Efficient Computational Methods for Wavelet Domain Signal Restoration Problems *

Eric L. Miller
235 Forsyth Building
Northeastern University
360 Huntington Ave.
Boston, MA 02215
Tel: (617) 373-8386
email: elmiller@cdsp.neu.edu

June 12, 1998

Abstract

We present an efficient, wavelet domain algorithm for computing the error variances associated with a wide class of linear inverse problems posed in a maximum *a posteriori* (MAP) estimation framework. Our method is based on the permutation and subsequent partitioning of the Fisher information matrix into a 2×2 block structure with the lower-right block well approximated as diagonal and significantly larger than the upper-left block. We prove that under appropriate conditions this diagonal approximation does in fact allow for the accurate recovery of the error variances, and we introduce a greedy-type method based on the optimization of a diagonal dominance criterion for determining the "best" partition. We demonstrate the speed of this technique and its accuracy for a set of inverse problems corresponding to a variety of blurring kernels, problem sizes, and noise conditions.

Submitted to *IEEE Trans. Signal Processing*
EDICS Category: SP 3.10

*This work was supported in part by NSF CAREER Grant MIP-9623721, subcontract GC123920NDG from Boston University under the AFOSR MURI Program on Reduced Signature Target Recognition, and DOE contract DE-FC07-95ID13395

1 Introduction

A common approach for recovering a signal given noisy, linear functionals of the original defines the reconstruction as the solution to a least-squares problem with a quadratic regularization term [2]. This formulation has the interpretation of solving a maximum *a posteriori* (MAP) estimation problem with additive Gaussian noise and a Gaussian prior model whose covariance structure is determined by the form of the regularizer [2]. The statistical perspective is useful because it provides a framework for performance analysis via the examination of error variance (EV) information. Unfortunately, computing the variances requires the inversion of the Fisher information matrix (FIM); a procedure whose $O(N^3)$ complexity is prohibitive for most multi-dimensional problems.

Here we present an efficient method for obtaining the EVs for a wide class of linear inverse problems posed in the wavelet transform domain. We are concerned particularly with problems characterized by space-varying blurs which provide primarily coarse scale information about the unknown signal along with limited fine scale detail. Such distortions arise in applications including ultrasonic imaging, geophysical prospecting, and non-destructive testing [4, 6, 7]. The work in this paper is motivated by our previous efforts [5, 6] where we observed that a wavelet formulation for such problems leads to a partitioning of the information matrix into a 2×2 block structure with the lower right block (the 22 block) well approximated as diagonal and significantly larger than the upper left block (the 11 block). In Section 3, we show how such a partitioning in principle can be used to extract error variances efficiently. Unfortunately as detailed in [7], constructing the partition actually required prior knowledge of the error variances.

We demonstrate here an efficient method for determining a partition and obtaining the variances in a manner which does not require this prior knowledge. We introduce and prove that a particular diagonal dominance criterion represents a useful means of evaluating a given partition. A greedy-type algorithm is employed to determine an approximation to the optimal partition for a given problem. Numerical experiments demonstrate that this method produces highly accurate EV information for a variety of image restoration problems at a fraction of the complexity required to directly invert the FIM.

2 Problem Formulation

We consider image restoration problems of the form

$$y(m_1, n_1) = \sum_{m_2,n_2} T(m_1, n_1, m_2, n_2) g(m_2, n_2) + w(m_1, n_1) \Longleftrightarrow \mathbf{y} = \mathbf{Tg} + \mathbf{w} \tag{1}$$

where $\mathbf{g}$ is the vector of lexicographically ordered pixels in $g(m, n)$, the image to be recovered; $\mathbf{T}$ is the matrix representing the blurring kernel, $T(m_1, n_1, m_2, n_2)$; and $\mathbf{w}$ is unit variance, zero mean, white Gaussian noise. Transformation of (1) to a multiscale domain is achieved by defining two orthonormal wavelet transform matrices, $\boldsymbol{\mathcal{W}}_y$ and $\boldsymbol{\mathcal{W}}_g$, and applying them to (1) as follows [5,6]

$$\boldsymbol{\mathcal{W}}_y \mathbf{y} = \left[\boldsymbol{\mathcal{W}}_y \mathbf{T} \boldsymbol{\mathcal{W}}_g^T\right] \left[\boldsymbol{\mathcal{W}}_g \mathbf{g}\right] + \boldsymbol{\mathcal{W}}_y \mathbf{w} \equiv \boldsymbol{\eta} = \boldsymbol{\Theta}\boldsymbol{\gamma} + \boldsymbol{\nu} \tag{2}$$

In (2) eg., $\boldsymbol{\mathcal{W}}_g$ takes $\mathbf{g}$ into a vector, $\boldsymbol{\gamma}$, comprised of all the wavelet coefficients and coarsest scale scaling coefficients in a two dimensional wavelet transform of the image.

