

# Reconstruction Algorithms for MRI

by

Berkin Bilgic

S.M. Massachusetts Institute of Technology (2010)

Submitted to the Department of Electrical Engineering & Computer  
Science in Partial Fulfillment of the Requirements for the Degree of Doctor  
of Philosophy at the Massachusetts Institute of Technology

February 2013

© 2013 Massachusetts Institute of Technology. All rights reserved.

Signature of Author:

---

Department of Electrical Engineering and Computer Science

January 11, 2013

Certified by:

---

Elfar Adalsteinsson  
Associate Professor of Electrical Engineering and Computer Science  
Associate Professor of Health Sciences and Technology  
Thesis Supervisor

Accepted by:

---

Leslie A. Kolodziejski  
Chair, Department Committee on Graduate Students



# Reconstruction Algorithms for MRI

by  
Berkin Bilgic

Submitted to the Department of Electrical Engineering & Computer Science on January 11, 2013, in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

## **Abstract**

This dissertation presents image reconstruction algorithms for Magnetic Resonance Imaging (MRI) that aims to increase the imaging efficiency. Algorithms that reduce imaging time without sacrificing the image quality and mitigate image artifacts are proposed. The goal of increasing the MR efficiency is investigated across multiple imaging techniques: structural imaging with multiple contrasts preparations, Diffusion Spectrum Imaging (DSI), Chemical Shift Imaging (CSI), and Quantitative Susceptibility Mapping (QSM). The main theme connecting the proposed methods is the utilization of prior knowledge on the reconstructed signal. This prior often presents itself in the form of sparsity with respect to either a prespecified or learned signal transformation.

Thesis Supervisor: Elfar Adalsteinsson

Title: Associate Professor of Electrical Engineering and Computer Science  
Associate Professor of Health Sciences and Technology



## Acknowledgments

This is inarguably the most important page in this thesis.

First, I would like express my gratefulness to my advisor Prof. Elfar Adalsteinsson, who gave me the opportunity to study at MIT. Over the last three years, he thought me not only how to be a scientist, but also a decent human being.

Second, I am thankful to Prof. Kawin Setsompop for his guidance and friendship. He has been the perfect role model for me. From him, I learnt how to look out for younger people around me. I am looking forward to learning more.

Third, I am grateful to Prof. Lawrence Wald for being a part of my committee and being one of my references during faculty application. I would like to extend my thanks to Prof. Pablo Parrilo for allocating time from his schedule to join my thesis committee.

I was lucky enough to be surrounded by great dudes in the lab. I learnt everything I know about MR from Borjan, Joonsung, Trina, Audrey and Daniel. Towards the end, I had the pleasure of knowing Itthi, Obaidah, Shaoying and Paula. Drinking at ISMRM, scanning until 1 am, eating chi-burger at lunch are integral parts of the graduate student experience. Thank you guys for putting up with me...

Mom and dad, thank you. You made sacrifices in your lives just so that I had a better opportunity to succeed. Hopefully your sacrifices didn't go to waste and I have made myself a useful person.

My wife, best friend and companion. You complete me, give me a purpose and make me a better person every day I'm alive. Though it is not worthy, this thesis is dedicated to you.

-b

Cambridge, MA

December 2012



# Contents

<b>1 Introduction</b>	<b>19</b>
1.1 Motivation	19
1.2 Outline and bibliographical notes	21
<b>2 Joint Reconstruction of Multi-Contrast Images</b>	<b>25</b>
2.1 Introduction	25
2.2 Theory	28
2.2.1 Compressed Sensing in MRI	28
2.2.2 Conventional Compressed Sensing from a Bayesian Standpoint	28
2.2.3 Extending Bayesian Compressed Sensing to Multi-Contrast MRI	30
2.2.4 Bayesian Framework to Estimate the Image Gradient Coefficients	32
2.2.5 Reconstructing the Images from Horizontal and Vertical Gradient Estimates	35
2.2.6 Extension to Complex-Valued Images	35
2.3 Methods	36
2.3.1 Reconstruction with Extended Shepp-Logan Phantoms	37
2.3.2 SRI24 Multi-Channel Brain Atlas Data	38
2.3.3 3T Turbo Spin Echo (TSE) Slices with Early and Late TE's	38
2.3.4 Complex-Valued Shepp-Logan Phantoms	39
2.3.5 Complex-Valued Turbo Spin Echo Slices with Early and Late TE's	39
2.4 Results	40
2.4.1 CS Reconstruction with Extended Shepp-Logan Phantoms	40
2.4.2 SRI24 Multi-Channel Brain Atlas Data	41
2.4.3 Turbo Spin Echo (TSE) Slices with Early and Late TE's	42
2.4.4 Impact of Spatial Misregistration on Joint Reconstruction	44
2.4.5 Complex-Valued Shepp-Logan Phantoms	45
2.4.6 Complex-Valued Turbo Spin Echo Slices with Early and Late TE's	47
2.5 Discussion	48

2.6 Joint Reconstruction with Prior Estimate	52
2.6.1 EM Algorithm for Joint Reconstruction with Prior Estimate	53
2.6.2 Methods	54
2.6.3 Results	54
2.6.4 Remarks on Reconstruction with Prior Estimate	55
2.7 Conclusion	55
<b>3 Regularized Quantitative Susceptibility Mapping</b>	<b>57</b>
3.1 Introduction	58
3.2 Methods	60
3.2.1 Susceptibility and MR signal phase	60
3.2.2 Background effect removal from the field map	61
3.2.3 Susceptibility inversion with $\ell_1$ regularization	62
3.2.4 Susceptibility inversion with $\ell_2$ regularization	63
3.2.5 Effect of regularization parameters $\lambda$ and $\beta$	63
3.2.6 Selection of regularization parameters $\lambda$ and $\beta$	66
3.2.7 Dataset acquired in younger and elderly adults used for comparison of regularized QSM and FDRI	66
3.3 Results	71
3.3.1 Correlations of FDRI and QSM values with postmortem iron concentrations	71
3.3.2 Correlations between in vivo QSM and FDRI iron concentration metrics	71
3.3.3 Age differences in regional iron concentration: QSM and FDRI	72
3.3.4 Age differences identified with regularized QSM	72
3.3.5 Age differences identified with FDRI	75
3.4 Discussion	75
3.5 Fast $\ell_2$ -regularized QSM	77
3.5.1 Methods	78
3.5.2 Results	79
3.5.3 Remarks on the Fast $\ell_2$ -regularized QSM	80
3.6 Conclusion	80

<b>4 Lipid Suppression in Chemical Shift Imaging</b>	<b>83</b>
4.1 Introduction	83
4.2 Theory	85
4.2.1 Dual-Density Reconstruction	85
4.2.2 Iterative Reconstruction with Lipid-Basis Penalty	86
4.2.3 The Basic Method: Combining 2-average, high-resolution data with high SNR, low-resolution data	86
4.2.4 The Refined Method: Combining 2-average, <i>undersampled</i> high-resolution data with high SNR, low-resolution data	87
4.3 Methods	88
4.3.1 Choosing an Optimal Regularization Parameter	90
4.4 Results	92
4.5 Discussion	98
4.6 Conclusion	101
<b>5 Accelerated Diffusion Spectrum Imaging</b>	<b>103</b>
5.1 Introduction	103
5.2 Theory	105
5.2.1 CS Recovery with Prespecified Transforms	105
5.2.2 Training an Adaptive Transform with K-SVD	105
5.2.3 CS Recovery with an Adaptive Transform using FOCUSS	106
5.3 Methods	106
5.4 Results	108
5.5 Discussion	114
5.6 Fast DSI Reconstruction with Trained Dictionaries	117
5.6.1 Theory	117
5.6.2 Methods	120
5.6.3 Results	121
5.6.4 Remarks on Fast DSI Reconstruction with Trained Dictionaries	127
5.7 Conclusion	130

<b>6 Future Directions</b>	<b>131</b>
6.1 Joint Reconstruction	131
6.2 Quantitative Susceptibility Mapping	131
6.3 Chemical Shift Imaging	132
6.4 Diffusion Spectrum Imaging	133
<b>7 Bibliography</b>	<b>135</b>

## List of Figures

Fig. 2.1. Joint image reconstruction begins with modifying the undersampled  $k$ -space data to obtain undersampled  $k$ -space representations of vertical and horizontal image gradients. After finding the hyperparameters via Maximum Likelihood (ML) estimation, the means of the posterior distributions are assigned to be the gradient estimates. Finally, images are integrated from gradient estimates via solving a Least Squares (LS) problem.

Fig 2.2 Reconstruction results with the extended Shepp-Logan phantoms after undersampling with acceleration  $R = 14.8$ , at  $128 \times 128$  resolution. (a) Phantoms at Nyquist rate sampling. (b) Undersampling patterns in  $k$ -space corresponding to each image. (c) CS reconstructions with Lustig *et al.*'s algorithm yielded 15.9 % RMSE (root-mean-square error). (d) Absolute error plots for Lustig *et al.*'s method. (e) Reconstructions obtained with the M-FOCUSS joint reconstruction algorithm have 8.8 % RMSE. (f) Absolute difference between the Nyquist sampled phantoms and the M-FOCUSS reconstruction results. (g) Joint Bayesian CS reconstruction resulted in 0 % RMSE. (h) Absolute error plots for the Bayesian CS reconstructions.

Fig. 2.3. Reconstruction results with SRI24 atlas after undersampling along the phase encoding direction with  $R = 4$ , at  $256 \times 256$  resolution. (a) Atlas images at Nyquist rate sampling. (b) Undersampling patterns in  $k$ -space corresponding to each image. (c) Applying the gradient descent algorithm proposed by Lustig *et al.* resulted in reconstructions with 9.4 % RMSE. (d) Absolute difference between the gradient descent reconstructions and the Nyquist rate images. (e) M-FOCUSS reconstructions have 3.2 % RMSE. (f) Absolute error plots for the M-FOCUSS algorithm. (g) Joint Bayesian reconstruction yielded images with 2.3 % RMSE. (h) Error plots for the joint Bayesian reconstructions.

Fig. 2.4. Reconstruction results with TSE after undersampling along the phase encoding direction with  $R = 2.5$ , at  $256 \times 256$  resolution. (a) TSE scans at Nyquist rate sampling. (b) Undersampling patterns used in this experiment. (c) Reconstructions obtained with Lustig *et al.*'s gradient descent algorithm have 9.4 % RMSE. (d) Plots of absolute error for the gradient descent reconstructions. (e) M-FOCUSS joint reconstruction yielded images with 5.1 % RMSE. (f) Error plots for the M-FOCUSS results. (g) Images obtained with the joint Bayesian CS reconstruction returned 3.6 % RMSE. (h) Error plots for the Bayesian CS reconstructions.

Fig. 2.5. To investigate the impact of spatial misalignments on joint reconstruction with Bayesian CS and M-FOCUSS, one of the TSE images was shifted relative to the other by 0 to 2 pixels with step sizes of  $\frac{1}{2}$  pixels using power law and phase encoding undersampling patterns. For speed, low resolution images with size  $128 \times 128$  were used. For joint Bayesian CS, reconstruction error increased from 2.1 % to 2.8 % at 2 pixels of vertical shift for power law sampling, and from 5.2 % to 6.4 % at 2 pixels of horizontal

shift for phase encoding sampling; for the M-FOCUSS method error increased from 4.7 % to 4.9 % for power law sampling, and from 6.2 % to 6.6 % for phase encoding sampling.

Fig. 2.6. Reconstruction results with the complex-valued Shepp-Logan phantoms after undersampling with acceleration  $R = 3.5$ , at  $128 \times 128$  resolution. (a) Magnitudes of phantoms at Nyquist rate sampling. (b) Symmetric undersampling patterns in  $k$ -space corresponding to each image. (c) Real and imaginary parts of the first phantom (on the left in (a)). (d) Real and imaginary parts of the second phantom (on the right in (a)). (e) CS reconstructions with Lustig *et al.*'s algorithm yielded 13.1 % RMSE. (f) Absolute error plots for Lustig *et al.*'s method. (g) Reconstructions obtained with the M-FOCUSS joint reconstruction algorithm have 5.4 % RMSE. (h) Absolute difference between the Nyquist sampled phantoms and the M-FOCUSS reconstruction results. (i) Joint Bayesian CS reconstruction resulted in 2.4 % RMSE. (h) Absolute error plots for the Bayesian CS reconstructions.

Fig. 2.7. Reconstruction results for complex-valued TSE images after undersampling along the phase encoding direction with  $R = 2$ , at  $128 \times 128$  resolution. (a) Magnitudes of the TSE scans at Nyquist rate sampling. (b) Symmetric undersampling patterns used in this experiment. (c) Real and imaginary parts of the early echo image (on the left in (a)). (d) Real and imaginary parts of the late echo image (on the right in (a)). (e) Reconstructions obtained with Lustig *et al.*'s gradient descent algorithm have 8.8 % RMSE. (d) Plots of absolute error for the gradient descent reconstructions. (e) M-FOCUSS joint reconstruction yielded images with 9.7 % RMSE. (f) Error plots for the M-FOCUSS results. (g) Images obtained with the joint Bayesian CS reconstruction returned 6.1 % RMSE. (h) Error plots for the Bayesian CS reconstructions.

Fig. 2.8. (a) Image gradients for the multi-contrast TSE scans demonstrate the similarity under the gradient transform. (b) To quantify this similarity, we computed the cumulative energy of the image gradient of early TSE scan ( $TSE1$  in  $TSE1$  order). Then we sorted the late TSE scan ( $TSE2$ ) in descending order, and computed the cumulative energy in  $TSE1$  corresponding to the sorted indices in  $TSE2$  which gave the curve  $TSE1$  in  $TSE2$  order. The similarity of the curves indicates similar sparsity supports across images.

Fig. 2.9. (a) Lustig *et al.*'s algorithm yielded 9.3% error (b) absolute error for (c) Bayesian CS with prior returned 5.8% error (d) error for Bayesian CS (e) fully-sampled prior (f)  $R=4$  sampling pattern.

Fig. 2.10. (a1-a2) Lustig *et al.*'s algorithm yielded 9.5% error (b1-b2) absolute error plots for Lustig *et al.* (c1-c2). Joint Bayesian CS with prior returned 4.3% error (d1-d2) error plots for Bayesian CS (e) fully-sampled PD weighted prior image (f)  $R=4$  random undersampling pattern in 1D.

Fig. 3.1. L-curve for  $\ell_1$ -regularized QSM results for a young subject. X-axis: data consistency term  $\|\delta - \mathbf{F}^{-1} \mathbf{DF} \chi\|_2$  in regularized reconstruction for varying values of the

smoothing parameter  $\lambda$ . Y-axis: regularization term  $\|\mathbf{G}\boldsymbol{\chi}\|_1$ . Setting  $\lambda = 5 \cdot 10^{-5}$  yielded an under-regularized susceptibility map with ringing artifacts (a), whereas using  $\lambda = 10^{-3}$  resulted an over-regularized reconstruction (c). For  $\lambda = 2 \cdot 10^{-4}$ , the operating point with the largest curvature on the L-curve was obtained (b). This setting was used for the reported  $\ell_1$ -regularized results.

Fig. 3.2. L-curve for  $\ell_2$ -regularized QSM results for a young subject. X-axis: data consistency term  $\|\boldsymbol{\delta} - \mathbf{F}^{-1} \mathbf{D} \mathbf{F} \boldsymbol{\chi}\|_2$  in regularized reconstruction for varying values of the smoothing parameter  $\beta$ . Y-axis: regularization term  $\|\mathbf{G}\boldsymbol{\chi}\|_2$ . Setting  $\beta = 3 \cdot 10^{-3}$  yielded an under-regularized susceptibility map with ringing artifacts (a), whereas using  $\beta = 7 \cdot 10^{-2}$  resulted an over-regularized reconstruction (c). For  $\beta = 1.5 \cdot 10^{-2}$ , the operating point with the largest curvature on the L-curve was obtained (b). This setting was used for the reported  $\ell_2$ -regularized results.

Fig. 3.3. Young (left) and elderly (right) group averages for FDRI (a),  $\ell_1$ -regularized QSM (b), and  $\ell_2$ -regularized QSM (c). Greater iron concentration yields brighter QSM and FDRI images. Splenium reference ROIs are indicated with a white box on the axial QSM slices.

Fig. 3.4. X-axis: Mean  $\pm$  SD iron concentration (mg/100 g fresh weight) determined *postmortem* in each ROI (1). Y-axis: Mean  $\pm$  SD  $\ell_1$ -regularized QSM in ppm (left) and FDRI in  $\text{s}^{-1}/\text{Tesla}$  (right) indices in all 23 subjects (black squares); the gray circles indicate the mean of the young group, and the open circles indicate the mean of the elderly group.

Fig. 3.5. Correlation between FDRI and  $\ell_1$ -regularized QSM results on the regions of interest. Results indicate strong relationship between the two methods ( $Rho = 0.976$ ,  $p = 0.0098$ ). Left: all 23 subjects; middle: young group; right: elderly group.

Fig. 3.6. Mean  $\pm$  S.E.M. of average susceptibility in ppm computed by the two methods ( $\ell_1$ -regularized QSM, top;  $\ell_2$ -regularized QSM, bottom) for each ROI in the young and elderly groups.

Fig. 3.7 Reconstruction experiment for the piece-wise constant numerical phantom with 3 compartments. (a) Noisy field map from which the susceptibility is estimated. (b) Closed-form QSM solution. (c) Difference between ground truth and closed-form reconstructions.

Fig. 3.8 In vivo reconstruction at 1.5T. (a) Tissue field map obtained after removing the background phase. (b) Closed-form QSM solution. (c) Difference between iterative and closed-form solutions.

Fig. 4.1. The L-curve traced by the data consistency and lipid-basis penalty terms as the regularization parameter  $\lambda$  varies. Summation over lipid frequencies for under-regularized (a), optimally regularized (b) and over-regularized reconstructions (c) are presented. Panel

(d) depicts the analytically computed L-curve curvature results for the sample points.

Fig. 4. 2. Comparing the different artifact reduction algorithms by taking projections over the lipid resonance frequencies (in dB scale). Gold standard reconstruction is obtained using 20 averages of high-resolution data without peripheral  $k$ -space undersampling ( $20 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$ , shown in (a)), while the basic proposed method is obtained using 2 averages of high-resolution data without undersampling ( $2 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$ , shown in (b)) and the refined proposed method uses 10-fold undersampled, 2 average high-resolution data ( $2 \text{ avg}_{\text{high}}, R_{\text{high}} = 10$ , shown in (c)). Lipid suppression results obtained by using only lipid-basis penalty method and only dual-density approach are depicted in panels (d) and (e), respectively. Applying no lipid suppression (f) results in severely corrupted spectra.

Fig. 4.3. Comparison between NRMSE values of NAA maps relative to the gold standard reconstruction.

Fig. 4.4. Comparison between NRMSE values of NAA maps computed within the  $9 \times 9 \text{ cm}^2$  excitation box relative to the NAA maps obtained with the OVS method. In (a), reconstruction results obtained using the gold-standard ( $20 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$ ) method (blue) and the OVS spectra (black) belonging to the region inside the red box are also overplotted. In (b), the basic proposed method (blue) and the OVS spectra are compared. The spectra obtained with the refined method (blue) and the OVS results (black) are overplotted in (c). Lipid-basis penalty and OVS spectra are compared in (d).

Fig. 4.5. Comparison of spectra inside the region of interest marked with the red box that were obtained with different lipid suppression methods. In (a), reconstruction results obtained using lipid-basis penalty method (blue) and the gold-standard reconstruction (black) are overplotted. In (b), the basic proposed method (blue) and the gold-standard spectra are presented. The spectra obtained with the refined method (blue) and the gold-standard results (black) are plotted in (c).

Fig. 4.6. Comparison of spectra inside the region of interest marked with the red box that were obtained with different lipid suppression methods. Panel (a) overplots reconstruction results using lipid-basis penalty method (blue) and the gold-standard reconstruction (black). In (b), the basic proposed method (blue) and the gold-standard spectra are compared. The spectra obtained with the refined method (blue) and the gold-standard results (black) are depicted in (c).

Fig. 4.7. Lipid and NAA maps and artifact-free spectra for the Cartesian synthetic phantom are shown in (a). In (b), spiral sampling trajectory at Nyquist rate and reconstruction results upon the application of lipid-basis penalty are depicted. Using the undersampled spiral trajectory in (c), a high-resolution lipid image was estimated using FOCUSS, from which a combined image was computed due to the dual-density method. Finally, lipid-basis penalty was applied to this combined image. Panel (d) shows lipid suppression results when the  $k$ -space is sampled only at half of the full resolution and lipid-basis penalty is applied. For the three reconstruction methods, the example spectra

(plotted in blue) belong to the region of interest marked with the red box, and are overplotted with the artifact-free spectra (in black) for comparison.

Fig. 4.8. Demonstration of approximate orthogonality between metabolite spectra obtained from *in vivo* OVS scan and lipid spectra from high resolution *in vivo* acquisition. In (a), the lipid and metabolite spectra with the highest orthogonality are plotted. In (b), the components of the metabolite spectrum that are orthogonal and parallel to the lipid spectrum for the best case in (a) are overplotted. The actual metabolite spectrum (in blue) is totally occluded by the orthogonal component (in orange). In (c), the lipid and metabolite spectra that are least orthogonal are depicted. In (d), the orthogonal and parallel components of the metabolite spectrum are overplotted for the worst case in (c). Panel (e) depicts the methodology used in computing the orthogonal and parallel metabolite components.

Fig. 5.1. RMSE at each voxel in slice 40 of subject A upon  $R = 3$  acceleration and reconstruction with Menzel et al.'s method (a),  $\ell_1$ -FOCUSS (b), Dictionary-FOCUSS trained on subjects A (c), B (d), and C (e). Dictionary-FOCUSS errors in (f), (g) and (h) are obtained at higher acceleration factor of  $R = 5$  with training on subjects A, B and C, respectively. Results for the reconstructions at  $R = 9$  are given in (i), (j) and (k).

Fig. 5.2. RMSE at each voxel in slice 25 of subject B upon  $R = 3$  acceleration and reconstruction with Menzel et al.'s method (a),  $\ell_1$ -FOCUSS (b), Dictionary-FOCUSS trained on subjects A (c), B (d), and C (e). Dictionary-FOCUSS errors in (f), (g) and (h) are obtained at higher acceleration factor of  $R = 5$  with training on subjects A, B and C, respectively. Results for the reconstructions at  $R = 9$  are given in (i), (j) and (k).

Fig. 5.3. Mean and standard deviation of RMSEs computed on various slices of subject A using  $\ell_1$ - and Dictionary-FOCUSS trained on subject B. Lower panel depicts RMSE maps for four selected slices.

Fig. 5.4. Mean and standard deviation of RMSEs computed on various slices of subject B using  $\ell_1$ - and Dictionary-FOCUSS trained on subject A. Lower panel depicts RMSE maps for four selected slices.

Fig. 5.5. Top panel shows RMSEs in 'missing'  $q$ -space directions that are estimated with Wavelet+TV,  $\ell_1$ -FOCUSS and Dictionary-FOCUSS with training on subjects A, B and C at  $R=3$ .  $q$ -space images at directions [5,0,0] (a) and [0,4,0] (c) are depicted for comparison of the reconstruction methods. In panels (b) and (d), reconstruction errors of Wavelet+TV,  $\ell_1$ -FOCUSS and dictionary reconstructions relative to the 10 average fully-sampled image at directions [5,0,0] and [0,4,0] are given.

Fig. 5.6. Panel on top depicts RMSEs of Wavelet+TV,  $\ell_1$ -FOCUSS and Dictionary-FOCUSS at  $R = 3$  and fully-sampled 1 average data computed in 5  $q$ -space locations relative to the 10 average data for subject A. Panel on the bottom shows the same comparison for the slice belonging to subject B.

Fig. 5.7. Axial view of white-matter pathways labeled from streamline DSI tractography in fully-sampled data (a) and Dictionary-FOCUSS reconstruction at  $R = 3$  (b). The following are visible in this view: corpus callosum - forceps minor (FMIN), corpus callosum - forceps major (FMAJ), anterior thalamic radiations (ATR), cingulum - cingulate gyrus bundle (CCG), superior longitudinal fasciculus - parietal bundle (SLFP), and the superior endings of the corticospinal tract (CST). Average FA (c) and volume in number of voxels (d) for each of the 18 labeled pathways, as obtained from the fully-sampled ( $R=1$ , green) and Dictionary-FOCUSS reconstructed with 3-fold undersampling ( $R=3$ , yellow) datasets belonging to subject A. Intra-hemispheric pathways are indicated by “L-” (left) or “R-” (right). The pathways are: corpus callosum - forceps major (FMAJ), corpus callosum - forceps minor (FMIN), anterior thalamic radiation (ATR), cingulum - angular (infracallosal) bundle (CAB), cingulum - cingulate gyrus (supracallosal) bundle (CCG), corticospinal tract (CST), inferior longitudinal fasciculus (ILF), superior longitudinal fasciculus - parietal bundle (SLFP), superior longitudinal fasciculus - temporal bundle (SLFT), uncinata fasciculus (UNC).

## List of Tables

Table 2.1. Summary of additional reconstruction results on the TSE and SRI 24 datasets using the three algorithms after retrospective undersampling with various patterns and acceleration factors.

Table 3.1a. Mean ( $\pm$ SD) of each measure by region for each group:  $\ell_1$ -regularized QSM results using phase-guided ROIs and FDRI-guided ROIs

Table 3.1b. Mean ( $\pm$ SD) of each measure by region for each group:  $\ell_2$  regularized QSM results using phase-guided ROIs and FDRI-guided ROIs

Table 3.1c. Mean ( $\pm$ SD) of each measure by region for each group: FDRI results using phase-guided ROIs and FDRI-guided ROIs



# Chapter 1

## Introduction

### 1.1 Motivation

Magnetic Resonance Imaging (MRI) is a non-invasive imaging modality that is capable of yielding high-resolution and high-contrast images of soft tissues of the body. Unlike Computed Tomography (CT) or X-ray imaging, MRI does not employ ionizing radiation. It also does not require the introduction of a radioactive agent as employed in Positron Emission Tomography (PET). Therefore, MRI is considered to be a safe imaging modality that finds important clinical use. However, a major drawback of MRI is that it is inherently a slow imaging modality, requiring the subjects to remain motionless within a tight, closed environment typically for half an hour or longer, depending on the imaging protocol. This constraint on the imaging time reduces subject compliance and raises challenges especially in pediatric and patient populations.

With the introduction of parallel imaging and compressed sensing (CS) methods and ultra high-field systems over the last decade, substantial progress has been made towards improved the image quality and reduced acquisition time. Parallel imaging relies on the information provided by multiple receive coils that are sensitive to different parts of the region of interest for accelerated imaging. Aliasing caused by subsampled acquisitions is disentangled with the help of multiple coil data to yield high quality images. Parallel imaging has made the transition from being a *technique* to becoming a *technology*, as 2 to 3-fold accelerated acquisitions in the clinical setting are ubiquitous. Parallel imaging methods can operate either in the image space (2), or in the Fourier space (*k*-space) of the object where the data are collected (3). Compressed sensing, on the other hand, is a less mature technique in the field of medical imaging. CS is a collection of algorithms that aim to recover signals from subsampled measurements by applying a sparsity-inducing prior over the signal coefficients. Even though the idea of using sparsity-promoting optimization techniques in signal processing and statistics is not new (e.g. (4,5)), it was not deployed in MR image reconstruction until recently (6). Because of the non-linear nature of the processing involved, CS reconstruction artifacts are not fully characterized. As such, the clinical translation of CS has not reached the same level as parallel imaging methods.

More recent developments aim to merge parallel imaging and CS techniques to allow further reduction in imaging time. In this domain, L1 SPIR-iT (7) is a popular algorithm that combines

the  $k$ -space data from multiple coils while enforcing sparsity of coil images with respect to the wavelet transform. Similarly, it is possible to combine image-domain parallel imaging with sparsity priors for improved reconstruction (8).

In the light of these recent developments, this thesis presents image reconstruction algorithms that aim to further increase the imaging efficiency of MRI. These algorithms achieve,

- i. Reduction of the total scan time without sacrificing the image quality, and
- ii. Mitigation image artifacts due to physiology or MR physics to improve the image quality.

Reduction of imaging time is a well-motivated research goal, leading to increased patient comfort and reduced costs. This goal is investigated for the following MR imaging techniques,

- i. Structural imaging with multiple contrast preparations: By exploiting image statistics and similarity between images obtained with different contrasts, improved image reconstruction from undersampled data is demonstrated.
- ii. Diffusion Spectrum Imaging (DSI): Diffusion Weighted Imaging (DWI) aims to explore the brain connectivity by mapping the water diffusion as a function of space. DSI is a particular DWI method that is able to generate a complete description of diffusion probability density functions (pdfs), but suffers from significantly long imaging times. This dissertation demonstrates that by learning the structure of pdfs from training data, it is possible to substantially reduce the scan time with small cost on the image quality.

Mitigation of image artifacts is yet a different way to achieve increased efficiency, as it increases the amount of meaningful data for further processing and diagnosis. Results on artifact mitigation are demonstrated within two contexts,

- i. Regularized Quantitative Susceptibility Mapping (QSM): The magnetic property of the tissues called magnetic susceptibility gives rise to the observed signal phase in MRI, which is estimated using an iterative background removal method and regularized inversion. Regularization helps reduce the streaking artifacts in the reconstructed susceptibility map, which stem from the ill-posed nature of the relation connecting the phase to the magnetic susceptibility.
- ii. Lipid artifact reduction in Chemical Shift Imaging (CSI): A major obstacle in CSI is the contamination of brain spectra by the strong lipid signals around the skull. Lipid artifacts are substantially reduced by employing an iterative reconstruction method that makes use of rapidly sampled high frequency content of lipid signals.

## 1.2 Outline and bibliographical notes

In the following, the organization of the thesis is presented with brief descriptions and bibliographical contributions of each section.

**Chapter 2:** is concerned with reconstruction of structural MR images from undersampled observations. Versatility of MRI allows images with multiple contrasts preparations to be acquired, wherein each contrast emphasizes certain tissue types. Collection of such multi-contrast data facilitates diagnosis and finds frequent clinical use. In this setting, it is assumed that data are acquired with a single receive coil, hence parallel imaging is outside the scope of this chapter. One option for recovery of undersampled multi-contrast images is to employ compressed sensing on each contrast independently. These images belong to the same underlying physiology, so they are expected to share common tissue boundaries. Focusing on this point, this chapter presents a joint reconstruction method capable of improving compressed sensing reconstruction quality by exploiting the shared information content across contrasts. This method is based on Bayesian compressed sensing, which interprets sparsity-inducing reconstruction within a probabilistic framework. An extension to joint reconstruction is also presented: since the imaging sequences involved in the multi-contrast protocol may have different acquisition speeds, it might be possible to obtain a fully-sampled dataset using a fast sequence in addition to the undersampled contrasts. By using the fully-sampled image to initialize the reconstruction, further improvement in joint reconstruction quality is demonstrated.

The proposed methods take place in,

- **B. Bilgic**, V.K. Goyal, E. Adalsteinsson; *Multi-contrast Reconstruction with Bayesian Compressed Sensing*; Magnetic Resonance in Medicine, 2011; 66(6):1601-1615.
- **B. Bilgic**, V.K. Goyal, E. Adalsteinsson; *Joint Bayesian Compressed Sensing for Multi-contrast Reconstruction*; International Society for Magnetic Resonance in Medicine 19th Scientific Meeting, Montreal, Canada, 2011, p. 71.
- **B. Bilgic**, E. Adalsteinsson; *Joint Bayesian Compressed Sensing with Prior Estimate*; International Society for Magnetic Resonance in Medicine 20th Scientific Meeting, Melbourne, Australia, 2012, p. 75.

**Chapter 3:** focuses on Quantitative Susceptibility Mapping (QSM) which is an MRI based imaging technique that provides valuable quantitation of tissue iron concentration and vessel oxygenation. However, susceptibility cannot be observed directly with MRI. Reconstruction of

underlying susceptibility maps from measured MR signal phase is a challenging problem that requires deconvolution of an ill-posed kernel. Hence, this problem benefits from regularization that reflects prior knowledge on the tissue susceptibility. As susceptibility is a feature tied to the paramagnetic properties of the underlying tissues, it is expected to vary smoothly within tissue compartments. Using regularization based on spatial gradients of the susceptibility maps facilitates the deconvolution. In a group study where the brain iron concentration in normal aging was investigated, this chapter shows that accurate quantification is possible with this regularized deconvolution approach. Further, an algorithm that solves the regularized inversion problem in less than 5 seconds is proposed, which is a significant speed up relative to proposed iterative methods that can take up to an hour.

The contents of this chapter are included in,

- **B. Bilgic**, A. Pfefferbaum, T. Rohlfing, E.V. Sullivan, E. Adalsteinsson; *MRI Estimates of Brain Iron Concentration in Normal Aging Using Quantitative Susceptibility Mapping*; NeuroImage, 2012; 59(3):2625-2635.
- **B. Bilgic**, A.P. Fan, E. Adalsteinsson; *Quantitative Susceptibility Map Reconstruction with Magnitude Prior*; International Society for Magnetic Resonance in Medicine 19th Scientific Meeting, Montreal, Canada, 2011, p. 746.
- **B. Bilgic**, I. Chatnuntaweck, A.P. Fan, E. Adalsteinsson; *Regularized QSM in Seconds*; submitted to International Society for Magnetic Resonance in Medicine 21st Scientific Meeting, Salt Lake City, Utah, USA, 2013.
- **B. Bilgic**, I. Chatnuntaweck, K. Setsompop, S.F. Cauley, L.L. Wald, E. Adalsteinsson; *Fast Regularized Reconstruction Tools for QSM and DSI*; ISMRM Workshop on Data Sampling & Image Reconstruction, Sedona, Arizona, USA, 2013, accepted.

**Chapter 4:** proposes two lipid artifact suppression methods for CSI. While MRI enables spatial encoding of the human tissue, CSI also provides encoding in magnetic resonance frequency. At each voxel, this yields a 1-dimensional spectrum of relative concentrations of biochemical metabolites, each with a slightly different resonant frequency. The ability to map biochemical metabolism proves to be critical in cancer, Alzheimer's disease and multiple sclerosis. The dominant challenge of CSI is in the low signal of the metabolites of interest. Since signal-to-noise ratio (SNR) is proportional to the voxel size due to averaging effect, large voxels are required to lower the noise threshold, thereby constraining the voxel sizes in spectroscopy to be much larger than those of MRI ( $1 \text{ cm}^3$  compared to  $1 \text{ mm}^3$ ). The resolution constraint poses a

significant difficulty to metabolite detection as it leads to signal contamination from the subcutaneous lipid layer. This chapter proposes two post-processing methods that exploit prior knowledge about lipid and metabolite signals to yield artifact-free metabolite spectra. These algorithms rely on two observations: lipid signals are constrained to reside around the skull, and metabolite and lipid spectra are approximately orthogonal. As the lipids are constrained to reside on a ring in space and within a certain range in resonance frequency, they can be well approximated from undersampled data using sparsity-enforcing reconstruction. This permits estimation of high-resolution lipid signals, effectively reducing the ringing artifacts. Combined with iterative reconstruction that enforces orthogonality among metabolites in the brain and the lipid spectra, artifact-free metabolite maps are thus obtained.

The contributions in this chapter can also be found in,

- **B. Bilgic**, B. Gagoski, E. Adalsteinsson; *Lipid Suppression in CSI with Spatial Priors and Highly-Undersampled Peripheral k-space*; Magnetic Resonance in Medicine, 2012; DOI: 10.1002/mrm.24399.
- **B. Bilgic**, B. Gagoski E. Adalsteinsson; *Lipid Suppression in CSI with Highly-Undersampled Peripheral k-space and Spatial Priors*; International Society for Magnetic Resonance in Medicine 20th Scientific Meeting, Melbourne, Australia, 2012, p. 4455.

**Chapter 5:** Diffusion Weighted Imaging (DWI) is a widely used method to study white matter connectivity of the brain. Diffusion Tensor Imaging (DTI) is an established DWI method that models the water diffusion in each voxel as a univariate Gaussian distribution. Fiber tractography algorithms are employed to follow the major eigenvector of the tensor fit across neighboring voxels. However, the diffusion tensor model is unable to characterize multiple fiber orientations within the same voxel, which constitute over 90% of white matter voxels. Rather than modeling the diffusion, Diffusion Spectrum Imaging (DSI) offers a complete description of the diffusion probability density function (pdf). This provides DSI with the capability to resolve complex distributions of fiber orientations, thus overcoming the single-orientation limitation of DTI. The tradeoff is that, while a typical DTI scan takes ~5 minutes, DSI suffers from prohibitively long imaging times of ~50 minutes. By relying on prior information extracted from a training dataset, this chapter demonstrates dramatic reduction in DSI scan time while retaining the image quality. This high quality reconstruction is made possible by the priors encoded in a dictionary (created from a separately acquired training DSI dataset) that captures the structure of

diffusion pdfs. Further, two efficient dictionary-based reconstruction methods that attain 1000-fold computation speed-up relative to iterative DSI compressed sensing algorithms are presented.

The methods introduced in this chapter can also be found in,

- **B. Bilgic**, K. Setsompop, J. Cohen-Adad, A. Yendiki, L.L. Wald, E. Adalsteinsson; *Accelerated Diffusion Spectrum Imaging with Compressed Sensing using Adaptive Dictionaries*; *Magnetic Resonance in Medicine*, 2012; 68(6):1747-1754.
- **B. Bilgic**, I. Chatnuntawech, K. Setsompop, S.F. Cauley, L.L. Wald, E. Adalsteinsson; *Fast Diffusion Spectrum Imaging Reconstruction with Trained Dictionaries*; submitted to *IEEE Transactions on Medical Imaging*.
- **B. Bilgic**, K. Setsompop, J. Cohen-Adad, V. Wedeen, L. Wald, E. Adalsteinsson; *Accelerated Diffusion Spectrum Imaging with Compressed Sensing using Adaptive Dictionaries*; 15th International Conference on Medical Image Computing and Computer Assisted Intervention, 2012; LNCS 7512:1-9.
- **B. Bilgic**, I. Chatnuntawech, K. Setsompop, S.F. Cauley, L.L. Wald, E. Adalsteinsson; *Fast DSI Reconstruction with Trained Dictionaries*; submitted to International Society for Magnetic Resonance in Medicine 21st Scientific Meeting, Salt Lake City, Utah, USA, 2013.
- **B. Bilgic**, I. Chatnuntawech, K. Setsompop, S.F. Cauley, L.L. Wald, E. Adalsteinsson; *Fast Regularized Reconstruction Tools for QSM and DSI*; ISMRM Workshop on Data Sampling & Image Reconstruction, Sedona, Arizona, USA, 2013, accepted.

**Chapter 6:** proposes potential extensions to the methods introduced throughout the dissertation. Higher acceleration factors may be achieved by extending the multi-contrast reconstruction idea to include parallel imaging. Multi-modality imaging (e.g. MR-PET) may also benefit from joint reconstruction. Employing magnitude information in QSM deconvolution may improve the conditioning of the inversion. Quantitative susceptibility venography with vessel tracking may be feasible with the help of tracking algorithms in fiber tractography literature. In the context of spectroscopic imaging, parametric signal models may provide further regularization in lipid artifact suppression. Finally, through the combination of parallel imaging and dictionary-based reconstruction, even higher acceleration factors in DSI acquisitions may become achievable.

## Chapter 2

### Joint Reconstruction of Multi-Contrast Images

Clinical imaging with structural MRI routinely relies on multiple acquisitions of the same region of interest under several different contrast preparations. This chapter presents a reconstruction algorithm based on Bayesian compressed sensing to jointly reconstruct a set of images from undersampled  $k$ -space data with higher fidelity than when the images are reconstructed either individually or jointly by a previously proposed algorithm, M-FOCUSS. The joint inference problem is formulated in a hierarchical Bayesian setting, wherein solving each of the inverse problems corresponds to finding the parameters (here, image gradient coefficients) associated with each of the images. The variance of image gradients across contrasts for a single volumetric spatial position is a single hyperparameter. All of the images from the same anatomical region, but with different contrast properties, contribute to the estimation of the hyperparameters, and once they are found, the  $k$ -space data belonging to each image are used independently to infer the image gradients. Thus, commonality of image spatial structure across contrasts is exploited without the problematic assumption of correlation across contrasts. Examples demonstrate improved reconstruction quality (up to a factor of 4 in root-mean-square error) compared to previous compressed sensing algorithms and show the benefit of joint inversion under a hierarchical Bayesian model.

#### 2.1 Introduction

In clinical applications of structural MRI, it is routine to image the same region of interest under multiple contrast settings to enhance the diagnostic power of T1, T2, and proton-density weighted images. Herein, a Bayesian framework that makes use of the similarities between the images with different contrasts is presented to jointly reconstruct MRI images from undersampled data obtained in  $k$ -space. This method applies the joint Bayesian compressive sensing (CS) technique of Ji *et al.* (9) to the multi-contrast MRI setting with modifications for computational and  $k$ -space acquisition efficiency. Compared to conventional CS algorithms that work on each of the images independently (e.g. (6)), this joint inversion technique is seen to improve the reconstruction quality at a fixed undersampling ratio and to produce similar reconstruction results at higher undersampling ratios (i.e., with less data).

Conventional CS produces images using sparse approximation with respect to an appropriate basis; with gradient sparsity or wavelet-domain sparsity, the positions of nonzero coefficients correspond directly to spatial locations in the image. A natural extension to exploit structural similarities in multi-contrast MRI is to produce an image for each contrast setting while keeping the transform-domain sparsity pattern for each image the same. This is called *joint* or *simultaneous* sparse approximation. One of the earliest applications of simultaneous sparse approximation was in localization and used an algorithm based on convex relaxation (10). An early greedy algorithm was provided by Tropp *et al.* (11). Most methods for simultaneous sparse approximation extend existing algorithms such as Orthogonal Matching Pursuit (OMP), FOCal Underdetermined System Solver (FOCUSS) (4), or Basis Pursuit (BP) (12) with a variety of ways for fusing multiple measurements to recover the nonzero transform coefficients. Popular joint reconstruction approaches include Simultaneous OMP (SOMP) (11), M-FOCUSS (13), and the convex relaxation algorithm in (14). All of these algorithms provide significant improvement in approximation quality, however they suffer from two important shortcomings for the current problem statement. First, they assume that the signals share a common sparsity support, which does not apply to the multi-contrast MRI scans. Even though these images have nonzero coefficients in similar locations in the transform domain, assuming perfect overlap in the sparsity support is too restrictive. Second, with the exception of (15), most methods formulate their solutions under the assumption that all of the measurements are made via the same observation matrix, which in this context would correspond to sampling the same  $k$ -space points for all of the multi-contrast scans. As demonstrated here, observing different frequency sets for each image increases the overall  $k$ -space coverage and improves reconstruction quality.

The general joint Bayesian CS algorithm recently presented by Ji *et al.* (9) addresses these shortcomings and fits perfectly to the multi-contrast MRI context. Given the observation matrices  $\Phi_i \in \mathbb{C}^{K_i \times M}$  with  $K_i$  representing the number of  $k$ -space points sampled for the  $i^{\text{th}}$  image and  $M$  being the number of voxels, the linear relationship between the  $k$ -space data and the unknown images can be expressed as  $\mathbf{y}_i = \Phi_i \mathbf{x}_i$  where  $i = 1, \dots, L$  indexes the  $L$  multi-contrast scans and  $\mathbf{y}_i$  is the vector of  $k$ -space samples belonging to the  $i^{\text{th}}$  image  $\mathbf{x}_i$ . Let  $\delta_i^x$  and  $\delta_i^y$  denote the vertical and the horizontal image gradients, which are approximately sparse since the MRI images are approximately piecewise constant in the spatial domain. In the Bayesian setting, the task is to provide a posterior belief for the values of the gradients  $\delta_i^x$  and  $\delta_i^y$ , with the prior assumption that these gradients should be sparse and the reconstructed images should be consistent with the acquired  $k$ -space data. Each image formation problem (for a single contrast) constitutes an inverse

problem of the form  $\mathbf{y}_i \rightarrow \mathbf{x}_i$ , and the joint Bayesian algorithm aims to share information among these tasks by placing a common hierarchical prior over the problems. Such hierarchical Bayesian models can capture the dependencies between the signals without imposing correlation, for example by positing correlation of variances between zero-mean quantities that are conditionally independent given the hyperparameters. Data from all signals contribute to learning the common prior (i.e., estimating the hyperparameters) in a maximum likelihood framework, thus making information sharing among the images possible. Given the hierarchical prior, the individual gradient coefficients are estimated independently. Hence, the solution of each inverse problem is affected by both its own measured data and by data from the other tasks via the common prior. The dependency through the estimated hyperparameters is essentially a spatially-varying regularization, so it preserves the integrity of each individual reconstruction problem.

Apart from making use of the joint Bayesian CS machinery to improve the image reconstruction quality, the proposed method presents several novelties. First, the Bayesian algorithm is reduced to practice on MRI data sampled in  $k$ -space with both simulated and *in vivo* acquisitions. In the elegant work by Ji *et al.* (9), their method was demonstrated on CS measurements made directly in the sparse transform domain as opposed to the  $k$ -space domain that is the natural source of raw MRI data. The observations  $\mathbf{y}_i$  were obtained via  $\mathbf{y}_i = \Phi_i \boldsymbol{\theta}_i$  where  $\boldsymbol{\theta}_i$  are the wavelet coefficients belonging to the  $i^{\text{th}}$  test image. But in all practical settings of MRI data acquisition, the observations are carried out in the  $k$ -space corresponding to the reconstructed images themselves, i.e. the  $k$ -space data belonging to the wavelet transform of the image is not accessible. In the method as presented here, measurements of the image gradients are obtained by a simple modification of the  $k$ -space data and thus it is possible to overcome this problem. After solving for the gradient coefficients with the Bayesian algorithm, images that are consistent with these gradients are recovered in a least-squares setting. Secondly, the presented version accelerates the computationally-demanding joint reconstruction algorithm by making use of the Fast Fourier Transform (FFT) to replace some of the demanding matrix operations in the original implementation by Ji *et al.* This makes it possible to use the algorithm with higher resolution data than with the original implementation, which has large memory requirements. Also, partially-overlapping undersampling patterns are exploited to increase the collective  $k$ -space coverage when all images are considered; herein it is reported that this flexibility in the sampling pattern design improves the joint CS inversion quality. Additionally, the algorithm is generalized to allow inputs that correspond to complex-valued images. Finally, these findings are compared with the popular method in (6) and with the M-FOCUSS joint reconstruction scheme.

In addition to yielding smaller reconstruction errors relative to either method, the proposed Bayesian algorithm contains no parameters that need tuning.

## 2.2 Theory

### 2.2.1 Compressed Sensing in MRI

Compressed sensing has received abundant recent attention in the MRI community because of its demonstrated ability to speed up data acquisition. Making use of CS theory to this end was first proposed by Lustig *et al.* (6), who formulated the inversion problem as

$$\hat{\mathbf{x}} = \arg \min_x \|\Psi^T \mathbf{x}\|_1 + \beta \cdot TV(\mathbf{x}) \quad \text{s.t.} \quad \|\mathbf{y} - \mathbf{F}_\Omega \mathbf{x}\|_2 < \varepsilon \quad (2.1)$$

where  $\Psi$  is the wavelet basis,  $TV(\cdot)$  is the  $\ell_1$  norm of discrete gradients as a proxy for total variation,  $\beta$  trades off wavelet sparsity and gradient sparsity,  $\mathbf{F}_\Omega$  is the undersampled Fourier transform operator containing only the frequencies  $\omega \in \Omega$ , and  $\varepsilon$  is a threshold parameter that needs to be tuned for each reconstruction task. This constrained inverse problem can be posed as an unconstrained optimization program (6)

$$\hat{\mathbf{x}} = \arg \min_x \|\mathbf{y} - \mathbf{F}_\Omega \mathbf{x}\|_2^2 + \lambda_{wavelet} \cdot \|\Psi^T \mathbf{x}\|_1 + \lambda_{TV} \cdot TV(\mathbf{x}) \quad (2.2)$$

where  $\lambda_{wavelet}$  and  $\lambda_{TV}$  are wavelet and total variation regularization parameters that again call for tuning.

### 2.2.2 Conventional Compressed Sensing from a Bayesian Standpoint

Before presenting the mathematical formulation that is the basis for the proposed method, this section briefly demonstrates that it is possible to recover the conventional CS formulation in Eq. 2.2 with a Bayesian treatment. For the moment, consider abstractly that a sparse signal  $\mathbf{x} \in \mathbb{R}^M$  that is observed by compressive measurements via the matrix  $\Phi \in \mathbb{R}^{K \times M}$ , where  $K < M$  is under consideration. The general approach of Bayesian CS is to find the most likely signal coefficients with the assumptions that the signal is approximately sparse and that the data are corrupted by noise with a known distribution. The sparsity assumption is reflected by the prior defined on the signal coefficients, whereas the noise model is expressed via the likelihood term.

As a means to justify Eq. (2.2), a commonly-used signal prior and noise distribution are presented. The data are modeled as being corrupted by additive white Gaussian noise with variance  $\sigma^2$  via  $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{n}$ . In this case, the probability of observing the data  $\mathbf{y}$  given the signal  $\mathbf{x}$  is a Gaussian probability density function (pdf) with mean  $\mathbf{\Phi}\mathbf{x}$  and variance  $\sigma^2$ ,

$$p(\mathbf{y} | \mathbf{x}) = (2\pi\sigma^2)^{-K/2} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|^2\right) \quad (2.3)$$

which constitutes the likelihood term. To formalize the belief that the signal  $\mathbf{x}$  is sparse, a sparsity-promoting prior is placed on it. A common prior is the separable Laplacian density function (16)

$$p(\mathbf{x}) = (\lambda/2)^M \exp\left(-\lambda \sum_{i=1}^M |x_i|\right) \quad (2.4)$$

Invoking Bayes' theorem, the posterior for the signal coefficients can be related to the likelihood and the prior as

$$p(\mathbf{x} | \mathbf{y}) = \frac{p(\mathbf{y} | \mathbf{x})p(\mathbf{x})}{p(\mathbf{y})} \quad (2.5)$$

The signal that maximizes this posterior probability via maximum *a posteriori* (MAP) estimation is sought for. Since the denominator is independent of  $\mathbf{x}$ , the MAP estimate can be found by minimizing the negative of the logarithm of the numerator:

$$\mathbf{x}_{MAP} = \arg \min_{\mathbf{x}} \|\mathbf{y} - \mathbf{\Phi}\mathbf{x}\|_2^2 + 2\sigma^2 \lambda \|\mathbf{x}\|_1 \quad (2.6)$$

This expression is very similar to the unconstrained convex optimization formulation in Eq. (2.2); it is possible obtain Eq. (2.2) with a slightly more complicated prior that the wavelet coefficients and gradient of the signal of interest follow Laplacian distributions. Therefore, it is possible to view the convex relaxation CS algorithms as MAP estimates with a Laplacian prior on the signal coefficients. It is possible to view many algorithms used in CS as MAP estimators with respect to some prior (17).

### 2.2.3 Extending Bayesian Compressed Sensing to Multi-Contrast MRI

The Bayesian analysis in the previous section has two significant shortcomings. First, it is assumed that the signal of interest is sparse with respect to the base coordinate system. To get the maximum benefit from estimation with respect to a separable signal prior, it is critical to change to coordinates in which the marginal distributions of signal components are highly peaked at zero (18). For MR image formation, we aim to take advantage of the highly peaked distributions of image-domain gradients, and show how to modify  $k$ -space data to obtain measurements of these gradients. Second, the optimal MAP estimation through Eq. (2.6) requires knowledge of parameters  $\lambda$  and  $\sigma$ . The proposed method eliminates the tuning of such parameters by imposing a *hierarchical* Bayesian model in which  $\lambda$  and  $\sigma$  are modeled as realizations of random variables; this introduces the need for “hyperpriors” at a higher level of the model, but as detailed below, it suffices to eliminate tuning of the hyperpriors using a principle of least informativeness. Along with addressing these shortcomings, modifications for joint reconstruction across contrast preparations are also discussed.

In the multi-contrast setting, the signals  $\{\mathbf{x}_i\}_{i=1}^L \in \mathbb{R}^M$  represent MRI scans with different image weightings, e.g. T1, T2 and proton density weighted images might have been obtained for the same region of interest. These are not sparse directly in the image domain. Therefore, it is beneficial to cast the MRI images into a sparse representation to make use of the Bayesian formalism. The fact that the observation matrices  $\mathbf{F}_{\Omega_i} \in \mathbb{C}^{K_i \times M}$  in MRI are undersampled Fourier operators makes it very convenient to use spatial image gradients as a sparsifying transform (19,20). To obtain the  $k$ -space data corresponding to vertical and horizontal image gradients, it is sufficient to modify the data  $\mathbf{y}_i$  according to

$$\mathbf{F}_{\Omega_i} \boldsymbol{\delta}_i^x(\omega, \nu) = (1 - e^{-2\pi j \omega / n}) \mathbf{y}_i(\omega, \nu) \equiv \mathbf{y}_i^x \quad (2.7)$$

$$\mathbf{F}_{\Omega_i} \boldsymbol{\delta}_i^y(\omega, \nu) = (1 - e^{-2\pi j \nu / m}) \mathbf{y}_i(\omega, \nu) \equiv \mathbf{y}_i^y \quad (2.8)$$

where  $j = \sqrt{-1}$ ;  $\boldsymbol{\delta}_i^x$  and  $\boldsymbol{\delta}_i^y$  are the  $i^{\text{th}}$  image gradients;  $\mathbf{y}_i^x$  and  $\mathbf{y}_i^y$  are the modified observations; and  $\omega$  and  $\nu$  index the frequency space of the  $n$  by  $m$  pixel images, with  $n \cdot m = M$ . To solve Eq. (2.2), Lustig *et al.* (6) proposes to use the conjugate gradient descent algorithm, for which it is relatively straightforward to incorporate the TV norm. But algorithms that do not explicitly try to minimize an objective function (e.g. OMP and Bayesian CS) will need to modify the  $k$ -space data according to Eqs. (2.7) and (2.8) to make use of the Total Variation penalty in the form of spatial derivatives.

Secondly, we need to express the likelihood term in such a way that both real and imaginary parts of the noise  $\mathbf{n}_i \in \mathbb{C}^{K_i}$  in  $k$ -space are taken into account. We rearrange the linear observations  $\mathbf{y}_i^x = \mathbf{F}_{\Omega_i} \boldsymbol{\delta}_i^x + \mathbf{n}_i$  as

$$\begin{bmatrix} \mathcal{R}e(\mathbf{y}_i^x) \\ \mathcal{I}m(\mathbf{y}_i^x) \end{bmatrix} = \begin{bmatrix} \mathcal{R}e(\mathbf{F}_{\Omega_i}) \\ \mathcal{I}m(\mathbf{F}_{\Omega_i}) \end{bmatrix} \boldsymbol{\delta}_i^x + \begin{bmatrix} \mathcal{R}e(\mathbf{n}_i) \\ \mathcal{I}m(\mathbf{n}_i) \end{bmatrix} \quad (2.9)$$

for  $i = 1, \dots, L$ , where  $\mathcal{R}e(\cdot)$  and  $\mathcal{I}m(\cdot)$  indicate real and imaginary parts with the understanding that we also have an analogous set of linear equations for the horizontal gradients  $\boldsymbol{\delta}_i^y$ . For simplicity, we adopt the notation

$$\mathbf{Y}_i^x = \boldsymbol{\Phi}_i \boldsymbol{\delta}_i^x + \mathbf{N}_i \quad (2.10)$$

where  $\mathbf{Y}_i^x, \mathbf{N}_i \in \mathbb{R}^{2K_i}$ , and  $\boldsymbol{\Phi}_i \in \mathbb{R}^{2K_i \times M}$  correspond to the respective concatenated variables in Eq. (2.9). With the assumption that both real and imaginary parts of the  $k$ -space noise are white Gaussian with some variance  $\sigma^2$ , the data likelihood becomes

$$p(\mathbf{Y}_i^x | \boldsymbol{\delta}_i^x, \sigma^2) = (2\pi\sigma^2)^{-K_i} \exp\left(-\frac{1}{2\sigma^2} \|\mathbf{Y}_i^x - \boldsymbol{\Phi}_i \boldsymbol{\delta}_i^x\|^2\right) \quad (2.11)$$

With these modifications, it is now possible to compute the MAP estimates for the image gradients by invoking Laplacian priors over them. Unfortunately, obtaining the MAP estimates for each signal separately contradicts with the ultimate goal to perform joint reconstruction. In addition, it is beneficial to have a full posterior distribution for the sparse coefficients rather than point estimates, since having a measure of uncertainty in the estimated signals leads to an elegant experimental design method. As argued in (16), it is possible to determine an optimal  $k$ -space sampling pattern that reduces the uncertainty in the signal estimates. But since the Laplacian prior is not a conjugate distribution to the Gaussian likelihood, the resulting posterior will not be in the same family as the prior, hence it will not be possible to perform the inference in closed form to get a full posterior. The work by Ji *et al.* (9) presents an elegant way of estimating the image gradients within a hierarchical Bayesian model. This approach allows information sharing between the multi-contrast scans, at the same yields a full posterior estimate for the sparse coefficients. In the following section, the algorithm used for finding this distribution is summarized and the complete image reconstruction scheme is depicted in Fig. 2.1.

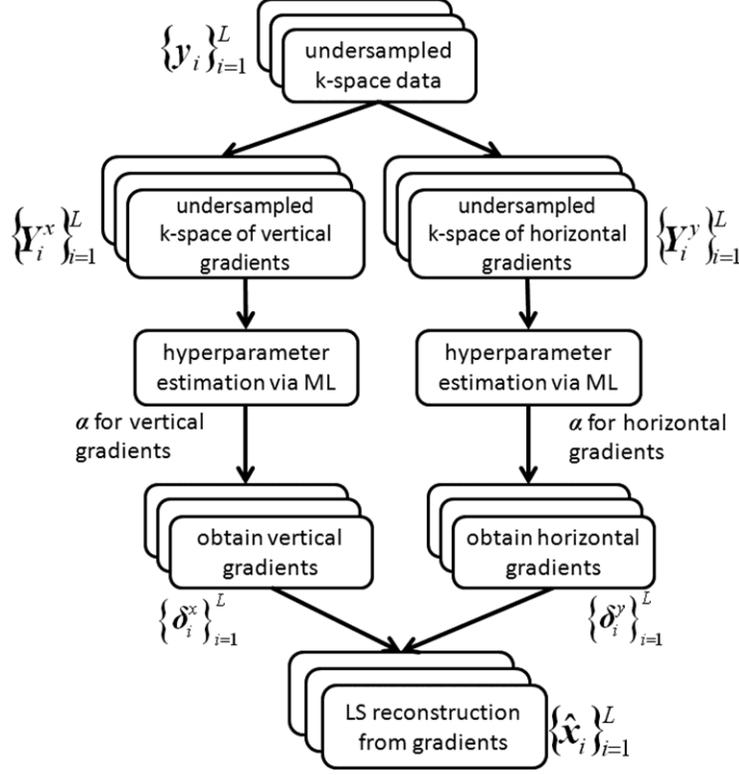


Fig. 2.1. Joint image reconstruction begins with modifying the undersampled k-space data to obtain undersampled  $k$ -space representations of vertical and horizontal image gradients. After finding the hyperparameters via Maximum Likelihood (ML) estimation, the means of the posterior distributions are assigned to be the gradient estimates. Finally, images are integrated from gradient estimates via solving a Least Squares (LS) problem.

## 2.2.4 Bayesian Framework to Estimate the Image Gradient Coefficients

Hierarchical Bayesian representation provides the ability to capture both the idiosyncrasy of the inversion tasks and the relations between them, while allowing closed form inference for the image gradients. According to this model, the sparse coefficients are assumed to be drawn from a product of zero mean normal distributions with variances determined by the hyperparameters

$$\boldsymbol{\alpha} = \{\alpha_j\}_{j=1}^M$$

$$p(\delta_i^x | \boldsymbol{\alpha}) = \prod_{j=1}^M \mathcal{N}(\delta_{i,j}^x | 0, \alpha_j^{-1}) \quad (2.12)$$

where  $\mathcal{N}(\cdot | 0, \alpha_j^{-1})$  is a zero mean Gaussian density function with variance  $\alpha_j^{-1}$ . In order to promote sparsity in the gradient domain, Gamma priors are defined over the hyperparameters  $\boldsymbol{\alpha}$

$$p(\boldsymbol{\alpha} | a, b) = \prod_{j=1}^M Ga(\alpha_j | a, b) = \prod_{j=1}^M \frac{b^a}{\Gamma(a)} \alpha_j^{a-1} \exp(-b\alpha_j) \quad (2.13)$$

where  $\Gamma(\cdot)$  is the Gamma function, and  $a$  and  $b$  are hyper-priors that parametrize the Gamma prior. To see why the combination of Gaussian and Gamma priors will promote a sparse representation, consider marginalizing over the hyperparameters  $\boldsymbol{\alpha}$  to obtain the *marginal priors* acting on the signal coefficients (9,16,21)

$$p(\boldsymbol{\delta}_{i,j}^x) = \int p(\boldsymbol{\delta}_{i,j}^x | \alpha_j) p(\alpha_j | a, b) d\alpha_j \quad (2.14)$$

which turn out to yield improper priors of the form  $p(\boldsymbol{\delta}_{i,j}^x) \propto 1/|\boldsymbol{\delta}_{i,j}^x|$  in the particular case of uniform hyper-priors  $a = b = 0$ . Similar to the analysis for the Laplacian prior, this formulation would introduce an  $\ell_1$  regularizer of the form  $\sum_{j=1}^M \log|\boldsymbol{\delta}_{i,j}^x|$  if a non-joint MAP solution was sought for. Here, it should also be noted that the hyperparameters  $\boldsymbol{\alpha}$  are shared across the multi-contrast images, each  $\alpha_j$  controlling the variance of all  $L$  gradient coefficients  $\{\boldsymbol{\delta}_{i,j}^x\}_{i=1}^L$  through Eq. (2.12). In this case,  $\alpha_j$ 's diverging to infinity implies that the pixels in the  $j^{\text{th}}$  location of all images are zero, due to the zero-mean, zero-variance Gaussian prior at this location. On the other hand, a finite  $\alpha_j$  does not constrain all  $L$  pixels in the  $j^{\text{th}}$  location to be non-zero, which allows the reconstruction algorithm to capture the diversity of sparsity patterns across the multi-contrast scans.

In practice, the noise variance  $\sigma^2$  would also need to be estimated as it propagates via the data likelihood term to the posterior distribution of gradient coefficients (Eq. 2.5). Even though it is not difficult to obtain such an estimate in image domain if the full  $k$ -space data were available, this would not be straightforward with undersampled measurements. Therefore, following Ji *et al.* (9), the formulation is slightly modified so that the noise variance can be analytically integrated out while computing the posterior. This is made possible by including the noise precision  $\alpha_0 = \sigma^{-2}$  in the signal prior,

$$p(\boldsymbol{\delta}_i^x | \boldsymbol{\alpha}, \alpha_0) = \prod_{j=1}^M \mathcal{N}(\boldsymbol{\delta}_{i,j}^x | 0, \alpha_j^{-1} \alpha_0^{-1}) \quad (2.15)$$

A Gamma prior over the noise precision parameter  $\alpha_0$  is defined as

$$p(\alpha_0 | c, d) = Ga(\alpha_0 | c, d) = \frac{d^c}{\Gamma(c)} \alpha_0^{c-1} \exp(-d\alpha_0) \quad (2.16)$$

In all of the following experiments, the hyper-priors are set to  $c = d = 0$  to express that no *a priori* noise precision is favored as they lead to the “least informative” improper prior  $p(\alpha_0 | c = 0, d = 0) \propto 1/\alpha_0$ . The choice of priors in Eqs. (2.15-16) allows analytical computation of the posterior for the image gradients  $p(\delta_i^x | \mathbf{Y}_i^x, \alpha)$ , which turns out to be a multivariate Student- $t$  distribution with mean  $\boldsymbol{\mu}_i = \boldsymbol{\Sigma}_i \boldsymbol{\Phi}_i^T \mathbf{Y}_i^x$  and covariance  $\boldsymbol{\Sigma}_i = (\boldsymbol{\Phi}_i^T \boldsymbol{\Phi}_i + \mathbf{A})^{-1}$  with  $\mathbf{A} = \text{diag}(\alpha_1, \dots, \alpha_M)$ . This formulation is seen to allow robust coefficient shrinkage and information sharing thanks to inducing a heavy-tail in the posterior (9). It is worth noting that placing a Gamma prior on the noise precision does not change the additive nature of observation noise, however a heavier-tailed  $t$ -distribution replaces the normal density function in explaining this residual noise. This has been seen to be more resilient in allowing outlying measurements (9).

Now that an expression for the posterior  $p(\delta_i^x | \mathbf{Y}_i^x, \alpha)$  is obtained, the remaining work is to find a point estimate for the hyperparameters  $\alpha \in \mathbb{R}^M$  in a maximum likelihood (ML) framework. This is achieved by searching for the hyperparameter setting that makes the observation of the  $k$ -space data most likely, and such an optimization process is called *evidence maximization* or *type-II maximum likelihood* method (9,16,21). Therefore, the hyperparameters that maximize

$$\mathcal{L}(\alpha) = \sum_{i=1}^L p(\mathbf{Y}_i^x | \alpha) = \sum_{i=1}^L \int p(\alpha_0 | a, b) p(\delta_i^x | \alpha, \alpha_0) p(\mathbf{Y}_i^x | \delta_i^x, \alpha_0) d\delta_i^x d\alpha_0 \quad (2.17)$$

are sought for. It should be noted that data from all  $L$  tasks contribute to the evidence maximization procedure via the summation over conditional distributions. Hence, the information sharing across the images occurs through this collaboration in the maximum likelihood estimation of the hyperparameters. Once the point estimates are constituted using all of the observations, the posterior for the signal coefficients  $\delta_i^x$  is estimated based only on its related  $k$ -space data  $\mathbf{Y}_i^x$  due to  $\boldsymbol{\mu}_i = \boldsymbol{\Sigma}_i \boldsymbol{\Phi}_i^T \mathbf{Y}_i^x$ . Thus, all of the measurements are used in the estimation of the hyperparameters, but only the associated data are utilized to constitute an approximation to the gradient coefficients.

Ji *et al.* show that it is possible to maximize Eq. (2.17) with a sequential greedy algorithm, in which the starting point is a single basis vector for each signal, then the basis function that yields the largest increase in the log likelihood is added at each iteration. Alternatively, a hyperparameter corresponding to a basis vector that is already in the dictionary of current bases can be updated or deleted, if this gives rise to the largest increase in the likelihood at that

iteration. A final refinement to Ji *et al.*'s Bayesian CS algorithm is added by replacing the observation matrices  $\{\Phi_i\}_{i=1}^L$  that are needed to be stored with the Fast Fourier Transform (FFT). This enables working with MRI images of practical sizes; otherwise each of the observation matrices would occupy 32GB of memory for a 256×256 image. The reader is referred to Appendix B in (9) for the update equations of this algorithm.

### 2.2.5 Reconstructing the Images from Horizontal and Vertical Gradient Estimates

Once the image gradients  $\{\delta_i^x\}_{i=1}^L$  and  $\{\delta_i^y\}_{i=1}^L$  are estimated with the joint Bayesian algorithm, the images  $\{\mathbf{x}_i\}_{i=1}^L$  consistent with these gradients and the undersampled measurements  $\{\mathbf{Y}_i\}_{i=1}^L$  need to be found. Influenced by (19), this is formulated as a least squares (LS) optimization problem

$$\hat{\mathbf{x}}_i = \arg \min_{\mathbf{x}_i} \left\| \partial_x \mathbf{x}_i - \delta_i^x \right\|_2^2 + \left\| \partial_y \mathbf{x}_i - \delta_i^y \right\|_2^2 + \lambda \left\| \mathbf{F}_{\Omega_i} \mathbf{x}_i - \mathbf{Y}_i \right\|_2^2 \quad (2.18)$$

for  $i=1, \dots, L$  where  $\partial_x \mathbf{x}_i$  and  $\partial_y \mathbf{x}_i$  represent vertical and horizontal image gradients. Using Eqs. (2.7) and (2.8) and invoking Parseval's Theorem, the optimization problem can be cast into  $k$ -space

$$\hat{\mathbf{X}}_i = \arg \min_{\mathbf{x}_i} \left\| (1 - e^{-2\pi j \omega / n}) \mathbf{X}_i - \mathbf{A}_i^x \right\|_2^2 + \left\| (1 - e^{-2\pi j \nu / m}) \mathbf{X}_i - \mathbf{A}_i^y \right\|_2^2 + \lambda \left\| \mathbf{X}_{\Omega_i} - \mathbf{Y}_i \right\|_2^2 \quad (2.19)$$

where  $\mathbf{X}_i$ ,  $\mathbf{A}_i^x$  and  $\mathbf{A}_i^y$  are the Fourier transforms of  $\mathbf{x}_i$ ,  $\delta_i^x$  and  $\delta_i^y$ , respectively and  $\mathbf{X}_{\Omega_i}$  is the transform of  $\mathbf{x}_i$  restricted to the frequency set  $\Omega_i$ . Based on this, the following solution is found by representing Eq. (2.19) as a quadratic polynomial and finding the root with  $\lambda \rightarrow \infty$

$$\hat{\mathbf{X}}_i(\omega, \nu) = \begin{cases} \mathbf{X}_{\Omega_i} & \text{if } (\omega, \nu) \in \Omega_i \\ \frac{(1 - e^{2\pi j \omega / n}) \mathbf{A}_i^x + (1 - e^{2\pi j \nu / m}) \mathbf{A}_i^y}{|1 - e^{-2\pi j \omega / n}|^2 + |1 - e^{-2\pi j \nu / m}|^2} & \text{otherwise} \end{cases} \quad (2.20)$$

Finally, taking the inverse Fourier transform gives the reconstructed images  $\{\hat{\mathbf{x}}_i\}_{i=1}^L$ .

### 2.2.6 Extension to Complex-Valued Images

In the general case where the underlying multi-contrast images are complex-valued, the linear observation model of Eq. (2.9) is no longer valid. Under the assumption that the support of the

frequency set  $\Omega_i$  is symmetric, it is possible to decouple the undersampled  $k$ -space observations belonging to the real and imaginary parts of the signals,

$$\text{if } \text{supp}(\Omega_i[k_x, k_y]) = \text{supp}(\Omega_i[(-k_x, -k_y)]) , \quad (2.21)$$

$$\mathbf{y}_i^{\text{Re}} \square \mathbf{F}_{\Omega_i} \mathcal{R}\mathbf{e}(\mathbf{x}_i) = \frac{1}{2} \cdot (\mathbf{y}_i[k_x, k_y] + \mathbf{y}_i^* [(-k_x, -k_y)]) \quad (2.22)$$

$$\mathbf{y}_i^{\text{Im}} \square \mathbf{F}_{\Omega_i} \mathcal{I}\mathbf{m}(\mathbf{x}_i) = \frac{-j}{2} \cdot (\mathbf{y}_i[k_x, k_y] - \mathbf{y}_i^* [(-k_x, -k_y)]) \quad (2.23)$$

Here,  $[k_x, k_y]$  index the frequency space and  $\mathbf{y}_i^* [(-k_x, -k_y)]$  is the complex conjugate of index-reversed  $k$ -space observations. In the case of one dimensional undersampling, the constraint on  $\Omega_i$  would simply correspond to an undersampling pattern that is mirror-symmetric with respect to the line passing through the center of  $k$ -space. After obtaining the  $k$ -space data  $\mathbf{y}_i^{\text{Re}}$  and  $\mathbf{y}_i^{\text{Im}}$  belonging to the real and imaginary parts of the  $i^{\text{th}}$  image  $\mathbf{x}_i$ ,  $\mathcal{R}\mathbf{e}(\mathbf{x}_i)$  and  $\mathcal{I}\mathbf{m}(\mathbf{x}_i)$  are solved for jointly in the gradient domain, in addition to the joint inversion of multi-contrast data, hence exposing a second level of simultaneous sparsity in the image reconstruction problem. Final reconstructions are then obtained by combining the real and imaginary channels into complex-valued images.

## 2.3 Methods

To demonstrate the inversion performance of the joint Bayesian CS algorithm, three data sets that include a numerical phantom, the SRI24 brain atlas, and *in vivo* acquisitions, were reconstructed from undersampled  $k$ -space measurements belonging to the *magnitude images*. In addition, two datasets including a numerical phantom and *in vivo* multi-contrast slices, both consisting of complex-valued images, were also reconstructed from undersampled measurements to test the performance of the method with complex-valued image-domain signals. The results were quantitatively compared against the popular implementation by Lustig *et al.* (6), which does not make use of joint information across the images, as well as the M-FOCUSS algorithm, which is an alternative joint CS reconstruction algorithm.

### 2.3.1 CS Reconstruction with Extended Shepp-Logan Phantoms

To generalize the Shepp-Logan phantom to the multi-contrast setting, two additional phantoms were generated by randomly permuting the intensity levels in the original  $128 \times 128$  image. Further, by placing 5 more circles with radii chosen randomly from an interval of  $[7, 13]$  pixels and intensities selected randomly from  $[0.1, 1]$  to the new phantoms, the idiosyncratic portions of the scans were aimed to be represented with different weightings. A variable-density undersampling scheme in  $k$ -space was applied by drawing three fresh samples from a power law density function, so that the three masks' frequency coverage was only partially overlapping. Power law sampling indicates that the probability of sampling a point in  $k$ -space is inversely proportional to the distance of that point to the center of  $k$ -space, which makes the vicinity of the center of  $k$ -space more densely sampled. To realize this pattern, again Lustig *et al.*'s software package (6) was used, which randomly generates many sampling patterns and retains the one that has the smallest sidelobe-to-peak ratio in the point spread function. This approach aims to create a sampling pattern that induces optimally incoherent aliasing artifacts (6). A high acceleration factor of  $R = 14.8$  was tested using the joint Bayesian CS, Lustig *et al.*'s gradient descent and the M-FOCUSS algorithm. For the gradient descent method, using wavelet and TV norm penalties were seen to yield better results than using only one of them. In all experiments, all combinations of regularization parameters  $\lambda_{TV}$  and  $\lambda_{wavelet}$  from the set  $\{10^{-4}, 10^{-3}, 10^{-2}, 0\}$  were tested and the setting that gave the smallest reconstruction error was retained as the optimal one. In the Shepp-Logan experiment, the parameter setting  $\lambda_{TV} = \lambda_{wavelet} = 10^{-3}$  was seen to yield optimal results for the gradient descent method. The number of iterations was taken to be 50 in all of the examples. The Bayesian algorithm continues the iterations until convergence, which is determined by

$$|\Delta \ell_k - \Delta \ell_{k-1}| < (\Delta \ell_{\max} - \Delta \ell_k) \cdot \eta \quad (2.24)$$

where  $\Delta \ell_k$  is the change in log likelihood at iteration  $k$  and  $\Delta \ell_{\max}$  is the maximum change in likelihood that has been encountered in all  $k$  iterations. The convergence parameter  $\eta$  was taken to be  $10^{-8}$  in this example. For the M-FOCUSS method, each image was undersampled with the same mask as phantom 1 in the joint Bayesian CS since M-FOCUSS does not admit different observation matrices.

### 2.3.2 SRI24 Multi-Channel Brain Atlas Data

This experiment makes use of the multi-contrast data extracted from the SRI24 atlas (22). The atlas features structural scans obtained with three different contrast settings at 3T,

- i. Proton density weighted images: obtained with a 2D axial dual-echo fast spin echo (FSE) sequence (TR = 10000 ms, TE = 14 ms)
- ii. T2 weighted images: acquired with the same sequence as the proton density weighted scan, except with TE = 98 ms.
- iii. T1 weighted images: acquired with a 3D axial IR-prep Spoiled Gradient Recalled (SPGR) sequence (TR = 6.5 ms, TE = 1.54 ms)

The atlas images have a resolution of 256×256 pixels and cover a 24-cm field-of-view (FOV). Since all three data sets are already registered spatially, no post-processing was applied except for selecting a single axial slice from the atlas. Prior to reconstruction, retrospective undersampling<sup>1</sup> was applied along the phase encoding direction with acceleration  $R = 4$  using a different undersampling mask for each image. Again a power law density function was utilized in selecting the sampled  $k$ -space lines. In this case, a 1-dimensional pdf was employed, so that it was more likely to acquire phase encoding lines close to the center of  $k$ -space. Reconstructions were performed using Lustig *et al.*'s conjugate gradient descent algorithm (with  $\lambda_{TV} = \lambda_{wavelet} = 10^{-3}$ ), joint Bayesian method (with  $\eta = 10^{-9}$ ) and the M-FOCUSS joint reconstruction algorithm.

### 2.3.3 3T Turbo Spin Echo (TSE) Slices with Early and Late TE's

T2-weighted axial multi-slice images of the brain of a young healthy male volunteer were obtained with two different TE settings using a TSE sequence (256×256 pixel resolution with 38 slices, 1×1 mm in-plane spatial resolution with 3 mm thick contiguous slices, TR = 6000 ms, TE<sub>1</sub> = 27 ms, TE<sub>2</sub> = 94 ms). Out of these, a single image slice was selected and its magnitude was retrospectively undersampled in  $k$ -space along the phase encoding direction with acceleration  $R = 2.5$  using a different mask for each image, again by sampling lines due to a 1-dimensional power law distribution. The images were reconstructed using Lustig *et al.*'s algorithm with an optimal

---

<sup>1</sup> We use the retrospective undersampling phrase to indicate that  $k$ -space samples are discarded synthetically from data obtained at Nyquist rate in software environment, rather than skipping samples during the actual scan.

parameter setting ( $\lambda_{TV} = \lambda_{wavelet} = 10^{-3}$ ), joint Bayesian CS algorithm (with  $\eta = 10^{-9}$ ) and the M-FOCUSS method.

### 2.3.4 Complex-Valued Shepp-Logan Phantoms

Using four numerical phantoms derived from the original Shepp-Logan phantom, two complex valued numerical phantoms were generated by combining the four images in real and imaginary pairs. Retrospective undersampling was applied along the phase encoding direction with acceleration  $R = 3.5$  using a different undersampling mask for each image. A 1-dimensional power law density function was utilized in selecting the sampled  $k$ -space lines, making it more likely to acquire phase encoding lines close to the center of  $k$ -space. Again many sampling patterns were randomly generated and the one that has the smallest sidelobe-to-peak ratio in the point spread function was retained, but also the sampling masks were constrained to be mirror-symmetric with respect to the center of  $k$ -space. This way, it was possible to obtain the undersampled  $k$ -space data belonging to the real and imaginary channels of the phantoms separately. The images were reconstructed using Lustig *et al.*'s algorithm ( $\lambda_{TV} = \lambda_{wavelet} = 10^{-3}$ ), joint Bayesian CS algorithm (reconstructing real & imaginary parts together, in addition to joint multi-contrast reconstruction) and the M-FOCUSS method. Further, non-joint reconstructions with the Bayesian CS method (doing a separate reconstruction for each image, but reconstructing real & imaginary channels of each image jointly) and the FOCUSS algorithm (non-joint version of M-FOCUSS) were conducted for comparison with Lustig *et al.*'s approach.

### 2.3.5 Complex-Valued Turbo Spin Echo Slices with Early and Late TE's

To test the performance of the algorithms on complex-valued *in vivo* images, axial multi-slice images of the brain of a young healthy female subject were obtained with two different TE settings using a TSE sequence (128×128 pixel resolution with 38 slices, 2×2 mm in-plane spatial resolution with 3 mm thick contiguous slices, TR = 6000 ms, TE<sub>1</sub> = 17 ms, TE<sub>2</sub> = 68 ms). Data were acquired with a body coil and both the magnitude and the phase of the images were recorded. To enhance SNR, 5 averages and a relatively large 2-mm in-plane voxel size were used. A single slice was selected from the dataset and its raw  $k$ -space data were retrospectively undersampled along the phase encoding direction with acceleration  $R = 2$  using a different mask for each image, again by sampling lines due to a 1-dimensional power law distribution. For the complex-valued image-domain case, the masks were constrained to be symmetric with respect to the line passing through the center of  $k$ -space. The images were reconstructed using Lustig *et al.*'s

algorithm ( $\lambda_{TV} = \lambda_{wavelet} = 10^{-3}$ ), our joint Bayesian CS algorithm (reconstructing real & imaginary parts and multi-contrasts together) and the M-FOCUSS method. In addition, non-joint reconstructions with the Bayesian CS method (using a separate reconstruction for each image, but reconstructing real & imaginary parts of each image together) and the FOCUSS algorithm were performed.

## 2.4 Results

### 2.4.1 CS Reconstruction with Extended Shepp-Logan Phantoms

Fig. 2.2 presents the reconstruction results for the three algorithms for the extended phantoms, along with the  $k$ -space masks used in retrospective undersampling. At acceleration  $R = 14.8$ , the Bayesian algorithm obtained perfect recovery of the noise-free numerical phantom, whereas the gradient descent algorithm by Lustig *et al.* returned 15.9 % root mean squared error (RMSE), which we define as

$$\mathbf{RMSE} = 100 \cdot \frac{\|\mathcal{Re}(\hat{\mathbf{x}}) - \mathbf{x}\|_2}{\|\mathbf{x}\|_2} \quad (2.25)$$

where  $\mathbf{x}$  is the vector obtained by concatenating all  $L$  images together, and similarly  $\hat{\mathbf{x}}$  is the concatenated vector of all  $L$  reconstructions produced by an inversion algorithm. The M-FOCUSS joint reconstruction algorithm yielded an error of 8.8 %. The reconstruction times were measured to be 5 minutes for gradient descent, 4 minutes for M-FOCUSS and 25 minutes for the joint Bayesian CS algorithm.

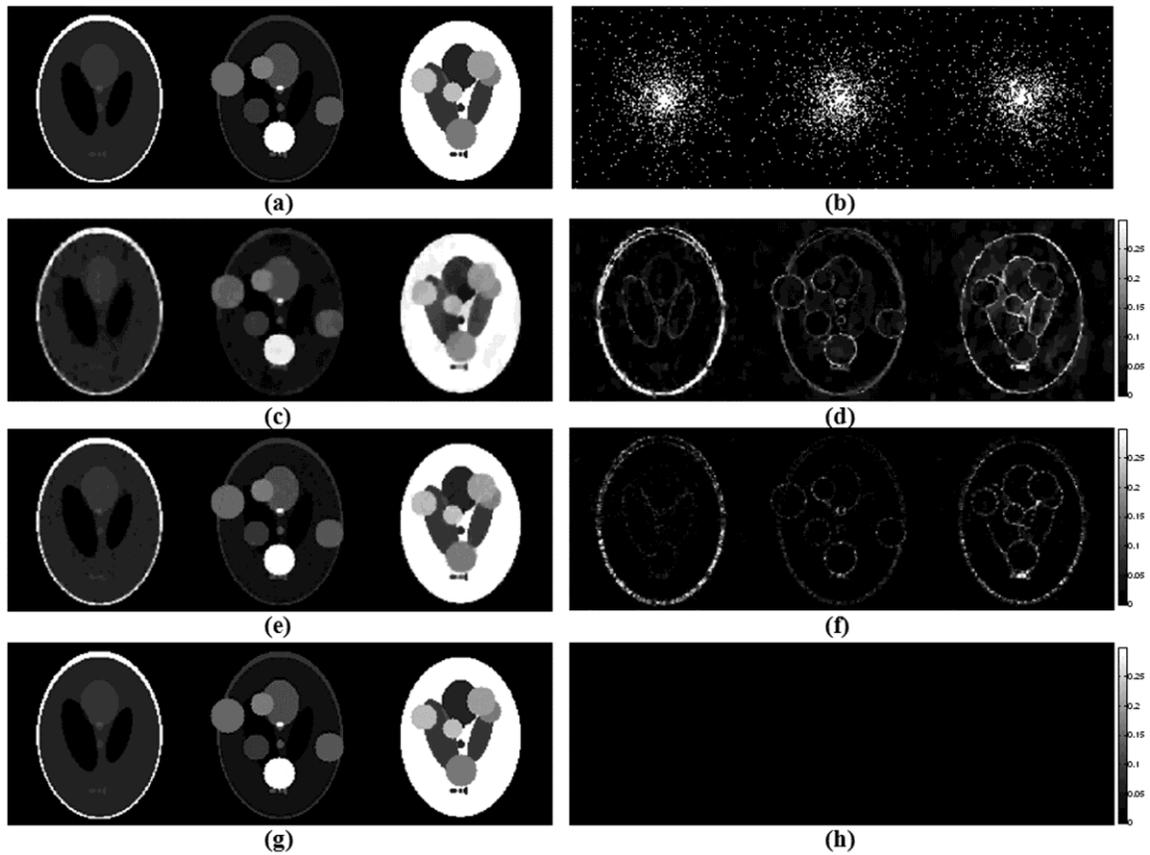


Fig 2.2 Reconstruction results with the extended Shepp-Logan phantoms after undersampling with acceleration  $R = 14.8$ , at  $128 \times 128$  resolution. (a) Phantoms at Nyquist rate sampling. (b) Undersampling patterns in  $k$ -space corresponding to each image. (c) CS reconstructions with Lustig *et al.*'s algorithm yielded 15.9 % RMSE (root-mean-square error). (d) Absolute error plots for Lustig *et al.*'s method. (e) Reconstructions obtained with the M-FOCUSS joint reconstruction algorithm have 8.8 % RMSE. (f) Absolute difference between the Nyquist sampled phantoms and the M-FOCUSS reconstruction results. (g) Joint Bayesian CS reconstruction resulted in 0 % RMSE. (h) Absolute error plots for the Bayesian CS reconstructions.

#### 2.4.2 SRI24 Multi-Channel Brain Atlas Data

The results for reconstruction upon phase encoding undersampling with acceleration  $R = 4$  are given in Fig. 2.3. In this case, Lustig *et al.*'s algorithm returned 9.4 % RMSE, while the error was 3.2 % and 2.3 % for M-FOCUSS and joint Bayesian CS methods, respectively. The reconstructions took 43 minutes for gradient descent, 5 minutes for M-FOCUSS and 26.4 *hours* for the Bayesian CS algorithm.

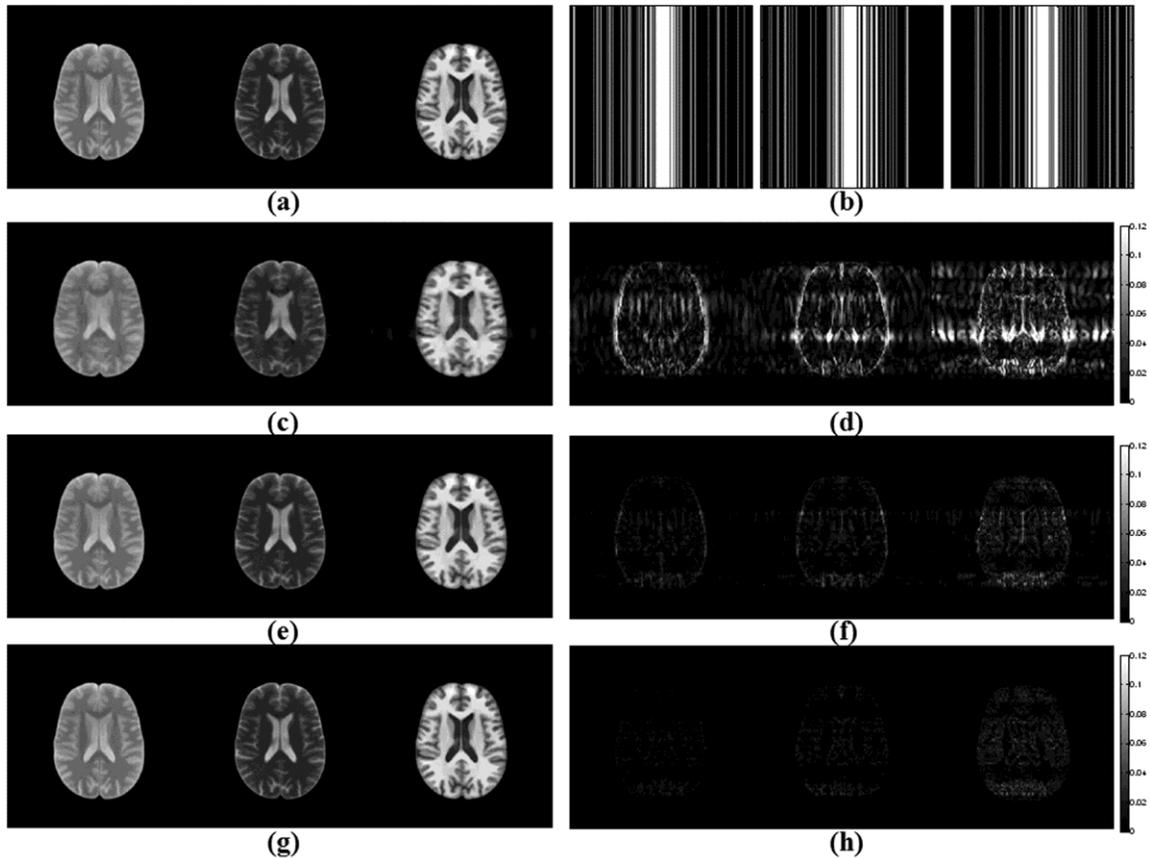


Fig. 2.3. Reconstruction results with SRI24 atlas after undersampling along the phase encoding direction with  $R = 4$ , at  $256 \times 256$  resolution. (a) Atlas images at Nyquist rate sampling. (b) Undersampling patterns in  $k$ -space corresponding to each image. (c) Applying the gradient descent algorithm proposed by Lustig *et al.* resulted in reconstructions with 9.4 % RMSE. (d) Absolute difference between the gradient descent reconstructions and the Nyquist rate images. (e) M-FOCUSS reconstructions have 3.2 % RMSE. (f) Absolute error plots for the M-FOCUSS algorithm. (g) Joint Bayesian reconstruction yielded images with 2.3 % RMSE. (h) Error plots for the joint Bayesian reconstructions.

### 2.4.3 Turbo Spin Echo (TSE) Slices with Early and Late TE's

Fig. 2.4 depicts the TSE reconstruction results obtained with the three algorithms after undersampling along phase encoding with acceleration  $R = 2.5$ . In this setting, Lustig *et al.*'s code returned a result with 9.4 % RMSE, whereas M-FOCUSS and joint Bayesian reconstruction had 5.1 % and 3.6 % errors, respectively. The total reconstruction times were 26 minutes for gradient descent, 4 minutes for M-FOCUSS and 29.9 hours for the Bayesian CS algorithm.

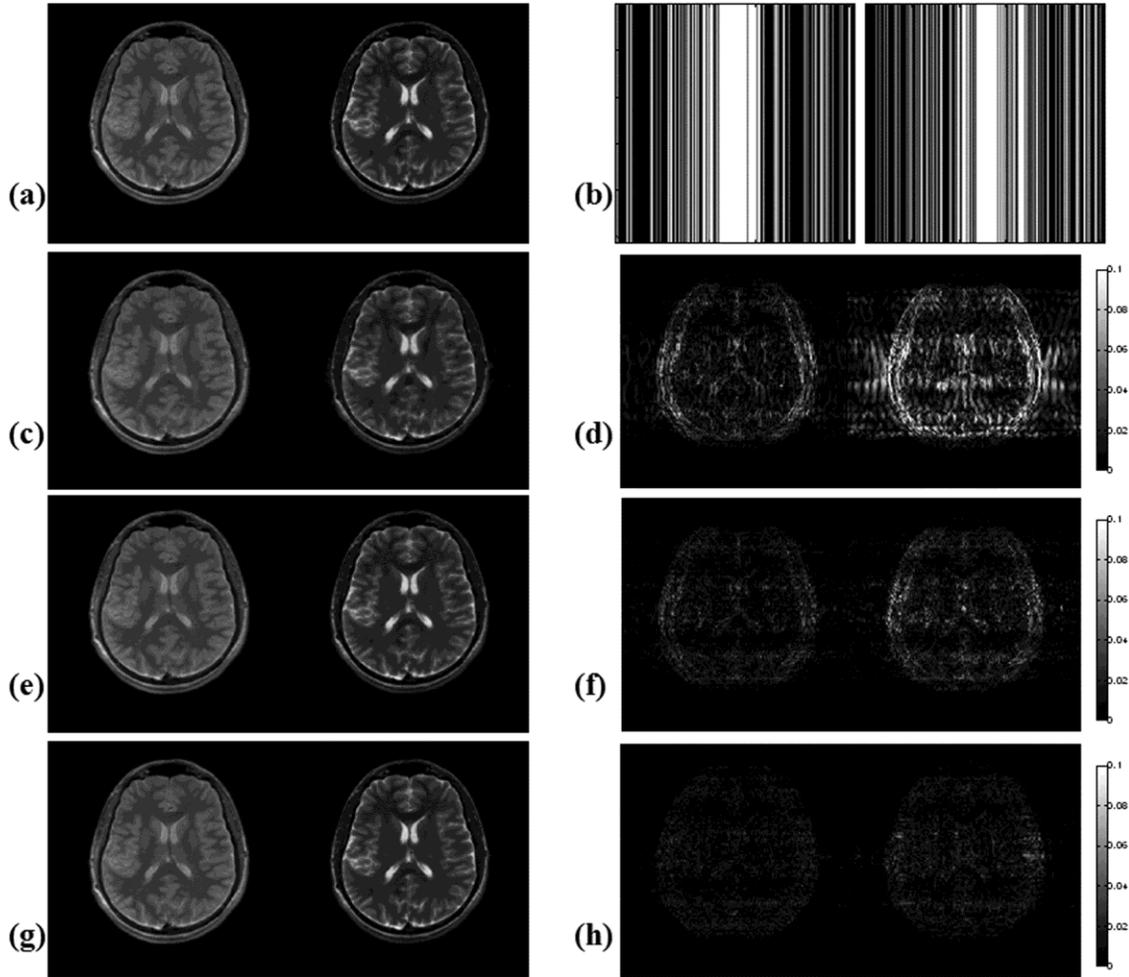


Fig. 2.4. Reconstruction results with TSE after undersampling along the phase encoding direction with  $R = 2.5$ , at  $256 \times 256$  resolution. (a) TSE scans at Nyquist rate sampling. (b) Undersampling patterns used in this experiment. (c) Reconstructions obtained with Lustig *et al.*'s gradient descent algorithm have 9.4 % RMSE. (d) Plots of absolute error for the gradient descent reconstructions. (e) M-FOCUSS joint reconstruction yielded images with 5.1 % RMSE. (f) Error plots for the M-FOCUSS results. (g) Images obtained with the joint Bayesian CS reconstruction returned 3.6 % RMSE. (h) Error plots for the Bayesian CS reconstructions.

These results are also included in Table 2.1 as “*PE, (Fig. 4)*” for comparison with reconstruction using the same undersampling pattern.

For brevity, additional results are presented in Table 2.1 from more extensive tests in which various undersampling patterns and accelerations were employed. To test the algorithms' performance at a different resolution, the TSE and atlas images were downsampled to size  $128 \times 128$  prior to undersampling, and similar RMSE results as the high resolution experiments

were noted. The table also includes an experiment with  $256 \times 256$  TSE scans accelerated along the phase encoding with  $R = 2.5$ , but using the *same* undersampling pattern for both images.

Dataset	Resolution	Undersampling method	Acceleration factor $R$	RMSE %		
				Lustig <i>et al.</i>	M-Focuss	Bayesian CS
TSE	256×256	Phase encoding (PE)	3	9.7	6.8	5.8
	256×256	Power law	6	8.1	7.8	6.3
	<b>256×256</b>	<b>PE (Fig. 2.4)</b>	<b>2.5</b>	<b>9.4</b>	<b>5.1</b>	<b>3.6</b>
	<b>256×256</b>	<b>PE, same pattern</b>	<b>2.5</b>			<b>4.7</b>
	128×128	PE	2	8.1	3.8	2.1
SRI 24	256×256	Radial	9.2	6.0	4.5	3.0
	128×128	PE	3	7.2	4.2	3.1

Table 2.1. Summary of additional reconstruction results on the TSE and SRI 24 datasets using the three algorithms after retrospective undersampling with various patterns and acceleration factors.

