Computing Optimal Low-Rank Matrix Approximations for Image Processing

Julianne Chung
Department of Mathematics
Virginia Tech
Blacksburg, VA 24061, USA
Email: jmchung@vt.edu

Matthias Chung
Department of Mathematics
Virginia Tech
Blacksburg, VA 24061, USA
Email: mcchung@vt.edu

***Abstract*—In this work, we describe a new framework for solving inverse problems, where training data is used, as a substitute for the forward model, to compute an optimal low-rank regularized inverse matrix *directly*, allowing for very fast computation of a regularized solution. An empirical Bayes risk minimization framework will be used to incorporate training data and to formulate the problem of computing an optimal low-rank regularized inverse matrix. We describe some theoretical results that motivate the development of numerical methods for computing an optimal low-rank regularized inverse matrix and demonstrate our approach on examples from image deconvolution.**

I. Introduction

In many imaging applications, such as image deconvolution, the underlying mathematical model can be written as a linear problem

$$\mathbf{A}\mathbf{x} + \boldsymbol{\delta} = \mathbf{b}, \tag{1}$$

where $\mathbf{b} \in \mathbb{R}^n$ is the observed data, $\mathbf{A} \in \mathbb{R}^{n\times n}$ models the forward process, $\boldsymbol{\delta} \in \mathbb{R}^n$ is additive noise, and $\mathbf{x} \in \mathbb{R}^n$ is the desired solution. Given $\mathbf{b}$ and $\mathbf{A}$, the goal of the inverse problem is to reconstruct $\mathbf{x}$. For example, in image deconvolution the goal of the inverse problem is to reconstruct an approximation of the vectorized true image $\mathbf{x}$, given a vectorized version of the observed, blurred and noisy image $\mathbf{b}$, and a large, sparse matrix that models the blurring process $\mathbf{A}$. The entries in $\mathbf{A}$ depend on the point spread functions (PSFs) of the imaging system, and under simplifying assumptions such as spatial invariance and certain boundary conditions, matrix $\mathbf{A}$ may be highly structured so that efficient algorithms can be used [1]. For problems where the PSF is unknown, blind deconvolution methods have been developed, but most rely heavily on strict assumptions regarding the imaging system, as well as a good inital guess of the PSF [2]. In this paper, we assume no knowledge regarding the PSFs and avoid the simplifying assumptions described above.

It is well known that the underlying model for image deconvolution is a Fredholm integral equation of the first kind, which is ill-posed, meaning the solution, $\mathbf{x}$, does not exist, is not unique, or does not depend continuously on the data [3]. A significant challenge in solving inverse problems is that small errors in the data may cause large errors in the reconstruction. To alleviate this, regularization is typically used to impose prior information or additional constraints on the solution. However, selecting an appropriate regularization and regularization parameter can be difficult [4], and knowledge of $\mathbf{A}$ along with its singular value decomposition (SVD) is typically required.

In this work, we use training data as a substitute for knowledge of the forward model by introducing a framework that avoids including $\mathbf{A}$ in the problem formulation. Furthermore, by computing an optimal regularized reconstruction matrix directly, computing reliable solutions to inverse problems can be very efficient, requiring only a matrix-vector multiplication, and tuning parameters is unnecessary since regularization is built into the framework.

Assume training data $\mathbf{b}^{(k)}, \mathbf{x}^{(k)}$ for $k = 1, \dots, K$, for (1) are given. First, our goal is to find a rank-r matrix $\mathbf{Z} \in \mathbb{R}^{n\times n}$ that gives a small reconstruction error for the given training set. That is, the sample mean squared error $\frac{1}{K}\sum_{k=1}^{K} \left\|\mathbf{Z}\mathbf{b}^{(k)} - \mathbf{x}^{(k)}\right\|_2^2$ should be small, where $\|\cdot\|_2$ corresponds to the Euclidean norm. In this paper, we consider efficient methods for obtaining an *optimal low-rank regularized inverse matrix*

$$\widehat{\mathbf{Z}} = \underset{\mathrm{rank}(\mathbf{Z})\le r}{\arg\min} \frac{1}{K}\sum_{k=1}^{K} \left\|\mathbf{Z}\mathbf{b}^{(k)} - \mathbf{x}^{(k)}\right\|_2^2. \tag{2}$$

Then, once matrix $\widehat{\mathbf{Z}}$ is computed, a simple matrix-vector multiplication is required to solve the inverse problem, $\mathbf{x} = \widehat{\mathbf{Z}}\mathbf{b}$. Matrix $\widehat{\mathbf{Z}}$ can be precomputed, and a regularized solution to the inverse problem can be obtained efficiently and accurately. Optimization problem (2) can be interpreted as an *empirical Bayes risk minimization problem*, c. f. [5].

