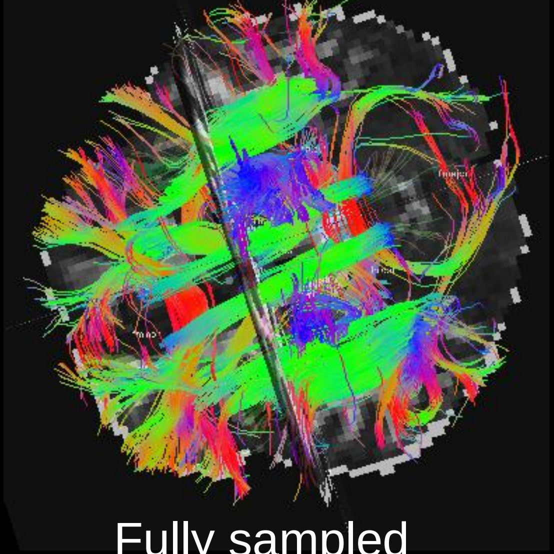
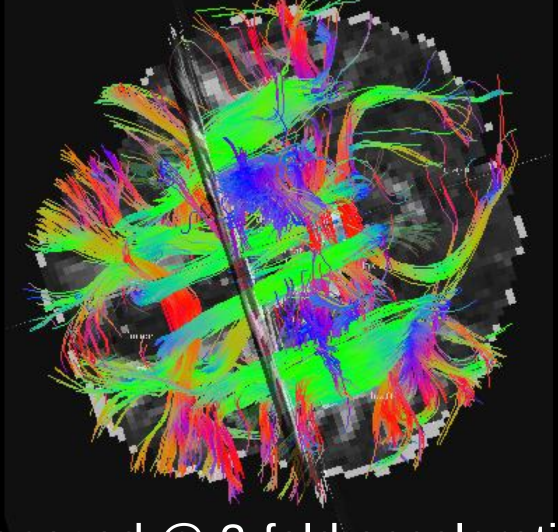


Fully sampled



Proposed @ 3-fold acceleration

Fast DSI Reconstruction with Trained Dictionaries

Berkin Bilgic[1], Itthi Chatnuntawech[2], Kawin Setsompop[1,3],
Stephen F. Cauley[1], Lawrence L. Wald[1,4], Elfar Adalsteinsson[2,4]

[1]Martinos Center for Biomedical Imaging, Charlestown, MA, USA
[2]MIT, Cambridge, MA USA
[3]Harvard Medical School, Boston, MA, USA
[4]Harvard-MIT Health Sciences and Technology, Cambridge, MA USA

ISMRM
21st Annual
Meeting & Exhibition
Salt Lake City, Utah, USA
20–26 April 2013
"Discovery, Innovation & Application – Advancing MR for Improved Health"

Declaration of Relevant Financial Interests or Relationships

Speaker Name: Berkin Bilgic

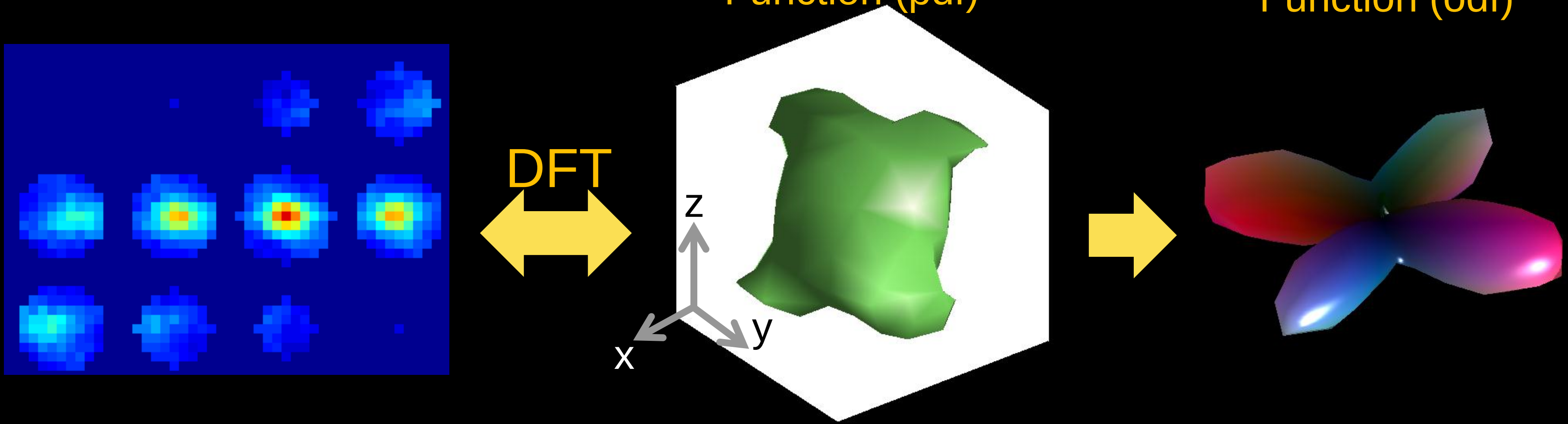
I have no relevant financial interest or relationship to disclose with regard to the subject matter of this presentation.

Diffusion Spectrum Imaging (DSI)

- DSI offers a complete description of water diffusion
- And reveals complex distributions of fiber orientations
- However, DSI requires full sampling of q-space

Diffusion Spectrum Imaging (DSI)

- DSI offers a complete description of water diffusion
- And reveals complex distributions of fiber orientations
- However, DSI requires full sampling of q-space



Q-space of a single voxel
515 directions

DFT

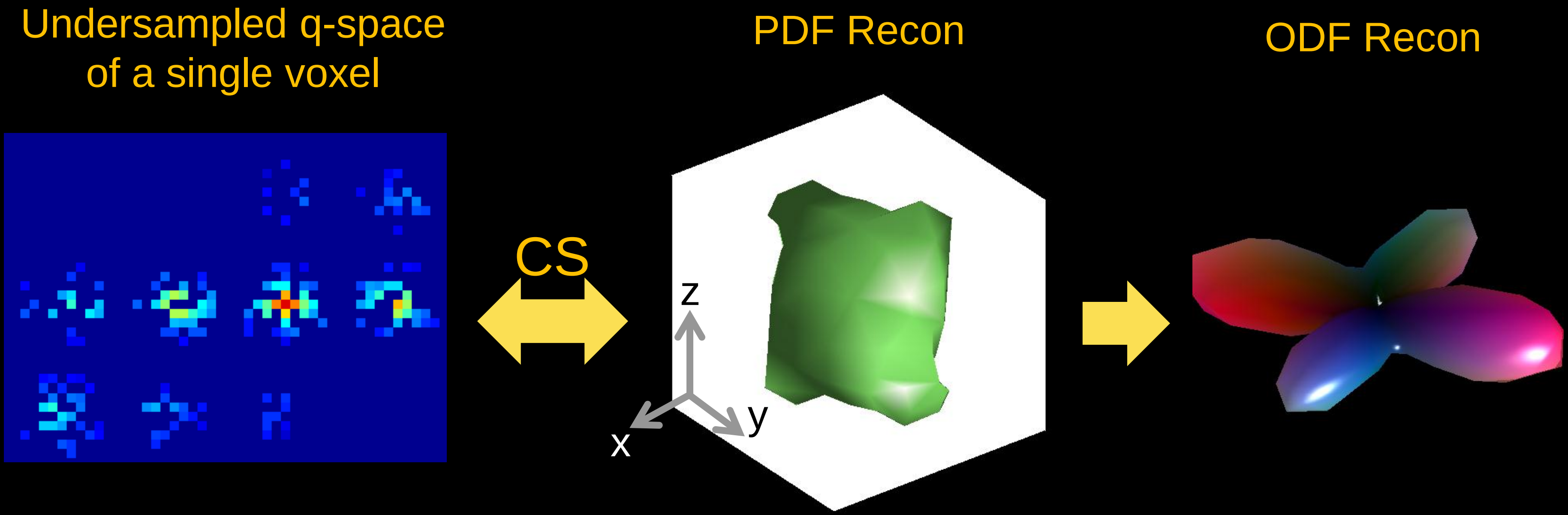
Probability Density
Function (pdf)

Orientation Distribution
Function (odf)

Sampling full q-space takes ~1 hour

Undersampled DSI

- To reduce scan time, undersample q-space
- Use sparsity prior to recon the pdfs via Compressed Sensing



Previous work on DSI recon

- To reduce scan time, undersample q-space
- Use sparsity prior to recon the pdfs via Compressed Sensing

i. **Wavelet & Total Variation [1]**

$$min_{\boldsymbol{p}} \, \|\underset{\substack{\swarrow \\ \text{undersampled} \\ \text{DFT}}}{\mathbf{F}_{\Omega}} \underset{\substack{\downarrow \\ \text{pdf}}}{\boldsymbol{p}} - \underset{\substack{\searrow \\ \text{q-samples}}}{\boldsymbol{q}}\|_2{}^2 + \alpha \cdot \|\underset{\substack{\downarrow \\ \text{wavelet}}}{\boldsymbol{\Psi}} \boldsymbol{p}\|_1 + \beta \cdot \underset{\substack{\downarrow \\ \text{total variation}}}{\mathrm{TV}}(\boldsymbol{p})$$

1. Menzel *et al* MRM 2011

Previous work on DSI recon

- To reduce scan time, undersample q-space
- Use sparsity prior to recon the pdfs via Compressed Sensing

i. **Wavelet & Total Variation [1]**

$$min_{\boldsymbol{p}} \|\mathbf{F}_{\Omega}\boldsymbol{p} - \boldsymbol{q}\|_2{}^2 + \alpha \cdot \|\boldsymbol{\Psi}\boldsymbol{p}\|_1 + \beta \cdot \mathrm{TV}(\boldsymbol{p})$$

ii. Our previous work: **Dictionary-FOCUSS [2]**

❑ Create a dictionary $\mathbf{D}$ from a training dataset of pdfs using K-SVD algorithm [3] → tailored for sparse representation

❑ Impose sparsity constraint via FOCUSS algorithm [4] by solving

$$min\|\boldsymbol{x}\|_1 \quad \text{such that} \quad \mathbf{F}_{\Omega}\mathbf{D}\boldsymbol{x} = \boldsymbol{q}$$

1. Menzel *et al* MRM 2011
2. Bilgic *et al* MRM 2012
3. Aharon *et al* IEEE TSP 2006
4. Gorodnitsky *et al* IEEE TSP 1997

Previous work on DSI recon

- To reduce scan time, undersample q-space
- Use sparsity prior to recon the pdfs via Compressed Sensing

i. **Wavelet & Total Variation [1]**

$$min_{\boldsymbol{p}} \|\mathbf{F}_{\Omega}\boldsymbol{p} - \boldsymbol{q}\|_2{}^2 + \alpha \cdot \|\boldsymbol{\Psi}\boldsymbol{p}\|_1 + \beta \cdot \mathrm{TV}(\boldsymbol{p})$$

ii. Our previous work: **Dictionary-FOCUSS [2]**

- ☐ Create a dictionary $\mathbf{D}$ from a training dataset of pdfs using K-SVD algorithm [3] → tailored for sparse representation
- ☐ Impose sparsity constraint via FOCUSS algorithm [4] by solving

$$min\|\boldsymbol{x}\|_1 \quad \text{such that} \quad \mathbf{F}_{\Omega}\mathbf{D}\boldsymbol{x} = \boldsymbol{q}$$

Dictionary transform coefficients

Previous work on DSI recon

- Dictionary-FOCUSS [1] yields up to 2-times RMSE reduction using compared to Wavelet & TV [2]

1. Bilgic *et al* MRM 2012
2. Menzel *et al* MRM 2011

Previous work on DSI recon

- Dictionary-FOCUSS [1] yields up to 2-times RMSE reduction using compared to Wavelet & TV [2]

- Both Compressed Sensing recons are iterative, with processing times up to 1 sec / voxel

1. Bilgic *et al* MRM 2012
2. Menzel *et al* MRM 2011

Previous work on DSI recon

- Dictionary-FOCUSS [1] yields up to 2-times RMSE reduction using compared to Wavelet & TV [2]

- Both Compressed Sensing recons are iterative, with processing times up to 1 sec / voxel

- Full-brain recon for 10^5 voxels: ~ **1 DAY of computation**

1. Bilgic *et al* MRM 2012
2. Menzel *et al* MRM 2011

Proposed methods

- We propose two L2-based methods:

i. **Proposed I: Principal Component Analysis (PCA)**
   - Summarize the training dataset with dominant eigenvectors
   - Simple training and recon : linear algebra

ii. **Proposed II: Dictionary-L2**
   - Instead of L1-, apply L2-regularization wrt dictionary
   - Fast recon with closed form solution

Proposed methods

- We propose two L2-based methods:

i. **Proposed I: Principal Component Analysis (PCA)**

ii. **Proposed II: Dictionary-L2**

- **Contributions**:
  - The two proposed methods are **100-fold faster** with image quality similar to Dictionary-FOCUSS

Proposed methods

- We propose two L2-based methods:

i. **Proposed I: Principal Component Analysis (PCA)**

ii. **Proposed II: Dictionary-L2**

- **Contributions**:
  - The two proposed methods are **100-fold faster** with image quality similar to Dictionary-FOCUSS
  - We show that crucial component is presence of dictionary, *not the sparsity constraint*

Proposed methods

- We propose two L2-based methods:

i. **Proposed I: Principal Component Analysis (PCA)**

ii. **Proposed II: Dictionary-L2**

- **Contributions**:
  - The two proposed methods are **100-fold faster** with image quality similar to Dictionary-FOCUSS
  - We show that crucial component is presence of dictionary, *not the sparsity constraint*
  - At 3-fold acceleration, proposed recons comparable to fully-sampled data in pdf, odf and fiber domains

Proposed I: PCA Reconstruction

- PCA: approximates data points using a linear combo of them to retain the maximum variance in the dataset

Proposed I: PCA Reconstruction

- PCA: approximates data points using a linear combo of them to retain the maximum variance in the dataset

- Start with a training set of pdfs $\mathbf{P}$
- Subtract the mean, diagonalize the covariance matrix:

$$\mathbf{Z} = \mathbf{P} - \boldsymbol{p}_{mean}$$

$$\mathbf{ZZ}^H = \mathbf{Q\Lambda Q}^H$$

Proposed I: PCA Reconstruction

- PCA: approximates data points using a linear combo of them to retain the maximum variance in the dataset

- Start with a training set of pdfs $\mathbf{P}$
- Subtract the mean, diagonalize the covariance matrix:

$$\mathbf{Z} = \mathbf{P} - \boldsymbol{p}_{mean}$$

$$\mathbf{Z}\mathbf{Z}^{H} = \mathbf{Q}\Lambda\mathbf{Q}^{H}$$

- Pick the first T columns of $\mathbf{Q}$ corresponding to largest eigvals: $\mathbf{Q}_T$

$$\underset{\substack{\downarrow \\ T\text{ - dimensional} \\ \text{pca coefficients}}}{\boldsymbol{pca}} = \mathbf{Q}_T^H(\boldsymbol{p} - \boldsymbol{p}_{mean})$$

Proposed I: PCA Reconstruction

- PCA: approximates data points using a linear combo of them to retain the maximum variance in the dataset

- Start with a training set of pdfs $\mathbf{P}$
- Subtract the mean, diagonalize the covariance matrix:

$$\mathbf{Z} = \mathbf{P} - \boldsymbol{p}_{mean}$$

$$\mathbf{Z}\mathbf{Z}^{H} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^{H}$$

- Pick the first T columns of $\mathbf{Q}$ corresponding to largest eigvals: $\mathbf{Q}_{T}$

$$\boldsymbol{pca} = \mathbf{Q}_{T}^{H}(\boldsymbol{p} - \boldsymbol{p}_{mean})$$

- The location of $\boldsymbol{pca}$ in the pdf space,

$$\boldsymbol{p}_{T} = \mathbf{Q}_{T}\boldsymbol{pca} + \boldsymbol{p}_{mean}$$

Proposed I: PCA Reconstruction

- PCA: approximates data points using a linear combo of them to retain the maximum variance in the dataset
- Least-squares approximation in T - dimensions,

$$min\ ||\mathbf{F}_{\Omega}\boldsymbol{p}_{\boldsymbol{T}} - \boldsymbol{q}||_2{}^2$$

Proposed I: PCA Reconstruction

- PCA: approximates data points using a linear combo of them to retain the maximum variance in the dataset

- Least-squares approximation in T - dimensions,

$$min \ \|\mathbf{F}_{\Omega}\boldsymbol{p}_{\boldsymbol{T}} - \boldsymbol{q}\|_2{}^2$$

- In PCA coordinates,

$$min_{\boldsymbol{pca}} \ \|\mathbf{F}_{\Omega}\mathbf{Q}_T\boldsymbol{pca} - (\boldsymbol{q} - \mathbf{F}_{\Omega}\boldsymbol{p}_{mean})\|_2^2$$

Proposed I: PCA Reconstruction

- PCA: approximates data points using a linear combo of them to retain the maximum variance in the dataset

- Least-squares approximation in T - dimensions,

$$min\ \|\mathbf{F}_\Omega \boldsymbol{p}_{\boldsymbol{T}} - \boldsymbol{q}\|_2{}^2$$

- In PCA coordinates,

$$min_{\boldsymbol{pca}}\ \|\mathbf{F}_\Omega \mathbf{Q}_T \boldsymbol{pca} - (\boldsymbol{q} - \mathbf{F}_\Omega \boldsymbol{p}_{mean})\|_2^2$$

- Closed-form solution:

$$\widetilde{\boldsymbol{pca}} = \underbrace{\mathrm{pinv}(\mathbf{F}_\Omega \mathbf{Q}_T)}_{\text{compute once}}(\boldsymbol{q} - \mathbf{F}_\Omega \boldsymbol{p}_{mean})$$

Proposed II: Dictionary-L2

- Dictionary-FOCUSS iteratively solves

$$min\|\boldsymbol{x}\|_1 \quad \text{such that} \quad \mathbf{F}_\Omega\mathbf{D}\boldsymbol{x} = \boldsymbol{q}$$

- Instead, consider

$$min\ \|\mathbf{F}_\Omega\mathbf{D}\boldsymbol{x} - \boldsymbol{q}\|_2{}^2 + \lambda \cdot \|\boldsymbol{x}\|_{\mathbf{2}}{}^{\mathbf{2}}$$

Proposed II: Dictionary-L2

- Dictionary-FOCUSS iteratively solves

$$min\|\boldsymbol{x}\|_1 \quad \text{such that} \quad \mathbf{F}_{\Omega}\mathbf{D}\boldsymbol{x} = \boldsymbol{q}$$

- Instead, consider

$$min\ \|\mathbf{F}_{\Omega}\mathbf{D}\boldsymbol{x} - \boldsymbol{q}\|_2{}^2 + \lambda \cdot \|\boldsymbol{x}\|_{\mathbf{2}}{}^{\mathbf{2}}$$

- Solution: $\widetilde{\boldsymbol{x}} = ((\mathbf{F}_{\Omega}\mathbf{D})^H\mathbf{F}_{\Omega}\mathbf{D} + \lambda\mathbf{I})^{-1}(\mathbf{F}_{\Omega}\mathbf{D})^H q$

Proposed II: Dictionary-L2

- Dictionary-FOCUSS iteratively solves

$$min\|\boldsymbol{x}\|_1 \quad \text{such that} \quad \mathbf{F}_\Omega\mathbf{D}\boldsymbol{x} = \boldsymbol{q}$$

- Instead, consider

$$min\ \|\mathbf{F}_\Omega\mathbf{D}\boldsymbol{x} - \boldsymbol{q}\|_2{}^2 + \lambda \cdot \|\boldsymbol{x}\|_{\mathbf{2}}{}^{\mathbf{2}}$$

- Solution: $\widetilde{\boldsymbol{x}} = ((\mathbf{F}_\Omega\mathbf{D})^H\mathbf{F}_\Omega\mathbf{D} + \lambda\mathbf{I})^{-1}(\mathbf{F}_\Omega\mathbf{D})^H q$

Singular Value Decomposition: $\mathbf{F}_\Omega\mathbf{D} = \mathbf{U\Sigma}\boldsymbol{V}^H$

Proposed II: Dictionary-L2

- Dictionary-FOCUSS iteratively solves

$$min\|\boldsymbol{x}\|_1 \quad \text{such that} \quad \mathbf{F}_\Omega \mathbf{D}\boldsymbol{x} = \boldsymbol{q}$$

- Instead, consider

$$min\ \|\mathbf{F}_\Omega \mathbf{D}\boldsymbol{x} - \boldsymbol{q}\|_2{}^2 + \lambda \cdot \|\boldsymbol{x}\|_2{}^{\mathbf{2}}$$

- Solution: $\tilde{\boldsymbol{x}} = ((\mathbf{F}_\Omega \mathbf{D})^H \mathbf{F}_\Omega \mathbf{D} + \lambda \mathbf{I})^{-1} (\mathbf{F}_\Omega \mathbf{D})^H q$

$$\mathbf{F}_\Omega \mathbf{D} = \mathbf{U\Sigma}\boldsymbol{V}^H$$

$$\mathbf{\Sigma}^+ = (\mathbf{\Sigma}^H \mathbf{\Sigma} + \lambda \mathbf{I})^{-1} \mathbf{\Sigma}^H$$

$$\tilde{\boldsymbol{x}} = \underbrace{\mathbf{V\Sigma}^+\mathbf{U}^H}_{\text{compute once}} q$$

Proposed II: Dictionary-L2

- Dictionary-FOCUSS iteratively solves

$$min||\boldsymbol{x}||_1 \quad \text{such that} \quad \mathbf{F}_\Omega \mathbf{D}\boldsymbol{x} = \boldsymbol{q}$$

Dictionary $\mathbf{D}$ obtained by K-SVD training

- Solution: $\widetilde{\boldsymbol{x}} = ((\mathbf{F}_\Omega \mathbf{D})^H \mathbf{F}_\Omega \mathbf{D} + \lambda \mathbf{I})^{-1} (\mathbf{F}_\Omega \mathbf{D})^H q$

Proposed II: Dictionary-L2

- Dictionary-FOCUSS iteratively solves

$$min\|\boldsymbol{x}\|_1 \quad \text{such that} \quad \mathbf{F}_\Omega \mathbf{D}\boldsymbol{x} = \boldsymbol{q}$$

Dictionary $\mathbf{D}$ obtained by K-SVD training

- Solution: $\widetilde{\boldsymbol{x}} = ((\mathbf{F}_\Omega \mathbf{D})^H \mathbf{F}_\Omega \mathbf{D} + \lambda \mathbf{I})^{-1} (\mathbf{F}_\Omega \mathbf{D})^H q$

Proposed: use training dataset as dictionary, $\mathbf{D} = \mathbf{P}$

Both recon and training simplified

DSI Acquisition

- 2.3 mm isotropic with b_{max} = 8000 s/mm^2 at 3T
- Connectom gradients and 64-chan head coil [1]
- 515 q-space points collected in 50 min
- Two subjects scanned → dictionary training is based on a subject different from the test subject

1. Keil *et al* MRM 2012

Comparison of methods

- Previous methods:
    i. CDF 9/7 Wavelet & TV [1,2]
    ii. Dictionary-FOCUSS [3]
- New methods:
    iii. Proposed I: PCA
    iv. Proposed II:Dictionary-L2

1. Menzel *et al* MRM 2011
2. Merlet et al ISMRM 2012
3. Bilgic *et al* MRM 2012

Comparison of methods

- Previous methods:
    1. CDF 9/7 Wavelet & TV [1,2]
    2. Dictionary-FOCUSS [3]
- New methods:
    3. **Proposed I: PCA**
    4. **Proposed II:Dictionary-L2**
- Recon experiments at accelerations R = 3, 5 and 9
- Compare to fully-sampled in terms of pdf, odf and fiber tracts

1. Menzel *et al* MRM 2011
2. Merlet et al ISMRM 2012
3. Bilgic *et al* MRM 2012

Comparison of methods

- Previous methods:
  i. CDF 9/7 Wavelet & TV [1,2]
  ii. Dictionary-FOCUSS [3]

- New methods:
  iii. Proposed I: PCA
  iv. Proposed II:Dictionary-L2

- Comparison to low-noise data:
  - Acquire 10 average data at 5 q-space points

1 avg

10 avg

  - Disentangle RMSE due to recon error and noise

PDF Recon Error Maps

Test data: Subject A, Slice 40

Training data: Subject B, Slice 30

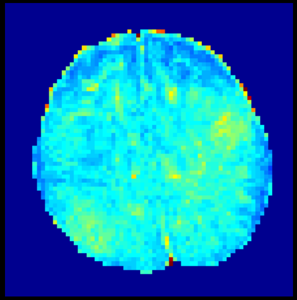
Acceleration
R = 3

Wavelet & TV

15.9% avg RMSE

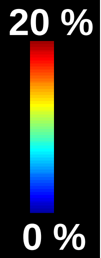
35 min

Dictionary-FOCUSS



7.6% avg RMSE

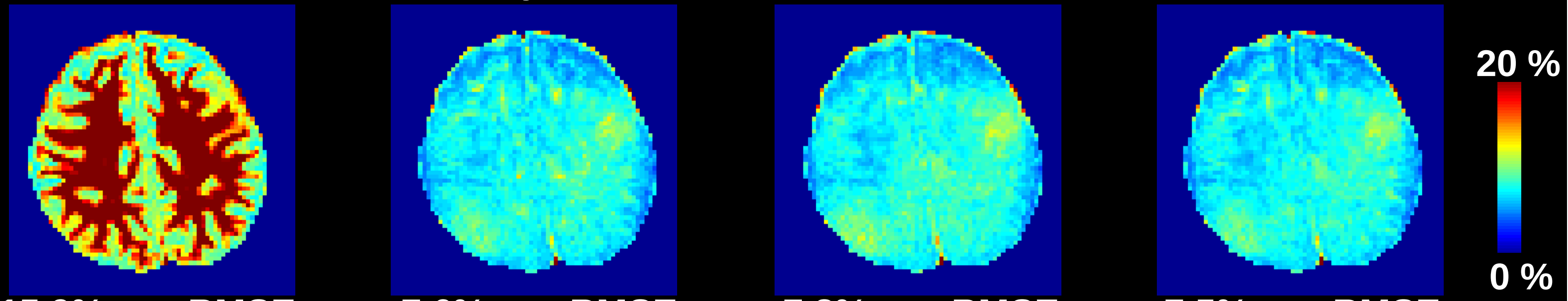
11 min



PDF Recon Error Maps

Test data: Subject A, Slice 40

Training data: Subject B, Slice 30



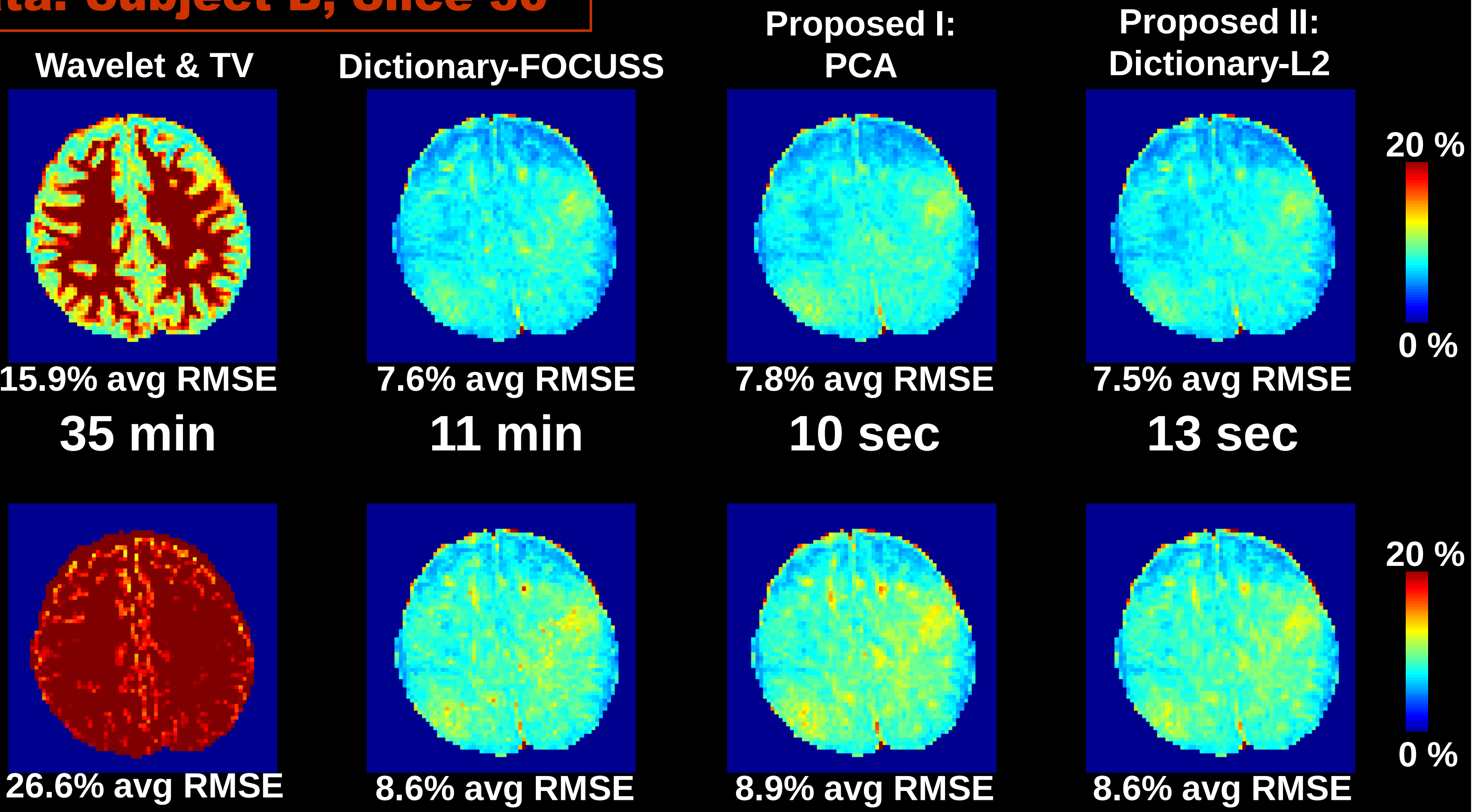
Acceleration
R = 3

Wavelet & TV	Dictionary-FOCUSS	Proposed I: PCA	Proposed II: Dictionary-L2
15.9% avg RMSE	7.6% avg RMSE	7.8% avg RMSE	7.5% avg RMSE
35 min	11 min	10 sec	13 sec

PDF Recon Error Maps

Test data: Subject A, Slice 40

Training data: Subject B, Slice 30

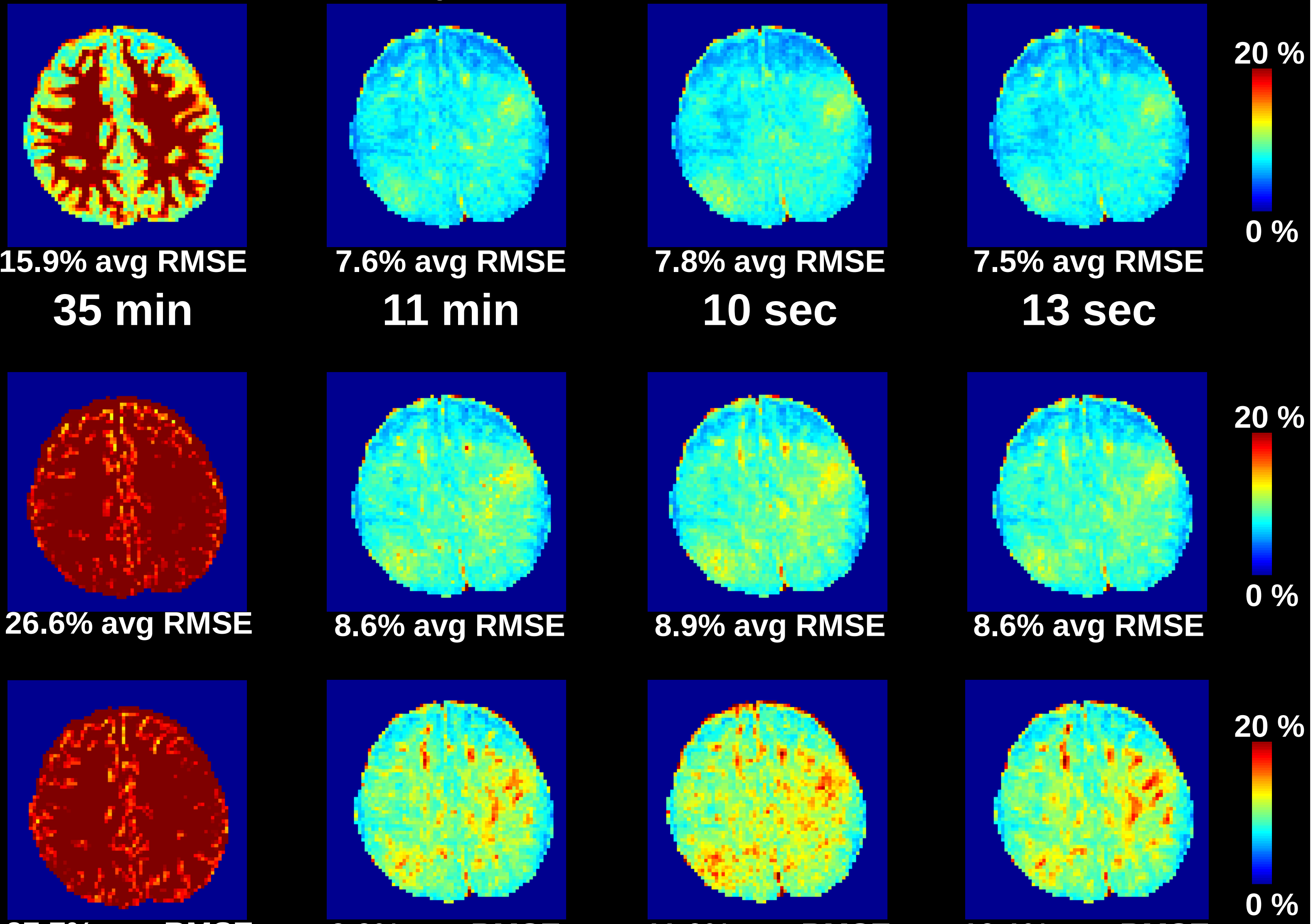


	Wavelet & TV	Dictionary-FOCUSS	Proposed I: PCA	Proposed II: Dictionary-L2
Acceleration R = 3	15.9% avg RMSE 35 min	7.6% avg RMSE 11 min	7.8% avg RMSE 10 sec	7.5% avg RMSE 13 sec
Acceleration R = 5	26.6% avg RMSE	8.6% avg RMSE	8.9% avg RMSE	8.6% avg RMSE

