Suboptimality of the Truncated SVD for Ill-Conditioned Inverse Problems

Jeffrey A. Fessler*and Alfred O. Hero †‡

(Accepted to IEEE T-IP in 1995 but withdrawn by authors.)

July 2, 2003

Abstract

The truncated singular value decomposition (SVD) is a popular method for computing regularized estimates in ill-posed inverse problems. This paper analyzes the bias-variance tradeoff of a class of linear estimators that includes the truncated SVD method. The bias-variance tradeoff of the truncated SVD method is shown to be suboptimal, whereas a penalized least-squares estimator is shown to be optimal within the linear class.

Keywords: Singular value decomposition, bias/variance tradeoff, regularization.)

I. Theory

Many inverse problems can be represented by the standard linear additive noise model:

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{\theta} + \epsilon,$$

where $\boldsymbol{\theta} \in \Re^p$ is the parameter to be estimated, $\boldsymbol{Y} \in \Re^n$ is the noisy measurement, and $\epsilon \in \Re^p$ is the zero-mean additive noise with positive-definite covariance matrix $\boldsymbol{\Pi}$, assumed known up to a scaling constant. For simplicity we assume $n \geq p$. For poorly conditioned inverse problems, the standard linear weighted least-squares estimator

$$\hat{\boldsymbol{\theta}}_{\mathrm{LS}} = (\boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{A})^{-1}\boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{Y}$$

is unbiased if $\boldsymbol{A}$ has full column rank, but has unacceptably high variance due to the small singular values of $\boldsymbol{A}$. Many methods exist for reducing the variance, all of which induce bias.

One popular method for regularizing the least-squares estimate is the truncated SVD [1–3]. In this paper, we derive the bias and variance properties of the class of *linear weighted SVD estimators*, which includes both the truncated SVD and the penalized least-squares estimator. We show that if the norm of the bias is constrained to not exceed a given value, then a penalized least-squares estimator has minimum variance, whereas the truncated SVD is suboptimal.

A. Linear Weighted SVD Estimators

Define the following SVD:

$$\boldsymbol{\Pi}^{-1/2}\boldsymbol{A} = \boldsymbol{U}\boldsymbol{D}\{\nu_k\}\boldsymbol{V}', \tag{1}$$

where $\boldsymbol{U} \in \Re^{n\times p}$ and $\boldsymbol{V} \in \Re^{p\times p}$ are each orthogonal, and $\boldsymbol{D}\{\nu_k\}$ is a $p \times p$ diagonal matrix with entries $\{\nu_k\}_{k=1}^p$. We consider the class of *linear weighted SVD estimators*, i.e., those that can be written in the following form:

$$\hat{\boldsymbol{\theta}}(\boldsymbol{w}) = \sum_{k=1}^p \boldsymbol{v}_k w_k \boldsymbol{u}_k' \boldsymbol{\Pi}^{-1/2}\boldsymbol{Y} = \boldsymbol{V}\boldsymbol{D}\{w_k\}\boldsymbol{U}'\boldsymbol{\Pi}^{-1/2}\boldsymbol{Y}, \tag{2}$$

where $\boldsymbol{v}_k$ and $\boldsymbol{u}_k$ are the columns of $\boldsymbol{V}$ and $\boldsymbol{U}$ respectively. The weights $\boldsymbol{w}$ control the tradeoff between bias and variance. The covariance of such an estimator is

$$\mathrm{Cov}\{\hat{\boldsymbol{\theta}}(\boldsymbol{w})\} = \boldsymbol{V}\boldsymbol{D}\{|w_k|^2\}\boldsymbol{V}'. \tag{3}$$

The bias of such an estimator is

$$\boldsymbol{b}(\boldsymbol{w}) = E\{\hat{\boldsymbol{\theta}}(\boldsymbol{w})\} - \boldsymbol{\theta} = -\boldsymbol{V}\boldsymbol{D}\{1 - w_k\nu_k\}\boldsymbol{V}'\boldsymbol{\theta}. \tag{4}$$

B. Bias-Variance Tradeoff

Regularization always involves a tradeoff between bias and variance. From (3) we see that to minimize variance we would like each w_k to be small, whereas from (4) to minimize bias we would like $w_k \approx 1/\nu_k$. These are conflicting requirements that epitomize the bias-variance tradeoff. We would like a strategy for choosing $\boldsymbol{w}$ that minimizes variance subject to a constraint on the allowable bias.