The paper is organized as follows. Section II establishes notation and provides some background and motivation on low-rank regularized inverse matrices. Then methods for solving the empirical Bayes risk minimization problem (2) are described in Section III, along with some remarks on computational concerns. Numerical results are presented in Section IV, and conclusions and future discussion are provided in Section V.

II. Background and Motivation

In the following we give some background on low-rank approximations and their connection to the singular value de-

composition (SVD). In particular, we show that a well-known method for solving inverse problems, namely the truncated-SVD (TSVD) approach, uses low-rank reconstruction matrices to produce regularized solutions.

For any real nonsingular $n \times n$ matrix $\mathbf{A}$, let $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$ be the singular value decomposition, where $\mathbf{\Sigma}$ is a diagonal matrix containing the singular values, $\sigma_1 \geq \ldots \geq \sigma_n > 0$, and orthogonal matrices $\mathbf{U} = [\mathbf{u}_1, \ldots, \mathbf{u}_n]$ and $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ contain the left and right singular vectors $\mathbf{u}_i$ and $\mathbf{v}_i$, $i = 1, \ldots, n$, respectively [6]. For a given positive integer $r \leq \text{rank}(\mathbf{A})$, define

$$\mathbf{A}_r = \mathbf{U}_r \mathbf{\Sigma}_r \mathbf{V}_r^\top,$$

where $\mathbf{U}_r = [\mathbf{u}_1, \ldots, \mathbf{u}_r] \in \mathbb{R}^{n \times r}$, $\mathbf{V}_r = [\mathbf{v}_1, \ldots, \mathbf{v}_r] \in \mathbb{R}^{n \times r}$, and

$$\mathbf{\Sigma}_r = \begin{bmatrix} \sigma_1 & & & \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_r \end{bmatrix} \in \mathbb{R}^{r \times r}.$$

Then, matrix $\mathbf{A}_r^\dagger = \mathbf{V}_r \mathbf{\Sigma}_r^{-1} \mathbf{U}_r^\top$ is the so called pseudo-inverse of $\mathbf{A}_r$ [7].

It is well-known that for ill-posed inverse problems, the singular values decay to and cluster at zero, with no noticeable gap in the spectrum, and singular vectors corresponding to small singular values are highly oscillatory. Spectral filtering methods provide solutions where components of the solution corresponding to the small singular values are damped or neglected [8].

Assume for a moment that $\mathbf{A}$ and its SVD are available, then the rank-r TSVD solution

$$\mathbf{x}_{\text{TSVD}} = \mathbf{A}_r^\dagger \mathbf{b}$$

provides a regularized solution to (1), see [9]. The behavior of the TSVD reconstruction for an image deblurring example is illustrated in Figures 1 and 2. Figure 1 shows TSVD reconstructions $\mathbf{x}_{\text{TSVD}}$ for ranks $r = 30$, 175, and 500, and Figure 2 provides a plot of the relative reconstruction errors $\frac{\|\mathbf{x}_{\text{TSVD}} - \mathbf{x}_{\text{true}}\|_2}{\|\mathbf{x}_{\text{true}}\|_2}$ for various ranks r.

rank-30 TSVD rank-175 TSVD rank-500 TSVD

Fig. 1. TSVD reconstructions for an image deblurring example corresponding to ranks $r = 30$, 175, and 500.

Notice that for small rank, the reconstruction remains blurry and the reconstruction error is high. This is due to the fact that the singular vectors corresponding to the large singular values typically contain low-frequency information, so the image detail has not yet fully resolved. On the other hand, for large values of r, the TSVD solution becomes contaminated with inverted noise. This is commonly referred to as semi-convergence behavior. Selecting an appropriate truncation or regularization parameter (i.e, rank r) can be difficult, and an imprecise estimate may result in poor solutions. A variety of methods have been proposed for estimating r. Some methods such as the discrepancy principle or unbiased predictive risk estimator (UPRE) require an estimate of the noise level, while other methods such as generalized cross-validation (GCV) or methods based on the normalized cumulative periodogram can estimate a parameter from the data [10], [4]. However, all of these methods rely on knowledge of $\mathbf{A}$ and require computational expensive tuning of the regularization parameter.

