Quantitative magnetic resonance imaging: From fingerprinting to integrated physics-based models

Guozhi Dong[1,2], Michael Hintermüller[1,2], Kostas Papafitsoros[1]

submitted: July 31, 2018 (revision: November 20, 2018)

[1] Weierstrass Institute
Mohrenstr. 39
10117 Berlin
Germany
E-Mail: guozhi.dong@wias-berlin.de
michael.hintermueller@wias-berlin.de
kostas.papafitsoros@wias-berlin.de

[2] Humboldt-Universität zu Berlin
Unter den Linden 6
10099 Berlin
Germany
E-Mail: hint@math.hu-berlin.de
guozhi.dong@hu-berlin.de

No. 2528

Berlin 2018

Quantitative magnetic resonance imaging: From fingerprinting to integrated physics-based models

Guozhi Dong, Michael Hintermüller, Kostas Papafitsoros

Abstract

Quantitative magnetic resonance imaging (qMRI) is concerned with estimating (in physical units) values of magnetic and tissue parameters e.g., relaxation times T_1, T_2, or proton density ρ. Recently in [Ma et al., Nature, 2013], Magnetic Resonance Fingerprinting (MRF) was introduced as a technique being capable of simultaneously recovering such quantitative parameters by using a two step procedure: (i) given a probe, a series of magnetization maps are computed and then (ii) matched to (quantitative) parameters with the help of a pre-computed dictionary which is related to the Bloch manifold. In this paper, we first put MRF and its variants into a perspective with optimization and inverse problems to gain mathematical insights concerning identifiability of parameters under noise and interpretation in terms of optimizers. Motivated by the fact that the Bloch manifold is non-convex and that the accuracy of the MRF-type algorithms is limited by the "discretization size" of the dictionary, a novel physics-based method for qMRI is proposed. In contrast to the conventional two step method, our model is dictionary-free and is rather governed by a single non-linear equation, which is studied analytically. This non-linear equation is efficiently solved via robustified Newton-type methods. The effectiveness of the new method for noisy and undersampled data is shown both analytically and via numerical examples for which also improvement over MRF and its variants is documented.

1 Introduction

1.1 Context

The current routine of magnetic resonance imaging (MRI) examinations typically provides qualitative images of nuclear magnetization of tissue accompanied by contrast "weights". Physicians then visually inspect these images, which, being qualitative only, may however not provide enough information for certain diagnostic purposes. In order to remedy this, *quantitative MRI* (qMRI) seeks to not only visualize the structure of the imaged object, but also to provide accurate parameters values (in physical units) that characterize different tissue types. Such quantities are typically the proton density ρ of Hydrogen atoms in water molecules, and the longitudinal and the transverse relaxation parameters T_1 and T_2, respectively, among others. These magnetic parameters are related to the evolution of the net magnetization $\mathfrak{m}$ through the renowned Bloch equations [4]:

$$
\begin{aligned}
\tfrac{\partial m}{\partial t}(t) &= m(t) \times \gamma B(t) - \left(\tfrac{m_x(t)}{T_2}, \tfrac{m_y(t)}{T_2}, \tfrac{m_z(t)-m_{eq}}{T_1} \right)^\top, \\
m(0) &= m_0.
\end{aligned}
\tag{1}
$$

Here m, yielding $\mathfrak{m} = \rho m$, is the macroscopic magnetization of (Hydrogen) proton of some unitary density in the tissue under an external magnetic field B, and the relaxation rates T_1 and T_2 are

associated model parameters. Further, m_0 represents an initial state. System (1) is instrumental in our quantification process established below and will be further described in Section 2.1.

Although qMRI techniques are still in their infancy, several interesting ideas and methods have already been conceived. Early approaches [21] are based on a set of spin echo or inversion recovery images that are reconstructed from k-space data with respect to various repetition times (TR) and echo times (TE). In that context, acquisitions are designed for each parameter individually. The overall technique is often referred to as parametric mapping method and consists of two steps: (i) reconstruct a sequence of images as in qualitative MRI, and (ii) for each pixel of those images fit its intensity to an ansatz curve characterized by the magnetic parameter associated to the tissue imaged at that pixel. Based on this idea, many improvements have been suggested in the literature; see for instance [14]. The associated approaches aim to simplify the physical model and handle tissue parameters separately, as these are considered to be time consuming for the patient.

Another line of research, initiated by Ma et al. in [23] and named Magnetic Resonance Fingerprinting (MRF), has recently gained considerable attention. First, in an offline phase, it builds a database (dictionary) consisting of all trajectories (fingerprints) of the evolution of the associated magnetization. Each of the latter is obtained by solving Bloch equations for some pre-selected combination of parameter values, typically those of T_1 and T_2 (but sometimes also others such as, e.g., the off-resonance frequency). The underlying parameter combinations stem from a (sufficiently large) selection within a region C_{ad} which is meaningful for human tissue. Hence, the outcome of this first step of the method is a physiologically informed dictionary $\mathrm{Dic}(C_{ad})$ (i.e., a look-up table) relating a set of feasible parameters to their associated solutions of Bloch equations. In a second phase, given a (sensed) magnetization trajectory that is assumed to be related to a solution of the Bloch equations, with the help of this look-up table, the method identifies the parameter values that fit best to this trajectory. This main principle behind MRF enables a simultaneous estimation of (quantitative) tissue parameters.

As our new method is inspired by MRF we further detail the MRF-workflow. Focusing on a thin slice Ω of the tissue of interest, its first step is to reconstruct a sequence of L images $\{X^{(\ell)}\}_{\ell=1}^L$ from data $\{D^{(\ell)}\}_{\ell=1}^L$ as in qualitative MRI, using a sufficiently rich excitation process through L fast radio pulses; see Section 2 for more details. At every time step, the data consists of a sub-sampling of the Fourier coefficients of the magnetization. Sub-sampling occurs due to the short time between each excitation. In a typical MRF routine, the reconstruction of the magnetization relies on the pseudo-inverse of the Fourier transform. This leads to noticeable artifacts in the magnetization images. However, the evolution of the magnetization of a specific tissue element (voxel) along the series of the reconstructed magnetization images can be assumed to correspond (approximately) to the solution of Bloch equations with parameters that correspond to this specific voxel. Hence, the second step of MRF matches the recorded trajectory of each voxel to a fingerprint in the pre-computed dictionary, typically through minimizing a least-squares distance. In this way, the parameter values that correspond to the "best" fingerprint are then assigned to that very voxel. Formally, the MRF procedure can be stated as follows:

- Compute $X^{(\ell)} \in \underset{X}{\mathrm{argmin}} \left\| P^{(\ell)} \mathcal{F} X - D^{(\ell)} \right\|_2^2, \quad \ell = 1, \ldots, L,$ (MRF-step 1)

- Compute $m^* \in \underset{m \in \mathsf{Dic}(C_{ad})}{\mathrm{argmin}} \; S(T_{x,y} m, X^*)$ with $X^* := (X^{(1)}, \ldots, X^{(L)})$. (MRF-step 2)

In (MRF-step 1) $D = (D^{(1)}, \ldots, D^{(L)})$ denotes the data obtained after each pulse, where $D^{(\ell)}$ is a sub-sampling of the Fourier coefficients of the magnetization, more precisely of its transverse component $T_{x,y}$, i.e., the first two components only. Here, P^ℓ is the ℓ-th sub-sampling operator, and the Fourier transform is denoted by $\mathcal{F}$. Accordingly, the first step computes L reconstructions of the magnetization of the tissue slice, i.e., $X^{(\ell)} = \mathcal{F}^{-1}(P^{(\ell)})^\top D^{(\ell)}$, where $\mathcal{F}^{-1}$ stands for the pseudo-

inverse. In (MRF-step 2), for every voxel r (in practice, for every image pixel) the best approximation is obtained via exhaustive search over the dictionary $\text{Dic}(C_{ad})$. We recall that $\text{Dic}(C_{ad}) = \{m^\theta : \theta \in C_{ad}\}$, where θ is a vector of tissue parameters – here, for simplicity, $\theta = (T_1, T_2)$ – and C_{ad} is the admissible domain for these parameters. By m^θ we denote the solution of Bloch equations with parameter θ, evaluated at the same time instances as for the magnetization responses. Hence, every element of $\text{Dic}(C_{ad})$ is a vector sequence of length L. The function $S(\cdot, \cdot)$ is a Euclidean distance of normalized quantities, in order to avoid the multiplicative effect that the density ρ has in the magnetization. Correspondingly, the minimization task in (MRF-step 2) has to be understood in a "voxel-wise" sense, i.e., it is performed as often as the number of voxels (in practice pixels). Finally, the spatial parameter maps are formed by assigning (to the corresponding voxels) the parameter values θ that correspond to the optimal matchings m^θ. For more details we refer to Section 2.3.

While first numerical results [23] show that MRF is a promising qMRI approach, several issues remain open from a mathematical viewpoint which motivate our work. For instance, with respect to stability one is interested in knowing whether two close trajectories yield similar parameter values. Assuming that this is the case, i.e., the method is stable, and L is fixed, then mainly two factors influence the accuracy of MRF: (i) the quality of the L magnetization reconstructions, and (ii) the completeness (fineness) of the dictionary. Clearly, one is interested in analysing and optimizing both aspects. In this vein, the available literature mostly focuses on improving (i), see, e.g., [2, 6], while (ii) remains critical and will always impose limitations depending on the quality of the dictionary. Moreover, both steps of the MRF procedure may benefit from each other when combined. For example, in order to obtain improved reconstructions in (MRF-step 1), it may be informed by the physics-based model built into (MRF-step 2). This motivates our approach of integrating the Bloch equations already into step (MRF-step 1), yielding a single-step qMRI approach upon fixing the best approximation problem.

Davies et al. [6] proposed a first approach in this direction which exhibits advantages over the original MRF scheme. They coined the name **BL**och response recovery through **I**terated **P**rojection (BLIP) for their solution scheme which relates to a projected Landweber-type iteration for reconstructing the magnetization. A key step of the procedure is to project, in every iteration, the current reconstruction onto a dictionary related to the *Bloch manifold*. This leads to an improved solution for the magnetization especially in the case of strongly sub-sampled data. The underlying constrained optimization problem reads

$$\min_X \quad \|P\mathcal{F}T_{x,y}X - D\|_2^2 \quad \text{subject to (s.t.)} \quad X \in \mathbb{R}^+\text{Dic}(C_{ad}). \tag{BLIP}$$

While we defer more details on the BLIP method to Section 2.3, we mention already here that the potential non-convexity of the positive cone of Bloch manifolds $\mathbb{R}^+\text{Dic}(C_{ad})$ represents a major complication as the projection may become non-unique; compare Proposition 4.2 below. This problem is even more concerning when data is corrupted by noise. In addition, as the projection (matching) is still dictionary based, the method can be memory consuming, especially when the dictionary is highly refined in order to have high accuracy.

1.2 Our contribution and the structure of the paper

Our work has two major focus points. In a first part, contained in Section 3, we perform sensitivity analysis to show that the matching process is a well-posed inverse problem when it is restricted to the Bloch manifold. This fact partially explains why the concept of involving a dictionary in MRF has been so successful. In particular in Theorem 3.6 we show that if two trajectories of the magnetization evolution as dictated by the Bloch equations are sufficiently close, then the same holds true for the

associated parameters, i.e., $\|\theta - \theta^\delta\| \le C\delta$ if $\|m - m^\delta\| \le \delta$. Here θ, θ^δ are the inferred parameters given the Bloch trajectories m, m^δ. The constant C is independent of δ, and the norms will be made precise later in the text. Furthermore, we also establish a mathematical understanding of why a large number of frames L yields a positive influence on the quality of the final result; compare Theorem 4.7.

In the second part of our work, with the goal of avoiding the potentially ill-posed projection step in BLIP, we aim to solve the parameter identification problem directly subject to the Bloch manifold. The associated new single-step model reads: Find (ρ, θ) such that

$$Q(\rho, \theta) := P\mathcal{F}(\rho T_{x,y} m(\theta)) = D, \quad \text{with } (\rho(r), \theta(r)) \in \mathbb{R}^+ \times C_{ad}, \quad \text{for every } r \in \Omega. \quad (2)$$

Here the qMRI-operator Q inserts the Bloch dynamics into the data acquisition, and by solving (2) we can recover both ρ and θ. However, the non-linearity of Q makes the problem rather challenging as also additional difficulties arise due to aspects like, e.g., sub-sampling and noise in MRI.

We propose a projected Levenberg-Marquardt regularized variant of the Gauss-Newton method as a numerical remedy under such adverse circumstances. Analytically, this requires a differentiability result for the map $\theta \mapsto m(\theta)$. Furthermore, as for many highly non-linear and non-convex problems, the initialization of the iteration turns out to be crucial. For this initialization, we suggest to use BLIP (or MRF) with a rather coarse dictionary for efficiency purposes, only. Overall it turns out that our approach allows to produce more accurate parameter maps in less CPU-time.

We mention that similar single-step dictionary-free approaches can be found in the recent papers [28] and [27]. In particular, the model in [28] abandons the Fourier space character of the data and asks for a relatively large number of data frames which leads to solving a very large non-linear system. As a result, the method is memory and CPU intense. The work of [27] focuses mainly on the experimental design, aiming at optimizing the excitation pulse sequences as well as the repetition times.

