Trust region methods for estimation of a complex exponential decay model in MRI with a single-shot or multi-shot trajectory (IN REVIEW)

Chenxi Hu, *Student Member, IEEE* and Stanley Reeves, *Senior Member, IEEE*

***Abstract*—Joint estimation of spin density, R_2^* decay and off-resonance frequency maps is very useful in many magnetic resonance imaging (MRI) applications. The standard multi-echo approach can achieve high accuracy but requires a long acquisition time for sampling multiple k-space frames. There are many approaches to accelerate the acquisition. Among them, single- or multi-shot trajectory based sampling has recently drawn attention due to its fast data acquisition. However, this sampling strategy destroys the Fourier relationship between k-space and images, leading to a great challenge for the reconstruction. In this work, we present two trust region methods based on two different linearization strategies for the nonlinear signal model. A trust region is defined as a local area in the variable space where a local linear approximation is trustable. In each iteration, the method minimizes a local approximation within a trust region so that the step size can be kept in a suitable scale. A continuation scheme is applied to gradually reduce the regularization over the parameter maps and facilitate convergence from poor initializations. The two trust region methods are compared to two other previously proposed methods—the nonlinear conjugate gradients and the gradual refinement algorithm. Experiments based on various synthetic data and real phantom data show that the two trust region methods have a clear advantage in both speed and stability.**

***Index Terms*—R_2^* estimation, field map estimation, trust region method, iterative algorithm**

I. Introduction

QUANTIFICATION of R_2^* and off-resonance frequency maps is very important in many MR applications. For example, R_2^* quantification can be applied to BOLD functional MRI [1], [2], iron deposition measurement [3], [4], and early detection of articular joint degeneration [5], [6]. Off-resonance quantification can be used to evaluate the severity of B_0 inhomogeneity [7], and it shows promise in MR thermometry [8]. In this paper, we focus on joint reconstruction of spin density, R_2^* decay and off-resonance frequency maps through a single- or multi-shot trajectory. We model the time-varying signal at each voxel as a mono-exponential function. Mono-exponential modeling has been extensively used in many relaxation time mapping applications such as T_2^* mapping [1]–[6], T_2 mapping [9], [10], T_1 mapping [11], and $T_{1\rho}$ mapping [10]. In this work, the initial value of the mono-exponential model is spin density. The decay rate of the mono-exponential model is a complex number whose real and imaginary parts are the R_2^* decay and off-resonance frequency, respectively. It should be noted that some other signal models [12]–[15] can fit the image series more accurately than mono-exponential in some applications. However, these models, such as multi-exponentials [13], are usually more complicated and are not addressed specifically in this work.

Among a variety of approaches to reconstruct R_2^* decay rate (reciprocal of T_2^*) or off-resonance frequency, the multi-echo based approach is an important one in many clinical and research applications [1]–[6]. Typically, these methods acquire multiple k-space frames each at a different echo time. After the acquisition, these k-space frames are inverse Fourier transformed and then a prior signal model such as the mono-exponential model is fitted to the time-varying signal on a voxel-by-voxel basis. The fitting gives an estimation of model parameters, which in our application are the spin density, R_2^* decay, and off-resonance frequency maps. This approach usually achieves high estimation accuracy and is therefore the mainstream method for relaxation time reconstruction. However, since multiple k-space frames must be fully sampled, this approach takes a relatively long data acquisition time, ranging from tens of seconds to minutes for one image slice [3]–[7].

A remedy for the long acquisition time is to undersample the multi-echo k-space frames [16]–[22]. Since each k-space frame is undersampled, the acquisition time for multiple k-space frames is reduced. However, the undersampling causes an underdetermined reconstruction problem for each k-space frame. Prior knowledge about the signal model is thus required to make the problem well-posed. A common approach is to exploit sparsity of the signal model in either spatial domain [22], or temporal domain [19], [20], or simultaneously in both domains [21]. For these techniques, the undersampling rate of k-space is usually around 4-6.

As a similar idea to undersampling, reconstruction based on single- or multi-shot trajectories has attracted increasing interest [23]–[26]. These methods discard the idea of acquiring multiple k-space frames; instead, they use a long readout time in each readout cycle to expand their sampling to cover a long time frame. At each time point, only one or a few k-space samples are acquired, depending on the number of single-shot trajectories used. However, the sampling density in the time domain becomes much larger compared to multi-echo sampling. Such sampling can be done in tens of milliseconds, dramatically reducing the acquisition time. The challenge,

C. Hu is with the Department of Electrical and Computer Engineering, Auburn University, Alabama, AL, 36830 USA e-mail: czh0032@auburn.edu

S. Reeves is with the Department of Electrical and Computer Engineering, Auburn University, Alabama, AL, 36830 USA e-mail: reevesj@auburn.edu

however, lies in the exceedingly difficult image reconstruction problem. We have observed the problem to be ill-conditioned, nonlinear, and of large scale. This work thus focuses on algorithms developed to stably and rapidly solve the problem.

Reference [23] may be one of the earliest papers that proposed to reconstruct spin density, R_2^* decay, and off-resonance frequency maps through a single-shot trajectory. Twieg reported that this method, named by PARSE, can be applied to BOLD functional MRI to increase the estimation precision and robustness. In this paper, Twieg explicitly formulated the nonquadratic cost function and minimized it with a steepest descend method which runs more than one hour to obtain a satisfactory result. Reference [24] is an extension of the previous paper, where Twieg and Reeves proposed a customized conjugate gradient method that requires several minutes for the same problem. They adopted a continuation framework on data length to avoid local minima and increase convergence speed. However, the robustness of the method to different initializations and objects is unclear. Olafsson et al. proposed a gradual refinement method to jointly reconstruct R_2^* decay and off-resonance maps [26]. The algorithm refines each iterate toward the solution by employing the Gauss-Newton method with linear conjugate gradients for each subproblem. However, this method does not have a line search in each iteration. It also requires a given spin density map as an input and only reconstructs the R_2^* decay and off-resonance maps. A good initialization for the two unknown maps is required for this method since a line search is missing.

Overall, the published literature on joint reconstruction of spin density, R_2^* decay and off-resonance maps through a single- or multi-shot trajectory is still sparse. Quite commonly, the previously presented methods use a conjugate gradient method or one of its variants to address the associated large-scale nonlinear optimization problem. A good initialization is also commonly used in previous methods [24], [26]. In this work, we propose two novel regularized trust region (TR) continuation methods to address the same optimization problem robustly. TR is a classical iterative method for nonlinear optimization [27]–[29]. A trust region is defined as a local area in the variable space where a local linear approximation is trustable. In each iteration, the method minimizes a local approximation within a trust region so that the step size can be kept in a suitable scale. We compare the trust region method to the nonlinear conjugate gradient method and the gradual refinement method. In order to test the robustness, we emphasize the capability of converging from a poor initialization. A continuation method is applied with all other algorithms to increase the range of initializations.

The rest of the paper is organized as follows. Section II introduces the dynamic signal model and the cost function. Section III elaborates on the two trust region methods and their implementation details. The two algorithms used for comparison and the continuation method are also covered in Section III. Section IV shows details associated with reconstruction based on real data. Section V and section VI draw various comparisons between the trust region methods and the two other methods based on various synthetic data and real data. Section VII discusses the origin of each iterative method as well as limitations of the two trust region methods and draws some conclusion.

II. Cost function formulation

Let $m(\vec{r})$, $d(\vec{r})$, $f(\vec{r})$ represent the spin density, the R_2^* decay and the off-resonance frequency of a voxel located at position $\vec{r}$ (relative to the center of the field of view). The time-varying signal at this voxel is modeled as

$$m(\vec{r})e^{(-d(\vec{r})+\iota f(\vec{r}))t_l} \tag{1}$$

where ι represents the imaginary unit. Plugging in $z(\vec{r}) \triangleq -d(\vec{r}) + \iota f(\vec{r})$, we rewrite (1) as

$$m(\vec{r})e^{z(\vec{r})t_l} \tag{2}$$

In the following parts, we call $z(\vec{r})$ the complex frequency at location $\vec{r}$. Let $\vec{k}_{lp}$ denote a multi-shot trajectory where $l = 1, 2, \ldots, L$ and $p = 1, 2, \ldots, P$ denote the lth time sample and the pth trajectory shot among L samples and P shots. When a single-shot trajectory is used, $P = 1$. The acquired k-space data from a multi-shot trajectory can then be represented by

$$y_{lp} = s_{lp}(m(\vec{r}), z(\vec{r})) + \varepsilon_{lp} \tag{3}$$

$$s_{lp}(m(\vec{r}), z(\vec{r})) = \int_{\vec{r}} m(\vec{r})e^{z(\vec{r})t_l}e^{-2\pi\iota(\vec{k}_{lp}\cdot\vec{r})}d\vec{r} \tag{4}$$

for $l = 1, 2, \ldots, L$ and $p = 1, 2, \ldots, P$. y_{lp}, s_{lp} and ε_{lp} represent the acquired k-space data, the predicted k-space signal, and the noise at time t_l and shot p, respectively. The dot operator in the exponential term represents inner product. Given data y_{lp} for L discrete time points and P trajectory shots, we want to estimate $m(\vec{r})$ and $z(\vec{r})$ for every voxel in the FOV. To proceed, (4) requires a suitable discretization in the spatial domain since data is discrete while unknown maps are (piecewise) continuous. We therefore introduce a finite-dimensional representation for $m(\vec{r})e^{z(\vec{r})t_l}$ so that