#### 2.4.4 Impact of Spatial Misregistration on Joint Reconstruction

Due to aliasing artifacts caused by undersampling, image registration prior to CS reconstruction across multi-contrast images is likely to perform poorly. The effect of spatial misalignments was investigated by shifting one of the images in the TSE dataset relative to the other by 0 to 2 pixels with step sizes of  $\frac{1}{2}$  pixels using two different undersampling patterns. The first pattern incurs  $R = 3$  acceleration by 2D undersampling with  $k$ -space locations drawn from a power law probability distribution. In this case, the effect of vertical misalignments was tested. The second pattern undersamples  $k$ -space at  $R = 2.5$  in the phase encoding direction, for which horizontal dislocations were tested. For speed, low resolution images at size  $128 \times 128$  were used. M-FOCUSS and joint Bayesian CS methods were tested for robustness against misregistration and that the effect of spatial misalignment was observed to be mild for both (Fig. 2.5). Even though Bayesian CS consistently had less reconstruction errors relative to M-FOCUSS on both undersampling patterns at all dislocations, the performance of M-FOCUSS was seen to change less relative to Bayesian CS with respect to the incurred translations. For joint Bayesian CS, reconstruction error increased from 2.1 % to 2.8 % at 2 pixels of vertical shift for power law sampling, and from 5.2 % to 6.4 % at 2 pixels of horizontal shift for phase encoding sampling; for the M-FOCUSS method error increased from 4.7 % to 4.9 % for power law sampling, and from 6.2 % to 6.6 % for phase encoding sampling.

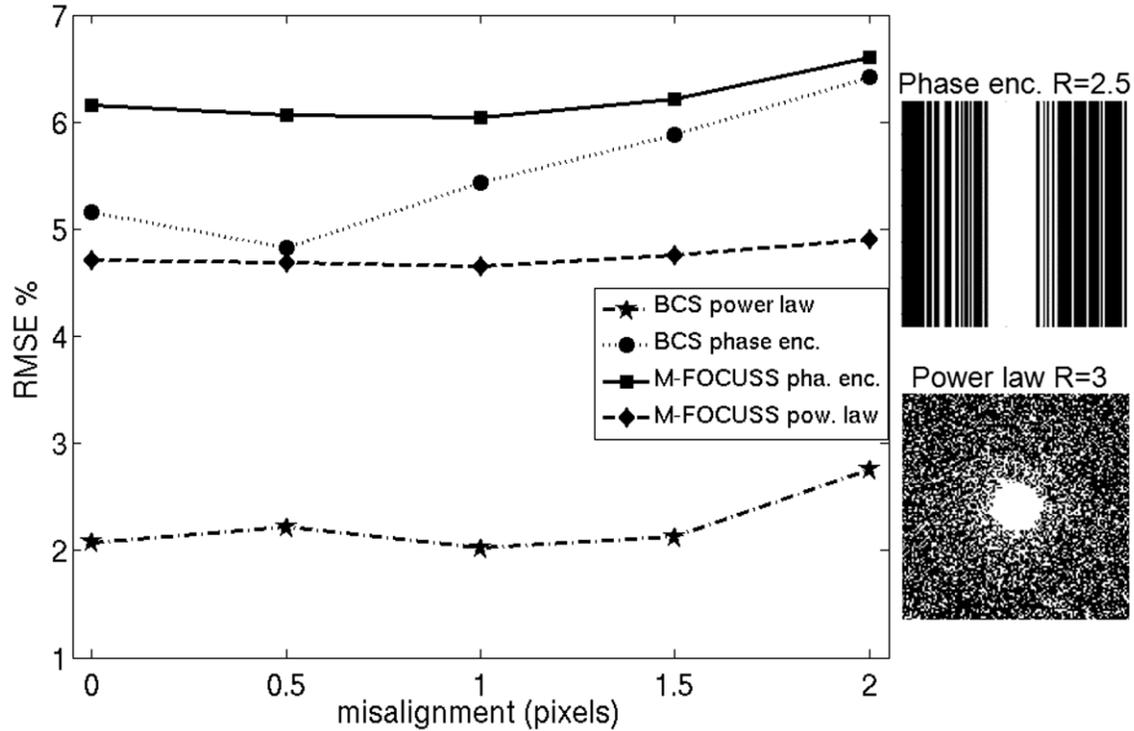


Fig. 2.5. To investigate the impact of spatial misalignments on joint reconstruction with Bayesian CS and M-FOCUSS, one of the TSE images was shifted relative to the other by 0 to 2 pixels with step sizes of  $\frac{1}{2}$  pixels using power law and phase encoding undersampling patterns. For speed, low resolution images with size  $128 \times 128$  were used. For joint Bayesian CS, reconstruction error increased from 2.1 % to 2.8 % at 2 pixels of vertical shift for power law sampling, and from 5.2 % to 6.4 % at 2 pixels of horizontal shift for phase encoding sampling; for the M-FOCUSS method error increased from 4.7 % to 4.9 % for power law sampling, and from 6.2 % to 6.6 % for phase encoding sampling.

#### 2.4.5 Complex-Valued Shepp-Logan Phantoms

Absolute values of the reconstruction results after undersampling with a symmetric mask with  $R = 3.5$  for the complex-valued phantoms are depicted in Fig. 2.6. For complex signals, the error metric  $RMSE = 100 \cdot \|\hat{\mathbf{x}} - \mathbf{x}\|_2 / \|\mathbf{x}\|_2$  is used. In this case, Lustig *et al.*'s algorithm returned a result with 13.1 % RMSE, whereas joint reconstructions with M-FOCUSS and joint Bayesian methods had 5.4 % and 2.4 % errors, respectively. The total reconstruction times were 21 minutes for gradient descent, 0.5 minutes for M-FOCUSS and 18 minutes for the Bayesian CS algorithm. On the other hand, reconstructing each complex-valued image separately with FOCUSS and Bayesian CS yielded 6.7 % and 4.6 % RMSE.

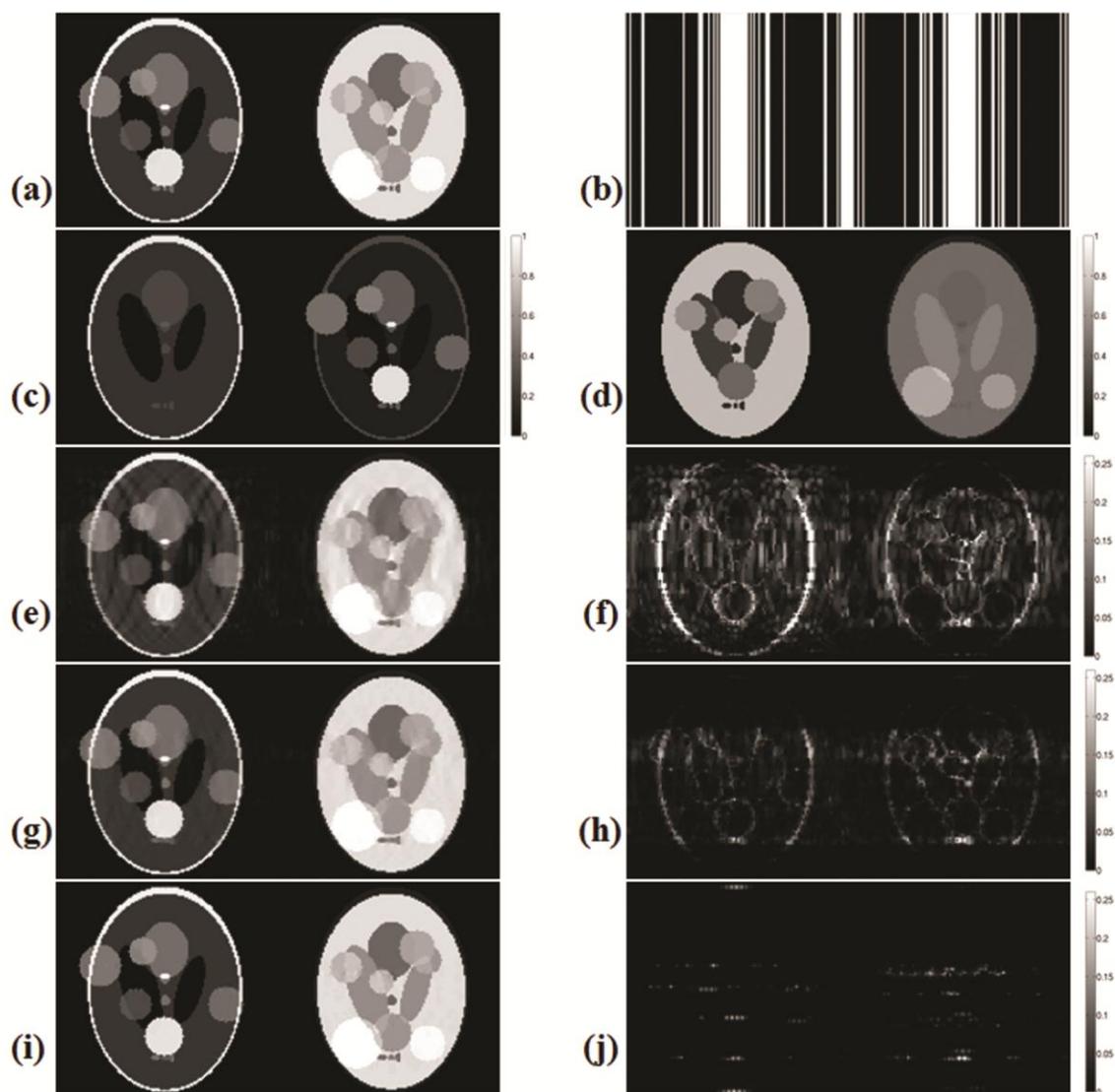


Fig. 2.6. Reconstruction results with the complex-valued Shepp-Logan phantoms after undersampling with acceleration  $R = 3.5$ , at  $128 \times 128$  resolution. (a) Magnitudes of phantoms at Nyquist rate sampling. (b) Symmetric undersampling patterns in  $k$ -space corresponding to each image. (c) Real and imaginary parts of the first phantom (on the left in (a)). (d) Real and imaginary parts of the second phantom (on the right in (a)). (e) CS reconstructions with Lustig *et al.*'s algorithm yielded 13.1 % RMSE. (f) Absolute error plots for Lustig *et al.*'s method. (g) Reconstructions obtained with the M-FOCUSS joint reconstruction algorithm have 5.4 % RMSE. (h) Absolute difference between the Nyquist sampled phantoms and the M-FOCUSS reconstruction results. (i) Joint Bayesian CS reconstruction resulted in 2.4 % RMSE. (j) Absolute error plots for the Bayesian CS reconstructions.

### 2.4.6 Complex-Valued Turbo Spin Echo Slices with Early and Late TE's

Reconstruction results are compared in Fig. 2.7 for the discussed algorithms. Lustig *et al.*'s method had 8.8 % error upon acceleration by  $R = 2$  with a symmetric pattern, whereas the joint reconstruction algorithms M-FOCUSS and joint Bayesian CS yielded 9.7 % and 6.1 % RMSE. The processing times were 20 minutes for gradient descent, 2 minutes for M-FOCUSS and 5.2 hours for the Bayesian CS algorithm. Non-joint reconstructions with FOCUSS and Bayesian CS returned 10.0 % and 8.6 % errors.

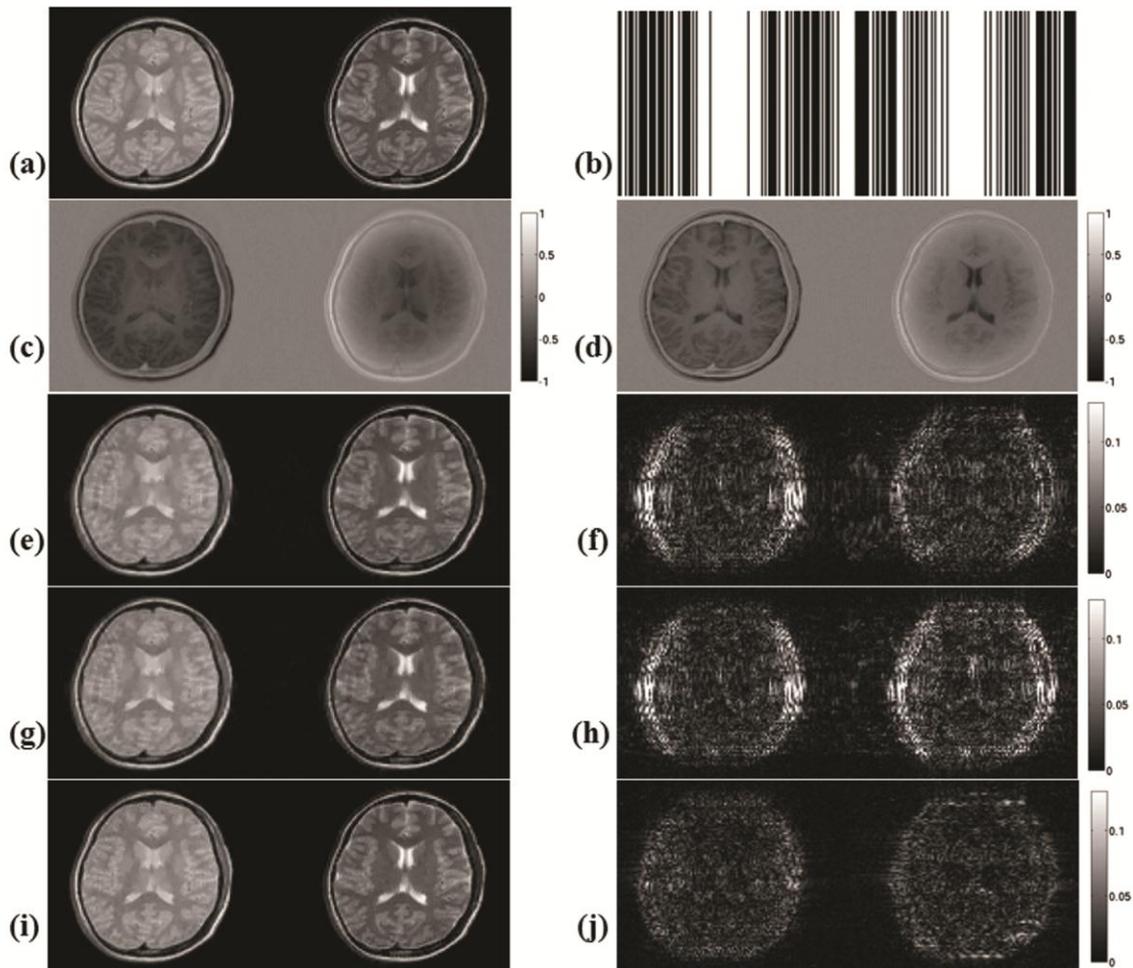


Fig. 2.7. Reconstruction results for complex-valued TSE images after undersampling along the phase encoding direction with  $R = 2$ , at  $128 \times 128$  resolution. (a) Magnitudes of the TSE scans at Nyquist rate sampling. (b) Symmetric undersampling patterns used in this experiment. (c) Real and imaginary parts of the early echo image (on the left in (a)). (d) Real and imaginary parts of the late echo image (on the right in (a)). (e) Reconstructions obtained with Lustig *et al.*'s gradient descent algorithm have 8.8 % RMSE. (d) Plots of absolute error for the gradient descent reconstructions. (e) M-FOCUSS joint reconstruction yielded images with 9.7 % RMSE. (f) Error plots for the M-FOCUSS results. (g) Images obtained with the joint Bayesian CS reconstruction returned 6.1 % RMSE. (h) Error plots for the Bayesian CS reconstructions.

With the same dataset, additional reconstructions were performed to quantify the effect of the symmetry constraint on the sampling masks. Both of the late and early TE images were reconstructed 5 times with freshly generated, random masks with  $R = 2$  (no symmetry constraints) and also 5 times with freshly generated symmetric masks again at  $R = 2$ . Using Lustig *et al.*'s method ( $\lambda_{TV} = 10^{-3}$ ) with the random masks yielded an average error of 10.5 %, whereas using symmetric masks incurred an average error of 11.5 %.

## 2.5 Discussion

The application of joint Bayesian CS MRI reconstruction to images of the same object acquired under different contrast settings was demonstrated to yield substantially higher reconstruction fidelity than either Lustig *et al.*'s (non-joint) algorithm or joint M-FOCUSS, but at the cost of substantially increased reconstruction times in this initial implementation. In contrast to M-FOCUSS, the proposed algorithm allows for different sampling matrices being applied to each contrast setting and unlike the gradient descent method, it has no parameters that need adjustments. The success of this algorithm is based on the premise that the multi-contrast scans of interest share a set of similar image gradients while each image may also present additional unique features with its own image gradients. In Fig. 2.8 the vertical image gradients belonging to the TSE scans are presented, where a simple experiment was conducted to quantify the similarity between them. After sorting the image gradient magnitudes of the early TSE scan in descending order, the cumulative energy in them was computed. Next, the late TSE gradient magnitude was sorted in descending order and the cumulative energy in the *early* TSE gradient was calculated by using the pixel index order belonging to the *late* TSE scan. This cumulative sum reached 95 % of the original energy, thus confirming the visual similarity of the two gradients.

It is important to note that in the influential work by Ji *et al.* (9), the authors also consider joint reconstruction of MRI images. However their dataset consists of five different slices taken from the *same scan*, so the motivation for their MRI work is different from what is presented here. Even though the multislice images have considerable similarity from one slice to the next, one would expect multi-contrast scans to demonstrate a yet higher correlation of image features and a correspondingly larger benefit in reconstruction fidelity.

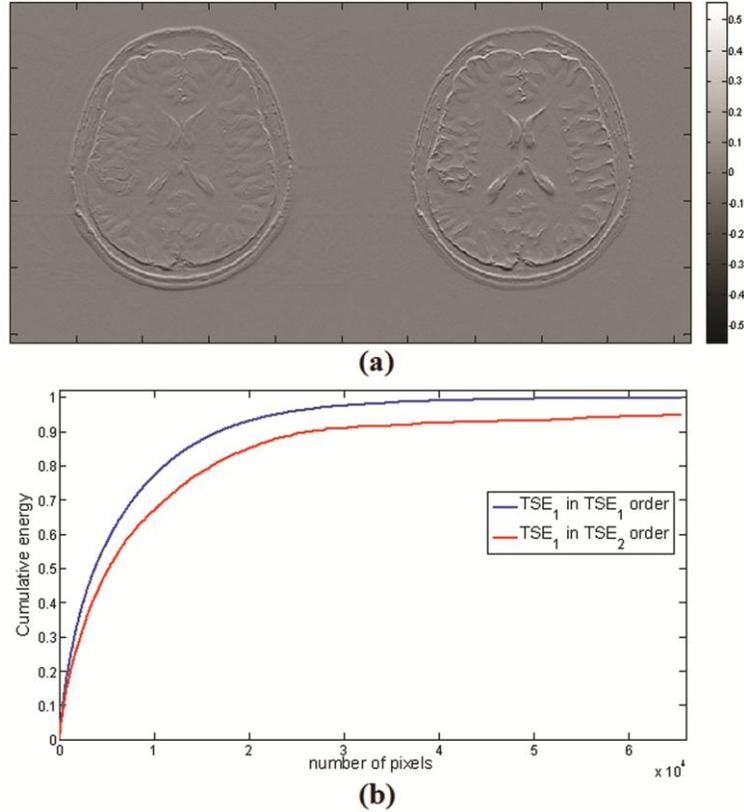


Fig. 2.8. (a) Image gradients for the multi-contrast TSE scans demonstrate the similarity under the gradient transform. (b) To quantify this similarity, we computed the cumulative energy of the image gradient of early TSE scan ( $TSE_1$  in  $TSE_1$  order). Then we sorted the late TSE scan ( $TSE_2$ ) in descending order, and computed the cumulative energy in  $TSE_1$  corresponding to the sorted indices in  $TSE_2$  which gave the curve  $TSE_1$  in  $TSE_2$  order. The similarity of the curves indicates similar sparsity supports across images.

Two aspects of the proposed Bayesian reconstruction algorithm demand further attention. First, relative to the other two algorithms we investigated, the Bayesian method is dramatically more time consuming. The reconstruction times can be on the order of hours, which is prohibitive for clinical use as currently implemented. As detailed in the Results section, the proposed algorithm is about 40 times slower than gradient descent, and about 300 times slower than M-FOCUSS for the *in vivo* data. Future implementations and optimizations that utilize specialized scientific computation hardware are expected to overcome this current drawback. Particularly, it is common to observe an order of magnitude speed-up with CUDA (Compute Unified Device Architecture) enabled Graphics Processing Units when the problem under consideration can be adapted to the GPU architecture (23). In a recent work, using CUDA architecture in compressed sensing was reported to yield accelerations up to a factor of 40 (24). It is expected that parallelizing matrix

operations and FFTs can yield significant performance boost. On the other hand, an algorithmic reformulation can be another source of performance increase. Solving the inference problem via variational Bayesian analysis (25) was seen to yield an order of magnitude speed-up relative to the greedy Bayesian CS method for non-joint image reconstruction.

A second aspect of this reconstruction method that requires further analysis is the potentially detrimental impact of source data that are not perfectly spatially aligned. To maximize the information sharing among the inversion tasks, it is crucial to register the multi-contrast scans before applying the joint reconstruction. To minimize the adverse consequences of such misalignment, future implementations might deploy either real-time navigators (e.g. (26)) or retrospective spatial registration among datasets based on preliminary CS reconstructions without the joint constraint. For some acquisitions, subtle, non-rigid spatial misregistration may occur due to eddy-current or  $B_0$  inhomogeneity induced distortions. To correct for such higher-order translation effects, several fast and accurate correction methods have been proposed (e.g. (27,28)) and could be applied for correction of undersampled images in joint Bayesian reconstruction. As the preliminary investigation in the Results section demonstrates, joint Bayesian CS algorithm is robust against misregistration effects up to shifts of 2 pixels, and it is believed that existing registration techniques can bring the images within this modest range. Alternatively, future work aimed at the simultaneous joint reconstruction and spatial alignment might pose an interesting and challenging research project in this area, which might be accomplished by introducing additional hidden variables.

Regarding real-valued image-domain datasets, the presented CS reconstructions obtained with Lustig *et al.*'s conjugate gradient descent method yielded 2 to 4 times of the RMSE returned by the joint Bayesian algorithm. Even though this error metric cannot be considered the sole criterion for "good" image reconstruction (29), making use of similarities between multi-contrast scans can be a first step in this direction. In the more general case where the methods were tested with complex-valued images, the improvement in RMSE reduced to about 1.5 times on the *in vivo* data with the joint Bayesian algorithm. When the individual images were reconstructed separately, but using their real & imaginary parts jointly, this non-joint version of the Bayesian algorithm outperformed both Lustig *et al.*'s method and M-FOCUSS on the complex-valued numerical data and the TSE scans. This might suggest that exploiting the similarity between real and imaginary channels of the images can also be source of performance increase. It is important to note that the current Bayesian algorithm requires the sampling patterns to be symmetric in order to handle complex-valued images, and this constraint might be reducing the incoherence of

the aliasing artifacts. As reported in the Result section, using symmetric patterns instead of unconstrained ones increased the error incurred by Lustig *et al.*'s algorithm from 10.5 % to 11.5 %, which seems to be a mild effect. Even though the proposed joint reconstruction algorithm increases the collective coverage of  $k$ -space by sampling non-overlapping data points across the multi-contrast images, this benefit might be dampened by the symmetry constraint.

For comparison, the M-FOCUSS joint reconstruction algorithm was implemented and it was noted that it also attained smaller RMSE figures compared to the gradient descent technique. Even though M-FOCUSS is seen to outperform other competing matching pursuit based joint algorithms (13), the Bayesian method proved to exploit the signal similarities more effectively in the presented experiments. This is made possible by the fact that the Bayesian framework is flexible enough to allow idiosyncratic signal parts, and strict enough to provide information sharing. Importantly, the Bayesian approach also permits the use of different observation matrices for each signal. This allows increased total  $k$ -space coverage across the multi-contrast scans, and its benefit can be seen from the two experiments conducted on the TSE scans with acceleration  $R = 2.5$  along the phase encoding direction. The Bayesian reconstruction results displayed in Fig. 2.4 are obtained by using a different undersampling pattern for  $k$ -space corresponding to each image, and this yielded 2.6 times less RMSE compared to Lustig *et al.*'s algorithm, demonstrating the benefits of variations in the sampling pattern for different contrast weightings. On the other hand, the experiment in Table 2.1 that uses the same pattern for both images returned 2 times smaller RMSE compared to the gradient descent method. However, M-FOCUSS has the advantage of being a much faster algorithm with only modest memory requirements. Interestingly, the performance of the M-FOCUSS algorithm deteriorated significantly when tested on the complex-valued signals, yielding poorer results relative to Lustig *et al.*'s method for the complex-valued TSE dataset. Even though the joint Bayesian algorithm also suffered a performance decrease, it still yielded significantly lower errors with the complex-valued signals.

A direction for future work is the application of the covariance estimates for the posterior distribution produced by the Bayesian algorithm, which could be used to design optimal undersampling patterns in  $k$ -space so as to reduce the uncertainty in the estimated signal (16,30). Also, it is possible to obtain SNR priors, which might be utilized in the Gamma prior  $p(\alpha_0 | c, d) = Ga(\alpha_0 | c, d)$  defined over the noise precision  $\alpha_0$  in the Bayesian algorithm. The setting  $c = d = 0$  was used to incur a non-informative noise prior which would not bias the reconstructions towards a particular noise power. In our informal experiments, smaller RMSE

scores were also obtained with this setting. Yet the optimal selection of  $c$  and  $d$  needs further investigation.

Results in this work do not cover parallel imaging considerations, yet combining compressive measurements with multichannel acquisitions has received considerable attention, e.g. (8,31). Even though exposing the Bayesian formalism to parallel imaging is beyond the current scope, treating the receiver channels as a similarity axis in addition to the contrast dimension might be a natural and useful extension of the work presented here.

In addition to the demonstration of the joint CS reconstruction of multiple different image contrasts, other applications lend themselves to the same formalism for joint Bayesian image reconstruction. These include, for instance,

- ❖ *Quantitative Susceptibility Mapping (QSM)*: In this setting, the aim is to solve an inverse problem of estimating a susceptibility map  $\chi$  related to the phase of a complex image  $|\mathbf{M}|e^{j\varphi}$  via an ill-posed inverse kernel. Since the magnitude part  $|\mathbf{M}|$  is expected to share common image boundaries with  $\chi$ , it might be possible to use it as a prior to guide the inversion task.
- ❖ *Magnetic Resonance Spectroscopic Imaging (MRSI)*: Combining spectroscopic data with high resolution structural scans might help reducing the lipid contamination due to the subcutaneous fat or enhance resolution of brain metabolite maps.
- ❖ *Multi-modal imaging techniques*: Simultaneous acquisitions with different modalities (e.g. PET-MRI) may benefit from joint reconstruction with this Bayesian formulation.

## 2.6 Joint Reconstruction with Prior Estimate

As acquisition times may vary among different contrasts in the multi-contrast protocol, the overall scan time can be minimized for a fixed amount of undersampling by modulating the degree of undersampling among the different contrast preparations. Here, the joint Bayesian framework is extended to asymmetric undersampling schemes where one contrast image is fully sampled while other contrasts are undersampled. By reformulating the inference problem, a new reconstruction method that is based on the Expectation-Maximization (EM) algorithm is also introduced. The EM approach permits the use of a prior image to facilitate the reconstruction, and is detailed in the following.

### 2.6.1 EM Algorithm for Joint Reconstruction with Prior Estimate

Given  $L$  undersampled images  $\{\mathbf{x}_i\}_{i=1,L} \in \mathbb{C}^N$  acquired with different contrasts and a fully-sampled image  $\mathbf{x}_{prior}$ , a sparse representation is again obtained for the undersampled contrasts by taking the spatial gradients in  $k$ -space:

$$\mathbf{F}_\Omega \boldsymbol{\delta}_i = (1 - e^{-2\pi j w/n}) \mathbf{y}_i \equiv \mathbf{z}_i \quad (2.26)$$

To simplify the expressions, the distinction between the vertical and horizontal gradients is now omitted and the corresponding superscripts are dropped in this section. The gradient of the prior image is directly computed as  $\boldsymbol{\delta}_{prior} = \mathbf{F}^{-1} \{(1 - e^{-2\pi j w/n}) \mathbf{y}_{prior}\}$ . The data are modeled to be corrupted by complex Gaussian noise with variance  $\sigma^2$ , yielding the data likelihood

$$p(\mathbf{z}_i | \boldsymbol{\delta}_i, \sigma^2) = \mathcal{N}(\mathbf{F}_\Omega \boldsymbol{\delta}_i, \sigma^2 \mathbf{I}) \quad (2.27)$$

A Gaussian prior across each pixel of the  $L$  images is placed to couple them,

$$p(\boldsymbol{\delta}_{:,t} | \gamma_t) = \mathcal{N}(0, \gamma_t \mathbf{I}) \quad (2.28)$$

where  $\boldsymbol{\delta}_{:,t} \in \mathbb{C}^L$  is the vector formed by taking the  $t^{\text{th}}$  pixel in each image and  $\gamma_t = 1/\alpha_t$  is the inverse of the hyperparameter  $\alpha_t$  controlling the variance. By multiplicative combination of all pixels, full prior distribution is obtained,

$$p(\boldsymbol{\delta} | \boldsymbol{\gamma}) = \prod_{t=1,N} p(\boldsymbol{\delta}_{:,t} | \gamma_t) \quad (2.29)$$

Combining the likelihood and the prior with the Bayes' rule, posterior for the  $i^{\text{th}}$  image becomes

$$p(\boldsymbol{\delta}_i | \mathbf{z}_i, \boldsymbol{\gamma}) = \mathcal{N}(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}) \quad (2.30)$$

$$\text{with } \boldsymbol{\Sigma} = \boldsymbol{\Gamma} - \boldsymbol{\Gamma} \mathbf{F}_\Omega^H \mathbf{B}^{-1} \mathbf{F}_\Omega \boldsymbol{\Gamma} \text{ and,} \quad (2.31)$$

$$\boldsymbol{\mu}_i = \boldsymbol{\Gamma} \mathbf{F}_\Omega^H \mathbf{B}^{-1} \mathbf{z}_i \quad (2.32)$$

where  $\mathbf{B} \equiv \sigma^2 \mathbf{I} + \mathbf{F}_\Omega \boldsymbol{\Gamma} \mathbf{F}_\Omega^H$  and  $\boldsymbol{\Gamma} \equiv \text{diag}(\boldsymbol{\gamma})$ . The posterior distribution is fully characterized if the (inverse) hyperparameters  $\boldsymbol{\gamma}$  are estimated, which can be done with an EM-type algorithm by iteratively applying Eqs. (2.31) and (2.32) followed by the update

$$\gamma_t^{new} = \|\boldsymbol{\mu}_{:,t}\|^2 / (L - L \boldsymbol{\Sigma}_{tt} / \gamma_t) \quad (2.33)$$

By using the prior image to initialize the EM iterations,  $\gamma_t^{initial} = |\delta_{prior,t}|^2$ , the known sparsity support of  $\boldsymbol{\delta}_{prior}$  facilitates the recovery of the undersampled images. After estimating the vertical and horizontal gradients, the images  $\{\mathbf{x}_i\}_{i=1,L}$  that are consistent with these and the  $k$ -space data  $\{\mathbf{y}_i\}_{i=1,L}$  are again found by solving a least squares problem.

## 2.6.2 Methods

Bayesian CS with prior was applied to the TSE and SRI24 datasets, which were also reconstructed with the CS algorithm by Lustig et al. (6) using total variation penalty with an optimal regularization parameter that yielded the smallest RMSE. In the TSE experiment, an early echo slice was retrospectively undersampled with a random 2D pattern using acceleration  $R = 4$  while the late echo image was kept fully sampled to serve as prior. Regarding the SRI24 dataset, single slices from the T2 and T1 weighted images were undersampled along phase encoding with acceleration  $R = 4$ , while the PD image was kept fully sampled to supply prior information. An approximate solution to the large-scale matrix inversion  $\mathbf{B}^{-1}$  in Eq. (2.31) was computed iteratively by Lanczos algorithm with partial reorthogonalization for the Bayesian CS algorithm.

## 2.6.3 Results

Fig. 2.9 depicts the TSE dataset reconstruction results, for which Lustig et al.'s algorithm yielded 9.3% RMSE, while Bayesian CS with prior information had 5.8% error. Results for the SRI24 dataset are given in Fig. 2.10. Here, Lustig et al.'s method yielded 9.5% NRMSE, and the error was 4.3% for Bayesian CS that jointly reconstructed T2 and T1 images with the help of fully-sampled PD image. Joint Bayesian CS *without* using a prior had 4.9% error (not shown).

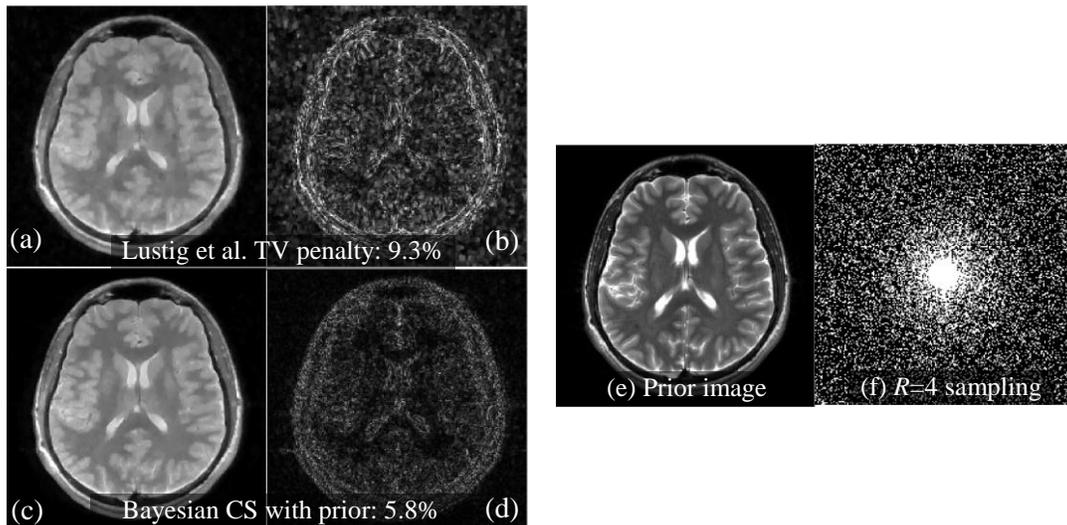


Fig. 2.9. (a) Lustig et al.'s algorithm yielded 9.3% error (b) absolute error for (c) Bayesian CS with prior returned 5.8% error (d) error for Bayesian CS (e) fully-sampled prior (f)  $R=4$  sampling pattern.

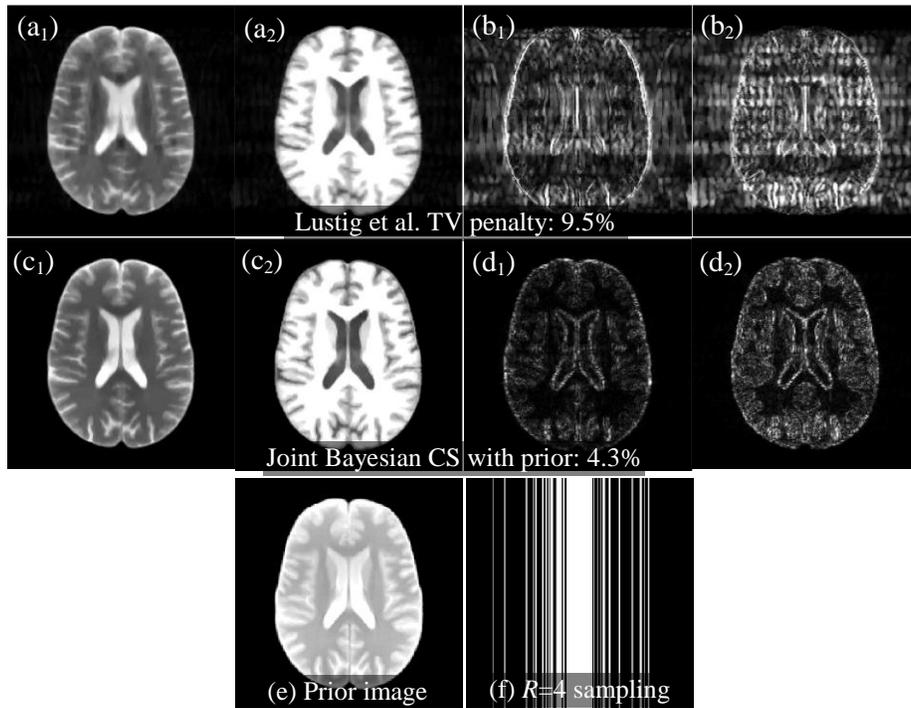


Fig. 2.10. (a<sub>1</sub>-a<sub>2</sub>) Lustig et al.'s algorithm yielded 9.5% error (b<sub>1</sub>-b<sub>2</sub>) absolute error plots for Lustig et al. (c<sub>1</sub>-c<sub>2</sub>). Joint Bayesian CS with prior returned 4.3% error (d<sub>1</sub>-d<sub>2</sub>) error plots for Bayesian CS (e) fully-sampled PD weighted prior image (f)  $R=4$  random undersampling pattern in 1D

#### 2.6.4 Remarks on Reconstruction with Prior Estimate

The presented method makes use of the known sparsity support of a fully-sampled image only to initialize Bayesian CS iterations, and hence avoids imposing this support on the reconstructed images. Acquiring a fully-sampled prior is desirable in cases where one imaging sequence is significantly faster than the other contrast weightings, e.g. an MP-RAGE acquisition along with other contrasts.

### 2.7 Conclusion

This chapter presented the theory and the implementation details of a Bayesian framework for joint reconstruction of multi-contrast MRI scans. By efficient information sharing among these similar signals, the Bayesian algorithm was seen to obtain reconstructions with smaller errors (up to a factor of 4 in RMSE) relative to two popular methods, Lustig *et al.*'s conjugate gradient

descent algorithm (6) and the M-FOCUSS joint reconstruction approach (13). In the presence of a fully-sampled image, it was shown that joint reconstruction can be further enhanced by using this image to supply prior information.

## Chapter 3

### Regularized Quantitative Susceptibility Mapping

Quantifying tissue iron concentration *in vivo* is instrumental for understanding the role of iron in physiology and in neurological diseases associated with abnormal iron distribution. In this chapter, the recently-developed Quantitative Susceptibility Mapping (QSM) methodology is used to estimate the tissue magnetic susceptibility based on MRI signal phase. To investigate the effect of different regularization choices,  $\ell_1$  and  $\ell_2$  norm regularized QSM algorithms are implemented and compared. These regularized approaches solve for the underlying magnetic susceptibility distribution, a sensitive measure of the tissue iron concentration, that gives rise to the observed signal phase. Regularized QSM methodology also involves a pre-processing step that removes, by dipole fitting, unwanted background phase effects due to bulk susceptibility variations between air and tissue and requires data acquisition only at a single field strength. For validation, performances of the two QSM methods were measured against published estimates of regional brain iron from *postmortem* and *in vivo* data. The *in vivo* comparison was based on data previously acquired using Field-Dependent Relaxation Rate Increase (FDRI), an estimate of MRI relaxivity enhancement due to increased main magnetic field strength, requiring data acquired at two different field strengths. The QSM analysis was based on susceptibility-weighted images acquired at 1.5T, whereas FDRI analysis used Multi-Shot Echo-Planar Spin Echo images collected at 1.5T and 3.0T. Both datasets were collected in the same healthy young and elderly adults. The *in vivo* estimates of regional iron concentration comported well with published *postmortem* measurements; both QSM approaches yielded the same rank ordering of iron concentration by brain structure, with the lowest in white matter and the highest in globus pallidus. Further validation was provided by comparison of the *in vivo* measurements,  $\ell_1$ -regularized QSM versus FDRI and  $\ell_2$ -regularized QSM versus FDRI, which again yielded perfect rank ordering of iron by brain structure. The final means of validation was to assess how well each *in vivo* method detected known age-related differences in regional iron concentrations measured in the same young and elderly healthy adults. Both QSM methods and FDRI were consistent in identifying higher iron concentrations in striatal and brain stem ROIs (i.e., caudate nucleus, putamen, globus pallidus, red nucleus, and substantia nigra) in the older than in the young group. The two QSM methods appeared more sensitive in detecting age differences in brain stem structures as they revealed differences of much higher statistical significance between

the young and elderly groups than did FDRI. However, QSM values are influenced by factors such as the myelin content, whereas FDRI is a more specific indicator of iron content. Hence, FDRI demonstrated higher specificity to iron yet yielded noisier data despite longer scan times and lower spatial resolution than QSM. The robustness, practicality, and demonstrated ability of predicting the change in iron deposition in adult aging suggest that regularized QSM algorithms using single-field-strength data are possible alternatives to tissue iron estimation requiring two field strengths.

Further, this chapter develops a closed-form expression for  $\ell_2$ -regularized QSM that can be computed in less than 5 seconds, which is a substantial speed-up compared to iterative methods that may take up to an hour of processing time.

### 3.1 Introduction

Excessive iron deposition in subcortical and brain stem nuclei occurs in a variety of degenerative neurological and psychiatric disorders, including Alzheimer's disease, Huntington's Chorea, multiple sclerosis, and Parkinson's disease (32). Further, *postmortem* (1) and *in vivo* (33-37) studies have revealed that deep gray matter brain structures accumulate iron at different rates throughout adult aging. Structures that exhibit iron accrual support components of cognitive and motor functioning (37-39). To the extent that excessive iron presence may attenuate neuronal function or disrupt connectivity, quantification and location of iron deposition may help explain age- and disease-related motor slowing and other selective cognitive decline.

Several MRI methods have been proposed for *in vivo* iron mapping and quantification. Bartzokis et al. (40) capitalized on the enhanced transverse relaxivity (R2) due to iron with increasing main field strength for the Field-Dependent Relaxation Rate Increase (FDRI) method. FDRI relies on the use of R2-weighted imaging at two different field strengths and attributes the relaxation enhancement at higher field to iron, which may be a specific measure of tissue iron stores (40).

Whereas FDRI relies on the modulation of signal intensity in MRI to infer iron concentration, MRI signal phase has also been proposed as a source signal for iron mapping, both by direct evaluation of phase images (41,42) and by reconstruction of magnetic susceptibility images that derive from the phase data (34,42). Local iron concentration is strongly correlated with the magnetic susceptibility values (43-45); therefore, quantification of this paramagnetic property presents a sensitive estimate of iron concentration, although possibly complicated by more

uncommon factors, such as pathological manganese deposition (46). Phase mapping yields high-resolution, high-SNR data that demonstrate correlation with iron (34), but as an estimate of the underlying magnetic susceptibility, it suffers from non-local effects and spatial modulation artifacts due to the non-trivial mapping from susceptibility to phase (47). To overcome these limitations, herein regularized Quantitative Susceptibility Mapping (QSM) algorithms are employed for robust estimation of the magnetic susceptibility  $\chi$  of tissues based on gradient-echo signal phase. The magnetic susceptibility  $\chi$  maps to the observed phase shift in MRI via a well-understood transformation, but the inverse problem, i.e., estimation of  $\chi$  from phase, is ill posed due to zeros on a conical surface in the Fourier space of the forward transform; hence,  $\chi$  inversion benefits from additional regularization. Recently, elegant regularization methods were proposed for deriving susceptibility inversion. In the work by de Rochefort et al. (2010), smooth regions in the susceptibility map are promoted to match those of the MR magnitude image by introducing a weighted  $\ell_2$  norm penalty on the spatial gradients of  $\chi$ . Likewise, Liu et al. (2010) regularized the inversion by minimizing the  $\ell_1$  norm of gradients of  $\chi$ , again weighted with a mask derived from the image magnitude. Kressler et al. (2010) experimented using  $\ell_1$  and  $\ell_2$  norm regularizations directly on the susceptibility values, rather than posing the minimization on the gradient coefficients. Another method to stabilize the susceptibility reconstruction problem is to acquire data at multiple orientations and invert them simultaneously *without* regularization. This approach was introduced by Liu *et al.* (2009) and also investigated by others such as Wharton and Bowtell (2010) and Schweser *et al.* (2011).

In this work, two different regularization schemes are investigated for susceptibility inversion; using  $\ell_1$ -regularized QSM that parallels the approach of Liu *et al.* (2010) and  $\ell_2$ -regularized QSM which was introduced by de Rochefort *et al.* (2010). Given that magnetic susceptibility is a property of the underlying tissue, in  $\ell_1$ -regularized QSM the underlying assumption is that susceptibility is approximately constant within regions of the same tissue type or within an anatomical structure. Based on this premise, the  $\ell_1$ -norm-penalized QSM algorithm regularizes the inversion by requiring the estimated  $\chi$  to be sparse in the image gradient domain. On the other hand, placing an  $\ell_2$  norm penalty on the spatial gradients of  $\chi$  does not promote sparsity, but results in a large number of small gradient coefficients and thus incurs a smooth susceptibility reconstruction. In addition to regularized susceptibility inversion, the presented approach incorporates a robust background phase removal technique based on effective dipole fitting (48), which addresses the challenging problem of removing phase variations in the data that arise primarily from bulk susceptibility variations between air and tissue rather than the more subtle changes of  $\chi$  within the brain. Dipole fitting contains no parameters that need tuning and

preserves the phase variations caused by internal susceptibility effects more faithfully than high-pass filtering, as employed in susceptibility-weighted imaging (SWI) (41,42). All susceptibility mapping methods require data acquired at only one field strength, thereby overcoming certain limitations of the FDRI approach, including long scan times and the need for spatial registration of image data acquired with different scanners at different field strengths.

Here, the  $\ell_1$  and  $\ell_2$  norm regularized QSM methods are described and applied to SWI data previously acquired in groups of younger and elderly, healthy adults (35). To validate the iron measures, the results of QSM methods were compared with values published from a *postmortem* study (1). As further validation, QSM results were compared with those based on FDRI collected in the same adults (35) to test the hypothesis that the iron deposition in striatal and brain stem nuclei, but not white matter or thalamic tissue, would be greater in older than younger adults. The chapter closes with a fast algorithm that achieves  $\ell_2$ -regularized susceptibility mapping in seconds.

## 3.2 Methods

### 3.2.1 Susceptibility and MR signal phase

The normalized magnetic field shift  $\delta$  measured in a gradient-echo sequence is related to the MR image phase  $\varphi$  via  $\delta = -\varphi/(B_0 \cdot \gamma \cdot TE)$ , where  $B_0$  is the main magnetic field strength,  $\gamma$  is the gyromagnetic ratio, and  $TE$  is the echo time. It follows from Maxwell's magnetostatic equations that the relationship between the underlying susceptibility distribution  $\chi$  and the observed field shift  $\delta$  is given by (47,49,50)

$$\mathbf{F} \delta = \left( \frac{1}{3} - \frac{k_z^2}{k_x^2 + k_y^2 + k_z^2} \right) \circ (\mathbf{F} \chi) \quad (3.1)$$

where  $\mathbf{F}$  is the discrete Fourier transform matrix,  $k_x$  and  $k_y$  are the in-plane frequency indices,  $k_z$  is the frequency index along  $B_0$ , and  $\circ$  denotes element-wise multiplication. Denoting with  $\mathbf{D}$  the kernel that relates the field map to the susceptibility, the relation can also be expressed as

$$\delta = \mathbf{F}^{-1} \mathbf{D} \mathbf{F} \chi \quad (3.2)$$

The spatial frequencies at which the kernel is zero define a conical surface in  $k$ -space, which effectively undersamples the Fourier transform of  $\chi$  and thereby gives rise to the ill-posed problem of susceptibility estimation from image phase. In addition, the susceptibility kernel is not defined at the center of  $k$ -space (the DC point), but one can choose a solution that vanishes at

infinity, which is obtained by setting the Fourier transform of the field to 0 at  $k = 0$  (47). This assignment of signal for the  $k$ -space origin causes the resulting  $\chi$  to have zero mean; but independent of the particular design choice for this DC signal, the susceptibility distribution is inherently a spatial map of *relative* susceptibilities. Under the assumption that the field map and the susceptibility distribution are differentiable along  $k_z$ , Li *et al.* (2011) derived that the convolution kernel equals  $-2/3$  at  $k = 0$ . In this work, the convention of assigning 0 to the DC value of the kernel is adopted. Thus, to achieve absolute quantification of  $\chi$ , some reference value needs to be established. For this study, the magnetic susceptibility value in splenium is chosen as a reference. This structure was preferred over taking as a reference the CSF, for which the susceptibility values were observed to differ substantially between the anterior and the posterior ventricles in this study.

