Nonlocal regularization of inverse problems: a unified variational framework

Zhili Yang, *Student Member, IEEE* and Mathews Jacob, *Member, IEEE*

***Abstract*—We introduce a unifying energy minimization framework for nonlocal regularization of inverse problems. In contrast to the weighted sum of square differences between image pixels used by current schemes, the proposed functional is an unweighted sum of inter-patch distances. We use robust distance metrics that promote the averaging of similar patches, while discouraging the averaging of dissimilar patches. We show that the first iteration of a majorize-minimize algorithm to minimize the proposed cost function is similar to current non-local methods. The reformulation thus provides a theoretical justification for the heuristic approach of iterating non-local schemes, which re-estimate the weights from the current image estimate. Thanks to the reformulation, we now understand that the widely reported alias amplification associated with iterative non-local methods are caused by the convergence to local minimum of the non-convex penalty. We introduce an efficient continuation strategy to overcome this problem. The similarity of the proposed criterion to widely used non-quadratic penalties (eg. total variation and ℓ_p semi-norms) opens the door to the adaptation of fast algorithms developed in the context of compressive sensing; we introduce several novel algorithms to solve the proposed non-local optimization problem. Thanks to the unifying framework, these fast algorithms are readily applicable for a large class of distance metrics.**

***Index Terms*—nonlocal means, compressed sensing, inverse problems, non-convex**

I. Introduction

The recovery of images from their few noisy linear measurements is an important problem in several areas, including remote sensing [1], biomedical imaging, astronomy [2], and radar imaging. The standard approach is to formulate the recovery as an optimization problem, where the linear combination of data consistency error and a regularization penalty is minimized. The regularization penalty exploits the apriori image information (eg.image smoothness [3], [4], transform domain sparsity [5]–[7]) to make the recovery problem well-posed.

Nonlocal means (NLM) denoising schemes have recently received much attention in image processing [8]–[10]. These methods exploit the similarity between rectangular patches in the image to reduce noise. Specifically, each pixel in the denoised image is recovered as a weighted linear combination of all the pixels in the noisy image. The weight between two pixels is essentially a measure of similarity between their patch neighborhoods (rectangular image regions, centered on the specified pixels) (see Fig.1). Recently, several authors have extended the non-local smoothing algorithm by reformulating it as a regularized reconstruction scheme. The regularization functional is the weighted sum of square differences between all the pixel pairs in the image [11]–[13]. The main challenge in applying this method to general inverse problems is the explicit dependence of the regularization penalty on pre-determined weights. In contrast to denoising and deblurring, good initial image guesses are often not available for many challenging inverse problems (eg. compressed sensing), which makes the reliable estimation of inter-pixel weights difficult. Some authors have suggested to iterate non-local schemes to improve the performance of deblurring and denoising algorithms; they re-estimate the weights from the current image iterate [14]–[16]. However, the use of this strategy to recover the image from its sparse Fourier samples results in the enhancement of alias patterns. Hence, this approach is not frequently used in such challenging inverse problems. Another limitation of current methods is that different optimization algorithms are required for each choice of regularization functional and weight [11], [14], [15], [17]; the algorithms designed for one penalty are often not readily applicable to other functionals.

To overcome the above mentioned problems, we introduce a unifying nonlocal regularization framework. We choose the regularization functional as an unweighted sum of non-Euclidean distances between patch pairs (see Fig. 1.a). We use robust distance metrics to promote the averaging of similar patches, while minimizing the averaging of dissimilar patches. Since the proposed criterion is not dependent on pre-estimated weights, the quality of the reconstructions is independent of the initial guess used for weight estimation. We show that a majorization of the proposed regularization penalty is very similar to current non-local regularization functionals [11], [14], [15], [17], when the robust distance metric is chosen appropriately. Thus, the fixed-weight NL schemes are similar to the first iteration of a majorize-minimize algorithm to solve the criterion. More importantly, the formulation provides a theoretical justification for the heuristic approach of iterating the NL algorithms by re-estimating the weights from the current image estimate [14], [15]. The availability of the global criterion, which does not change with iterations, enables us to analyze the convergence and design efficient algorithms; this approach is different from earlier methods that analyzed and optimized only one step of the above iterative scheme [11]–[15]. The practical benefits of the proposed reformulation are as follows:

Z. Yang is with the Department of Electrical and Computer Engineering, University of Rochester, NY, USA.

M. Jacob is with the Department of Electrical and Computer Engineering, University of Iowa, IA, USA.

e-mail: (see http://www.engineering.uiowa.edu/ jcb/). This work is supported by NSF awards CCF-0844812 and CCF-1116067.

- We now understand that the reason behind enhancement of alias artifacts, which are commonly reported in the context of current iterative weight update schemes, is caused by the convergence to the local minimum of the proposed criterion. We introduce homotopy continuation schemes to minimize such local minima problems, inspired by similar methods in compressed sensing [18]. Our experiments show that this approach eliminates the local minima issues in all the cases that we considered.
- The similarity of the proposed criterion to similar penalties in compressive sensing makes it possible to exploit the extensive literature in non-quadratic optimization (e.g. [19], [20]) to significantly improve computational efficiency. We introduce three majorize minimize (MM) algorithms, which rely on **(a)** the majorization of the penalty term (denoted as PM scheme), **(b)** the majorization of data term (indicated as DM algorithm), and **(c)** majorization of both data and penalty terms (termed as DPM scheme), respectively. Our experiments show that the PM scheme requires fewer computationally expensive weight computations and hence is computationally much more efficient than the DM and DPM schemes.
- Thanks to the unified perspective, it is possible to use the efficient PM algorithm for all non-local distance metrics. Previous methods required customized algorithms for each flavor of NL regularization.

The proposed framework is related to generalized nonlocal denoising schemes introduced in [21], [30]; they show that the iterative re-estimation of the weights results in improved reconstructions. However, the algorithms in [21], [30] are specifically designed for the denoising setting and are not applicable to general inverse problems, which is the main focus of this paper. This work is also related to convex patch based regularization scheme in [17], which is published in the same proceedings as the conference version of this paper [22]. The PM-CG algorithm provides faster convergence compared to DPM algorithm used in [17] (see the results section). In addition, we observe that non-convex nonlocal distance functions, along with homotopy continuation, provide significantly ameliorated results over the convex ℓ_1 metric considered in [17].

II. Background

A. Current nonlocal algorithms

The classical NL means algorithm was originally designed for denoising. It derives each pixel in the denoised image as the weighted average of all the pixels in the noisy image[1] $f:\Omega\to\mathbb{R}$:

$$\hat{f}(\mathbf{x}) = \frac{\sum_{\mathbf{y}} w(\mathbf{x},\mathbf{y}) f(\mathbf{x})}{\sum_{\mathbf{y}} w(\mathbf{x},\mathbf{y})} \tag{1}$$

The weight function $w(\mathbf{x},\mathbf{y})$ is estimated from the noisy image or its smoothed version g as the similarity between the patch neighborhoods of the specific pixels:

$$w(\mathbf{x},\mathbf{y}) = \exp\left(-\frac{\|P_{\mathbf{x}}(g)-P_{\mathbf{y}}(g)\|_\eta^2}{\sigma^2}\right). \tag{2}$$

Here, $P_{\mathbf{x}}(g)$ is a $(2N_p+1)\times(2N_p+1)$ image patch of g, centered at $\mathbf{x}$:

$$P_{\mathbf{x}}(f) = f(\mathbf{x}+\mathbf{p});\ \mathbf{p}\in \underbrace{[-N_p,..N_p]\times[-N_p,..N_p]}_{\mathcal{B}_{\mathbf{x}}}, \tag{3}$$

and $\|\cdot\|_\eta$ denotes the weighted ℓ_2 metric defined as

$$\|f\|_\eta^2 = \sum_{\mathbf{p}\in\mathcal{B}_{\mathbf{x}}} |f(\mathbf{p})|^2\ \eta(\mathbf{p}). \tag{4}$$

$\mathcal{B}_{\mathbf{x}}$ denote the pixels in the patch $P_{\mathbf{x}}(f)$ and $\eta(\mathbf{p})$ the window function. η is often chosen as a Gaussian function ($\eta(\mathbf{p}) = \exp\left(-\|\mathbf{p}\|^2/2N_p^2\right)$) to give more weight to the center pixel.

