MATHEMATICAL MODELING AND FAST NUMERICAL ALGORITHMS IN MEDICAL IMAGING APPLICATIONS

By

XIAOJING YE

A DISSERTATION PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

2011

To my parents

ACKNOWLEDGMENTS

Foremost, my greatest thankfulness goes to my advisor, Yunmei Chen, for her wise guidance and support throughout my time as a graduate researcher. I am very grateful for her encouragement and efforts to my academic career.

Many thanks to Feng Huang for the fruitful discussion on the interesting problems in magnetic resonance imaging. Several projects related to partially parallel imaging reconstruction were originally motivated by these clinical applications.

I would like to thank Lei Zhang for valuable discussion about my academic career as well as his consistent support to my study and life these years.

Thanks to William Hager and his students Dung Phan and Hongchao Zhang for the helpful discussion on optimization techniques which I have been using for several of my projects.

Many thanks to my other committee members Jay Gopalakrishnan, Scott McCullough, Jonathan Li and Murali Rao, who have been very supportive to my graduate research.

I would like to thank Hani Doss, Liqing Yan, and Mark Yang for supervising my master thesis at the Department of Statistics. Although this is somehow an area quite different from my own, I learned a lot from them.

TABLE OF CONTENTS

LIST OF TABLES

Table page

LIST OF FIGURES

Abstract of Dissertation Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Doctor of Philosophy

MATHEMATICAL MODELING AND FAST NUMERICAL ALGORITHMS IN MEDICAL IMAGING APPLICATIONS

By

Xiaojing Ye

May 2011

Chair: Yunmei Chen
Major: Mathematics

A series of projects are presented to provide advanced mathematical models and numerical algorithms that improve the accuracy, robustness and efficiency of the compressed sensing reconstruction technique in magnetic resonance imaging and inverse consistent image regisration.

Chapter one introduces a novel variational model that enforces the sparsity of the underlying image in terms of its spatial finite differences and representation with respect to a dictionary. The dictionary is trained using prior information to improve accuracy in reconstruction. In the meantime the proposed model enforces the consistency of the underlying image with acquired data by using the maximum likelihood estimator of the reconstruction error in partial k-space to improve the robustness to parameter selection. Moreover, a simple and fast numerical scheme is provided to solve this model.

In chapters two and three we develop fast numerical algorithms for solving total variation and ℓ_1 (TVL1) based image reconstruction with application in partially parallel MR imaging. Our algorithms use variable splitting method to reduce computational cost. Moreover, the Barzilai-Borwein step size selection method is adopted in our algorithms for much faster convergence. Theoretical and experimental results on clinical partially parallel imaging data demonstrate that the proposed algorithm requires much fewer iterations and/or less computational cost than recently developed operator splitting and

Bregman operator splitting methods, which can deal with a general sensing matrix in reconstruction framework, to get similar or even better quality of reconstructed images.

Chapter four introduces a novel variational model for inverse consistent deformable image registration. The proposed model deforms both source and target images simultaneously, and aligns the deformed images in the way that the forward and backward transformations are inverse consistent. To avoid the direct computation of the inverse transformation fields, our model estimates two more vector fields by minimizing their invertibility error using the deformation fields. Moreover, to improve the robustness of the model to the choice of parameters, the dissimilarity measure in the energy functional is derived using the likelihood estimation.

The experimental results on clinical data indicate the efficiency of the proposed method with improved robustness, accuracy and inverse consistency. These methods are aimed to benefit the practical usage of medical imaging techniques.

CHAPTER 1
A NOVEL METHOD AND FAST ALGORITHM FOR MR IMAGE RECONSTRUCTION VIA LEARNT DICTIONARIES

Outline

The aim of this work is to improve the accuracy, robustness and efficiency of the compressed sensing reconstruction technique in magnetic resonance imaging. We propose a novel variational model that enforces the sparsity of the underlying image in terms of its spatial finite differences and representation with respect to a dictionary. The dictionary is trained using prior information to improve accuracy in reconstruction. In the meantime the proposed model enforces the consistency of the underlying image with acquired data by using the maximum likelihood estimator of the reconstruction error in partial k-space to improve the robustness to parameter selection. Moreover, a simple and fast numerical scheme is provided to solve this model. The experimental results on both synthetic and in vivo data indicate the improvement of the proposed model in preservation of fine structures, flexibility of parameter decision, and reduction of computational cost.

1.1 Backgrounds in Compressive Magnetic Resonance Imaging

Magnetic resonance (MR) imaging is a technique that allows visualization of structures and functions of a body by non-invasive and non-ionizing means. It provides better contrast between the different soft tissues than most other modalities. However, MR imaging takes much longer acquisition time than many other imaging modalities, which limits its applications. To reduce acquisition time, the most common and feasible approach is by acquiring only partial k-space data, followed by adequate reconstruction techniques to obtain images with well-preserved quality.

The idea of reconstructing images from partial data coincides with the compressed sensing (CS), a technique used in signal/image processing. CS can accurately recover a signal/image using data with significantly less measurements than regular, provided the sparsity of the underlying signal/image and a sophisticated reconstruction procedure.

Recently, the application of this technique in medical imaging has become a hot research topic, and shown promising results [16–18, 21, 22, 27, 42, 46, 50, 68]. In particular, the redundancy of the MR data acquired in the frequency domain, i.e. the k-space, and implicit sparsity in MR images have motivated many researchers to study the application of CS to fast MR imaging (CS-MRI).

CS-MRI has the advantage of producing high quality reconstruction of MR images from partial Fourier data. Recent study has shown that the key to the success of CS-MRI is a combination of the sparsity of the underlying image under an appropriate domain, the k-space trajectory that provides incoherent undersampling artifacts, and an adequate nonlinear reconstruction method that enforces both the sparsity and data consistency of the underlying image [26, 38, 50]. A great progress of researches on CS-MRI has been made. However, for clinical applications, radiologists often demand improvements on accuracy, robustness, and efficiency of the current CS-MRI algorithms. The desired improvements include the ability of removing artifacts while preserving important diagnostic information (in particular, sharp edges and fine structures), the robustness to the choice of parameters, and the speed of reconstructions.

In this paper, we propose a novel variational model and a fast numerical algorithm for MR image reconstruction with highly undersampled data, which tackles the three problems mentioned above as follows.

- *Accuracy*
  The proposed model enforces the sparsity of the underlying image in terms of its spatial finite differences and representation by a dictionary trained using prior information. Thus, improvement on accuracy of reconstruction can be achieved.

- *Robustness*
  The proposed model enhances the data consistency by the approach of maximum likelihood estimation for the discrepancy between the reconstruction and acquired data in k-space. This leads to an automatically optimized weighting parameter which makes the parameter selection more flexible.

- *Efficiency*

To make the proposed model clinically applicable, we also provide a simple and fast numerical algorithm to solve the model. The main computations involve only shrinkage, matrix-vector multiplication and fast Fourier transform (FFT).

The background and brief introduction of our contributions to these issues are provided in the following three subsections.

1.1.1 Trained Dictionaries as Sparsifying Transforms

Since sparsity is the key to the success of CS and consequent reconstructions, many researches exploited the transforms under which images have their sparse representations [50]. The theory of CS indicates that once such transforms were found, an image can be accurately recovered using a set of random measurements with cardinality much less than the original resolution of the image [18, 26].

In recent years, finite difference operator and wavelet transforms have been widely used as such sparsifying transforms for MR images [34, 50]. In [34], the authors proposed a total variation (TV) based model to reconstruct MR images from partially acquired k-space data. Their model works well for piecewise constant or very homogeneous images [59]. For images with inhomogeneous intensity, TV based models may not work well when the undersampling rate is high. In [50], Lustig *et al.* proposed a model that minimizes the Besov together with TV norms of the underlying image, subjected to a data consistency constraint:

$$\min_{u} TV(u) + \mu\|\Psi^\top u\|_1, \quad \text{s.t. } \|\mathcal{F}_p u - f_p\|_2 < \sigma, \tag{1–1}$$

where $\|\cdot\|_1$ is the ℓ_1 norm, $TV(u) \triangleq \|Du\|_1$ is the (anisotropic) TV semi-norm of u, Ψ is the Haar wavelet transform, the superscript $^\top$ denotes (conjugate) transpose of matrices. In the constraint, $\mathcal{F}_p$ denotes the undersampled Fourier operator corresponding to the customized k-space sampling pattern, f_p is the partially acquired k-space data, and σ is an estimate of acquisition error. As proved in [18, 26], minimizing ℓ_1 norm subjected to data consistency yields sparse solutions. Therefore, model (1–1) in fact leads to a reconstructed image that has sparse gradient and wavelet transform

coefficients. Simulations in [50] showed very promising results using model (1–1). However, image quality degrading and the loss of diagnostic information may happen in reconstructions using TV and wavelet transforms as they may eliminate some fine structures and/or useful local information in the recovered images. As an alternate, we propose to use the dictionaries trained using prior information as sparsifying transforms to tackle this problem.

A recent boost of the study on dictionary design shows the great potential of using dictionaries in signal/image processing. Dictionary is usually formed as a set of overcomplete bases and its elements/atoms have much smaller sizes than the image size. On the contrary, wavelet transform has a set of complete bases with elements of the same size as image itself, and therefore can be treated as a special case of dictionary. Furthermore, a dictionary can be properly trained such that its prototype signal-atoms are more adequate to sparsely represent objective signals than wavelet. A number of researches have shown the benefits of using dictionaries for sparse representation; see, e.g. [46, 62].

In this work, we train a dictionary by applying K-SVD algorithm to a database consisting of patches extracted from images acquired from the same sequence but perhaps different subjects. Then the trained dictionary A, as shown in Figure 1-1, is used to reconstruct other MR images under the same acquisition sequence with similar structures. The database used for training consists of 4096 patches extracted from four MR brain images (but excludes the image to be reconstructed). Each block represents an atom of size 8×8. Atoms are sorted by ascending the variances of intensities. Details on the training and reconstruction processes are provided in the following sections. Comparison of the accuracy of sparse representations using wavelet transform and the trained dictionary A is shown in Figure 1-2. Results by using dictionary has has better preserved edges and fine structures because dictionary absorbed prior knowledge by learning features of this type of images during the training process [2].

Moreover, the dictionary training and representation processes are stable and robust as shown in our experimental results in section 1.4.

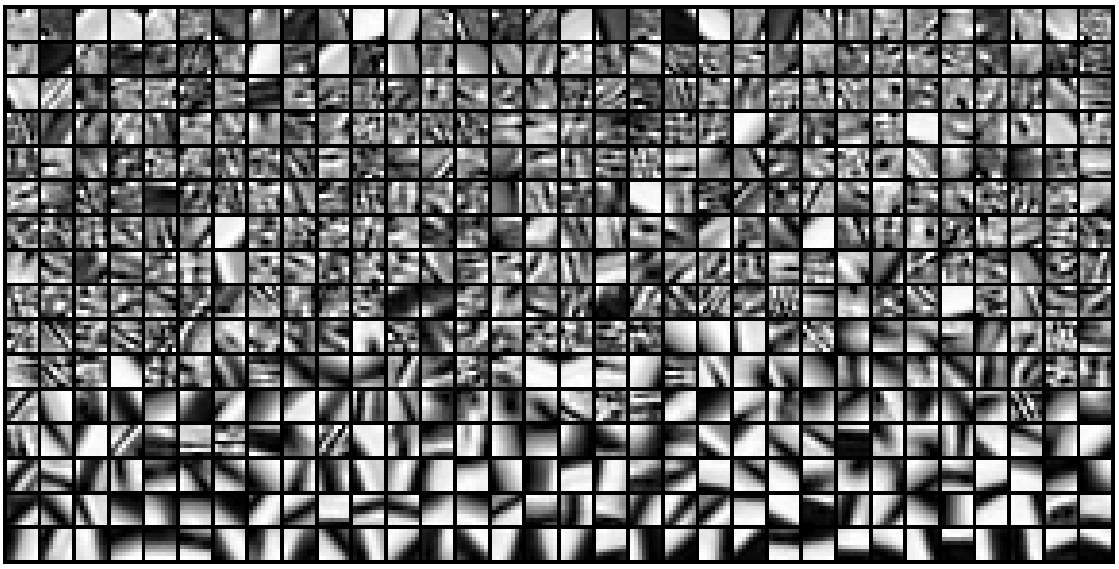


Figure 1-1. Dictionary trained by K-SVD algorithm.

In both sparse representations by wavelet and trained dictionary in Figure 1-2, images are represented by picking up the largest 12.5% transform coefficients. Bottom images are corresponding zoomed-in square area shown on the top left image. Left column: reference. Middle column: representation by wavelet, with RMSE 5.74% and SNR 22.3. Right column: representation by trained dictionary A shown in Figure 1-1, with RMSE 4.36% and SNR 24.7.

1.1.2 Likelihood Estimate as Data Fidelity Measure

To improve the robustness to the parameter selection, we use the likelihood estimation of the reconstruction error as the data fidelity measure. The reconstruction error is the difference between the partially acquired data and the Fourier transform of the reconstruction at sampled k-space locations, i.e. $\mathcal{F}_p u - f_p$ in (1–1). In previously proposed CS-MRI algorithms, least squares, i.e. the sum of squared difference (SSD) $\|\mathcal{F}_p u - f_p\|_2^2$, is a commonly used data consistency measure. For instance,

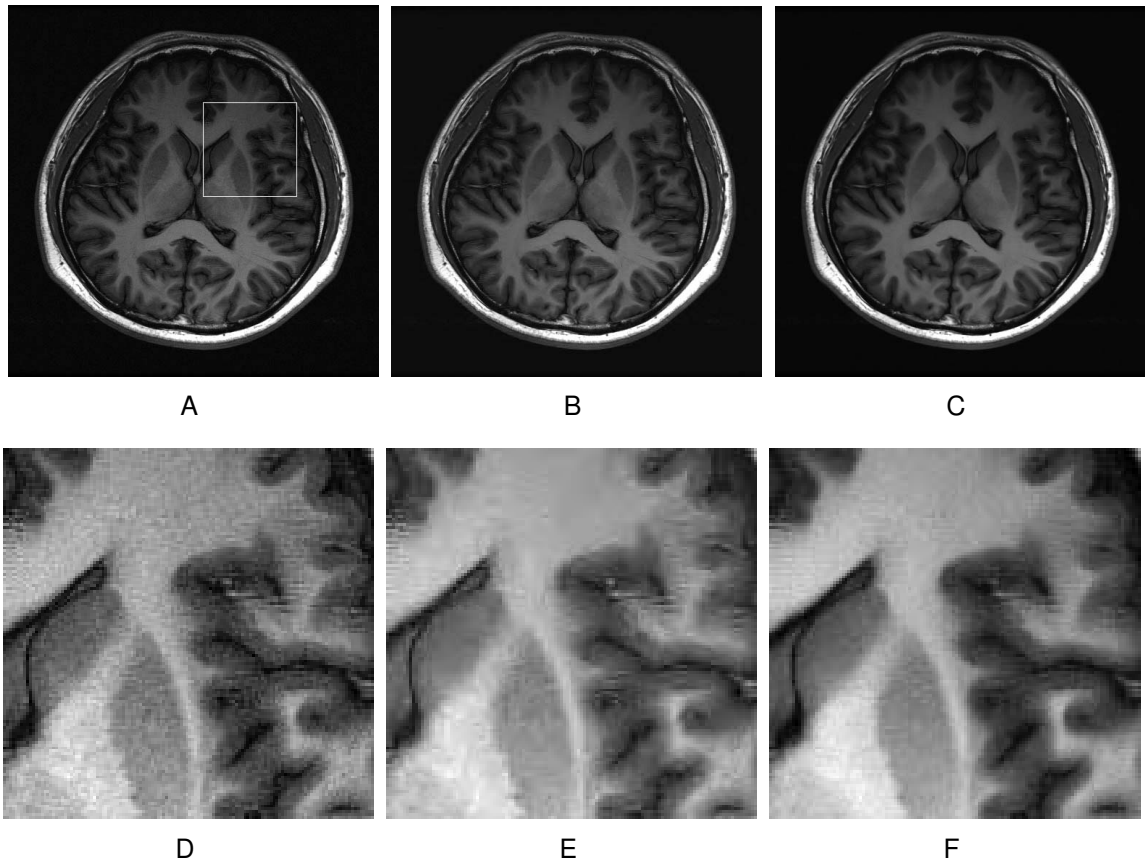


Figure 1-2. Compare the accuracy of sparse representations by wavelet and trained dictionary.

the unconstrained version of model (1–1) solves for the reconstruction by

$$\min_u TV(u) + \mu\|\Psi^\top u\|_1 + \frac{\lambda}{2}\|\mathcal{F}_p u - f_p\|_2^2, \tag{1–2}$$

where the parameter λ is crucial to the reconstruction results: an improperly large weight for the data fidelity term results in serious residual artifacts, whereas an improperly small weight results in damaged edges and/or fine structures. In this work, we tackle this problem by treating the reconstruction errors at all pixels as samples independently drawn from a Gaussian distribution with mean zero and variance σ^2 to be optimized. By maximum likelihood estimate (MLE) approach, the weight on $\|\mathcal{F}_p u - f_p\|_2^2/2$ becomes λ/σ^2 rather than a prescribed λ. Since σ is to be optimized, it is updated during iterations (in fact, it is the standard deviation of the reconstruction

error, see (1–17) below). When the reconstruction error reduces, the weight λ/σ^2 on $\|\mathcal{F}_p u - f_p\|_2^2$ increases, and hence the accuracy is improved. This automatically optimized weighting feature makes the choice of λ much more flexible.

1.1.3 Fast Numerical Algorithms for Solving CS-MRI Models

Despite that dictionaries are more adequate in signal/image reconstructions, the computational cost is higher than that using wavelet transform due to the redundancy of dictionaries and overlapping of patches to be represented. Also, the non-differentiability of TV and ℓ_1 norms brings difficulties to fast solutions of CS-MRI models. There have been many numerical algorithms developed to solve TV and ℓ_1 regularized minimization problems, more recent developments can be found in [19, 34, 64, 67, 78, 79] and [11, 14, 36, 41, 52, 66] and references therein. Our approach in this work is closely related to the algorithm developed in [67], in which Yang *et al.* introduced a simple and fast method to solve model (1–2) with isotropic TV norm of u defined by

$$TV(u) \triangleq \sum_{i=1}^{N} \|D_i u\|_2. \tag{1–3}$$

In (1–3) $u \in \mathbb{R}^N$ is the vector formed by stacking all columns of the image vertically, N is the total number of pixels in the image, and $D_i \in \mathbb{R}^{2\times N}$ represents the gradient operator at the i-th element in the vector of u. To overcome the non-differentiability of TV and ℓ_1 norms, they introduced auxiliary variables and used a classical quadratic penalty method which yields an alternating minimization scheme. By diagonalizing the gradient operator using Fourier transform, they made the main computation of the algorithm involving only soft shrinkage and fast Fourier transform. However, their algorithm cannot be directly applied to the models using dictionary A since it requires the orthogonality of Ψ in (1–2). Therefore the development of efficient algorithms involving the use of a dictionary is still a remaining problem. In this paper, we show that a simple trick on selecting patches from the underlying image can be used to overcome this difficulty.

Based on the method in [67] and this trick, we provide a simple and fast numerical algorithm that can be applied to reconstruction models involving dictionaries.

1.1.4 Organization

The rest of this paper is organized as follows. A detailed description of the proposed model is given in 4.2. In section 4.4, a fast algorithm to solve the proposed model and its derivation are provided. Experimental results on synthetic and in vivo data are presented in section 1.4. The last section concludes the paper.

1.2 Proposed Model

Before going into details of the proposed model, we address the notations used throughout the rest of the paper. First of all, all vectors in this paper are column vectors. Let $u \in \mathbb{R}^N$ be the underlying reconstruction as in (1–3), and $\mathcal{F}$ be the discrete Fourier transform, which can be treated as an $N \times N$ unitary matrix. Let $P \in \mathbb{R}^{p\times N}$ denote the binary matrix that selects certain rows of $\mathcal{F}$ according to the k-space sampling pattern. Then $\mathcal{F}_p \triangleq P\mathcal{F}$ is the undersampled Fourier transform. Let $f_p \in \mathbb{C}^p$ be the partially acquired k-space data and use $\|\cdot\|$ to denote Euclidean norm $\|\cdot\|_2$ of vectors henceforth. The notation $(\cdot\,;\cdot)$ represents a matrix formed by stacking its arguments vertically.

In this paper, all patches have size $\sqrt{n}\times\sqrt{n}$ and are often to be treated as n-vectors unless otherwise noted (in our experiments $n = 64$). The dictionary $A \in \mathbb{R}^{n\times K}$ consists of K n-vectors as atoms. Binary matrix $R_j \in \mathbb{R}^{n\times N}$ extracts the j-th patch of u, and forms the patch $R_j u$ as an n-vector. All patches $\{R_j u\}_{j=1}^J$ cover the entire image u and may overlap.

1.2.1 Sparse Representation Using Trained Dictionary

To improve the accuracy of reconstruction, especially the ability of preserving diagnostic information, we here propose to use a trained dictionary instead of wavelet as the sparsifying transform in MR image reconstructions. We chose the recently proposed dictionary design method, termed as K-SVD algorithm, to perform the dictionary

training. K-SVD is an iterative method that alternates between sparse coding of the examples based on the current dictionary and a process of updating the dictionary atoms to better fit the given database. The output is a trained dictionary that can represent all signals in the database under strict sparsity constraints and error tolerance. Interested readers are referred to [2, 30] for details.

Our procedure of forming a database and applying K-SVD algorithm to train an adaptive dictionary for MR image reconstruction is depicted as follows.

1. Collect a number of MR images acquired using the same sequence as that for the image to be reconstructed, but from different subjects. The training images and the image to be reconstructed are preferred to be the same body parts to get a better sparse representation. Using the same acquisition sequence ensures that they have similar contrast.

2. Decompose the training images to $\sqrt{n} \times \sqrt{n}$ patches, and discard those patches with constant intensity. Then randomly choose $8K$ patches from the remaining patches, where K is the number of atoms in the dictionary to be trained.

3. Train a dictionary A by applying K-SVD algorithm to that $8K$ patches with the overcomplete DCT matrix as the initial. The resulting trained dictionary has K elements, i.e. $A \in \mathbb{R}^{n \times K}$.

In our experiments, we set n to 64 and K to $512/256$ for brain/chest MR data. The dictionary trained for brain image reconstruction is illustrated in Figure 1-1. The dictionary A we obtain from this training procedure can adequately represent any patches of brain MR images (e.g. from different subjects or the same subject in different periods) acquired under the same sequence. In particular, each patch $R_j u$ of u can be sparsely represented by A. Namely, there exist representation coefficients $\alpha_j \in \mathbb{R}^K$ such that

$$\|\alpha_j\|_0 < n < K \quad \text{s.t.} \quad A\alpha_j \approx R_j u, \quad j = 1, \cdots, J,$$

where $\|\cdot\|_0$ counts the number of nonzero components of its argument. Therefore, the sparsity of u under the representation of A can be used as a regularization in the

reconstruction. That is, we enforce the following into our model

$$\min_{\alpha} \sum_{j=1}^{J} \left(\|\alpha_j\|_1 + \frac{\nu}{2} \|A\alpha_j - R_j u\|^2 \right), \tag{1–4}$$

where $\alpha = (\alpha_1; \cdots; \alpha_J) \in \mathbb{R}^{KJ}$. This is in fact the relaxed form of sparse-land problem with ℓ_0-norm substituted by ℓ_1-norm. The reason why we use ℓ_1 instead of ℓ_0 is that minimizing the non-convex ℓ_0 is generally a NP-hard problem and hence is not tractable in practice. Moreover, it has been proved that minimization problems with ℓ_1 and ℓ_0 share the same solution under certain conditions [18, 26].

Note that if $J = 1$, $R_1 = I$ (the identity matrix) and $A = \Psi$ (the wavelet transform), then (1–4) reduces to $\|\Psi^\top u\|_1$ as in (1–2) when the difference in the quadratic is exactly 0. Namely, wavelet is a special case of dictionary.

1.2.2 Likelihood Estimate for the Data Consistency

One difficulty of applying the unconstrained energy minimization problem (1–2) for MR image reconstruction is in determining the weighting parameter that balances the data consistency and image sparsity. The reconstruction results are sensitive to the choice of this parameter. To tackle this problem, we derive the data consistency measure, the so-called data fidelity, from maximum likelihood estimate (MLE) approach.

Let $\omega = (\omega_1, \cdots, \omega_p)^\top \in \mathbb{C}^p$ be the reconstruction error in k-space, which is the difference between the Fourier transform of the reconstruction u and the acquired data f_p at the sampled k-space locations:

$$f_p = \mathcal{F}_p u + \omega.$$

Consider $\{\omega_l\}_{l=1}^p$ as independent random samples drawn from a normal distribution of mean zero and variance σ^2 to be determined. Therefore, the joint probability density function (pdf) of $\{\omega_l\}_{l=1}^p$, which is also the likelihood of σ, becomes

$$\mathcal{L}(\sigma|\omega) = \prod_{l=1}^{p} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-\omega_l^2/2\sigma^2} \right) = \left(2\pi\sigma^2\right)^{-p/2} e^{-\|\omega\|^2/2\sigma^2}.$$

Thus, the negative log-likelihood is

$$-\log \mathcal{L}(\sigma|\omega) = \|\omega\|^2/2\sigma^2 + p\log\sqrt{2\pi}\sigma. \tag{1–5}$$

Substituting ω by $\mathcal{F}_p u - f_p$, and omitting the constant $p\log\sqrt{2\pi}$, we obtain a MLE based consistency estimation with the partially acquired data:

$$F(u,\sigma,f_p) = \|\mathcal{F}_p u - f_p\|^2/2\sigma^2 + p\log\sigma. \tag{1–6}$$

This is a generalization of the least square estimation, which is just the case where $\sigma \equiv 1$. We will use (1–6) as data fidelity term in our energy functional.

1.2.3 Variational Model for MR Image Reconstruction from Undersampled Data

Now we are ready to present our model. We propose to use TV and sparse representation by trained dictionary as regularization and MLE (1–6) as data consistency measure. Our model is formulated as an unconstrained minimization problem

$$\min_{u,\alpha,\sigma} TV(u) + \mu\sum_{j=1}^{J}\left(\|\alpha_j\|_1 + \frac{\nu}{2}\|A\alpha_j - R_j u\|^2\right) + \lambda F(u,\sigma,f_p), \tag{1–7}$$

where $TV(u)$ is the TV norm of u defined as in (1–3), the summation over α is the regularization of u using the sparsity under representation by dictionary A, and $F(u,\sigma,f_p)$ is MLE data consistency measure (1–6). By using MLE based approach, σ is also optimized along with u. In (1–7) the weight on $\|\mathcal{F}_p u - f_p\|^2$ versus the sparsity of the underlying image is λ/σ^2 rather than λ only. In the Euler-Lagrange (EL) equations associated with the proposed energy function below, one can see that σ is the standard deviation of the reconstruction error ω. Hence, when the construction error ω decreases, the weight λ/σ^2 on minimizing ℓ_2 norm of ω increases automatically, which makes the choice of the initial weighting parameter λ more flexible. This flexibility dramatically reduces the difficulty of parameter decision and improves the applicability of CS. Moreover, our experimental results below show that this automatically updated weighting makes faster convergence, and better accuracy in reconstruction.

1.3 Algorithm

There minimization problem (1–7) is closely related to the well-known TV and ℓ_1 based signal/image reconstruction problems. Since the non-differentiability of TV and ℓ_1 terms bring a difficulty in computations, there have been a number of numerical algorithms developed to efficiently solve this type of problems. The algorithm provided in this section is inspired by the work in [64, 67], which uses the variable splitting and classical quadratic penalty technique in optimization to make the computation fast and stable.

1.3.1 A Fast Algorithm for Solving the Proposed Model

We first introduce two auxiliary variables $w = (w_1^\top; w_2^\top; \cdots; w_N^\top) \in \mathbb{R}^{N\times 2}$ and $\beta = (\beta_1; \beta_2; \cdots; \beta_J) \in \mathbb{R}^{KJ}$ where $w_i \in \mathbb{R}^2$ and $\beta_j \in \mathbb{R}^K$ for all $i = 1, \cdots, N$ and $j = 1, \cdots, J$. Then we consider the minimization problem equivalent to (1–7):

$$\begin{aligned} \min_{u,w,\alpha,\beta,\sigma} \quad & \sum_{i=1}^{N} \|w_i\| + \mu \sum_{j=1}^{J} \left(\|\beta_j\|_1 + \frac{\nu}{2} \|A\alpha_j - R_j u\|^2 \right) + \lambda F(u, \sigma, f_p) \\ \text{s.t.} \quad & w_i = D_i u, \ \beta_j = \alpha_j, \ \forall\, i = 1, \cdots, N, j = 1, \cdots, J. \end{aligned} \tag{1–8}$$

Relaxing the equality constraint and penalizing their violations by quadratic functions, we obtain an unconstrained version of (1–8):

$$\min_{u,w,\alpha,\beta,\sigma} \sum_{i=1}^{N} \phi(w_i, D_i u) + \mu\psi(\beta, \alpha) + \sum_{j=1}^{J} \frac{\mu\nu}{2} \|A\alpha_j - R_j u\|^2 + \lambda F(u, \sigma, f_p) \tag{1–9}$$

where functions ϕ and ψ are defined as

$$\phi(s, t) = \|s\| + \frac{\eta}{2}\|s - t\|^2, \quad s, t \in \mathbb{R}^2$$

and

$$\psi(s, t) = \|s\|_1 + \frac{\theta}{2}\|s - t\|^2, \quad s, t \in \mathbb{R}^{KJ}$$

for given $\eta, \theta > 0$. With η, θ gradually increasing, solving (1–9) lead to approximations to the solution of (1–8).

The minimization (1–9) can be carried out in a much faster and more stable manner than (1–7): first, for fixed u and α, the minimization with respect to w and β can be carried out in parallel:

$$w_i = \mathcal{S}_2(D_i u), \quad i = 1, \cdots, N, \tag{1–10}$$

where $\mathcal{S}_2(t)$ is the two-dimensional (2D) shrinkage that minimizes $\phi(s, t)$ for fixed t:

$$\mathcal{S}_2(t) \triangleq \max\left\{\|t\| - \frac{1}{\eta}, 0\right\} \cdot \frac{t}{\|t\|}, \quad t \in \mathbb{R}^2.$$

Moreover, we have

$$\beta = \mathcal{S}_c(\alpha) \tag{1–11}$$

where $\mathcal{S}_c(t)$ is the componentwise shrinkage that minimizes $\psi(s, t)$ for fixed $t = (t_1, \cdots, t_{KJ})^\top \in \mathbb{R}^{KJ}$:

$$\mathcal{S}_c(t) = (\mathcal{S}(t_1), \cdots, \mathcal{S}(t_{KJ}))^\top$$

and

$$\mathcal{S}(x) = \max\{x - 1/\theta, 0\} \cdot \mathrm{sign}(x), \quad x \in \mathbb{R},$$

with assumption $0 \cdot (0/0) = 0$. Both computational costs for w and β are linear in N.

Secondly, for fixed u and β, we can have $\alpha = (\alpha_1; \cdots; \alpha_J)$ by solving the following minimization problem:

$$\min_{\alpha} \sum_{j=1}^{J} \left(\theta\|\alpha_j - \beta_j\|^2 + \nu\|A\alpha_j - R_j u\|^2\right). \tag{1–12}$$

The solution can be obtained by setting α_j as

$$\alpha_j = V(\theta I + \nu\Lambda)^{-1} V^\top \left(\theta\beta_j + \nu A^\top R_j u\right) \tag{1–13}$$

where the diagonal matrix Λ and orthogonal matrix V come from the eigendecomposition $A^\top A = V\Lambda V^\top$. This decomposition does not drag the computation since the dictionary

A is prepared before any experiments, and hence, V and Λ can be pre-computed. Also the largest dimension K of A is usually much less than N, and it can be seen that the number of nonzero eigenvalues can never exceed n. As a result, the computations of α_j's can be carried out in parallel, and each of them only involves matrix-vector multiplication.