### 3.2.2 Background effect removal from the field map

In addition to the relatively subtle internal effects of the tissue iron on the MRI phase, background artifacts caused by air-tissue boundaries contribute the vast majority of signal variation in the observed phase. While the susceptibility difference between air and water is about 9.4 ppm (parts per million) (51), the largest within-brain variation due to tissue iron is more than an order of magnitude smaller. Assuming that the average human tissue susceptibility is similar to that of water, it is clear that background effects dominate the observed phase and this undesired signal component is a challenge to robust susceptibility inversion. Because the background effects usually vary slowly across space, various methods have been proposed to filter them out based on this frequency characteristic, such as polynomial fitting (44) and forward modeling to estimate the phase from the air/tissue interface (52). Even though these methods are effective for background phase removal, their impact on the internal phase variations due to tissue iron is unclear. A recent background field removal algorithm, *effective dipole fitting* (48), aims to estimate the background susceptibility distribution that optimally matches the field inside the region of interest (ROI), and removes this contribution to recover the foreground field map. This is achieved by solving a least-squares problem

$$\chi_{out} = \operatorname{argmin}_{\chi} \left\| \mathbf{M} \left( \delta - \mathbf{F}^{-1} \mathbf{D} \mathbf{F} \tilde{\mathbf{M}} \chi \right) \right\|_2^2 \quad (3.3)$$

where  $\mathbf{M}$  is the brain mask that marks the ROI and  $\tilde{\mathbf{M}}$  is the complement of  $\mathbf{M}$ , thus marking the background. After solving for  $\chi_{out}$ , the field map induced only by the internal local effects is obtained by

$$\delta_{in} = \delta - \mathbf{F}^{-1} \mathbf{D} \mathbf{F} \tilde{\mathbf{M}} \chi_{out} \quad (3.4)$$

Compared with high-pass filtering, effective dipole fitting was seen to yield 1/3 to 1/7 times the root-mean-square error relative to the true field maps obtained from reference scans (48). Another elegant background removal technique called SHARP (45), with results comparable to those of the dipole fitting method (53), involves removing the harmonic contributions to the phase inside the region of interest by filtering.

### 3.2.3 Susceptibility inversion with $\ell_1$ regularization

The final step in the proposed algorithm is to estimate the susceptibility distribution that gives rise to  $\delta_{in}$ . Hence, the aim is to solve

$$\delta_{in} = \mathbf{F}^{-1} \mathbf{DF} \chi_{in} \quad (3.5)$$

Because some of the spatial frequencies are undersampled by the kernel  $\mathbf{D}$ , the inversion of  $\chi_{in}$  benefits from regularization that imposes prior knowledge on the reconstructed susceptibility map. The susceptibility values are tied to the paramagnetic properties of the underlying tissue structure; hence they vary smoothly across space within anatomical boundaries and can be approximated to be piece-wise constant. In this case, the susceptibility map is expected to be sparsely represented in the spatial image gradient domain. To formulate this belief, the aim is to find the  $\chi$  distribution that matches the field map  $\delta_{in}$ , and that also has sparse image gradients

$$\chi_{in} = \operatorname{argmin}_{\chi} \left\| \delta_{in} - \mathbf{F}^{-1} \mathbf{DF} \chi \right\|_2^2 + \lambda \cdot \|\mathbf{G} \chi\|_1 \quad \text{with} \quad \mathbf{G} = \begin{bmatrix} \mathbf{G}_x \\ \mathbf{G}_y \\ \mathbf{G}_z \end{bmatrix} \quad (3.6)$$

where  $\|\mathbf{G} \chi\|_1$  is the  $\ell_1$  norm of image gradients in all three dimensions, and  $\lambda$  is a regularization parameter that trades off data consistency and spatial smoothness. This convex program is very similar to the objective function in the compressed sensing (CS) MRI literature, where the aim is to reconstruct MR images from undersampled  $k$ -space data. According to CS theory, if the underlying image can be approximated to be sparse in a transform domain, then it can be recovered from randomly undersampled  $k$ -space data via a nonlinear recovery scheme, and the reconstruction quality depends on the number of observed frequency samples as well as the coherence of the aliasing artifacts in the transform domain (54). The nonlinear recovery method usually involves penalizing the  $\ell_1$  norm of the transformed image. Based on this, Eq. (3.6) can be viewed as CS reconstruction with a modified observation matrix  $\mathbf{DF}$  instead of the undersampled Fourier transform.

An objective function similar to Eq. (3.6) has been previously proposed in Liu et al. (2010), which included a smoothing term of the form  $\|\mathbf{W}_G \mathbf{G} \boldsymbol{\chi}\|_L$ . Here,  $\mathbf{W}_G$  is a weighting matrix derived from the MRI image magnitude, and  $L$  denotes the choice of the norm, which can be either  $\ell_1$  or a homotopic approximation to the  $\ell_0$  norm. Apart from the magnitude weighting, the presented method parallels this approach.

### 3.2.4 Susceptibility inversion with $\ell_2$ regularization

Another way of introducing regularization to the inversion problem is by penalizing the  $\ell_2$  norm of spatial gradients of the susceptibility distribution,

$$\boldsymbol{\chi}_{in} = \operatorname{argmin}_{\boldsymbol{\chi}} \|\boldsymbol{\delta}_{in} - \mathbf{F}^{-1} \mathbf{D} \mathbf{F} \boldsymbol{\chi}\|_2^2 + \beta \cdot \|\mathbf{G} \boldsymbol{\chi}\|_2^2 \quad (3.7)$$

In contrast with the  $\ell_1$  regularization that promotes sparse spatial gradients (i.e. a small number of non-zero gradient coefficients),  $\ell_2$ -regularized inversion favors a large number of small gradient coefficients. Regularized QSM with  $\ell_2$  norm penalty was introduced in de Rochefort et al. (2010), which also included a weighting matrix  $\mathbf{W}_1$  derived from the signal magnitude in the regularization term to yield  $\|\mathbf{W}_1 \mathbf{G} \boldsymbol{\chi}\|_2^2$ . To investigate the effect of the regularization norm selection in susceptibility inversion, QSM results with both regularization styles are presented.

### 3.2.5 Effect of regularization parameters $\lambda$ and $\beta$

The regularization parameter  $\lambda$  in Eq. (3.6) determines the smoothness of the reconstructed susceptibility map such that larger values of  $\lambda$  yield smoother image results than do smaller ones (Fig. 3.1). This flexibility permits controlling the scale of spatial features present in the  $\boldsymbol{\chi}$  reconstruction.

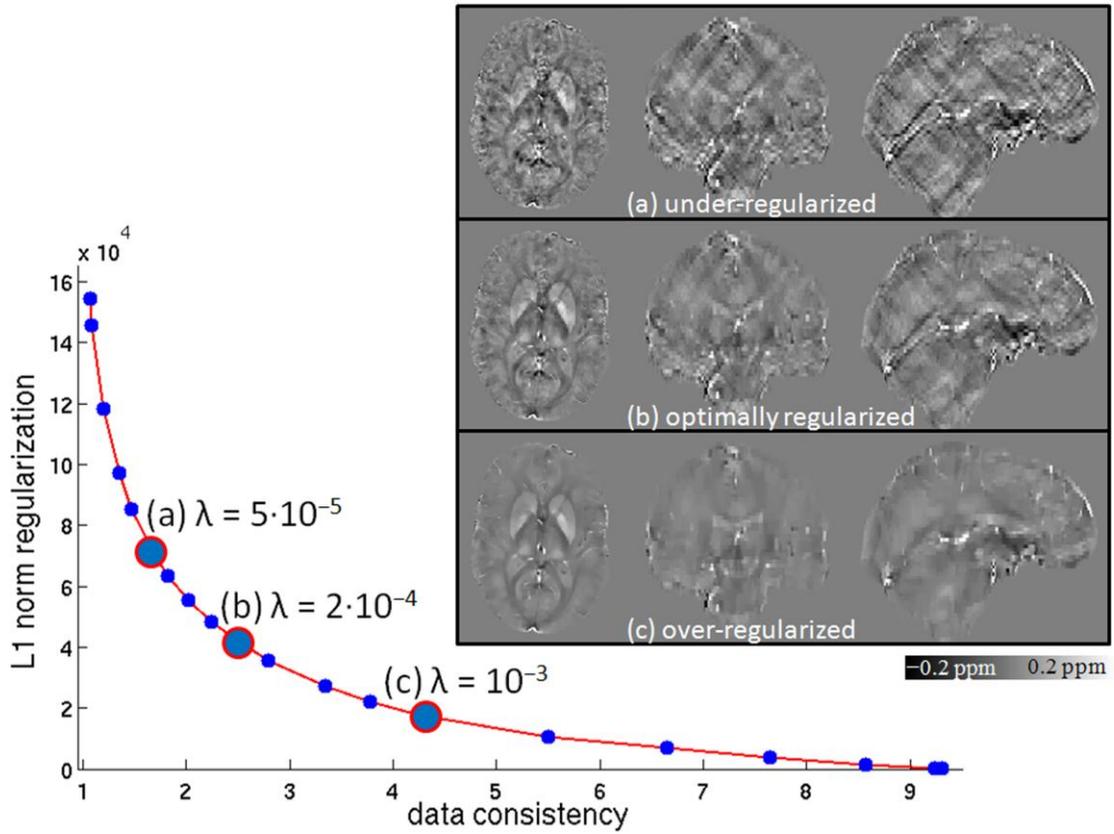


Fig. 3.1. L-curve for  $\ell_1$ -regularized QSM results for a young subject. X-axis: data consistency term  $\|\delta - \mathbf{F}^{-1} \mathbf{D} \mathbf{F} \chi\|_2$  in regularized reconstruction for varying values of the smoothing parameter  $\lambda$ . Y-axis: regularization term  $\|\mathbf{G} \chi\|_1$ . Setting  $\lambda = 5 \cdot 10^{-5}$  yielded an under-regularized susceptibility map with ringing artifacts (a), whereas using  $\lambda = 10^{-3}$  resulted an over-regularized reconstruction (c). For  $\lambda = 2 \cdot 10^{-4}$ , the operating point with the largest curvature on the L-curve was obtained (b). This setting was used for the reported  $\ell_1$ -regularized results.

In terms of imposing prior belief on the susceptibility distribution, it is possible to recover Eq. (3.6) by assuming that the normalized field map  $\delta_{in}$  is corrupted by white Gaussian noise with some variance  $\sigma^2$  and by placing a sparsity-promoting Laplacian prior distribution on the gradient coefficients of the  $\chi$  map,

$$p(\partial\chi) = \left(\frac{\lambda}{4\sigma^2}\right)^M \exp\left(-\frac{\lambda}{2\sigma^2} \sum_{i=1}^M |\partial\chi_i|\right) \quad (3.8)$$

where  $\partial\chi$  represents the spatial gradient of  $\chi$ , and  $M$  is the total number of voxels in  $\chi$ . With these noise and prior models, invoking the *maximum a posteriori* (MAP) estimate reduces to Eq. (3.6). From this point of view, using a large  $\lambda$  will produce a highly peaked prior distribution at zero, inducing sparser image gradient solutions, and smoother susceptibility maps.

Again from a Bayesian perspective, the  $\ell_2$  norm regularization corresponds to computing the MAP estimate after placing a multivariate Gaussian prior on the gradient coefficients of the susceptibility map,

$$p(\hat{\chi}) = \frac{1}{(2\pi\sigma^2 / \beta)^{M/2}} \exp\left(-\frac{1}{2\sigma^2 / \beta} \sum_{i=1}^M |\hat{\chi}_i|^2\right) \quad (3.9)$$

where  $\sigma^2$  is the data noise in the field map and  $\beta$  is the regularization parameter in Eq. (3.7). Hence, the variance of the gradient coefficients ( $\sigma^2 / \beta$ ) is inversely proportional to the  $\ell_2$  regularization parameter  $\beta$ . Accordingly, a large regularization parameter will limit the variation in the gradient coefficients and induce smaller values (Fig. 3.2).

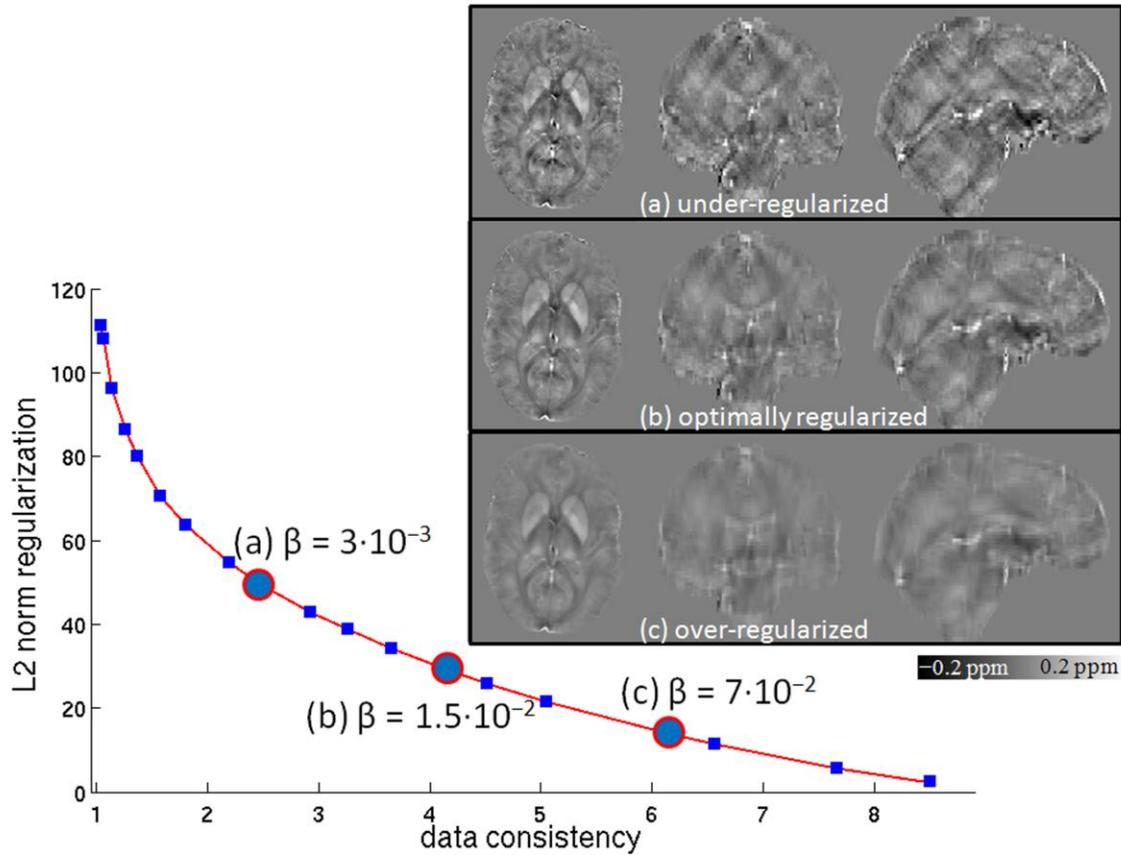


Fig. 3.2. L-curve for  $\ell_2$ -regularized QSM results for a young subject. X-axis: data consistency term  $\|\delta - \mathbf{F}^{-1} \mathbf{D} \mathbf{F} \chi\|_2$  in regularized reconstruction for varying values of the smoothing parameter  $\beta$ . Y-axis: regularization term  $\|\mathbf{G} \chi\|_2$ . Setting  $\beta = 3 \cdot 10^{-3}$  yielded an under-regularized susceptibility map with ringing artifacts (a), whereas using  $\beta = 7 \cdot 10^{-2}$  resulted an over-regularized reconstruction (c). For  $\beta = 1.5 \cdot 10^{-2}$ , the operating point with the largest curvature on the L-curve was obtained (b). This setting was used for the reported  $\ell_2$ -regularized results.

### 3.2.6 Selection of regularization parameters $\lambda$ and $\beta$

To choose appropriate regularization parameters that balance data consistency and the amount of regularization, the L-curve method was employed (55). The corners of the L-curves were not sharp for  $\ell_1$ - and  $\ell_2$ -regularized reconstructions (Figs. 3.1&2), and optimal regularization parameters were determined by finding the operating points with the largest curvature. L-curve tests were performed on a young and an elderly subject from the *in vivo* dataset and the optimal operating points were found to be  $\lambda = 2 \cdot 10^{-4}$  for  $\ell_1$ -regularized QSM and  $\beta = 1.5 \cdot 10^{-2}$  for  $\ell_2$ -regularized reconstructions on both the young and the elderly subjects.

### 3.2.7 Dataset acquired in younger and elderly adults used for comparison of regularized QSM and FDRI

To examine consistency with our previous study that investigated the performance of FDRI (35), the proposed iron quantification algorithm was tested on the same dataset, as summarized below.

#### *Subjects*

Two groups of healthy, highly educated, right-handed adults were studied: 11 younger adults (mean $\pm$ S.D. age = 24.0  $\pm$  2.5, range = 21 to 29 years, 15.9 years of education; 5 men, 6 women) and 12 elderly adults (mean $\pm$ S.D. age = 74.4  $\pm$  7.6, range = 64 to 86 years, 16.3 years of education; 6 men, 6 women). The younger subjects included laboratory members and volunteers recruited from the local community. All older participants were recruited from a larger ongoing study of normal aging and scored well within the normal range on the Dementia Rating Scale (56): mean = 140.6, range = 132 to 144 out of 144, cutoff for dementia = 124. Mean (and range) of days between 1.5T and 3.0T scan acquisition were 16.5 (0 to 56) days for the young and 9.3 (0 to 42) days for the elderly group; for 2 of the young and 8 of the elderly both sets of scans were acquired on the same day.

#### *Image acquisition protocols*

MRI data were acquired prospectively on 1.5T and 3.0T General Electric (Milwaukee, WI) Signa human MRI scanners (gradient strength = 40 mT/m; slew rate = 150 T/m/s).

#### *FDRI acquisition*

At 1.5T, after auto shimming for the session, the following sequences were acquired for 62 axial slices, each 2.5 mm thick:

1) 3D SPoiled Gradient Recalled Echo (SPGR) for structural imaging and registration (TR/TE=8.1/3.3 ms, FA=30°);

2) multi-shot Echo Planar Spin Echo (EPSE) (TR/TE 6000/17, FA=90°, 256×192 in-plane, FOV=24 cm, 4 NEX, 24 interleaves with 8 phase-encode lines per TR, 9:40 min);

3) multi-shot EPSE (TR/TE 6000/60, FA=90°, 256×192 in-plane, FOV=24 cm, 6 NEX, 24 interleaves, 14:20 min).

At 3.0T, after auto shimming for the session, the following sequences were acquired in the axial plane:

1) 3D SPGR for structural imaging and registration (TR/TE=8.1/3.3 ms, FA=15°, 124 slices, 1.25 mm thick);

2) multi-shot EPSE (TR/TE 6000/17, FA=90°, 256×192 in-plane, FOV=24 cm, 3 NEX, 24 interleaves, 62 slices, 2.5 mm thick, 7:10 min);

3) multi-shot EPSE (TR/TE 6000/60 ms, FA=90°, 256×192 in-plane, FOV=24 cm, 6 NEX, 24 interleaves, 62 slices, 2.5 mm thick, 14:20 min).

### ***Susceptibility-Weighted Image acquisition***

At 1.5T, after auto shimming for the session, the following sequences were acquired for 62 axial slices, each 2.5 mm thick:

1) 3D SPGR for structural imaging and registration (TR/TE=28/10 ms, FA=30°, 256×256 in-plane, 24 cm FOV);

2) susceptibility-weighted 3D SPGR (TR/TE=58 ms/40 ms, FA=15°, 512×256 in-plane, 24 cm FOV, 12:20 min, with flow compensation) (34,57);

3) 2D gradient-recalled echo sequence (TR/TE=600/3 ms, FA=20°);

4) 2D gradient-recalled echo sequence (TR/TE=600/7 ms, FA=20°).

Phase images were constructed from the real and imaginary components of the SWI-SPGR data after the phase had been unwrapped with FSL PRELUDE (Phase Region Expanding Labeler for Unwrapping Discrete Estimates (58)). The magnitude and phase-unwrapped SWI data were down-sampled from 512×256 to 256×256 via averaging to match the FDRI resolution. Brain masks were generated with the FSL Brain Extraction Tool, BET (59), to be used in the dipole fitting step for background phase removal. After estimating the foreground field maps from the

unwrapped phase data with the down-sampled size  $256 \times 256$ , susceptibility maps were generated with the two QSM algorithms.

### ***Image registration***

As previously described (35), for each subject and for 1.5T and 3.0T separately, the late-echo EPSE data were nonrigidly registered (60) [<http://nitrc.org/projects/cmtk/>] to the early-echo EPSE data. This was necessary because the two echoes arose from separate acquisitions, rather than a single dual-echo acquisition, and were, therefore, not always perfectly aligned with each other. The 1.5T early-echo EPSE image of each subject was registered to the 3.0T early-echo EPSE image of the same subject, which was then registered nonrigidly to the subject's 3.0T SPGR image. The 3.0T SPGR image from each subject, after brain extraction using BET, finally was registered nonrigidly to the SPGR channel of the SRI24 atlas (22) [<http://nitrc.org/projects/sri24/>]. Via concatenation of the aforementioned registration transformations, the 1.5T and 3.0T early-echo and late-echo images were all reformatted into 1-mm isotropic SRI24 space, each using a single interpolation with a 5-pixel-radius cosine-windowed sinc kernel. Reformatting both 1.5T and 3.0T data from each subject into SRI24 coordinates via that subject's 3.0T SPGR image (rather than separately via the early-echo EPSE images at each field strength) ensures that the unavoidable inter-subject registration imperfections are consistent for images from both field strengths. The 1.5T SWI magnitude images were rigidly registered to a contemporaneously acquired structural SPGR image, which was then registered nonrigidly to the same subject's 3.0T SPGR image. The SWI-SPGR registration was limited to a rigid transformation because signal dropouts in magnitude SWI due to B0 field inhomogeneities prevented nonrigid correction of the relatively small distortions between SWI and SPGR. Again, via concatenation of transformations, the phase images were reformatted into SRI24 space, again with a 5-pixel radius cosine sinc kernel. All data were analyzed in common 1-mm isotropic SRI24 atlas space.

### ***Region-of-Interest (ROI) identification***

Voxel-by-voxel FDRI images ( $FDRI = (R2_{3T} - R2_{1.5T}) / 1.5T$ ) were created for each subject and used to make a group FDRI average, comprising all young and elderly subjects. A similar group average was made for the QSM images, and separate young and elderly group averages were made for display purposes (Fig. 3.3).

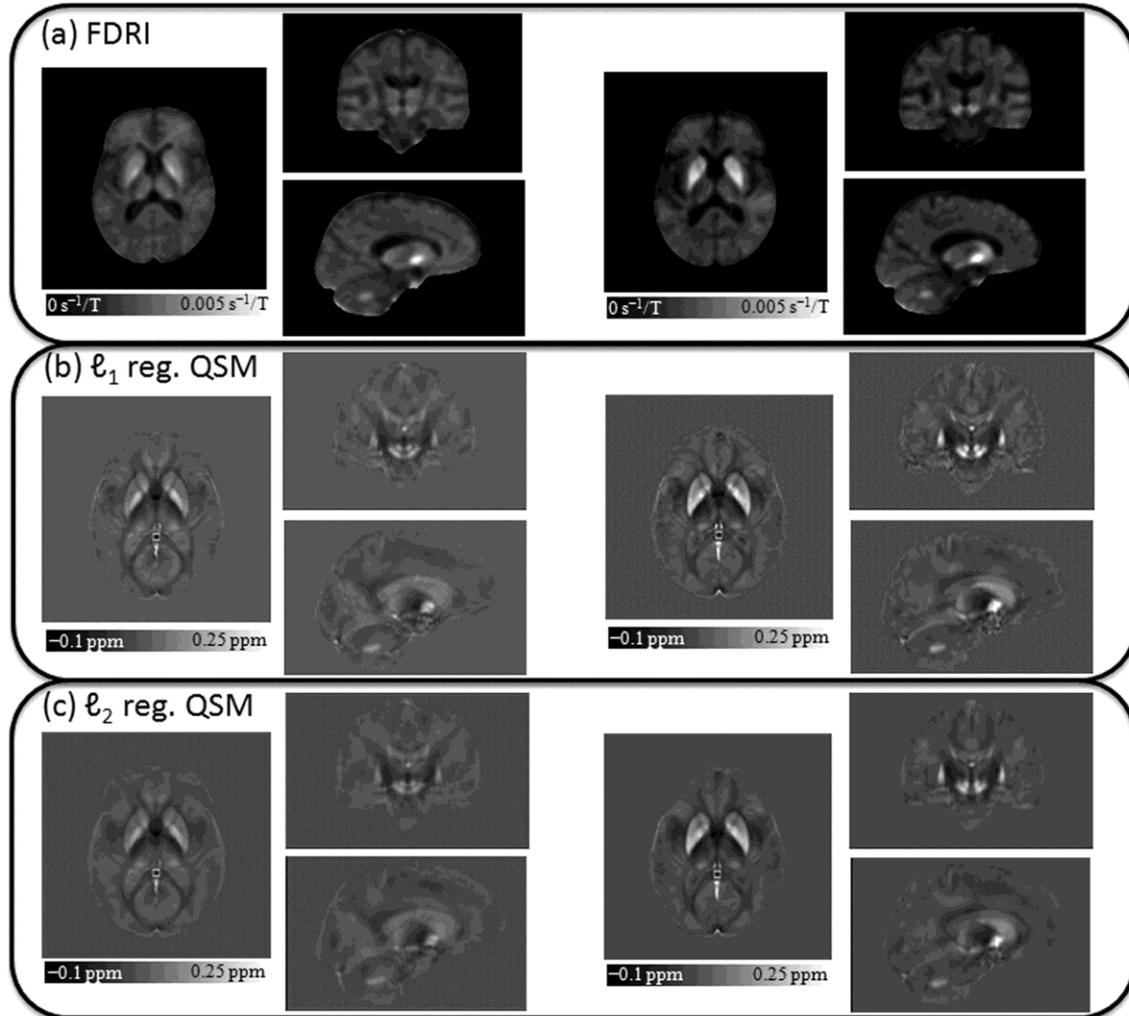


Fig. 3.3. Young (left) and elderly (right) group averages for FDR (a),  $\ell_1$ -regularized QSM (b), and  $\ell_2$ -regularized QSM (c). Greater iron concentration yields brighter QSM and FDR images. Splenium reference ROIs are indicated with a white box on the axial QSM slices.

As previously described (35), bilateral caudate, globus pallidus, putamen, thalamus, and white matter sample regions of interest (ROIs) were drawn on the group-average (all young plus all elderly subjects) FDR images in common SRI24 space, reformatted in the coronal plane. The globus pallidus, putamen, caudate, and white matter sample were drawn on 10 contiguous, 1-mm thick slices at an anterior–posterior location that maximized the presence of all three basal ganglia structures in the same slices. The thalamus was drawn on the next 10 contiguous slices posterior to the basal ganglia. The caudate was eroded one pixel and thalamus was eroded two pixels on a slice-by-slice basis to avoid partial voluming of CSF. Substantia nigra and red nucleus ROIs were also identified, based on their FDR intensities. The same ROIs were also manually identified on

the group-average phase data (all young and all elderly combined), reformatted in the *axial* plane (61), and guided by phase conspicuity. When drawing ROIs on the phase data, an effort was made to exclude the bright rims around the globus pallidus and putamen as well as the division between them. Although this approach biases the data towards more negative phase (i.e., lower values reflecting less iron), its purpose was to maximize the sensitivity of phase to age effects. Thus, iron estimates were conducted on both sets of ROI identifications, the phase-guided and the FDRI-guided.

For each subject and for each ROI at each field strength, the mean intensity of all voxels in an ROI for the early- and late-echo EPSE were used to compute  $R_{2_{3T}}$  and  $R_{2_{1.5T}}$  and the FDRI. QSM values were computed as the magnetic susceptibility in parts per million (ppm) for all voxels identified in each ROI projected onto each individual's QSM dataset. Thus, both FDRI intensity and phase conspicuity were each used to guide ROI delineation. The average susceptibility of splenium in each subject was used as a reference for that subject's reported QSM results. This was preferred over taking the CSF susceptibility as a reference, as it was seen to differ substantially between the anterior and the posterior regions. Although the raw averages in the splenium did not differ significantly between the young and the elderly groups ( $p=0.2359$  for  $\ell_1$ -regularized and  $p=0.2016$  for  $\ell_2$ -regularized QSM), they were larger in the elderly group than the young group ( $\chi_{splenium}^{elderly} = -0.0378$  ppm and  $\chi_{splenium}^{young} = -0.0479$  ppm for  $\ell_1$ -regularized and  $\chi_{splenium}^{elderly} = -0.0297$  ppm and  $\chi_{splenium}^{young} = -0.0374$  ppm for  $\ell_2$ -regularized QSM). This should induce a bias *against* observing young-elderly group susceptibility differences in the regularized QSM reconstructions.

### ***Statistical analysis***

It was predicted in this study that the ROI iron values would correlate positively with published *postmortem* iron values (1) and with FDRI values. Comparisons of the two *in vivo* iron indices with each other and also with published *postmortem* values were based on nonparametric (Spearman) correlations. The hypotheses that, relative to the young group, the elderly group would have higher QSM and FDRI values in striatal and brain stem ROIs, but not in thalamic or white matter ROIs was tested. Because a directional hypotheses was posed, group differences were considered significant at  $p \leq 0.0125$ , the one-tailed, family-wise Bonferroni-corrected p-value at  $\alpha=0.05$  for 8 measures. All measurements were conducted twice: once with FDRI-guided ROI identification, and once with phase-guided ROI identification.

### 3.3 Results

#### 3.3.1 Correlations of FDRI and QSM values with postmortem iron concentrations

Fig. 3.4 presents the mean  $\pm$  SD iron concentration determined *postmortem* in each ROI (1) on the *x*-axis and the mean  $\pm$  SD FDRI values in  $s^{-1}$ /Tesla and  $\ell_1$ -regularized QSM values in ppm for young plus elderly subjects on the *y*-axis. The correlations between  $\ell_1$ -regularized QSM and *postmortem* ( $Rho = 0.881$ ,  $p = 0.0198$ ), between  $\ell_2$ -regularized QSM and *postmortem* ( $Rho = 0.881$ ,  $p = 0.0198$ ), and between FDRI and *postmortem* iron indices ( $Rho = 0.952$ ,  $p = 0.0117$ ) were high.

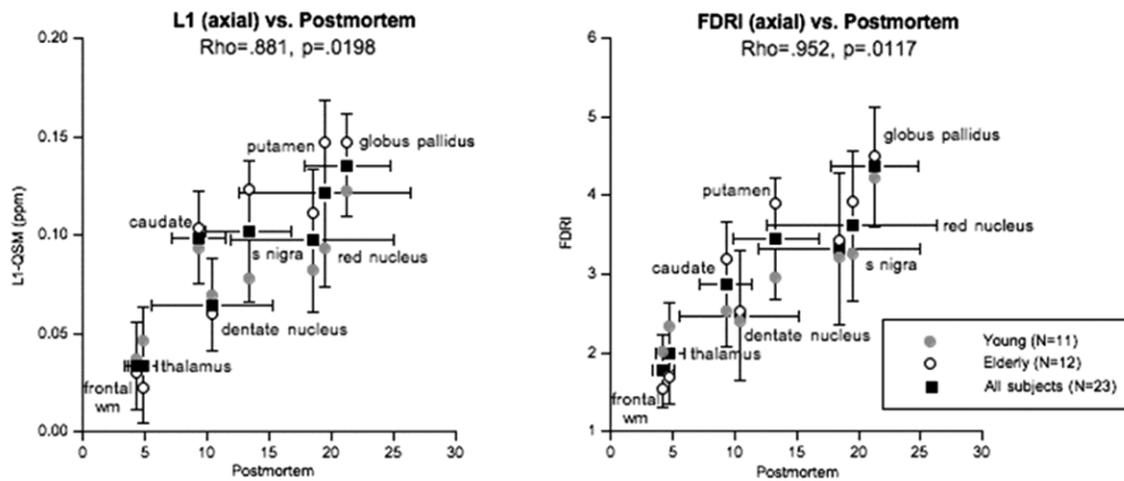


Fig. 3.4. *X*-axis: Mean  $\pm$  SD iron concentration (mg/100 g fresh weight) determined *postmortem* in each ROI (1). *Y*-axis: Mean  $\pm$  SD  $\ell_1$ -regularized QSM in ppm (left) and FDRI in  $s^{-1}$ /Tesla (right) indices in all 23 subjects (black squares); the gray circles indicate the mean of the young group, and the open circles indicate the mean of the elderly group.

#### 3.3.2 Correlations between in vivo QSM and FDRI iron concentration metrics

To investigate the consistency between the iron concentrations predicted by the two QSM methods and FDRI, the three metrics in each ROI belonging to the 23 subjects were correlated. The correlation parameters indicate strong agreement between  $\ell_1$ -regularized QSM and FDRI ( $Rho = 0.976$ ,  $p = 0.0098$ ) (Fig. 3.5) and between  $\ell_2$ -regularized QSM and FDRI ( $Rho = 0.976$ ,  $p = 0.0098$ ) (not shown).

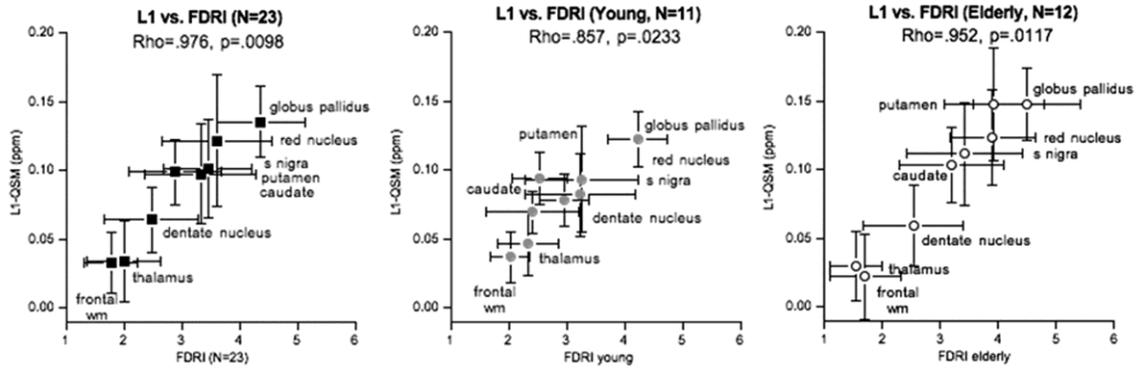


Fig. 3.5. Correlation between FDR1 and  $\ell_1$ -regularized QSM results on the regions of interest. Results indicate strong relationship between the two methods ( $Rho = 0.976$ ,  $p = 0.0098$ ). Left: all 23 subjects; middle: young group; right: elderly group.

### 3.3.3 Age differences in regional iron concentration: QSM and FDR1

All ROI and statistical analyses were conducted on both phase-guided and FDR1-guided ROIs. Based on the initial FDR1 data analysis, which reported lack of consistent cerebral hemisphere asymmetries across iron-rich structures (35), all analyses herein used bilateral data, expressed as the mean of the left and right measures for each ROI (Table 1). The three methods produced essentially the same results. All  $t$ -test and  $p$ -values are presented in Table 1.

### 3.3.4 Age differences identified with regularized QSM

Analysis of the QSM results indicated that the elderly group had significantly more iron than the young group in striatal regions of the putamen and globus pallidus for both  $\ell_1$ - and  $\ell_2$ -norm regularized results. Even though the elderly tended to have more iron in the caudate nucleus than the young, the difference was not significant in either of the QSM methods. Likewise,  $\ell_1$ - and  $\ell_2$ -regularized QSM values indicated significantly more iron in the elderly than young group in the red nucleus and substantia nigra, but not the dentate nucleus. The only exception was the  $\ell_1$ -regularized substantia nigra results on the phase-guided ROIs, for which the group difference was not significant using family-wise Bonferroni correction

Average susceptibility values in the thalamus tended to be lower in the elderly relative to the young (indicating less iron in the elderly group) for both types of regularization, and this difference was significant for  $\ell_2$  norm regularized QSM under phase-guided ROIs. Likewise, the elderly had smaller susceptibility values in the white matter sample, but the difference was not significant (Fig. 3.6).

**Table 1a. Mean ( $\pm$ SD) of each measure by region for each group:  $\ell_1$ -regularized QSM results using phase-guided ROIs and FDRI-guided ROIs**

Region	$\ell_1$ -regularized QSM (ppm), phase-guided ROIs			$\ell_1$ -regularized QSM (ppm), FDRI-guided ROIs		
	Young (N=11)	Elderly (N=12)	$t$ (elderly>young)	Young (N=11)	Elderly (N=12)	$t$ (elderly>young)
Frontal WM	0.0367 (0.0187)	0.02982 (0.0251)	$t=-0.7505^a$ $p=0.2307$	0.0349 (0.0190)	0.0275 (0.0194)	$t=-0.9182^a$ $p=0.1844$
Thalamus	0.0464 (0.0230)	0.0220 (0.0129)	$t=-2.1336^a$ $p=0.0224$	0.0420 (0.0210)	0.0208 (0.0317)	$t=-1.8805^a$ $p=0.0370$
Caudate	0.0937 (0.0189)	0.1033 (0.0274)	$t=0.9689$ $p=0.1718$	0.0763 (0.0224)	0.1038 (0.0356)	$t=2.1970$ $p=0.0197$
Putamen	0.0779 (0.0188)	0.1233 (0.0343)	$t=3.8807$ $p=0.0004$	0.0683 (0.0205)	0.1134 (0.0369)	$t=3.5777$ $p=0.0009$
Globus Pallidus	0.1224 (0.0200)	0.1472 (0.0261)	$t=2.5420$ $p=0.0095$	0.1422 (0.0172)	0.1961 (0.0318)	$t=4.9807$ $p=0.0001$
Substantia Nigra	0.0820 (0.0299)	0.1113 (0.0372)	$t=2.0712$ $p=0.0254$	0.1045 (0.0426)	0.1524 (0.0331)	$t=3.0319$ $p=0.0031$
Red Nucleus	0.0933 (0.0379)	0.1473 (0.0413)	$t=3.2568$ $p=0.0019$	0.0927 (0.0395)	0.1435 (0.0458)	$t=2.8404$ $p=0.0049$
Dentate Nucleus	0.0693 (0.0151)	0.0595 (0.0292)	$t=-1.0000^a$ $p=0.1643$	0.0544 (0.0174)	0.0487 (0.0225)	$t=-0.6703^a$ $p=0.2550$

p-values are 2-tailed. Numbers in bold indicate significant differences, family-wise Bonferroni corrected based on one-tailed directional hypotheses, requiring  $p \leq 0.0125$  for 8 comparisons.

<sup>a</sup> Negative  $t$  values indicate a group difference with the elderly having less iron than the young.

**Table 1b. Mean ( $\pm$ SD) of each measure by region for each group:  $\ell_2$  regularized QSM results using phase-guided ROIs and FDRI-guided ROIs**

Region	$\ell_2$ -regularized QSM (ppm), phase-guided ROIs			$\ell_2$ -regularized QSM (ppm), FDRI-guided ROIs		
	Young (N=11)	Elderly (N=12)	$t$ (elderly>young)	Young (N=11)	Elderly (N=12)	$t$ (elderly>young)
Frontal WM	0.0240 (0.0146)	0.0191 (0.0143)	$t=-0.8163^a$ $p=0.2118$	0.0228 (0.0156)	0.0187 (0.0124)	$t=-0.7029^a$ $p=0.2449$
Thalamus	0.0388 (0.0214)	0.0155 (0.0194)	$t=-2.738^a$ $p=0.0061$	0.0344 (0.0199)	0.0139 (0.0211)	$t=-2.3931^a$ $p=0.0131$
Caudate	0.0814 (0.0164)	0.0897 (0.0195)	$t=1.1032$ $p=0.1412$	0.0653 (0.0211)	0.0888 (0.0276)	$t=2.2814$ $p=0.0166$
Putamen	0.0677 (0.0168)	0.1101 (0.0248)	$t=4.7501$ $p=0.0001$	0.0568 (0.0176)	0.0976 (0.0264)	$t=4.3091$ $p=0.0002$
Globus Pallidus	0.1069 (0.0188)	0.1341 (0.0233)	$t=3.0833$ $p=0.0028$	0.1221 (0.0153)	0.1740 (0.0298)	$t=5.1724$ $p=0.0001$
Substantia Nigra	0.0656 (0.0280)	0.0939 (0.0246)	$t=2.5812$ $p=0.0087$	0.0832 (0.0354)	0.1210 (0.0227)	$t=3.0743$ $p=0.0029$
Red Nucleus	0.0740 (0.0333)	0.1184 (0.0331)	$t=3.2024$ $p=0.0021$	0.0738 (0.0339)	0.1141 (0.0379)	$t=2.6751$ $p=0.0071$
Dentate Nucleus	0.0570 (0.0137)	0.0509 (0.0178)	$t=-0.9161^a$ $p=0.1850$	0.04314 (0.0146)	0.0400 (0.0147)	$t=-0.5076^a$ $p=0.3085$

**Table 1c. Mean ( $\pm$ SD) of each measure by region for each group: FDRI results using phase-guided ROIs and FDRI-guided ROIs**

Region	FDRI ( $s^{-1}$ /Tesla), phase-guided ROIs			FDRI ( $s^{-1}$ /Tesla), FDRI-guided ROIs		
	Young (N=11)	Elderly (N=12)	$t$ (elderly>young) $p$	Young (N=11)	Elderly (N=12)	$t$ (elderly>young) $p$
Frontal WM	2.02 (0.3268)	1.545 (0.4522)	$t=-2.8643^a$ $p=0.0093$	2.0732 (0.6149)	1.5976 (0.6144)	$t=-1.8535^a$ $p=0.0779$
Thalamus	2.331 (0.5172)	1.698 (0.6105)	$t=-2.6712^a$ $p=0.0143$	2.2635 (0.5353)	1.6767 (0.6229)	$t=-2.4115^a$ $p=0.0251$
Caudate	2.531 (0.4752)	3.198 (0.9042)	$t=2.1812$ $p=0.0407$	2.5384 (0.3842)	2.9789 (1.0421)	$t=1.3198$ $p=0.2011$
Putamen	2.954 (0.4282)	3.904 (0.738)	$t=3.7284$ $p=0.0012$	2.8900 (0.4137)	3.9732 (0.7612)	$t=4.1820$ $p=0.0004$
Globus Pallidus	4.223 (0.5178)	4.497 (0.9267)	$t=0.8642$ $p=0.3972$	4.8961 (0.4369)	5.5338 (1.0121)	$t=1.9285$ $p=0.0674$
Substantia Nigra	3.225 (0.9541)	3.421 (0.9988)	$t=0.4804$ $p=0.6359$	3.1479 (0.9576)	3.9619 (0.9641)	$t=2.0290$ $p=0.0553$
Red Nucleus	3.268 (0.9763)	3.932 (0.8528)	$t=1.7415$ $p=0.0962$	3.1284 (0.8765)	3.99916 (0.7634)	$t=2.5240$ $p=0.0197$
Dentate Nucleus	2.41 (0.7971)	2.533 (0.8682)	$t=0.3546$ $p=0.7264$	2.0137 (0.5972)	1.9244 (0.5801)	$t=-0.3637$ $p=0.7196$

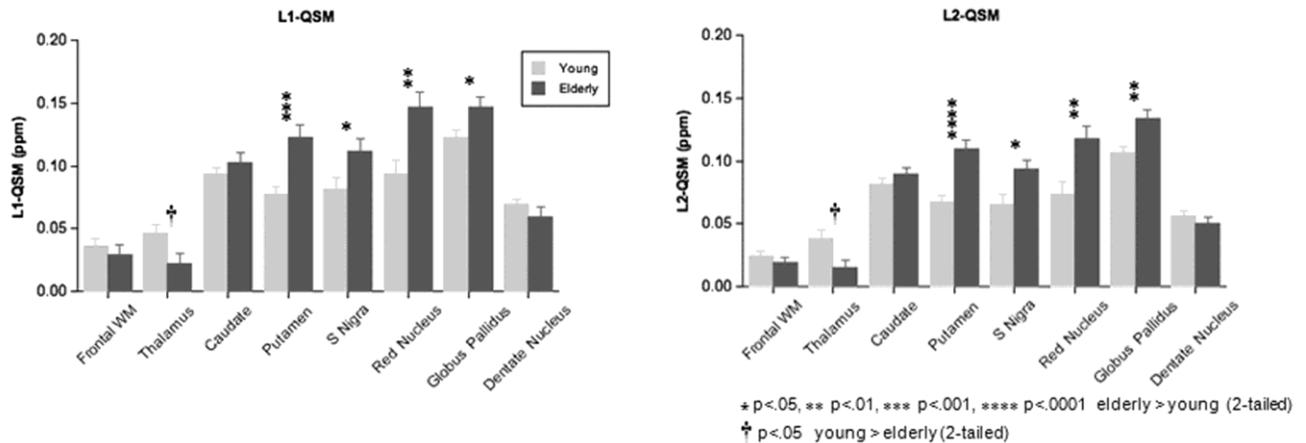


Fig. 3.6. Mean  $\pm$  S.E.M. of average susceptibility in ppm computed by the two methods ( $\ell_1$ -regularized QSM, top;  $\ell_2$ -regularized QSM, bottom) for each ROI in the young and elderly groups.

### 3.3.5 Age differences identified with FDRI

The elderly group had a significantly higher FDRI than the young group in the putamen but not the caudate nucleus or the very iron-rich globus pallidus. Although the elderly tended to have higher FDRI values in the red nucleus and substantia nigra, the differences were not significant; the groups did not differ significantly in FDRI of the dentate nucleus. By contrast, the FDRI values in the thalamic and white matter samples were significantly lower (indicative of less iron) in the elderly than the young group.

## 3.4 Discussion

This study presented regularized QSM methods with two different choices of regularization, namely  $\ell_1$  and  $\ell_2$  norm penalties, for quantifying susceptibility-weighted imaging data, and established their ability to measure iron concentration in regional striatal and brain stem nuclei of young and elderly adults. The *in vivo* estimates of regional iron concentration comported well with published *postmortem* measurements (1), with both approaches yielding the same rank ordering of iron concentration by brain structure, from lowest in white matter to highest in globus pallidus. Further validation was provided by comparison of the *in vivo* measurements, the two QSM methods and FDRI, which again yielded perfect rank ordering of iron by structure. The final means of validation was to assess how well each *in vivo* method detected known age-related differences in regional iron concentrations measured in the same young and elderly healthy adults. Results from all three methods were consistent in identifying higher iron concentrations in striatal and brain stem ROIs (i.e., caudate nucleus, putamen, globus pallidus, red nucleus and substantia nigra) in the older than the young group. With the exception of  $\ell_1$ -regularized results for the substantia nigra averaged under phase-guided ROIs, QSM values in the globus pallidus, red nucleus and substantia nigra were significantly larger in the elderly than the young based on both FDRI- and phase-guided ROIs using  $\ell_1$  or  $\ell_2$  regularization. For the FDRI metric, significant difference was observed only in the putamen for FDRI- and phase-guided delineation. Therefore, QSM appeared more sensitive than FDRI in detecting age differences in brain stem structures by producing much smaller *p*-values in the statistical tests. Although both measurement approaches identified the globus pallidus as being the most iron-rich structure regardless of age, only QSM found that the concentration in the elderly was significantly higher than that in the young adults. The average susceptibility value in the globus pallidus of young subjects has been reported to be around 0.20 ppm by several groups, e.g. (45,62) (taking CSF as reference, with isotropic voxels),

which is larger than the group averages reported in this study (0.10 – 0.14 ppm, taking splenium as reference) . This difference might stem from averaging across subjects and partial volume issues considering the 2.5 mm slice thickness used in data acquisition.

The two regularized QSM methods produced iron concentration estimates consistent with the well-established FDRI metric. In addition to yielding strongly correlated results to both FDRI and *postmortem* data, the susceptibility mapping approach possesses several other favorable qualities. First, the data acquisition step for QSM is completed at a single field strength, whereas acquisitions at two field strengths are required to compute the FDRI values. Working at a single field strength also eliminates the need for spatial registration, and thus a potential source of measurement error. Second, the susceptibility maps estimated with the QSM algorithms have a higher spatial resolution than the FDRI images. This has the additional benefit of enabling the quantification of vessel oxygenation ratios, because the individual vessels can be clearly resolved in the produced  $\chi$  maps. However, the presented QSM algorithms produce relative maps of tissue susceptibility, which requires the selection of a reference susceptibility value for absolute quantification. In this study, the average susceptibility of splenium in each subject was taken as reference, but a point to note is that white matter samples have been reported to have anisotropic susceptibility (63), i.e., their susceptibility values depend on the orientation relative to the main magnetic field.

The regularized QSM algorithms can be considered a refinement of the pioneering work by Haacke (34,41,42) on Susceptibility-Weighted Imaging (SWI), which estimates local iron concentration by inspecting the changes in gradient-echo image phase. Because the background phase constitutes the major part of the observed phase, high-pass filtering is applied to obtain an estimate of the phase accrued by the tissue iron while removing the slowly-varying background effects. Although practical, filtering also removes some tissue phase information (48).

The proposed method addresses this problem by using an optimization approach called *dipole fitting* (48) that estimates and subtracts the background phase without affecting the tissue phase. In addition to yielding high-quality tissue field maps, dipole fitting only requires the solution of a least-squares problem, which can be done using a variety of gradient or conjugate direction optimization methods. As opposed to the high-pass filtering approach, which requires optimal selection of filter size, and polynomial fitting that depends on the order of the polynomial, dipole fitting contains no parameters that need tuning. On the other hand, high-pass filtering methods are dramatically faster than iterative optimization methods employed in the dipole fitting approach. In addition, rather than relying only on the image phase, which produces a spatially distorted

measure of tissue iron concentration, the proposed method solves for the underlying paramagnetic property of the tissue and produces a regularized measure of  $\chi$ , which in turn is a sensitive estimate of iron concentration.

Other susceptibility mapping algorithms have demonstrated robust results. An elegant approach by Schweser et al. (2011) estimated the  $\chi$  distribution without employing regularization. This approach, however, requires data to be acquired at three different orientations with respect to the main magnetic field, thereby providing challenges to subjects in terms of scan time and head positioning and challenges to post-acquisition processing in terms of spatial registration. Another influential QSM algorithm using regularization was introduced by de Rochefort et al. (2010) and it forms the basis of the  $\ell_2$ -regularized method used in our work. After obtaining the tissue field map by solving a least squares problem similar to the dipole fitting formulation of Liu et al. (2010), this QSM algorithm places a weighted  $\ell_2$  norm penalty on the spatial gradients of  $\chi$ . However, posing the reconstruction problem with an  $\ell_1$  norm penalty that promotes sparsity in the spatial gradient domain of the susceptibility distribution may be a better fit to the nature of the problem. As the susceptibility kernel effectively undersamples the  $k$ -space of the tissue field map, the inversion problem is inherently an under-determined system similar to the one encountered in the compressed-sensing literature (54). The demonstrated ability of sparsity-inducing priors in undersampled image reconstruction makes the  $\ell_1$  norm an excellent candidate for susceptibility mapping (64), and the  $\ell_1$ -regularized algorithm in this study parallels this effort. An interesting comparison in (62) between the  $\ell_2$ -regularized approach similar to that of (47) against a multiple-orientation reconstruction strategy should also be noted. These results indicate that  $\ell_2$ -regularized single-orientation susceptibility maps yield iron estimates of quality comparable to those calculated using data acquired at multiple orientations.

### 3.5 Fast $\ell_2$ -regularized QSM

This section presents a solution to the regularized QSM formulation that is computed in less than 5 *seconds*, which yields the exact minimizer of the optimization problem unlike time-consuming iterative methods. The proposed method is straightforward to implement and can be coded in a single line of Matlab code. Results are presented on a numerical phantom with known susceptibility and on in vivo data.

### 3.5.1 Methods

$\ell_2$ -regularized reconstruction involves the minimization of  $\|\delta_{in} - \mathbf{F}^H \mathbf{D} \mathbf{F} \chi\|_2^2 + \beta \cdot \|\mathbf{G} \chi\|_2^2$ , as introduced in Eq. (3.7). The minimizer can be evaluated in closed-form by taking the gradient and setting it to zero,

$$\chi_{in} = (\mathbf{F}^H \mathbf{D}^2 \mathbf{F} + \beta \cdot \mathbf{G}^H \mathbf{G})^{-1} \mathbf{F}^H \mathbf{D} \mathbf{F} \delta_{in} \quad (3.10)$$

Eq. (3.10) can be computed efficiently given that the matrix inversion is rapidly performed. The gradient along the x-axis can be expressed as

$$\mathbf{G}_x = \mathbf{F}^H \mathbf{E}_x \mathbf{F} \quad (3.11)$$

where  $\mathbf{E}_x$  is a diagonal matrix with entries  $\mathbf{E}_x(i, i) = 1 - e^{(-2\pi\sqrt{-1}k_x(i,i)/N_x)}$ , which is the  $k$ -space representation of the difference operator  $\delta_x - \delta_{x-1}$ . Here,  $k_x$  is the  $k$ -space index and  $N_x$  is the matrix size along x, and  $\mathbf{G}_y$  and  $\mathbf{G}_z$  are similarly defined. With this formulation, the closed-form solution becomes,

$$\chi_{in} = \mathbf{F}^H \mathbf{D} [\mathbf{D}^2 + \beta \cdot (\mathbf{E}_x^2 + \mathbf{E}_y^2 + \mathbf{E}_z^2)]^{-1} \mathbf{F} \delta_{in} \quad (3.12)$$

The total cost is two FFTs and multiplication of diagonal matrices. For comparison, the objective function is minimized iteratively using nonlinear conjugate gradient (CG) (6). 100 CG iterations were used for all results. Experiments were performed on two datasets;

- i. The first set is a numerical phantom with 3-compartments (gray and white matter, CSF). Within each compartment,  $\chi$  is constant and equal to  $\chi_{gray} = -0.023$ ,  $\chi_{white} = 0.027$ ,  $\chi_{CSF} = -0.018$  ppm (65). The field map  $\delta$  (Fig.3.7a) is computed from the ground truth  $\chi$  map using the forward dipole model and Gaussian noise with peak-SNR = 100 was added, so that the normalized RMSE of the noisy field map was 5.9% relative to the noise free phase.  $\beta$  was chosen to minimize the RMSE in the reconstructed  $\chi$ , and was found to be  $\beta = 2 \cdot 10^{-4}$ . The same  $\beta$  was used for both the closed form and iterative reconstructions.
- ii. The second dataset is a 3D SPGR on a healthy subject at 1.5T with resolution  $0.94 \times 0.94 \times 2.5 \text{mm}^3$  and TR/TE = 58ms/40ms. Background phase (Fig.3.8a) was removed using dipole fitting (48).  $\beta = 1.5 \cdot 10^{-2}$  was chosen based on the L-curve heuristic. Data were zero-padded to twice the size to avoid aliasing with circular convolution.

### 3.5.2 Results

Fig. 3.7 shows closed-form QSM reconstruction and the error relative to the ground-truth  $\chi$  for the numerical phantom. Using Matlab running on a standard workstation, the proposed method took 3.3 seconds and yielded 17.4% RMSE, while the iterative algorithm gave 18.0% error in 65 minutes.

In vivo reconstruction results are presented in Fig. 3.8, where the processing time was 1.3 seconds for the proposed method and 29 minutes for the iterative CG algorithm. The difference between the closed-form and iterative solutions was computed to be 0.3% RMSE, and is depicted at 250-times scaling in Fig.3.8c.

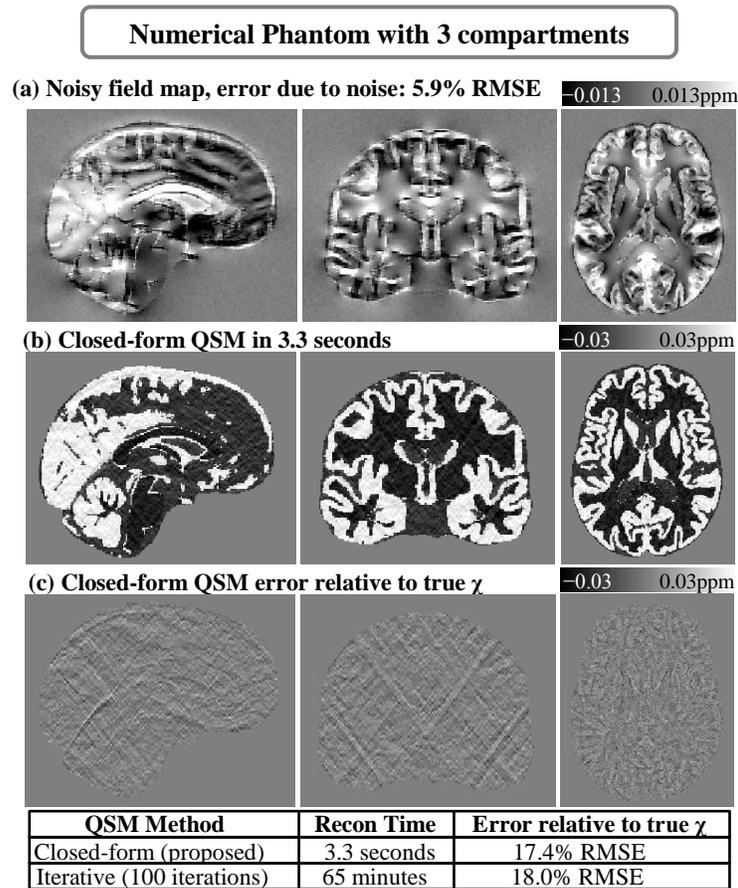


Fig. 3.7 Reconstruction experiment for the piece-wise constant numerical phantom with 3 compartments. (a) Noisy field map from which the susceptibility is estimated. (b) Closed-form QSM solution. (c) Difference between ground truth and closed-form reconstructions.

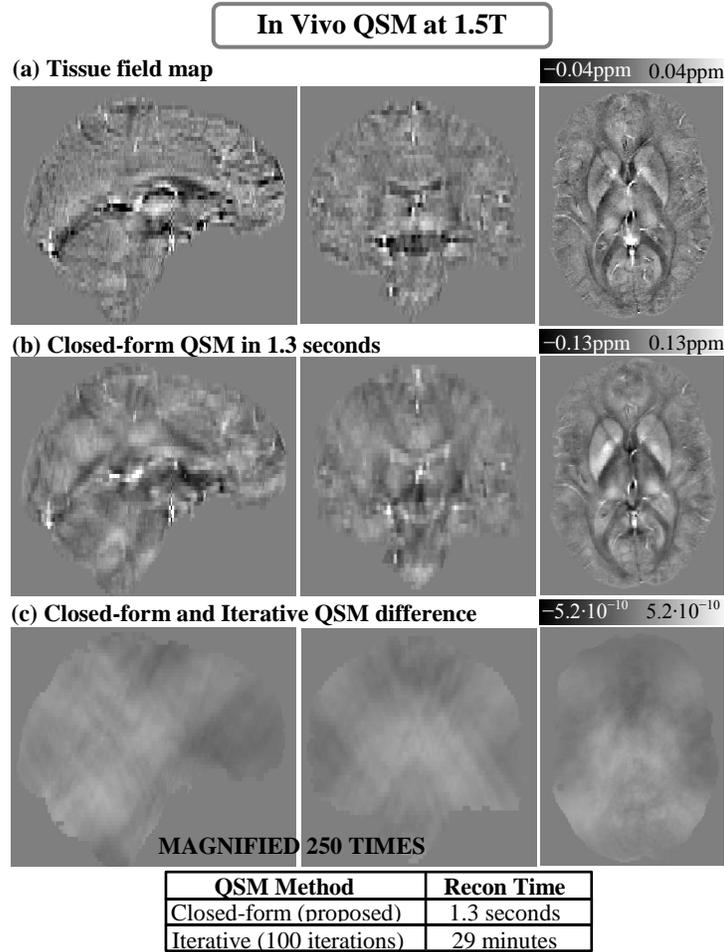


Fig. 3.8 In vivo reconstruction at 1.5T. (a) Tissue field map obtained after removing the background phase. (b) Closed-form QSM solution. (c) Difference between iterative and closed-form solutions.

### 3.5.3 Remarks on the Fast $\ell_2$ -regularized QSM

The proposed closed-form solution is demonstrated to yield much faster and more accurate results than its iterative counterpart. This QSM solver is expected to facilitate online reconstruction of susceptibility maps.

## 3.6 Conclusion

Herein are presented two regularized Quantitative Susceptibility Mapping algorithms, employing  $\ell_1$  and  $\ell_2$  norm regularization, which successfully remove background phase effects via dipole fitting and solve for the tissue susceptibility distribution via convex optimization. The

performance of these algorithms was favorable when compared with other published *in vivo* and *postmortem* estimates of regional tissue iron concentrations. Because the accumulation of iron in the brain can have untoward effects on motor and cognitive function in normal aging (38,39) and can be disproportionately greater in degenerative diseases (66-72), quantitative assessment of this accumulation has the potential of providing a tool for monitoring or even diagnosis. The robustness, practicality, and demonstrated ability of predicting the change in iron deposition in adult aging suggest that the presented QSM algorithms using single-field-strength data is a possible alternative for FDRI tissue iron estimation requiring two field strengths. Further, a closed form expression for  $\ell_2$ -regularized QSM is developed, which leads to estimation of susceptibility maps within seconds.



## Chapter 4

### Lipid Suppression in Chemical Shift Imaging

Mapping  $^1\text{H}$  brain metabolites using chemical shift imaging (CSI) is hampered by the presence of subcutaneous lipid signals, which contaminate the metabolites by ringing due to limited spatial resolution. Even though CSI at spatial resolution high enough to mitigate the lipid artifacts is infeasible due to signal-to-noise (SNR) constraints on the metabolites, the lipid signals have orders of magnitude higher concentration, which enables the collection of high-resolution lipid maps with adequate SNR. The previously proposed dual-density approach exploits this high-SNR property of the lipid layer to suppress truncation artifacts using high-resolution lipid maps. Another recent approach for lipid suppression makes use of the fact that metabolite and lipid spectra are approximately orthogonal, and seeks sparse metabolite spectra when projected onto lipid-basis functions. The present work combines and extends the dual-density approach and the lipid-basis penalty, while estimating the high-resolution lipid image from 2-average  $k$ -space data to incur minimal increase on the scan time. Further, the spectral-spatial sparsity of the lipid ring is exploited to estimate it from substantially undersampled (acceleration  $R = 10$  in the peripheral  $k$ -space) 2-average *in vivo* data using compressed sensing, and improved lipid suppression relative to using dual-density or lipid-basis penalty alone is still obtained.

#### 4.1 Introduction

The spatial resolution in proton spectroscopic imaging is constrained by the low SNR of the metabolite signals and the total scan time required for encoding in both chemical shift and space. Poor spatial resolution with impulse response functions of either square or circular  $k$ -space sampling leads to significant spatial ringing artifacts, which in the case of large and undesirable signals from subcutaneous lipid layer in spectroscopic imaging of the brain can significantly contaminate the desired metabolite spectra throughout the brain. Considering that the lipid signals are several orders of magnitude stronger than the biochemical spectra, the diagnostic quality of spectroscopic data is severely limited if the truncation artifacts are not mitigated by some means of lipid suppression.

Standard means of lipid suppression include outer-volume suppression (OVS) (73-75), inversion recovery (76-78), and selective brain-only excitation (79,80). Although these methods provide effective artifact reduction, their inevitable tradeoff and common drawback is the associated loss of brain metabolite signals, either through signal loss in peripheral brain regions (e.g. OVS, PRESS) or throughout the brain (IR). Another proposal for lipid artifact reduction is to acquire CSI data with a variable sampling density pattern and apply SNR-optimal apodization in the  $k$ -space to reduce the side-lobes of the point spread function (81). Optimal filters specifically designed to reduce the lipid contamination inside the brain yield further improvement over the variable density approach (82). An alternative approach acquires high-resolution lipid maps in addition to highly oversampled, low-resolution CSI data. This dual-density method (83-85) exploits the fact that the lipid signals have high SNR, so a high-resolution lipid estimate can be obtained with adequate SNR for subsequent processing, which includes spatial lipid masking and combination with low-resolution CSI data. Another research direction involves  $k$ -space extrapolation with prior knowledge of spatial boundaries of the brain (86,87). In particular, effective lipid suppression is demonstrated at a relatively short TE of 50 ms in (87). A yet different method of lipid suppression was recently proposed (88) by relying on the approximation that the metabolite and lipid spectra are orthogonal, and seeks sparse metabolite spectra when projected onto lipid-basis functions selected from the lipid layer.

The present work combines and extends the dual-density approach and the iterative lipid-basis reconstruction. A method to estimate the high-resolution lipid image from 2-average  $k$ -space data in fast spiral CSI is proposed and demonstrated, wherein these data are combined with the low-resolution CSI image while imposing the lipid-basis penalty. This way, the truncation artifacts are substantially reduced at the expense of minimal increase in total scan time. This method is then refined by incorporating the observation that the high-resolution lipid ring is sparse in both space and chemical shift. This leads to successful recovery of the lipid image via compressed sensing (4,6) using highly-undersampled peripheral  $k$ -space data.

To demonstrate the performance of the proposed methods, single-slice, high-resolution (0.16 cc) CSI data were acquired *in vivo* at 3T with 20 averages, requiring 33 min of scan time. Applying the lipid-basis penalty to this high-resolution data yielded virtually artifact-free spectra, which were taken to be the gold-standard results. To apply the basic method with fully-sampled lipid data, 20 averages of low-resolution (0.56 cc, corresponding to 10 min of scan time) CSI data were combined with 2 averages of high-resolution data while imposing lipid-basis penalty, and reduced-artifact metabolite spectra were obtained with normalized RMSE (NRMSE) of 8.5 % in

the NAA maps relative to the gold-standard reconstruction. However, using the lipid-basis penalty approach (88) with 20 averages of 0.56 cc data yielded 41.3 % NRMSE in the NAA maps. Moreover, using the refined method, a high-resolution lipid layer was estimated via the FOCUSS algorithm (4) from 2-average, highly undersampled ( $R_{\text{high}}=10$  in the peripheral  $k$ -space) data, which was combined with the 0.56 cc CSI image followed by lipid-basis penalty reconstruction to yield 17.0 % NRMSE in the NAA map. By incurring only a minimal increase in the scan time, 4.9- and 2-fold error reduction in metabolite maps are demonstrated relative to (88) using the basic and refined versions of the proposed method, respectively. Further, validation for the application of undersampling and compressed sensing recovery using variable density spirals is presented with 10-fold undersampling on a synthetic phantom.

## 4.2 Theory

### 4.2.1 Dual-Density Reconstruction

Let  $\mathbf{y}_{\text{low}}$  denote the  $k$ -space representation of low-resolution CSI data, and  $\mathbf{y}_{\text{high}}$  denote the  $k$ -space representation of high-resolution data from which the lipid image will be estimated due to

$$\mathbf{x}_{\text{lipid}} = \mathbf{M}_{\text{lipid}} \mathbf{F}_{\text{high}}^{-1} \mathbf{y}_{\text{high}} \quad (4.1)$$

where  $\mathbf{x}_{\text{lipid}}$  is the high-resolution, masked lipid layer image,  $\mathbf{M}_{\text{lipid}}$  is a binary mask marking the location of the lipid layer, and  $\mathbf{F}_{\text{high}}$  is the Fourier Transform operator that samples the full extent of high-resolution  $k$ -space. Since  $\mathbf{y}_{\text{high}}$  usually has low SNR, the masking operation aims to select only the lipid layer and reduce the amount of noise that will propagate from the rest of the data.

Next, the low-resolution data is combined with the high-resolution lipid image via

$$\mathbf{x}_{\text{dual}} = \mathbf{F}_{\text{high}}^{-1} \{ (\mathbf{F}_{\text{high}} - \mathbf{F}_{\text{low}}) \mathbf{x}_{\text{lipid}} + \mathbf{y}_{\text{low}} \} \quad (4.2)$$

Here,  $\mathbf{F}_{\text{low}}$  is the Fourier Transform operator that samples only the lower frequency indices corresponding to  $\mathbf{y}_{\text{low}}$ . Eq. 4.2 can be interpreted as extending the low-resolution  $k$ -space data using the high frequency content of the masked lipid image, which helps reducing the ringing artifact (83-85).

### 4.2.2 Iterative Reconstruction with Lipid-Basis Penalty

Again starting with the low-resolution CSI  $k$ -space data  $\mathbf{y}_{low}$ , the artifact reduction algorithm in (88) aims to solve the convex programming problem

$$\mathbf{x}_{LB} = \underset{\mathbf{x}}{\operatorname{argmin}} \underbrace{\|\mathbf{F}_{low}\mathbf{x} - \mathbf{y}_{low}\|_2^2}_{\text{data consistency}} + \lambda \underbrace{\sum_{i \in \mathbf{M}_{brain}} \|\mathbf{L}_{low}^H \mathbf{x}_i\|_1}_{\text{lipid-basis penalty}} \quad (4.3)$$

where  $\mathbf{M}_{brain}$  is a binary mask that indicates the metabolite region,  $\mathbf{x}_i$  is the spectrum at voxel  $i$ ,  $\lambda$  is a regularization parameter that needs to be determined and  $\mathbf{x}_{LB}$  is the artifact-suppressed image. After denoting the initial image with truncation artifacts with  $\mathbf{x}_{low}$  (where  $\mathbf{x}_{low} = \mathbf{F}_{low}^{-1}\mathbf{y}_{low}$ ), a lipid-basis matrix  $\mathbf{L}_{low}$  can be formed using the spectra inside the lipid layer of  $\mathbf{x}_{low}$  as column vectors. Hence, to generate the lipid-basis  $\mathbf{L}_{low}$ , the initial image with artifacts  $\mathbf{x}_{low}$  is masked to retain only the lipid ring voxels. Next, each lipid spectra is assigned to be a column of the lipid-basis  $\mathbf{L}_{low}$ . This way, the lipid-basis matrix will have  $n$  columns, where  $n$  is the number of voxels in the lipid mask, and each of its columns will be a lipid spectrum. Eq. 4.3 then aims to find spectra that match the acquired  $k$ -space data, but at the same time impose the constraint that no lipid signals arise from the brain itself.

The cost function in the iterative lipid-basis penalty reconstruction is composed of *data consistency* and *lipid-basis penalty* terms (Eq.4.3) which penalize the deviation from the  $k$ -space samples and the projection onto the lipid-basis, respectively. As the cost is composed of a linear combination of the convex  $\ell_2$  and  $\ell_1$  norms, the optimization problem is an unconstrained convex programming problem, which has the important feature that all local minima are also global (89).

### 4.2.3 The Basic Method: Combining 2-average, high-resolution data with high SNR, low-resolution data

The first proposal in this chapter is to combine the two orthogonal lipid suppression approaches: the dual-density method and the lipid-basis penalty. An additional assumption that the high-resolution  $k$ -space  $\mathbf{y}_{high}$  is obtained with only 2 averages is made, hence it has low metabolite SNR while having a rapid acquisition time, and that the low-resolution  $\mathbf{y}_{low}$  is acquired with multiple averages to yield decent metabolite SNR. The combined image  $\mathbf{x}_{dual}$  is

then formed by the application of Eqs. 4.1 and 4.2. Imposing lipid-basis penalty on  $\mathbf{x}_{dual}$  yields the final result,

$$\mathbf{x}_{basic} = \operatorname{argmin}_{\mathbf{x}} \|\mathbf{F}_{high}\mathbf{x} - \mathbf{y}_{dual}\|_2^2 + \lambda \sum_{i \in \mathbf{M}_{brain}} \|\mathbf{L}_{dual}^H \mathbf{x}_i\|_1 \quad (4.4)$$

where  $\mathbf{y}_{dual} = \mathbf{F}_{high}\mathbf{x}_{dual}$  is the  $k$ -space representation of the high-resolution combined image and  $\mathbf{x}_{basic}$  is the artifact-suppressed spectra obtained with the first proposed method. In this case,  $\mathbf{L}_{dual}$  contains the lipid spectra collected from the combined image  $\mathbf{x}_{dual}$ . After masking  $\mathbf{x}_{dual}$  to retain only the lipid ring voxels, each lipid spectrum is assigned to be a column of the lipid-basis matrix  $\mathbf{L}_{dual}$ . This way, the lipid-basis is formed by using the high-frequency lipid information present in the combined image  $\mathbf{x}_{dual}$ .

#### 4.2.4 The Refined Method: Combining 2-average, *undersampled* high-resolution data with high SNR, low-resolution data

Differently from the first method,  $\mathbf{y}_{high}$  now represents undersampled, 2-average, high-resolution  $k$ -space data. Owing to the fact the lipid layer is sparse in both spatial and spectral domains, this section proposes to estimate it using the sparsity-enforcing, iteratively reweighted least-squares algorithm, FOCUS (4):

For iteration number  $t = 1, \dots, T$ ,

$$\mathbf{W}_{j,j}^t = \operatorname{diag}(|x_j^t|^{1/2}) \quad (4.5)$$

$$\mathbf{q}^t = \operatorname{argmin}_{\mathbf{q}} \|\mathbf{q}\|_2^2 \quad \text{such that} \quad \mathbf{M}_{\Omega} \mathbf{F}_{high} \mathbf{W}^t \mathbf{q} = \mathbf{y}_{high} \quad (4.6)$$

$$\mathbf{x}^{t+1} = \mathbf{W}^t \mathbf{q}^t \quad (4.7)$$

Here,  $\mathbf{W}^t$  is a diagonal weighting matrix whose  $j^{\text{th}}$  diagonal entry is denoted as  $\mathbf{W}_{j,j}^t$ ,  $\mathbf{x}^t$  is the lipid layer estimate at iteration  $t$  whose  $j^{\text{th}}$  entry is  $x_j^t$  and  $\mathbf{M}_{\Omega}$  is the undersampling mask in  $(k_x, k_y, k_f)$ . Masking out the background yields the final lipid image estimate,  $\mathbf{x}_{lipid}^{FOCUSS} = \mathbf{M}_{lipid} \mathbf{x}^{T+1}$ . Now, the combined image  $\mathbf{x}_{dual}^{FOCUSS}$  is formed using the compressed sensing-estimated lipid image,

$$\mathbf{x}_{dual}^{FOCUSS} = \mathbf{F}_{high}^{-1} \{ (\mathbf{F}_{high} - \mathbf{F}_{low}) \mathbf{x}_{lipid}^{FOCUSS} + \mathbf{y}_{low} \} \quad (4.8)$$

and iterative lipid-basis reconstruction is applied as

$$\mathbf{x}_{refined} = \operatorname{argmin}_{\mathbf{x}} \left\| \mathbf{F}_{high} \mathbf{x} - \mathbf{y}_{dual}^{FOCUSS} \right\|_2^2 + \lambda \sum_{i \in \mathbf{M}_{brain}} \left\| (\mathbf{L}_{dual}^{FOCUSS})^H \mathbf{x}_i \right\|_1 \quad (4.9)$$

to yield the artifact-suppressed image  $\mathbf{x}_{refined}$ . Here,  $\mathbf{y}_{dual}^{FOCUSS}$  is the  $k$ -space representation of the combined image  $\mathbf{x}_{dual}^{FOCUSS}$  due to  $\mathbf{y}_{dual}^{FOCUSS} = \mathbf{F}_{high} \mathbf{x}_{dual}^{FOCUSS}$  and  $\mathbf{L}_{dual}^{FOCUSS}$  is the lipid-basis matrix collected from the compressed sensing reconstructed combined image. In other words, lipid ring voxels in the combined image  $\mathbf{x}_{dual}^{FOCUSS}$  are selected with masking, then each lipid spectrum is assigned to be a column of the lipid-basis matrix  $\mathbf{L}_{dual}^{FOCUSS}$ . Hence, the lipid-basis is formed by the lipid spectra in the compressed sensing reconstructed image,  $\mathbf{x}_{dual}^{FOCUSS}$ .

### 4.3 Methods

A healthy volunteer was scanned at a Siemens 3T scanner using 32-channel receive coil with high spatial resolution, single-slice, constant density spiral CSI (voxel size = 0.16 cc,  $\text{FOV}_{xy} = 24$  cm, slice thickness = 1 cm, TE = 50 ms, TR = 2 s, number of averages = 20, acquisition time = 33 min, CHESS pulse applied for water suppression, PRESS-box excites entire FOV, including the skull). While the large number of averages at such high resolution made the total scan time significantly long, it enabled the reconstruction of the artifact suppressed gold-standard image. At the scanner, this spiral acquisition was coil-combined after being gridded onto a Cartesian grid, on which all subsequent processing was performed. The final gridded matrix size was  $(x,y,f) = (64,64,512)$ . To reduce processing times, only the frequencies beyond the water peak were reconstructed. Lipid layer and brain masks ( $\mathbf{M}_{lipid}, \mathbf{M}_{brain}$ ) were generated *manually* based on the high-resolution CSI image. In particular, projection of the CSI image over the lipid frequencies served as a guide in determining the lipid mask. Additional data were collected by using a  $9 \times 9$  cm<sup>2</sup> PRESS-box to excite the interior of the brain (voxel size = 0.5 cc, number of averages = 20, acquisition time = 11 min, with water suppression), and outer-volume suppression bands were placed around the skull to null the lipid signals.

Next, the lipid suppression methods that were applied to the *in vivo* data are detailed and enumerated:

- i. **Lipid-basis penalty method:** A low-resolution, 20-average CSI  $k$ -space  $\mathbf{y}_{low}$  was generated by sampling only the center 32-pixel diameter in  $k_x$ - $k_y$  plane corresponding to the operator  $\mathbf{F}_{low}$ . The voxel size of this low-resolution image was

0.56 cc (with 1cm slice thickness), corresponding to a 10 min scan. This image was then processed using the lipid-basis penalty method (88).

- ii. **Gold-standard reconstruction:** To obtain the gold-standard spectra, a lipid image was obtained from the high-resolution 20-average data which was masked with  $\mathbf{M}_{lipid}$  to retain only the lipid ring, and then combined with the low-resolution 20-average CSI image as per the dual-density approach (83-85) in Eq. 4.2 and iterative reconstruction with lipid-basis penalty (88) was applied to this combined image to yield the gold-standard spectra.
- iii. **The basic method:** For this method, masked high-resolution lipid image was obtained from 2-average high-resolution data, and combined with the low-resolution 20-average CSI image. Lipid basis penalty reconstruction was then applied to this combined image.
- iv. **The refined method:** Here, the high-resolution lipid image was estimated from significantly undersampled 2-average data. In addition to the fully sampled center 32-pixel diameter  $k$ -space, the peripheral  $k$ -space region was substantially undersampled ( $R_{high} = 10$ ). In particular, Cartesian undersampling was applied to the gridded data in all 3-dimensions by generating a randomly-undersampled  $k_x$ - $k_y$  sampling mask at each  $k_f$  sample. High-resolution lipid image was reconstructed with the FOCUSS algorithm (4) using the undersampled  $k$ -space data. This lipid layer estimate was then combined with the low-resolution CSI image, and lipid basis penalty was applied to further reduce the ringing artifacts.
- v. **Dual-density method:** Finally, the dual density method (83-85) was applied *without* using lipid-basis penalty, by obtaining a masked high-resolution lipid image obtained from 2-average high-resolution data, and combining it with the low-resolution 20-average CSI image.

To provide a more practical undersampling example, a synthetically generated phantom was also studied. A Cartesian CSI phantom was formed by using metabolite data from a spectroscopic phantom scanned at 3T with a voxel size of 0.16 cc, and surrounding the phantom with *in vivo* lipid spectra sampled from the 20-average, 0.16 cc human subject dataset (Fig. 4.7). Hence, the

metabolite spectra in the numerical phantom are derived from a spectroscopic phantom where no lipids are present, resulting in metabolite signals free of any lipid contamination. Also, each lipid spectrum in the lipid layer of the numerical phantom is unique and comes from an *in vivo* acquisition where spatial variations of lipids occur naturally. No synthetic noise was added to the numerical phantom, the only noise present is due to the acquisition of the source signals. The peak-to-peak NAA – lipid amplitude ratio was selected to be 1:100. Since the Cartesian phantom demonstrates no lipid ringing artifacts by design, it also serves as the gold-standard image. First, a constant density spiral sampling pattern at Nyquist rate was generated using time-optimal gradient design toolbox (90), from which spiral  $k$ -space data was generated using the Non-Uniform FFT (NUFFT) toolbox (91). Artifact suppression with lipid-basis penalty was applied to obtain a high-resolution lipid-suppressed image based on the spiral  $k$ -space samples. Second, a variable-density spiral trajectory with Nyquist rate sampling in the first half of the  $k$ -space, and undersampling with  $R_{\text{high}} = 10$  in the second half of the  $k$ -space was generated. High-resolution lipid image estimate was generated using FOCUSS algorithm with NUFFT based on the undersampled spiral data. Next, a combined image was formed using the high-resolution lipid estimate and the fully-sampled portion of the  $k$ -space iteratively. Lipid-basis penalty was applied to yield an artifact suppressed image. Finally, a low-resolution image was generated by using only the first of the spiral  $k$ -space, which was then processed with the lipid-basis penalty. The Cartesian image without artifacts serves as a substitute for the gold-standard *in vivo* reconstruction, the Nyquist-rate sampled spiral data represent the *in vivo* basic method reconstruction, and the undersampled spiral data stand for the *in vivo* refined reconstruction. Likewise, the low-resolution spiral image is intended to represent the low-resolution *in vivo* image with lipid-basis penalty.

### 4.3.1 Choosing an Optimal Regularization Parameter

To choose an optimal regularization parameter  $\lambda$  for the lipid-basis penalty that balances the data consistency and artifact suppression, the L-curve approach was employed (55) for the *in vivo* study. After running the iterative reconstruction to compute the gold-standard image for several different regularization parameters, the resulting data consistency  $\|\mathbf{F}_{\text{high}}\mathbf{x}_{\text{gold}} - \mathbf{y}_{\text{dual}}\|_2$  and lipid-basis norms  $\sum_{i \in \mathbf{M}_{\text{brain}}} \|\mathbf{L}_{\text{dual}}^H \mathbf{x}_{\text{gold},i}\|_1$  traced a curve from which the data point with the largest curvature was chosen to be the optimal  $\lambda$ . Analytical curvature computation became possible by expressing the data consistency and lipid-basis penalty as functions of  $\lambda$  by cubic

spline fitting. The optimal value of  $\lambda = 10^{-3}$  that is determined from the gold-standard dataset was then used for all iterative reconstructions in this work, where the optimization problems were solved using the conjugate gradient algorithm (89). Fig. 4.1 depicts the resulting L-curve and projections over the lipid frequencies for various  $\lambda$  values, as well as the curvature values at the sample points.

In the phantom study,  $\lambda = 10^{-1}$  was taken to be the value of the regularization parameter for all of iterative reconstructions.

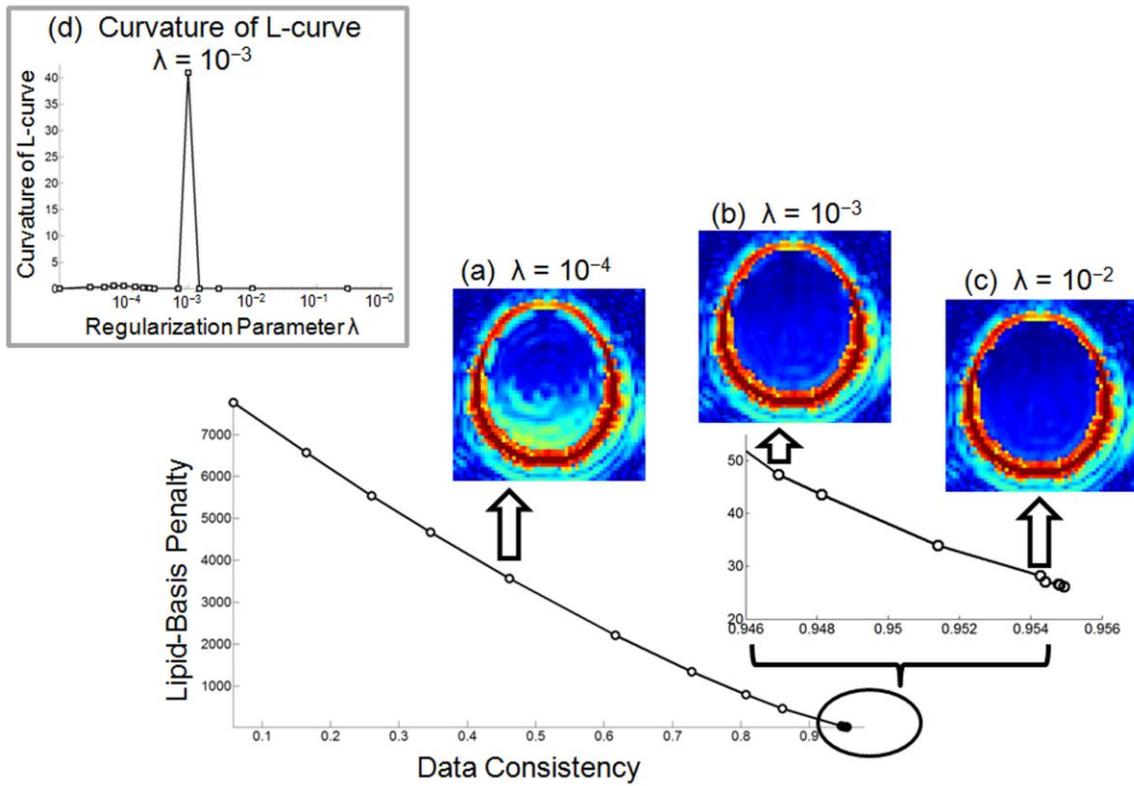


Fig. 4.1. The L-curve traced by the data consistency and lipid-basis penalty terms as the regularization parameter  $\lambda$  varies. Summation over lipid frequencies for under-regularized (a), optimally regularized (b) and over-regularized reconstructions (c) are presented. Panel (d) depicts the analytically computed L-curve curvature results for the sample points.

## 4.4 Results

Artifact reduction performances of the five methods under evaluation, as well as spectra without any lipid suppression are compared by taking projections over the lipid resonance frequencies in Fig. 4.2. High-quality lipid images are obtained with the gold-standard ( $20 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$ , denoting that 20 averages of high-resolution data are used without peripheral  $k$ -space undersampling, shown in Fig. 4.2a), and the basic and refined methods ( $2 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$  in Fig. 2b and  $2 \text{ avg}_{\text{high}}, R_{\text{high}} = 10$  in Fig. 2c). Iterative reconstruction with lipid-basis penalty (88) also demonstrates substantial artifact reduction (Fig. 2d) while not being able to completely remove the ringing inside the brain. Using the dual-density approach (83-85) without lipid-basis penalty (Fig. 2e) provides partial artifact reduction relative to the low-resolution CSI image with no lipid suppression (Fig. 2f).

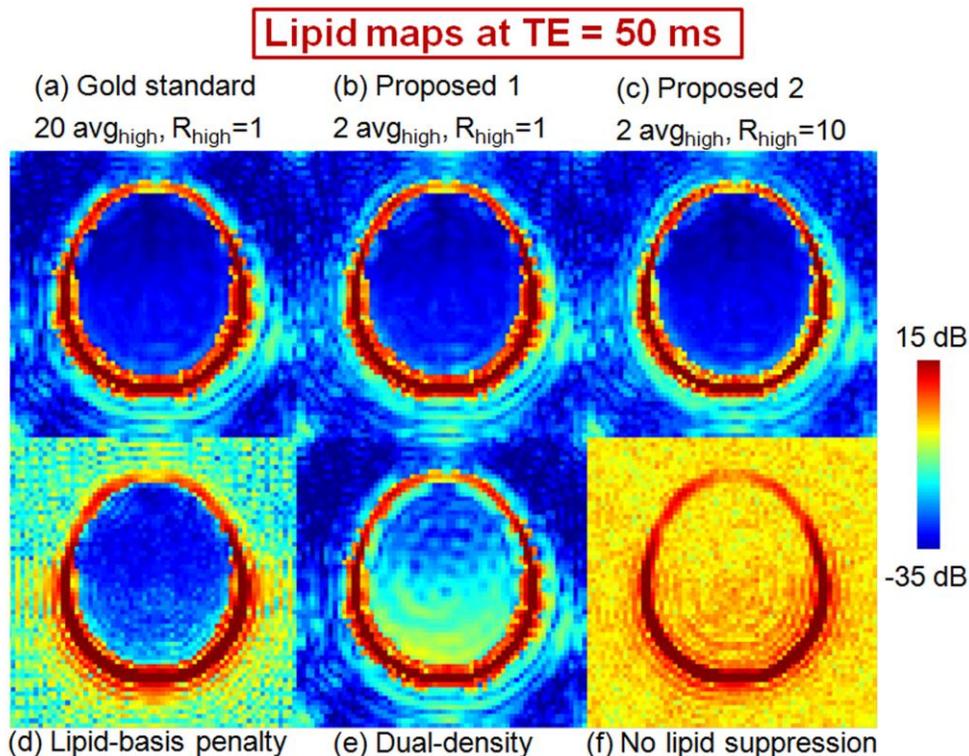


Fig.4. 2. Comparing the different artifact reduction algorithms by taking projections over the lipid resonance frequencies (in dB scale). Gold standard reconstruction is obtained using 20 averages of high-resolution data without peripheral  $k$ -space undersampling ( $20 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$ , shown in (a)), while the basic proposed method is obtained using 2 averages of high-resolution data without undersampling ( $2 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$ , shown in (b)) and the refined proposed method uses 10-fold undersampled, 2 average high-resolution data ( $2 \text{ avg}_{\text{high}}, R_{\text{high}} = 10$ , shown in (c)). Lipid suppression results obtained by using only lipid-basis penalty method and only dual-density approach are depicted in panels (d) and (e), respectively. Applying no lipid suppression (f) results in severely corrupted spectra.

Fig. 4.3 validates the observation seen in Fig. 2 in terms of NRMSE by comparing the lipid-basis penalty algorithm and the two proposed artifact reduction methods with the gold-standard NAA map. All maps are generated by simple integration of NAA peaks over a 37.5 Hz bandwidth. While the lipid-basis algorithm (88) has 41.3% error in the NAA maps relative to the gold-standard, the basic method ( $2 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$ ) reduces the error by 4.9 times to yield 8.5% error, and the refined method ( $2 \text{ avg}_{\text{high}}, R_{\text{high}} = 10$ ) by 2 times to give 17.0% error relative to lipid-basis penalty approach.

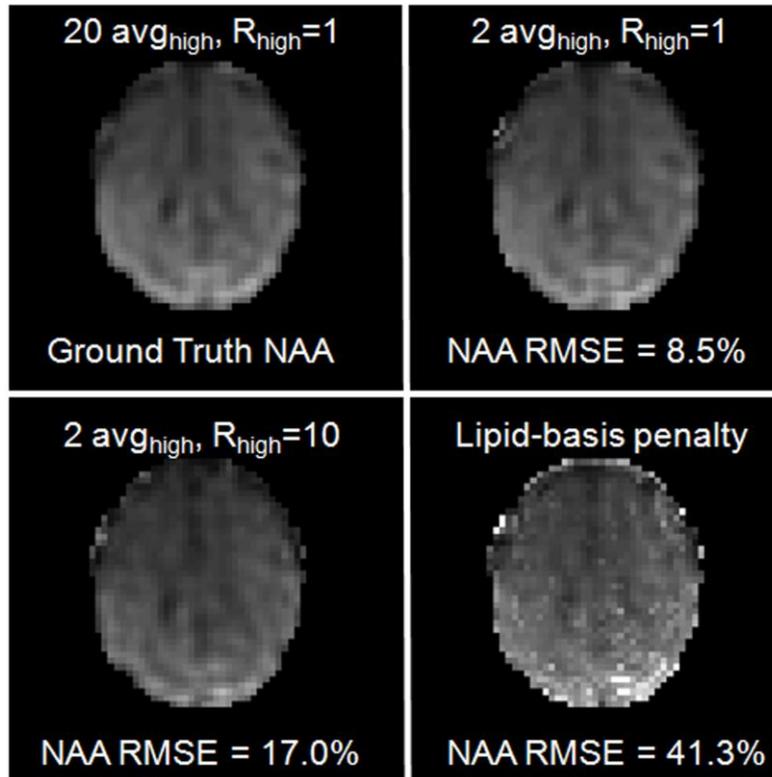


Fig. 4.3. Comparison between NRMSE values of NAA maps relative to the gold standard reconstruction.

Fig. 4.4 presents the NAA maps computed within the  $9 \times 9 \text{ cm}^2$  excitation box used in the OVS acquisition. By taking the OVS NAA images as ground truth, the relative errors were found to be 11.1% for the gold-standard ( $20 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$ , shown in (a)), 11.5% for the basic ( $20 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$ , shown in (b)), 12.9% in the refined method ( $20 \text{ avg}_{\text{high}}, R_{\text{high}} = 10$ , shown in (c)) and 14.7% in the NAA map produced by the lipid-basis penalty algorithm (shown in (d)). Reconstructed spectra are also overplotted with the OVS spectra for the four methods.

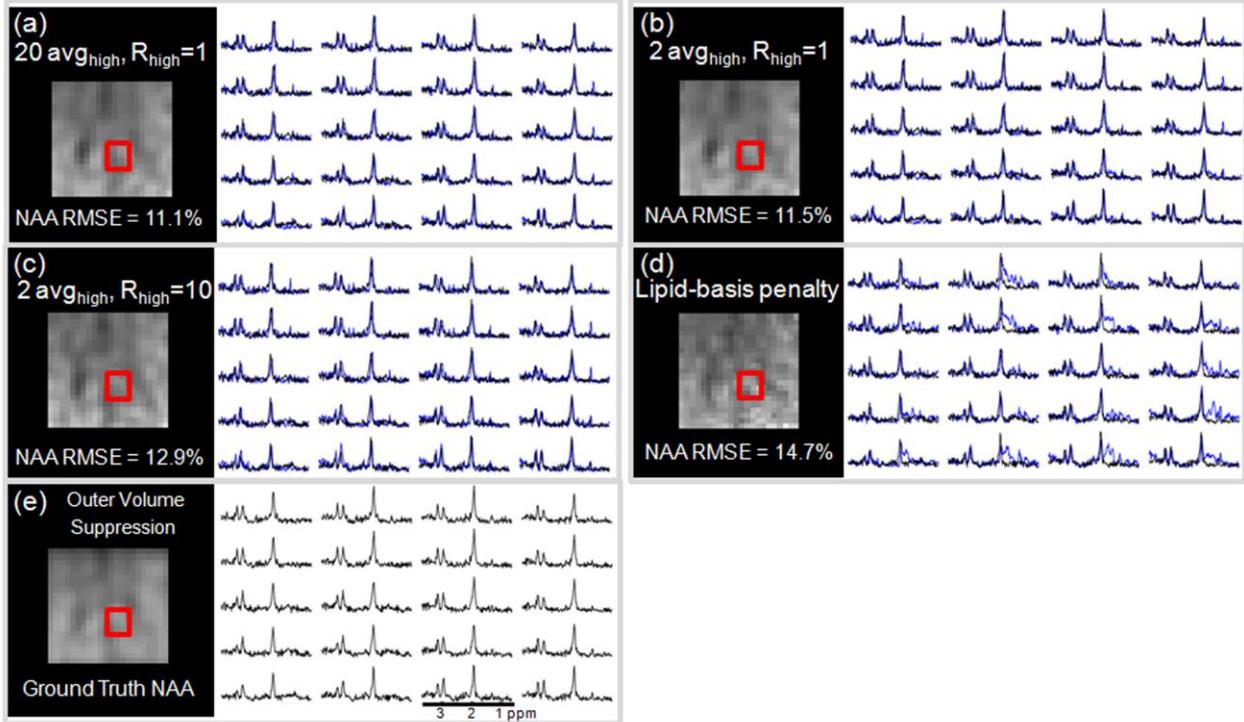


Fig. 4.4. Comparison between NRMSE values of NAA maps computed within the  $9 \times 9 \text{ cm}^2$  excitation box relative to the NAA maps obtained with the OVS method. In (a), reconstruction results obtained using the gold-standard ( $20 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$ ) method (blue) and the OVS spectra (black) belonging to the region inside the red box are also overplotted. In (b), the basic proposed method (blue) and the OVS spectra are compared. The spectra obtained with the refined method (blue) and the OVS results (black) are overplotted in (c). Lipid-basis penalty and OVS spectra are compared in (d).

Figs. 4.5 and 4.6 show the performances of the lipid-basis algorithm (88), the proposed methods and the gold-standard reconstruction by comparing representative spectra in the vicinity of two sides of the skull. Panels a, b and c in Figs. 4.5 and 4.6 overplot the spectra from the gold-standard with lipid-basis method (88), the basic method ( $2 \text{ avg}_{\text{high}}, R_{\text{high}} = 1$ ) and the refined method ( $2 \text{ avg}_{\text{high}}, R_{\text{high}} = 10$ ), respectively.

Lipid suppression experiment performed with the synthetic phantom is depicted in Fig. 4.7. Panel a depicts the NAA and lipid maps from the Cartesian, artifact-free phantom and includes spectra free of contamination. In panel b, lipid-basis penalty is applied to the phantom that was sampled on a spiral trajectory at Nyquist rate, to yield 41.9% error in the NAA map. In c, lipid suppression results with undersampled spiral trajectory are presented. In this case, NAA map was recovered with 41.7% error. Panel d depicts the performance of lipid-basis penalty method when the  $k$ -space was sampled at half of the full resolution to yield 104.1% NRMSE in the NAA map.

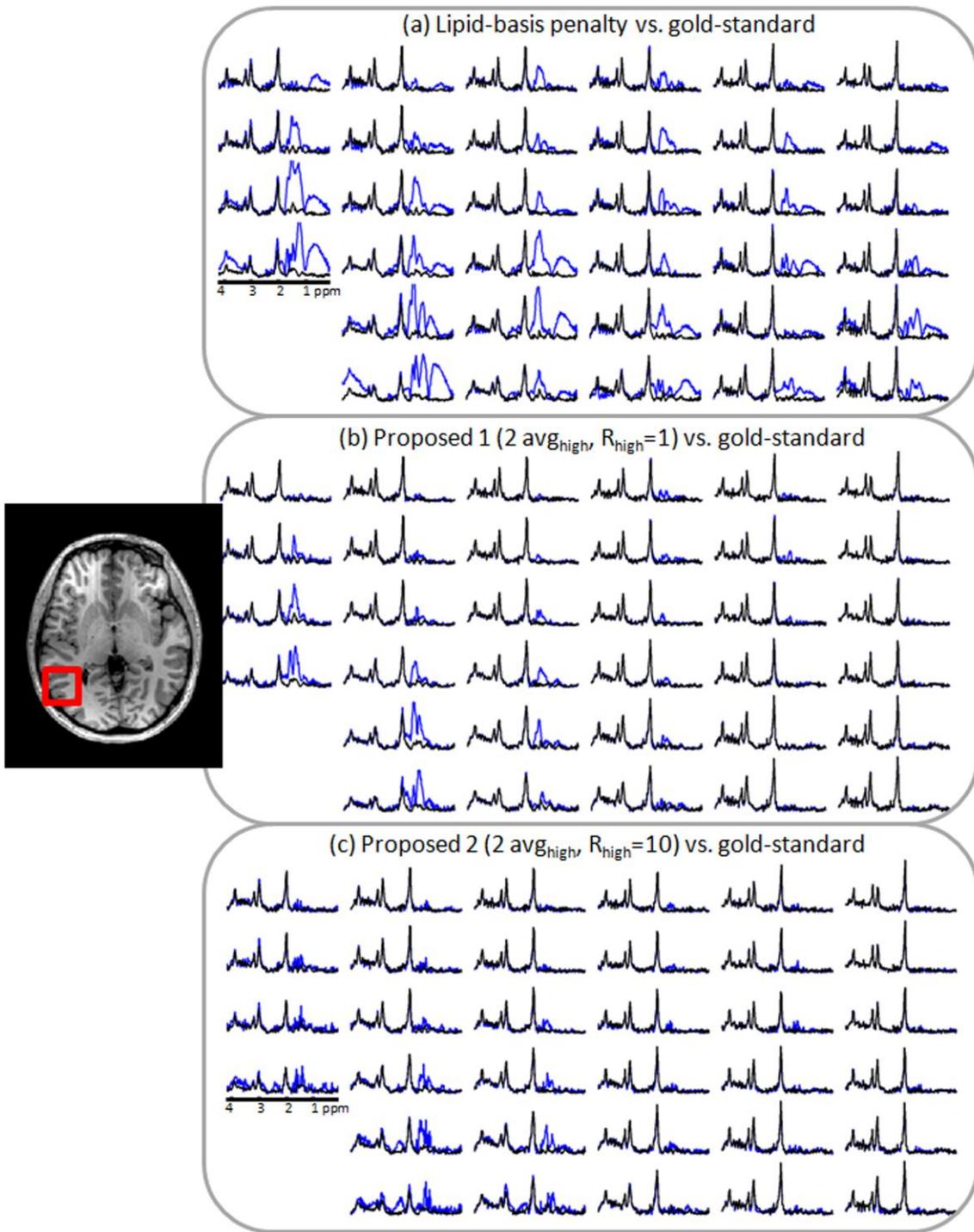


Fig. 4.5. Comparison of spectra inside the region of interest marked with the red box that were obtained with different lipid suppression methods. In (a), reconstruction results obtained using lipid-basis penalty method (blue) and the gold-standard reconstruction (black) are overplotted. In (b), the basic proposed method (blue) and the gold-standard spectra are presented. The spectra obtained with the refined method (blue) and the gold-standard results (black) are plotted in (c).

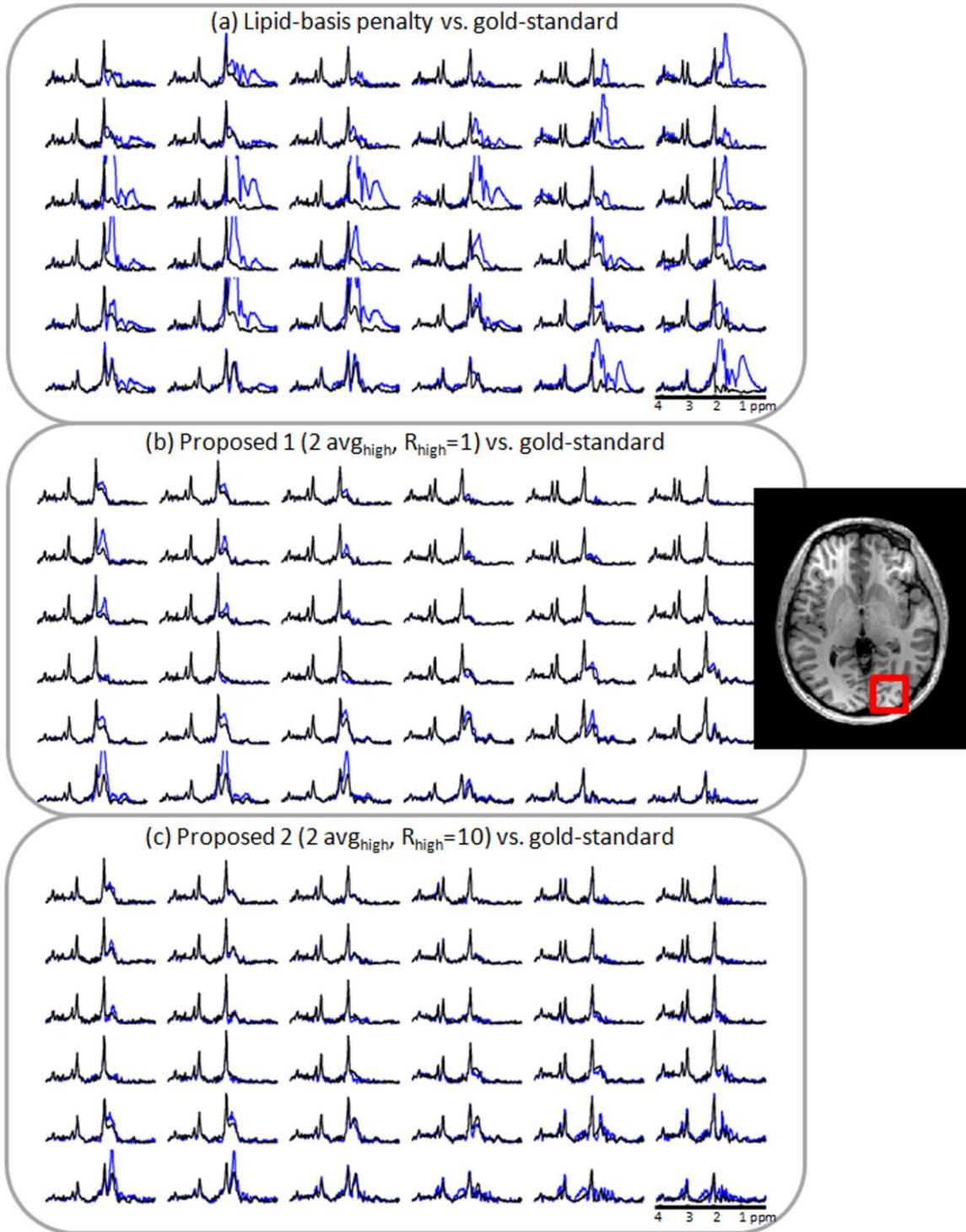


Fig. 4.6. Comparison of spectra inside the region of interest marked with the red box that were obtained with different lipid suppression methods. Panel (a) overplots reconstruction results using lipid-basis penalty method (blue) and the gold-standard reconstruction (black). In (b), the basic proposed method (blue) and the gold-standard spectra are compared. The spectra obtained with the refined method (blue) and the gold-standard results (black) are depicted in (c).

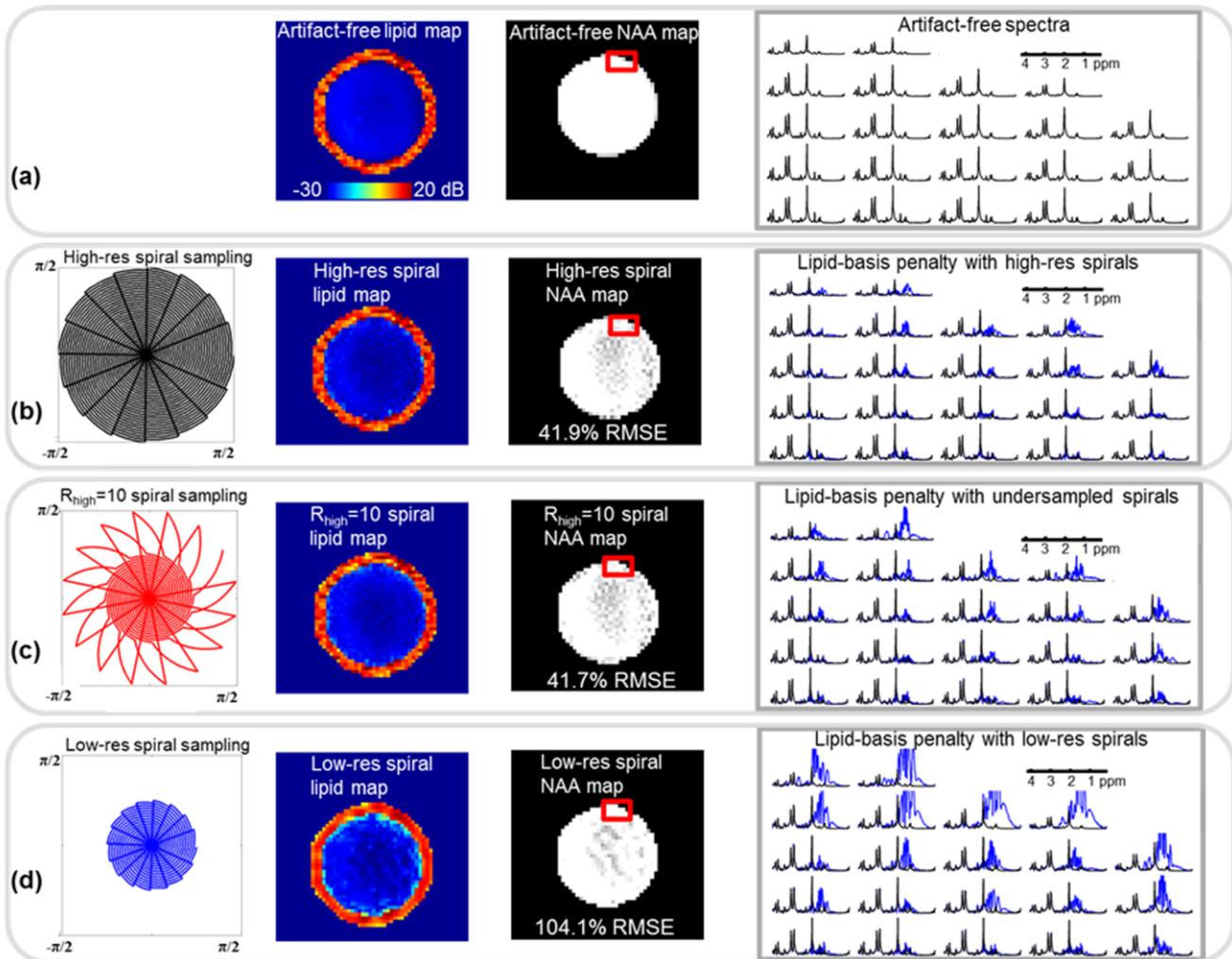


Fig. 4.7. Lipid and NAA maps and artifact-free spectra for the Cartesian synthetic phantom are shown in (a). In (b), spiral sampling trajectory at Nyquist rate and reconstruction results upon the application of lipid-basis penalty are depicted. Using the undersampled spiral trajectory in (c), a high-resolution lipid image was estimated using FOCUSS, from which a combined image was computed due to the dual-density method. Finally, lipid-basis penalty was applied to this combined image. Panel (d) shows lipid suppression results when the  $k$ -space is sampled only at half of the full resolution and lipid-basis penalty is applied. For the three reconstruction methods, the example spectra (plotted in blue) belong to the region of interest marked with the red box, and are overlotted with the artifact-free spectra (in black) for comparison.

The total reconstruction time for the *in vivo* dataset was 7 min for the iterative lipid-basis penalty algorithm and 4 min for compressed sensing reconstruction of the high-resolution lipid image with the FOCUSS algorithm on a workstation running Matlab with 48 GB memory and 12 processors.

## 4.5 Discussion

The dual-density method makes use of the fact that subcutaneous lipid signals have several orders of magnitude higher amplitudes than the brain metabolites, which enables their estimation from single-average, high-resolution data. In this study, the high-resolution lipid images had a voxel size of 0.16 cc, while the previous implementations of the dual-density method enjoyed a smaller voxel size (128×128 matrix size in (85) and 0.076 cc voxels with 128×128 matrix size in (83)). Naturally, the dual-density method is expected to perform better with increased lipid image resolution, however at the cost of increased scan time. Since the dual-density idea constitutes an important part of the proposed methods, their performances are also expected to increase when even higher resolution lipid priors are available. The optimal selection of the high-resolution voxel size to balance the impact on total scan time and lipid artifact reduction remains an open problem.

In the current work, selection of the lipid mask was performed manually, with the guidance of the projection over lipid resonance frequencies. The brain mask was then assigned to be region remaining inside the lipid mask. A more elegant approach can involve a pilot structural scan acquired at the same resolution as the lipid image, which can be then segmented (e.g. using FreeSurfer (92)) to yield the skull and brain regions. A similar idea was also implemented in (88).

A similar approach that also restricts the space in which the metabolite signals reside is by Eslami and Jacob (93), where the spectrum at each voxel is parameterized as a sparse linear combination of spikes and polynomials to capture the metabolite and baseline components, respectively. Their elegant method is a holistic framework that performs field map compensation, noise reduction and lipid artifact reduction simultaneously. In particular, their lipid suppression performance was seen to be comparable with extrapolation methods (87). The methods proposed here involve no parametric signal modeling, but they simply minimize projection onto lipid spectra. Hence, it might be possible to combine Eslami and Jacob's method synergistically with the proposed schemes to further refine the metabolite spectra.

The L-curve analysis employed for selecting an optimal regularization parameter  $\lambda$  revealed that the operating points on the curve map virtually to the same point for a wide range of parameters (Fig. 4.1). In particular, the data consistency increases only 0.05 % and the regularization decreases only 3.8 % as  $\lambda$  increases from  $10^{-2}$  to  $10^2$ . Hence, if a slightly over-regularized reconstruction is acceptable, the selection of  $\lambda$  does not pose a problem as the reconstruction results are insensitive to its selection.

From a sequence design point of view, the 3-dimensional Cartesian undersampling pattern used in the *in vivo* dataset will not be feasible within the spiral CSI framework, as the samples were randomly removed in the Cartesian  $k$ -space of the *gridded* CSI data. Design of undersampled trajectories that will make the refined method feasible for *in vivo* acquisitions is under progress with initial results reported in (94).

Lipid suppression results obtained with the synthetic phantom demonstrates the feasibility of spiral undersampling. Relative to the conventional, low-resolution spiral reconstruction in Fig. 4.7d, the example spectra obtained with undersampled spirals in Fig. 4.7c exhibit substantially reduced lipid ringing artifacts in the vicinity of the lipid ring. Relative to the Nyquist rate spiral reconstruction, compressed sensing reconstruction with 10-fold accelerated spirals yielded comparable NAA maps and spectra. While the current work focused on undersampled spiral trajectory, other families of readout trajectories can be deployed in the proposed scheme, e.g. a trajectory that continues along the tangent of the spiral at the end of the low-resolution  $k$ -space (spiral+radial).

*In vivo* reconstructions at TE = 50 ms with the basic and refined methods exhibit successful artifact suppression in the cortical region (Figs. 4.5 and 4.6). Relative to the gold-standard reconstruction (20  $\text{avg}_{\text{high}}$ ,  $R_{\text{high}} = 1$ ) corresponding to a 33 min scan, the proposed methods yielded comparable NAA maps (Fig. 4.3) with substantial savings in the imaging time. While using the lipid-basis penalty at 0.56 cc voxel size (corresponding to a 10 min scan) gives effective lipid suppression, the presence of residual lipid artifacts is visible in the lipid and NAA maps (Figs. 4.2 and 4.3) and the cortical spectra (Figs. 4.5 and 4.6).

For additional validation, the reconstruction methods were also compared with a commercially available lipid suppression method, OVS. Taking the NAA maps obtained with OVS as ground truth, the four methods, namely, gold-standard (20  $\text{avg}_{\text{high}}$ ,  $R_{\text{high}} = 1$ ), basic (2  $\text{avg}_{\text{high}}$ ,  $R_{\text{high}} = 1$ ) refined (20  $\text{avg}_{\text{high}}$ ,  $R_{\text{high}} = 10$ ) and lipid-basis penalty, yielded similar fidelity where the gold-standard gave the smallest error (11.1%) and the lipid-basis penalty method had the largest mismatch (14.7%). Since the OVS method is obtained by exciting a  $9 \times 9 \text{ cm}^2$  box inside the brain surrounded by suppression bands to null the lipid signal, the comparison is limited to the interior of the brain where the lipid ringing artifacts are milder than the periphery of the cortex. It is seen that the spectra reconstructed with the lipid-basis method still demonstrate residual artifacts while the proposed methods are free of lipid ringing (Fig. 4.4). To compute the RMSEs relative to the NAA map obtained from the OVS acquisition, all methods were masked in  $k$ -space to match the

resolution of the OVS scan and the mean intensities of the NAA images were scaled to match the mean intensity of the OVS map.

Relative to the lipid-basis penalty method (88), the drawback of the proposed basic algorithm is the additional scan time required for collecting the peripheral  $k$ -space information. The refined method addresses this problem by aggressively undersampling the high  $k$ -space and exploiting the spatial and spectral sparsity of the lipid ring. While this entails an additional iterative reconstruction step for the FOCUSS (4) algorithm, the computational requirements of the refined method is not prohibitive for *in vivo* applications, taking only 11 min of processing time on a workstation.

The validity of the approximation that lipid and metabolite spectra are orthogonal is demonstrated in Fig. 4.8. All lipid spectra inside the lipid mask were selected from the 33 min, 20 average *in vivo* scan (578 lipid spectra in total), and all metabolite spectra were chosen from the *in vivo* OVS acquisition (521 metabolite spectra in total). For each lipid spectrum, the parallel and orthogonal components of each metabolite spectrum were computed. Based on this, the worst and best case situations were identified, where the ratio of energy in the parallel and orthogonal components were highest and lowest, respectively. Fig. 4.8a overplots the lipid and metabolite spectra in the best case scenario. Even though the NAA peak completely overlaps with the lipid signal in resonance frequency, the component of the metabolite spectra parallel to the lipid signal has almost no energy compared to the orthogonal component. Fig. 4.8c shows the worst case scenario for the lipid and metabolite spectra with the least degree of orthogonality. In this case, parallel and orthogonal components have comparable signal energy. When averaged over all lipid and metabolite spectra, the ratio of parallel component energy to orthogonal component energy was 15.6%, showing that the orthogonality assumption is a reasonable one in practice.

Limitations of this study include that,

- i. No B<sub>0</sub>-correction was employed in post-processing, but simply the data acquired at the scanner were used as input to the proposed lipid suppression methods. Therefore, more refined metabolite images can be obtained when local B<sub>0</sub> shifts are taken into account, e.g. Fig. 2 in (95) and Fig. 3 in (77).
- ii. The practical implementation of the dual-density method is considerably challenging, but this has been addressed adequately by previous investigators, e.g. (83,85). Similarly, practical realization of prospective undersampling with spiral readout is challenging.

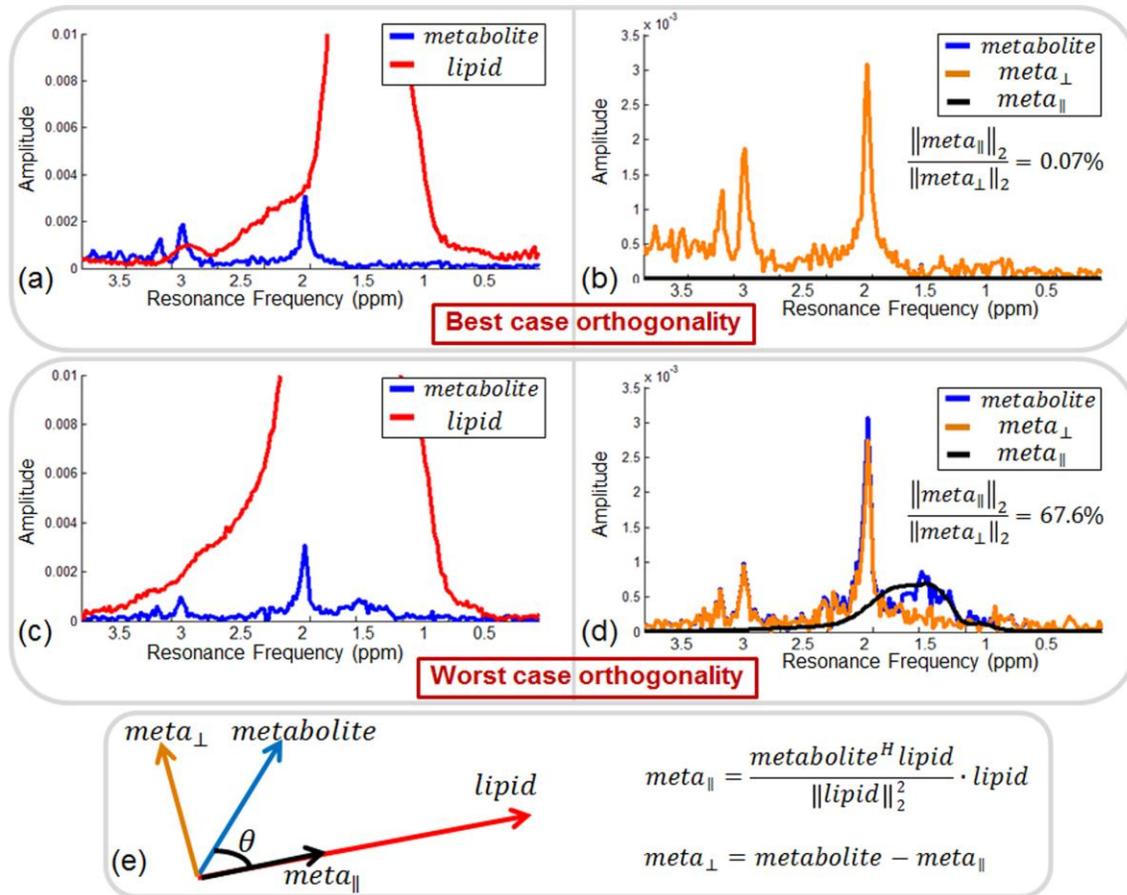


Fig. 4.8. Demonstration of approximate orthogonality between metabolite spectra obtained from *in vivo* OVS scan and lipid spectra from high resolution *in vivo* acquisition. In (a), the lipid and metabolite spectra with the highest orthogonality are plotted. In (b), the components of the metabolite spectrum that are orthogonal and parallel to the lipid spectrum for the best case in (a) are overlotted. The actual metabolite spectrum (in blue) is totally occluded by the orthogonal component (in orange). In (c), the lipid and metabolite spectra that are least orthogonal are depicted. In (d), the orthogonal and parallel components of the metabolite spectrum are overlotted for the worst case in (c). Panel (e) depicts the methodology used in computing the orthogonal and parallel metabolite components.

## 4.6 Conclusion

The proposed lipid suppression algorithms combine and extend two previously proposed approaches, dual-density sampling and lipid-basis orthogonality, with minimal increase on the total scan time by collecting only 2-average high-resolution data and aggressive undersampling ( $R = 10$ ) of high frequency  $k$ -space. Successful *in vivo* lipid-suppression performance was

demonstrated with artifact-free observation of metabolite spectra even in the peripheral cortical regions without any other means of lipid suppression during acquisition at TE = 50 ms.

## Chapter 5

### Accelerated Diffusion Spectrum Imaging

Diffusion Spectrum Imaging (DSI) offers detailed information on complex distributions of intravoxel fiber orientations at the expense of extremely long imaging times (~1 hour). Recent work by Menzel et al. (96) demonstrated successful recovery of diffusion probability density functions (pdfs) from sub-Nyquist sampled  $q$ -space by imposing sparsity constraints on the pdfs under wavelet and Total Variation (TV) transforms. As the performance of compressed sensing (CS) reconstruction depends strongly on the level of sparsity in the selected transform space, a dictionary specifically tailored for diffusion pdfs can yield higher fidelity results. This chapter presents the first application of adaptive dictionaries in DSI, whereby the scan time of whole brain DSI acquisition is reduced from 50 to 17 min while retaining high image quality. *In vivo* experiments were conducted with the 3T Connectome MRI. The RMSE of the reconstructed ‘missing’ diffusion images were calculated by comparing them to a gold standard dataset (obtained from acquiring 10 averages of diffusion images in these missing directions). The RMSE from the proposed reconstruction method is up to 2 times lower than that of Menzel et al.’s method, and is actually comparable to that of the fully-sampled 50 minute scan. Comparison of tractography solutions in 18 major white-matter pathways also indicated good agreement between the fully-sampled and 3-fold accelerated reconstructions. Further, it is demonstrated that a dictionary trained using pdfs from a single slice of a particular subject generalizes well to other slices from the same subject, as well as to slices from other subjects.

#### 5.1 Introduction

Diffusion weighted MR imaging is a widely used method to study white matter structures of the brain. Diffusion Tensor Imaging (DTI) is an established diffusion weighted imaging method, which models the diffusion as a univariate Gaussian distribution (97). One limitation of this model arises in the presence of fiber crossings, and this can be addressed by using a more involved imaging method (98,99). Diffusion Spectrum Imaging (DSI) results in magnitude representation of the full  $q$ -space and yields a complete description of the diffusion probability density function (pdf) (100,101). While DSI is capable of resolving complex distributions of

intravoxel fiber orientations, full  $q$ -space coverage comes at the expense of substantially long scan times (~1 hour).

Compressed sensing (CS) comprises algorithms that recover data from undersampled acquisitions by imposing sparsity or compressibility assumptions on the reconstructed images (6). In the domain of DSI, acceleration with CS was successfully demonstrated by Menzel et al. (102) by imposing wavelet and Total Variation (TV) penalties in the pdf space. Up to an undersampling factor of 4 in  $q$ -space, it was reported that essential diffusion properties such as orientation distribution function (odf), diffusion coefficient, and kurtosis were preserved (102). A recent study focused on the problem of finding the best wavelet basis to represent the diffusion pdf by comparing various wavelet transforms (103).

The performance of CS recovery depends on the level of sparsity of the signal in the selected transform domain, as well as the incoherence of the aliasing artifacts in the transform domain and the amount of acceleration in the sampling space (6). While prespecified transformations such as wavelets and spatial gradients yield sparse signal representation, tailoring the sparsifying transform based on the characteristics of the particular signal type may offer even sparser results. K-SVD is an algorithm that designs a dictionary that achieves maximally sparse representation of the input training data (104). The benefit of using data-driven, adaptive dictionaries trained with K-SVD was also demonstrated in CS reconstruction of structural MR imaging (105,106).

In this chapter, the K-SVD algorithm is employed to design a sparsifying transform that yields a signal representation with increased level of sparsity. Coupling this adaptive dictionary with the Focal Underdetermined System Solver (FOCUSS) algorithm (4), a parameter-free CS algorithm is obtained. With 3-fold undersampling of  $q$ -space in in vivo experiments, up to 2-fold reduced pdf reconstruction errors are demonstrated relative to the CS algorithm that uses wavelets and variational penalties by Menzel et al. (102). At higher acceleration factors of 5 and 9, up to 1.8-fold and 1.6-fold reduced errors are still obtained relative to wavelet and TV reconstruction *at the lower acceleration factor of 3*. For additional validation, the RMSE of the reconstructed ‘missing’ diffusion images were calculated by comparing them to a gold standard dataset obtained with 10 averages. In this case, dictionary-based reconstructions were seen to be comparable to the fully-sampled 1 average data. For further validation, average Fractional Anisotropy (FA) and tract volume metrics obtained from 18 major white-matter pathways were compared between the fully-sampled and 3-fold accelerated dictionary reconstructions to yield good agreement. Additionally, it is shown that a dictionary trained on data from a particular subject generalizes well to reconstruction of another subject’s data, still yielding up to 2-fold reduced reconstruction errors

relative to using prespecified transforms. Hence, application of the proposed method might reduce a typical 50-minute DSI scan to 17 minutes (upon  $3\times$  acceleration) while retaining high image quality. In addition, using a simple  $\ell_1$ -norm penalty in the pdf space with the FOCUSS algorithm is also investigated, and it is shown that this approach gives comparable results to the more involved wavelet- and TV-based reconstruction by Menzel et al. (102), while being computationally more efficient.

## 5.2 Theory

### 5.2.1 CS Recovery with Prespecified Transforms

Letting  $\mathbf{p} \in \mathbb{C}^N$  represent the 3-dimensional diffusion pdf at a particular voxel as a column vector, and  $\mathbf{q} \in \mathbb{C}^M$  denote the corresponding undersampled  $q$ -space information, CS recovery with wavelet and TV penalties aim to solve the convex optimization problem at a single voxel,

$$\min_{\mathbf{p}} \|\mathbf{F}_\Omega \mathbf{p} - \mathbf{q}\|_2^2 + \alpha \cdot \|\Psi \mathbf{p}\|_1 + \beta \cdot \text{TV}(\mathbf{p}) \quad (5.1)$$

where  $\mathbf{F}_\Omega$  is the undersampled Fourier transform operator,  $\Psi$  is a wavelet transform operator,  $\text{TV}(\cdot)$  is the Total Variation penalty, and  $\alpha$  and  $\beta$  are regularization parameters that need to be determined. CS recovery is applied on a voxel-by-voxel basis to reconstruct all brain voxels.

### 5.2.2 Training an Adaptive Transform with K-SVD

Given an ensemble  $\mathbf{P} \in \mathbb{C}^{N \times L}$  formed by concatenating  $L$  example pdfs  $\{\mathbf{p}_i\}_{i=1}^L$  collected from a training dataset as column vectors, the K-SVD algorithm (104) aims to find the best possible dictionary for the sparse representation of this dataset by solving,

$$\min_{\mathbf{P}, \mathbf{D}} \sum_{i=1}^L \|\mathbf{x}_i\|_0 \quad \text{subject to} \quad \|\mathbf{P} - \mathbf{D}\mathbf{X}\|_F^2 \leq \varepsilon \quad (5.2)$$

where  $\mathbf{X}$  is the matrix that contains the transform coefficient vectors  $\{\mathbf{x}_i\}_{i=1}^L$  as its columns,  $\mathbf{D}$  is the adaptive dictionary,  $\varepsilon$  is a fixed constant that adjusts the data fidelity, and  $\|\cdot\|_F$  is the Frobenius norm. The K-SVD works iteratively, first by fixing  $\mathbf{D}$  and finding an optimally sparse  $\mathbf{X}$  using orthogonal matching pursuit, then by updating each column of  $\mathbf{D}$  and the transform coefficients corresponding to this column to increase data consistency.

### 5.2.3 CS Recovery with an Adaptive Transform using FOCUSS

The FOCUSS algorithm aims to find a sparse solution to the underdetermined linear system  $\mathbf{F}_\Omega \mathbf{D} \mathbf{x} = \mathbf{q}$ , where  $\mathbf{x}$  is the vector of transform coefficients in the transform space defined by the dictionary  $\mathbf{D}$  using the following iterations,

For iteration number  $t = 1, \dots, T$ ,

$$\mathbf{W}_{j,j}^t = \text{diag}(|x_j^t|^{1/2}) \quad (5.3)$$

$$\mathbf{s}^t = \text{argmin}_s \|\mathbf{s}\|_2^2 \text{ such that } \mathbf{F}_\Omega \mathbf{D} \mathbf{W}^t \mathbf{s} = \mathbf{q} \quad (5.4)$$

$$\mathbf{x}^{t+1} = \mathbf{W}^t \mathbf{s}^t \quad (5.5)$$

Here,  $\mathbf{W}^t$  is a diagonal weighting matrix whose  $j^{\text{th}}$  diagonal entry is denoted as  $\mathbf{W}_{j,j}^t$ ,  $\mathbf{x}^t$  is the estimate of transform coefficients at iteration  $t$  whose  $j^{\text{th}}$  entry is  $x_j^t$ . The final reconstruction in diffusion pdf space is obtained via the mapping  $\mathbf{p} = \mathbf{D} \mathbf{x}^{T+1}$ .

It is also possible to impose sparsity-inducing  $\ell_1$  penalty directly on the pdf coefficients by taking  $\mathbf{D}$  to be the identity matrix  $\mathbf{I}$ .

## 5.3 Methods

Diffusion EPI acquisitions were obtained from three healthy volunteers (subjects A, B and C) using a novel 3T system (Magnetom Skyra Connectom, Siemens Healthcare, Erlangen, Germany) equipped with the AS302 ‘‘CONNECTOM’’ gradient with  $G_{\text{max}} = 300$  mT/m (here  $G_{\text{max}} = 200$  mT/m was used) and Slew = 200 T/m/s. A custom-built 64-channel RF head array (107) was used for reception with imaging parameters of 2.3 mm isotropic voxel size, FOV = 220×220×130, matrix size = 96×96×57,  $b_{\text{max}} = 8000$  s/mm<sup>2</sup>, 514 directions full sphere  $q$ -space sampling (corners of  $q$ -space were zero-padded since they were not sampled) organized in a Cartesian grid with interspersed  $b=0$  images every 20 TRs (for motion correction, 25  $b=0$  images in total), in-plane acceleration = 2× (using GRAPPA algorithm), TR/TE = 5.4 s / 60 ms, total imaging time ~50 min. In addition, at 5  $q$ -space points ([1,1,0], [0,2,−1], [0,0,3], [0,4,0], and [5,0,0]) residing on 5 different shells, 10 averages were collected for noise quantification. The corresponding  $b$ -values for these 5 points were 640, 1600, 2880, 5120, and 8000 s/mm<sup>2</sup>. Eddy current related distortions were corrected using the reversed polarity method (108). Motion correction (using interspersed  $b=0$ ) was performed using FLIRT (109) with sinc interpolation.

Variable-density undersampling (using a power-law density function (6)) with  $R = 3$  acceleration was applied in  $q$ -space on a  $12 \times 12 \times 12$  grid. Three different adaptive dictionaries were trained with data from slice 30 of subjects A, B and C. Reconstruction experiments were applied on test slices that are different than the training slices. In particular, two reconstruction experiments were performed:

- i. First, voxels in slice 40 of subject A were retrospectively undersampled in  $q$ -space, and reconstructed using 5 different methods: wavelet+TV method of Menzel et al. (102),  $\ell_1$ -regularized FOCUSS, and Dictionary-FOCUSS with the three dictionaries trained on three different subjects.
- ii. Second, voxels in slice 25 of subject B were retrospectively undersampled with the same  $R = 3$  sampling pattern, and again reconstructed with wavelet+TV,  $\ell_1$ -FOCUSS, and the three dictionaries trained on three different subjects.

Slice 30 was selected for training and slices 25 and 40 were chosen for test based on their anatomical location, so that the test slices would reside on lower and upper parts of the brain, while the training slice was one of the middle slices. For Menzel et al.'s method, Haar wavelets in MATLAB's wavelet toolbox were used. The regularization parameters  $\alpha$  and  $\beta$  in Eq. 5.1 were chosen by parameter sweeping with values  $\{10^{-4}, 3 \cdot 10^{-4}, 10^{-3}, 3 \cdot 10^{-3}\}$  to minimize the reconstruction error of 100 randomly selected voxels in slice 40 of subject A. The optimal regularization parameters were found to be  $\alpha = 3 \cdot 10^{-4}$  for wavelet and  $\beta = 10^{-4}$  for the TV term. By taking the fully-sampled data as ground-truth, the fidelity of the five methods were compared using root-mean-square error (RMSE) normalized by the  $\ell_2$ -norm of ground-truth as the error metric both in pdf domain and  $q$ -space.

Since the fully-sampled data are corrupted by noise, computing RMSEs relative to them will include contributions from both reconstruction errors and additive noise. To address this, the additional 10 average data acquired at the selected 5  $q$ -space points were used. As a single average full-brain DSI scan takes  $\sim 50$  min, it was not practical to collect 10 averages for all of the undersampled  $q$ -space points. As such, both error metrics are utilized for performance quantification, namely: the RMSE relative to one average fully-sampled dataset and the RMSE relative to gold standard data for 5  $q$ -space points.

To compare the fully-sampled and 3-fold accelerated Dictionary-FOCUSS reconstructions in terms of tractography solutions, streamline deterministic DSI tractography on the two datasets

was performed in trackvis (<http://trackvis.org>) and 18 white-matter pathways were labeled. The labeling was performed following the protocol described in (110), where two regions of interest (ROIs) are drawn for each pathway in parts of the anatomy that the pathway is known to traverse. To eliminate variability due to manual labeling in the two data sets and make the comparison as unbiased as possible, the ROIs used here were not drawn manually on the fully-sampled and 3-fold accelerated data. Instead, the ROIs were obtained from a different data set of 33 healthy subjects, where the same pathways had been previously labeled (111). The respective ROIs from the 33 subjects were averaged in MNI space (112) and the average ROIs were mapped to the native space of the fully-sampled and  $R = 3$  datasets using affine registration. In each data set the tractography streamlines going through the respective ROIs were isolated to identify the 18 pathways.

## 5.4 Results

Fig. 5.1 depicts the error of the different reconstruction methods in the pdf domain for each voxel in slice 40 of subject A. At  $R = 3$  acceleration, reconstruction error of Menzel et al.'s method averaged over brain voxels in the slice was 15.8%, while the error was 15.0% for  $\ell_1$ -regularized FOCUSS. Adaptive dictionary trained on subject A yielded 7.8% error. Similarly, reconstruction with dictionaries trained on pdfs of the other subjects B and C returned 7.8% and 8.2% RMSE, respectively. At  $R = 5$ , Dictionary-FOCUSS returned 8.9%, 8.9% and 9.3% error with training on subjects A, B and C, respectively. At  $R = 9$  dictionary reconstruction with training on subjects A, B and C returned 10.0%, 10.0% and 10.4% RMSE.

In Fig. 5.2, reconstruction errors at  $R = 3$  on slice 25 of subject B are presented. In this case, Menzel et al.'s method yielded 17.5% average RMSE, and  $\ell_1$ -FOCUSS had 17.3% error. Dictionary trained on slice 40 of subject A returned 11.4% RMSE, while adaptive transforms trained on subjects B and C had 11.4% and 11.8% error, respectively. At a higher acceleration factor of  $R = 5$ , Dictionary-FOCUSS with training on subjects A, B and C returned 13.1%, 13.3% and 13.5% error. At  $R = 9$  dictionary reconstruction with training on subjects A, B and C yielded 14.2%, 14.2% and 14.4% RMSE, respectively.

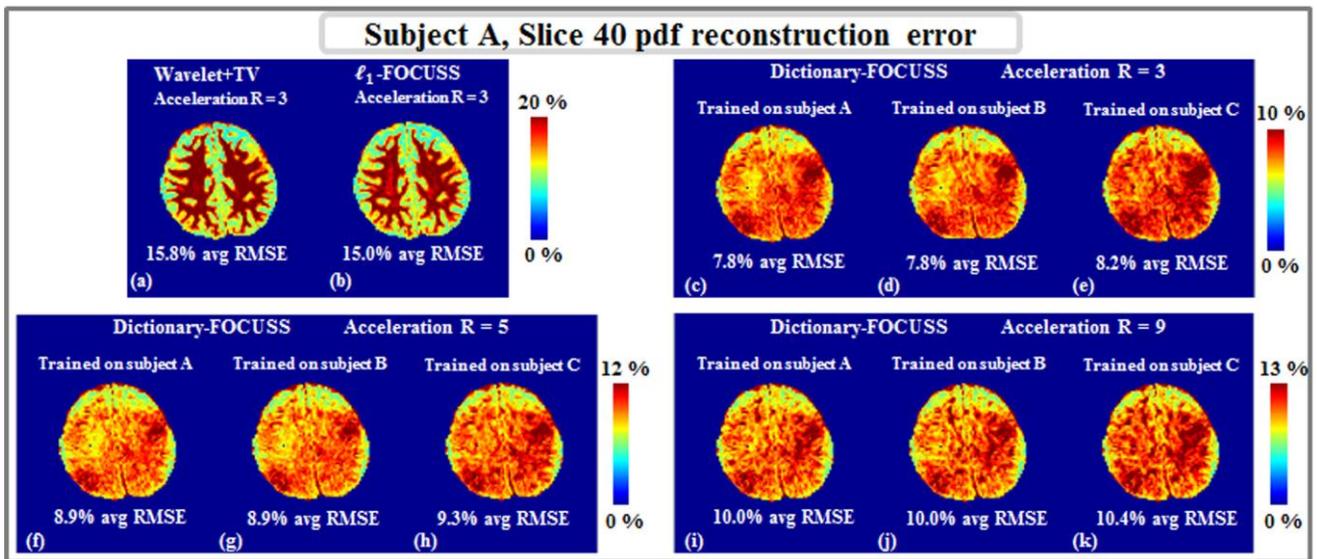


Fig. 5.1. RMSE at each voxel in slice 40 of subject A upon  $R = 3$  acceleration and reconstruction with Menzel et al.'s method (a),  $\ell_1$ -FOCUSS (b), Dictionary-FOCUSS trained on subjects A (c), B (d), and C (e). Dictionary-FOCUSS errors in (f), (g) and (h) are obtained at higher acceleration factor of  $R = 5$  with training on subjects A, B and C, respectively. Results for the reconstructions at  $R = 9$  are given in (i), (j) and (k).

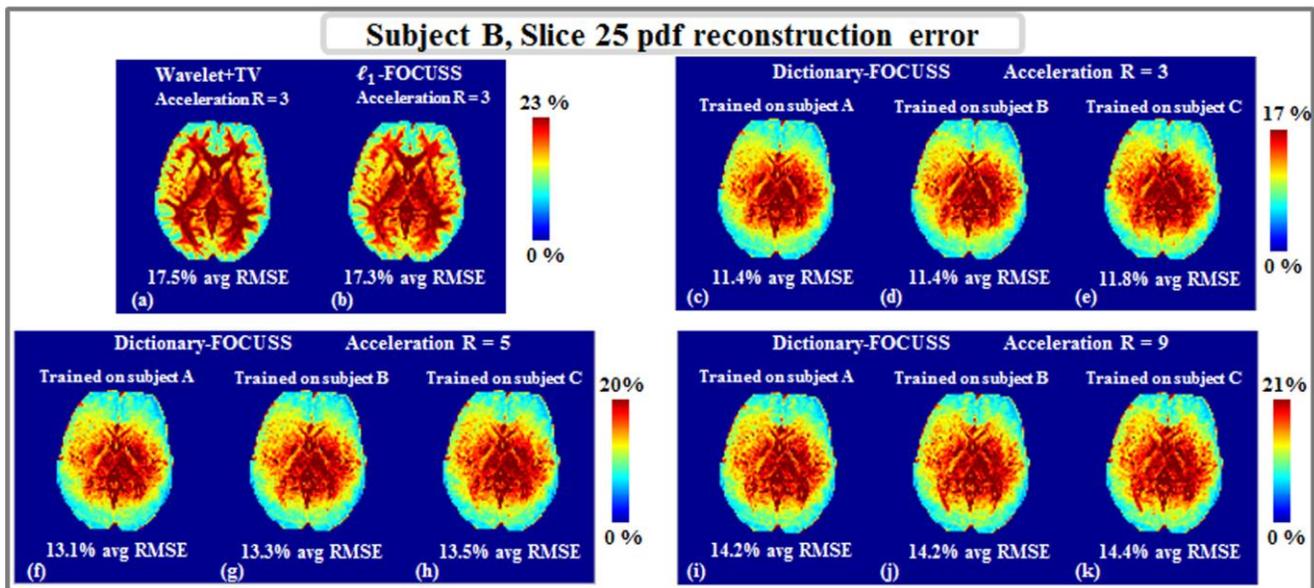


Fig. 5.2. RMSE at each voxel in slice 25 of subject B upon  $R = 3$  acceleration and reconstruction with Menzel et al.'s method (a),  $\ell_1$ -FOCUSS (b), Dictionary-FOCUSS trained on subjects A (c), B (d), and C (e). Dictionary-FOCUSS errors in (f), (g) and (h) are obtained at higher acceleration factor of  $R = 5$  with training on subjects A, B and C, respectively. Results for the reconstructions at  $R = 9$  are given in (i), (j) and (k).

Fig. 5.3 presents RMSEs obtained on various slices of subject A using Dictionary- and  $\ell_1$ -FOCUSS. Error bars that show the variation of the reconstruction errors are also included. RMSE maps on four selected slices are plotted for comparison. The same analysis is carried out on various slices of subject B, and the results are depicted in Fig 5.4.

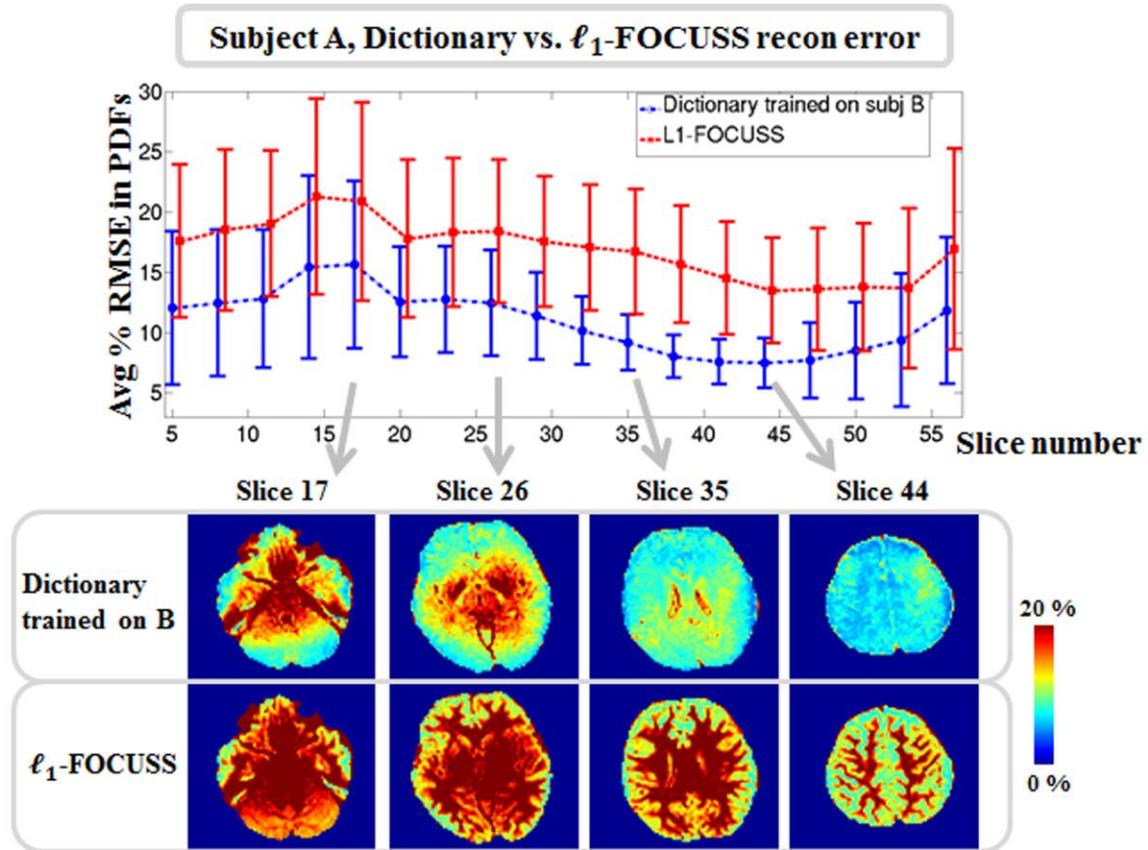


Fig. 5.3. Mean and standard deviation of RMSEs computed on various slices of subject A using  $\ell_1$ - and Dictionary-FOCUSS trained on subject B. Lower panel depicts RMSE maps for four selected slices.

