Application of denoising methods to regularization of ill-posed problems

Tristan A. Hearn* Lothar Reichel†

Abstract

Linear systems of equations and linear least-squares problems with a matrix whose singular values "cluster" at the origin and with an error-contaminated data vector arise in many applications. Their numerical solution requires regularization, i.e., the replacement of the given problem by a nearby one, whose solution is less sensitive to the error in the data. The amount of regularization depends on a parameter. When an accurate estimate of the norm of the error in the data is known, this parameter can be determined by the discrepancy principle. This paper is concerned with the situation when the error is white Gaussian and no estimate of the norm of the error is available, and explores the possibility of applying a denoising method to both reduce this error and to estimate its norm. Applications to image deblurring are presented.

1 Introduction

We consider the approximate solution of linear least-squares problems

$$\min_{x\in\mathbb{R}^n} \|Ax-b\|_2, \tag{1}$$

with a matrix $A\in\mathbb{R}^{m\times n}$ whose singular values "cluster" at the origin. In particular, A is severely ill-conditioned and may be singular. Such least-squares problems arise from the discretization of Fredholm integral equations of the first kind and, in particular, in image deblurring problems. They are commonly referred to as discrete ill-posed problems. For notational simplicity, we will assume that $m\geq n$, however, this constraint easily can be removed. Throughout this paper $\|\cdot\|_2$ denotes the Euclidean vector norm or the spectral matrix norm.

*Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail: tristan.a.hearn@nasa.gov.

†Department of Mathematical Sciences, Kent State University, Kent, OH 44242, USA. E-mail: reichel@math.kent.edu. Research supported in part by NSF grant DMS-1115385.

The vector $b \in \mathbb{R}^m$ in (1) represents available data and is assumed to be contaminated by an additive error $e \in \mathbb{R}^m$, i.e.,

$$b = b_t + e,$$

where $b_t \in \mathbb{R}^n$ denotes the unknown error-free data. Such error contamination can occur if b is recorded by a process subject to measurement and transmission errors. The error e is assumed to be white Gaussian in this paper. We sometimes will refer to e as "noise". The least-squares problem (1) with b replaced by b_t is assumed to be consistent, but consistency of the available problem (1) is not required.

Let $A^\dagger$ denote the Moore-Penrose pseudoinverse of A. Then the solution of the least-squares problem (1) can be expressed as

$$x = A^\dagger b = A^\dagger b_t + A^\dagger e = \hat{x} + A^\dagger e, \tag{2}$$

where $\hat{x} = A^\dagger b_t$ is the least-square solution of minimal Euclidean norm of the error-free problem associated with (1). We would like to determine an accurate approximation of $\hat{x}$. Due to the ill-conditioning of A and the error e in b, the vector (2) typically is dominated by the propagated error $A^\dagger e$, and therefore not a useful approximation of $\hat{x}$; see Section 2.1 for further details. Regularization aims to remedy this situation.

Regularization methods replace the given problem (1) by a nearby one, whose solution is less sensitive to the error e in b. Popular regularization methods include the Truncated Singular Value Decomposition (TSVD), Tikhonov regularization, and truncated iteration. In the TSVD method all but the k largest singular values of the matrix A are set to zero, and the solution of minimal Euclidean norm of the modified least-squares problem so obtained is computed. The parameter k is referred to as a regularization parameter and determines how much the given problem (1) is modified and how sensitive the solution of the modified problem is to the error e. Tikhonov regularization replaces the problem (1) by a penalized least-squares problem, and a regularization parameter determines how much the given problem (1) is modified. Finally, in truncated iteration regularization is achieved by carrying out sufficiently few iterations with an iterative solution method. The number of iterations is the regularization parameter. We will illustrate the performance of these regularization methods when the regularization parameter is determined by a novel denoising method.

Let $\hat{x}_{\text{Reg.}}(p)$ denote the approximation of $\hat{x}$ determined by a regularization method with regularization parameter p and assume that $\|A\hat{x}_{\text{Reg.}}(p) - b\|_2$ decreases as p increases. If a fairly tight bound for the norm of the error in b,

$$\|e\|_2 \leq \delta,$$

is available, then it is convenient to apply the discrepancy principle to determine p. The discrepancy principle selects the smallest value of the regularization parameter p such that

$$\|A\hat{x}_{\text{Reg.}}(p) - b\|_2 \leqslant \tau\delta, \tag{3}$$

where $\tau \geq 1$ is a user-chosen parameter independent of the norm of e; see, e.g., Engl et al. [9] and Hansen [12] for discussions. The main drawback of the discrepancy principle is that a fairly accurate estimate of $\|e\|_2$ is required in order for this parameter choice method to determine a value of p that yields an as accurate approximation $\hat{x}_{\text{Reg.}}(p)$ of $\hat{x}$ as possible. A significant overestimate of $\|e\|_2$ results in an unnecessarily large value of the regularization parameter and, therefore, in a needlessly large error $\hat{x}_{\text{Reg.}}(p) - \hat{x}$; conversely a significant underestimate of $\|e\|_2$ gives a computed solution that is severely contaminated by propagated error.