Structure of the paper. The rest of the paper is structured as follows: In Section 2, we provide a general background of MRI, particularly to the Bloch equations and MRF. In Section 3, we relate MRF-type algorithms to inverse problems. We also perform stability analysis for inversion of the Bloch mapping. Our new integrated physics approach leading to a single-step model in form of a non-linear operator equation is the focus of Section 4. We analyse the differentiability of the associated operator and show the non-convexity of the Bloch manifold. Subsequently, we discuss several Newton-type methods for its numerical solution. Here, we particularly focus on the case of undersampled and noisy data. In order to illustrate the efficiency of the proposed method for qMRI, numerical tests and comparisons are presented in Section 5. A short description on solutions of Bloch equations in different cases is given in the Appendix.

2 Background on MRI and MRF

We provide here a brief summary of the principles underlying MRI as they are useful for our purpose of generating an integrated physics-based model for reconstruction; see [31] for more details. Also, a mathematical description of MRF and the BLIP algorithms are given.

2.1 Bloch equations

The Bloch equations [4] characterize the key physics in nuclear magnetic resonance. For the sake of their derivation, let Ω be a domain in $\mathbb{R}^2$ modelling a thin slice of tissue. Every element point (or

voxel) of Ω is denoted by r. The main principles of MRI lie in the interaction between an externally applied dynamic magnetic field $B = (B_x, B_y, B_z)^\top$ and the (net or bulk) magnetization which is equal to all the individual dipole moments of the proton spins within a voxel. This net magnetization is proportional to the hydrogen proton density ρ. Correspondingly, letting $m = (m_x, m_y, m_z)^\top$ denote the magnetization per unit density element, the net magnetization in a voxel of density ρ equals ρm.

In the case of a static magnetic field B_0, which is typically regarded to lie in z-direction, the net magnetization is aligned to that field with its longitudinal component m_z reaching an equilibrium m_{eq}. This alignment is not achieved instantaneously but it is controlled by the *longitudinal relaxation time* T_1 (or $T_1(r)$ emphasizing the dependence on a specific voxel). The longitudinal magnetization evolves according to $m_z(t) = m_{eq}(1 - e^{-(t/T_1)})$. Furthermore, the part of the magnetization orthogonal to B_0, which is called the transverse magnetization $(m_x, m_y)^\top$, precesses about the z-axis at a frequency equal to $\gamma|B_0|$ where γ denotes the gyromagnetic ratio. This precession emits an electromagnetic signal which can be detected and measured by the coils of the MR machine. The transverse magnetization decays exponentially at a rate T_2, the *transverse relaxation time*.

The overall macroscopic dynamics that dictate the relation between the magnetization m, the magnetic field B and the relaxation times T_1, T_2, are governed by the Bloch equations, which is a system of linear ordinary differential equations (ODEs):

$$\begin{array}{rcl} \frac{\partial m(t,r)}{\partial t} & = & m(t,r) \times \gamma B(t,r) - \Theta(r) \bullet (m(t,r) - m_e), \\ m(0,r) & = & m_0(r), \end{array} \tag{3}$$

where $m_e = (0, 0, m_{eq})^\top$ (without loss of generality we assume $m_e = (0, 0, 1)^\top$ in what follows), and "$\times$" denotes the outer product between vectors. For the ease of notation we use

$$\Theta(r) := (\Theta_1(r), \Theta_2(r), \Theta_3(r))^\top := \left(1/T_2(r), 1/T_2(r), 1/T_1(r)\right)^\top,$$

and the operation $\bullet$ in (3) denotes Hadamard product (component-wise multiplication of vectors). As introduced above, $m : (0, \tau) \times \Omega \to \mathbb{R}^3$, for some time horizon $\tau > 0$, denotes the magnetization in a unit volume per unit proton density, and m_0 is a given initial state. Note that the dependence on r is here intrinsic and does not enter the equation. As the latter is linear, one can simply multiply (3) by ρ in order to get the net magnetization.

The total magnetic field $B(t,r)$ can be typically decomposed into

$$B(t,r) = B_0(r) + B_1(t,r) + (0, 0, G(t) \cdot r)^\top. \tag{4}$$

Here B_0 denotes the external constant magnetic field that points into the positive z direction, and it is generally assumed to be spatially homogeneous. For the sake of generality, we, however, keep here the dependence on r. The summand $B_1(t,r) = (B_{1,x}(t,r), B_{1,y}(t,r), 0)^\top$ corresponds to a radio frequency (RF) pulse, which is sent periodically and lasts only for a very short time. It is used to excite the magnetization from its equilibrium by turning the magnetization precession away from the direction of the main magnetic field with the so-called flip angle

$$\alpha(t) = \gamma \int_0^t |B_1(s)| \, ds.$$

These pulses usually last only very briefly compared to T_1 and T_2. Therefore, RF sequences can be completely characterized by sequences of flip angles, and time is normally omitted. The interval between two consecutive pulses is called repetition time (TR). As we shall see later in Section 4, we consider a specific flip angle sequence pattern referred to as *Inversion Recovery balanced Steady*

State Free Precession (IR-bSSFP) [29]. Through this choice, the solution of Bloch equations can be simulated by a discrete linear dynamical system; see Section 4.1. In the Appendix, we provide case discussions concerning the discrete Bloch dynamics and the solution of Bloch equations. The factor $G(t)$ in (4) is a magnetic gradient field designed to differentiate the point-wise information from the detected signal.

In brief, the measured signal can be expressed by

$$S(t) = \int_\Omega \rho(x, y)(T_{x,y}m(t, x, y))e^{-\mathrm{i}\gamma|B_0|t}e^{-\mathrm{i}\gamma \int_0^t (xG_x + yG_y)d\tau}dxdy,$$

where $T_{x,y}m := m_x(x, y) + \mathrm{i}m_y(x, y)$ stands for the transverse magnetization, and i is the imaginary unit. Alternatively, one can think of $T_{x,y}m$ as a pair of real-valued components. The third component of m can usually not be measured due to the position of coils. Finally, up to a demodulation by $e^{\mathrm{i}\gamma|B_0|t}$, the MR signal $D^{(t)}$ can mathematically be modelled as a collection of coefficients of a Fourier transform of the transverse magnetization, i.e.,

$$P^{(t)}\mathcal{F}(\rho T_{x,y}m^{(t)}) = D^{(t)},$$

where $\mathcal{F}$ denotes the Fourier transform and $P^{(t)}$ a sub-sampling operator.

2.2 Sub-sampling

In MRI, and in particular in MRF, one does not wait for the signal to return to equilibrium between two excitation pulses and due to time constraints only a small proportion of the k-space is sampled. Reconstruction of the magnetization under such circumstances leads to the occurrence of aliasing artifacts, especially when this reconstruction uses the basic (but fast to apply) pseudo-inverse of $\mathcal{F}$.

In the literature, three different sub-sampling schemes are designed and are often practically employed: spiral, radial and Cartesian sub-sampling. Each of these corresponds to a different variation in time of the selection gradients G_x and G_y. In the original version of MRF, the former two patterns were preferred as the associated aliasing artifacts appear to be uncorrelated, respectively, and can be roughly treated as random noise. The latter is not the case for Cartesian sub-sampling; see for instance the numerical examples in [6]. The BLIP method, reviewed in the next section and improving over MRF, however perfectly fits to Cartesian sub-sampling. As our starting point is the BLIP method, we thus focus here on Cartesian sub-sampling based on multishot echo-planar imaging (EPI) [24]; see Section 5.2 for details. We note however that this choice is not limiting as other sub-sampling patterns may be used as well.

2.3 MRF and BLIP in some detail

In MRF one initially considers a pre-designed excitation pattern of L flip angles $\{\alpha_\ell\}_{\ell=1}^L$ separated by a repetition time TR. Here, for simplicity, we consider TR to be constant but this is not necessary. Also a subset $C_{ad} \subset \mathbb{R}^m$ of the space of tissue parameters to be estimated is predefined. For the ease of exposition, here we consider C_{ad} to contain admissible $\theta = (T_1, T_2)$-values, yielding $m = 2$. For example, values for T_1 would typically range from $685\mathrm{ms}$ (white matter on brain) to $4880\mathrm{ms}$ (cerebrospinal fluid), with the corresponding range for T_2 to be $65\mathrm{ms}$–$550\mathrm{ms}$ [23]. As we shall see below, in our dictionary-free approach we choose C_{ad} to be a convex subset of $\mathbb{R}^+ \times \mathbb{R}^+$, in particular a box, thus admitting values between a minimum and a maximum value. Dictionary based methods then

replace C_{ad} by a sufficiently fine discretization yielding J parameter values $\{\theta_j\}_{j=1}^J$. For simplicity, in this section we write C_{ad} also for the discretization. Using this set of J parameter values, the specific excitation pattern, the sequence of flip angles $\{\alpha_\ell\}_{\ell=1}^L$ and the repetition time TR, one can simulate the Bloch equations by using a discrete linear dynamical system. In this context, the solutions of the Bloch equations are evaluated at discrete times $t_1, t_2, \ldots, t_L$; see Section 4.1 for details. This generates a dictionary $\mathrm{Dic}(C_{ad})$ of J magnetization responses (i.e., trajectories of the solutions of Bloch equations evaluated at times $t_1, t_2, \ldots, t_L$) $\{m^{\theta_j}\}_{j=1}^J$:

$$\mathrm{Dic}(C_{ad}) = \{m^{\theta_j} : \theta_j \in C_{ad}, \ j = 1, \ldots, J\} \subset \left(\left(\mathbb{R}^3\right)^L\right)^J.$$

Next, MR data are collected at the respective L read-out times. Each component $D^{(\ell)}$ of the data $D = (D^{(1)}, \ldots, D^{(L)})$, corresponds to a sub-sampling (resulting by $P^{(\ell)}$) of the Fourier coefficients of the net magnetization $X^{(\ell)}$. Here, the reconstruction of the transverse magnetization image is done via the least square solution and hence these images suffer from aliasing artifacts. This step therefore consists of solving L least square solutions (using the pseudo-inverse Fourier transform $\mathcal{F}^{-1}(P^{(\ell)})^\top$) to obtain $X^* = (X^{(1)}, \ldots, X^{(L)})$, where $X^{(\ell)} : \Omega \to \mathbb{R}^2$. Note that, instead of $\mathbb{R}^2$, one can also use the complex number representation of the reconstructed magnetization X and the Bloch response m, i.e., $m = m_x + \mathrm{i}m_y$. Observe that in this section Ω denotes a set of discrete voxels, which in practice are represented by pixels $i : 1, \ldots, N$. Summarizing, we have

Step 1 of the MRF process: Reconstruction of the magnetizations
Reconstruct the vector of L net magnetizations $X^* = (X^{(1)}, \ldots, X^{(L)})$ by solving

$$X^{(\ell)} \in \operatorname*{argmin}_{X:\Omega\to\mathbb{R}^2} \|P^{(\ell)}\mathcal{F}X - D^{(\ell)}\|_2^2 \quad \text{using } X^{(\ell)} = \mathcal{F}^{-1}(P^{(\ell)})^\top D^{(\ell)}, \quad \ell = 1, \ldots, L$$

The second and final step of MRF identifies the transverse component of m^{θ_j} (denoted by $T_{x,y}m^{\theta_j}$) in the dictionary $\mathrm{Dic}(C_{ad})$ that best matches the reconstructed magnetization at every voxel. The desired parameter map $\theta : \Omega \to \mathbb{R}^2$ is then obtained by mapping every discrete voxel i to the θ-value that corresponds to the matched m^θ, and the reconstructed magnetization sequence at voxel i, i.e., $(X_i^\ell)_{\ell=1}^L$, contributes with density ρ_i that is associated with this particular tissue element. Utilizing normalization and an ℓ_2-projection onto the discrete Bloch manifold, the best approximation and following density computation yield

Step 2 of the MRF process: Matching of the magnetizations to the dictionary
For every discrete voxel $i = 1, \ldots, N$, compute the projected magnetization $X_i = (X_i^\ell)_{\ell=1}^L$ according to

$$m^{\theta_{j_i}} = \operatorname*{argmin}_{m^\theta \in \mathrm{Dic}(C_{ad})} \left\| \frac{T_{x,y}m^\theta}{\|T_{x,y}m^\theta\|_2} - X_i \right\|_2^2.$$

Then extract $\{\theta_{j_i}\}_{i=1}^N = \{(T_1(i), T_2(i))\}_{i=1}^N$ from a look-up table, and compute the density map $\{\rho_i\}_{i=1}^N$ as

$$\rho_i = \frac{\|X_i\|_2}{\|T_{x,y}m^{\theta_{j_i}}\|_2}.$$

One may notice that Step 1 very likely has non-unique minimizers due to sub-sampling. In [23] the specific minimizer $X^{(\ell)} = \mathcal{F}^{-1}((P^{(\ell)})^\top D^{(\ell)})$ was chosen, which, however, may not be suitable; compare, e.g., [6]. BLIP is an alternative MRF-approach and is due to [6]. It operates in the following way:

Algorithm 1. *BLIP [6].*

   1 Generate a dictionary $\mathrm{Dic}(C_{ad})$.

   2 Initialize the magnetization vector $X = 0$ and choose an initial step size μ_1.

   3 For $n = 1, 2, 3, \ldots$ iterate as follows:

      3.1 For every $\ell = 1, \ldots, L$, perform a gradient decent step yielding

$$\left(X^{(\ell)}\right)_{n+1} = \left(X^{(\ell)}\right)_n - \mu_n \mathcal{F}^{-1}(P^{(\ell)})^\top \left(P^{(\ell)}\mathcal{F}(X^{(\ell)})_n - D^{(\ell)}\right).$$

      3.2 Project each $(X_i)_{n+1} = \left(\left(X_i^{(\ell)}\right)_{\ell=1}^L\right)_{n+1}$ onto the dictionary $\mathrm{Dic}(C_{ad})$ to obtain, as Step 2 in MRF,

$$\left(m^{\theta_{j_i}}\right)_{n+1} = \left(\left((m^{\theta_{j_i}})^{(\ell)}\right)_{\ell=1}^L\right)_{n+1} \text{ and } (\rho_i)_{n+1}$$

      for every voxel $i, = 1, \ldots, N$.

