



Published in final edited form as:

*IEEE Trans Med Imaging*. 2015 December ; 34(12): 2417–2428. doi:10.1109/TMI.2015.2398466.

## Iterative Shrinkage Algorithm for Patch-Smoothness Regularized Medical Image Recovery

**Yasir Q. Mohsin [Student Member, IEEE],**

Department of Electrical and Computer Engineering, Univ. Iowa, IA, USA

**Gregory Ongie [Student Member, IEEE], and**

Department of Mathematics, Univ. Iowa, IA, USA

**Mathews Jacob [Senior Member, IEEE]**

Department of Electrical and Computer Engineering, Univ. Iowa, IA, USA

### Abstract

We introduce a fast iterative shrinkage algorithm for patch-smoothness regularization of inverse problems in medical imaging. This approach is enabled by the reformulation of current non-local regularization schemes as an alternating algorithm to minimize a global criterion. The proposed algorithm alternates between evaluating the denoised inter-patch differences by shrinkage and computing an image that is consistent with the denoised inter-patch differences and measured data. We derive analytical shrinkage rules for several penalties that are relevant in non-local regularization. The redundancy in patch comparisons used to evaluate the shrinkage steps are exploited using convolution operations. The resulting algorithm is observed to be considerably faster than current alternating non-local algorithms. The proposed scheme is applicable to a large class of inverse problems including deblurring, denoising, and Fourier inversion. The comparisons of the proposed scheme with state-of-the-art regularization schemes in the context of recovering images from undersampled Fourier measurements demonstrate a considerable reduction in alias artifacts and preservation of edges.

### Index Terms

MRI; non-local means; shrinkage; compressed sensing; denoising; non-Cartesian

### I. Introduction

The recovery of images from noisy measurements acquired using an ill-posed measurement scheme is a central problem in many medical imaging modalities such as magnetic resonance imaging (MRI), positron emission tomography (PET), and computed tomography (CT) [1]. Several regularized optimization schemes were introduced to make the recovery of images well-posed [2], [3]. We focus on patch based regularization in this paper, which is motivated by non-local algorithms. Non-local means (NLM) denoising algorithms recover

each pixel as a weighted linear combination of all pixels in the noisy image; the inter-pixel weights were estimated as the similarity between patch neighborhoods [4], [5], [6]. This algorithm was extended to deblurring problems [7], [8], [9] by posing the recovery as an optimization scheme. However, this algorithm is not readily applicable to MRI recovery from undersampled data since the weights estimated from aliased images preserve alias patterns. Some authors have iterated between denoising and weight estimation steps [10]. However, this scheme had limited success in heavily undersampled Fourier inversion problems. This alternating NLM scheme has been recently shown to be a majorize-minimize algorithm to solve for a penalized optimization problem, where the penalty term is the sum of unweighted robust distances between image patches [11], [12], [13]. While convex distance metrics such as  $\ell_1$  distances may be used, non-convex metrics that correspond to the classical NLM choices are seen to provide significantly improved image quality [11]. The availability of the global cost function enabled homotopy continuation strategies to encourage the convergence to global minima, when non-convex metrics are used [11]. The main challenge associated with the implementation in [11] is the high computational complexity of the alternating minimization algorithm.

The main focus of this paper is to introduce an iterative nonlocal shrinkage (NLS) algorithm, which directly minimizes the robust non-local criterion in [11]. This algorithm alternates between **(a)** evaluating the denoised inter-patch differences by shrinkage, and **(b)** computing an image that is consistent (in least-square sense) with the denoised inter-patch differences and measured data. This approach is based on additive half quadratic majorization of the patch based penalty term [14], [15], [16], [17]. Unlike the majorization used in our previous work [11], the weights of the quadratic terms are identical for all patch pairs, but now involve new auxiliary variables that may be interpreted as the denoised inter-patch differences. We derive analytical shrinkage expressions for approximate versions of a range of distance functions that are relevant for non-local regularization; this generalizes the shrinkage formulae derived by Chartrand in the context of  $\ell_p$  penalties [18]. The proposed approximations are the same as the Huber-like corner-rounded penalties [19], [20], [21], [22]. However, the key difference with our approach is that the corner-rounded approximations emerge as convex relaxations, which shows that this approximation is indeed the best possible one for the original distance function that allows a valid quadratic majorization. Note that each step of the iterative shrinkage algorithm is fundamentally different from classical non-local schemes that solve an weighted quadratic optimization at each step [4], [5], [6], [7], [8], [9]. The direct evaluation of the patch shrinkages are computationally expensive. We propose to exploit the redundancies in the shrinkages at adjacent pixels using separable convolution operations, thereby considerably reducing computational complexity. Computing the image that is consistent with the measured data and the denoised inter patch differences is a quadratic subproblem. We re-express the 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.

We validate the proposed scheme in the context of recovering MR images from undersampled measurements. This is an active research area with several applications [23], [24], [25] and several algorithms [2], [3]. In this context, Akcakaya et al., proposed to form

clusters of patches, which are denoised using hard-thresholding in appropriate transform domain; these denoised patches are later averaged to provide the recovery [25]. This approach implicitly uses different basis sets for different clusters and hence is conceptually similar to [26] and recent methods that exploit low-rank structure of patch clusters [27], [28]. While these methods are very powerful, they are also considerably more complex than the proposed scheme. In addition, since the above alternating schemes lack a global energy function, their convergence properties cannot be commented upon. Wong et al., uses the similarity of nearby pixels using homotopic  $\ell_1$  minimization [29], which is also related to non-local means regularization used in [30]. The proposed optimization strategy also has similarities to the recent approach for regularized reconstruction based on nonlocal operators [31].

The rest of this paper is organized as follows. We briefly describe the background in Section II. The proposed iterative non-local shrinkage algorithm is detailed in Section III, while the details of the implementation are outlined in Section IV. Section V demonstrates the utility of the algorithm in the context of recovering MR images from undersampled Fourier measurements.

## II. Background

### A. Problem Formulation

We consider the recovery of a complex image  $f: \Omega \rightarrow \mathbb{C}$  from its measurements  $\mathbf{b}$ . Here,  $\Omega \subset \mathbb{Z}^2$  is the spatial support of the complex image. We model the acquisition scheme by the linear degradation operator  $\mathbf{A}$ :

$$\mathbf{b} = \mathbf{A}\mathbf{f} + \mathbf{n}.$$

For example,  $\mathbf{A} \in \mathbb{C}^{M \times N}$  corresponds to a convolution in the deconvolution setting, a Fourier matrix in the case of Fourier inversion, or identity matrix in the case of denoising. Here,  $\mathbf{f} \in \mathbb{C}^N$  is a vector obtained by the concatenating the rows in a 2-D image  $f(\mathbf{x})$  and  $\mathbf{b} \in \mathbb{C}^M$  is the vector of measurements. We assume  $\mathbf{n}$  to be Gaussian distributed white noise process of a specified standard deviation  $\sigma$ .

### B. Unified Non-Local Regularization

The iterative algorithm that alternates between classical non-local image recovery [32] and the re-estimation of weights was shown [11] to be a majorize-minimize (MM) algorithm:

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

where  $\|\mathbf{A}\mathbf{f} - \mathbf{b}\|^2$  is the data fidelity term, and the regularization functional  $\mathcal{G}(\mathbf{f})$  is specified by:

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

Here  $\varphi$  is an appropriately chosen potential function and  $P_{\mathbf{x}}(\mathbf{f})$  is a patch extraction operator which extracts an image patch centered at the spatial location  $\mathbf{x}$  from the image  $\mathbf{f}$ ; i.e.  $P_{\mathbf{x}}(\mathbf{f})$  can be written as  $\mathbf{f}(\mathbf{x} + \mathbf{p})$ , where  $\mathbf{p} \in \mathcal{B}$  which denotes the indices in the patch. For example, if we choose a square patch of size  $(2N + 1)$ , the set  $\mathcal{B} = [-N, \dots, N] \times [-N, \dots, N]$ . Similarly,  $\mathcal{N}$  are the indices of the search neighborhood; the patch  $P_{\mathbf{x}}(\mathbf{f})$  is compared to all the patches whose centers are specified by  $\mathbf{x} + \mathcal{N}$ . For example, if we choose a square shaped neighborhood of size  $2M + 1$ , the set  $\mathcal{N} = [-M, \dots, M] \times [-M, \dots, M]$ . The shape of the patches and the search neighborhood may be easily changed by re-defining the sets  $\mathcal{N}$  and  $\mathcal{B}$ .

In this paper, we focus on potential functions of the form

$$\varphi(\mathbf{g}) = \phi(\|\mathbf{g}\|), \quad (3)$$

where  $\phi: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is an appropriately chosen distance metric and  $\|\mathbf{g}\|^2 = \sum_{\mathbf{p} \in \mathcal{B}} |g(\mathbf{p})|^2$ .

### C. Solution Using Iterative Reweighted Algorithm

We showed in [11] that (1) can be solved using a majorize-minimize algorithm, where the regularization term is majorized by the weighted sum of patch differences:

$$\mathcal{G}(\mathbf{f}) \leq \underbrace{\sum_{\mathbf{x}} \sum_{\mathbf{p} \in \mathcal{N}} w_n(\mathbf{x}, \mathbf{x} + \mathbf{p}) \|P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{x} + \mathbf{p}}(\mathbf{f})\|^2}_{\mathcal{G}_n(\mathbf{f})} \quad (4)$$

The above relation is essentially the classical multiplicative half-quadratic majorization [15], [16], [33] of the penalty in (1). The weights  $w_n(\mathbf{x}, \mathbf{x} + \mathbf{p})$  in (4) are specified by:

$$w_n(\mathbf{x}, \mathbf{y}) = \frac{\phi'(\|P_{\mathbf{x}}(\mathbf{f}_n) - P_{\mathbf{y}}(\mathbf{f}_n)\|)}{2\|P_{\mathbf{x}}(\mathbf{f}_n) - P_{\mathbf{y}}(\mathbf{f}_n)\|}. \quad (5)$$

Here,  $\mathbf{f}_n$  is an image vector at the  $n^{\text{th}}$  iteration. Each iteration of the MM algorithm involves the minimization of the criterion

$$\mathbf{f}_{n+1} = \arg \min_{\mathbf{f}} \|\mathbf{A}\mathbf{f} - \mathbf{b}\|^2 + \lambda \mathcal{G}_n(\mathbf{f}). \quad (6)$$

Note from (4) that  $\mathcal{G}_n$  is a weighted linear combination of patch differences, where the weights  $w_n(\mathbf{x}, \mathbf{x} + \mathbf{p})$  are spatially varying. This makes it difficult to develop a closed form expression to solve (6). As discussed above, one of the main challenges of the algorithm is its high computational complexity. Specifically, the conjugate gradients algorithm to solve (6) converges slowly, especially as the weights increase.

### III. Proposed Algorithm

#### A. Majorization of the Penalty Term

In this work, we will consider the additive half-quadratic majorization [14], [17], [33] of the potential function  $\varphi$ , specified by

$$\varphi(\mathbf{t}) \triangleq \min_{\mathbf{s}} \left\{ \psi(\mathbf{s}) + \frac{\beta}{2} \|\mathbf{s} - \mathbf{t}\|^2 \right\} \quad (7)$$

This additive half-quadratic majorization scheme was originally introduced for edge preserving gradient regularization; using this scheme has not been reported in patch regularization to the best of our knowledge. With additional simplifications below, using this scheme results in a fast algorithm. The function  $\psi$  in (7) depends on  $\varphi$  and  $\beta$ , while  $\mathbf{s}$  is an auxiliary variable. Since  $\varphi$  is the minimum of  $\psi(\mathbf{s}) + \frac{\beta}{2} \|\mathbf{s} - \mathbf{t}\|^2$ , we will also solve for  $\mathbf{s}$  along with  $\mathbf{t}$ , as shown in Section III-B. Note that unlike (4), the quadratic term on the right hand side of (7) is weighted by the scalar  $\beta/2$ , which is the same for all spatial locations. The spatial invariance of the weights enables us to analytically solve a key subproblem as shown in Section III-C, which is not possible with (4). Using (7), we express the original problem in (1) as:

$$\hat{\mathbf{f}} = \arg \min_{\mathbf{f}} \min_{\{\mathbf{s}_{\mathbf{x},\mathbf{q}}\}} \|\mathbf{A}\mathbf{f} - \mathbf{b}\|^2 + \lambda \sum_{\mathbf{x}} \sum_{\mathbf{q} \in \mathcal{N}} \psi(\mathbf{s}_{\mathbf{x},\mathbf{q}}) + \lambda \frac{\beta}{2} \sum_{\mathbf{x}} \sum_{\mathbf{q} \in \mathcal{N}} \|P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{q}}(\mathbf{f}) - \mathbf{s}_{\mathbf{x},\mathbf{q}}\|^2. \quad (8)$$

Note that the solution involves the minimization with respect to both  $\mathbf{f}$  and the variables  $\{\mathbf{s}_{\mathbf{x},\mathbf{q}}\}$ . We use an alternating minimization algorithm, where we alternate between two steps. In the first step, we minimize (8) with respect to  $\{\mathbf{s}_{\mathbf{x},\mathbf{q}}\}$ , assuming  $\mathbf{f}$  to be fixed. In the second step, we solve for  $\mathbf{f}$ , assuming  $\{\mathbf{s}_{\mathbf{x},\mathbf{q}}\}$  to be fixed. We show in Section III-B and Section III-C that these steps can be solved analytically. The efficient implementation of these subproblems is the main reason why the proposed algorithm is faster than the iterative reweighted implementation. Similar speed-ups were reported with using the half-quadratic regularization scheme in the context of other regularization penalties [14], [17], [33].

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

We focus on minimizing (8) with respect to the auxiliary variables  $\{\mathbf{s}_{\mathbf{x},\mathbf{q}}\}$ , assuming  $\mathbf{f}$  to be a constant in this subsection. In this case, determining each of the auxiliary variables  $\mathbf{s}_{\mathbf{x},\mathbf{q}}$  corresponding to different values of  $\mathbf{x}$  and  $\mathbf{q}$  can be treated independently:

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

We will show in the subsection III-D (see (21)) that with a convex hull approximation of the function  $\beta t^2/2 - \phi(t)$ , we can analytically determine  $\mathbf{s}_{\mathbf{x},\mathbf{q}}$  as a shrinkage for all penalties of interest:

$$\bar{\mathbf{s}}_{\mathbf{x},\mathbf{q}} = [P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{q}}(\mathbf{f})] \nu(\|P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{q}}(\mathbf{f})\|), \quad (9)$$

where  $\nu: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is a function that depends on the distance metric  $\phi$ . Note that the structure of the algorithm is exactly the same for different choices of distance function; only the analytical expressions for the shrinkage steps will change depending on the specific choice. We will determine the shrinkage rules corresponding to the useful penalties in Section III-D.

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

In this subsection, we focus on minimizing (8) with respect to  $\mathbf{f}$ , assuming the auxiliary variables  $\{\mathbf{s}_{\mathbf{x},\mathbf{q}}\}$  to be fixed. We show that this quadratic sub-problem can be solved analytically in the Fourier domain for several measurement operators.

Minimizing (8) with respect to  $\mathbf{f}$ , assuming  $\{\mathbf{s}_{\mathbf{x},\mathbf{q}}\}$  fixed, simplifies to:

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

The above quadratic penalty may be solved using the conjugate gradients algorithm. However, we will now simplify it to an expression that can be solved analytically, which is considerably more efficient. The quadratic penalty term involves differences between multiple patches in the image, each of which is a linear combination of quadratic differences between image pixels. The differences between two specific pixels are thus involved in different patch differences. We show in Appendix A that the pixel differences from several patches can be combined to obtain the following pixel-based penalty:

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

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  $\mathbf{h}_q(\mathbf{x})$ ,  $q \in \mathcal{N}$ , are obtained from the patch shrinkages in Section III-B as:

$$\mathbf{h}_q = (\mathbf{D}_q \mathbf{f}) \bullet \mathbf{v}_q. \quad (13)$$

Here,  $\bullet$  denotes the entrywise multiplication of the vectors, and  $\mathbf{v}_q$  is obtained by summing the shrinkage weights of nearby patch pairs

$$\mathbf{v}_q(\mathbf{x}) = \sum_{\mathbf{p} \in \mathcal{B}} \underbrace{\nu(\|P_{\mathbf{x}+\mathbf{p}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{p}+q}(\mathbf{f})\|)}_{u_q(\mathbf{x})}. \quad (14)$$

We solve (11) in the Fourier domain for measurement operators  $\mathbf{A}$  that are diagonalizable in the Fourier domain, as shown section IV.

#### D. Determination of Shrinkage Rules

The existence of an analytical solution to (6) is key to the fast implementation of the alternating minimization scheme. In this subsection, we determine analytical shrinkage rules for a larger class of potential functions. The main idea is to use properties of convex conjugate functions to derive the dual function  $\psi(\mathbf{s})$  in (7), which is required for the solution of (6). However, this approach requires certain convexity requirements which are not met by many potential functions of practical interest. Therefore, we propose a procedure for approximating potential functions that yields an analytical shrinkage rule. This quality of this approximation is controlled by a parameter  $\beta$ , becoming exact as  $\beta \rightarrow \infty$ .

We rewrite (7) as:

$$\varphi(\mathbf{t}) = \min_{\mathbf{s}} \left\{ \psi(\mathbf{s}) + \frac{\beta}{2} \left( \|\mathbf{s}\|^2 - 2\Re(\langle \mathbf{s}, \mathbf{t} \rangle) + \|\mathbf{t}\|^2 \right) \right\},$$

where  $\Re(x)$  is the real part of  $x$ . The above equation is further rearranged as:

$$\underbrace{\frac{\|\mathbf{t}\|^2}{2} - \frac{1}{\beta} \varphi(\mathbf{t})}_{r(\mathbf{t})} = \max_{\mathbf{s}} \left\{ \Re(\langle \mathbf{s}, \mathbf{t} \rangle) - \underbrace{\left( \frac{1}{\beta} \psi(\mathbf{s}) + \frac{\|\mathbf{s}\|^2}{2} \right)}_{g(\mathbf{s})} \right\} \quad (15)$$

From the theory in [34], the above relation is satisfied when  $r(\mathbf{t})$  is a convex function, in which case  $g = r^*$ , the Legendre-Fenchel dual (or convex conjugate) of  $r$ :

$$r^*(\mathbf{s}) = \max_{\mathbf{t}} \{ \Re(\langle \mathbf{s}, \mathbf{t} \rangle) - r(\mathbf{t}) \}. \quad (16)$$

However, the function  $r$  is not convex for most penalties  $\varphi$  that we are interested in, especially for small values of  $\beta$ . When  $r$  is not convex, we propose to approximate  $r$  by a convex function  $\hat{r}$  so that relation (15) is satisfied. We choose  $\hat{r}$  such that the epigraph of  $\hat{r}$  is the convex hull of the epigraph of  $r$ ;  $\hat{r}$  is thus the closest convex function to  $r$  (see Fig. 2.b). For  $\varphi$  functions of the form (3), we have  $r(\mathbf{t}) = q(\|\mathbf{t}\|)$ , where the function  $q: \mathbb{R}^+ \rightarrow \mathbb{R}^+$  is specified by  $q(t) = t^2/2 - \phi(t)/\beta$ . In all the cases we consider in this paper (see Appendix B), we can obtain the convex hull approximation of  $r$  as

$$\hat{r}(\mathbf{t}) = \begin{cases} q(\|\mathbf{t}\|) & \text{if } q'(\|\mathbf{t}\|) > 0 \\ c & \text{else} \end{cases}, \quad (17)$$

where  $c$  is an appropriately chosen constant to ensure continuity of  $\hat{r}$ , see Fig. (2 b) for an example.

The convex hull approximation of  $r(\mathbf{t})$  is equivalent to approximating the original penalty  $\varphi(\mathbf{t})$  as:

$$\hat{\varphi}(\mathbf{t}) := \beta \left( \frac{\|\mathbf{t}\|^2}{2} - \hat{r}(\mathbf{t}) \right);$$

see Fig. (2 c). For the potentials considered in this paper, this ‘‘Huber-like’’ approximation:

$$\hat{\varphi}(\mathbf{t}) = \begin{cases} \frac{\beta\|\mathbf{t}\|^2}{2} - c & \text{if } \|\mathbf{t}\| < L \\ \phi(\|\mathbf{t}\|) & \text{else} \end{cases}, \quad (18)$$

amounts to approximating the cusp of the original potential function at the origin by a quadratic function. Note that  $L = L(\beta) \rightarrow 0$  as  $\beta \rightarrow \infty$ , when the approximation of  $\varphi$  by  $\hat{\varphi}$  is exact. In particular we have  $\hat{\varphi} \rightarrow \varphi$  uniformly as  $\beta \rightarrow \infty$ ; see Fig. (2 d). This approximation enables us to derive the analytical solution for (6), which is termed as a shrinkage rule.

The shrinkage rule in (6) involves computing  $\bar{\mathbf{s}}$  specified by:

$$\bar{\mathbf{s}}(\mathbf{t}) = \arg \min_{\mathbf{s}} \left\{ \psi(\mathbf{s}) + \frac{\beta}{2} \|\mathbf{s} - \mathbf{t}\|^2 \right\}$$

which is often called the proximal mapping of  $\psi$ . The above equation can be simplified as



$$\begin{aligned}
\bar{\mathbf{s}}(\mathbf{t}) &= \arg \min_{\mathbf{s}} \left\{ \beta \left( \frac{1}{\beta} \psi(\mathbf{s}) + \frac{\|\mathbf{s}\|^2}{2} + \frac{\|\mathbf{t}\|^2}{2} - \langle \mathbf{s}, \mathbf{t} \rangle \right) \right\} \\
&= \arg \max_{\mathbf{s}} \left\{ \langle \mathbf{s}, \mathbf{t} \rangle - \underbrace{\left( \frac{\|\mathbf{s}\|^2}{2} + \frac{1}{\beta} \psi(\mathbf{s}) \right)}_{\hat{r}^*(\mathbf{s})} \right\} \quad (19)
\end{aligned}$$

Differentiating the right hand side of (19) with respect to  $\mathbf{s}$  and setting it to zero, we obtain  $\mathbf{t} - \hat{r}^*(\bar{\mathbf{s}}) \ni 0$ , or equivalently,

$$\bar{\mathbf{s}}(\mathbf{t}) \in (\partial \hat{r}^*)^{-1}(\mathbf{t}).$$

Since the subgradients of Legendre-Fenchel duals satisfy  $(\hat{r}^*)^{-1}(\mathbf{t}) = (\hat{r})(\mathbf{t})$ , we have

$$\bar{\mathbf{s}}(\mathbf{t}) \in \partial \hat{r}(\mathbf{t}). \quad (20)$$

Considering the expression for the convex hull approximation of  $\hat{r}$  in (17), we have:

$$\bar{\mathbf{s}}(\mathbf{t}) = \begin{cases} q'(\|\mathbf{t}\|) \frac{\mathbf{t}}{\|\mathbf{t}\|} & \text{if } q'(\|\mathbf{t}\|) > 0 \\ 0 & \text{else.} \end{cases}$$

Setting  $q(t) = t^2/2 - \phi(t)/\beta$  in the above equation, we obtain the shrinkage rule as:

$$\begin{aligned}
\bar{\mathbf{s}}(\mathbf{t}) &= \frac{\mathbf{t}}{\|\mathbf{t}\|} \left( \|\mathbf{t}\| - \frac{1}{\beta} \phi'(\|\mathbf{t}\|) \right)_+ \\
&= \mathbf{t} \underbrace{\left( 1 - \frac{\phi'(\|\mathbf{t}\|)}{\beta \|\mathbf{t}\|} \right)}_{\nu(\|\mathbf{t}\|)}_+, \quad (21)
\end{aligned}$$

where  $(\cdot)_+ := \max\{\cdot, 0\}$ . Here,  $\nu(\|\mathbf{t}\|)$  is a scalar between 0 and 1, which when multiplied by  $\mathbf{t}$  will yield the shrinkage of  $\mathbf{t}$ . Setting  $\mathbf{t} = P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{q}}(\mathbf{f})$  in the above equation, we obtain the shrinkage rules to be used in (9). Note that the above approach can be adapted to most penalties. We determine the shrinkage rules and the associated  $\nu$  functions for common penalty functions  $\phi$  in non-local regularization in Appendix B. Figure (1) shows the penalty functions for different metrics and the corresponding shrinkage weights. We observe that the derived shrinkage rule for the  $\mathbf{s}_{\mathbf{x},\mathbf{q}}$  sub-problem is only exact for the convex hull approximation  $\hat{r}$  in (17). However, note the Huber-like approximations of the penalties corresponding to  $\hat{r}$  specified by (18) approach the original penalty as  $\beta \rightarrow \infty$ .

## IV. Implementation

We now focus on the implementation of the sub-problems. Specifically, we show that all of the above steps can be solved analytically for many penalties and measurement operators  $\mathbf{A}$  that are diagonalizable by the Fourier domain (e.g. Fourier sampling, deblurring). This enables us to realize a computationally efficient algorithm. We also introduce a continuation scheme to improve the convergence of the algorithm.

### A. Analytical Solution of (10) in the Fourier Domain

The Euler-Lagrange equation for (11) is given by:

$$\left( 2\mathbf{A}^H \mathbf{A} + \lambda\beta \sum_{\mathbf{q} \in \mathcal{N}} \mathbf{D}_{\mathbf{q}}^H \mathbf{D}_{\mathbf{q}} \right) \mathbf{f} = 2\mathbf{A}^H \mathbf{b} + \lambda\beta \sum_{\mathbf{q} \in \mathcal{N}} \mathbf{D}_{\mathbf{q}}^H \mathbf{h}_{\mathbf{q}} \quad (22)$$

Here  $\mathbf{B}^H$  denotes the Hermitian transpose of matrix  $\mathbf{B}$ . We now aim to solve for  $\mathbf{f}$ , assuming  $\mathbf{h}_{\mathbf{q}}$ ;  $\mathbf{q} \in \mathcal{N}$  in the left hand side of (22) to be pre-determined from the previous iterate  $\mathbf{f}_{n-1}$ . Thus, this step involves the solution to a linear system of equations. Motivated by analytical solutions to similar problems in the context of total variation minimization [35], we propose to solve for (22) in the Fourier domain. Specifically, the measurement operator  $\mathbf{A}$  is diagonalizable in the Fourier domain in many inverse problems of interest (e.g. Fourier sampling, deblurring). In these cases, we may write  $\mathbf{A}^H \mathbf{A}$  as a pointwise multiplication in the Fourier domain. For instance, in the particular case when  $\mathbf{A}$  is a Cartesian Fourier undersampling operator, we may write

$$\mathbf{A}^H \mathbf{A} \mathbf{f} = \mathcal{F}^{-1}(\mathbf{a} \bullet \mathcal{F}(\mathbf{f})) \quad (23)$$

where  $\mathcal{F}$  discrete Fourier transform and  $\mathbf{a}$  is a vector of ones and zeros corresponding to the Fourier sample locations. It is well-known that the operators  $\mathbf{D}_{\mathbf{q}}^H \mathbf{D}_{\mathbf{q}}$  are circulant under periodic boundary conditions on  $\mathbf{f}$  (see [35] for example) and hence are diagonalizable in the Fourier domain as

$$\mathbf{D}_{\mathbf{q}}^H \mathbf{D}_{\mathbf{q}} \mathbf{f} = \mathcal{F}^{-1}(|\mathbf{d}_{\mathbf{q}}|^2 \bullet \mathcal{F}(\mathbf{f})), \quad (24)$$

where  $|\mathbf{d}_{\mathbf{q}}|^2$  is the pointwise modulus squared of the Fourier multiplier  $\mathbf{d}_{\mathbf{q}}$  corresponding to  $\mathbf{D}_{\mathbf{q}}$ . Hence, taking the DFT of both sides of (22) we have

$$\left( 2\mathbf{a} + \lambda\beta \sum_{\mathbf{q} \in \mathcal{N}} |\mathbf{d}_{\mathbf{q}}|^2 \right) \bullet \mathcal{F}(\mathbf{f}) = 2\mathbf{b}_0 + \lambda\beta \mathcal{F} \left( \sum_{\mathbf{q} \in \mathcal{N}} \mathbf{D}_{\mathbf{q}}^H \mathbf{h}_{\mathbf{q}} \right)$$

where  $\mathbf{b}_0 = \mathcal{F}^{-1}(\mathbf{A}^H \mathbf{b}) \in \mathbb{C}^M$  is a zero-padded version of the Fourier samples  $\mathbf{b} \in \mathbb{C}^N$ . Solving for  $\mathbf{f}$  gives

$$\hat{\mathbf{f}} = \mathcal{F}^{-1} \left[ \frac{2\mathbf{b}_0 + \lambda\beta \mathcal{F} \left( \sum_{\mathbf{q} \in \mathcal{N}} \mathbf{D}_{\mathbf{q}}^H \mathbf{h}_{\mathbf{q}} \right)}{2\mathbf{a} + \lambda\beta \sum_{\mathbf{q} \in \mathcal{N}} |\mathbf{d}_{\mathbf{q}}|^2} \right] \quad (25)$$

where the division occurs entrywise. Note that we have omitted the iteration step  $n$  from the above equation. Strictly speaking, it is an update for  $\hat{\mathbf{f}}$ , assuming  $\mathbf{h}_{\mathbf{q}}$  to be determined from  $\mathbf{f}_{n-1}$ .

In inverse problems such as non-Cartesian MRI and parallel MRI, where the measurement operator  $\mathbf{A}$  is not diagonalizable in the Fourier domain, we solve (22) efficiently using conjugate gradient (CG) algorithm [36]. A few CG steps at each iteration are often sufficient for good convergence since the algorithm is initialized by the previous iterate.

## B. Efficient Evaluation of Shrinkage Weights

We now focus on the efficient evaluation of  $\mathbf{v}_{\mathbf{q}}(\mathbf{x})$ ;  $\forall \mathbf{q} \in \mathcal{N}$  in (14). Note that  $\mathbf{u}_{\mathbf{q}}(\mathbf{x})$  involves the comparison of the patches  $P_{\mathbf{x}}(\mathbf{f})$  and  $P_{\mathbf{x}+\mathbf{q}}(\mathbf{f})$ ; since these quantities have to be computed for all spatial locations  $\mathbf{x}$  and different shifts  $\mathbf{q}$ , the direct evaluation of (14) is computationally expensive. We propose to exploit the redundancies between  $\mathbf{v}_{\mathbf{q}}(\mathbf{x})$  to considerably accelerate their computation. From (14), we have

$$\mathbf{u}_{\mathbf{q}}(\mathbf{x}) = \nu \left( \sqrt{\underbrace{\sum_{\mathbf{p} \in \mathcal{B}} \|\mathbf{f}(\mathbf{x}-\mathbf{p}) - \mathbf{f}(\mathbf{x}-\mathbf{p}+\mathbf{q})\|^2}_{(|\mathbf{D}_{\mathbf{q}}\mathbf{f}|^2 * \eta)(\mathbf{x})}} \right). \quad (26)$$

Here  $\eta$  is a moving average filter with the size of the patch. The above equation implies that computing  $\mathbf{u}_{\mathbf{q}}(\mathbf{x})$ ;  $\forall \mathbf{x}$  can be obtained efficiently by simple pointwise operations and a computationally efficient convolution operation. Combining the above result with (14), we obtain

$$\mathbf{v}_{\mathbf{q}} = \underbrace{\left[ \nu \left( \sqrt{|\mathbf{D}_{\mathbf{q}}\mathbf{f}|^2 * \eta} \right) \right]}_{\mathbf{u}_{\mathbf{q}}} * \eta \quad (27)$$

We realize the convolutions  $|\mathbf{D}_{\mathbf{q}}\mathbf{f}|^2 * \eta$  and  $\mathbf{u}_{\mathbf{q}} * \eta$  using separable moving average convolution operations. This approach has some similarities to [37], where moving average filters were used to speed up the non-local means algorithm.

### C. Continuation Strategy to Improve Convergence

The quality of the majorization in (7) depends on the parameter  $\beta$ . It is known that high values of  $\beta$  results in poor convergence. However, since we require the convex-hull approximation of the original penalty (see Section III-D) for the majorization, the solution of the proposed scheme corresponds to that of the original problem only when  $\beta \rightarrow \infty$ . Hence we use a continuation strategy to improve the convergence rate, where  $\beta$  is initialized with a small value and is increased gradually to a high value [11]. As discussed in [11], we use homotopy continuation on the penalties to encourage convergence to global minima. For example, with truncated penalties we start with a large threshold  $T$  and gradually decrease it until it attains a small value. In all the experiments, we initialize  $\beta$  to 1e-2 and set  $\beta_{\text{incfactor}} = 2$  while we set  $T_{\text{decfactor}} = 1.1$ . The pseudo-code of the algorithm is shown below.

#### Algorithm IV.1

NonLocal Shrinkage( $\mathbf{A}, \mathbf{b}, \lambda$ )

---

**Input :**  $\mathbf{b}$  = k-space measurements  
 $\beta = \beta_{\text{init}}; T = T_{\text{init}};$   
**while**  $i < \#$  Outer Iterations  
     **while**  $j < \#$  Inner Iterations  
         **do**  $\left\{ \begin{array}{l} \text{Shrinkage: Compute } \mathbf{v}_q \text{ using (27)} \\ \text{Compute } \mathbf{h}_q \text{ using (13)} \\ \text{Update } \mathbf{f} \text{ using (25)} \end{array} \right.$   
          $\beta \leftarrow \beta * \beta_{\text{incfactor}}$   
          $T \leftarrow T * T_{\text{decfactor}}$   
**return** ( $\mathbf{f}$ )

---

We observe that the proposed algorithm is similar in structure to our previous algorithm [11], shown below.

#### Algorithm IV.2

Rewighted MM ( $\mathbf{A}, \mathbf{b}, \lambda, \sigma$ )

---

**Input :**  $\mathbf{b}$  = k-space measurements  
 $T = T_{\text{init}};$   
**while**  $i < \#$  Outer Iterations  
     **while**  $j < \#$  Inner Iterations  
         **do**  $\left\{ \begin{array}{l} \text{Update } \mathbf{w}_{x,y} \text{ using (16) in [11]} \\ \text{Compute } \gamma_{x,y} \text{ using (20) in [11]} \\ \text{Update } \mathbf{f} \text{ by solving (26) in [11]} \end{array} \right.$   
          $T \leftarrow T * T_{\text{decfactor}}$   
**return** ( $\mathbf{f}$ )

---

Note from the pseudo-code that the non-local shrinkage algorithm requires two moving-average convolution operations per  $\mathbf{q}$  value to evaluate (27). For a  $3 \times 3$  neighborhood, this translates to 16 moving convolution operations. In addition, evaluating  $\mathbf{f}$  according to (25) requires one FFT and one IFFT. We typically need 10 – 20 inner iterations and about 30 – 40 outer iterations for the best convergence and recovery.

All the experiments were performed in MATLAB 2012 on a Linux Intel Xeon workstation machine with four cores, 3.2 GHz CPU and 32 GB RAM.

## V. Results

We focus on the application of this scheme for the recovery of MR images from undersampled measurements. All the datasets used in this paper were acquired on the Siemens 3T Tim Trio scanner at the University of Iowa. The datasets were collected under a protocol approved by the Institutional Review Board (IRB) at the University of Iowa and an informed consent was obtained from the subjects prior to the scan.

### A. Convergence Rate

We first compare the proposed scheme with our previous iterative reweighted non-local algorithm [11]. We considered the recovery of an MR brain image used in Fig. 4 from its retrospectively undersampled Fourier measurements using a variable density random sampling pattern (see Fig. 4.(e) for an example). The regularization parameters of both algorithms were set to  $\lambda = 10^{-4}$ ; this parameter was chosen to obtain the best possible reconstructions. The number of inner iterations and outer iterations in both the algorithms were set to 5 and 45, respectively. The maximum number of CG iterations to solve each quadratic subproblem in IRW scheme was set to 10. The tolerance values for all loops in both algorithms were set to  $1e-8$ .

The convergence plots of the two algorithms as functions of computation time and number of iterations are shown in Fig. 3. We observe that both algorithms converge to almost the same final cost. However, the non-local shrinkage (NLS) algorithm is around ten times faster than the iterative reweighted (IRW) scheme in terms of computation time; the NLS scheme took around 17 seconds, while the IRW required 172 seconds to converge. All the weight updates in IRW together took a total of 9.4 seconds, while the shrinkage steps in NLS took a total of 12.0 seconds. The main difference in complexity between the algorithms can be attributed to the analytical solution of  $f$  in the NLS scheme, which took only 0.4 seconds for all 225 inner iterations. At the same time, solving the quadratic subproblems in IRW using CG took around 163 seconds for all 225 inner iterations. We observe that the condition number of the quadratic subproblem in iterative reweighting [11] grows with iterations, resulting in slow convergence of the CG algorithms especially in later iterations. The speedup offered by using an additive half-quadratic majorization specified by (7) is consistent with using this method in non-patch regularization schemes [33].

### B. Impact of the Distance Metric

The proposed scheme can be adapted to most non-local distance metrics by simply changing the shrinkage rule. The shrinkage rules for different non-local penalties are shown in Fig. 1.

In Table I we compare the different metrics in the context of recovering three MR images from five fold randomly undersampled data. Here we quantify the reconstruction quality by the signal-to-noise ratio (SNR), defined as

$$\text{SNR}=20\log_{10}\left(\frac{\|\Gamma_{\text{orig}}\|_F}{\|\Gamma_{\text{rec}}-\Gamma_{\text{orig}}\|_F}\right),$$

where  $\Gamma_{\text{orig}}$  is the original image,  $\Gamma_{\text{rec}}$  is the recovered image, and  $\|\cdot\|_F$  is the Frobenius norm. The parameters of all the algorithms are optimized to provide the best possible SNR. The first column corresponds to the convex  $\ell_1$  differences between patches. The second and third columns correspond to alternating H1 and NLTV penalties [11], respectively.

All of the penalties saturate with inter-patch distances except the  $\ell_1$  distance function. This explains the poor performance of the convex  $\ell_1$  penalty compared to the non-convex counterparts. Unlike local total variation, which only compares a particular pixel with a few other pixels, several pixel comparisons are involved in non-local regularization. Saturating priors are needed to avoid the averaging of dissimilar patches, which may result in blurring. Since the saturating  $\ell_1$  metric provides the best overall reconstructions, we use this prior for remaining comparisons.

### C. Comparisons With State-of-the-Art Algorithms

We compare the proposed scheme with local total variation regularization (TV) and the dictionary learning MRI (DLMRI) scheme [26] using retrospectively undersampled MRI data. Specifically, the Fourier samples of the images on the specified sampling mask are used for reconstruction using different algorithms. These reconstructions are compared to the original image. We relied on the MATLAB implementation of DLMRI available from the authors webpage, which was adapted to account for complex MR images. The regularization parameters of all the algorithms have been optimized to yield the best SNR. The comparison of the above methods in the context of random sampling with 5 fold undersampled data in the absence of noise is shown in Fig. 4. This fully sampled  $128 \times 128$  MR brain image was acquired using a Turbo spin echo (TSE) sequence, FOV= $22 \times 22$  cm<sup>2</sup>, slice thickness=5.0 mm. The under sampling pattern in (e) was generated using a Monte-Carlo algorithm [2], which may be realized in 3D imaging by choosing the readout to be orthogonal to the image plane. We observe that the proposed non-local algorithm provides better preservation of edge details. The quantitative comparisons of different methods on the retrospective under sampling of more MR images in the absence of noise and five fold random undersampling are reported in Table II. We observe that NLS provides a consistent 1–4 dB improvement over other methods.

### D. Performance with noise

We study the performance of the proposed algorithm in the context of recovering MR images from their retrospectively undersampled measurements using different sampling trajectories in the presence of noise. The reconstructions of  $512 \times 512$  MR head complex image from its three-fold Cartesian retrospectively undersampled Fourier data, corrupted

with zero mean complex Gaussian noise are shown in Fig. 5. The SNR of the noisy measurements was 25.0 dB. This is a really challenging case since the 1-D downsampling pattern is considerably less efficient than 2-D random sampling. We observe that the non-local algorithm provides better reconstructions than the other schemes. Specifically, the TV scheme results in patchy artifacts and are over-smoothed. The DLMRI results in blurring and loss of details. By contrast to the classical algorithms, the degradation in performance of the non-local algorithm is comparatively small. The quantitative comparisons of the algorithms on this setting using different MR images are shown in the top section of Table III.

We also consider the recovery of five various MR images from their pseudo-radial samples acquired with 70 spokes/frame, which approximately corresponds to an acceleration factor of 4.2. The radial samples are approximated by the nearest Cartesian samples. The quantitative results in this setting for those MR images are shown in the bottom section of Table III. The Fourier measurements are corrupted with zero mean complex Gaussian noise of a specific variance. The SNR of the corresponding k-space measurements is reported in the second column. All methods are observed to result in loss of subtle image features since the acceleration factor and the noise level are high. But we also observe that the NLS scheme provides better recovery than the competing methods. The SNR improvement offered by NLS over the other methods in this experiment is not as high as in the previous cases, mainly due to the considerable noise in the data and the high acceleration. All of the above experiments were conducted at high acceleration factors to demonstrate the performance improvement offered by the proposed scheme. We show in Fig. 6 the recovery of three MR images from Fourier samples corresponding to low accelerations, contaminated with zero mean complex Gaussian noise. These experiments show that the NLS scheme can be used to obtain good quality reconstructions at moderate acceleration factors and noise levels.

### E. Validation using non-Cartesian MRI data

We consider the recovery of multichannel multi-shot spiral MRI data using the proposed scheme and TV regularization in this subsection. These datasets were acquired using a spin-echo variable density multi shot spiral acquisition with 22 interleaves,  $192 \times 192$  matrix and a 12 channel head array. The fully sampled dataset is recovered from all 22 interleaves and four of the important coils using total variation regularization. Specifically, each of the channels were reconstructed independently from the measured k-space data and the coil sensitivities were estimated from them. These coil sensitivities were used to recover the images from undersampled data. The forward and backward models are implemented using the non-uniform fast Fourier transform (NUFFT) [38], [39]. Since the forward model is a non-Cartesian Fourier transform, we are no longer able to use the analytical solution (25). We instead use the conjugate gradient algorithm (CG) to solve (22). We set the maximum number of CG iterations to 20 and the the previous iterate was set as the the initial guess. We observe that very few CG iterations (average of 3–4) are needed in each inner iteration, thanks to the good conditioning of (22), especially at later iterations.

The undersampled data was recovered from the measured k-space data corresponding to a random subset of 7 interleaves of the above four channels. We recovered the images using the different algorithms. The forward model in this case is the non-Cartesian Fourier transform of the coil sensitivity weighted images, using the coil sensitivities estimated from fully sampled data<sup>1</sup>. We use the CG algorithm with the same settings as above to solve (22). We did not use any pre-conditioners in the present study; we expect to further reduce the computational complexity in the future using efficient preconditioners. We only compare our recovery against TV as it was difficult to modify the DLMRI scheme to the non-Cartesian setting. We show the acquisition from two subjects in Fig. 7. The trajectories corresponding to the datasets are shown in the first column. The data from the second subject (bottom slice) had considerable off-resonance losses since the slice was close to the frontal sinuses and the ear regions. In addition, it also suffered from inter-shot motion that resulted in inconsistencies between the interleaves. The comparisons clearly show the benefit of the proposed scheme in a practical setting. Specifically, the NLS reconstructions preserve fine details better than TV regularization.

## VI. Conclusion

We introduced a fast iterative non-local shrinkage algorithm to recover MR image data from undersampled Fourier measurements. This approach is enabled by the reformulation of current non-local schemes as an iterative re-weighting algorithm to minimize a global criterion [11]. The proposed algorithm alternates between a non-local shrinkage step and a quadratic subproblem, which can be solved analytically and efficiently. We derived analytical shrinkage rules for several penalties that are relevant in non-local regularization. We accelerated the non-local shrinkage step, whose direct evaluation involves expensive non-local patch comparisons, by exploiting the redundancy between the terms at adjacent pixels. The resulting algorithm is observed to be considerably faster than our previous implementation. The comparison of different penalties demonstrated the benefit in using distance functions that saturate with distant patches. The comparisons of the proposed scheme with state of the art algorithms show a considerable reduction in alias artifacts and better preservation of edges.

## Acknowledgments

We thank the anonymous reviewers for their constructive feedback, which significantly improved the paper.

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.

## References

1. Prince, Jerry L., Links, Jonathan M. Medical imaging signals and systems. Pearson Prentice Hall; Upper Saddle River, NJ: 2006.
2. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid mr imaging. *Magnetic Resonance in Medicine*. 2007; 58(6):1182–1195. [PubMed: 17969013]

---

<sup>1</sup>This is justified since many methods such as fMRI can acquire fully sampled anatomical images, while the functional data needs to be acquired rapidly.



3. Block KT, Uecker M, Frahm J. Undersampled radial MRI with multiple coils. iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine*. 2007; 57(6): 1086–1098. [PubMed: 17534903]
4. Buades, A., Coll, B., Morel, JM. Denoising image sequences does not require motion estimation. *Advanced Video and Signal Based Surveillance, 2005. AVSS 2005. IEEE Conference on; IEEE; 2006*. p. 70-74.
5. Awate SP, Whitaker RT. Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Trans Pattern Recognition*. 2006; 28:364.
6. Buades A, Coll B, Morel JM. A review of image denoising algorithms, with a new one. *Multiscale Modeling and Simulation*. 2006; 4(2):490–530.
7. Cohen, LD., Bougleux, S., Peyré, G. Non-local regularization of inverse problems. *European Conference on Computer Vision (ECCV'08); 2008*;
8. Gilboa G, Darbon J, Osher S, Chan T. Nonlocal convex functionals for image regularization. *UCLA CAM Report*. 2006:06–57.
9. Lou Y, Zhang X, Osher S, Bertozzi A. Image recovery via nonlocal operators. *Journal of Scientific Computing*. 2010; 42(2):185–197.
10. Peyré, Gabriel, Bougleux, Sébastien, Cohen, Laurent. *Computer Vision–ECCV 2008*. Springer; 2008. Non-local regularization of inverse problems; p. 57-68.
11. Yang, Zhili, Jacob, Mathews. Nonlocal regularization of inverse problems: a unified variational framework. *IEEE Transactions on Image Processing*. 2013; 22(8):3192–3203. [PubMed: 23014745]
12. Yang Z, Jacob M. A unified energy minimization framework for nonlocal regularization. *IEEE ISBI*. 2011
13. Wang, Guobao, Qi, Jinyi. Penalized likelihood pet image reconstruction using patch-based edge-preserving regularization. *Medical Imaging, IEEE Transactions on*. 2012; 31(12):2194–2204.
14. Geman D, Yang Chengda. Nonlinear image recovery with half-quadratic regularization. *IEEE Transactions on Image Processing*. 1995; 4(7):932–946. [PubMed: 18290044]
15. Charbonnier, Pierre, Blanc-Féraud, Laure, Aubert, Gilles, Barlaud, Michel. Deterministic edge-preserving regularization in computed imaging. *Image Processing, IEEE Transactions on*. 1997; 6(2):298–311.
16. Delaney, Alexander H., Bresler, Yoram. Globally convergent edge-preserving regularized reconstruction: an application to limited-angle tomography. *Image Processing, IEEE Transactions on*. 1998; 7(2):204–221.
17. Nikolova, Mila, Ng, Michael. Fast image reconstruction algorithms combining half-quadratic regularization and preconditioning. *Image Processing, 2001 Proceedings 2001 International Conference on IEEE*. 2001; 1:277–280.
18. Chartrand R. Exact reconstruction of sparse signals via nonconvex minimization. *Signal Processing Letters, IEEE*. 2007; 14(10):707–710.
19. Böhning, Dankmar, Lindsay, Bruce G. Monotonicity of quadratic-approximation algorithms. *Annals of the Institute of Statistical Mathematics*. 1988; 40(4):641–663.
20. Erdogan, Hakan, Fessler, Jeffrey A. Ordered subsets algorithms for transmission tomography. *Physics in medicine and biology*. 1999; 44(11):2835. [PubMed: 10588288]
21. Erdogan, Hakan, Fessler, Jeffrey A. Monotonic algorithms for transmission tomography. *Biomedical Imaging, 2002. 5th IEEE EMBS International Summer School on; IEEE; 2002*. p. 14
22. Huber, Peter J. *Robust statistics*. Springer; 2011.
23. Ying, Leslie, Liang, Zhi-Pei. Parallel mri using phased array coils. *Signal Processing Magazine, IEEE*. 2010; 27(4):90–98.
24. Akçakaya, Mehmet, Nam, Seunghoon, Hu, Peng, Moghari, Mehdi H., Ngo, Long H., Tarokh, Vahid, Manning, Warren J., Nezafat, Reza. Compressed sensing with wavelet domain dependencies for coronary mri: a retrospective study. *Medical Imaging, IEEE Transactions on*. 2011; 30(5):1090–1099.
25. Akçakaya, Mehmet, Basha, Tamer A., Goddu, Beth, Goepfert, Lois A., Kissinger, Kraig V., Tarokh, Vahid, Manning, Warren J., Nezafat, Reza. Low-dimensional-structure self-learning and

- thresholding: Regularization beyond compressed sensing for mri reconstruction. *Magnetic Resonance in Medicine*. 2011; 66(3):756–767. [PubMed: 21465542]
26. Ravishankar, Saiprasad, Bresler, Yoram. MR image reconstruction from highly undersampled k-space data by dictionary learning. *Medical Imaging, IEEE Transactions on*. 2011; 30(5):1028–1041.
  27. Chen, Xiao, Salerno, Michael, Yang, Yang, Epstein, Frederick H. Motion-compensated compressed sensing for dynamic contrast-enhanced mri using regional spatiotemporal sparsity and region tracking: Block low-rank sparsity with motion-guidance (blosm). *Magnetic Resonance in Medicine*. 2013
  28. Yoon, Huisu, Kim, K., Kim, Daniel, Bresler, Yoram, Ye, J. Motion adaptive patch-based low-rank approach for compressed sensing cardiac cine mri. *IEEE Transactions on Medical Imaging*. 2014 in press.
  29. Wong, Alexander, Mishra, Akshaya, Fieguth, Paul, Clausi, David A. Sparse reconstruction of breast mri using homotopic minimization in a regional sparsified domain. *Biomedical Engineering, IEEE Transactions on*. 2013; 60(3):743–752.
  30. Adluru, G., Tasdizen, T., Whitaker, R., DiBella, E. Improving un-ersampled MRI reconstruction using non-local means. *Proceedings of the International Conference on Pattern Recognition*; 2010; p. 4000-4004.
  31. Venkatakrishnan, S., Bouman, C., Wohlberg, B. Plug-and-play priors for model based reconstruction. *Proceedings of the IEEE GlobalSIP*; 2013;
  32. Osher, Stanley, Solé, Andrés, Vese, Luminita. Image decomposition and restoration using total variation minimization and the  $H^{-1}$  norm. *Multiscale Modeling & Simulation*. 2003; 1(3):349–370.
  33. Nikolova, Mila, Ng, Michael. *Journal on Scientific Computing*. SIAM; 2005. Analysis of half-quadratic minimization methods for signal and image recovery; p. 937-966.
  34. Lewis, Adrian S. The convex analysis of unitarily invariant matrix functions. *Journal of Convex Analysis*. 1995; 2(1):173–183.
  35. Wang, Yilun, Yang, Junfeng, Yin, Wotao, Zhang, Yin. A new alternating minimization algorithm for total variation image reconstruction. *SIAM Journal on Imaging Sciences*. 2008; 1(3):248–272.
  36. Pruessmann, Klaas P., Weiger, Markus, Börnert, Peter, Boesiger, Peter. Advances in sensitivity encoding with arbitrary k-space trajectories. *Magnetic Resonance in Medicine*. 2001; 46(4):638–651. [PubMed: 11590639]
  37. Condat, Laurent. Tech Rep 00512801. HAL; 2010. A simple trick to speed up and improve the non-local means.
  38. Yang, Zhili, Jacob, Mathews. Mean square optimal nufft approximation for efficient non-cartesian mri reconstruction. *Journal of Magnetic Resonance*. 2014; 242:126–135. [PubMed: 24637054]
  39. Jacob, Mathews. Optimized least-square nonuniform fast fourier transform. *Signal Processing, IEEE Transactions on*. 2009; 57(6):2165–2177.
  40. Peyré G, Bougleux S, Cohen LD. Non-local regularization of inverse problems. *Inverse Problems and Imaging*. 2011:511–530.
  41. Lou, Yifei, Zhang, Xiaoqun, Osher, Stanley, Bertozzi, Andrea. Image recovery via nonlocal operators. *Journal of Scientific Computing*. 2010; 42(2):185–197.

## Appendix A: Simplification of Eq (10)

Using the formula for the shrinkage from (9), specified by

$$s_{x,q} = (P_x(\mathbf{f}) - P_{x+q}(\mathbf{f})) \nu (\|P_x(\mathbf{f}) - P_{x+q}(\mathbf{f})\|),$$

we obtain

$$\mathcal{R}(\mathbf{f}) = \sum_{\mathbf{x}; \mathbf{q} \in \mathcal{N}} \|P_{\mathbf{x}}(\mathbf{f}) - P_{\mathbf{x}+\mathbf{q}}(\mathbf{f}) - \mathbf{s}_{\mathbf{x}, \mathbf{q}}\|^2. \quad (28)$$

Expanding the above expression:

$$\mathcal{R}(\mathbf{f}) = \sum_{\mathbf{x}; \mathbf{q} \in \mathcal{N}} \sum_{\mathbf{p} \in \mathcal{B}} |f(\mathbf{x}+\mathbf{p}) - f(\mathbf{x}+\mathbf{q}+\mathbf{p}) - \mathbf{s}_{\mathbf{x}, \mathbf{q}}(\mathbf{p})|^2 \quad (29)$$

We use a change of variables  $\mathbf{x} = \mathbf{x} + \mathbf{p}$  to obtain:

$$\begin{aligned} \mathcal{R}(\mathbf{f}) &= \sum_{\mathbf{x}; \mathbf{q} \in \mathcal{N}} \sum_{\mathbf{p} \in \mathcal{B}} \left| \overbrace{f(\mathbf{x}) - f(\mathbf{x}+\mathbf{q})}^{e(\mathbf{x}, \mathbf{q})} - \mathbf{s}_{\mathbf{x}-\mathbf{p}, \mathbf{q}}(\mathbf{p}) \right|^2 \\ &= \sum_{\mathbf{x}; \mathbf{q} \in \mathcal{N}} |e(\mathbf{x}, \mathbf{q})|^2 - 2 \sum_{\mathbf{x}; \mathbf{q} \in \mathcal{N}} e(\mathbf{x}, \mathbf{q}) \underbrace{\sum_{\mathbf{p} \in \mathcal{B}} \mathbf{s}_{\mathbf{x}-\mathbf{p}, \mathbf{q}}(\mathbf{p})}_{h_{\mathbf{q}}(\mathbf{x})} + c \\ &= \sum_{\mathbf{x}; \mathbf{q} \in \mathcal{N}} |f(\mathbf{x}) - f(\mathbf{x}+\mathbf{q}) - h_{\mathbf{q}}(\mathbf{x})|^2 + c - d. \end{aligned} \quad (30)$$

In the above equations,  $c = \sum_{\mathbf{x}; \mathbf{q} \in \mathcal{N}} \sum_{\mathbf{p} \in \mathcal{B}} |\mathbf{s}_{\mathbf{x}-\mathbf{p}, \mathbf{q}}(\mathbf{p})|^2$  and  $d = \sum_{\mathbf{x}; \mathbf{q} \in \mathcal{N}} |h_{\mathbf{q}}(\mathbf{x})|^2$ . Since the solution to (10) does not depend on the constants, we ignore these terms. Thus, (29) can be rewritten using (30) as

$$\mathcal{R}(\mathbf{f}) = \sum_{\mathbf{x}; \mathbf{q} \in \mathcal{N}} \underbrace{\|f(\mathbf{x}) - f(\mathbf{x}+\mathbf{q}) - h_{\mathbf{q}}(\mathbf{x})\|^2}_{\mathcal{D}_{\mathbf{q}}f}$$

Here,  $\mathcal{D}_{\mathbf{q}}f(\mathbf{x}) = f(\mathbf{x} + \mathbf{q}) - f(\mathbf{x})$  is the finite difference operator. We observe that the expression for  $h_{\mathbf{q}}(\mathbf{x})$

$$h_{\mathbf{q}}(\mathbf{x}) = \sum_{\mathbf{p} \in \mathcal{B}} \mathbf{s}_{\mathbf{x}-\mathbf{p}, \mathbf{q}}(\mathbf{p}), \quad (31)$$

can be further simplified. From (9), we have the patch  $\mathbf{s}$  specified as

$$\mathbf{s}_{\mathbf{x}, \mathbf{q}} = (P_{\mathbf{x}}f - P_{\mathbf{x}+\mathbf{q}}f) \underbrace{\nu(\|P_{\mathbf{x}}f - P_{\mathbf{x}+\mathbf{q}}f\|)}_{u_{\mathbf{q}}(\mathbf{x})}$$

Here,  $u_{\mathbf{q}}(\mathbf{x}) = \nu(\|P_{\mathbf{x}}f - P_{\mathbf{x}+\mathbf{q}}f\|)$  is the factor between 0 and 1, which is multiplied by the patch to get the shrunk patch. Hence,

$$s_{\mathbf{x},\mathbf{q}}(\mathbf{r})=[f(\mathbf{x}+\mathbf{r})-f(\mathbf{x}+\mathbf{q}+\mathbf{r})] \cdot u_{\mathbf{q}}(\mathbf{x}); \mathbf{r} \in \mathcal{B}.$$

Thus, we have

$$s_{\mathbf{x}-\mathbf{p},\mathbf{q}}(\mathbf{p})=[f(\mathbf{x})-f(\mathbf{x}+\mathbf{q})] u_{\mathbf{q}}(\mathbf{x}-\mathbf{p}). \quad (32)$$

Substituting in (31), we get

$$h_{\mathbf{q}}(\mathbf{x})=(f(\mathbf{x})-f(\mathbf{x}+\mathbf{q})) \underbrace{\sum_{\mathbf{p} \in \mathcal{B}} u_{\mathbf{q}}(\mathbf{x}-\mathbf{p})}_{v_{\mathbf{q}}(\mathbf{x})}. \quad (33)$$

## Appendix B: Shrinkage rules for useful non-local distance functions

### 1) Thresholded $\ell_p$ ; $p \geq 1$ metric

We now consider the saturating  $\ell_p$  metric, specified by

$$\phi(t)=\begin{cases} |t|^p/p & \text{if } |t| < T \\ T^p/p & \text{else} \end{cases} \quad (34)$$

Computing the shrinkage rule for this mapping according to (21), we obtain

$$\nu(t)=\begin{cases} 0 & \text{if } |t| < \beta^{1/(p-2)} \\ 1-\frac{1}{\beta}|t|^{p-2} & \text{if } \beta^{1/(p-2)} \leq |t| < T \\ 1 & \text{else} \end{cases} \quad (35)$$

Setting  $T = \infty$  we get the shrinkage rule for the unthresholded  $\ell_p$  metric as

$$\nu(t)=\begin{cases} 0 & \text{if } |t| < \beta^{1/(p-2)} \\ 1-\frac{1}{\beta}|t|^{p-2} & \text{else} \end{cases}, \quad (36)$$

which is consistent with [18].

### 2) Penalty corresponding to alternating $H_1$ non-local scheme

We now consider the  $H_1$  metric, specified by

$$\varphi(t) = 1 - \exp\left(-\frac{t^2}{2\sigma^2}\right). \quad (37)$$

From (21) we obtain

$$\nu(t) = \begin{cases} 0 & \text{if } \exp\left(-\frac{t^2}{2\sigma^2}\right) > \beta\sigma^2 \\ 1 - \frac{\exp\left(-\frac{t^2}{2\sigma^2}\right)}{\beta\sigma^2} & \text{else} \end{cases} \quad (38)$$

### 3) Penalty corresponding to Peyre's non-local scheme

We now consider the penalty corresponding to Peyre's alternating scheme [11], [40]:

$$\phi(t) = 1 - \exp\left(-\frac{t}{\sigma}\right). \quad (39)$$

From (21) we obtain

$$\nu(t) = \begin{cases} 0 & \text{if } \exp\left(-\frac{t}{\sigma}\right) > \beta\sigma t \\ 1 - \frac{\exp\left(-\frac{t}{\sigma}\right)}{\beta\sigma t} & \text{else} \end{cases} \quad (40)$$

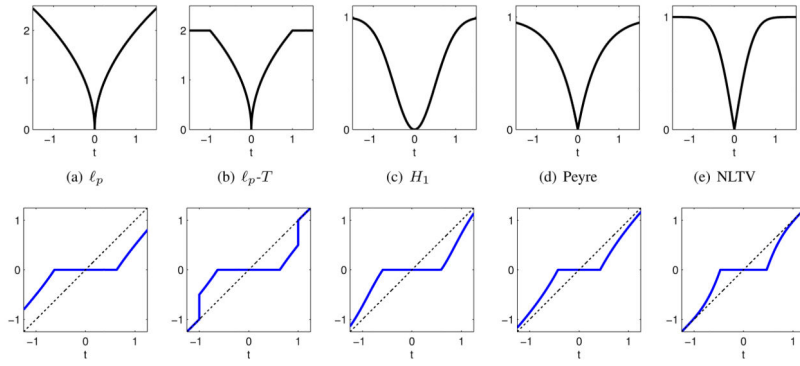
### 4) Penalty corresponding to alternating non-local TV scheme

The penalty function for the alternating non-local TV scheme is specified by [11], [41]:

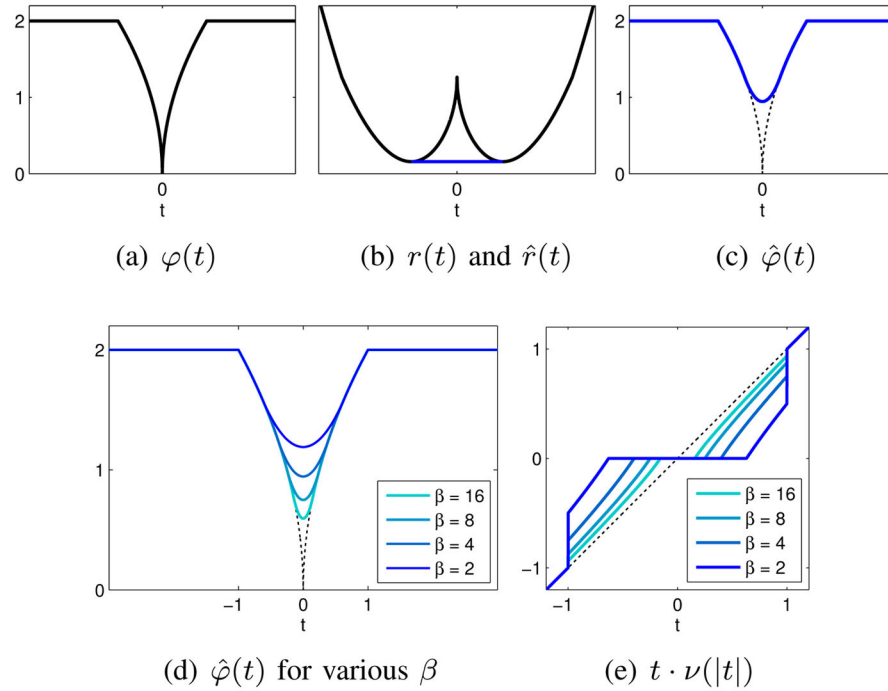
$$\phi(t) = \operatorname{erf}\left(\frac{t}{\sigma}\right). \quad (41)$$

From (21) we obtain

$$\nu(t) = \begin{cases} 0 & \text{if } \frac{2}{\sqrt{\pi}} \exp\left(-\frac{t^2}{\sigma^2}\right) > \beta\sigma t \\ 1 - \frac{2}{\sqrt{\pi}} \frac{\exp\left(-\frac{t^2}{\sigma^2}\right)}{\beta\sigma t} & \text{else} \end{cases} \quad (42)$$

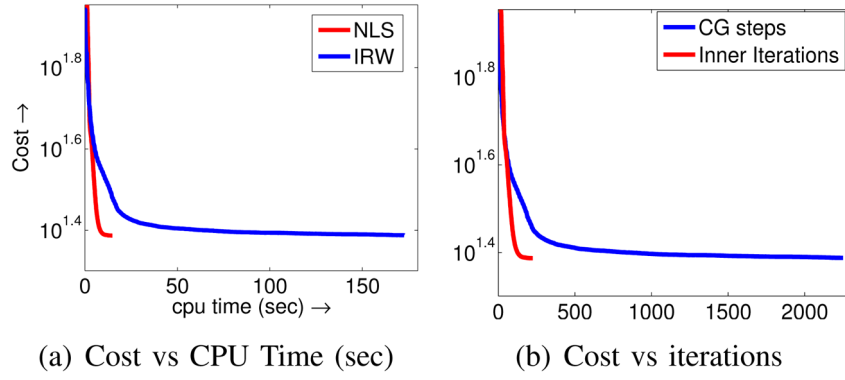


**Fig. 1.** Distances functions  $\phi(t)$  that are relevant in non-local regularization (first row) and the associated shrinkage rules  $t \cdot \nu(|t|)$  (second row); see Appendix B for the corresponding formulas that relies on a convex hull approximation (see Fig. 2) of the original penalty. Here we illustrate the shrinkage rules in 1-D for the parameter choices  $\beta = 2$ ,  $p = 0.5$ ,  $T = 1$ , and  $\sigma = 0.5$ . The approach introduced in the paper enables evaluating the shrinkage rules for a much larger class of penalties, generalizing the results in [18] for  $\ell_p$  penalties shown in the first column.

**Fig. 2.**

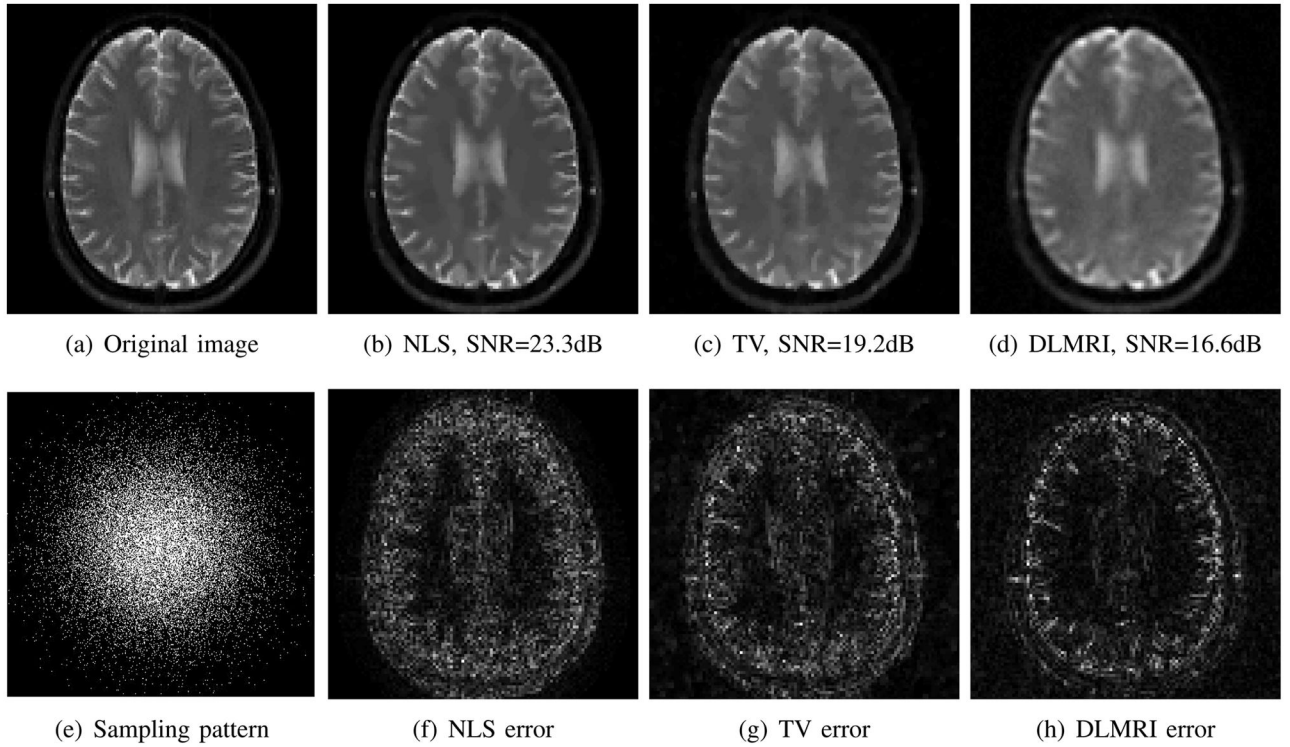
Approximation of the potential function: (a) shows the original potential function  $\varphi(t)$  in 1-D, which is the truncated  $\ell_p$ ;  $p = 0.5$  penalty,  $T = 1$ . (b) indicates the corresponding

$r(t) = t^2/2 - \frac{1}{\beta}\varphi(t)$  function with  $\beta = 2$ , shown in black. Note that this function is non-convex. Hence, we approximate this function by  $\hat{r}(t)$  shown in blue, which is the best convex approximation of  $r(t)$ . The corresponding modified potential function is shown in blue in (c). (d) indicates the approximations for different values of  $\beta$ . Note that the approximations converge uniformly to  $\varphi$ . (e) shows the corresponding shrinkage rules. The potential functions and shrinkage rules for different penalties are shown in Fig. 1.

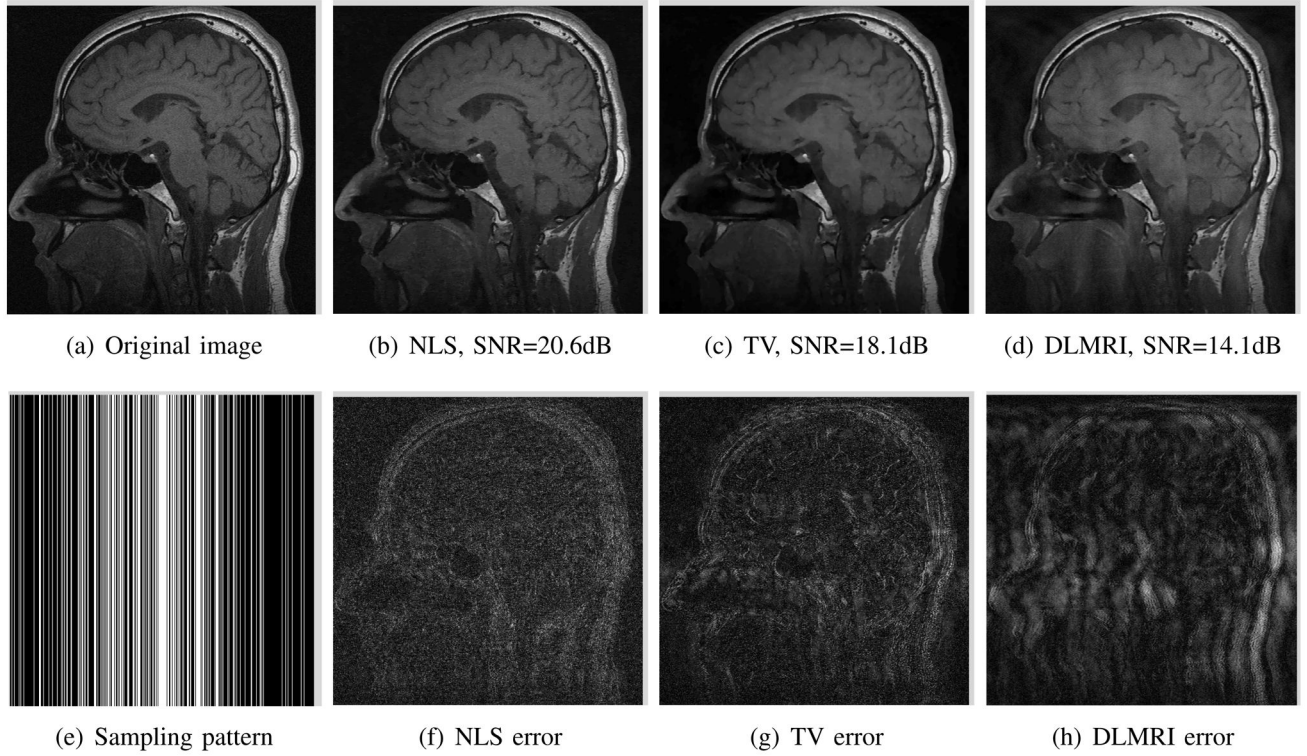
**Fig. 3.**

Comparison of the convergence rate of the iterative reweighted (IRW) algorithm and the proposed iterative non-local shrinkage (NLS) algorithm. The plots indicate the evolution of the cost function in (1) as functions of (a) the CPU time and (b) number of inner iterations in NLS and CG steps in IRW. Both NLS and IRW algorithms converged in around 225 inner iterations. However, the IRW scheme needed around 9 CG steps/inner iteration on average, requiring a total of 2200 (see Algorithms IV.1 and IV.2). Since each inner iteration in NLS is considerably faster than the corresponding one in [11], we obtain a speedup of approximately ten fold.

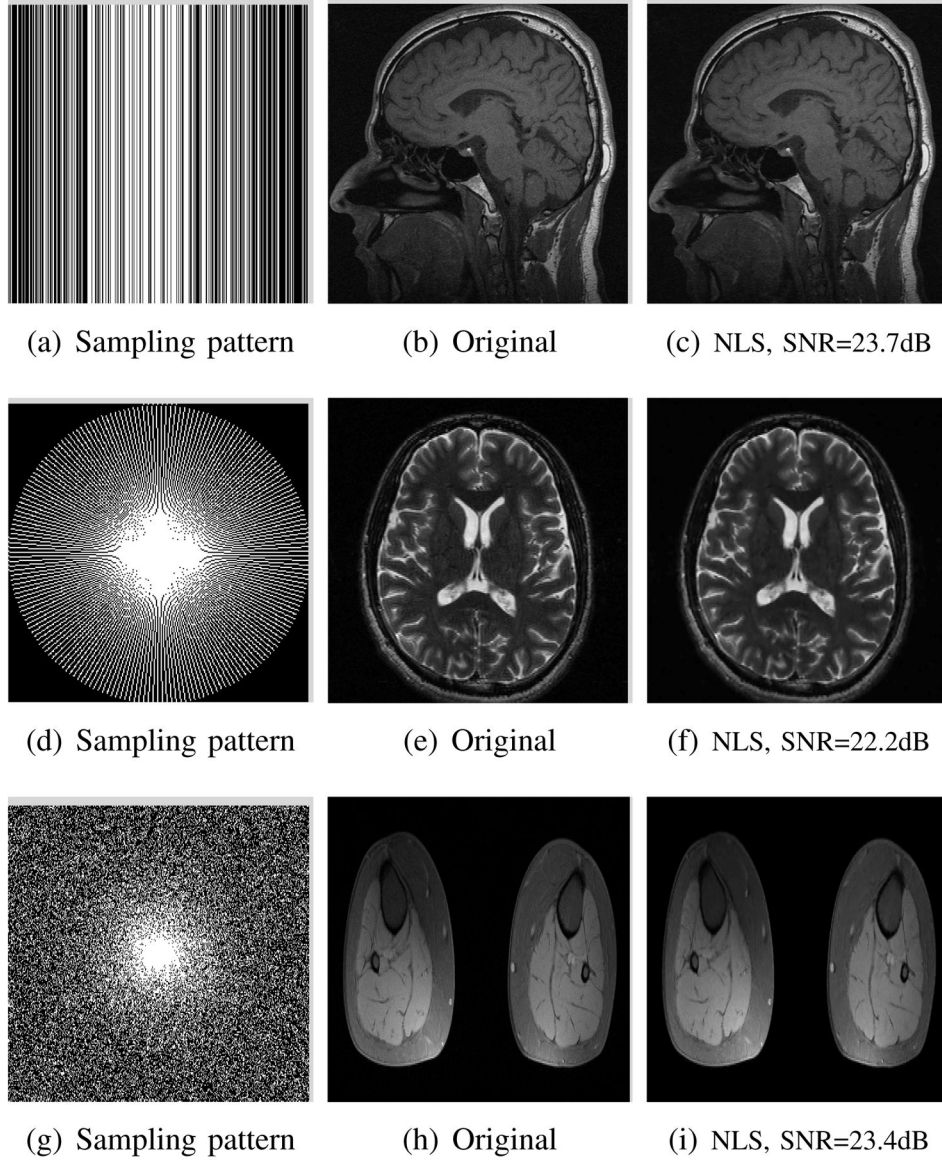


**Fig. 4.**

Comparison of the algorithms in the absence of noise. We consider the recovery of a  $128 \times 128$  MR brain image from 5 fold undersampled Fourier samples, acquired using a random sampling pattern shown in (e) using non-local shrinkage scheme (NLS), DLMRI and local TV. The reconstructions are shown in (b)-(d). The corresponding error images, scaled by a factor of 5 for better visualization, are shown in the bottom row. The reconstructions show that the NLS scheme is capable of better preserving the edges and details, resulting in less blurred reconstructions. Note that this example was used as an illustration; the proposed 2-D under sampling pattern on the dataset acquired using a 3-D sequence is not very realistic. We also used a high acceleration factor to demonstrate differences between the methods; thus the resulting images may not be of diagnostic quality.

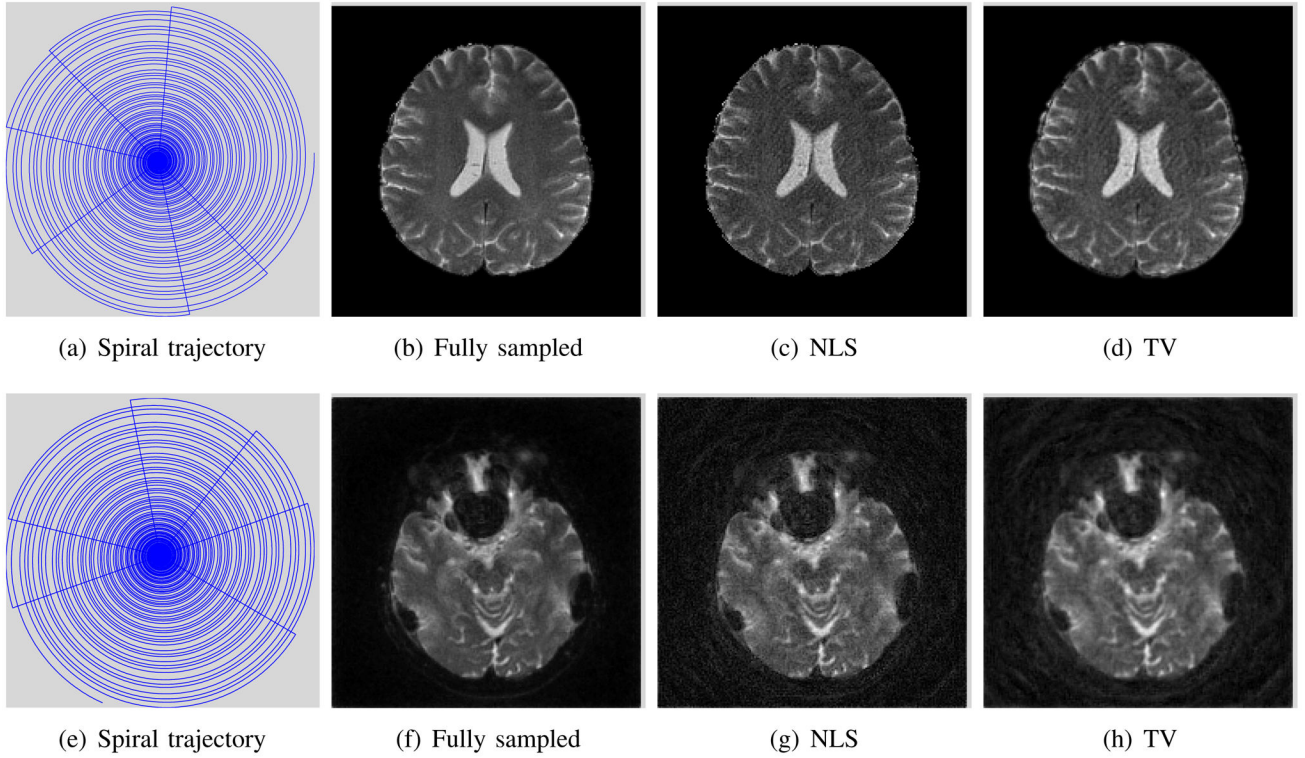
**Fig. 5.**

Comparison of the algorithms in the presence of noise. We consider the recovery of a  $512 \times 512$  MR head complex image from three-fold undersampled k-space data, acquired using a Cartesian sampling pattern contaminated by zero mean complex Gaussian such that the SNR value after adding the noise is 25.0 dB. The top row shows the original and reconstructed images using non-local shrinkage scheme (NLS), DLMRI and local TV. The reconstructions are shown in (b)-(d). The corresponding error images as well as the sampling pattern 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 better reconstructions with minimal aliasing artifacts. Note that this example was used as an illustration; the proposed 2-D under sampling pattern on the dataset acquired using a 3-D spin-echo sequence is not very realistic. Note that we used a high acceleration factor to demonstrate differences between the methods. The resulting images may not be of diagnostic quality.

**Fig. 6.**

Comparison of three different MR images using NLS scheme in the presence of noise. The Fourier samples are contaminated by zero mean complex Gaussian noise such that the SNR of the data corresponding to the head, brain2 and calf images are 32.3, 24.7 and 27.6 dB, respectively. The top row shows the recovery from 2 fold acceleration using the Cartesian sampling pattern, while the middle one shows the recovery from 100 radial spokes ( $\approx 3$  fold acceleration). The bottom row shows the recovery from 3 fold acceleration using random pattern. We observe that the NLS scheme is capable of preserving the key image details.



**Fig. 7.**

Comparison of MR brain images using NLS and TV algorithms from undersampled multichannel multi-shot spiral data. The fully sampled data was acquired using a multishot spiral sequence with 22 interleaves acquired using 12 channels. All the interleaves were used to recover the fully sampled datasets, while only 7 randomly chosen interleaves were used to recover the undersampled datasets. The retained interleaves are shown in the first column. We only used the data corresponding to four important coils for the recovery. The two rows corresponds to two slices in the acquisition. The recovery of the lower slice was considerably challenging due to field inhomogeneity losses and subtle physiological motion between interleaves; this explains the poor recovery of the datasets. The two rows show the spiral trajectory, original and reconstructed images. Both the proposed and TV regularized reconstructions are seen to have lower SNR, mainly due to decreased number of measurements. However, we observe that the proposed scheme provides sharper reconstructions.

TABLE I

(SNR in dB) Impact of the distance metric on the reconstructions. We compare the reconstructions obtained using the non-local shrinkage algorithm using  $\ell_1$ ,  $H_1$ , NLTV, thresholded  $\ell_1$  and thresholded  $\ell_2$ ;  $p = 0.5$  metrics. All the metrics, except the convex  $\ell_1$  scheme are saturating priors. We observe that saturation is key to good performance of non-local algorithms. Among the different metrics, the thresholded  $\ell_2$  penalty is observed to provide the best results in all the examples.

Image	$\ell_1$	$H_1$	NLTV	$\ell_1$ -T	$\ell_2$ -T
Brain1	19.8	22.8	23.0	21.4	23.4
Brain2	18.0	20.2	21.1	19.5	22.9
Head	19.1	19.2	19.5	19.3	19.9

**TABLE II**

(SNR in dB) Quantitative comparison of the proposed iterative non-local shrinkage (NLS) algorithm using the saturating  $\ell_p$ ;  $p = 0.5$  penalty with dictionary learning MRI (DLMRI) [26] and local total variation regularization (TV) schemes in the absence of noise. We considered five-fold random undersampling.

Image	DLMRI	TV	NLS
Brain	16.6	19.3	<b>23.4</b>
Brain2	17.5	21.0	<b>22.9</b>
Thigh	16.3	22.0	<b>24.0</b>
Calf	19.1	21.2	<b>22.5</b>
Head	18.6	19.6	<b>19.9</b>

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**TABLE III**

Quantitative comparison of the algorithms in the presence of noise. The top part shows the SNR of the reconstructions obtained from 3 fold Cartesian undersampled data. The bottom part shows the SNR of the reconstructions from radial undersampled data with 70 spokes. Both experiments are contaminated by zero mean complex Gaussian noise such that the SNR of the noisy images are reported in the table. The quantitative results show that the proposed iterative NLS scheme provides consistently better reconstructions for the above cases.

Image	k-space SNR	DLMRI	TV	NLS
<b>Brain</b>	12.9	13.9	15.0	<b>16.8</b>
<b>Brain2</b>	18.7	15.9	16.1	<b>18.5</b>
<b>Thigh</b>	23.7	12.6	18.7	<b>21.7</b>
<b>Calf</b>	21.6	15.9	18.5	<b>20.6</b>
<b>Head</b>	25.0	14.1	18.1	<b>20.3</b>
<b>Brain</b>	11.6	14.6	<b>17.6</b>	17.3
<b>Brain2</b>	17.4	17.0	17.7	<b>18.0</b>
<b>Thigh</b>	22.4	13.8	17.2	<b>20.0</b>
<b>Calf</b>	20.3	9.3	16.7	<b>19.0</b>
<b>Head</b>	23.4	17.9	18.1	<b>18.2</b>