It is the purpose of the present paper to investigate the application of denoising methods both for improving the data used in the computation and for estimating the norm of the error in the data. We present applications to digital image deblurring problems. Three data driven denoising methods are considered for this purpose: the wavelet-shrinkage denoising algorithm BayesShrink [3], the wavelet-shrinkage-based Residual Kurtosis Minimization Denoising (RKMD) method described in [15], and a Total-Variation (TV) denoising method that uses RKMD to determine the regularization parameter. We are interested both in denoising the data b without removing significant features of the image and in determining an accurate estimate of the norm of the error.

Another approach to the estimation of $\|e\|_2$ is described in [21, 22], where an evolution equation associated with a regularization method based on the Perona-Malik diffusivity is integrated a few time steps. This removes the highest frequencies in b, and the difference between the integrated data and b is used as an estimate of $\|e\|_2$. Computed results reported in [21, 22] are promising, but the integrated data is not an accurate approximation of the error-free data b_t; the integrated data generally cannot replace the vector b in (1) without loss of accuracy. The BayesShrink method has recently been applied in [23] to estimate the noise in the data on each level of an alternating multilevel method. As we will see in Section 5, the BayesShrink method does not always determine $\|e\|_2$ accurately.

Methods for determining a suitable value of the regularization parameter without using a bound for $\|e\|_2$ are commonly referred to as heuristic; see, e.g., [2, 9, 14, 18, 20, 27] for discussions and illustrations. All of these methods implicitly or explicitly estimate $\|e\|_2$, but none of them seek to remove the noise to determine more accurate data for the deblurring problem.

Let p be a value of the regularization parameter determined by a heuristic method and let $\hat{x}_{\text{Reg.}}(p)$ be the associated approximate solution of (1). Then the discrepancy principle suggests that $\|A\hat{x}_{\text{Reg.}}(p) - b\|_2$ is an estimate of the norm of e. If this estimate is close to but not smaller than $\|e\|_2$, then $\hat{x}_{\text{Reg.}}(p)$ is likely to be a useful approximation of $\hat{x}$; however, if the estimate is much smaller than $\|e\|_2$, then $\hat{x}_{\text{Reg.}}(p)$ might not be close to $\hat{x}$.

This paper is organized as follows. Section 2 gives a short review of the regularization methods applied in the computed examples. They are TSVD, Tikhonov regularization, and truncated iteration with a minimal residual method [25]. In Section 3, we briefly review the L-curve criterion for determining the regularization parameter p in the regularization methods considered. The L-curve criterion was first proposed by Lawson and Hanson [19], and has subsequently been advocated by Hansen [11, 12]. We will compare approaches for determining p based on denoising methods to the L-curve criterion. Section 4 discusses denoising by the BayesShrink method, TV-norm minimization, and the RKMD data-driven procedure. Numerical examples that illustrate the application of these denoising methods to the determination of the regularization parameter for the regularization methods described in Section 2 via the discrepancy principle and compares them to the L-curve criterion are presented in Section 5. Concluding remarks can be found in Section 6.

2 Regularization methods

This section provides an overview of the TSVD, Tikhonov, and minimal residual methods for computing regularized solution to discrete ill-posed problems.

2.1 TSVD

The Singular Value Decomposition (SVD) of a matrix $A \in \mathbb{R}^{m\times n}$ is a factorization of the form

$$A = U\Sigma V^T,$$

where

$$\Sigma = \mathrm{diag}\,[\sigma_1, \sigma_2, \ldots, \sigma_n] \in \mathbb{R}^{m\times n}$$

is a (possibly rectangular) diagonal matrix whose diagonal entries $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n \geq 0$ are the singular values of A, and $U = [u_1, u_2, \ldots, u_m] \in \mathbb{R}^{m\times m}$ and $V = [v_1, v_2, \ldots, v_n] \in \mathbb{R}^{n\times n}$ are orthogonal matrices. The superscript ^{T} denotes transposition. An optimal rank-p approximation A_p of A in the spectral matrix norm $\|\cdot\|_2$ is given by

$$A_p = \sum_{i=1}^{p} \sigma_j u_j v_j^T.$$

The pseudoinverse $A_p^\dagger$ of A_p can be expressed as

$$A_p^\dagger = \sum_{i=1}^{p} \frac{v_i u_i^T}{\sigma_i},$$

where we assume that p is chosen small enough so that $\sigma_p > 0$. The solution of minimal Euclidean norm of the least-squares problem (1) with A replaced by A_p is given by

$$\hat{x}_{\mathrm{TSVD}} = A_p^\dagger b = A_p^\dagger (b_t + e) = \sum_{i=1}^{p} \left(\frac{u_i^T b_t}{\sigma_i} v_i + \frac{u_i^T e}{\sigma_i} v_i \right). \tag{4}$$

If A has tiny positive singular values, then the terms $\frac{u_i^T e}{\sigma_i} v_i$ in (4) associated with these singular values may dominate the other terms and give rise to a large propagated error; see, e.g., [12] for a discussion. Regularization by TSVD resolves this issue by omitting terms associated with the smallest singular values. The integer p is a regularization parameter. In order for $\hat{x}_{\text{Reg.}}(p) = \hat{x}_{\text{TSVD}}$ to be a meaningful approximate solution of (1), p must be chosen small enough so that $\hat{x}_{\text{TSVD}}$ is not dominated by propagated error, yet large enough that fidelity to the underlying problem is maintained.

