ACCELERATED WAVELET-REGULARIZED DECONVOLUTION FOR 3-D FLUORESCENCE MICROCOPY

İ. Bayram, M. Guerquin-Kern, R. Terrés-Cristofani and M. Unser

Biomedical Imaging Group, EPFL, CH-1015 Lausanne, Switzerland

ABSTRACT

Modern deconvolution algorithms are often specified as minimization problems involving a non-quadratic regularization functional. When the latter is a wavelet-domain ℓ_1-norm that favors sparse solutions, the problem can be solved by a simple iterative shrinkage/thresholding algorithm (ISTA). This approach provides state-of-the-art results in 2-D, but is harder to deploy in 3-D because of its slow convergence.

In this paper, we propose an acceleration scheme that turns wavelet-regularized deconvolution into a competitive solution for 3-D fluorescence microscopy. A significant speed-up is achieved though a synergistic combination of subband-adapted thresholds and sequential TwIST updates. We provide a theoretical justification of the procedure together with an experimental evaluation, including the application to real 3-D fluorescence data.

Index Terms— fluorescent microscopy, deconvolution, wavelet regularization, accelerated shrinkage/thresholding.

1. INTRODUCTION

Modern biology makes extensive use of fluorescent markers which can be visualized in 3-D using optical microscopy. The optical sectioning, which is the key to accessing the third physical dimension (depth), is either achieved physically using a confocal microscope (which is a high-end imaging device), or numerically by applying deconvolution techniques on focal series of images (z-stacks) acquired on a conventional widefield fluorescence microscope [11]. The main difficulty with "deconvolution microscopy" lies in the size of the data sets that need to be acquired for reconstructing 3-D images. This puts a strong limitation on the computational complexity of potential deconvolution procedures, and also explains why the algorithms used in current commercial systems are still rudimentary in comparison to the current state of research on inverse problems.

There is recent evidence that imposing sparsity constraints on the wavelet decomposition of the reconstructed image can improve the results of deconvolution, at least in 2-D [8, 6]. To specify the reconstruction problem, we consider the image formation model expressed in matrix/vector notation

$$\mathbf{y} = \mathbf{H}\,\mathbf{x} + \mathbf{n}, \tag{1}$$

where $\mathbf{x}$ represents the 3D specimen we are trying to recover; $\mathbf{H}$, the 3D linear space-invariant imaging operator (diffraction-limited point spread function of the microscope); $\mathbf{n}$, the measurement noise; and $\mathbf{y}$, the data acquired by the microscope. The ill-conditioning of $\mathbf{H}$ and the presence of noise hinders the estimation of $\mathbf{x}$ from the observations $\mathbf{y}$ and requires the incorporation of additional information about $\mathbf{x}$. To that end, we utilize the property that natural images tend to have sparse wavelet representations. Specifically, given a wavelet frame $\mathbf{W}$—which consists of a (possibly overcomplete) collection of wavelet vectors—we assume that one can approximate $\mathbf{x}$ well using a linear combination of just a few vectors of $\mathbf{W}$; i.e., $\mathbf{x} \approx \mathbf{W}\mathbf{c}$ with $\|\mathbf{c}\|_0$ small. Based on the property that the ℓ_1-norm $\|\mathbf{c}\|_1$ is a good proxy for $\|\mathbf{c}\|_0$ (the number of non-zero wavelet coefficients), we specify the solution of our deconvolution problem as $\mathbf{x}^* = \mathbf{W}\mathbf{c}^*$ where $\mathbf{c}^*$ is the minimizer of the convex functional

$$J(\mathbf{c}) = \|\mathbf{y} - \mathbf{H}\mathbf{W}\mathbf{c}\|_2^2 + \lambda\|\mathbf{c}\|_1. \tag{2}$$

The regularization parameter $\lambda \geq 0$ controls the strength of the sparsity requirement (for $\lambda = 0$, we revert to the classical unconstrained least squares formulation). Several algorithms have been proposed to minimize generic functionals of the form (2) where $\mathbf{H}$ is not assumed to have any particular structure. The best known is the "iterative shrinkage/thresholding algorithm" (ISTA), derived by several authors using different point of views (see for example [7, 8, 6]). The slow convergence of ISTA has also led to attempts to accelerate it by employing multi-step variations known as TwIST [4], FISTA [3] and continuation schemes that vary λ with iterations [10].