The MAP estimate of $\boldsymbol{\gamma}$ from $\boldsymbol{\eta}$ is the solution to the following set of normal equations [5]

$$\left(\boldsymbol{\Theta}^T \boldsymbol{\Theta} + \kappa^{-2} \mathbf{P}_0^{-1}\right) \hat{\boldsymbol{\gamma}} = \boldsymbol{\Theta}^T \boldsymbol{\eta} \tag{3}$$

where $\kappa^2 \mathbf{P_0}$ is the prior covariance matrix for $\boldsymbol{\gamma}$, $\mathbf{L} = \boldsymbol{\Theta}^T \boldsymbol{\Theta} + \mathbf{P}_0^{-1}$ is the FIM, and $\mathbf{P}$, the error covariance matrix (ECM), is $\mathbf{L}^{-1}$. We model $\boldsymbol{\gamma}$ using a separable, two-dimensional $1/f$-type model where the wavelet coefficients are zero mean, uncorrelated, Gaussian random variables (hence $\mathbf{P_0}$ is diagonal) with a geometrically decreasing variance progression depending only on the scale index with finer coefficients having smaller variances [5]. The problem considered here is the computation of the error variances i.e. the diagonal elements of $\mathbf{L}^{-1}$, without the direct inversion of the FIM.

3 Variance Extraction

Our method for extracting the error variances is motivated by the results in [5–7] which indicated that the error variance analysis associated with a wavelet domain form of ill-posed problems can be used to divide $\hat{\gamma}$ into two sub-vectors: $\hat{\gamma}_1$ corresponding to those coefficients for which the data provides significant information and $\hat{\gamma}_2$ corresponding to those degrees of freedom for which the data are not informative. Given this decomposition, the rows and columns of (3) are re-ordered to obtain a permuted form of the normal equations:

$$\begin{bmatrix} \tilde{\mathbf{L}}_{11} & \tilde{\mathbf{L}}_{12} \\ \tilde{\mathbf{L}}_{12}^T & \tilde{\mathbf{L}}_{22} \end{bmatrix} \begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \end{bmatrix} = \begin{bmatrix} \mathbf{v}_1 \\ \mathbf{v}_2 \end{bmatrix} \tag{4}$$

where $\mathbf{v} = \mathbf{\Pi}\mathbf{\Theta}^T\mathbf{R}^{-1}\boldsymbol{\eta}$ and $\tilde{\mathbf{L}} = \mathbf{\Pi}\mathbf{L}\mathbf{\Pi}^T$ is the information matrix permuted using $\mathbf{\Pi}$. The block matrix inversion formula [3, Sec. 0.7.3] provides the means of computing $\tilde{\mathbf{P}} \equiv \mathbf{\Pi}\mathbf{P}\mathbf{\Pi}^T$ as

$$\tilde{\mathbf{P}} = \begin{bmatrix} \tilde{\mathbf{L}}_{11} & \tilde{\mathbf{L}}_{12} \\ \tilde{\mathbf{L}}_{12}^T & \tilde{\mathbf{L}}_{22} \end{bmatrix}^{-1} = \begin{bmatrix} \tilde{\mathbf{P}}_{11} & -\tilde{\mathbf{L}}_{11}^{-1}\tilde{\mathbf{L}}_{12}\tilde{\mathbf{P}}_{22} \\ -\tilde{\mathbf{P}}_{22}\tilde{\mathbf{L}}_{12}^T\tilde{\mathbf{L}}_{11}^{-1} & \tilde{\mathbf{P}}_{22} \end{bmatrix} \tag{5}$$

where $\tilde{\mathbf{P}}_{11} = \left(\tilde{\mathbf{L}}_{11} - \tilde{\mathbf{L}}_{12}\tilde{\mathbf{L}}_{22}^{-1}\tilde{\mathbf{L}}_{12}^T\right)^{-1}$ and $\tilde{\mathbf{P}}_{22} = \tilde{\mathbf{L}}_{22}^{-1} - \tilde{\mathbf{L}}_{22}^{-1}\tilde{\mathbf{L}}_{12}\tilde{\mathbf{P}}_{11}\tilde{\mathbf{L}}_{12}^T\tilde{\mathbf{L}}_{22}^{-1}$. It was observed in [7] that $\tilde{\mathbf{L}}_{11}$ was small and dense, $\tilde{\mathbf{L}}_{12}$ was rectangular and sparse, and $\tilde{\mathbf{L}}_{22}$ was large but well approximated as diagonal. Thus, the error variances can be computed by (a) approximating $\tilde{\mathbf{L}}_{22}$ as diagonal and trivially inverting this matrix, (b) constructing the small $\tilde{\mathbf{P}}_{11}$ matrix directly, and (c) calculating only the diagonal elements of $\tilde{\mathbf{P}}_{22}$. The diagonal approximation to $\tilde{\mathbf{L}}_{22}$ and the small size of $\tilde{\mathbf{P}}_{11}$ makes such a procedure far more efficient than computing the full inverse of $\mathbf{L}$. The key difficulty is that the initial specification of $\mathbf{\Pi}$ required knowledge of the error variances. The objective of the work here is the efficient construction of $\mathbf{\Pi}$ in a way which is *not* based on this prior knowledge.

The goal of permuting the columns and rows of $\mathbf{L}$ is to find a high dimensional $\tilde{\mathbf{L}}_{22}$ block which can be well approximated as diagonal. Specifically, we seek an $\tilde{\mathbf{L}}_{22}$ which is maximally diagonally dominant. A matrix $\mathbf{A}$ is strictly diagonally dominant [3, Sec. 6.1.9] if for all j, $[\mathbf{s}(\mathbf{A})]_j \equiv \sum_{j\neq i} \left|[\mathbf{A}]_{ij}\right| < \left|[\mathbf{A}]_{jj}\right|$. To gauge the degree of diagonal dominance, we define the non-negative quantity $\Delta(\mathbf{A}) = \max_j [\mathbf{s}(\mathbf{A})]_j \,/ \left|[\mathbf{A}]_{jj}\right|$ so that $\mathbf{A}$ is strictly diagonally dominant iff $\Delta < 1$.