2.2 Tikhonov regularization

In its simplest form, Tikhonov regularization replaces (1) by the penalized least-squares problem

$$\min_{x\in\mathbb{R}^n} \left\{ \|Ax - b\|_2^2 + \frac{1}{p} \|x\|_2^2 \right\}, \tag{5}$$

where $p > 0$ is a regularization parameter. This parameter, much like the regularization parameter p in TSVD, acts as a selectable trade-off between smoothness of the solution and fidelity to the underlying problem (1). The solution of (5) can be determined inexpensively for many values of p when the SVD of A is available

$$\hat{x}_{\text{Reg.}}(p) = \hat{x}_{\text{Tik.}} = V\left(\Sigma^T\Sigma + \frac{1}{p}I\right)^{-1} \Sigma^T U^T b = \sum_{i=1}^{n} \frac{\sigma_i}{\sigma_i^2 + 1/p} \left(u_i^T b\right) v_i. \tag{6}$$

Alternative implementations are discussed in [1, 8, 16].

2.3 A minimal residual iterative method

In this subsection, we assume the matrix $A \in \mathbb{R}^{n\times n}$ in (1) to be symmetric. It is well known that iterative methods for the solution of minimization problems (1) with a symmetric matrix should determine an approximate solution in the range of A, because then the approximate solution is orthogonal to the null space of A. We will use the implementation described in [25, Algorithm 3.1] of a minimal residual Krylov subspace method with this property in some of the examples of Section 5 using the software [26]. Thus, beginning with the initial iterate $x_0 = 0$, the p^{th} iterate x_p determined by this method satisfies

$$\|Ax_p - b\|_2 = \min_{x\in\mathbb{K}_p(A,Ab)} \|Ax - b\|_2, \qquad x_p \in \mathbb{K}_p(A, Ab),$$

where $K_p(A, Ab) = \text{span}\{Ab, A^2b, \ldots, A^pb\}$ is a Krylov subspace in the range of A. We assume that p is chosen small enough so that $\dim(K_p(A, Ab)) = p$.

The use of this iterative method requires that the number of iterations, p, be specified. This parameter is a regularization parameter, i.e., $\hat{x}_{\text{Reg.}}(p) = x_p$.

When p is (too) small, the number of iterations may not be sufficient to accurately approximate the desired solution $\hat{x}$, while if p is (too) large, then the computed approximate solution x_p may be dominated by propagated error.

3 The L-curve criterion

The L-curve criterion determines the regularization parameter p by balancing the sizes of the computed solution $\hat{x}_{\text{Reg.}}(p)$ and of the associated residual error

$$\hat{r}_{\text{Reg.}}(p) = b - A\hat{x}_{\text{Reg.}}(p).$$

It can be applied with all regularization methods of Section 2. We briefly review the L-curve criterion and first consider its application to the determination of the regularization parameter in Tikhonov regularization. Thus, $\hat{x}_{\text{Reg.}}(p)$ is given by (6). The L-curve criterion considers the curve

$$p \to \{\ln \|x_{\text{Reg.}}(p)\|_2, \ln \|r_{\text{Reg.}}(p)\|_2\}. \tag{7}$$

This curve is typically L-shaped in a vicinity of suitable values of the regularization parameter p and therefore is referred to as the L-curve. One selects the value $p = p_L$ that corresponds to a point of largest curvature in the L-shaped part of the curve. We refer to this point as the "vertex" of the L-curve. The required computations are inexpensive when the SVD of the matrix A is available. Software for determining p_L is provided in Regularization Tools [13]. For many linear discrete ill-posed problems, the point $\{\ln \|x_{\text{Reg.}}(p_L)\|_2, \ln \|r_{\text{Reg.}}(p_L)\|_2\}$ is near the point of the L-curve (7) closest to the origin.

The L-curve criterion also can be applied in conjunction with the TSVD and minimal residual iterative methods. Then the regularization parameter achieves integer values and the analogue of the L-curve (7) is a discrete point set, which we refer to as the "discrete L-curve". We will chose $p = p_L$ so that the point $\{\ln \|x_{\text{Reg.}}(p_L)\|_2, \ln \|r_{\text{Reg.}}(p_L)\|_2\}$ is closest to the origin among the points that make up the discrete L-curve. This point is easy to determine. When using the minimal residual iterative method, we terminate the iterations within a few steps after the function $p \to \ln \|x_{\text{Reg.}}(p)\|_2 + \ln \|r_{\text{Reg.}}(p)\|_2$ does not decrease any more. The point $\{\ln \|x_{\text{Reg.}}(p_L)\|_2, \ln \|r_{\text{Reg.}}(p_L)\|_2\}$ is for many discrete ill-posed problems at or close to a "vertex" determined by visual inspection.

4 Noise level estimation via denoising

This section outlines three parameter-free data-driven denoising methods. They will be applied both to denoise the available data b and to estimate the norm of the noise in b.