Reconstruction errors in  $q$ -space images of subject A obtained with Wavelet+TV,  $\ell_1$ -FOCUSS and Dictionary-FOCUSS trained on the three subjects for the undersampled  $q$ -space directions are plotted in Fig. 5.5. For two particular diffusion directions  $[0,4,0]$  and  $[5,0,0]$ ,  $q$ -space reconstructions obtained with the three methods are also presented. In Fig. 5.5a,  $q$ -space images obtained with Wavelet+TV,  $\ell_1$ -FOCUSS and Dictionary-FOCUSS (with training on subject B) are compared with the 10 average fully-sampled image at  $[5,0,0]$ . Fig. 5.5b presents the error images relative to the 10 average data for the three methods. Figs. 5.5c and d depict the same analysis at direction  $[0,4,0]$ .

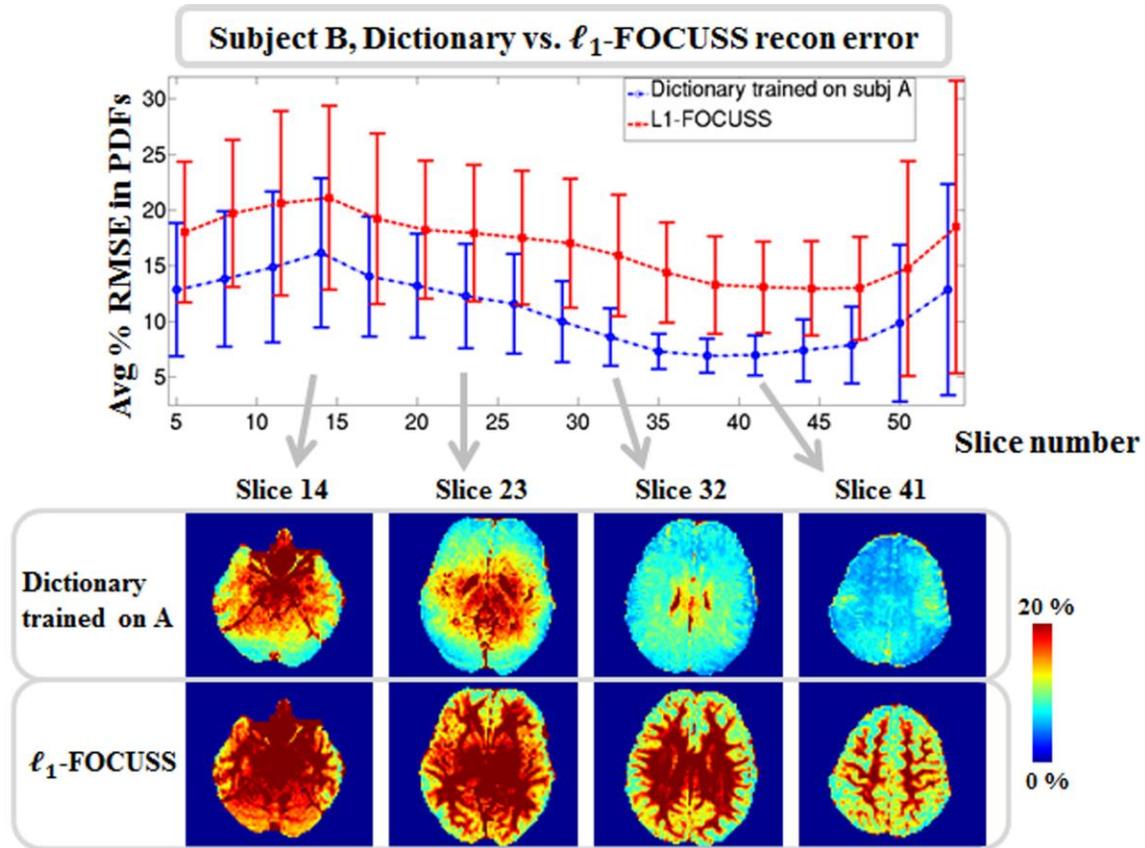


Fig. 5.4. Mean and standard deviation of RMSEs computed on various slices of subject B using  $\ell_1$ - and Dictionary-FOCUSS trained on subject A. Lower panel depicts RMSE maps for four selected slices.

In an attempt to quantify the noise in  $q$ -space and separate it from CS reconstruction error, the 10 average data acquired at 5  $q$ -space directions were taken as ground truth and the RMSEs were computed relative to them. Fig. 5.6 shows the error plots for the 1 average fully sampled data, Wavelet+TV,  $\ell_1$ -FOCUSS, and Dictionary-FOCUSS reconstructions relative to the 10 average data for slices from subjects A and B.

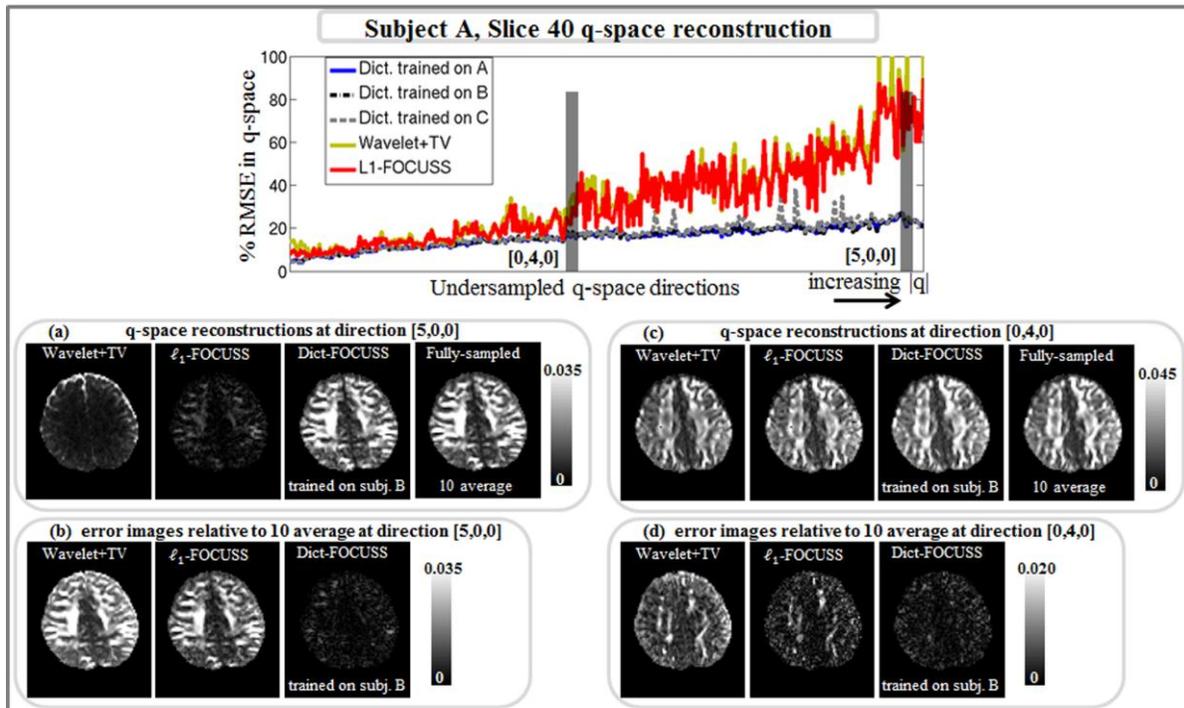


Fig. 5.5. Top panel shows RMSEs in ‘missing’  $q$ -space directions that are estimated with Wavelet+TV,  $\ell_1$ -FOCUSS and Dictionary-FOCUSS with training on subjects A, B and C at  $R=3$ .  $q$ -space images at directions [5,0,0] (a) and [0,4,0] (c) are depicted for comparison of the reconstruction methods. In panels (b) and (d), reconstruction errors of Wavelet+TV,  $\ell_1$ -FOCUSS and dictionary reconstructions relative to the 10 average fully-sampled image at directions [5,0,0] and [0,4,0] are given.

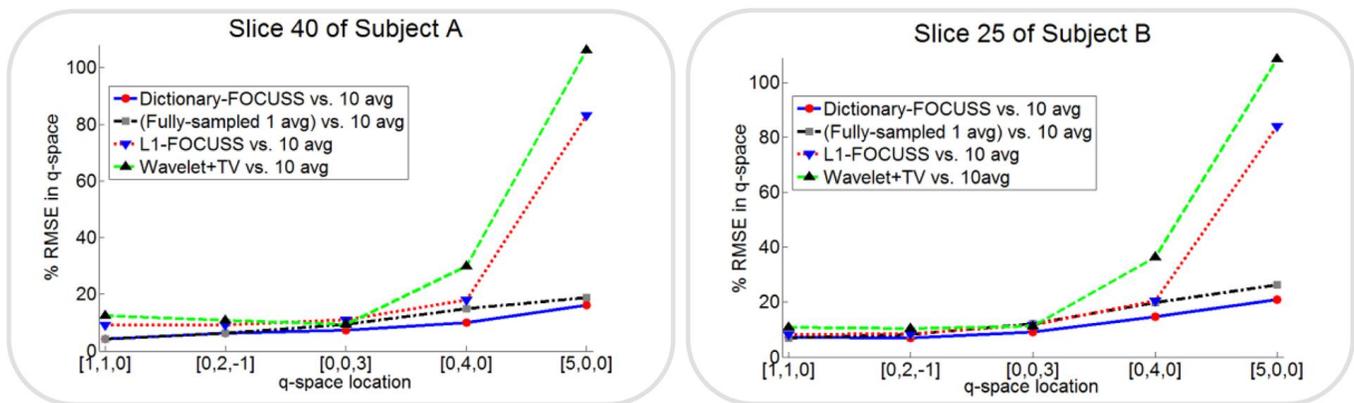


Fig. 5.6. Panel on top depicts RMSEs of Wavelet+TV,  $\ell_1$ -FOCUSS and Dictionary-FOCUSS at  $R = 3$  and fully-sampled 1 average data computed in 5  $q$ -space locations relative to the 10 average data for subject A. Panel on the bottom shows the same comparison for the slice belonging to subject B.

Fig. 5.7a and b show tractography results of subject A for the labeled white-matter pathways in the fully-sampled and 3-fold accelerated Dictionary-FOCUSS reconstructions. Fig. 5.7c and d show plots of the average FA and volume of each pathway for the 18 white-matter pathways, as calculated from the two reconstructions.

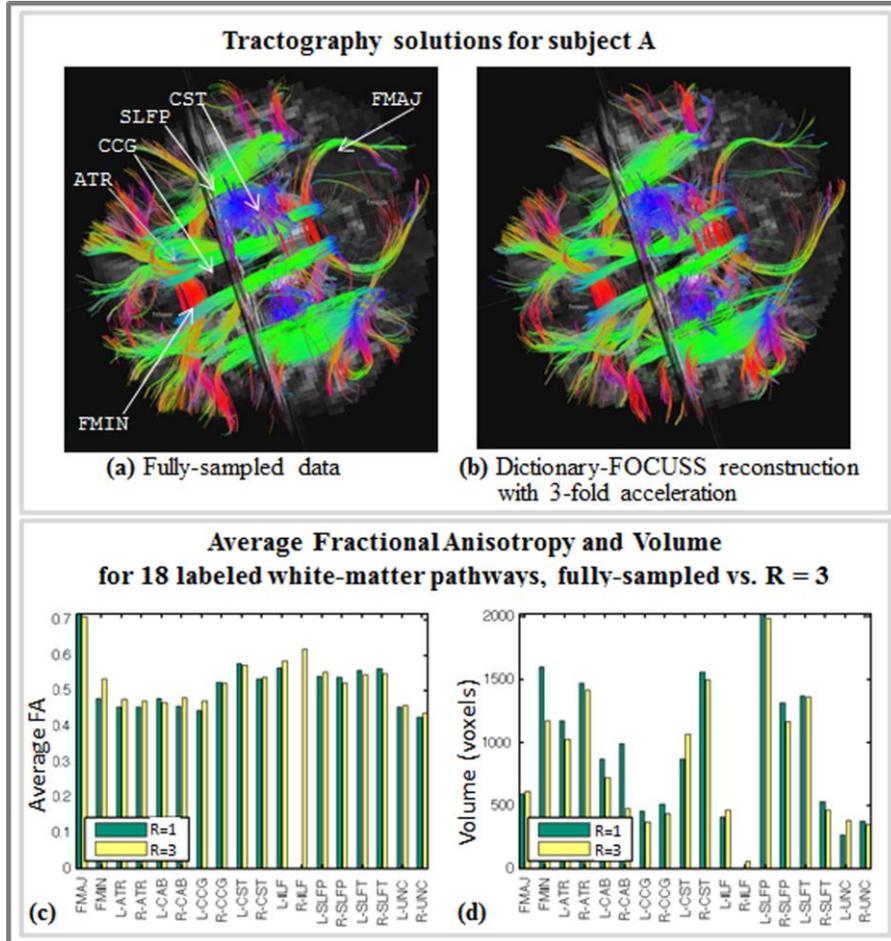


Fig. 5.7. Axial view of white-matter pathways labeled from streamline DSI tractography in fully-sampled data (a) and Dictionary-FOCUSS reconstruction at  $R = 3$  (b). The following are visible in this view: corpus callosum - forceps minor (FMIN), corpus callosum - forceps major (FMAJ), anterior thalamic radiations (ATR), cingulum - cingulate gyrus bundle (CCG), superior longitudinal fasciculus - parietal bundle (SLFP), and the superior endings of the corticospinal tract (CST). Average FA (c) and volume in number of voxels (d) for each of the 18 labeled pathways, as obtained from the fully-sampled ( $R=1$ , green) and Dictionary-FOCUSS reconstructed with 3-fold undersampling ( $R=3$ , yellow) datasets belonging to subject A. Intra-hemispheric pathways are indicated by “L-” (left) or “R-” (right). The pathways are: corpus callosum - forceps major (FMAJ), corpus callosum - forceps minor (FMIN), anterior thalamic radiation (ATR), cingulum - angular (infracallosal) bundle (CAB), cingulum - cingulate gyrus (supracallosal) bundle (CCG), corticospinal tract (CST), inferior longitudinal fasciculus (ILF), superior longitudinal fasciculus - parietal bundle (SLFP), superior longitudinal fasciculus - temporal bundle (SLFT), uncinate fasciculus (UNC).

## 5.5 Discussion

This chapter presented the first application of adaptive transforms to voxel-by-voxel CS reconstruction of undersampled  $q$ -space data. Relative to reconstruction with prespecified transforms, i.e. wavelet and TV, the proposed algorithm has up to 2 times reduced error in the pdf domain at the same acceleration factor ( $R = 3$ ), while requiring no regularization parameter tuning. When the undersampling ratio was increased to  $R = 5$  and even up to  $R = 9$ , the proposed method still demonstrated substantial improvement relative to using prespecified transforms at a lower acceleration factor of  $R = 3$  (Figs. 5.1 and 5.2). As demonstrated, a dictionary trained with pdfs from a single slice of a particular subject generalizes to other slices of the same subject, as well as to different subjects. However, further tests are needed to see if dictionaries can generalize across healthy and patient populations, or across age groups.

Since the acquired 1 average DSI data is corrupted by noise (especially in the outer shells), it is desired to obtain noise-free data for more reliable computation of CS reconstruction errors. Because even the 1 average full-shell acquisition takes  $\sim 50$  min, it is practically not possible to collect multiple-average data at all  $q$ -space points. To address this, one representative  $q$ -space sample at each shell was collected with 10 averages to serve as “(approximately) noise-free” data. When the noise-free data were taken to be ground-truth, the dictionary reconstruction with 3-fold undersampling was comparable to the fully-sampled 1 average data for both subjects (Fig. 5.6).

RMSE in Fig. 5.2 was overall higher than in Fig. 5.1. A possible explanation is the inherently lower signal-noise-ratio (SNR) in the lower axial slice, particularly in the center area of the brain which is further away from the receive coils. In particular, the error is higher in the central region of the image where the SNR is expected to be lowest. Since the noisy 1 average datasets were taken to be the reference in RMSE computations in Figs. 5.1 and 2, the errors are expected to be influenced by noise in these lower SNR regions. As seen in Figs. 5.1 and 2, Wavelet+TV and  $\ell_1$ -FOCUSS tend to yield larger error in the white matter, where the information is more critical for fiber tracking. Error maps from the dictionary reconstruction is more homogenous across white and gray matter, especially results on Fig. 5.2 resemble SNR maps where the middle of the brain is further away from the receive coils. As Fig. 5.6 demonstrates, dictionary reconstruction has a certain degree of denoising property, since it yields lower RMSE than the 1 average data relative to the 10 average data. This might be one possible explanation why the RMSE is relatively higher in the middle of the brain, which should be explored in future investigation.

As seen in Fig. 5.5, wavelet and TV penalized reconstruction and  $\ell_1$ -FOCUSS yield especially poor quality results in estimating the high  $q$ -space samples. In particular, as depicted in Fig. 5.5a and b, these CS methods tend to underestimate the high  $q$ -space content. However, this is not a simple scaling problem, as they yield either flat (Wavelet+TV) or grainy ( $\ell_1$ -FOCUSS) results. Because Wavelet+TV reconstruction imposes piece-wise smoothness assumption in compressed sensing reconstruction, it leads to loss of high frequency content. In the context of DSI, this corresponds to attenuated high  $q$ -space information (flat, underestimated outer shell).  $\ell_1$ -penalized reconstruction encourages small number of non-zero coefficients, which is seen to be insufficient to model the diffusion pdfs. This also leads to underestimated high  $q$ -space content, but since there is no smoothness constraint in pdf domain, the reconstructed  $q$ -space is not flat. The RMSE plot in Fig. 5.5 also demonstrates that Wavelet+TV and  $\ell_1$ -FOCUSS results are comparable to the adaptive reconstruction at lower  $q$ -space (Fig. 5.5c and d), and the difference becomes more pronounced as  $|q|$  increases.

Visual inspection of the tractography solutions from the fully sampled and 3-fold accelerated Dictionary-FOCUSS data sets (Fig. 5.7a and b) showed that the white-matter pathways reconstructed from the two acquisitions were very similar. When comparing average FA over each pathway, as calculated from the two reconstructions, there are two potential sources of differences: the tractography streamlines could be different, visiting different voxels in the brain for each data set, and/or the tensor, from which the FA value is calculated, could be different at same voxel for each data set. However, good agreement was found between the average FA values in the fully-sampled and 3-fold accelerated reconstructions (Fig. 5.7c and d). Some differences are to be expected in weaker pathways that only consist of very few streamlines and thus are more sensitive to noise and have lower test-retest reliability than the stronger pathways. This was the case particularly for the right inferior longitudinal fasciculus (R-ILF), which did not have any streamlines in the fully-sampled data set (Fig. 5.7d). Therefore it was not possible to extract an average FA value for the R-ILF from the fully sampled data. Apart from this pathway, the mean difference between the average FA values in the fully sampled and 3-fold accelerated data, as a percentage of the value in the fully sampled data, was 3%. For the volume estimates, the mean error was 16%. It is possible that even more stable FA and volume measurements could be obtained by manual labeling of the paths directly on each data set, instead of using the average ROIs. This is because the averaging of ROIs in MNI space is susceptible to misregistration errors, leading to average ROIs that are typically much larger than the individual ROIs than a rater would draw directly on the images. Thus the bundles that we obtained with the average ROIs are more likely to contain stray streamlines that would be eliminated in a careful individual manual

labeling, leading to less noisy volume and FA estimates. However, the average ROIs were used here to avoid introducing variability due to manual labeling.

In a previous study the intra-rater and inter-rater reliability of the manual labeling procedure was evaluated by performing manual labeling several times on the same data set. It was found that the average distance between pathways labeled by the same and different raters to be, respectively, in the order of 1 voxel and 2 voxels (111). In the present study, it is found that the distance between the pathways obtained from the fully-sampled and  $R = 3$  data sets were comparable (median distance: 2.37mm, mean distance: 2.74mm with acquisition voxel size of 2.3mm isotropic). Further investigation with test-retest scans is warranted to determine how the differences between the fully sampled and 3-fold undersampled results compare to the test-retest reliability of each type of reconstruction.

In this study, fully-sampled  $q$ -space data were collected for comparison with the CS reconstruction methods. With the fully-sampled dataset, it was simple to apply the reverse polarity approach (108) to get good eddy current correction. It should be noted that in a real random undersampling case where reverse pairs are not present, such eddy current correction method will not be applicable. However, various approaches exist in performing eddy current correction, such as linearly fitting the eddy-current distortions parameters (translation, scaling, shearing) using the available data, and then estimating the transformation for any given  $q$ -space data.

In the current implementation, *per voxel* processing time of  $\ell_1$ -FOCUSS was 0.6 seconds, while this was 12 seconds for Dictionary-FOCUSS and 27 seconds for Wavelet+TV method on a workstation with 12GB memory and 6 processors. Hence, full-brain reconstruction using the  $\ell_1$ -FOCUSS algorithm would still take *days*. Because each voxel can be processed independently, parallel implementation is likely to be a significant source of performance gain. Dictionary training step (for subject A, using 3200 voxels inside the brain mask from a single slice) took 12 minutes. An additional research direction is to evaluate the change in reconstruction quality when multiple slices are used for training. Increased processing times due to employing a larger dictionary may become a practical concern in this case.

The proposed CS acquisition/reconstruction can be combined with other techniques to further reduce the acquisition time. In particular, combining the proposed method with the Blipped-CAIPI Simultaneous MultiSlice (SMS) acquisition (113) could reduce a 50 minute DSI scan to 5.5 minutes upon 9-fold acceleration ( $3 \times 3$  CS-SMS).

## 5.6 Fast DSI Reconstruction with Trained Dictionaries

Wavelet and TV reconstruction (96) and the dictionary-based compressed sensing method introduced in this chapter operate iteratively, therefore they require processing times on the order of *hours* per imaging slice using Matlab running on a workstation. This section presents two dictionary-based reconstruction techniques that use analytical solutions, and are 3 orders of magnitude faster than the Dictionary-FOCUSS approach. The first method also employs a dictionary trained with the K-SVD algorithm, but instead of using iterative CS reconstruction, Tikhonov regularization is applied on the dictionary transform coefficients. This admits a closed-form expression, which is shown to be equivalent to the regularized pseudoinverse solution. The second proposal is to apply Principal Component Analysis (PCA) on training data to derive a lower dimensional representation of diffusion pdfs. This way, fewer PCA coefficients are required to represent individual pdfs, effectively reducing the acceleration factor of the undersampled acquisition. Both methods require a single matrix-vector multiplication per voxel, hence attaining 3-orders of magnitude speed up in computation relative to iterative CS algorithms. Computation times on the order of *seconds per slice* in Matlab are reported for the methods proposed in this section, and it is shown that the reconstruction qualities are comparable to that of Dictionary-FOCUSS on in vivo datasets. In particular, the proposed methods yield up to 2 times less reconstruction error relative to the Wavelet+TV method at the acceleration factor of 3, and similar results to those of Dictionary-FOCUSS algorithm. Even when the acceleration factor is increased to 9, the proposed methods have up to 1.5 times less error compared to the wavelet and TV results obtained at the *lower acceleration factor of 3*.

### 5.6.1 Theory

The proposed algorithms with closed-form solutions are detailed in the following.

#### i. **Proposal I: Dictionary-based Reconstruction with Regularized Pseudoinverse**

Given a dictionary  $\mathbf{D}$  trained with the K-SVD algorithm, Tikhonov regularized reconstruction of dictionary coefficients at a particular voxel  $\mathbf{x}$  are found by solving,

$$\min_{\mathbf{x}} \|\mathbf{F}_{\Omega} \mathbf{D} \mathbf{x} - \mathbf{q}\|_2^2 + \lambda \cdot \|\mathbf{x}\|_2^2 \quad (5.6)$$

Since this objective function is strictly convex, the unique global optimizer is found by setting the gradient equal to zero,

$$\tilde{\mathbf{x}} = (\mathbf{D}^H \mathbf{F}_\Omega^H \mathbf{F}_\Omega \mathbf{D} + \lambda \cdot \mathbf{I})^{-1} \mathbf{D}^H \mathbf{F}_\Omega^H \mathbf{q} \quad (5.7)$$

Alternatively, we can relate Eq. 5.7 to the singular values of the observation matrix  $\mathbf{F}_\Omega \mathbf{D}$  by letting  $\mathbf{F}_\Omega \mathbf{D} = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H$ ,

$$\begin{aligned} \tilde{\mathbf{x}} &= ((\mathbf{U} \mathbf{\Sigma} \mathbf{V}^H)^H \mathbf{U} \mathbf{\Sigma} \mathbf{V}^H + \lambda \cdot \mathbf{I})^{-1} (\mathbf{U} \mathbf{\Sigma} \mathbf{V}^H)^H \mathbf{q} \\ &= (\mathbf{V} (\mathbf{\Sigma}^H \mathbf{\Sigma} + \lambda \cdot \mathbf{I}) \mathbf{V}^H)^{-1} \mathbf{V} \mathbf{\Sigma}^H \mathbf{U}^H \mathbf{q} \\ &= \mathbf{V} (\mathbf{\Sigma}^H \mathbf{\Sigma} + \lambda \cdot \mathbf{I})^{-1} \mathbf{\Sigma}^H \mathbf{U}^H \mathbf{q} \\ &= \mathbf{V} \mathbf{\Sigma}^+ \mathbf{U}^H \mathbf{q} \end{aligned} \quad (5.8)$$

Hence, the Tikhonov regularized solution can be found by applying singular value decomposition (SVD) to the observation matrix  $\mathbf{F}_\Omega \mathbf{D}$  and modifying  $i^{\text{th}}$  singular value due to  $\sigma_i^+ = \sigma_i / (\sigma_i^2 + \lambda)$ . Defining  $\mathbf{\Sigma}^+ = (\mathbf{\Sigma}^H \mathbf{\Sigma} + \lambda \cdot \mathbf{I})^{-1} \mathbf{\Sigma}^H$  to be the diagonal matrix with entries  $\sigma_i^+$ , the solution matrix  $\mathbf{V} \mathbf{\Sigma}^+ \mathbf{U}^H$  in the last line of Eq.8 needs to be computed only once. The regularization parameter  $\lambda$  can be selected to optimize the reconstruction performance on the training dataset that was used to generate the dictionary  $\mathbf{D}$ . This point is addressed in more detail under the Methods section. The result in pdf space is finally computed as  $\widetilde{\mathbf{pdf}} = \mathbf{D} \tilde{\mathbf{x}}$ . Regularized pseudoinverse reconstruction is denoted as PINV in the remainder of the text.

## ii. Proposal II: Reconstruction with Principal Component Analysis

PCA is a technique that seeks the best approximation of a given set of data points using a linear combination of a set of vectors which retain maximum variance along their directions. PCA starts by subtracting the mean from the data points, which becomes the virtual origin of the new coordinate system (114). Again beginning with a collection of pdfs  $\mathbf{P} \in \mathbb{C}^{N \times L}$ , whose  $i^{\text{th}}$  column is a training pdf  $\mathbf{p}_i \in \mathbb{C}^N$ , we obtain a modified matrix  $\mathbf{Z} \in \mathbb{C}^{N \times L}$  by subtracting the mean pdf from each column,

$$\mathbf{z}_i = \mathbf{p}_i - \frac{1}{L} \cdot \sum_{j=1}^L \mathbf{p}_j = \mathbf{p}_i - \mathbf{p}_{mean} \quad (5.9)$$

where  $\mathbf{z}_i$  is the  $i^{\text{th}}$  column of  $\mathbf{Z}$ . Next, the covariance matrix  $\mathbf{Z} \mathbf{Z}^H$  is diagonalized to produce an orthonormal matrix  $\mathbf{Q}$  and a matrix of eigenvalues  $\mathbf{\Lambda}$ ,

$$\mathbf{z}\mathbf{z}^H = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H \quad (5.10)$$

It is possible to obtain a reduced-dimensionality representation by generating the matrix  $\mathbf{Q}_T$  which contains the  $T$  columns of  $\mathbf{Q}$  that correspond to the  $T$  largest eigenvalues in  $\mathbf{\Lambda}$ . For a given pdf  $\mathbf{p}$ , the PCA projection becomes,

$$\mathbf{pca} = \mathbf{Q}_T^H(\mathbf{p} - \mathbf{p}_{mean}) \quad (5.11)$$

The location of the projected point  $\mathbf{pca}$  in the larger dimensional pdf space is,

$$\begin{aligned} \mathbf{p}_T &= \mathbf{Q}_T \mathbf{pca} + \mathbf{p}_{mean} \\ &= \mathbf{Q}_T \mathbf{Q}_T^H(\mathbf{p} - \mathbf{p}_{mean}) + \mathbf{p}_{mean} \end{aligned} \quad (5.12)$$

With the preceding definitions, the target pdf can be estimated from undersampled q-space in the least-square sense,

$$\min_{\mathbf{p}_T} \|\mathbf{F}_\Omega \mathbf{p}_T - \mathbf{q}\|_2^2 \quad (5.13)$$

Expressed in terms of the PCA coefficients,

$$\begin{aligned} &\min_{\mathbf{pca}} \|\mathbf{F}_\Omega(\mathbf{Q}_T \mathbf{pca} + \mathbf{p}_{mean}) - \mathbf{q}\|_2^2 \\ &= \min_{\mathbf{pca}} \|\mathbf{F}_\Omega \mathbf{Q}_T \mathbf{pca} - (\mathbf{q} - \mathbf{F}_\Omega \mathbf{p}_{mean})\|_2^2 \end{aligned} \quad (5.14)$$

The solution to the least squares problem in Eq. 5.14 is computed using the pseudoinverse,  $\text{pinv}(\cdot)$ ,

$$\widetilde{\mathbf{pca}} = \text{pinv}(\mathbf{F}_\Omega \mathbf{Q}_T)(\mathbf{q} - \mathbf{F}_\Omega \mathbf{p}_{mean}) \quad (5.15)$$

The result in pdf space is finally given by

$$\begin{aligned} \widehat{\mathbf{pdf}} &= \mathbf{Q}_T \widetilde{\mathbf{pca}} + \mathbf{p}_{mean} \\ &= \mathbf{Q}_T \text{pinv}(\mathbf{F}_\Omega \mathbf{Q}_T)(\mathbf{q} - \mathbf{F}_\Omega \mathbf{p}_{mean}) + \mathbf{p}_{mean} \end{aligned} \quad (5.16)$$

The reconstruction matrix  $\mathbf{Q}_T \text{pinv}(\mathbf{F}_\Omega \mathbf{Q}_T)$  needs to be computed only once. The dimension of the PCA space  $T$  is a parameter that needs to be determined. A possible way to choose this parameter is by optimizing the reconstruction performance with respect to an error metric on the training dataset. This point will be discussed in more detail under the Methods subsection.

### 5.6.2 Methods

To compare the performance of the proposed reconstruction algorithms Tikhonov regularization (PINV) and PCA against Dictionary-FOCUSS and Wavelet+TV, data from subjects A and B were analyzed. Variable-density undersampling with  $R=3$  acceleration was applied in  $q$ -space on a  $12 \times 12 \times 12$  grid. Two different dictionaries were trained with data from slice 30 of subjects A and B. Reconstruction experiments were applied on test slices that are different than the training slices. In particular, two reconstruction experiments were performed,

- i. First, voxels in slice 40 of subject A were retrospectively undersampled in  $q$ -space, and reconstructed using four different methods: Wavelet+TV method of Menzel et al. (96), Dictionary-FOCUSS, PINV and PCA. The training data were sampled from slice 30 of subject B.
- ii. Second, voxels in slice 25 of subject B were retrospectively undersampled with the same  $R=3$  sampling pattern, and again reconstructed with Wavelet+TV, Dictionary-FOCUSS, PINV and PCA. In this experiment, the training data were obtained from slice 30 of subject A.

In these experiments, slice 30 was selected for training and slices 25 and 40 were chosen for test based on their anatomical location, so that the test slices would reside on lower and upper parts of the brain, while the training slice was one of the middle slices.

By taking the fully-sampled data as ground-truth, the fidelity of the four methods were compared using RMSE normalized by the  $\ell_2$ -norm of ground-truth as the error metric both in pdf domain and  $q$ -space. To observe the performance of the dictionary-based methods, additional reconstructions were performed at the higher acceleration factors of  $R=5$  and 9.

Tikhonov regularization parameter  $\lambda$  for PINV and the dimension of the PCA space  $T$  were determined using the training data. In particular, reconstruction experiments were performed on the training dataset with the same undersampling pattern used for the test dataset, and the parameter that yielded the lowest RMSE was selected as the “optimal” regularization parameter. At all acceleration factors and for both of subjects A and B,  $\lambda=0.03$  was seen to yield the lowest RMSE values on the training set. For subject A, the optimal dimension of the PCA space was found to be  $T=(50, 26, 22)$  at accelerations  $R=(3, 5, 9)$ , respectively. For subject B, the corresponding values were  $T=(45, 27, 13)$ .

To map the performance of the methods across the brain, reconstruction experiments on multiple slices across the whole brain were performed using, Dictionary-FOCUSS, PINV and PCA. Mean and standard deviation of RMSE in pdf domain for each slice were computed for subjects A and B.

Since the fully-sampled data are corrupted by noise, computing RMSEs relative to them will include contributions from both reconstruction errors and additive noise. To address this, the additional 10 average data acquired at the selected five q-space points were used again.

To explore how reconstruction error varies as a function of q-space location, at acceleration  $R=3$ , q-space images at the “missing” (not sampled) directions were computed using the pdfs estimated by the four methods. RMSE values were obtained for all missing q-space directions by taking the fully-sampled 1 average dataset as ground truth.

### 5.6.3 Results

Wavelet+TV, Dictionary-FOCUSS, Tikhonov regularized (PINV) and PCA reconstruction errors relative to fully-sampled pdfs in slice 40 of subject A are presented in Fig. 5.8. All dictionary-based methods use the same training pdfs that were collected from slice 30 of subject B. At acceleration factor  $R=3$ , Wavelet+TV yielded 15.8% average RMSE in the reconstructed pdfs. The dictionary-based methods Dictionary-FOCUSS, PINV and PCA had 7.8%, 8.1% and 8.7% average error, respectively. At the higher acceleration factor of  $R=5$ , Dictionary-FOCUSS, PINV and PCA yielded 8.9%, 8.9%, and 9.6% RMSE, respectively. At  $R=9$ , the average RMSE figures were 10.0%, 10.2% and 11.2% for Dictionary-FOCUSS, PINV and PCA. The computation times were 1190 min for Wavelet+TV, 530 min for Dictionary-FOCUSS, 0.6 min for PINV and 0.4 min for PCA.

Fig. 5.9 compares pdf reconstruction errors obtained with the four methods for slice 25 of subject B. The training data for the dictionary-based methods were the pdfs in slice 30 of subject A. At acceleration  $R=3$ , Wavelet+TV yielded 17.5%, while Dictionary-FOCUSS, Tikhonov regularization (PINV) and PCA returned 11.4%, 11.8% and 12.8% average RMSE, respectively. At the higher acceleration factor of  $R=5$ , Dictionary-FOCUSS, PINV and PCA yielded 13.1%, 12.8% and 13.8% RMSE, respectively. At  $R=9$ , the errors were 14.2%, 14.1% and 15.5% for Dictionary-FOCUSS, PINV and PCA. The computation times were 1450 min for Wavelet+TV, 645 min for Dictionary-FOCUSS, 0.8 min for PINV and 0.6 min for PCA.

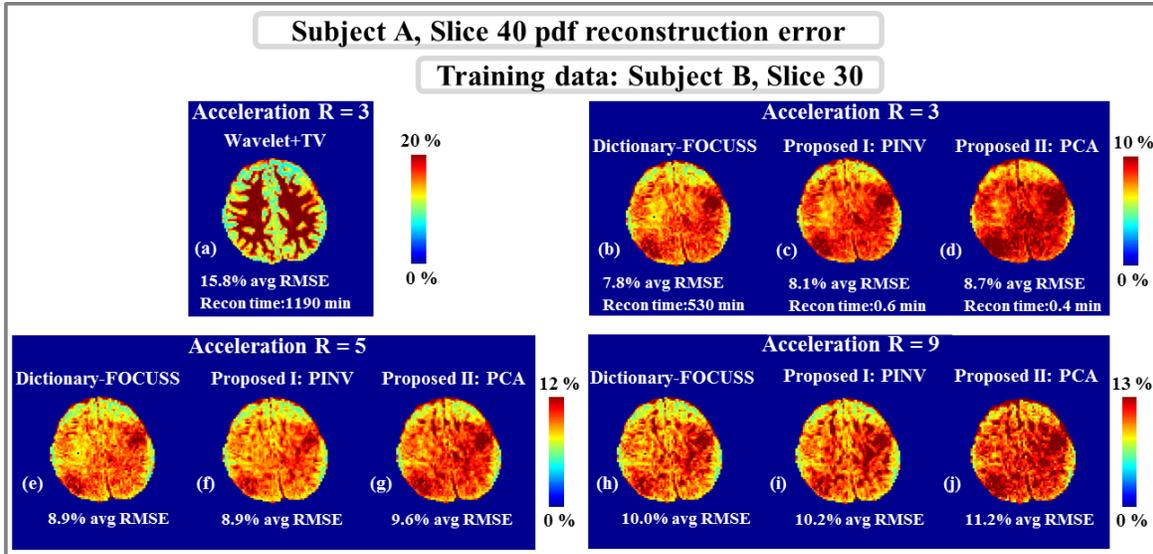


Fig. 5.8. Pdf reconstruction error maps for slice 40 of subject A. **a**: At acceleration  $R=3$ , Wavelet+TV reconstruction returned 15.8% average RMSE with a computation time of 1190 min. **b**, **c** and **d**: At  $R=3$ , Dictionary-FOCUSS yielded 7.8% error in 530 min, Tikhonov regularized reconstruction (PINV) had 8.1% error in 0.6 min and the PCA method resulted in 8.7% error in 0.4 min. **e**, **f**, and **g**: At  $R=5$ , the three dictionary-based methods yielded 8.9%, 8.9% and 9.6% RMSE. **h**, **i** and **j**: At  $R=9$ , the reconstruction errors were 10.0%, 10.2% and 11.2% for Dictionary-FOCUSS, PINV and PCA, respectively.

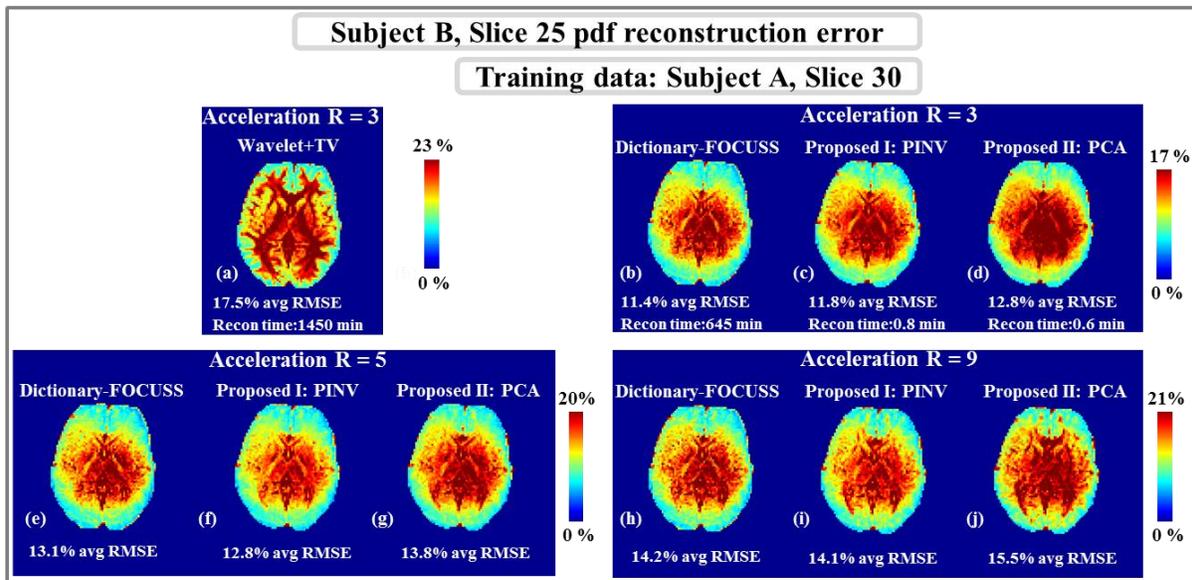


Fig. 5.9. Pdf reconstruction error maps for slice 25 of subject B. **a**: At acceleration  $R=3$ , Wavelet+TV reconstruction returned 17.5% average RMSE with a computation time of 1450 min. **b**, **c** and **d**: At  $R=3$ , Dictionary-FOCUSS yielded 11.4% error in 645 min, Tikhonov regularized reconstruction (PINV) had 11.8% error in 0.8 min and the PCA method resulted in 12.8% error in 0.6 min. **e**, **f**, and **g**: At  $R=5$ , the three dictionary-based methods yielded 13.1%, 12.8% and 13.8% RMSE. **h**, **i** and **j**: At  $R=9$ , the reconstruction errors were 14.2%, 14.1% and 15.5% for Dictionary-FOCUSS, PINV and PCA, respectively.

Pdf reconstruction errors for subject A across slices are plotted for Dictionary-FOCUSS, PINV and PCA in Fig. 5.10. At four slices, RMSE maps are also depicted for comparison. The mean RMSE averaged across all slices was found to be 11.0% for Dictionary-FOCUSS, 11.3% for PINV and 12.1% for PCA. Results from the same analysis are presented in Fig. 5.11 for subject B. In this case, the mean RMSE averaged across all slices was found to be 11.0% for Dictionary-FOCUSS, 11.3% for PINV and 12.3% for PCA.

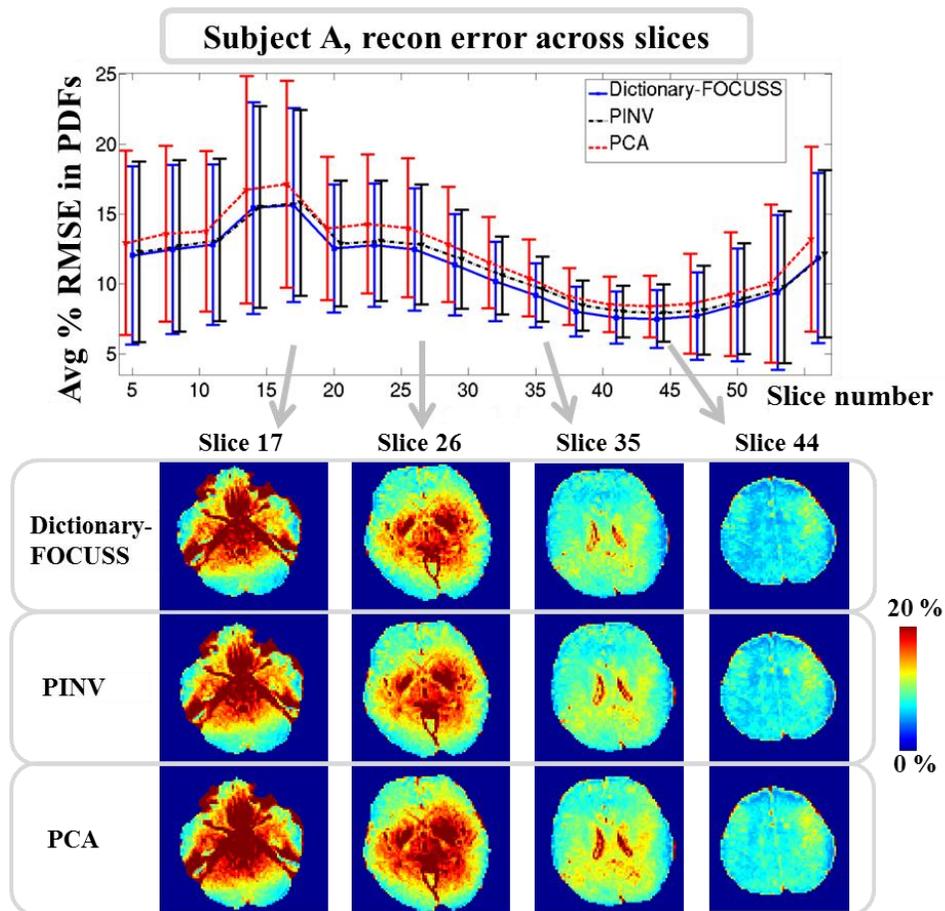


Fig. 5.10. **Upper panel:** average and standard deviation of pdf reconstruction errors in each slice for subject A. **Lower panel:** comparison of Dictionary-FOCUSS, PINV and PCA error maps at four different slices.

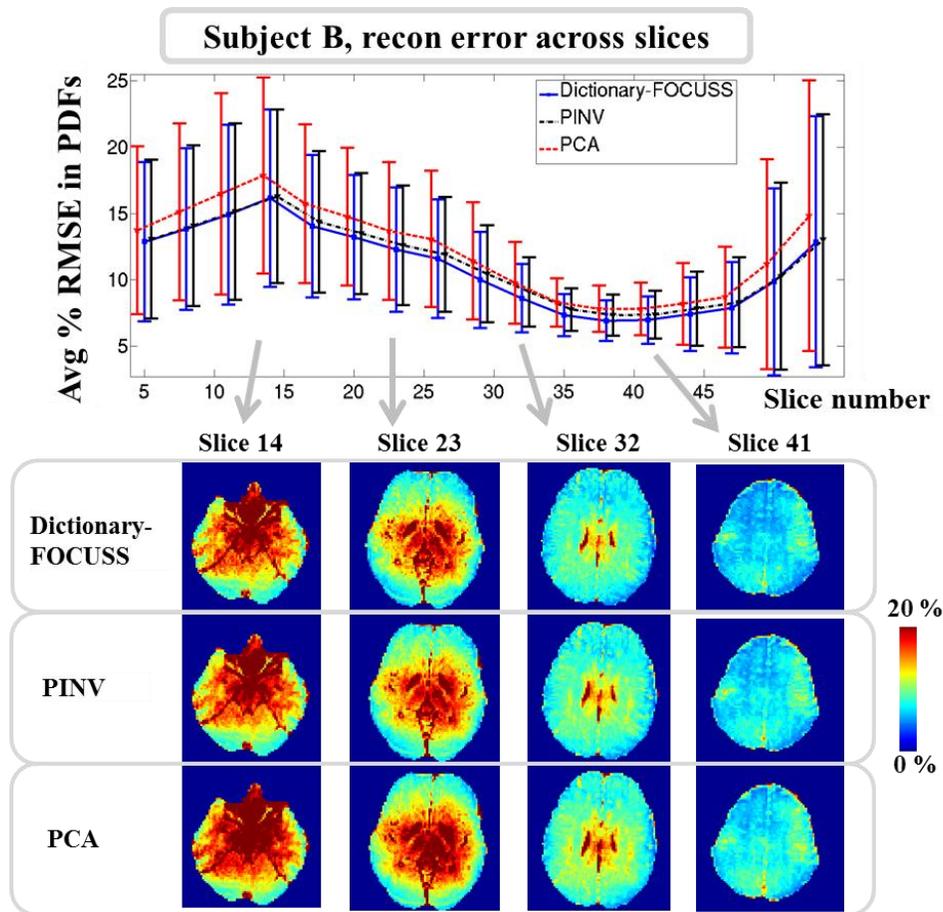


Fig. 5.11. **Upper panel:** average and standard deviation of pdf reconstruction errors in each slice for subject B. **Lower panel:** comparison of Dictionary-FOCUSS, PINV and PCA error maps at four different slices.

To isolate the reconstruction error from the contribution of noise to the RMSE figures, comparison against the 10 average dataset collected at 5 different q-space points are presented in Fig. 5.12. The comparison is based on slice 40 of subject A, while the data used for dictionary training were slice 30 of subject B. The average error for each of the curves in Fig. 5.12 were 33.7% for Wavelet+TV, 10.7% for fully-sampled 1 average data, 8.9% for Dictionary-FOCUSS, 8.4% for PINV and 9.0% for PCA reconstruction.