Recently, several authors have extended the nonlocal smoothing scheme [8]–[10] to deblurring and denoising problems by posing the image recovery as an optimization scheme [12], [13]:

$$\hat{f} = \arg\min_f \|\mathcal{A}f-b\|_2^2 + \lambda J_w(f). \tag{5}$$

Here, the noisy measurements of the image are acquired by the ill-conditioned linear operator $\mathcal{A}$. λ is the regularization parameter and $\mathcal{J}_w(f)$ is the nonlocal regularization functional. The subscript w is used to indicate that $\mathcal{J}_w(f)$ is explicitly dependent on pre-specified weights w. Several flavors of NL regularization penalties have been recently introduced. For example, H_1 nonlocal regularization uses the regularization functional:

$$J_{H_1}(f) = \sum_{\mathbf{x}\in\Omega}\sum_{\mathbf{y}\in\mathcal{N}_{\mathbf{x}}} w(\mathbf{x},\mathbf{y})\ |f(\mathbf{x})-f(\mathbf{y})|^2, \tag{6}$$

where the weights are specified as in (2). Note that the search window is restricted to the square neighborhood of $\mathbf{x}$, denoted as $\mathcal{N}_{\mathbf{x}}$. This restriction is often used to keep the computational complexity manageable. Gilboa et. al., have suggested to replace the penalty term in (6) as

$$J_{\mathrm{TV}}(f) = \sum_{\mathbf{x}\in\Omega}\sum_{\mathbf{y}\in\mathcal{N}_{\mathbf{x}}} w(\mathbf{x},\mathbf{y})\ |f(\mathbf{x})-f(\mathbf{y})|, \tag{7}$$

while keeping the expression for the weights as in (2), to improve the quality of the reconstructions [11], [12]. This cost function is termed as nonlocal total variation (TV) penalty. Similarly, Peyre has introduced the regularization functional [15]:

$$J_{\mathrm{peyre}}(f) = \sum_{\mathbf{x}\in\Omega}\sum_{\mathbf{y}\in\mathcal{N}_{\mathbf{x}}} w(\mathbf{x},\mathbf{y})\,\|P_{\mathbf{x}}(f)-P_{\mathbf{y}}(f)\|_\eta, \tag{8}$$

where the weights are chosen as

$$w(\mathbf{x},\mathbf{y}) = \exp\left(-\frac{\|P_{\mathbf{x}}(g)-P_{\mathbf{y}}(g)\|_\eta}{\sigma}\right). \tag{9}$$

Custom designed non-linear iterative algorithms are introduced to solve the regularized reconstruction problems for each choice of regularization penalty [12], [13], [15]. The algorithms that are designed for one specific penalty are often not

[1] $\Omega\subset\mathbb{R}^2$ is the spatial support of the complex image; it is often chosen as the rectangular region $[0,T_1]\times[0,T_2]$

readily applicable for other functionals. The main challenge with the above formulations is the dependence of the cost function on pre-specified weights. The popular approach is to derive g using other algorithms (e.g. Tikhonov regularization, local TV regularization). While this approach works well in denoising and deblurring, it often results in poor weights in challenging inverse problems. Some researchers have proposed to iterate the NL framework by re-estimating the weights from the previous iterations [14], [15]. However, this approach is often not used in image recovery from sparse Fourier samples, for the fear of the weights learning the alias patterns. Since the weight-dependent cost function changes from iteration to iteration, it is difficult to analyze the convergence of this scheme.

B. Majorize Minimize (MM) optimization framework

We propose to use the majorize-minimize (MM) framework to develop fast algorithms to solve the proposed optimization problem. MM algorithms are widely used in the context of compressed sensing [23], [24]. The practice is to reformulate the original problem as the solution to a sequence of simpler quadratic surrogate problems. The surrogate criteria, denoted by $\mathcal{C}_n(f)$, majorize the original objective function $\mathcal{C}(f)$, and are dependent on the current iterate f_n:

$$\mathcal{C}(f) \leq \mathcal{C}_n\left(f\right), \forall f; \quad \mathcal{C}_n(f_{(n)}) = \mathcal{C}\left(f_{(n)}\right). \tag{10}$$

Thus, the mth iteration of the MM algorithm involves the following two steps

1) evaluate the majorizing functional $\mathcal{C}_n(f)$ that satisfy (10), and
2) solve for $f_{n+1} = \arg\min_f \mathcal{C}_n(f)$ using an appropriate quadratic solver (e.g. CG algorithm).

The above two-step approach is guaranteed to monotonically decrease the cost function $\mathcal{C}(f)$. If $\theta(v) = \phi(\sqrt{v})$ is a concave function, it can be majorized by:

$$\phi\left(\sqrt{v}\right) \leq c(v)\, v + b(v), \tag{11}$$

where $c(v) = \theta'(v) = \frac{\phi'(\sqrt{v})}{2\sqrt{v}}$ and $b(v) = \theta\left(\theta'^{-1}(v)\right) - v\,\theta'^{-1}(v)$ [25]; since θ is concave, θ' is monotonic and hence invertible. The general practice in majorize-minimize/half-quadratic algorithms is to assume that $c(v)$ and $b(v)$ are constants at each iteration. Note that the left hand side is the equation of a straight line, when c and b are assumed to be constants. We use this relation to derive efficient MM algorithms for nonlocal regularization.

III. Unified non-local regularization

We now introduce a unified NL regularization framework, which is independent of pre-specified weights. We also illustrate the similarity of the majorization of the proposed penalty and current NL regularization schemes.

A. Robust nonlocal regularization

We pose the nonlocal regularized reconstruction of the complex image $f : \Omega \to \mathbb{C}$, supported on Ω, as the optimization problem:

$$\hat{f} = \arg\min_f \|\mathcal{A}f - b\|_2^2 + \lambda\, \mathcal{G}(f), \tag{12}$$

where, the regularization penalty $\mathcal{G}(f)$ is specified by

$$\mathcal{G}(f) = \sum_{\mathbf{x}\in\Omega} \sum_{\mathbf{y}\in\mathcal{N}_\mathbf{x}} \phi(\|P_\mathbf{x}(f) - P_\mathbf{y}(f)\|_\eta). \tag{13}$$

Here, $\phi : \mathbb{R} \to \mathbb{R}$ is an appropriately chosen robust distance metric, which weight large differences less heavily than small differences. One possible choice is the class of $\ell_p; p \leq 1$ semi-norms:

$$\phi(x) = |x|^p. \tag{14}$$

If ϕ is strictly convex, the solution of (12) is unique. However, our experiments show that non-convex metrics gives reconstructions with less blurring than convex metrics. The term $P_\mathbf{x}(f)$ in (13) denotes a square patch, which is centered at $\mathbf{x}$. The size of the patch is assumed to be smaller than that of the search window $\mathcal{N}_\mathbf{x}$ (see Fig. 1). While the search window can be chosen as the support of the image (i.e., $\mathcal{N}_x = \Omega$), it is often chosen as a local neighborhood in the interest of computational efficiency. The proposed penalty is illustrated in Fig. 1.a. Note that the nonlocal penalty $\mathcal{G}(f)$ in (13) is only specified by the distance metric ϕ; it is not dependent on any apriori selected weight function. This property makes the proposed scheme independent of the specific algorithm used to derive the initial image guess g, which is used to estimate the weights. More importantly, the new framework can be readily applied to ill-conditioned inverse problems, where good initial guesses are difficult to derive. We will now demonstrate the relation between the proposed formulation and current one-step NL schemes and NL methods that re-estimate the weights.

B. Majorization of the robust non-local penalty

We majorize (13) by a simpler quadratic surrogate functional. Setting $v = \|P_\mathbf{x}(f) - P_\mathbf{y}(f)\|^2$ in (11), we obtain

$$\phi\left(\|P_\mathbf{x} - P_\mathbf{y}\|_\eta\right) \leq\ w_f\left(\mathbf{x},\mathbf{y}\right)\ \|P_\mathbf{x} - P_\mathbf{y}\|_\eta^2 + b, \tag{15}$$

where $w_f\left(\mathbf{x},\mathbf{y}\right)$ is specified by

$$w_f(\mathbf{x},\mathbf{y}) = \frac{\phi'(\|P_\mathbf{x}(f) - P_\mathbf{y}(f)\|_\eta)}{2\|P_\mathbf{x}(f) - P_\mathbf{y}(f)\|_\eta}. \tag{16}$$

Note that the weights $w_f(\mathbf{x},\mathbf{y})$ are obtained as a non-linear function of the Euclidean distance between patches $\psi(\|P_\mathbf{x}(f) - P_\mathbf{y}(f)\|_\eta)$, where

$$\psi(x) = \frac{\phi'(x)}{2x}. \tag{17}$$

As discussed previously, the weights and the parameter b are assumed to be constants in each iteration of the corresponding majorize minimize algorithm. The parameter b can hence be safely ignored in the optimization process. We thus obtain

$$\mathcal{G}(f) \leq \underbrace{\sum_{\mathbf{x}\in\Omega} \sum_{y\in\mathcal{N}_\mathbf{x}} w_f\left(\mathbf{x},\mathbf{y}\right) \|P_\mathbf{x}(f) - P_\mathbf{y}(f)\|_\eta^2}_{\mathcal{G}_w(f)}. \tag{18}$$

The non-linear function ψ used to estimate the weights (see (16)) is a monotonically decreasing function of its argument, for all robust distance metrics ϕ that are of practical importance (see Fig. 1.b and Fig. 2). Thus, $\psi(\|P_{\mathbf{x}}(f) - P_{\mathbf{y}}(f)\|)$ is the measure of similarity between the patches $P_{\mathbf{x}}(f)$ and $P_{\mathbf{y}}(f)$.

Note that (18) involves the weighted norm of the patch differences. This expression is the sum of pixel differences:

$$\begin{aligned}\mathcal{G}_{w_n}(f) &= \sum_{\mathbf{x}\in\Omega}\sum_{y\in\mathcal{N}_{\mathbf{x}}} w_f(\mathbf{x},\mathbf{y}) \overbrace{\sum_{\mathbf{p}\in\mathcal{B}_{\mathbf{x}}} \eta(\mathbf{p})\,|f(\mathbf{x}+\mathbf{p}) - f(\mathbf{y}+\mathbf{p})|^2}^{\|P_{\mathbf{x}}(f)-P_{\mathbf{y}}(f)\|_\eta^2} \\ &= \sum_{\mathbf{x}\in\Omega}\sum_{y\in\mathcal{N}_{\mathbf{x}}} \gamma_f(\mathbf{x},\mathbf{y})\,|f(\mathbf{x}) - f(\mathbf{y})|^2 \end{aligned} \tag{19}$$

We used a change of variables $\mathbf{x} = \mathbf{x}+\mathbf{p}$ and $\mathbf{y} = \mathbf{y}+\mathbf{p}$ and the symmetry of $\eta(\mathbf{p})$ to derive the second step in (19). The weights γ_f are specified by

$$\gamma_f(\mathbf{x},\mathbf{y}) = \sum_{\mathbf{p}\in\mathcal{B}_{\mathbf{x}}} \eta(\mathbf{p})\, w_f(\mathbf{x}-\mathbf{p},\mathbf{y}-\mathbf{p}). \tag{20}$$

Note that the surrogate criterion $\mathcal{G}_w$ in (19) is similar to the H_1 NL penalty. The only difference is that the weights $\gamma_f(\mathbf{x},\mathbf{y})$ are obtained as the sum of the similarity measures $w_f(\mathbf{x},\mathbf{y})$ between all the patch pairs that contain the pixels $\mathbf{x}$ and $\mathbf{y}$ (See Fig. 1.b). This summation is required to ensure that the algorithm is consistent with the minimization of the cost-function (12). We will now show the similarity of the current methods to the first iteration of the proposed scheme.

