ACCELERATED MRI USING ITERATIVE NON-LOCAL SHRINKAGE

Yasir Q. Mohsin, Gregory Ongie, and Mathews Jacob

University of Iowa, IA, USA

ABSTRACT

We introduce a fast iterative non-local shrinkage algorithm to recover MRI data from undersampled Fourier measurements. This approach is enabled by the reformulation of current non-local schemes as an alternating algorithm to minimize a global criterion. The proposed algorithm alternates between a non-local shrinkage step and a quadratic subproblem. The resulting algorithm is observed to be considerably faster than current alternating non-local algorithms. We use efficient continuation strategies to minimize local minima issues. The comparisons of the proposed scheme with state-of-the-art regularization schemes show a considerable reduction in alias artifacts and preservation of edges.

Index Terms— *MRI, non-local means, shrinkage, compressed sensing, denoising.*

1. INTRODUCTION

Non-local means (NLM) is a class of denoising methods which exploits the similarity between patches in an image to suppress noise [1, 2]. These methods recover the denoised image as a weighted linear combination of all the pixels in the noisy image. The weights between any two pixels are estimated from the noisy image as the measure of similarity between their patch neighborhoods. One of the difficulties in applying this scheme to MRI recovery from undersampled data is the dependence of the criterion on pre-specified weights; the use of the weights estimated from aliased images often preserve the alias patterns rather than suppressing them. Some authors have shown that alternating between the denoising and weight-estimation step improves the quality of the images in deblurring applications [3], but often had limited success in heavily undersampled Fourier inversion problems.

This alternating scheme for NLM has been recently shown to be equivalent to a majorize-minimize (MM) algorithm to optimize a regularized global cost function, where the regularization term is the sum of unweighted robust distances between image patches [4, 5, 6]. The formulation as the optimization of a global criterion enabled us to devise efficient continuation strategies to overcome the local minima problems; this enabled the application of the algorithm to MR image recovery from heavily under sampled measurements in a previous work of ours [4]. The main challenge associated with this implementation is the high computational complexity of the alternating minimization algorithm.

In this work we introduce a novel iterative algorithm to directly minimize the robust non-local criterion. This approach is based on a quadratic majorization of the patch based penalty term. Unlike the majorization used in our previous work, the weights of the quadratic terms are identical for all patch pairs, but now involves a new auxiliary variable. The proposed algorithm alternates between two main steps: (**a**) non-local shrinkage, and (**b**) a quadratic optimization problem. We re-express the quadratic penalty involving the sum of patch differences as one involving sum of pixel differences, which enables us to solve for the quadratic sub-problem analytically. Each step of the iterative shrinkage algorithm is fundamentally different from classical non-local schemes that solve an weighted quadratic optimization at each step. We show that compared to the classical non-local schemes, our proposed scheme converges approximately ten times faster. Our comparisons in the results section against local total variation (TV) regularization and a recent patch dictionary learning algorithm [7] demonstrate the considerable benefits of using non-local regularization.

2. PROPOSED ALGORITHM

2.1. Unified Non-Local Formulation

The iterative algorithm that alternates between classical non-local image recovery [8] and the re-estimation of weights was shown in [4] to be a majorize-minimize algorithm to solve for

$$\widehat{\mathbf{f}} = \arg\min_{\mathbf{f}} \underbrace{\|\mathbf{A}\mathbf{f} - \mathbf{b}\|^2 + \lambda \mathcal{G}(\mathbf{f})}_{\mathcal{C}(\mathbf{f})}, \tag{1}$$

where $\mathbf{f} \in \mathbb{C}^N$ is a vector obtained by the concatenating the rows in a 2-D image $f(\mathbf{x}), \mathbf{x} \in \mathbb{Z}^2$; $\mathbf{A} \in \mathbb{C}^{M\times N}$ is a matrix that models the measurement process; and $\mathbf{b} \in \mathbb{C}^M$ is the vector of measurements. While the first term in the cost function enforces data fidelity in k-space, the second term enforces sparsity. The regularization functional $\mathcal{G}(\mathbf{f})$ is specified

This work is supported by grants NSF CCF-0844812, NSF CCF-1116067, NIH 1R21HL109710-01A1, ACS RSG-11-267-01-CCE, and ONR grant N00014-13-1-0202.

by:

$$\mathcal{G}(\mathbf{f}) = \sum_{\mathbf{x}} \sum_{\mathbf{y}\in\mathbf{x}+\mathcal{N}} \varphi\left(P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{y}}(\mathbf{f})\right). \quad (2)$$

