Published in final edited form as:

Med Image Comput Comput Assist Interv. 2016 October ; 9902: 561–569. doi:
10.1007/978-3-319-46726-9_65.

Tight Graph Framelets for Sparse Diffusion MRI q -Space Representation

Pew-Thian Yap¹, Bin Dong², Yong Zhang³, and Dinggang Shen¹

¹Department of Radiology and BRIC, University of North Carolina, Chapel Hill, U.S.A.

²Beijing International Center for Mathematical Research, Peking University, Beijing, China

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

Abstract

In diffusion MRI, the outcome of estimation problems can often be improved by taking into account the correlation of diffusion-weighted images scanned with neighboring wavevectors in q -space. For this purpose, we propose in this paper to employ tight wavelet frames constructed on non-flat domains for multi-scale sparse representation of diffusion signals. This representation is well suited for signals sampled regularly or irregularly, such as on a grid or on multiple shells, in q -space. Using spectral graph theory, the frames are constructed based on quasi-affine systems (i.e., generalized dilations and shifts of a finite collection of wavelet functions) defined on graphs, which can be seen as a discrete representation of manifolds. The associated wavelet analysis and synthesis transforms can be computed efficiently and accurately without the need for explicit eigen-decomposition of the graph Laplacian, allowing scalability to very large problems. We demonstrate the effectiveness of this representation, generated using what we call *tight graph framelets*, in two specific applications: denoising and super-resolution in q -space using ℓ_0 regularization. The associated optimization problem involves only thresholding and solving a trivial inverse problem in an iterative manner. The effectiveness of graph framelets is confirmed via evaluation using synthetic data with noncentral chi noise and real data with repeated scans.

1 Introduction

Diffusion-weighted MR signals, in addition to the diffusion time, are controlled by a quantity called wavevector. The wavevector is dependent on the length, strength, and orientation of the gradient pulses during the measurement sequence. It is often denoted as vector $\mathbf{q}$, which can be separated into a scalar wavenumber $|\mathbf{q}|$ and a diffusion encoding direction $\hat{\mathbf{q}} = \mathbf{q}/|\mathbf{q}|$. The effects of both diffusion time, t , and wavenumber are summarized using a quantity called b -value, which is defined as $b = t|\mathbf{q}|^2$. A diffusion MRI protocol normally acquires multiple diffusion-weighted images, each corresponding to a wavevector $\mathbf{q}$ in q -space. For example, in shell acquisition schemes, b is fixed by fixing both t and $|\mathbf{q}|$ and only the gradient direction varies among measurements. This can be extended to multi-shell acquisition, where the measurements are collected at shells of different b -values. These

[✉] ptyap@med.unc.edu.

images afford microstructural information that can be harnessed for tissue characterization and axonal tracing.

Due to the noisy nature of diffusion MRI, it is often beneficial to leverage the correlation between diffusion-weighted images acquired with neighboring wavevectors in q -space for improving estimation robustness in applications such as denoising and compressed-sensing reconstruction. To achieve this, a common practice is by fitting a diffusion model to the data, concurrently taking into account all q -space measurements at each voxel. Such approach, however, is dependent on specific assumptions the model makes regarding the data. For example, the diffusion tensor model assumes that the signals at each voxel reflect only axon bundles that are aligned in one coherent direction, hence falling short in capturing the complexity of the water diffusion patterns in the human brain.

In this paper, we take a wavelet-based approach, which is less dependent on specific biophysical assumptions, to constructing sparse representations for diffusion signals. To cater to the fact that q -space sampling schemes can differ significantly, ranging from Cartesian sampling in diffusion spectrum imaging (DSI) to multi-shell sampling with multiple distinct b -values, we propose to represent the q -space signals using tight wavelet frames defined on graphs [1, 2], which can be flexibly adapted to different sampling domains. 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. A wide range of tight wavelet frames can be generated conveniently using the unitary extension principle (UEP) [3]. We use the term *tight graph framelets* to denote basic functions that, when dilated and shifted, form tight wavelet frames on graphs. Tight graph framelets have compact support, are not restricted to signals of a certain shape (e.g., the “donut” [4]), and provide a natural bridge between continuous and discrete representations [1].