4.1 BayesShrink

BayesShrink is a level-adaptive wavelet-shrinkage denoising method [3]. The method first applies a multi-level discrete wavelet transform to the noisy data. The BayesShrink method assumes that the coefficients of each detail sub-band of the noise-free data conform to a generalized Gaussian distribution, i.e., they have a density function of the form

$$G_{\sigma_X,\beta}(t) = C\left(\sigma_X,\beta\right)\exp\left\{-\left[\alpha\left(\sigma_X,\beta\right)|t|\right]^{\beta}\right\}$$

with support $-\infty < t < \infty$ and parameters $\sigma_X > 0$ and $\beta > 0$. Here X denotes the noise-free signal and σ_X the standard deviation of X. The functions $C\left(\sigma_X,\beta\right)$ and $\alpha\left(\sigma_X,\beta\right)$ are given by

$$C\left(\sigma_X,\beta\right) = \tfrac{\beta\alpha(\sigma_X,\beta)}{2\Gamma(\beta^{-1})} \quad \text{and} \quad \alpha\left(\sigma_X,\beta\right) = \sigma_X^{-1}\sqrt{\tfrac{\Gamma(3/\beta)}{\Gamma(1/\beta)}}.$$

The BayesShrink method proceeds by soft-thresholding each wavelet detail sub-band i using the threshold

$$\lambda_i^* = \frac{\hat{\sigma}_e^2}{\hat{\sigma}_{b,i}},$$

where $\hat{\sigma}_{b,i}$, an estimate of the signal variance in sub-band i, is computed according to

$$\hat{\sigma}_{b,i} = \sqrt{\max\left\{\hat{\sigma}_i^2 - \hat{\sigma}_e^2, 0\right\}}.$$

The BayesShrink method requires an a priori estimate of the standard deviation of the noise $\hat{\sigma}_e$. Chang [3] proposed to estimate $\hat{\sigma}_e$ by using the median of the wavelet coefficients in the first level HH_1 sub-band,

$$\hat{\sigma}_e = \frac{\text{median}\,|HH_1|}{0.6745}. \tag{8}$$

This estimate is commonly used as an a priori noise-level estimator in wavelet denoising methods; see [3, 6, 7]. We used the CDF 9/7 wavelet with four levels of decomposition in the numerical examples reported in Section 5.

4.2 Residual kurtosis minimization denoising

Residual kurtosis minimization denoising (RKMD) is a new data-driven methodology for reducing Gaussian noise in given data b [15]. Let a denoising method be parametrized by the scalar t. The RKMD method determines denoised data $\breve{b}\left(t\right)$ by choosing t so that the normalized residual kurtosis

$$\kappa_2\left(r\left(t\right)\right) = \sqrt{\frac{mn}{24}}\left[mn\left(\frac{\|r\left(t\right)\|_4}{\|r\left(t\right)\|_2}\right)^4 - 3\right]$$

of the residual vector

$$r(t) = \breve{e}(t) - \text{mean}(\breve{e}(t))$$

is minimized, where

$$\breve{e}(t) = b - \breve{b}(t).$$

The reasoning behind the RKMD methodology is as follows. Consider a random normal sample $X = (x_1, \ldots, x_n) \in \mathbb{R}^n$, where $x_1, \ldots, x_n \overset{\text{i.i.d}}{\sim} N(0, \sigma^2)$. The kurtosis κ_1 of X is a sample statistic defined as

$$\kappa_1 = \frac{m_4}{m_2^2},$$

where m_4 and m_2 are the sample moments

$$m_4 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^4, \quad m_2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2.$$

Based on earlier work by Fisher [10], it has been shown by D'Agostino, Pearson, and Tietjen [4, 5] that in the case of normal samples, a normalized statistic for kurtosis is given by

$$\kappa_2 = \sqrt{\frac{n}{24}}\left(\frac{m_4}{m_2^2} - 3\right) \overset{\text{distr.}}{\longrightarrow} N(0, 1) \tag{9}$$

with convergence as $n \to \infty$.

Let Φ be the probability that a value of κ_2 (or a value further away from zero) is attained in the normal sample X. Interpreted as a measure of normality of the sample X, this probability can be computed as

$$\Phi = 2P(y > |\kappa_2|) = 2\int_{|\kappa_2|}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}} dy = 1 - \text{erf}\left(\frac{|\kappa_2|}{\sqrt{2}}\right),$$

where y is a standard normal random variable $y \sim N(0, 1)$ and $\text{erf}(x)$ denotes the Gauss error function. This probability can be computed using standard tables or by numerical quadrature.

We should mention that the robustness of using the kurtosis statistic as a measure of normality is dependent on the sample size. This has been observed in the contexts of other types of mathematical problems. For instance, consider the problem of linear blind source separation of either the columns or rows of a matrix $A \in \mathbb{R}^{m \times n}$. This problem is solved within the framework of independent component analysis (ICA) by decreasing a selected numerical measure of normality of the columns $a_1, \ldots, a_n$ of A or of the rows $a_1^T, \ldots, a_m^T$ of A. It has been noted, see, e.g., [17], that kurtosis may not be a robust enough measure of normality for the purpose of ICA when the columns or rows of A do not have sufficiently many entries.

In the present paper, we assume that the sample distribution of κ_2 is accurately characterized by the asymptotic distribution in (9). Then kurtosis is a robust measure of normality. The justification for this assumption is that the sample sizes, i.e., the number of pixels of digital images to be denoised, typically is large, even when the image is represented by a fairly small number of pixels in the horizontal and vertical directions. For instance, a digital image comprised of 64×64 pixels, yields a sample size of $64^2 = 4096$ entries. Compared to the sample sizes considered in [4, 5] (from 30 to 200), this sample size is large. Moreover, many images that arise in applications are represented by many more than 64×64 pixels.

