THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

A Compressed Sensing Framework for Magnetic Resonance Fingerprinting

Citation for published version:
Davies, M, Puy, G, Vandergheynst, P & Wiaux, Y 2014, 'A Compressed Sensing Framework for Magnetic Resonance Fingerprinting' Siam journal on imaging sciences, vol 7, no. 4, pp. 2623.

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Author final version (often known as postprint)

Published In:
Siam journal on imaging sciences

OPEN ACCESS

A COMPRESSED SENSING FRAMEWORK FOR MAGNETIC RESONANCE FINGERPRINTING

MIKE DAVIES, GILLES PUY, PIERRE VANDERGHEYNST AND YVES WIAUX

Abstract. Inspired by the recently proposed Magnetic Resonance Fingerprinting (MRF) technique, we develop a principled compressed sensing framework for quantitative MRI. The three key components are: a random pulse excitation sequence following the MRF technique; a random EPI subsampling strategy and an iterative projection algorithm that imposes consistency with the Bloch equations. We show that theoretically, as long as the excitation sequence possesses an appropriate form of persistent excitation, we are able to accurately recover the proton density, T1, T2 and off-resonance maps simultaneously from a limited number of samples. These results are further supported through extensive simulations using a brain phantom.

Key words. Compressed sensing, MRI, Bloch equations, manifolds, Johnson-Linderstrauss embedding

1. Introduction. While traditional Magnetic Resonance Imaging (MRI) has had great success as a medical research and diagnostic tool, the images that are created are essentially qualitative. Quantitative MR imaging aims to provide additional physiological information by estimating the spatial variation of one or more of the physical parameters that control the imaging process. This can help in the discrimination of different tissue types, provide further valuable diagnostic information and is of use in numerous application areas, such as diffusion and perfusion imaging. However, quantitative imaging can be prohibitively slow and therefore there is a major challenge to design imaging protocols that can estimate the quantitative information within a reasonable time, and with an acceptable signal-to-noise ratio (SNR) and resolution.

A number of researchers have attempted to address, this problem taking a combination of model-based and compressed sensing (CS) approaches [6, 17, 39, 37, 25, 29] with some success. Such techniques accelerate the parameter acquisition by partially sampling k-space and then using model-based reconstruction algorithms that constrain the solution to fit some quantitative signal model. However, while motivated by the theory of CS, these methods are essentially heuristic and little work has been done to explore acquisition protocols that provide a suitable fit from a CS perspective.

Most recently, a radically new procedure for quantitative imaging called Magnetic Resonance Fingerprinting (MRF) [29] has been proposed. MRF uses a combination of random excitation pulse sequences and partial k-space sampling to simultaneously acquire all relevant quantitative information at once. While MRF was also inspired by the recent growth of CS techniques in MRI [29], the exact link to CS was not made explicit and the paper does not consider a full CS formulation. Indeed the role of sparsity, random excitation and sampling are not clarified. The goal of this paper is to make the links with CS explicit, shed light on the appropriate acquisition and reconstruction procedures and hence to develop a full compressed sensing strategy for quantitative MRI.

In particular, we identify separate roles for the pulse excitation and the subsampling of k-space. We identify the Bloch response manifold as the appropriate low dimensional signal model on which the CS acquisition is performed, and interpret the model-based dictionary of [29] as a natural discretization of this response manifold. We also discuss what is necessary in order to have an appropriate CS-type acquisition scheme.

Having identified the underlying signal model we next turn to the reconstruction process. In [29] this was performed through pattern matching using a matched filter based on the model-based dictionary. However, this does not offer the opportunity for exact reconstruction, even if the signal is hypothesised to be 1-sparse in this dictionary due to the undersampling of k-space. This suggests that we should look to a model-based CS framework that directly supports such manifold models [4]. Recent algorithmic work in this direction has been presented by Iwen and Maggioni [26], however, their approach is not practical in the present context as the computational cost of their scheme grows exponentially with the dimension of the manifold. Instead, we leverage recent results from [11] and develop a recovery algorithm based on the Projected Landweber Algorithm (PLA). This method also has the appealing interpretation of an iterated refinement of the original MRF scheme.

The remainder of the paper is set out as follows. We begin by giving a brief overview of MRI acquisition. Then we discuss the challenges of quantitative imaging in MRI and review the recently proposed MRF scheme [29]. We next develop the detailed mathematical model associated with MRF acquisition which leads us to the voxel-wise Bloch response manifold model observed through a sequence of partially sampled k-space measurements. In §4, using the MRF acquisition model, we set out a framework for a compressed sensing solution to the quantitative MRI problem followed by a simple extension that provides a degree of spatial regularization.

In the simulation section we demonstrate the efficacy of our methods on an anatomical brain phantom [15], available at the BrainWeb repository [12]. Our results show that our CS method offers substantial gains in reconstruction accuracy over the original MRF matched filter scheme [29]. We also demonstrate the efficiency of the proposed algorithm in terms of speed of convergence and the empirical trade-off between undersampling in k-space and excitation sequence length.

Finally, we summarize what we have learnt by placing the MRF procedure within a CS framework and highlight a number of open questions and research challenges.

2. Magnetic Resonance Imaging Principles. MRI, with its ability to image soft tissue, provides a very powerful imaging tool for medicine. The basic principles of MRI lie in the interaction of proton spins with applied magnetic fields. While a full review of these principles is beyond the scope of this paper, following [38], we now introduce the basics required to understand the motivation for the proposed acquisition scheme and the subsequent mathematical models. For a more detailed treatment of MRI from a signal processing perspective we refer the reader to one of the excellent reviews on the subject, such as [38, 20].

2.1. Bloch Equations. The main source of the measured signal in MRI comes from the magnetic moments of the proton spins. In a single volume element (voxel) the net magnetization $\mathbf{m} = (m^x, m^y, m^z)^T$ is the vector sum of all the individual dipole moments within the voxel. If there is no magnetic field then at equilibrium the net magnetization is zero.

If a static magnetic field, $\mathbf{B}_0$ (usually considered to lie in the $[0, 0, 1]^T$ direction), is then applied the spins align with this field and the net magnetization at equilibrium, $\mathbf{m}_{\mathrm{eq}}$, is proportional to the proton density ρ within the volume. However, equilibrium is not achieved immediately after the field is applied, but is controlled by the longitudinal relaxation time, T1, such that the net magnetization at time t is given by: $m^z(t) = m_{\mathrm{eq}}(1 - \exp(-t/\mathrm{T1}))$.

If there is magnetization in the plane orthogonal to $\mathbf{B}_0$ then the magnetization, $\{m^x, m^y\}$, precesses about the z axis at a frequency called the Lamor frequency,

2

$\omega_L = \gamma|\mathbf{B}_0|$ (approximately 42.6MHz per Tesla), where the quantity γ is called the gyromagnetic ratio. This in turn emits an electromagnetic signal which is the signal that is measured. As the individual dipoles dephase the net transverse magnetization decays exponentially at a rate T2, called the transverse relaxation time.

In MRI the magnetic field is composed of a static magnetic field and a dynamic component which is manipulated through a radio frequency (RF) coil aligned with the x direction. When a transverse magnetic field is applied via an RF pulse the proton dipoles rotate about the applied magnetic field. The overall macroscopic dynamics of the net magnetization can be summarized by a set of linear differential equations called the Bloch equations [38]:

$$(2.1) \qquad \frac{\partial \mathbf{m}(t)}{\partial t} = \mathbf{m}(t) \times \gamma \mathbf{B}(t) - \begin{pmatrix} m^x(t)/\,\mathrm{T2} \\ m^y(t)/\,\mathrm{T2} \\ (m^z - m_{\mathrm{eq}})/\,\mathrm{T1} \end{pmatrix}$$

The response at a given readout time (TE) from an initial RF pulse can be determined by integrating these equations over time. When a specific sequence of pulses is applied (assuming pulse length $\ll \mathrm{T1}, \mathrm{T2}$) then the dynamics of the magnetization from pulse to pulse or readout to readout can be described simply by a three dimensional discrete time linear dynamical system [27].

2.2. Spatial Encoding and Image Formation. In order to produce an image it is necessary to spatially encode the magnetization in the received signal. This is done through the application of various magnetic gradients. First, a slice can be selected through the application of a magnetic gradient along the z direction, while appropriately restricting the frequency band of the excitation pulses. The gradient changes the Larmor frequency as a function of z, and only those positions that are excited by the pulses generate a magnetization in the transverse plane.

In order to encode the transverse magnetization spatially at the acquisition time (called the echo time (TE)) the magnetic field can be modified further to have gradients G_x and G_y in the x and y directions. For example, if a linear gradient is applied along the x direction so that $B^z = (B_0 + G_x x)$, then the spatial variation of the transverse magnetization is encoded in the Larmor frequency and hence in the frequency of the received signal (it is usually assumed that the duration of the signal read out time is sufficiently short such that the magnetization can be treated as stationary). The received signal therefore corresponds to a line in the spatial Fourier transform, known as k-space, of the transverse magnetization. By careful selection of G_x and G_y it is possible to sample different lines of k-space until it is adequately sampled. The most popular technique is to take measurements, which we denote $\mathbf{y}$, that sample k-space on a Cartesian grid, so that the image can be formed by the application of the inverse 2D Discrete Fourier transform (DFT), F. Thus we can generate a discrete image $\mathbf{x}$ (represented here in vector form) by using $\mathbf{x} = F^H \mathbf{y}$, where H denotes the conjugate transpose. For simplicity, unless stated otherwise, we will work with this discrete representation and assume that samples in k-space have been taken on the Cartesian grid.

2.3. Rapid Imaging. A key challenge in MRI is acquiring the signals in a reasonably short time. Long scan times are costly, unpopular with patients and can introduce additional complications such as motion artefacts. However, the set up described so far for MRI requires the application of repeated excitation pulses and gradients to acquire the multiple lines of k-space. Furthermore, after each acquisition sufficient time must be left for the magnetization to achieve equilibrium once again.

One way to accelerate the imaging is to acquire more samples from k-space per acquisition. By varying the transverse gradients G_x and G_y as a function of time it is possible to generate more sophisticated sampling patterns. For example, in echo-planar imaging (EPI) [30] multiple lines of k-space are acquired at each pulse. Another strategy is to generate spiral trajectories in k-space. However, in both cases as the readout time gets longer artefacts are introduced by variation in the transverse magnetization over the read out time. Furthermore, in the case of spiral and other non-Cartesian trajectories there is the added complication of requiring more complicated image formation algorithms, such as gridding techniques [13], that attempt to approximate the pseudo-inverse of the non-uniform Fourier transform [19].

A second approach to rapid imaging is to take fewer samples. Since the emergence of compressed sensing in MRI [28], the idea of subsampling k-space has become very popular. Compressed Sensing exploits the fact that the image being acquired can be approximated by a low dimensional model, e.g. sparse in the spatial or wavelet domain. Then, under certain circumstances, the image can be recovered from a subsampled k-space using an appropriate iterative reconstruction algorithm.

Parallel imaging (multiple receive coils) techniques, e.g. [32], provide another means to accelerate imaging. Such techniques are typically complementary with the above strategies but require additional calibration to estimate the coil sensitivity maps. To keep the discussion simple we will concentrate on the single coil scenario here and leave discussion of possible extensions to parallel imaging until the conclusions.

2.4. Quantitative MRI. Rather than simply forming an image that measures the transverse magnetization response from a single excitation, the aim of quantitative imaging is to provide additional physiological information by estimating the spatial variation of one or more of the physical parameters that control the Bloch equations: T1, T2, off-resonance, proton density, and more generally in the Bloch-Torey equations, diffusivity. These can help in the discrimination of different tissue types and provide useful information in numerous application areas, such as diffusion and perfusion imaging.

The standard approach to parameter estimation is to acquire a large sequence of images in such a way that for each voxel the sequence of values is dependent on either T1 and/or T2, as well as certain nuisance parameters. For example, the most common techniques acquire a sequence of images at different echo times from an initial excitation pulse. For T1 estimation, this is typically an inversion recovery pulse (full 180° rotation of the magnetic field) and for T2 it is a spin-echo pulse (90° rotation). The image sequences encode the exponential relaxation and the parameter of interest can be estimated by fitting an exponential curve to each voxel sequence. Another approach [16] uses a set of well tailored steady state sequences, such that each voxel sequence encodes the relevant parameter values. Such techniques require the acquisition of multiple lines for multiple images, and it is very challenging to achieve within a reasonable time and with an acceptable signal-to-noise ratio (SNR) and resolution.

Recently there have been a number of papers attempting to address this problem taking a compressed sensing approach [6, 17, 39, 37, 25]. These techniques accelerate the parameter acquisition using a fit to some quantitative signal model, such as the relaxation exponential, combined with a partial sampling of k-space for each image in the sequence. Model-based optimization algorithms [20] are then used to retrieve the parameter values. However, while such approaches take their inspiration from compressed sensing, these techniques mainly focus on the development of novel re-

construction algorithms, and do not tackle the fundamental issue of how to design the acquisition in such a way as to meet the compressed sensing sampling criteria.

In contrast to this previous body of work, here we will set out a principled compressed sensing approach for the simultaneous determination of all parameters of interest. Like previous model-based techniques [20] our algorithm iteratively minimises the measurement error while imposing consistency with a quantitative signal model. However, unlike previous work, we develop an acquisition framework that can be shown to satisfy the compressed sensing criteria and a parameter estimation algorithm that enjoys exact recovery guarantees as well as robustness to model imperfections and measurement errors. The basis of our acquisition scheme is the recently proposed 'magnetic resonance fingerprinting' technique [29] which we describe next.

3. Magnetic Resonance Fingerprinting. In the recent paper [29] a new type of MRI acquisition scheme is presented that enables the quantification of multiple tissue properties simultaneously through a single acquisition process. The procedure is composed of 4 key ingredients:

1. The material magnetization is excited through a sequence of random RF pulses. There is no need to wait for the signal to return to equilibrium between pulses or for the response to reach a steady state condition as in other techniques.
2. After each pulse the response is recorded through measurements in k-space. Due to the time constraints only a proportion of k-space can be sampled between each pulse. In [29] this is achieved through Variable Density Spiral (VDS) sampling.
3. A sequence of magnetization response images is formed using gridding to approximate the least square solution. These images suffer from significant aliasing due to the high level of undersampling.
4. Parameter maps (proton density, ρ, T1, T2 and off-resonance,[1] δf) are formed through a non-iterative pattern matching algorithm that attempts to match the (alias-distorted) magnetization response sequences per voxel to the response predicted from the Bloch equations. This is distinct from the usual model-based reconstruction algorithms.

Below we develop the relevant mathematical models for the MRF acquisition system that will allow us to develop a full CS strategy for quantitative MRI.

3.1. Pulse excitation and the Bloch response manifold. The MRF process is based upon an Inversion Recovery Steady State Free Precession (IR-SSFP) pulse sequence.[2] The dynamics of the magnetization for each voxel, assuming a single chemical composition, are described by the response of the Bloch equations when 'driven' by the excitation parameters.

Let $i = 1, \ldots, N$ index the voxels of the imaged slice. The MRF excitation generates a magnetization response that can be observed (or at least partially observed) at each excitation pulse. The magnetization at a given voxel at the lth echo time is then a function of the excitation parameters of the lth excitation pulse, the magnetization at the $(l - 1)$th echo time, the overall magnetic field and the unknown parameters associated with the given voxel. The overall dynamics can be described

---

[1] The off-resonance frequency is a additional parameter that can be incorporated into the Bloch equations and measures local field inhomogeneity and chemical shift effects [21].

[2] As the excitation pulses in MRF are random the term steady state is now somewhat of a misnomer and we should possibly call these Inversion Recovery Randomly Excited Free Precession.

by a parametrically excited linear system and are summarized in appendix A.

The magnetization dynamics at voxel i are parameterized by the voxel's parameter set $\theta_i = \{\mathrm{T1}_i, \mathrm{T2}_i, \delta f_i\} \in \mathcal{M}$, where $\mathcal{M} \subset \mathbb{R}^3$ denotes the set of feasible values for θ_i, and the voxel's proton density, ρ_i. The magnetization response dynamics are also characterized by the excitation parameters of the lth pulse, namely the flip angle, α_l, and the repetition time, TR_l.

Now and subsequently we will denote the magnetization image sequence by the matrix X, with $X_{i,l}$ denoting the magnetization for voxel i at the lth read out time. Note we are representing the response image at the lth readout by a column vector which we denote as: $X_{:,l}$, using a Matlab style notation for indexing. Similarly, we will denote the magnetization response sequence for a given voxel i as $X_{i,:}$.

Given the initial magnetic field, the initial magnetization of any voxel is known up to the unknown scaling by its proton density ρ_i. Thus the magnetization response at any voxel can be written as a parametric nonlinear mapping from $\{\rho_i, \theta_i\}$ to the sequence, $X_{i,:}$:

$$(3.1) \qquad\qquad X_{i,:} = \rho_i B(\theta_i; \alpha, \mathrm{TR}) \in \mathbb{C}^{1 \times L}.$$

Here $\rho_i \in \mathbb{R}_+$ is the proton density at voxel i, L is the excitation sequence length and B is a smooth mapping induced by the Bloch equation dynamics: $B : \mathcal{M} \to \mathbb{C}^{1 \times L}$, where its smoothness can be deduced by the smooth dependence of the dynamics (A.1) and (A.3) with respect to θ_i.

In order to be able to retrieve the Bloch parameters θ_i and proton density from $X_{i,:}$ it is necessary that the excitation sequence is "sufficiently rich" such that the voxel's magnetization response (3.1) can be distinguished from a response with different parameters. Mathematically this means that there is an embedding of $\mathbb{R}_+ \times \mathcal{M}$ into $\mathbb{C}^L$.[3] We will call $\mathcal{B} = B(\mathcal{M}; \alpha, \mathrm{TR}) \subset \mathbb{C}^L$ the Bloch response manifold and denote its cone by $\mathbb{R}_+\mathcal{B}$.

REMARK 1. *Note that this component of the MRF procedure is not compressive, as the mapping (3.1) will typically need to map to a higher dimension than $\dim(\mathbb{R}_+\mathcal{M})$ in order to induce an embedding. The primary role of the excitation sequence is therefore to ensure identifiability and this can typically be achieved through random excitation as is commonly used in system identification. We will see, however, that the excitation sequence will also need to induce a sufficiently persistent excitation for it to be observed in a compressive manner.*

REMARK 2. *The aim of a good excitation sequence should be to minimize the time taken to acquire the necessary data rather than minimizing the total number of samples. To this end, the total acquisition time for the sequences, $\sum_l \mathrm{TR}_l$ is the relevant cost. Here, while more samples may be taken in MRF in comparison with other quantitative techniques the benefit comes in not having to wait for the magnetization to relax to its equilibrium state between samples.*

REMARK 3. *While it is clear that the proton density, ρ_i, will necessarily be real valued and non-negative, it is common practice in MRI to allow this quantity to absorb additional phase terms due to, for example, coil sensitivity or timing errors. Therefore ρ_i is often allowed to take a complex value. In this work we will retain the idealized model, treating it as non-negative real, however, we note that the subsequent theory presented here can typically be easily modified to work with $\rho_i \in \mathbb{C}$ instead of $\rho_i \in \mathbb{R}_+$,*

---

[3]Strictly speaking we can only consider this to be an embedding for $\rho_i > 0$ otherwise θ_i is not observable.

albeit with an increase in the dimensionality of the unknown parameter set. We will highlight specific differences along the way.

3.2. MRF imaging and k-space sampling. So far we have considered the signal model for a single voxel. For a complete spatial image, assuming a discretization into N voxels and treating each voxel as independent we have $\theta \in \mathcal{M}^N$ and $\rho \in \mathbb{R}_+^N$. Similary $X \in \mathbb{C}^{N \times L}$. We can therefore define the full response mapping, $X = f(\rho, \theta)$, $f : \mathbb{R}_+^N \times \mathcal{M}^N \to (\mathbb{R}_+ \mathcal{B})^N \subset \mathbb{C}^{N \times L}$, as:

$$(3.2) \qquad X = f(\rho, \theta) = [\rho_1 B(\theta_1; \alpha, \mathrm{TR}), \ldots, \rho_N B(\theta_N; \alpha, \mathrm{TR})]^T.$$

Unfortunately, it is impractical to observe the full spatial magnetization (via k-space) at each repetition time within a sufficiently small time for the magnetization to remain approximately constant. It is therefore necessary to resort to some form of undersampling. Let us denote the observed sequence of k-space samples as $Y \in \mathbb{C}^{M \times L}$, such that the samples taken at the lth read out, $Y_{:,l} \in \mathbb{C}^M$ are given by:

$$(3.3) \qquad Y_{:,l} = P(l) F X_{:,l}$$

where F again denotes the 2D discrete Fourier transform and $P(l)$ is the projection onto a subset of coefficients measured at the lth read out (although the original MRF scheme used a sequence of spiral trajectories, for simplicity we will assume that the Fourier samples are only taken from a Cartesian grid). We can finally define the full linear observation map from the spatial magnetization sequence to the observation sequence as $Y = h(X)$ where h is given by:

$$(3.4) \qquad Y = h(X) = [P(1) F X_{:,1}, \ldots, P(N) F X_{:,N}].$$

Together (3.2) and (3.4) define the full MRF acquisition model from the parameter maps $\mathrm{T1}, \mathrm{T2}, \delta f$ and ρ to the observed data Y.

3.3. MRF matched filter reconstruction. In [29] the image sequence is first reconstructed using the regridding method [13] which approximates the least squares estimate for $X_{i,:}$ given $Y_{i,:}$:

$$(3.5) \qquad \hat{X}_{:,l} = F^H P(t)^T Y_{:,l}$$

or equivalently $\hat{X} = h^H(Y)$. Due to the high level of undersampling, each reconstructed image contains significant aliasing. However, it is argued in [29] that accurate estimates of the parameter maps can still be obtained by matching each voxel sequence to a predicted Bloch response sequence using a set of matched filters. This essentially averages the aliasing across the sequence, treating the aliasing as noise. While the technique provides impressive results, it ignores the main tenet of compressed sensing - that aliasing is interference and under the right circumstances can be completely removed (we explore this idea in detail in §4).

Mathematically, it will be convenient to view the matched filter solution as the projection of the voxel sequence onto a discretization of the Bloch response manifold as follows.

3.3.1. Sampling the Bloch response manifold. Suppose that we wished to approximate the projection of the sequence $X_{i,:}$ onto the cone of the Bloch response manifold. One way to do this is to first take a discrete set of samples of the parameter space, $\mathcal{M}$, $\theta_i^{(k)} = \{\mathrm{T1}_i^{(k)}, \mathrm{T2}_i^{(k)}, \delta f_i^{(k)}\}$, $k = 1, \ldots, P$ and construct a 'dictionary' of

Algorithm 1 MRF reconstruction
> **Given:** Y
> Reconstruct X:
> $\hat{X} = h^H(Y)$
> MF parameter estimation:
> **for** $i = 1 : N$ **do**
> $\hat{k}_i = \text{argmax}_k \, \text{real}\langle D_k, \hat{X}_{i,:}\rangle / \|D_k\|_2$
> $\hat{\theta}_i = \text{LUT}_{\mathcal{B}}(\hat{k}_i)$
> $\hat{\rho}_i = \max\{0, \text{real}\langle D_{\hat{k}_i}, \hat{X}_{i,:}\rangle / \|D_{\hat{k}_i}\|_2^2\}$
> **end for**
> **Return:** $\hat{\theta}, \hat{\rho}$

magnetization responses, $D = \{D_k\}$, $D_k = B(\theta_i^{(k)}; \alpha, \text{TR})$, $k = 1, \ldots, P$. The density of such samples controls the accuracy of the final approximation of the projection operator.

We can similarly construct a look-up table (LUT) to provide an inverse for $B(\theta_i; \alpha, \text{TR})$ on the discrete samples such that $\theta_i^{(k)} = \text{LUT}_B(k)$.

The projection onto the cone of the discretized response manifold, D, can then be calculated using:

$$(3.6) \qquad \hat{k}_i = \underset{k}{\text{argmax}} \, \frac{\text{real}\langle D_k, X_{i,:}\rangle}{\|D_k\|_2}$$

to select the closest sample $D_{\hat{k}_i}$ and

$$(3.7) \qquad \hat{\rho}_i = \max\{\text{real}\langle D_{\hat{k}_i}, X_{i,:}\rangle / \|D_{\hat{k}_i}\|_2^2, 0\}$$

for the proton density, where the real and max operations are necessary to select only positive correlations since negative ρ_i are not admissible.

If we allow ρ_i to be complex valued (see Remark 3) then the projection equations become:

$$(3.8) \qquad \hat{k}_i = \underset{k}{\text{argmax}} \, \frac{|\langle D_k, X_{i,:}\rangle|}{\|D_k\|_2}$$

and

$$(3.9) \qquad \hat{\rho}_i = \langle D_{\hat{k}_i}, X_{i,:}\rangle / \|D_{\hat{k}_i}\|_2^2$$

Equations (3.8) and (3.9) are precisely the matched filter equations used in [29], applied to the distorted voxel sequences. We therefore see that one interpretation of matched filtering with the MRF dictionary model is to provide an approximate projection onto the cone of the Bloch response manifold for each voxel sequence.

A summary of the full MRF parameter map recovery algorithm (with a real valued proton density model) is given in Algorithm 1.

Computational cost and accuracy. Given that the discretized MRF dictionary can be very large ($\approx 500{,}000$ samples in [29]), it is useful to consider the computational complexity of the above calculations as a function of parameter accuracy as this is the major computational bottleneck that we will encounter.

The accuracy with which we can estimate the parameters for a given voxel will depend on the accuracy of the approximate projection operator and the Lipschitz constants of the inverse mapping, $LUT_{\mathcal{B}}$. We can achieve an approximate projection by generating an ϵ-cover of $\mathcal{B}$ with D_k. As the dimension of $\mathcal{B}$ is 3, this requires choosing $P \sim C\epsilon^{-3}$ atoms in our dictionary. Furthermore, as the projection operation described in (3.6) takes the form of a nearest neighbour search, we can use fast nearest neighbour search strategies, such as the cover tree method [5], to quickly solve (3.6) in $\mathcal{O}(L \ln(1/\epsilon))$ computations per voxel, instead of the $\mathcal{O}(L\epsilon^{-3})$ necessary for exhaustive search. This effectively makes the speed of each application of D on a par with that of a traditional fast transform. Similarly, the approximate inverse using $LUT_{\mathcal{B}}$ can also be computed in $\mathcal{O}(\ln(1/\epsilon))$.

We could also consider enhancing such an estimate by exploiting the smoothness of the response manifold, either by using local linear approximations of the manifold [26] or by further locally optimizing the projection numerically around the selected parameter set, once we are assured global convergence. Such an enhancement could allow either for increased accuracy or reduced computation through the use of fewer parameter samples, however, we do not pursue these ideas further here.

4. Compressed Quantitative Imaging. In order to generate a full compressed sensing framework for MRF we will identify sufficient conditions on the excitation pulse sequences and the k-space sampling, along with a suitable reconstruction algorithm, to guarantee recovery of the parameter maps from the observed k-space samples. As the dimension of our problem is large, $\dim((\mathbb{R}_+ \times \mathcal{M})^N) = 4N$, we do not consider the manifold reconstruction algorithms in [26] as these scale poorly with the dimension of the manifold. Instead, we propose a CS solution based around the iterative projection algorithm of Blumensath [11] which we will see has computational cost that is linear in the voxel dimension. Our approach, which we call BLIP (BLoch response recovery via Iterated Projection), has three key ingredients: a random pulse excitation sequence following the original MRF technique; a random subsampling strategy that can be shown to induce a low distortion embedding of $\mathbb{R}_+^N \times \mathcal{M}^N$ and an efficient iterated projection algorithm [11] that imposes consistency with the Bloch equations. Moreover, the projection operation is the same nearest neighbour search described in section 3.3.1.

We first describe the iterative projection method and then consider the implications for the appropriate excitation and sampling strategies.

4.1. Reconstruction by Iterated Projection. In [11] a general reconstruction algorithm, the Projected Landweber Algorithm (PLA) was proposed as an extension of the popular Iterated Hard Thresholding Algorithm [7, 9]. PLA is applicable to arbitrary union of subspace models as long as we have access to a computationally tractable projection operator onto the union of subspace model within the complete signal space. The algorithm is given by:

$$(4.1) \qquad X^{(n+1)} = \mathcal{P}_{\mathcal{A}}(X^{(n)} + \mu h^H(Y - hX^{(n)}))$$

where $\mathcal{P}_{\mathcal{A}}$ is the projection onto the signal model $\mathcal{A}$ (possibly non-unique) such that

$$(4.2) \qquad \mathcal{P}_{\mathcal{A}}(X) \in \operatorname*{argmin}_{\tilde{X} \in \mathcal{A}} \|X - \tilde{X}\|_F$$

and μ is the step size.

The current theory for PLA [11] states that a sufficient condition for stable recovery of X given Y is that h is a stable embedding - a so-called Restricted Isometry

Algorithm 2 BLoch response recovery via Iterative Projection (BLIP)

Given: Y
Initialization: $X^{(0)} = \mathbf{0}$, $\mu = N/M$
Image sequence reconstruction
for $n = 1; n := n + 1$ **until stopping criterion do**
   Gradient step:
   for $l = 1 : L$ **do**
     $X^{(n+1/2)}_{:,l} = X^{(n)}_{:,l} + \mu F^H P(l)^T (Y_{:,l} - P(l) F X^{(n)}_{:,l})$;
   end for
   Projection step:
   for $i = 1 : N$ **do**
     $\hat{k}_i = \mathrm{argmax}_k \, \mathrm{real} \langle D_k, X^{(n+1/2)}_{i,:} \rangle / \|D_k\|_2$
     $\hat{\rho}_i = \max\{0, \mathrm{real} \langle D_{\hat{k}_i}, X^{(n+1/2)}_{i,:} \rangle / \|D_{\hat{k}_i}\|_2^2\}$
     $X^{(n+1)}_{i,:} = \hat{\rho}_i D_{\hat{k}_i}$
   end for
end for
Parameter map estimation:
for $i = 1 : N$ **do**
   $\hat{\theta}_i = \mathrm{LUT}_{\mathcal{B}}(\hat{k}_i)$
end for
Return: $\hat{\theta}, \hat{\rho}$

Property (RIP) or bi-Lipshitz embedding - for the signal model, $\mathcal{A}$. A mapping, h, is said to have the RIP (be a bi-Lipschitz embedding) for the signal model $\mathcal{A}$ if there exists a sufficiently small constant $\delta > 0$ such that:

$$(4.3) \qquad (1 - \delta)\|X - \tilde{X}\|_2^2 \leq \frac{N}{M} \|h(X - \tilde{X})\|_2^2 \leq (1 + \delta)\|X - \tilde{X}\|_2^2$$

for all pairs X and $\tilde{X}$ in $\mathcal{A}$. How to achieve such an embedding will be considered later in section 4.2.

The theory [11] states that it is sufficient that h satisfy the RIP with $\frac{M}{N}(1 + \delta) < 1/\mu < \frac{3M}{2N}(1 - \delta)$ for the guaranteed recovery. If h is essentially 'optimal', e.g. a random ortho-projector, then we should set the step size $\mu \approx N/M$ since in the large system limit $\delta \to 0$.

For our compressed sensing scenario the signal model $\mathcal{A}$ is the product set $(\mathbb{R}_+ \mathcal{B})^N$ or, more precisely, its discrete approximation $(\mathbb{R}_+ D)^N$ and the projection operator $\mathcal{P}_{\mathcal{A}}$ can be realized by separately projecting the individual voxel sequences $X^n_{i,:}$ onto the cone of the Bloch response manifold using the equations (3.6) and (3.7). Although $(\mathbb{R}_+ \mathcal{B})^N$ is not itself a union of subspace model it can easily be extended to $(\mathbb{R}\mathcal{B})^N$, which forms an uncountably infinite union of lines (1D subspaces). In fact, the theory of [11] does not require $\mathcal{A}$ to be a union of subspace [11] and is directly applicable to $\mathcal{A} = (\mathbb{R}_+ \mathcal{B})^N$. We therefore appear to have all the ingredients for a full compressed sensing recovery algorithm. This is summarized in Algorithm 2.

REMARK 4. *Note that the above procedure has separated out the parameter map estimation (by inverting the estimated Bloch responses) and the reconstruction of the magnetization image sequence (via the PLA). Indeed, as long as the partial k-space sampling provides a bi-Lipschitz embedding for all possible magnetization responses*

then the CS component of the imaging is well defined even if the Bloch response is not invertible.

4.1.1. Step size selection. Selection of the correct step size is crucial in order to attain good performance from these iterative projection based algorithms [10, 11]. Note that the original parameter estimation in [29] can be interpreted as an application of a single iteration of PLA with a step size $\mu = 1$ and iterating PLA with this step size tends to only deliver a modest improvement over the matched filter (single iteration). The matched filter also has the effect of underestimating the magnitude of X, and hence also the proton density map, as h tends to shrink vectors uniformly (when it provides a stable embedding).

In contrast, when using the substantially more aggressive step size proposed by the theory we will see that significant improvements are observed in signal recovery and often in a very small number of iterations.

In practice, it is also beneficial to select the step size for PLA adaptively to ensure stability. Following the work on adaptive step size selection for IHT [10] we adopt the following heuristic. We begin each iteration by choosing $\mu = N/M$ as is suggested from the CS theory. Then after calculating a new proposed value for X^{n+1} we calculate the quantity:

$$(4.4) \qquad \omega = \kappa \frac{\|X^{n+1} - X^n\|_2^2}{\|h(X^{n+1} - X^n)\|_2^2}$$

for some $\kappa < 1$. If $\mu > \omega$ we reject this update, shrink the step size, $\mu \mapsto \mu/2$ and calculate a new proposed value for X^{n+1}. As with the Normalized IHT algorithm [10], this form of line search is sufficient to ensure convergence of the solution irrespective of conditions on the measurement operator, and we will use this form of step size selection in all subsequent experiments.

4.2. Strategies for subsampling k-space. We now consider what properties of the excitation response sequences and the k-space sampling pattern will ensure that the sufficient RIP conditions in the PLA theory are satisfied.

First note that, as the signal model treats each voxel as independent, we need to take at least $N \dim(\mathbb{R}_+ \mathcal{M})$ measurements as this is the dimension of our model. Furthermore, since we only take a small number of measurements at each repetition time, we cannot expect to achieve a stable embedding without imposing further constraints on the excitation response. For example, if the embedding was induced in the first few repetition times and all further responses were non-informative we would not have taken sufficient measurements from the informative portion of the response. Therefore we consider responses that somehow spread the information across the repetition times. We will assume that the excitation sequence induces an embedding for the response map (3.2) (here random sequences seem to suffice), and identify additional conditions that enable us to develop a random k-space subsampling strategy with an appropriate RIP condition. Our approach will follow the technique of random sampling as is common in compressed sensing measurement design, along with a pre-conditioning technique that has been used in the Fast Johnson-Lindenstrauss Transform [2] and in spread spectrum compressed sensing [34]. It is also reminiscent of the Rauhut's bounded orthonormal systems [35] and has a similar aim of ensuring that information is sufficiently spread within the measurement domain

The key vectors of interest are those that discriminate between pairs of possible signals within our model, namely the *chords* of $\mathbb{R}_+ \mathcal{B}$, which are the vectors of the

form $u = X_{i,:} - \tilde{X}_{i,:}$ with $X_{i,:}, \tilde{X}_{i,:} \in \mathbb{R}_+\mathcal{B}$ and $X_{i,:} \neq \tilde{X}_{i,:}$. We will quantify the pre-conditioning requirement for the excitation response through the *flatness* of such vectors which we define as follows.

DEFINITION 1. *Let U be a collection of vectors $\{u\}$ in $\mathbb{C}^L$. We denote the* flatness, *λ, of the these vectors by:*

$$(4.5) \qquad \lambda := \max_{u \in U} \frac{\|u\|_\infty}{\|u\|_2}.$$

Note that from standard norm inequalities $L^{-1/2} \leq \lambda \leq 1$.

We will consider the chords of an excitation response to be sufficiently flat up to a log penalty if $\lambda \sim L^{-1/2} \log^\alpha L$ for $U = \{\mathbb{R}_+\mathcal{B} - \mathbb{R}_+\mathcal{B}\}\backslash\{0\}$.

In constructing our measurement function we also note that the signal model contains no spatial structure, and therefore we should expect to have to uniformly sample k-space in order to achieve a sufficient RIP. Note this is in contrast with the variable density sampling strategy proposed by [29] which concentrated samples at the centre of k-space. It turns out that we can achieve this using a remarkably simple random subsampling pattern based on multi-shot Echo-planar Imaging (EPI) [30].

Let $F \in \mathbb{C}^{N \times N}$ denote the 2D discrete Fourier transform (assuming an image size of $\sqrt{N} \times \sqrt{N}$) with $F_{i,:}$, $i = 1, \ldots, N$ denoting the N 2D discrete Fourier basis vectors associated with the spatial frequencies $k_x(i), k_y(i) \in \{0, \ldots, \sqrt{N} - 1\}$. Without loss of generality we assume that the vectors are ordered such that $k_x(i) = (i-1) \mod \sqrt{N}$, and $k_y(i) = \lfloor (i-1)/\sqrt{N} \rfloor$. We can now define a *random Echo-Planar Imaging* measurement operator by $Y_{:,l} = P(\zeta_l)FX_{:,l}$, where ζ_l is a sequence of independent random variables uniformly drawn from $\{0, \ldots, p-1\}$ and $P(\zeta) \in \mathbb{R}^{M \times N}$ is defined as follows:

$$(4.6) \qquad P_{i,j} = \begin{cases} 1 & \text{if } j = i + \sqrt{N}\big(\zeta + (p-1)\lfloor (i-1)/\sqrt{N} \rfloor\big) \\ 0 & \text{otherwise.} \end{cases}$$

where for convenience we have assumed that N is exactly divisible by p so that $M = N/p$ is an integer. In words, we uniformly subsample k_y by a factor of p with random shifts across time in k_y of the set of k-space samples. This is illustrated in figure 4.2.

Random EPI, along with an excitation response with appropriate chord flatness, λ, is then sufficient to provide us with a measurement operator, h, that is a bi-Lipschitz embedding on our signal model. In appendix B we prove the following theorem:

THEOREM 1 (RIP for random EPI). *Given an excitation response cone, $\mathbb{R}_+\mathcal{B}$ of dimension $d_\mathcal{B}$, whose chords have a flatness λ and a random EPI operator $h : (\mathbb{R}_+\mathcal{B})^N \to \mathbb{C}^{M \times L}$, then, with probability at least $1 - \eta$, h is a restricted isometry on $(\mathbb{R}_+\mathcal{B})^N - (\mathbb{R}_+\mathcal{B})^N$ with constant δ as long as:*

$$(4.7) \qquad \lambda^{-2} \geq C\delta^{-2}p^2 d_\mathcal{B} \log(N/\delta\eta)$$

for some constant C independent of $p, N, d_\mathcal{B}, \delta$ and η.

Specifically, if $\lambda = \mathcal{O}(L^{-1/2} \log^\alpha L)$ then we require:

$$(4.8) \qquad L = \mathcal{O}(\delta^{-2}p^2 d_\mathcal{B} \log(N/\delta\eta) \log^\alpha(L))$$

excitation pulses.

Theorem 1 indicates to the practical reader that the sequence length $L \geq Cp^2$ is sufficient for the measurement operator to possess the RIP, where C is a universal
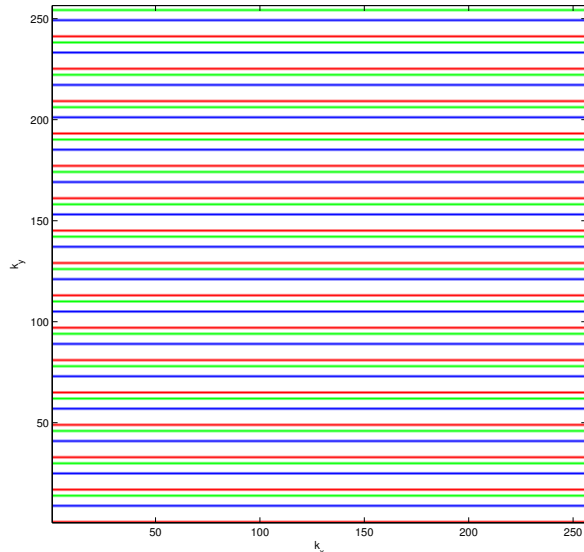
FIG. 4.1. *The plot shows an instance of random EPI k-space sampling for three time frames: red, green and blue respectively. A colored pixel indicate that that (k_x, k_y) frequency is sampled at the associated time frame through the projection operator, $P(\zeta_l)$. In this instance $p = 16$.*

constant. While we might hope to get L of the order of $pd_{\mathcal{B}}$ it appears that this is not possible, at least for a worst case RIP analysis based on the flatness criterion alone. Furthermore, in the experimental section we will provide evidence to suggest that $L \sim p^2$ is indeed the scaling behaviour that we empirically observe. The value of the constant C can be estimated from (B.24), however the obtained value seems too pessimistic. As in the usual compressed sensing scenario, a more realistic value can be estimated by simulation. From the experimental section, we will see that good recovery can be obtained for $L/p^2 \approx 0.3$.

REMARK 5. *It might seem surprising that the proposed scheme uses uniform random sampling in k-space whereas it is usually advisable to use a variable density sampling strategy for compressed sensing solutions for MRI. Indeed, there is good theoretical justification for variable density sampling patterns [1, 33]. Our theory above is not inconsistent with such results. Variable density sampling is advantageous because the underlying signal model - sparsity in the wavelet domain - is not incoherent with the Fourier basis [33, 1]. However, the Fourier basis is incoherent with a voxel-wise signal model as used above. This is not to say that spatial structure cannot be effectively exploited within a compressed quantitative imaging scheme or that variable density sampling would not then be of benefit. However, as the basic MRF based model does not exploit spatial structure we argue that uniform random sampling is appropriate here.*

The challenge of incorporating spatial regularity into the signal model is discussed next.

4.3. Extending the Bloch response model. Our current compressed sensing model takes no account of additional structure within the parameter maps. This

structure could, for example, be the piecewise smoothness of the parameter maps or the magnetization response maps, or an imposed segmentation of the image into different material compositions. In general, it is not clear how such additional regularization can be included in a principled manner, although many heuristic approaches could of course be adopted, as for example in [17]. This is because the parameter values are encoded within the samples of the Bloch response manifold, and therefore the spatial regularity would need to be mapped through the Bloch response leading to a non-separable high dimensional nonlinear signal model.

The one exception, which we consider here, is the regularization of the proton density map, or at least a close relative. We note, however, that in this instance the theory relies on the real non-negative proton density model and does not directly extend to the complex case.

Let us define the *pseudo-density*, $\tilde{\rho}$ as the proton density map scaled by the norm of the Bloch response vector, so that:

$$(4.9) \qquad \tilde{\rho}_i = \rho_i \|B(\theta_i; \alpha, \mathrm{TR})\|_2.$$

Similarly we can define the normalized Bloch response as:

$$(4.10) \qquad \eta_{i,:} = \tilde{B}(\theta_i; \alpha, \mathrm{TR}) \triangleq B(\theta_i; \alpha, \mathrm{TR})/\|B(\theta_i; \alpha, \mathrm{TR})\|_2$$

and the normalized Bloch response manifold, $\tilde{\mathcal{B}}$ as:

$$(4.11) \qquad \tilde{\mathcal{B}} = \left\{ \eta_{i,:} = \tilde{B}(\theta_i; \alpha, \mathrm{TR}) \ \ \text{for some } \theta_i \in \mathcal{M} \right\}$$

The pseudo-density will be roughly the same as the density, as long as the Bloch response sequences are all of approximately the same magnitude. The transform to $\{\tilde{\rho}, \eta\}$ normalizes the manifold $\tilde{\mathcal{B}}$ so that we can more easily calculate projections onto product signal models of the form $\{\tilde{\rho}, \eta\} \in \Sigma \times \tilde{\mathcal{B}}^N$, where Σ denotes the set of spatially regularized pseudo-density maps. To do this we will find the following proposition useful:

PROPOSITION 1. *Given an $X \in \mathbb{C}^{N \times L}$, suppose that the projection onto the signal model $\Sigma \times \tilde{\mathcal{B}}^N$ is given by $\hat{\tilde{\rho}} \in \Sigma$ and $\hat{\eta}_{i,:} \in \tilde{\mathcal{B}}$ and results in $\hat{\tilde{\rho}}_i \geq 0$ for all i, then:*

$$(4.12) \qquad \hat{\eta}_{i,:} = \underset{\eta_{i,:} \in \tilde{\mathcal{B}}}{\operatorname{argmax}} \, z_i$$

and

$$(4.13) \qquad \hat{\tilde{\rho}} = \underset{\tilde{\rho} \in \Sigma}{\operatorname{argmin}} \, \|\tilde{\rho} - z\|_2^2$$

where $z_i = \operatorname{real}\langle \eta_{i,:}, X_{i,:} \rangle$.

Proof. By definition of the orthogonal projection we have:

$$(4.14) \qquad \{\hat{\eta}, \hat{\tilde{\rho}}\} = \underset{\eta, \tilde{\rho}}{\operatorname{argmin}} \sum_i \sum_j |X_{i,j} - \tilde{\rho}_i \eta_{i,j}|^2$$

Expanding (4.14), substituting in z_i and noting that $\|\eta_{i,:}\|_2 = 1$ we have:

$$(4.15) \qquad \{\hat{\eta}, \hat{\tilde{\rho}}\} = \underset{\eta \in \mathcal{B}, \tilde{\rho} \in \Sigma}{\operatorname{argmin}} \sum_i \left(\tilde{\rho}_i^2 - 2\tilde{\rho}_i z_i \right).$$

14

By assumption $\hat{\tilde{\rho}}_i$ is non-negative so the expression is minimized with respect to $\eta_{i,:}$ by (4.12) independently of $\tilde{\rho}_i$. Finally we note that (4.13) holds since:

$$(4.16) \qquad \sum_i \left(\tilde{\rho}_i^2 - 2\tilde{\rho}_i z_i \right) = \|\tilde{\rho} - z\|_2^2 + \text{const.}$$

□

One way to impose spatial regularity on $\tilde{\rho}$ is to force it to be sparse in the wavelet domain for some appropriate orthogonal wavelet representation, W, such that $c = W\tilde{\rho}$. In this case, the projection (4.13) can be written as $\hat{\tilde{\rho}} = W^T \hat{c}$ with:

$$(4.17) \qquad \hat{c} = \mathcal{H}_k(Wz)$$

where $\mathcal{H}_k$ denotes an element-wise hard thresholding [7, 9] that retains only the largest k elements.

Under the non-negativity assumption the projection operator can be formed by applying (4.12) followed by (4.17). This results in a simple algorithm for incorporating a degree of spatial regularization within the compressed quantitative imaging framework. In the next section we will see, however, that the inclusion of this addition spatial constraint adds little to the performance of the compressed sensing approach, suggesting that the Bloch equation constraint dominates the performance.

REMARK 6. *The above calculation is only guaranteed to be valid when the resulting pseudo-density map is non-negative. In theory, applying such an operator when we incur negative values of pseudo-density could give a projection that is not optimal. However, in practice we have found that this is not a problem as, when we assume ρ to be real valued, we always impose non-negativity on both the pseudo-density and the correlations with the Bloch response, z_i, in order to ensure that the projection is physically meaningful.*

5. Experiments. In order to test the efficacy of BLIP for compressed quantitative imaging we performed a set of simulations using an anatomical brain phantom, segmented into various material compositions. This provided a well defined ground truth and enabled us to demonstrate image sequence recovery and parameter map estimation as a function of the k-space subsampling factor and the excitation sequence lengths.

5.1. Experimental Set up. The key ingredients of the experimental set up are described below.

Anatomical Brain Phantom. To develop realistic simulations that also provide a solid ground truth we have adapted the anatomical brain phantom of [15], available at the BrainWeb repository [12]. A 217×181 slice (slice 40) of the crisp segmented anatomical brain was used and restricted to contain only 6 material components, listed in table 5.1. The phantom was further zero padded to make a 256×256 image to simplify the computations. Since we are using the crisp segmentation the model is somewhat idealized and does not address inaccuracies associated with partial volume effects or many of the other issues with real MRI. However, it serves as a useful test bed to provide a good proof-of-concept for our proposed techniques.

The material properties were chosen to be both representative of the correct tissue type [24] and challenging: the proton densities values were chosen to give little discrimination for individual parameters and were set so that there is not an exact match to the sampling of the Bloch response manifold.

The segmented brain is shown, colored by index, in figure 5.1.

TABLE 5.1
Tissue types used from MNI segmented brain phantom

Tissue	index	proton density	T1 (ms)	T2 (ms)
Background	0	0	-	-
CSF	1	100	5012	512
Grey matter	2	100	1545	83
White matter	3	80	811	77
Adipose	4	80	530	77
Skin/Muscle	5/6	80	1425	41
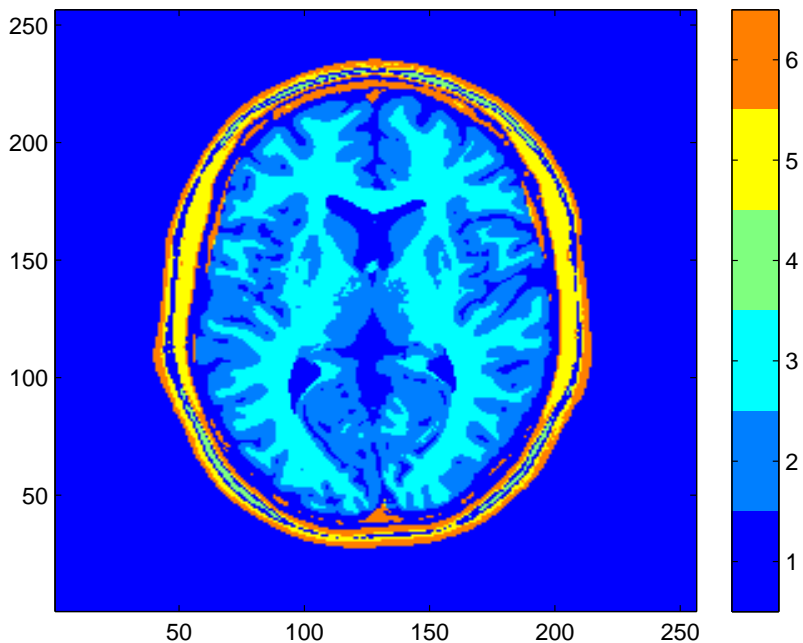
FIG. 5.1. *The MNI segmented anatomical brain phantom [15] colored by index: 0 = Background, 1 = CSF, 2 = Grey Matter, 3 = White Matter, 4 = Fat, 5 = Muscle/Skin, 6 = Skin.*

Pulse excitation. For the excitation sequences we use IR-SSFP sequences (exemplar code can be found in the supplementary material of [29]) with random flip angles drawn from an independent and identically distributed Gaussian distribution:

$$\alpha_l \sim \mathcal{N}(0, \sigma_\alpha^2) \tag{5.1}$$

with a standard deviation, $\sigma_\alpha = 10$ degrees. The repetition times were uniformly spaced at an interval of 10 ms. While we also experimented with randomizing repetition times, we did not find that these significantly changed the performance of the techniques. Constant repetition time intervals also mean that we can directly assess the imaging speed in terms of the sequence length, L.

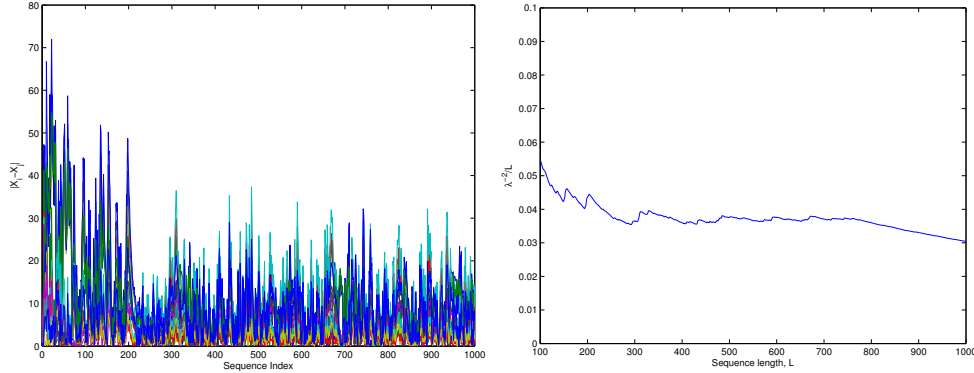The value of σ_α was chosen empirically to provide reasonable persistence of exci-

FIG. 5.2. *Left: examples of the response differences for pairs of tissue types given in table 5.1 when using IR-SSFP pulse sequence excitation with random flip angles. Right: λ^{-2}/L as a function of sequence length for the repsonse differences plotted on the left. From this plot it can be deduced that λ^{-2} grows roughly proportionally to L.*

tation for the expected T1 and T2 responses. Figure 5.2 (left) shows the magnitude of the response differences for the set of tissue types listed in table 5.1. It can be seen that the difference in the responses does indeed persist over time. Using these differences we can also estimate their flatness. Figure 5.2 (right) shows how the flatness varies as a function of sequence length. We see that λ^{-2} roughly scales proportionally to L, as desired, with a slight downward sublinear trend.

Discretized Bloch response. The Bloch response manifold was sampled in a similar manner to [29], however, to simplify things we have only considered variation in T1 and T2 here, assuming the off resonance frequency is equal to zero. Similar to [29], discrete samples for T1 were selected to go between 100 and 2000 in increments of 20 and from 2300 to 6000 in increments of 300. T2 was sampled between 20 and 100 in increments of 5, from 110 to 200 in increments of 20 and from 400 to 1000 in increments of 200. This results in a dictionary of size $3379 \times L$. This range of T1 and T2 values clearly spans the anticipated range for the tissue types listed in table 5.1.

Subsampling strategy. For the k-space subsampling we use the random EPI sampling scheme detailed in section 4.2. Specifically, we fully sample the k-space in the k_x direction while regularly subsampling the k_y direction by a factor of p. This deterministic sampling pattern was then cyclically shifted by a random number of k_y lines at each repetition time. In most experiments p is set to 16 (sampling at 6.25% of Nyquist).

5.1.1. Reconstruction algorithms. In the experiments below we compare three distinct algorithms for reconstructing the magnetization image sequences. These are: (1) the original MRF algorithm; (2) BLIP algorithm presented in Algorithm 2; and (3) BLIP with spatial regularization as detailed in section 4.3. For both iterative algorithms we use the adaptive step size strategy set out in section 4.1.1 with $\kappa = 0.99$. For the spatial regularization we use a Haar wavelet representation with hard thresholding as detailed in section 4.3, retaining only the largest 12000 wavelet coefficients at each iteration.

As the MRF algorithm (with step size equal to 1) underestimates the value of the image sequence (and also the proton density) we include in the appropriate plots the performance of a rescaled MRF algorithm where the step size is $\mu = N/M$.

Finally, in some of the plots we also include the performance for an oracle estimator. This oracle is given the fully sampled image sequence data as an input and then projects each voxel sequence onto the discretized Bloch response. In this way we can differentiate between errors associated with the Bloch response discretization and the image sequence reconstruction.

5.2. Results. All the experiments were evaluated using a signal-to-error-ratio (SER) in decibels (dBs), calculated as $20 \log_{10} \frac{\|x\|_2}{\|x-\hat{x}\|_2}$ for a target signal x with the estimate $\hat{x}$. For T1 and T2 this corresponds to the measures T_1NR and T_2NR that has been used to gauge the efficiency of relaxation time acquisition schemes [16]. To avoid issues of estimates associated with empty voxels the errors are only calculated over regions with a non-zero proton density value.

In all experiments, unless stated otherwise, the following parameters were used: the undersampling ratio for the operator $h(\cdot)$ was fixed at 1/16 and for both the iterative algorithms a maximum of 20 iterations was allowed, though in many cases fewer iterations would have sufficed.

5.2.1. Performance as a function of excitation sequence length. Our first experiment evaluates the performance of the algorithms in terms of the sequence length, which was varied between 10 and 1000 pulses. Here we can separately evaluate the performance of the compressed sensing component and the recovery of the parameter maps.

The compressed sensing recovery performance, evaluated by the SER of the image sequence reconstruction, X, is shown in figure 5.3 (a).

First, note that the strange behaviour of the oracle estimator for small sequence lengths is probably due to the failure of $f(\cdot)$ to achieve a low distortion embedding. This would result in it being easier to approximate voxel sequences with a given element of the Bloch response approximation. Beyond this the performance reaches a plateau at approximately SER = 27 dB which can be considered to be the error associated with the discretization of the Bloch response.

The performance of both BLIP algorithms is roughly equivalent. They both sharply increase in performance at a sequence length of 100 and then tend to a plateau beyond this with an SER of about 0.5 dB below that of the oracle estimator. This suggests that we can achieve near perfect compressed sensing reconstruction with a sequence containing as few as 100 pulses. In this simulation there was no significant gain from the additional inclusion of the spatial regularization.

The performance of MRF is significantly worse. We first highlight that the non-rescaled MRF performance is terrible, however, as noted earlier, this is mainly due to the shrinkage effect of the subsampling operator, $h(\cdot)$. Correcting for this with appropriate rescaling leads to significantly improved estimation. However, we see that the SER increases slowly as a function of sequence length, which is consistent with the argument that the matched filter is averaging over the aliasing rather than cancelling it, as presented in section 3.3. Furthermore, even for a sequence length of 1000 the SER still only reaches 12dB.

Subfigures 5.3 (b), (c) and (d) show the SER for the estimation of the parameter maps, proton density, T1 and T2 respectively, and reflects the combined performance of inverting both $h(\cdot)$ and $f(\cdot)$. In each case the two iterative algorithms approach the oracle performance for sequence lengths of $L \geq 200$, indicating successful parameter map recovery. Furthermore, the performance for the ρ estimates and T2 estimates do not improve substantially beyond the $L = 200$ value as L is increased reaching a
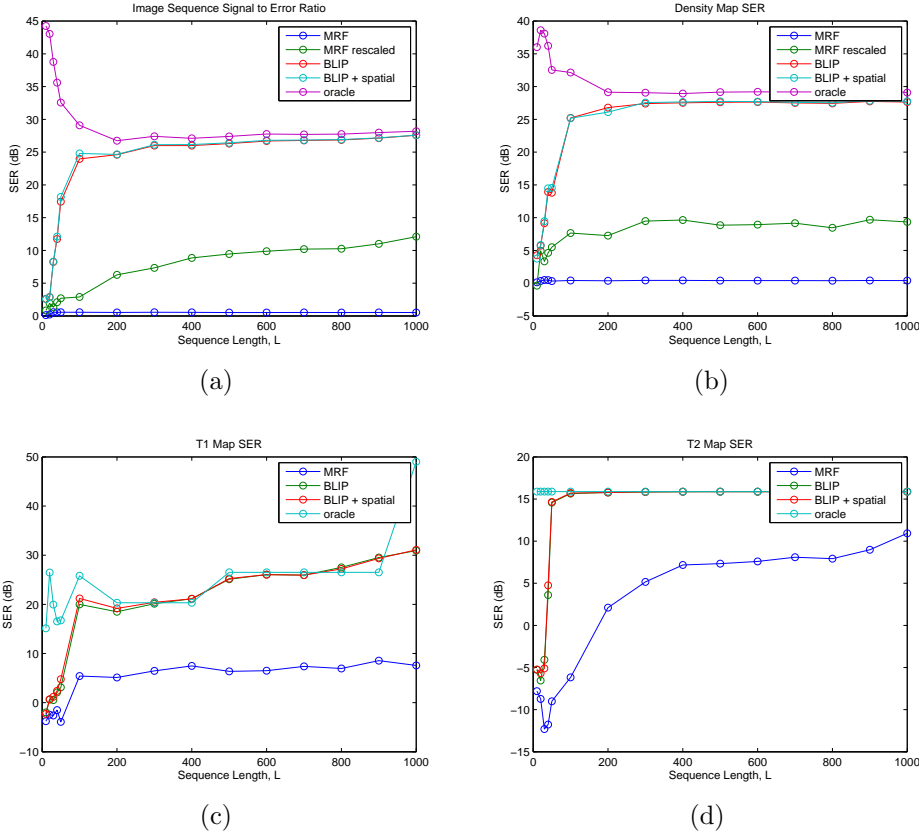
Fig. 5.3. *Reconstruction performance as a function of sequence length. (a) SER for image sequence reconstruction; (b) SER for density map estimation; (c) SER for* T1 *map estimation; and (d) SER for* T2 *map estimation. Results are shown for the following algorithms: MRF, BLIP, BLIP with spatial regularization. Also shown is the performance of an oracle estimator given the full image sequence data. Finally subfigures (a) and (b) also include the performance of a rescaled MRF estimator.*

plateau at approximately 16dB which corresponds to a root mean squared (rms) error of approximately 30ms. In contrast, the T1 estimation performance does increase from roughly 20dB (213ms rms error) at $L = 200$ to 30dB (67ms rms error) at $L = 1000$. This may be a function of the isometry properties (in the T1 direction) for the Bloch response embedding, and is possibly related to the longer time constants of T1. It is an open question as to whether a better excitation sequence can be designed to improve the T1 estimates for small L.

5.2.2. Visual Comparison. To get a visual indication of the performance of the BLIP approach over the MRF reconstruction at low sequence lengths, images of the 3 different parameter estimates for $L = 300$ are given in figures 5.4, 5.5 and 5.6. The left hand column shows the ground truth parameter maps while the middle row shows the MRF reconstruction (scaled) and the right hand column shows the BLIP estimates (with spatial regularization). While the main aspects of the parameter maps are visible in the MRF reconstructions, there are still substantial aliasing artefacts. These are most prominent in the T1 and T2 estimates. In contrast, the BLIP estimates

are virtually distortion-free, indicating that good spatial parameter estimates can be obtained with as little as 300 excitation pulses.

5.2.3. Convergence rates for BLIP. The convergence of the iterative algorithms is shown in figure 5.7 as a function of the relative data consistency error at each iteration k, which we define as $\|Y - h(X^k)\|_2^2/\|Y\|_2^2$. Results for three different sequence lengths, 100, 200 and 500, are shown in the figure. It is clear that in all cases the algorithms converge rapidly and for sequence lengths of 200 or more have effectively converged within 20 iterations (note the log scale along the y-axis). Indeed, this is predicted by the compressed sensing theory for IPA: when the sequence length increases, so that compressed sensing task becomes easier (smaller isometry constant) the rate of convergence also increases. Thus BLIP can be considered to be reasonably computationally efficient.

5.3. Subsampling versus sequence length. In our next experiment we investigate the dependencies of the undersampling ratio and the sequence length on the reconstruction performance. In this experiment we evaluate the image sequence SER as a function of L and p. Recall that the theory presented in section 4.2 suggested that this performance might degrade roughly as a function of p^2/L. However, as we noted earlier, the analysis in that section is of a 'worst case' type and may be highly conservative. Figure 5.8 shows a plot of the image sequence SER as a function of L/p^2 for three different subsampling rates: $p = 16$ (green), $p = 32$ (red) and $p = 64$ (blue). From the plot we can see that the rapid growth of the SER that we associate with successful recovery occurs in each case at roughly the same value of L/p^2. This seems to suggest that the predicted scaling behaviour for L and p in random EPI to achieve RIP is of the right order. This in turn suggests that to maximize efficiency we should attempt to minimize p (all other design criteria being equal).

5.4. Using a complex density model. The simulations, so far, have used the somewhat idealized model that the density map is real and non-negative. In this experiment we demonstrate that the algorithm works just as well when the density map is allowed to be complex and to absorb sensitivity maps and other phase terms. Here we repeat the first experiment but we modify the density map to have a quadratic phase that is zero at the centre of the image and $pi/4$ at the corners. A plot of the phase is shown on the left hand side in figure 5.9.

We then ran the MRF reconstruction algorithm and BLIP with equations (3.6) and (3.7) replaced by (3.8) and (3.9) in both algorithms. The resulting performance was very similar to that in the real valued case. For brevity we only show a plot of the the T2 SER in figure 5.9. We see that the parameter estimation behaves identically to that in the first experiment. Similar behaviour can be observed for the other parameters. Therefore, it seems that there is no significant difference in using the real or complex model for proton density.

5.5. Uniform versus non-uniform sampling. In §4.2 we asserted that as the Bloch response model does not include any spatial structure it is preferable to take uniformly random samples of k-space in order to achieve the RIP rather than use a variable density scheme. In this final experiment we examine the effect of replacing the (uniform) random EPI sampling with a sampling pattern that weights the lower frequencies more, as is common in compressed sensing schemes for MRI [28]. Specifically, we choose a non-uniform sampling pattern with an equivalent undersampling ratio, $M/N = 1/16$, that always samples $k_y = 0, 1, 2, \sqrt{N} - 3, \sqrt{N} - 2$ and $\sqrt{N} - 1$ (the centre of k_y-space), and then samples the remainder of k-space uniformly at
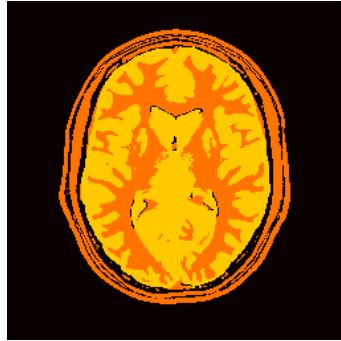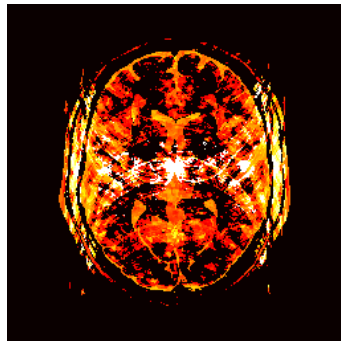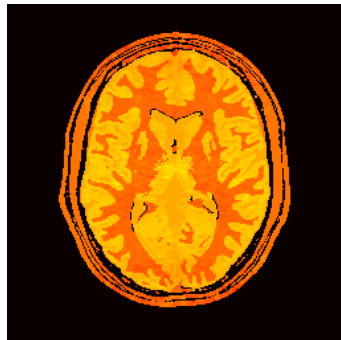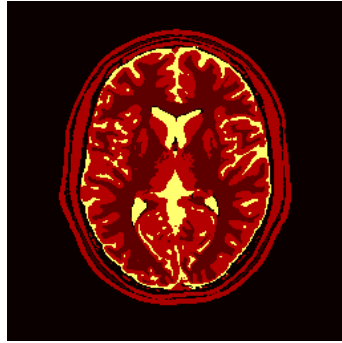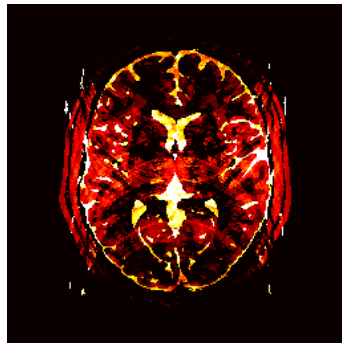
Fig. 5.4. *A visual comparison of the density map estimates from a sequence of length $L = 300$. The top plot shows the original density map. The middle image is the MRF estimate and the bottom image is the BLIP estimate.*

Original T1

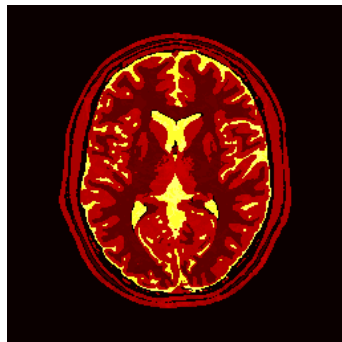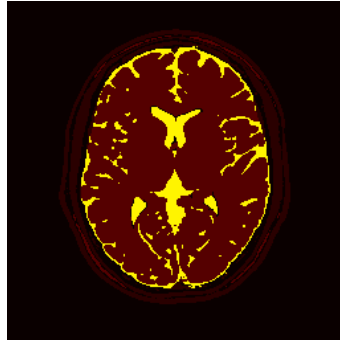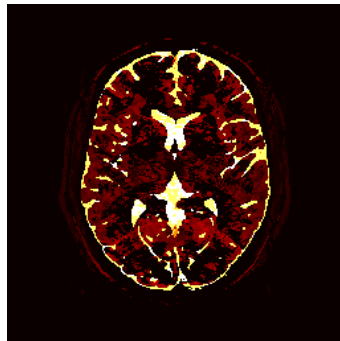MRF T1 estimate

BLIP T1 estimate

Fig. 5.5. *A visual comparison of the* T1 *map estimates from a sequence of length* $L = 300$. *The top plot shows the original* T1 *map. The middle image is the MRF estimate and the bottom image is the BLIP estimate.*
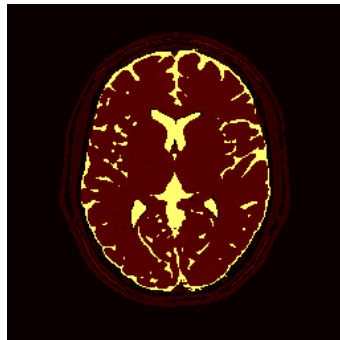
FIG. 5.6. *A visual comparison of the* T2 *map estimates from a sequence of length* $L = 300$. *The top plot shows the original* T2 *map. The middle image is the MRF estimate and the bottom image is the BLIP estimate.*
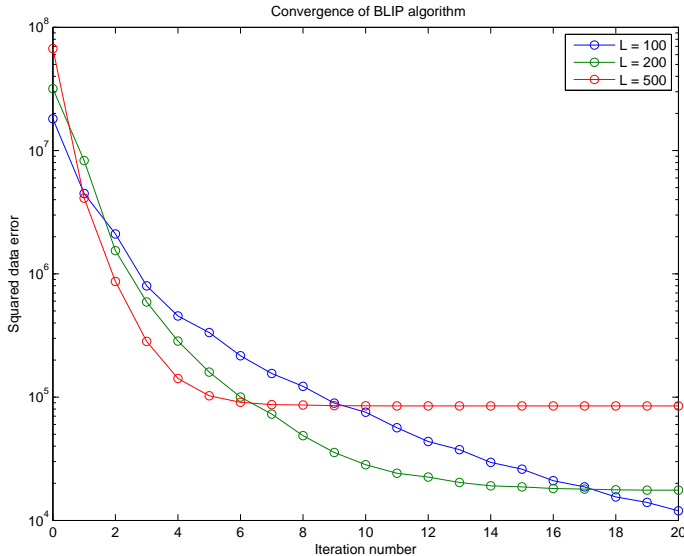
FIG. 5.7. *Plots of the data consistency error at each iteration for BLIP using a varying sequence length. The convergence rate increases as the sequence length increases. This is consistent with theory as the increased sequence length is likely to reduce the isometry constant.*

random (with the remaining 10 samples). While we have not tried to optimize this non-uniform sampling strategy we have found that other variable density sampling strategies performed similarly.

We repeated the first experiment and compared the random EPI sampling to using non-uniform sampling with the sequence length varied between 10 and 300. Again we focus on the T2 reconstruction, although similar behaviour was observed for the density and T1 estimation (not shown). The T2 results are plotted in figure 5.10. It is clear from the figure that BLIP does not perform well with the non-uniform sampling of k-space, and it never achieves the near oracle performance that we observe with the random EPI sampling strategy. Indeed, we observed no non-uniform sampling strategy to do this. Other simulations (not shown) have indicated that uniform i.i.d. undersampling in k_y also performs well, although we have yet to prove this has the RIP.

Interestingly, the MRF reconstruction does benefit from the non-uniform sampling, however the reconstruction quality is still very poor. We believe that this can be explained by the fact that in both cases the MRF reconstructions exhibit significant aliasing. However, in the non-uniform case the aliasing is concentrated more in the high frequencies where the signal has less energy and therefore introduces less distortion.

6. Conclusions and open questions. We have presented a principled mathematical framework for compressed quantitative MRI based around the recently proposed technique of Magnetic Resonance Fingerprinting [29]. The sensing process can be considered in two separate stages. First, the embedding of the parameter information into the magnetization response sequences through the mapping $f(\cdot)$. Second, the compressive imaging of the induced magnetization image sequence. The key elements of our approach have been: the characterization of the signal model through
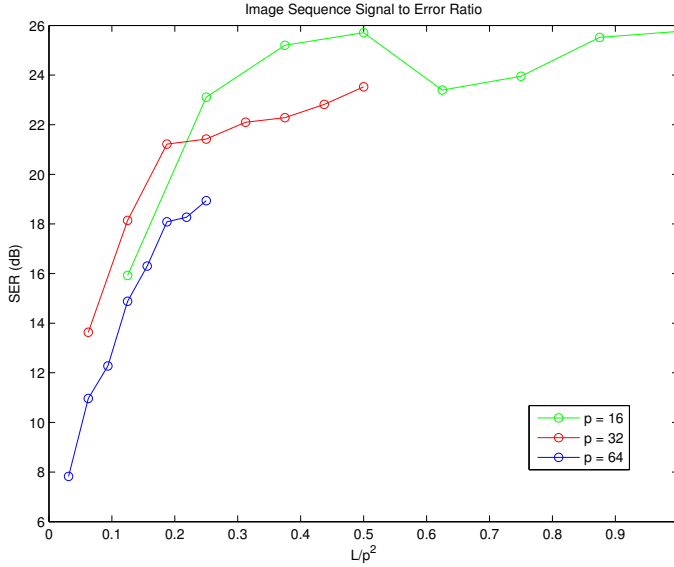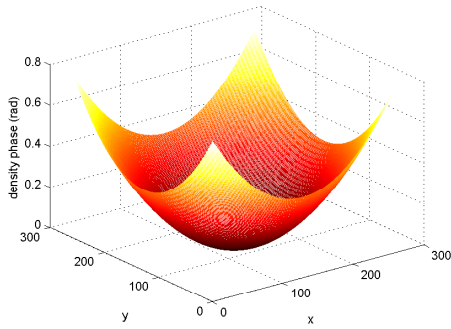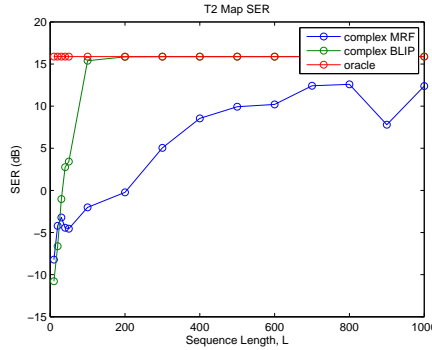
24

FIG. 5.8. *A plot of the Image sequence SER (dB) against L/p^2 for three different levels of undersampling: $p = 16$ (green), $p = 32$ (red) and $p = 64$ (blue). The rapid increase in SER appears to occur at roughly the same value of L/p^2 in each case suggesting that the RIP result in Theorem 1 is of the right order.*

FIG. 5.9. *Reconstruction performance for the* T2 *map using a complex density model. (a) The quadratic phase applied to the density map; (b) SER for* T2 *map estimation as a function of sequence length. Results are shown for the following algorithms: complex MRF, complex BLIP and the complex oracle estimate.*

the Bloch response manifold; the identification of a provably good image sequence reconstruction algorithm based on iterative projection; an excitation response condition based on a newly introduced measure of *flatness* to quantify the persistence of the excitation; and a random EPI k-space sampling scheme that can be shown to have the necessary RIP condition when the excitation is suitably flat. The simulations presented in §5 show that this proposed technique is capable of achieving good parameter map reconstruction with very short pulse sequences.
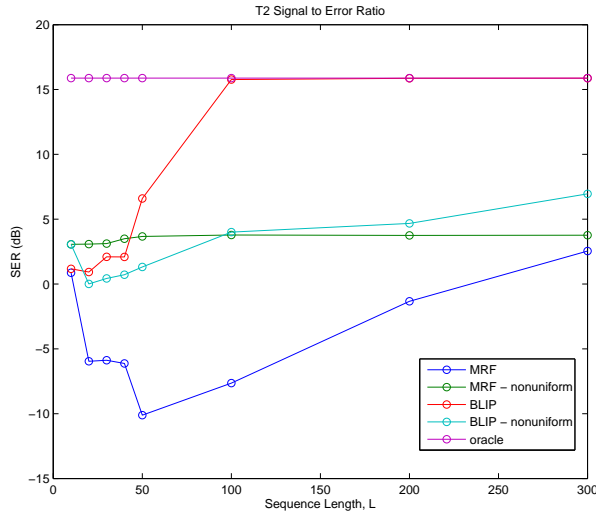
Fig. 5.10. *A plot of the* T2 *estimate SER (dB) against L for reconstruction algorithms MRF and BLIP using uniform random (EPI) sampling with $p = 16$ and a non-uniform random sampling with an equivalent undersampling ratio $M/N = 16$. Only in the case of BLIP with uniform random sampling does the* T2 *estimate performance approach that of the oracle estimator.*

The next step will be to make a thorough comparison on an MRI scanner with MRF and other existing quantitative MRI techniques such as [16]. Furthermore, it is important to acknowledge that any competitive strategy would also need to incorporate some form of parallel (multi-coil) imaging. Given the great similarity between our proposed random EPI sampling and the multi-coil subsampling used in SENSE [32], we anticipate that a parallel version of BLIP should be relatively straight forward to develop. The main additional complication would be the need to estimate the coil sensitivities. It will be interesting to see whether these can be acquired simultaneously as an integral part of the quantitative imaging. We leave this for future work.

While the current work is specifically targeted at a compressed sensing framework for MRF, we believe that many elements of it should be more broadly applicable. Specifically, the RIP condition for randomized EPI may well have applications in other MR imaging strategies and the characterization of excitation response in terms of flatness could prove a useful tool for the analysis of other compressed sensing schemes involving some form of active sensing.

Finally, the use of parametric physical models (through appropriate discretisation) could be applicable to many areas of compressed sensing beyond MRI. The experience we have gained here suggests that such models can be more powerful than traditional spatial image models, such as wavelet sparsity, that are often found in compressive imaging.

6.1. Open Questions. In setting out this compressed sensing framework a number of questions have arisen that we feel should be addressed. We conclude by briefly describing these below.

Excitation sequences. What are the key requirements for the excitation sequences? We have introduced the flatness condition, however, we have so far not exploited randomness in the excitation. This raises the question: does the excitation sequence need to be random? Although randomness seems a natural way to obtain flat responses, it is not clear that it is necessary or even preferable. Previous model based quantitative techniques have reported successful performance without random excitation so the real question is do random excitation sequences enable faster imaging? Random excitations may also be able to provide less stringent sampling conditions that provide the RIP. Furthermore, whether deterministic or random, how should we optimize the excitation sequences in order to maximise the performance of the parameter map estimation? This seems to be very much a system identification problem.

Improved signal models. A key question for the Bloch response model is: how densely do we need to sample $\mathcal{M}$? This will depend on the response mapping f, the undersampling operator h and the performance of the recovery algorithm. It would be interesting to try to quantify these errors using the existing union of subspace compressed sensing theory [8, 11].

A second question is: how should we best include additional modelling information? It is clearly desirable to include spatial regularization. However, we have seen in §5 that the inclusion of our limited spatial regularization within the signal model did not significantly improve performance. On the other hand, this only regularized the density map, whereas, ideally we would like to impose spatial regularity on each of the parameter maps. Unfortunately, a naive construction of such a model would lead to a complex non-separable representation that we cannot easily project onto. Alternatively, we might try to impose block spatial regularity on the image sequence on top of the Bloch response model. This form of spatial regularization was used in [17] and appears to have only provided modest performance improvements. Therefore the question is how to best combine these models to maximize reconstruction performance and can we back this up theoretically?

The current signal model is also somewhat idealised. We have treated the proton density values, ρ_i, as nonnegative, following the physics. However, in MRI it is more common to treat ρ_i as a complex value, absorbing various phase factors into the quantity. While our framework easily extends to the complex case as highlighted in §3.3.1 and evaluated in §5.4, it would be interesting to see whether there was a more principled way to deal with such additional phase factors.

Another idealization that is made both here and in the original MRF is that the read out time is assumed negligible with respect to the relaxation times. Depending on the level of undersampling this may not be true. This might introduce significant artefacts. If so, can we modify the signal model to account for this?

Finally, our model does not account for partial volume effects. These were briefly touched on in the supplementary material of [29], where it was proposed to model individual voxels as a composition of different material components. Such a model is reminiscent of the spatial abundance maps used in hyperspectral imaging. In such a case we are in the realms of compressive source separation [22]. Can we formulate a compressive MRF problem that accounts for partial volume effects in a similar manner?

Subsampling k-space. We have identified certain conditions that guarantee the RIP for random EPI sampling. This allows us to trade off the k-space subsampling factor $p = N/M$ with the length of the excitation sequence, L. Unfortunately the trade off scales as $L \sim p^2$. It is not clear whether similar guarantees could be achieved from

a deterministic sampling sequence or whether this is indeed optimal. It would be more desirable to have a proportional trade off $L \sim p$. Is such a scaling possible? If so, what is the appropriate combination of excitation sequence and sampling strategy?

Finally, if we can successfully incorporate spatial structure into our signal model, as suggested above, it is very likely that a variable density sampling would be preferable. If so, can we leverage existing theory for variable density sampling [1, 33] to develop principled designs for variable density sampling for compressive MRF?

Appendix A. Dynamics of balanced SSFP sequences. Balanced SSFP sequences are popular in MRI and were the basis of the excitation sequences used in MRF [29], although the term 'steady state' is somewhat of a misnomer as this refers to the steady state conditions arrived at following periodic excitation with constant α and TR [36].

In fact, here we are explicitly interested in the transient dynamics of a non-periodic excitation sequence. This is in contrast with traditional SSFP sequences where transient oscillations are seen as undesirable as they can introduce imaging artefacts [23]. In this work, as in [29], we will regard the transient behaviour as essential in enabling us to distinguish between different quantitative behaviour.

The transient response can be formally described in terms of a 3-dimensional linear discrete time dynamical system that we summarize below, see [23, 21, 36] for further details. To keep things simple we will assume there is no phase increment between pulses and also that the lth echo time, TE_l, is half the lth repetition time TR_l.

Following [23], let $\mathbf{m}_l = (m_l^x, m_l^y, m_l^z)^T \in \mathbb{R}^3$ represent the 3-dimensional magnetization vector for a voxel at the lth excitation pulse. In Inversion Recovery SSFP sequences the equilibrium magnetization, $\mathbf{m}_{\text{eq}} = [0, 0, 1]^T$, is initially inverted so that $\mathbf{m}_0 = [0, 0, -1]^T$. Then the magnetization after the lth RF-excitation is given by the following linear discrete time dynamical system:

$$(A.1) \qquad \mathbf{m}_{l+1} = R_x(\alpha_l) R_z(\phi_l) E_l \mathbf{m}_l + R_x(\alpha_l)(\text{Id} - E_l)\mathbf{m}_{\text{eq}}$$

where $R_u(\phi)$ denotes a rotation about the $u \in \{x, y, z\}$ axis by an angle ϕ, $\phi_l = 2\pi\delta f \, \text{TR}_l$ is the off-resonance phase associated with local field variations and chemical shift effects [21] and E_l is the diagonal matrix characterizing the relaxation process:

$$(A.2) \qquad E_l := \begin{pmatrix} e^{-\text{TR}_l / \text{T2}} & & \\ & e^{-\text{TR}_l / \text{T2}} & \\ & & e^{-\text{TR}_l / \text{T1}} \end{pmatrix}$$

where the T1 relaxation time controls the rate of relaxation along the z-axis, while the T2 relaxation time controls the relaxation onto the z-axis.

Finally let $\hat{\mathbf{m}}_l$ denote the magnetization at the echo time, TE_l. Then this is given by [23]:

$$(A.3) \qquad \hat{\mathbf{m}}_l = R_z(\phi_l/2) E_l^{1/2} \mathbf{m}_l + (\text{Id} - E_l^{1/2})\mathbf{m}_{\text{eq}},$$

with the readout coil measuring $\hat{m}_l^x + j\hat{m}_l^y$. Thus the magnetization dynamics in response to a sequence of RF pulses with flip angles, α_l, and repetition times, TR_l, is given by (A.1) and (A.3) which apart from the input parameters is solely a function of the tissue parameters T1, T2, and the off-resonance frequency, δf.

Appendix B. Proof of Theorem 1. We first introduce the key lemmas that form the main ingredients of the proof. Our approach will follow the standard route

of concentration of measure, ϵ-net and union bound. To this end we will need the following well known Chernoff bound [18]:

LEMMA 1. *Let* $X = X_1 + X_2 + \ldots + X_n$, $0 \leq X_i \leq 1$ *with* $\mu = \mathbb{E}(X)$. *Then*

$$(\text{B.1}) \qquad \mathbb{P}(|X - \mu| > \epsilon\mu) \leq 2\exp\left(-\frac{\epsilon^2\mu}{3}\right)$$

The next lemma establishes a near isometry for a single aliased voxel sequence.

LEMMA 2. *Let* $z \in \mathbb{C}^L$ *be a random vector given by:*

$$(\text{B.2}) \qquad z_i = \frac{1}{p}\sum_k U_{k,i} e^{-j2\pi\zeta_i k/p}$$

where ζ_i *are independent random variables drawn uniformly from* $\{0, \ldots, p-1\}$ *and* $U \in \mathbb{C}^{p \times L}$ *is a matrix whose rows have flatness* λ. *Then, with probability at least* $1 - 2e^{-\epsilon^2/(3p\lambda^2)}$, z *satisfies:*

$$(\text{B.3}) \qquad (1 - \epsilon)\|U\|_F^2 \leq p^2\|z\|_2^2 \leq (1 + \epsilon)\|U\|_F^2$$

Proof. We first show that $\mathbb{E}\|z\|_2^2 = \frac{1}{p^2}\|U\|_F^2$ and then derive the necessary tail bounds.

Let $W_{a,k} = \frac{1}{\sqrt{p}}e^{-j2\pi ak/p}$, $a, k = 0, \ldots, p-1$, denote the unitary Discrete Fourier transform in $\mathbb{C}^p$. We can then write

$$(\text{B.4}) \qquad \mathbb{E}\|z\|_2^2 = \sum_{a=0}^{p-1}\frac{1}{p}\left(\sum_{i=1}^L \frac{1}{p}|W_{a,:}U_{:,i}|^2\right)$$

$$(\text{B.5}) \qquad = \frac{1}{p^2}\sum_i\sum_a |W_{a,:}U_{:,i}|^2$$

$$(\text{B.6}) \qquad = \frac{1}{p^2}\sum_i \|U_{:,i}\|_2^2$$

$$(\text{B.7}) \qquad = \frac{1}{p^2}\|U\|_F^2,$$

Now note that $\|z\|_2^2$ is the sum of L independent random variables, $\|z\|_2^2 = \sum_i \xi_i$ with $\xi_i = \frac{1}{p}|W_{\zeta_i,:}U_{:,i}|^2$. Furthermore the ξ_i satisfy:

$$0 \leq \xi_i \leq \frac{1}{p}\|U_{:,i}\|_2^2$$

$$\leq \frac{1}{p}\sum_k \max_i |U_{k,i}|^2$$

$$(\text{B.8}) \qquad \leq \frac{1}{p}\sum_k \lambda^2\|U_{k,:}\|_2^2$$

$$= \frac{\lambda^2}{p}\|U\|_F^2$$

We can therefore apply the Chernoff bound from Lemma 1 to $\sum \xi_i$ rescaled by $\frac{\lambda^2}{p}\|U\|_F^2$ to give:

$$(\text{B.9}) \qquad \mathbb{P}(|\|z\|_2^2 - \frac{1}{p^2}\|U\|_F^2| > \epsilon\frac{1}{p^2}\|U\|_F^2) \leq 2\exp\left(-\frac{\epsilon^2}{3p\lambda^2}\right)$$

29

Rearranging this expression completes the proof. ☐

Next we extend Lemma 2 to a near isometry for groups of aliased voxels under the action of h. Since h is an ortho-projector, $\|h(X)\|_2^2 = \|h^H h(X)\|_2^2$ and so we can equivalently consider the isometry properties of $h^H h$.

Let us denote $Z = h^H(h(X))$ such that $Z_{:,l} = F^H P(\zeta_l)^T Y_{:,l}$. Recall that h is a partially sampled 2D discrete Fourier transform that is fully sampled in the k_x direction and periodically subsampled by a factor of $p = N/M$ in the k_y direction. Therefore each $Z_{i,l}$ is the sum of p aliases taken from $X_{:,l}$:

$$(B.10) \qquad Z_{i,l} = \frac{1}{p} \sum_{k=0}^{p-1} X_{\tau_i(k),l} \; e^{-j2\pi\zeta_l k/p}$$

where $\tau_i(k)$ gives the index of the kth alias for the ith voxel (with $\tau_i(0) = i$). We can therefore partition the set $\{1,\ldots,N\}$ into M disjoint index sets $\Lambda_1,\ldots,\Lambda_M$ with each set associated with p aliases, such that $h^H h$ is separable over $\{\Lambda_i\}$ and $Z_{\Lambda_i,:} = [h^H h]_{\Lambda_i} X_{\Lambda_i,:}$. Since each $Z_{\Lambda_i,:}$ contains p copies of the same combination of aliases (up to a phase shift) we can conclude that:

$$(B.11) \qquad \|Z_{\Lambda_i,:}\|_F^2 = p\|Z_{k,:}\|_2^2, \; \forall k \in \Lambda_i$$

Applying Lemma 2 then gives us:

LEMMA 3. *Let* $Z_{\Lambda_i,:} = [h^H h]_{\Lambda_i} X_{\Lambda_i,:}$ *for some* $X_{\Lambda_i,:} \in \mathbb{C}^{p \times L}$ *whose rows have a flatness* λ *where* $[h^H h]_{\Lambda_i}$ *is defined above. Then with probability at least* $1 - 2e^{-\epsilon^2/(3p\lambda^2)}$ *we have*

$$(B.12) \qquad (1-\epsilon)\|X_{\Lambda_i,:}\|_F^2 \le p\|Z_{\Lambda_i,:}\|_F^2 \le (1+\epsilon)\|X_{\Lambda_i,:}\|_F^2$$

The final ingredient guarantees a near isometry for low dimensional subsets of the unit sphere (for a more sophisticated but slightly different result in this direction see [14])

LEMMA 4. *Let* $S \subset \mathbb{S}^{n-1}$ *have box counting dimension* d *such that for any* $\epsilon > 0$ *there exists an* ϵ-*cover of* S *of size* $C_S \epsilon^{-d}$. *Let* $P : \mathbb{C}^n \to \mathbb{C}^k$ *be a random projection such that for any* $\delta > 0$ *and a fixed* $x \in S$,

$$(B.13) \qquad 1 - \delta \le \frac{n}{k}\|Px\|_2^2 \le 1 + \delta$$

holds with probability at least $1 - c_0 e^{-c_1 \delta^2}$. *Then* P *satisfies* (B.13) *for all* $x \in S$ *with probability at least* $1 - \eta$ *as long as:*

$$(B.14) \qquad c_1 \ge 72\delta^{-2} \left(d\log(36n/\delta k) + \log C_S c_0/\eta \right)$$

Proof. Consider an ϵ-cover S_ϵ of S with $\epsilon = \delta'/(2\sqrt{n/k})$ and suppose that P satisfies

$$(B.15) \qquad 1 - \delta'/2 \le \frac{n}{k}\|Px\|_2^2 \le 1 + \delta'/2$$

for all $x \in S_\epsilon$ with a constant $0 < \delta' < 1$. Then there exists a $u \in S_\epsilon$ such that:

$$(\text{B.16}) \qquad \sqrt{\frac{n}{k}}\|Px\|_2 \leq \sqrt{\frac{n}{k}}\|Pu\|_2 + \sqrt{\frac{n}{k}}\|P(x-u)\|_2$$

$$(\text{B.17}) \qquad \leq 1 + \delta'/2 + \sqrt{\frac{n}{k}}\epsilon$$

$$(\text{B.18}) \qquad = 1 + \delta'$$

where in (B.17) we have used the fact that $(1 + \delta'/2)^2 > (1 + \delta'/2)$.

We can similarly show that $\sqrt{\frac{n}{k}}\|Px\|_2 \geq 1 - \delta'$. Then finally noting that the "non-squared" RIP implies the squared RIP in (B.13) with $\delta = 3\delta'$ gives us the required isometry.

It only remains to bound the probability of failure. Let p_f be the probability that P fails to satisfy (B.13) on S. By the union bound:

$$(\text{B.19}) \qquad pf \leq |S_\epsilon| c_0 e^{-c_1(\delta'/2)^2}$$

$$(\text{B.20}) \qquad \leq C_S c_0 \left(\frac{\delta'}{2\sqrt{n/k}}\right)^{-d} e^{-c_1(\delta'/2)^2}$$

Therefore it is sufficient to choose η so that:

$$(\text{B.21}) \qquad \frac{\eta}{C_S c_0} \geq \left(\frac{\delta}{6\sqrt{n/k}}\right)^{-d} e^{-c_1(\delta/6)^2}$$

Re-arranging the above gives:

$$(\text{B.22}) \qquad c_1 \geq 72\delta^{-2}\left(d\log(36n/\delta k) + \log C_S c_0/\eta\right)$$

as required. □

We are now ready to prove the main theorem.

Proof. [Proof of Theorem 1]

First, note that $\mathbb{R}_+\mathcal{B} \subset \mathbb{R}\mathcal{B}$ which is an infinite union of subspace model, as is its p-product, $(\mathbb{R}\mathcal{B})^p$ associated with a group of aliased voxels, Λ_i. To guarantee that h_{Λ_i} possesses the necessary RIP on $(\mathbb{R}\mathcal{B})^p - (\mathbb{R}\mathcal{B})^p$ it is sufficient to consider the RIP on the normalized difference set S given by:

$$(\text{B.23}) \qquad S = \{x \in ((\mathbb{R}\mathcal{B})^p - (\mathbb{R}\mathcal{B})^p), \|x\|_2 = 1\},$$

due to the linearity of h.

By construction we have $\dim(S) = 2pd_\mathcal{B} - 1$ and we can therefore apply Lemma 4 to S together with Lemma 3. This guarantees for all $X_{\Lambda_i,:} \in (\mathbb{R}\mathcal{B})^p - (\mathbb{R}\mathcal{B})^p$ that h satisfied (B.12) with probability at least $1 - \eta$ as long as:

$$(\text{B.24}) \qquad \lambda^{-2} \geq (3p) \times 72\delta^{-2}\left((2pd_\mathcal{B} - 1)\log(36p/\delta) + \log C_S c_0/\eta\right)$$

To ensure this holds for all aliased voxel groups Λ_i, $i = 1, \ldots, M$ we can again apply the union bound and replace η by $M\eta$. Noting that $p, \delta^{-1}, \eta^{-1} > 1$ we can collect together the constants and simplify to finally give:

$$(\text{B.25}) \qquad \lambda^{-2} \geq C\delta^{-2}p^2 d_\mathcal{B} \log(N/\delta\eta)$$

for some constant C independent of $p, N, d_\mathcal{B}, \delta$ and η which gives the required conditions of the theorem. □

REFERENCES

[1] B. Adcock and A. C. Hansen, Generalized sampling and infinite-dimensional compressed sensing. DAMTP Tech. Rep. 2011/NA12, 2011.

[2] N. Ailon and B. Chazelle, The Fast Johnson-Lindenstrauss Transform and Approximate Nearest Neighors. SIAM J. Computing, vol. 39, No. 1, pp. 302-322, 2009.

[3] R. Baraniuk, M. Davenport, R. De Vore, and M. Wakin, A simple proof of the restricted isometry property for random matrices, Constructive Approx., vol. 28, pp. 253–263, 2008.

[4] R. G. Baraniuk and M. B. Wakin, Random Projections of Smooth Manifolds. Foundations of Computational Mathematics, vol. 9(1), pp. 51-77, 2009.

[5] A Beygelzimer, S. Kakade and J. Langford, Cover Trees for Nearest Neighbor. In Proceedings of the 23rd International Conference on Machine Learning (ICML), Pittsburgh, PA, pp. 97–104, 2006.

[6] K. T. Block, M. Uecker, and J. Frahm, Model-based Iterative Reconstruction for Radial Fast Spin-Echo MRI. IEEE Trans. Med. Imag., vol. 28(11), pp. 1759–1769, 2009.

[7] T. Blumensath and M. E. Davies, Iterative Hard Thresholding for Sparse Approximation, J. Fourier Analysis and Applications, vol. 14, no. 5, pp. 629–654, 2008.

[8] T. Blumensath and M. E. Davies, Sampling Theorems for Signals From the Union of Finite-Dimensional Linear Subspaces. IEEE Trans. Inf. Theory, vol. 55(4), pp. 1872–1882, 2009.

[9] T. Blumensath and M. E. Davies, Iterative Hard thresholding for Compressed sensing. Applied Computational Harmonic Analysis, vol. 27, no. 3, pp. 265-274, 2009.

[10] T. Blumensath, M. E. Davies, Normalised Iterative Hard Thresholding; guaranteed stability and performance, IEEE Journal of Selected Topics in Signal Processing, vol. 4(2), pp. 298-309, 2010.

[11] T. Blumensath, Sampling and Reconstructing Signals From a Union of Linear Subspaces. IEEE Trans. Inf. Theory, vol. 57(7), pp. 4660–4671, 2011.

[12] Brainweb data repository, available at: `http://brainweb.bic.mni.mcgill.ca/brainweb/`

[13] M. Bydder, A. A. Samsonov, and J. Du, Evaluation of optimal density weighting for regridding. Mag. Res. Im., vol. 25(5), pp. 695-702, 2007.

[14] K. Clarkson, Tighter Bounds for Random Projections of Manifolds. Proceedings of the 24th annual symposium on Computational geometry (SCG'08), pp. 39-48, 2008.

[15] D.L. Collins, A.P. Zijdenbos, V. Kollokian, J.G. Sled, N.J. Kabani, C.J. Holmes and A.C. Evans, Design and Construction of a Realistic Digital Brain Phantom. IEEE Trans. on Medical Imaging, vol.17(3), pp.463–468, 1998.

[16] S.C.L. Deoni, B.K. Rutt, and T.M. Peters, Rapid Combined T1 and T2 Mapping Using Gradient Recalled Acquisition in the Steady State. Magn. Reson. Med. vol. 49, pp. 515-526, 2003.

[17] M. Doneva, P. Bornert, H. Eggers, C. Stehning, J. Senegas and A. Mertins, Compressed sensing reconstruction for magnetic resonance parameter mapping, Magn. Reson. Med., vol 64, pp. 1114-1120, 2010.

[18] D.P. Dubhashi and A. Panconesi, *Concentration of Measure for the Analysis of Randomized Algorithms.* Cambridge University Press, 2009.

[19] J. A. Fessler and B. P. Sutton, Nonuniform fast Fourier transform using min-max interpolation. IEEE Trans. Sig. Proc. vol. 51(2) pp. 560–574, 2003.

[20] J. A. Fessler, Model-based image reconstruction for MRI. IEEE Sig. Proc. Mag., vol. 27(4), pp. 81–89, 2010.

[21] C. Ganter, Off-Resonance Effects in the Transient Response of SSFP Sequences. Magn. Reson. Med. vol. 52, pp. 368-375, 2004.

[22] M. Golbabaee, S. Arberet and P. Vandergheynst, Compressive Source Separation: Theory and Methods for Hyperspectral Imaging. IEEE Trans. Image Proc., vol. 22(12), pp. 5096–5110, 2013.

[23] B.A. Hargreaves, S.S. Vasanawala, J.M. Pauly and D.G. Nishmura, Characterization and Reduction of the Transient Response in Steady-State MR Imaging. Magn, Reson. Med. vol 46, pp. 149–158, 2001.

[24] J.P. Hornak, *The Basics of MRI.* Webbook, available on-line at: `http://www.cis.rit.edu/htbooks/mri/`.

[25] C. Huang, A. Bilgin, T. Barr and M. I. Altbach, T2 relaxometry with indirect echo compensation from highly undersampled data. Magn Reson Med., vol. 70, pp. 1026-1037, 2013.

[26] M. Iwen and M. Maggioni, Approximation of Points on low-dimensional manifolds via random linear projections. arXiv:1204.3337.

[27] E. T. Jaynes, Matrix treatment of nuclear induction. The Physics Review, vol 98(4), pp.

1099–1105, 1955.

[28] M. Lustig, D. L. Donoho, J.M. Santos, and J.M. Pauly, Compressed sensing MRI. IEEE Sig. Proc. Mag., vol. 25(2), pp. 72-82, 2008.

[29] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. L. Sunshine, J. L. Duerk and M. A. Griswold, Magnetic Resonance Fingerprinting. Nature, vol. 145, pp. 187–192, 2013.

[30] G.C. McKinnon, Ultrafast interleaved gradient-echo-planar imaging on a standard scanner. Magn. Reson. Med. vol. 30, pp. 609–616, 1993.

[31] P. Niyogi, S. Smale, and S. Weinberger, Finding the Homology of Submanifolds with High Confidence from Random Samples. Discrete Comput. Geom. vol. 39(1), pp. 419–441, 2008.

[32] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, SENSE: Sensitivity Encoding for Fast MRI. Magn. Reson. Med. vol. 42, pp. 952–962, 1999.

[33] G. Puy, P. Vandergheynst and Y. Wiaux, On Variable Density Compressive Sampling. IEEE Sig. Proc. Lett., vol. 18(10), pp. 595–598, 2011.

[34] G. Puy, P. Vandergheynst, R. Gribonval and Y. Wiaux, Universal and efficient compressed sensing by spread spectrum and application to realistic Fourier imaging techniques. EURASIP Journal on Advances in Signal Processing, 2012, 2012:6.

[35] H. Rauhut, Compressive sensing and structured random matrices. Radon Series Comp. Appl. Math., vol. 9, pp. 1–92, 2010.

[36] K. Sheffler and S. Lehnhardt, Principles and applications of balanced SSFP techniques. Eur. Radiol. vol. 13, pp. 2409-2418, 2003.

[37] J. Tran-Gia, D. Stab, T. Wech, D. Hahn, and H. Kostler, Model-based Acceleration of Parameter mapping (MAP) for saturation prepared radially acquired data. Magn Reson Med., vol. 70, pp. 1524-1534, 2013.

[38] G. A. Wright, Magnetic Resonance Imaging. IEEE Sig. Proc. Mag., vol.14(1), pp. 56–66, 1997.

[39] B. Zhao, F. Lam, W. Luy and Z.-P. Liang, Model-based MR parameter mapping with sparsity constraint. IEEE Int. Symp. Biomed. Imag. (ISBI), pp. 1–4, 2013.