Thirdly, for fixed w, α and σ, the minimization of u is

$$\min_u \|w_x - D_x u\|^2 + \|w_y - D_y u\|^2 + \sum_{j=1}^{J} \gamma \|A\alpha_j - R_j u\|^2 + \xi \|\mathcal{F}_p u - f_p\|^2, \tag{1–14}$$

Here D_x, D_y are N-square matrices formed by the top and bottom rows of $D_i \in \mathbb{R}^{2\times N}$, $i = 1, \cdots, N$, and hence $D_x u, D_y u$ represent the gradient of u along the x and y directions, respectively. w_x and w_y are the first and second column of w, respectively, and $\gamma = \mu\nu/\eta$, $\xi = \xi(\sigma) = \lambda/\eta\sigma^2$. Thus the Euler-Lagrange equation of (1–14) yields

$$Lu = r, \tag{1–15}$$

where

$$L = D_x^\top D_x + D_y^\top D_y + \sum_{j=1}^{J} \gamma R_j^\top R_j + \xi \mathcal{F}_p^\top \mathcal{F}_p$$

and

$$r = D_x^\top w_x + D_y^\top w_y + \sum_{j=1}^{J} \gamma R_j^\top A\alpha_j + \xi \mathcal{F}_p^\top f_p.$$

Under the periodic boundary condition for u, the finite difference operators D_x and D_y are block circulant matrices with circulant blocks and hence can be diagonalized by Fourier matrix $\mathcal{F}$. Thus, $\hat{D}_x = \mathcal{F} D_x \mathcal{F}^\top$ and $\hat{D}_y = \mathcal{F} D_y \mathcal{F}^\top$ are diagonal. Also, periodic boundary condition enables us to extract patches that cover each pixel m times, where $m = n/d^2$ and d is the sliding distance between all concatenated patches. Usually, we fix the patch size as 8×8, i.e. $n = 64$, and choose $(d, m) = (8, 1)$ or $(4, 4)$. Since $\sum_j R_j^\top R_j$ is a diagonal matrix with the i-th diagonal entry counting the number of times the i-th

pixel was covered by patches, we have $\sum_j R_j^\top R_j = mI$, where I is the identity matrix. So multiplying $\mathcal{F}$ on both sides of (1–15) gives

$$\hat{L}\mathcal{F}(u) = \hat{r}, \tag{1–16}$$

where

$$\hat{L} = \hat{D}_x^\top \hat{D}_x + \hat{D}_y^\top \hat{D}_y + m\gamma I + \xi P^\top P$$

is a diagonal matrix since $P^\top P$ is diagonal, and

$$\hat{r} = \hat{D}_x^\top \mathcal{F}(w_x) + \hat{D}_y^\top \mathcal{F}(w_y) + \gamma \mathcal{F}(u_\alpha) + \xi P^\top f_p$$

where $u_\alpha = \sum_j R_j^\top A\alpha_j$ is an "image" assembled using patches that are represented by dictionary A and α.

Finally, the computation of first variation of $F(u, \sigma, f_p)$ gives an update of σ in each iteration:

$$\sigma = \sqrt{\|\mathcal{F}_p u - f_p\|^2 / p}. \tag{1–17}$$

Therefore, similar to the algorithm developed in [67], the main computation of our algorithm for the CS model using a dictionary also involves only shrinkage, matrix-vector multiplication and fast Fourier transform.

1.3.2 Numerical Algorithm and Convergence Analysis

Based on the discussion above, we are ready to propose the fast algorithm used for solving model (1–7). Note that the solution to (1–7) can be approximated by solving (1–9) with continuation on the penalty parameters η and θ [36]. For stopping criterion, we let "res" be the maximum absolute/norm value of increments of w, α, β, u, and terminate each inner loop once $\text{res} < \epsilon$ for a prescribed error tolerance ϵ, then update u and start a new loop with doubled η and θ. The upper bound 2^{12} for η, θ is chosen empirically so it is sufficiently large that solutions to (1–9) is a close approximation

of (1–7). Based on derivations above, we summarize the algorithm for our model as Algorithm 1.

Algorithm 1 MR Image Reconstruction via Sparse Representation (recMRI)

Input P, f_p, and $\mu, \nu, \lambda, \epsilon > 0$. Initialize $u = \mathcal{F}_p^\top f_p$, $\eta = \theta = 2^6$ and $\alpha = 0$.
while $\eta, \theta < 2^{12}$ **do**
  repeat
    Given u and α, compute w and β using (1–10) and (1–11).
    for $j = 1$ to J **do**
      Given u and β, compute α_j using (1–13).
    end for
    Given w and α, compute u by solving (1–16) and update σ by (1–17).
  until res $< \epsilon$
  return $u^{\eta,\theta}$
  $u \leftarrow u^{\eta,\theta}$, $(\eta, \theta) \leftarrow (2\eta, 2\theta)$
end while

The proof of the convergence of the proposed algorithm 5 is similar to the one given in [64] with slight modifications, and thus is omitted here.

1.4 Experimental Results

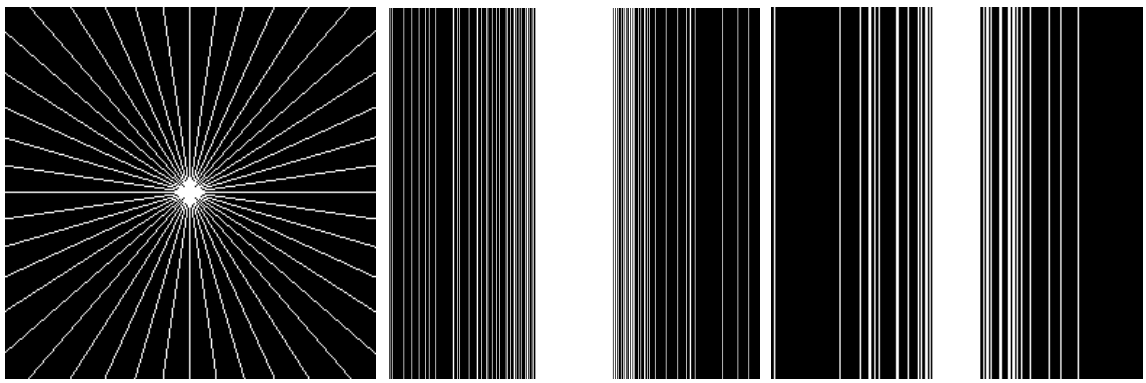
In this section, we present the experimental results of the proposed model (1–7) using Algorithm 5 and the comparisons with that resulting from using wavelet transform on both synthetic and in vivo MR data. All implementations involved in the experiments were coded in MATLAB®v7.3 (R2006b), except the shrinkage and wavelet transform operators, which were coded in C++. Computations were performed on a GNU/Linux (version 2.6.16) workstation with Intel®Core 2 CPU at 1.86GHz and 2GB memory.

1.4.1 Improvement on Accuracy by Using Dictionaries

To show the improvement on the accuracy of reconstructions by using dictionaries as sparsifying transforms, we applied model (1–2), which uses wavelet as the sparsifying transform, and the proposed model (1–7) to three data sets. The three data sets are: the default Shepp-Logan phantom image provided by MATLAB®, a 2D axial brain MR image and a 2D chest MR image. The sampling masks used for these three data sets are depicted in Figure 3-1, where white pixels indicate sampled locations

in k-space. In Figure 3-1, the masks are for (a) phantom (b) brain image and (c) chest image, respectively. We used pseudo radial mask for phantom, and Cartesian mask undersampling phase encoding (PE) lines for in vivo data to simulate random acquisition. All of the k-space data in the simulated pseudo-radial trajectory is located on Cartesian grid.

In practice, CS-MRI algorithm prefers random acquisition trajectory that can lead to incoherent artifacts aliasing. However, the trajectory of the acquired data in each echo time (TE) is limited by the MR system, and hence true random acquisition is not possible. In recent years, acquisition schemes that are feasible and can produce incoherent aliasing artifacts are developed, e.g. radial and spiral trajectories. In our experiments, for simplicity, we used pseudo sampling masks which can simulate randomness in acquisition for demonstration purpose.



A SR=8.4% B SR=34.0% C SR=25.0%

Figure 1-3. Sampling masks used with sampling ratios (SR).

1.4.1.1 Results of Phantom Reconstruction

The default Shepp-Logan phantom of size 256×256 is shown in Figure 1-4A. Then a full k-space data was simulated by the 2D Fast Fourier transform (`fft2` in MATLAB) of the phantom. We used the pseudo radial mask (shown in Figure 1-3A) to the full k-space data and added complex valued Gaussian random noise with standard deviation (of magnitude) 3.2 to simulate the partially acquired data f_p. Direct using

FFT of zero filling unscanned k-space locations results in notorious artifact aliasing, as shown in Figure 1-4B.

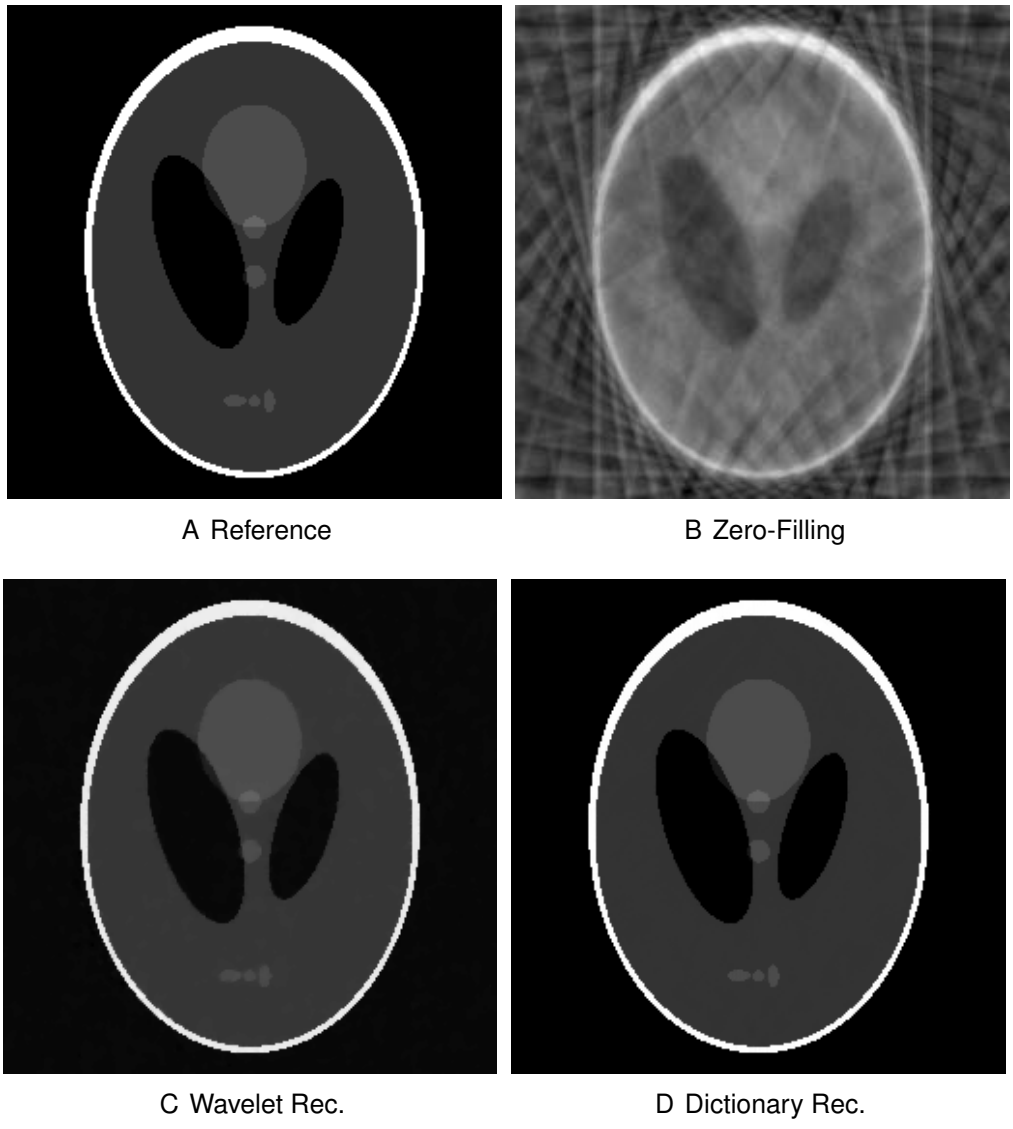
Then we applied model (1–2) with Haar wavelet and model (1–7) with an overcomplete discrete cosine transform (DCT) consisting of 256 atoms of size 8×8 [2] as the dictionary to the partial data f_p. The parameters we used for both models were $(\mu, \lambda, \epsilon) = (1, 10^3, 10^{-3})$ and the parameter ν was set to 1. Figure 1-4 shows the following: (a) Reference image. (b) Zero-filling. (c) Reconstruction by model (1–2) with Haar wavelet (d) Reconstruction using overcomplete DCT dictionary. The results by using wavelet and dictionary have corresponding RMSEs are 2.47% and 2.18%, and SNR are 32.9 and 34.5, respectively.

1.4.1.2 Results of Brain Image Reconstruction

The second test is on an axial brain MR image. The 2-dimentional multi-slice data set was collected on a 3T GE system (GE Healthcare, Waukesha, Wisconsin, USA) using the T1 FLAIR sequence (FOV 220mm, matrix size 512×512, TR 3060ms, TE 126ms, flip angle 90°, slice thickness 5mm, number of averages 1) with an 8-channel head coil (Invivo Corporation, Gainesville, FL, USA). Phase encoding direction was anterior-posterior. The imaging process output a high resolution and SNR image of size 512×512, which were used as the reference image in our experiment, as shown in Figure 1-5A.

The simulated full k-space data in our experiment was obtained by the Fourier transform of this reference image, and then was artificially undersampled using the Cartesian mask shown in Figure 1-3B, which led to the partial k-space data f_p. The zoomed-in of the square area in Figure 1-5A is shown in Figure 1-5B.

Figure 1-5 shows the following: (a) Reference. (b) Zoomed-in of square area in the reference image. (c) Zoomed-in of reconstruction by zero-filling unscanned k-space locations. (d) Zoomed-in of low resolution (LR) image reconstructed by 34.0% central PE lines, with RMSE 10.32% and SNR 17.2. (e) Zoomed-in of reconstruction obtained



A Reference B Zero-Filling

C Wavelet Rec. D Dictionary Rec.

Figure 1-4. Reconstructed phantom image from simulated partial k-space data.

using wavelet as sparsifying transform, RMSE is 8.52% and SNR is 20.7. (f) Zoomed-in of reconstruction obtained using trained dictionary shown in Figure 1-1, RMSE is 7.74% and SNR is 22.0.

With only 34.0% data for reconstruction, strong aliasing artifacts can be observed in the image reconstructed by zero-filling unscanned k-space locations, and the zoomed-in is shown in Figure 1-5C.

In this experiment the database used for training a dictionary consists of 4096 8×8 patches extracted randomly from four 2D brain MR images of different normal subjects (excluding the one to be reconstructed one) with the same acquisition sequence. The trained dictionary $A \in \mathbb{R}^{64\times 512}$, as shown in Figure 1-1, consists of 512 atoms of size 8×8.

We applied model (1–2) with Haar wavelet and model (1–7) with this trained dictionary to the partial Fourier data f_p for brain MR image reconstruction. The parameters λ, μ and ϵ were set to 2e+3, 1 and 5e-4 in model (1–2) and (1–7), respectively. The parameter ν in model (1–7) was set to 10^6. The zoomed-in area of the reconstructed images by model (1–2) and proposed model (1–7) are shown in Figure 1-5E and 1-5F, respectively. The RMSEs of reconstructions are 8.52% for model (1–2) and 7.74% for proposed model (1–7), and SNRs are 20.7 and 22.0, respectively. It can be seen that the image reconstructed by model (1–2) has oil-painting effect. On the contrary, the image reconstructed by the proposed model has better preserved fine structures. This further confirms the higher accuracy obtained by the proposed method.

We also simulated a low resolution (LR) image by using the 34.0% central PE lines (i.e. all white vertically lines in the middle in Figure 1-3B), which has RMSE 10.32% and SNR 17.2.

1.4.1.3 Results of Chest Image Reconstruction

We also validate the proposed method on chest MR images. In this experiment the dictionary was trained by slices extracted from a three-dimensional (3D) MR chest data set, that consists of 19 adjacent 2D image slices of size 256×256 near the thorax region. Our procedure of training a dictionary is as follows: we randomly chose four slices and decomposed them into non-overlapping patches of size 8×8, and discarded those patches with homogeneous intensities, and then use K-SVD on the remaining patches to train a dictionary $A \in \mathbb{R}^{8^2\times 256}$ with 256 atoms. It is worth noting that if additional data (e.g. chest MR data scanned from different subjects but the same sequence, which

are usually available in clinical applications) were available, one can readily construct a dictionary that is comparable to the one we trained using adjacent slices, and obtain similar reconstruction results as we showed below.

To demonstrate the improved accuracy by using dictionary, we randomly chose a 2D slice shown in Figure 1-6A, which is different from the four used as training slices. Then we artificially downsampled its k-space data using a Cartesian mask with 25% sampling ratio, as shown in Figure 1-3C. Zero-filling the unscanned k-space locations results in severe artifacts as shown in Figure 1-6B with RMSE 15.59%. We again simulated a low resolution (LR) image, as shown in Figure 1-6D, by using the 25.0% central PE lines, which has RMSE 14.44% and SNR 15.4. From the corresponding error map Figure 1-6G, i.e. the absolute difference to the reference image, we can see the potential loss of edges and diagnostic information.

The reconstructions performed by using model (1–2) with Haar wavelet and model (1–7) with the trained dictionary are shown in Figure 1-6E and 1-6F, and the corresponding RMSEs are 12.04% and 8.48%, and SNRs are 17.3 and 20.1. In both cases λ, μ and ϵ were set to 1e+4, 2.5 and 1e-5, respectively. Parameter ν in model (1–7) was set to 10^5. The error maps of these two reconstructions are shown in Figure 1-6H and 1-6I, respectively. It can be seen that the image reconstructed by the proposed model (1–7) has lower artifacts level and better preserved edges. This experiment demonstrates again the advantages of using prior information to define the sparsifying transformation, which results in higher accuracy of reconstructions.

In Figure 1-6: (a) Reference image. (b) Zero-filling unscanned locations, RMSE is 15.59% and SNR is 14.2. (d) LR image obtained by using central PE lines, RMSE is 14.44% and SNR is 15.4. (e) Reconstruction by model (1–2) with Haar wavelet, RMSE is 12.09% and SNR is 17.1. (f) Reconstruction by proposed model (1–7) with trained dictionary, RMSE is 8.48% and SNR is 20.1. Figures (g), (h) and (i) are corresponding error maps of (d), (e) and (f) to the reference image (a).

1.4.2 Improvement on Robustness (to Parameter Selection) and Efficiency

To demonstrate the improvement on the robustness of the proposed model (1–7) with respect to the choice of parameter λ, we tested the reconstruction of the Shepp-Logan phantom using model (1–2) with SSD and the proposed model (1–7) using MLE as as data consistency measures on various choices of λ. The resulting RMSEs are shown in Table 1-1. From the changes of RMSEs, we can see the model with MLE as data consistency measure generated similarly good results whereas the model with SSD failed when λ went improperly large. This result shows that the proposed model with MLE data consistency measure is much less sensitive to the choice of λ and hence makes the reconstruction more robust.

Table 1-1 also shows the CPU time (in seconds) for phantom reconstruction using three different algorithms: nonlinear conjugate gradient (CG) algorithm for model (1–2), algorithm 5 recMRI for model (1–7) with the term involving dictionary

$$\sum_{j=1}^{J}\left(\|\alpha_j\|_1+\frac{\nu}{2}\|A\alpha_j-R_ju\|^2\right)$$

replaced by Haar wavelet term $\|\Psi^\top u\|_1$, and algorithm 5 recMRI for model (1–7) with overcomplete DCT as dictionary. Here the wavelet transforms were generated using optimized DWT package for MATLAB®It can be seen that the proposed numerical method was over 2.6 times faster than conjugate gradient based method. The dictionary based sparse representation consistently produced images with lower RMSE than wavelet based method, but it takes longer reconstruction time due to the redundancy of dictionaries. Moreover, using optimized discrete wavelet transform (DWT) package in MATLAB®makes the computation for wavelet based model even faster. Therefore, we expect an improvement on speed by code optimization when using dictionaries.

1.4.3 Robustness of Dictionary Training and Reconstruction

In the experiment on brain MR image reconstruction in section 1.4.1.2, the patches used by K-SVD algorithm were randomly extracted from the four training images.

Table 1-1. Comparison of results of phantom reconstructions using nonlinear conjugate gradient (CG) for model (1–2) with Haar wavelet, algorithm recMRI with Haar wavelet, and recMRI for model (1–7) with overcomplete DCT dictionary.

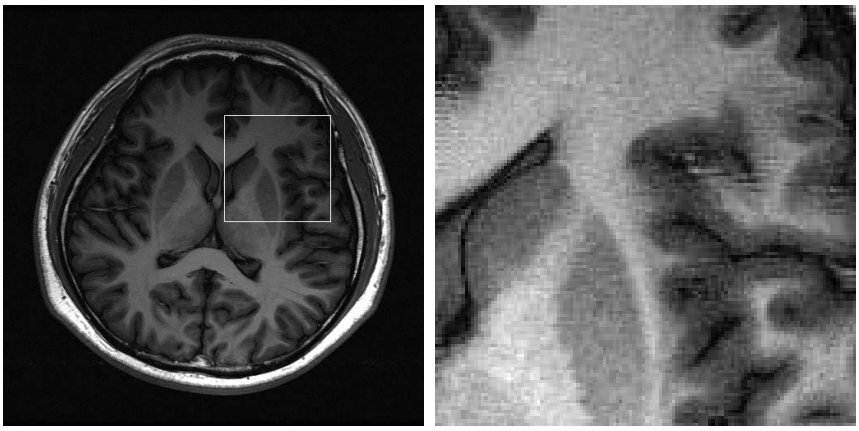
Method	CG(Wavelet*)			recMRI(Wavelet*)			recMRI(Dictionary)		
λ	RMSE	Obj	CPU	RMSE	Obj	CPU	RMSE	Obj	CPU
1e+2	5.93%	12.13	86.3	7.93%	13.98	28.2	5.21%	11.79	211
1e+3	2.47%	11.92	71.6	2.52%	12.11	27.7	2.18%	10.92	199
1e+4	5.05%	3.147	71.4	4.98%	3.025	26.9	3.47%	2.540	198
1e+5	25.9%	2.271	87.1	5.93%	1.375	27.0	3.67%	1.116	201
1e+6	37.0%	2.165	81.2	6.16%	1.129	28.7	5.52%	1.091	212

Different patches may lead to different trained dictionary using K-SVD algorithm, and hence may impact the consequent reconstructions. Therefore it is important to verify that the dictionary training and reconstruction process are robust to certain level of variability on the training database used in K-SVD algorithm. In this experiment, we repeated 10 times of the entire process from forming a data set of training image to using trained dictionary in brain MR image reconstruction as described in section 1.4.1.2. The difference is that, in each run, the 4096 training images are randomly chosen from a pool of patches (around $50,000$) extracted from images acquired under the same sequence as that used for the image to be reconstructed. Therefore, the training patches used in one run are different from those in another. The RMSEs and SNRs of reconstruction results are shown in Table 1-2. Meanwhile, when we directly

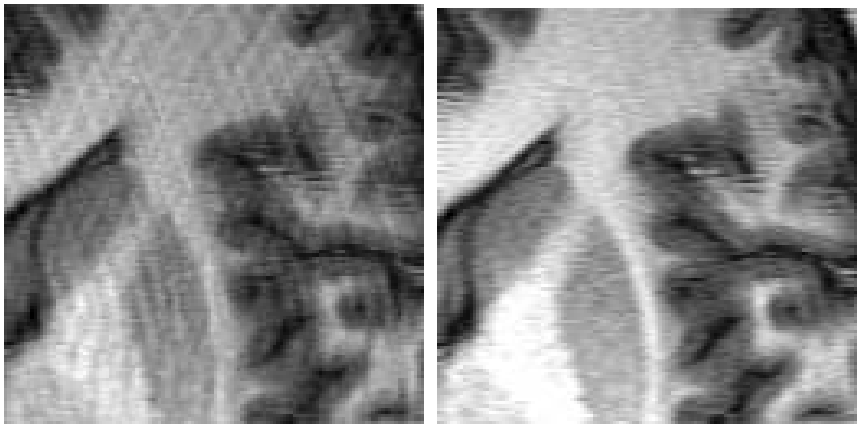
Table 1-2. Experimental results of 10 runs of dictionary training and brain image reconstruction as in 1.4.1.2.

Runs	1	2	3	4	5	6	7	8	9	10
RMSE(%)	7.74	7.77	7.65	7.81	7.83	7.78	7.79	7.79	7.72	7.68
SNR	22.0	21.8	22.6	21.6	21.6	21.8	21.7	21.7	22.1	22.5

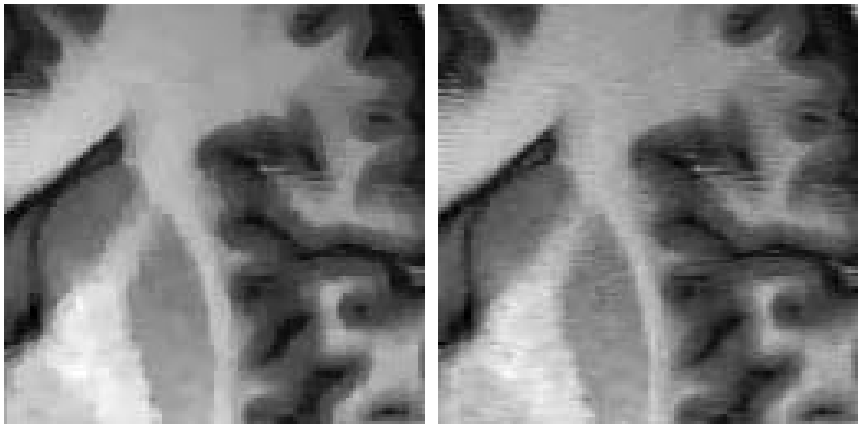
used the overcomplete DCT instead of the trained dictionary in the reconstruction (1–7), the reconstruction had RMSE 8.27% and SNR 21.1. Table 1-2 indicates that the K-SVD algorithm and the consequent reconstructions using trained dictionaries can consistently generate good results despite that the training patches may vary. Therefore, the proposed scheme using dictionaries trained by K-SVD algorithm in MR image reconstruction is stable and robust, and hence has great practical potential.



A Reference B Zoomed-In Reference



C Zoomed-In Zero-Filling D Zoomed-In LR



E Zoomed-In Wavelet F Zoomed-In Dictionary

Figure 1-5. Reconstruction of brain MR Image.

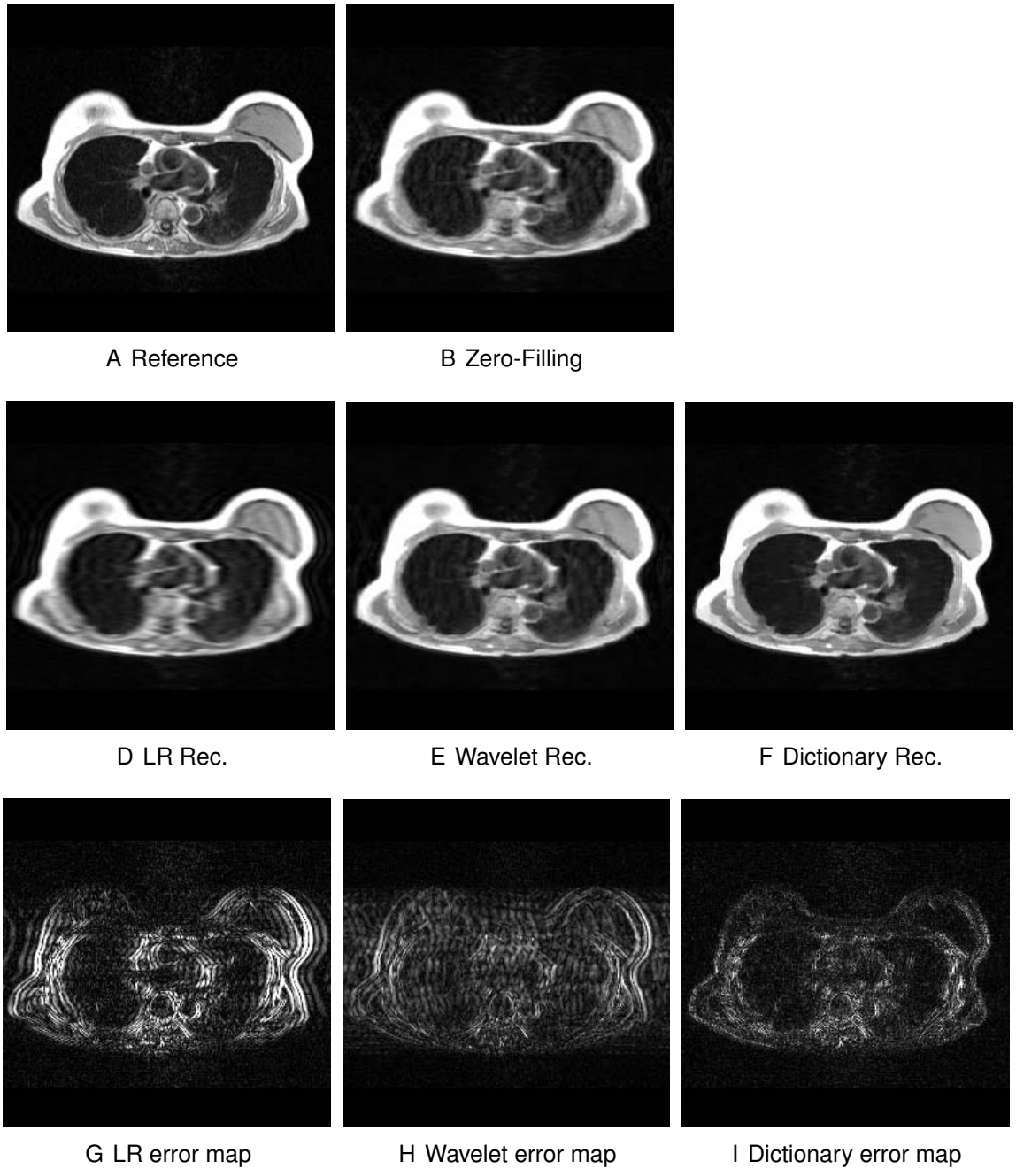


Figure 1-6. Reconstructions of chest MR image using model (1–2) and (1–7).

CHAPTER 2
COMPUTATIONAL ACCELERATION FOR MR IMAGE RECONSTRUCTION IN PARTIALLY PARALLEL IMAGING

Outline

In this paper, we present a fast numerical algorithm for solving total variation and ℓ_1 (TVL1) based image reconstruction with application in partially parallel MR imaging. Our algorithm uses variable splitting method to reduce computational cost. Moreover, the Barzilai-Borwein step size selection method is adopted in our algorithm for much faster convergence. Experimental results on clinical partially parallel imaging data demonstrate that the proposed algorithm requires much fewer iterations and/or less computational cost than recently developed operator splitting and Bregman operator splitting methods, which can deal with a general sensing matrix in reconstruction framework, to get similar or even better quality of reconstructed images.

2.1 Backgrounds in Total Variation Based Image Reconstruction

In this paper we develop a novel algorithm to accelerate the computation of total variation (TV) and/or ℓ_1 based image reconstruction. The general form of such problems is

$$\min_u \left\{ \alpha \|u\|_{TV} + \beta \|\Psi^\top u\|_1 + \frac{1}{2}\|Au - f\|^2 \right\}, \tag{2–1}$$

where $\|\cdot\|_{TV}$ is the total variation, $\|\cdot\|_1$ and $\|\cdot\| \equiv \|\cdot\|_2$ are the ℓ_1 and ℓ_2 norms (Euclidean norms), respectively. For notation simplicity we only consider two dimensional (2D) images in this paper, whereas the method can be easily extended to higher dimensional cases. Following the standard treatment we will vectorize an (2D) image u into one-dimensional column vector, i.e. $u \in \mathbb{C}^N$ where N is the total number of pixels in u. Then, the (isotropic) TV norm is defined by

$$\|u\|_{TV} = \int_\Omega |Du| = \sum_{i=1}^{N} \|D_i u\| \tag{2–2}$$

where for each $i = 1, \cdots, N$, $D_i \in \mathbb{R}^{2\times N}$ has two nonzero entries in each row corresponding to finite difference approximations to partial derivatives of u at the i-th pixel along the coordinate axes. In (2–1), $\alpha, \beta \geq 0$ ($\alpha + \beta > 0$) are parameters corresponding to the relative weights of the data fidelity term $\|Au - f\|^2$ and the terms $\|u\|_{TV}$ and $\|\Psi u\|_1$. Model (2–1) has been widely applied to image reconstruction problems. Solving (2–1) yields a restored clean image u from an observed noisy or blurred image f when $A = I$ or a blurring matrix, respectively. In compressive sensing (CS) applications, A is usually a large and ill-conditioned matrix depending on imaging devices or data acquisition patterns, and f represents the under-sampled data. In CS $\Psi = [\psi_1, \cdots, \psi_N] \in \mathbb{C}^{N\times N}$ is usually a proper orthogonal matrix (e.g. wavelet) that sparsifies underlying image u.

2.1.1 Partially Parallel MR Imaging

The CS reconstruction via TVL1 minimization (2–1) has been successfully applied to an emerging MR imaging application known as partially parallel imaging (PPI). PPI uses multiple RF coil arrays with separate receiver channel for each RF coil. A set of multi-channel k-space data from each radiofrequency (RF) coil array is acquired simultaneously. The imaging is accelerated by acquiring a reduced number of k-space samples. Partial data acquisition increases the spacing between regular subsequent read-out lines, thereby reducing scan time. However, this reduction in the number of recorded Fourier components leads to aliasing artifacts in images. There are two general approaches for removing the aliasing artifacts and reconstructing high quality images: image domain-based methods and k-space based methods. Various models in the framework of (2–1) have been employed as image domain-based reconstruction methods in PPI [15, 29, 54–57, 60, 70, 73]. Sensitivity encoding (SENSE) [54, 55] is one of the most commonly used methods of such kind. SENSE utilizes knowledge of the coil sensitivities to separate aliased pixels resulted from undersampled k-space.

The fundamental equation for SENSE is as follows: In a PPI system consisting of J coil arrays, the under-sampled k-space data f_j from the j-th channel relates to the underlying image u by

$$P\mathcal{F}(S_j \odot u) = f_j, \quad j = 1, \cdots, J, \tag{2–3}$$

where $\mathcal{F}$ is the Fourier transform, P is a binary matrix representing the under-sampling pattern (mask), and $S_j \in \mathbb{C}^N$ is the sensitivity map of the j-th channel in the vector form as u. The symbol $\odot$ is the Hadamard (or componentwise) product between two vectors. In early works on SENSE, the reconstruction was obtained by solving a least squares problem

$$\min_{u \in \mathbb{C}^N} \sum_{j=1}^{J} \|\mathcal{F}_p(S_j \odot u) - f_j\|^2, \tag{2–4}$$

where $\mathcal{F}_p$ is the undersampled Fourier transform defined by $\mathcal{F}_p \triangleq P\mathcal{F}$. Denote

$$A = \begin{pmatrix} \mathcal{F}_p \mathbf{S}_1 \\ \mathcal{F}_p \mathbf{S}_2 \\ \vdots \\ \mathcal{F}_p \mathbf{S}_J \end{pmatrix} \quad \text{and} \quad f = \begin{pmatrix} f_1 \\ f_2 \\ \vdots \\ f_J \end{pmatrix}, \tag{2–5}$$

where $\mathbf{S}_j \triangleq \text{diag}(S_j) \in \mathbb{C}^{N \times N}$ is the diagonal matrix with $S_j \in \mathbb{C}^N$ on the diagonal, $j = 1, \cdots, J$. Then problem (3–48) can be expressed as

$$\min_{u \in \mathbb{C}^N} \|Au - f\|^2, \tag{2–6}$$

and then solved by conjugate gradient (CG) algorithm. However, due to the ill-conditioning of the encoding matrix A, it has been shown in [57] that the CG iteration sequence often exhibits a "semi-convergence" behavior, which can be characterized as initially converging toward the exact solution and later diverging. Moreover, the convergence speed is low, when the acceleration factor is high.

Recently, total variation (TV) based regularization has been incorporated into SENSE to improve reconstructed image quality and convergence speed over the

un-regularized CG method ([15, 73]). TV based regularization can be also considered as forcing the reconstructed image to be sparse with respect to spatial finite differences. This sparsity along with the sparsity of MR signals under wavelet transforms have been exploited in [44], where the framework (2–1) has been employed to reconstruct MR images from under-sampled k-space data.

There have been several fast numerical algorithms for solving (2–1) that will be briefly reviewed in the next section. However, computational acceleration is still an important issue for certain medical applications, such as breath-holding cardiac imaging. For the application in PPI the computational challenging is not only from the lack of differentiability of the TV and ℓ_1 terms , but also the inversion matrix A in (2–5) which has large size and is severely ill-conditioned.

The main contribution of this paper is to develop a fast numerical algorithm for solving (2–1) with general A. The proposed algorithm incorporates the Barzilai-Borwein (BB) method into a variable splitting framework for optimal step size selection. The numerical results on partially parallel imaging (PPI) problems demonstrate much improved performance on reconstruction speed for similar image quality.

2.1.2 Previous Work

In reviewing the prior work on TVL1-based image reconstruction, we simplify (2–1) by taking $\beta = 0$. It is worth pointing out here that TV has much stronger practical performance than ℓ_1 in image reconstructions, yet harder to solve because the gradient operators involved are not invertible as Ψ in the ℓ_1 term.

With $\beta = 0$, the image reconstruction is equivalent to solving the problem

$$\min_{u\in\mathbb{C}^N} \left\{ \alpha \sum_{i=1}^{N} \|D_i u\| + \frac{1}{2}\|Au - f\|^2 \right\}. \tag{2–7}$$

Early work on algorithms for (3–6) used gradient descent methods with explicit [59] or semi-implicit schemes [43, 63] in which the TV norm was replaced by a smooth

approximation

$$\|u\|_{TV,\varepsilon} = \sum_{i=1}^{N} \sqrt{\|D_i u\|^2 + \varepsilon}. \tag{2–8}$$

However, the choice of $\varepsilon > 0$ is crucial to the reconstruction results and convergence speed. A large ε encourages fast convergence rate, but fails to preserve high quality details such as edges in the restored image; a small ε better preserves fine structure in the reconstruction at the expense of slow convergence.

In [64, 67], a method is developed based on the following reformulation of (3–6):

$$\min_{u,w} \left\{ \alpha \sum_{i=1}^{N} \|w_i\| + \frac{1}{2}\|Au - f\|^2 : w_i = D_i u, \forall\, i \right\} \tag{2–9}$$

Then the linear constraint was treated with a quadratic penalty

$$\min_{u,w} \left\{ \alpha \sum_{i=1}^{N} \|w_i\| + \rho\|Du - w\|^2 + \frac{1}{2}\|Au - f\|^2 \right\}, \tag{2–10}$$

where $w \in \mathbb{C}^{2N}$ is formed by stacking the two columns of $(w_1, \cdots, w_N)^\top$, and $D = (D^x; D^y) \in \mathbb{C}^{2N\times N}$. D^x and D^y are the horizontal and vertical global finite difference matrices (N-by-N), i.e. they consist of the first and second rows of all D_i's, respectively. For any fixed ρ, (3–10) can be solved by alternating minimizations. If both $D^\top D$ and $A^\top A$ can be diagonalized by the Fourier matrix, as they would if A is either the identity matrix or a blurring matrix with periodic boundary conditions, then each minimization involves shrinkage and two fast Fourier transforms (FFTs). A continuation method is used to deal with the slow convergence rate associated with a large value for ρ. The method, however, is not applicable to more general A.

In [34] Goldstein and Osher developed a split Bregman method for (3–9). The resulting algorithm has similar computational complexity to the algorithm in [64]; the convergence is fast and the constraints are exactly satisfied. Later the split Bregman method was shown to be equivalent to the alternating direction method of multipliers

(ADMM) [13, 28, 32, 33] applied to the augmented Lagrangian $L(w, u; p)$ defined by

$$\alpha \sum_{i=1}^{N} \|w_i\| + \frac{1}{2}\|Au - f\|^2 + \langle p, Du - w\rangle + \frac{\rho}{2}\|Du - w\|^2, \tag{2–11}$$

where $p \in \mathbb{C}^{2N}$ is the Lagrangian multiplier. Nonetheless, the algorithms in [34, 64, 67] benefit from the special structure of A, and they lose efficiency if $A^\top A$ cannot be diagonalized by fast transforms. To treat a more general A, the Bregman operator splitting (BOS) method [75] replaces $\|Au - f\|^2$ by a proximal-like term $\delta\|u - (u^k - \delta^{-1}A^\top(Au^k - f))\|^2$ for some $\delta > 0$. BOS is an inexact Uzawa method that depends on the choice of δ. The advantage of BOS is that it can deal with general A and does not require the inversion of $A^\top A$ during the computation. However, BOS is relatively less efficient than the method presented in this paper, even if δ is chosen optimally. The comparison of our method with the BOS algorithm will be presented in Section 2.4.

There are also several methods developed to solve the associated dual or primal-dual problems of (3–6) based on the dual formulation of the TV norm:

$$\|u\|_{TV} = \max_{p \in X} \langle p, Du\rangle, \tag{2–12}$$

where $X = \{p \in \mathbb{C}^{2N} : p_i \in \mathbb{C}^2, \|p_i\| \leq 1, \forall\, i\}$ and p_i extracts the i-th and $(i+N)$-th entries of p. Consequently, (3–6) can be written as a minimax problem

$$\min_{u \in \mathbb{C}^N} \max_{p \in X} \left\{ \alpha\langle p, Du\rangle + \frac{1}{2}\|Au - f\|^2 \right\}. \tag{2–13}$$

In [20], Chan *et al.* proposed to solve the primal-dual Euler-Lagrange equations using Newton's method. This leads to a quadratic convergence rate and highly accurate solutions; however, the cost per iteration is much higher since the method explicitly uses second-order information and the inversion of a Hessian matrix is required. In [19], Chambolle used the dual formulation of the TV denoising problem (3–6) with $A = I$, and provided an efficient semi-implicit gradient descent algorithm for the dual. However, the method does not naturally extend to the case with more general A. Recently, Zhu and

Chan [78] proposed a primal-dual hybrid gradient (PDHG) method. PDHG alternately updates the primal and dual variables u and p. Numerical results show that PDHG outperforms methods in [19, 34] for denoising and deblurring problems, but its efficiency again relies on the fact that $A^\top A$ can be diagonalized by fast transforms. Later, several variations of PDHG, referred to as projected gradient descent algorithms, were applied to the dual formulation of image denoising problem in [79] to make the method more efficient. Further enhancements involve different step-length rules and line-search strategies, including techniques based on the Barzilai-Borwein method [9].

Another approach that can be applied to (3–6) in the imaging context (2–1) with a general A is the operator splitting (OS) method. In [51] the OS idea of [47] is applied to image reconstruction in compressed magnetic resonance imaging. The OS scheme rewrites (3–6) as

$$\min_u \alpha \sum_i h(D_i u) + \frac{1}{2}\|Au - f\|^2 \tag{2–14}$$

where $h(\cdot) \triangleq \|\cdot\|$. Then the optimal conditions for (2–14) are

$$w_i^* \in \partial h(D_i u^*), \quad \delta_1 \alpha D_i^\top w_i^* + \delta_1 A^\top (Au^* - f) = 0, \tag{2–15}$$

where $\partial h(z)$ is the subdifferential of h at some point z defined by a set

$$\partial h(z) \triangleq \{d \in \mathbb{C}^N : h(y) - h(z) \geq \langle d, y - z\rangle, \forall\, y\}.$$

The theory of conjugate duality gives the equivalency $y \in \partial h(z) \Leftrightarrow z \in \partial h^*(y), \forall y, z$, where $h^*(y) \triangleq \sup_v \{\langle y, v\rangle - h(v)\}$. Hence the first condition in (2–15) can be written as

$$0 \in \delta_2 h^*(w_i^*) + (w_i^* - t_i^*), \quad t_i^* = \delta_2 D_i u^* + w_i^* \tag{2–16}$$

and then the first one leads to

$$w_i^* \in \partial h\left((t_i^* - w_i^*)/\delta_2\right) = \partial h(t_i^* - w_i^*), \tag{2–17}$$

where the equality is due to $h(\cdot) = \|\cdot\|$. (2–17) is equivalent to

$$w_i^* = \arg\min_{w_i} \left\{ h(t_i^* - w_i) + \frac{1}{2}\|w_i\|^2 \right\} \tag{2–18}$$

that projects t_i^* onto the unit ball in $\mathbb{R}^2$. Then, combining (2–18) and the last equalities in (2–15) and (2–16), the OS scheme iterates the following for a fixed point (which is also a solution to (3–6)):

$$\begin{cases} t_i^{k+1} = w_i^k + \delta_2 D_i u^k, \quad \forall i \\ w_i^{k+1} = \arg\min_{w_i} \left\{ \|t_i^{k+1} - w_i\| + \frac{1}{2}\|w_i\|^2 \right\}, \quad \forall i \\ u^{k+1} = \delta_1 \alpha \sum_i D_i^\top w_i^{k+1} + \delta_1 A^\top (A u^k - f) + u^k \end{cases}$$

OS is efficient for solving (2–1) with general A when all the parameters are carefully chosen. However it is still not as efficient as our method even under its optimal settings. The comparison of our method with the OS algorithm [51] will be given in Section 2.4.

2.1.3 Organization

The rest of this paper is organized as follows. In Section 2.2 we present the proposed algorithm with detailed derivations. Section 2.3 describes our experiment design and the clinical data used in this paper. Section 2.4 compares our algorithm to BOS [75] and operator splitting [51] on PPI data. Finally, we conclude the paper in Section 2.5.

2.2 Proposed Algorithm

In this paper, we develop a fast and simple algorithm to numerically solve problem (2–1). Note that the computational challenge of (2–1) comes from the combination of two issues: one is possibly huge size and of the inversion matrix A, and the other one is the non-differentiability of the TV and ℓ_1 terms.

As discussed earlier, despite that there were some fast algorithms proposed recently to solve image restoration problems similar to (2–1), their efficiency relies on a very special structure of A such that $A^\top A$ can be diagonalized by fast transforms,

which is not the case in most medical imaging problems, such as that in (2–5) in PPI application.

To tackle the computational problem of (2–1), we first introduce auxiliary variables w_i and z_i to transform $D_i u$ and $\psi_i^\top u$ out of the non-differentiable norms:

$$\begin{aligned} &\min_{w,z,u} \left\{ \alpha \sum_i \|w_i\| + \beta \sum_i |z_i| + \frac{1}{2}\|Au - f\|^2 \right\}, \\ &w_i = D_i u, z_i = \psi_i^\top u, \forall\, i = 1, \cdots, N, \end{aligned} \tag{2–19}$$

which is clearly equivalent to the original problem (2–1) as they share the same solutions u. To deal with the constraints in (2–19) brought by variable splitting, we form the augmented Lagrangian defined by

$$\begin{aligned} &L(w, z, u; b, c) \\ &= \alpha \sum_i \left(\|w_i\| - \rho\langle b_i, w_i - D_i u\rangle + \frac{\rho}{2}\|w_i - D_i u\|^2 \right) \\ &+ \beta \sum_i \left(|z_i| - \rho c_i (z_i - \psi_i^\top u) + \frac{\rho}{2}|z_i - \psi_i^\top u|^2 \right) \\ &+ \frac{1}{2}\|Au - f\|^2, \end{aligned} \tag{2–20}$$

where $b \in \mathbb{C}^{2N}$ and $c = (c_1, \cdots, c_N)^\top \in \mathbb{C}^N$ are Lagrangian multipliers. Here $b_i \in \mathbb{C}^2$ extracts the i-th and $(i+N)$-th entries of b. For notation simplicity we used the same parameter $\rho > 0$ for all constraints in (2–20). The method of multipliers iterates the minimizations of Lagrangian L in (2–20) with respect to (w, z, u) and the updates of the multipliers b and c:

$$\begin{cases} (w^{k+1}, z^{k+1}, u^{k+1}) = \arg\min_{w,z,u} L(w, z, u; b^k, c^k) \\ b_i^{k+1} = b_i^k - (w_i^{k+1} - D_i u^{k+1}), \quad \forall\, i \\ c_i^{k+1} = c_i^k - (z_i^{k+1} - \psi_i^\top u^{k+1}), \quad \forall\, i \end{cases} \tag{2–21}$$

It is proved that the sequence $\{(w^k, z^k, u^k)\}_k$ generated by (2–21) converges to the solution of (2–19) with any $\rho > 0$.

Since the updates of b^k and c^k are merely simple calculations, we now focus on the minimization of $L(w, z, u; b^k, c^k)$ in (2–21). First we introduce functions

$$\phi_1(s, t) = |s| + (\rho/2) \cdot |s - t|^2, \ s, t \in \mathbb{C} \tag{2–22}$$

and

$$\phi_2(\mathbf{s}, \mathbf{t}) = \|\mathbf{s}\| + (\rho/2) \cdot \|\mathbf{s} - \mathbf{t}\|^2, \ \mathbf{s}, \mathbf{t} \in \mathbb{C}^2. \tag{2–23}$$

By completing the squares in (2–20), we find the equivalency

$$\begin{aligned} &\arg\min_{w,z,u} L(w, z, u; b^k, c^k) \equiv \\ &\arg\min_{w,z,u} \left\{ \alpha \sum_i \phi_2(w_i, D_i u + b_i^k) + \beta \sum_i \phi_1(z_i, \psi_i^\top u + c_i^k) \right. \\ &\qquad \left. + \frac{1}{2}\|Au - f\|^2 \right\} \end{aligned} \tag{2–24}$$

because the objective functions in these two minimizations are equal up to a constant independent of (w, z, u).

To solve (2–24) we first rewrite the objective function in a simpler way. Let $x = (w; z; u)$ and $B = (0, 0, A)$, and define functions $J_k(x) \equiv J_k(w, z, u)$ by

$$J_k(x) \triangleq \alpha \sum_i \phi_2(w_i, D_i u + b_i^k) + \beta \sum_i \phi_1(z_i, \psi_i^\top u + c_i^k), \tag{2–25}$$

and data fidelity $H(x)$ by

$$H(x) = (1/2) \cdot \|Bx - f\|^2. \tag{2–26}$$

Then problem (2–24) (or equivalently, the minimization subproblem in (2–21)) can be expressed as

$$x^{k+1} = \arg\min_x \{J_k(x) + H(x)\}, \tag{2–27}$$

We further introduce $Q_\delta(x, y)$ defined by

$$Q_\delta(x, y) \triangleq H(y) + \langle \nabla H(y), x - y \rangle + \frac{\delta}{2}\|x - y\|^2, \tag{2–28}$$

which is a linearization of $H(x)$ at point y plus a proximity term $\|x-y\|^2/2$ penalized by parameter $\delta > 0$. It has been shown in [66] that the iterative sequence $\{x^{k+1,l}\}_l$ generated by

$$x^{k+1,l+1} = \arg\min_x \left\{ J_k(x) + Q_{\delta_{k+1,l}}(x, x^{k+1,l}) \right\} \tag{2–29}$$

converges to the solution x^{k+1} of (2–27) with any initial $x^{k+1,0}$ and proper choice of $\delta_{k+1,l}$ for $l = 0, 1, \cdots$ [1] . Interestingly, we found that in practice the optimal performance can be consistently achieved if only iterating (2–29) once to approximate the solution x^{k+1} in (2–27).

Therefore, we substitute the first subproblem in (2–21) by

$$x^{k+1} = \arg\min_x \left\{ J_k(x) + Q_{\delta_k}(x, x^k) \right\}, \tag{2–30}$$

where δ_k is chosen based on the Barzilai-Borwein (BB) method as suggested in [66]. BB method handles ill-conditioning much better than gradient methods with a Cauchy step [3]. In the BB implementation, the Hessian of the objective function is approximated by a multiple of the identity matrix. We employ the approximation

$$\delta_k = \arg\min_\delta \left\| \left(\nabla H(x^k) - \nabla H(x^{k-1})\right) - \delta(x^k - x^{k-1}) \right\|^2, \tag{2–31}$$

and get

$$\delta_k = \langle \nabla H(x^k) - \nabla H(x^{k-1}), x^k - x^{k-1} \rangle / \|x^k - x^{k-1}\|^2. \tag{2–32}$$

This makes the iteration (2–30) exhibit a certain level of quasi-Newton convergence behavior.

[1] e.g. for fixed k, any limit point of $\{x^{k+1,l}\}_l$ is a solution of (2–27) when $\delta_{k+1,l}$ was chosen such that the objective function $J_k(x^{k+1,l}) + H(x^{k+1,l})$ monotonically decreases as $l \to \infty$ [66].

$$\begin{cases} w_i^{k+1} = \arg\min\limits_{w_i} \left\{ \|w_i\| + \dfrac{\rho}{2}\|w_i - D_i u^k - b_i^k\|^2 + \dfrac{\delta_k}{2\alpha}\|w - w^k\|^2 \right\}, \quad \forall\, i; \\ z_i^{k+1} = \arg\min\limits_{z_i} \left\{ |z_i| + \dfrac{\rho}{2}|z_i - \Psi_i^\top u^k - c_i^k|^2 + \dfrac{\delta_k}{2\beta}|z_i - z_i^k|^2 \right\}, \quad \forall\, i; \\ u^{k+1} = \arg\min\limits_{u} \left\{ \alpha\rho\|Du - w^{k+1}\|^2 + \beta\rho\|\Psi^\top u - z^{k+1}\|^2 + \delta_k \left\| u - \left(u^k - \delta_k^{-1} A^\top (Au^k - f)\right)\right\|^2 \right\}; \\ b_i^{k+1} = b_i^k - (w_i^{k+1} - D_i u^{k+1}), \quad \forall\, i; \\ c_i^{k+1} = c_i^k - (z_i^{k+1} - \psi_i u^{k+1}), \quad \forall\, i; \\ \delta_{k+1} = \|A(u^{k+1} - u^k)\|^2 / \left(\|w^{k+1} - w^k\|^2 + \|z^{k+1} - z^k\|^2 + \|u^{k+1} - u^k\|^2\right). \end{cases} \tag{2–35}$$

From the definition of J_k and Q_{δ_k}, (2–30) is equivalent to

$$(w^{k+1}, z^{k+1}, u^{k+1}) = \arg\min_{w,z,u} \Phi_k(w, z, u) \tag{2–33}$$

where the objective $\Phi_k(w, z, u)$ is defined by

$$\begin{aligned} &\Phi_k(w, z, u) \\ &\triangleq \alpha \sum_i \phi_2(w_i, D_i u + b_i^k) + \beta \sum_i \phi_1(z_i, \psi_i^\top u + c_i^k) \\ &+ \frac{\delta_k}{2}\left(\|w - w^k\|^2 + \|z - z^k\|^2 + \left\|u - u^k + \delta_k^{-1} A^\top (Au^k - f)\right\|^2\right) \end{aligned} \tag{2–34}$$

Theoretically, an iterative scheme can be applied to obtain the solution $(w^{k+1}, z^{k+1}, u^{k+1})$ of (2–33). However, here we propose only to do *one* iteration followed by the updates of b^k, c^k and δ_k in (2–21). This is an analogue to the split Bregman method and ADMM applied to the augmented Lagrangians, and leads to the optimal performance of (2–21). In summary, we propose a scheme as in (2–35) for solving the minimization problem (2–19).

The updates of b^{k+1} and c^{k+1} in (2–35) are merely simple calculations. In (2–35), δ_{k+1} is derived from (2–32) with H defined in (2–26), and also has an explicit form that

can be quickly computed [2] . Next, we show that w_i^{k+1} and z_i^{k+1} can be obtained by soft shrinkages by the following theorem.

Theorem 2.1. *For given d-vectors $\mathbf{t}_1, \mathbf{t}_2 \in \mathbb{R}^d$ and positive numbers $a_1, a_2 > 0$, the solution to minimization problem*

$$\min_{\mathbf{s}\in\mathbb{R}^d} \left\{ \|\mathbf{s}\| + \frac{a_1}{2}\|\mathbf{s}-\mathbf{t}_1\|^2 + \frac{a_2}{2}\|\mathbf{s}-\mathbf{t}_2\|^2 \right\} \tag{2–36}$$

is given by the shrinkage of a weighted sum of $\mathbf{t}_1$ and $\mathbf{t}_2$:

$$\mathcal{S}_d(\mathbf{t}_1, \mathbf{t}_2; a_1, a_2) \triangleq \texttt{shrink}_d\left(\frac{a_1\mathbf{t}_1 + a_2\mathbf{t}_2}{a_1+a_2}, \frac{1}{a_1+a_2}\right) \tag{2–37}$$

where $\texttt{shrink}_d$ is the d-dimensional soft shrinkage operator defined by

$$\texttt{shrink}_d(\mathbf{t}, \mu) \triangleq \max\{\|\mathbf{t}\| - \mu, 0\} \cdot \frac{\mathbf{t}}{\|\mathbf{t}\|}. \tag{2–38}$$

with convention $0 \cdot (\mathbf{0}/\|\mathbf{0}\|) = \mathbf{0}$.

Proof. By completing the squares, the minimization problem (2–36) is equivalent to

$$\min_{\mathbf{s}\in\mathbb{R}^d} \left\{ \|\mathbf{s}\| + \left(\frac{a_1+a_2}{2}\right) \cdot \left\| \mathbf{s} - \frac{a_1\mathbf{t}_1 + a_2\mathbf{t}_2}{a_1+a_2} \right\|^2 \right\}, \tag{2–39}$$

because the objective functions are the same up to a constant independent of $\mathbf{s}$. Minimizations of form (2–39) have a well known explicit solver $\texttt{shrink}_d$ and hence the conclusion follows. □

According to Theorem 2.1, w_i^{k+1} and z_i^{k+1} in (2–35) can be obtained by

$$w_i^{k+1} = \mathcal{S}_2\left(D_i u^k + b_i^k, w_i^k; \rho, \alpha_k\right) \tag{2–40}$$

[2] The main computations for updating δ_k are norm evaluations (no A operation needed since Au^k has been computed in the u-step and can be saved for use in δ-step in (2–35)).

and

$$z_i^{k+1} = \mathcal{S}_1\left(\psi_i^\top u^k + c_i^k, z_i^k; \rho, \beta_k\right) \tag{2–41}$$

where $\alpha_k = \delta_k/\alpha$ and $\beta_k = \delta_k/\beta$. Therefore the computational costs for (2–40) and (2–41) are linear in terms of N.

The u-subproblem in (2–35) is a least squares problem. The optimal condition of this problem reads

$$L_k u = r_k \tag{2–42}$$

where

$$L_k = \alpha\rho D^\top D + \beta\rho I + \delta_k I \tag{2–43}$$

and

$$r_k = \alpha\rho D^\top w^{k+1} + \beta\rho\Psi^\top z^{k+1} + \delta_k u^k - A^\top(Au^k - f). \tag{2–44}$$

Under periodic boundary condition, the matrix $D^\top D$ is block circulant and hence can be diagonalized by Fourier matrix $\mathcal{F}$. Let $\Lambda = \mathcal{F}^\top D^\top D\mathcal{F}$ which is a diagonal matrix, then apply $\mathcal{F}$ on both sides of (2–42) to obtain

$$\hat{L}_k \mathcal{F} u = \hat{r}_k \tag{2–45}$$

where

$$\hat{L}_k = \alpha\rho\Lambda + \beta\rho I + \delta_k I \quad \text{and} \quad \hat{r}_k = \mathcal{F} r_k. \tag{2–46}$$

Note that $\hat{L}_k$ can be "trivially" inverted because it is diagonal and positive definite. Therefore, u^{k+1} can be easily obtained by

$$u^{k+1} = \mathcal{F}^\top(\hat{L}_k^{-1}\mathcal{F} r_k). \tag{2–47}$$

As all variables in (2–35) can be quickly solved, we propose Algorithm 2, called TVL1rec, to solve problem (2–19). As discussed above, w and z can be updated using soft shrinkages and hence the computational costs are linear in terms of N. The update of u involves two fast Fourier transforms (FFTs) which has computational complexity

Algorithm 2 TVL1 Reconstruction Algorithm (TVL1rec)

Input α, β, ϵ and ρ. Set $u^0 = c = 0$, $b = 0$, $\delta_0 = 1$, $k = 0$.
repeat
  Given u^k, compute w^{k+1} and z^{k+1} using (2–40) and (2–41);
  Given w^{k+1} and z^{k+1}, compute u^{k+1} using (2–47);
  Update b^k, c^k and δ_k as in (2–35);
  $k \leftarrow k + 1$
until $\|u^k - u^{k-1}\|/\|u^k\| < \epsilon$.
return u^k

$N \log N$ and two operations of A (one is $A^\top$). Therefore, unlike most recently developed algorithms, our algorithm can deal with arbitrary matrix A and even more general H with nonlinear constraint (as long as H is convex and ∇H is computable). Also, the per iteration computation of the proposed algorithm is very cheap, and thanks to the BB step size δ_k, the convergence speed is significantly improved compared to other two modern methods BOS and OS, as shown in Section 2.4.

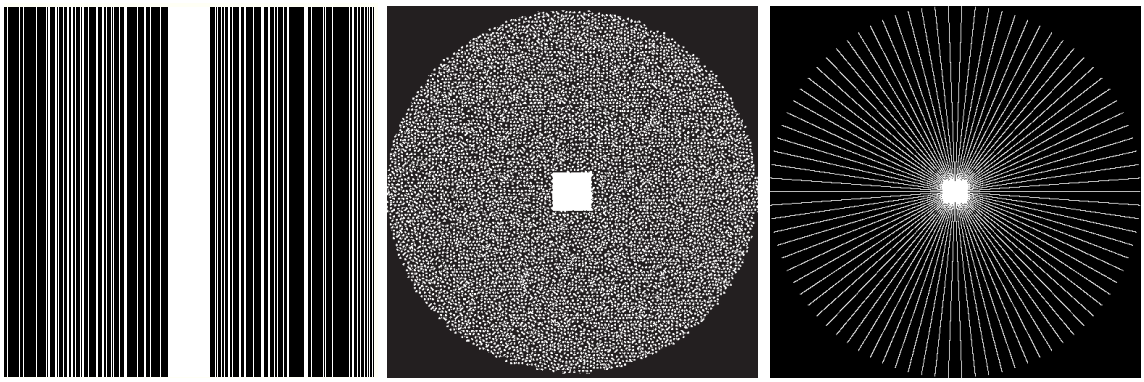
2.3 Method

Experiments were designed to test the effectiveness of the proposed algorithm TVL1rec on PPI reconstructions. To demonstrate the potential in clinic applications, the three data sets used in the experiments were acquired by commercially available 8-element head coils. For comparison, we also implemented the Bregman operator splitting algorithm (BOS) [75] and a compressive MR image reconstruction algorithm based on operator splitting (OS) [51] for solving (2–1).

2.3.1 Data Acquisition

The first data set (top left in Figure 2-2), termed by data1, is a set of sagittal Cartesian brain images acquired on a 3T GE system (GE Healthcare, Waukesha, Wisconsin, USA). The data acquisition parameters were FOV 220mm^2, size $512\times512\times8$, TR 3060ms, TE 126ms, slice thickness 5mm, flip angle 90°, and phase encoding direction was anterior-posterior.

The second data set (left in Figure 2-4) is a Cartesian brain data set acquired on a 3.0T Philips scanner (Philips, Best, Netherlands) using T2-weighted turbo spin echo (T2



A B C

Figure 2-1. The k-space masks used for the three data sets tested in the experiments.

TSE) sequence. The acquisition parameters were FOV 205mm^2, matrix $512 \times 500 \times 8$, TR 3000ms, TE 85ms, and the echo train length was 20. To avoid similar comparison plot due to the same data size, we reduce the image to $256 \times 250 \times 8$ and obtain full k-space data of this same size, termed by data2.

The last one (right of Figure 2-4), denoted by data3, is a radial brain data set acquired on a 1.5T Siemens Symphony system (Siemens Medical Solutions, Erlangen, Germany). The acquisition parameters were FOV 220mm^2, matrix $256 \times 512 \times 8$ (256 radial lines), slice thickness 5mm, TR 53.5ms, TE 3.4ms, and flip angle 75°.

All three data sets were fully acquired, and then artificially down-sampled using the masks in Figure 3-1 for reconstruction. Figure 3-1 shows, from left to right, the Cartesian mask with net reduction factor 3, pseudo random mask with reduction factor 4, and radial mask with 43 (out of 256) projections, i.e. reduction factor 6. The reference images in our experiments were obtained by fully acquired k-space. A summary of the data information is in Table 2-1. In Table 2-1, "Cart.Sag." means "Cartesian sagittal brain image", and "Rad.Axi." stands for "radial axial brain image". The column P in Table 2-1 present the mask number (refer to Figure 3-1).

Table 2-1. Tests number, data information and parameters.

No.	Image	Abbrev.	Size($\times 8$)	P	(α, β)
1	Cart.Sag.	data1	512×512	1	(1e-5$\sim$1e-2,0)
2	Cart.Sag.	data2	256×250	2	(1e-4,5e-5)
3	Rad.Axi.	data3	256×512	3	(1e-4,5e-5)

2.3.2 Test Environment

All algorithms were implemented in the MATLAB®programming environment (Version R2009a, MathWorks Inc., Natick, MA, USA). The sparsifying operator Ψ is set to Haar wavelet transform using Rice wavelet toolbox with default settings. The experiments were performed on a Dell Optiplex desktop with Intel®Dual Core 2.53 GHz processors (only 1 core was used in computation), 3GB of memory and Windows™operating system.

Theoretically the choice of ρ does not effect the convergence of TVL1rec. This is also demonstrated by our experiments since the results are not sensitive to ρ for a large range. Therefore in all experiments we set ρ to a moderate value 10. Algorithm 2 is automatically terminated if the relative change of u^k is less than a prescribed tolerance ϵ. In all of our experiments, we set $\epsilon = 10^{-3}$. Note that smaller ϵ leads to slightly better accuracy at the cost of more iterations and longer computational time. Other parameter settings are shown in the next section. For all algorithms tested in this paper, the sensitivity maps S_j's were estimated from the central 32×32 k-space data (which was a subset of the acquired partial data) and then fixed during the reconstructions, and the initial u^0 was set to 0.