In this paper, we propose a novel variation of ISTA that builds upon these previous works while taking advantage of the special convolutional structure of $\mathbf{H}$. It is sufficiently fast to handle the large data sets produced by 3-D fluorescent microscopes. The primary contributions are:

- The proposal of a fast-converging version of ISTA that combines two acceleration strategies: subband-adaptive update steps/thresholds and sequential TwIST

This work was partially supported by the Indo-Swiss joint research program, the Center for Biomedical Imaging (CIBM) and the foundations Leenaards and Louis-Jeantet.

updates. This subband-adaptive version of TwIST (dubbed STwIST) preserves the form of ISTA and is almost as simple to implement. The adaptation of the thresholds makes the algorithm progress faster and more evenly in each wavelet subband.

- A theoretical analysis that explains the synergistic effect of the two acceleration strategies.
- An experimental evaluation and validation of the algorithm. We also present deconvolution results with real 3D fluorescence microscopy data.

2. ALGORITHM DESCRIPTION

Before describing our approach, we briefly review ISTA[8, 7]. ISTA minimizes the functional $J(\cdot)$ in (2) by performing a Landweber step, followed by soft thresholding, which is defined for a given vector $\mathbf{z}$, and threshold t by

$$\mathrm{soft}(\mathbf{z}, t) = \mathrm{sgn}(\mathbf{z})\ \max\{|\mathbf{z}| - t, 0\}. \tag{3}$$

Algorithm 1 (ISTA)**.** *Initialize* $\mathbf{c}^0$. *Repeat until some convergence criterion is met,*

(I) Set $\mathbf{c}^{k+\frac{1}{2}} := \mathbf{c}^k + \frac{1}{\alpha}\,(\mathbf{HW})^*\ \big(\mathbf{y} - \mathbf{H}\,\mathbf{W}\,\mathbf{c}^k\big)$.

(II) Set $\mathbf{c}_i^{k+1} := \mathrm{soft}(\mathbf{c}_i^{k+\frac{1}{2}}, \lambda/(2\alpha))$.

Intuitively, we can regard the two steps as attempts to minimize the two terms in $J(\cdot)$ in sucession (see for example the forward-backward derivation of ISTA [6]). The critical parameter that makes ISTA work and conditions its converge speed is the step size $\tau = 1/\alpha$ in *(I)* which should be chosen as large as possible. The condition for convergence is that $2\alpha\mathbf{I} \geq (\mathbf{HW})^*\mathbf{HW}$ which imposes an upper bound on τ (resp., lower bound on α) and therefore on the overall speed of convergence ISTA.

We now propose a combination of two strategies to accelerate ISTA. Our first ingredient is threshold adaption. It is motivated by the observation that our minimization problem is approximately decoupled across wavelet subbands since the wavelets in the synthesis matrix $\mathbf{W}$ essentially act as bandpass filters. Now, if the effect of $\mathbf{H}$ is substantially different for different subbands, we expect that treating the subbands differently might lead to an acceleration. To make this precise, we partition $\mathbf{W}$ and $\mathbf{c}$ into s subbands; i.e., $\mathbf{W} = [\mathbf{W}_1\,\mathbf{W}_2\,\ldots\,\mathbf{W}_s]$ and $\mathbf{c} = (\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_s)$. We also define the vector $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_s)$ and the matrix $\Lambda_{\boldsymbol{\alpha}}$ (acting on $\mathbf{c}$) which multiplies the j^{th} subband by α_j, i.e. $(\Lambda_{\boldsymbol{\alpha}}\mathbf{c})_j = \alpha_j\,\mathbf{c}_j$. With this, the subband adaptive iterative shrinkage/thresholding algorithm (SISTA) is stated as,

Algorithm 2 (SISTA)**.** *Initialize* $\mathbf{c}^0$. *Repeat until some convergence criterion is met,*

(I) Set $\mathbf{c}^{k+\frac{1}{2}} := \mathbf{c}^k + \boldsymbol{\Lambda}_{\boldsymbol{\alpha}}^{-1}\,\big[(\mathbf{HW})^*(\mathbf{y} - \mathbf{H}\,\mathbf{W}\,\mathbf{c}^k)\big]$.

(II) Set, for all subbands $j = 1, 2, \ldots, s$,

$$\mathbf{c}_j^{k+1} := \mathrm{soft}(\mathbf{c}_j^{k+\frac{1}{2}}, \lambda/(2\alpha_j)). \tag{4}$$

Similar to ISTA, the convergence rate of SISTA is dependent upon α_i. To that end, we choose the "smallest" possible values such that the condition for convergence is met: $2\boldsymbol{\Lambda}_\alpha \geq (\mathbf{HW})^*\,\mathbf{HW}$ (see [2] for a more precise description). We remark that the form of the algorithm is essentially the same as ISTA, and that the idea of a subband adaption of the thresholds was first proposed by Vonesch and Unser [12], albeit in the more restrictive context of Shannon wavelets. In the sequel, we will use the notation $\mathbf{c}^{k+1} = \mathrm{SISTA}(\mathbf{c}^k)$ (resp., $\mathrm{ISTA}(\mathbf{c}^k)$) to represent one iteration of the corresponding algorithm.

The second ingredient is the sequential TwIST update rule which, when used on its own, already provides an effective acceleration of ISTA [4]. The idea in TwIST is to combine the previous estimates to obtain a better update, similar to two step methods used to accelerate the solution of linear system of equations (see Section 5.2 in [1]). At the k^{th} iteration, TwIST updates $\mathbf{c}^k$ as,

$$\mathbf{c}^{k+1} := (1-\gamma)\,\mathbf{c}^{k-1} + (\gamma - \beta)\,\mathbf{c}^k + \beta\,\mathrm{ISTA}(\mathbf{c}^k), \tag{5}$$

where γ and β are suitable constants. Since SISTA has basically the same form as ISTA, we propose to use the same recipe to devise an accelerated version of Algorithm 2. In the following section we provide a new theoretical justification of this proposal as well.

Algorithm 3 (SISTA+TwIST=STwIST)**.** *Initialize* $\mathbf{c}^0$. *Set* $\mathbf{c}^1 = \mathrm{SISTA}(\mathbf{c}^0)$. *Repeat until some convergence criterion is met,*

$$\mathbf{c}^{k+1} := (1-\gamma)\,\mathbf{c}^{k-1} + (\gamma - \beta)\,\mathbf{c}^k + \beta\,\mathrm{SISTA}(\mathbf{c}^k). \tag{6}$$

In the sequel, we will demonstrate that the payoff for combining the two acceleration schemes is quite substantial and sufficient to meet our goal: making sparsity promoting methods as fast as the basic deconvolution algorithms currently used in microscopy (e.g. Richardson-Lucy).

3. COMPLEMENTARY ACCELERATION PROPERTIES OF SISTA AND TWIST

We now provide an analysis on the combination of SISTA and TwIST. This analysis (which was not provided in [2]) justifies the proposed STwIST algorithm by showing that the combination of ISTA and TwIST is especially favorable. We start with a brief description of a standard two-step scheme for linear problems (see [1] for more details).

Consider a linear system of equations, $\mathbf{Ax=b}$. We are given $\mathbf{A}$, $\mathbf{b}$ and asked to recover $\mathbf{x}$. Consider now a splitting of $\mathbf{A}$ as $\mathbf{A= C}$ - $\mathbf{R}$ and devise an iterative method as

$$\mathbf{x}^{k+1} = \mathbf{C}^{-1}\,\mathbf{R}\mathbf{x}^k + \mathbf{C}^{-1}\mathbf{b} \tag{7}$$

$$= \mathbf{x}^k + \mathbf{C}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}^k). \tag{8}$$