We will apply the kurtosis statistic to all of the components of the residual image, following an application of a denoising procedure parametrized by t. Then, clearly, Φ depends on this value. When a value of t is found that maximizes Φ, this value maximizes the likelihood of normality of the residual $\breve{e}$. This observation is the basis for our denoising method.

The RKMD approach has in [15] been successfully paired with denoising via wavelet coefficient thresholding using soft-shrinkage operators. The soft-shrinkage thresholds are chosen to minimize $\kappa_2(r(t))$ for each wavelet decomposition level, while keeping the thresholds optimized for previous levels fixed. Thus, there is a scalar parameter t for each wavelet decomposition level. The implementation of the RKMD method used for the numerical examples of Section 5 applies the Newton-CG method [24] for the minimization of $\kappa_2\left(r\left(t\right)\right)$ on each wavelet decomposition level.

4.3 Total variation minimization

In Total Variation (TV) image denoising a restoration $\breve{F}^{TV}$ is computed by minimizing a discretization of the functional

$$\breve{b}^{TV} = \breve{F}^{TV}\left(b;\lambda\right) = \underset{\breve{b}\in\mathbb{R}^{m\times n}}{\arg\min}\left\{\left\|b-\breve{b}\right\|_2 + \lambda\int\limits_{\Omega}\left|\nabla\breve{b}\right|d\Omega\right\} \tag{10}$$

for some parameter $\lambda > 0$. Here Ω is the rectangular image domain. In computations a discrete version of the integral in (10) is evaluated. The minimum is typically computed with the aid of an iterative method applied to the Euler-Lagrange equation associated with (10); see, e.g., [28, 29].

TV-norm-based denoising methods effectively smooth noise within "flat" regions of a digital image while preserving edges. They require a value of the parameter λ in (10) to be chosen. For a fair comparison with the RKMD method and BayesShrink, the implementation of the TV-norm-based denoising method should be entirely data-driven. To this end, we determine a value $\lambda = \lambda^*$ that minimizes the normalized residual kurtosis by the Newton-CG method [24]. Thus,

$$\lambda^* = \arg\min_{\lambda}\left|k_2\left(b-\breve{b}^{TV}\left(b,\lambda\right)\right)\right|.$$

This is an extension of the RKMD methodology beyond the framework of wavelet coefficient thresholding.

5 Numerical examples

We present a few numerical examples where we restore images that have been contaminated by white Gaussian noise and Gaussian blur. First the image is denoised by one of the methods of Section 4. Besides reducing the noise in the given image, the denoising methods provide estimates of the norm of the noise in the image. These estimates allow us to use the discrepancy principle when solving the discrete ill-posed deblurring problems. Each image to be restored is a digital gray-scale image that has been contaminated by Gaussian blur and white Gaussian noise.

Let $\hat{X} \in \mathbb{R}^{m\times n}$ denote the unknown uncontaminated desired gray-scale image. This image is represented by an array of $m \times n$ pixels. Our model for image contamination is

$$B = B_t + E = K * \hat{X} + E,$$

where $K \in \mathbb{R}^{m\times n}$ is a known kernel of the blurring operation (often referred to as a point spread function (PSF)) and $E \in \mathbb{R}^{m\times n}$ is a matrix whose entries are white Gaussian noise. The matrix $B_t \in \mathbb{R}^{m\times n}$ represents the unavailable blurred and noise-free image, and $B \in \mathbb{R}^{m\times n}$ is the recorded blurred and noisy image. The operation $*$ denotes discrete convolution. We have

$$[B_t]_{i,j} = F * K = \sum_{h=1}^{m} \sum_{l=1}^{n} F_{h,l} K_{i-h,j-l}, \qquad 1 \leq i \leq m, \quad 1 \leq j \leq n.$$

Convolution is a linear operation on $\hat{X}$ and can be written as

$$\mathrm{vec}\,(B_t) = A\,\mathrm{vec}\left(\hat{X}\right),$$

where the matrix A encodes the discrete convolution operation, and $\mathrm{vec}\,(B_t)$ and $\mathrm{vec}\left(\hat{X}\right)$ are column vectors formed by successively stacking the columns of the matrices B_t and $\hat{X}$, respectively. The singular values of A typically "cluster" at the origin. Let

$$b = \mathrm{vec}\,(B) = \begin{bmatrix} B_1 \\ B_2 \\ \vdots \\ B_n \end{bmatrix} = \begin{bmatrix} B_{11} \\ \vdots \\ B_{m1} \\ \vdots \\ B_{mn} \end{bmatrix}, \quad \hat{x} = \mathrm{vec}\left(\hat{X}\right) = \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \\ \vdots \\ \hat{x}_n \end{bmatrix} = \begin{bmatrix} \hat{f}_{11} \\ \vdots \\ \hat{f}_{m1} \\ \vdots \\ \hat{f}_{mn} \end{bmatrix}.$$

Figure 1: The original noise-free *Kent* and *Barbara* images.

Figure 2: Blurred and noisy *Kent* and *Barbara* images.

Thus, the problem of determining an approximation of $\hat{X}$ from B and K, using the matrix A and the vector b, is a discrete ill-posed problem of the form (1).