The reconstruction results were evaluated qualitatively by zoomed-in regions of the reconstructed images, and quantitatively by relative error (to the reference image) and CPU times. Reference and reconstructed images corresponding to data1 and data3 were brightened by 3 times, and those corresponding to data2 were brightened by 2 times, to help visual justifications.

2.4 Comparison Algorithms and Results

2.4.1 Comparison with BOS

In the first experiment, we use data1 with a Cartesian sampling pattern (left in Figure 3-1) to undersample k-space data. We compare TVL1rec with BOS which also solves (2–1) via a variable splitting framework (2–19). To simplify comparison, we here set $\beta = 0$ in (2–1) and focus on the computational efficiency of two algorithms in solving (2–1).

The BOS algorithm solves (2–1) by iterating

$$\begin{cases} s^{k+1} = u^k - \delta^{-1} A^\top (Au^k - f) \\ w_i^{k+1} = \arg\min_{w_i} \left\{ \|w_i\| + \frac{\rho}{2} \|w_i - D_i u^k - b_i^k\|^2 \right\}, \quad \forall\, i \\ u^{k+1} = \arg\min_{u} \{ \alpha\rho \|Du - w^{k+1} + b^k\|^2 + \delta \left\| u - s^{k+1} \right\|^2 \} \\ b_i^{k+1} = b_i^k - (w_i^{k+1} - D_i u^{k+1}), \quad \forall\, i \end{cases} \tag{2–48}$$

and converges if $\delta \geq \|A^\top A\|_2$, i.e. the largest eigenvalue of $A^\top A$. In SENSE applications, the magnitudes of sensitivity maps are usually normalized into $[0, 1]$. Therefore from the definition of A in (2–5), we have $\delta \geq \|A^\top A\|_2 = 1$ and hence set $\delta = 1$ for optimal performance of BOS. With $\beta = 0$, TVL1rec only updates w, u, b and δ in (2–35). As can be seen, the per iteration computational costs for BOS and TVL1rec are almost identical: the main computations consist of one shrinkage, A, $A^\top$ and two FFTs (including one inverse FFT). Therefore the computation cost for a complete reconstruction is nearly proportional to the number of iterations required by BOS and TVL1rec. In this paper, we set the stopping criterion of BOS the same as TVL1rec, namely the computation will be automatically terminated when the relative change of the iterate u^k is less than $\epsilon = 10^{-3}$.

Table 2-2 shows the performance results of TVL1rec and BOS on data1 for different values of TV regularization parameter α. In Table 2-2, we list the following quantities: the relative error of the reconstructed images to the reference image (Err), the final objective

function values (Obj), the number of iterations (Iter), and the CPU time in seconds (CPU).

Table 2-2. Results of BOS and TVL1rec on data1.

	BOS				TVL1rec			
α	Err	Obj	Iter	CPU	Err	Obj	Iter	CPU
1e-5	8.1%	.281	33	75.1	7.2%	.252	7	18.6
1e-4	7.4%	1.01	17	38.9	7.1%	.860	11	26.7
1e-3	7.4%	6.00	39	88.2	7.3%	5.98	7	16.0
1e-2	11.5%	41.0	63	142.1	10.6%	40.7	7	15.9

From Table 2-2, we can see that both BOS and TVL1 are able to stably recover the image from 34% k-space data. This is further demonstrated by Figure 2-2, where both method generated images very close to the reference image. Despite that there are few observable aliasing artifacts due to Cartesian undersampling, the details such as edges and fine structures were well preserved in both reconstructions, as can be seen in the zoomed-ins in the right column of Figure 2-2. In terms of accuracy, TVL1rec gives slightly better reconstruction quality in the sense of lower relative error and objective values.

In terms of efficiency, we found that TVL1rec significantly outperforms BOS by requiring much fewer iterations (and hence less CPU time) to obtain the similar or even better image quality, as shown in Table 2-2. Compared to BOS, TVL1rec is up to 9 times faster and hence has much higher efficiency. Although two algorithms have almost the same computational costs per iteration, TVL1rec benefits from the adaptive choice of step sizes and readily outperforms BOS which uses fixed step size $\delta = \|A^\top A\|_2$ throughout the computations. The adaptive step size selection makes TVL1rec exhibits a quasi-Newton convergence behavior in some sense because $\delta_k I$ implicitly uses partial Hessian (second order) information.

The adaptive step size selection not only leads to higher efficiency but also better stableness of TVL1rec. As shown in Table 2-2, for a large range of α in $[10^{-5}, 10^{-2}]$, TVL1rec always requires 11 or fewer iterations to recover high quality images. In

comparison, BOS appears to be quite sensitive to the choice of α: this is exemplified by the last row ($\alpha = 10^{-2}$) of Table 2-2, where BOS required much more iterations than usual; meanwhile, TVL1rec benefits from the optimal step size in each iteration and readily approximates the solution in only few iterations.

In Figure 2-2, the right column shows the zoomed-in (square in data1) of images in the left column. From top to bottom of Figure 2-2, they are the reference image, reconstructed image using BOS (Err=7.4%), and reconstructed image using TVL1rec (Err=7.1%), respectively.

The better performance of TVL1rec over BOS relies on two phases: one is that TVL1rec imposes proximity terms not only for u but also for w and z in (2–35), which lead to better choices of the updates w^{k+1} and z^{k+1}; the other one is the adoption of BB method for optimal penalty parameters δ_k selection, which affects the updates of all variables as in (2–35) and leads much improved convergence speed.

2.4.2 Comparison with OS

For data2 and data3, we compare TVL1 with OS [51] for solving (2–1) with both TV and ℓ_1 terms ($\alpha = 10^{-4}, \beta = \alpha/2$). The OS scheme of [51], with a minor correction, is as follows:

$$\begin{cases} s^{k+1} = \Psi\left(u^k - (d_1/\lambda)\cdot\left(D^\top w^k + \lambda A^\top(Au^k - f)\right)\right), \\ t_i^{k+1} = w_i^k + d_2 D_i u^k, \ \forall i, \\ u^{k+1} = \Psi\left(\text{sign}(s^{k+1}) \odot \max\{|s^{k+1}| - d_1\tau/\lambda, 0\}\right), \\ w_i^{k+1} = \min(1, \|t_i^{k+1}\|)\cdot t_i^{k+1}/\|t_i^{k+1}\|_2, \ \forall i. \end{cases} \tag{2–49}$$

where $s^k \in \mathbb{C}^N$, t_i^k and $w_i^k \in \mathbb{C}^2$, $i = 1, \cdots, N$, $w^k \in \mathbb{C}^{2N}$ is formed by stacking the two columns of matrix $(w_1^k, \cdots, w_N^k)^\top$, and the "max" and "sign" operations in the computation of u^{k+1} are componentwise operations corresponding to shrinkage. The main computational cost per iteration in the OS scheme corresponds to the following operations: a 2D shrinkage during the computation of w^{k+1}, a projection during the computation of u^{k+1}, two wavelet transforms during the computation of s^{k+1} and u^{k+1}, A

and $A^\top$ during the computation of s^{k+1}. In [51] it is shown that for $d_1, d_2 > 0$ in certain ranges, the OS scheme converges to a fixed point which is also a solution of (2–1). The iterations were stopped when either the following conditions were satisfied:

$$\begin{aligned} \|u^{k+1} - u^k\|_2 / \max\{1, \|u^k\|_2\} &< \epsilon_1 \\ (f^k - f^{k+1}) / \max\{1, f^k\} &< \epsilon_2 \sqrt{\tau_c/\tau_t}, \end{aligned} \tag{2–50}$$

where f^k is the objective value of (2–1) at u^k, τ_c and τ_t are the current and target values of τ respectively and ϵ_1 and ϵ_2 are prescribed stopping tolerances.

Since OS has multiple tuning parameters that affect the convergence speed and image quality: larger d_i's and ϵ_i's lead to faster convergence but result in larger relative error, whereas smaller d_i's and ϵ_i's yield monotonic decreases in objective values and better image quality at the cost of much longer computation. Based on the selection by the authors and several tries, we chose moderate values $d_1 = d_2 = 1$, $\epsilon_1 = 10^{-4}$ and $\epsilon_2 = 10^{-3}$ which appear to give a best compromise between convergence speed and image quality of the OS scheme. The results on data2 and data3 are shown in Figure 2-4, and the comparison on relative errors and objective values (both in logarithmic) are plotted in logarithmic scale in Figure 2-5. In Figure 2-4, the left column and right column correspond to the results for data2 and data3, respectively. In the left column (results of data2), they are reference, reconstructed images by OS (Err=7.6%) and TVL1rec (Err=4.6%). In the right column (results of data3) from top to bottom, they are reference, reconstructed images by OS (Err=6.7%) and TVL1rec (Err=6.1%). The horizontal label is chosen as CPU time because the per iteration computational costs for OS and TVL1rec are slightly different.

From Figures 2-4 and 2-5 we can see that TVL1rec converges much faster than OS, and achieved lower relative errors and objective values than OS overall. Therefore, it is evident that TVL1rec can outperform OS scheme in efficiency as the former requires much less computational time to reach the similar or even better image quality. It is also

worth pointing out that both algorithms can further reduce the relative error slightly by setting a tighter stopping criterion at the cost of more iteration numbers. Nevertheless, the TVL1rec still can maintain lower relative error and objective value than OS during the reconstruction process.

2.5 Concluding Remarks

This paper presents a fast numerical algorithm, called TVL1rec, for TVL1 minimization problem (2–1) arising from CS reconstruction problems. The proposed algorithm incorporates the Barzilai-Borwein (BB) method into a variable splitting framework to optimize the selection of step sizes. The optimal step sizes exploit partial Hessian information and hence lead to a quasi-Newton convergence behavior of TVL1rec. Experimental results demonstrate the outstanding efficiency of the proposed algorithm in CS-PPI reconstruction.

We compared TVL1rec to another two recently developed algorithms BOS [75] and OS [51] which also solve the minimization problem (2). The common property of these algorithms is that they can deal with general sensing matrix A, and even nonlinear data fidelity term $H(u)$ other than $\|Au - f\|^2$ as long as H is convex and ∇H is computable. Meanwhile, TVL1rec significantly outperforms the other two algorithms by taking advantages of the optimal step size selection based on BB method. We hope TVL1rec can be beneficial to PPI and other medical imaging applications.

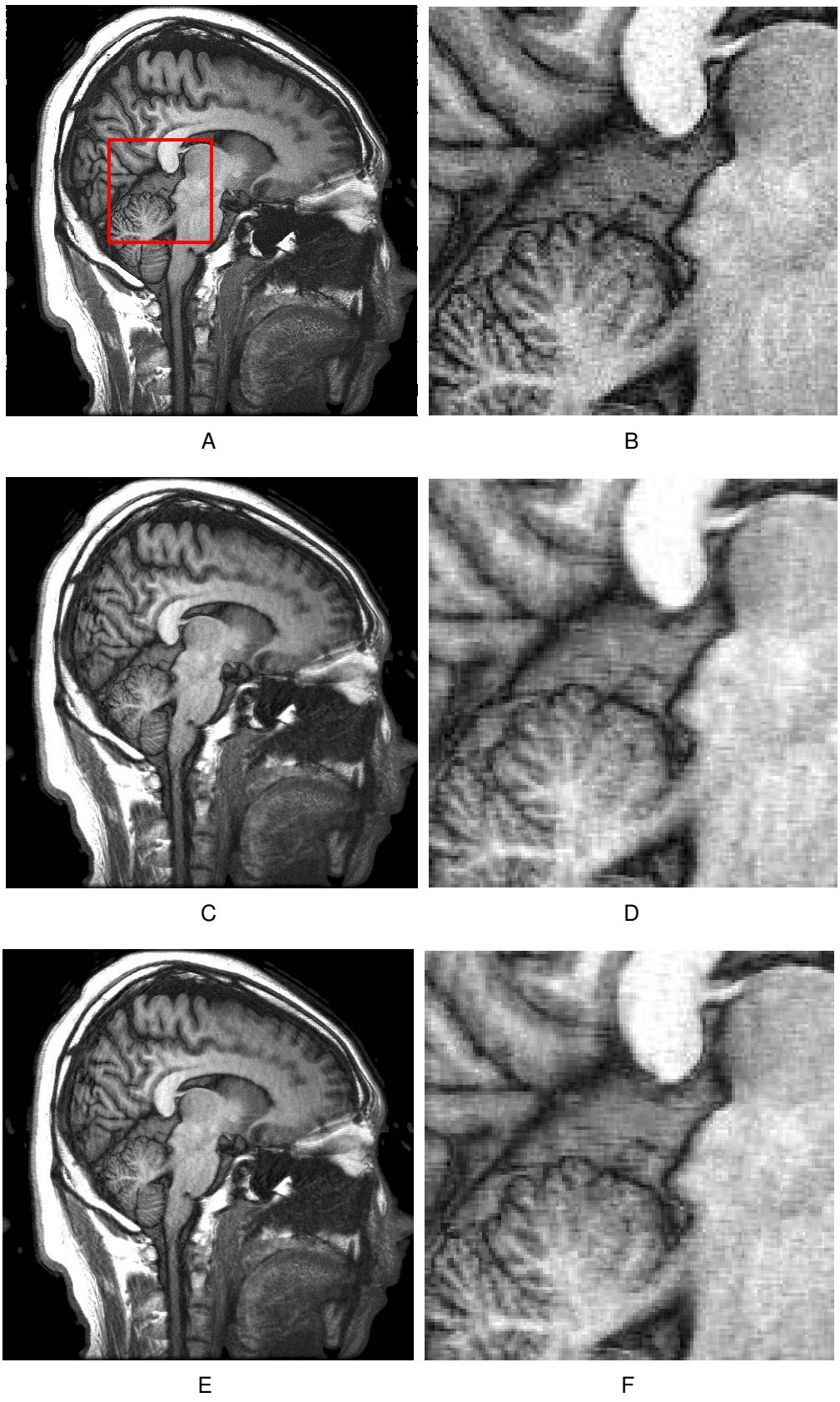
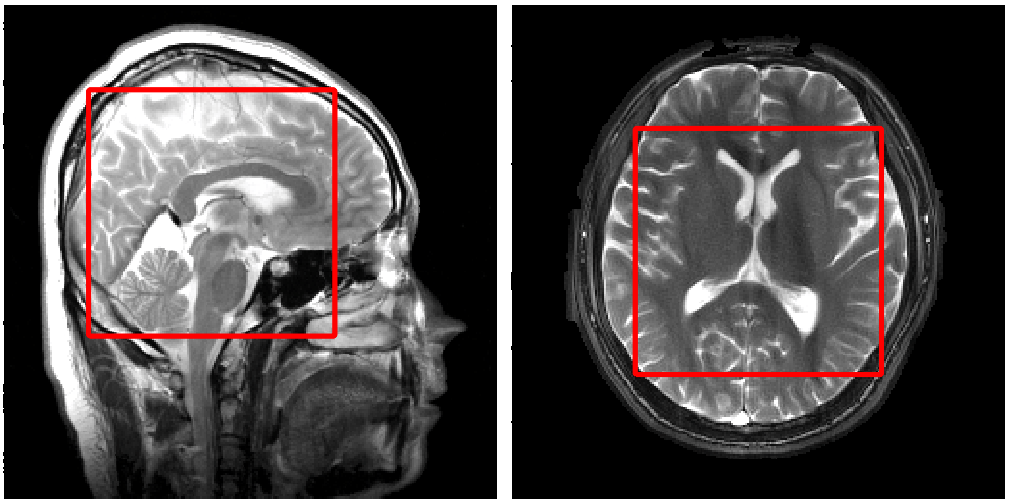


Figure 2-2. Comparison of BOS and TVL1rec on data1.



A B

Figure 2-3. Testing data used for the comparison of OS and TVL1rec for data2 and data3.

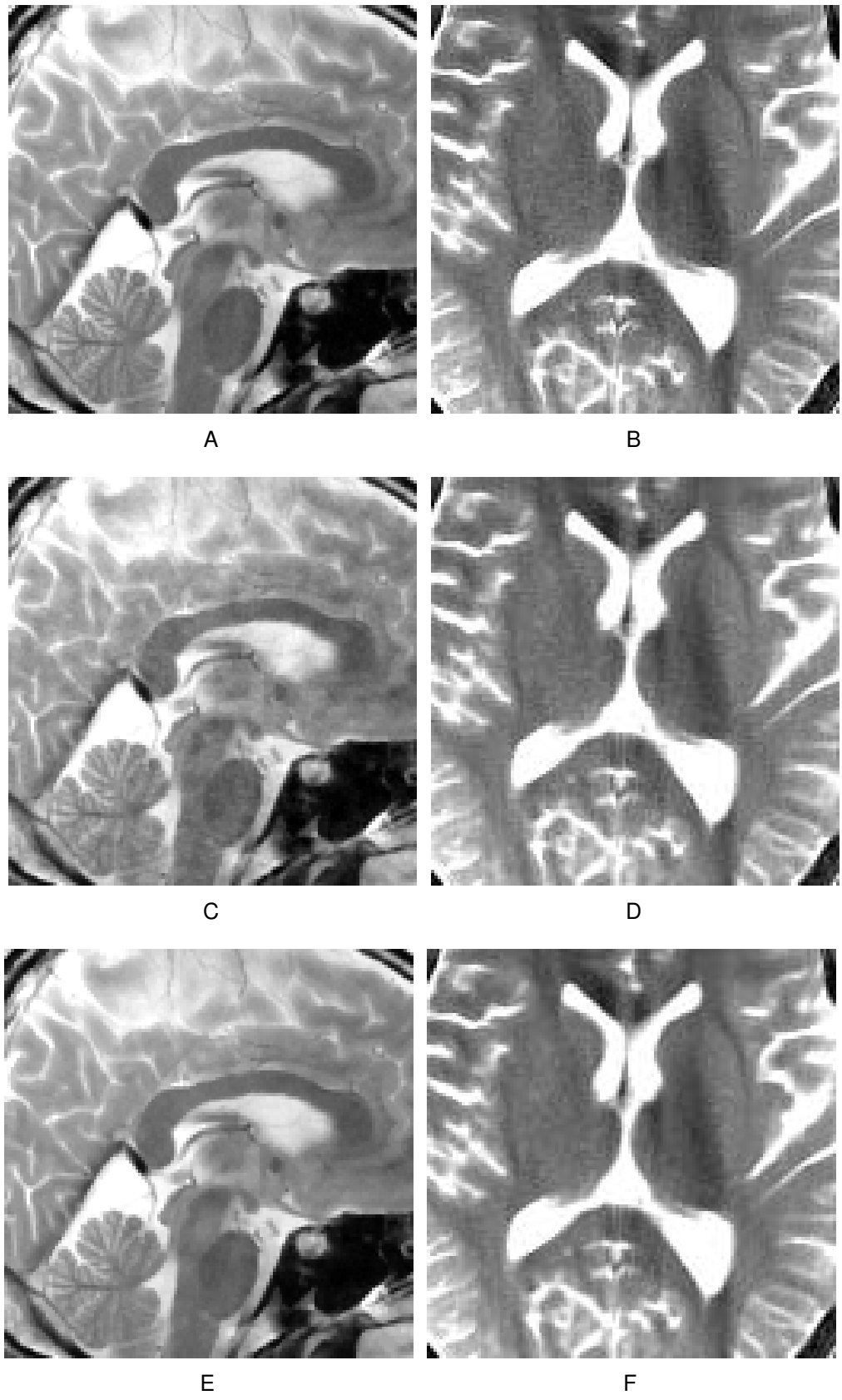
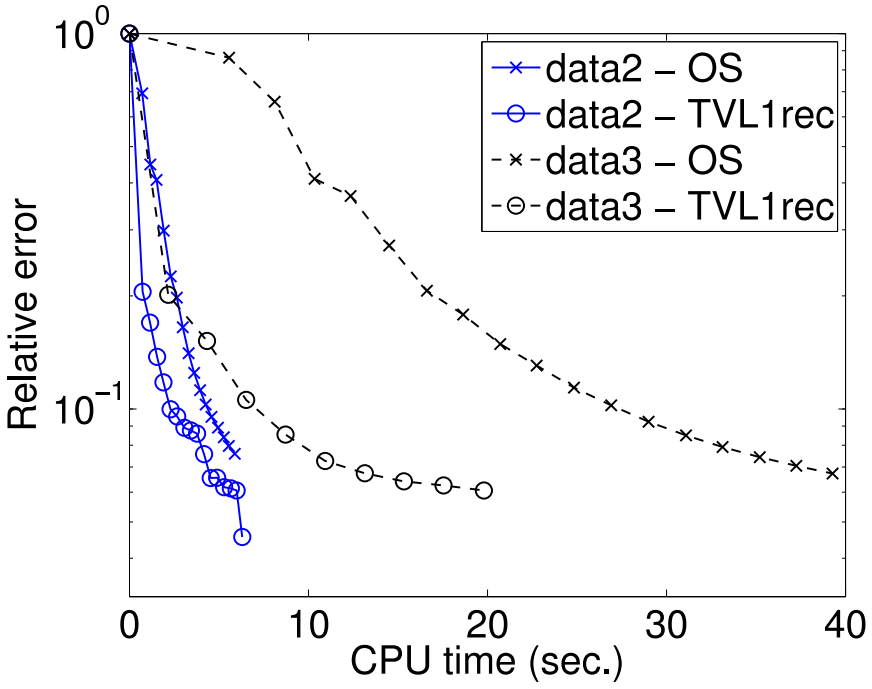
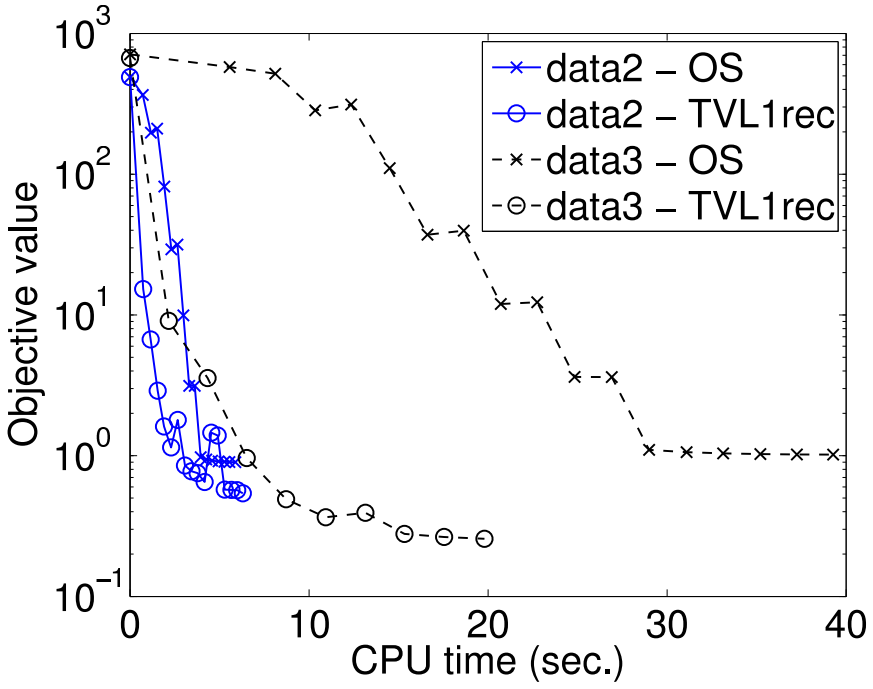


Figure 2-4. Reconstructions of data2 and data3 by OS and TVL1rec.



A



B

Figure 2-5. Comparisons of OS and TVL1rec on data2 (blue solid lines) and data3 (black dashed lines).

CHAPTER 3
A FAST ALGORITHM FOR TV IMAGE RECONSTRUCTION WITH APPLICATION TO PARTIALLY PARALLEL MR IMAGING

Outline

This paper presents a fast algorithm for total variation-based image reconstruction. The proposed method combines variable splitting and the classic penalty technique. This reduces the image reconstruction problem to an unconstrained minimization problem, which is solved by an alternating proximal minimization algorithm. One phase of the algorithm solves a total variation (TV) denoising problem, and second phase solves an ill-conditioned linear system. Linear and sublinear convergence results are given, and an implementation based on a split Bregman scheme for the TV problem and a Barzilai-Borwein method for the linear system is proposed. The algorithm is applied to image reconstruction problems that arise in partially parallel magnetic resonance imaging (PPI). Performance is compared to that of an operator splitting scheme.

3.1 Backgrounds in Optimization Methods for Nonsmooth Based Image Reconstruction

In this paper we develop a new algorithm for total variation (TV) based image reconstruction. The general form of such problems is

$$\min_{u\in\mathbb{C}^N} J(u) + H(u), \tag{3–1}$$

where J is a convex and possibly nondifferentiable function, and H is convex and continuously differentiable. In TV-based image reconstruction problems, J and H often have the form

$$J(u) = \|u\|_{TV} + \mu\|\Psi u\|_1 \quad \text{and} \quad H(u) = \lambda\|Au - f\|_2^2, \tag{3–2}$$

where $\|\cdot\|_{TV}$ is the total variation, $\|\cdot\|_1$ is the 1-norm, $\|\cdot\|_2$ is the 2-norm (Euclidean norm), A is a possibly large and ill-conditioned matrix describing the imaging device or the data acquisition pattern, f is the measured data, and $\Psi \in \mathbb{C}^{N\times N}$ is an orthogonal

sparsifying matrix. λ and $\mu \geq 0$ are parameters corresponding to the relative weights of the data fidelity term $\|Au - f\|_2^2$ and the terms $\|u\|_{TV}$ and $\|\Psi u\|_1$ which control the solution sparsity. Ψ projects u along a set of basis functions described by, for example, wavelets or a dictionary [50, 69]. It is expected that many components of Ψu will vanish at a solution of (4–15)–(3–2) for magnetic resonance images.

TV-based regularization was originally introduced by Rudin, Osher and Fetami for image denoising in their pioneering work [59]. A significant advantage of TV regularization is that the solution yields a restored clean image with well-preserved edges. The TV and 1-norm terms in (3–2) lead to an underlying sparse solution of $Au = f$. The lack of smoothness in both the TV and 1-norm terms makes the solution of (4–15) difficult.

To cope with the lack of smoothness in J, we introduce an auxiliary variable v to obtain the equivalent constrained problem

$$\min J(v) + H(u) \quad \text{subject to} \quad u = v, \quad u, \;\; v \in \mathbb{C}^N. \tag{3–3}$$

The equality constrained problem is converted to an unconstrained problem using a quadratic penalty:

$$\min_{u,v \in \mathbb{C}^N} J(v) + H(u) + \alpha \|v - u\|_2^2, \tag{3–4}$$

where $\alpha > 0$ is a parameter. The additional variable v allows us to treat the smooth term H and the nondifferentiable term J somewhat independently. Starting from an initial guess u^0, we solve the penalized problem by first minimizing over v with u fixed, and then minimizing over u with v fixed:

$$\left.\begin{aligned} v^{k+1} &= \mathcal{T}(u^k), \quad \mathcal{T}(u) \triangleq \arg\min_{v \in \mathbb{C}^N} J(v) + \alpha \|v - u\|_2^2 \quad \text{(TV)} \\ u^{k+1} &= \mathcal{L}(v^{k+1}), \quad \mathcal{L}(v) \triangleq \arg\min_{u \in \mathbb{C}^N} H(u) + \alpha \|v - u\|_2^2 \quad \text{(LS)} \end{aligned}\right\} \tag{3–5}$$

Since J is convex, the (TV) subproblem is strongly convex in v. Likewise, since H is convex, the (LS) subproblem is strongly convex in u. Hence, for any starting guess u^0,

the iteration sequence (v^k, u^k), $k \geq 1$, exists and is unique. In the imaging context, the first subproblem, denoted (TV), has the same form as TV-wavelet based image denoising which has been extensively studied in the literature, and second subproblem (LS) is a least squares problem. Both subproblems can be solved quickly.

In the literature, algorithms of the form (3–5) are called alternating proximal minimization algorithms. References include [1, 7, 10]. The iterates converge to a solution of (3–4), if a solution exists, according to [10, Cor. 4.5], for example. In general, one needs to let α tend to infinity to obtain the solution of (4–15). However, our numerical experience in partially parallel image reconstruction indicates that in this application, a suitable approximation to the solution of (4–15) is generated using a fixed, not very large α.

Our paper is organized as follows. In Section 3.2 we give an overview of TV-based image reconstruction techniques. Section 3.3 studies the convergence rate of (3–5). In Section 3.4 we present the algorithms that we have used to solve each of the minimization problems in (3–5) when J and H are given by (3–2). Section 3.5 gives an overview of an emerging magnetic resonance imaging technology known as partially parallel imaging (PPI). Finally, Section 3.6 compares our algorithm to operator splitting [51] using PPI generated images.

Notation. For a differentiable function, ∇f denotes the gradient of f, a row vector. More generally, $\partial J(x)$ denotes the subdifferential set at x, a set of row vectors. For any matrix M, $\mathcal{N}(M)$ is the null space of M. $x^\top$ denotes the conjugate transpose of the vector x and $\langle x, y\rangle = x^\top y$ is the Euclidean inner product. $\|\cdot\|_p$ is the p-norm, and $\|\cdot\|_{TV}$ is the discrete total variation semi-norm.

3.2 Related Work

In reviewing the prior work on TV-based image reconstruction, we simplify (3–2) by taking $\mu = 0$. With this simplification, the image reconstruction problem is equivalent to

solving the problem

$$\min_{u\in\mathbb{C}^N} \|u\|_{TV} + \lambda\|Au - f\|_2^2, \tag{3–6}$$

where $\|\cdot\|_{TV}$ is the discrete (isotropic) TV semi-norm defined by

$$\|u\|_{TV} \triangleq \sum_{i=1}^{N} \|D_i u\|_2, \tag{3–7}$$

where $D_i \in \mathbb{R}^{2\times N}$ has two nonzero entries in each row corresponding to finite difference approximations to partial derivatives along the coordinate axes, and N is the number of pixels in the image. The early work on algorithms for (3–6) used gradient descent methods with explicit [59] or semi-implicit schemes [43, 63] in which the TV norm was replaced by a smooth approximation

$$\|u\|_{TV,\epsilon} = \sum_{i=1}^{N} \sqrt{\|D_i u\|_2^2 + \epsilon}. \tag{3–8}$$

The choice of $\epsilon > 0$ was crucial to the reconstruction results and convergence speed. A large ϵ encourages fast convergence rate, but fails to preserve high quality details such as edges in the restored image; a small ϵ better preserves fine structure in the reconstruction at the expense of slow convergence.

In [64, 67], a method is developed based on the following reformulation of (3–6):

$$\min_{u,w} \sum_{i=1}^{N} \|w_i\|_2 + \lambda\|Au - f\|_2^2, \text{ subject to } \quad w_i = D_i u,\ i = 1, \cdots, N. \tag{3–9}$$

The linear constraint was treated with a quadratic penalty

$$\min_{u,w} \sum_{i=1}^{N} \|w_i\|_2 + \beta\|Du - w\|_2^2 + \lambda\|Au - f\|_2^2, \tag{3–10}$$

where $w \in \mathbb{C}^{2N}$ and D is obtained by stacking the D_i matrices. For any fixed β, (3–10) can be solved by alternating minimizations. If both $D^\top D$ and $A^\top A$ can be diagonalized by the Fourier matrix, as they would if A is either the identity matrix or a blurring matrix with periodic boundary conditions, then each minimization involves shrinkage and

a fast Fourier transform (FFT). A continuation method is used to deal with the slow convergence rate associated with a large value for β. The method, however, is not applicable to more general A.

In [34] Goldstein and Osher develop a split Bregman method for (3–10). The resulting algorithm has similar computational complexity to the algorithm in [64]; the convergence is fast and the constraints are exactly satisfied. Later the split Bregman method was shown to be equivalent to the alternating direction method of multipliers (ADMM) [13, 28, 32, 33] applied to the augmented Lagrangian

$$L(w, u, p) \triangleq \sum_{i=1}^{N} \|w_i\|_2 + \lambda \|Au - f\|_2^2 + \langle p, Du - w \rangle + \beta \|Du - w\|_2^2. \tag{3–11}$$

Nonetheless, the algorithms in [34, 64, 67] benefit from the special structure of A, and they lose efficiency if $A^\top A$ cannot be diagonalized by fast transforms. To treat a more general A, the Bregman operator splitting (BOS) method [75] replaces $\|Au - f\|_2^2$ by a proximal-like term $\|u - (u^k - \delta A^\top (Au^k - f))\|^2/\delta$ for some $\delta > 0$. BOS is an inexact Uzawa method that depends on the choice of δ. It is generally less efficient than split Bregman.

