Published in final edited form as:

Comput Diffus MRI (2016). 2016 October ; 2016: 49–59. doi:10.1007/978-3-319-54130-3_4.

Denoising Diffusion-Weighted Images Using Grouped Iterative Hard Thresholding of Multi-Channel Framelets

Jian Zhang^{1,5}, Geng Chen^{2,5}, Yong Zhang³, Bin Dong⁴, Dinggang Shen⁵, and Pew-Thian Yap⁵

¹School of Information and Electrical Engineering, Hunan University of Science & Technology, Xiangtan, China

²Data Processing Center, Northwestern Polytechnical University, Xi'an, China

³Department of Psychiatry & Behavioral Sciences, Stanford University, U.S.A

⁴Beijing International Center for Mathematical Research, Peking University, Beijing, China

⁵Department of Radiology and Biomedical Research Imaging Center, University of North Carolina, Chapel Hill, U.S.A

Abstract

Noise in diffusion-weighted (DW) images increases the complexity of quantitative analysis and decreases the reliability of inferences. Hence, to improve analysis, it is often desirable to remove noise and at the same time preserve relevant image features. In this paper, we propose a tight wavelet frame based approach for edge-preserving denoising of DW images. Our approach (i) employs the unitary extension principle (UEP) to generate frames that are discrete analogues to differential operators of various orders; (ii) introduces a very efficient method for solving an ℓ_1 denoising problem that involves only thresholding and solving a trivial inverse problem; and (iii) groups DW images acquired with neighboring gradient directions for collaborative denoising. Experiments using synthetic data with noncentral chi noise and real data with repeated scans confirm that our method yields superior performance compared with denoising using state-of-the-art methods such as non-local means.

Keywords

Diffusion-weighted imaging; tight wavelet frames; multi-channel framelets; sparse representation; noncentral chi noise

1 Introduction

Diffusion MRI is a powerful neuroimaging technique due to its unique ability to extract microstructural information by utilizing restricted and hindered diffusion to probe compartments that are much smaller than the voxel size. One important goal of diffusion MRI is to estimate axonal orientations, tracing of which will allow one to gauge connectivity

between brain regions and will provide *in vivo* information on white matter pathways for neuroscience studies involving development, aging, and disorders [1–5]. In order to capture orientation information, the brain has to be scanned using a range of diffusion-sensitizing gradient directions that are ideally distributed uniformly on the unit sphere.

As shown in Figure 1, DW images that are scanned with similar gradient directions share a lot of commonalities. However, these commonalities diminish very quickly if the difference between the gradient directions increases. As can also be seen from the figure, the images are typically very noisy and can benefit greatly from denoising. Denoising performance can be improved by borrowing information between images scanned at similar gradient directions; however, images scanned at a very different direction have to be avoided in this process to reduce artifacts.

In this paper, we take advantage of the correlation between DW images scanned with neighboring gradient directions in a group ℓ_1 minimization denoising framework that is based on tight wavelet frames. The power of tight wavelet frames lies in their ability to sparsely approximate piecewise smooth functions and the existence of fast decomposition and reconstruction algorithms associated with them. In contrast, total-variation (TV) based methods are effective in restoring images that are piecewise constant, e.g., binary or cartoon-like images. They will, however, cause staircasing effects in images that are not piecewise constant [6].

Instead of the more conventional ℓ_1 regularization, which has been shown in the theory of compressed sensing [7] to produce sparse solutions, we opted to use ℓ_2 regularization. In [8], both iterative soft and hard thresholding algorithms were adopted and the latter was found to achieve better image quality. In [9], wavelet frame based ℓ_2 regularization also shows better edge-preserving quality compared with the conventional ℓ_1 regularization. In contrast to previous works, we propose a group version of ℓ_1 minimization to take advantage of the correlation between DW images.

Evaluations performed using synthetic data with noncentral chi noise distribution as well as real data with repeated scans indicate that the proposed method is superior to its ℓ_1 counterpart and non-local means denoising.

2 Approach

We will first provide a brief introduction to framelets, followed by details on how framelets can be incorporated into an ℓ_1 minimization framework for DWI denoising.

2.1 Tight Framelets

A system $X \subset L_2(\mathbb{R})$ is called a tight wavelet frame of $L_2(\mathbb{R})$ if

$$f = \sum_{g \in X} \langle f, g \rangle g, \quad \forall f \in L_2(\mathbb{R}), \quad (1)$$

where $\langle \cdot, \cdot \rangle$ is the inner product of $L_2(\mathbb{R})$. It is clear that an orthonormal basis is a tight frame, since the identity hold for arbitrary orthonormal bases in $L_2(\mathbb{R})$. Tight frames are generalization of orthonormal bases with greater redundancy — a property central to sparse representation and often desirable in applications such as denoising [10].

Given a set of generators $\Psi := \{\psi_1, \dots, \psi_R\} \subset L_2(\mathbb{R}^d)$, which are desirably (anti)symmetric and compact functions, the corresponding quasi-affine system $X(\Psi)$ generated by Ψ is the collection of dilations and shifts of Ψ : $X(\Psi) = \{\psi_{l,r,k} : 1 \leq r \leq R; l, k \in \mathbb{Z}\}$, where $\psi_{l,n,k}$ is defined by

$$\psi_{l,r,k} := \begin{cases} 2^{\frac{l}{2}} \psi_r(2^l \cdot -k), & l \geq J, \\ 2^{\frac{l-J}{2}} \psi_r(2^l \cdot -2^{l-J} k), & l < J. \end{cases} \quad (2)$$

When $X(\Psi)$ forms an orthonormal basis of $L_2(\mathbb{R})$, $X(\Psi)$ is called an orthonormal wavelet basis. When $X(\Psi)$ forms a (tight) frame of $L_2(\mathbb{R})$, each function ψ_r , $r = 1, \dots, R$, is called a (tight) framelet and the whole system $X(\Psi)$ is called a (tight) wavelet frame system. A tight wavelet frame is also called a Parseval frame. Note that in the literature, the affine (or wavelet) system, which corresponds to the decimated wavelet (frame) transforms, is commonly used. The quasi-affine system, introduced and analyzed in [11], corresponds to the undecimated wavelet (frame) transforms and essentially over samples the wavelet frame system starting from level $J-1$ and downwards. In this paper, we focus on the quasi-affine system because it has been shown to work better in image restoration [12]. We set $J=0$ and consider only $l < 0$.

The construction of Ψ is usually based on a multiresolution analysis (MRA) [12] that is generated by some refinable function ϕ with refinement mask $a_0 \in \ell_2(\mathbb{Z})$ satisfying

$$\phi = 2 \sum_{k \in \mathbb{Z}} a_0[k] \phi(2 \cdot -k). \quad (3)$$

The key is to find the masks $a_r \in \ell_2(\mathbb{Z})$ that gives

$$\psi_r = 2 \sum_{k \in \mathbb{Z}} a_r[k] \phi(2 \cdot -k), \quad r=1, 2, \dots, R. \quad (4)$$

The finite sequences, $a_1, \dots, a_R$ are called wavelet frame masks, or the high pass filters of the system. The refinement mask a_0 is also known as the low pass filter. The two equations above can be combined by defining $\psi_0 := \phi$.

The unitary extension principle (UEP) [11] provides a general theory for constructing MRA-based tight wavelet frames. As long as $\{a_1, \dots, a_R\}$ are finitely supported and their Fourier series satisfy

$$\sum_{r=0}^R |\hat{a}_r(\xi)|^2 = 1 \quad \text{and} \quad \sum_{r=0}^R \hat{a}_r(\xi) \overline{\hat{a}_r(\xi + \nu)} = 0, \quad (5)$$

for all $\nu \in \{0, \pi\}$ and $\xi \in [-\pi, \pi]$, the quasi-affine system $X(\Psi)$ forms a tight frame in $L_2(\mathbb{R})$.

Consider the centered B-splines of order p , i.e.,

$$\hat{\phi}(\xi) = e^{-ij\frac{\xi}{2}} \left(\frac{\sin(\xi/2)}{\xi/2} \right)^p, \quad (6)$$

with $j = 0$ when p is even and $j = 1$ when p is odd. The corresponding refinement mask is given as

$$\hat{a}_0(\xi) = e^{-ij\frac{\xi}{2}} \cos^p(\xi/2), \quad (7)$$

and the p wavelet masks as

$$\hat{a}_r(\xi) := -i^r e^{-ij\frac{\xi}{2}} \sqrt{\binom{p}{r}} \sin^r(\xi/2) \cos^{p-r}(\xi/2), \quad r = 1, 2, \dots, p. \quad (8)$$

It is straightforward to show that the UEP conditions (5) are satisfied. Wavelet frame masks for $p = 1, 2, 4$ are shown in Table 1. It is worth noting that these masks correspond to differential operators of various orders. For example, for piecewise linear B-spline, the masks a_1 and a_2 correspond to the first order and second order difference operators respectively up to a scaling factor.

When a tight wavelet frame is used, the given data is considered to be sampled as a local average $u[k] = \langle f, \phi(\cdot - k) \rangle$. Noting that [12]

$$\langle f, \psi_{l-1,r,k} \rangle = \sum_{k' \in \mathbb{Z}} a_{l,r}[k'] \langle f, \psi_{l,0,k+k'} \rangle, \quad (9)$$

where the dilated sequence is defined as

$$a_{l,r}[k] = \begin{cases} a_r[2^l k], & k \in 2^{-l}\mathbb{Z}, \\ 0, & k \notin 2^{-l}\mathbb{Z}, \end{cases} \quad (10)$$

the decomposition and reconstruction down to level $-L$ [12], i.e.,

$$\mathcal{P}_0 f = \mathcal{P}_{-L} f + \sum_{r=1}^R \sum_{j=-L}^{-1} \sum_{k \in \mathbb{Z}} \langle f, \psi_{r,j,k} \rangle \psi_{r,j,k}, \quad (11)$$

where

$$\mathcal{P}_l f = \sum_{r=1}^R \sum_{j < l} \sum_{k \in \mathbb{Z}} \langle f, \psi_{r,j,k} \rangle \psi_{r,j,k}, \quad (12)$$

can be realized with convolution using the masks. We denote by W the L -level framelet decomposition, i.e.,

$$Wu = (\dots, W_{l,r}u, \dots)^\top \quad \text{for } (l, r) \in \mathcal{B}_L, \quad (13)$$

with $\mathcal{B}_L := \{(1, 1), (1, 2), \dots, (1, R), (2, 1), \dots, (L, R)\} \cup \{(L, 0)\}$, where the level l and band r framelet decomposition is given by

$$W_{l,r}u = a_{-l,r} \otimes a_{-l+1,0} \otimes \dots \otimes a_{0,0} \otimes u. \quad (14)$$

If we use $W^\top$ to denote the framelet reconstruction, we have $W^\top W = I$, i.e., $u = W^\top Wu$.

Given a 1-dimensional framelet system for $L_2(\mathbb{R})$, the d -dimensional tight wavelet frame system for $L_2(\mathbb{R}^d)$ can be easily constructed by using tensor products of 1-dimensional framelets [12].

2.2 Problem Formulation

Given a vector-valued image f of an arbitrary dimension with pixel $i \in \{1, \dots, N\}$ consisting of vector $f_i \in \mathbb{R}^M$, we are interested in restoring its denoised counterpart u by solving the following problem:

$$\min_u \left\{ \phi(u) = \|u - f\|_2^2 + \sum_{i,g,l,r} \lambda_{g,l,r} \sqrt{\sum_m w_{g,m}^2 \| (W_{l,r} u^{(m)})_i \|_2^2} \right\}. \quad (15)$$

The regularization term is in fact a sum of G regularization terms, each of which grouping a set of images. The g -th grouping (with associated tuning parameter $\lambda_{g,l,r}$, where $g = \{1, 2, \dots, G\}$, is defined according to a set of weights $\{w_{g,m}\}$, where $m \in \{1, \dots, M\}$. Channels with $w_{g,m} = 0$ are included in the grouping and their weighted framelet coefficients are jointly considered via ℓ_2 -norm for penalization. The different groupings can possibly overlap, implying each image can be at the same time considered in different groups. This is in similar spirit as the overlapped group LASSO [13]. We set $\lambda_{g,l,r} = \lambda (\sum_m w_{g,m}^2)^{\frac{1}{2}}$ if $l, r = 0$ or $\lambda_{g,l,r} = 0$ if otherwise. Here λ is a constant that can be set independent of the weights.

2.3 Optimization

Problem (15) can be solved effectively using penalty decomposition (PD) [14]. Defining auxiliary variables $(v_{g,m,l,r})_i := w_{g,m} (W_{l,r} u^{(m)})_i$, this amounts to minimizing the following objective function with respect to u and $v := \{v_{g,m,l,r}\}$:

$$L_\mu(u, v) = \|u - f\|_2^2 + \sum_{i,g,l,r} \lambda_{g,l,r} \sqrt{\sum_m \| (v_{g,m,l,r})_i \|_2^2} + \frac{\mu}{2} \sum_{i,g,m,l,r} \|w_{g,m} (W_{l,r} u^{(m)})_i - (v_{g,m,l,r})_i\|_2^2, \quad (16)$$

In PD, we (i) alternate between solving for u and v using block coordinate descent (BCD). Once this converges, we (ii) increase $\mu > 0$ by a multiplicative factor that is greater than 1 and repeat step (i). This is repeated until increasing μ does not result in further changes to the solution [14].

First Subproblem—We solve for v in the first problem, i.e., $\min_v L_\mu(u, v)$. This is a group ℓ_2 problem and the solution can be obtained via hard-thresholding:

$$(v_{g,m,l,r})_i = \begin{cases} w_{g,m} (W_{l,r} u^{(m)})_i, & (h_{g,l,r})_i \geq \frac{2\lambda_{g,l,r}}{\mu} \\ 0, & \text{otherwise,} \end{cases} \quad (17)$$

where

$$(h_{g,l,r})_i = \sum_{m'} \|w_{g,m'} (W_{l,r} u^{(m')})_i\|_2^2. \quad (18)$$

An ℓ_1 version of the algorithm can be obtained by using soft-thresholding instead.

Second Subproblem—By taking the partial derivative with respect to $u^{(m)}$, the solution to the second subproblem, i.e., $\min_u L_\mu(u, v)$, is for each m

$$\left(I + \frac{\mu}{2} \sum_{g,l,r} w_{g,m}^2 W_{l,r}^\top W_{l,r} \right) u^{(m)} = f^{(m)} + \frac{\mu}{2} \sum_{g,l,r} w_{g,m} W_{l,r}^\top v_{g,m,l,r}, \quad (19)$$

where we have dropped the subscript i for notation simplicity. Note that since we have

$\sum_{l,r} W_{l,r}^\top W_{l,r} = I$, the the problem can be simplified to become

$$\left(1 + \frac{\mu}{2} \sum_g w_{g,m}^2 \right) u^{(m)} = f^{(m)} + \frac{\mu}{2} \sum_{g,l,r} w_{g,m} W_{l,r}^\top v_{g,m,l,r}. \quad (20)$$

Solving the above equation for $u^{(m)}$ is trivial and involves only simple division.

2.4 Setting the Weights

In setting the weights $\{w_{g,m}\}$, we note that the weights should decay with the dissimilarity between gradient directions associated with a pair of diffusion-weighted images. To reflect this, we let $G = M$ and set for $g, m \in \{1, \dots, M\}$ $w_{g,m} = e^{\kappa[(\nu_m^\top \nu_g)^2 - 1]}$ if $|\nu_m^\top \nu_g| < \cos(\theta)$ or 0 otherwise, where $\kappa > 0$ is a parameter that determines the rate of decay of the weight. The exponential function is in fact modified from the probability density function of the Watson distribution [15] with concentration parameter κ . Essentially, this implies that for the g -th diffusion-weighted image acquired at gradient direction ν_g , there is a corresponding group of images with associated weights $\{w_{g,m}\}$. The weight is maximal at $w_{g,g} = 1$ and is attenuated when $m \neq g$. To reduce computation costs, weights of images scanned at gradient directions deviating more than θ from ν_g are set to 0, and the respective images are hence discarded from the group. We set $\theta = 30^\circ$.

3 Experimental Results

In all experiments, we used the piecewise linear tight wavelet frame and set the level L to 2. The optimal λ was chosen based on grid search.

3.1 Datasets

Synthetic Data—A synthetic dataset of a spiral was generated for quantitative evaluation. The parameters used for generating synthetic data simulation were consistent with the real data described next: $b = 2000 \text{ s/mm}^2$, 48 gradient directions, $64 \times 64 \times 16$ voxels with resolution $2 \times 2 \times 2 \text{ mm}^3$. Three levels of 32-channel noncentral Chi noise [16] was added: $\sigma =$

5, 7.5, and 10, corresponding to $\text{SNR} = 30, 20, 10$ with respect to the white matter non-diffusion-weighted signal.

Real Data—The real datasets were acquired using Siemens 3T TRIO MR scanner with the same gradient directions and b -value as the synthetic dataset. The imaging protocol is as follows: 128×96 imaging matrix, voxel size of $2 \times 2 \times 2 \text{ mm}^3$, $\text{TE} = 97 \text{ ms}$, $\text{TR} = 11,300 \text{ ms}$. Imaging acquisition was repeatedly performed on the same subject for 8 times. We averaged the 8 sets of DW images and removed the noncentral chi noise bias to obtain the ground truth for evaluation.

3.2 Results

The numerical results for the synthetic data, shown in Fig. 2, indicate that the proposed ℓ_2 framelet denoising method gives the best performance for all noise levels, showing improvements over ℓ_1 framelet denoising and the non-local means (NLM) algorithm [17]. The DW images, shown in Fig. 3, indicate that both ℓ_1 and ℓ_2 give sharper edges compared with NLM. Noise, however, is not totally removed for the case of ℓ_1 denoising. Only ℓ_2 denoising is able to effectively remove noise and preserve edges. Note that, for both synthetic and real data, noncentral chi bias was removed using the method described in [16]. The best tuning parameters for both ℓ_1 and ℓ_2 were selected based on grid search.

For the real data, we used the average image as the ground truth for PSNR computation. The results, shown in Fig. 4, are consistent with Fig. 2, indicating that ℓ_2 gives the best performance. The visual results in Fig 5 indicate that the results given by ℓ_2 is closest to the ground truth. In contrast, NLM over-smooths the image and edge information is hence lost.

4 Conclusion

In this paper, we have proposed a denoising method by using multi-channel framelet grouped iterative hard thresholding, which not only takes advantage of inter-image correlations but yields good edge-preserving property. Experiments on synthetic data with noncentral chi noise and real data with repeated scans confirm that the proposed method outperforms state-of-the-art methods such as non-local means.

Acknowledgments

This work was supported in part by NIH grants (NS093842, EB006733, EB009634, AG041721, MH100217, and AA012388) and Hunan Provincial Education Department grant (15A066).

References

1. Yap PT, Fan Y, Chen Y, Gilmore J, Lin W, Shen D. Development trends of white matter connectivity in the first years of life. *PLoS ONE*. 2011; 6(9):e24678. [PubMed: 21966364]
2. Wee CY, Yap PT, Li W, Denny K, Brownyke J, Potter G, Welsh-Bohmer K, Wang L, Shen D. Enriched white-matter connectivity networks for accurate identification of MCI patients. *NeuroImage*. 2011; 54(3):1812–1822. [PubMed: 20970508]
3. Wee CY, Yap PT, Zhang D, Denny K, Browndyke JN, Potter GG, Welsh-Bohmer KA, Wang L, Shen D. Identification of MCI individuals using structural and functional connectivity networks. *NeuroImage*. 2012; 59(3):2045–2056. [PubMed: 22019883]

4. Wee CY, Wang L, Shi F, Yap PT, Shen D. Diagnosis of autism spectrum disorders using regional and interregional morphological features. *Human Brain Mapping*. 2014; 35(7):3414–3430. in press. [PubMed: 25050428]
5. Jin Y, Wee CY, Shi F, Thung KH, Yap PT, Shen D. Identification of infants at high-risk for autism spectrum disorder using multi-parameter multi-scale white matter connectivity networks. *Human Brain Mapping*. 2015; 36(12):4880–4896. [PubMed: 26368659]
6. Ring W. Structural properties of solutions to total variation regularization problems. *Mathematical Modelling and Numerical Analysis*. 2000; 34:799–810.
7. Candès EJ, Romberg J, Tao T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*. 2006; 52(2): 489–509.
8. Chan RH, Chan TF, Shen L, Shen Z. Wavelet algorithms for high-resolution image reconstruction. *SIAM Journal on Scientific Computing*. 2003; 24(4):1408–1432.
9. Zhang Y, Dong B, Lu Z. ℓ_1 minimization for wavelet frame based image restoration. *Mathematics of Computation*. 2013; 82:995–1015.
10. Mallat, S. *A Wavelet Tour of Signal Processing: The Sparse Way*. 3. Academic Press; 2008.
11. Ron A, Shen Z. Affine systems in $\ell_2(\mathbb{R}^d)$: The analysis of the analysis operator. *Journal of Functional Analysis*. 1997; 148(2):408–447.
12. Dong, B., Shen, Z. IAS Lecture Notes Series, Summer Program on “The Mathematics of Image Processing”. Park City Mathematics Institute; 2010. MRA-based wavelet frames and applications.
13. Jacob L, Obozinski G, Vert JP. Group lasso with overlap and graph lasso. *International Conference on Machine Learning*. 2009
14. Lu Z, Zhang Y. Sparse approximation via penalty decomposition methods. *SIAM Journal on Optimization*. 2013
15. Schwartzman A, Dougherty RF, Taylor JE. False discovery rate analysis of brain diffusion direction maps. *Annals of Applied Statistics*. 2008; 2(1):153–175.
16. Koay CG, Özarslan E, Basser PJ. A signal transformational framework for breaking the noise floor and its applications in MRI. *Journal of Magnetic Resonance*. 2009; 197(2):108–119. [PubMed: 19138540]
17. Coupé P, Yger P, Prima S, Hellier P, Kervrann C, Barillot C. An optimized blockwise nonlocal means denoising filter for 3-D magnetic resonance images. *IEEE transactions on medical imaging*. 2008; 27(4):425–441. [PubMed: 18390341]

Fig. 1.

Diffusion-weighted images scanned at different gradient directions. The left and middle images were scanned with similar gradient directions. The right image was scanned at a nearly perpendicular gradient direction with respect to the reference.

Fig. 2. Performance comparison between NLM and our method using synthetic data.

Fig. 3.
Comparison of denoised DW images given by different methods ($\sigma = 5$).

Fig. 4.
Performance comparison between NLM and our method using real data.

Fig. 5.
Comparison of denoised DW images using the real data.

Table 1

Wavelet frame masks.

Piecewise Constant, $p = 1$	Piecewise Linear, $p = 2$	Piecewise Cubic, $p = 4$
$a_0 = \frac{1}{2}[1, 1]$	$a_0 = \frac{1}{4}[1, 2, 1]$	$a_0 = \frac{1}{16}[1, 4, 6, 4, 1]$
$a_1 = \frac{1}{2}[1, -1]$	$a_1 = \frac{1}{4}[-1, 2, -1]$	$a_1 = \frac{1}{16}[1, -4, 6, -4, 1]$
	$a_2 = \frac{\sqrt{2}}{4}[1, 0, -1]$	$a_2 = \frac{1}{8}[-1, 2, 0, -2, 1]$
		$a_3 = \frac{\sqrt{6}}{16}[1, 0, -2, 0, 1]$
		$a_4 = \frac{1}{8}[-1, -2, 0, 2, 1]$