The digital images used in the numerical experiments were the *Kent* and *Barbara* images shown in Figure 1. Each image was blurred using a Gaussian PSF

$$K_{i,j} = \exp\left\{-\frac{1}{2\theta^2}\left[\left(i-\frac{m}{2}\right)^2+\left(j-\frac{n}{2}\right)^2\right]\right\}$$

with scale parameter $\theta = 1.25$, and contaminated with 5% additive white Gaussian noise, i.e.,

$$\frac{\|A\hat{x}-b\|_2}{\|A\hat{x}\|_2} = \frac{\|e\|_2}{\|A\hat{x}\|_2} = 0.05.$$

The blurred and noisy images are shown in Figure 2.

5.1 Results

The blurred and noisy images were restored using TSVD, Tikhonov regularization, and the minimal residual iterative method described in Subsection 2.3. For each restoration, the regularization parameter was selected with the aid of the discrepancy principle (3) with $\tau = 1$, where an estimate δ of the norm of the error e in b is determined with TV-denoising, BayesShrink, and

Method	Image	PSNR_B	$\mathrm{PSNR}_{\breve{X}_L}$	$\mathrm{PSNR}_{\breve{X}_{opt}}$	err_δ	err_p	$\mathrm{PSNR}_{\breve{X}}$	$\mathrm{PSNR}_{\breve{X}(\breve{B})}$
TV denoising	Kent	28.59	31.17	31.32	29.24	4.15	27.06	25.23
	Barbara	27.67	28.29	28.98	14.61	3.128	26.81	25.94
BayesShrink	Kent	28.59	31.17	31.32	0.1273	0.333	30.19	31.88
	Barbara	27.67	28.29	28.98	0.1702	0.6	28.29	29.5
RKMD	Kent	28.59	31.17	31.32	0.017	0.0	31.32	32.84
	Barbara	27.67	28.29	28.98	0.0812	0.0	28.98	29.79

Table 1: Tabulation of the err_δ, err_p, and PSNR restoration values for the *Kent* and *Barbara* images, using TSVD.

Method	Image	PSNR_B	$\mathrm{PSNR}_{\breve{X}_L}$	$\mathrm{PSNR}_{\breve{X}_{opt}}$	err_δ	err_p	$\mathrm{PSNR}_{\breve{X}}$	$\mathrm{PSNR}_{\breve{X}(\breve{B})}$
TV denoising	Kent	28.59	31.66	31.83	29.24	5.187	28.06	27.23
	Barbara	27.67	29.19	29.65	14.61	3.254	27.76	27.62
BayesShrink	Kent	28.59	31.66	31.83	0.1873	0.1267	31.19	31.96
	Barbara	27.67	29.19	29.65	0.1202	0.1042	29.08	29.81
RKMD	Kent	28.59	31.66	31.83	0.0275	0.0132	31.82	33.02
	Barbara	27.67	29.19	29.65	0.0812	0.0293	29.63	30.82

Table 2: Tabulation of the err_δ, err_p, and PSNR restoration values for the *Kent* and *Barbara* images, using Tikhonov regularization.

RKMD. For BayesShrink and RKMD, a four-level decomposition with the CDF 9/7 wavelet was used.

We tabulate the Peak-Signal-to-Noise-Ratio of the given blurred and noise-contaminated image B defined by

$$\mathrm{PSNR}_B = 20 \log_{10} \frac{255}{\|B - \hat{X}\|_F}. \tag{11}$$

Here $\|\cdot\|_F$ denotes the Frobenius norm. Thus, $\|B - \hat{X}\|_F = \|b - \hat{x}\|_2$. The numerator $255 = 2^8 - 1$ of (11) stems from the fact that every pixel is represented by 8 bits. The PSNR-value increases with the quality of restoration.

We first apply the L-curve criterion outlined in Section 3. It yields restored images that we denote by $\breve{X}_L$. The columns labeled $\mathrm{PSNR}_{\breve{X}_L}$ of Tables 1, 2, and 3 show the PSNR-values for the images $\breve{X}_L$ determined by TSVD, Tikhonov regularization, and truncated iteration with the minimal residual method of Subsection 2.3, respectively. The PSNR-values for the restored

Method	Image	PSNR_B	$\mathrm{PSNR}_{\breve{X}_L}$	$\mathrm{PSNR}_{\breve{X}_{opt}}$	err_δ	err_p	$\mathrm{PSNR}_{\breve{X}}$	$\mathrm{PSNR}_{\breve{X}(\breve{B})}$
TV denoising	Kent	28.59	31.72	31.98	29.24	15.35	27.44	22.43
	Barbara	27.67	28.79	29.02	14.61	10.65	26.05	24.10
BayesShrink	Kent	28.59	31.72	31.98	0.1273	0.349	31.32	32.58
	Barbara	27.67	28.79	29.02	0.1702	0.223	28.72	29.97
RKMD	Kent	28.59	31.79	31.98	0.017	0.0	32.06	34.09
	Barbara	27.67	28.79	29.02	0.0812	0.0	29.63	30.88

Table 3: Tabulation of the err_δ, err_p, and PSNR restoration values for the *Kent* and *Barbara* images, using a minimal residual iterative method.

images $\breve{X}_L$ are larger than for the corresponding blurred and noisy image B. The increase is particularly large for the image *Kent*.