$$m(\vec{r})e^{z(\vec{r})t_l} \approx \sum_n m_n e^{z_n t_l} g(\vec{r} - \vec{r_n}) \tag{5}$$

where r_n is the nth voxel geometric center, and $m_n \triangleq m(\vec{r_n})$, $z_n \triangleq z(\vec{r_n})$. Function $g(\vec{r})$ is chosen as a linear interpolation basis function in our work. With the approximation (5), we can rewrite (4) as

$$s_{lp}(m_n, z_n) \approx \tilde{g}_{lp} \sum_{n=0}^{N-1} m_n e^{z_n t_l} e^{-2\pi\iota(\vec{k}_{lp}\cdot\vec{r}_n)} \tag{6}$$

where $\tilde{g}_{lp} \triangleq \tilde{g}(\vec{k}_{lp})$ represents the sample of the Fourier transform of $g(\vec{r})$ at $\vec{k}_{lp}$. Because the noise in (4) is Gaussian, we estimate the discrete unknowns based on minimizing a least-squares objective function

$$\|\vec{y} - \vec{s}(\vec{m}, \vec{z})\|^2 \tag{7}$$

where

$$\begin{aligned}
\vec{y} &\triangleq [y_{11}, y_{21}, \ldots, y_{L1}, y_{12}, \ldots, y_{L2}, \ldots, y_{LP}]^T \\
\vec{s} &\triangleq [s_{11}, s_{21}, \ldots, s_{L1}, s_{12}, \ldots, s_{L2}, \ldots, s_{LP}]^T \\
\vec{m} &\triangleq [m_1, m_2, \ldots, m_N]^T \\
\vec{z} &\triangleq [z_1, z_2, \ldots, z_N]^T
\end{aligned}$$

Since the function (7) is usually ill-conditioned [30], we apply regularization with respect to the unknown maps, which changes (7) to:

$$\|\vec{y} - \vec{s}(\vec{m}, \vec{z})\|^2 + \lambda_1 \|D_1 \vec{m}\|^2 + \lambda_2 \|D_2 \vec{z}\|^2 \tag{8}$$

where the matrices D_1 and D_2 are the regularization matrices corresponding to $\vec{m}$ and $\vec{z}$. For simplicity, we use the first-order smoothness penalty for both D_1 and D_2 in this work. Separate regularizations on the real and imaginary parts of $\vec{z}$ have also been utilized by other authors [26]. In the next section, we focus on algorithms that can stably and rapidly minimize (8) with respect to $\vec{m}$ and $\vec{z}$.

III. Proposed trust region methods

In this section, we present two trust region (TR) methods [27] namely the ordinary trust region method (OTR) and the the change-of-variable trust region method (CVTR) due to their different linearization strategies,

A. The ordinary trust region method

1) Approximating the local cost function: Finding a quadratic approximation of the nonquadratic cost function is the first step in utilizing a trust region method. Let m_{0n}, z_{0n} represent the parameters at a reference voxel and Δm_n, Δz_n the offsets. Taking the first-order Taylor expansion about m_{0n} and z_{0n} in the discrete version of (2) leads to the approximation

$$m_n e^{z_n t_l} \approx m_{0n} e^{z_{0n} t_l} + \Delta m_n e^{z_{0n} t_l} + m_{0n} e^{z_{0n} t_l} \Delta z_n t_l \tag{9}$$

where Δm_n and Δz_n are sufficiently small. With this approximation, one then minimizes the following function in each iteration:

$$\begin{aligned} &\min_{\Delta\vec{m}, \Delta\vec{z}} \ \|\vec{\zeta} - [\Theta_1 \ \Theta_2][\Delta\vec{m}^T \ \Delta\vec{z}^T]^T\|^2 \\ &\text{s.t. } \|\Delta\vec{m}\|_2 \le \eta_1; \|\Delta\vec{z}\|_2 \le \eta_2 \end{aligned} \tag{10}$$

where $\eta_1, \eta_2 \in \mathbb{R}^+$ and

$$\begin{aligned} [\vec{\zeta}]_{(l,p)} &\triangleq y_{lp} - \sum_n m_{0n} e^{z_{0n} t_l} e^{-2\pi\iota(\vec{k}_{lp} \cdot \vec{r}_n)}; \\ [\Theta_1]_{(l,p),n} &\triangleq \tilde{g}_{lp} e^{z_{n0} t_l} e^{-2\pi\iota(\vec{k}_{lp} \cdot \vec{r}_n)}; \\ [\Theta_2]_{(l,p),n} &\triangleq \tilde{g}_{lp} t_l m_{0n} e^{z_{n0} t_l} e^{-2\pi\iota(\vec{k}_{lp} \cdot \vec{r}_n)} \end{aligned}$$

where $(l,p) \triangleq (p-1)L + l$. This is a constrained linear optimization problem.

2) Solving the sub-problem: Equation (10) with the regularization in (8) is equivalent to

$$\begin{aligned} \min_{\Delta\vec{m}, \Delta\vec{z}} \ &\|\vec{\zeta} - [\Theta_1 \ \Theta_2][\Delta\vec{m}^T \ \Delta\vec{z}^T]^T\|^2 \\ &+ \lambda_1 \|D_1(\vec{m}_0 + \Delta\vec{m})\|^2 + \lambda_2 \|D_2(\vec{z}_0 + \Delta\vec{z})\|^2 \\ &+ \sigma_1 \|\Delta\vec{m}\|^2 + \sigma_2 \|\Delta\vec{z}\|^2 \end{aligned} \tag{11}$$

where σ_1 and σ_2 are properly chosen positive numbers. To solve (11), we choose preconditioned linear conjugate gradients (PCG) [31]. We use diagonal preconditioners which are defined as the inverse of the diagonal of the Hessian in (11). The maximal number of iterations for the sub-problem is set to 40. However, convergence is usually reached within this number with either a small gradient or a low iterate variation.

3) Summarizing OTR: The following procedure defines one iteration of OTR. Throughout our work, we use $\mu_{1,2,3,4} = 0.60, 2, 0.99, 0.7$.

Procedure 1

- If gradient or iterate variation is sufficiently small, the algorithm is stopped.
- Minimize (11) with PCG. Let the solution be $\Delta\vec{m}$, $\Delta\vec{z}$.
- Calculate the ratio γ between the decrease in (10) and the decrease in (8) caused by the new iterate.
- If $\gamma > 0$ (descent), $\vec{m} \leftarrow \vec{m} + \Delta\vec{m}$ and $\vec{z} \leftarrow \vec{z} + \Delta\vec{z}$.
- If $\gamma < \mu_1$, then $\sigma_1 \leftarrow \mu_2\sigma_1$, $\sigma_2 \leftarrow \mu_2\sigma_2$. If $\gamma > \mu_3$, then $\sigma_1 \leftarrow \mu_4\sigma_1$, and $\sigma_2 \leftarrow \mu_4\sigma_2$.

B. The change-of-variable trust region method with regularization

1) Introducing CVTR: The difference between OTR and CVTR lies in the local approximation. For CVTR, the nonlinear signal model is first approximated by

$$m_n e^{z_n t_l} \approx m_n e^{z_{0n} t_l}(1 + \Delta z_n t_l) = m_n e^{z_{0n} t_l} + e^{z_{0n} t_l} m_n \Delta z_n t_l \tag{12}$$

and then followed by combining $m_n \Delta z_n$ into a new variable c_n

$$m_n e^{z_n t_l} \approx e^{z_{0n} t_l} m_n + t_l e^{z_{0n} t_l} c_n \tag{13}$$

Equation (13) also explains how the method is named. The approximation leads to a different data fidelity term from the one in (10):

$$\|\vec{y} - [\Theta_1 \ \Theta_3][\vec{m}^T \ \vec{c}^T]^T\|^2 \tag{14}$$

where

$$[\Theta_3]_{(l,p),n} = \tilde{g}_{lp} t_l e^{z_{n0} t_l} e^{-2\pi\iota(\vec{k}_{lp} \cdot \vec{r}_n)}$$

2) Reformulating the cost function for CVTR: A problem arises when constructing the approximated cost function for CVTR. The regularization of $\vec{z}$ would introduce a nonlinear term in the cost function since $z_n = c_n/m_n$ for the nth voxel. A straightforward solution is to regularize the point-wise product of $\vec{m}$ and $\vec{z}$. This solution leads to a new cost function

$$\|\vec{y} - \vec{s}(\vec{m}, \vec{z})\|^2 + \lambda_3 \|D_1 \vec{m}\|^2 + \lambda_4 \|D_2(\vec{m} \odot \vec{z})\|^2 \tag{15}$$

where $\lambda_{3,4} \in \mathbb{R}^+$ and $\odot$ represents the point-wise product between two vectors. The main difference compared to (8) lies in the regularization over $\vec{z}$. A new sub-problem also arises for CVTR

$$\begin{aligned} &\min_{\vec{m}, \vec{c}} \ \|\vec{y} - [\Theta_1 \ \Theta_3][\vec{m}^T \ \vec{c}^T]^T\|^2 \\ &+\lambda_3 \|D_1 \vec{m}\|^2 + \lambda_4 \|D_2(\vec{m} \odot \vec{z}_0 + \vec{c})\|^2 \\ &\text{s.t. } \|\vec{m} - \vec{m}_0\|_2 \le \eta_3; \|\vec{c}\|_2 \le \eta_4 \end{aligned} \tag{16}$$

where $\eta_3, \eta_4 \in \mathbb{R}^+$. Similar to (11), we use PCG to solve the equivalent problem of (16) in each iteration:

$$\begin{aligned} \min_{\vec{m}, \vec{c}} \ &\|y_l - [\Theta_1 \ \Theta_3][\vec{m}^T \ \vec{c}^T]^T\|^2 \\ &+ \lambda_3 \|D_1 \vec{m}\|^2 + \lambda_4 \|D_2(\vec{m} \odot \vec{z}_0) + D_2 \vec{c}\|^2 \\ &+ \sigma_3 \|\vec{m} - \vec{m}_0\|^2 + \sigma_4 \|\vec{c}\|^2 \end{aligned} \tag{17}$$

where σ_3 and σ_4 are properly chosen positive numbers.

3) Specifying the trust region for CVTR: The trust region of CVTR is less intuitive than OTR. Since $|\Delta b_n| = |c_n/m_n|$ must be bounded above, $|c_n|$ must be bounded above and $|m_n|$ must be bounded below. However, a lower bound for the modulus of a complex number is hard to implement. A more practical method is to set a bound for the variation of $\vec{m}$ like the bound for $\Delta\vec{m}$ in OTR. This bounding method gives the constraints in (16).

The two trust regions associated with the two linearization methods bear further discussion. Figure 1 illustrates their differences. The trust region of CVTR is nonconvex in the $(\vec{m}, \vec{z})$ domain due to the quotient c_n/m_n. This trust region makes CVTR either slower than OTR when the trust region is small or causes unacceptable updates when the trust region is large. An advantage of the CVTR trust region is that it can be much broader in the spin density direction. This advantage may help stabilize the convergence of the spin density since the variation of the spin density has a small influence on the convergence. This advantage may also speed up the convergence of the spin density when decay and frequency have good initial guesses.

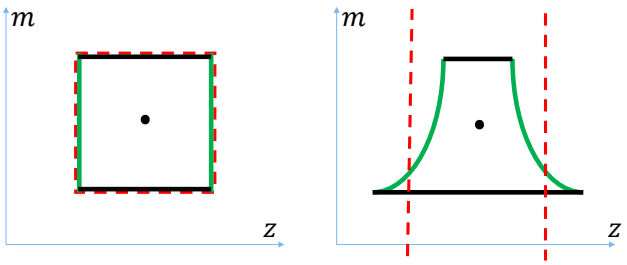


Fig. 1: Comparison between OTR (left) and CVTR (right) trust regions in a simplified case where m and z are both real scalars. The black and green lines represent trust region constraints associated with m and z, respectively. The red dash line represents the support for the local linearization.

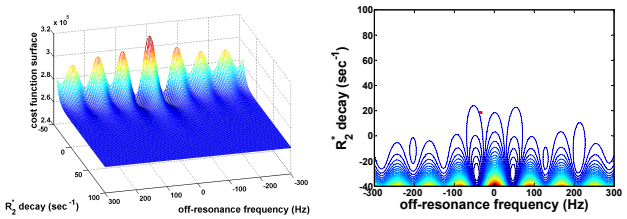
4) Summarizing CVTR: The following defines an iteration of CVTR. $\mu_{1,2,3,4} = 0.60, 2, 0.99, 0.7$ as for OTR.

Procedure 2

- If gradient or iterate variation is sufficiently small, then the algorithm is stopped.
- Minimize function (17) with PCG. Let the solution be $\vec{m}_{est}$, $\vec{c}_{est}$.
- Calculate the ratio γ between the decrease in (16) and the decrease in (15) caused by the new iterate.
- If $\gamma > 0$ (descent), $\vec{m} \leftarrow \vec{m}_{est}$ and $\vec{z} \leftarrow \vec{z} + \frac{\vec{c}_{est}}{\vec{m}_{est}}$.
- If $\gamma < \mu_1$, then $\sigma_1 \leftarrow \mu_2\sigma_1$, $\sigma_2 \leftarrow \mu_2\sigma_2$. If $\gamma > \mu_3$, then $\sigma_1 \leftarrow \mu_4\sigma_1$, and $\sigma_2 \leftarrow \mu_4\sigma_2$.

C. Ill-conditioning and nonlinearity

We observed the conditioning of jointly estimating the spin density, R_2^* decay and the off-resonance frequency maps out of a single-shot trajectory is very poor [30]. The severe ill-conditioning as well as the nonlinearity of the signal model (2) result in a great challenge for each iterative method. This situation is illustrated in Figure 2, where the cost function without regularization is plotted by varying R_2^* decay and off-resonance frequency at a reference voxel and holding all other values constant. On one hand, the ill-conditioning mainly exists in the R_2^* decay direction, since a large variation of the R_2^* decay introduces a small variation of the function value. On the other hand, the nonlinearity of the signal model is reflected by the multiple local minima in the axis of off-resonance frequency of this subspace. Such a function profile dramatically increases the challenge for an iterative method.



(a) (b)

Fig. 2: Cost function without regularization in a neighborhood of $m_n = 0.5, z_n = 0$ for all n. (a) 3d plot of the cost function, and (b) level sets of the cost function in (a). The actual complex frequency z_{n_0} at this voxel is $z_{n_0} = -17.19 \text{ sec}^{-1} - \iota 24.04$ Hz, located by the red square in (b).

D. Alternative methods for comparison

We use a nonlinear conjugate gradient (NCG) method that we previously developed for the same problem [32]. In each iteration, the method uses an interpolating polynomial to make an inexact line search, followed by a check to see whether the residual decreases. If the residual does not decrease, an exact line search based on a bisection method is applied. This procedure improves the efficiency of the NCG method. For this particular problem, rescaling of the variables was reported to be necessary in order to hold m_n and z_n at the same scale [16]. However, the rescaling parameter is usually empirically selected, and any scale mismatch can cause a poor optimization [18]. In this work, we replace rescaling with a diagonal preconditioning which is defined for this nonlinear forward operator as $\mathbf{J}^H\mathbf{J}$, where $\mathbf{J}$ represents the Jacobian matrix at each iterate and H represents conjugate transpose. We observe that the method speed is significantly improved by the preconditioning. Moreover, preconditioning makes the comparison between NCG and trust region methods fairer since both use the same preconditioner.

The gradual refinement (GR) method used in our context is inspired by the method developed in [26]. The reference method only reconstructs R_2^* decay and off-resonance frequency maps. In our work, we extend the reference method to simultaneously reconstruct the three parameter maps. In each iteration, the GR method minimizes the approximated cost function in (10) without the constraint. The method then accepts the minimizer of the approximated function without a line search (i.e. step length is always set to 1). [26] states that the method converges well when a good initialization is present. In our work, we observe that the GR method

has a similar behavior. However, GR can sometimes fail the reconstruction when initialization is poor, which makes the algorithm behavior unpredictable. This characteristic is compared to the trust region methods to demonstrate their advantage in stability for this problem.

E. Continuation methods

Solving an ill-conditioned nonlinear equation system from poor initializations usually requires a continuation method [33], [34]. The continuation method embeds the original problem into a one-parameter series of problems where the complexity of the function surface and therefore the estimation difficulty monotonically increase. The original problem is set to be the last problem of this series with the previous solution used as the initialization, so the sequence of solutions converges to the solution of the original problem. Examples of this method include the fixed-point continuation method (FPC) devised for sparse reconstruction [35] and the progressive data length method in [24].

In our work, the regularization parameters must be set high when the algorithm starts from a distant initial guess so that local minima can be avoided (Figure 2). The strong regularization causes severe oversmoothing artifacts in the estimates. These artifacts are then reduced by gradually reducing the regularization parameters. Each set of the parameters is called a continuation phase. The continuation method dramatically increases the probability of convergence from a distant initialization for all algorithms. For all our experiments, the algorithms actually used are the trust region continuation methods, conjugate gradient continuation method (CGC), and the gradual refinement continuation method (GRC). The trust region continuation methods include the ordinary trust region continuation method (OTRC) and the change-of-variable continuation method (CVTRC).

How to set the continuation scheme turns out to be nontrivial. In this work, we use the same continuation scheme for all methods and this continuation scheme is found in a trial-and-error manner. Specifically, we first decide the regularization parameters based on the object smoothness and the noise level (e.g., running OTR with a good initialization for varying regularization parameters). We then choose a suitable number of continuation phases based on the quality of the initialization. The closer the initialization is to the minimizer, the fewer continuation phases we need for convergence. We then find out the reduction factor for all regularization parameters by trial-and-error.

Procedure 3

- Set λ_1, λ_2 (λ_3, λ_4 for CVTRC)
- For $j = 1, \ldots, J$ (outer loop)
  - Set $\sigma_{1,2}$
  - For $i = 1, \ldots, I(j)$ (middle loop)
    - Run Procedure 1 or Procedure 2
  - End
  - $\lambda_1 \leftarrow \lambda_1/\xi_1$; $\lambda_2 \leftarrow \lambda_2/\xi_2$ or $\lambda_3 \leftarrow \lambda_3/\xi_1$; $\lambda_4 \leftarrow \lambda_4/\xi_2$
- End

F. Choices of parameters

CVTRC and OTRC are nested algorithms of three program loops. The outer and middle loop are shown in Procedure 3; the inner loop is the PCG loop in Procedure 1 & 2. In the outer loop, we need to initialize $\lambda_{1,2}$ or $\lambda_{3,4}$ (regularization parameters), assign ξ_1 and ξ_2 (regularization reductions in each continuation phase), and decide J (number of continuation phases). We also need to initialize $\sigma_{1,2}$ before each middle loop. $\lambda_{1,2}$ or $\lambda_{3,4}$ are somewhat dependent on the object and noise level; however, the dependence is continuous and rather insensitive to many different objects. For example, a phantom and a human brain slice require about the same regularization parameters based on our experiments, since they are similarly smooth. $\xi_{1,2}$ and J can be predetermined; in fact, in all our experiments $\xi_1 = 10$, $\xi_2 = 6$ and $J = 4$. σ_1 and σ_2 are initialized to 1e4 and 1e2 respectively in the first continuation phase for all experiments. After that, they are automatically updated based on Procedure 1 & 2 and passed on to the next continuation phase when j increases.

The middle loop consists of $I(j)$ iterations of Procedure 1 (for OTRC) or Procedure 2 (for CVTRC). In our work, $I(1), I(2), I(3), I(4)$ are set to be (30, 10, 10, 5) for both OTRC and CVTRC. We set $I(1)$-$I(4)$ differently to avoid too many iterations for each continuation phase. Usually, the two algorithms activate the stopping rule before the iteration limits. Even when that does not happen, these numbers are large enough to obtain a significant improvement in the estimate. Parameters in the inner loop are primarily $\mu_{1,2,3,4}$. Choices of these parameters can be found in standard literature on trust region methods [27], [29], since they are a part of the algorithm. Since NCGC and GRC address the same cost function (8) with OTRC, we use identical regularization parameters and continuation parameters for these two comparison algorithms.

.

IV. Real data reconstruction

The ultimate goal of the trust region algorithms is to reconstruct from real MRI data with a minimal number of trajectories. An important problem with real data is lack of ground truth. In this work, we used a multi-echo gradient echo sequence to estimate the three parameter maps for a cylindrical phantom and treated the estimates as the gold standard [1]–[6]. The parameters for the multi-echo gradient echo sequence are: TE = 5, 6, 7, 8, $\cdots$ 82 ms, TR = 200 ms, image size = 64 $\times$ 64, bandwidth = 390 Hz/Pixel, FOV = 120 mm $\times$ 120 mm, slice thickness = 2 mm, flip angle = 15°. We used a long echo train for phantoms to reduce noise. We employed a 32-channel head coil but only used one channel of data for the reconstruction. We chose the channel with the most uniform sensitivity. After acquiring all k-space frames, we applied an inverse Fourier transform to every frame and then a curve fitting on a voxel-by-voxel basis to reconstruct the three parameter maps.

Theoretically, the data from any single-shot trajectory should be approximately equal to $\vec{s}(\vec{m}, \vec{z})$ in (7) with $\vec{m}$, $\vec{z}$ given by the estimates from the multi-echo approach. However, the difference between the actual k-space data and model k-space data is very large in our findings. The large model

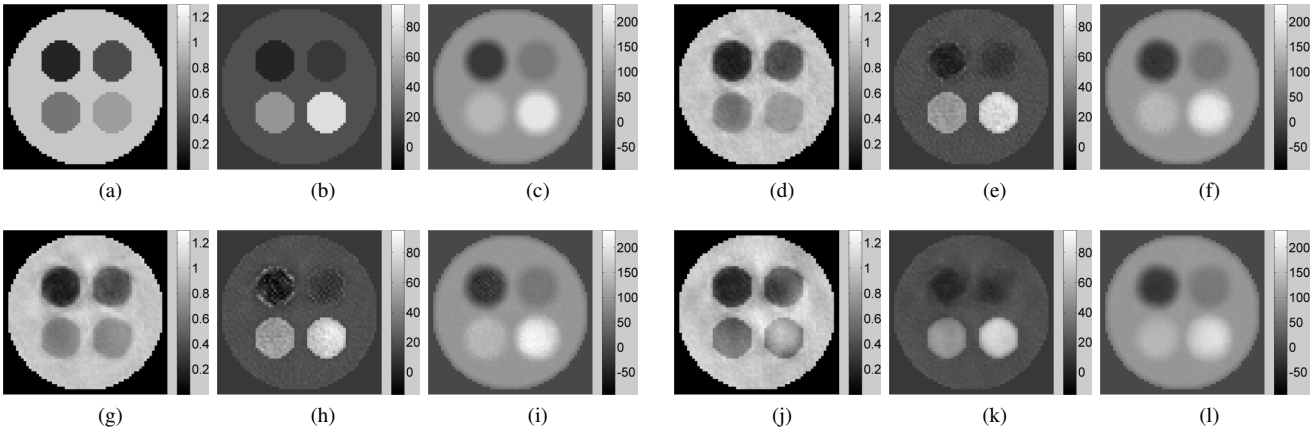


Fig. 3: Simulation results: (a-c) ground truth for the spin density, R_2^* decay and the off-resonance frequency, (d-f) OTRC reconstruction results, (g-i) CVTRC reconstruction results, and (j-l) NCGC reconstruction results.

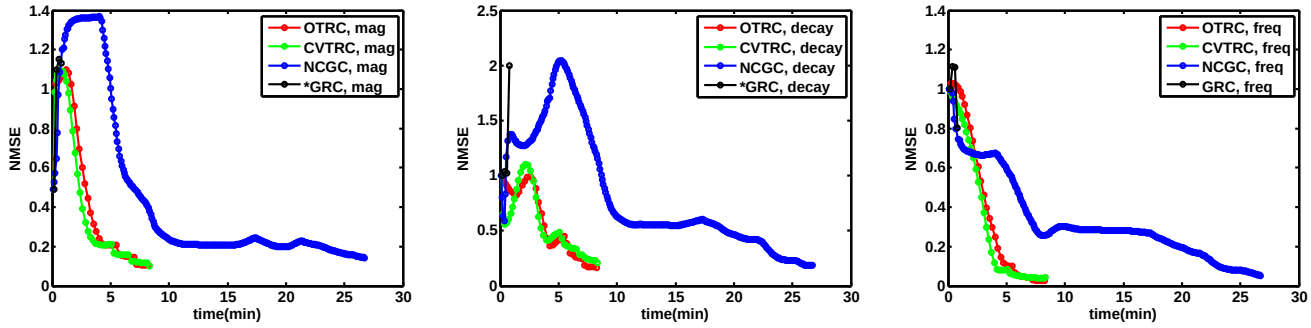


Fig. 4: Convergence profiles of spin density (left), R_2^* decay (middle) and off-resonance frequency (right). Red, green, blue, and black lines represent OTRC, CVTRC, NCGC, and GRC, respectively. * indicates the algorithm did not converge .

mismatch causes the reconstruction for all algorithms to be suboptimal. The reason is still under investigation. Possible reasons include hidden inconsistency between the multi-echo gradient echo sequence and the single-shot rosette trajectory, trajectory miscalibration, and perhaps some model limitations such as imperfect modeling for intra-voxel gradients. We have especially paid attention to trajectory miscalibration because it is a crucial factor within the model (6). We used the calibration technique proposed in [36] to reduce the miscalibration, but the precision level may be insufficient for this reconstruction problem.

The trajectory we used in our experiment is a rosette trajectory. This trajectory has been verified by others to be superior in reconstruction quality to other single-shot trajectories such as spirals and echo planar imaging (EPI) trajectories for this problem [23], [37]. Its formulation is shown below

$$\vec{k}_l = \frac{1}{2} k_{max} \sin(w_{osc} t_l) e^{\iota w_{rot} t_l + \theta} \tag{18}$$

where k_{max} denotes the range of k-space, w_{osc} the oscillation frequency, w_{rot} the rotation frequency, and θ the initial angle of the trajectory in the complex domain. In our work, w_{osc} = 3196 rad/sec, w_{rot} = 1577 rad/sec, time span of the trajectory is 81.92 ms and totally 8192 samples are acquired. When a single-shot rosette is used, θ=0°. However, since the model mismatch is large, we used multi-shot rosette trajectories for reconstruction to improve the conditioning and make the result less sensitive to modeling error. In particular, we used four interleaved single-shot rosette of which θ equals to $0°$, $22.5°$, $45°$, and $67.5°$, respectively. We emphasize that the focus of the real data reconstruction is on comparing the four algorithms on their convergence performance rather than their absolute accuracy relative to a gold standard.

V. Simulations

In this section, we show reconstructions based on a simulated phantom and a human brain slice. We used a single-shot rosette trajectory to synthesize k-space data, and the trajectory is specified in the previous section. The model of the k-space data is given by (6). The noise was white Gaussian and the signal-to-noise ratio (SNR) in this work is defined as

$$\text{SNR} = \frac{\|s\|_2}{\|s - s_0\|_2}$$

where s is the noisy data and s_0 is the noiseless data. Normalized mean square error (NMSE) was used as a metric for the accuracy of the reconstructions. The NMSE is defined

Table 1: Convergence accuracy and time of OTRC, CVTRC, NCGC, and GRC for different data sets

Data	Criteria	OTRC	CVTRC	NCGC	GRC
SNR = 100 (reference regularization)	Time(min)	**8.2**	8.3	26.7	--
	NMSE(SD, Decay, Freq.)	**(0.09, 0.14, 0.03)**	(0.09, 0.20, 0.04)	(0.14, 0.19, 0.05)	--
SNR = 20 (10x regularization)	Time(min)	**7.9**	8.2	27.2	--
	NMSE(SD, Decay, Freq.)	**(0.13, 0.26, 0.06)**	(0.14, 0.34, 0.08)	(0.21, 0.47, 0.23)	--
SNR = 10 (100x regularization)	Time(min)	7.8	9.7	25.7	**7.1**
	NMSE(SD, Decay, Freq.)	(0.18, 0.35, 0.10)	(0.20, 0.58, 0.10)	(0.26, 0.54, 0.33)	**(0.18, 0.34, 0.10)**
Double Resolution (SNR = 100, reference regularization)	Time(min)	10.9	**10.8**	34.2	--
	NMSE(SD, Decay, Freq.)	**(0.17, 0.29, 0.07)**	(0.16, 0.39, 0.10)	(0.22, 0.27, 0.07)	--