PDF Recon Error Maps

Test data: Subject A, Slice 40

Training data: Subject B, Slice 30



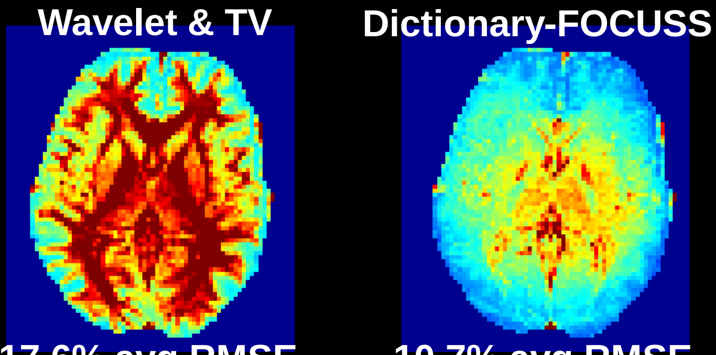
	Wavelet & TV	Dictionary-FOCUSS	Proposed I: PCA	Proposed II: Dictionary-L2
Acceleration R = 3	15.9% avg RMSE 35 min	7.6% avg RMSE 11 min	7.8% avg RMSE 10 sec	7.5% avg RMSE 13 sec
Acceleration R = 5	26.6% avg RMSE	8.6% avg RMSE	8.9% avg RMSE	8.6% avg RMSE
Acceleration R = 9	27.7% avg RMSE	9.9% avg RMSE	11.0% avg RMSE	10.1% avg RMSE

PDF Recon Error Maps

Test data: Subject B, Slice 25

Training data: Subject A, Slice 30

Acceleration
R = 3



Wavelet & TV

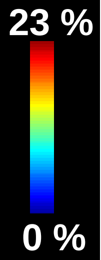
17.6% avg RMSE

43 min

Dictionary-FOCUSS

10.7% avg RMSE

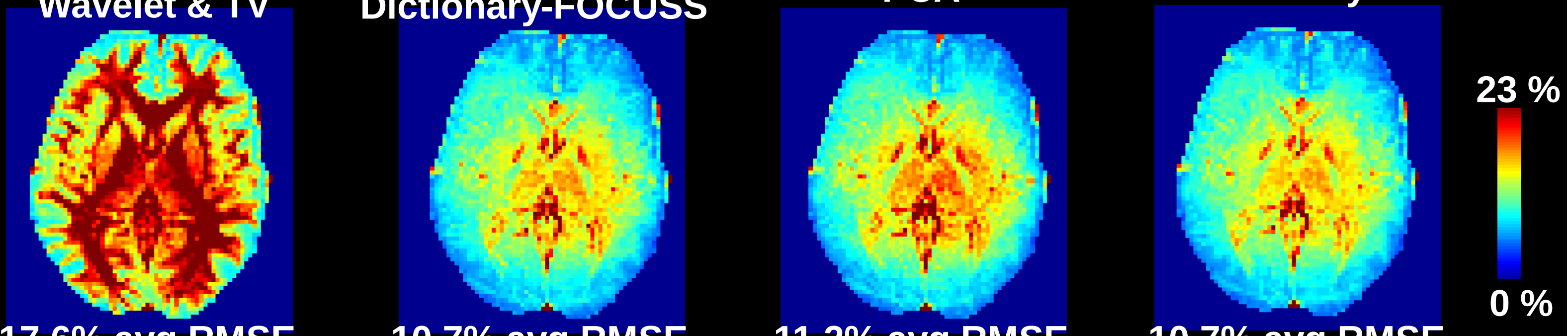
13 min



PDF Recon Error Maps

Test data: Subject B, Slice 25

Training data: Subject A, Slice 30



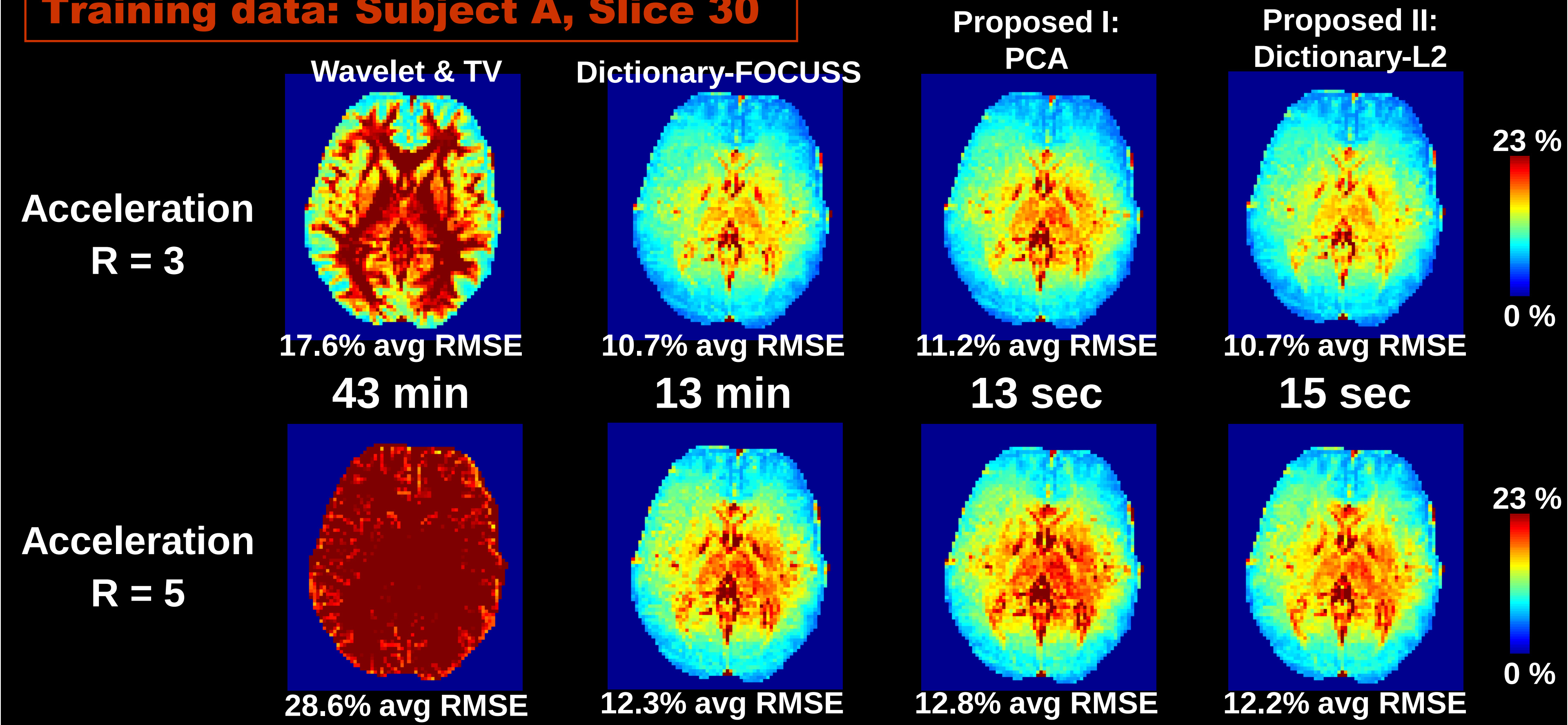
Acceleration R = 3

Wavelet & TV	Dictionary-FOCUSS	Proposed I: PCA	Proposed II: Dictionary-L2
17.6% avg RMSE	10.7% avg RMSE	11.2% avg RMSE	10.7% avg RMSE
43 min	13 min	13 sec	15 sec

PDF Recon Error Maps

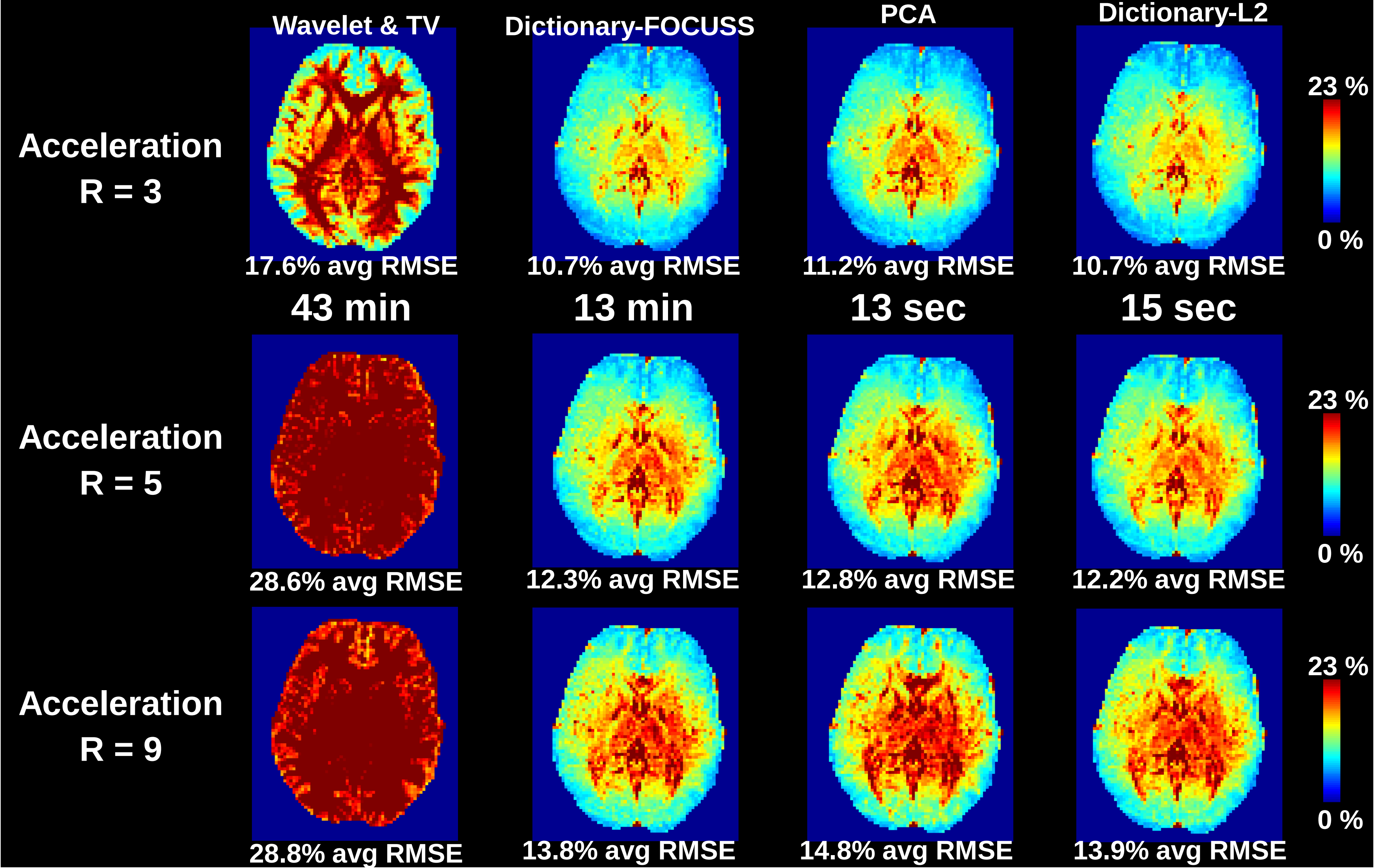
Test data: Subject B, Slice 25

Training data: Subject A, Slice 30



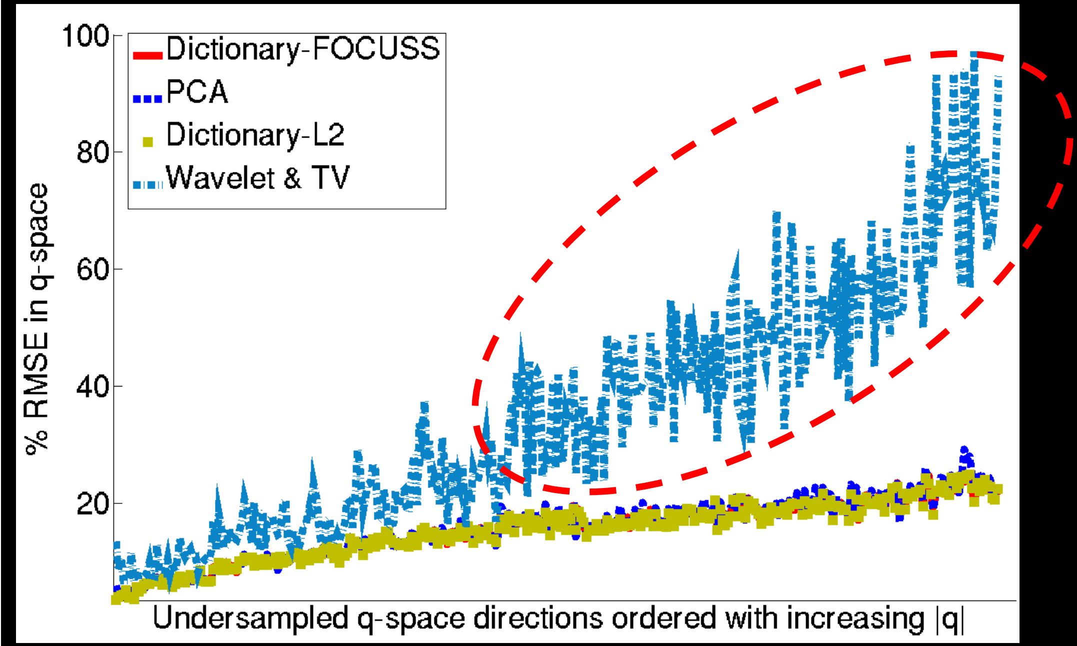
PDF Recon Error Maps

Test data: Subject B, Slice 25
Training data: Subject A, Slice 30



Q-space recon error

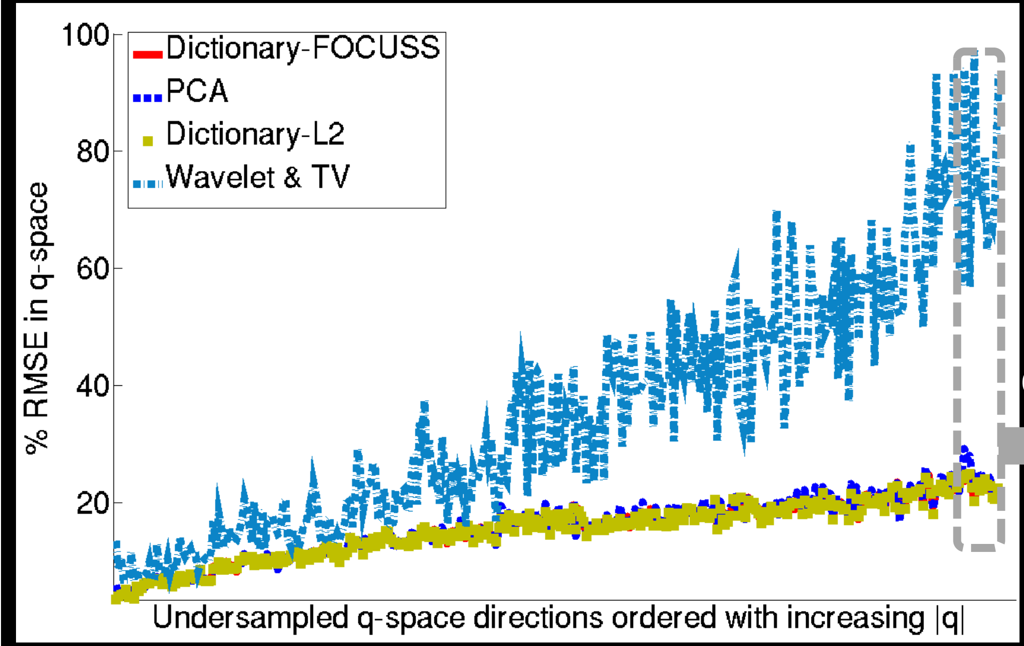
- Compute RMSE in missing q-space relative to fully-sampled data

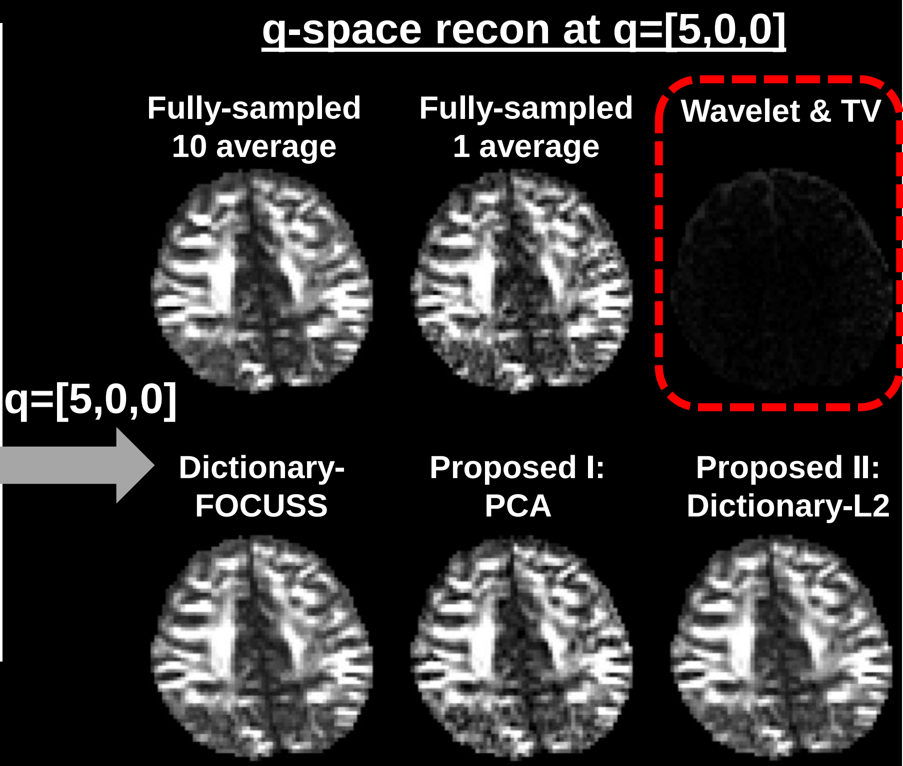


- Wavelet & TV : large error at outer q-space
- Dictionary-based methods : mild increase in error

Q-space recon error

- Compute RMSE in missing q-space relative to fully-sampled data

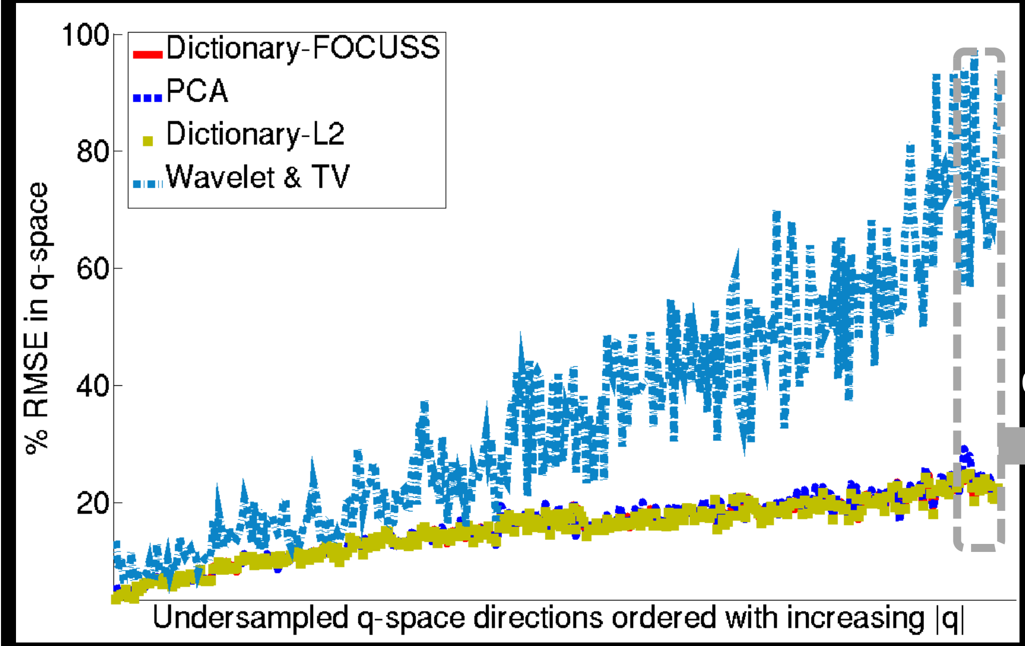




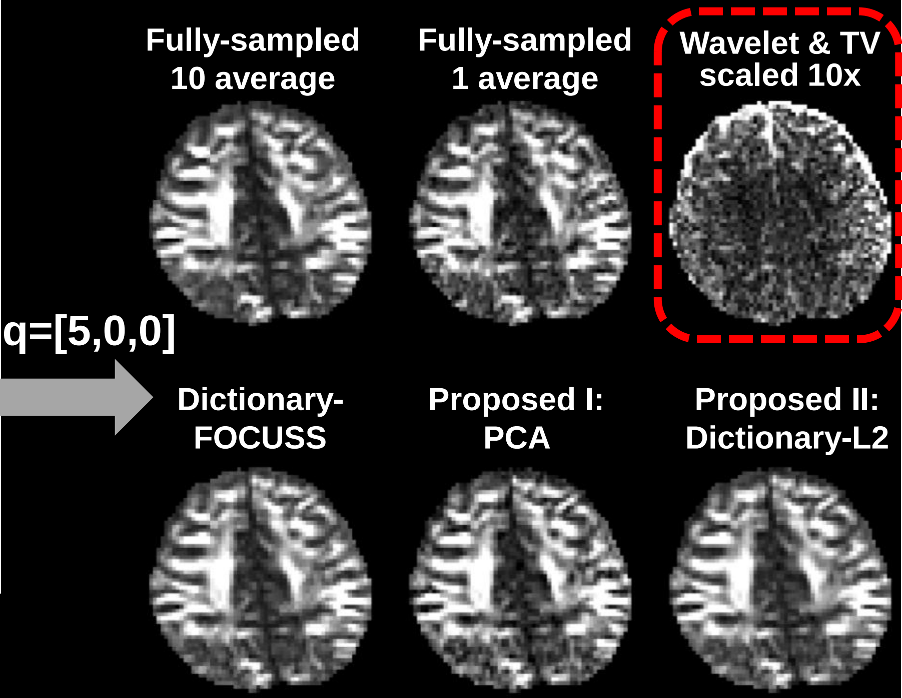
- Wavelet & TV : large error at outer q-space
- Dictionary-based methods : mild increase in error

Q-space recon error

- Compute RMSE in missing q-space relative to fully-sampled data



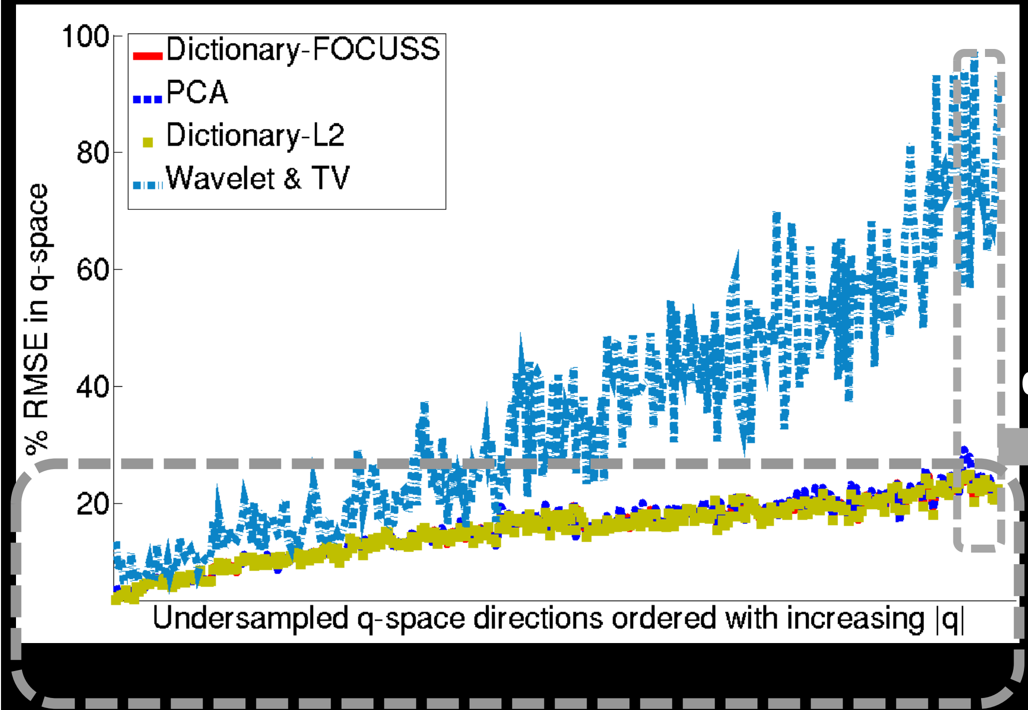
q-space recon at q=[5,0,0]



- Wavelet & TV : large error at outer q-space
- Dictionary-based methods : mild increase in error

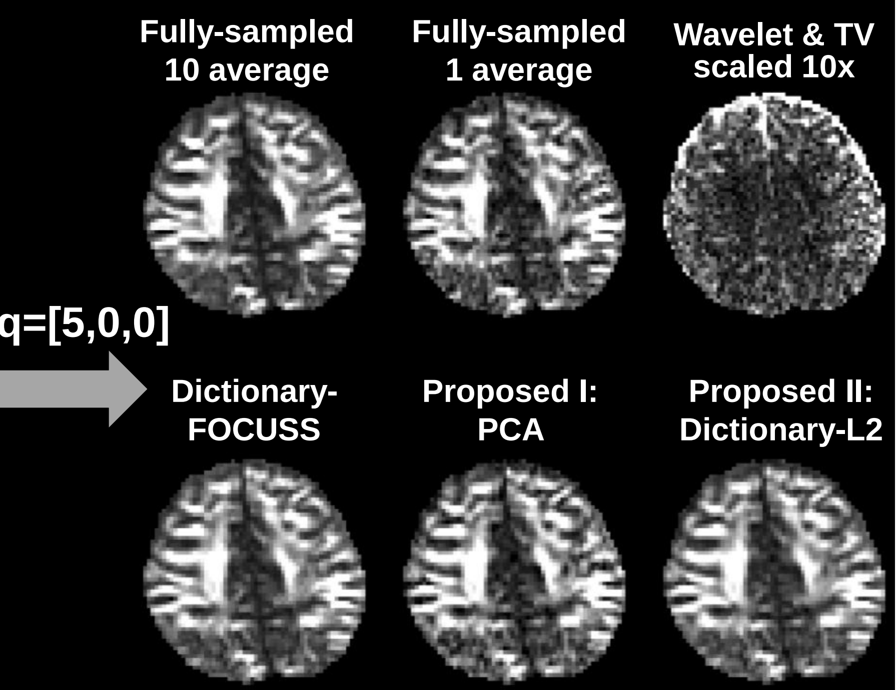
Q-space recon error

- Compute RMSE in missing q-space relative to fully-sampled data



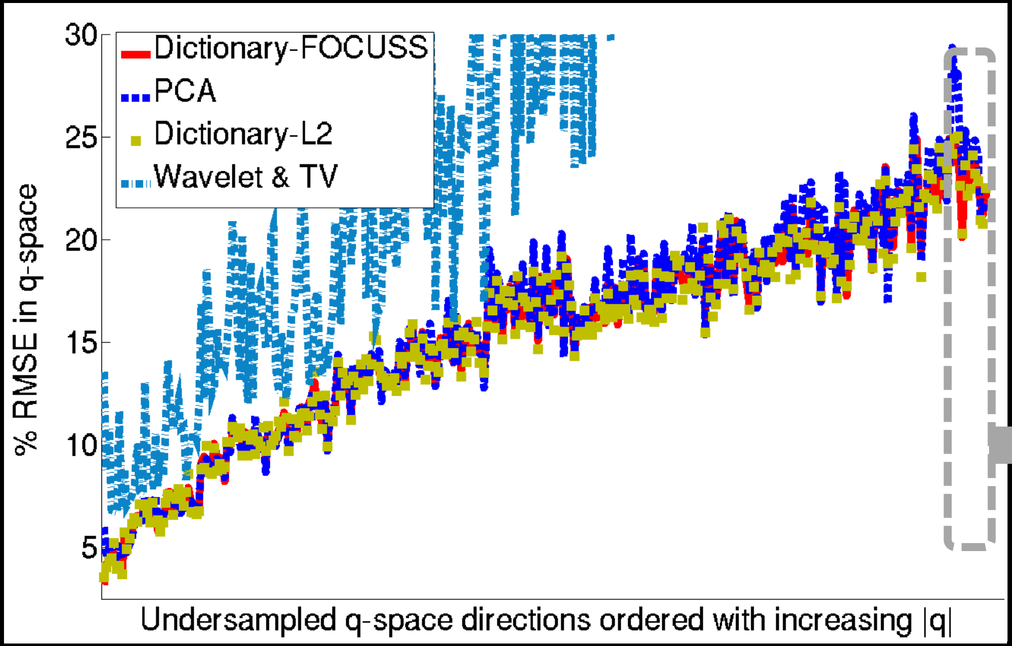
magnify

q-space recon at q=[5,0,0]

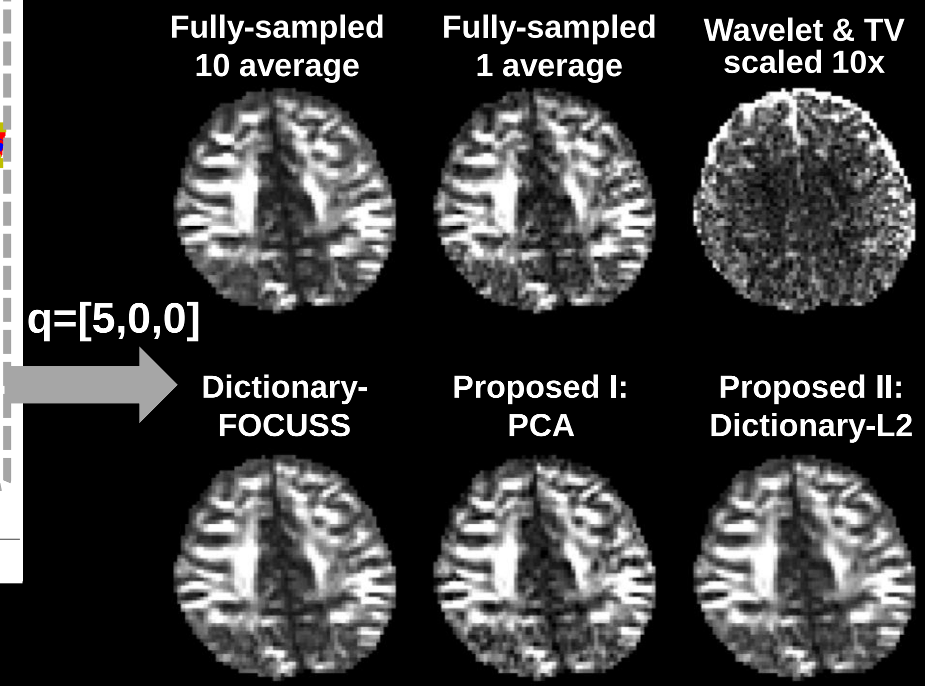


Q-space recon error

- Compute RMSE in missing q-space relative to fully-sampled data



q-space recon at q=[5,0,0]



Fully-sampled 10 average

Fully-sampled 1 average

Wavelet & TV scaled 10x

Dictionary-FOCUSS

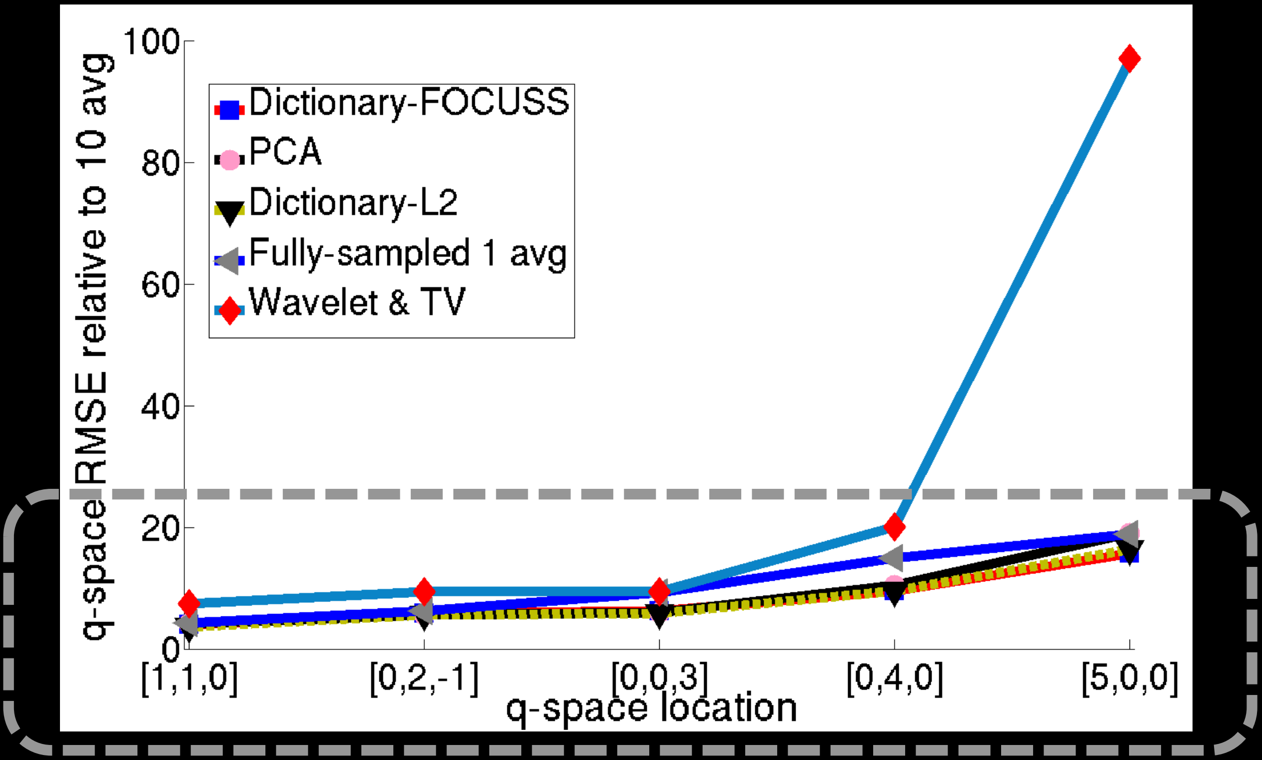
Proposed I: PCA

Proposed II: Dictionary-L2

- Dictionary-FOCUSS and the proposed methods have comparable performance

Comparison to Low-Noise dataset

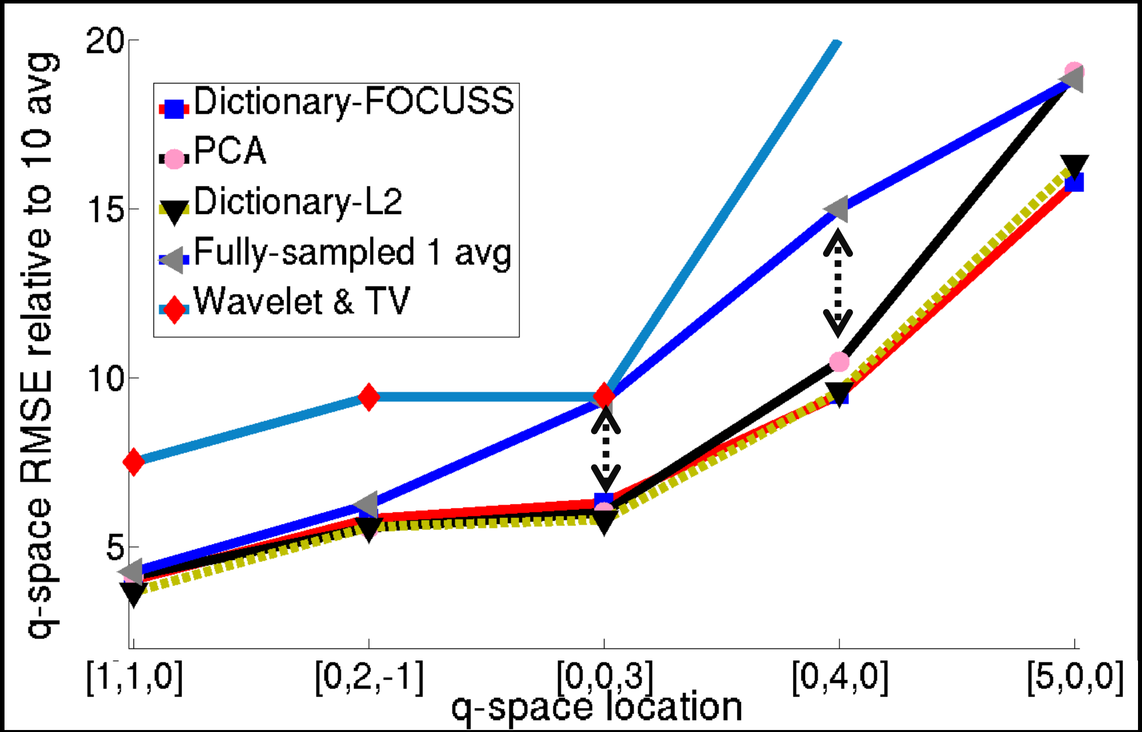
- Compute RMSEs relative to 10 average fully-sampled data



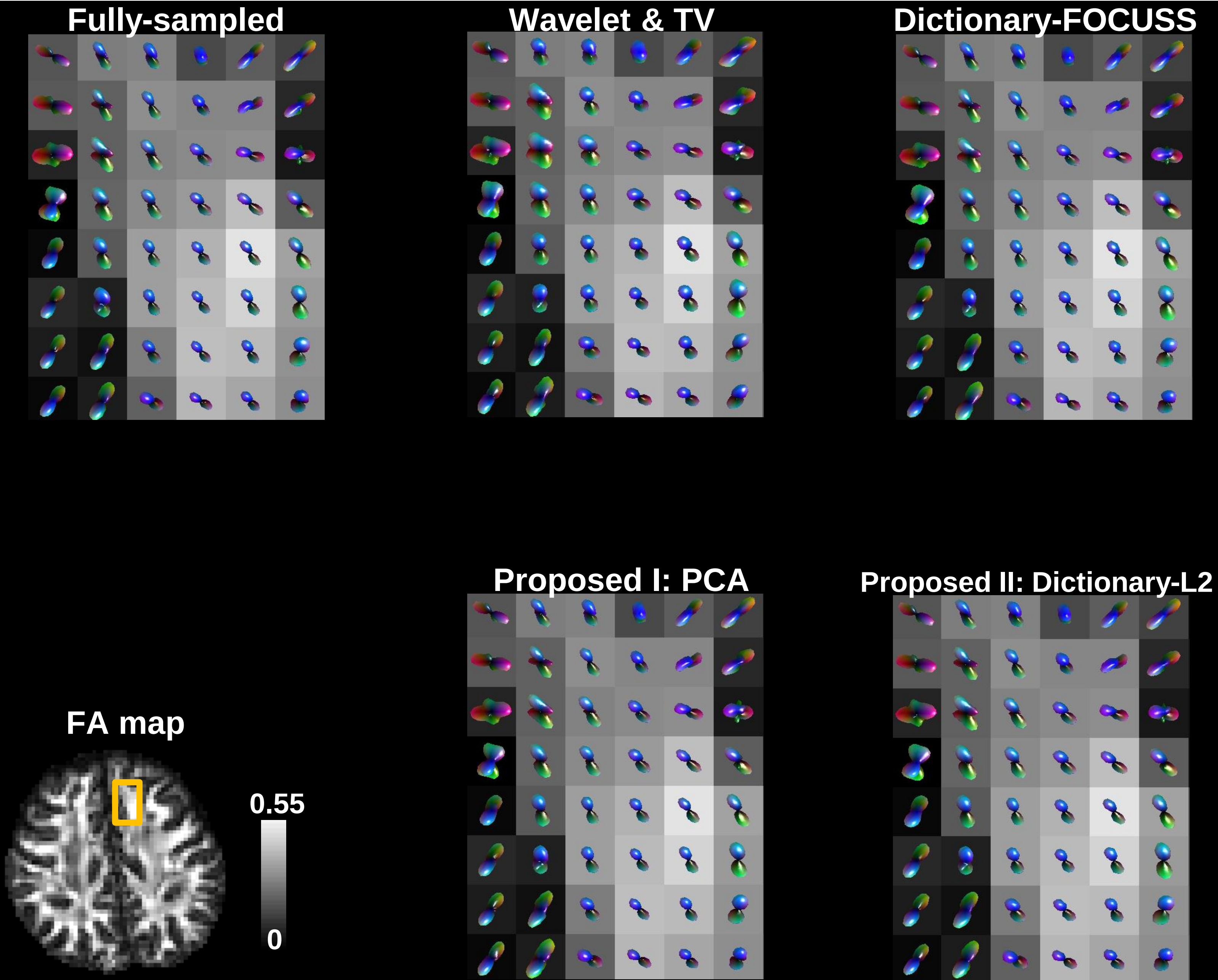
magnify

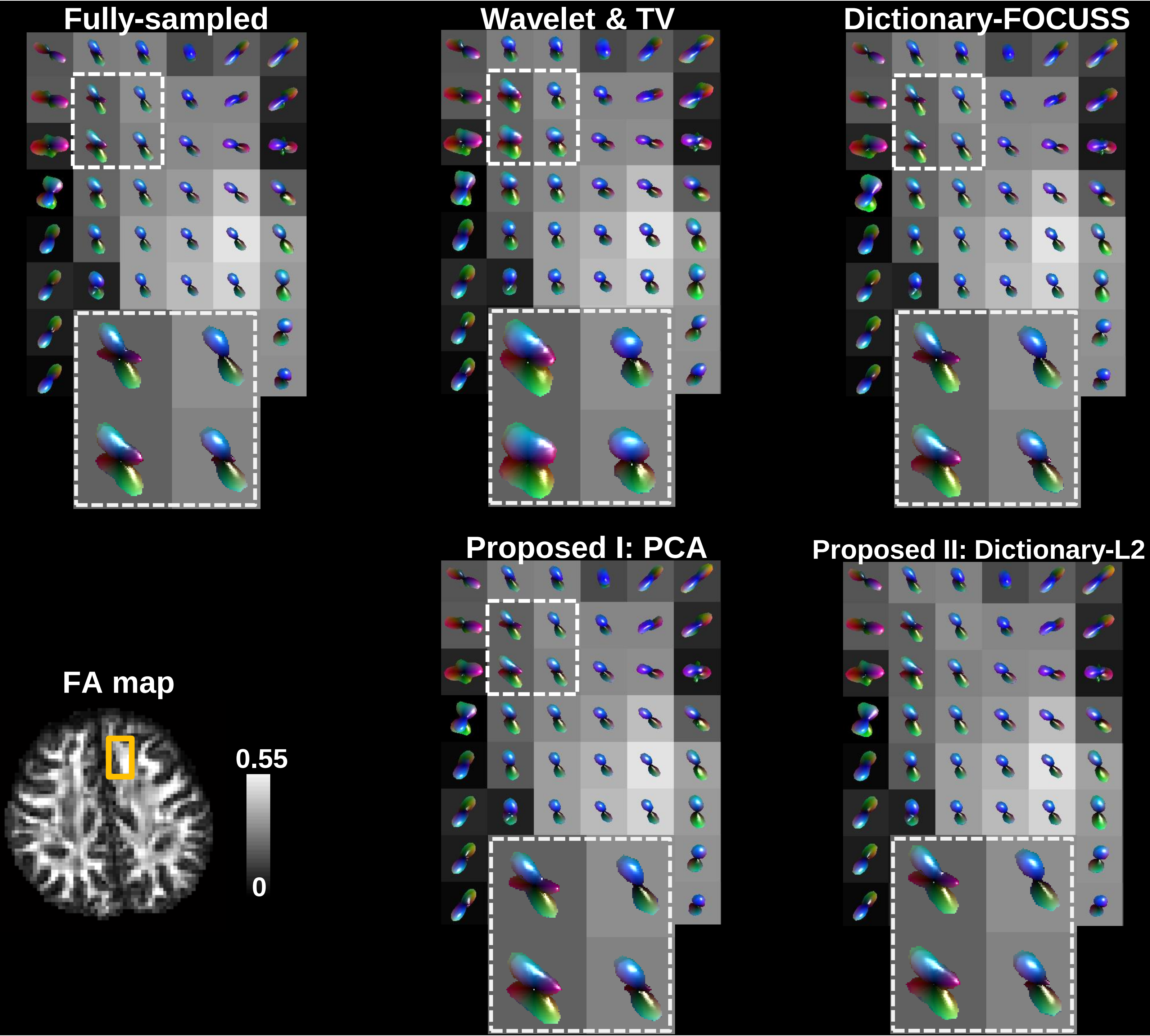
Comparison to Low-Noise dataset

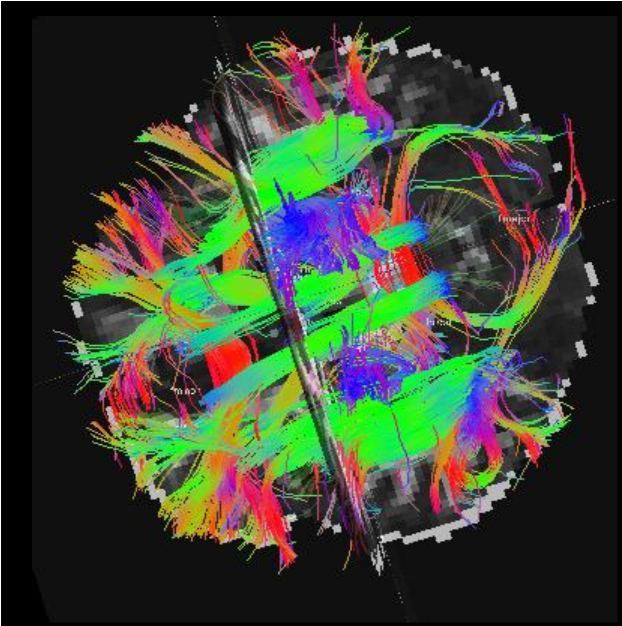
- Compute RMSEs relative to 10 average fully-sampled data



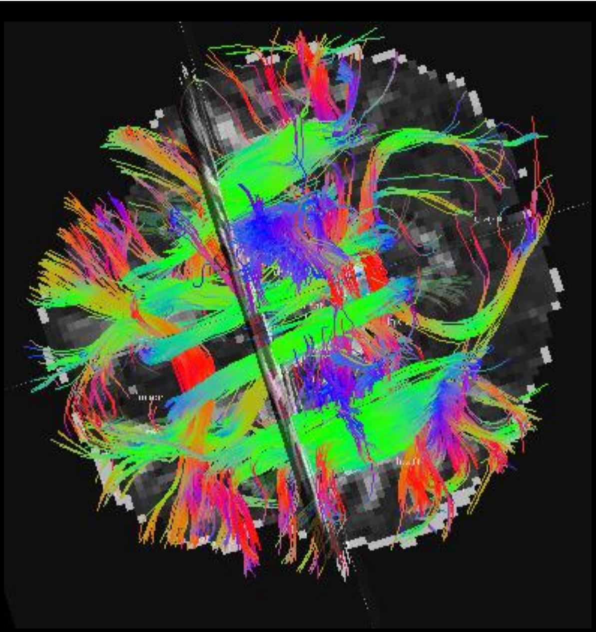
- At R = 3, all dictionary-based methods have less error than 1 average fully-sampled data


Fully-sampled
Wavelet & TV
Dictionary-FOCUSS
FA map
0.55
0
Proposed I: PCA
Proposed II: Dictionary-L2

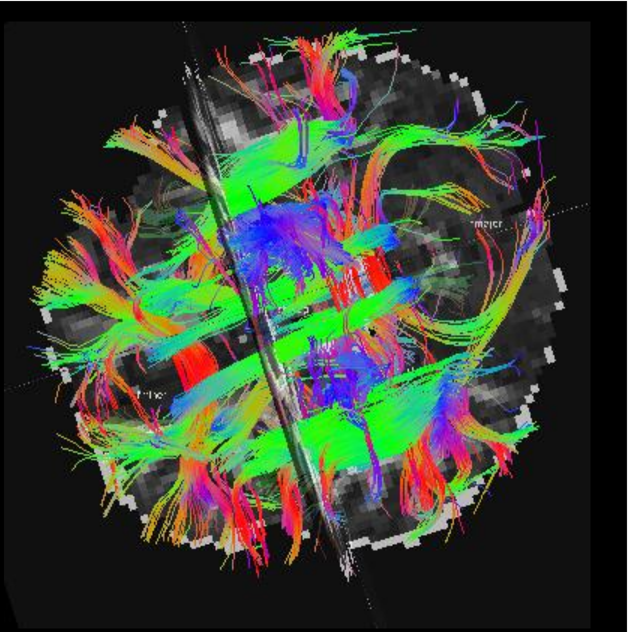

Fully-sampled
Wavelet & TV
Dictionary-FOCUSS
FA map
0.55
0
Proposed I: PCA
Proposed II: Dictionary-L2



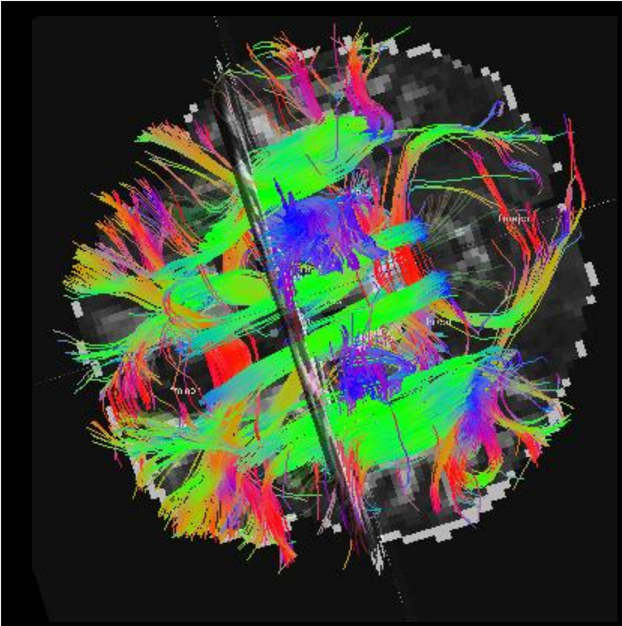
Fully-sampled



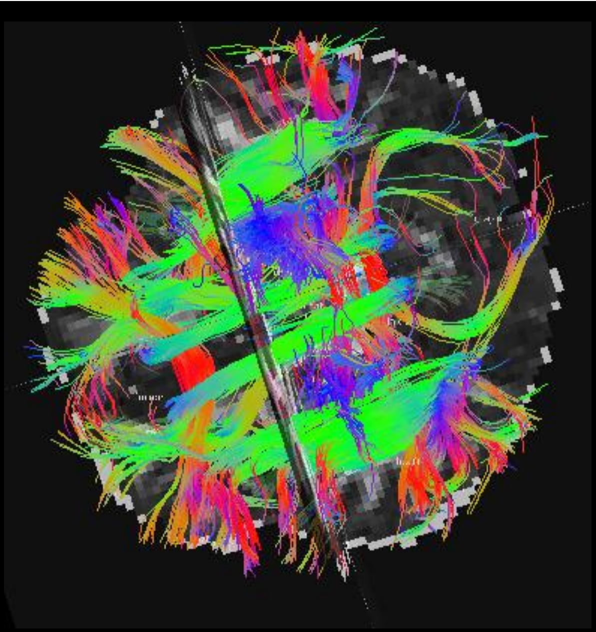
PCA at R = 3



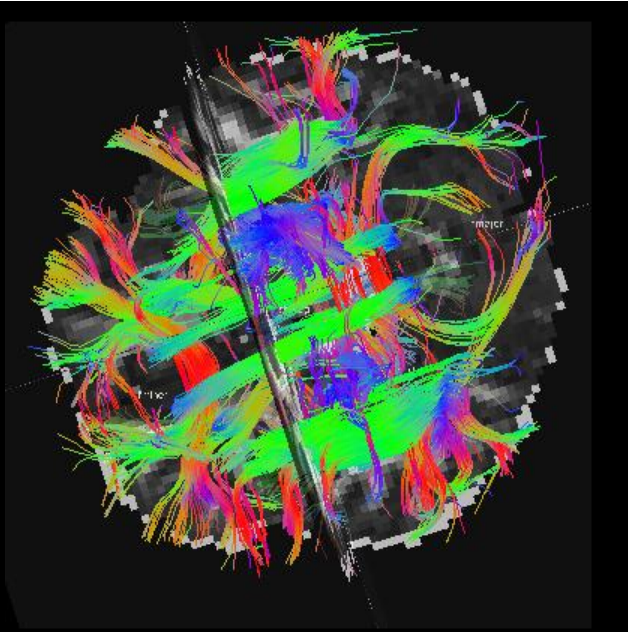
Dictionary-L2 at R = 3



Fully-sampled



PCA at R = 3



Dictionary-L2 at R = 3

- Based on deterministic DSI tractography, 18 white matter pathways were automatically labeled [1]

Fractional Anisotropy average error

PCA: 5.0%

Dict-L2: 6.0%

Mean Diffusivity average error

PCA: 3.2%

Dict-L2: 3.4%

1. Yendiki *et al* Front Neuroinform 2011

Conclusion

- Presence of dictionary is the crucial prior,
  not the sparsity constraint

Conclusion

- Presence of dictionary is the crucial prior,
  not the sparsity constraint

- The two proposed L2-regularized dictionary-based methods:
  simple, fast recon and trivial training step

Conclusion

- Presence of dictionary is the crucial prior,
  not the sparsity constraint

- The two proposed L2-regularized dictionary-based methods:
  simple, fast recon and trivial training step

- Proposed methods are comparable to L1-based Dictionary-FOCUSS in terms of pdf, odf and q-space recon quality

Conclusion

- Presence of dictionary is the crucial prior,
  not the sparsity constraint

- The two proposed L2-regularized dictionary-based methods:
  simple, fast recon and trivial training step

- Proposed methods are comparable to L1-based Dictionary-FOCUSS in terms of pdf, odf and q-space recon quality

- At 3-fold acceleration, proposed L2-based recons are comparable to fully-sampled acquisition

Conclusion

- Presence of dictionary is the crucial prior,
  not the sparsity constraint

- The two proposed L2-regularized dictionary-based methods:
  simple, fast recon and trivial training step

- Proposed methods are comparable to L1-based Dictionary-FOCUSS in terms of pdf, odf and q-space recon quality

- At 3-fold acceleration, proposed L2-based recons are comparable to fully-sampled acquisition

- Matlab code online:

  http://web.mit.edu/berkin/www/software.html

Acknowledgments

Grants:

- K99EB012107, U01MH093765,
- R01EB006847, R01EB007942,
- R01EB000790, P41RR14075

Sponsors:

- MIT-CIMIT Medical Engineering Fellowship
- Siemens Healthcare
- Siemens-MIT Alliance