We demonstrate the utility and effectiveness of tight graph framelets in two specific applications: denoising and super-resolution reconstruction in q -space. To take advantage of the high redundancy of graph framelets, we regularize the associated problems using a sparse-inducing norm. Instead of the more conventional ℓ_1 regularization, which has been shown in the theory of compressed sensing to produce sparse solutions, we opted to use ℓ_0 regularization. In [5], both iterative soft and hard thresholding algorithms were adopted and the latter was found to achieve better image quality. In [6], wavelet frame based ℓ_0 regularization also shows better edge-preserving quality compared with the conventional ℓ_1 regularization. Evaluation performed using synthetic data with noncentral chi signal distribution as well as real data with repeated scans indicates that the proposed method is superior to denoising and interpolation using spherical radial basis functions (sRBF) [7].

2 Approach

2.1 Graph Framelets

In q -space, sampling points and their relationships can be encoded, respectively, using vertices and edges of a graph. We denote a graph by $\mathcal{G} := \{\mathcal{V}, \mathcal{E}, w\}$, where $\mathcal{V} := \{v_k \in \mathcal{M} : k = 1, \dots, K\}$ is a set of vertices representing points on a manifold $\mathcal{M}$, $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is a set of

edges relating the vertices, and $w: \mathcal{E} \mapsto \mathbb{R}^+$ is a weight function. The associated adjacency matrix $\mathcal{A} = (w_{k,k'})$ is symmetric with $w_{k,k'} > 0$ if v_k and $v_{k'}$ are connected by an edge in $\mathcal{E}$; otherwise $w_{k,k'} = 0$. Given the degree matrix $D = \text{diag}\{d[1], d[2], \dots, d[K]\}$, where $d[k] = \sum_{k'} w_{k,k'}$, the graph Laplacian, defined as $\mathcal{L} = D - \mathcal{A}$, is consistent with the Laplace-Beltrami operator of the manifold [8]. Denote by $\{\lambda_k, u_k\}_{k=0}^{K-1}$ the pairs of eigenvalues and eigenvectors of $\mathcal{L}$ with $0 = \lambda_0 < \lambda_1 < \lambda_2 < \dots < \lambda_{K-1} = \lambda_{\max}$. The eigenvectors form an orthonormal basis for all functions on the graph: $\langle u_k, u_{k'} \rangle := \sum_{n=1}^K u_k[n] u_{k'}[n] = \delta_{k,k'}$. The Fourier transform of a function $f_{\mathcal{G}}: \mathcal{V} \mapsto \mathbb{R}$ on the graph $\mathcal{G}$ is given by

$$\widehat{f_{\mathcal{G}}}[k] := \sum_{n=0}^{K-1} f_{\mathcal{G}}[n] u_k[n].$$

The key idea involved in constructing wavelet frames on a graph is to view eigenvectors of the graph Laplacian as Fourier basis on graphs and the associated eigenvalues as frequency components [1,2]. One then slices the frequency spectrum in a multi-scale fashion by using a set of masks $\{\hat{a}_r(\cdot) : r = 0, \dots, R\}$, where $\hat{a}_0(\cdot)$ acts as a low-pass filter and $\hat{a}_r(\cdot)$ with $0 < r < R$ as band-pass or high-pass filters. More specifically, the graph framelet analysis transform is defined as $\mathbf{W} f_{\mathcal{G}} := \{W_{l,r} f_{\mathcal{G}} : (l, r) \in \mathcal{B}_L\}$, with $\mathcal{B}_L := \{(1, 1), (1, 2), \dots, (1, R), (2, 1), \dots, (L, R)\} \cup \{(L, 0)\}$ and