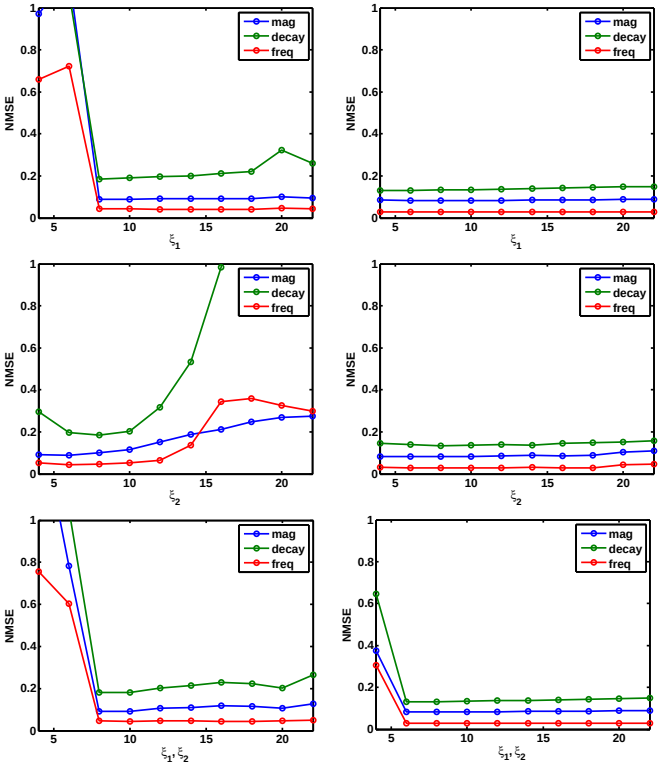


Fig. 5: Robustness of OTRC and CVTRC to variation of $\xi_{1,2}$. The top, middle and bottom lines show NMSE of CVTRC (left) and OTRC (right) for varying ξ_1 only, varying ξ_2 only, and simultaneously varying ξ_1 and ξ_2, respectively.

as:

$$\text{NMSE} = \frac{\|f - f_0\|_2}{\|f_0\|_2}$$

where f and f_0 represent the estimate and the ground truth. All methods started at a trivial initialization: $m_n = 0.5$ and $z_n = 0$ for all n. For all experiments, we used a masking technique

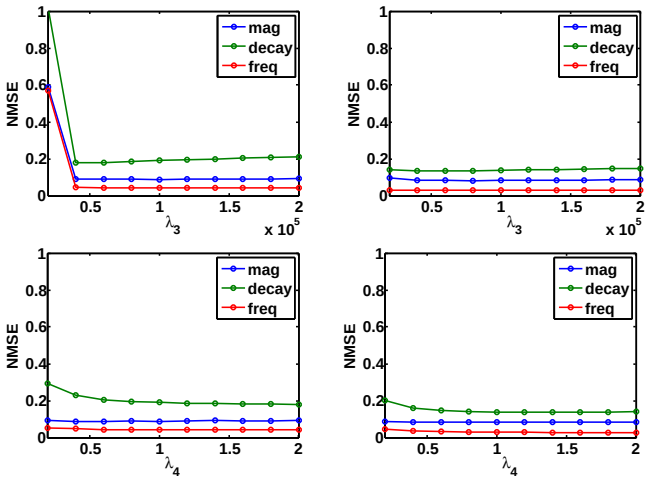
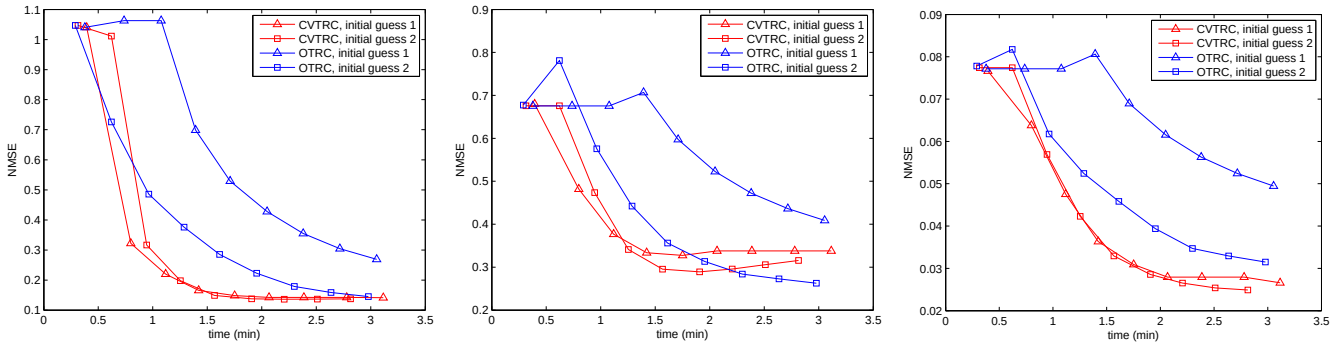


Fig. 6: Robustness of OTRC and CVTRC to regularization parameter variation. The top and bottom lines show NMSE of CVTRC (left) and OTRC (right) for varying λ_3 only and varying λ_4 only, respectively.

to mask out voxels with low spin density before applying any iterative algorithm. This technique has also commonly been used by some other authors [26], [38]. All algorithms were implemented using MATLAB and run with an Intel Core i7-4700MQ CPU.

A. *Simulated phantom*

We simulated a piecewise continuous cylinder phantom to mimic four small cylinders each filled with different materials and placed within a large cylinder container. Since the off-resonance map is often modeled as a smooth map [39], we have smoothed the off-resonance map by a circular averaging filter with a radius of 5 voxels. We use a triple to represent the maps of the spin density, R_2^* decay, and the off-resonance



(a) spin density (b) R_2^* decay (c) off-resonance frequency

Fig. 7: Verification of the faster convergence of CVTRC than OTRC when good initial guesses for R_2^* decay and off-resonance frequency are present.

frequency for each material. The units for the R_2^* decay and off-resonance frequency are sec^{-1} and Hz. The parameters of the material in the large container are (1, 20, 100). The parameters of the four small containers in the middle are (0.2, 2, -20), (0.4, 10, 60), (0.6, 50, 140), and (0.8, 80, 200) in a left-right, top-bottom order. Figure 3 (a-c) shows the three parameter maps for the phantom.

Figure 3 shows the estimation results obtained through OTRC, CVTRC, and NCGC with SNR $= 100$. All three algorithms obtained reasonable reconstructions. Single-shot reconstruction commonly manifests edge artifacts. We conjecture the artifacts are caused by insufficient sampling of the k-space high-frequency band for a single-shot trajectory. These artifacts are present for every method and do not influence the comparison. Also, these artifacts can be reduced by increasing the overall sampling using a technique such as a multi-shot acquisition.

Figure 4 shows the convergence profiles associated with OTRC, CVTRC, NCGC, and GRC for the same k-space data. OTRC has about the same profile with CVTRC. CVTRC appears to be slightly faster than OTRC and slightly less accurate than OTRC especially on the R_2^* decay reconstruction. However, both of them are much faster than NCGC. In addition, GRC does not converge. It stops because the residual and gradient become infinite in the next iterate. This numerical instability is caused by the fact that a line search is missing in GRC. Without a line search, the algorithm updates based on a local quadratic approximation, which may increase the actual cost function. Trust region methods never have this issue because 1) minimization associated with the sub-problem is always applied within a trust region and 2) a check is carried out in each iteration to guarantee the residual reduction.

1) Robustness to different SNRs and discretization: Table 1 shows the accuracy and time required by different methods in processing data sets with different SNR and different discretization resolution. With different SNRs (100, 20, and 10), all methods except GRC are stable. However, OTRC has higher accuracy than NCGC and also CVTRC, and both OTRC and CVTRC are much faster than NCGC. GRC only converges for SNR = 10, and its accuracy is roughly the same as OTRC. Since GRC and OTRC have the same cost function, similar accuracy is not surprising. However, the stability of GRC is a major problem for the algorithm in the case of a poor initialization. We need to point out that when SNR decreases, the regularization parameters for all methods increase in order to guarantee the convergence. However, since all methods have the same regularization parameters for a given SNR, a direct comparison of performance can still be made.

As real k-space data arises physically from continuous parameter maps, a higher discretization resolution was used when synthesizing k-space data to test the robustness of the algorithms. As shown in Table 1, double resolution in the data synthesis leads to longer reconstruction time and moderately worse accuracy for all algorithms. OTRC and CVTRC still achieve reasonable accuracy and still outperform NCGC and GRC.

2) Robustness to ξ and λ parameter sets : Figure 5 shows the different sensitivity of OTRC and CVTRC to variation of ξ_1 only, of ξ_2 only, and of ξ_1 and ξ_2 simultaneously. When only ξ_1 or only ξ_2 changes, CVTRC displays a large variation of accuracy while OTRC does not. When both ξ_1 and ξ_2 change by the same amount, CVTRC displays much better robustness than it does when a single parameter changes. This experiment demonstrates that the current implementation of CVTRC is susceptible to relative variation between ξ_1 and ξ_2. It also shows that OTRC is very robust for these parameters. Notice what $\xi_{1,2}$ influence are the two regularization parameters. Hence, CVTRC relies more on a good match of the two regularization parameters than OTRC. Figure 6 shows the performance of the two algorithms when only λ_1 or only λ_2 varies at the final continuation phase. The two algorithms worked well for a wide range of variations for $\lambda_{1,2}$.