Here, φ is an appropriately chosen potential function and $P_{\mathbf{x}}$ is a patch extraction operator which extracts an image patch centered at the spatial location $\mathbf{x}$ from the image $\mathbf{f}$:

$$P_{\mathbf{x}}(\mathbf{f}) = f(\mathbf{x}+\mathbf{p}), \quad \mathbf{p}\in\mathcal{B}, \quad (3)$$

where $\mathcal{B}$ denotes the indices in the patch. In this paper, we focus on potential functions of the form

$$\varphi(\mathbf{g}) = \phi\left(\|\mathbf{g}\|\right), \quad (4)$$

where $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ is an appropriately chosen distance metric and $\|\mathbf{g}\|^2 = \sum_{\mathbf{p}\in\mathcal{B}} |g(\mathbf{p})|^2$. Based on our extensive comparisons (not shown here), we choose ϕ as the thresholded ℓ^p, $0 < p \leq 1$, metric:

$$\phi(t) = \begin{cases} |t|^p/p & \text{if } t < T \\ T^p/p & \text{if } t \geq T. \end{cases} \quad (5)$$

However, the framework presented in this paper is general enough to work for other potential functions.

2.2. Majorization of the Penalty Term

The above problem is often solved using the iterative reweighed algorithm [4, 6]. However, in this work we consider an alternate majorization of the cost function $\mathcal{C}$ in (1) as:

$$\mathcal{C}_\beta(\mathbf{f}) = \min_{\{\mathbf{s}_{\mathbf{x},\mathbf{q}}\}} \|\mathbf{A}\mathbf{f}-\mathbf{b}\|^2 + \lambda \sum_{\mathbf{x}} \sum_{\mathbf{q}\in\mathcal{N}} \psi(\mathbf{s}_{\mathbf{x},\mathbf{q}}) + \frac{\lambda\beta}{2} \sum_{\mathbf{x},\mathbf{q}} \|P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{p}}(\mathbf{f}) - \mathbf{s}_{\mathbf{x},\mathbf{q}}\|^2. \quad (6)$$

Here β is a parameter that controls the quality of the majorization; as $\beta \to \infty$ the majorization approaches the original cost in (1). Additionally, ψ is a function that depends on ϕ and β such that $\psi \to \phi$ as $\beta \to \infty$, and $\{\mathbf{s}_{\mathbf{x},\mathbf{q}}\}$ are auxiliarly variables. We use an alternating minimization algorithm to optimize (6). Specifically, we alternate between the determination of the optimal variables $\{\mathbf{s}_{\mathbf{x},\mathbf{q}}\}$, assuming $\mathbf{f}$ to be fixed and the determination of the optimal $\mathbf{f}$, assuming $\{\mathbf{s}_{\mathbf{x},\mathbf{q}}\}$ to be fixed.

2.3. The s Sub-Problem: solve for $\mathbf{s}_{\mathbf{x},\mathbf{q}}$, assuming f fixed

If the variable $\mathbf{f}$ is assumed to be a constant, the determination of each of the auxiliary variables $\mathbf{s}_{\mathbf{x},\mathbf{q}}$ corresponding to different values of $\mathbf{x}$ and $\mathbf{y}$ can be treated independently:

$$\widehat{\mathbf{s}}_{\mathbf{x},\mathbf{q}} = \min_{\mathbf{s}_{\mathbf{x},\mathbf{q}}} \frac{\beta}{2} \|P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{q}}(\mathbf{f}) - \mathbf{s}_{\mathbf{x},\mathbf{q}}\|^2 + \psi(\mathbf{s}_{\mathbf{x},\mathbf{q}}). \quad (7)$$

It can be shown that $\widehat{\mathbf{s}}_{\mathbf{x},\mathbf{q}}$ is given as a shrinkage for all penalties of interest

$$\widehat{\mathbf{s}}_{\mathbf{x},\mathbf{q}} = \left[P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{q}}(\mathbf{f})\right] \nu\left(\|P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{q}}(\mathbf{f})\|\right), \quad (8)$$

where $\nu : \mathbb{R}^+ \to \mathbb{R}^+$ is a function that is dependent on the distance metric ϕ. In particular, when ϕ is the thresholded ℓ^p metric (5), we have

$$\nu(t) = \begin{cases} 0 & \text{if} \quad |t| < \beta^{1/(p-2)} \\ 1 - \frac{1}{\beta}|t|^{p-2} & \text{if} \quad \beta^{1/(p-2)} \leq |t| < T \\ 1 & \text{else.} \end{cases} \quad (9)$$