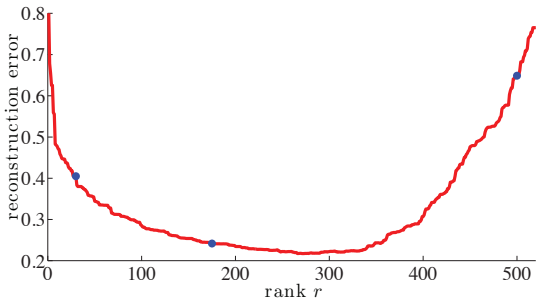


Fig. 2. Relative reconstruction errors for TSVD reconstructions for various ranks r. Blue dots indicate errors corresponding to the reconstructions presented in Figure 1 for $r = 30$, 175, and 500 .

III. Computing Optimal Low-Rank Inverse Matrices

As illustrated in the previous section, the low-rank TSVD reconstruction matrix $\mathbf{A}_r^\dagger$ can provide regularized solutions for inverse problems in imaging applications such as image deconvolution. In this paper, we seek an *optimal* low-rank regularized inverse matrix that minimizes the sample mean squared errors for the training data, and we assume that we do not have knowledge of the matrix $\mathbf{A}$, i.e., we want to solve (2).

Before addressing the problem of computing a solution for (2), we want to make the reader aware that problem (2) can be interpreted as an approximation of an underlying *Bayes risk minimization problem*. Assume $\mathbf{x}$ and $\boldsymbol{\delta}$ are realizations of random variables $\mathcal{X}$ and $\mathcal{D}$, then the Bayes risk minimization problem is given by

$$\min_{\text{rank}(\mathbf{Z}) \leq r} \mathbb{E}_{\mathcal{X},\mathcal{D}} \left\{ \|\mathbf{Z}(\mathbf{A}\mathbf{x} + \boldsymbol{\delta}) - \mathbf{x}\|_2^2 \right\},$$

where $\mathbb{E}$ is the expected value. If we assume that the training data $\mathbf{x}^{(k)}$ and $\mathbf{b}^{(k)} = \mathbf{A}\mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)}$ are obtained by Monte Carlo samples $\mathbf{x}^{(k)}$ and $\boldsymbol{\delta}^{(k)}$ drawn from the corresponding probability distributions of the Bayes problem, then we get the empirical Bayes risk minimization problem (2). For details on Bayes and empirical Bayes risk estimators, as well as numerical methods for these estimators, the reader is referred

to Vapnik's book on statistical learning theory [5] and Shapiro et al.'s book on stochastic programming [11].

Next, we describe a method for computing a solution to problem (2), thereby providing us with an optimal low-rank regularized matrix inverse approximation that can be used to solve inverse problems efficiently. Let us define $\mathbf{B} = [\mathbf{b}^{(1)}, \ldots, \mathbf{b}^{(K)}]$ and $\mathbf{C} = [\mathbf{x}^{(1)}, \ldots, \mathbf{x}^{(K)}]$, where $\mathbf{B}$ and $\mathbf{C}$ are $n \times K$ matrices. Then using relationships between the 2-norm of a vector and the Frobenius norm of a matrix, we can recast problem (2) as an equivalent minimization problem,

$$\min_{\mathrm{rank}(\mathbf{Z}) \leq r} \frac{1}{K} \|\mathbf{ZB} - \mathbf{C}\|_{\mathrm{F}}^2 , \tag{3}$$

where $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm of a matrix.

Problem (3) has been studied in the matrix analysis community, and the following theorem provides a closed form solution to the minimizer of our problem.

Theorem 3.1: Let matrices $\mathbf{B}, \mathbf{C} \in \mathbb{R}^{n \times K}$ be given. Further let $\tilde{\mathbf{V}}$ be the matrix of right singular vectors of $\mathbf{B}$ and define $\mathbf{P} = \mathbf{C}\tilde{\mathbf{V}}_s(\tilde{\mathbf{V}}_s)^\top$ where $s = \mathrm{rank}\,(\mathbf{B})$. Then

$$\widehat{\mathbf{Z}} = \mathbf{P}_r \mathbf{B}^\dagger$$

is a solution to the minimization problem

$$\widehat{\mathbf{Z}} = \operatorname*{arg\,min}_{\mathrm{rank}(\mathbf{Z}) \leq r} \frac{1}{K} \|\mathbf{ZB} - \mathbf{C}\|_{\mathrm{F}}^2 .$$

This solution is unique if and only if either $r \geq \mathrm{rank}\,(\mathbf{P})$ or $1 \leq r \leq \mathrm{rank}\,(\mathbf{P})$ and $\bar{\sigma}_r > \bar{\sigma}_{r+1}$, where $\bar{\sigma}_r$ and $\bar{\sigma}_{r+1}$ denote the r and $(r+1)$-st singular values of $\mathbf{P}$.