The motivation for the use of Δ is as follows. Let $\tilde{\mathbf{L}}$ be a symmetric positive definite FIM partitioned as in (5), $\mathbf{D}$ be diagonal with $[\mathbf{D}]_{ii} = \left[\tilde{\mathbf{L}}_{22}\right]_{ii}$, and set $\bar{\mathbf{L}}$ equal to $\tilde{\mathbf{L}}$ with $\tilde{\mathbf{L}}_{22}$ replaced by $\mathbf{D}$. Letting $\tilde{\sigma}_i^2$ (resp. $\bar{\sigma}_i^2$) be the ith diagonal element of $\tilde{\mathbf{L}}^{-1}$ (resp. $\bar{\mathbf{L}}^{-1}$) then if $\bar{\mathbf{L}}$ remains a symmetric positive definite FIM, we show in Appendix A that $\sum_i \left| \tilde{\sigma}_i^2 - \bar{\sigma}_i^2 \right| \leq \Delta\left(\tilde{\mathbf{L}}_{22}\right) \Upsilon\left(\tilde{\mathbf{L}}_{22}\right)$ where $\Upsilon\left(\tilde{\mathbf{L}}_{22}\right)$ is explicitly a function of $\tilde{\mathbf{L}}_{22}$ and implicitly a function of Δ. Ignoring Υ for a moment, the error in computing the variances which arises from the diagonal approximation to $\tilde{\mathbf{L}}_{22}$ decreases as Δ increases. $\Upsilon\left(\tilde{\mathbf{L}}_{22}\right)$ depends on $\tilde{\mathbf{L}}_{22}$, and hence $\Delta\left(\tilde{\mathbf{L}}_{22}\right)$, in a complicated manner not particularly amenable to analysis. In Appendix A we show that if $\Delta\left(\tilde{\mathbf{L}}_{22}\right)$ is decreased by lowering κ in (3) for those coefficients in the 22 block, (i.e. by increasing regularization for those coefficients for which the data are uninformative), then as $\kappa \to 0$, $\Upsilon\left(\tilde{\mathbf{L}}_{22}\right) \to 0$. Thus, at least for ill-posed problems requiring a relatively large amount of regularization, small $\Delta\left(\tilde{\mathbf{L}}_{22}\right)$ corresponds to small errors in the variance computation. This is the behavior seen in the examples of Section 4 and holds even for problems with relatively little regularization.

The combinatorial nature of determining the optimal $\mathbf{L}_{22}$ knowing neither the appropriate size nor the best collection of columns/rows for a given size prohibits an exhaustive search. Here we employ a greedy type method to find an approximate solution to this problem. We take the best 1×1 $\tilde{\mathbf{L}}_{22}$ as that with the smallest Δ value. We next cycle through the remaining rows to generate that 2×2 submatrix with the smallest Δ. The procedure is the repeated for $n = 3, 4, \ldots$ and is terminated when Δ exceeds some threshold < 1. Although we make no claim as to the optimality of this process, it is fast, guaranteed to terminate, and in practice has very strong performance. Finally, to further reduce the complexity, we avoid a full search at each stage. Rather, we accept the first column which yields an acceptably small (here taken as 0.02) increase in Δ.

4 Examples

Here we demonstrate the performance of the algorithm described in Section 3 for a variety of image restoration kernels and noise conditions. The signal-to-noise ratio (SNR) is determined by appropriate choice of κ in (3). The computational load of our algorithm is compared against direct inversion of the information matrix using the MATLAB **inv** function. The complexity of each approach is measured using the MATLAB **flops** facility. The accuracy is evaluated using the relative error in the variance computation defined as $\sum_i \left| \sigma_i^2 - \bar{\sigma}_i^2 \right| / \sum_i \left| \sigma_i^2 \right|$ where σ_i^2 are the exact error variances and $\bar{\sigma}_i^2$ are the error variances produced by our algorithm. In all examples, we use that $\tilde{\mathbf{L}}_{22}$ corresponding to a Δ value of 0.5. Finally, $\mathbf{\Theta}^T\mathbf{\Theta}$ is sparsified by setting to zero all elements smaller than $\frac{1e-2}{N}\|\mathbf{\Theta}^T\mathbf{\Theta}\|_\infty$ with N the number of columns in $\mathbf{\Theta}^T\mathbf{\Theta}$ [1].

We first consider the performance of our algorithm for the non-convolutional kernels of the form

$$T\left(m_1, n_1, m_2, n_2\right) = \exp\left[-\frac{\left(m_1 - n_1\right)^2}{2\sigma^2(m_1)}\right] \exp\left[-\frac{\left(m_2 - n_2\right)^2}{2\sigma^2(m_2)}\right] \tag{6}$$

with $m_1, m_2, n_1, n_2 = 1, 2, \ldots N$ and $\sigma^2(m) = \zeta\left[1 + \beta\cos\left(2\pi f m\right)\right]$. This choice of T represents convolution with separable Gaussian kernels whose widths vary sinusoidally. The degree of variation is set by β. Choosing $\beta = 0$ corresponds to simple convolution with a Gaussian blurring function. The data obtained with larger β contain information at a variety of resolutions at different points in space. For the experiments below, we use $\zeta = 2$ and $f = 1/16$.