Define the vector $\boldsymbol{x} = \boldsymbol{V}'\boldsymbol{\theta}$, then the norm of the bias vector $\boldsymbol{b}(\boldsymbol{w})$ in (4) can be written

$$\begin{aligned} \|\boldsymbol{b}(\boldsymbol{w})\|^2 &= \|\boldsymbol{D}\{1 - w_k\nu_k\}\boldsymbol{V}'\boldsymbol{\theta}\|^2 \\ &= \sum_k |1 - w_k\nu_k|^2 |x_k|^2. \end{aligned}$$

For a given amount of bias, we would like to minimize the corresponding variance. Since in general the x_k's are unknown since they depend on the unknown signal $\boldsymbol{\theta}$, we take as our goal **(BIG LEAP HERE)** the following optimization problem: minimize the trace of the covariance matrix, subject to the constraint:

$$\sum_k |1 - w_k\nu_k|^2 q_k \leq C, \tag{5}$$

for $C \in (0, \sum_k q_k)$. The nonnegative weights q_k are design parameters that control the relative importance of biases in each

---

*J. Fessler is with the Division of Nuclear Medicine of the University of Michigan, 3480 Kresge III, Box 0552, Ann Arbor, MI, 48109-0552. E-mail: fessler@umich.edu, phone: 734-763-1434.

†A. Hero is with the Dept. of Electrical Engineering and Computer Science of the University of Michigan.

‡This work was supported in part by DOE Grant DE-FG02-87ER65061 and NIH Grants CA60711, CA54362, and CA52880.

of the directions of the singular vectors. For the moment we leave $\{q_k\}$ unspecified. In the discussion we argue that in the absence of prior information, the equally-weighted choice $q_k = 1$ is sensible.

From (3), the trace of the covariance matrix is:

$$\operatorname{trace}\{\boldsymbol{V}\boldsymbol{D}\{|w_k|^2\}\boldsymbol{V}'\} = \operatorname{trace}\{\boldsymbol{D}\{|w_k|^2\}\} = \sum_k w_k^2.$$

Using the method of Lagrange:

$$\min_{w_1,\ldots,w_p} \sum_{k=1}^{p} |w_k|^2 + \lambda \left(\sum_{k=1}^{p} |1 - w_k \nu_k|^2 q_k - C \right),$$

where λ is the Lagrangian multiplier, the optimal solution for w_k is easily shown to be:

$$w_k(\hat{\lambda}) = \frac{\nu_k^\star}{\hat{\lambda}^{-1} q_k^{-1} + |\nu_k|^2}, \tag{6}$$

where $\hat{\lambda} \in (0, \infty)$ is the solution to

$$C = \sum_{k=1}^{p} |1 - w_k(\hat{\lambda})\nu_k|^2 q_k = \sum_{k=1}^{p} \left(\frac{1}{1 + \hat{\lambda} q_k |\nu_k|^2} \right)^2.$$

C. Truncated SVD Estimator

For the truncated SVD estimator with $m \le p$ retained components, the form for w_k is:

$$w_k = \begin{cases} 1/\nu_k, & k = 1, \ldots, m \\ 0, & k = m+1, \ldots, p \end{cases}. \tag{7}$$

Comparing this to (6), one sees immediately that the truncated SVD estimator correponds to the *particular* case where $q_k = \infty$ for $k = 1, \ldots, m$, and $q_k = 0$ for $k = m+1, \ldots, p$. When the system matrix $\boldsymbol{A}$ is circulant and $\boldsymbol{U}$ and $\boldsymbol{V}$ correspond to the Fourier basis, then this “hard threshold” choice for the weights q_k forces the signal to be band-limited. But for more general systems $\boldsymbol{A}$, it is perhaps unlikely that this choice of q_k will be well-matched to the signal’s x_k, so for a given amount of bias, the variance could be unnecessarily large. **PRETTY WEAK CONCLUSION. THE BOTTOM LINE IS THAT SVD *IS* OPTIMAL FOR SOME CHOICE OF THE QK’S, SO UNFORTUNATELY IT JUST BOILS DOWN TO THAT SUBJECTIVE CHOICE, ALTHOUGH I GUESS ONE COULD ARGUE THAT IT IS BETTER TO DECIDE UP FRONT WHAT BIAS WEIGHTING YOU WANT AND THEN FIND THE APPROPRIATE ESTIMATOR RATHER THAN “DEFAULATING” TO WHATEVER SVD GIVES...**