      3.3 For every $\ell = 1, \ldots, L$, update $\left(X^{(\ell)}\right)_{n+1}$ as follows

$$\left(X_i^{(\ell)}\right)_{n+1} \leftarrow (\rho_i)_{n+1} \left((T_{x,y}m^{\theta_{j_i}})^{(\ell)}\right)_{n+1}, \quad i = 1, \ldots, N.$$

      3.4 Update the step size μ_n (see [6] for some rules).

   4 Upon termination of the iteration with outcome X, as in MRF, construct parameter maps from X by using a look-up table.

As mentioned above, BLIP aims to compute an approximate solution to

$$\min_X \quad \|P\mathcal{F}T_{x,y}X - D\|_2^2, \text{ s.t. } X \in \mathbb{R}^+\mathrm{Dic}(C_{ad}) \tag{BLIP}$$

by employing a projected gradient descent method. Note that in contrast to MRF, BLIP integrates the dictionary constraint into a single minimization step and is shown in [6] to be superior to MRF, in particularly for Cartesian sub-samping.

3 MRF as an inverse problem and its stability analysis

3.1 Towards a coupled inverse problem

For the sake of generality, our starting point is the time continuous version of the Bloch equations. In order to fix our setting, let $\mathcal{Y} := [L^2(\Omega)]^3$ and $\mathcal{Z} := [L^\infty(\Omega)]^3$. The initial magnetization is given by $m_0 \in \mathcal{Y}$, and $B \in L^\infty(0, \tau; \mathcal{Z})$ denotes a given external magnetic field for some time horizon $\tau > 0$. We recall the Bochner space

$$L^\infty(0, \tau; \mathcal{Z}) := \{f : (0, \tau) \to \mathcal{Z} : \|f\|_{L^\infty(0,\tau;\mathcal{Z})} < +\infty\},$$

with $\|f\|_{L^\infty(0,\tau;\mathcal{Z})} = \operatorname*{ess\,sup}_{0 < t < \tau} \|f(t)\|_{\mathcal{Z}}$.

The space $L^1(0, \tau; \mathcal{Y})$ is defined similarly. The space $W^{1,1}(0, \tau; \mathcal{Y})$ consists of all the functions $f : (0, \tau) \rightarrow \mathcal{Y}$ such that both f and $\frac{\partial f}{\partial t}$ belong to $L^1(0, \tau; \mathcal{Y})$. We refer to [8] for more on Lebesgue, Sobolev and Bochner spaces.

A natural space for the parameter $\theta = (T_1, T_2)$ is $[L^\infty(\Omega)]^2$, and we also consider this parameter to be bounded uniformly away from zero. Consequently, we have $\Theta = (1/T_2, 1/T_2, 1/T_1)^\top$ belongs to $[L^\infty(\Omega)]^3$, as well. Finally, recall that $m_e \equiv (0, 0, 1)^\top$.

For our further analysis, it is convenient to introduce the operator

$$\mathcal{B}_{m_0, B} : [L^\infty(\Omega)]^2 \rightarrow \{m : (0, \tau) \rightarrow \mathcal{Y}\},$$

where $\mathcal{B}_{m_0, B}(\theta)$ denotes the solution mapping of the Bloch equations (3) up to time τ. Equipped with this notation, we now state the following family of inverse problems which represents a continuous version of the MRF process:

- Problem 1: For some $t_\ell \in (0, \tau)$, $\ell = 1, \ldots, L$, in order to obtain $X^{(t_\ell)} \in L^2(\Omega)$ solve the linear equation

$$P^{(t_\ell)} \mathcal{F} X^{(t_\ell)} = D^{(t_\ell)}, \tag{5}$$

  where $D^{(t_\ell)} \in [L^2(K)]^2$, K is a bounded frequency domain which is usually called k-space, $\mathcal{F} : [L^2(\Omega)]^2 \rightarrow [L^2(K)]^2$, $P^{(t_\ell)} : [L^2(K)]^2 \rightarrow [L^2(K)]^2$.

- Problem 2: For every $r \in \Omega$, to obtain $\theta = \theta(r) \in \mathbb{R}^+ \times \mathbb{R}^+$ solve

$$\rho(r) T_{x,y}(\mathcal{B}_{m_0, B}(\theta))(\cdot, r) = X^{(\cdot)}(r), \tag{6}$$

where $\rho \in L^\infty(\Omega)$, and $T_{x,y}$ is the transverse projection. Note that, strictly speaking, the coupling of (5) and (6) makes sense only when $P^{(t_\ell)} = \mathrm{Id}$, i.e., there is no sub-sampling. This is because of the fact that under sub-sampling, uniqueness of solutions for (5) is not guaranteed, and $X^{(\cdot)}(r)$ may not belong to the Bloch manifold.

Here Problem 1 corresponds to the first step in MRF and aims to invert the Fourier transform for sub-sampled (and potentially noisy) data. This type of problem is the central mathematical problem in standard MRI and has been extensively studied in the literature. In particular, variational methods e.g., sparse regularization methods and optimal weighted total variational methods, have been successfully applied towards that [11, 13, 19, 22], to mention only a few recent results.

In view of the parameter identification problem involving the Bloch equations in the second step of MRF we now focus on equation (6). But for the sake of ease of demonstration, we neglect the effect of the density map ρ and the transverse projection operator $T_{x,y}$, i.e., we study

$$\mathcal{B}_{m_0, B}(\theta) = m. \tag{7}$$

3.2 Stability analysis on inverting the Bloch mapping

With the aim of quantifying the influence of inaccuracies or noise in the solution of (5) on solving (6), we next analyse stability of (7). This is of relevance for both, MRF and BLIP.

In order to simplify the discussion, in this section we consider the time domain $(0, \tau)$ to be the period between two consecutive pulses. From a modelling point of view, m_0 will be the magnetization right after the first pulse, i.e., after the application of the flip angle displacement and, $m(\tau, \cdot)$ will be the

magnetization right before the next pulse. In that case the magnetic field $B \neq 0$ is considered to be time independent which means that B is a constant function in $L^\infty(0, \tau; \mathcal{Y})$ with respect to time and with (possibly spatially varying) values in $\mathcal{Y}$. Also, the effect of the gradient field G is ignored here, as it only encodes the MRI signal.

From a classical result for evolutionary equations in Banach spaces (see e.g. [3, Proposition 3.3]), we infer existence of a solution m of (3) in $W^{1,1}(0, \tau; \mathcal{Y})$. In fact, m enjoys even higher regularity, but for our purposes $W^{1,1}(0, \tau; \mathcal{Y})$ turns out to be sufficient. Hence, we consider

$$\mathcal{B}_{m_0, B} : [L^\infty(\Omega)]^2 \to W^{1,1}(0, \tau; \mathcal{Y})$$

for given m_0 and B.

Given the existence of solutions, our further analysis relies on the following assumptions:

Assumption 3.1. *Let* $\theta(r) \in C_{ad}$ *for all* $r \in \Omega$, *where* $C_{ad} \subset \mathbb{R}^+ \times \mathbb{R}^+$ *denotes a feasible domain which is convex and bounded away from zero.*

For our next assumption, we define the range of the Bloch map, i.e.,

$$R(\mathcal{B}_{m_0, B}) := \left\{ m : \ m = \mathcal{B}_{m_0, B}(\theta) \text{ with } \theta \in [L^\infty(\Omega)]^2 \text{ and } \theta(r) \in C_{ad} \text{ for all } r \in \Omega \right\}.$$

Assumption 3.2. *Let*

$$m \in R(\mathcal{B}_{m_0, B}) \subset W^{1,1}(0, \tau; \mathcal{Y}), \tag{8}$$

be a solution of the Bloch equations (3). *Then the quantity* $(\omega_\tau^1(r), \omega_\tau^2(r), \omega_\tau^3(r))^\top := \int_0^\tau m(t, r) dt - m_e \tau$ *is bounded away from zero, i.e., there is a constant* $c_\tau > 0$ *such that*

$$\inf_{r \in \Omega} \left| \omega_\tau^i(\tau) \right| \geq c_\tau, \quad \text{for } i = 1, 2, 3. \tag{9}$$

Remark 3.3. *Assumption 3.1 implies no factual limitation in practice. Assumption 3.2 is also justified in practice as we consider* $(0, \tau)$ *to be the time between two consecutive pulses which roughly equals to repetition time. In this period, the net magnetization always satisfies* $m_x > c_\tau' > 0, m_y > c_\tau'' > 0$, *and* $m_z < m_e$, *and these give the estimate* (9). *Since, in an MRI experiment, the time domain consists of the repetition of periodic radio pulses, (the excitation time of the pulse is usually very short) our assumption is satisfied during the entire experiment.*

Theorem 3.4. *Let Assumption 3.2 hold, the magnetic field satisfy* $B \neq 0$, *and let* $m \in R(\mathcal{B}_{m_0, B})$ *for some* θ. *Then the* θ-value associated with m is unique.

Proof. Observe that by integrating the Bloch equations over the time domain $(0, \tau)$, we have

$$\Theta(r) = \left(m(0, r) - m(\tau, r) + \int_0^\tau m(t, r) \times \gamma B(t, r) dt \right) ./ \omega_\tau, \tag{10}$$

where "./" denotes a component-wise quotient of vectors. Note that the integrals are well-defined, since for almost every r, $m(r, \cdot) \in L^1(0, \tau)$. Also, due to Assumption 3.2 we have $\omega_\tau^i \neq 0$ for $i \in \{1, 2, 3\}$. The uniqueness of $\theta = \left(\frac{1}{\Theta_3}, \frac{1}{\Theta_1} \right)^\top$ follows readily. $\qquad \square$

We immediately have the next corollary.

Corollary 3.4.1. *Let Assumption 3.2 hold, and* $B \neq 0$. *Then the Bloch mapping satisfies*

$$\mathcal{B}_{m_0, B}(\theta_1) = \mathcal{B}_{m_0, B}(\theta_2) \iff \theta_1 = \theta_2.$$

Proof. For every fixed $\theta(r) \in C_{ad}$, the Bloch mapping is well-defined under Assumption 3.2. The mapping is injective by Theorem 3.4. The other direction follows from the Picard–Lindelöf Theorem (see, e.g., [30]). □

Remark 3.5. *The uniqueness result indicates that enforcing the magnetization function to be in the range of a Bloch mapping also guarantees a unique parameter. This explains the idea behind BLIP which aims at an improved solution (when compared to MRF) from undersampled data by using projection steps onto the Bloch manifold.*

The main stability result of this section is stated next.

Theorem 3.6. *Let the Assumption 3.1 be satisfied, and let $m, m^\delta \in R(\mathcal{B}_{m_0, B})$ with corresponding parameters θ, θ^δ. If Assumption 3.2 holds for both m and m^δ, and $\left\| m - m^\delta \right\|_{W^{1,1}(0,\tau;\mathcal{Y})} \leq \delta$ for $\delta > 0$, then we have*

$$\left\| \theta - \theta^\delta \right\|_{[L^1(\Omega)]^2} \leq C(\tau, \theta, B)\delta,$$

where $C(\tau, \theta, B)$ is a constant depending on τ, θ and B, but not on δ.

Proof. Using equation (10) with the obvious definition of ω_τ^δ, we have

$$
\begin{aligned}
\Theta - \Theta^\delta &= \left(m_0 - m(\tau, r) + \int_0^\tau m(t, r) \times \gamma B(t, r) dt \right) ./\omega_\tau \\
&\quad - \left(m_0^\delta - m^\delta(\tau, r) + \int_0^\tau m^\delta(t, r) \times \gamma B(t, r) dt \right) ./\omega_\tau^\delta \\
&= \Theta \bullet \left((\omega_\tau^\delta - \omega_\tau) ./\omega_\tau^\delta \right) - \left(m_0^\delta - m_0 - m^\delta(\tau, r) + m(\tau, r) \right) ./\omega_\tau^\delta \\
&\quad - \left(\int_0^\tau m^\delta(t, r) \times \gamma B(t, r) - m(t, r) \times \gamma B(t, r) dt \right) ./\omega_\tau^\delta.
\end{aligned}
$$

Note that $\int_0^\tau m^\delta(t, r) - m_e^\delta(r) dt = \int_0^\tau m^\delta(t, r) dt - m_e^\delta \tau$, and $m_e^\delta = m_e$. Due to (9), we have

$$\inf_{r \in \Omega} \left| \int_0^\tau m(t, r) - m_e(r) dt \right| \geq c_\tau, \quad \text{and} \quad \inf_{r \in \Omega} \left| \int_0^\tau m^\delta(t, r) - m_e(r) dt \right| \geq c_\tau.$$

As a consequence, we obtain the estimate

$$
\begin{aligned}
\left\| \Theta - \Theta^\delta \right\|_{[L^1(\Omega)]^3} &\leq \frac{1}{c_\tau} \int \left| \Theta(r) \bullet \int_0^\tau (m^\delta(t, r) - m(t, r)) dt \right| dr \\
&\quad + \frac{1}{c_\tau} \int \left| \int_0^\tau \left(\frac{\partial m(t, r)}{\partial t} - \frac{\partial m^\delta(t, r)}{\partial t} \right) dt \right| dr \\
&\quad + \frac{1}{c_\tau} \int \left| \int_0^\tau (m^\delta(t, r) - m(t, r)) \times \gamma B(t, r) dt \right| dr \\
&\leq \frac{1}{c_\tau} C \| m^\delta - m \|_{L^1(0,\tau;[L^1(\Omega)]^3)} + \frac{1}{c_\tau} \left\| \frac{\partial m^\delta}{\partial t} - \frac{\partial m}{\partial t} \right\|_{L^1(0,\tau;[L^1(\Omega)]^3)} \\
&\quad + \frac{1}{c_\tau} C \| m^\delta - m \|_{L^1(0,\tau;[L^1(\Omega)]^3)},
\end{aligned}
$$

with generic constants C depending on γ, B and Θ. Here we have used the facts that $\Theta \in [L^\infty(\Omega)]^2$ and the outer product with $B(t, r)$ can be written as the application of a linear operator with bounded norm (in t and r) as B is bounded in $L^\infty(0, \tau; \mathcal{Z})$ and also independent of time.