In 1(a) the FLOP counts are displayed as a function of the problem size for three different SNRs. At an SNR $= 0$ exact inversion requires about 33 times more FLOPS than our approach for a 16×16 image with this number increasing to 180 for the 64×64 case. As the SNR rises one can recover reliably a larger number of coefficients in $\boldsymbol{\gamma}$. Thus, the size of the $\tilde{\mathbf{L}}_{11}$ matrix rises and the efficiency of our approach declines somewhat. Nonetheless, even at an SNR of 10, exact inversion still requires well over an order of magnitude more operations than the technique described here for the 64×64 problem. In Fig. 1(b)–(c) it is seen that the relative error in the variances is below 10^{-3} for all cases thereby demonstrating the accuracy of this approach.

Comparable results are displayed in Fig. 2 for $\beta = 0.8$. At an SNR of 0, exact inversion requires 33 time more FLOPS for the $N = 16$ image with this number rising to 160 for $N = 64$. Even at an SNR of 10, the partitioned inversion method is over 30 times less costly than exact inversion. For all SNRs and problem sizes the algorithm is accurate to within 0.3%.

The second type of reconstruction problem considered here is a linearized inverse electrical conductivity problem described in [6, 7]. A detailed description of the highly non-convolutional $\mathbf{T}$ may be found in [4]. The length of $\mathbf{y}$ is 1152 and we consider performance for the discretization of A into 16×16 and 32×32 arrays of square pixels. Finer discretization typically is unwarranted given the frequencies of interest in this problem. The error variance results for this problem are illustrated in Fig. 3. At an SNR of 10, our approach is almost 50 times less intensive than direct inversion for the 32×32 case. In Fig. 3(b)–(c) we see that while the accuracy of our algorithm is somewhat lower as compared to the Gaussian case the worst case relative error is still only 1.2%.

5 Conclusion

We have developed efficient methods for the computation of error variance information for a class of linear restoration problems specified in the transform domain. The approach is based on the permutation and partitioning of the FIM to maximize a diagonal dominance criterion. A suboptimal greedy approach has been employed to obtain the required permutation in practice.

While the work here has concentrated exclusively on multiscale inverse problems, we note that the use of wavelets is in fact not necessary. The only requirements are a diagonal regularization matrix, a sparse transform domain matrix, and the property that the unknowns can in fact be partitioned into sets for which the data do and do not provide significant information. Application of our methods to other problems possessing these characteristics represents an interesting and useful avenue for future effort.

A Error Due to Diagonal Approximation of $\tilde{\mathbf{L}}_{22}$

Here we explore the error caused by diagonal approximation to $\tilde{\mathbf{L}}_{22}$. Dropping the tilde notation, let $\mathbf{P} = \mathbf{L}^{-1}$ (resp. $\bar{\mathbf{P}} = \bar{\mathbf{L}}^{-1}$) be the error covariance matrix associated with $\mathbf{L}$ (resp. $\bar{\mathbf{L}}^{-1}$). Because the diagonal elements of $\mathbf{P}$ and $\bar{\mathbf{P}}$ correspond to error variances

$$\sum_i \left| \sigma_i^2 - \bar{\sigma}_i^2 \right| = \mathrm{tr} \left| \mathbf{P} - \bar{\mathbf{P}} \right| = \mathrm{tr} \left| \mathbf{P}_{11} - \bar{\mathbf{P}}_{11} \right| + \mathrm{tr} \left| \mathbf{P}_{22} - \bar{\mathbf{P}}_{22} \right| \tag{7}$$

where $|\mathbf{A}|$ is the matrix constructed of the absolute values of the elements of $\mathbf{A}$ and

$$\mathbf{P}_{11}^{-1} = \mathbf{L}_{11} - \mathbf{L}_{12}\mathbf{L}_{22}^{-1}\mathbf{L}_{12}^T \qquad \bar{\mathbf{P}}_{11}^{-1} = \mathbf{L}_{11} - \mathbf{L}_{12}\mathbf{D}^{-1}\mathbf{L}_{12}^T \tag{8a}$$

$$\mathbf{P}_{22}^{-1} = \mathbf{L}_{22} - \mathbf{L}_{12}^T\mathbf{L}_{11}^{-1}\mathbf{L}_{12} \qquad \bar{\mathbf{P}}_{22}^{-1} = \mathbf{D} - \mathbf{L}_{12}^T\mathbf{L}_{11}^{-1}\mathbf{L}_{12}. \tag{8b}$$