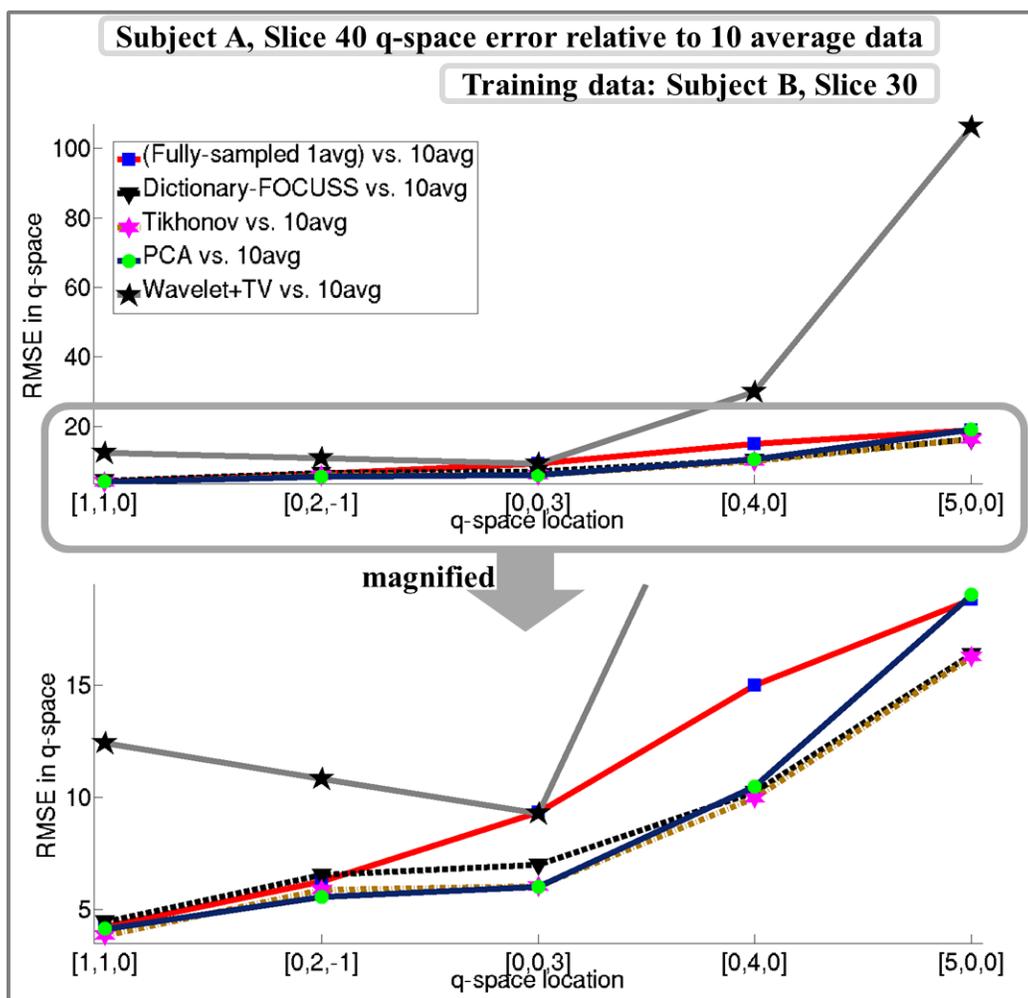


Fig. 5.12. **Upper panel:** q-space reconstruction errors relative to the 10 average data collected in five q-locations. On average, the RMSE figures were 33.7% for Wavelet+TV, 10.7% for 1 average fully-sampled data, 8.9% for Dictionary-FOCUSS, 8.4% for PINV and 9.0% for PCA. **Lower panel:** zoomed-in version for detailed comparison of the methods.

Reconstruction errors at acceleration  $R=3$  for slice 40 of subject A at the “missing” q-space points are plotted in Fig. 5.13. When averaged over all missing q-space points, the RMSEs were found to be 33.8% for Wavelet+TV, 15.7% for Dictionary-FOCUSS, 15.6% for PINV and 15.8% for PCA. The lower panel shows the q-space images at the location [5,0,0] estimated by the four reconstruction methods, as well as the fully-sampled 10 average and 1 average images.

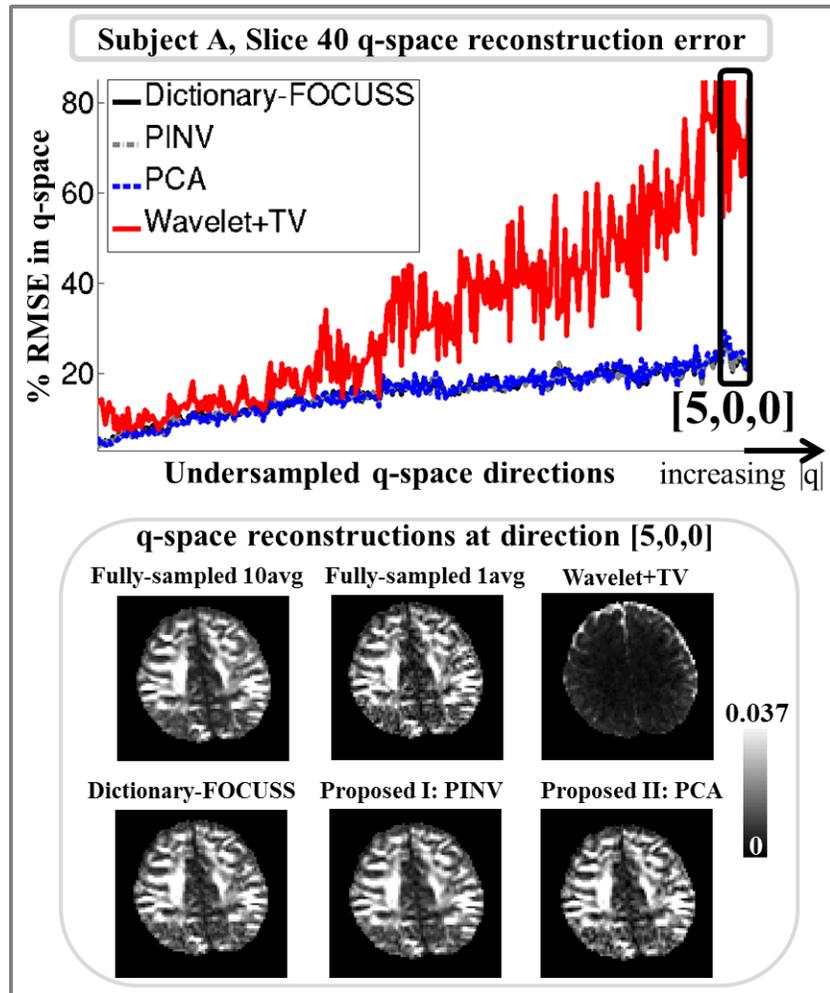


Fig. 5.13. **Upper panel:** q-space reconstruction errors at the missing directions relative to the 1 average fully sampled data. When averaged over the missing q-space points, the RMSEs were found to be 33.8% for Wavelet+TV, 15.7% for Dictionary-FOCUSS, 15.6% for PINV and 15.8% for PCA. **Lower panel:** Comparison of the q-space reconstructions at the point  $q=[5,0,0]$ .

Fig. 5.14 depicts the effect of applying  $\ell_1$  regularization on the dictionary coefficient vectors via Dictionary-FOCUSS and  $\ell_2$  regularization via PINV. Fig.7a-c show  $\ell_1$  and  $\ell_2$  regularized coefficient vectors at three example voxels. Fig. 5.14d shows the cumulative sum of  $\ell_1$  and  $\ell_2$  regularized solution vectors averaged over all voxels in the slice.  $\ell_1$  penalized coefficients reach 90% of total energy using 19% of the coefficients, whereas 63% of all coefficients are required to reach the same level with  $\ell_2$  regularization.

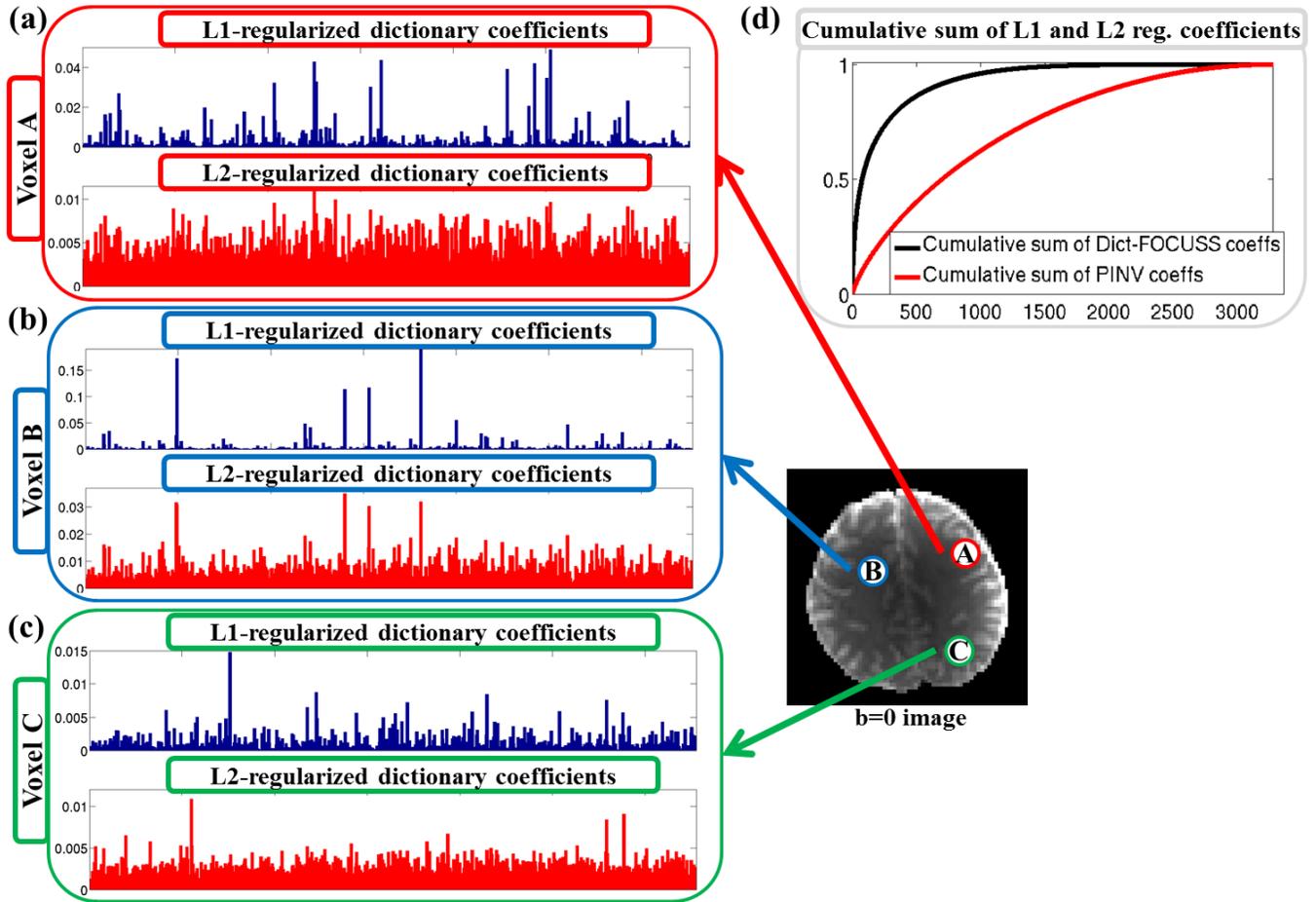


Fig. 5.14. **a**, **b** and **c**:  $\ell_1$ regularized dictionary transform coefficients obtained with Dictionary-FOCUSS and  $\ell_2$ regularized coefficients obtained with the PINV method compared at three voxels A, B and C. **d**: Cumulative sums of dictionary coefficient vectors from Dictionary-FOCUSS and PINV reconstructions averaged over the voxels in the slice.

#### 5.6.4 Remarks on Fast DSI Reconstruction with Trained Dictionaries

The two proposed methods PINV and PCA were demonstrated to have reconstruction quality comparable to that of Dictionary-FOCUSS in pdf domain and  $q$ -space based on Figs. 5.8 through 14. At the same time, they attained 3 orders of magnitude reduction in computation time relative to both of the previously proposed algorithms Wavelet+TV (96) and Dictionary-FOCUSS. With this initial implementation, which reconstructs each voxel sequentially and runs on Matlab, processing time on the order of seconds per slice is already achievable. While being feasible for

clinical application of accelerated DSI, the presented methods do not sacrifice reconstruction quality for computation speed.

An alternative approach to improving reconstruction speed is to increase the convergence rate of iterative CS algorithms through Nesterov-type gradient descent algorithms (115). These optimal methods rely on a weighted combination of all previous gradients, and reduce the number of iterations required to reach a certain solution precision. With the FISTA algorithm (115), wavelet-based deblurring was seen to require 10 times fewer number of iterations compared to gradient descent. However, even with optimal descent techniques, it would be challenging to reduce the processing time of CS algorithms from *days* to a clinically feasible interval. Further,  $\ell_1$  regularization parameters need to be determined for such optimal CS algorithms.

For both of the proposed methods, there is also one parameter that needs to be tuned, namely, the  $\ell_2$  regularization parameter for PINV and the number of columns kept in PCA. Herein, the fully-sampled training dataset was used to determine “optimal” parameters with respect to the RMSE metric. Using the same undersampling pattern that will be applied on the test dataset, the parameter setting that yields the lowest reconstruction RMSE on the training dataset is determined by parameter sweeping. It should be noted that there are other established ways to determine these regularization parameters (e.g., L-curve method, cross-validation). With the assumption that fully-sampled data exist for dictionary training, this set was further utilized for parameter extraction. For subject A, the dimension of the PCA space was determined by minimizing the reconstruction error on the training set from subject B to obtain the optimal values  $T=(50, 26, 22)$  at  $R=(3, 5, 9)$ . Instead, if the parameters were determined by optimizing the performance on subject A, the optimal values would be  $T=(49, 24, 23)$ . However, both sets of parameters yield the same RMSE on subject A, namely 8.7%, 9.6% and 11.2% at  $R=(3, 5, 9)$  in Fig. 5.8. Similarly for subject B, the optimal parameter setting was found to be  $T=(45, 27, 13)$  based on the training dataset from subject A. Assuming that it was possible to optimize the parameters with data from subject B,  $T=(45, 22, 13)$  would have been obtained. In this case, the only difference is at  $R=5$ , where the RMSE in Fig. 5.9g would decrease from 13.8% to 13.7% if  $T=22$  was used instead of  $T=27$ . Hence, the PCA parameters extracted from the training dataset generalize very well to the test dataset and yield close to optimal reconstruction performance.

Because the optimal PCA dimension was seen to differ across subjects (for A,  $T=(50, 26, 22)$  and for B,  $T=(45, 27, 13)$ ), the effect of applying the optimal  $T$  determined for subject A while reconstructing B and vice versa was tested. Regarding slice 40 of subject A, the same RMSE values were obtained at  $R=3$  and 5. At  $R=9$ , the error increased from 11.2% (with the optimal

$T=22$ ) to 11.9% (with  $T=13$  from B). Regarding the reconstruction of slice 25 of subject B, the RMSE value did not change at  $R=5$ , however it increased from 12.8% (with the optimal  $T=45$ ) to 12.9% (with  $T=50$  from A) at  $R=3$ , and from 15.5% (with the optimal  $T=13$ ) to 16.3% (with  $T=22$  from A) at  $R=9$ . These results suggest that optimal number of columns for PCA reconstruction generalizes across subjects, except when very high acceleration factors are employed.

The number of PCA columns that yielded the lowest error was seen to decrease as the acceleration factor increased. PCA reconstruction in Eq. 5.16 involves the solution of a least squares problem via the pseudoinverse of  $\mathbf{F}_\Omega \mathbf{Q}_T$ , and this problem is ill-conditioned if the condition number of  $\mathbf{F}_\Omega \mathbf{Q}_T$  is large. In this case, small errors in the entries of this matrix can lead to large errors in the solution vector. While keeping the number of columns  $T$  fixed, it was observed that the condition number increased as the acceleration factor increased. For instance, letting  $T=100$ ,  $\text{cond}(\mathbf{F}_\Omega \mathbf{Q}_T)$  was computed to be 6.8 at  $R=3$ , 113.2 at  $R=5$  and  $2.0 \cdot 10^{14}$  at  $R=9$ . This indicates that smaller number of columns needs to be used at higher accelerations to keep the least squares problem well-conditioned.

For Dictionary-FOCUSS, dictionary trained on one subject was shown to generalize to other subjects (116). Since PINV and PCA attain similar reconstruction quality, the same observation can be made for the proposed methods. Whether the dictionaries generalize across age groups or patient populations remains an open question.

When 10 average data were taken as ground truth (Fig. 5.12), all three dictionary-based methods at 3-fold acceleration were seen to yield lower error than the fully-sampled 1 average data. This could indicate that dictionary-based techniques successfully estimate the missing  $q$ -space samples as well as denoise the  $q$ -space. In accordance with this conclusion, K-SVD was recently proposed as a denoising tool for high-angular diffusion imaging (HARDI) (117), where training and denoising were performed on  $q$ -space images.

Also evident in Figs. 5.12 and 13 is the observation that the performance of all reconstruction methods deteriorates at higher  $|q|$  values. While this effect is milder in the dictionary-based techniques, the RMSE values of Wavelet+TV exceed 100% at the outer shells. As seen in the lower panel of Fig. 5.13, Wavelet+TV fails to estimate the  $q$ -space content at large  $|q|$  and yields coefficients with much lower amplitude than the fully-sampled reconstruction.

Regarding the PCA method, using a lower dimensional space reduces the number of coefficients that need to be estimated from the sampled  $q$ -space points. Considering the case at  $R=3$  with  $T=50$  principal components, the projected pdfs reside in a 50-dimensional space whose

coordinates need to be determined using 171 q-space samples (at 3-fold undersampling for 515 directions). Rather than operating in the pdf space with  $12 \times 12 \times 12 = 1728$  dimensions, PCA method seeks 50 coefficients, thus substantially reduces the effective undersampling factor of the problem. This is thanks to the prior information encoded in the training pdf dataset.

Results obtained with Dictionary-FOCUSS and PINV indicate that with a dictionary trained by K-SVD, using either  $\ell_1$  or  $\ell_2$  penalty on the dictionary transform coefficients yield comparable results. In the case of  $\ell_2$  regularization, the computational advantage is that there is a closed form expression for the solution vector. As expected, Fig. 5.14 shows that  $\ell_1$  penalty yielded several large coefficients and many smaller ones, while  $\ell_2$  regularization gave a more spread-out coefficient pattern in terms of amplitude.

## 5.7 Conclusion

By using a data-driven transform specifically tailored for sparse representation of diffusion pdfs, up to 2-fold reduction in reconstruction errors were obtained relative to using either prespecified wavelet and gradient transforms, or  $\ell_1$ -norm penalty. Further, it was demonstrated that an adaptive dictionary trained on a particular subject generalizes well to other subjects, still yielding significant benefits in CS reconstruction performance. Coupled with the parameter-free FOCUSS algorithm, the proposed method can help accelerate DSI scans in the clinical domain.

In addition to the Dictionary-FOCUSS algorithm, two dictionary-based methods that yield closed-form solutions were also presented. These methods decreased the computation time by 3 orders of magnitude relative to the iterative CS techniques, while retaining the high reconstruction quality.

## Chapter 6

### Future Directions

This chapter proposes potential extensions and future research directions related to the methods presented in this dissertation.

#### 6.1 Joint Reconstruction

Bayesian compressed sensing algorithm developed for reconstruction of multi-contrast structural MRI images from undersampled data can be extended in two directions,

- i. *Parallel imaging*: MR scanners are equipped with multiple receivers that are sensitive to signals in their vicinity. Data obtained from multiple views of the object provide extra spatial information, and thereby can also be used to facilitate reconstruction from undersampled datasets. Joint reconstruction can be augmented with parallel imaging to yield even higher savings in imaging time without loss of image quality. In addition to the capability of recent methods e.g. (7,8,118) to combine parallel imaging with sparsity techniques, joint reconstruction will also exploit the similarity across contrasts.
- ii. *Multimodal imaging*: Image similarity is not limited to multi-contrast MRI. Developments in hardware bring modalities such as MRI, PET (Positron Emission Tomography) and CT (Computed Tomography) together in one scanner. Combined MR-PET and PET-CT imaging may benefit from joint reconstruction, wherein quantitative physiology and high-resolution structural imaging complete each other.

#### 6.2 Quantitative Susceptibility Mapping

The work presented on QSM can be extended both on the application and algorithm development fronts in directions such as,

- i. *Susceptibility venography*: The paramagnetic nature of deoxyhemoglobin in cerebral veins allows quantification of vessel oxygen saturation from susceptibility, which

provides critical information for monitoring therapy in stroke and tumor (119). Since oxygenation is expected to vary slowly along a vessel, QSM-based vessel-tracking should be feasible. Tracking is a mature field in the context of *fiber tractography*, which aims to reveal white matter connectivity in the brain. The QSM tools developed herein can be combined with the techniques in the literature on fiber tractography to generate susceptibility venograms.

- ii. *QSM with magnitude prior*: While MR signal phase is processed to yield susceptibility maps, the magnitude of the signal is often overlooked. The magnitude image may provide critical information for localization of vessels and iron-rich basal ganglia structures. It might be possible to exploit this prior during deconvolution to improve the conditioning of the inverse problem.

### 6.3 Chemical Shift Imaging

While the focus of thesis on spectroscopic imaging was limited to lipid artifact reduction, constrained reconstruction techniques may also be able to alleviate some of the major challenges faced in CSI, such as low metabolite SNR and low spatial resolution. Further research directions in this domain may include,

- i. *Model based spectroscopy with lipid prior*: Parametric modeling of spectroscopy signal at each voxel allows mitigation of magnetic field inhomogeneities and facilitates improved mapping of the metabolites (93). As the spectra are characterized by a small number of parameters that need to be determined, reconstruction from undersampled data is also made easier. Extending the model to include lipid basis functions may lead to further reduction in lipid contamination, while taking advantage of the benefits of mitigated artifacts from field inhomogeneity and allowing accelerated acquisition.
- ii. *Metabolite estimation with joint reconstruction*: CSI acquisition is commonly preceded by a structural MRI for localization of the field of view. The structural images and metabolite maps share the same tissue boundaries, while the metabolite images enjoy much lower resolution than that of the structural images. Similar to the multi-contrast MRI setting wherein structural images were reconstructed jointly, it might be possible to

regularize and enhance the resolution of metabolite maps with the help of structural priors.

- iii. *Lactate imaging*: Lactate is a metabolite that plays a critical role in brain pathologies such as tumor, stroke and cerebral ischemia (120), whose detection is made especially difficult by lipid signals that reside over the same resonance frequency. In addition to facilitating the detection of NAA, regularized spectroscopic imaging with effective lipid suppression may be deployed as a critical tool in lactate imaging.

## 6.4 Diffusion Spectrum Imaging

Model-based reconstruction of diffusion pdfs from undersampled  $q$ -space was demonstrated to reduce DSI scan times without compromising the quality of tractography solutions. Another source of reduction in imaging time is parallel imaging, which can be extended and enhanced in the following ways,

- i. *Simultaneous Multi-Slice (SMS) imaging* (113,121): While the proposed dictionary-based method skips a subset of diffusion directions entirely, SMS accelerates imaging by excitation of multiple image slices at the same time. The overlapping slices are then disentangled using information from multiple receivers. SMS demonstrated 3-fold reduction in acquisition time with little impact on the image quality. As dictionary reconstruction and SMS undersample data in orthogonal directions, they may synergistically combine to yield 9-fold ( $3 \times 3$ ) reduction in scan time, rendering clinical DSI feasible for the first time.
- ii. *SMS with low-rank prior*: When DSI data are arranged into a matrix where each column is the vector form of a diffusion weighted image, this data structure can be well approximated with a low-dimensional representation. This stems from the fact that all diffusion weighted images share common structures. Enforcing this prior during SMS recovery may improve the reconstruction.



## 7. Bibliography

1. Hallgren B, Sourander P. The effect of age on the non-haemin iron in the human brain. *Journal of Neurochemistry* 1958;3(1):41-51.
2. Pruessmann KP, Weiger M, Scheidegger MB, Boesiger P. SENSE: sensitivity encoding for fast MRI. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 1999;42(5):952-962.
3. Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 2002;47(6):1202-1210.
4. Gorodnitsky IF, Rao BD. Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm. *Ieee T Signal Proces* 1997;45(3):600-616.
5. Tibshirani R. Regression shrinkage and selection via the Lasso. *J Roy Stat Soc B Met* 1996;58(1):267-288.
6. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magn Reson Med* 2007;58(6):1182-1195.
7. Lustig M, Pauly JM. SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 2010;64(2):457-471.
8. Liang D, Liu B, Wang JJ, Ying L. Accelerating SENSE Using Compressed Sensing. *Magnetic Resonance in Medicine* 2009;62(6):1574-1584.
9. Ji SH, Dunson D, Carin L. Multitask Compressive Sensing. *Ieee T Signal Proces* 2009;57(1):92-106.
10. Malioutov D, Cetin M, Willsky AS. A sparse signal reconstruction perspective for source localization with sensor arrays. *Ieee T Signal Proces* 2005;53(8):3010-3022.
11. Tropp JA, Gilbert AC, Strauss MJ. Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit. *Signal Process* 2006;86(3):572-588.
12. Chen SSB, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. *Siam Rev* 2001;43(1):129-159.
13. Cotter SF, Rao BD, Engan K, Kreutz-Delgado K. Sparse solutions to linear inverse problems with multiple measurement vectors. *Ieee T Signal Proces* 2005;53(7):2477-2488.
14. Duarte MF, Sarvotham S, Baron D, Wakin MB, Baraniuk RG. Distributed compressed sensing of jointly sparse signals. 2005 39th Asilomar Conference on Signals, Systems and Computers, Vols 1 and 2 2005:1537-1541.
15. Zelinski AC, Goyal VK, Adalsteinsson E. Simultaneously Sparse Solutions to Linear Inverse Problems with Multiple System Matrices and a Single Observation Vector. *Siam J Sci Comput* 2010;31(6):4553-4579.

16. Ji SH, Xue Y, Carin L. Bayesian compressive sensing. *Ieee T Signal Proces* 2008;56(6):2346-2356.
17. Rangan S, Fletcher AK, Goyal VK. Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing. arXiv:09063234v1 2009.
18. Mallat SG. A wavelet tour of signal processing : the Sparse way. Amsterdam ; Boston: Elsevier/Academic Press; 2009. xx, 805 p. p.
19. Maleh R. Efficient sparse approximation methods for medical imaging. PhD Thesis 2009.
20. Candes EJ, Romberg JK, Tao T. Stable signal recovery from incomplete and inaccurate measurements. *Commun Pur Appl Math* 2006;59(8):1207-1223.
21. Tipping ME. Sparse Bayesian learning and the relevance vector machine. *J Mach Learn Res* 2001;1(3):211-244.
22. Rohlfing T, Zahr NM, Sullivan EV, Pfefferbaum A. The SRI24 multichannel atlas of normal adult human brain structure. *Human Brain Mapping* 2010;31(5):798-819.
23. Murphy M, Keutzer K, Vasanaawala S, Lustig M. Clinically Feasible Reconstruction Time for L1-SPIRiT Parallel Imaging and Compressed Sensing MRI. 18th Annual ISMRM Scientific Meeting and Exhibition, 2010 2010.
24. AGILE: An open source library for image reconstruction using graphics card hardware acceleration. submitted to 19th Annual ISMRM Scientific Meeting and Exhibition 2011.
25. He LH, Chen HJ, Carin L. Tree-Structured Compressive Sensing With Variational Bayesian Analysis. *Ieee Signal Proc Let* 2010;17(3):233-236.
26. van der Kouwe AJW, Benner T, Dale AM. Real-time rigid body motion correction and shimming using cloverleaf navigators. *Magn Reson Med* 2006;56(5):1019-1032.
27. Duyn JH, Yang YH, Frank JA, van der Veen JW. Simple correction method for k-space trajectory deviations in MRI. *J Magn Reson* 1998;132(1):150-153.
28. Jezzard P, Balaban RS. Correction for Geometric Distortion in Echo-Planar Images from B-0 Field Variations. *Magn Reson Med* 1995;34(1):65-73.
29. Sharma SD, Fong, C., Tzung, B., Nayak, K. S., and Law, M. Clinical Image Quality Assessment of CS-Reconstructed Brain Images. 18th Annual ISMRM Scientific Meeting and Exhibition 2010.
30. Seeger M, Nickisch H, Pohmann R, Scholkopf B. Optimization of k-space trajectories for compressed sensing by Bayesian experimental design. *Magn Reson Med* 2010;63(1):116-126.
31. Weller DS, Polimeni JR, Grady LJ, Wald LL, Adalsteinsson E, Goyal VK. Combining nonconvex compressed sensing and GRAPPA using the nullspace method. 18th Annual ISMRM Scientific Meeting and Exhibition 2010 2010.
32. Hallgren B, Sourander P. The non-haemin iron in the cerebral cortex in Alzheimer's disease. *J Neurochem* 1960;5:307-310.
33. Bartzokis G, Tishler TA, Lu PH, Villablanca P, Altshuler LL, Carter M, Huang D, Edwards N, Mintz J. Brain ferritin iron may influence age- and gender-related risks of neurodegeneration. *Neurobiol Aging* 2007;28(3):414-423.
34. Haacke EM, Ayaz M, Khan A, Manova ES, Krishnamurthy B, Gollapalli L, Ciulla C, Kim I, Petersen F, Kirsch W. Establishing a baseline phase behavior in magnetic

- resonance imaging to determine normal vs. abnormal iron content in the brain. *Journal of Magnetic Resonance Imaging* 2007;26(2):256-264.
35. Pfefferbaum A, Adalsteinsson E, Rohlfing T, Sullivan EV. MRI estimates of brain iron concentration in normal aging: Comparison of field-dependent (FDRI) and phase (SWI) methods. *Neuroimage* 2009;47(2):493-500.
  36. Pfefferbaum A, Adalsteinsson E, Rohlfing T, Sullivan EV. Diffusion tensor imaging of deep gray matter brain structures: Effects of age and iron concentration. *Neurobiology of Aging* 2010;31(3):482-493.
  37. Raz N, Rodrigue KM, Haacke EM. Brain aging and its modifiers: insights from in vivo neuromorphometry and susceptibility weighted imaging. *Ann N Y Acad Sci* 2007;1097:84-93.
  38. Bartzokis G, Lu PH, Tingus K, Mendez MF, Richard A, Peters DG, Oluwadara B, Barrall KA, Finn JP, Villablanca P, Thompson PM, Mintz J. Lifespan trajectory of myelin integrity and maximum motor speed. *Neurobiol Aging* 2010;31(9):1554-1562.
  39. Sullivan EV, Adalsteinsson E, Rohlfing T, Pfefferbaum A. Relevance of Iron Deposition in Deep Gray Matter Brain Structures to Cognitive and Motor Performance in Healthy Elderly Men and Women: Exploratory Findings. *Brain Imaging Behav* 2009;3(2):167-175.
  40. Bartzokis G, Aravagiri M, Oldendorf WH, Mintz J, Marder SR. Field dependent transverse relaxation rate increase may be a specific measure of tissue iron stores. *Magnetic Resonance in Medicine* 1993;29(4):459-464.
  41. Haacke EM, Xu Y, Cheng YC, Reichenbach JR. Susceptibility weighted imaging (SWI). *Magn Reson Med* 2004;52(3):612-618.
  42. Haacke EM, Cheng NY, House MJ, Liu Q, Neelavalli J, Ogg RJ, Khan A, Ayaz M, Kirsch W, Obenaus A. Imaging iron stores in the brain using magnetic resonance imaging. *Magnetic Resonance Imaging* 2005;23:1-25.
  43. Liu T, Liu J, de Rochefort L, Ledoux J, Zhang Q, Prince MR, Wu J, Wang Y. Measurement of iron concentration in human brain using Quantitative Susceptibility Mapping (QSM): correlation with age. *International Society for Magnetic Resonance in Medicine 18th Scientific Meeting* 2010.
  44. Duyn JH, van Gelderen P, Li TQ, de Zwart JA, Koretsky AP, Fukunaga M. High-field MRI of brain cortical substructure based on signal phase. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104(28):11796-11801.
  45. Schweser F, Deistung A, Lehr BW, Reichenbach JR. Quantitative imaging of intrinsic magnetic tissue properties using MRI signal phase: An approach to in vivo brain iron metabolism? *Neuroimage* 2011;54(4):2789-2807.
  46. Hazell AS, Butterworth RF. Hepatic encephalopathy: An update of pathophysiologic mechanisms. *Proc Soc Exp Biol Med* 1999;222(2):99-112.
  47. de Rochefort L, Liu T, Kressler B, Liu J, Spincemaille P, Lebon V, Wu JL, Wang Y. Quantitative Susceptibility Map Reconstruction from MR Phase Data Using Bayesian Regularization: Validation and Application to Brain Imaging. *Magnetic Resonance in Medicine* 2010;63(1):194-206.

48. Liu T, Khalidov I, de Rochefort L, Spincemaille R, Liu J, Wang Y. Improved background field correction using effective dipole fitting. International Society for Magnetic Resonance in Medicine 18th Scientific Meeting 2010.
49. Marques JP, Bowtell R. Application of a fourier-based method for rapid calculation of field inhomogeneity due to spatial variation of magnetic susceptibility. Concepts in Magnetic Resonance Part B-Magnetic Resonance Engineering 2005;25B(1):65-78.
50. Salomir R, De Senneville BD, Moonen CTW. A fast calculation method for magnetic field inhomogeneity due to an arbitrary distribution of bulk susceptibility. Concepts in Magnetic Resonance Part B-Magnetic Resonance Engineering 2003;19B(1):26-34.
51. Schenck JF. The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds. Medical Physics 1996;23(6):815-850.
52. Neelavalli J, Cheng YCN, Jiang J, Haacke EM. Removing Background Phase Variations in Susceptibility-Weighted Imaging Using a Fast, Forward-Field Calculation. Journal of Magnetic Resonance Imaging 2009;29(4):937-948.
53. Schweser F, Atterbury M, Deistung A, Lehr BW, Sommer K, Reichenbach JR. Harmonic phase subtraction methods are prone to B1 background components. International Society for Magnetic Resonance in Medicine 19th Scientific Meeting 2011.
54. Lustig M, Donoho D, Pauly JM. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magnetic Resonance in Medicine 2007;58(6):1182-1195.
55. Hansen PC. The L-Curve and its Use in the Numerical Treatment of Inverse Problems. Computational inverse problems in electrocardiology 2000:119-142.
56. Mattis S. Dementia rating scale. Psychological Assessment Resources, Odessa, FL 1988.
57. Haacke EM, Chengb NYC, House MJ, Liu Q, Neelavalli J, Ogg RJ, Khan A, Ayaz M, Kirsch W, Obenaus A. Imaging iron stores in the brain using magnetic resonance imaging. Magnetic Resonance Imaging 2005;23(1):1-25.
58. Jenkinson M. Fast, automated, N-dimensional phase-unwrapping algorithm. Magnetic Resonance in Medicine 2003;49(1):193-197.
59. Smith SM. Fast robust automated brain extraction. Human Brain Mapping 2002;17(3):143-155.
60. Rohlfing T, Maurer CR, Jr. Nonrigid image registration in shared-memory multiprocessor environments with application to brains, breasts, and bees. IEEE Trans Inf Technol Biomed 2003;7(1):16-25.
61. Ogg RJ, Langston JW, Haacke EM, Steen RG, Taylor JS. The correlation between phase shifts in gradient-echo MR images and regional brain iron concentration. Magn Reson Imaging 1999;17(8):1141-1148.
62. Wharton S, Bowtell R. Whole-brain susceptibility mapping at high field: A comparison of multiple- and single-orientation methods. Neuroimage 2010;53(2):515-525.
63. Lee J, Shmueli K, Fukunaga M, van Gelderen P, Merkle H, Silva AC, Duyn JH. Sensitivity of MRI resonance frequency to the orientation of brain tissue microstructure. Proc Natl Acad Sci U S A 2010;107(11):5130-5135.
64. Liu J, Liu T, de Rochefort L, Khalidov I, Price M, Wang Y. Quantitative susceptibility mapping by regulating the field to source inverse problem with a sparse prior derived

- from the Maxwell Equation: validation and application to brain. International Society for Magnetic Resonance in Medicine 18th Scientific Meeting 2010.
65. Duyn JH, van Gelderen P, Li TQ, de Zwart JA, Koretsky AP, Fukunaga M. High-field MRI of brain cortical substructure based on signal phase. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104(28):11796-11801.
  66. Bartzokis G, Cummings JL, Markham CH, Marmarelis PZ, Treciokas LJ, Tishler TA, Marder SR, Mintz J. MRI evaluation of brain iron in earlier- and later-onset Parkinson's disease and normal subjects. *Magn Reson Imaging* 1999;17(2):213-222.
  67. Bartzokis G, Lu PH, Tishler TA, Fong SM, Oluwadara B, Finn JP, Huang D, Bordelon Y, Mintz J, Perlman S. Myelin Breakdown and Iron Changes in Huntington's Disease: Pathogenesis and Treatment Implications. *Neurochem Res* 2007;32(10):1655-1664.
  68. Bartzokis G, Tishler TA. MRI evaluation of basal ganglia ferritin iron and neurotoxicity in Alzheimer's and Huntington's disease. *Cell Mol Biol (Noisy-le-grand)* 2000;46(4):821-833.
  69. Brass SD, Chen NK, Mulkern RV, Bakshi R. Magnetic resonance imaging of iron deposition in neurological disorders. *Top Magn Reson Imaging* 2006;17(1):31-40.
  70. Granholm E, Bartzokis G, Asarnow RF, Marder SR. Preliminary associations between motor procedural learning, basal ganglia T2 relaxation times, and tardive dyskinesia in schizophrenia. *Psychiatry Res* 1993;50(1):33-44.
  71. Martin WR, Roberts TE, Ye FQ, Allen PS. Increased basal ganglia iron in striatonigral degeneration: in vivo estimation with magnetic resonance. *Can J Neurol Sci* 1998;25(1):44-47.
  72. Michaeli S, Oz G, Sorce DJ, Garwood M, Ugurbil K, Majestic S, Tuite P. Assessment of brain iron and neuronal integrity in patients with Parkinson's disease using novel MRI contrasts. *Mov Disord* 2007;22(3):334-340.
  73. Duyn JH, Gillen J, Sobering G, van Zijl PC, Moonen CT. Multisection proton MR spectroscopic imaging of the brain. *Radiology* 1993;188(1):277-282.
  74. Le Roux P, Gilles RJ, McKinnon GC, Carlier PG. Optimized outer volume suppression for single-shot fast spin-echo cardiac imaging. *J Magn Reson Imaging* 1998;8(5):1022-1032.
  75. Luo Y, de Graaf RA, DelaBarre L, Tannus A, Garwood M. BISTRO: an outer-volume suppression method that tolerates RF field inhomogeneity. *Magn Reson Med* 2001;45(6):1095-1102.
  76. Bydder GM, Young IR. MR imaging: clinical use of the inversion recovery sequence. *J Comput Assist Tomogr* 1985;9(4):659-675.
  77. Ebel A, Govindaraju V, Maudsley AA. Comparison of inversion recovery preparation schemes for lipid suppression in 1H MRSI of human brain. *Magn Reson Med* 2003;49(5):903-908.
  78. Spielman DM, Pauly JM, Macovski A, Glover GH, Enzmann DR. Lipid-suppressed single- and multisection proton spectroscopic imaging of the human brain. *J Magn Reson Imaging* 1992;2(3):253-262.
  79. Spielman D, Meyer C, Macovski A, Enzmann D. 1H spectroscopic imaging using a spectral-spatial excitation pulse. *Magn Reson Med* 1991;18(2):269-279.

80. Bottomley PA. Spatial localization in NMR spectroscopy in vivo. *Ann N Y Acad Sci* 1987;508:333-348.
81. Adalsteinsson E, Star-Lack J, Meyer CH, Spielman DM. Reduced spatial side lobes in chemical-shift imaging. *Magn Reson Med* 1999;42(2):314-323.
82. Lee J, Adalsteinsson E. Algorithm for Lipid Suppression by Real-Time Isotropic Filter Design in Spectroscopic Brain Imaging. *International Society for Magnetic Resonance in Medicine 20th Scientific Meeting 2011*:Page:1419.
83. Sarkar S, Heberlein K, Hu X. Truncation artifact reduction in spectroscopic imaging using a dual-density spiral k-space trajectory. *Magn Reson Imaging* 2002;20(10):743-757.
84. Hu X, Stillman AE. Technique for reduction of truncation artifact in chemical shift images. *IEEE Trans Med Imaging* 1991;10(3):290-294.
85. Metzger G, Sarkar S, Zhang X, Heberlein K, Patel M, Hu X. A hybrid technique for spectroscopic imaging with reduced truncation artifact. *Magn Reson Imaging* 1999;17(3):435-443.
86. Plevritis SK, Macovski A. MRS imaging using anatomically based k-space sampling and extrapolation. *Magn Reson Med* 1995;34(5):686-693.
87. Haupt CI, Schuff N, Weiner MW, Maudsley AA. Removal of lipid artifacts in 1H spectroscopic imaging by data extrapolation. *Magn Reson Med* 1996;35(5):678-687.
88. Lee J, Adalsteinsson E. Iterative CSI Reconstruction with High-Resolution Spatial Priors for Improved Lipid Suppression. *International Society for Magnetic Resonance in Medicine 19th Scientific Meeting 2010*:Page: 965.
89. Bertsekas DP. *Nonlinear programming*. Belmont, Mass.: Athena Scientific; 1999. xiv, 777 p. p.
90. Lustig M, Kim SJ, Pauly JM. A fast method for designing time-optimal gradient waveforms for arbitrary k-space trajectories. *IEEE Trans Med Imaging* 2008;27(6):866-873.
91. Fessler JA, Sutton BP. Nonuniform fast Fourier transforms using min-max interpolation. *Ieee T Signal Proces* 2003;51(2):560-574.
92. Segonne F, Dale AM, Busa E, Glessner M, Salat D, Hahn HK, Fischl B. A hybrid approach to the skull stripping problem in MRI. *Neuroimage* 2004;22(3):1060-1075.
93. Eslami R, Jacob M. Robust Reconstruction of MRSI Data Using a Sparse Spectral Model and High Resolution MRI Priors. *IEEE transactions on medical imaging* 2010;29(6):1297-1309.
94. Gagoski B, Bilgic B, Adalsteinsson E. Accelerated Spiral Chemical Shift Imaging with Compressed Sensing. *International Society for Magnetic Resonance in Medicine 21st Scientific Meeting, 2012*.
95. Ebel A, Maudsley AA, Schuff N. Correction of local B0 shifts in 3D EPSI of the human brain at 4 T. *Magn Reson Imaging* 2007;25(3):377-380.
96. Menzel MI, Tan ET, Khare K, Sperl JI, King KF, Tao XD, Hardy CJ, Marinelli L. Accelerated Diffusion Spectrum Imaging in the Human Brain Using Compressed Sensing. *Magnetic Resonance in Medicine* 2011;66(5):1226-1233.

97. Basser PJ, Mattiello J, LeBihan D. MR diffusion tensor spectroscopy and imaging. *Biophys J* 1994;66(1):259-267.
98. Tuch DS. Q-ball imaging. *Magn Reson Med* 2004;52(6):1358-1372.
99. Tournier JD, Calamante F, Gadian DG, Connelly A. Direct estimation of the fiber orientation density function from diffusion-weighted MRI data using spherical deconvolution. *Neuroimage* 2004;23(3):1176-1185.
100. Callaghan PT, Eccles CD, Xia Y. Nmr Microscopy of Dynamic Displacements - K-Space and Q-Space Imaging. *J Phys E Sci Instrum* 1988;21(8):820-822.
101. Wedeen VJ, Hagmann P, Tseng WY, Reese TG, Weisskoff RM. Mapping complex tissue architecture with diffusion spectrum magnetic resonance imaging. *Magn Reson Med* 2005;54(6):1377-1386.
102. Menzel MI, Tan ET, Khare K, Sperl JI, King KF, Tao X, Hardy CJ, Marinelli L. Accelerated diffusion spectrum imaging in the human brain using compressed sensing. *Magn Reson Med* 2011;66(5):1226-1233.
103. Saint-Amant E, Descoteaux M. Sparsity Characterisation of the Diffusion Propagator. 19th Annual ISMRM Scientific Meeting and Exhibition 2011, Montreal, Canada.
104. Aharon M, Elad M, Bruckstein A. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *Ieee T Signal Proces* 2006;54(11):4311-4322.
105. Ravishankar S, Bresler Y. MR image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans Med Imaging* 2011;30(5):1028-1041.
106. Otazo R, Sodickson DK. Adaptive Compressed Sensing MRI. 18th Annual ISMRM Scientific Meeting and Exhibition 2010, Stockholm, Sweden.
107. Keil B, Blau JM, Hoecht P, Biber S, Hamm M, Wald LL. A 64-channel brain array coil for 3T imaging. 20th Annual ISMRM Scientific Meeting and Exhibition 2012, Melbourne, Australia.
108. Bodammer N, Kaufmann J, Kanowski M, Tempelmann C. Eddy current correction in diffusion-weighted imaging using pairs of images acquired with opposite diffusion gradient polarity. *Magnet Reson Med* 2004;51(1):188-193.
109. Jenkinson M, Bannister P, Brady M, Smith S. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage* 2002;17(2):825-841.
110. Wakana S, Caprihan A, Panzenboeck MM, Fallon JH, Perry M, Gollub RL, Hua K, Zhang J, Jiang H, Dubey P, Blitz A, van Zijl P, Mori S. Reproducibility of quantitative tractography methods applied to cerebral white matter. *Neuroimage* 2007;36(3):630-644.
111. Yendiki A, Panneck P, Srinivasan P, Stevens A, Zollei L, Augustinack J, Wang R, Salat D, Ehrlich S, Behrens T, Jbabdi S, Gollub R, Fischl B. Automated probabilistic reconstruction of white-matter pathways in health and disease using an atlas of the underlying anatomy. *Front Neuroinform* 2011;5:23.
112. Talairach J, Tournoux P. Co-planar stereotaxic atlas of the human brain : 3-dimensional proportional system : an approach to cerebral imaging. Stuttgart ; New York: G. Thieme ; New York : Thieme Medical Publishers; 1988. viii, 122 p. p.
113. Setsompop K, Gagoski BA, Polimeni JR, Witzel T, Wedeen VJ, Wald LL. Blipped-controlled aliasing in parallel imaging for simultaneous multislice echo planar imaging

- with reduced g-factor penalty. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 2012;67(5):1210-1224.
114. Miranda AA, Le Borgne YA, Bontempi G. New routes from minimal approximation error to principal components. *Neural Process Lett* 2008;27(3):197-207.
  115. Beck A, Teboulle M. A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems. *Siam J Imaging Sci* 2009;2(1):183-202.
  116. Bilgic B, Setsompop K, Cohen-Adad J, Yendiki A, Wald LL, Adalsteinsson E. Accelerated diffusion spectrum imaging with compressed sensing using adaptive dictionaries. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 2012.
  117. Patel V, Shi YG, Thompson PM, Toga AW. K-Svd for Hardi Denoising. *I S Biomed Imaging* 2011:1805-1808.
  118. Weller DS, Polimeni JR, Grady L, Wald LL, Adalsteinsson E, Goyal VK. Denoising sparse images from GRAPPA using the nullspace method. *Magnetic Resonance in Medicine* 2012;68(4):1176-1189.
  119. Fan AP, Benner T, Bolar DS, Rosen BR, Adalsteinsson E. Phase-based regional oxygen metabolism (PROM) using MRI. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine* 2012;67(3):669-678.
  120. Lange T, Dydak U, Roberts TPL, Rowley HA, Bieljac M, Boesiger P. Pitfalls in lactate measurements at 3T. *Am J Neuroradiol* 2006;27(4):895-901.
  121. Setsompop K, Cohen-Adad J, Gagoski BA, Raij T, Yendiki A, Keil B, Wedeen VJ, Wald LL. Improving diffusion MRI using simultaneous multi-slice echo planar imaging. *NeuroImage* 2012;63(1):569-580.