As $(\theta_1, \theta_2) = (\frac{1}{\Theta_3}, \frac{1}{\Theta_1})$ and by Assumption 3.1 we can find a constant $C_1 > 0$ such that

$$\left\| \theta - \theta^\delta \right\|_{[L^1(\Omega)]^2} \le C_1 \left\| \Theta - \Theta^\delta \right\|_{[L^1(\Omega)]^3}.$$

This follows from the fact that the function $h : [a, b] \to \mathbb{R}$, with $h(\beta) = 1/\beta$, is Lipschitz for $0 < a < b < \infty$. The proof is completed by combining the last two inequality relations and the fact that the $[L^1(\Omega)]^3$ norm is bounded by $[L^2(\Omega)]^3$ norm. $\qquad\square$

The above result can be interpreted as follows. Theorem 3.6 shows that the inverse problem (7) is well-posed by restricting the right hand side to the range of the Bloch mapping. That is, if the reconstructed magnetization is in the Bloch manifold (more precisely the positive cone of the manifold), then the values of the tissue parameters θ recovered from the dictionary should in principle be not too far away from the exact solutions.

The analytical properties of the Bloch mapping and its inverse not only support the application of MRF-type schemes, but they also motivate us to find yet more accurate solution techniques for quantitative MRI. This is our target in the next section.

4 An integrated physics-based method for qMRI

We now propose a method for qMRI that integrates the physics model into the reconstruction process. In contrast to the previously discussed two-step procedures, it consists of a single step only.

On an abstract level, our model is associated with the non-linear operator equation

$$Q(\mathbf{x}) = D, \tag{11}$$

where $\mathbf{x}(r) = (\rho(r), \theta(r)) \in \tilde{C}_{ad} := \mathbb{R}^+ \times C_{ad}$ for all $r \in \Omega$, D is the acquired MRI signal, and the qMRI-operator Q is defined by

$$Q(\mathbf{x}) := P\mathcal{F}(\rho T_{x,y} M(\theta)). \tag{12}$$

It integrates the Bloch mapping within the data acquisition procedure.

Anticipating our subsequent development, $M(\theta)$ represents the discrete Bloch dynamics, which corresponds to the time continuous version $m(\theta)$ previously discussed.

4.1 Bloch mapping as discrete dynamics

With the aim of employing a fast imaging protocol for absolute quantification of T_1 and T_2 post-contrast (e.g., upon administering Gadolinium (Gd)), we focus here on Inversion Recovery balanced Steady-State Free Precession (*IR-bSSFP*) flip angle sequence patterns; see [29] and compare also [18]. IR-bSSFP is a specific MRI excitation pulse sequence widely used in applications and it allows for a simple approximation of the solutions of the Bloch equations at the read out times. In our subsequent analysis and numerical examples, we always use the associated discrete dynamics approximating the continuous Bloch equations.

To simplify the presentation, we will ignore the factor of off-resonance and only consider the homogeneous case of the flip angles and off-resonance frequency. In this case, the magnetization after each

n-th excitation pulse is simulated by the following recursion formula

$$
\begin{cases}
M_\ell & = & E_1(TR_\ell, \theta) R_{\phi_\ell} R_x(\alpha_\ell) R_{\phi_\ell}^\top M_{\ell-1} + E_2(TR_\ell, \theta) M_e, \\
M_e & = & (0, 0, 1)^\top, \\
M_0 & = & -M_e = (0, 0, -1)^\top.
\end{cases}
\tag{13}
$$

Here $\{\alpha_\ell\}_{\ell=1}^L$ and $\{TR_\ell\}_{\ell=1}^L$ are the flip angles and repetition time sequences, and $\{M_\ell\}_{\ell=1}^L$ are the magnetizations at the middle of each TR_ℓ time interval. Moreover we denote

$$
E_1(TR_\ell, \theta) = \begin{pmatrix} e^{-\frac{TR_\ell}{T_2}} & 0 & 0 \\ 0 & e^{-\frac{TR_\ell}{T_2}} & 0 \\ 0 & 0 & e^{-\frac{TR_\ell}{T_1}} \end{pmatrix}, \qquad E_2(TR_\ell, \theta) = \left(1 - e^{-\frac{TR_\ell}{T_1}}\right)
$$

and also

$$
R_{\phi_\ell} = \begin{pmatrix} \cos(\phi_\ell) & \sin(\phi_\ell) & 0 \\ -\sin(\phi_\ell) & \cos(\phi_\ell) & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \text{and} \quad R_x(\alpha_\ell) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha_\ell) & \sin(\alpha_\ell) \\ 0 & -\sin(\alpha_\ell) & \cos(\alpha_\ell) \end{pmatrix}.
$$

The angle ϕ_ℓ denotes a phase shift by the gradient magnetic fields [29] and is assumed to be known.

Writing (13) in a compact form, we are able to derive the evolution of the discrete system for the magnetization vectors

$$
\begin{aligned}
M_\ell = {} & \left(\prod_{k=1}^\ell E_1(TR_k, \theta) R(\alpha_k) \right) M_0 + E_2(TR_\ell, \theta) M_e \\
& + \sum_{k=1}^{\ell-1} \left(E_2(TR_k, \theta) \prod_{j=k+1}^\ell E_1(TR_j, \theta) R(\alpha_j) \right) M_e,
\end{aligned}
\tag{14}
$$

where we use the matrix notation $R(\alpha_\ell) := R_{\phi_\ell} R_x(\alpha_\ell) R_{\phi_\ell}^\top$. Note that (14) establishes a mapping between θ and $\{M_\ell\}_{\ell=1}^L$ yielding a discrete (in time) version of the operator $\mathcal{B}_{m_0, B}$ associated with the IR-bSSFP pulse sequence. For utilizing Gauss-Newton-type algorithms for solving (11), it is of interest to study differentiability of this mapping. This and further properties are therefore the subjects of the following section.

4.2   Properties of the Bloch mapping and the qMRI-operator

Consider the discrete Bloch mapping as defined in (13):

$$
M : V \to \left[\mathcal{Y} := [L^2(\Omega)]^3 \right]^L, \quad M(\theta) := \{M_\ell(\theta)\}_{\ell=1}^L,
$$

where V is the open subset of $[L^\infty(\Omega)]^2$ that consists of all functions with strictly positive values almost everywhere. In the later of our text, we constantly use the notation $o(\sigma)$ for small positive real value σ satisfying $\frac{o(\sigma)}{\sigma} \to 0$ as $\sigma \to 0$.

Proposition 4.1. *Let $\{M_\ell(\theta)\}_{\ell=1}^L$ be the sequence given in* (14), *$\{\alpha_\ell\}_{\ell=1}^L$ the sequence of flip angles with $\alpha_\ell \in (0, \pi)$ for every $\ell = 1, \ldots, L$, and $\{TR_\ell\}_{\ell=1}^L$ the sequence of repetition times with $TR_\ell > 0$ for every ℓ. Given $M_0 \in \mathcal{Y}$, then the following statements hold true:*

(i) M is Fréchet differentiable with bounded derivative. Moreover, given sufficiently small $h \in [L^\infty(\Omega)]^2$ we have the general estimate for all $q \geq 2$ and for $q = +\infty$:

$$\|M(\theta + h) - M(\theta) - M'(\theta)h\|_{[\mathcal{Y}]^L} = o\left(\|h\|_{[L^q(\Omega)]^2}\right). \tag{15}$$

In addition if $M_0 \in [L^\infty(\Omega)]^3$, then there exist some constant C independent of h for the following estimate

$$\|M'(\theta)h\|_{[\mathcal{Y}]^L} \leq C \|h\|_{L^2(\Omega)]^2}. \tag{16}$$

(ii) Let $M_e = (0, 0, 1)^\top$ and either $M_0 = M_e$ or $M_0 = -M_e$, then the operator M is injective, i.e., given θ^a, $\theta^b \in V$, we have

$$M(\theta^a) = M(\theta^b) \implies \theta^a = \theta^b.$$

Proof. Due to the recursive nature of M_ℓ, it suffices to analyse M_1:

$$M_1(\theta) = E_1(TR_1, \theta)R_\phi R_x(\alpha_1)R_\phi^\top M_0 + E_2(TR_1, \theta)M_e.$$

(i) We start by considering the differentiability of $M_1(\theta)$. This is readily derived when using the differentiability of $x \mapsto e^{-\frac{TR}{x}}$ for $x > 0$. We denote by $M_1'(\theta)$ the Fréchet derivative of the map M_1 evaluated at θ, that is $M_1'(\theta) : [L^\infty(\Omega)]^2 \to \mathcal{Y}$ bounded, linear such that

$$\lim_{h \to 0} \frac{\|M_1(\theta + h) - M_1(\theta) - M_1'(\theta)h\|_{\mathcal{Y}}}{\|h\|_{[L^\infty(\Omega)]^2}} = 0. \tag{17}$$

To simplify the formulas, for every $\ell = 1, \ldots, L$, we denote

$$U_1(\ell) := \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \frac{TR_\ell}{(T_1)^2}e^{-\frac{TR_\ell}{T_1}} \end{pmatrix} R(\alpha_\ell),$$

and

$$U_2(\ell) := \begin{pmatrix} \frac{TR_\ell}{(T_2)^2}e^{-\frac{TR_\ell}{T_2}} & 0 & 0 \\ 0 & \frac{TR_\ell}{(T_2)^2}e^{-\frac{TR_\ell}{T_2}} & 0 \\ 0 & 0 & 0 \end{pmatrix} R(\alpha_\ell).$$

We compute

$$M_1'(\theta) = \left(M_{1,1}'(\theta),\ M_{1,2}'(\theta)\right) := \left(U_1(1)M_0 - \frac{TR_1}{T_1^2}e^{-\frac{TR_1}{T_1}}M_e,\ U_2(1)M_0\right). \tag{18}$$

Note that $M_1'(\theta) \in \mathcal{Y} \times \mathcal{Y} = [L^2(\Omega)]^3 \times [L^2(\Omega)]^3$. It can be regarded as a bounded linear operator from $[L^\infty(\Omega)]^2 \to \mathcal{Y} \times \mathcal{Y}$ which is defined for every $h = (h_1, h_2) \in [L^\infty(\Omega)]^2$ as

$$\begin{aligned} M_1'(\theta)h &= M_{1,1}'(\theta)h_1 + M_{1,2}'(\theta)h_2 \\ &:= \left([M_{1,1}'(\theta)]_x h_1, [M_{1,1}'(\theta)]_y h_1, [M_{1,1}'(\theta)]_z h_1\right) \\ &\quad + \left([M_{1,2}'(\theta)]_x h_2, [M_{1,2}'(\theta)]_y h_2, [M_{1,2}'(\theta)]_z h_2\right), \end{aligned}$$

where $[\cdot]_x$, $[\cdot]_y$, $[\cdot]_z$ denote components of a vector, and the multiplication of $L^2(\Omega)$- and $L^\infty(\Omega)$-functions is understood in a pointwise sense. The resulting product is in $L^2(\Omega)$. Using the fact that

$(e^{-\frac{TR}{x}})' = \frac{TR}{x^2}e^{-\frac{TR}{x}}$ is Lipschitz continuous over $x \in (0, \infty)$ for every fixed $TR > 0$ (actually the derivative of every order of $e^{-\frac{TR}{x}}$ is Lipschitz continuous), then the following pointwise estimate holds true:

$$|M_1(\theta(r) + h(r)) - M_1(\theta(r)) - M_1'(\theta(r))h(r)| \le C\,|h(r)|^2, \quad \text{for all } r \in \Omega, \qquad (19)$$

where C is a constant independent of h and θ, as well independent on r. Note that

$$\left\||h|^2\right\|_{L^2(\Omega)} \le C_q \left\||h|^2\right\|_{L^q(\Omega)}, \quad \text{for all } q \ge 2 \text{ and } q = +\infty,$$

where C_q is a also constant independent of h and θ. Then for all $h \in [L^\infty(\Omega)]^2$ sufficiently small

$$\frac{\|M_1(\theta + h) - M_1(\theta) - M_1'(\theta)h\|_{\mathcal{Y}}}{\|h\|_{[L^q(\Omega)]^2}} \le \frac{C\left\||h|^2\right\|_{L^2(\Omega)}}{\|h\|_{[L^q(\Omega)]^2}} \le \frac{CC_q\left\||h|^2\right\|_{L^q(\Omega)}}{\|h\|_{[L^q(\Omega)]^2}},$$

and with this we get (15) for M_1.

The Frechét differentiability of M_1 in the space $[L^\infty(\Omega)]^2$ is then a consequence of the above estimate, where the constant $C_q = C_\infty = |\Omega|$. In this case, it then implies (17) which gives us the conclusion.

The derivative of $M_\ell(\theta)$ for $\ell > 1$ can then be calculated by applying the chain rule to the recursion formula (13), i.e.,

$$M_\ell'(\theta) = \begin{pmatrix} \left(U_1(\ell)M_{\ell-1}(\theta) + E_1(TR_\ell, \theta)R(\alpha_\ell)M_{\ell-1,1}'(\theta) - \frac{TR_\ell}{T_1^2}e^{-\frac{TR_\ell}{T_1}}M_e\right)^\top \\ \left(U_2(\ell)M_{\ell-1}(\theta) + E_1(TR_\ell, \theta)R(\alpha_\ell)M_{\ell-1,2}'(\theta)\right)^\top \end{pmatrix}^\top . \qquad (20)$$

We get the boundedness of the derivatives because all the quantities $U_a(\ell)$, $R(\alpha_\ell)$, $E_1(TR_\ell, \theta)$ and $e^{-\frac{TR_\ell}{T_a}}\frac{TR_\ell}{T_a^2}$ for $a = 1, 2$ and $\ell = 1, \ldots, L$ are uniformly bounded. If in addition we have $M_0 \in [L^\infty(\Omega)]^3$, the iteration (14) will assure that $M_\ell \in [L^\infty(\Omega)]^3$. Then the estimate (16) immediately follows.

(ii) We show that the map $M_1 : [L^\infty(\Omega)]^2 \to \mathcal{Y}$ is injective for some non-zero α_1 and TR_1. We first note that $R = R(\alpha_1) := R_\phi R_x(\alpha_1)R_\phi^\top$ is unitary, and $E_1(TR_1, \theta)$ and $E_2(TR_1, \theta)$ are contraction operators. Assume now that $M_1(\theta^a) = M_1(\theta^b)$ for $\theta^a \ne \theta^b$. Then we have

$$\big(E_1(TR_1, \theta^a) - E_1(TR_1, \theta^b)\big)RM_0 + (E_2(TR_1, \theta^a) - E_2(TR_1, \theta^b))M_e = 0 \quad \text{in } \mathcal{Y}.$$

Assume further that $T_1^a \ne T_1^b$, then in those points of Ω where this occurs we have (suppressing spatial dependence r)

$$M_e = \begin{pmatrix} \frac{e^{-\frac{TR_1}{T_2^a}} - e^{-\frac{TR_1}{T_2^b}}}{e^{-\frac{TR_1}{T_1^a}} - e^{-\frac{TR_1}{T_1^b}}} & 0 & 0 \\ 0 & \frac{e^{-\frac{TR_1}{T_2^a}} - e^{-\frac{TR_1}{T_2^b}}}{e^{-\frac{TR_1}{T_1^a}} - e^{-\frac{TR_1}{T_1^b}}} & 0 \\ 0 & 0 & \frac{e^{-\frac{TR_1}{T_1^a}} - e^{-\frac{TR_1}{T_1^b}}}{e^{-\frac{TR_1}{T_1^a}} - e^{-\frac{TR_1}{T_1^b}}} \end{pmatrix} RM_0. \qquad (21)$$

Suppose now that $M_0 = -M_e$. Then, since R is unitary, (21) is satisfied if and only if $T_1^a = T_2^a$, $T_1^b = T_2^b$, $-RM_e = M_e$ and $\alpha_1 = \pi$. This, however, contradicts $\alpha \in (0, \pi)$. The case $M_0 = M_e$ is similar. If $T_1^a = T_1^b$ but $T_2^a \ne T_2^b$, then one uses the inverse relation of (21) and arrives at the same conclusion. Thus we have injectivity for M_1 and hence also of M. $\qquad\square$

Regarding non-convexity of the Bloch manifold we have the following result.

Proposition 4.2. *Suppose that the assumptions of Proposition 4.1 hold true. Furthermore, let the operator M be restricted to some feasible set $\mathcal{C}_{ad}$ which is connected and convex:*

$$\mathcal{C}_{ad} := \{\theta \in [L^\infty(\Omega)]^2 \mid \theta(r) \in C_{ad}, \text{ for every } r \in \Omega\},$$

where C_{ad} is a convex subset of $\mathbb{R}^+ \times \mathbb{R}^+$ (typically a box) which is bounded and bounded away from zero. Then the image $M[\mathcal{C}_{ad}]$ of $M : \mathcal{C}_{ad} \to \mathcal{Y}^L$ is a non-convex subset of $\mathcal{Y}^L$.

Proof. Suppose that $M[\mathcal{C}_{ad}]$ is a convex subset of $\mathcal{Y}^L$. Then, for arbitrary $\theta^a \neq \theta^b \in \mathcal{C}_{ad}$ and for every $\lambda \in (0,1)$, there exist $\theta^\lambda \in \mathcal{C}_{ad}$ such that

$$\lambda M_\ell(\theta^a) + (1-\lambda)M_\ell(\theta^b) = M_\ell(\theta^\lambda) \text{ for all } \ell \in \{1,\ldots,L\}. \tag{22}$$

We focus on the first two components $M_1, M_2 \in \mathcal{Y}$ and recall

$$M_1(\theta) = E_1(TR_1, \theta)R(\alpha_1)M_0 + E_2(TR_1, \theta)M_e,$$

$$M_2(\theta) = \left(\prod_{k=1}^2 E_1(TR_k, \theta)R(\alpha_k)\right)M_0$$
$$+ \left(E_2(TR_1, \theta) + E_2(TR_2, \theta)E_1(TR_1, \theta)R(\alpha_1)\right)M_e.$$

Upon some straightforward calculations, the convexity condition (22) can be equivalently written as the following system of equations (where $\theta^a = (T_1^a, T_2^a)$, $\theta^b = (T_1^b, T_2^b)$, $\theta^\lambda = (T_1^\lambda, T_2^\lambda)$):

$$\lambda e^{-\frac{TR_1}{T_1^a}} + (1-\lambda)e^{-\frac{TR_1}{T_1^b}} = e^{-\frac{TR_1}{T_1^\lambda}}, \tag{23}$$

$$\lambda e^{-\frac{TR_1}{T_2^a}} + (1-\lambda)e^{-\frac{TR_1}{T_2^b}} = e^{-\frac{TR_1}{T_2^\lambda}}, \tag{24}$$

$$\lambda e^{-\frac{TR_2}{T_1^a}}e^{-\frac{TR_1}{T_1^a}} + (1-\lambda)e^{-\frac{TR_2}{T_1^b}}e^{-\frac{TR_1}{T_1^b}} = e^{-\frac{TR_2}{T_1^\lambda}}e^{-\frac{TR_1}{T_1^\lambda}}, \tag{25}$$

$$\lambda e^{-\frac{TR_2}{T_2^a}}e^{-\frac{TR_1}{T_2^a}} + (1-\lambda)e^{-\frac{TR_2}{T_2^b}}e^{-\frac{TR_1}{T_2^b}} = e^{-\frac{TR_2}{T_2^\lambda}}e^{-\frac{TR_1}{T_2^\lambda}}, \tag{26}$$

$$\lambda e^{-\frac{TR_1}{T_1^a}}\left(1 - e^{-\frac{TR_2}{T_1^a}}\right) + (1-\lambda)e^{-\frac{TR_1}{T_1^b}}\left(1 - e^{-\frac{TR_2}{T_1^b}}\right) = e^{-\frac{TR_1}{T_1^\lambda}}\left(1 - e^{-\frac{TR_2}{T_1^\lambda}}\right), \tag{27}$$

$$\lambda e^{-\frac{TR_1}{T_2^a}}\left(1 - e^{-\frac{TR_2}{T_1^a}}\right) + (1-\lambda)e^{-\frac{TR_1}{T_2^b}}\left(1 - e^{-\frac{TR_2}{T_1^b}}\right) = e^{-\frac{TR_1}{T_2^\lambda}}\left(1 - e^{-\frac{TR_2}{T_1^\lambda}}\right), \tag{28}$$

where (23)–(24) come from (22) for M_1, and (25)–(28) from (22) for M_2.

Simplifying we get:

$$\lambda e^{-\frac{TR_1}{T_1^a}} + (1-\lambda)e^{-\frac{TR_1}{T_1^b}} = e^{-\frac{TR_1}{T_1^\lambda}},$$

$$\lambda e^{-\frac{TR_1}{T_2^a}} + (1-\lambda)e^{-\frac{TR_1}{T_2^b}} = e^{-\frac{TR_1}{T_2^\lambda}},$$

$$\lambda e^{-\frac{TR_2+TR_1}{T_1^a}} + (1-\lambda)e^{-\frac{TR_2+TR_1}{T_1^b}} = e^{-\frac{TR_2+TR_1}{T_1^\lambda}},$$

$$\lambda e^{-\frac{TR_2+TR_1}{T_2^a}} + (1-\lambda)e^{-\frac{TR_2+TR_1}{T_2^b}} = e^{-\frac{TR_2+TR_1}{T_2^\lambda}},$$

$$\lambda e^{-\frac{TR_1}{T_2^a}}e^{-\frac{TR_2}{T_1^a}} + (1-\lambda)e^{-\frac{TR_1}{T_2^b}}e^{-\frac{TR_2}{T_1^b}} = e^{-\frac{TR_1}{T_2^\lambda}}e^{-\frac{TR_2}{T_1^\lambda}}.$$

Since $TR > 0$, this system has a solution only if $\theta^a = \theta^b = \theta^\lambda$ which gives a contradiction. $\square$

The asserted non-convexity in Proposition 4.2 yields a disadvantage for methods based on projections of the magnetization reconstruction sequences onto the Bloch manifold, as projections need no longer be unique. One specific instance of such a method is BLIP.

Concerning the MRI data, it is quite nature to assume $D \in ([L^2(K)]^2)^L$ where K denotes a compact frequency domain. Now we are in a position to show the Fréchet differentiability of the qMRI-operator.

Lemma 4.3. *Let* $\mathbf{x} = (\rho, \theta) \in \tilde{V} \subset L^\infty(\Omega) \times [L^\infty(\Omega)]^2$, *where* $\tilde{V}$ *is the open subset with functions of strictly positive values only. Then the qMRI-operator*

$$Q : \tilde{V} \to \left([L^2(K)]^2\right)^L$$

is Fréchet differentiable. Similarly, given $M_0 \in [L^\infty(\Omega)]^3$ *we have the following general estimate for sufficiently small* $\mathbf{h} \in L^\infty(\Omega) \times [L^\infty(\Omega)]^2$:

$$\left\| Q(\mathbf{x} + \mathbf{h}) - Q(\mathbf{x}) - Q'(\mathbf{x})\mathbf{h} \right\|_{([L^2(K)]^2)^L} = o\left(\|\mathbf{h}\|_{[L^2(\Omega)]^3} \right). \tag{29}$$

Proof. Recall that

$$Q(\mathbf{x}) = \{Q^{(\ell)}(\mathbf{x})\}_{\ell=1}^L = \left\{ P^{(\ell)} \mathcal{F}(\rho T_{x,y} M_\ell(\theta)) \right\}_{\ell=1}^L.$$

To see the Fréchet differentiability, we first notice that all $P^{(\ell)}$, $\mathcal{F}$ and $T_{x,y}$ are bounded, linear operators. Then we consider

$$
\begin{aligned}
Q^{(\ell)}(\mathbf{x} + \mathbf{h}) - Q^{(\ell)}(\mathbf{x}) =& P^{(\ell)} \mathcal{F}((\rho + h_\rho) T_{x,y} M_\ell(\theta + h_\theta)) - P^{(\ell)} \mathcal{F}(\rho T_{x,y} M_\ell(\theta + h_\theta)) \\
&+ P^{(\ell)} \mathcal{F}\left(\rho T_{x,y}\left(M_\ell(\theta + h_\theta) - M_\ell(\theta)\right)\right).
\end{aligned}
$$

For every $\mathbf{x} = (\rho, \theta) \in \tilde{V}$, and $\mathbf{h} = (h_\rho, h_\theta) \in L^\infty(\Omega) \times [L^\infty(\Omega)]^2$ small enough, applying the Fréchet differentiability of each M_ℓ from Proposition 4.1, and using the estimates (15) and (16), we get the estimates below:

$$
\begin{aligned}
& \left\| P^{(\ell)} \mathcal{F}((\rho + h_\rho) T_{x,y} M_\ell(\theta + h_\theta)) - P^{(\ell)} \mathcal{F}(\rho T_{x,y} M_\ell(\theta + h_\theta)) \right\|_{[L^2(K)]^2} \\
=& \left\| P^{(\ell)} \mathcal{F}(h_\rho T_{x,y} M_\ell(\theta + h_\theta)) \right\|_{[L^2(K)]^2} \\
\leq& \left\| P^{(\ell)} \mathcal{F}(h_\rho T_{x,y} M_\ell(\theta)) \right\|_{[L^2(K)]^2} + C \|h_\rho h_\theta\|_{[L^2(\Omega)]^2} + o(\|h_\theta\|_{[L^2(\Omega)]^2})
\end{aligned}
$$

and

$$
\begin{aligned}
\left\| P^{(\ell)} \mathcal{F}\left(\rho T_{x,y}\left(M_\ell(\theta + h_\theta) - M_\ell(\theta)\right)\right) \right\|_{[L^2(K)]^2} \leq& \left\| P^{(\ell)} \mathcal{F}(\rho T_{x,y} M_\ell'(\theta) h_\theta) \right\|_{[L^2(K)]^2} \\
&+ o(\|h_\theta\|_{[L^2(\Omega)]^2}).
\end{aligned}
$$

The two inequalities indicate for sufficiently small $\mathbf{h} \in L^\infty(\Omega) \times [L^\infty(\Omega)]^2$

$$\left\| Q^{(\ell)}(\mathbf{x} + \mathbf{h}) - Q^{(\ell)}(\mathbf{x}) - A^{(\ell)} \mathbf{h} \right\|_{[L^2(K)]^2} = o\left(\|\mathbf{h}\|_{[L^2(\Omega)]^3} \right), \tag{30}$$

where

$$
\begin{aligned}
A^{(\ell)} &: [L^\infty(\Omega)]^3 &\to& \quad [L^2(K)]^2, \\
&: \mathbf{h} &\mapsto& \quad P^{(\ell)} \mathcal{F}(h_\rho T_{x,y} M_\ell(\theta)) + P^{(\ell)} \mathcal{F}(\rho T_{x,y} M_\ell'(\theta) h_\theta),
\end{aligned}
$$

is a bounded linear operator. Using (30) and the fact that

$$\|\mathbf{h}\|_{[L^2(\Omega)]^3} \leq C \|\mathbf{h}\|_{[L^\infty(\Omega)]^3}$$

we show that $Q^{(\ell)}$ is Fréchet differentiable, and $A^{(\ell)}$ is the derivative. The derivative of Q is obtained from derivatives of each $Q^{(\ell)}$ for $\ell \in \{1, \ldots, L\}$. Finally the estimate (29) is obtained from (30). □

The above proof also presents a way of how to calculate the derivative of Q.