Next we consider application of the discrepancy principle using the exact norm of the noise E in B, i.e.,

$$\delta = \|E\|_F. \tag{12}$$

The restored images using TSVD, Tikhonov regularization, and truncated iteration described in Subsection 2.3 are denoted by $\breve{X}_{opt}$ and the columns labeled $\mathrm{PSNR}_{\breve{X}_{opt}}$ of Tables 1-3 display the PSNR-values for the restored images so obtained. In all tables the images $\breve{X}_{opt}$ have a somewhat larger PSNR-value than the corresponding images $\breve{X}_L$. Thus, the restorations obtained with the discrepancy principle are of somewhat higher quality than those determined with the L-curve criterion.

We now apply each denoising method of Section 4 to determine a denoised image, which we refer to as $\breve{B}$. The quantity

$$\breve{\delta} = \|B - \breve{B}\|_F \tag{13}$$

provides an estimate of the noise in B. The relative error in this estimate is

$$\mathrm{err}_\delta = \frac{|\delta - \breve{\delta}|}{\delta}.$$

For Tikhonov regularization, the discrepancy principle applied with δ and the noise estimate $\breve{\delta}$ gives the regularization parameters $p(\delta)$ and $p(\breve{\delta})$, respectively. We tabulate the relative errors

$$\mathrm{err}_p = \frac{|p(\delta) - p(\breve{\delta})|}{p(\delta)}. \tag{14}$$

For TSVD and the iterative method of Subsection 2.3, the numerator of (14) is a nonnegative integer.

The columns labeled $\mathrm{PSNR}_{\breve{X}}$ show PSNR-values for restored images determined when the discrepancy principle is applied with a noise estimate $\breve{\delta}$ obtained from one of the denoising methods of Section 4. These PSNR-values are larger than those obtained with the L-curve criterion only when $\breve{\delta}$ is determined by the RKMD method. This is commented on further below.

The last column of Tables 1-3 is labeled $\mathrm{PSNR}_{\breve{X}(\breve{B})}$. It displays the PSNR-values of the restored images obtained by using the discrepancy principle with the noise estimate $\breve{\delta}$ and restoration is applied to the denoised blurry image, $\breve{B}$, rather than to the given blurred and noisy image B. When $\breve{B}$ and $\breve{\delta}$ are determined by the RKMD method, the PSNR-values for the restorations $\breve{X}(\breve{B})$ are the largest for all restoration methods considered. Selected restorations are shown in Figures 3-5.

Our examples show TV denoising to overly smooth each blurred and noisy image. This resulted in a significant over-estimation of the norm of the noise

(a) Restored, RKMD.

(b) Restored, TV denoising.

(c) Restored, BayesShrink.

(d) Denoised blurry image.

(e) Restoration from denoised image.

Figure 3: The *Kent* image blurred and contaminated with 5% noise, and restored using TSVD. The discrepancy principle was applied using each of the indicated denoising methods to estimate the noise level δ. Image (e) shows the restoration achieved by deblurring image (d), which is a denoised version of (a) determined with the RKMD method.

(a) Restored, RKMD.

(b) Restored, TV denoising.

(c) Restored, BayesShrink.

(d) Restoration from denoised image.

Figure 4: The *Barbara* image blurred and contaminated with 5% noise, and restored using Tikhonov regularization. The discrepancy principle was applied using each of the indicated denoising methods to estimate the noise level δ. Image (d) shows the restoration achieved by deblurring a denoised version of the *Barbara* image in Figure 2 with the denoising done by the RKMD method.

(a) Restored, RKMD.

(b) Restored, TV denoising.

(c) Restored, BayesShrink.

(d) Restoration from denoised image.

Figure 5: The *Kent* image blurred and contaminated with 5% noise, and restorations determined by a minimal residual iterative method. The discrepancy principle was applied using each of the indicated denoising methods to estimate the noise level δ. Image (d) shows the restoration achieved by deblurring a denoised version of the *Kent* image in Figure 2 with the denoising done by the RKMD method.

in B. The computed estimate $\breve{\delta}$ of δ was much too large to yield an accurate restoration when applied in the discrepancy principle.

In contrast, the BayesShrink method under-estimated the norm of the noise for each example, though the estimate was far closer to the true δ-value (12) than the estimates determined by TV-denoising. When using the $\breve{\delta}$-value obtained from (13) in the discrepancy principle, the resulting restorations were of larger norm than the optimal restoration, $\breve{X}_{opt}$. A modification of the BayesShrink method in which an initial estimate different from (8) is used might mitigate this issue.

The smallest relative error err_δ in the computed noise-estimate was obtained by the RKMD method. This led to computed regularization parameter values with very small relative errors err_p. Moreover, the resulting restorations had the highest PSNR-values among the three denoising methods. These values were very close to the $\text{PSNR}_{\breve{X}_{opt}}$-values for all computed examples. Further, the $\text{PSNR}_{\breve{X}(\breve{B})}$-values were larger than the corresponding $\text{PSNR}_{\breve{X}_{opt}}$-values for the RKMD method. This indicates that not only does the RKMD method allow for a data-driven implementation of the discrepancy principle but, moreover, the denoising can lead to restorations of higher quality than those obtained from the discrepancy principle with exact knowledge of the norm of the noise.