Theorem 3.1 is a special case of the main result (Theorem 2.1) of Friedland and Torokhti's paper on rank-constrained matrix approximations [12] and is proven there.

In this work, we propose to use Theorem 3.1 to compute an *optimal* low-rank regularized inverse matrix, $\widehat{\mathbf{Z}}$. The main computational burden consists of computing two SVDs (one of $\mathbf{B}$ and one of $\mathbf{P}$). However, both SVDs can be computed efficiently since we only require s singular values and corresponding singular vectors of $\mathbf{B}$ and the largest r singular values and corresponding singular vectors of $\mathbf{P}$. It is worth noting that these computations can be computed off-line. Once computed, $\widehat{\mathbf{Z}}$ can be used to solve inverse problems. Furthermore, the ability to solve hundreds of inverse problems in real time, each requiring only *one* matrix-vector multiplication, enables us to amortize the initial cost of constructing $\widehat{\mathbf{Z}}$.

IV. Numerical Results

In this section, we provide numerical results for an image deblurring problem. Our training set $\{\mathbf{x}^{(k)}, \mathbf{b}^{(k)}\}_{k=1}^{K}$ consisted of 2400 satellite images, each image being 128×128. True images $\mathbf{x}^{(k)}$ were obtained using random translation, rotation, and zoom of eight images (see top row of Figure 3 for 5 sample images). To generate the corresponding noisy, blurred images, we computed $\mathbf{b}^{(k)} = \mathbf{A}\mathbf{x}^{(k)} + \boldsymbol{\delta}^{(k)}$ where $\mathbf{A}$ represents a spatially invariant blurring matrix, defined by a symmetric Gaussian point spread function with mean 0 and covariance matrix equal to 5 times the identity matrix. The noise realizations $\boldsymbol{\delta}^{(k)}$ were drawn from a Gaussian normal distribution with uniform noise ranges between 0.1 and 0.15. Five of the corresponding blurred, noisy images can be found in the bottom row of Figure 3.

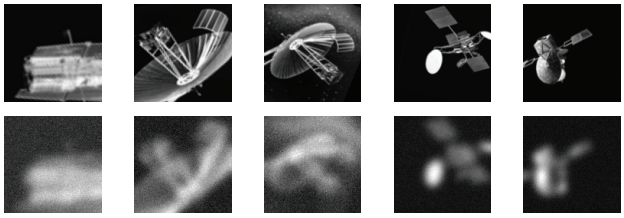


Fig. 3. Image deblurring example. Sample true training images can be found in the top row. Images in the bottom row are the corresponding blurred, noisy images.

The validation set of images was generated in a similar manner but with eight different satellite images, resulting in 2400 different satellite images for validating our methods.

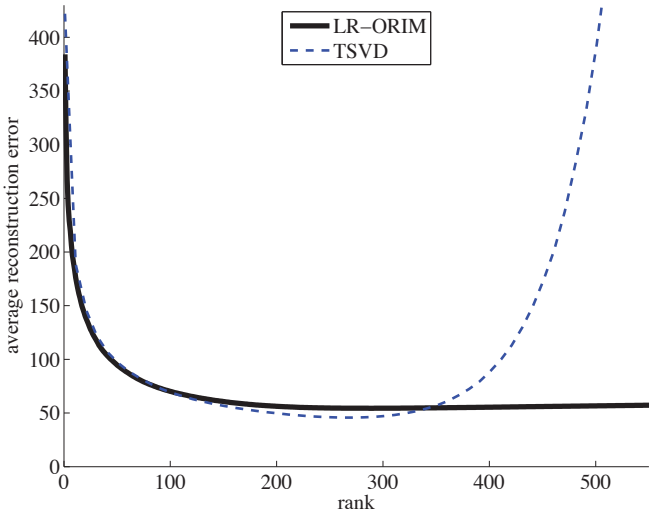


Fig. 4. Average reconstruction errors of the validation set for LR-ORIM and TSVD for various ranks r.