Remark 4.4. *The estimates* (15) *and* (29) *do not guarantee the differentiability of M and Q in the whole space $[L^2(\Omega)]^3$ and $[L^2(\Omega)]^3$, respectively. However, the Fréchet derivative $Q'(\mathbf{x})$ at $\mathbf{x} \in \tilde{V}$ as an operator from $[L^2(\Omega)]^3$ to $([L^2(K)]^2)^L$ is well defined given $M_0 \in [L^\infty(\Omega)]^3$ which is applicable in practice.*

4.3 (Gauss-) Newton method for ideal data

Next we turn towards iterative methods for computing a solution to the non-linear equation

$$Q(\mathbf{x}) = D, \tag{31}$$

which in fact represents a system of equations

$$P^{(\ell)}\mathcal{F}(\rho T_{x,y}M_\ell(\theta)) = D^{(\ell)}, \quad \ell = 1, \ldots, L.$$

Because of the regularity of the operator Q, a first idea to solve the non-linear operator equation (31) is using a Gauss-Newton method, which, given some approximate solution $\mathbf{x}_n$, is based on the first-order approximation

$$Q(\mathbf{x}_{n+1}) \simeq Q(\mathbf{x}_n) + Q'(\mathbf{x}_n)(\mathbf{x}_{n+1} - \mathbf{x}_n) = D. \tag{32}$$

By letting $D_n := D - Q(\mathbf{x}_n) + Q'(\mathbf{x}_n)\mathbf{x}_n$, (32) becomes

$$Q'(\mathbf{x}_n)\mathbf{x}_{n+1} - D_n = 0. \tag{33}$$

Note that since $D = \{D^{(\ell)}\}_{\ell=1}^L$ is a sequence of data frames of length L, so is D_n. Typically, we have $L \geq 2$. Therefore (the space discrete version of) (33) in general contains redundant equations. Thus, one considers (33) in a least-squares sense. Taking into account also the physical constraint of the tissue parameters, we introduce the feasible set $\tilde{\mathcal{C}}_{ad} \subset [L^\infty(\Omega)]^3$ which is a connected and convex set (typically a box) and contains all feasible values for $\mathbf{x} = (\rho, \theta)$. Finally, it leads to computing $\mathbf{x}_{n+1}$ by solving

$$\mathbf{x}_{n+1} = \underset{\mathbf{x} \in \tilde{\mathcal{C}}_{ad}}{\operatorname{argmin}} \|Q'(\mathbf{x}_n)\mathbf{x} - D_n\|^2_{([L^2(K)]^2)^L}, \quad n = 0, 1, 2 \ldots \tag{34}$$

The solution of the problem in (34) can be approximated by a projection step to $\tilde{\mathcal{C}}_{ad}$, resulting to the following projected Gauss-Newton iteration:

$$D_n = D - Q(\mathbf{x}_n) + Q'(\mathbf{x}_n)\mathbf{x}_n, \tag{35}$$

$$\mathbf{y}_{n+1} = (Q')^\dagger(\mathbf{x}_n)D_n := \left((Q'(\mathbf{x}_n))^\top Q'(\mathbf{x}_n)\right)^{-1}(Q'(\mathbf{x}_n))^\top D_n, \tag{36}$$

$$\mathbf{x}_{n+1} = P_{\tilde{\mathcal{C}}_{ad}}\mathbf{y}_{n+1}. \tag{37}$$

We point out that the step in (36) is regarded in a Hilbert space setting, i.e.,

$$Q'(\mathbf{x}_n) : [L^2(\Omega)]^3 \to \left([L^2(K)]^2\right)^L, \quad \text{for } n \in \mathbb{N},$$

and $(Q'(\mathbf{x}_n))^\top$ is the Hermitian adjoint of the linear operator $Q'(\mathbf{x}_n)$. This can be done since, as we have mentioned that $Q'(\mathbf{x}_n)$ is a well defined linear operator for functions in $[L^2(\Omega)]^3$, and (36) will give a solution $\mathbf{y}_{n+1} \in [L^2(\Omega)]^3$. The subsequent projection step (37) assures that $\mathbf{x}_{n+1} \in \tilde{\mathcal{C}}_{ad} \subset$

$[L^\infty(\Omega)]^3$. Supposing that $\tilde{\mathcal{C}}_{ad} := \{\mathbf{x} \in [L^2(\Omega)]^3 : x_p(r) \in [\underline{C}_p, \overline{C}_p] \text{ for } p \in \{1,2,3\} \text{ a.e. } r \in \Omega\}$ for $\underline{C}, \overline{C} \in \mathbb{R}^3$ with $\underline{C}_p < \overline{C}_p$ for $p \in \{1,2,3\}$ and $\mathbf{x} = (x_1, x_2, x_3)^\top$, the projection can be realised by

$$(P_{\tilde{\mathcal{C}}_{ad}}\mathbf{x})_p(r) = \begin{cases} \underline{C}_p & \text{for } x_p(r) \leq \underline{C}_p, \\ x_p(r) & \text{for } \underline{C}_p < x_p(r) < \overline{C}_p, \\ \overline{C}_p & \text{for } \overline{C}_p \leq x_p(r) \end{cases} \tag{38}$$

for every $r \in \Omega$. Different from the projection in BLIP algorithm, the projection in (38) is uniquely defined because of the convexity of the feasible domain. In particular, for an exact solution $\mathbf{x}^*$ of (31), we assume that $\mathbf{x}^* \in \tilde{\mathcal{C}}_{ad}$. It is obvious that the non-expansiveness holds for the projection operator:

$$\|\mathbf{x}_{n+1} - \mathbf{x}^*\|_{[L^2(\Omega)]^3} \leq \|\mathbf{y}_{n+1} - \mathbf{x}^*\|_{[L^2(\Omega)]^3}. \tag{39}$$

We show here the superlinear convergence rate of the projected Gauss-Newton iteration (35)–(37) given the Fréchet differentiability of the non-linear operator Q and the general estimate (29). The proof is based on a classical result for unconstrained problems, see, e.g., [15].

Theorem 4.5. *Let $\mathbf{x}^* \in \tilde{\mathcal{C}}_{ad}$ be an exact solution of* (31)*, and assume there exists a neighbourhood $N(\mathbf{x}^*) \subset [L^\infty(\Omega)]^3$ of $\mathbf{x}^*$ such that $(Q')^\dagger(\mathbf{x})$ is uniformly bounded for all $\mathbf{x} \in N(\mathbf{x}^*)$. Then there exists a potentially smaller neighbourhood such that for every initial guess $\mathbf{x}_0$ belonging there, for the iterates in* (36) *and* (37) *we have that $\mathbf{x}_n \to \mathbf{x}^*$ with a superlinear rate of convergence, i.e.,*

$$\|\mathbf{x}_{n+1} - \mathbf{x}^*\|_{[L^2(\Omega)]^3} = o\left(\|\mathbf{x}_n - \mathbf{x}^*\|_{[L^2(\Omega)]^3}\right) \quad \text{for all } n \in \mathbb{N}. \tag{40}$$

Proof. Since $D = Q(\mathbf{x}^*)$, for every $\mathbf{x}_n \in N(\mathbf{x}^*)$, the forms in (35) and (36) give us

$$\|\mathbf{y}_{n+1} - \mathbf{x}^*\|_{[L^2(\Omega)]^3} \leq \left\|(Q')^\dagger(\mathbf{x}_n)\right\| \left\|Q(\mathbf{x}^*) - Q(\mathbf{x}_n) - Q'(\mathbf{x}_n)(\mathbf{x}^* - \mathbf{x}_n)\right\|_{([L^2(K)]^2)^L}. \tag{41}$$

Let $B(\mathbf{x}^*, \varepsilon)$ be an $[L^2(\Omega)]^3$ ball of radius ε around $\mathbf{x}^*$, then there exist $W > 0$ such that for every $\mathbf{x} \in B(\mathbf{x}^*, \varepsilon) \cap N(\mathbf{x}^*)$ it holds that $\left\|(Q')^\dagger(\mathbf{x})\right\| \leq W$. Using the estimate (29), for those $\mathbf{x} \in B(\mathbf{x}^*, \varepsilon) \cap N(\mathbf{x}^*)$ where ε small enough, we have

$$\|Q(\mathbf{x}^*) - Q(\mathbf{x}) - Q'(\mathbf{x})(\mathbf{x}^* - \mathbf{x})\|_{([L^2(K)]^2)^L} = o\left(\|\mathbf{x}^* - \mathbf{x})\|_{[L^2(\Omega)]^3}\right). \tag{42}$$

Let $\eta \in (0,1]$ be arbitrary, because of (42), we can find $\rho \in (0, \varepsilon)$, such that for all $\mathbf{h} \in [L^\infty(\Omega)]^3$, and $\|\mathbf{h}\|_{[L^2(\Omega)]^3} < \rho$:

$$\|Q(\mathbf{x}^*) - Q(\mathbf{x}^* + \mathbf{h}) + Q'(\mathbf{x}^* + \mathbf{h})\mathbf{h}\|_{([L^2(K)]^2)^L} < \frac{\eta}{W}\|\mathbf{h}\|_{[L^2(\Omega)]^3} \leq \frac{\|\mathbf{h}\|_{[L^2(\Omega)]^3}}{W}.$$

This shows that if $\mathbf{x}_0 \in B(\mathbf{x}^*, \rho) \cap N(\mathbf{x}^*)$, when applying the iteration steps (36) and (37), the non-expansive property (39) and the estimate (41) gives $\{\mathbf{x}_n\}_{n \in \mathbb{N}} \subset B(\mathbf{x}^*, \rho) \cap N(\mathbf{x}^*)$. Therefore the iteration (36) is always well defined. Now by replacing $\mathbf{x}$ by $\mathbf{x}_n$ in (42) and using the relation (41) and (39), then we get the final relation:

$$\|\mathbf{x}_{n+1} - \mathbf{x}^*\|_{[L^2(\Omega)]^3} = o\left(\|\mathbf{x}^* - \mathbf{x}_n)\|_{[L^2(\Omega)]^3}\right).$$

$\square$

However, due to the non-linearity of the map Q and non-convexity of $Q(\tilde{\mathcal{C}}_{ad})$, the iteration in (36) will only converge for initial values $\mathbf{x}_0 \in \tilde{\mathcal{C}}_{ad}$ in a certain neighbourhood of the exact solution $\mathbf{x}^*$, provided that the data D contains no noise.

For undersampled and noisy data, it is even crucial to choose a good initial guess in order to obtain a robust and efficient numerical algorithm for solving the problem (11). This would be the main task of the next section.

4.4 A projected Levenberg-Marquardt method for undersampled and noisy data

Undersampling is often unavoidable in the acquisition process of MRI due to time constraints. The main problem caused by undersampling is ill-posedness of the equation (11) due to the properties of the operator P composed into the qMRI-operator Q. As a consequence, the solution of (11) may be unreliable, even when the data is contaminated by noise of small intensity.

In order to address the problem of undersampling and noise, and to solve (11) robustly, we turn to a projected *Levenberg-Marquardt* (L-M) method instead of the aforementioned projected Gauss-Newton scheme. It essentially relies on a Tikhonov type regularization. Suppose that the ideal data D has been corrupted by some noise, leading to perturbed data D^δ. Then the projected L-M iteration reads: Given $\mathbf{x}_0 \in \tilde{\mathcal{C}}_{ad}$ and a sequence $\{\lambda_n\}_{n \in \mathbb{N}}$ of positive real numbers, iterate for $n = 0, 1, 2, \ldots$:

$$\tilde{D}_n^\delta = D^\delta - Q(\mathbf{x}_n), \tag{43}$$

$$\mathbf{h}_n^\delta = \operatorname*{argmin}_{\mathbf{h}} \ \left\| Q'(\mathbf{x}_n)\mathbf{h} - \tilde{D}_n^\delta \right\|_{([L^2(K)]^2)^L}^2 + \lambda_n \left\| \mathbf{h} \right\|_{[L^2(\Omega)]^3}^2, \tag{44}$$

$$\mathbf{x}_{n+1} = P_{\tilde{\mathcal{C}}_{ad}}(\mathbf{x}_n + \mathbf{h}_n^\delta). \tag{45}$$

where $P_{\tilde{\mathcal{C}}_{ad}}$ is the projection as defined in (38).

From a regularization point of view, the L-M iteration (44) is nothing else but an iterative Tikhonov regularization for solving a non-linear equation [12, 16]. Note that if $\lambda_n = 0$ for every n, then the L-M method becomes a Gauss-Newton method. The convergence and convergence rates of L-M methods in the sense of regularization have been shown in many works; see, e.g., [12]. There, general rules of choosing the parameter of a form $\lambda_n = \lambda_0 \beta^n$ for some $\lambda_0 > 0$, $\beta \in (0, 1)$ are discussed, as well as a discrepancy principle of terminating the iterations at step $n = n_e$ where n_e is the first iteration index such that the condition

$$\left\| Q(\mathbf{x}_{n_e}) - D^\delta \right\|_{([L^2(K)]^2)^L} \leq \varrho\delta$$

holds. It is also shown that with these choices, the solution of the L-M method converges to a solution of the original non-linear equation. In our case this yields $\mathbf{x}_{n_e} \to \mathbf{x}^*$ as $\delta \to 0$.

The local and global convergence as well as rates of convergence of (projected) L-M algorithms have also been intensively studied; we refer to [7, 9, 32, 17] for instance. In the absence of additive noise and with proper initial values, the optimal convergence rates of the L-M algorithm are determined by the rates of the updated parameters λ_n, i.e., $\|\mathbf{x}_n - \mathbf{x}^*\|_{[L^2(\Omega)]^3} = \mathcal{O}(\lambda_n)$. In [17], quadratic convergence rate of projected L-M algorithm for convex constraint has been proved in finite dimensional spaces. For non-zero residual problems, i.e., in the presence of additive noise, a standard L-M method with no projection usually only achieves a linear convergence rate $\|\mathbf{x}_{n+1} - \mathbf{x}_n\|_{[L^2(\Omega)]^3} \leq C \|\mathbf{x}_n - \mathbf{x}_{n-1}\|_{[L^2(\Omega)]^3}$ for some constant $C < 1$. With an additional convex constraint, in the case of non-zero residual problem, we expect that the projected L-M method will keep the convergence rate as the non-projected L-M for unconstrained problems, even though the convergence result seems to be more complicated than the zero residual problem. We ignore the discussion in detail in this paper.

As for the (projected) Gauss-Newton iteration, initialization is crucial for the (projected) L-M method. Unfortunately, there is no general way to produce good initial guesses, rather this is a problem-dependent task. Here we suggest to use a very fast version of MRF and BLIP in order to produce initial points in a neighbourhood of a solution. The low run-time of the initialization scheme is related to using a relatively coarse dictionary only. In this way, the dictionary is no longer refined in the L-M iterations. Having clarified this, our main proposed algorithm is summarized in Algorithm 2.

Algorithm 2. *Projected L-M iteration with MRF/BLIP-based initialization*

- *Input and setting:*

  - *MRI data $D^\delta \in ([L^2(K)]^2)^L$;*

  - *Parameters for the physical setting of MRI, e.g., flip angle and repetition time sequences, $\{\alpha_\ell\}_{\ell=1}^L, \{TR_\ell\}_{\ell=1}^L$;*

  - *A coarse discretization of the set $C_{ad} = [T_1^{\min}, T_1^{\max}] \times [T_2^{\min}, T_2^{\max}]$.*

- *Initialization:*

  - *Generate a dictionary $\mathrm{Dic}(C_{ad})$, using the coarse discretization of C_{ad}, the flip angles and the repetition times, with the help of the IR-bSSFP, formula (14) for magnetization;*

  - *Use the BLIP method, Algorithm 1, to produce an initialization: $\mathbf{x}_0 = (\rho_0, \theta_0) \in \tilde{\mathcal{C}}_{ad} \subset [L^\infty(\Omega)]^3$;*

  - *Choose an initial parameter $\lambda_0 \geq 1$.*

- *Projected L-M iteration:*

  - *(1) Do the projected L-M iteration step (43)–(45);*

  - *(2) If stopping criteria are not fulfilled, set $n \leftarrow n+1$, update $\lambda_n = \max\{\lambda_0\beta^n, \mu_n\}$, where $\beta \in (0,1)$, and $\mu_n \geq 0$ and go back to (1); otherwise, give the output.*

- *Output: The estimated parameter map $\mathbf{x}_{n_e} = (\rho_{n_e}, \theta_{n_e})$, for some final iteration index n_e.*

There $(\mu_n)_{n\in\mathbb{N}}$ is a sequence of parameters that depend on the noise level in the data, and λ_0 depends on the sub-sampling rate. In our numerical examples below, we set $\lambda_0 = s^2$, where $1/s$ is the undersampling rate of the data. A typical choice for μ_n is $\mu_n = \epsilon \left\| Q(\mathbf{x}_n) - D^\delta \right\|_{([L^2(K)]^2)^L}$ where $\epsilon \in (0,1)$.

4.5 Why more data frames can help

In the original MRF approach, in order to handle the problem of noisy data, the use of a large number L of consecutive pulse sequences and acquisitions is proposed. Conceptionally, this technique should average out noise and thus support better reconstructions. We borrow this idea here and justify it theoretically in what follows. In this part we consider problems after discretization, that is, in finite dimensional spaces.

For this purpose, we first recall the so called Chebyshev's inequality for vector-valued random variables (see e.g. [10, 26]). In its formulation, $\mathbb{P}(\cdot)$ stands for the probability of an event and $\| \cdot \|_{\mathbb{R}^p}$ denotes the Euclidean norm in $\mathbb{R}^p$.

Lemma 4.6 (Chebyshev's inequality)**.** *Let $\phi = (\phi_1, \phi_2, \dots, \phi_p)$ be a vector-valued random variable, for some $p \in \mathbb{N}$, with expected value and variance $E(\phi) = \chi = (\chi_1, \chi_2, \dots, \chi_p)$, $V(\phi) = \Sigma^2 = (\sigma_1^2, \sigma_2^2, \dots, \sigma_p^2)$, respectively. Then, for every $\epsilon > 0$, we have*

$$\mathbb{P}(\|\phi - \chi\|_{\mathbb{R}^p} > \epsilon) \leq \frac{\|\Sigma^2\|_{\mathbb{R}^p}}{\epsilon^2}. \tag{46}$$

The following, main theorem of this section, states that if a family of L linear systems has a common solution and the right hand sides are perturbed by noise, then by solving a least-squares problem one can get an approximation of the common solution, with a certain probability that gets improved as the number L increases. Later we shall see how this applies to our proposed algorithm to qMRI.

Theorem 4.7. *Let $A_\ell \zeta = b_\ell$, $\ell = 1, \ldots, L$, be a family of L linear systems of equations, where $\{b_\ell\}_{\ell=1}^L$, with $b_\ell \in \mathbb{R}^p$ for every ℓ, and $\{A_\ell\}_{\ell=1}^L$, with $A_\ell \in \mathbb{R}^{d \times p}$, $p \leq d$, and $\mathrm{rank}(A_\ell) = p$ for every ℓ, are given sequences of data and system matrices, respectively. Assume also that the singular values of all A_ℓ have a uniform lower and upper bound $\sqrt{c}$ and $\sqrt{C}$, respectively, which are both independent of L. Further suppose that this family of equations has a common solution $\zeta^* \in \mathbb{R}^p$. If $\tilde{b}_\ell = b_\ell + \delta_\ell$, where $\{\delta_\ell\}_{\ell=1}^L$ are independently identically distributed (i.i.d.) random variables with expected value $(0, \ldots, 0) \in \mathbb{R}^p$, and variance $(\sigma^2, \ldots, \sigma^2) \in \mathbb{R}^p$, then the least-squares solution*

$$\zeta_{ls} = \underset{\zeta \in \mathbb{R}^p}{\arg\min} \left\| A\zeta - \tilde{b} \right\|_{\mathbb{R}^{Ld}}^2, \tag{47}$$

where

$$A = (A_1, A_2, \cdots, A_L)^\top \quad and \quad \tilde{b} = (\tilde{b}_1, \tilde{b}_2, \cdots, \tilde{b}_L)^\top$$

approximates the solution ζ^* with the following probability estimate

$$\mathbb{P}(\|\zeta_{ls} - \zeta^*\|_{\mathbb{R}^p} > \epsilon) < \frac{\sigma^2}{\epsilon^2} \mathcal{O}\left(\frac{p}{L}\right), \quad \text{for every } \epsilon > 0. \tag{48}$$

Proof. From (47), we get $\zeta_{ls} = (A^\top A)^{-1} A^\top \tilde{b}$, which is also a random variable. Since A is not random, we can compute the expected value of ζ_{ls} as follows:

$$E(\zeta_{ls}) = E((A^\top A)^{-1} A^\top \tilde{b}) = (A^\top A)^{-1} A^\top E(\tilde{b}) = (A^\top A)^{-1} A^\top b = \zeta^*.$$

Therefore $E(\zeta_{ls} - \zeta^*) = 0$. Similarly, for the variance (diagonal of the covariance matrix) we have

$$V(\zeta_{ls} - \zeta^*) = \sigma^2 \mathsf{diag}\left((A^\top A)^{-1}\right),$$

where "diag" denotes the diagonal of a matrix. Denoting by Tr the trace operator, i.e., the summation of the diagonal values and using Lemma 4.6, we get that for every $\epsilon > 0$

$$\mathbb{P}(\|\zeta_{ls} - \zeta^*\|_{\mathbb{R}^p} > \epsilon) < \frac{\sigma^2 \|\mathsf{diag}\left((A^\top A)^{-1}\right)\|_{\mathbb{R}^p}}{\epsilon^2} \leq \frac{\sigma^2 \, \mathrm{Tr}\left((A^\top A)^{-1}\right)}{\epsilon^2}. \tag{49}$$

Here we have used the fact that the matrix $(A^\top A)^{-1}$ is positive definite and hence it has strictly positive diagonal elements, together with the fact that the ℓ_1 norm in $\mathbb{R}^p$ is larger than the Euclidean one.

From the form of A we have $A^\top A = \sum_{\ell=1}^L A_\ell^\top A_\ell$ with trace

$$\mathrm{Tr}(A^\top A) = \sum_{\ell=1}^L \mathrm{Tr}(A_\ell^\top A_\ell).$$

Since every $A_\ell^\top A_\ell$ is positive definite, so is $A^\top A$. Let $\{S_j\}_{j=1}^p$ be the eigenvalues of $A^\top A$ allowing for the decomposition

$$A^\top A = USU^{-1} \quad and \quad (A^\top A)^{-1} = US^{-1}U^{-1}, \tag{50}$$

where S is the diagonal matrix with entries $\{S_j\}_{j=1}^p$, and U is a unitary matrix. Then, for the traces we have

$$\mathrm{Tr}(A^\top A) = \sum_{j=1}^p S_j \quad \text{and} \quad \mathrm{Tr}\left((A^\top A)^{-1}\right) = \sum_{j=1}^p \frac{1}{S_j}.$$

Due to the uniform lower and upper bounds on the singular values of $\{A_\ell\}_{\ell=1}^L$, we get a corresponding uniform bound on the eigenvalues of the matrices $\{A_\ell^\top A_\ell\}_{\ell=1}^L$, i.e.,

$$cL \leq S_j \leq CL, \quad \text{for all} \quad j = 1, 2, \ldots, p.$$

Consequently, we have

$$\frac{1}{CL} \leq \frac{1}{S_j} \leq \frac{1}{cL} \quad \Longrightarrow \quad \frac{1}{S_j} = \mathcal{O}\left(\frac{1}{L}\right), \quad \text{for} \quad j = 1, 2, \ldots, p.$$

From this we infer the following estimate

$$\mathrm{Tr}\left((A^\top A)^{-1}\right) = \sum_{j=1}^p \frac{1}{S_j} = \mathcal{O}\left(\frac{p}{L}\right),$$

and combined with (49) it proves the assertion

$$\mathbb{P}(\|\zeta_{ls} - \zeta^*\|_{\mathbb{R}^p} > \epsilon) = \frac{\sigma^2}{\epsilon^2}\mathcal{O}\left(\frac{p}{L}\right).$$

$\square$

Theorem 4.7 relates to our qMRI algorithm in several ways:

(i) Observe that regarding the setting of qMRI problems, the noise in the data obtained after each pulse sequence can be considered as realisations of i.i.d. random variables.

(ii) In Newton-type methods, if there is no sub-sampling for the qMRI-operator, then Q' is non-degenerate on the effective domain Ω, i.e., on the part of the slices where the biological tissue is imaged. In this case, we may consider $A_\ell = (Q^{(\ell)})'$, and $b_\ell = (D_k^{(\ell)})^\delta$, with both quantities satisfying the assumptions of Theorem 4.7 given that the data contains Gaussian noise. This indicates that the result of Theorem 4.7 can be applied to every Newton-type step for a given $\epsilon > 0$, and an initial value $\|\mathbf{x}_0 - \mathbf{x}^*\| \leq \epsilon$. Since we have restricted to a small neighbourhood of the exact solution $\mathbf{x}^*$, we can take roughly the common solution ζ^* corresponding to (47) of Theorem 4.7 as the exact solution of the least-squares problem (34).

(iii) In the case of the Levenberg-Marquardt method, with the sub-sampling operators $P^{(\ell)}$, the results of Theorem 4.7 can still be applied as the involved matrices become $A_\ell = \left(((Q^{(\ell)})')^\top, \sqrt{\lambda_n}\mathrm{Id}\right)^\top$, and $b_\ell = \left(((\tilde{D}_n^{(\ell)})^\delta)^\top, 0\right)^\top$. Note that in this case ζ corresponds to $\mathbf{h}$. Further the matrices A_ℓ will always be of full rank with uniformly bounded singular values, whenever we let the sequence $(\lambda_n)_{n \in \mathbb{N}}$ be uniformly bounded away from zero. Such a uniform lower bound is indeed usually in place at the presence of noise. In such a case, we can treat $\zeta^* \equiv 0$ as the common solution of (44).