D. Penalized Weighted Least Squares Estimator

Another natural regularized estimator for the linear additive noise problem is the penalized weighted least-squares (PWLS) estimator:

$$\hat{\boldsymbol{\theta}}_{\mathrm{PLS}} = \arg\min_{\boldsymbol{\theta}} (\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\theta})'\boldsymbol{\Pi}^{-1}(\boldsymbol{y} - \boldsymbol{A}\boldsymbol{\theta}) + \beta \boldsymbol{\theta}'\boldsymbol{R}\boldsymbol{\theta}, \tag{8}$$

where $\boldsymbol{R}$ is a symmetric nonnegative definite regularization matrix. In the absence of constraints, the solution is:

$$\hat{\boldsymbol{\theta}}_{\mathrm{PLS}} = (\boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{A} + \beta\boldsymbol{R})^{-1}\boldsymbol{A}'\boldsymbol{\Pi}^{-1}\boldsymbol{Y}. \tag{9}$$

If $\boldsymbol{R} = \boldsymbol{V} \operatorname{diag}\{q_k\}^{-1} \boldsymbol{V}'$, then we can rewrite (9) in the form (2) with

$$w_k = \frac{\nu_k^\star}{\beta q_k^{-1} + |\nu_k|^2}. \tag{10}$$

Comparing (10) to (6), we see that by identifying β with $\hat{\lambda}^{-1}$, the PWLS estimator has exactly the optimal form. In other words, there is a value for β such that the PWLS estimator satisfies the constraint (5) and has minimum variance, as measured by the trace of the covariance, over the class of linear weighted SVD estimators. The choice of the $\{q_k\}$’s directly affects the regularization matrix $\boldsymbol{R}$.

II. Example

To illustrate the difference between the truncated SVD and PWLS estimators, consider a simple deconvolution problem where the $\boldsymbol{A}$ matrix corresponds to circular convolution with the $(1, 2, 1)$ kernel, and the noise covariance $\boldsymbol{\Pi}$ is simply the identity matrix. For the signal $\boldsymbol{\theta}$ shown in Fig. 1, we constrained the total variance to be $\sum_k w_k^2 \le 43$, which corresponded to retaining $m = 18$ of the $p = 30$ components for the truncated SVD estimator, and $\beta = 0.0804$ for the PWLS estimator, i.e., we selected m and β so that the total variance of the two estimators was matched. Since both estimators are linear, we can analytically compute their means, which are shown in Fig. 1. Due to the ringing of the truncated SVD estimator, its total bias $\|\boldsymbol{b}(\boldsymbol{w})\|$ is about 20% higher than the total bias of the PWLS estimator. Fig. 2 displays the weights w_k for the two estimators. The truncated SVD weights are significantly different from the optimal choice used by the PWLS estimator.

III. Discussion

We have shown that given a bias norm constraint, the truncated SVD has suboptimally high variance (except for one particular choice of the weights $\{q_k\}$), whereas a PWLS estimator attains the minimum variance over the linear class (2). The primary issue is then how to choose the weights $\{q_k\}$. If $\boldsymbol{A}$ is circulant, then one can choose $\boldsymbol{V}$ and $\boldsymbol{U}$ in (1) to be the Fourier basis. **THIS IS NONSTANDARD TO HAVE COMPLEX SINGULAR VECTORS?)** If the object $\boldsymbol{\theta}$ is a point source $\boldsymbol{\theta} = \boldsymbol{e}_j$, then the elements of the vector $\boldsymbol{V}'\boldsymbol{\theta}$ will all be complex exponentials, which have equal magnitude ($|x_k|^2 = 1$). When the object is a point source, the norm of the bias of the estimator is a measure of resolution. Therefore, by using $q_k = 1$ in (5), we are essentially *constraining the minimum resolution* of the estimator. This can be formalized by examining the bias gradient [4, 5]. In other words, in imaging problems, for a given resolution, the PWLS estimator will have less noise than the truncated SVD estimator. These results can be viewed more broadly in terms of lower bounds on achievable variance for biased estimators [4, 5].

Sometimes one may have the additional side information that for certain k the inner product $x_k = \boldsymbol{v}_k'\boldsymbol{\theta}$ is small. In such cases, one can choose a smaller q_k for such k, which will reduce the corresponding w_k. However, if one does not know for which k the product $\boldsymbol{v}_k'\boldsymbol{\theta}$ is small, and if the object may contain small features of interest (such as point sources), then using PWLS with $q_k = 1$ will yield smaller bias (or variance) over SVD with its choice of q_k.

One can obtain some intuition about why the truncated SVD is suboptimal by comparing (7) with (6) (see Fig. 2). Whereas the truncated SVD (7) completely discards all information in components $n+1$ through p, the optimal choice (6) gracefully diminishes the contribution of the components corresponding to small singular values, thereby reducing the variance relative to the conventional least-squares estimator, but yet still retaining some of the information in the upper singular vectors.

From a practical view, the PWLS estimator also has the advantage that there are very fast converging algorithms [6, 7] for performing the minimization in (8). The coordinate ascent algorithms in [6, 7] use much less computation than required by the singular value decomposition. In addition, unlike the truncated SVD, one can easily incorporate a non-negativity constraint on the parameters into the coordinate ascent algorithms.

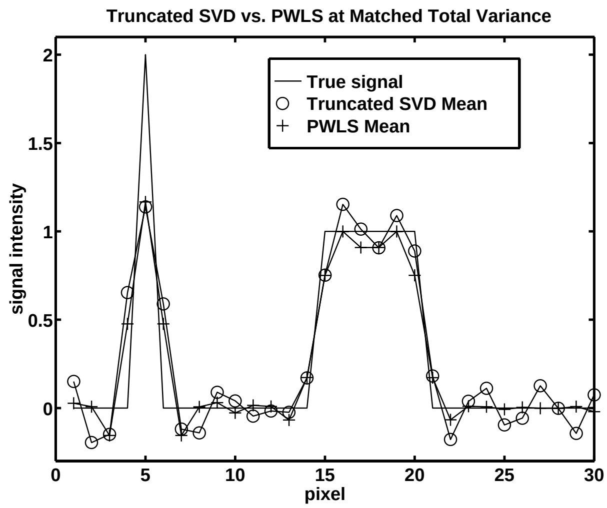

Figure 1: Mean response for truncated SVD estimator and PWLS estimator at a matched total variance. The truncated SVD estimator has more bias due to the ringing.

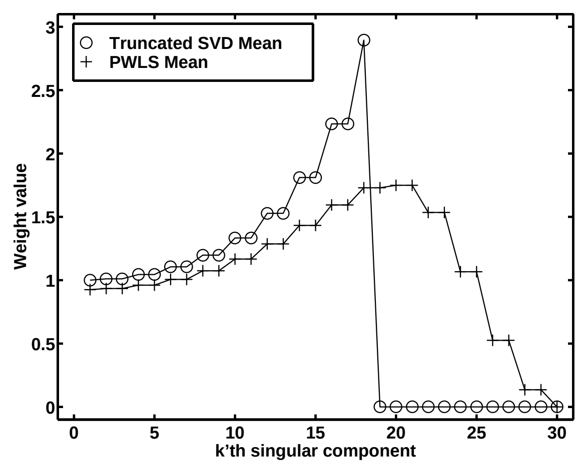

Figure 2: Comparison of the weights w_k for the truncated SVD estimator and for the PWLS estimator.

REFERENCES

[1] Y. S. Shim and Z. H. Cho. SVD pseudoinversion image reconstruction. *IEEE Tr. Acoust. Sp. Sig. Proc.*, 29(4):904–909, August 1981.

[2] J. Llacer. Tomographic image reconstruction by eigenvector decomposition: Its limitations and areas of applicability. *IEEE Tr. Med. Im.*, 1(1):34–42, July 1982.

[3] M. F. Smith, C. E. Floyd, R. J. Jaszczak, and R. E. Coleman. Reconstruction of SPECT images using generalized matrix inverses. *IEEE Tr. Med. Im.*, 11(2):?, June 1992.

[4] A. O. Hero. A Cramer-Rao type lower bound for essentially unbiased parameter estimation. Technical Report 890, Lincoln Laboratory, MIT, ?, January 1992.

[5] J. A. Fessler and A. O. Hero. Cramer-Rao lower bounds for biased image reconstruction. In *Proc. Midwest Symposium on Circuits and Systems*, volume 1, pages 253–256, 1993.

[6] K. Sauer and C. Bouman. A local update strategy for iterative reconstruction from projections. *IEEE Tr. Sig. Proc.*, 41(2):534–48, February 1993.

[7] J. A. Fessler. Penalized weighted least-squares image reconstruction for positron emission tomography. *IEEE Tr. Med. Im.*, 13(2):290–300, June 1994.