3) Advantage of CVTRC: This experiment aims to show the speed advantage of CVTRC over OTRC when an educated $\vec{z}$ initialization exists. In this scenario, the fact that CVTRC has a larger approximation support in the spin density dimension should give it an advantage. We used the same simulated phantom and applied two initializations to the parameter maps. Both initializations are close to the ground truth of $\vec{z}$ and distant from that of $\vec{m}$. The first initialization adds

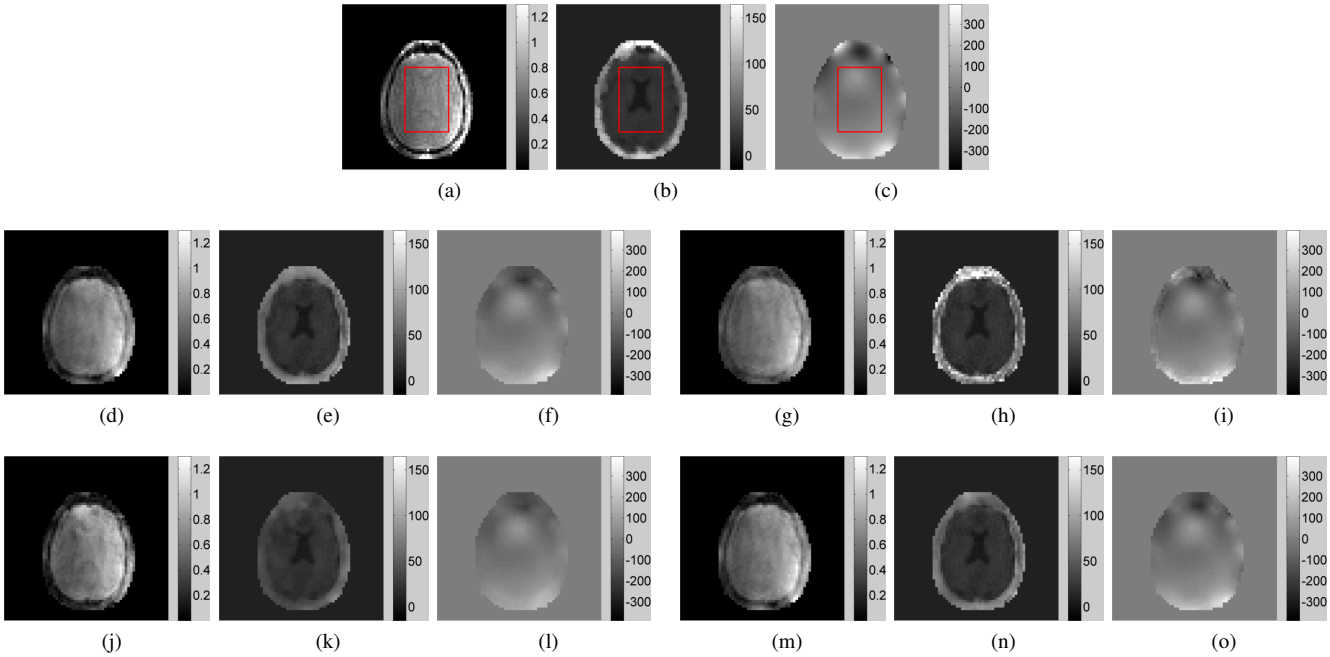


Fig. 8: A human brain transversal slice with synthetic k-space data. White noise with SNR = 100 added to the data. (a-c) Gold standard for the spin density, R_2^* decay and the off-resonance frequency, (d-f) OTRC reconstruction results, (g-i) CVTRC reconstruction results, (j-l) NCGC reconstruction results, and (m-o) GRC reconstruction results. NMSE of the rectangle area for the four algorithms: (5.0%, 5.2%, 2.7%) for OTRC, (4.3%, 6.1%, 3.6%) for CVTRC, (13.7%, 7.9%, 13.5%) for NCGC, and (4.2%, 4.4%, 2.4%) for GRC.

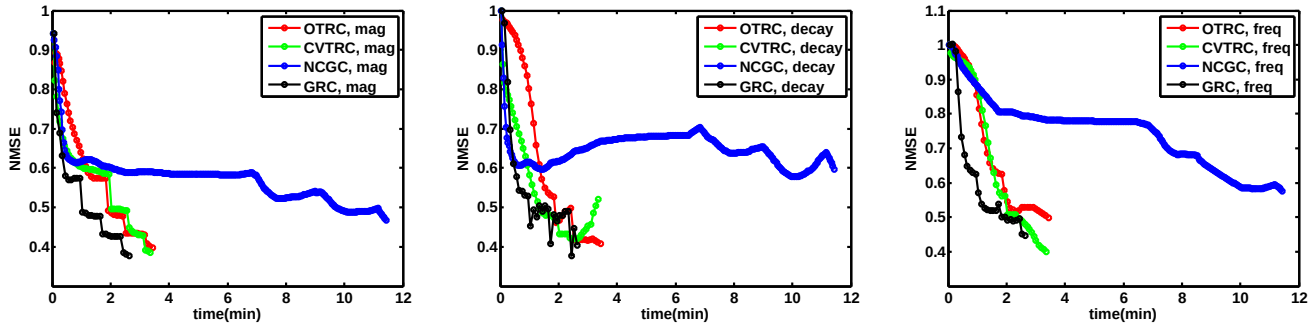


Fig. 9: Convergence profiles of the spin density (left), R_2^* decay (middle) and the off-resonance frequency (right). Red, green, blue, and black lines represent OTRC, CVTRC, NCGC, and GRC, respectively.

a small amount of noise to the ground truth of the $\vec{z}$ map, and the second initialization randomly rotates the $\vec{z}$ map by a small angle in space. Since the two initializations are only statistically similar in initializing $\vec{z}$, this experiment can well verify that CVTRC converges faster than OTRC when $\vec{z}$'s initialization is good. The results are shown in Figure 7. Despite the two initialization differences, CVTRC has a more uniform and faster convergence profile than OTRC, demonstrating the advantage of CVTRC.

B. Brain slice

Figure 8 shows the results of CVTRC, OTRC, NCGC, and GRC for a transversal slice of a human brain and synthetic k-space data. The slice crosses the orbitofrontal area of the brain cortex where a large off-resonance ranging from -414 Hz to 378 Hz is present. The three parameter maps were obtained through the aforementioned multi-echo approach. We used SNR = 100 in this simulation. All four methods converged by activating the stopping criterion, including GRC, which does not converge for the simulated phantom with the same SNR. In fact, as shown in Figure 9, GRC is even faster than OTRC by roughly 20%. The speed gain is due to the fact that GRC does not impose any trust region control and therefore the variables vary freely in every iteration. This can bring a faster convergence, as in this example, but it can also destroy the convergence, as shown in the last one. Thus, GRC is unpredictable. In comparison, OTRC uses the same cost function but with a trust region control and maintains a good

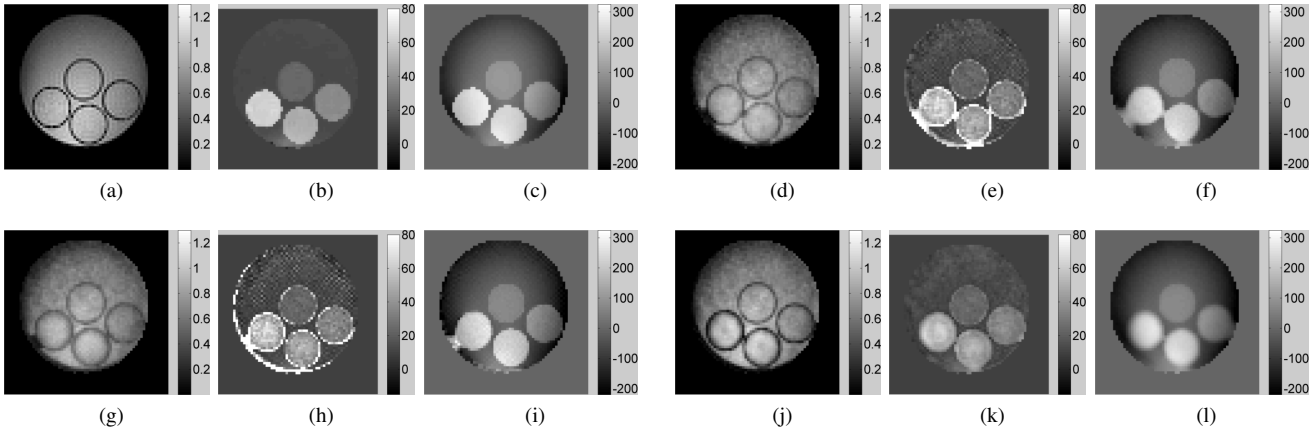


Fig. 10: Reconstruction of a real four-cylinder phantom with acquired k-t space data. (a-c) Gold standard for spin density, R_2^* decay and off-resonance frequency, (d-f) OTRC reconstruction results, (g-i) CVTRC reconstruction results, and (j-l) NCGC reconstruction results.