2.4. The f Sub-Problem: solve for f, assuming $\mathbf{s}_{\mathbf{x},\mathbf{q}}$ fixed

In this step, we assume the auxiliary variables $\mathbf{s}_{\mathbf{x},\mathbf{q}}$ to be fixed. Hence, the minimization of (6) simplifies to:

$$\min_{\mathbf{f}} \|\mathbf{A}\mathbf{f}-\mathbf{b}\|^2 + \frac{\lambda\beta}{2} \sum_{\mathbf{x},\mathbf{q}\in\mathcal{N}} \|P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{q}}(\mathbf{f}) - \mathbf{s}_{\mathbf{x},\mathbf{q}}\|^2 \quad (10)$$

It can be shown that the above penalty can be re-expressed in terms of pixel differences as

$$\min_{\mathbf{f}} \|\mathbf{A}\mathbf{f} - \mathbf{b}\|^2 + \frac{\lambda\beta}{2} \sum_{\mathbf{q}\in\mathcal{N}} \|\mathbf{D}_{\mathbf{q}}\mathbf{f} - \mathbf{h}_{\mathbf{q}}\|^2. \quad (11)$$

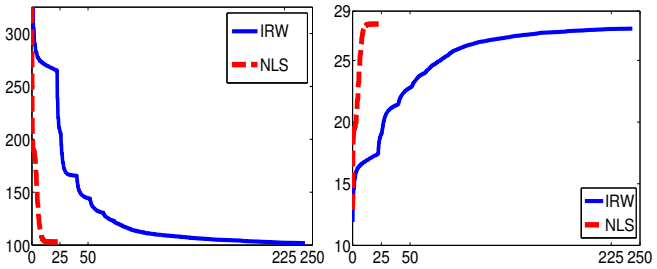
Here, $\mathbf{D}_{\mathbf{q}}$ is the finite difference operator

$$(\mathbf{D}_{\mathbf{q}}\mathbf{f})(\mathbf{x}) = \mathbf{f}(\mathbf{x}) - \mathbf{f}(\mathbf{x}+\mathbf{q}). \quad (12)$$

The images $\mathbf{h}_{\mathbf{q}}(\mathbf{x}), \mathbf{q} \in \mathcal{N}$, are obtained by shrinking the finite difference terms $\mathbf{D}_{\mathbf{q}}\mathbf{f}$:

$$\mathbf{h}_{\mathbf{q}} = (\mathbf{D}_{\mathbf{q}}\mathbf{f}) \bullet \mathbf{v}_{\mathbf{q}}, \quad (13)$$

where $\bullet$ denotes the entrywise multiplication of the vectors, and the pixel shrinkage weights $\mathbf{v}_{\mathbf{q}}$ for a specified spatial location $\mathbf{x}$ are obtained by the sum of the shrinkage weights for



(a) Cost vs Computation Time (b) SNR vs Computation Time

Fig. 1: Comparison of the convergence of the iterative reweighted non-local algorithm and the proposed iterative non-local shrinkage algorithm. The plots indicate the evolution of the cost function specified by (1) and the signal to noise ratio (SNR) as a function of the computational time.

(a) Sampling pattern (b) Original (c) DLMRI, SNR=12.96 (d) TV, SNR=15.02 (e) NLS, SNR=18.52

Fig. 2: Comparison of the recovery schemes in the presence of noise. We consider the recovery of a 256×256 MRI ankle image from its Cartesian Fourier sampling pattern (shown in (e)), contaminated by zero-mean complex Gaussian noise with standard deviation $\sigma = 10$. The top row shows the original and reconstructed images, while the error images scaled by a factor of five are shown in the bottom row. This is a challenging case due to the high 1-D undersampling factors and noise. We observe that the NLS scheme provides the best reconstructions with minimal aliasing artifacts.

the nearby patch pairs

$$\mathbf{v_q}(\mathbf{x}) = \sum_{\mathbf{p}\in\mathcal{B}} \nu\left(\|P_{\mathbf{x}+\mathbf{p}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{p}+\mathbf{q}}(\mathbf{f})\|\right). \qquad (14)$$

The solution of (11) satisfies

$$\left(2\mathbf{A}^H\mathbf{A} + \lambda\beta \sum_{\mathbf{q}\in\mathcal{N}} \mathbf{D}_\mathbf{q}^H\mathbf{D}_\mathbf{q}\right)\mathbf{f} = 2\mathbf{A}^H\mathbf{b} + \lambda\beta \sum_{\mathbf{q}\in\mathcal{N}} \mathbf{D}_\mathbf{q}^H\mathbf{h}_\mathbf{q}. \qquad (15)$$

Note that under appropriate boundary conditions $\mathbf{D}_\mathbf{q}^H\mathbf{D}_\mathbf{q}$ is diagonalizable in the Fourier domain. When the operator $\mathbf{A}^H\mathbf{A}$ is also diagonalizable in the Fourier domain, as is the case in Cartesian MRI, we may solve (15) exactly with one FFT and one IFFT.