First, we provide a comparison of the average reconstruction error for various ranks. That is, for various choices of r, we computed a low-rank optimal regularized inverse matrix (LR-ORIM) $\widehat{\mathbf{Z}}$ as described in Theorem 3.1. Then reconstructions for the validation set were computed and reconstruction errors were recorded. The average error, computed according to (2), is reported in Figure 4 as a function of rank. As a comparison, we provide the average reconstruction errors for the rank-r TSVD solution, which exhibits semi-convergence behavior similar to that of Figure 2.

TSVD requires full knowledge of $\mathbf{A}$. Figure 4 illustrates that LR-ORIM can achieve similar reconstruction errors as

TSVD without that knowledge, but by instead incorporating training data. Another important remark is that with a sufficient number of training images, the LR-ORIM reconstructions can avoid bias to the training set and simultaneously incorporate appropriate regularization, thereby avoiding semi-convergence behavior.

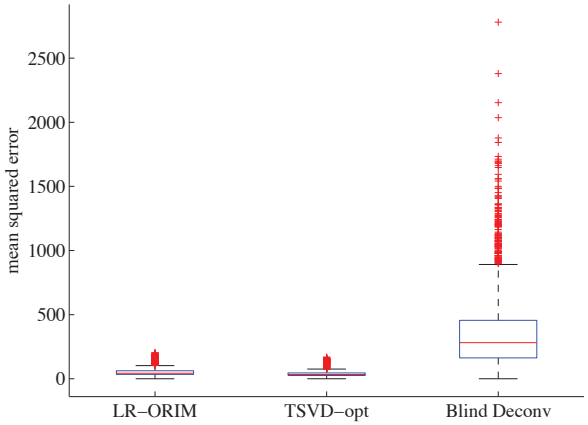


Fig. 5. Mean square errors for LR-ORIM, TSVD-opt, and Blind Deconv are presented for the validation images using box plots. LR-ORIM refers to our newly proposed method. TSVD-opt is TSVD where the rank was chosen to minimize the mean squared errors for the training data. Blind deconv refers to a blind deconvolution algorithm. The central mark refers to the median, the edges of the box correspond to the 25th and 75th percentiles, and the whiskers include outliers while extreme outliers are plotted individually.

If training data and knowledge of the PSF is available, then the training data could be used to select an optimal regularization parameter (i.e., rank) for TSVD [8]. In particular, the rank for TSVD can be selected such that it minimizes the reconstruction error for the training data. Here, we refer to this method as TSVD-opt, and for this example, the selected rank was $r = 280$. Since TSVD requires knowledge of the PSF, we also compare to an approach for blind deconvolution, where the PSF is unknown. Matlab's built-in `deconvblind` function from the Image Processing Toolbox was used. Their approach assumes spatially invariant blur and periodic boundary conditions, and they use a maximum likelihood approach. A significant limitation of this approach is that a good initial guess of the point spread function is required. In particular, an accurate estimate of the support of the PSF is important. For this experiment, we selected the initial guess of the PSF to be a box-car blur whose size matched the variance of the Gaussian PSF. That is, we used a 21×21 box car blur as an initial guess. The performance of the three methods on the validation set is illustrated in Figures 5 and 6 by using box plots of the mean squared errors of the reconstructions, and some reconstructions for one of the validation images can be found in Figure 7.

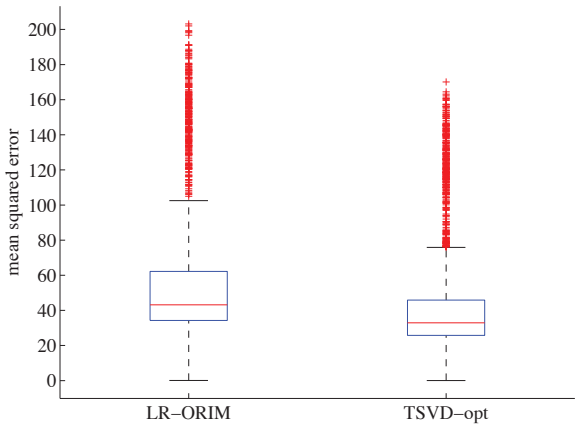
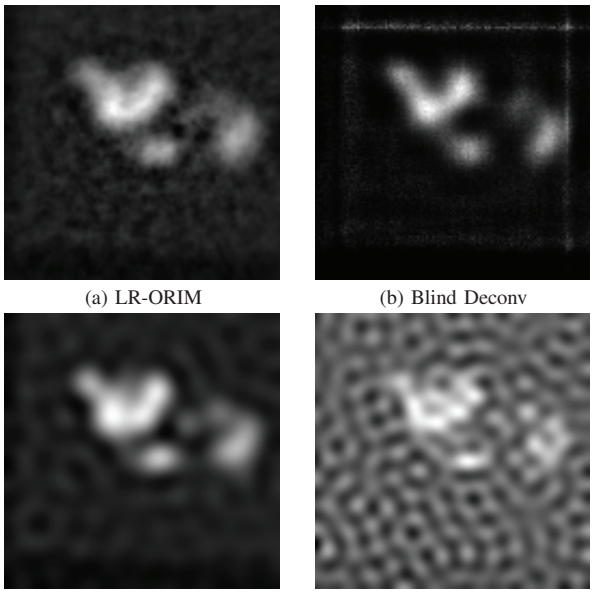


Fig. 6. For better comparison of LR-ORIM and TSVD-opt, this figure excludes Blind Deconv from Figure 5.



(a) LR-ORIM (b) Blind Deconv

(c) TSVD-opt (d) TSVD-500

Fig. 7. Reconstructed images in (a) and (b) correspond to LR-ORIM and Matlab's built-in blind deconvolution algorithm respectively. The TSVD reconstruction in (c) uses the optimally selected rank using the training data, and (d) corresponds to the rank-500 TSVD solution.

The authors acknowledge that a fair comparison of algorithms is difficult. Our goal is that the comparison with TSVD provides the reader insight into how our approach compares to other low-rank reconstruction methods. In particular, our framework can produce similar results to TSVD, where the main difference is that LR-ORIM needs training data and TSVD additionally requires full knowledge of $\mathbf{A}$. Then, to provide a comparison with a method that does not require knowledge of the PSF, we present results for a blind deconvolution algorithm. However, the success of blind deconvolution methods relies heavily on a good initial guess

of the PSF and other constraints on the PSF. We remark that work on multi-frame blind deconvolution [13], where the PSF is estimated from multiple images, may provide a more appropriate comparison, but additional assumptions on the PSF are required here as well.

V. Conclusions and Discussions

In this paper, we presented an approach for solving inverse problems, where training data is used as a substitute for knowledge of the forward model. We formulated the problem of computing an optimal low-rank regularized inverse matrix and described numerical approaches for computing such a matrix. Once computed, the recontruction matrix can be used to efficiently solve inverse problems, with only a matrix-vector mutiplication. Numerical examples from image deblurring illustrate the sucess of this approach.

Although we have focused mainly on the image deconvolution problem, algorithms developed here are applicable to a wide range of inverse problems.

Acknowledgment

We are grateful to Dianne P. O'Leary for helpful discussions. The satellite images used here were obtained from NASA's website: www.nasa.gov.

References

[1] P. C. Hansen, J. G. Nagy, and D. P. O'Leary, *Deblurring Images: Matrices, Spectra and Filtering*. Philadelphia: SIAM, 2006.

[2] P. Campisi and K. Egiazarian, *Blind image deconvolution: theory and applications*. CRC press, 2007.

[3] J. Hadamard, *Lectures on Cauchy's Problem in Linear Differential Equations*. New Haven: Yale University Press, 1923.

[4] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*. Philadelphia: SIAM, 2010.

[5] V. Vapnik, *Statistical Learning Theory*. San Francisco: Wiley, 1998.

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

[7] G. W. Stewart, *Matrix Algorithms: Volume 1, Basic Decompositions*. SIAM, 1998.

[8] J. Chung, M. Chung, and D. P. O'Leary, "Designing optimal filters for ill-posed inverse problems," *SIAM Journal on Scientific Computing*, vol. 33, no. 6, pp. 3132–3152, 2011.

[9] C. R. Vogel, "Optimal choice of a truncation level for the truncated SVD solution of linear first kind integral equations when data are noisy," vol. 23, pp. 109–117, 1986.

[10] ——, *Computational Methods for Inverse Problems (Frontiers in Applied Mathematics)*. Philadelphia: SIAM, 1987.

[11] A. Shapiro, D. Dentcheva, and R. Ruszczynski, *Lectures on Stochastic Programming: Modeling and Theory*. Philadelphia: SIAM, 2009.

[12] S. Friedland and A. Torokhti, "Generalized rank-constrained matrix approximations," *SIAM Journal on Matrix Analysis and Applications*, vol. 29, no. 2, pp. 656–659, 2007.

[13] T. J. Schulz, "Multiframe blind deconvolution of astronomical images," *JOSA A*, vol. 10, no. 5, pp. 1064–1073, 1993.