We first consider $\mathrm{tr} \left| \mathbf{P}_{22} - \bar{\mathbf{P}}_{22} \right|$ where

$$\begin{aligned} \mathrm{tr} \left| \mathbf{P}_{22} - \bar{\mathbf{P}}_{22} \right| &= \mathrm{tr} \left| \mathbf{P}_{22} \left(\mathbf{I} - \mathbf{P}_{22}^{-1} \bar{\mathbf{P}}_{22} \right) \right| = \sum_i \left| \sum_j \left[\mathbf{P}_{22} \right]_{ij} \left[\mathbf{I} - \mathbf{P}_{22}^{-1} \bar{\mathbf{P}}_{22} \right]_{ji} \right| && (9) \\ &\le \sum_i \sum_j \left| \left[\mathbf{P}_{22} \right]_{ji} \right| \left| \left[\mathbf{I} - \mathbf{P}_{22}^{-1} \bar{\mathbf{P}}_{22} \right]_{ji} \right| && (10) \\ &\le \max_{ij} \left| \left[\mathbf{P}_{22} \right]_{ij} \right| \sum_{ij} \left| \left[\mathbf{I} - \mathbf{P}_{22}^{-1} \bar{\mathbf{P}}_{22} \right]_{ji} \right| \le \alpha N_{22} \| \mathbf{I} - \mathbf{P}_{22}^{-1} \bar{\mathbf{P}}_{22} \|_1 \end{aligned}$$

where (10) follows from (9) via the triangle and Schwartz inequalities, $\alpha \equiv \left| \max_{ij} \left[\mathbf{P}_{22} \right]_{ij} \right|$, and $\mathbf{P}_{22}$ is of size $N_{22} \times N_{22}$. Straightforward linear algebra and (8b) allows $\mathbf{I} - \mathbf{P}_{22}^{-1} \bar{\mathbf{P}}_{22}$ to be written as $\left(\mathbf{I} - \mathbf{L}_{22} \mathbf{D}^{-1} \right) \left(\mathbf{I} - \mathbf{C} \mathbf{D}^{-1} \right)$ where $\mathbf{C} = \mathbf{L}_{12}^T \mathbf{L}_{11}^{-1} \mathbf{L}_{12}$. Thus,

$$\begin{aligned} \mathrm{tr} \left| \mathbf{P}_{22} - \bar{\mathbf{P}}_{22} \right| &\le \alpha N_{22} \| \mathbf{I} - \mathbf{C} \mathbf{D}^{-1} \|_1 \| \mathbf{I} - \mathbf{L}_{22} \mathbf{D}^{-1} \|_1 && (11) \\ &= \alpha N_{22} \| \mathbf{I} - \mathbf{C} \mathbf{D}^{-1} \|_1 \max_j \sum_i \left| \delta_{ij} - \frac{\left[\mathbf{L}_{22} \right]_{ij}}{\left[\mathbf{L}_{22} \right]_{jj}} \right| = \Delta \left(\mathbf{L}_{22} \right) \alpha N_{22} \| \mathbf{I} - \mathbf{C} \mathbf{D}^{-1} \|_1. && (12) \end{aligned}$$

which is the desired bound for the 22 block.

Next, consider $\mathrm{tr} \left(\mathbf{P}_{11} - \bar{\mathbf{P}}_{11} \right)$. Using the matrix inversion lemma we have $\mathbf{P}_{11} = \mathbf{L}_{11}^{-1} + \mathbf{A} \mathbf{P}_{22} \mathbf{A}^T$ and $\bar{\mathbf{P}}_{11} = \mathbf{L}_{11}^{-1} + \mathbf{A} \bar{\mathbf{P}}_{22} \mathbf{A}^T$ where $\mathbf{A} = \mathbf{L}_{11}^{-1} \mathbf{L}_{12}$. Hence,

$$\begin{aligned} \mathrm{tr} \left| \mathbf{P}_{11} - \bar{\mathbf{P}}_{11} \right| &= \mathrm{tr} \left| \mathbf{A} \left(\mathbf{P}_{22} - \bar{\mathbf{P}}_{22} \right) \mathbf{A}^T \right| \le \sum_{ikj} \left| \left[\mathbf{A} \right]_{ij} \left[\mathbf{A} \right]_{ik} \right| \left| \left[\mathbf{P}_{22} - \bar{\mathbf{P}}_{22} \right]_{jk} \right| \\ &\le \beta \sum_{jk} \left| \left[\mathbf{P}_{22} - \bar{\mathbf{P}}_{22} \right]_{jk} \right| \le \beta N_{22} \| \mathbf{P}_{22} - \bar{\mathbf{P}}_{22} \|_1 \le \beta N_{22} \| \mathbf{P}_{22} \|_1 \| \mathbf{I} - \mathbf{P}_{22}^{-1} \bar{\mathbf{P}}_{22} \| \\ &\le \Delta \left(\mathbf{L}_{22} \right) \beta N_{22} \| \mathbf{P}_{22} \|_1 \| \mathbf{I} - \mathbf{C} \mathbf{D}^{-1} \|_1. && (13) \end{aligned}$$

where $\beta \equiv \max_{ijk} \left| \left[\mathbf{A} \right]_{ij} \left[\mathbf{A} \right]_{ik} \right|$.

Finally, from (12) and (13),

$$\mathrm{tr} \left| \mathbf{L}^{-1} - \bar{\mathbf{L}}^{-1} \right| \le \Delta \left(\mathbf{L}_{22} \right) N_{22} \| \mathbf{I} - \mathbf{L}_{12}^T \mathbf{L}_{11}^{-1} \mathbf{L}_{12} \mathbf{D}^{-1} \|_1 \left[\alpha + \beta \| \mathbf{P}_{22} \|_1 \right] \tag{14}$$

which completes the derivation with $\Upsilon \left(\tilde{\mathbf{L}}_{22} \right) = N_{22} \| \mathbf{I} - \mathbf{L}_{12}^T \mathbf{L}_{11}^{-1} \mathbf{L}_{12} \mathbf{D}^{-1} \|_1 \left[\alpha + \beta \| \mathbf{P}_{22} \|_1 \right]$.