2.5. Continuation Strategy to Improve Convergence

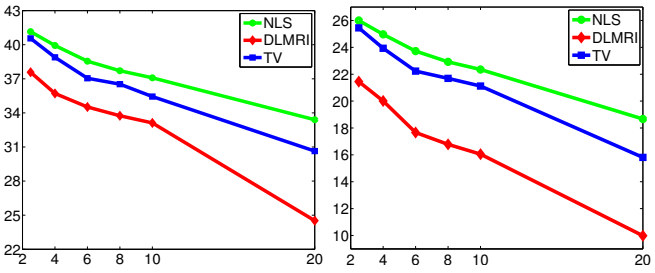
The solution of the proposed scheme corresponds to that of the original problem only when the parameter $\beta \to \infty$. However, it is known that high values of β result in poor convergence. Hence we use a continuation strategy to improve the convergence rate, where we initialize β with a small value β_0 and iteratively increase it to a high value according to $\beta_{i+1} = c \cdot \beta_i$ with $1 < c \leq 2$ constant. With each update of β we run the above proposed alternating scheme to convergence, and then warm-start the next iteration with the previously obtained solution. We also use continuation to truncate distance metric (5) in which we start with a large threshold T and gradually decrease it until it attains a small value; this means that we are not concerned with the distant patches that are likely to be dissimilar.

3. RESULTS

3.1. Convergence Rate

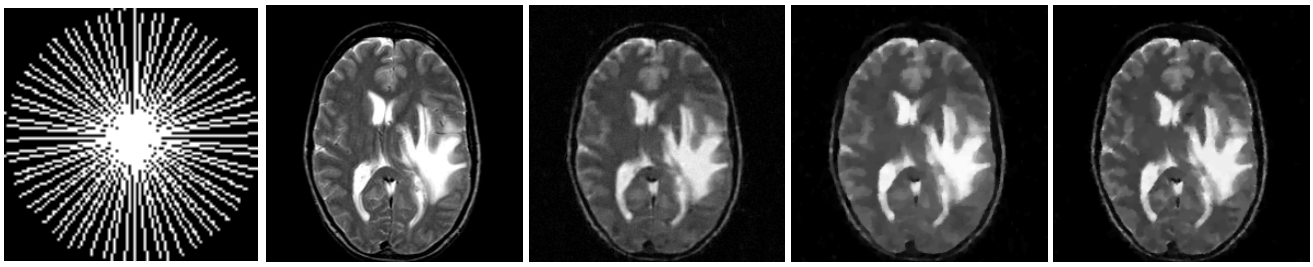
We first compare the proposed scheme with our previous iterative reweighted non-local algorithm [4]. We consider the recovery of a 256×256 MR brain image using a five-fold undersampled random sampling pattern. The regularization parameters of both algorithms were set to $\lambda = 10^{-4}$; this parameter was chosen to obtain the best possible reconstruction by comparing with the original image. The convergence plots of the algorithm as a function of the CPU time are shown in Fig. 1. We observe that both the algorithms converge to almost the same result. However, the non-local shrinkage algorithm converges around ten times faster than the iterative reweighted scheme. The reconstructions demonstrate the quality improvement offered by the proposed scheme for a specified computation time. One of the reasons for the faster convergence of the proposed algorithm can be attributed to the fast inversion of the quadratic sub-problems. The condition number of the quadratic subproblem in iterative reweighting [4] grows with iterations, resulting in slow convergence of the CG algorithms that were used to solve it.

3.2. Comparisons With State-of-the-Art Algorithms



(a) PSNR vs. Acceleration (b) SNR vs. Acceleration

Fig. 4: Performance of the recovery schemes at different accelerations. We used a 512×512 MRI brain image, sampled using a random sampling operator at different acceleration factors ($R = 2.5$, 4, 6, 8, 10 and 20). The measurements were contaminated with complex white Gaussian noise of $\sigma = 10.2$. The SNR of the reconstructions obtained using the three algorithms are plotted. These results show that the NLS scheme is capable of providing better reconstructions at a range of accelerations.



(a) Sampling pattern (b) Original (c) DLMRI, SNR=17.46 dB (d) TV, SNR=17.43 dB (e) NLS, SNR=18.46 dB

Fig. 3: Comparison of the recovery schemes in the presence of noise. We consider the recovery of a 256×256 original MRI brain image from its radial trajectory with 40 spokes, contaminated by Gaussian noise with standard deviation $\sigma = 18.8$. The error images are magnified by a scale of 5-fold for the best visibility. This is a challenging case due to the high undersampling factor and high measurement noise. We observe that the NLS scheme provides the best overall reconstructions.

We compare our proposed non-local shrinkage (NLS) scheme with local TV regularization (TV) and a dictionary learning scheme (DLMRI) [7]. The latter scheme was reported to provide considerably better reconstructions than the sparse recovery combining wavelet and TV regularization [9]. A key difference with the results reported in [7] is that we used the complex version of the code distributed by the authors. This was required to make the comparisons fair to TV and our scheme, both of which do not use this constraint.

In Fig. 2 we show the reconstructions of an ankle image from its four-fold Cartesian undersampled k-space measurements, corrupted with zero-mean complex Gaussian noise with $\sigma = 10$. We observe that the NLS scheme provides better reconstructions than the other schemes. Specifically, the TV scheme results in patchy artifacts, and the DLMRI results in blurring and loss of details close to the heel. By contrast, the degradation in performance of the NLS is comparatively small. The quantitative comparisons of the algorithms on this setting using different images are shown in the top section of Table 1.

Image	**DLMRI**		**TV**		**NLS**	
	SNR	PSNR	SNR	PSNR	SNR	PSNR
Brain1	13.55	22.82	14.81	24.67	**18.29**	**28.45**
Brain2	14.38	24.74	16.10	27.12	**18.63**	**29.83**
Brain3	13.10	26.82	16.19	30.37	**19.73**	**33.80**
Ankle	12.96	25.00	15.02	27.64	**18.52**	**31.13**
Spine	16.33	26.57	18.38	29.29	**20.49**	**31.57**
Willis'	14.56	26.35	16.08	28.53	**18.14**	**30.45**
Brain1	**12.59**	**21.87**	11.84	21.27	12.35	21.81
Brain2	17.46	27.83	17.43	28.14	**18.46**	**29.54**
Brain3	14.13	27.85	16.98	31.00	**18.00**	**31.97**
Ankle	15.80	27.85	16.17	28.57	**16.81**	**29.57**
Spine	18.54	28.77	19.86	30.50	**20.30**	**30.82**
Willis'	14.18	25.97	14.55	26.69	**15.61**	**27.69**

Table 1: Quantitative comparison of the recovery schemes in the presence of noise. The top part shows the SNR of the reconstructions obtained from 4-fold Cartesian undersampled data, contaminated by zero-mean complex Gaussian noise with $\sigma = 10$. The bottom part shows the SNR of the reconstructions from radial under sampled data with 40 spokes, contaminated by zero-mean complex Gaussian noise with $\sigma = 18.8$. The quantitative results show that the proposed iterative NLS scheme provides consistently better reconstructions for the above cases.

The reconstructions of a 256×256 brain image from its radial samples acquired with an 40 spoke trajectory are shown in Fig. 3. The measurements are corrupted with zero mean complex Gaussian noise of $\sigma = 18.8$. All methods result in loss of subtle image features since the acceleration factor and the noise level are high; but we observe that the NLS scheme provides better recovery than the competing methods. The quantitative results in this setting for various MR images are shown in the bottom section of Table 1. We observe that the SNR improvement offered by NLS over the other methods are not as high as in the previous case, mainly due to the considerable noise in the data and the high acceleration.

We also study the performance of the recovery schemes as a function of acceleration in the presence of noise in Fig. 4. We used a 512×512 MRI brain image, sampled using a random sampling operator at different acceleration factors ($R = 2.5$, 4, 6, 8, 10 and 20). The measurements were contaminated with complex white Gaussian noise of $\sigma = 10.2$. The PSNR and SNR as a function of acceleration are plotted in Fig. 4, where we compare our method against DLMRI and TV. We observe that the proposed scheme provides a consistent improvement in the presence of noise.

4. CONCLUSION

We introduced a fast iterative algorithm for MRI recovery using non-local means. The proposed algorithm alternates between a non-local shrinkage step and a quadratic subproblem, both of which are solved analytically and efficiently. The resulting algorithm is considerably faster than classical non-local methods that alternate between weight estimation and non-local regularization. Our scheme compares favorably to other state-of-the-art MRI recovery schemes in terms of the quality of the recovered images.