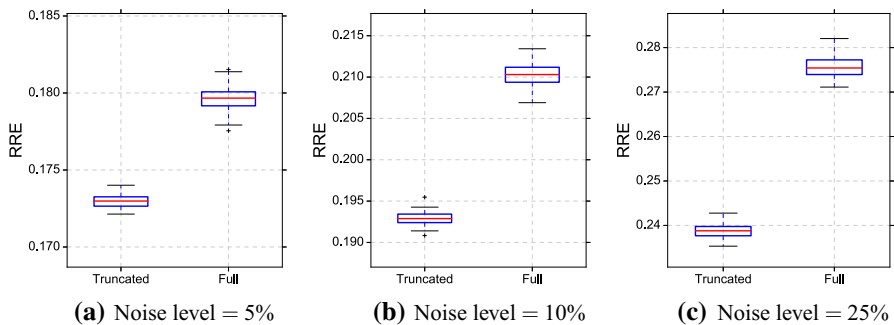
**Fig. 10** Box plots showing the index  $k_{opt}$  produced by Algorithm 1 for problem HST computed from 100 runs for noise levels 5%, 10%, and 25%. The number of terms  $k$  in the TSVD that provide useful information decreases as the noise level increases

and severe Gaussian blurring. The results, summarized in Figs. 10, 11, 12 are consistent with the results for test case Satellite.

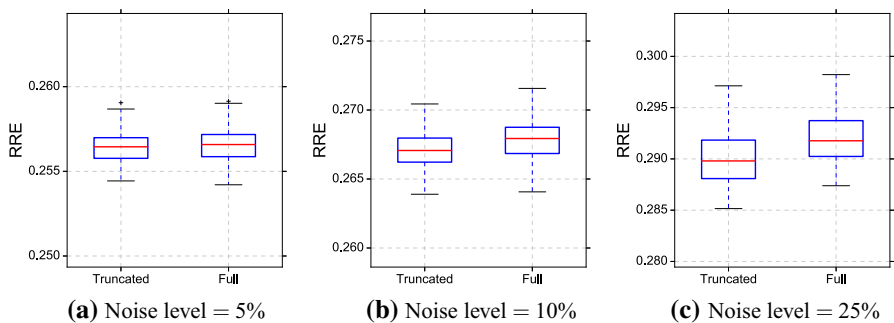
In summary, given a TSVD or SVD, an optional estimate of  $\ell$ , and suitable parameters determined by the ill-posedness of the problem, Algorithm 1 is able to effectively determine a regularization parameter  $\alpha_{k_{opt}}$  obtained by UPRE minimization over the TSVD subspace of size  $k_{opt}$ , such that the regularized truncated solution  $\mathbf{x}$  has consistently lower RRE than the full UPRE solution.

## 5 Conclusions

We have demonstrated that the regularization parameter obtained using the UPRE estimator converges with increasing number of terms used from the TSVD for the



**Fig. 11** Box plots of RRE comparing solutions using truncated UPRE with parameter  $\alpha_{k_{\text{opt}}}$  and solutions using full UPRE with parameter  $\alpha_r$  for problem HST with mild blur computed from 100 runs for noise levels 5%, 10%, and 25%. Regularization parameter  $\alpha_{k_{\text{opt}}}$  obtained by UPRE on a TSVD has consistent lower error, as evident from Truncated UPRE plots being vertically shifted downwards relative to full UPRE boxplots. Note that the limits on the y-axes vary across subplots to better visualize the spread of the distributions across noise levels



**Fig. 12** Box plots of RRE comparing solutions using truncated UPRE with parameter  $\alpha_{k_{\text{opt}}}$  and solutions using full UPRE with parameter  $\alpha_r$  for problem HST with severe blur computed from 100 runs for noise levels 5%, 10%, and 25%. Regularization parameter  $\alpha_{k_{\text{opt}}}$  obtained by UPRE on a TSVD generally has lower error for noise levels 10% and 25%, with comparable error for noise level 5%. Note that the limits on the y-axes vary across subplots to better visualize the spread of the distributions across noise levels

solution. For a severely ill-posed problem the convergence occurs very quickly and is independent of the size of the problem due to the fast contamination of data coefficients by practical levels of noise. Practically-relevant problems are often, however, only moderately or mildly ill-posed, e. g. [6,14,36,38], and it is therefore important to accurately and efficiently find both  $k_{\text{opt}}$  and  $\alpha_{k_{\text{opt}}}$ .

Theoretical results have been presented that demonstrate the convergence of the regularization parameter  $\alpha_k$  with  $k$ , increasing from below to  $\alpha_{k_{\text{opt}}} \leq \alpha_r$ , the optimal value for the full SVD. The posterior covariance thus decreases with  $k$ , leveling at approximately  $\sigma^2/(4\alpha_{k_{\text{opt}}}^2)$ . Thus the method naturally finds a solution which has increasing smoothness with increasing  $k$  and solutions obtained without truncation will exhibit larger error due to increased smoothing. An effective and practical algorithm that implements the theory has also been provided, and validated for 2D image deblurring. These results expand on recent research on the characterization of the

regularization parameter as closely dependent on the size of the singular subspace represented in the solution [9,30,31]. As there is a resurgence of interest in using a TSVD solution for the solution of ill-posed problems due to increased feasibility of finding a good approximation of a dominant singular subspace using techniques from randomization, e.g. [7,8,13,26,28,32], the results are more broadly relevant for more efficient estimates of the TSVD. Implementation of the algorithm in these contexts is a topic for future work.

## References

1. Abascal, J.-F.P.J., Arridge, S.R., Bayford, R.H., Holder, D.S.: Comparison of methods for optimal choice of the regularization parameter for linear electrical impedance tomography of brain function. *Physiol. Meas.* **29**, 1319–1334 (2008)
2. Aster, R.C., Borchers, B., Thurber, C.H.: *Parameter Estimation and Inverse Problems*, 2nd edn. Elsevier, Amsterdam (2013)
3. Bakushinskii, A.B.: Remarks on choosing a regularization parameter using the quasi-optimality and ratio criterion. *USSR Comput. Math. Math. Phys.* **24**(4), 181182 (1984)
4. Bauer, F., Lukas, M.A.: Comparing parameter choice methods for regularization of ill-posed problems. *Math. Comput. Simul.* **81**, 1795–1841 (2011)
5. Björck, A.: *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics, Philadelphia (1996)
6. Chung, J., Nagy, J.G., O’Leary, D.P.: A weighted GCV method for Lanczos hybrid regularization. *Electron. Trans. Numer. Anal.* **28**, 149–167 (2008)
7. Drineas, P., Mahoney, M.W.: RandNLA: randomized numerical linear algebra. *Commun. ACM* **59**, 80–90 (2016)
8. Drineas, P., Mahoney, M.W., Muthukrishnan, S., Sarlós, T.: Faster least squares approximation. *Numer. Math.* **117**, 219–249 (2011)
9. Fenu, C., Reichel, L., Rodrigues, G., Sadok, H.: GCV for Tikhonov regularization by partial SVD. *BIT Numer. Math.* **57**, 1019–1039 (2017)
10. Gazzola, S., Hansen, P.C., Nagy, J.G.: IR tools: a MATLAB package of iterative regularization methods and large-scale test problems. *Numer. Algorithms*. <https://doi.org/10.1007/s11075-018-0570-7> (2018)
11. Golub, G.H., Heath, M., Wahba, G.: Generalized cross-validation as a method for choosing a good ridge parameter. *Technometrics* **21**, 215–223 (1979)
12. Golub, G.H., van Loan, C.F.: *Matrix Computations*, 3rd edn. Johns Hopkins Press, Baltimore (1996)
13. Gower, R.M., Richtárik, P.: Randomized iterative methods for linear systems. *SIAM J. Matrix Anal. Appl.* **36**, 1660–1690 (2015)
14. Hämäläinen, K., Harhanen, L., Kallonen, A., Kujanpää, A., Niemi, E., Siltanen, S.: Tomographic X-ray data of a walnut. [arXiv:1502.04064](https://arxiv.org/abs/1502.04064) (2015)
15. Hämarik, U., Palm, R., Raus, T.: On minimization strategies for choice of the regularization parameter in ill-posed problems. *Numer. Funct. Anal. Optim.* **30**, 924–950 (2009)
16. Hämarik, U., Palm, R., Raus, T.: A family of rules for parameter choice in Tikhonov regularization of ill-posed problems with inexact noise level. *J. Comput. Appl. Math.* **236**, 2146–2157 (2012)
17. Hanke, M., Hansen, P.C.: Regularization methods for large-scale problems. *Surv. Math. Ind.* **3**, 253–315 (1993)
18. Hansen, J.K., Hogue, J.D., Sander, G.K., Renaut, R.A., Popat, S.C.: Non-negatively constrained least squares and parameter choice by the residual periodogram for the inversion of electrochemical impedance spectroscopy data. *J. Comput. Appl. Math.* **278**, 52–74 (2015)
19. Hansen, P.C.: The discrete Picard condition for discrete ill-posed problems. *BIT Numer. Math.* **30**, 658–672 (1990)
20. Hansen, P.C.: Regularization tools—a Matlab package for analysis and solution of discrete ill-posed problems. *Numer. Algorithms* **46**, 189–194 (1994)
21. Hansen, P.C.: *Rank-Deficient and Discrete Ill-Posed Problems*. Society for Industrial and Applied Mathematics, Philadelphia (1998)

22. Hansen, P.C.: Discrete Inverse Problems. Society for Industrial and Applied Mathematics, Philadelphia (2010)
23. Hofmann, B.: Regularization for Applied Inverse and Ill-posed Problems: A Numerical Approach, Teubner-Texte zur Mathematik. Teubner, B.G., Berlin (1986)
24. Levin, E., Meltzer, A.Y.: Estimation of the regularization parameter in linear discrete ill-posed problems using the Picard parameter. *SIAM J. Sci. Comput.* **39**, A2741–A2762 (2017)
25. Lin, Y., Wohlberg, B., Guo, H.: UPRE method for total variation parameter selection. *Signal Process.* **90**, 2546–2551 (2010)
26. Mahoney, M.W.: Randomized algorithms for matrices and data. *Found. Trends® Mach. Learn.* **3**, 123–224 (2011)
27. Mead, J.L., Renaut, R.A.: A Newton root-finding algorithm for estimating the regularization parameter for solving ill-conditioned least squares problems. *Inverse Probl.* **25**, 025002 (2009)
28. Meng, X., Saunders, M.A., Mahoney, M.W.: LSRN: a parallel iterative solver for strongly over- or underdetermined systems. *SIAM J. Sci. Comput.* **36**, C95–C118 (2014)
29. Morozov, V.A.: On the solution of functional equations by the method of regularization. *Sov. Math. Dokl.* **7**, 414–417 (1966)
30. Renaut, R.A., Horst, M., Wang, Y., Cochran, D., Hansen, J.: Efficient estimation of regularization parameters via downsampling and the singular value expansion. *BIT Numer. Math.* **57**, 499–529 (2017)
31. Renaut, R.A., Vatankehah, S., Ardestani, V.E.: Hybrid and iteratively reweighted regularization by unbiased predictive risk and weighted GCV for projected systems. *SIAM J. Sci. Comput.* **39**, B221–B243 (2017)
32. Rokhlin, V., Tygert, M.: A fast randomized algorithm for overdetermined linear least-squares regression. *Proc. Natl. Acad. Sci.* **105**, 13212–13217 (2008)
33. Stein, C.M.: Estimation of the mean of a multivariate normal distribution. *Ann. Stat.* **9**, 1135–1151 (1981)
34. Taroudaki, V., O’Leary, D.P.: Near-optimal spectral filtering and error estimation for solving ill-posed problems. *SIAM J. Sci. Comput.* **37**, A2947–A2968 (2015)
35. Toma, A., Sixou, B., Peyrin, F.: Iterative choice of the optimal regularization parameter in tv image restoration. *Inverse Probl. Imaging* **9**, 1171 (2015)
36. Vatankehah, S., Ardestani, V.E., Renaut, R.A.: Automatic estimation of the regularization parameter in 2D focusing gravity inversion: application of the method to the Safa manganese mine in the northwest of Iran. *J. Geophys. Eng.* **11**, 045001 (2014)
37. Vatankehah, S., Ardestani, V.E., Renaut, R.A.: Application of the  $\chi^2$  principle and unbiased predictive risk estimator for determining the regularization parameter in 3-D focusing gravity inversion. *Geophys. J. Int.* **200**, 265–277 (2015)
38. Vatankehah, S., Renaut, R.A., Ardestani, V.E.: 3-D projected  $\ell_1$  inversion of gravity data using truncated unbiased predictive risk estimator for regularization parameter estimation. *Geophys. J. Int.* **210**, 1872–1887 (2017)
39. Vatankehah, S., Renaut, R.A., Ardestani, V.E.: A fast algorithm for regularized focused 3-D inversion of gravity data using the randomized SVD. *Geophysics* (2018)
40. Vatankehah, S., Renaut, R.A., Ardestani, V.E.: Total variation regularization of the 3-D gravity inverse problem using a randomized generalized singular value decomposition. *Geophys. J. Int.* **213**, 695–705 (2018)
41. Vogel, C.: Computational Methods for Inverse Problems. Society for Industrial and Applied Mathematics, Philadelphia (2002)

## Affiliations

**Rosemary A. Renaut<sup>1</sup>**  · **Anthony W. Helmstetter<sup>1</sup>** · **Saeed Vatankhah<sup>2</sup>**

Anthony W. Helmstetter  
Anthony.Helmstetter@asu.edu

Saeed Vatankhah  
svatan@ut.ac.ir

<sup>1</sup> School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287-1804, USA

<sup>2</sup> Institute of Geophysics, University of Tehran, Tehran, Iran