C. Similarity to current non-local methods

We now illustrate the similarity of current nonlocal regularization penalties to the surrogate functional $\mathcal{G}_w(f)$. Specifically, we choose the robust distance function $\phi(\cdot)$ such that the expressions of $\mathcal{G}_w(f)$ and $w_f(\mathbf{x},\mathbf{y})$ match the current schemes.

1) H_1 nonlocal regularization: If the distance metric is chosen as

$$\phi(z) = \left(1 - \exp\left(-\frac{z^2}{\sigma^2}\right)\right), \tag{21}$$

we obtain $\psi(z) = \exp\left(-\frac{z^2}{\sigma^2}\right)/\sigma^2$ using (16). Note that this choice of weights is very similar to the classical H_1 NL regularization. The main difference between the majorization and the H_1 scheme is that the weights in (19) are obtained as the summation of the similarity measures (see (20)). In contrast, the similarity measures themselves are used as weights in classical H_1 regularization. For convenience, we will refer to the metric in (21) as the H_1 distance function.

2) Peyre's scheme: The NL penalty term (8) in Peyre's NL method involves a weighted sum of ℓ_1 norms of patch differences. Peyre's scheme can also be expressed as a majorization of the proposed penalty by setting $\phi(v) = \varphi(\sqrt{v})$ and using (11):

$$\varphi\left(\sqrt{\|P_{\mathbf{x}}f - P_{\mathbf{y}}f\|}\right) \le \underbrace{\left(\frac{\varphi'(\sqrt{\|P_{\mathbf{x}}f - P_{\mathbf{y}}f\|})}{2\sqrt{\|P_{\mathbf{x}}f - P_{\mathbf{y}}f\|}}\right)}_{\eta(\mathbf{x},\mathbf{y})} \|P_{\mathbf{x}}f - P_{\mathbf{y}}f\| + b, \tag{22}$$

Here, b is a constant that is safely ignored. The right hand side of the above expression is the same form as (8). From the above expression, the weights are obtained as $\eta(\mathbf{x},\mathbf{y}) = (\varphi'(v)/2v)$, where $v = \sqrt{\|P_{\mathbf{x}}f - P_{\mathbf{y}}f\|}$. Comparing with the expression of the weights used in Peyre's scheme, specified by (9), we have $\frac{\varphi'(v)}{2v} = e^{-v^2/\sigma}$. This equation is satisfied if $\varphi(v) = \sigma\left(1 - e^{-v^2/\sigma}\right)$. Thus, the equivalent distance metric in (12) is given by

$$\phi(x) = \varphi(\sqrt{x}) = \left(1 - e^{-x/\sigma}\right). \tag{23}$$

We ignored the constant factor σ to ensure consistency between the different metrics. For convenience, we will refer to the metric in (23) as Peyre's distance function in the rest of the paper.

3) Nonlocal TV (NLTV) scheme: The nonlocal TV penalty in (7) involves the weighted ℓ_1 norm of pixel differences. Since the ℓ_1 norm of differences between two patches cannot be expressed as the sum of ℓ_1 norms of the corresponding pixel differences, the NLTV scheme cannot be expressed as a special case of the proposed framework. However, if the penalty involves the ℓ_1 norm of the patch differences as in (8), the corresponding NL scheme can be viewed as the majorization of the proposed scheme. Proceeding as in the Peyre's case, we have $\frac{\varphi'(\sqrt{v})}{2\sqrt{v}} = e^{-v^2/\sigma^2}$. If we set $\phi(v) = \varphi(\sqrt{v})$, we have $\phi'(v) = \frac{\varphi'(\sqrt{v})}{2\sqrt{v}} = e^{-v^2/\sigma^2}$. This relationship is satisfied when the distance metric is specified by

$$\phi(x) = \operatorname{erf}\left(\frac{x}{\sigma}\right) \tag{24}$$

For convenience, we will refer to the metric in (24) as NL TV distance function in the rest of the paper.

4) Convex nonlocal regularization: All of the above distance metrics are non-convex. Hence, the corresponding algorithms are not guaranteed to converge to the global minimum. The distance function can be chosen as a convex function to overcome this problem [17]:

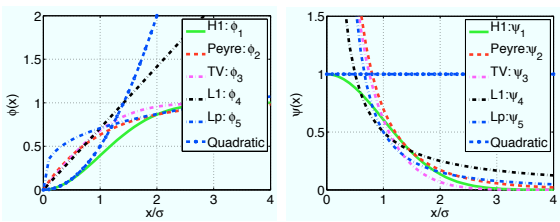
$$\phi(x) = \frac{|x|}{\sigma}. \tag{25}$$

Our experiments show that use of such convex cost functions result in blurring at high acceleration factors, compared to the non-convex choices considered above.

We list the current NL schemes, which are specified by the regularization functional $\mathcal{J}_w$ and the specific formula to compute the weights, in Table 1. We re-interpret these methods as the first iteration of a MM scheme to solve for the minimum of (12). The corresponding penalty functions ϕ are also shown in Table 1. We also show in next section that the proposed criterion can be efficiently minimized using MM algorithms, which rely on the re-computation of weights using ψ; the expressions for ψ is also listed in Table 1. The ϕ and the ψ functions are plotted in Fig. 2. We also plot the quadratic penalty $\phi(x) = |x|^2$ and the corresponding weights for comparison. It is seen that the quadratic choice encourages the averaging of all patches in the neighborhood. In contrast, the equivalent ϕ functions of current nonlocal schemes (H_1, TV, and Peyre's scheme) saturate with increasing inter-patch

TABLE I
REINTERPRETATION OF CURRENT NONLOCAL SCHEMES

Current methods		**Re-interpretation**	
Reference	Current algorithm $\mathcal{J}_w(f)$ and $w(\mathbf{x},\mathbf{y})$	Penalty function $\phi(x)$	Weight function $\psi(x)$
H_1 [11], [14]	(6) and (2)	$\left(1-\exp\left(-x^2/2\sigma^2\right)\right)$	$\exp\left(-x^2/2\sigma^2\right)$
Peyre et al., [15]	(8) and (9)	$\left(1-e^{-x/\sigma}\right)$	$\frac{\exp(-x/\sigma)}{2\sigma x}$
nonlocal TV [11], [12]	(7) and (2)	$\mathrm{erf}\left(\frac{x}{\sigma}\right)$	$\frac{\exp(-x^2/\sigma^2)}{\sqrt{\pi}\sigma\ x}$
nonlocal $L1$ [17]	-	$\frac{x}{\sigma}$	$\frac{1}{2\sigma x}$



(a) distance metric (b) weight function

Fig. 2. Comparison of the distance metrics ϕ and corresponding weight functions ψ: The different distance metrics and the weight functions are plotted in (a) and (b), respectively. Note that the convex distance functions, shown by the blue and black curves, do not saturate with the Euclidean inter-patch distances. In contrast, the distance functions corresponding to the current nonlocal schemes saturate as the patches become dissimilar. This ensures that the distances between dissimilar patches are not penalized, thus minimizing the blurring compared to the convex choices. This is also observed by the weights in the sum in (6). Note that the weights associated with the current schemes decay to zero for large distances. The slow decay of the weights associated with the convex metrics can result in blurred reconstructions. The $\ell_p; p = 0.3$ norms saturate rapidly, when compared to other non-local metrics, resulting in reduced averaging of similar patches.

distance. This behaviour discourages the averaging of dissimilar patches, thus minimizing the blurring, compared to quadratic schemes that encourage uniform smoothing. Note that the above metrics saturate faster than the convex ℓ_1 metric, thus providing reduced blurring. Note that this case is very different from convex local smoothness regularization, where the penalty only involves distances a pixel and its immediate neighbors. Since the NL penalty involves the distances between each patch and several other patches, this penalty can result in significant blurring if the functional does not saturate with the Euclidean distance.

IV. NUMERICAL ALGORITHM

In the previous section, we used the majorization of the penalty term in (12) to illustrate the similarity of the proposed scheme to current NL methods. We now realize efficient algorithms based on alternate majorizations of (12). Since the penalty term is majorized in Section III-B, we term the resulting scheme as a penalty majorization (PM) algorithm. We also consider algorithms based on the majorization of the data consistency term (denoted by DM), and majorization of both data consistency and penalty terms (denoted by DPM). Thanks to the unified treatment, all of these algorithms can be used with any distance function ϕ. This is an advantage over current schemes, which develop customized algorithms for each specific choice of $\mathcal{J}_w$ and ψ [11], [14], [15].

A. Majorization of the penalty term (PM)

We use the majorization of the regularization penalty, introduced in Section III-B, to develop a two-step alternating algorithm. The algorithm alternates between the solution of the quadratic surrogate problem:

$$\hat{f}_{n+1} = \underbrace{\arg\min_f \|\mathbf{A}\mathbf{f}-\mathbf{b}\|_2^2 + \lambda \overbrace{\mathbf{f}^T \mathbf{\Gamma}_n \mathbf{f}}^{\mathcal{G}_w(f)}}_{\mathcal{C}_{n+1}}. \tag{26}$$

and the estimation of the weights from the current iterate as $\mathbf{w}_{f_n}(\mathbf{x},\mathbf{y}) = \psi\left(\|P_{\mathbf{x}}(f_n) - P_{\mathbf{y}}(f_n)\|_\eta\right)$. Here, $\mathbf{f}$ is the vectorized image, $\mathbf{A}$ is the matrix operator corresponding to the forward model, and $\mathbf{\Gamma}_n$ is the sparse matrix with the entries: $\mathbf{\Gamma}_n(\mathbf{x},\mathbf{y}) = \gamma_{f_n}(\mathbf{x},\mathbf{y})$. Here, $\gamma_{f_n}(\mathbf{x},\mathbf{y})$ is the sum of similarity measures $\mathbf{w}(\mathbf{x},\mathbf{y})$ using (20). This approach is inspired by the iterative reweighted least squares (IRLS) methods widely used in total variation and ℓ_1 minimization schemes [26]. Note that the optimization criterion is a quadratic. The gradient of this criterion is obtained as

$$\nabla\mathcal{C}_{n+1} = 2\left(\mathbf{A}^T\left(\mathbf{A}\mathbf{f}-\mathbf{b}\right) + \lambda\mathbf{\Gamma}_n\mathbf{f}\right). \tag{27}$$

We propose to minimize the sub-criteria $\mathcal{C}_{n+1}(f)$ using conjugate gradients algorithm (CG). We observe the CG algorithm provides significantly faster convergence compared to the steepest descent scheme used in [13] to solve (26). The CG scheme may be further accelerated using efficient preconditioning steps.

This two-step alternating scheme is similar to iterating the classical H_1 NL algorithm, followed by the re-estimation of the weights from the current image iterate. However, the interpretation of this approach as a specific algorithm to solve the global optimization problem (12) provides useful insights on the alternating strategy. For example, it enables us to monitor whether the algorithm is trapped in local minima and device continuation strategies to overcome such problems. This reinterpretation may also enable the development of novel algorithms to directly minimize (12). Moreover, since the different NL regularization penalties (NL TV, ℓ_1, and Peyre's schemes) can be interpreted as a special case of the global penalty, the same algorithms can be used for a range of NL penalties.

B. Majorization of the data term (DM)

The data majorization approach was originally introduced by Combettes et al., [27] for ℓ_1 regularized problems. The main idea is to majorize the data term in (12) as

$$\|\mathbf{A}\mathbf{f}-\mathbf{b}\|^2 \leq \tau\|\mathbf{f}-\mathbf{z}_n\|^2, \tag{28}$$

where

$$\mathbf{z}_n = \mathbf{f}_n - \mathbf{A}^T\left(\mathbf{A}\mathbf{f}_n - \mathbf{b}\right) \tag{29}$$

and $\tau\mathbf{I} \leq \mathbf{A}^T\mathbf{A}$, for an appropriately chosen τ. For example, when $\mathbf{A}$ is the one dimensional Cartesian Fourier undersampling operator, we have $\tau = \frac{M}{N}$, where M is the number of the samples that are retained and N is the length of the signal. Using this majorization, we obtain the surrogate criterion of (12) at the nth iteration as

$$\mathcal{C}_n(f) = \|f - z_n\|^2 + \frac{\lambda}{\tau}\,\mathcal{G}(f) \tag{30}$$

The minimization of $\mathcal{C}_n(f)$ is essentially a denoising problem, which can be solved using the fixed point iterative algorithm [21]. Specifically, the Euler-Lagrange equation of the above criterion is given by

$$z_n - f = \frac{\lambda}{2\tau}\sum_{\mathbf{y}\in\mathcal{N}_\mathbf{x}} \underbrace{\frac{\phi'(\|P_\mathbf{x}(f)-P_\mathbf{y}(f)\|_\eta)}{\|P_\mathbf{x}(f)-P_\mathbf{y}(f)\|_\eta}}_{w_f(\mathbf{x},\mathbf{y})=\psi(\|P_\mathbf{x}(f)-P_\mathbf{y}(f)\|_\eta)} \left(P_\mathbf{x}(f)-P_\mathbf{y}(f)\right)_\eta \tag{31}$$

This equation can be rewritten in terms of pixel differences as

$$f(\mathbf{x}) - z_n(\mathbf{x}) + \frac{\lambda}{2\tau}\sum_{y\in\mathcal{N}_\mathbf{x}} \nu_f(\mathbf{x},\mathbf{y})\left(f(\mathbf{x}) - f\left(\mathbf{y}\right)\right) = 0, \tag{32}$$

where ν_{f_n} is defined as in (20). Thus, the solution to (30) can be obtained using the following fixed point iterations [21]

$$f_{n,m+1}(\mathbf{x}) = \frac{z_n(\mathbf{x}) + \frac{\lambda}{2\tau}\sum_{\mathbf{y}\in\mathcal{N}_\mathbf{x}} \nu_{f_{n,m}}(\mathbf{x},\mathbf{y}) f_{n,m}\left(\mathbf{y}\right)}{1 + \frac{\lambda}{2\tau}\sum_{\mathbf{y}\in\mathcal{N}_\mathbf{x}} \nu_{f_{n,m}}(\mathbf{x},\mathbf{y})}. \tag{33}$$

Here, m is the index for the inner loop. Thus, the each step of the algorithm involves one steepest descent step (29), followed by nonlocal denoising using several fixed point iterations. Note that each iteration of the fixed point algorithm requires the re-estimation of the weights $\nu_{f_{n,m}}$, which is computationally expensive.

C. Majorization of both the data and penalty terms (DPM)

Following the data majorization in (30), we now additonally majorize the penalty term to obtain

$$\mathcal{D}_n(f) = \|f - z_n\|^2 + \frac{\lambda}{\tau}\sum_{\mathbf{x}\in\Omega}\sum_{y\in\mathcal{N}_\mathbf{x}} \nu_{f_n}(\mathbf{x},\mathbf{y})\|f(\mathbf{x}) - f(\mathbf{y})\|^2 \tag{34}$$

Solving this criterion as before, we get

$$f_{n+1}(\mathbf{x}) = \frac{z_n(\mathbf{x}) + \frac{\lambda}{2\tau}\sum_{\mathbf{y}\in\mathcal{N}_\mathbf{x}} \nu_{f_n}(\mathbf{x},\mathbf{y}) f\left(\mathbf{y}\right)}{1 + \frac{\lambda}{2\tau}\sum_{\mathbf{y}\in\mathcal{N}_\mathbf{x}} \nu_{f_n}(\mathbf{x},\mathbf{y})}. \tag{35}$$

This equation is similar to (33), except that only one fixed point update is required at each iteration. Thus, this algorithm alternates between one steepest-descent step and one fixed point step. As in the DM case, the weights have to be re-estimated for each steepest descent step. This approach is similar to the algorithms used in [17] and [16].

D. Summary of the MM algorithms

Each iteration of the above four algorithms involves the following steps:

- **PM-CG**: One weight update, followed by iterative CG optimization, until convergence; this algorithm is introduced in this paper.
- **PM-SD**: One weight update, followed by several steepest descent updates; this is the approach followed in [13] to solve conventional NL regularization problems.
- **DM**: One steepest descent update, followed by several fixed point steps (each involving one weight evaluation). The fixed point iteration to solve denoising problems is introduced in [21].
- **PDM** One steepest descent update, followed by one fixed point step (each fixed point step involving one weight evaluation); this is the approach followed in [17].

The computation of the weights $w_f(\mathbf{x},\mathbf{y})$ involves the evaluation of the distances between all patch pairs (see (16) and hence is computationally expensive. In contrast, each of the CG steps involves three FFTs (in Fourier inversion and deblurring applications) and a weighted linear combination of pixel values (evaluation of $\mathbf{\Gamma}_n\mathbf{f}$) per iteration; these steps are relatively inexpensive compared to the evaluation of the weights. Since the **PM-CG** scheme relies on one weight computation, followed by several CG updates, we expect this algorithm to be more efficient than other methods. Note that the **MD** and **MPD** schemes require at least one weight computation per one steepest descent step.

E. Expression of the weights for specific distance metrics

All of the above algorithms require the repeated evaluation of $\nu_{f_n}(\mathbf{x},\mathbf{y})$, which are computed from the similarity measures $w_n(\mathbf{x},\mathbf{y}) = \psi(\|P_\mathbf{x}(f) - P_\mathbf{x}(f)\|_\eta)$ using (20). We focus on the specific distance function (ϕ) considered earlier and derive the expression for the corresponding ψ functions.

1) H_1 distance metric: Applying (16) to the distance metric ϕ, specified by (21), we obtain

$$\psi(x) = \frac{\phi'(x)}{x} = \frac{1}{\sigma^2}\exp\left(-x^2/\sigma^2\right).$$

This Gaussian weight function is widely used in nonlocal H_1 algorithms.

2) Peyre's distance metric: When $\phi(x)$ is specified by (23), we obtain

$$\psi(x) = \frac{\exp\left(-x/\sigma\right)}{2\sigma x}.$$

3) Nonlocal TV distance metric: In this case, we obtain the the weights as

$$\psi(x) = \frac{\exp\left(-x^2/\sigma^2\right)}{\sqrt{\pi}\sigma\; x}.$$

The iterative reweighted quadratic minimization scheme is considerably simpler than the current non-linear minimization schemes used for nonlocal TV [11].

4) ℓ_1 distance metric: In the ℓ_1 case, we obtain

$$\psi(x) = \frac{1}{2\sigma x}.$$

We plot the four ψ functions in Fig. 2(b), along with $\psi(x) = 1$, which corresponds to $\phi(x) = x^2$. Note that this choice results in the averaging of all the patches in the specified neighborhood, irrespective of the similarity. In contrast, the weights corresponding to other NL penalties decay with Euclidean inter-patch distances, ensuring that dissimilar patches are not averaged. The smoothing properties are determined by the rate of decay of ψ. For example, the slow decay of the the weight function corresponding to the convex ℓ_1 metric results in blurred reconstructions. In contrast, the weights corresponding to the non-convex distances decay rapidly, resulting in less blurred reconstructions. This explains the desirable properties of the non-convex distance measures, which are also confirmed by our experimental results.

F. Continuation to improve convergence to global minimum

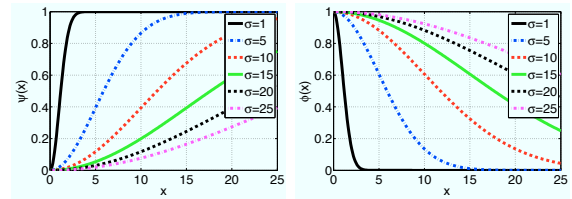
Since non-convex distance functions have narrow valleys, they discourage the averaging of dissimilar patches, while promoting the averaging of similar patches, resulting in improved reconstructions. However, the main challenge associated with non-convex distance metrics is the convergence of the alternating algorithm to the local minima of (12). To improve the convergence to the global minimum, we propose to use a continuation strategy.

Note from Fig.8 that the non-convex distance functions closely approximate quadratic or convex ℓ_1 functions when $x/\sigma << 1$. The distance functions saturate only when $x/\sigma > 1$. We use this property to realize the continuation scheme. Specifically, we start with large values of σ and reduce it to the desired value in several steps. At each step, we use the image iterate corresponding to the σ value at the previous iteration as the initialization. This approach is similar to the homotopy continuation scheme used in non-convex compressive sensing [28], which were reported to be very effective. We initialize the algorithm with $\sigma_{\text{final}} + \sigma_{\text{incfactor}} \times \text{MaxOuter}$. We decrement σ by $\sigma_{\text{incfactor}}$ at each outer iteration. Thus, in the final iteration, we algorithm uses $\sigma = \sigma_{\text{final}}$. In this work, we choose $\sigma_{\text{incfactor}} = 5$. The pseudo-code of the proposed PM-CG scheme with continuation is shown below.

Algorithm IV.1: PM-CG(A, b, λ, σ)

$i \leftarrow 1; j \leftarrow 0$
$\sigma \leftarrow \sigma_{.\text{final}} + \sigma_{\text{factor}} \times \text{MaxOuter}$
$f^{(0,j+1)} \leftarrow A^*(b)$
while i $<$ MaxOuter
do
- $f^{(i,1)}(\mathbf{r}) \leftarrow f^{(i-1,j+1)}(\mathbf{r})$;
- $j \leftarrow 1$;
- **while** (j $<$ MaxInner) && Error not saturated
  - **do**
    - Update similarity measures using (16)
    - Compute weights using (20)
    - CG : determine $f^{(i,j+1)}(\mathbf{r})$ using (26)
- $\sigma \leftarrow \sigma - \sigma_{\text{factor}}$
- $i \leftarrow i + 1$

return (f)



(a) continuation: distance metrics (b) continuation: weight fuctions

Fig. 3. Illustration of the continuation scheme used to minimize the convergence of the algorithm to local minimum. We start with $\sigma = 25$, when the distance metric is convex for most patch pairs in the image. We then gradually reduce the value of σ by a step size of 5 until it achieves the desired value $\sigma = 1$. Note that the distance metrics with small values of σ are non-convex.

V. Experimental results

We introduced a unifying non-local regularization framework, where the regularization penalty is the unweighted sum of robust distances between image patch pairs. In addition to providing a justification for the heuristic approach of iterating current NL algorithms, the novel framework enabled us to understand and mitigate the alias amplification issues that are widely reported in the context of Fourier inversion. In this section, we will determine the utility of the continuation scheme, introduced in Section IV-F, to overcome the local minima problems. The similarity of the proposed framework with regularization schemes in compressive sensing enabled us to develop computationally efficient algorithms. We compare the computational efficiency of the proposed algorithms in this section. Thanks to the unifying framework, these fast methods are applicable to a wide range of NL penalties. We also study the impact of the various parameters and distance metrics on image quality, and compare the non-local schemes with local TV algorithms.

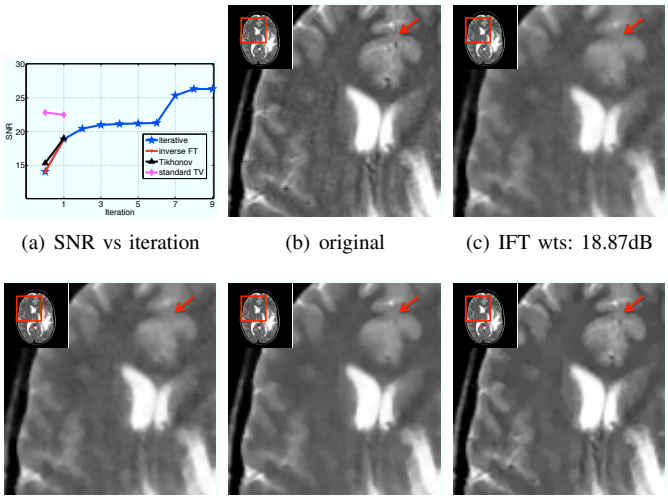
A. Utility of continuation

We aim to recover a 128x128 Shepp Logan phantom from ten lines and a 256x256 brain image from its five fold under sampled measurements in the Fourier domain. We used the nonlocal H1 and TV distance functions. We assume that the measurements are corrupted by complex Gaussian noise, such that the SNR of the noise k-space data is 50dB. The original images are shown in Fig. 4.(a) and Fig. 4.(f), while the initial guesses obtained by computing zero filled inverse Fourier transform are shown in Fig. 4.(b) and Fig. 4.(g), respectively. The heuristic alternating algorithm results in reconstructions with amplified alias artifacts (see (c) and (h)). We observe that the continuation strategy, illustrated in Fig. 3.(a) & (b), is capable of removing the alias artifacts completely. This convergence can also be appreciated by the plot of the cost in Fig. 4.(e). It is seen that the cost function saturate to high values (red dotted and solid lines) because the algorithms get stuck in local minima of (12), when no continuation is used. In contrast, they converge to smaller costs with the continuation scheme. We confirmed that this is indeed the global minimum of the criterion by using multiple initializations, including the original image. The improved

performance can also be appreciated by the SNR plots shown in Fig. 4.(j). This experiment shows that non-convex metrics can provide improved reconstructions in challenging inverse problems, but continuation schemes are essential to minimize the convergence to local minima.

B. Utility of iterating between weight and image estimation

Since alternating schemes result in alias amplification in the context of Fourier inversion, current NL schemes use fixed weights with such challenging problems. The weights are often estimated from the noisy/smoothed image [13], [15], [29] and Tikhonov or total variation regularized reconstructions in deblurring applications [13]. We demonstrate the improvement offered by the proposed scheme with continuation, compared to these fixed weight methods in Fig.5. We consider the recovery of a brain MRI image from five fold randomly under-sampled Fourier data. We consider three different fixed weights, which are estimated from different initial estimates: (a) zero padded inverse Fourier transform $\mathcal{A}^*$ of the measurements, (b) Tikhonov regularized reconstruction, and (c) local TV regularized reconstruction. We initialize the PM-CG scheme with weights obtained from zero-padded IFFT reconstruction. We observe that the the iterative strategy provides a 4 dB improvement over the best fixed weight scheme. These experiments demonstrate that the quality of the reconstructions can be significantly improved by the iterative framework with appropriate continuation strategies to ensure convergence to global minima.



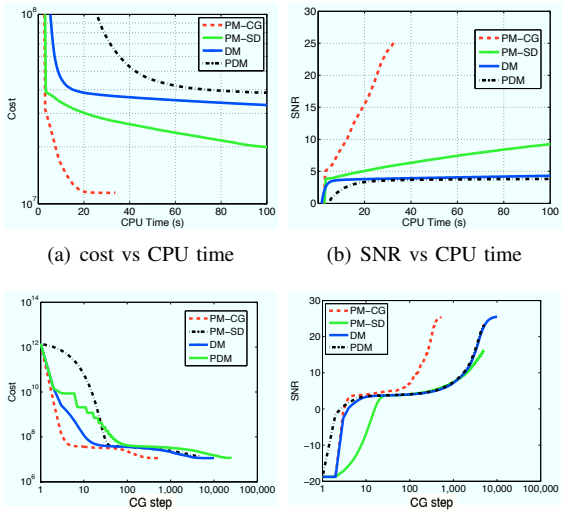
(a) SNR vs iteration (b) original (c) IFT wts: 18.87dB

(d) Tikh wts: 19.00dB (e) TV wts: 22.48dB (f) Iterative: 26.32dB

Fig. 5. Comparison of current algorithms with pre-computed weights and the proposed iterative framework: We consider the reconstruction of the MRI brain image from 20% of its random Fourier samples using the H_1 distance metric. In Fig. 5(a), we plot the improvement in signal to noise ratio as a function of the number of iterations. The zeroth iteration correspond to the different initial guesses. These guesses are derived with three different algorithms: zero filled inverse FFT (red), Tikhonov reconstruction (black), and local TV (majenta). Since the current NL schemes do not re-estimate the weights, only one iteration is involved. The blue curve corresponds to the proposed iterative scheme, which re-estimates the weights from the current image estimate. Since this scheme is also initialized with the zero filled IFFT, the red and the blue curves overlap. Note that the proposed iterative method provides significantly improved reconstructions.

C. Comparison of optimization algorithms

We compare the convergence rate and computation complexity of the four MM algorithms, described in Section IV, in Fig. 6. Here, we consider the reconstruction of a 128 $\times$128 Shepp-Logan phantom from 16 radial lines in the Fourier domain/k-space. We choose ϕ as the H_1 distance metric, specified by (21). We set $\lambda = 50$ and $\sigma = 10$, since this choice of parameters gives the best reconstruction. Note that the PM-CG scheme requires around 20 times fewer CG steps to converge, compared with other algorithms. In addition, the computational complexity of each iteration of the PM scheme is lower than DM and DPM methods, since the computationally expensive weight computations are performed only in the outer loop. In this specific example, the PM-CG scheme requires around 33 seconds to converge, while the DM and PDM algorithms require 5000 seconds or more. The experiments are performed on a laptop with 2.4 GHz Duo Core CPU processor and 4 GB memory in MATLAB R2010a. Since the PM-CG scheme is significantly more computationally efficient than other methods, we will use this method for all the other experiments considered in this paper. Thanks to the unification offered by the proposed scheme, we can now use this algorithm for all the distance metrics.



(a) cost vs CPU time (b) SNR vs CPU time

(c) cost vs CG steps (d) SNR vs CG steps

Fig. 6. Comparison of optimization algorithms: We focus on the recovery of a 128 $\times$ 128 Shepp-Logan phantom from 16 radial lines in the Fourier domain by minimizing (12). The decay of the cost function (12) as a function of the number of conjugate gradient steps (inner iterations) and CPU time are shown in (a) and (b), respectively. The improvement in SNR as a function of iterations and CPU time are shown in (c) and (d), respectively. Note that all the optimization algorithms converge to the same minima. However, the **PM-CG** scheme requires much fewer (20 fold) number of iterations. Moreover, the computational complexity per iteration of this method is also lower since the expensive weight computation is only performed in the outer loop.

D. Choice of parameters

There are various parameters in the proposed nonlocal algorithms, such as regularization parameter λ, size of the search window $\mathcal{N}_{\mathbf{x}}$, and patch size $P_{\mathbf{x}}$. We now study the effect of these parameters on the quality of the reconstructions. The parameter λ controls the trade off between data fidelity and regularization. To ensure fair comparisons, we search the

optimal λ for each algorithm, measurement, and noise level. We pick the value of λ that provides the best reconstruction in each case. To keep the computational complexity manageable, most NL schemes do not search the entire image for similar patches.

We study the effect of patch size N_p and neighborhood size N_w in the context of image denoising[2] in Fig.7. We add Gaussian noise to the Shepp-Logan phantom such that the SNR of the noisy image is 16dB. We then use nonlocal H1 and TV penalties to recover the image. The results are shown in Fig. 7. We observe that the performance of the algorithms degrade with increasing patch size. This is expected since increasing the size of the patches will result in decreased number of patches similar to it. The more interesting observation is that increasing the size of the neighborhood results in decreased SNR with only one exception. Specifically, the SNR improves slightly while the neighborhood size is increased from $N_w = 5$ to $N_w = 7$ in NL TV with patch size $N_p = 3$. These findings can be explained using the unified formulation. Note that the metric in (13) involves the unweighted sum of distances between all patch pairs in the neighborhood. As we increase N_w, we compare more dissimilar patches. If the distance function $\phi(x)$ does not saturate to one with large values of x, the functional in (13) will continue to penalize the distances between dissimilar patches and hence result in blurring. Note that the TV penalty saturates much faster than the H_1 penalty, which explains the improvement in performance when N_w is increased from 5 to 7. The comparisons in Fig.7 show that choosing $N_p \times N_p = 3 \times 3$ and $N_w \times N_w = 5 \times 5$ are sufficient to obtain good results. Even though the results may change slightly depending on the image and the type of application, we will use these parameters in the rest of the paper.

The reason for obtaining good reconstructions with smaller search window sizes is due to the ability of iterative NL schemes to exploit the similarities between patch pairs, even when they are not in each others neighborhood. It is sufficient that they are linked by at least one similar patch that is in both the neighborhoods. i.e, if $P_x(f) \approx P_y(f) \approx P_z(f)$ and $y \in \mathcal{N}_x$ and $y \in \mathcal{N}_z$, then the algorithm can exploit the similarity between $P_x(f)$ and $P_z(f)$, even when $z \notin \mathcal{N}_x$. This is enabled by the terms $\phi(\|P_x(f) - P_y(f)\|)$ and $\phi(\|P_y(f) - P_z(f)\|)$ in the proposed NL penalty. Since conventional NL means smoothing filters are non-iterative, they are not capable of exploiting the similarity between patches that are not in each others neighborhoods.

E. Image reconstruction with different distance metrics

We compare the proposed NL scheme with different non-local distance metrics (see Table I), in the context of image recovery from sparse Fourier samples. We use popular test images (e.g., Lena and cameraman) as well as multiple typical MRI images (e.g., brain and cardiac MRI images). We also compare the non-local schemes against local TV and wavelet algorithms to bench mark the performance. The comparisons of the methods on a MRI brain image is shown in Fig. 8. Here

[2]We focus on denoising to keep the computational complexity manageable, while dealing with large number of comparisons.

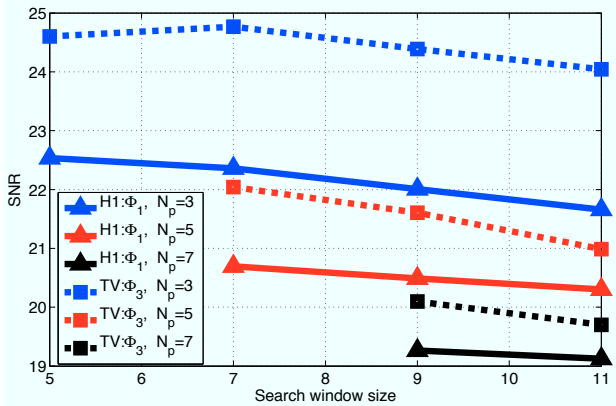


Fig. 7. Impact of the parameters on reconstruction accuracy: We consider the denoising of noisy Shepp-Logan phantom image with a SNR of 16 dB. We plot the change in the SNR of the denoised image as a function of the search window size N_w. The blue, red, and black curves correspond to $N_p = 3$, $N_p = 5$, and $N_p = 7$, respectively. The plot shows that SNR of the images degrades with increasing N_p. This is expected since the number of similar patches decrease with increasing patch size. We observe that the SNR also decreases with increasing window size, except for the non-local TV distance metric with $N_w = 7$. See text for more explanation.

we consider a downsampling factor of 4 and the downsampling pattern is random with non-uniform k-space density. We add complex Gaussian white noise to the measured data such that the SNR of measurements is 40dB. Continuation schemes are used with non-convex non-local methods to minimize the local minima effects. We observe that the non-convex algorithms provide the best reconstructions. In contrast, the proposed non-local scheme with the L_1 metric results in reconstructions with significant blurring, while the local TV scheme provides patchy reconstructions. The nonlocal TV metric gives the highest SNR, followed by nonlocal H1. A similar comparison is performed in Fig. 9, where we recover another MRI brain image with a tumor from its five fold under-sampled Fourier data set. We use the polar trajectory with 52 radial lines to sample. Gaussian noise is added to the k-space data such that the noise level of the measurements is 80dB. The zoomed versions of different regions of the image is shown in Fig. 9. We observe that proposed NL method using the non-convex metrics (TV, H1, and Peyre's schemes) provides better results than the local TV scheme, and non-local scheme using the L1 metric results.

The signal to noise ratios of the reconstructions of different images using the different algorithms are shown in Table II. We consider an acceleration of 5 and the Fourier space is sampled randomly using a non-uniform k-space density. We observe that the proposed nonlocal algorithm with non-convex penalties consistently outperform the local TV algorithm, resulting in an improvement of more than 2-3 dB. In contrast, the SNR of the non-local scheme with convex ℓ_1 distance metric is only comparable to the local TV scheme.

VI. Conclusion

We introduced a unifying energy minimization framework for the nonlocal regularization of inverse problems. The proposed functional is the unweighted sum of inter-patch distances. We showed that the first iteration of a MM algorithm to

Image	local TV	nonlocal L1	H1 metrix	NLTV metric	Peyre's metric
lena	20.46	20.97	23.37	23.93	**24.05**
cameraman	23.31	23.04	24.84	25.88	**25.99**
brain1	23.24	22.48	26.12	**27.51**	27.45
brain2	22.24	22.12	27.15	**28.06**	27.99
ankle	21.41	22.23	23.43	24.56	**24.70**
abdomen	16.45	16.90	18.51	19.46	**19.70**
heart	20.32	21.56	22.26	24.47	**24.90**

TABLE II
SIGNAL TO NOISE RATIO OF THE RECONSTRUCTED IMAGES USING DIFFERENT ALGORITHMS.

minimize the proposed cost function is similar to the classical non-local means algorithms. Thus, the reformulation provided a theoretical justification for the heuristic approach of iterating non-local schemes, which re-estimate the weights from the current image estimate. Thanks to the reformulation, we now understand that the widely reported alias enhancement issues with iterative non-local methods are caused by the convergence of these algorithms to the local minimum of the proposed non-convex penalty. We introduce an efficient continuation strategy to overcome this problem. We introduced several fast majorize-minimize algorithms. Thanks to the unifying framework, all of the novel fast algorithms are readily applicable for a large class of non-local formulations.

ACKNOWLEDGEMENT

We thank the anonymous reviewers for the valuable suggestions that significantly improved the quality of the manuscript.

REFERENCES

[1] J. Ma and L. Dimet, "Deblurring from highly incomplete measurements for remote sensing," *IEEE Trans. Geoscience and Remote Sensing*, vol. 47, no. 3, pp. 792–802, 2009.

[2] C Vogel and M Oman, "Fast, robust total variation-based reconstruction of noisy, blurredimages," *IEEE Transactions on Image Processing*, Jan 1998.

[3] S. Osher, A. Solé, and L. Vese, "Image decomposition and restoration using total variation minimization and the $h-1$ norm," *Multiscale Modeling and Simulation*, vol. 1, pp. 349–370, 2003.

[4] A. Chambolle and P.L. Lions, "Image recovery via total variation minimization and related problems," *Numerische Mathematik*, vol. 76, no. 2, pp. 167–188, 1997.

[5] M Lustig, D Donoho, and J Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *MRM*, vol. 58, no. 6, pp. 1182–95, Dec 2007.

[6] C Vonesch and M Unser, "A fast multilevel algorithm for wavelet-regularized image restoration," *IEEE Transactions on Image Processing*, vol. 18, no. 3, pp. 509–523, 2009.

[7] M. Guerquin-Kern, D. Van De Ville, C. Vonesch, J.C. Baritaux, KP Pruessmann, and M. Unser, "Wavelet-regularized reconstruction for rapid MRI," in *ISBI 2007*. IEEE, 2009, pp. 193–196.

[8] A. Buades, B. Coll, and J.M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 490–530, 2006.

[9] S.P Awate and R.T. ; Whitaker, "Unsupervised, information-theoretic, adaptive image filtering for image restoration," *IEEE Trans. Pattern Recognition*, vol. 28, pp. 364, 2006.

[10] A. Buades, B. Coll, and J.M. Morel, "Nonlocal image and movie denoising," *International Journal of Computer Vision*, vol. 76, no. 2, pp. 123–139, 2008.

[11] L.D. Cohen, S. Bougleux, and G. Peyré, "Non-local regularization of inverse problems," in *European Conference on Computer Vision (ECCV'08)*, 2008.

[12] G. Gilboa, J. Darbon, S. Osher, and T. Chan, "Nonlocal convex functionals for image regularization," *UCLA CAM Report*, pp. 06–57, 2006.

[13] Y. Lou, X. Zhang, S. Osher, and A. Bertozzi, "Image recovery via nonlocal operators," *Journal of Scientific Computing*, vol. 42, no. 2, pp. 185–197, 2010.

[14] X. Zhang, M. Burger, X. Bresson, and S. Osher, "Bregmanized nonlocal regularization for deconvolution and sparse reconstruction," *SIAM Journal on Imaging Sciences*, vol. 3, pp. 253, 2010.

[15] G. Peyré, S. Bougleux, and L.D. Cohen, "Non-local regularization of inverse problems," *Inverse Problems and Imaging*, pp. 511–530, 2011.

[16] G. Adluru, T. Tasdizen, M.C. Schabel, and E.V.R. DiBella, "Reconstruction of 3D dynamic contrast-enhanced magnetic resonance imaging using nonlocal means," *Journal of Magnetic Resonance Imaging*, vol. 32, no. 5, pp. 1217–1227, 2010.

[17] G Wang and J Qi, "Patch-based regularization for iterative pet image reconstruction," in *IEEE ISBI*, 2011.

[18] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *Signal Processing Letters, IEEE*, vol. 14, no. 10, pp. 707–710, 2007.

[19] M Figueiredo, J Dias, and R Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2980–2991, 2007.

[20] J. Yang and Y. Zhang, "Alternating direction algorithms for ell_1 problems in compressive sensing," *Arxiv preprint arXiv:0912.1185*, 2009.

[21] L. Pizarro, P. Mrázek, S Didas, S. Grewenig, and J. Weickert, "Generalized nonlocal image smoothing," *International Journal of Computer Vision*, vol. 90, pp. 62–87, 2010.

[22] Z Yang and M Jacob, "A unified energy minimization framework for nonlocal regularization," in *IEEE ISBI*, 2011.

[23] M. Figueiredo, J. Bioucas-Dias, J. Oliveira, and R. Nowak, "On total variation denoising: A new majorization-minimization algorithm and an experimental comparison with wavalet denoising," in *IEEE International Conference on Image Processing*, 2006, p. 2633.

[24] M Figueiredo, J Dias, J Oliveira, and R Nowak, "On total variation denoising: A new majorization-minimization algorithm and an experimental comparisonwith wavalet denoising," *IEEE International Conference on Image Processing, 2006*, pp. 2633–2636, 2006.

[25] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud, "Deterministic edge-preserving regularization in computed imaging," *Image Processing, IEEE Transactions on*, vol. 6, no. 2, pp. 298–311, 1997.

[26] R. Chartrand and Wotao Yin, "Iteratively reweighted algorithms for compressive sensing," in *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on*, 31 2008-april 4 2008, pp. 3869 –3872.

[27] P. L. Combettes and V.R. Wajs, "Signal recovery by proximal forward-backward splitting," *Multiscale Modeling*, vol. 4, no. 4, pp. 1168–1200.

[28] J. Trzasko and A. Manduca, "Highly undersampled magnetic resonance image reconstruction via homotopic ℓ_0-minimization," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 1, pp. 106–121, 2009.

[29] G. Gilboa and S. Osher, "Nonlocal operators with applications to Image Processing," *Multiscale Model. Simul*, vol. 7, no. 3, pp. 1005–1028, 2008.

[30] T. Brox, O. Kleinschmidt, and D. Cremers, "Efficient nonlocal means for denoising of textural patterns," *IEEE Transactions on Image Processing*, vol. 17, no. 7, pp. 1083 –1092, july 2008.

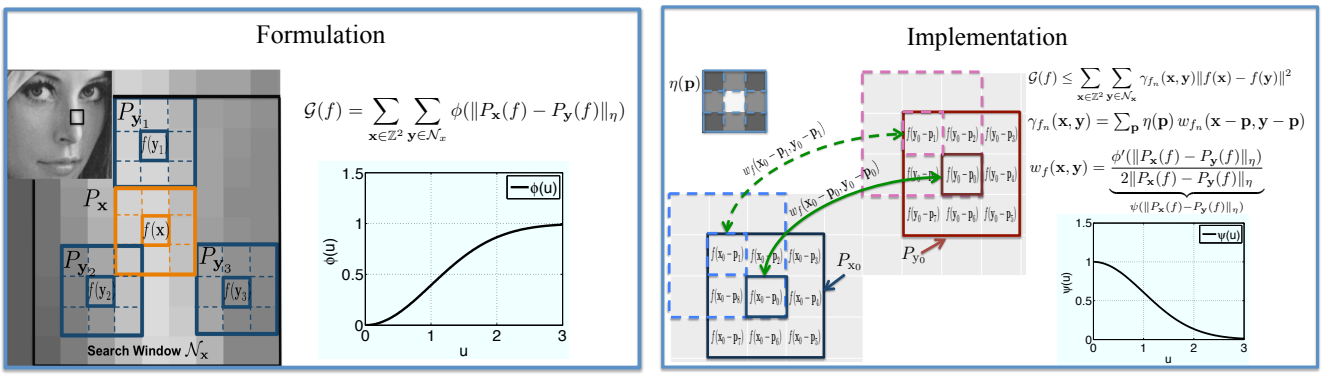
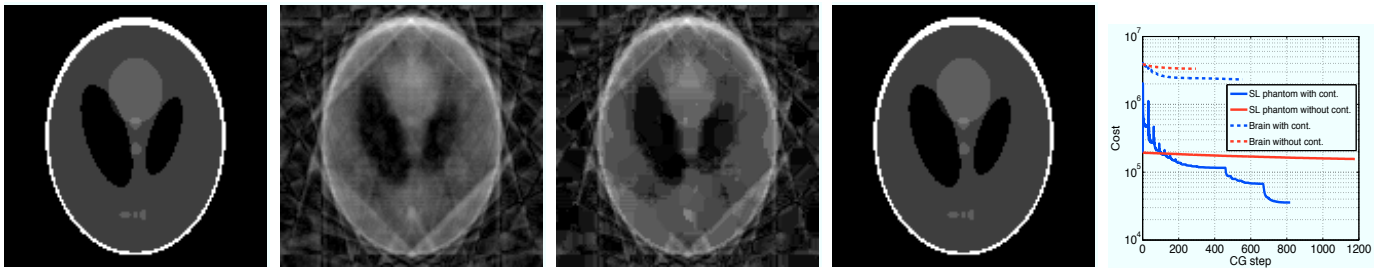
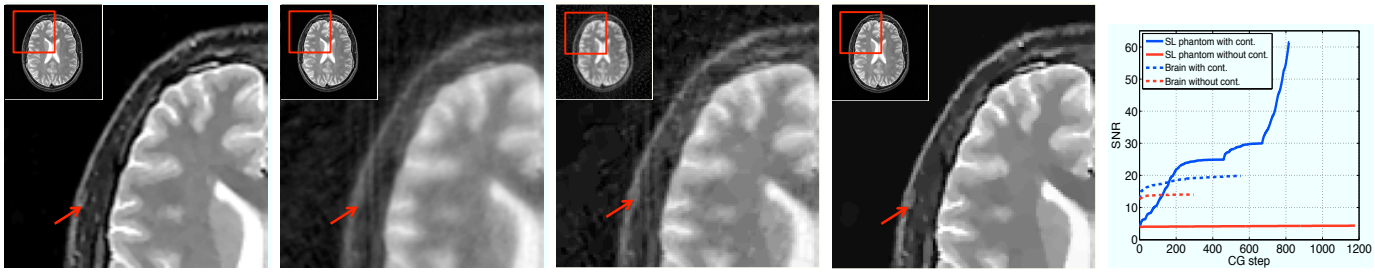


Fig. 1. Illustration of the proposed nonlocal framework: We illustrate the proposed regularization functional, specified by (13), in the left box. The regularization penalty is the sum of distances between patch pairs in the image.For each pixel $\mathbf{x}$, we consider the distances between the patch centered at $\mathbf{x}$ (specified by $P_{\mathbf{x}}(f)$) and patches centered on the neighboring pixels $y \in \mathcal{N}_x$; $\mathcal{N}_x$ is a square shaped window in which the algorithm searches for similar patches.We use the robust distance metric ϕ, which saturates with the the inter-patch distance. This property make the regularization penalty insensitive to large inter-patch distances, thus minimizing the averaging between dissimilar patches. The surrogate penalty, obtained by the majorization of $\mathcal{G}(f)$ in illustrated the right box. The surrogate penalty is essentially a weighted sum of Euclidean distances between pixel intensities. This criterion is very similar to the classical H_1 non-local penalty. The weights $\gamma_{f_n}(\mathbf{x},\mathbf{y})$ is obtained as the sum of the similarity measures as in (20). The similarity measures between patches are computed as a monotonically decreasing non-linear function (ψ) of the inter patch distances. This property ensures that dissimilar patch pairs result in low inter pixel weights, thus encouraging the averaging of similar pixels.

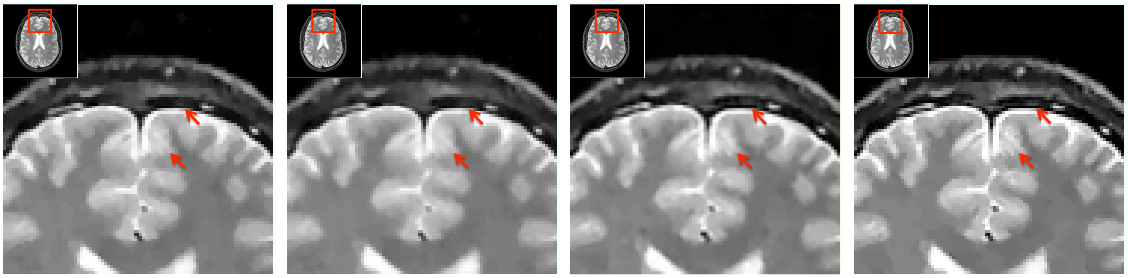


(a) original image (b) initial guess, SNR=3.99 (c) NL-H1,no cont.;SNR=4.25 (d) NL-H1,cont.: SNR=61.56 (e) Cost vs CG step

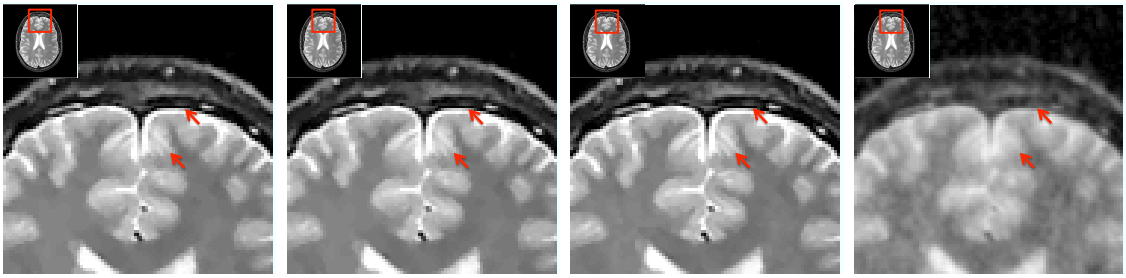


(f) original image (g) initial guess, SNR=12.71 (h) NL-TV, no cont.;SNR=14.03 (i) NL-TV, cont.: SNR=19.84 (j) SNR vs CG step

Fig. 4. Utility of continuation: We consider the recovery of a 128x128 Shepp Logan phantom from its ten radial lines in the Fourier domain in (a)-(d). The classical alternating algorithm emphasizes the alias artifacts in the initial guess in (b), which is obtained by zero-filled IFFT. Specifically, the result in (c) correspond to a local minimum of (12). The use of the continuation scheme mitigated these issues as seen in (d). We also consider the reconstruction of a 256 $\times$ 256 MRI brain image from 46 uniformly spaced radial lines in the Fourier domain. We observe that the alternating algorithm results in the enhancement of the alias artifacts (see (h)). These problems are eliminated by the continuation scheme, shown in (i). The behavior of the algorithms can be better understood from the cost vs iterations shown in (e). It is observed that the continuation schemes shown in blue solid and dotted lines (corresponding to Shepp-Logan and brain images, respectively) converge to a lower minimum than the methods without continuation (red solid and dotted lines). The improvement in performance is also SNR vs CG step plot in (j).

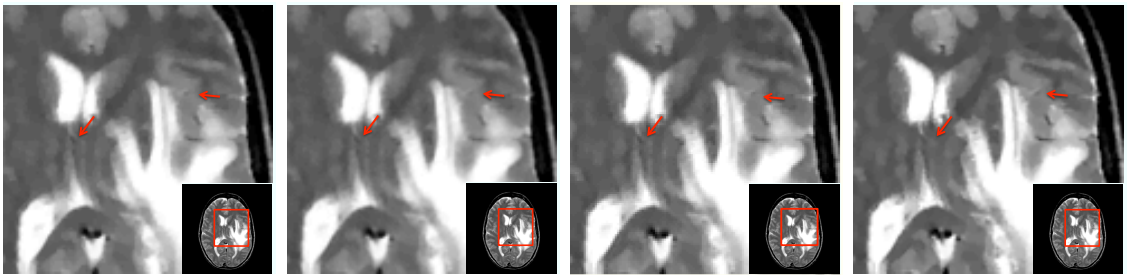


(a) Local TV, SNR=23.87 dB (b) NL-L1 metric,SNR=22.49 dB (c) NL-Lp metric, p=0.3, SNR=25.24 dB (d) NL-H1 metric, SNR=27.93 dB

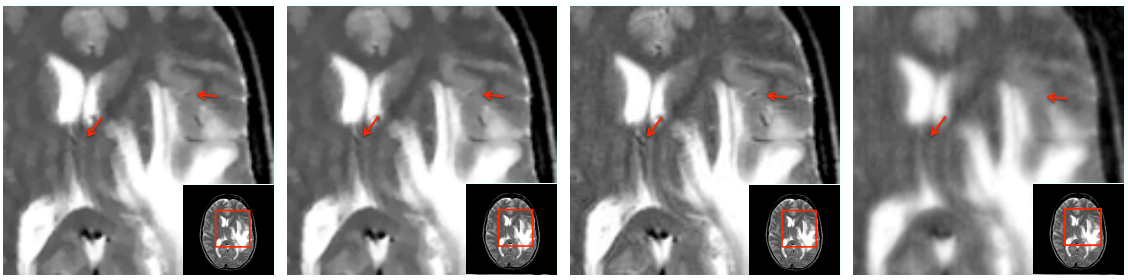


(e) NL-TV metric, SNR=28.09 dB (f) NL-Peyre's scheme, SNR=28.11 dB (g) Original image (h) Initial guess, SNR=12.03

Fig. 8. Comparison of the proposed algorithm with different metrics against classical schemes: We reconstruct the 256 $\times$ 256 MRI brain image from its sparse Fourier samples. We consider an under sampling factor of 4. We choose the standard deviation of the complex noise that is added to the measurements such that the SNR of the measurements is 40dB. We observe that the proposed schemes that use non-convex distance metrics provide the best SNR, which is around 4.22 dB better than local TV. The use of convex distance metrics can only provide reconstructions than are comparable to local TV. The $\ell_p; p = 0.3$ penalty provides improved results than the $p = 1$ case. However, the performance improvement is not as significant as the other non-local metrics, probably due to the reduced averaging of similar patches. The arrows indicate the details preserved by the non-convex schemes, but missed by local TV and convex non-local algorithm.



(a) Local TV, SNR=23.56 (b) NL-L1 metric, SNR=23.29 (c) NL-Lp metric, p=0.3, SNR=24.30 (d) NL-H1 metric, SNR=25.18



(e) NL-TV metric, SNR=26.44 (f) Peyre's scheme, SNR=26.42 (g) Original image (h) Initial guess, SNR=15.18

Fig. 9. Comparison of the proposed algorithm with different metrics against classical schemes: We reconstruct the 256 $\times$ 256 MRI brain image from its sparse Fourier samples. We consider an under sampling factor of 5. We choose the standard deviation of the gaussian noise that is added to the measurements such that the SNR of the measurements is 40dB. We observe that the proposed schemes that use non-convex distance metrics provide the best SNR, which is around 3 dB better than local TV. The arrows indicate the details preserved by the non-convex schemes, but missed by local TV and convex non-local algorithm.