Defining the error vector at iteration k by $\mathbf{e}^k = \mathbf{x}^k - \mathbf{x}$, it can be shown that $\mathbf{e}^{k+1} = \mathbf{C}^{-1}\mathbf{R}\mathbf{e}^k$. Therefore, the linear convergence factor[1] of this method depends on the spectral radius of $\mathbf{C}^{-1}\mathbf{R}$. To accelerate the convergence, a two-step scheme can be employed, which is defined by

$$\mathbf{x}^{k+1} = \gamma\,\mathbf{x}^k + (1-\gamma)\mathbf{x}^{k-1} + \beta\,(\mathbf{C}^{-1}(\mathbf{b} - \mathbf{A}\mathbf{x}^k)). \quad (9)$$

The error term $\mathbf{e}^k$ is now given by,

$$\mathbf{e}^{k+1} = (\gamma\,\mathbf{I} + \beta\,\mathbf{C}^{-1}\mathbf{R})\mathbf{e}^k + (1-\gamma)\mathbf{e}^{k-1}. \quad (10)$$

Here, the convergence factor depends on the eigenvalues of $\mathbf{C}^{-1}\mathbf{R}$ as well as γ and β. In particular, for the optimal choice of γ, β, the convergence factor is given by $(1-\sqrt{\kappa})/(1+\sqrt{\kappa})$ where κ is the ratio of the smallest and largest eigenvalues of $\mathbf{C}^{-1}\mathbf{R}$ [1]. The important point for us is that, as κ gets closer to unity, the error decays faster.

Let us now consider ISTA/SISTA and focus on a special case where the matrix $\mathbf{HW}$ is invertible. In this setting, the first term of $J(\cdot)$ is strictly convex, and therefore the minimizer of $J(\cdot)$, which we denote by $\hat{\mathbf{c}}$, is unique. An iteration of ISTA is then,

$$\mathbf{c}^{k+1} = \text{soft}\left(\mathbf{c}^k + \frac{1}{\alpha}\,(\mathbf{HW})^*\,\left(\mathbf{y} - \mathbf{H}\,\mathbf{W}\,\mathbf{c}^k\right)\right). \quad (11)$$

We also have that $\hat{\mathbf{c}}$ is a fixed point of the mapping on the right hand side of (11). Following the idea in the appendix of [4], we define $\mathbf{e}^k = \mathbf{c}^k - \hat{\mathbf{c}}$ and apply the generalized mean value theorem for nonsmooth mappings [5] to write

$$\mathbf{e}^{k+1} = \mathbf{D}_k\left(\mathbf{I} - \frac{1}{\alpha}(\mathbf{HW})^*(\mathbf{HW})\right)\mathbf{e}^k, \quad (12)$$

where $\mathbf{D}_k$ is a matrix that depends on $\mathbf{c}^k$, whose spectral radius is less than or equal to one. By the same argument, we obtain that, for SISTA, $\mathbf{e}^k$ satisfies,

$$\mathbf{e}^{k+1} = \tilde{\mathbf{D}}_k\left(\mathbf{I} - \mathbf{\Lambda}_\alpha^{-1}(\mathbf{HW})^*(\mathbf{HW})\right)\mathbf{e}^k, \quad (13)$$

again with an iteration-dependent $\tilde{\mathbf{D}}_k$, whose spectral radius is less than unity. Neglecting the effect of $\mathbf{D}_k$ and $\tilde{\mathbf{D}}_k$, we argue that the spectrum of $\mathbf{I} - \mathbf{\Lambda}_\alpha^{-1}(\mathbf{HW})^*(\mathbf{HW})$ is more concentrated than that of $\mathbf{I} - \frac{1}{\alpha}(\mathbf{HW})^*(\mathbf{HW})$ for a proper choice of $\mathbf{\Lambda}_\alpha$. In other words, for SISTA, the ratio of the smallest and largest eigenvalues is expected to be closer to unity compared to that of ISTA, at each iteration. Therefore, recalling that the improvement of the two-step method described above is a function of the mentioned ratio κ, we argue that SISTA conditions the iterations of ISTA, making it more suitable for a two step scheme like TwIST.

[1]The error decays as a power of the convergence factor, so the smaller this number, faster the convergence.

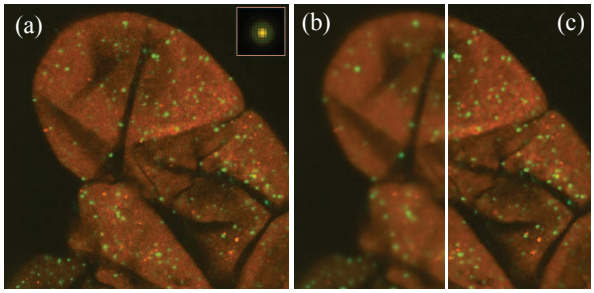

Fig. 1. 2-D deconvolution of defocused widefield microscopy image (Experiment 1). (a) Original image with the PSF on the top-right corner, (b) simulated widefield image (left half), (c) deconvolution result after 30 iterations of STwIST (right half).

4. EXPERIMENTS

In this section, we evaluate the proposed deconvolution algorithm experimentally. In order to demonstrate the acceleration in convergence rate, we work with simulated 2D data in Experiment 1. Following this, in Experiment 2, we present some deconvolution results with a fixed time budget on real 3D fluorescence micrographs to compare the proposed algorithm with a state-of-the-art algorithm.

Experiment 1. In this experiment, we simulate the widefield imaging of a flat fluorescent sample slightly out of focus. The reference image[2] is shown in Fig. 1. Two channels were acquired, corresponding to the fluorescent markers GFP (green: $488\ nm$) and Alexa568 (red: $568\ nm$). We blurred the channels with normalized PSFs (shown in Fig.1a, top right) and added noise, to obtain the 'observation' shown in Fig.1b. For the reconstruction, we used separable $9/7$ wavelets with 4 levels of decomposition. For SISTA, combined with the PSF of the two channels, the bounds are tabulated in Table 1. A close estimate of the minimizer of the functional $J(\cdot)$ for a λ value of 8 obtained by running ISTA with 10000 iterations is shown in the rightmost panel of Fig. 1. The distance between the current estimate and the minimizer vs. the number of iterations is plotted in Fig. 3 for three different algorithms. Here, SISTA combined with TwIST is clearly faster than the other algorithms. We remark that TwIST is currently one of the fastest algorithms, and therefore provides a fair comparison with the state-of-the-art.

Experiment 2. Our second experiment involves a dataset composed of three stacks of images of a C. Elegeans embryo. Images are acquired at EPFL using a 100X/1.4 oil-immersion objective and a z-step of $200\ nm$ (111 slices). Three channels were acquired, corresponding to the fluorescent markers

[2]The images were acquired at the Department of Biology in Chinese University of Hong Kong. We are especially grateful to Prof. Liwen Jiang and Dr. Yansong Miao for the preparation of the biological sample and their help on acquiring the images.

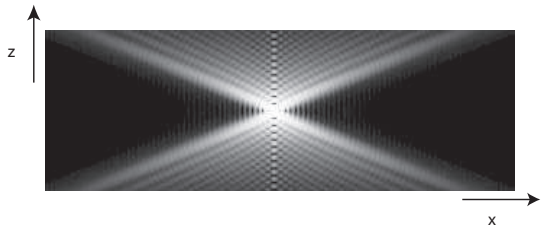

Fig. 2. A cross section along $y = 0$ of the theoretically estimated PSF (histogram equalized to improve visualization). The 3D PSF is symmetric wrt to the z axis and is determined by this 2D function.

Level	1		2		3		
Sub	LH/HL	HH	LH/HL	HH	LH/HL	HH	L
Red	2.4e-2	3.8e-3	1.9e-2	3.6e-3	8.3e-1	6.1e-1	1.41
Green	1.6e-2	5.0e-3	1.9e-2	2.2e-2	8.6e-1	6.4e-1	1.46

Table 1. Bounds obtained for Experiment 1 (9/7 wavelets).

DAPI (blue : 377 nm), FITC (green : 485 nm) and CY3 (red : 560 nm). The PSF was computed theoretically using the three-dimensional coherent image formation model [9]. Fig. 2 shows a cross-section of the PSF. More information about the data including the 3-D PSF (as well as the parameters used to compute it) can be found on the web link `http://bigwww.epfl.ch/deconvolution/`.

Figure 4 shows the original data using maximum intensity projection in each channel and the deconvolution result obtained by TwIST and STwIST after 30 iterations. We remark that the output of STwIST is sharper than that of TwIST alone, in line with our claim that subband adaptation leads to significant acceleration.

5. CONCLUSION

The practical challenge of 3-D fluorescent microscopy deconvolution is the necessity of handling huge data sets. Until very recently, this has precluded the use of algorithms that optimize non-quadratic criteria because they require many iterations to converge. The present work shows that it is possible to design wavelet-regularized schemes that are sufficiently fast to be applicable to real 3-D data.

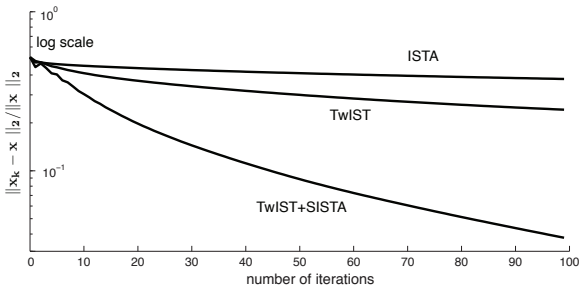

Fig. 3. 2-D deconvolution of the green channel : Convergence rates of ISTA, TwIST and STwIST=TwIST+SISTA.

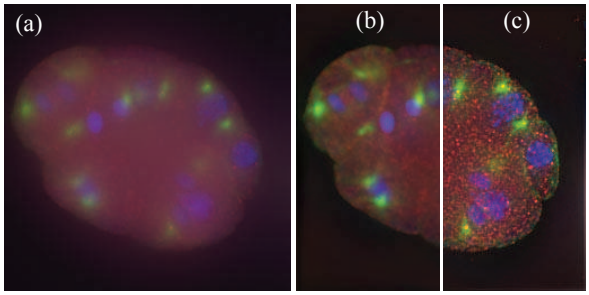

Fig. 4. 3D Fluorescent Deconvolution Results for real data (Experiment 2). (a) Original acquired image, (b) output of TwIST after 30 iterations (left half), (c) output of STwIST after 30 iterations (right half).

6. REFERENCES

[1] O. Axelsson. *Iterative Solution Methods*. Cambridge University Press, 1993.

[2] İ. Bayram and I.W. Selesnick. A subband adaptive iterative shrinkage/thresholding algorithm. To appear, IEEE Trans. Signal Processing, 2010.

[3] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sciences*, 2(1):183–202, 2009.

[4] J. M. Bioucas-Dias and M. A. T. Figueiredo. A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration. *IEEE Trans. Image Processing*, 16(12):2992–3003, December 2007.

[5] F. H. Clarke. *Optimization and Nonconvex Analysis*. SIAM, 1987.

[6] P. L. Combettes and V. R. Wajs. Signal recovery by proximal forward-backward splitting. *SIAM J. Multiscale Model. Simul.*, 4(4):1168–1200, November 2005.

[7] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Commun. Pure Appl. Math.*, 57(11):1413–1457, August 2004.

[8] M. A. T. Figueiredo and R. D. Nowak. An EM algorithm for wavelet-based image restoration. *IEEE Trans. Image Processing*, 12(8):906–916, August 2003.

[9] M. Gu. *Principle of Three-Dimensional Image Formation in Confocal Microscopes*. World Scientific, 1996.

[10] E. T. Hale, W. Yin, and Y. Zhang. Fixed-point continuation for l_1 minimization: Methodology and convergence. *SIAM J. Optim.*, 19(3):1107–1130, November 2008.

[11] C. Vonesch, F. Aguet, J.-L. Vonesch, and M. Unser. The coloured revolution of bioimaging. *IEEE Signal Processing Magazine*, 23(3):20–31, May 2006.

[12] C. Vonesch and M. Unser. A fast thresholded Landweber algorithm for wavelet-regularized multidimensional deconvolution. *IEEE Trans. Image Processing*, 17(4):539–549, April 2008.