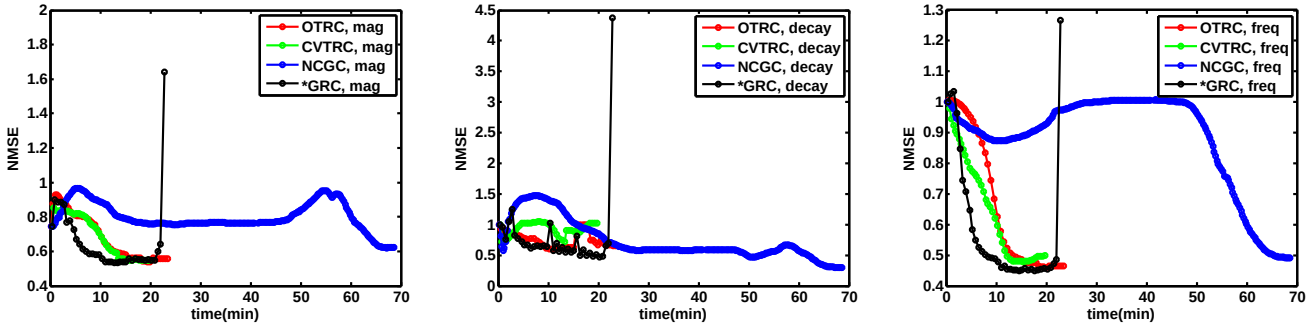


Fig. 11: Convergence profiles of spin density (left), R_2^* decay (middle) and off-resonance frequency (right). Red, green, blue, and black lines represent OTRC, CVTRC, NCGC, and GRC, respectively.

balance between stability and speed.

From Figure 8 and Figure 9, we can also see that the accuracy of reconstruction is generally worse than the simulated phantom. This is because in the surrounding area of the brain slice has either a fast R_2^* decay or a low spin density. A low spin density makes the estimation of the complex frequency $z(\vec{r})$ in these voxels challenging because the time signal is not sensitive to variations in $z(\vec{r})$ [40]. A large R_2^* decay makes estimation difficult for every parameter, since the amount of data is limited. In this example, the intensity of many voxels in this area becomes insignificant after 10 ms, leading to a very small number of useful samples compared to an 80 ms trajectory. These factors together cause a high estimation bias in the surrounding area, which increases the MSE of the entire image. We calculated the NMSEs of algorithms for a rectangular area in the middle, and these NMSEs are consistent with findings in the previous experiment.

CVTRC generally has a larger estimation error than OTRC in R_2^* decay reconstruction in the presence of poor initializations. We conjecture that this is caused by the difference in regularization used in CVTRC. In order to accommodate the change of variables, we regularize $\vec{m} \odot \vec{z}$ in the CVTRC cost function rather than $\vec{z}$ alone. This causes a suboptimal regularization over $\vec{z}$, especially when regularization parameters are small or when spin density is low.

VI. Real k-space data

As we state in section IV, we employed reconstructions from acquired multi-echo data to obtain a gold standard. A multi-echo rosette instead of a single-shot rosette was used to acquire more data and reduce the error caused by noise, model mismatch, and other possible unidentified factors. Figure 10 shows the reconstruction results of OTRC, CVTRC, and NCGC. GRC is not shown because it does not converge. The other three methods lead to a reasonable reconstruction, although considerable noise and boundary artifacts are present. Figure 11 shows the convergence profiles of the four methods. GRC failed to converge at the end, demonstrating its unpredictability. OTRC and CVTRC converged much faster than NCGC, demonstrating their speed advantage. More severe boundary artifacts were present in the OTRC and CVTRC results than the GRC results. In fact, all our experiments indicate that NCGC leads to a smoother result than other methods. This may be caused by the well-known regularization property of the conjugate gradient method when it stops early [41]. However, the long time required to converge dramatically

reduces this advantage of NCGC in the presence of a poor initialization.

VII. Conclusions and Discussions

We have presented a new approach to joint reconstruction of the spin density, R_2^* decay and off-resonance frequency maps through single- and multi-shot trajectories. This approach adopts the trust region framework commonly used to solve nonlinear equation systems. Due to different local approximation strategies, two trust region methods OTRC and CVTRC were proposed. Experiments show that the two methods outperform an NCG method in speed and the GR method in stability. The four methods were applied to synthetic k-space data based on simulated phantoms and an orbitofrontol brain slice as well as real phantom data. Between the two trust region methods OTRC and CVTRC, OTRC usually has smaller boundary artifacts, but CVTRC can be faster in the presence of a good complex frequency initialization. To our knowledge, this is the first paper focusing on algorithm stability and convergence speed in joint estimation of the three parameter maps.

The origins of the three methods NCG, TR, and GR can be traced to some of the classical methods in dealing with nonlinear optimization. NCG itself is a classical method in dealing with nonlinear optimization [42]. NCG holds a superlinear convergence rate and is therefore faster than steepest descent but slower than Newton-like methods. However, for a large system, even a superlinear convergence rate can lead to an intolerable computation time especially when a good preconditioning is absent (cf. [31] P.150). GR can be categorized as a truncated Gauss-Newton (TGN) method [43]. The method is rooted in the Gauss-Newton method, yet in each iteration it searches for the Gauss-Newton direction by iteratively solving an approximated linear system with methods such as conjugate gradients. This strategy makes TGN suitable for large-scale systems. Since TGN is a Newton-like method, it can possibly approach the quadratic convergence rate that the Newton method possesses. In fact, TGN has been reported to be faster than NCG with a suitable preconditioning in the inner problem [43]. However, an important aspect for a general unconstrained optimization problem is the line search in each iteration, which is lacking in the current GR method and its source method in [26]. The lack of line search causes the algorithm to be unstable since the residual at the next iterate may be far away from its predicted value. The trust region method solves this issue by confining the variable space for calculating the Gauss-Newton direction to be within a local domain called trust region. The algorithm examines the residual at next iterate for every iteration to make sure that the residual is actually reduced. In addition, the trust region size can vary based on the behavior of the previous iteration. This flexibility greatly preserves the speed advantage of the Newton method. Overall, these strategies allow TR to be faster than NCG and stabler than the GR method. More discussion of the three methods can be found in [31].

Experiments show that the full advantage of CVTRC in utilizing a change-of-variable linearization is reduced by the restriction on variations in $\vec{m}$ and the indirect regularization of $\vec{z}$. In particular, the restriction of the variation in $\vec{m}$ in the trust region setup limits the convergence speed, and the regularization over $\vec{m} \odot \vec{z}$ instead of $\vec{z}$ allows room for boundary artifacts. It would be advantageous if these restrictions concerning the trust region setup and the regularization can be removed or replaced. However, this would be not trivial to accomplish, since simply removing these restrictions will cause the method to be unstable. With the current approach, CVTRC works better than OTRC only when initialization of $\vec{z}$ is good.

Although the two methods can handle a wide range of objects for joint reconstruction, there are still some limitations in their capabilities. First, there seems to be a "workable" range for each parameter map within which the convergence can be guaranteed [24]. Currently, this range is estimated empirically. For example, the largest R_2^* decay rate should be comparable to the readout time, and the spin density dynamic range should be limited. Otherwise, artifacts would increase in regions of low spin density. A deeper understanding of the off-resonance frequency range is still needed. However, we emphasize that these ranges are present for all reconstruction methods and are dependent on the number of k-space samples. Second, parameter tuning including the regularization parameters and the continuation related parameters are demanding for the two methods. For most objects, these parameters do not need to be changed much. However, we have observed that a small variation of the parameters may help reduce the boundary artifacts. In addition, the continuation scheme must be changed when the initialization quality changes to maintain the time efficiency of the two methods. This is challenging, since various initializations are possible in a given application. We are currently working on an auxiliary variable method that provides a good initialization to the reconstruction very efficiently [44]. This method is very promising as a way to replace the continuation method used in the current methods. Third, although the two trust region methods are faster than the commonly used NCG method, they are still time-consuming under the metric for clinically applicable methods. A C++ implementation instead of the current MATLAB implementation might achieve a speed gain of 10-50. Some fast computational methods have also been proposed, such as using nonuniform FFTs and time segmentation [38] or using Toeplitz-based fast matrix multiplication [45]. These methods can be easily applied to the two trust region methods, where the acceleration could be several fold. Overall, the two proposed trust region methods take some vital steps toward improving clinical MRI methods for fast joint estimation of the spin density, R_2^* decay and off-resonance frequency maps.

References

[1] M. Barth, A. Metzler, M. Klarhöfer, S. Röll, E. Moser, and D. Leibfritz, "Functional MRI of the human motor cortex using single-shot, multiple gradient-echo spiral imaging," *Magnetic resonance imaging*, vol. 17, no. 9, pp. 1239–1243, 1999.

[2] A. B. A. Wennerberg, T. Jonsson, H. Forssberg, and T.-Q. Li, "A comparative fMRI study: T2*-weighted imaging versus R2* mapping," *NMR in Biomedicine*, vol. 14, no. 1, pp. 41–47, 2001.

[3] J. C. Wood, C. Enriquez, N. Ghugre, J. M. Tyzka, S. Carson, M. D. Nelson, and T. D. Coates, "MRI R2 and R2* mapping accurately estimates hepatic iron concentration in transfusion-dependent thalassemia and sickle cell disease patients," *Blood*, vol. 106, no. 4, pp. 1460–1465, 2005.

[4] L. Anderson, S. Holden, B. Davis, E. Prescott, C. Charrier, N. Bunce, D. Firmin, B. Wonke, J. Porter, J. Walker *et al.*, "Cardiovascular T2-star (T2*) magnetic resonance for the early diagnosis of myocardial iron overload," *European Heart Journal*, vol. 22, no. 23, pp. 2171–2179, 2001.

[5] T. C. Mamisch, T. Hughes, T. J. Mosher, C. Mueller, S. Trattnig, C. Boesch, and G. H. Welsch, "T2 star relaxation times for assessment of articular cartilage at 3 T: a feasibility study," *Skeletal radiology*, vol. 41, no. 3, pp. 287–292, 2012.

[6] A. M. Ellingson, H. Mehta, D. W. Polly, J. Ellermann, and D. J. Nuckley, "Disc degeneration assessed by quantitative T2*(T2 star) correlated with functional lumbar mechanics," *Spine*, vol. 38, no. 24, pp. E1533–40, 2013.

[7] Y. P. Du, Z. Jin, Y. Hu, and J. Tanabe, "Multi-echo acquisition of MR angiography and venography of the brain at 3 Tesla," *Journal of Magnetic Resonance Imaging*, vol. 30, no. 2, pp. 449–454, 2009.

[8] V. Rieke and K. Butts Pauly, "MR thermometry," *Journal of Magnetic Resonance Imaging*, vol. 27, no. 2, pp. 376–390, 2008.

[9] A. Watanabe, L. M. Benneker, C. Boesch, T. Watanabe, T. Obata, and S. E. Anderson, "Classification of intervertebral disk degeneration with axial T2 mapping," *American Journal of Roentgenology*, vol. 189, no. 4, pp. 936–942, 2007.

[10] X. Li, C. B. Ma, T. M. Link, D.-D. Castillo, G. Blumenkrantz, J. Lozano, J. Carballido-Gamio, M. Ries, and S. Majumdar, "In vivo T 1ρ and T 2 mapping of articular cartilage in osteoarthritis of the knee using 3T MRI," *Osteoarthritis and Cartilage*, vol. 15, no. 7, pp. 789–797, 2007.

[11] C. Schwarzbauer, J. Syha, and A. Haase, "Quantification of regional blood volumes by rapid T1 mapping," *Magnetic resonance in medicine*, vol. 29, no. 5, pp. 709–712, 1993.

[12] K. P. Whittall, A. L. Mackay, D. A. Graeb, R. A. Nugent, D. K. Li, and D. W. Paty, "In vivo measurement of T2 distributions and water contents in normal human brain," *Magnetic Resonance in Medicine*, vol. 37, no. 1, pp. 34–43, 1997.

[13] K. P. Whittall, A. L. MacKay, and D. K. Li, "Are mono-exponential fits to a few echoes sufficient to determine T2 relaxation for in vivo human brain?" *Magnetic resonance in medicine*, vol. 41, no. 6, pp. 1255–1257, 1999.

[14] S. C. Deoni, B. K. Rutt, and T. M. Peters, "Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state," *Magnetic Resonance in Medicine*, vol. 49, no. 3, pp. 515–526, 2003.

[15] T. J. Sumpf, M. Uecker, S. Boretius, and J. Frahm, "Model-based nonlinear inverse reconstruction for T2 mapping using highly undersampled spin-echo MRI," *Journal of Magnetic Resonance Imaging*, vol. 34, no. 2, pp. 420–428, 2011.

[16] K. Block, M. Uecker, and J. Frahm, "Model-based iterative reconstruction for radial fast spin-echo MRI," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 11, pp. 1759–1769, Nov 2009.

[17] T. Knopp, H. Eggers, H. Dahnke, J. Prestin, and J. Senegas, "Iterative off-resonance and signal decay estimation and correction for multi-echo MRI," *Medical Imaging, IEEE Transactions on*, vol. 28, no. 3, pp. 394–404, 2009.

[18] C. Huang, C. G. Graff, E. W. Clarkson, A. Bilgin, and M. I. Altbach, "T2 mapping from highly undersampled data by reconstruction of principal component coefficient maps using compressed sensing," *Magnetic Resonance in Medicine*, vol. 67, no. 5, pp. 1355–1366, 2012.

[19] M. Doneva, P. Börnert, H. Eggers, C. Stehning, J. Sénégas, and A. Mertins, "Compressed sensing reconstruction for magnetic resonance parameter mapping," *Magnetic Resonance in Medicine*, vol. 64, no. 4, pp. 1114–1120, 2010.

[20] J. V. Velikina, A. L. Alexander, and A. Samsonov, "Accelerating MR parameter mapping using sparsity-promoting regularization in parametric dimension," *Magnetic Resonance in Medicine*, vol. 70, no. 5, pp. 1263–1273, 2013.

[21] A. Majumdar and R. K. Ward, "Joint reconstruction of multiecho MR images using correlated sparsity," *Magnetic resonance imaging*, vol. 29, no. 7, pp. 899–906, 2011.

[22] B. Zhao, F. Lam, W. Lu, and Z.-P. Liang, "Model-based MR parameter mapping with sparsity constraint," in *Biomedical Imaging (ISBI), 2013 IEEE 10th International Symposium on*. IEEE, 2013, pp. 1–4.

[23] D. B. Twieg, "Parsing local signal evolution directly from a single-shot MRI signal: A new approach for fMRI," *Magnetic Resonance in Medicine*, vol. 50, no. 5, pp. 1043–1052, 2003.

[24] D. B. Twieg and S. J. Reeves, "Basic properties of SS-PARSE parameter estimates," *Medical Imaging, IEEE Transactions on*, vol. 29, no. 5, pp. 1156–1172, 2010.

[25] B. P. Sutton, S. J. Peltier, J. A. Fessler, and D. C. Noll, "Simultaneous estimation of I_0, R_2^*, and field map using a multi-echo spiral acquisition." Proc. Intl. Soc. Mag. Reson. Med., Oct 2002, p. 1323.

[26] V. T. Olafsson, D. C. Noll, and J. A. Fessler, "Fast joint reconstruction of dynamic R_2^* and field maps in functional MRI," *Medical Imaging, IEEE Transactions on*, vol. 27, no. 9, pp. 1177–1188, Sept 2008.

[27] T. Steihaug, "The conjugate gradient method and trust regions in large scale optimization," *SIAM Journal on Numerical Analysis*, vol. 20, no. 3, pp. 626–637, 1983.

[28] M. Powell, *Nonlinear optimization, 1981*, ser. NATO conference series: Systems science. Academic Press in cooperation with NATO Scientific Affairs Division, 1982.

[29] T. Coleman and Y. Li, "An interior trust region approach for nonlinear minimization subject to bounds," *SIAM Journal on Optimization*, vol. 6, no. 2, pp. 418–445, 1996.

[30] C. Hu and S. Reeves, "On the characterization and reduction of ill-conditioning in jointly estimating proton density, T_2^* decay and the field map," in *International Symposium on Biomedical Imaging*. http://hdl.handle.net/11200/44672, 2014.

[31] P. E. Gill, W. Murray, and M. H. Wright, *Practical optimization*. Academic press London, 1981, vol. 5.

[32] W. Tang, *Reconstruction of Parametric Image Maps in Single- and Multiple-Coil Functional Magnetic Resonance Imaging*, 2009.

[33] E. Wasserstrom, "Numerical solutions by the continuation method," *SIAM Review*, vol. 15, no. 1, pp. 89–119, 1973.

[34] K.-M. Ng, "A continuation approach for solving nonlinear optimization problems with discrete variables," Ph.D. dissertation, stanford university, 2002.

[35] E. T. Hale, W. Yin, and Y. Zhang, "A fixed-point continuation method for l1-regularized minimization with applications to compressed sensing," *CAAM TR07-07, Rice University*, 2007.

[36] Y. Zhang, H. P. Hetherington, E. M. Stokely, G. F. Mason, and D. B. Twieg, "A novel k-space trajectory measurement technique," *Magnetic Resonance in Medicine*, vol. 39, no. 6, pp. 999–1004, 1998.

[37] S. Lee, D. Noll, and J. Fessler, "EXTended Rosette ACquisition technique (EXTRACT): A dynamic R2* mapping method using extended rosette trajectory," in *ISMRM Tenth Scientific Meeting*, 2004, p. 2128.

[38] B. P. Sutton, D. C. Noll, and J. A. Fessler, "Fast, iterative image reconstruction for MRI in the presence of field inhomogeneities," *Medical Imaging, IEEE Transactions on*, vol. 22, no. 2, pp. 178–188, 2003.

[39] B. P. Sutton, D. C. Noll, and J. A. Fessler, "Dynamic field map estimation using a spiral-in/spiral-out acquisition," *Magnetic Resonance in Medicine*, vol. 51, no. 6, pp. 1194–1204, 2004.

[40] A. K. Funai, J. A. Fessler, D. Yeo, V. T. Olafsson, and D. C. Noll, "Regularized field map estimation in MRI," *Medical Imaging, IEEE Transactions on*, vol. 27, no. 10, pp. 1484–1494, 2008.

[41] M. Hanke, *Conjugate gradient type methods for ill-posed problems*. CRC Press, 1995, vol. 327.

[42] W. W. Hager and H. Zhang, "A survey of nonlinear conjugate gradient methods," *Pacific journal of Optimization*, vol. 2, no. 1, pp. 35–58, 2006.

[43] S. G. Nash, "Truncated-newton methods," DTIC Document, Tech. Rep., 1982.

[44] C. Hu and S. Reeves, "An efficient auxiliary variable method for quantification of spin density, R2* decay and field inhomogeneity maps in magnetic resonance imaging." IEEE symposium on biomedical imaging (ISBI), 2015. [Online]. Available: http://hdl.handle.net/11200/48505

[45] J. A. Fessler, S. Lee, V. T. Olafsson, H. R. Shi, and D. C. Noll, "Toeplitz-based iterative image reconstruction for MRI with correction for magnetic field inhomogeneity," *Signal Processing, IEEE Transactions on*, vol. 53, no. 9, pp. 3393–3402, 2005.