Eq. (14) indicates that the error made in approximating the 22 block in $\mathbf{L}$ is proportional to $\Delta\left(\mathbf{L}_{22}\right)$; however, Υ is a function of $\mathbf{L}_{22}$ through α and the 1-norms of $\mathbf{P}_{22}$ and $\mathbf{I}-\mathbf{L}_{12}^T\mathbf{L}_{11}^{-1}\mathbf{L}_{12}\mathbf{D}^{-1}$. To partially examine the behavior of Υ, we argue that decreasing $\Delta\left(\mathbf{L}_{22}\right)$ in a natural manner decreases this quantity as well. Assume that $\mathbf{\Theta}$ is fixed and Δ is decreased by decreasing the value of the κ parameter in the prior model specifically for the coefficients in the 22 block. Now, it is not hard to show that $\mathbf{D}^{-1} \xrightarrow{\kappa\rightarrow 0} \mathbf{0}$ so that with $\mathbf{B}=\mathbf{L}_{12}^T\mathbf{L}_{11}^{-1}\mathbf{L}_{12}$

$$\begin{aligned}
\|\mathbf{I}-\mathbf{B}\mathbf{D}^{-1}\|_1 &= \|\mathbf{I}-\mathbf{B}\mathbf{D}^{-1}+\mathbf{D}^{-1}-\mathbf{D}^{-1}\|_1 \leq \|\mathbf{I}-\mathbf{D}^{-1}\|_1+\|\mathbf{D}^{-1}-\mathbf{B}\mathbf{D}^{-1}\|_1 \\
&\leq \|\mathbf{I}-\mathbf{D}^{-1}\|_1+\|\mathbf{D}^{-1}\|_1\|\mathbf{I}-\mathbf{B}\|_1 \xrightarrow{\kappa\rightarrow 0} \|\mathbf{I}\|_1=1.
\end{aligned}$$

Hence, asymptotically, $\|\mathbf{I}-\mathbf{B}\mathbf{D}^{-1}\|_1$ is independent of $\tilde{\mathbf{L}}_{22}$. Referring to (8b), it is not difficult to show that as κ decreases $\mathbf{P}_{22}\rightarrow\kappa^2\mathbf{P}_{0,22}$ where $\mathbf{P}_{0,22}$ is the 22 block of the appropriately permuted form of $\mathbf{P}_0$. So as $\kappa\rightarrow 0$ both α and $\|\mathbf{P}_{22}\|_1$ go to 0. Thus, we conclude that decreasing $\Delta\left(\tilde{\mathbf{L}}_{22}\right)$ by varying the degree of regularization will cause $\Upsilon\rightarrow 0$.

References

[1] B. Alpert, G. Beylkin, R. Coifman, and V. Rokhlin. Wavelets for the fast solution of second-kind integral equations. *SIAM J. Sci. Comput.*, 14(1):159–184, 1993.

[2] Guy Demoment. Image reconstruction and restoration: Overview of common estimation structures and problems. *IEEE Trans on ASSP*, 37(12):2024–2036, 1989.

[3] Roger A. Horn and Charles R.. Johnson. *Matrix Analysis.* Cambridge Univ. Press, Cambridge, England, 1985.

[4] Eric L. Miller. *The Application of Multiscale and Statistical Techniques to the Solution of Inverse Problems.* PhD thesis, MIT, August 1994.

[5] Eric L. Miller and Alan S. Willsky. A multiscale approach to sensor fusion and the solution of linear inverse problems. *Appl. Comput. Harmonic Anal.*, 2:127–147, 1995.

[6] Eric L. Miller and Alan S. Willsky. Multiscale, statistically-based inversion scheme for the linearized inverse scattering problem. *IEEE Trans. Geosc. Remote Sens.*, 34(2):346–357, March 1996.

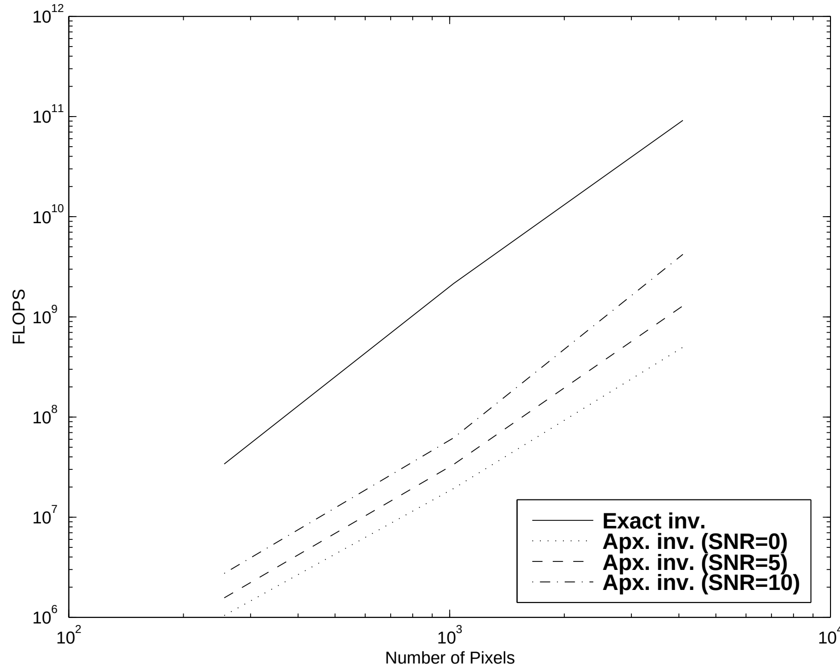
[7] Eric L. Miller and Alan S. Willsky. Wavelet-based methods for the nonlinear inverse scattering problem using the Extended Born Approximation. *Radio Sci.*, 31(1):51–67, Jan–Feb 1996.