There are also several methods developed to solve the associated dual or primal-dual formulations of (3–6) based on the dual formulation of the TV norm:

$$\|u\|_{TV} = \max_{p \in X} \langle p, Du \rangle, \text{ where } X = \{p = (p_1, \dots, p_N) : p_i \in \mathbb{C}^2, \quad 1 \le i \le N\} \tag{3–12}$$

Consequently, (3–6) can be written as a minimax problem

$$\min_{u \in \mathbb{C}^N} \max_{p \in X} \langle p, Du \rangle + \lambda \|Au - f\|_2^2. \tag{3–13}$$

In [20], Chan *et al.* proposed to solve the primal-dual Euler-Lagrange equations using Newton's method. This leads to a quadratic convergence rate and highly accurate solutions; however, the cost per iteration is much higher since the method explicitly uses second-order information and the inversion of a Hessian matrix is required. In [19],

Chambolle used the dual formulation of the TV denoising problem (3–6) with $A = I$, and provided an efficient semi-implicit gradient descent algorithm for the dual. However, the method does not naturally extend to the case with more general A. Recently, Zhu and Chan [78] proposed a primal-dual hybrid gradient (PDHG) method. PDHG alternately updates the primal and dual variables u and p. Numerical results show that PDHG outperforms methods in [19, 34] for denoising and deblurring problems, but its efficiency again relies on the fact that $A^\top A$ can be diagonalized by fast transforms. Later, several variations of PDHG, referred to as projected gradient descent algorithms, were applied to the dual formulation of image denoising problem in [79] to make the method more efficient. Further enhancements involve different step-length rules and line-search strategies, including techniques based on the Barzilai-Borwein method [9].

Another approach that can be applied to (4–15) in the imaging context (3–2) with a general A is the operator splitting (OS) method. In [51] the OS idea of [47] is applied to image reconstruction in compressed magnetic resonance imaging. The scheme is based on the first-order optimality condition at a local minimizer u^*:

$$0 \in \partial J(u^*) + \lambda A^\top (Au^* - f).$$

This is rewritten in the form

$$0 \in \partial J(u^*) + \frac{1}{\delta}(u^* - s^*), \quad s^* = u^* - \delta\lambda A^\top (Au^* - f).$$

The iterative scheme is

$$\begin{aligned} s^k &= u^k - \delta\lambda A^\top (Au^k - f), \\ u^{k+1} &= \arg\min_u J(u) + \frac{1}{2\delta}||u - s^k||_2^2. \end{aligned}$$

The computation of u^{k+1}, given s^k, is a TV-denoising problem. If this problem is solved using a split Bregman method [34], then this is equivalent to the Bregman operator splitting method [75]. In [51] a fixed point iteration is applied to solve the u subproblem;

the iterative scheme combines the OS idea with conjugate duality. The comparison of our method with the algorithm of [51] is given in Section 3.6.

3.3 Convergence Analysis

In this section, we examine the convergence rate of the alternating proximal minimization scheme (3–5). Since H is convex, there exists a constant $\sigma \geq 0$ such that the following monotonicity condition holds for all u and $v \in \mathbb{C}^n$:

$$(\nabla H(u) - \nabla H(v))(u - v) \geq \sigma \|u - v\|_2^2 \tag{3–14}$$

Here, ∇H denotes the gradient, which is a row vector throughout this paper. If $\sigma > 0$, then H is strongly convex. As shown below in Corollary 1, strong convexity of H and convexity of J imply that the objective function in the penalized problem (3–4) is strongly convex, which ensures the existence of a unique minimizer.

Theorem 3.1. *If (3–4) has minimizers v^* and u^*, then for each k we have*

$$\|v^{k+1} - v^*\|_2 \leq \frac{2\alpha}{2\alpha + \sigma} \|v^k - v^*\|_2 \quad \textit{and} \quad \|u^{k+1} - u^*\|_2 \leq \frac{2\alpha}{2\alpha + \sigma} \|u^k - u^*\|_2. \tag{3–15}$$

Proof. It is well-known that the operators $\mathcal{T}$ and $\mathcal{L}$ in (3–5) are nonexpansive relative to the Euclidean norm. That is, for all u and v, we have

$$\|\mathcal{T}(v) - \mathcal{T}(u)\|_2 \leq \|v - u\|_2 \quad \text{and} \quad \|\mathcal{L}(v) - \mathcal{L}(u)\|_2 \leq \|v - u\|_2.$$

This follows from the first-order optimality conditions characterizing the minimizers in (3–5). For example, if $v_i = \mathcal{T}(u_i)$ for $i = 1, 2$, then $2\alpha(u_i - v_i)^{\mathsf{T}} \in \partial J(v_i)$, where ∂ denotes the subdifferential. By the convexity of J, it follows that

$$J(v_2) \geq J(v_1) + 2\alpha(u_1 - v_1)^{\mathsf{T}}(v_2 - v_1). \tag{3–16}$$

Likewise, interchanging v_1 and v_2 gives

$$J(v_1) \geq J(v_2) + 2\alpha(u_2 - v_2)^{\mathsf{T}}(v_1 - v_2). \tag{3–17}$$

We add (3–16) and (3–17) to obtain

$$\|v_2 - v_1\|^2 \leq (u_2 - u_1)^{\mathsf{T}}(v_2 - v_1) \leq \|u_2 - u_1\|_2 \|v_2 - v_1\|_2. \tag{3–18}$$

Hence, $\|v_2 - v_1\|_2 = \|\mathcal{T}(u_2) - \mathcal{T}(u_1)\|_2 \leq \|u_2 - u_1\|_2$, which yields the nonexpansive property.

Since v^* and u^* achieve the minimum in (3–5), we have $v^* = \mathcal{T}(u^*)$. Subtracting this identity from the equation $v^{k+1} = \mathcal{T}(u^k)$ and utilizing the nonexpansive property gives

$$\|v^{k+1} - v^*\|_2 \leq \|\mathcal{T}(u^k) - \mathcal{T}(u^*)\|_2 \leq \|u^k - u^*\|_2. \tag{3–19}$$

The first-order optimality conditions for u^k and u^* are

$$\begin{aligned} \nabla H(u^k) - 2\alpha(v^k - u^k)^{\mathsf{T}} &= 0, \\ \nabla H(u^*) - 2\alpha(v^* - u^*)^{\mathsf{T}} &= 0. \end{aligned}$$

We subtract the second equation from the first and multiply by $(u^k - u^*)$ to obtain

$$\begin{aligned} (\nabla H(u^k) - \nabla H(u^*))(u^k - u^*) + 2\alpha\|u^k - u^*\|_2^2 &= 2\alpha(v^k - v^*)^{\mathsf{T}}(u^k - u^*) \\ &\leq 2\alpha\|v^k - v^*\|_2 \, \|u^k - u^*\|_2. \end{aligned} \tag{3–20}$$

Utilizing the monotonicity condition (3–14) on the left side of (3–20) gives

$$(\sigma + 2\alpha)\|u^k - u^*\|_2^2 \leq 2\alpha\|v^k - v^*\|_2 \, \|u^k - u^*\|_2,$$

which yields

$$\|u^k - u^*\|_2 \leq \left(\frac{2\alpha}{\sigma + 2\alpha}\right) \|v^k - v^*\|_2. \tag{3–21}$$

Combining this with (3–19) gives

$$\|v^{k+1} - v^*\|_2 \leq \left(\frac{2\alpha}{\sigma + 2\alpha}\right) \|v^k - v^*\|_2,$$

the first inequality in (3–15). Combining (3–21), with k replaced by $k + 1$, and the nonexpansive property (3–19) gives the second inequality in (3–15). □

Corollary 1. *If $\sigma > 0$, then the iterates generated by (3–5) converge linearly to the unique minimizer of (3–4).*

Proof. We first observe that when $\sigma > 0$, the objective function in (3–4) strongly convex. Let $F(u, v) = H(u) + \alpha\|v - u\|_2^2$ be the part of the objective which excludes J. By the convexity inequality (3–14), we have

$$\begin{aligned}(\nabla F(u_1, v_1) - \nabla F(u_2, v_2)) \begin{bmatrix} \delta u \\ \delta v \end{bmatrix} &= (\nabla H(u_1) - \nabla H(u_2))(u_1 - u_2) + 2\alpha\|\delta u - \delta v\|_2^2 \\ &\geq \sigma\|\delta u\|_2^2 + 2\alpha\|\delta u - \delta v\|_2^2, \end{aligned} \tag{3–22}$$

where $\delta u = u_1 - u_2$ and $\delta v = v_1 - v_2$. The matrix corresponding to the quadratic in (3–22) is

$$2\begin{bmatrix} \alpha + \sigma/2 & -\alpha \\ -\alpha & \alpha \end{bmatrix}.$$

Since the eigenvalues of this matrix are strictly positive, F is strongly convex. The objective function in (3–4) is the sum $J + F$ of a convex function J and a strongly convex function F. Hence, it is strongly convex and there exists a unique minimizer (u^*, v^*). By Theorem 3.1, the iterates generated by (3–5) converge to (u^*, v^*) linearly. □

In the case $\sigma = 0$, Theorem 3.1 only yields

$$\|v^{k+1} - v^*\|_2 \leq \|v^k - v^*\|_2 \quad \text{and} \quad \|u^{k+1} - u^*\|_2 \leq \|u^k - u^*\|_2, \tag{3–23}$$

which does not imply convergence. On the other hand, by the theory for the alternating proximal minimization algorithm, we know that the iterates do converge. We now observe that the inequalities in (3–23) are strict except when convergence is achieved in a finite number of steps. This result is based on the following property.

Lemma 1. *If $\mathcal{P} : \mathbb{C}^n \to \mathbb{C}^n$ satisfies*

$$\|\mathcal{P}(u) - \mathcal{P}(v)\|_2^2 \leq \langle \mathcal{P}(u) - \mathcal{P}(v), u - v \rangle \tag{3–24}$$

for all u and $v \in \mathbb{C}^n$, then

$$\|\mathcal{P}(u) - \mathcal{P}(v)\|_2 \leq \|u - v\|_2 \tag{3–25}$$

for all u and $v \in \mathbb{C}^n$ with equality only if $\mathcal{P}(u) - \mathcal{P}(v) = u - v$.

Operators satisfying (3–24) are called *firmly nonexpansive*. The fact that the proximal maps $\mathcal{T}$ or $\mathcal{L}$ are firmly nonexpansive is implied by (3–18).

Proof. The inequality (3–25) is a consequence of the Schwarz inequality applied to (3–24). Moreover, by (3–24) we have

$$\begin{aligned} \|(u - v) - (\mathcal{P}(u) - \mathcal{P}(v))\|_2^2 &= \|u - v\|_2^2 - 2\langle \mathcal{P}(u) - \mathcal{P}(v), u - v\rangle + \|\mathcal{P}(u) - \mathcal{P}(v)\|_2^2 \\ &\leq \|u - v\|_2^2 - \|\mathcal{P}(u) - \mathcal{P}(v)\|_2^2. \end{aligned} \tag{3–26}$$

If (3–25) is an equality, then the right side of (3–26) vanishes, which implies that the left side vanishes:

$$(u - v) - (\mathcal{P}(u) - \mathcal{P}(v)) = 0.$$

□

Theorem 3.2. *Suppose that u^* and v^* are optimal in (3–4). If for some k, the iterates of the alternating proximal minimization algorithm (3–5) satisfy $\|u^{k+1} - u^*\|_2 = \|u^k - u^*\|_2$, then $u^j = u^k$ and $v^{j+1} = v^{k+1}$ for all $j > k$. If $\|v^{k+1} - v^*\|_2 = \|v^k - v^*\|_2$ for some k, then $v^j = v^k$ and $u^j = u^k$ for all $j > k$.*

Proof. Suppose that $\|u^{k+1} - u^*\|_2 = \|u^k - u^*\|_2$. Since v^* and u^* are optimal in (3–4), we have

$$(\mathcal{L}\mathcal{T})(u^*) = \mathcal{L}(\mathcal{T}(u^*)) = \mathcal{L}(v^*) = u^*. \tag{3–27}$$

By (3–5), it follows that $u^{k+1} = (\mathcal{L}\mathcal{T})(u^k)$. Hence, the equality $\|u^{k+1} - u^*\|_2 = \|u^k - u^*\|_2$ coupled with the nonexpansive properties of $\mathcal{L}$ and $\mathcal{T}$ yield

$$\begin{aligned} \|u^k - u^*\|_2 &= \|(\mathcal{L}\mathcal{T})(u^k) - (\mathcal{L}\mathcal{T})(u^*)\|_2 = \|\mathcal{L}(\mathcal{T}(u^k)) - \mathcal{L}(\mathcal{T}(u^*))\|_2 \\ &\leq \|\mathcal{T}(u^k) - \mathcal{T}(u^*)\|_2 \\ &\leq \|u^k - u^*\|_2. \end{aligned} \tag{3–28}$$

Since the right and left sides of (3–28) are equal, all the inequalities in (3–28) are equalities. The equality $\|\mathcal{T}(u^k) - \mathcal{T}(u^*)\|_2 = \|u^k - u^*\|_2$ and Lemma 1 imply that

$$\mathcal{T}(u^k) - \mathcal{T}(u^*) = u^k - u^*. \tag{3–29}$$

The equality $\|\mathcal{L}(\mathcal{T}(u^k)) - \mathcal{L}(\mathcal{T}(u^*))\|_2 = \|\mathcal{T}(u^k) - \mathcal{T}(u^*)\|_2$ and Lemma 1 imply that

$$(\mathcal{L}\mathcal{T})(u^k) - (\mathcal{L}\mathcal{T})(u^*) = \mathcal{L}(\mathcal{T}(u^k)) - \mathcal{L}(\mathcal{T}(u^*)) = \mathcal{T}(u^k) - \mathcal{T}(u^*). \tag{3–30}$$

Together, (3–29) and (3–30) yield

$$(\mathcal{L}\mathcal{T})(u^k) - (\mathcal{L}\mathcal{T})(u^*) = u^k - u^*. \tag{3–31}$$

We combine (3–27) and (3–31) to obtain

$$u^{k+1} = (\mathcal{L}\mathcal{T})(u^k) = u^k.$$

Hence, u^k is a fixed point of $(\mathcal{L}\mathcal{T})$ and $u^j = u^k$ for all $j > k$. Since $v^{j+1} = \mathcal{T}(u^j)$, we conclude that $v^{j+1} = v^{k+1}$ for all $j > k$. The equality $\|v^{k+1} - v^*\|_2 = \|v^k - v^*\|_2$ is treated in the same way except that $\mathcal{L}$ and $\mathcal{T}$ are interchanged. □

By the convergence theory for the alternating proximal minimization algorithm, we know that the iterates converge to a solution (u^*, v^*) of (3–4) provided a solution exists. Theorem 3.2 implies that

$$\|u^{k+1} - u^*\|_2 / \|u^k - u^*\|_2 < 1$$

except when $u^k = u^*$. Likewise

$$\|v^{k+1} - v^*\|_2 / \|v^k - v^*\|_2 < 1$$

except when $v^k = v^*$. This implies at least sublinear convergence of the alternating proximal minimization algorithm (3–5).

For any fixed α, the solution of (3–4) generates an approximation to a solution of (4–15). Let α_k, $k \geq 0$, denote an increasing sequence of values for the penalty parameter tending to infinity, and let (U^k, V^k) denote associated solutions of (3–4), assuming they exist. By the theory describing the convergence of the penalty scheme (see [53, Thm. 17.1]), convergent subsequences of the iterates approach a solution of (4–15). We now show in the context (3–2) of image reconstruction that the iterates (U^k, V^k) are bounded.

Theorem 3.3. *Suppose that J and H are given by (3–2). If $\mu \geq 0$, $\lambda > 0$, and $\mathcal{N}(D) \cap \mathcal{N}(A) = 0$, where $\mathcal{N}$ denotes null space, then for each $\alpha_0 > 0$, there exists a compact set K which contains the solutions of (3–4) for all $\alpha \geq \alpha_0$. Moreover, as α tends to infinity, any convergent subsequence of the iterates approaches a solution of either (4–15) or the equivalent problem (3–3).*

Proof. In the special case (3–2), $J(0) = 0$ and $H(0) = \lambda \|f\|_2^2$. Let $\rho = \lambda \|f\|_2^2$ be the value of the objective function value in (3–4) corresponding to $u = v = 0$. For any choice of α, the optimal objective function value in (3–4) is bounded by ρ. Hence, for any choice of α, when minimizing the objective function in (3–4), we should restrict our attention to those u and v satisfying

$$J(v) + H(u) + \alpha \|v - u\|_2^2 \leq \rho. \tag{3–32}$$

Since $J(v) = \|v\|_{TV} + \mu\|\Psi v\|_1 \geq 0$ and $H(u) = \|Au - f\|_2^2 \geq 0$, it follows from (3–32) that

$$\|v - u\|_2 \leq \sqrt{\rho/\alpha}, \tag{3–33}$$

$$\|v\|_{TV} \leq \rho, \tag{3–34}$$

$$\|Au - f\|_2 \leq \sqrt{\rho/\lambda}. \tag{3–35}$$

Decompose $u = u_n + u_p$ where $u_n \in \mathcal{N}(A)$ and u_p is orthogonal to $\mathcal{N}(A)$. By (3–7), (3–33), and (3–34), we have

$$\begin{aligned} \rho \geq \|v\|_{TV} &= \sum_{i=1}^{N} \|D_i v\|_2 \geq \|Dv\|_2 \geq \|Du\|_2 - \|D(v-u)\|_2 \\ &\geq \|Du_n\|_2 - \|Du_p\|_2 - \|D\|_2\|v-u\|_2 \\ &\geq \|Du_n\|_2 - \|Du_p\|_2 - \|D\|_2\sqrt{\rho/\alpha}. \end{aligned} \tag{3–36}$$

Since $\mathcal{N}(D) \cap \mathcal{N}(A) = 0$, there exists $\gamma_1 > 0$ such that

$$\|Du\|_2 \geq \gamma_1\|u\|_2 \text{ for all } u \in \mathcal{N}(A).$$

Hence, by (3–36)

$$\|u_n\|_2 \leq \left(\rho + \|Du_p\|_2 + \|D\|_2\sqrt{\rho/\alpha}\right)/\gamma_1. \tag{3–37}$$

Similarly, there exists $\gamma_2 > 0$ such that

$$\|Au_p\|_2 \geq \gamma_2\|u_p\|_2.$$

Hence, by (3–35), we have

$$\gamma_2\|u_p\|_2 \leq \|Au\|_2 \leq \|f\|_2 + \|Au - f\|_2 \leq \|f\|_2 + \sqrt{\rho/\lambda}. \tag{3–38}$$

Combine (3–37) and (3–38) to deduce that $u = u_n + u_p$ lies in a compact set. By (3–33), we have

$$\|v\|_2 \leq \|u\|_2 + \sqrt{\rho/\alpha},$$

which yields a bound for $\|v\|_2$. As α increases, the level set of (3–4) corresponding to the objective function value ρ can only shrink. Hence, this level set is bounded for any $\alpha \geq \alpha_0$. Let α_k, $k = 0, 1, ...$, denote an increasing sequence of values for the penalty tending to infinity, and let (U^k, V^k) denote associated solutions of (3–4). By [53, Thm. 17.1], every convergent subsequence of the minimizers (U^k, V^k) approaches a solution of (3–3). □

Remark. If H is strongly convex, then (4–15) has a unique solution; hence, any sequence of solutions to (3–4) approaches the unique solution of (4–15) as α tends to infinity.

3.4 Algorithms for the TV and LS Subproblems

We now provide implementations for the (TV) and (LS) subproblems of the alternating proximal minimization algorithm (3–5) in the imaging context (3–2). One of the reasons that the splitting (3–3) worked well was that each of the subproblems could be solved quickly. For fixed u, subproblem (TV) is a TV-wavelet image denoising problem. As discussed earlier, there are many fast algorithms for this problem that take advantage of the simplicity of the $\|v - u\|_2^2$ term. Recent work includes the dual approach in [19], variable splitting and continuation [64, 67], split Bregman [34], primal-dual hybrid gradient [78, 79]. In the numerical experiments of Section 3.6, we used a split Bregman method which is among the fastest methods for TV-wavelet image denoising.

We now explain in detail the split Bregman scheme that we use for the TV subproblem in (3–5). Based on the representation (3–7) for the TV norm, the TV subproblem has the form

$$\min_v \quad \mu\|\Psi v\|_1 + \alpha\|v - u\|_2^2 + \sum_{i=1}^{N} \|D_i v\|_2.$$

Introducing additional variables w and z, where $w_i = D_i v$ and $\Psi v = z$, this is rewritten

$$\min_{v,w,z} \quad \mu\|z\|_1 + \alpha\|v - u\|_2^2 + \sum_{i=1}^{N} \|w_i\|_2 \quad \text{subject to} \quad Dv = w, \quad \Psi v = z.$$

We apply the split Bregman scheme given as Algorithm (A_1) in [76], but without the proximal terms. Let b^k and c^k denote approximations to multipliers for the constraints $Dv = w$ and $\Psi v = z$, and let β_1 and β_2 denote corresponding constraint penalties. The iteration has the following form:

Algorithm 3 Split Bregman [76] for TV Subproblem

$$
\begin{aligned}
v^{k+1} &= \arg\min_{v\in\mathbb{C}^N} \alpha\|v-u\|_2^2 + (Dv - w^k)^{\mathsf{T}} b^k + (\Psi v - z^k)^{\mathsf{T}} c^k \\
&\qquad +\beta_1\|Dv - w^k\|_2^2 + \beta_2\|\Psi v - z^k\|_2^2 && \text{(3–39)}\\
w_i^{k+1} &= \arg\min_{w_i\in\mathbb{C}^2} \|w_i\|_2 + (D_i v^{k+1} - w_i)^{\mathsf{T}} b_i^k + \beta_1\|D_i v^{k+1} - w_i\|_2^2, \quad 1\le i\le N && \text{(3–40)}\\
z^{k+1} &= \arg\min_{z\in\mathbb{C}^N} \|z\|_1 + (\Psi v^{k+1} - z)^{\mathsf{T}} c^k + \beta_2\|\Psi v^{k+1} - z\|_2^2 && \text{(3–41)}\\
b^{k+1} &= b^k + 2\beta_1(Dv^{k+1} - w^{k+1}) && \text{(3–42)}\\
c^{k+1} &= c^k + 2\beta_2(\Psi v^{k+1} - z^{k+1}) && \text{(3–43)}
\end{aligned}
$$

In Algorithm 1, (3–39) is a well-conditioned least squares problem; it can be solved by an iterative method such as Gauss-Seidel or by an FFT if the associated image satisfies periodic boundary conditions [64]. For the steps (3–40) and (3–41), there are closed form solutions [34, 64, 67], the 2D and componentwise shrinkage operators. The final steps (3–42) and (3–43) are the first-order multiplier updates. Based on the results given in [34], Algorithm 1 is expected to be very efficient.

The LS subproblem in (3–5) is a least-squares problem in u. This could be solved by a conjugate gradient method, however, we have obtained comparable or better performance using the Barzilai-Borwein [9] method (BB), which handles ill-conditioning much better than gradient methods with a Cauchy step [4]. The LS subproblem has the form

$$\min_u \lambda\|Au - f\|_2^2 + \alpha\|v-u\|^2. \qquad \text{(3–44)}$$

In the standard implementation of the BB method, the Hessian of the objective function is approximated by a multiple of the identity matrix. For the LS problem, however, the Hessian of $\|v-u\|^2$ with respect to u already a multiple of the identity. Hence, we only

approximate the Hessian of $\|Au - f\|_2^2$ by a multiple of the identity. More precisely, if u^k is the current BB iterate, then we employ the approximation

$$\|Au - f\|_2^2 \approx \|Au^k - f\|_2^2 + 2(Au^k - f)^\mathsf{T} A(u - u^k) + \delta^k \|u - u^k\|_2^2, \tag{3–45}$$

where

$$\delta^k = \|A(u^k - u^{k-1})\|_2^2 / \|u^k - u^{k-1}\|_2^2.$$

Since the $\|Au^k - f\|_2^2$ term in (3–45) does not depend on u, the BB method for the LS subproblem is as follows:

Algorithm 4 BB method [9] for LS Subproblem

$$u^{k+1} = \arg\min_{u \in \mathbb{C}^N} \lambda \left(2(Au^k - f)^\mathsf{T} A(u - u^k) + \delta^k \|u - u^k\|_2^2\right) + \alpha \|v - u\|^2. \tag{3–46}$$

The iteration (3–46) converges linearly to a solution of (3–44) by [24, 25, 31]. Each iteration involves multiplication by A and A^T. In the application to partially parallel imaging developed in the next section, A is represented as a product $M\mathcal{F}S_j$ where the matrix M is the identity matrix with some rows removed, $\mathcal{F}$ is a Fourier transform, and S_j is a diagonal matrix. The time to multiply by M or S_j is bounded by a constant times N, while the Fourier transform can be performed in time proportional to $N \log(N)$. Hence, each iteration of Algorithm 4 can be performed quickly in our target application.

3.5 Partially Parallel Imaging (PPI)

Magnetic resonance (MR) imaging is a medical imaging technique commonly used in radiology to visualize the internal structure and function of the body by non-invasive and non-ionizing means. It provides better contrast between the different soft tissues than most other modalities. MR images are obtained through an inversion of Fourier data acquired by the receiver coil(s).

Magnetic resonance images are obtained by placing an object in a strong magnetic field and then turning on and off a radio frequency electromagnetic field. Different body parts produce different signals which are detected by receivers. The resulting data is

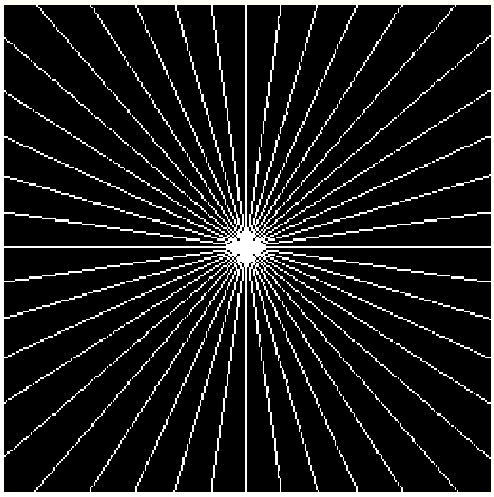


Figure 3-1. PPI mask

then inverted to obtain an image of the scanned object. To improve the quality of the image, a number of receivers are pointed at the scanned object from different directions, and data is collected in parallel. To accelerate the imaging process, only part of the Fourier components are recorded by the receiver. This technology based on collecting in parallel partial Fourier data from different coil arrays is called *partially parallel imaging* or PPI.

The undersampling patterns of the Fourier coefficients are often described by a mask. Figure 3-1 shows a simulated radial mask for a 2D image. The white pixels correspond to the Fourier component which are measured. The white region in the center of the mask indicates that the low frequency Fourier components are all measured, while the white rays in the surrounding darker region shows the spacing between the higher frequency Fourier components that are measured.

Partial data acquisition increases the spacing between regular subsequent read-out lines, thereby reducing scan time, however, this reduction in the number of recorded Fourier components leads to aliasing artifacts in images. There are two general approaches for removing the aliasing artifacts and reconstructing high quality images, image domain-based methods and k-space based methods. The k-space based

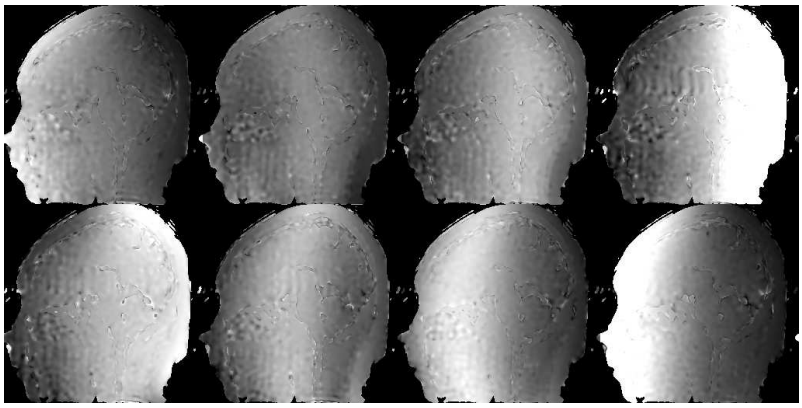


Figure 3-2. Sensitivity maps for an 8-channel coil

methods use coil sensitivity variations to reconstruct the missing k-space data, and then apply the Fourier transform to the original and reconstructed data to obtain the unaliased image [6, 35, 48]. In this paper, we employ image domain methods and coil sensitivity maps to reconstruct the underlying image [15, 29, 54–57, 60, 70, 73].

Sensitivity Encoding (SENSE) is the most common image domain-based parallel imaging method. It is based on the following equation which relates the partial k-space data f_j, acquired by the j-th channel, to the sensitivity map S_j and the mask M:

$$M\mathcal{F}(S_j \odot u) = f_j \tag{3–47}$$

Here $\odot$ is the Hadamard (or componentwise) product between two vectors, f_j is the vector of measured Fourier coefficients at receiver j, M is the mask which is obtained by extracting from the identity those rows corresponding to the measured Fourier components, $\mathcal{F}$ is the Fourier transform, $S_j \in \mathbb{C}^N$ is the sensitivity map for receiver j, and $u \in \mathbb{C}^N$ is the underlying image gotten by stacking all columns of the image to form a one dimensional vector. The sensitivity map is an estimate of the impact of a pixel in the image on the measured Fourier coefficients. Pixels closest to a receiver may have more impact on the signal than pixels far away from the receiver. An example of the sensitivity map for an 8-channel coil appears in Figure 3-2.

Based on (3–47), the reconstruction of the image u could be accomplished by solving the least squares problem

$$\min_{u\in\mathbb{C}^N} \sum_{j=1}^{J} \|M\mathcal{F}(S_j \odot u) - f_j\|_2^2, \tag{3–48}$$

where J is the number of channels. However, (3–48) often does not have a unique solution and the minimization problem can be ill-conditioned. To alleviate the effect of the ill-conditioning, the SENSE model (3–48) has been improved recently by incorporating regularization terms into the energy functional to take advantage of the underlying sparsity of MR images in the finite difference domain and wavelet transform domain [50]. The images are recovered by solving an optimization problem of the form

$$\min_{u\in\mathbb{C}^N} \|u\|_{TV} + \mu\|\Psi u\|_1 + \lambda \sum_{j=1}^{J} \|M\mathcal{F}(S_j \odot u) - f_j\|_2^2. \tag{3–49}$$

The first two terms in (3–49) correspond to J in (3–2) while the last term corresponds to $H(u) = \lambda\|Au - f\|_2^2$.

3.6 Numerical Experiments

3.6.1 Data Acquisition and Experimental Setup

In this section we give results for three PPI reconstructions based on the algorithm (3–5). We compare performance to that of the operator splitting (OS) in [51]. All k-space data were fully acquired with 8-channel head coil as illustrated in Figure 3-3. By full acquisition we mean that each receiver coil obtains the complete k-space data and hence a high resolution image. One of the data sets, denoted SB512, was a set of sagittal Cartesian brain images acquired on a 3T GE system (GE Healthcare, Waukesha, Wisconsin, USA). The data acquisition parameters were FOV 220mm^2, size $512 \times 512 \times 8$, TR 3060ms, TE 126ms, slice thickness 5mm, flip angle 90°, and phase encoding direction was anterior-posterior. The second data set was a radial brain data set acquired on a 1.5T Siemens Symphony system (Siemens Medical Solutions, Erlangen, Germany). The acquisition parameters were FOV 220mm^2, size $256 \times 512 \times 8$,

slice thickness 5mm, TR 53.5ms, TE 3.4ms, and flip angle 75°. A Cartesian data set with full k-space, denoted AB512, with size $512 \times 512 \times 8$, was generated by GRAPPA operator griding [61] which can shift non-Cartesian radial data to Cartesian grids. The third data set of size $256 \times 256 \times 8$, denoted AB256, has full k-space simulated by the complete central k-space of AB512. We simulated a PPI scan by undersampling the actual data using a radial mask similar to that shown in Figure 3-1. We used 88 radial lines corresponding to sampling ratios of 33.5% for AB256 and of 16.8% for SB512 and AB512.

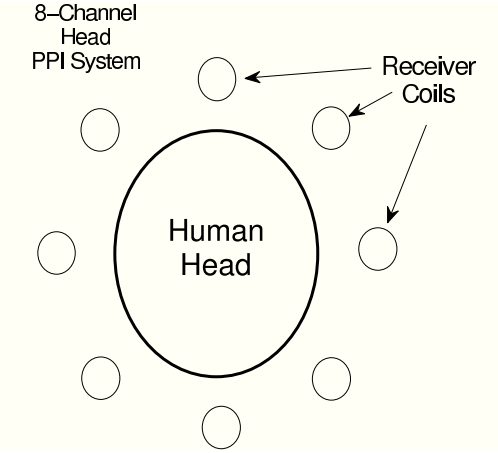


Figure 3-3. An illustration of a head PPI system with eight receiver coils.

The sensitivity maps were estimated using center k-space data of size 32×32, which is a subset of the partially sampled data. The estimated sensitivity maps for SB512 are shown in Figure 3-2. In all experiments, the sensitivity maps were obtained in the same way; the maps were fixed during the reconstruction process. Algorithms were implemented in MATLAB®, Version R2009b. All the experiments were performed on a Lenovo laptop with Intel®Dual Core 2 Duo 2.53 GHz processors and a Windows™operating system. Only 1 core was used in the computations.

3.6.2 Comparison Algorithm

Many of the algorithms in Section 3.2 are not very effective for PPI imaging due to the complicated structure of A. For comparison, we chose the operator splitting (OS)

Table 3-1. Tested Data and Model Parameters

Data	Image	Abbrev.	Size	Sample Ratio	(μ, λ)
1	Axial Brain	AB256	$256 \times 256 \times 8$	33.5%	(.25, 500)
2	Sagittal Brain	SB512	$512 \times 512 \times 8$	16.8%	(.25, 500)
3	Axial Brain	AB512	$512 \times 512 \times 8$	16.8%	(.25, 500)

scheme from [51], which is relatively efficient and only requires the computations of A and $A^\top$ in each iteration. The OS scheme of [51], with a minor correction, is as follows:

$$\begin{aligned} s^{k+1} &= \Psi u^k - (\delta_1/\lambda)\Psi\left(D^\top w^k + \lambda A^\top(Au^k - f)\right), \\ t_i^{k+1} &= w_i^k + \delta_2 D_i u^k, \quad 1 \le i \le N, \\ u^{k+1} &= \Psi\left(\text{sign}(s^{k+1}) \odot \max\{|s^{k+1}| - \delta_1\tau/\lambda, 0\}\right), \\ w_i^{k+1} &= \min(1, \|t_i^{k+1}\|) \cdot t_i^{k+1}/\|t_i^{k+1}\|_2 \quad 1 \le i \le N, \end{aligned}$$

where $s^k \in \mathbb{C}^N$, t_i^k and $w_i^k \in \mathbb{C}^2$, $i = 1, \cdots, N$, $w^k \in \mathbb{C}^{2N}$ is formed by stacking the two columns of matrix $(w_1^k, \cdots, w_N^k)^\top$, and the "max" and "sign" operations in the computation of u^{k+1} are componentwise operations corresponding to shrinkage. The main computational cost per iteration in the OS scheme corresponds to the following operations: a 2D shrinkage during the computation of w^{k+1}, a projection during the computation of u^{k+1}, two wavelet transforms during the computation of s^{k+1} and u^{k+1}, and two (inverse) Fourier transforms during the computation of s^{k+1}. In [51] it is shown that for $\delta_1, \delta_2 > 0$ in certain ranges, the OS scheme converges to a fixed point which is also a solution of (4–15). The iterations were stopped when either the following conditions were satisfied:

$$\|u^{k+1} - u^k\|_2 / \max\{1, \|u^k\|_2\} < \epsilon_1 \text{ or } (f^k - f^{k+1})/\max\{1, f^k\} < \epsilon_2\sqrt{\tau_c/\tau_t}, \qquad (3\text{–}50)$$

where f^k is the objective value of (4–15) at u^k, τ_c and τ_t are the current and target values of τ respectively and ϵ_1 and ϵ_2 are prescribed stopping tolerances.

3.6.3 Experimental Results

The parameters values and test problems are summarized in Table 3-1. In all

experiments, we set $\delta_1 = \delta_2 = 1$, $\epsilon_1 = 10^{-4}$ and $\epsilon_2 = 10^{-3}$ for the OS scheme. For the alternating minimization (AM) scheme (3–5), we set $\alpha = 2^5$ and we used the stopping criterion $(f^k - f^{k-1})/f^k \leq 10^{-3}$. For this stopping criterion, the image quality for the OS and the AM schemes are comparable. The reconstructed images for SB512 are shown in Figures 3-4B and 3-4C. The root mean squared error (RMSE) of the image u, given by $\|u - \bar{u}\|_2/\|\bar{u}\|_2$ where $\bar{u}$ is the reference image reconstructed from fully acquired data, was 13.1% for the OS scheme and 12.5% for the AM scheme. In Figures 3-4E and 3-4F, we zoom into the square shown in Figure 3-4A. It is seen that both methods adequately recovered the image while AM has slightly better preserved edges. The reconstructions for AB512 are shown in Figures 3-5B and 3-5C. The RMSE was 14.7% for OS and 13.9% for AM. Since AB512 does not contain obvious fine structures when compared to SB512 (see Figure 3-4D), we compare the differences of these two reconstructions to the reference image (Figure 3-5A) under the same contrast level. Figure 3-5A shows the followings: reference image, reconstruction by OS with RMSE=14.7%, reconstruction by AM with RMSE=13.9%, and the differences of reconstructions by OS (left) and AM (right) to the reference image displayed under the same contrast level. From Figure 3-5D, we can see the reconstruction by AM (right) has much less systematic error than that of OS (left), and hence is more likely to prevent loss of important diagnostic information such as edges.

To examine the efficiency of the proposed algorithm AM compared to OS, we tracked their objective function values and reconstruction errors during the computation processes for data sets AB256 and SB512. As the per iteration costs for two algorithms OS and AM are quite different, we compared the normalized RMSE and objective function values versus CPU time, which are plotted in Figures 3-6A and 3-6B, respectively. From Figures 3-6A and 3-6B, we can see both AM and OS converged faster for the smaller image AB256 than for SB512. For both of AB256 and

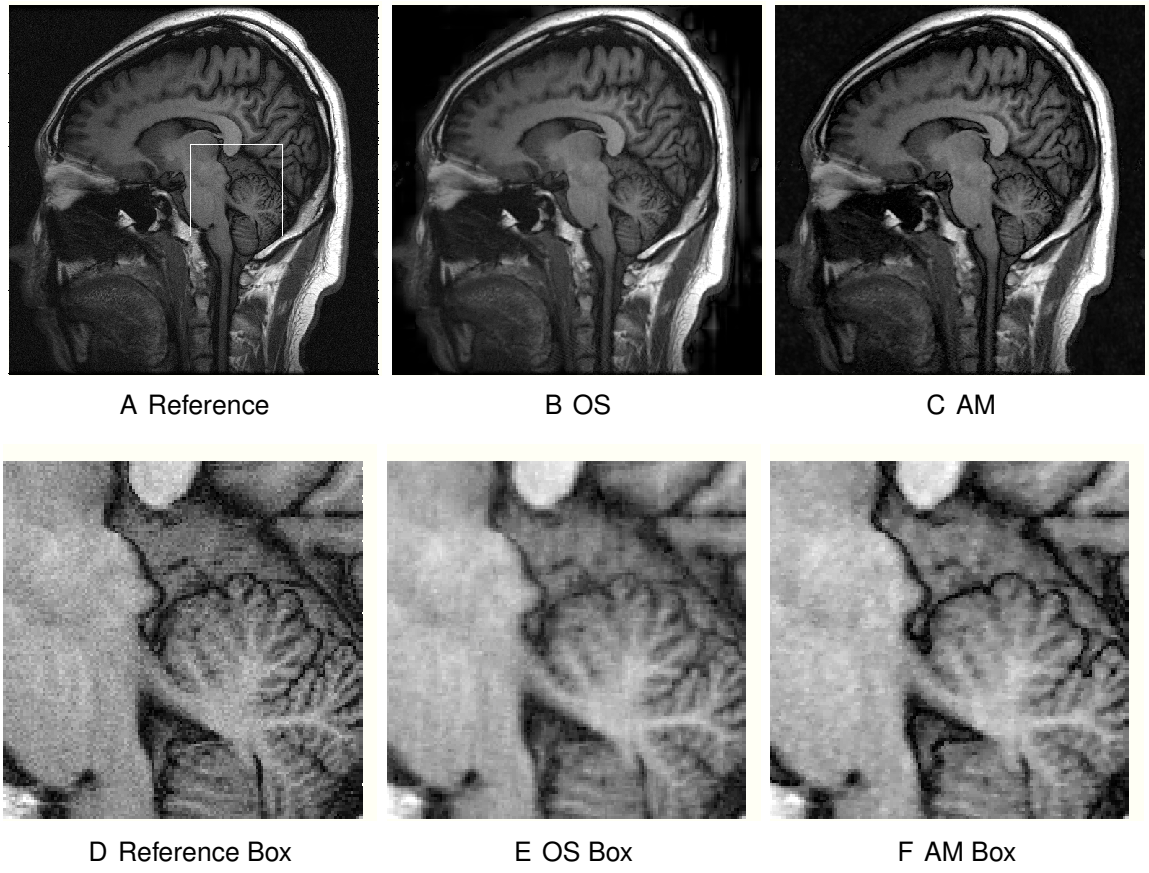
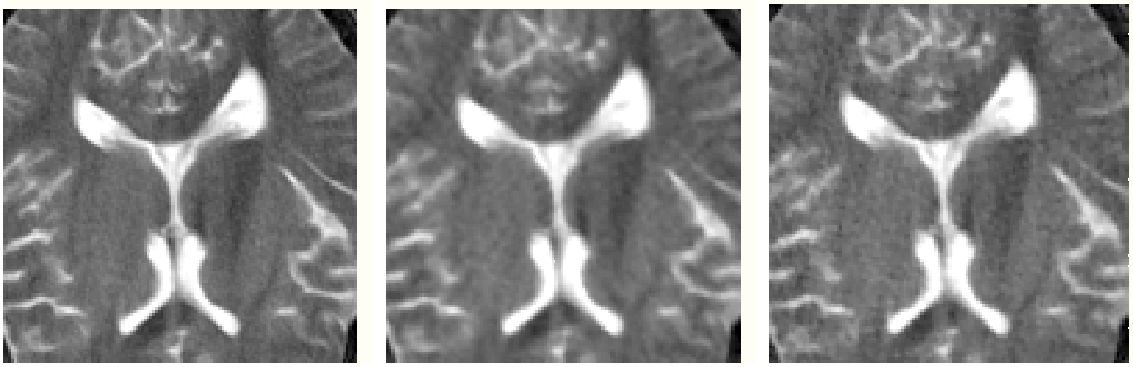


Figure 3-4. Reconstructed images by OS and AM for SB512.

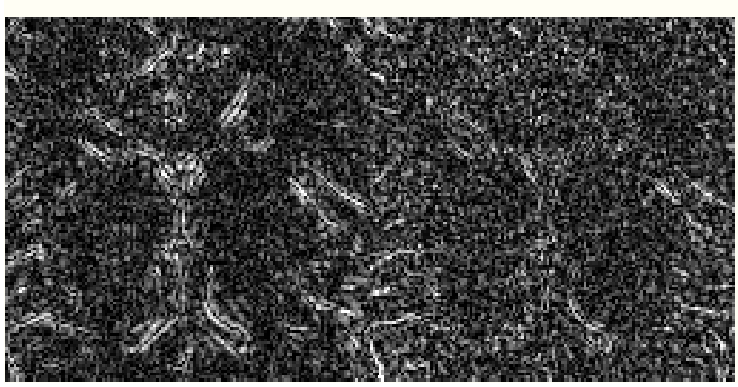
SB512, AM consistently reached and maintained the same or lower RMSE and objective functions than OS in much less CPU time.

3.7 Concluding Remarks

A fast numerical algorithm for total variation-based image reconstruction was developed and analyzed. The proposed method employs variable splitting, a quadratic penalty, and an alternating proximal minimization algorithm (AM). Linear convergence was established when the smooth part of the objective function was strongly convex, while the convergence was sublinear under a weaker convexity assumption. One phase of the alternating proximal minimization algorithm represents a total variation denoising problem, and the other phase is an ill-conditioned linear system. An implementation based on a split Bregman scheme for the TV problem and a Barzilai-Borwein method



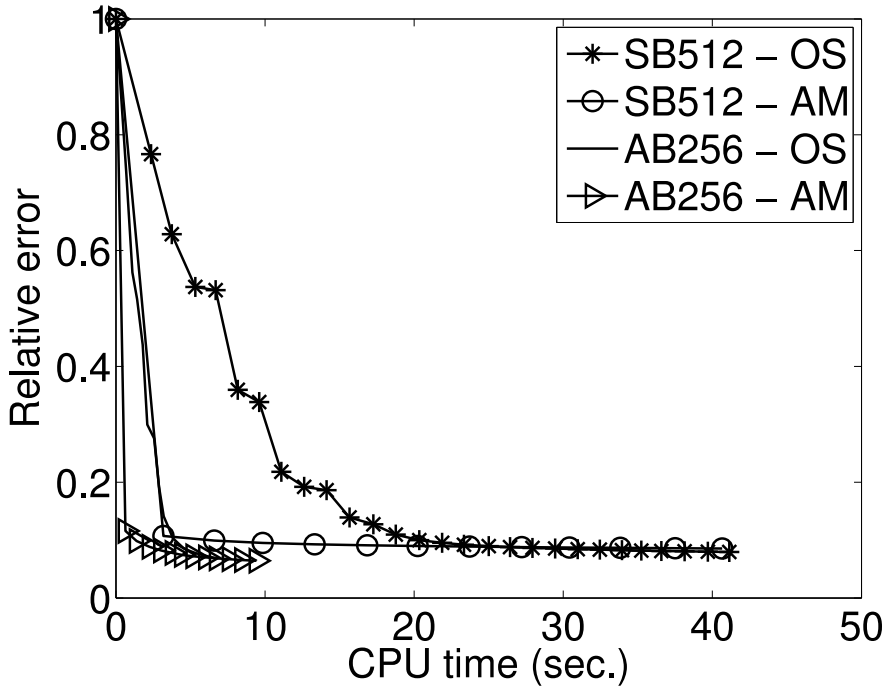
A Reference B OS C AM



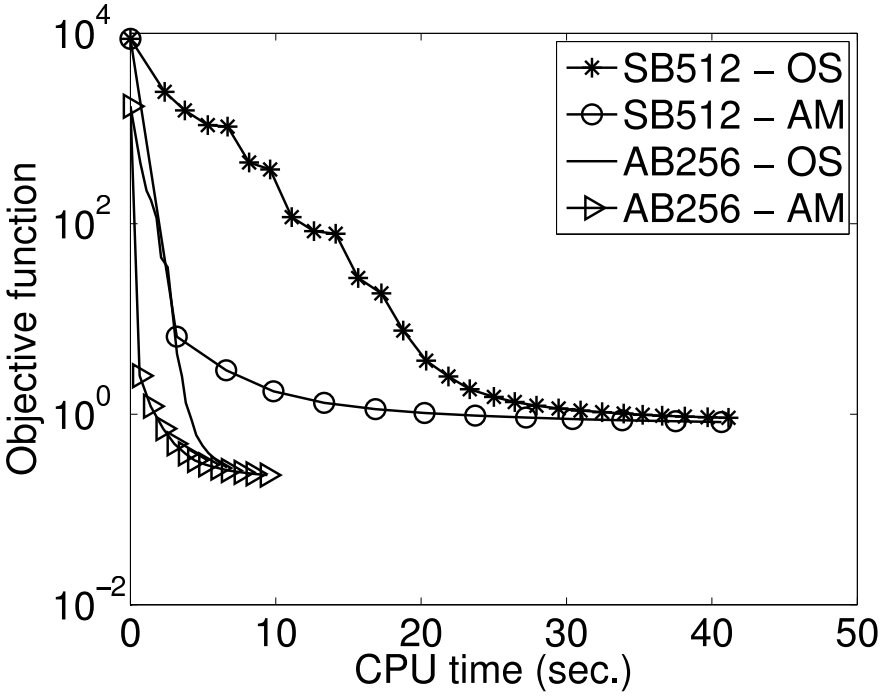
D Differences to Reference image

Figure 3-5. Reconstructions by OS and AM for AB512.

for the linear system was proposed. Numerical performance was evaluated using image reconstruction problems that arose from clinical applications of partially parallel magnetic resonance imaging (PPI). The performance of the alternating proximal minimization algorithm was compared to that of an operator splitting algorithm. The numerical results show excellent performance for the proposed algorithm in terms of efficiency and accuracy in reconstruction, which suggests its great potential for practical use.



A Relative errors versus CPU time.



B Objective functions versus CPU time.

Figure 3-6. Comparison of OS and AM on data sets AB256 and SB512.

CHAPTER 4
INVERSE CONSISTENT DEFORMABLE IMAGE REGISTRATION

Outline

This paper presents a novel variational model for inverse consistent deformable image registration. The proposed model deforms both source and target images simultaneously, and aligns the deformed images in the way that the forward and backward transformations are inverse consistent. To avoid the direct computation of the inverse transformation fields, our model estimates two more vector fields by minimizing their invertibility error using the deformation fields. Moreover, to improve the robustness of the model to the choice of parameters, the dissimilarity measure in the energy functional is derived using the likelihood estimation. The experimental results on clinical data indicate the efficiency of the proposed method with improved robustness, accuracy and inverse consistency.

4.1 Backgound of Consistent Image Registration

Image registration is a very important subject that has been widely applied in medical research and clinical applications. The task of image registration is to find a transformation field that relates points in the source image to their corresponding points in the target image. Deformable image registration allows localized transformations, and is able to account for internal organ deformations. Therefore, it has been increasingly used in health care to assist diagnosis and treatments. In particular, deformable image registration has become a critical technique for image guided radiation therapy. It allows more precise tumor targeting and normal tissue preservation. A comprehensive review of image registration in radiation therapy can be found in [40].

A deformable image registration is called inverse consistent, if the correspondence between two images is invariant to the order of the choice of source and target. More precisely, let S and T be the source and target images, and h and g be the forward and

backward transformations, respectively, i.e.

$$S \circ h = T \text{ and } T \circ g = S,$$

then an inverse consistent registration satisfies $h \circ g = id$ and $g \circ h = id$, where id is the identity map. This can be illustrated by the following diagram with constraints $g = h^{-1}, h = g^{-1}$:

$$\boxed{S} \underset{g}{\overset{h}{\leftrightarrows}} \boxed{T}, \tag{4–1}$$

where each of the two squares in (4–1) represents the domain on which the labeled image is defined. By applying an inverse consistent registration, measurements or segmentations on one image can be precisely transferred to the other. In imaging guided radiation therapy, the inverse consistent deformable registration technique provides the voxel-to-voxel mapping between the reference phase and the test phase in four-dimensional (4D) radiotherapy [49]. This technique is referred to ”automatic re-contouring”.

Inverse consistent deformable image registration has been an active subject of study in the literature. There has been a group of work developed in the context of large deformation by diffeomorphic metric mapping, e.g. [8, 12, 37, 39]. The main idea of this method is modeling the forward and backward transformations as a one-parameter diffeomorphism group. Then, a geodesic path connecting two images is obtained by minimizing an energy functional symmetric to the forward and backward transformations. This type of models produce a very good registration results. However, it take long time to compute, since strong regularization of the mappings are required.

Variational method is one of the popular approaches for inverse consistent deformable image registration. This method minimizes an energy functional(s) symmetric to the forward and backward transformations, and in general, consists of three parts: regularization of deformation fields, dissimilarity measure of the target and deformed source images, and penalty of inverse inconsistency [5, 23, 58, 77]. In [23],

Christensen and Johnson proposed to minimize the following coupled energy functionals with respect to h and g alternately:

$$\begin{cases} E(h) = & \lambda E_s(S \circ h, T) + E_r(u) + \rho \|h - g^{-1}\|^2_{L^2(\Omega)} \\ E(g) = & \lambda E_s(T \circ g, S) + E_r(v) + \rho \|g - h^{-1}\|^2_{L^2(\Omega)} \end{cases}, \tag{4–2}$$

where u and v are forward and backward deformation fields corresponding to h and g, respectively, i.e. $h(x) = x + u(x)$ and $g(x) = x + v(x)$. The dissimilarity measure E_s and the regularization of the deformation field E_r are defined by

$$E_s(S \circ h, T) = \|S \circ h - T\|^2_{L^2(\Omega)}, \quad E_r(u) = \|a\Delta u + b\nabla(\operatorname{div} u) - cu\|^2_{L^2(\Omega)}$$

with positive constants $a, b, c > 0$. The last term in both energy functionals enforces the inverse consistency of h and g. The solution (u, v) to (4–2) is obtained by iteratively solving a system of two evolution equations associated with their Euler-Lagrange (EL) equations. This model gives considerably good results with parameters chosen carefully. However, it needs to compute the inverse mappings g^{-1} and h^{-1} explicitly in each iteration, which is computationally intensive can cause cumulated numerical errors in the estimation of inverse mappings.

The variational models developed in [5] and [77] have the same framework as in [23], but with different representations of E_s, E_r, and inverse consistent constraints. In [5] and [77] the terms $\|h \circ g(x) - x\|^2_{L^2(\Omega)}$ and $\|g \circ h(x) - x\|^2_{L^2(\Omega)}$ are used in the energy functional to enforce the inverse consistency. By using these terms the explicit computation of the inverse transforms of h and g can be avoided during the process of finding optimal forward and backward transformations. The similarity measure in [77] is mutual information for multi-modal image registration. The $E_s(S \circ h, T)$ in [5] is $\|S \circ h - T\|^2_{L^2(\Omega)} / \max|DT|$. The regularization term $E_r(u)$ in [77] is a function of Du, and that in [5] is a tensor based smoothing which is designed to prevent the transformation fields from being smoothed across the boundaries of features. In [71, 72] the proposed

models incorporated stochastic errors in the inverse consistent constraints for both forward and backward transformations.

In [45], Leow *et al.* proposed a non-variational approach that updates the forward and backward transformations simultaneously by a force that reduces the first two terms in $E(h)$ and $E(g)$ in (4–2) and preserves the inverse consistency. However, in order to simplify the computation this algorithm only takes linear order terms in the Taylor expression to approximate the inverse consistent conditions for updated transformation fields. As a consequence, the truncating errors can be accumulated and exaggerated during iterations. This can lead to large inverse consistent error, despite that it can produce a good matching quickly [74].

In this paper we propose a novel variational model to improve the accuracy, robustness and efficiency of inverse consistent deformable registration. As an alternate to the current framework of variational methods which finds the forward and backward transformations that deform a source image S to match a target image T and vice versa, we propose to deform S and T simultaneously, and let the registration align the deformed source and deformed target images. It is clear that the disparity between deformed S and deformed T is smaller than that between deformed S and fixed T or deformed T and fixed S. Therefore, the deformation by the bidirectional simultaneous deformations is in general smaller than the deformation by unidirectional deformation that deforms S full way to T or T full way to S. As shown in section 4.5, deforming S and T simultaneously leads to a faster and better alignment than deforming S to the fixed T or vice versa. Let ϕ and $\tilde{\phi}$ represent the transformation fields such that $S \circ \phi$ matches $T \circ \tilde{\phi}$. It is not difficult to verify that if ϕ and $\tilde{\phi}$ are invertible, then the registrations from S to T, and T to S are inverse consistent. To avoid the direct computation of the inverse transformations of ϕ and $\tilde{\phi}$, our model seeks for two additional deformation fields ψ, $\tilde{\psi}$ such that ϕ and ψ are inverse to each other, and the same for $\tilde{\phi}$ and $\tilde{\psi}$. Moreover, the registration process enforces certain regularization of

these four deformation fields, and aligns the deformed S and deformed T. Then, the optimal inverse consistent transformations from S to T, and T to S can be obtained simply by appropriate compositions of these four transformations.

The idea of deforming S and T simultaneously has been adopted in the models where the forward or backward transformation is modeled as a one-parameter diffeomorphism group [8]. However, our model finds regularized invertible deformation fields by minimizing the L^2 norms of the deformation fields and inverse consistent errors rather than a one-parameter diffeomorphism group, whose computational cost is very expensive and hence hinders its application in clinical use. Moreover, our model allows parallel computations for all the deformation fields to significantly reduce the computational time.

Furthermore, to improve the robustness of the model to noises and the choice of the parameter λ that balances the goodness of matching and smoothness of the deformation fields (see the λ in $E(h)$ and $E(g)$ of (4–2)), we adopt the maximum likelihood estimate (MLE) that is able to accommodate certain degree of variability in matching to improve the robustness and accuracy of the registration. By using MLE, the ratio of weighting parameters on the sum of squared distance (SSD) of the residue image $S \circ \phi - T \circ \tilde{\phi}$ and the regularization term is not a fixed λ, but λ/σ^2 (see (4–18) below). This results in a self-adjustable weighting factor that makes the choice of λ more flexible, and also speeds up the convergence to the optimal deformation field.

The rest of the paper is organized as follows. In section 4.2, we present a detailed description of the proposed model. The existence of solutions to the proposed model is shown in section 4.3. The calculus of variation and an outline of a fast algorithm for solving the proposed model numerically are provided in section 4.4. In section 4.5, we present the experimental results on clinical data, and the application in auto re-contouring. The last section concludes the paper.

4.2 Proposed Method

Let S and T be the source and target images defined on Ω_S and Ω_T in $\mathbb{R}^d$, respectively. Note that, in real applications, Ω_S and Ω_T are usually fully overlapped. For simplicity we assume that images S and T are real-valued functions with continuous derivatives. Let $|\cdot|$ denote the absolute value (length) of a scaler (vector) in Euclidean spaces, and $\|\cdot\|$ denote $\|\cdot\|_{L^2(\Omega)}$ henceforth. We also extend this notation to vector-valued functions whose components are in L^2 or H^1: $u = (u_1, \cdots, u_d)^\top$ with each component $u_j \in H^1(\Omega)$, $j = 1, \cdots, d$, there is

$$\|u\|_{H^1(\Omega)} \triangleq \left(\|u\|^2 + \|Du\|^2\right)^{1/2}$$

and

$$\|u\| \triangleq \left(\sum_{j=1}^{d} \|u_j\|^2\right)^{1/2}, \quad \|Du\| \triangleq \left(\sum_{j=1}^{d} \|Du_j\|^2\right)^{1/2},$$

where

$$\|u_j\| = \left(\int_\Omega |u_j(x)|^2 dx\right)^{1/2} \quad \text{and} \quad \|Du_j\| = \left(\int_\Omega |Du_j(x)|^2 dx\right)^{1/2},$$

for $j = 1, \cdots, d$.

4.2.1 Motivation and Ideas of Proposed Method

In this paper, we propose a novel variational model for inverse consistent deformable registration to improve its efficiency and robustness. Our idea differs from the current framework which deforms source image S to target image T, or vice versa: as an alternate, we propose to deform S and T simultaneously, and match both deformed images. This means that ideally we pursuit for a pair of half-way transforms $\phi : \Omega_S \to \Omega_M$ and $\tilde{\phi} : \Omega_T \to \Omega_M$ such that $S \circ \phi = T \circ \tilde{\phi}$, where Ω_M is the region where $S \circ \phi$ and $T \circ \tilde{\phi}$ have overlap. To ensure the transformations from S to T and T to S are inverse consistent, the transforms ϕ and $\tilde{\phi}$ are required to be invertible (but not necessarily to be inverse to each other). Hence, our purpose is to find the transformations ϕ and $\tilde{\phi}$ such

that

$$S \circ \phi = T \circ \tilde{\phi}, \quad \phi, \tilde{\phi} \text{ invertible.} \tag{4–3}$$

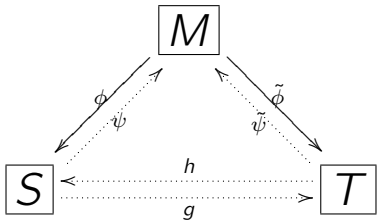
To avoid direct computation of inverses of ϕ and $\tilde{\phi}$ during iterations, we enforce the invertibility of ϕ and $\tilde{\phi}$ by finding another two transformations $\psi : \Omega_M \rightarrow \Omega_S$ and $\tilde{\psi} : \Omega_M \rightarrow \Omega_T$ such that

$$\begin{aligned} \psi \circ \phi &= id, \qquad \phi \circ \psi = id, \\ \tilde{\psi} \circ \tilde{\phi} &= id, \qquad \tilde{\phi} \circ \tilde{\psi} = id. \end{aligned} \tag{4–4}$$

Once we obtained such ψ and $\tilde{\psi}$, we can construct the objective full-way transformations h and g as follows,

$$h = \phi \circ \tilde{\psi}, \quad g = \tilde{\phi} \circ \psi.$$

It is easy to see that h and g satisfy the inverse consistent constraints $h \circ g = g \circ h = id$. This idea is illustrated by the following diagram, where M is an intermediate image.



(4–5)

Since by deforming S and T simultaneously the difference between deformed S and deformed T at each iteration, in general, is smaller than that between deformed S and fixed T, or deformed T and fixed S, the computational cost of deforming both S and T is much less than the conventional one that deform S all the way to T and T to S. In particular, if the underlying deformations of h and g are large, deforming both S and T can make the each deformation of ϕ and $\tilde{\phi}$ in the proposed model almost half smaller than that of h and g, and achieve a faster convergence for the computation of ϕ and $\tilde{\phi}$. Also, seeking ψ and $\tilde{\psi}$ along with ϕ and $\tilde{\phi}$ avoids direct computation of inverse

transformations in each iteration as that in (4–4), which usually causes cumulated errors during iterations if using approximations of the inverses.

Moreover, regularizing the deformation fields is very important to obtain physically meaningful and accurate registrations. Also, if the energy functional consists of only dissimilarity measures and invertible constraints, it is ill-posed in general. Therefore, we propose the following framework for deformable inverse consistent registration:

$$\min_{\phi,\tilde{\phi},\psi,\tilde{\psi}} R(\phi,\tilde{\phi},\psi,\tilde{\psi}) + \mathrm{dis}(S\circ\phi,\, T\circ\tilde{\phi}), \quad \text{s.t. condition (4–4) holds} \tag{4–6}$$

where R is a regularization operator of its arguments, $\mathrm{dis}(S\circ\phi,\, T\circ\tilde{\phi})$ measures the dissimilarity between $S\circ\phi$ and $T\circ\tilde{\phi}$.

4.2.2 Alternative Formulation of (4–4) Using Deformation Fields

Let the functions u, $\tilde{u}$, v and $\tilde{v}$ represent the corresponding deformation fields of the transformations ϕ, $\tilde{\phi}$, ψ and $\tilde{\psi}$, respectively. That is,

$$\begin{aligned} \phi(x) &= x + u(x), \quad \tilde{\phi}(x) = x + \tilde{u}(x), \\ \psi(x) &= x + v(x), \quad \tilde{\psi}(x) = x + \tilde{v}(x). \end{aligned} \tag{4–7}$$

Then, the constraints in (4–4) can be rewritten as

$$\begin{aligned} u + v(x+u) = v + u(x+v) = 0, \\ \tilde{u} + \tilde{v}(x+\tilde{u}) = \tilde{v} + \tilde{u}(x+\tilde{v}) = 0. \end{aligned} \tag{4–8}$$

4.2.3 MLE based derivation for $\mathrm{dis}(S\circ\phi,\, T\circ\tilde{\phi})$

To improve the robustness of the algorithm for deformable image registration, we use the negative log-likelihood of the residue image as a measure of mismatching. Consider voxel intensities of the residue image defined by

$$W(x) \triangleq S\circ\phi(x) - T\circ\tilde{\phi}(x), \quad x\in\Omega_M,$$

as independent samples drawn from a Gaussian distribution of mean zero and variance σ^2 to be optimized (see remark below for the reason of this assumption), whose probability density function (pdf) is denoted by $P(\cdot|\sigma)$. Then the likelihood of the residual image $W(x)$ can be computed as

$$\mathcal{L}(\sigma|\{W(x), x \in \Omega\}) = \prod_{x\in\Omega} P(W(x)|\sigma) = \prod_{x\in\Omega} \left(\frac{1}{\sqrt{2\pi}\sigma} e^{-|S\circ\phi - T\circ\tilde{\phi}|^2/2\sigma^2} \right). \tag{4–9}$$

Then, by writing the summation over all $x \in \Omega$ as an integral over Ω the negative log-likelihood function is given as follows:

$$\|S \circ \phi - T \circ \tilde{\phi}\|^2/2\sigma^2 + |\Omega| \log \sqrt{2\pi}\sigma.$$

Omitting the constant $\Omega \log \sqrt{2\pi}$, we define the dissimilarity term as

$$\text{dis}(S \circ \phi, T \circ \tilde{\phi}) \triangleq \|S \circ \phi - T \circ \tilde{\phi}\|^2/2\sigma^2 + |\Omega| \log \sigma. \tag{4–10}$$

which can be rewritten as our MLE fitting term F by using corresponding deformation fields u and $\tilde{u}$:

$$F(u, \tilde{u}, \sigma) \triangleq \text{dis}(S(x+u), T(x+\tilde{u})) = \|S(x+u) - T(x+\tilde{u})\|^2/2\sigma^2 + |\Omega| \log \sigma. \tag{4–11}$$

remark Let $\hat{P}$ be the estimation of the pdf for the random variable $X \triangleq W(x)$, $x \in \Omega$. We show below why it is reasonable to assume $\hat{P}$ to be a Gaussian distribution of zero mean and variance σ^2.

In fact, $\hat{P}$ is a function in $C_0(\mathbb{R})$, the space of all the continuous functions on real line vanishing at infinity with the supreme norm. Let $H_0(\mathbb{R})$ be the Hilbert space consisting of all linear combinations of $\kappa(x_l, x)$ for finite many of $x_l \in \mathbb{R}$, where

$$\kappa(x_l, x) = \left(2\pi\sigma^2\right)^{-1/2} e^{-(x_l - x)^2/2\sigma^2}, \quad \forall\, x \in \mathbb{R}. \tag{4–12}$$

Define an inner product on $H_0(\mathbb{R})$ by

$$\left\langle \sum_{i=1}^{m} a_i \kappa(x_i, \cdot), \sum_{j=1}^{n} b_j \kappa(y_j, \cdot) \right\rangle_{H_0(\mathbb{R})} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j \kappa(x_i, y_j).$$

We claim that

$$H_0(\mathbb{R}) \text{ is dense in } C_0(\mathbb{R}). \tag{4–13}$$

In fact, if the claim (4–13) is not true, by Hahn-Banach theorem there exists a bounded signed measure $\mathfrak{m}$ in the dual space of $C_0(\mathbb{R})$, such that

$$\int_{\mathbb{R}} \hat{P} d\mathfrak{m} \neq 0, \tag{4–14}$$

but $\int_{\mathbb{R}} f d\mathfrak{m} = 0$, for all $f \in H_0(\mathbb{R})$. In particular, for any $x \in \mathbb{R}$,

$$\int_{\mathbb{R}} \kappa(x, y) d\mathfrak{m}_y = 0,$$

where $\kappa(\cdot, \cdot)$ is as in (4–12), and hence,

$$\int_{\mathbb{R}\times\mathbb{R}} \kappa(x, y) d\mathfrak{m}_x d\mathfrak{m}_y = 0.$$

This implies $\mathfrak{m} = 0$, which contradicts (4–14). Therefore, the claim holds.

By this claim it is easy to see that

$$\hat{P}(z) \approx \sum_{l=1}^{k} \alpha_l \kappa(x_l, z) = \left(2\pi\sigma^2\right)^{-1/2} \sum_{l=1}^{k} \alpha_l e^{-(x_l - z)^2/2\sigma^2} \tag{4–15}$$

for some $\{x_l; \alpha_l\}_{l=1}^{k}$. Since a good registration requires the the intensities of the residue image $W(x)$ close to zero. Hence, in (4–15) the only dominate term in the sum should be the one corresponding to $x_l = 0$, and other terms are negligible. This means that $\hat{P}$ is approximately $\mathcal{N}(0, \sigma^2)$, the Gaussian distribution with mean 0 and variance σ^2.

4.2.4 Proposed model

Base on the discussion above, we are ready to present the proposed model. We define the regularization term $R(\phi, \tilde{\phi}, \psi, \tilde{\psi})$ in (4–6) using their corresponding

deformation fields as

$$R(\phi, \tilde{\phi}, \psi, \tilde{\psi}) = R(u, \tilde{u}, v, \tilde{v}) \triangleq \|Du\|^2 + \|D\tilde{u}\|^2 + \|Dv\|^2 + \|D\tilde{v}\|^2. \tag{4–16}$$

By plugging (4–16) and (4–11) into (4–6), and replacing the constraint in (4–6) by (4–8), the proposed model can be written as:

$$\min_{u,\tilde{u},v,\tilde{v},\sigma} R(u, \tilde{u}, v, \tilde{v}) + \lambda F(u, \tilde{u}, \sigma), \quad \text{s.t. condition (4–8) holds,} \tag{4–17}$$

where $R(u, \tilde{u}, v, \tilde{v})$ and $F(u, \tilde{u}, \sigma)$ are defined in (4–16) and (4–11), respectively.

To solve problem (4–17), we relax the equality constraints of inverse consistency, and penalize their violation using quadratic functions, then write it as an unconstrained energy minimization problem

$$\min_{u,\tilde{u},v,\tilde{v},\sigma} R(u, \tilde{u}, v, \tilde{v}) + \lambda F(u, \tilde{u}, \sigma) + \mu \left(\mathcal{I}(u, v) + \mathcal{I}(\tilde{u}, \tilde{v})\right), \tag{4–18}$$

where and $\mathcal{I}(u, v)$ is the cost of inverse inconsistency of u and v:

$$\mathcal{I}(u, v) = \mathcal{I}_v(u) + \mathcal{I}_u(v), \tag{4–19}$$

with

$$\mathcal{I}_v(u) = \|u + v(x + u)\|^2 \quad \text{and} \quad \mathcal{I}_u(v) = \|v + u(x + v)\|^2. \tag{4–20}$$

Similarly, we have $\mathcal{I}(\tilde{u}, \tilde{v})$. With sufficiently large μ, solving (4–18) gives an approximation to the solution of (4–17).

The term $F(u, \tilde{u}, \sigma)$ is from the negative log-likelihood of the residual image (4–11). Minimizing this term forces the mean of the residue image to be zero, but allows it to have a variance to accommodate certain variability. This makes the model more robust to noise and artifacts, and less sensitive to the choice of the parameter λ than the model using the SSD, i.e. the squared L^2-norm, of the residue image as a dissimilarity measure as in (4–2). The parameter λ balances the smoothness of deformation fields and goodness of alignments, and affects the registration result significantly. In the

proposed model, the ratio of the SSD of the residue image over the smoothing terms is λ/σ^2 rather than a prescribed λ. Since σ is to be optimized, and from its EL equation σ is the standard deviation of the residue image. Therefore, in the proposed model the weight on the matching term updates during iterations. When the alignment gets better, σ the standard deviation of the residue as shown in (4–35) decreases, and hence the weight on the matching term automatically increases. This self-adjustable feature of the weight not only enhances the accuracy of alignment, but also makes the choice of λ flexible, and results in a fast convergence.

As shown earlier, the final forward and backward transforms h and g can be obtained by

$$h = \phi \circ \tilde{\psi} = x + \tilde{v} + u(x + \tilde{v}) \quad \text{and} \quad g = \tilde{\phi} \circ \psi = x + \tilde{u} + v(x + \tilde{u}).$$

Thus, the corresponding final full-way forward and backward deformation fields $\bar{u}$ and $\bar{v}$ are given as

$$\bar{u} = \tilde{v} + u(x + \tilde{v}) \quad \text{and} \quad \bar{v} = \tilde{u} + v(x + \tilde{u}), \tag{4–21}$$

respectively. Then the inverse consistent constraints (4–4) can be represented using $\bar{u}, \bar{v}$ as follows:

$$\bar{u} + \bar{v}(x + \bar{u}) = \bar{v} + \bar{u}(x + \bar{v}) = 0. \tag{4–22}$$

4.3 Existence of Solutions

In this section we prove the existence of solutions $(u, \tilde{u}, v, \tilde{v}, \sigma)$ to the proposed model (4–18). For simplicity, we assume that both S and T defined on the same domain Ω, which is simply connected, closed and bounded in $\mathbb{R}^d$ with Lipschitz boundary $\partial\Omega$. Also $S, T \in C^1(\Omega)$. As in reality, deformation field cannot be unbounded, we restrict $u, \tilde{u}, v, \tilde{v}$ to be in a closed subset of $L^\infty(\Omega)$:

$$\mathcal{B} \triangleq \{u \in L^\infty(\Omega) : \|u\|_{L^\infty(\Omega)} \leq B,\ B \in \mathbb{R}_+ \text{ only depends on } \Omega\}$$

Then, we seek solutions $(u, \tilde{u}, v, \tilde{v}, \sigma)$ to the problem (4–18) in the spaces $u, \tilde{u}, v, \tilde{v} \in H^1(\Omega) \cap \mathcal{B}$ and $\sigma \in \mathbb{R}_+$. For short notations, we let w denote the quaternion $(u, \tilde{u}, v, \tilde{v})$. Then, we show the existence of solutions to the following minimization problem:

$$\min_{(w,\sigma)\in(H^1\cap\mathcal{B})\times\mathbb{R}_+} E(w, \sigma) \tag{4–23}$$

where

$$E(w, \sigma) = \|Dw\|^2 + \lambda F(w, \sigma) + \mu\mathcal{I}(w)$$

and F and $\mathcal{I}$ are defined correspondingly in (4–18) using the simplified notation of w, i.e.

$$\begin{aligned} \|Dw\|^2 &= \|Du\|^2 + \|D\tilde{u}\|^2 + \|Dv\|^2 + \|D\tilde{v}\|^2, \\ F(w, \sigma) &= \|S(x + u) - T(x + \tilde{u})\|^2/\sigma^2 + |\Omega| \log \sigma, \\ \mathcal{I}(w) &= \mathcal{I}_v(u) + \mathcal{I}_u(v) + \mathcal{I}_{\tilde{v}}(\tilde{u}) + \mathcal{I}_{\tilde{u}}(\tilde{v}). \end{aligned}$$

and the terms on the right side of $\mathcal{I}(w)$ are defined as in (4–20). The λ and μ are prescribed positive constants.

Theorem 4.1. *The minimization problem (4–23) admits solutions* $(w, \sigma) \in \left(H^1 \cap \mathcal{B}\right) \times \mathbb{R}_+$.

Proof. For $(w, \sigma) \in \left(H^1 \cap \mathcal{B}\right) \times \mathbb{R}_+$, $E(w, \sigma)$ is bounded below. Hence, there exists a minimizing sequence $\{(w_k, \sigma_k)\}_{k=1}^{\infty} \subset \left(H^1 \cap \mathcal{B}\right) \times \mathbb{R}_+$ such that

$$\lim_{k\to\infty} E(w_k, \sigma_k) = \inf_{(H^1\cap\mathcal{B})\times\mathbb{R}_+} E(w, \sigma).$$

Therefore $\{\|Dw_k\|\}_{k=1}^{\infty}$ are uniformly bounded. Along with $w_k \in \mathcal{B}$ we know that $\{w_k\}_{k=1}^{\infty}$ is a bounded sequence in H^1. By the weak compactness of H^1 and the fact that H^1 is precompact in L^2, there exists a convergent subsequence, which is still denoted by $\{w_k\}_{k=1}^{\infty}$, and a function $\hat{w} \in H^1$, such that

$$w_k \rightharpoonup \hat{w} \quad \text{weakly in } H^1, \tag{4–24}$$

$$w_k \to \hat{w} \quad \text{strongly in } L^2, \text{ and a.e. in } \Omega. \tag{4–25}$$

Moreover, since $E(w_k, \sigma_k) \to \infty$ if $\sigma_k \to 0$ or ∞, there is a constant $C > 0$ such that $\{\sigma_k\}_{k=1}^{\infty}$ are bounded below and above by $1/C$ and C respectively. Hence, there is a subsequence of $\{\sigma_k\}_{k=1}^{\infty}$ and a scaler $\hat{\sigma} \in \mathbb{R}_+$, without changing the notation for the subsequence we have

$$\sigma_k \to \hat{\sigma} \in \mathbb{R}_+. \tag{4–26}$$

From the weak lower semi-continuity of norms and (4–24), we know

$$\|D\hat{w}\|^2 \leq \lim_{k\to\infty} \|Dw_k\|^2. \tag{4–27}$$

Also, as $\mathcal{I}(w) \leq 8B$ for any $w \in H^1 \cap \mathcal{B}$ and $w_k \to \hat{w}$ a.e. in Ω, we get, by dominant convergence theorem, that

$$\lim_{k\to\infty} \mathcal{I}(w_k) = \mathcal{I}(\hat{w}). \tag{4–28}$$

By the same argument with the smoothness of S and T, the convergence of $\{\sigma_k\}_{k=1}^{\infty}$, and the fact that $w_k \to \hat{w}$ a.e. in Ω, we can also have

$$\lim_{k\to\infty} F(w_k, \sigma_k) = F(\hat{w}, \hat{\sigma}) \tag{4–29}$$

Combining (4–27), (4–28) with (4–29), we obtain that

$$E(\hat{w}, \hat{\sigma}) \leq \lim_{k\to\infty} E(w_k, \sigma_k) = \inf_{(H^1\cap\mathcal{B})\times\mathbb{R}_+} E(w, \sigma).$$

Furthermore, since $\{w_k\}_{k=1}^{\infty} \subset \mathcal{B} \subset L^\infty(\Omega)$, we know

$$w_k \rightharpoonup^{\mathbf{w}*} \hat{w} \quad \text{weakly* in } L^\infty$$

and hence $\hat{w} \in \mathcal{B}$. Therefore, $(\hat{w}, \hat{\sigma}) \in \left(H^1 \cap \mathcal{B}\right) \times \mathbb{R}_+$. Hence

$$E(\hat{w}, \hat{\sigma}) = \inf_{(H^1\cap\mathcal{B})\times\mathbb{R}_+} E(w, \sigma).$$

which implies that $(\hat{w}, \hat{\sigma})$ is a solution to the minimization problem (4–23). □

4.4 Numerical Scheme

In this section, we provide the numerical scheme for solving (4–18). Since the compositions in the inverse consistency constraints $\mathcal{I}_u(v)$ and $\mathcal{I}_v(u)$ bring a difficulty in getting an explicit form of the EL equations for the deformation fields and their inverses, in our computation, instead of directly solving (4–18), we solve the following two coupled minimization problems alternately:

$$\begin{cases} \min\limits_{u,\tilde{u}} E_{v,\tilde{v},\sigma}(u,\tilde{u}) \\ \min\limits_{v,\tilde{v}} E_{u,\tilde{u}}(v,\tilde{v}) \end{cases} \tag{4–30}$$

where

$$E_{v,\tilde{v},\sigma}(u,\tilde{u},\sigma) = \|Du\|^2 + \|D\tilde{u}\|^2 + \lambda F(u,\tilde{u},\sigma) + \mu\left(\mathcal{I}_v(u) + \mathcal{I}_{\tilde{v}}(\tilde{u})\right) \tag{4–31}$$

and

$$E_{u,\tilde{u}}(v,\tilde{v}) = \|Dv\|^2 + \|D\tilde{v}\|^2 + \mu\left(\mathcal{I}_u(v) + \mathcal{I}_{\tilde{u}}(\tilde{v})\right). \tag{4–32}$$

By taking first variation with respect to u, $\tilde{u}$, v, $\tilde{v}$, we get the EL equations:

$$\begin{cases} -\Delta u + \dfrac{\lambda}{\sigma^2} W_{u,\tilde{u}} DS(x+u) + \mu\left\langle I + Dv(x+u), u + v(x+u)\right\rangle = 0 \\ -\Delta v + \mu\left\langle I + Du(x+v), v + u(x+v)\right\rangle = 0 \\ -\Delta \tilde{u} - \dfrac{\lambda}{\sigma^2} W_{u,\tilde{u}} DT(x+\tilde{u}) + \mu\left\langle I + D\tilde{v}(x+\tilde{u}), \tilde{u} + \tilde{v}(x+\tilde{u})\right\rangle = 0 \\ -\Delta \tilde{v} + \mu\left\langle I + D\tilde{u}(x+\tilde{v}), \tilde{v} + \tilde{u}(x+\tilde{v})\right\rangle = 0 \end{cases}, \tag{4–33}$$

in Ω, with free Neumann boundary conditions for each of them on $\partial\Omega$:

$$\langle Du, n\rangle = \langle D\tilde{u}, n\rangle = \langle Dv, n\rangle = \langle D\tilde{v}, n\rangle = 0, \quad \text{on } \partial\Omega, \tag{4–34}$$

where $W_{u,\tilde{u}} \triangleq S(x+u) - T(x+\tilde{v})$, I is the identity matrix of size d, and n is the outer normal of $\partial\Omega$. Also, the first variation of σ gives

$$\sigma = \|S(x+u) - T(x+\tilde{u})\|/|\Omega|^{1/2}. \tag{4–35}$$

The solution to the EL equations (4–33) can be obtained by finding the stationary solution to the evolution equations associated with the EL equations. In numerical implementation, we use semi-implicit discrete form of the evolution equations. The additive operator splitting (AOS) scheme was applied to solve the problem faster [65]. An alternative way of AOS to solve the semi-implicit discrete evolution equation in this case can be obtained by applying discrete cosine transforms (DCT) to diagonalize the Laplace operator with the assumption that the deformation fields have symmetric boundary condition, which is compatible with (4–34).

In two-dimensional (2D) case, the semi-implicit discrete form of (4–33) with fixed step sizes τ_u, τ_v for the evolution equations of $u^{(k+1)}$ as

$$\frac{u_{i,j}^{(k+1)} - u_{i,j}^{(k)}}{\tau_u} = \Delta_{i,j} u^{(k+1)} - D_{i,j} \left(\lambda F \left(u^{(k)}, \tilde{u}^{(k)}, \sigma^{(k)} \right) + \mu \mathcal{I}_{v^{(k)}} \left(u^{(k)} \right) \right), \tag{4–36}$$

and $v^{(k+1)}$ as

$$\frac{v_{i,j}^{(k+1)} - v_{i,j}^{(k)}}{\tau_v} = \Delta_{i,j} v^{(k+1)} - \mu \, D_{i,j} \mathcal{I}_{u^{(k)}} \left(v^{(k)} \right), \tag{4–37}$$

where $\Delta_{i,j}$ and $D_{i,j}$ represent the discrete Laplacian and gradient operators at the pixel indexed by (i, j), respectively. The 3D case is a simple analogue with one more subscript in indices. Similarly, we have the discrete evolution equation for $\tilde{u}$ and $\tilde{v}$ with the two components within each of the three pairs $(u, \tilde{u})$, $(v, \tilde{v})$ and (S, T) switched in (4–36) and (4–37). With AOS scheme being applied, the computation for each update of u involves of solving d tridiagonal systems whose computational costs are linear in N, where N is the total number of pixels in S (or T). Also, in each iteration of updating u and v, there needs $2(d + 1)$ interpolations with size N. It is important to point out that, in each iteration, the computations of $u, \tilde{u}, v, \tilde{v}$ can be carried out in parallel. We summarize icDIR in Algorithm 5, where the maximum inverse consistency error (ICE) δ_c is defined by

$$\delta_c = \max_x \{ |\bar{u} + \bar{v}(x + \bar{u})|, |\bar{v} + \bar{u}(x + \bar{v})| \}, \tag{4–38}$$

Algorithm 5 Inverse Consistent Deformable Image Registration (icDIR)

Input S, T, and $\tau_u, \tau_v, \lambda, \mu > 0, \epsilon = .5, \delta_c = 1$. Initialize $(u^{(0)}, \tilde{u}^{(0)}, v^{(0)}, \tilde{v}^{(0)}) = 0$, $k = 0$.
while $\delta_c \geq \epsilon$ **do**
  repeat
    {All terms in $(u^{(k+1)}, \tilde{u}^{(k+1)}, v^{(k+1)}, \tilde{v}^{(k+1)})$ can be calculated in parallel}
    Calculate $(u^{(k+1)}, v^{(k+1)})$ using (4–36) and (4–37).
    Calculate $(\tilde{u}^{(k+1)}, \tilde{v}^{(k+1)})$ using (4–36) and (4–37) with $(u^{(k+1)}, v^{(k+1)})$ replaced by $(\tilde{u}^{(k+1)}, \tilde{v}^{(k+1)})$.
    update $\sigma^{(k+1)}$ by (4–35).
    $k \leftarrow k + 1$
  until convergence
  return $(u, \tilde{u}, v, \tilde{v})^{\mu}$
  $(u, \tilde{u}, v, \tilde{v}) \leftarrow (u, \tilde{u}, v, \tilde{v})^{\mu}, \mu \leftarrow 2\mu$.
  Compute $\bar{u}$ and $\bar{v}$ using (4–21) and then δ_c using (4–38).
end while

and $\bar{u}$ and $\bar{v}$ are the final full-way deformation fields shown in (4–21). That is, it measures the maximum ICE of deformations obtained by quaternion $(u, \tilde{u}, v, \tilde{v})$. The parameter μ in (4–18) may increase during iterations to ensure smaller ICE. In each inner loop with fixed μ, the computation is terminated when the mean of $CC(S(x+\bar{u}), T)$ and $CC(T(x+\bar{v}), S)$ converges. We set a stopping tolerance $\epsilon = .5$ and terminate the whole computation once δ_c is lower than ϵ, in which case the maximum ICE is less than half of the grid size between two concatenate pixels/voxels and hence the inverse consistency is exactly satisfied with respect to the original resolution of the images.

4.5 Experimental Results

In this section, we present the experimental results of proposed model using algorithm 1 (icDIR). All implementations involved in the experiments were coded in MATLAB®v7.3 (R2006b), except the Thomas tridiagonal solver, which was coded in C++. We used build-in functions `interp2/interp3` of Matlab with default settings for interpolations. All Computations were performed on a GNU/Linux (version 2.6.16) workstation with Intel®Core 2 CPU at 1.86GHz and 2GB memory.

We first test the accuracy of registration and auto re-contouring of the proposed algorithm on a clinical data set of 100 2D-prostate MR images. Each image, called a

phase, is a 2D image of dimension 288×192 that focuses on the prostate area. The first phase is used as a source image S, as shown in Figure 4-1A. The boundaries of the regions of interests (ROI) in S were delineated by contours and superimposed by medical experts, as enlarged and shown in Figure 4-4A. The rest 99 phases were considered as targets. In this experiment we applied the proposed model (4–18) with parameters (λ, μ, τ) set to be $(.05, .2, .05)$ to S and Ts. For demonstration, we only showed the result using the 21st phase as T, as depicted in Figure 4-1B. The deformed T and deformed S, i.e. $T(x + \bar{v})$ and $S(x + \bar{u})$, are shown in the Figure 4-1C and 4-1D respectively, where $\bar{u}$ and $\bar{v}$ are defined in (4–21) using the optimal $(u, \tilde{u}, v, \tilde{v})$ obtained by model (4–18). The errors of the alignments, $|T(x + \bar{v}) - S|$ and $|S(x + \bar{u}) - T|$, on the squared area (shown in Figure 4-1A) are displayed in Figure 4-2A and 4-2C, respectively. With comparison to the original error $|S - T|$ shown in Figure 4-2B, we can see the errors of alignments are significantly reduced. This indicates that the proposed registration model (4–18) has high accuracy in matching two images.

The final optimal forward and backward deformation fields $\bar{u}$ and $\bar{v}$ are displayed by applying them to a domain of regular grids, shown in Figure 4-3A and 4-3C, respectively. Furthermore, to validate the accurate inverse consistency obtained by our model (4–18), we applied $\bar{u} + \bar{v}(x + \bar{u})$ on a domain with regular grids, and plotted the resulting grids in Figure 4-3B. The resulting grids by $\bar{v} + \bar{u}(x + \bar{v})$ had the same pattern so we omitted it here. From Figure 4-3B, we can see that the resulting grids are the same as the original regular grids. This indicates that the inverse consistent constraints $\bar{u} + \bar{v}(x + \bar{u}) = \bar{v} + \bar{u}(x + \bar{v}) = 0$ are well preserved. We also computed the maximum ICE δ_c using $\bar{u}, \bar{v}$ and (4–38) and the result was $.46$. The mean ICE $(\|\bar{u} + \bar{v}(x + \bar{u})\| + \|\bar{v} + \bar{u}(x + \bar{v})\|)/2|\Omega|$ versus the number of iterations is plotted in Figure 4.5, which shows the inverse consistency is preserved during the registration. These imply that the proposed algorithm provides an accurate inverse consistent registration.

An accurate inverse consistent registration can transform segmentations from one image to another accurately. One of the applications is auto re-contouring, that deforms the expert's contours from a planning image to new images during the course of radiation therapy. In this experiment, we had expert's contours superimposed on the source image S as shown in Figure 4-4A. Then by applying the deformation field $\bar{u}$ on this contours we get the deformed contours on the target image T as shown in Figure 4-4B. The accuracy in auto re-contouring is evident.

Figure 4-3 shows, from left to right, the followings: $\bar{u}$, $\bar{u} + \bar{v}(x + \bar{u})$, which demonstrates the inverse consistency is well preserved, and $\bar{v}$, respectively.

The second experiment was aimed to test the efficiency of the proposed model (4–18) in registering 3D images. We applied (4–18) to a pair of 3D chest CT images of dimension $64 \times 83 \times 48$ taken from the same subject but at different periods. The parameters (λ, μ, τ) were set to be $(.05, .1, .004)$. The registration was performed in 3D, but for demonstration, we only show the corresponding axial (xy plane with $z = 33$), sagittal (yz plane with $x = 25$) and coronal (zx plane with $y = 48$) slices. The registration results are plotted in Figures 4-5, 4-6 and 4-7, respectively. In each figure, the images in the upper row are S and T, respectively, and the images in the middle row are deformed T and S, i.e. $T(x + \bar{v})$ and $S(x + \bar{u})$, respectively. The bottom row shows the residual images $|S(x + \bar{u}) - T|$, $|S - T|$ and $|T(x + \bar{v}) - S|$. The mean of $CC(S(x + \bar{u}), T)$ and $CC(T(x + \bar{v}), S)$ reached $.998$ after 50 iterations, and the mean of inverse consistency errors was $.015$. The results shows the high accuracy of proposed model (4–18) and the well preserved inverse consistency.

The third experiment was aimed to compare the effectiveness of model (4–18) with the following conventional full-way inverse consistent deformable registration model:

$$\min_{u,v,\sigma_u,\sigma_v} \|Du\|^2 + \|Dv\|^2 + \lambda J(u, v, \sigma_u, \sigma_v) + \mu \left(\mathcal{I}_v(u) + \mathcal{I}_u(v)\right) \tag{4–39}$$

where u and v are forward and backward deformation fields, respectively, and the term J is defined by

$$J(u, v, \sigma_u, \sigma_v) = \|S(x+u) - T\|^2/2\sigma_u^2 + \|T(x+v) - S\|^2/2\sigma_v^2 + |\Omega| \log \sigma_v \sigma_v.$$

The comparison is made on the efficiency and accuracy of matching, as well as the preservation of inverse consistency. The accuracy of matching is measured by correlation coefficients (CC) between the target image and deformed source image with the optimal forward and backward deformations obtained by model (4–39) and proposed model (4–18), respectively. Recall that for any two images S and T both with N pixels, the CC of S and T is defined by

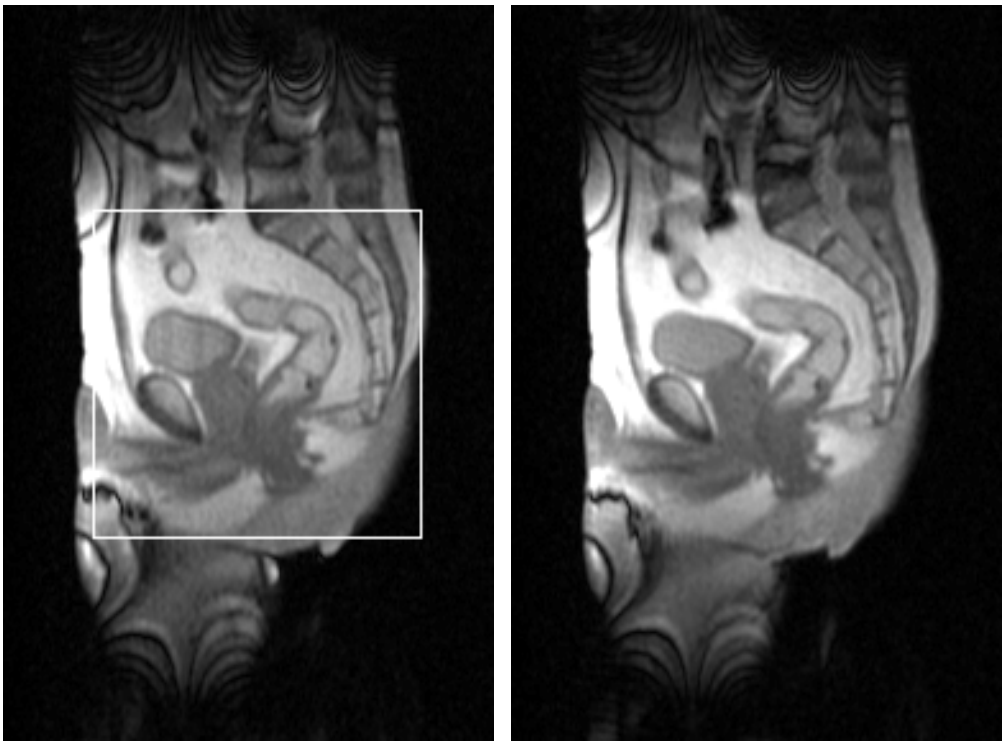
$$CC(S, T) = \frac{\sum_{i=1}^{N}(S_i - \bar{S})(T_i - \bar{T})}{\sqrt{\sum_{i=1}^{N}(S_i - \bar{S})^2 \sum_{i=1}^{N}(T_i - \bar{T})^2}},$$

where S_i and T_i are the intensities at the ith pixels of S and T, respectively, $\bar{S}$ and $\bar{T}$ are the mean intensities of S and T, respectively. The maximum value of CC is 1, in which case S and T are (positively) linearly related. In this experiment we applied models (4–39) and (4–18) to the images in the first experiment shown in Figure 4-1 with the same parameters (λ, μ, τ) to be $(.05, .2, .05)$. In Figure 4.5, we plotted the CC obtained by model (4–39) and proposed model (4–18) at each iteration. One can observed that the CC obtained by model (4–18) is higher and increases faster than model (4–39). This demonstrates that proposed model (4–18) is more efficient than the conventional full-way model. The reason is that the disparity between deformed S and deformed T is smaller than that between deformed S and fixed T or deformed T and fixed S. When S and T are deformed simultaneously, the two directional deformation fields are not necessarily to be large even if the underlying deformation field is large, which usually makes it difficult for the full-way based registration model to reach a satisfactory alignment in short time.

Table 4-1. Number of iterations used for convergence and the final CC obtained by proposed model with σ updated/fixed.

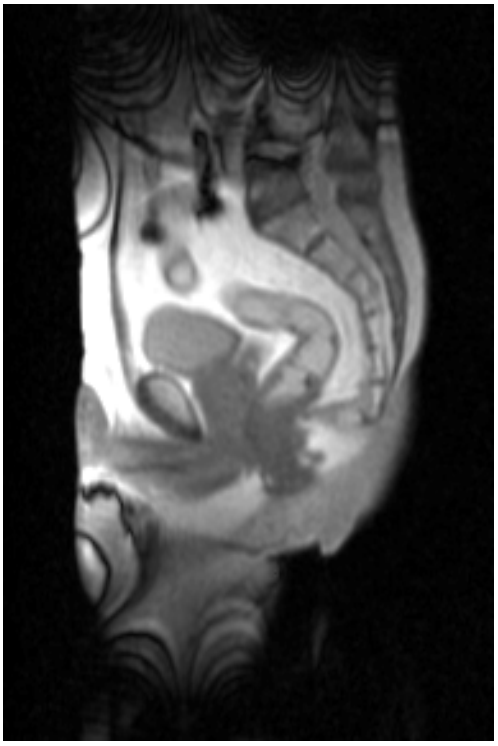
	Update σ		Fix σ	
λ	CC	Iter	CC	Iter
1e2	.962	48	.955	89
1e1	.962	97	.946	420
1e0	.960	356	.933	1762

The last experiment is aim to test the robustness of the model to noises and the choice of the parameter λ with the use of MLE based approach (4–11) for measuring the goodness of matching. The images S and T in Figure 4-1 with additive Gaussian noises (standard deviation is 3% of largest intensity value of S) were used in this experiment. The CC between S and T before registration is $CC(S, T) = .901$. We applied model (4–18) with σ to be updated/optimized by its EL equation (4–35), and σ to be set $\sigma = 1$, that is the same as using SSD as similarity measure, respectively, to the noise data mentioned above. We proceeded the registration with various values of λ, but kept other parameters fixed. Then the numbers of iterations (Iter) for convergence and the final CC were recorded and shown in Table 4-1. One can see that while λ decreases, the accuracy of model (4–18) using fixed σ reduces as the final CC become much smaller, and it also takes much longer time for the algorithm to converge. On the other hand, with σ being updated (whose computational cost is extremely cheap) model (4–18) can obtain good matching in much less iterations for a large range of λ. This shows that model with MLE fitting is much less sensitive to noise and the choice of λ, and can achieve fast and accurate results compared with the model using SSD to measure mismatching.

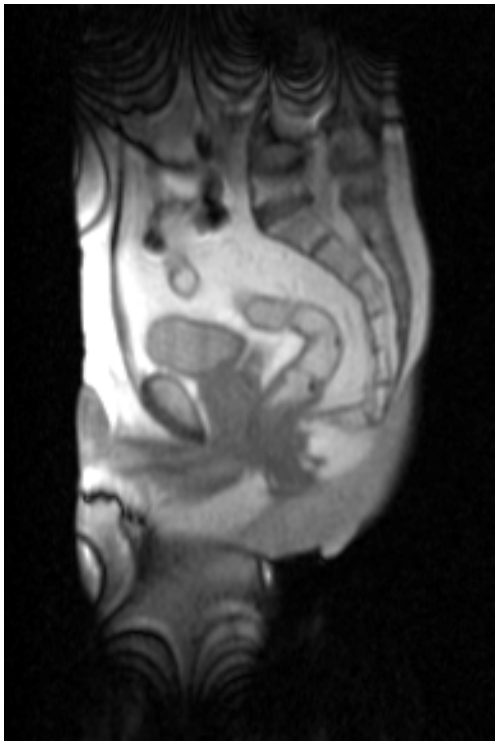


A Source image S

B Target image T

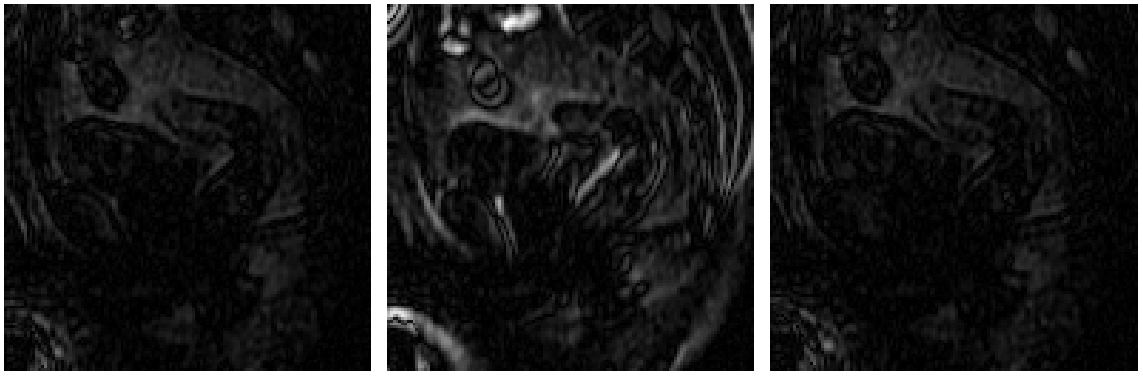


C Deformed T



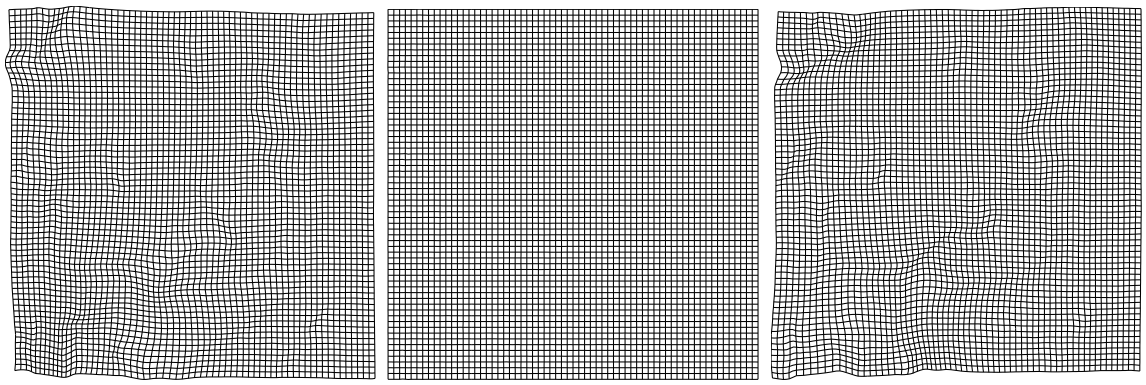
D Deformed S

Figure 4-1. Inverse consistent registration result by proposed model (4–18).



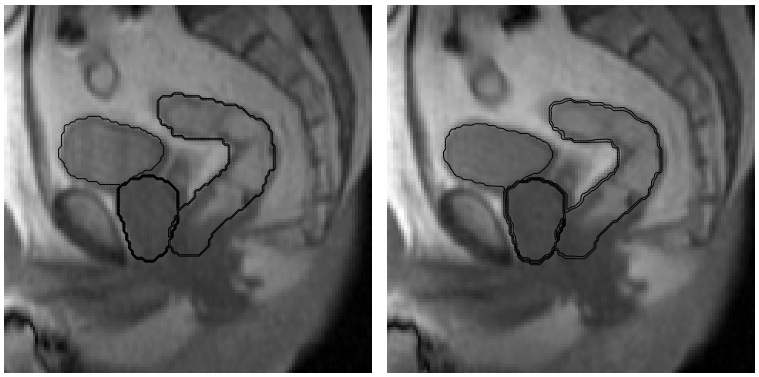
A $|S(x+\bar{u})-T|$ B $|S-T|$ C $|T(x+\bar{v})-S|$

Figure 4-2. Residue image obtained by proposed model (4–18).



A $\bar{u}$ B $\bar{u}+\bar{v}(x+\bar{u})$ C v

Figure 4-3. Deformation fields obtained by proposed model (4–18) in the zoomed-in area applied on regular grid with half of original resolution of images.



A Original Contour on S B Re-contouring on T

Figure 4-4. Auto re-contouring result using the deformation field $\bar{u}$ obtained by proposed model (4–18).

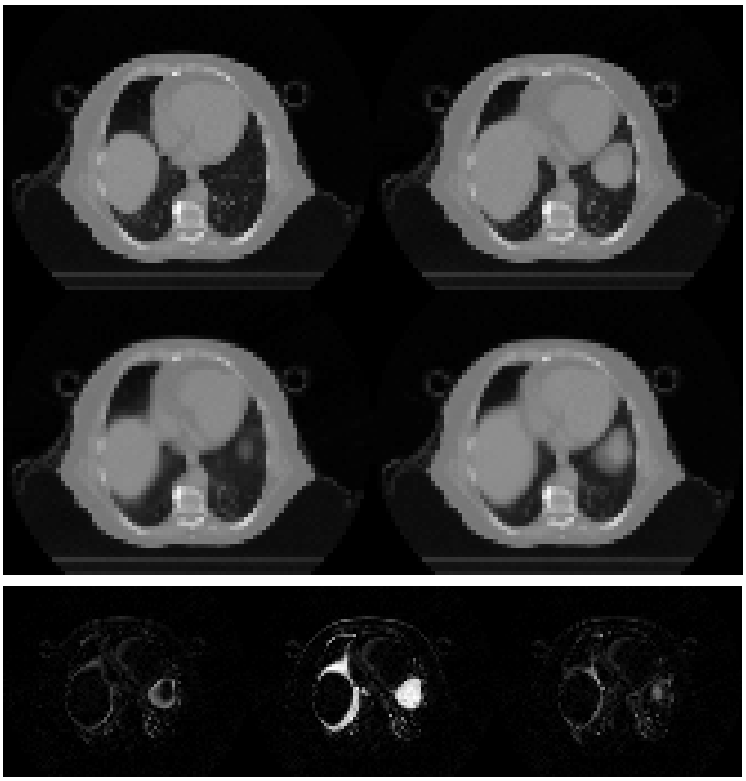


Figure 4-5. Registration result of proposed model (4–18) applied to 3D chest CT image. This figure shows the $z = 33$ slice at axial direction.

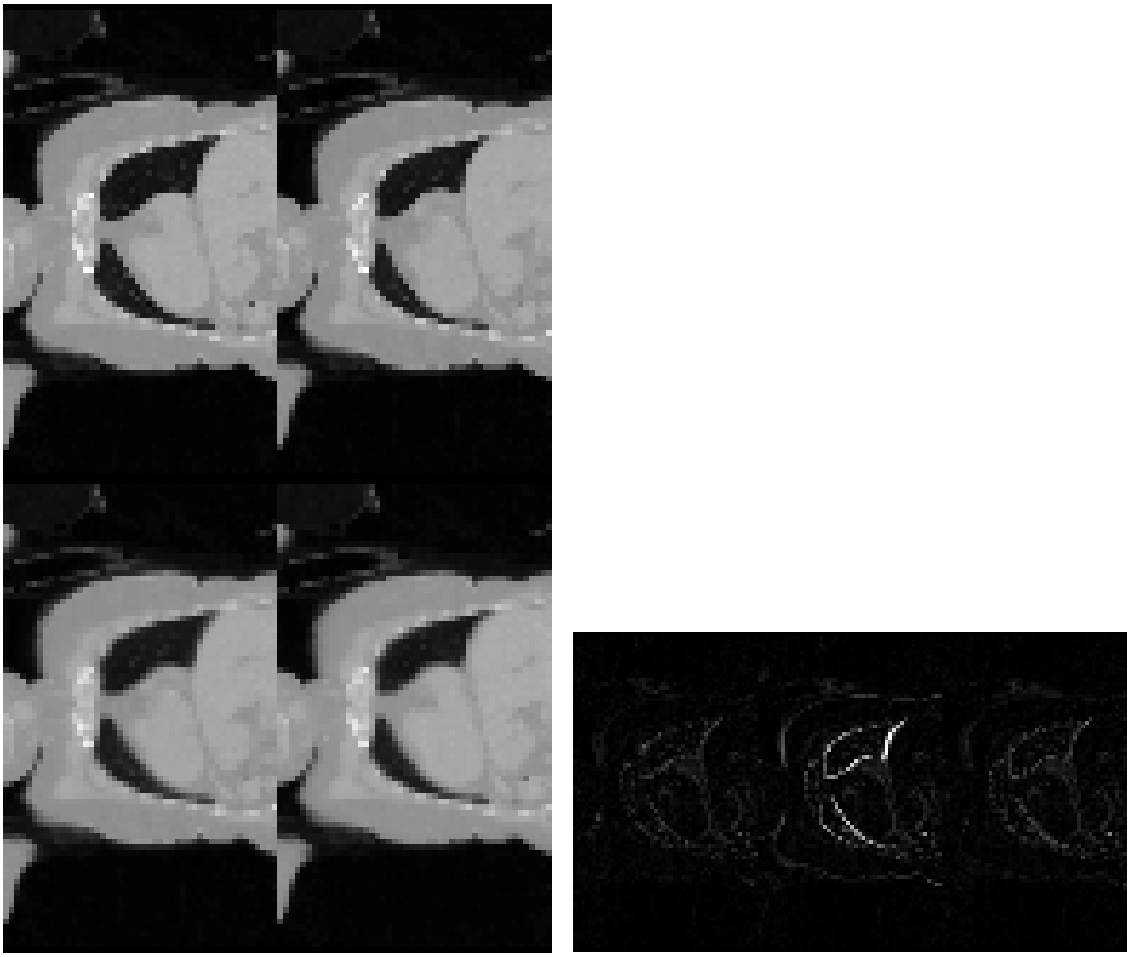


Figure 4-6. Registration result of proposed model (4–18) applied to 3D chest CT image. This figure shows the $x = 25$ slice at sagittal direction.

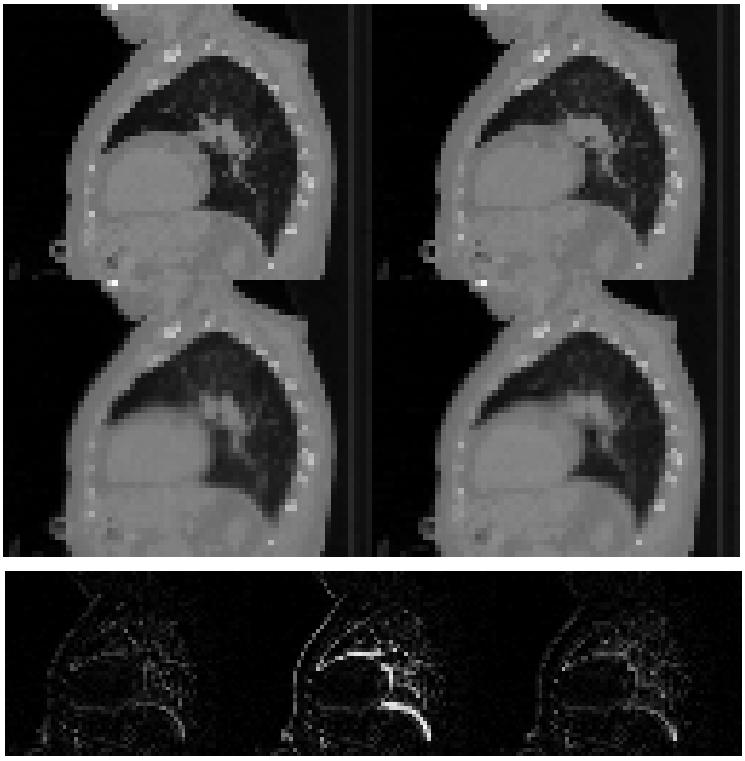


Figure 4-7. Registration result of proposed model (4–18) applied to 3D chest CT image. This figure shows the $y = 48$ slice at coronary direction.

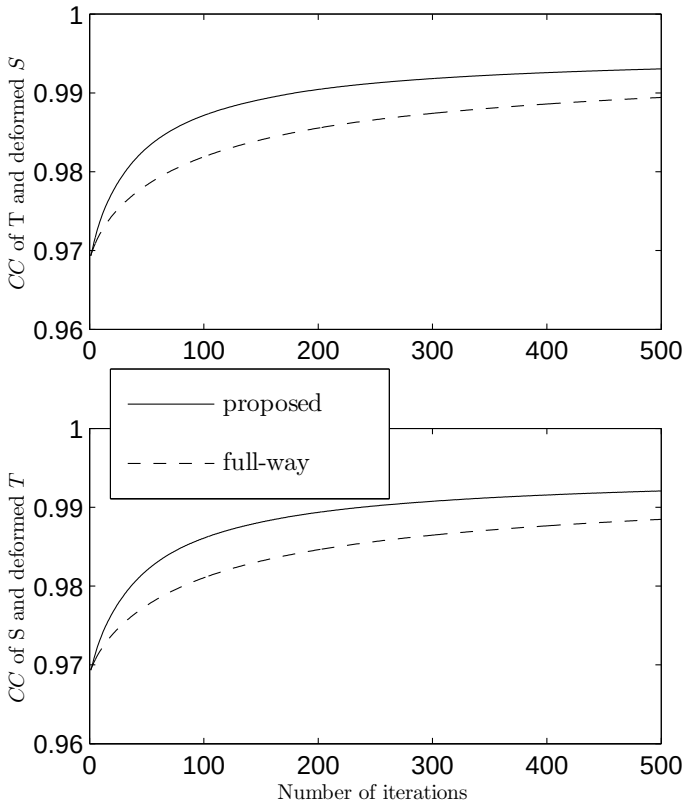


Figure 4-8. CC in each iteration obtained by full-way model (4–39) and proposed model (4–18).

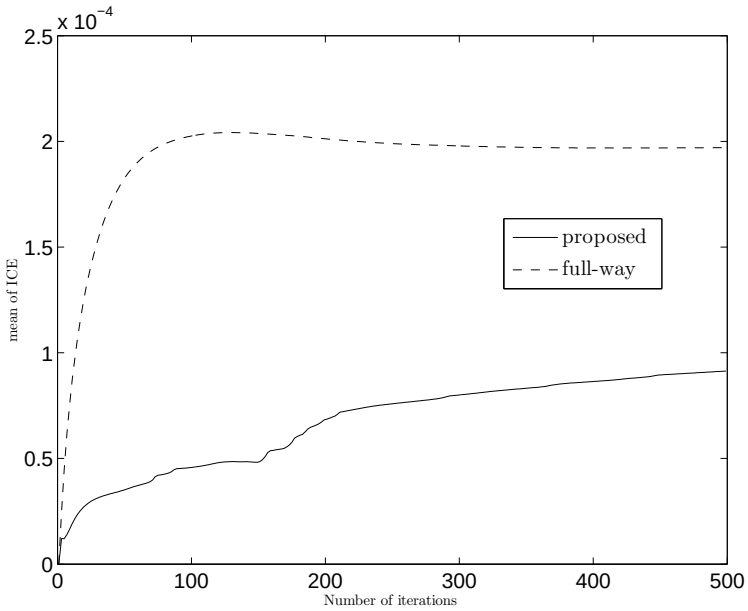


Figure 4-9. Mean of inverse consistent errors (ICE) of the final deformation fields obtained by using full-way model (4–39) and proposed model (4–18).

REFERENCES

[1] Acker, F. and Prestel, M.-A. “Convergence d’un schéma de minimisation alternée.” *Ann. Fac. Sci. Toulouse Math. (5)* 2 (1980): 1–9.

[2] Aharon, Michal, Elad, Michael, and Bruckstein, Alfred. “The K-SVD: An algorithm for Designing of Overcomplete Dictionaries for Sparse Representation.” *IEEE Trans. Signal Process.* 54 (2006).11: 4311–4322.

[3] Akaike, H. “On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method.” *Ann. Inst. Statist. Math. Tokyo* 11 (1959): 1–17.

[4] ———. “On a successive transformation of probability distribution and its application to the analysis of the optimum gradient method.” *Ann. Inst. Statist. Math. Tokyo* 11 (1959): 1–17.

[5] Alvarez, L., Deriche, R., Papadopoulo, T., and Sanchez, J. “Symmetrical dense optical flow estimation with occlusions detection.” *Proceedings of European Conference on Computer Vision* (2002): 721–735.

[6] Arunachalam, A., Samsonov, A., and Block, W.F. “Self-calibrated GRAPPA method for 2D and 3D radial data.” *Magn. Reson. Med.* 57 (2007).5: 931–938.

[7] Attouch, H., Redont, P., and Soubeyran, A. “A new class of alternating proximal minimization algorithms with cost-to-move.” *SIAM J. Optim.* 18 (2007): 1061–1081.

[8] Avants, B. B., Grossman, M., and Gee, J. C. “Symmetric diffeomorphic image registration: Evaluating automated labeling of elderly and neurodegenerative cortex and frontal lobe.” *Proceedings of Biomedical Image Registration* 4057 (2006): 50–57.

[9] Barzilai, J. and Borwein, J. M. “Two point step size gradient methods.” *IMA J. Numer. Anal.* 8 (1988): 141–148.

[10] Bauschke, H. H., Combettes, P. L., and Noll, D. “Joint minimization with alternating Bregman proximity operators.” *Pac. J. Optim.* 2 (2006): 401–424.

[11] Beck, Amir and Teboulle, Marc. “A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems.” *SIAM J. Imag. Sci.* 2 (2009).1: 183–202.

[12] Beg, Mirza Faisal and Khan, Ali. “Symmetric Data Attachment Terms for Large Deformation Image Registration.” *IEEE Transactions on Medical Imaging* 26 (2007).9: 1179–1189.

[13] Bertsekas, D. *Parallel and Distributed Computation*. Prentice Hall, 1989.

[14] Bioucas-Dias, J. and Figueiredo, M. "A new TwIST: Two-step iterative shrinkage/thresholding algorithms for image restoration." *IEEE Trans. Image Process.* 16 (2007).12: 2992–3004.

[15] Block, K., Uecker, M., and Frahm, J. "Undersampled Radial MRI with Multiple Coils: Iterative Image Reconstruction Using a Total Variation Constraint." *Magn. Reson. Med.* 57 (2007): 1086–1098.

[16] Block, K.T., Uecker, M., and Frahm, J. "Undersampled Radial MRI with Multiple Coils. Iterative Image Reconstruction Using a Total Variation Consraint." *Magn. Reson. Med.* 57 (2007): 1086–1098.

[17] Candes, Emmanuel J. and Romberg, Justin K. "Signal Recovery from Random Projections." *Proceedings of SPIE Computational Imaging III* 5674 (2005): 76–86.

[18] Candes, Emmanuel J., Romberg, Justin K., and Tao, Terence. "Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information." *IEEE Trans. Inform. Theory.* 52 (2006).2: 489–509.

[19] Chambolle, A. "An algorithm for total variation minimization and applications." *J. Math. Imaging Vis.* 20 (2004): 89–97.

[20] Chan, T. F., Golub, G. H., and Mulet, P. "A nonlinear primal-dual method for total variationbased image restoration." *SIAM J. Optim.* 20 (1999): 1964–1977.

[21] Chang, T.C., He, L., and Fang, T. "MR Image Reconstruction from Sparse Radial Samples Using Bregman Iteration." *Proc. Intl. Soc. Mag. Reson. Med.* (2006): 696.

[22] Chen, S., Donoho, D., and Saunders, M. "Atomic Decomposition by Basis Pursuit." *SIAM J. Sci. Comput.* 20 (1999): 33–61.

[23] Christensen, G. E. and Johnson, H. J. "Consistent Image Registration." *IEEE Transactions on Medical Imaging* 20 (2001).7: 721–735.

[24] Dai, Y. H., Hager, W. W., Schittkowski, K., and Zhang, H. "The cyclic Barzilai-Borwein method for unconstrained optimization." *IMA J. Numer. Anal.* 26 (2006): 604–627.

[25] Dai, Y. H. and Yuan, Y. "Alternate minimization gradient method." *IMA J. Numer. Anal.* 23 (2003): 377–393.

[26] Donoho, David L. "For Most Large Underdetermined Systems of Linear Equations, the Minimal l1-norm Solution is Also the Sparsest Solution." *Commun. Pure Appl. Math.* 59 (2006).7: 907–934.

[27] Donoho, David L., Elad, M., and Temlyakov, V. "Stable Recovery of Sparse Overcomplete Represenation in the Presence of Noise." *IEEE Trans. Inf. Theory* 52 (2006): 6–18.

[28] Eckstein, J. and Bertsekas, D. “On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators.” *Mathematical Programming* 55 (1992).1-3: 293–318.

[29] Eggers, H. and Boesiger, P. “Gridding- and Convolution-Based Iterative Reconstruction with Variable k-space Resolution for Sensitivity-Encoded Non-Cartesian Imaging.” *Proc. Intl. Soc. Mag. Reson. Med.*. 2003, 2346.

[30] Elad, Michael and Aharon, Michal. “Image Denoising Via Sparse and Redundant Representations Over Learned Dictionaries.” *IEEE Trans. Image Process.* 15 (2006).12: 3736–3745.

[31] Friedlander, A., Martínez, J. M., Molina, B., and Raydan, M. “Gradient method with retards and generalizations.” *SIAM J. Numer. Anal.* 36 (1999): 275–289.

[32] Gabay, D. and Mercier, B. “A dual algorithm for the solution of nonlinear variational problems via finite-element approximations.” *Comput. Math. Appl.* 2 (1976): 17–40.

[33] Glowinski, R. and Marrocco, A. “Sur l’approximation par éléments finis d’ordre un, et la résolution par pénalisation-dualité d’une classe de problèmes de Dirichlet nonlinéaires.” *RAIRO Analyse Numérique* 9 (1975): 41–76.

[34] Goldstein, T. and Osher, S. “The Split Bregman Method for L1 Regularized Problems.” *SIAM J. Imag. Sci.* 2 (2009): 323–343.

[35] Griswold, M.A., Jakob, P.M., Heidemann, R.M., Mathias, N., Jellus, V., Wang, J., Kiefer, B., and Haase, A. “Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA).” *Magn. Reson. Med.* 47 (2002): 1202–1210.

[36] Hale, E.T., Yin, W., and Zhang, Y. “Fixed-Point Continuation Method for ℓ_1-Minimization with Applications to Compressed Sensing.” *SIAM J. Optim.* 19 (2008).3: 1107–1130.

[37] He, J. C. and Christensen, G. E. “Large deformation inverse consistent elastic image registration.” *Proceedings of Information Processing in Medical Imaging* 2732 (2003): 438–449.

[38] He, Lin, Chang, Ti-Chiun, Osher, Stanley, Fang, Tong, and Speier, Peter. “MR Image Reconstruction by Using the Iterative Refinement Method and Nonlinear Inverse Scale Space Methods.” Tech. Rep. 06-35, CAM UCLA, 2006.

[39] Joshiand, S., Davis, B., Jomier, M., and Gerig, G. “Unbiased diffeomorphic atlas construction for computational anatomy.” *Neuroimage, Supplement* 23 (2004).1: 151–160.

[40] Kessler, M. L. “Image registration and data fusion in radiation therapy.” *British Journal on Radiology* 79 (2006): 99–108.

[41] Kim, Seung-Jean, Koh, K., Lustig, M., Boyd, Stephen, and Gorinevsky, Dimitry. "An Interior-Point Method for Large-Scale ℓ_1-Regularized Least Squares." *SIAM J. Optim.* 1 (2007).4: 606–617.

[42] Kim, S.J., Koh, K., Lustig, M., and Boyd, S. "An Efficient Method for Compressed Sensing." *Proceedings of IEEE Intl. Conf. Image Process.(ICIP)* 3 (2007): 117–120.

[43] Krishnan, D., Lin, P., and Tai, X. "An efficient operator splitting method for noise removal in images." *Commun. Comput. Phys.* 1 (2006): 847–858.

[44] Larson, A.C., White, R.D., Laub, G., McVeigh, E.R., Li, D., and Simonetti, O.P. "Self-gated cardiac cine MRI." *Magn. Reson. Med.* 51 (2004): 93–102.

[45] Leow, A.D., Huang, S.C., Geng, A., Becker, J., Davis, S., Toga, A., and Thompson, P. "Inverse consistent mapping in 3D deformable image registration: its construction and statistical properties." *Proceedings of Information Processing in Medical Imaging* (2005): 493–503.

[46] Liao, H.Y. and Sapiro, Guillermo. "Sparse Representation for Limited Data Tomography." *Proceedings of IEEE Intl. Symp. Biomed. Imag. (ISBI)* (2008): 1375–1378.

[47] Lions, P. L. and Mercier, B. "Splitting algorithms for the sum of two nonlinear operators." *SIAM J. Numer. Anal.* 16 (1979): 964–979.

[48] Liu, C., Bammer, R., and Moseley, M. "Parallel Imaging Reconstruction for Arbitrary Trajectories Using k-Space Sparse Matrices (kSPA)." *Magn. Reson. Med.* 58 (2007): 1071–1081.

[49] Lu, W., Olivera, G. H., Chen, Q., Chen, M., and Ruchala, K. "Automatic re-contouring in 4D radiotherapy." *Physics in Medicine and Biology* 51 (2006): 1077–1099.

[50] Lustig, Michael, Donoho, David, and Pauly, John M. "Sparse MRI: The Application of Compressed Sensing for Rapid MR Imaging." *Magn. Reson. Med.* 58 (2007).6: 1182–1195.

[51] Ma, Shiqian, Yin, Wotao, Zhang, Yin, and Chakraborty, A. "An efficient algorithm for compressed MR imaging using total variation and wavelets." *IEEE Proc. Conf. on Comp. Vis. Patt. Recog.* (2008): 1–8.

[52] Nesterov, Yu. "Gradient methods for minimizing composite objective function." CORE Discussion Papers 2007/76, Universit catholique de Louvain, Center for Operations Research and Econometrics (CORE), 2007.

URL http://ideas.repec.org/p/cor/louvco/2007076.html

[53] Nocedal, J. and Wright, S. J. *Numerical Optimization*. New York: Springer, 1999.

[54] Pruessmann, K.P., Weiger, M., Bornert, P., and Boesiger, P. "Advances in Sensitivity Encoding With Arbitrary k-Space Trajectories." *Magn. Reson. Med.* 46 (2001): 638–651.

[55] Pruessmann, K.P., Weiger, M., Scheidegger, M.B., and Boesiger, P. "SENSE: Sensitivity encoding for fast MRI." *Magn. Reson. Med.* 42 (1999): 952–962.

[56] Qian, Y., Zhang, Z., Stenger, V.A., and Wang, Y. "Self-Calibrated Spiral SENSE." *Magn. Reson. Med.* 52 (2004): 688–692.

[57] Qu, P., Zhong, K., Zhang, B., Wang, J., and Shen, G.-X. "Convergence Behavior of Iterative SENSE Reconstruction with Non-Cartesian Trajectories." *Magn. Reson. Med.* 54 (2005): 1040–1045.

[58] Rogelj, P. and Kovacic, S. "Symmetric image registration." *Medical Image Analysis* 10 (2006).3: 484–494.

[59] Rudin, L., Osher, S., and Fatemi, E. "Non-linear Total Variation Noise Removal Algorithm." *Physics D.* 60 (1992): 259–268.

[60] Samsonov, A. and Johnson, C. "Non-Cartesian POCSENSE." *Proc. Intl. Soc. Mag. Reson. Med.*. 2004, 2648.

[61] Seiberlich, N., Breuer, F.A., Blaimer, M., Barkauskas, K., and Griswold, P.M. Jakob M.A. "Non-Cartesian Data Reconstruction Using GRAPPA Operator Gridding (GROG)." *Magn. Reson. Med.* 58 (2007): 1257–1265.

[62] Starck, J., Elad, M., and Donoho, D. "Image Decomposition via the Combination of Sparse Representations and a Variational Approach." *IEEE Trans. Image Process.* 14 (2005): 1570–1582.

[63] Vogel, C. R. and Oman, M. E. "Iterative methods for total variation denoising." *SIAM J. Sci. Comput.* 17 (1996): 227–238.

[64] Wang, Y., Yang, J., Yin, W., and Zhang, Y. "A New Alternating Minimization Algorithm for Total Variation Image Reconstruction." *SIAM J. Imag. Sci.* 1 (2008).3: 248–272.

[65] Weickert, Joachim, ter Haar Romeny, Bart M., and Viergever, Max A. "Efficient and Reliable Schemes for Nonlinear Diffusion Filtering." *IEEE Transactions on Image Processing* 7 (1998).3: 398–410.

[66] Wright, Stephen J., Nowak, Robert D., and Figueiredo, M. "Sparse Reconstruction by Separable Approximation." *IEEE Trans. Signal Process.* 57 (2009).7: 2479–2493.

[67] Yang, Junfeng, Zhang, Yin, and Yin, Wotao. "A Fast TVL1-L2 Minimization Algorithm for Signal Reconstruction from Partial Fourier Data." Tech. Rep. 08-29, CAAM Rice Univ., 2008.

[68] Ye, J.C., Tak, S., Han, Y., and Park, H.W. "Projection Reconstruction MR Imaging Using FOCUSS." *Magn. Reson. Med.* 57 (2007): 764–775.

[69] Ye, X., Chen, Y., and Huang, F. "MR Image Reconstruction via Sparse Representation: Modeling and Algorithm." *International Conference on Image Processing, Computer Vision, and Pattern Recognition (IPCV)* (2009): 10–16.

[70] Yeh, E.N., Stuber, M., McKenzie, C.A., RM, R.M. Botnar, Leiner, T., Ohliger, M.A., Grant, A.K., Willig-Onwuachi, J.D., and Sodickson, D.K. "Inherently Self-Calibrating Non-Cartesian Parallel Imaging." *Magn. Reson. Med.* 54 (2005): 1–8.

[71] Yeung, Sai-Kit, Tang, Chi-Keung, Shi, Pengcheng, Pluim, Josien P.W., Viergever, Max A., Chung, Albert C.S., and Shen, Helen C. "Enforcing stochastic inverse consistency in non-rigid image registration and matching." *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (2008): 1–8.

[72] Yeung, S.K. and Shi, P.C. "Stochastic inverse consistency in medical image registration." *International Conference on Medical Image Computing and Computer-Assisted Intervention* 8 (2005).2: 188–196.

[73] Ying, L., Liu, B., Steckner, M.C., Wu, G., and M. Wu, S.-J. Li. "A Statistical Approach to SENSE Regularization With Arbitrary k-Space Trajectories." *Magn. Reson. Med.* 60 (2008): 414–421.

[74] Zeng, Q. and Chen, Y. "Accurate Inverse Consistent Non-rigid Image Registration and Its Application on Automatic Re-contouring." *Proceedings of International Symposium on Bioinformatics Research and Applications* (2008): 293–304.

[75] Zhang, Xiaoqun, Burger, Martin, Bresson, Xavier, and Osher, Stanley. "Bregmanized Nonlocal Regularization for Deconvolution and Sparse Reconstruction." *SIAM J. Imag. Sci.* 3 (2010).09-03: 253–276.

[76] Zhang, Xiaoqun, Burger, Martin, and Osher, Stanley. "A unified primal-dual algorithm framework based on Bregman iteration." Tech. Rep. 09-99, CAM UCLA, 2009.

[77] Zhang, Z., Jiang, Y., and Tsui, H. "Consistent multi-modal non-rigid registration based on a variational approach." *Pattern Recognition Letters* (2006): 715–725.

[78] Zhu, M. and Chan, T.F. "An Efficient Primal-Dual Hybrid Gradient Algorithm for Total Variation Image Restoration." Tech. Rep. 08-34, CAM UCLA, 2008.

[79] Zhu, M., Wright, S.J., and Chan, T.F. "Duality-Based Algorithms for Total-Variation-Regularized Image Restoration." *Comput. Optim. Appl.* 47 (2010): 377–400.

BIOGRAPHICAL SKETCH

Xiaojing Ye was born in Beijing, China, on August 7th, 1983. He received his bachelor's degree in pure and applied mathematics from the School of Mathematical Sciences at Peking University, Beijing, China, in July 2005. He received the Master of Statistics and Doctor of Philosophy in Mathematics in December 2009 and May 2011, respectively.