6 Conclusions

The numerical examples shown in Section 5.1 as well as numerous other computed examples suggest that the use of data-driven denoising methods has a promising potential informing the choice of regularization parameter value for discrete ill-posed problems. Interestingly, these methods aid not only through the estimation of norm of the Gaussian noise in the data, but also in the denoising of the input data as well. Of the denoising methods compared, only the RKMD method was robust enough to accurately estimate the amount of Gaussian noise in each one of the computed examples.

Acknowledgement

We would like to thank a referee for comments.

References

[1] Å. Björck, A bidiagonalization algorithm for solving ill-posed system of linear equations, BIT, 28 (1988), pp. 659–670.

[2] C. Brezinski, G. Rodriguez, and S. Seatzu, Error estimates for the regularization of least squares problems, Numer. Algorithms, 51 (2009), pp. 61–76.

[3] S. G. Chang, Adaptive wavelet thresholding for image denoising and compression, IEEE Transactions on Image Processing, 9 (2000), pp. 1532–1546.

[4] R. D'Agostino and E. S. Pearson, Tests for departure from normality. Empirical results for the distributions of b2 and $\sqrt{b1}$, Biometrika, 60 (1973), pp. 613–622.

[5] R. D'Agostino and G. L. Tietjen, Simulation probability points of b2 for small samples, Biometrika, 58 (1971), pp. 669–672.

[6] D. Donoho, De-noising by soft-thresholding, IEEE Transactions on Information Theory, 41 (1995), pp. 613–627.

[7] D. Donoho and I. M. Johnstone, Adapting to unknown smoothness via wavelet shrinkage, Journal of the American Statistical Association, 91 (1995), pp. 1200–1224.

[8] L. Eldén, Algorithms for the regularization of ill-conditioned least squares problems, BIT, 17 (1977), pp. 134–145.

[9] H. W. Engl, M. Hanke, and A. Neubauer, Regularization of Inverse Problems, Kluwer, Dordrecht, 1996.

[10] R. A. Fisher, The moments of the distribution for normal samples of measures of departure from normality, Proceedings of the Royal Society of London, Series A, 130 (1930), pp. 16–28.

[11] P. C. Hansen, Analysis of discrete ill-posed problems by means of the L-curve, SIAM Rev., 34 (1992), pp. 561–580.

[12] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems, SIAM, Philadelphia, 1988.

[13] P. C. Hansen, Regularization tools version 4.0 for Matlab 7.3, Numer. Algorithms, 46 (2007), pp. 189–194.

[14] P. C. Hansen, M. E. Kilmer, and R. Kjeldsen, Exploiting residual information in the parameter choice for discrete ill-posed problems, BIT, 46 (2006), pp. 41–59.

[15] T. A. Hearn and L. Reichel, Image denoising via residual kurtosis minimization, submitted for publication.

[16] M. E. Hochstenbach, N. McNinch, and L. Reichel, Discrete ill-posed least-squares problems with a solution norm constraint, Linear Algebra Appl., 436 (2012), pp. 3801–3818.

[17] A. Hyvärinen, J. Karhunen, and E. Oja, Independent Component Analysis, Wiley, New York, 2001.

[18] S. Kindermann, Convergence analysis of minimization-based noise level-free parameter choice rules for linear ill-posed problems, Electron. Trans. Numer. Anal., 38 (2011), pp. 233–257.

[19] C. L. Lawson and R. J. Hanson, Solving Least Squares Problems, SIAM, Philadelphia, 1995. First published by Prentice-Hall in 1974.

[20] J. Mead and R. A. Renaut, A Newton root-finding algorithm for estimating the regularization parameter for solving ill-conditioned least squares problems, Inverse Problems, 25 (2009) 025002.

[21] S. Morigi, L. Reichel, and F. Sgallari, Noise-reducing cascadic multilevel methods for linear discrete ill-posed problems, Numer. Algorithms, 53 (2010), pp. 1–22.

[22] S. Morigi, L. Reichel, and F. Sgallari, Cascadic multilevel methods for fast nonsymmetric blur- and noise-removal, Appl. Numer. Math., 60 (2010), pp. 378–396.

[23] S. Morigi, L. Reichel, and F. Sgallari, A cascadic alternating Krylov subspace image restoration method, in Scale Space and Variational Methods in Computer Vision 2013, eds. A. Kuijper et al., Lecture Notes in Computer Science, vol. 7893, Springer, Berlin, 2013, pp. 98–108.

[24] J. Nocedal and S. Wright, Numerical Optimization, Springer, New York, 1999.

[25] A. Neuman, L. Reichel, and H. Sadok, Implementations of range restricted iterative methods for linear discrete ill-posed problems, Linear Algebra Appl., 436 (2012), pp. 3974–3990.

[26] A. Neuman, L. Reichel, and H. Sadok, Algorithms for range restricted iterative methods for linear discrete ill-posed problems, Numer. Algorithms, 59 (2012), pp. 325–331.

[27] L. Reichel and G. Rodriguez, Old and new parameter choice rules for discrete ill-posed problems, Numer. Algorithms, 63 (2013), pp. 65–87.

[28] L. Rudin, S. Osher, and E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D, 60 (1992), pp. 259–268.

[29] C. Vogel and M. Oman, Iterative methods for total variation denoising, SIAM Journal on Scientific Computing, 17 (1996), pp. 227–238.