B Figure and Table Captions

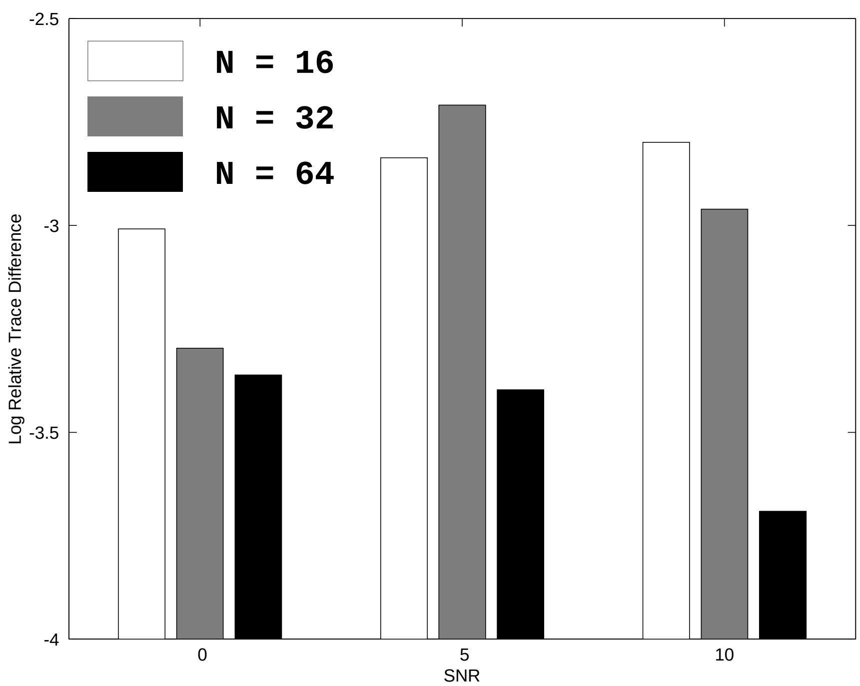
Figure 1 Inversion results for space varying Gaussian blur with $\beta = 0.2$

Figure 2 Inversion results for space varying Gaussian blur with $\beta = 0.8$

Figure 3 Inversion results for linearized inverse scattering example

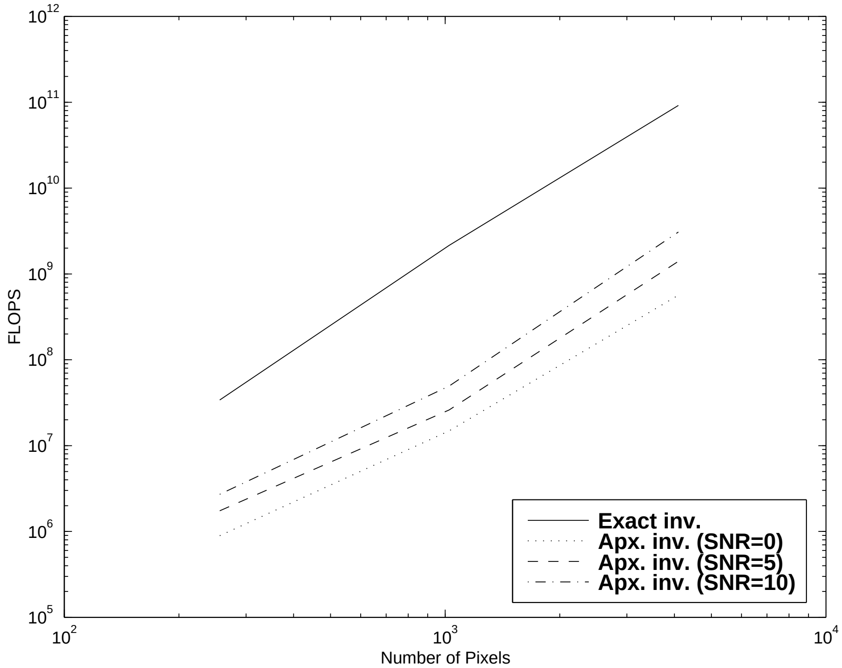


(a) FLOP Comparison

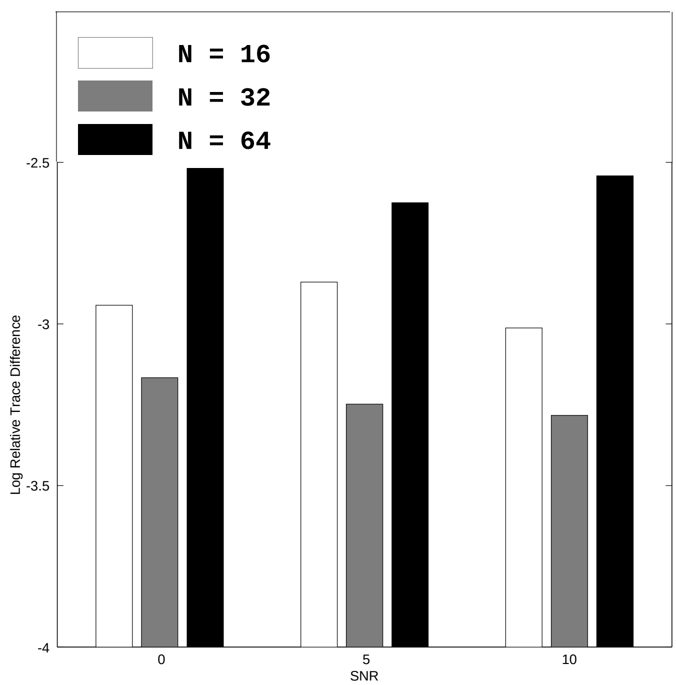


(b) Relative trace difference

Figure 1:

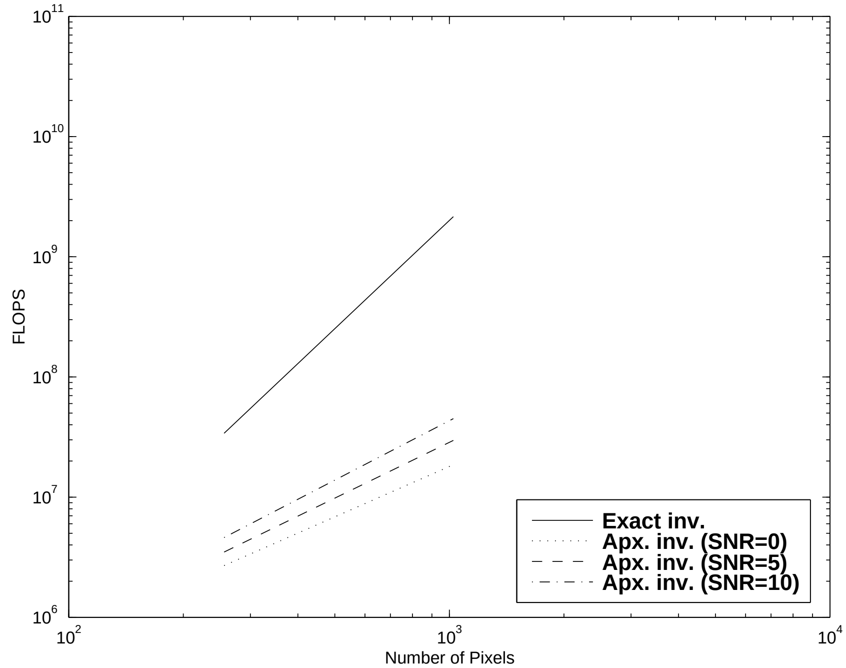


(a) FLOP Comparison

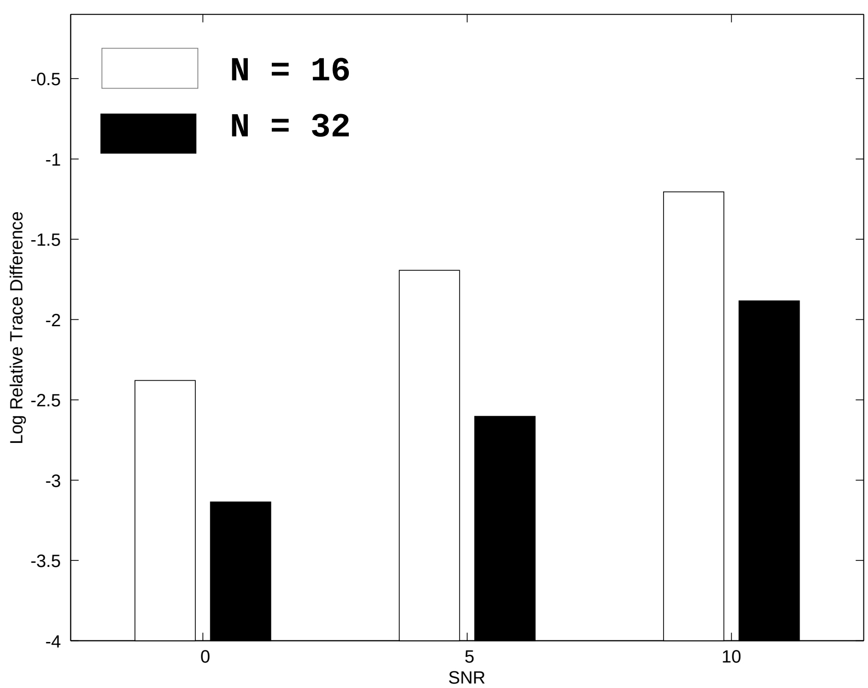


(b) Relative trace difference

Figure 2:



(a) FLOP Comparison



(b) Relative trace difference

Figure 3: