

BLIND DECONVOLUTION FOR SPARSE MOLECULAR IMAGING

Kyle Herrity^{1*}, Raviv Raich², and Alfred O. Hero III¹

¹ Department of EECS, University of Michigan, Ann Arbor, MI 48109-2122, USA

² School of EECS, Oregon State University, Corvallis, OR 97331-5501, USA

{kherrity,hero}@umich.edu, raich@eecs.oregonstate.edu

ABSTRACT

This paper considers the image reconstruction problem when the original image is assumed to be sparse and when limited information of the point spread function (PSF) is available. In particular, we are interested in reconstructing the magnetization density given Magnetic Resonance Force Microscopy (MRFM) image data, and an alternating iterative algorithm is presented to solve this problem. Simulations demonstrate its performance not only in the reconstruction of the original image, but also in the recovery of the partially known PSF. In addition, we suggest the introduction of a smoothing penalty on allowable PSFs to improve the reconstruction.

Index Terms— Image restoration, Blind deconvolution, Magnetic resonance force microscopy, Sparseness regularization, Optimization transfer

1. INTRODUCTION

The objective of image deconvolution is to recover a noisy and blurred version of an image. Common techniques of achieving this include Maximum-Likelihood (ML) and least squares formulations, when the statistical properties of the noise are at hand, and appropriate regularizing measures are taken.

In MRFM, a relatively new imaging technology capable of atomic-level resolution, the PSF of the system is sometimes known up to some prescribed error tolerance. In the previous formulation of this problem [1], pathological estimates of the true PSF were allowed. However, from the physical nature of a MRFM experiment, it makes sense to restrict our attention to a space of PSFs which have certain smooth characteristics, to be defined more precisely later. This effectively reduces the search space of the estimated image and gives rise to a more realistic estimate of the true PSF. We consider a parametric model of the PSF, $h(\theta)$, in [2], which includes parameters such as the externally applied magnetic field, B_{ext} , and the magnitude of the rf field, B_1 . We then linearize about a nominal PSF, i.e. $h(\theta) = h(\theta_0) + h'(\theta)(\theta - \theta_0)$. By linearizing

***Acknowledgement:** This work is partially supported by the ARO MURI grant W911NF-05-1-0403.

the PSF about its nominal value, we can recast the norm of the error as a non-isotropic quadratic criterion.

The topic of image deconvolution has been explored in detail [3]. Satisfactory reconstruction results have been obtained in [4] by employing a total variation (TV) based reconstruction method in an alternating minimization algorithm, which also simultaneously reconstructs the unknown PSF and image. The authors make minimal a priori assumptions on the PSF, but state the motivation for using TV is due to the fact that PSFs can have edges.

In this paper, certain modifications are made to the classical deconvolution problem. Namely, we make the assumptions that the original image is sparse in the image domain, and that partial knowledge of the PSF of the imaging system is available. The hope is that this a priori information will bolster the performance of our reconstruction algorithm.

2. PROBLEM STATEMENT

Here, we recall the problem of reconstructing a blurred and noisy image that is sparse. We assume the following model

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \quad (1)$$

where the original image is denoted as $\mathbf{x} \in \mathbb{R}^n$, the blurring matrix as $\mathbf{H} \in \mathbb{R}^{m \times n}$, and the noise vector is denoted by $\mathbf{n} \in \mathbb{R}^m$. When \mathbf{H} describes a convolution, it is Toeplitz, and if we zero-pad the image appropriately, \mathbf{H} is circulant, and hence diagonalizable by the discrete Fourier transform (DFT) matrix.

If \mathbf{n} is a zero-mean white Gaussian noise vector, then the maximum likelihood (ML) estimator of \mathbf{x} is the minimizer of the cost function

$$J(\mathbf{x}) = \|\mathbf{H}\mathbf{x} - \mathbf{y}\|^2. \quad (2)$$

We use $\|\cdot\|$ to denote the l_2 norm $\|\cdot\|_2$, where $\|\mathbf{x}\|_2 \triangleq \sum_{i=1}^n x_i^2$. Other norms such as l_1 , which is defined as $\|\mathbf{x}\|_1 \triangleq \sum_{i=1}^n |x_i|$, will be written explicitly. We assume the matrix \mathbf{H} is partially known, i.e., $\mathbf{H} = \mathbf{H}_0 + \epsilon\mathbf{\Delta}$, where

$$\|\mathbf{\Delta}\|_{\mathbf{w}} \triangleq \|\mathbf{W}\mathbf{\Delta}\| \leq \epsilon, \quad (3)$$

and \mathbf{W} is a non-identity smoothing matrix. The l_2 norm of a

matrix \mathbf{C} will be defined as $\|\mathbf{C}\| \triangleq \max_{\mathbf{s} \neq \mathbf{0}} \frac{\|\mathbf{C}\mathbf{s}\|}{\|\mathbf{s}\|}$.

Previously, we considered the case when Δ had unity l_2 norm, i.e., $\|\Delta\| \leq 1$. Intuitively, we now rule out bad estimates \mathbf{H}_0 of the PSF which may have satisfied the previous constraint of $\|\mathbf{H} - \mathbf{H}_0\| \leq \epsilon$, e.g., if we contaminate \mathbf{H} with white noise. Thus the cost function in (2) can be expressed as

$$J(\mathbf{x}) = \|(\mathbf{H}_0 + \epsilon\Delta)\mathbf{x} - \mathbf{y}\|^2. \quad (4)$$

We study the minimax criterion in order to remove the dependence on Δ . That is, we look for the \mathbf{x} which minimizes (4) so that the cost function becomes

$$J(\mathbf{x}) = \max_{\Delta} \|(\mathbf{H}_0 + \epsilon\Delta)\mathbf{x} - \mathbf{y}\|^2 \quad \text{s.t.} \quad \|\Delta\|_{\mathbf{w}} \leq \epsilon. \quad (5)$$

Since we view the linear operator \mathbf{H} acting on \mathbf{x} to be the convolution operator, by the commutative property of convolution, we may write $\mathbf{H}\mathbf{x} = \mathbf{X}\mathbf{h}$, where now \mathbf{X} is assumed to be a circulant matrix and hence diagonalizable by the DFT matrix, which we will denote by \mathbf{F} . Thus, we write

$$\mathbf{X} = \mathbf{F}\text{diag}(\mathbf{F}^H \mathbf{x})\mathbf{F}^H, \quad (6)$$

where \mathbf{F}^H denotes the conjugate transpose of \mathbf{F} . Similarly,

$$\begin{aligned} \mathbf{H} &= \mathbf{F}\text{diag}(\mathbf{F}^H \mathbf{h})\mathbf{F}^H \\ \Delta &= \mathbf{F}\text{diag}(\mathbf{F}^H \delta)\mathbf{F}^H. \end{aligned}$$

Here, $\text{diag}(\mathbf{x})$ represents a diagonal matrix whose entries are the elements of \mathbf{x} . This treatment is amenable to analysis and yields computational savings, as matrix multiplication becomes convolution, which may be efficiently implemented via the FFT.

2.1. The Sparse Constraint

Recall that the sparsity of an image \mathbf{x} is the number of nonzero elements of \mathbf{x} . More precisely, the sparsity is defined in terms of the l_0 norm as

$$\|\mathbf{x}\|_0 \triangleq \sum_{i=1}^n I(\mathbf{x}_i \neq 0), \quad (7)$$

where $I(\cdot)$ is the indicator function. Suppose we are given $\tilde{\mathbf{H}}$, and we wish to minimize the following objective function

$$\min_{\mathbf{x}} \|\tilde{\mathbf{H}}\mathbf{x} - \mathbf{y}\|^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq p. \quad (8)$$

This problem is combinatorial in nature, with a total of $\sum_{i=0}^p \binom{n}{i}$ possible solutions, and can be shown to be NP-hard. A well-known method for solving problems of this type is by employing convex relaxation to the problem. That is, we replace the l_0 constraint with an l_1 constraint, and under certain conditions on $\tilde{\mathbf{H}}$, each formulation yields the same solution. Therefore, we instead seek to minimize

$$\min_{\mathbf{x}} \|\tilde{\mathbf{H}}\mathbf{x} - \mathbf{y}\|^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_1 \leq \tilde{p}. \quad (9)$$

By introducing a sparse constraint on the image, we now seek to find the optimal sparse image. We use the method of La-

grange multipliers to express (9) as

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}} \|\tilde{\mathbf{H}}\mathbf{x} - \mathbf{y}\|^2 + \lambda \|\mathbf{x}\|_1. \quad (10)$$

We consider two approaches for solving (9). In the first approach, the solution is obtained via an iterative thresholding technique described in [5]. If we let s be the largest singular value of $\tilde{\mathbf{H}}$, the minimization can be achieved by minimizing with respect to each \mathbf{x}_i independently as

$$\hat{\mathbf{x}}_i^{(n+1)} = f_{\frac{\lambda}{2}}([\hat{\mathbf{x}}^{(n)} + \tilde{\mathbf{H}}^T(\frac{\mathbf{y}}{s^2} - \tilde{\mathbf{H}}\hat{\mathbf{x}}^{(n)})]_i), \quad (11)$$

where $f_{\theta}(\cdot)$ is the (nonlinear) soft thresholding operator, which acts element-wise on a vector \mathbf{x} as follows

$$f_{\theta}(x) = (x - \text{sgn}(x)\frac{\theta}{2})I(|x| \geq \frac{\theta}{2}). \quad (12)$$

The second approach utilizes the concept of optimization transfer, described in [1]. If we denote our optimality criterion by $F(\mathbf{x})$, the idea is to find a non-negative function $Q(\mathbf{x}, \mathbf{x}')$, i.e., $Q(\mathbf{x}, \mathbf{x}') \geq 0$, and such that $Q(\mathbf{x}, \mathbf{x}') = 0$ if and only if $\mathbf{x} = \mathbf{x}'$. Then the iterations

$$\hat{\mathbf{x}}^{(n+1)} = \arg \min_{\mathbf{x}} F(\mathbf{x}) + Q(\mathbf{x}, \hat{\mathbf{x}}^{(n)}), \quad (13)$$

are such that $F(\hat{\mathbf{x}}^{(n)})$ is a non-increasing function of n . Let $F(\mathbf{x}) = \|\tilde{\mathbf{H}}\mathbf{x} - \mathbf{y}\|^2 + \lambda \|\mathbf{x}\|_1$ in (10). Now observe that for all x, x' in \mathbb{R} , where $x' \neq 0$,

$$|x| \leq \frac{x^2}{2|x'|} + \frac{|x'|}{2}, \quad (14)$$

Thus, for all \mathbf{x}' whose components are nonzero,

$$F(\mathbf{x}) \leq \|\tilde{\mathbf{H}}\mathbf{x} - \mathbf{y}\|^2 + \frac{\lambda}{2}(\mathbf{x}^T \text{diag}(\frac{1}{|\mathbf{x}'|})\mathbf{x} + \|\mathbf{x}'\|_1). \quad (15)$$

Additional care must be taken when one or more of the components of \mathbf{x}' is zero. Therefore, we set

$$\hat{\mathbf{x}}^{(n+1)} = \arg \min_{\mathbf{x}} \|\tilde{\mathbf{H}}\mathbf{x} - \mathbf{y}\|^2 + \frac{\lambda}{2}(\mathbf{x}^T \text{diag}(\frac{1}{|\hat{\mathbf{x}}^{(n)}|})\mathbf{x}). \quad (16)$$

The surrogate function introduced for $\|\mathbf{x}\|_1$ was inspired by [6]. By differentiating with respect to \mathbf{x} , it can be shown that

$$\hat{\mathbf{x}}^{(n+1)} = \left(\tilde{\mathbf{H}}^T \tilde{\mathbf{H}} + \lambda \cdot \text{diag}(\frac{1}{|\hat{\mathbf{x}}^{(n)}|}) \right)^{-1} \tilde{\mathbf{H}}^T \mathbf{y}. \quad (17)$$

Assuming the inverse above exists, we compute it via a series of Landweber iterations. As the sparsity increases, the computation time of (17) decreases rapidly.

2.2. The Smoothness Constraint

As mentioned in the introduction, we are imposing a smoothness penalty on allowable PSFs. Thus, we seek to maximize the following cost function, while taking advantage of the conventions set out earlier, toward a minimax criterion.

Hence, Equation (5) becomes

$$\hat{\delta} = \arg \max_{\delta} \|\mathbf{e} + \epsilon \mathbf{X} \delta\|^2 \quad \text{s.t.} \quad \|\mathbf{W} \delta\|^2 \leq \epsilon, \quad (18)$$

where \mathbf{e} denotes the observation error vector, i.e., $\mathbf{e} = \mathbf{H}_0 \mathbf{x} - \mathbf{y}$. By employing Lagrange multipliers, (18) may be expressed as

$$\hat{\delta} = \arg \min_{\delta} -\|\mathbf{e} + \epsilon \mathbf{X} \delta\|^2 + \gamma \|\mathbf{W} \delta\|^2. \quad (19)$$

By differentiating (19) with respect to δ , and equating to zero, the minimizer becomes

$$\hat{\delta} = \epsilon (\gamma \mathbf{W}^T \mathbf{W} - \epsilon^2 \mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{e}. \quad (20)$$

By incorporating the sparse constraint in Section 2.1, the new regularized cost function becomes

$$\begin{aligned} J(\mathbf{x}; \delta) &= \|(\mathbf{H}_0 + \epsilon \mathbf{\Delta}) \mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{X}(\mathbf{h}_0 + \epsilon \delta) - \mathbf{y}\|^2 \\ \text{s.t.} \quad &\|\mathbf{x}\|_1 \leq \tilde{p} \text{ and } \|\mathbf{W} \delta\|^2 \leq \epsilon. \end{aligned} \quad (21)$$

3. ALTERNATING ALGORITHM

The alternating algorithm to solve the regularized cost function (21) proceeds as follows

Alternating algorithm

1. Initialize $\mathbf{x}^{(0)}$ to a suitable first estimate (e.g. via thresholding), $\mathbf{y}^{(0)} = \mathbf{y}$, and $\mathbf{H}^{(0)} = \mathbf{H}_0$.

2. Update $\mathbf{x}^{(n)}$ by solving

$$\mathbf{x}^{(n+1)} = \arg \min_{\mathbf{x}} \|\mathbf{H}^{(n)} \mathbf{x} - \mathbf{y}^{(n)}\|^2 \quad \text{s.t.} \quad \|\mathbf{x}\|_0 \leq p.$$

3. Update $\mathbf{H}^{(n)}$ and $\mathbf{y}^{(n)}$ by solving

$$\arg \max_{\mathbf{H}} \|\mathbf{y} - \mathbf{H} \mathbf{x}^{(n)}\|^2 \quad \text{s.t.} \quad \|\mathbf{W} \mathbf{\Delta}\|^2 \leq \epsilon.$$

4. When the stopping criterion is met, output the estimated image $\mathbf{x}^{(n)}$ and the estimated PSF $\mathbf{H}^{(n)}$.

The steps of the algorithm warrant some elaboration. In step two, we employ the methods discussed in Section 2.1 to arrive at Equations (11) and (17), with $\tilde{\mathbf{H}}$ and \mathbf{y} replaced by $\mathbf{H}^{(n)}$ and $\mathbf{y}^{(n)}$ respectively. In step three, we follow the discussion in Section 2.2, arriving at Equation (20), where we replace $\hat{\delta}$ by $\delta^{(n)}$. By substitution, we have

$$\|\mathbf{e} + \epsilon \mathbf{X} \delta\|^2 = \|(\mathbf{I} + \epsilon^2 \mathbf{X} (\gamma \mathbf{W}^T \mathbf{W} - \epsilon^2 \mathbf{X}^T \mathbf{X}^{-1}) \mathbf{X}^T) \mathbf{e}\|^2,$$

which allows us to update $\mathbf{H}^{(n)}$ and $\mathbf{y}^{(n)}$ as

$$\mathbf{H}^{(n+1)} = \left(\mathbf{I} + \epsilon^2 \mathbf{X}^{(n)} (\gamma \mathbf{W}^T \mathbf{W} - \epsilon^2 \mathbf{X}^{(n)T} \mathbf{X}^{(n)})^{-1} \mathbf{X}^{(n)T} \right) \mathbf{H}_0,$$

$$\mathbf{y}^{(n+1)} = \left(\mathbf{I} + \epsilon^2 \mathbf{X}^{(n)} (\gamma \mathbf{W}^T \mathbf{W} - \epsilon^2 \mathbf{X}^{(n)T} \mathbf{X}^{(n)})^{-1} \mathbf{X}^{(n)T} \right) \mathbf{y}.$$

3.1. Reconstructing the PSF

An alternative approach is to set the smoothing penalty γ sufficiently high in the alternating algorithm. After the completion of the algorithm, with the reconstructed image $\hat{\mathbf{x}}$ fixed,

we solve the minimization problem as before to recover \mathbf{H} .

$$\arg \min_{\delta} \|\hat{\mathbf{X}} \mathbf{h} - \mathbf{y}\|^2 + \tilde{\gamma} \|\mathbf{W} \delta\|^2, \quad (22)$$

where $\tilde{\gamma}$ is the smoothing penalty we enforce. It should be emphasized that while we reconstruct \mathbf{H} this way, we do not substitute the reconstructed \mathbf{H} back into the algorithm.

4. SIMULATIONS

Here, we test the effectiveness of the alternating algorithm in recovering both the original image \mathbf{x} and the true PSF. As before, we generate 33×33 images in Matlab[®] with all but 10 nonzero pixels. The nonzero pixel locations are chosen uniformly at random and the values are set to 1. For purposes of illustration, the image in Fig. 2(a) was artificially constructed, i.e., non-random.

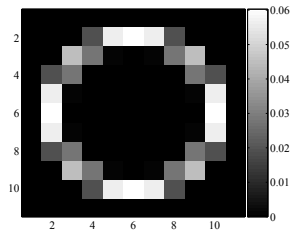
As mentioned in the Introduction, since we have motivated this paper based on MRFM, the PSF of \mathbf{H} that we use, is an idealized two-dimensional realization of an MRFM PSF. \mathbf{H} and \mathbf{H}_0 are depicted in Fig. (1), where \mathbf{H}_0 is chosen to be a smooth approximation of \mathbf{H} and such that it satisfies the criterion $\|\mathbf{W} \mathbf{\Delta}\| \leq \epsilon$. We base the smoothing operator \mathbf{W} on a Gaussian kernel, and choose ϵ to be 1. The blurred noisy image is generated by convolving the PSF corresponding to \mathbf{H} and adding zero-mean white Gaussian noise with standard deviation $\sigma = 0.2$.

By experimentally adjusting the smoothing penalty γ and the sparsifying penalty λ , we are able to find a realization of these parameters which yields a satisfactory reconstruction of the image. For the image in Fig. 2(a), we set $\gamma = 1 \times 10^7$ and $\lambda = 0.02$. The reconstructed image is displayed in Fig. 2(c).

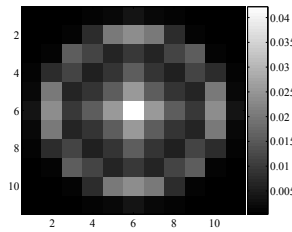
Next, we implemented a test to see how closely we could approximate the true PSF by running 1000 iterations of the alternating algorithm with a large smoothing penalty until the error between the true image and the approximant became tolerable. Empirical evidence suggests the algorithm exhibits nice convergence properties. We then fixed this estimate of \mathbf{x} and tried to recover the true PSF by relaxing the smoothing penalty ($\gamma = 1$). The sum of squared error (SSE) $\|\hat{\mathbf{x}} - \mathbf{x}\|^2$ is depicted in Fig. 3. In Fig. 1(c) the recovered \mathbf{H} is shown. It is evident from this figure that by relaxing the smoothing penalty, we are better able to approximate the true PSF.

5. CONCLUSIONS AND FUTURE DIRECTIONS

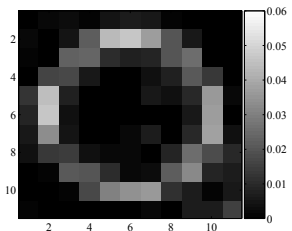
This paper introduced a non-identity smoothing penalty on allowable PSFs. A new alternating iterative algorithm was introduced in closed form which accomplishes this added consideration. Empirical evidence suggests gains in performance are created with this new formulation. An added advantage is that we are able to approximate the true PSF of the system better. Techniques of determining the proper values of γ and λ remain to be explored, as well as exploring the convergence



(a) The true PSF H .



(b) The approximate PSF H_0 .



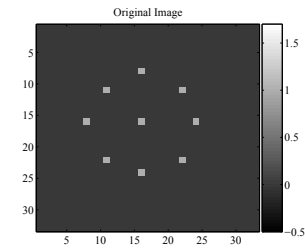
(c) The reconstructed PSF. This PSF was created by imposing a smoothing penalty of $\tilde{\gamma} = 0.5$ after the completion of the algorithm.

Fig. 1. The true, approximate and reconstructed PSFs.

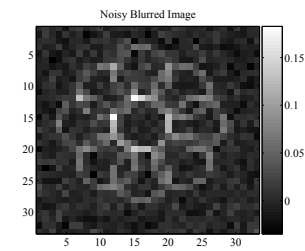
issues of the alternating algorithm. It might also be worthwhile to investigate the performance of the algorithm when hard thresholding is implemented versus soft thresholding.

6. REFERENCES

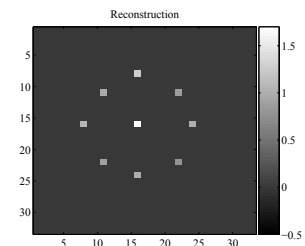
- [1] R. Raich and A. O. Hero III, "Sparse image reconstruction for partially unknown blur functions," in *Proc. IEEE Int. Conf. Image Processing*, (Atlanta, GA), pp. 637–640, Oct. 2006.
- [2] J. Mamin, R. Budakian, and D. Rugar, "Point response function of an MRFM tip," tech. rep., IBM Research Division, 2003.
- [3] H. Andrews and B. Hunt, *Digital Image Restoration*. Englewood Cliffs, NJ: Prentice-Hall, 1977.
- [4] T. Chan and C. Wong, "Total variation blind deconvolution," *IEEE Transactions on Image Processing*, vol. 7, pp. 370–375, March 1998.



(a) The original sparse image.



(b) The noisy blurred image.



(c) The reconstructed image.

Fig. 2. The original, noisy and reconstructed images.

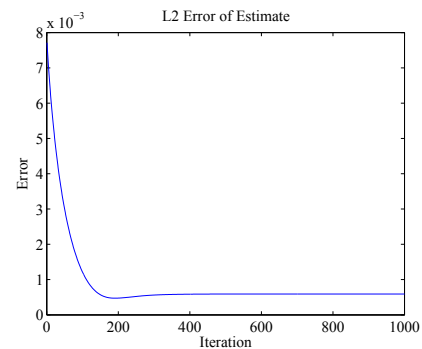


Fig. 3. The SSE of the reconstructed image, where $\lambda = 0.02$ and $\gamma = 1 \times 10^7$

- [5] I. Daubechies, M. Defrise, and C. D. Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," *Comm. Pure App. Math.*, vol. 57, pp. 1413–1457, Nov. 2004.
- [6] M. A. T. Figueiredo and R. D. Nowak, "A bound optimization approach to wavelet-based image deconvolution," in *Proc. IEEE Int. Conf. Image Processing*, vol. 2, pp. 782–785, Sept. 2005.