$$\widehat{W_{l,r} f_{\mathcal{G}}}[k] := \begin{cases} \hat{a}_r(\gamma^{-L+1} \tilde{\lambda}_k) \widehat{f_{\mathcal{G}}}[k] & l=1, \\ \hat{a}_r(\gamma^{-L+l} \tilde{\lambda}_k) \hat{a}_0(\gamma^{-L+l-1} \tilde{\lambda}_k) \dots \hat{a}_0(\gamma^{-L+1} \tilde{\lambda}_k) \widehat{f_{\mathcal{G}}}[k] & 2 \leq l \leq L, \end{cases} \quad (1)$$

where $\tilde{\lambda}_k = (\lambda_k / \lambda_{\max}) \pi$ and $\gamma > 1$ is the dilation factor. Noting the eigen-decomposition $\mathcal{L} = U \Lambda U^T$, where $\Lambda = \text{diag}\{\lambda_0, \lambda_1, \dots, \lambda_{K-1}\}$, we have

$$W_{l,r} := \begin{cases} U \hat{\Omega}_r(\gamma^{-L+1} \tilde{\Lambda}) U^T & l=1, \\ U \hat{\Omega}_r(\gamma^{-L+l} \tilde{\Lambda}) \hat{\Omega}_0(\gamma^{-L+l-1} \tilde{\Lambda}) \dots \hat{\Omega}_0(\gamma^{-L+1} \tilde{\Lambda}) U^T & 2 \leq l \leq L, \end{cases} \quad (2)$$

where $\tilde{\Lambda} = \text{diag}\{\tilde{\lambda}_0, \tilde{\lambda}_1, \dots, \tilde{\lambda}_{K-1}\}$ and

$$\hat{\Omega}_r(\beta \tilde{\Lambda}) = \text{diag}\{\hat{a}_r(\beta \tilde{\lambda}_0), \hat{a}_r(\beta \tilde{\lambda}_1), \dots, \hat{a}_r(\beta \tilde{\lambda}_{K-1})\}. \quad (3)$$

Letting $\mathbf{a} = \mathbf{W} f_{\mathcal{G}} := \{a_{r,l} := W_{l,r} f_{\mathcal{G}} : (l, r) \in \mathcal{B}_L\}$ and if the masks satisfy $\sum_{r=0}^R |\hat{a}_r(\xi)|^2 = 1$, which is one of the requirements of the unitary extension principle (UEP) [1, 3]), it is easy to show that the synthesis transform $\mathbf{W}^T \mathbf{a}$ gives $\mathbf{W}^T \mathbf{a} = \mathbf{W}^T \mathbf{W} f_{\mathcal{G}} = I f_{\mathcal{G}} = f_{\mathcal{G}}$. To avoid explicit decomposition of a potentially large Laplacian matrix, we approximate the masks

using Chebyshev polynomials, i.e., $\hat{a}_r(\xi) = \mathcal{T}_r^n(\xi)$, where $\mathcal{T}_r^n(\xi) = \frac{1}{2} c_{r,0} + \sum_{k=1}^{n-1} c_{r,k} T_k(\xi)$ with $T_k(\xi)$ being the k -th order Chebyshev polynomial and

$$c_{r,k} = \frac{2}{\pi} \int_0^\pi \cos(k\theta) \hat{a}_r \left(\frac{\pi}{2} (\cos \theta + 1) \right) d\theta. \quad (4)$$

Equation (2) can be shown to be equivalent to

$$W_{r,l} := \begin{cases} \mathcal{T}_r^n(\gamma^{-L+1} \mathcal{L}) & l=1, \\ \mathcal{T}_r^n(\gamma^{-L+l} \mathcal{L}) \mathcal{T}_r^n(\gamma^{-L+l-1} \mathcal{L}) \dots \mathcal{T}_r^n(\gamma^{-L+1} \mathcal{L}) & 2 \leq l \leq L. \end{cases} \quad (5)$$

Note the interchange between scalar-valued and matrix-valued polynomials. See [1, 2] for more details. Some examples of framelet masks are shown in Table 1.

2.2 Sparse Recovery

Given the observation vector $f \in \mathbb{R}^{K'}$, we are interested in recovering $u \in \mathbb{R}^K$ by solving the following problem:

$$\min_u \left\{ \phi(u) = \|Au - f\|_{2,\mathcal{G}}^2 + \sum_{(l,r) \in \mathcal{B}_L} \lambda_{l,r} \|W_{l,r}u\|_{0,\mathcal{G}} \right\}, \quad (6)$$

where $\|\bullet\|_{2,\mathcal{G}}^2 := \sum_i^{K'} (\bullet[i])^2 d[k_i]$ and $\|\bullet\|_{0,\mathcal{G}} := \sum_{k=1}^K \mathcal{I}(\bullet[k]) d[k]$ with $\mathcal{I}(z)$ being an indicator function that returns 1 when $z \neq 0$ or 0 otherwise. Here, u is defined on $\mathcal{G}$ with degree matrix $D := \text{diag}\{d[1], d[2], \dots, d[K]\}$. The observation vector f is defined in relation to a subset of vertices of $\mathcal{G}$, indexed by $\{k_i: i = 1, \dots, K'\}$. We let $D' := \text{diag}\{d[k_1], d[k_2], \dots, d[k_{K'}]\}$. K' and K are in general not necessarily equal. We require $K' \leq K$.

The problem (6) can be solved effectively using penalty decomposition (PD) [6, 9]. Defining auxiliary variables $v := (v_{l,r}) := (W_{l,r}u)$, this amounts to minimizing the following objective function with respect to u and v :

$$L_\mu(u, v) = \|Au - f\|_{2,\mathcal{G}}^2 + \sum_{(l,r) \in \mathcal{B}_L} \lambda_{l,r} \|v_{l,r}\|_{0,\mathcal{G}} + \frac{\mu}{2} \sum_{(l,r) \in \mathcal{B}_L} \|W_{l,r}u - v_{l,r}\|_2^2. \quad (7)$$

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 [6, 9].

First Subproblem—We solve for v in the first problem, i.e., $\min_v L_\mu(u, v)$:

$$\min_v \sum_{(l,r) \in \mathcal{B}_L} \left\{ \frac{\lambda_{l,r}}{\mu} \|v_{l,r}\|_{0,\mathcal{G}} + \frac{1}{2} \|W_{l,r}u - v_{l,r}\|_2^2 \right\}. \quad (8)$$

The solution can be obtained via hard thresholding [10, 11]:

$$v_{l,r}[k] = \begin{cases} (W_{l,r}u)[k] & ((W_{l,r}u)[k])^2 \geq \frac{2\lambda_{l,r}d[k]}{\mu}, \\ 0 & \text{otherwise.} \end{cases} \quad (9)$$

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

$$\left(A^{\top} D' A + \frac{\mu}{2} \sum_{(l,r) \in \mathcal{B}_L} W_{l,r}^{\top} W_{l,r} \right) u = A^{\top} D' f + \frac{\mu}{2} \sum_{(l,r) \in \mathcal{B}_L} W_{l,r}^{\top} v_{l,r}. \quad (10)$$

Since $\sum_{(l,r) \in \mathcal{B}_L} W_{l,r}^{\top} W_{l,r} = \mathbf{W}^{\top} \mathbf{W} = I$, the problem can be simplified to become

$$\left(A^{\top} D' A + \frac{\mu}{2} I \right) u = A^{\top} D' f + \frac{\mu}{2} \sum_{(l,r) \in \mathcal{B}_L} W_{l,r}^{\top} v_{l,r}, \quad (11)$$

which is a set of linear equations that can be solved, for example, using the conjugate gradient method.

2.3 Denoising and Super-Resolution

We are interested in the super-resolution problem of recovering signals when given only a subset of noisy measurements. We partition u as $u = [u_1; u_2] \in \mathbb{R}^K$, where $u_1 \in \mathbb{R}^{K'}$ is defined on vertices whose observed values are given by $f \in \mathbb{R}^{K'}$. No observed values are available for u_2 . To recover u , we set in (6) $A = [I_{K' \times K'} \ 0] \in \mathbb{R}^{K' \times K}$, where $I_{K' \times K'}$ is an identity matrix of size $K' \times K'$. When $K' = K$, i.e., $u = u_1$ and $A = I_{K' \times K'}$, the problem reduces to a denoising problem, where we are interested in recovering the noiseless version of f . Since

$$A^{\top} D' A = [I_{K' \times K'} \ 0]^{\top} D' [I_{K' \times K'} \ 0] = \begin{bmatrix} D' & 0 \\ 0 & 0 \end{bmatrix}, \quad (12)$$

the resulting matrix on the left of (11) is diagonal and the solution can hence be obtained easily without matrix inversion.

3 Experiments

3.1 Parameter Settings

For all experiments, we used the quadratic masks, set the decomposition level to $L = 5$, and set the maximum order of Chebyshev polynomials to $n = 8$. No significant gain in improvement was obtained beyond these values. The dilation factor was set to a standard value $\gamma = 2$. The tuning parameters for the sparse recovery problem were set as

$$\lambda_{l,r} = \begin{cases} \eta^{-l+1}\lambda & 1 \leq l \leq L, r \neq 0, \\ 0 & l=L, r=0, \end{cases} \quad (13)$$

with $\eta = 0.75$ and $\lambda = \sigma_{\text{noise}}^2 \log(K|\mathcal{B}_L|)$ (i.e., the universal penalty level [12]). We define the adjacency matrix $\mathcal{A} := (w_{k,k'})$ of the q -space samples by letting

$$w_{k,k'} = \exp \left\{ -\frac{3(1 - \hat{\mathbf{q}}_k^\top \hat{\mathbf{q}}_{k'})}{(1 - \cos^2(\theta))} \right\} \exp \left\{ -\frac{(b_{k'} - b_k)^2}{2\sigma_b^2} \right\}, \quad (14)$$

where $\sigma_b = 10$ and θ is set to the maximum of the angles between neighboring gradient directions.

3.2 Results

Synthetic Data—We generated a synthetic dataset consisting of a set of voxels with white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) signal profiles. Combinations of these signal profiles were also included to simulate partial volume effects. To generate the WM signal profiles, mixtures of up to two tensor models were used. The diffusivities for WM, GM, and CSF are respectively $[1.7, 0.3, 0.3] \times 10^{-3} \text{ mm}^2/\text{s}$, $0.9 \times 10^{-3} \text{ mm}^2/\text{s}$, and $2.5 \times 10^{-3} \text{ mm}^2/\text{s}$. The number of non-collinear gradient directions were 21, 81, and 321, each generated by subdividing the faces of an icosahedron a number of times. Three shells were generated, i.e., $b = 1000, 2000, 3000 \text{ s/mm}^2$. For evaluation, two levels of 32-channel noncentral chi noise [13] was added, corresponding to $\text{SNR} = 20$ and 30 with respect to the non-diffusion-weighted signal of WM.

Results for denoising (21 – 21, 81 – 81, 321 – 321) and super-resolution (21 – 81, 81 – 321), averaged over 10 realizations of noise, are shown in Fig. 1. The comparison baseline is the spherical radial basis function (sRBF) method described in [7]. This basically amounts to weighted averaging of signals on the shells using the weights given by (14). The results confirm that our method yields consistent PSNR improvements. Note that, for both synthetic and real data, noncentral chi bias was removed from the solution of (6) using the method described in [13].

Real Data—The real datasets were acquired using Siemens 3T TRIO MR scanner with $b = 2000 \text{ s/mm}^2$ and 48 non-collinear gradient directions. 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 diffusion-weighted images and removed the bias caused by the noncentral chi noise to obtain the ground truth for evaluation.

The performance statistics for the denoising of the 8 sets of diffusion-weighted images are shown in Fig. 2, again confirming that our method yields results that are in close agreement with the ground truth with higher PSNRs. We also show in Fig. 3 the white matter fiber orientation distribution functions (ODFs) to demonstrate the benefit of q -space super-resolution. We used the measurements from the original 48 gradient directions to generate the data for 81 directions. It can be observed from the figure that increasing the angular resolution helps resolve some fiber crossings.

4 Conclusion

In this paper, we present preliminary results confirming that tight graph framelets can be employed for sparse representation of diffusion signals in q -space. The effectiveness of this representation is demonstrated with applications involving denoising and super-resolution. Future work will be directed to incorporating this representation with spatial framelet representation to obtain a joint representation for both spatial and wavevector dimensions.

Acknowledgments

This work was supported in part by NIH grants (NS093842, EB006733, EB008374, EB009634, AG041721, and MH100217).

References

1. Dong B. Sparse representation on graphs by tight wavelet frames and applications. *Applied and Computational Harmonic Analysis*. 2015
2. Hammond DK, Vandergheynst P, Gribonval R. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*. 2011; 30(2):129–150.
3. Ron A, Shen Z. Affine systems in $L_2(\mathbb{R}^d)$: The analysis of the analysis operator. *Journal of Functional Analysis*. 1997; 148(2):408–447.
4. Michailovich O, Rathi Y. On approximation of orientation distributions by means of spherical ridgelets. *IEEE Transactions on Image Processing*. 2010; 19(2):461–476. [PubMed: 19887312]
5. 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.
6. Zhang Y, Dong B, Lu Z. ℓ_0 minimization for wavelet frame based image restoration. *Mathematics of Computation*. 2013; 82:995–1015.
7. Tuch DS. Q-ball imaging. *Magnetic Resonance in Medicine*. 2004; 52:1358–1372. [PubMed: 15562495]
8. Belkin M, Niyogi P. Towards a theoretical foundation for Laplacian-based manifold methods. *Learning Theory*. 2005; (486–500)
9. Lu Z, Zhang Y. Sparse approximation via penalty decomposition methods. *SIAM Journal on Optimization*. 2013; 23(4):2448–2478.
10. Lu Z. Iterative hard thresholding methods for ℓ_0 regularized convex cone programming. *Mathematical Programming: Series A and B*. 2014; 147(1–2):125–154.
11. Yap PT, Zhang Y, Shen D. Diffusion compartmentalization using response function groups with cardinality penalization. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. 2015; 9349:183–190.

12. Donoho DL. De-noising by soft-thresholding. IEEE Transactions on Information Theory. 1995; 41(3):613–627.
13. 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(108–119)

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Fig. 1.

Performance evaluation in terms of PSNR for two levels of noise, i.e., SNR=20 and 30. We have used here the notation $K' - K$, where K' and K are the numbers of input and output gradient directions, respectively. The standard errors are negligible and are hence not shown here.

Fig. 2. Performance statistics of the denoising of the real data consisting of 8 repeated sets of diffusion-weighted images.

Fig. 3.

Fiber ODFs of the low-angular-resolution data with 48 gradient directions and high-angular resolution data with 81 gradient directions. Visible improvements in resolving fiber crossings are marked with red arrows.

Table 1

Framelet masks.

Haar	Linear	Quadratic
$\hat{a}_0(\xi) = \cos(\xi/2)$	$\hat{a}_0(\xi) = \cos^2(\xi/2)$	$\hat{a}_0(\xi) = \cos^3(\xi/2)$
$\hat{a}_1(\xi) = \sin(\xi/2)$	$\hat{a}_1(\xi) = \frac{1}{\sqrt{2}} \sin(\xi)$	$\hat{a}_1(\xi) = \sqrt{3} \sin(\xi/2) \cos^2(\xi/2)$
	$\hat{a}_2(\xi) = \sin^2(\xi/2)$	$\hat{a}_2(\xi) = \sqrt{3} \sin^2(\xi/2) \cos(\xi/2)$
		$\hat{a}_3(\xi) = \sin^3(\xi/2)$

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript