

# A UNIFIED ENERGY MINIMIZATION FRAMEWORK FOR NONLOCAL REGULARIZATION

*Zhili Yang and Mathews Jacob*

Dept. of Biomedical Engineering, University of Rochester

## ABSTRACT

We introduce a unifying framework for the non-local regularization of biomedical inverse problems. We choose the regularization functional as the sum of distances between pairs of patches in the image. We introduce a novel majorize minimize algorithm to minimize the proposed criterion. We observe that the first iteration of the algorithm to be very similar to the classical non-local regularization schemes. In addition to providing a novel interpretation for heuristic iterative non-local regularization schemes, the proposed scheme enables us to develop efficient optimization algorithms, design novel non-local schemes by choosing the distance metric, and minimize local minima problems. We demonstrate the benefits of the unified framework in deblurring and compressed sensing.

**Index Terms**— Non-local means, compressed sensing, inverse problems, conjugate gradient

## 1. INTRODUCTION

Non local means (NLM) is a recent denoising method that has received a lot of attention in image processing [1]. This scheme exploits the similarity between rectangular patches in an image to suppress noise. Specifically, each pixel in the denoised image is recovered as a weighted linear combination of all the pixels in the noisy image. The weights are estimated from the noisy image as nonlinear functions of the similarity of the pixel neighborhood with other neighborhoods.

Recently, several authors have shown that the non-local filtering scheme is the solution of a variational problem, where a criterion involving the weighted sum of differences between the pixels is minimized [2, 3]; the weights are often estimated from an initial guess of the solution. This formulation enables the extension of the framework to inverse problems such as deblurring, where the weights are determined from the blurred image. The direct application of this framework to inverse problems such as compressed sensing is challenging since the initial guess often suffers from significant aliasing; the weights estimated from the initial guess often results in poor reconstructions. Recently some authors have proposed to iterate the above variational scheme, where the weights estimated from the recovered image is used to

update the criterion [4]. The main challenge of such heuristic approaches is the changing nature of the cost function with the iteration. This makes it difficult to characterize the solution based on the regularization functional, design efficient optimization algorithms, and avoid local minima problems.

We introduce a novel unified framework for the non-local regularization of inverse problems. In contrast to current non-local regularization schemes that penalize the weighted sum of pixel intensity differences [2, 3], the proposed regularization functional is an unweighted linear combination of distances between image patches. The main difference with the classical framework is that the functional it is not dependent on any specific weight function. Since the criterion is only specified by a robust distance metric, the properties of the solution is directly dependent on the metric. We introduce a novel majorize minimize (MM) algorithm to derive the solution of the proposed criterion. Each iteration of the MM algorithm involves the minimization of a quadratic cost function, where the regularization functional is a weighted sum of the  $\ell_2$  norms between image patches. The weights are obtained as nonlinear functions of the distances between the image patches in the previous iteration. We observe that the first iteration of this scheme proposed algorithm is essentially the classical non-local regularization scheme [1]. Thus, the proposed algorithm can be seen as a reinterpretation of heuristic schemes that iterate the non-local regularized recovery [4].

We show that the iterative reweighted non-local  $H^1$  and the non-local TV schemes introduced in [4] are special cases of our general scheme, when the distance metric is chosen appropriately. The unified framework enables us to develop a fast algorithm to significantly improve the convergence. We observe that most of the current non-local schemes correspond to non-convex metrics, which provide significantly improved results over convex cost functions. However, one challenge with these schemes is the convergence to local minima. We introduce homotopy continuation schemes to eliminate the convergence issues associated with these non-convex cost functions.

## 2. BACKGROUND

Several authors have proposed to generalize the non-local means smoothing scheme to general inverse problems by

---

This work is supported by NSF award CCF-0844812.

posing it as an optimization scheme [2, 3]

$$\hat{f} = \arg \min_f \|Af - b\|^2 + \lambda J_w(f) \quad (1)$$

where  $\lambda$  is the regularization parameter and  $J_w(f)$  is the non-local regularization functional:

$$J_w(f) = \int_{\Omega \times \Omega} w(\mathbf{x}, \mathbf{y}) \|f(\mathbf{x}) - f(\mathbf{y})\|^2 d\mathbf{x}d\mathbf{y}. \quad (2)$$

Here,  $w(x, y)$  is a weight function that is specified apriori; it is often estimated from the noisy image in denoising or blurred image in deblurring problems (denoted by  $g(\mathbf{x})$ ) as

$$w(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|N_{\mathbf{x}}(g) - N_{\mathbf{y}}(g)\|^2}{2\sigma^2}\right). \quad (3)$$

Here,  $g(\mathbf{x})$  is the image that is used to estimate the weight function, while  $N_{\mathbf{x}}(g)$  is a square image patch of a specified size centered at  $\mathbf{x}$ . This scheme is termed as non-local  $H_1$  regularization [2]. As discussed previously, the algorithm specified by (1) is iterated, following the redefinition of the weights based on the current solution.

### 3. PROPOSED SCHEME

We pose the non-local regularized reconstruction scheme as

$$\hat{f} = \arg \min_f \|Af - b\|^2 + \lambda \mathcal{G}(f), \quad (4)$$

where, the regularization functional is specified by

$$\mathcal{G}(f) = \int_{\Omega \times \Omega} \varphi(\|N_{\mathbf{x}}(f) - N_{\mathbf{y}}(f)\|) d\mathbf{x}d\mathbf{y} \quad (5)$$

Here,  $\varphi$  is an appropriately chosen distance metric. Note that the proposed function is not dependent on an apriori selected weight function.

We will now show that a majorize-minimize algorithm to solve for (4) is very similar to the common sense approach of iterating (1). If  $\varphi(\sqrt{v})$  is a concave function, it can be majorized by the straight line:

$$\varphi(\sqrt{v}) \leq \frac{\varphi'(\sqrt{v})}{2\sqrt{v}}v + b(v), \quad (6)$$

where  $b(v)$  is a constant. Setting,  $v = \|N_{\mathbf{x}}(f) - N_{\mathbf{y}}(f)\|^2$ , we have

$$\varphi(\|N_{\mathbf{x}} - N_{\mathbf{y}}\|) \leq \frac{\varphi'(\|N_{\mathbf{x}} - N_{\mathbf{y}}\|)}{2\|N_{\mathbf{x}} - N_{\mathbf{y}}\|} \|N_{\mathbf{x}} - N_{\mathbf{y}}\|^2 + b, \quad (7)$$

Using this property, we majorize the regularization term at each iteration as

$$\mathcal{G}(f) \leq \underbrace{\int_{\Omega \times \Omega} w_n(\mathbf{x}, \mathbf{y}) \|N_{\mathbf{x}}(f) - N_{\mathbf{y}}(f)\|^2 d\mathbf{x}d\mathbf{y}}_{\mathcal{G}_{w_n}(f)}, \quad (8)$$

where the weights are specified by

$$w_n(\mathbf{x}, \mathbf{y}) = \frac{\varphi'(\|N_{\mathbf{x}}(f_n) - N_{\mathbf{y}}(f_n)\|)}{2\|N_{\mathbf{x}}(f_n) - N_{\mathbf{y}}(f_n)\|}. \quad (9)$$

Here,  $f_n$  is the  $n^{\text{th}}$  iterate. At each iteration, we solve the minimization problem

$$\hat{f}_{n+1} = \arg \min_f \|Af - b\|^2 + \lambda \mathcal{G}_{w_n}(f) \quad (10)$$

Thus, the proposed scheme is very similar to the common sense approach of iterating (1), with the re-estimation of the weights based on the current iterate. The only difference is that (8) involves the weighted differences between the image patches as opposed to the differences between pixels as in (2). Our numerical results show that both of these methods offer similar qualitative performance. The new framework is attractive since the iterative reweighted quadratic minimization algorithm is guaranteed to converge to the local minimum of (4). This enables us to design novel distance metrics ( $\varphi$  functions) that are more efficient for the recovery of the image. Moreover, this approach also enables the development of novel algorithms to accelerate the convergence of the algorithm.

Note that the subproblem specified by (8) is quadratic; we propose to solve it efficiently using the conjugate gradients algorithm. This approach is considerably faster than the steepest descend approaches with heuristic selection of step sizes used in [3]. We will now show that two of the classical iterative non-local regularization schemes are special cases of the proposed scheme by appropriately choosing  $\varphi$ .

#### 3.1. Special case: $H_1$ non-local regularization

We choose the robust distance metric as

$$\varphi(x) = 2\sigma^2 \left(1 - \exp\left(-\frac{x^2}{2\sigma^2}\right)\right), \quad (11)$$

when  $w(\mathbf{x}, \mathbf{y})$  simplifies to (3). Thus, the majorize maximize algorithm to solve (4) using the above distance metric is very similar to the iterative version of the non-local  $H_1$  regularization scheme [3]. The main difference is that the criterion involves the weighted differences between the image patches as opposed to pixels in the classical methods. We will show in our results that these methods offer the same performance qualitatively and quantitatively.

#### 3.2. Special case: non-local TV regularization

Gilboa et al., proposed to change the  $\ell_2$  norm in (2) to  $\ell_1$  norm to obtain the non-local TV regularization scheme [2]:

$$\hat{f}_{TV} = \arg \min_f \|Af - b\|^2 + \lambda \int_{\Omega \times \Omega} w_n(\mathbf{x}, \mathbf{y}) \|N_{\mathbf{x}} - N_{\mathbf{y}}\| d\mathbf{x}d\mathbf{y}, \quad (12)$$

where  $w_n(\mathbf{x}, \mathbf{y})$  is chosen as (3). Since the criterion in (12) is different from the  $H_1$  scheme, it is realized using specialized

non-linear iterative algorithms in [4]. We show in the Appendix that the iterated version of this scheme is essentially a specific majorize minimize algorithm to solve (4), where

$$\varphi(x) = \sqrt{\frac{\pi\sigma^2}{2}} \operatorname{erf}\left(\frac{x}{\sqrt{2}\sigma}\right) + c, \quad (13)$$

where  $c$  is a constant. The reinterpretation enables us to solve it using exactly the same MM algorithm used in the  $H_1$  setup. Since  $\varphi$  is different from the  $H_1$  scheme, we obtain the the weights as:

$$w_n(\mathbf{x}, \mathbf{y}) = \frac{\exp(-\|N_{\mathbf{x}}(f_n) - N_{\mathbf{y}}(f_n)\|^2/2\sigma^2)}{\|N_{\mathbf{x}}(f_n) - N_{\mathbf{y}}(f_n)\|}. \quad (14)$$

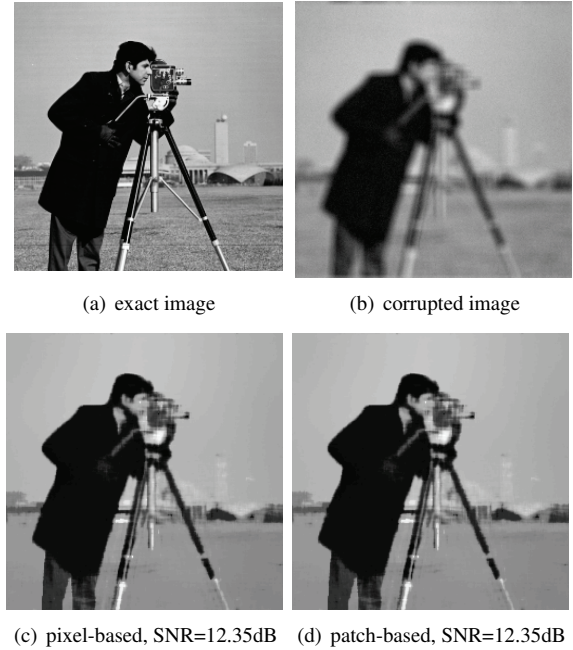
The iterative reweighted quadratic minimization scheme is considerably simpler than the current non-linear minimization schemes used for non-local TV.

We observe that the current distance metrics specified by (13) and (11) are highly non-convex. It is also seen that the TV scheme penalizes dissimilar image patches more heavily than the  $H_1$  scheme. This minimizes blurring introduced by the averaging of dissimilar patches, which explains the superior performance of this scheme (see Fig. 4). However, this approach has a higher probability to be affected by local minima. This is especially a serious concern in compressed sensing applications, where the initial guess is poor. We propose to use homotopy continuation schemes to minimize local minima problems. Specifically, we start with a large value of  $\sigma$  and gradually decrease it to improve convergence.

#### 4. EXPERIMENTAL RESULTS

We now present several experimental results to demonstrate the utility of the unified non-local regularization framework. We start by showing that the proposed patch based framework provides qualitatively similar results to conventional pixel based schemes. We compare the two algorithms in the context of deblurring in Fig. 1. The original cameraman image is corrupted using Gaussian low-pass blurring (standard deviation of ten pixels) and white Gaussian noise with  $\sigma = 10$ . Only one iteration of the two algorithms were considered; the weights were obtained from the blurred image with  $\sigma = 10$  and the images are recovered using (1) and (10), respectively. We used  $3 \times 3$  image patches and search windows of  $5 \times 5$ . We chose the regularization parameters of both schemes such that the signal to noise ratio is maximized. Note that both schemes provide qualitatively similar results, demonstrating the equivalence of the methods.

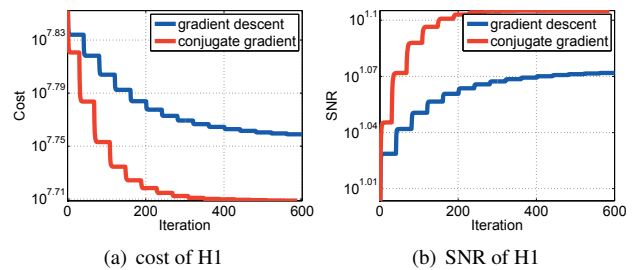
We compare the convergence rate of the proposed conjugate gradient algorithm to minimize (4) with the current iterative reweighted non-linear minimization schemes [2, 3] in the context of compressed sensing in Fig. 2. Specifically, we aim to recover the MRI image, shown in Fig. 3.a from 60 radial lines. We plot the decrease in the cost function and the improvement in signal to noise ratio as a function of the number of iterations in Fig. 2. It is seen that the proposed CG



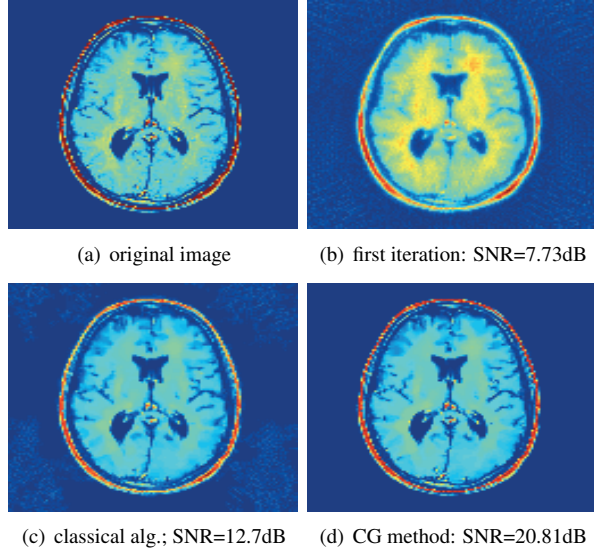
**Fig. 1.** Comparison of conventional pixel-based NL regularization scheme with proposed patch-based algorithm in the context of deblurring. Note that both the algorithms give comparable results when the distance function  $\varphi$  are chosen to match the weights.

scheme provides a much faster convergence to the minimum of 4, compared to the conventional algorithm. The images recovered after 600 iterations are shown in Fig. 3.(c) and (d) respectively. Note that the CG scheme resulted in considerably reduced aliasing artifacts compared to the classical scheme, resulting in a 8dB improvement in SNR. It is also seen that the image recovered using only one iteration (using weights derived from the aliased image) result in a very poor estimate; Unlike deblurring, iterating the conventional NL regularized algorithm, which is equivalent to (4), is absolutely necessary in challenging inverse problems such as compressed sensing.

We compare the utility of the distance metrics specified by (13) and (11) in Fig. 4. Here, we recover the Shepp Logan phantom from 30 radial lines using these algorithms. The regularization parameter was chosen to maximize the SNR. It is observed that the  $H_1$  scheme recovers the image with some



**Fig. 2.** Comparison of optimization speed of GD and CG. Clearly, CG speeds up the optimization significantly.



**Fig. 3.** Comparison of reconstructed images using gradient descent and conjugate gradient after 600 iterations. Note that the CG method in (d) provides superior reconstructions compared to the current iterative reweighted scheme shown in (c). The first iteration of the algorithm is shown in (b).

blurring of the small lesions. The straightforward implementation of (4) with (13) results in crisper details, but with some residual aliasing (see Fig. 4.(c)). This is due to the convergence of the scheme to local minimum. We use a continuation strategy, where  $\sigma$  is initialized with a large value and gradually decreased to the specified value. We observe that this approach significantly improved the convergence, resulting in an 11 dB improvement over Fig. 4.(b).

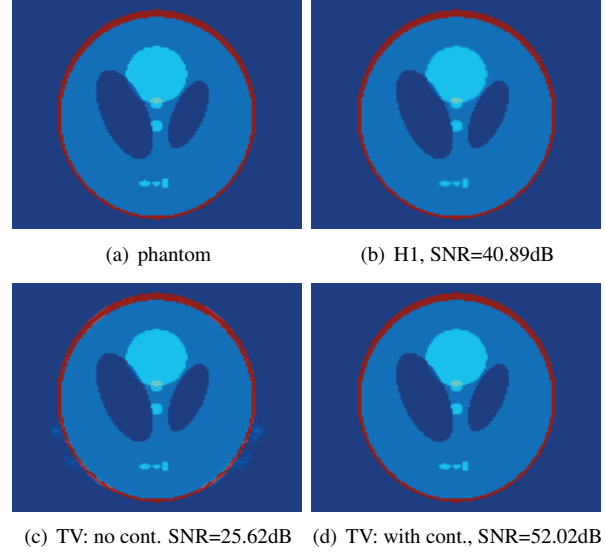
## 5. DISCUSSION AND CONCLUSION

We introduced a unified framework for the non-local regularized recovery of inverse problems. We show that a majorize minimize algorithm to minimize the proposed criterion simplifies to the common sense approach of iterating current non-local regularization technique. In addition to providing a novel interpretation for such heuristic methods, the proposed scheme enabled us to develop efficient optimization algorithms, design novel non-local schemes by choosing the distance metric, and minimize local minima problems.

## 6. APPENDIX

We consider the cost function specified by

$$\hat{f} = \arg \min_f \|Af - b\|^2 + \lambda \int_{\Omega \times \Omega} \psi \left( \sqrt{\|N_{\mathbf{x}} - N_{\mathbf{y}}\|} \right) dx dy. \quad (15)$$



**Fig. 4.** Comparison of reconstructed images using non-local  $H^1$  and TV regularization in highly undersampled case. Non-local TV reduces more artifacts than H1 and has higher SNR.

Note that we set  $\varphi(x) = \psi(\sqrt{x})$ . Using (6), we have

$$\psi \left( \sqrt{\|N_{\mathbf{x}} - N_{\mathbf{y}}\|} \right) \leq \underbrace{\frac{\psi'(\sqrt{\|N_{\mathbf{x}} - N_{\mathbf{y}}\|})}{2\sqrt{\|N_{\mathbf{x}} - N_{\mathbf{y}}\|}}}_{w(\mathbf{x}, \mathbf{y})} \|N_{\mathbf{x}} - N_{\mathbf{y}}\| + b.$$

Thus, the MM scheme to solve (15) will simplify to the non-local TV scheme specified by (12), where the weights  $w_n(\mathbf{x}, \mathbf{y})$  are related to  $\psi$  as shown above. The authors used the same weights as (3). Converting to the standard form, we have  $\varphi'(x) = \psi'(\sqrt{x})/2\sqrt{x} = \exp(-x^2/\sigma^2)$ , which results in (13).

## 7. REFERENCES

- [1] 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.
- [2] G. Gilboa, J. Darbon, S. Osher, and T. Chan, “Nonlocal convex functionals for image regularization,” *UCLA CAM Report*, pp. 06–57, 2006.
- [3] 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.
- [4] 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.