5 Numerical results

Now we report on numerical results obtained by our Algorithm 2 when applied to synthetic data. Our setting also allows for an extensive quantitative comparison with Algorithm 1 (BLIP), which was shown in [6] to be superior to the original MRF.

5.1 Generating test data

Our tests are based on synthetic data from an anatomical brain phantom, publicly available from the Brain Web Simulated Brain Database [1, 5]. We use a 217×181 slice completed by zero fill-in order to generate a 256×256-pixel image. The selected ranges for $\theta = (T_1, T_2)^\top$ and ρ reflect natural values encountered in the human body [6], with T_1 ranging from $530ms$–$5012ms$, T_2 from $41ms$–$512ms$, and ρ between 80–100. As pixel units in practical images very likely contain multiple tissue types rather than only a pure one in a single volume of the observed pixels, we interpolate the values of each parameter T_1, T_2 and ρ of the 256×256 phantom, respectively, by averaging the values of every four neighboured pixels with non-zero parameter values. This average process shrinks the 256×256 image to a 128×128 image. In Figure 1, we display the interpolated parameters of T_1, T_2 and ρ as coloured images. These serve as the ground truth for our numerical tests.
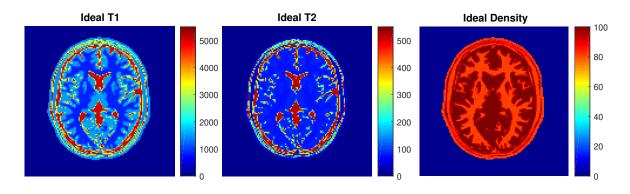

Figure 1: The interpolated parameters serve as the ground truth for our algorithm. From left to right: T_1, T_2 and ρ.

The IR-bSSFP pulse sequence scheme introduced in Section 4.1 is applied to generate MRI data. It is based on constant flip angles α and repetition time TR sequences of length L. The data D are generated by using the prescribed parameters T_1, T_2 and ρ with the pulse sequences characterized by α and TR. With this setting, we first simulate the magnetization, and then use FFT to generate the Fourier space data from it. Sub-sampling is implemented by using the scheme described in the next section. We also note that for the generation of the magnetization, we rely on (14) and take advantage of the MATLAB code provided in [23]. For simplicity, we set the phase shift $\phi \equiv 0$ in (14).

5.2 Sub-sampling pattern

Here we focus on Cartesian sub-sampling which is frequently used in practice; see, e.g., [24]. This choice implies a specific form of the sub-sampling operator $P^{(\ell)}$ for $\ell = 1, \ldots, L$. In the discrete setting, the full k-space data are given by a dense matrix of complex values or, equivalently, two real-valued dense matrices, respectively of size $N \times N$. According to our set-up above, we have $N = 128$. More specifically, we use here an n multishot Echo Planar Imaging (EPI) scheme, which means that at every read-out time, n rows of k-space are simultaneously filled. Hence, in every acquisition there will be $n < N$ rows of the matrix filled with Fourier coefficients. To simplify the discussion, we consider $(N \mod n) \equiv 0$, and further set $s := N/n$, which gives a sub-sampling rate of $1/s$. The sampling pattern $P^{(\ell)}$ is described in detail as follows:

(i) For every ℓ-th acquisition, define $\xi_\ell := (\ell \mod s)$ for $\ell = 1, \ldots, L$.

(ii) P^ℓ will include those rows of the full k-space matrices, indexed by numbers from the set ι with

$$\iota := \left\{ i \in \{1, \ldots, N\} \ : \ (i \mod s) \equiv \xi_\ell \right\}.$$

Thus, at every read-out time, $P^{(\ell)}$ samples n rows from the full Fourier space to simulate the n multishot EPI. A simple example of such a sub-sampling pattern is shown in Figure 2.

Figure 2: A sub-sampling pattern example for $s = 4$, $L = 4$, and $N = 16$. The first image depicts an example of fully sampled k-space data. The second to the fifth images are example frames of the undersampled data, where the information indicated by blank rows is not collected in that frame. The sub-sampling pattern follows the order periodically if $L > s$.

We note that this sub-sampling strategy differs from the one in [6]. There, ξ_ℓ is defined as a uniformly distributed random number in $\{0, 1, \ldots, s - 1\}$, whereas here we consider a deterministic periodical order. After experimentation and when compared to the pseudo-random strategy of [6], we found that the deterministic order is more stable and gives comparable or better results when the BLIP algorithm for Cartesian sub-sampled data is used. Therefore, in our implementation of the BLIP algorithm we use the deterministic strategy as described above.

We consider different sub-sampling rates on the data using the above strategy. By taking into account that longer processing time is needed for higher sampling rates, it follows that the flip angles and the repetition times must be increased proportionally. Accordingly, we use the following repetition times $TR = (TR_1, TR_2, \ldots, TR_L)$ and flip angles $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_L)$:

(a) Fully sampled data: Repetition time $TR_\ell = 40ms$ and flip angles $\alpha_\ell = \frac{40\pi}{180}$ for all $\ell = 1, \ldots, L$.

(b) $1/4$ sampled data (sub-sampling rate 25%), e.g., a 32 multi-shot EPI: Repetition time $TR_\ell = 20ms$ and flip angles $\alpha_\ell = \frac{20\pi}{180}$ for all $\ell = 1, \ldots, L$.

(c) $1/8$ sampled data (sub-sampling rate 12.5%), e.g., a 16 multi-shot EPI: a shorter repetition time $TR_\ell = 10ms$ and smaller flip angles $\alpha_\ell = \frac{10\pi}{180}$ are applied for all $\ell = 1, \ldots, L$.

It is reported in the literature [23] that adding random noise to the flip angles and to repetition times may improve the final results of MRF (and BLIP). However, in our experiments we did not find significant differences. Therefore, we do not add noise to the angles and repetition times in our numerical tests.

In order to compare our method and the BLIP algorithm, we used for the latter a very fine dictionary where T_1 was discretized from $15ms$ to $5500ms$ with increments of $15ms$, and T_2 was discretized

from $1.5ms$ to $550ms$ with increments of $1.5ms$. This means that the discretized feasible domain C_{ad} for $\theta = (T_1, T_2)$ was (in MATLAB notation)

$$C_{ad} = \{[15 : 15 : 5500] \times [1.5 : 1.5 : 550]\}.$$

In this case, the dictionary had $366 \times 366 = 133956$ entries, and it required memory for a storage matrix of dimension $133956 \times L$. The deterministic sub-sampling pattern was used in all numerical examples for the BLIP algorithms, including the generation of the initial values. For the projection onto the feasible set $\tilde{C}_{ad}$ we use the following thresholds for each parameter values: (the constants $\underline{C}$ and $\overline{C}$ here refer to (38))

$$
\begin{array}{ccc}
 & \underline{C} & \overline{C} \\
T_1 & 0 & 5500 \\
T_2 & 0 & 550 \\
\rho & 0 & 100
\end{array}
$$

Note that the value 0 is assigned to the marginal area in the tested images in Figure 1, where there is no tissue information. Further, the linear systems in the L-M iterations were solved by employing MAT-LAB's backslash command. For our test runs, we used a CPU with an Intel Core i5-7500, 3.40GHz, 2 cores, and RAM of 8GB DDR4, 2400 MHz, as well as MATLAB of version 2018a under the operating system openSUSE 42.3.

5.3 Undersampled data with no additive noise

The first set of examples addresses noiseless undersampled data (Cartesian sub-sampling at rate $1/8$), and totally $L = 80$ data frames. In these tests, we used a coarse dictionary for initializing Algorithm 2. Here T_1 was discretized from $200ms$ to $5500ms$ with increments of $200ms$, and T_2 was discretized from $20ms$ to $550ms$ with increments of $20ms$. Note we not only compare our results to the solutions of the BLIP algorithm, but we also plot the initial guesses produced by BLIP. Concerning BLIP, following the findings in [6] we applied 20 steps of a Landweber iteration. On the other hand, our method was stopped after 25 Levenberg-Marquardt steps as then no significant change in the iterates was observed. The regularization parameters had the following values: $\mu = \mu_n = 0$, for every $n \in \mathbb{N}$, $\lambda_0 = s^2$, and $\beta = 0.01$.

The reconstructed parameter maps are presented in Figure 3. In the first row we depict the parameter maps T_1, T_2, ρ of the BLIP algorithm, computed with the coarse dictionary. These quantities were subsequently used for the initialization of our new algorithm. In the second row, the corresponding results for the fine dictionary are shown. These are the ones that should be compared with the images of the third row, which are the results of our algorithm. In order to make the differences clearer, we also provide the corresponding error maps in Figure 4. In fact, we show the pointwise error maps $|\theta_{computed} - \theta_{gt}|$, where θ_{gt} are the ground-truth parameter maps shown in Figure 1, and draw the reader to observing the scale of error as depicted in the vertical bar. We observe that the accuracy of the estimated parameters, especially for T_1, is much higher in our method when compared to BLIP. Note that the error in BLIP is actually larger than the dictionary mesh size, which indicates that this is not a matter of the fineness of dictionary, but it could also be due to the projection onto a non-nonvex set as discussed above.

The rate of convergence of the proposed algorithm turns out to be linear for this example; see Figure 5. The figure depicts the ratio $\frac{\|x_{n+1}-x_n\|_2}{\|x_n-x_{n-1}\|_2}$ versus the number of iterations. Note that x stands here either for T_1, T_2, or ρ.
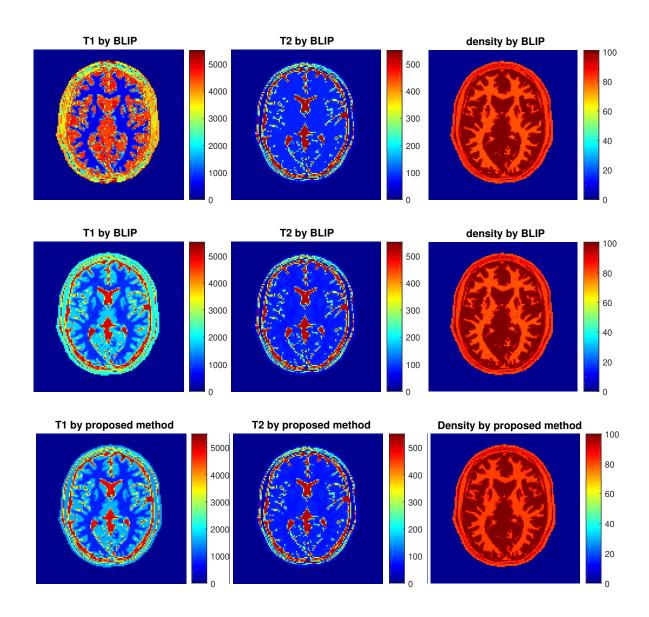
Figure 3: Experiment with noiseless undersampled data. First row: Initialization of our algorithm, computed by BLIP with a coarse dictionary. Middle row: Result by BLIP with fine dictionary. Last row: Solution by proposed algorithm.

5.4 Undersampled data with additive noise

Now we present results for undersampled noisy data with a sub-sampling rate of 1/4 and additive Gaussian white noise of variance $\sigma^2 = 0.8$ and mean 0. The total signal to noise ratio of the Fourier data is $SNR = 35$. As before, we use here $L = 80$ data frames.

The coarse dictionary employed in order to generate the initial value $\mathbf{x}_0$ used T_1 discretized from $400ms$ to $5500ms$ with increments of $400ms$, and T_2 from $40ms$ to $550ms$ with increments of $40ms$. This resulted in a dictionary with 169 entries only, and needed a complex-valued matrix of dimension $169 \times L$ for its representation. As in the previous example, the refined dictionary had a dimension $133956 \times L$. Again, we used 20 Landweber iterations for BLIP, and 25 iterations for our L-M algorithm. The regularization parameters were chosen as $\mu_n = 10^{-8} \left\| Q\mathbf{x}_n - D^\delta \right\|_2$ for every $n \in \mathbb{N}$, $\lambda_0 = s^2$, and $\beta = 0.01$. Note that because of noise, here we used a fixed μ strictly larger than zero.
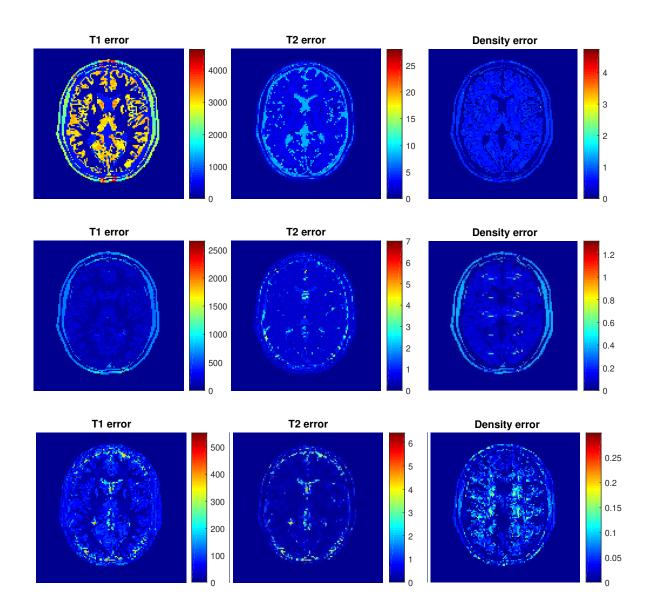
Figure 4: Experiment with noiseless undersampled data. Pointwise distance of the solutions of Figure 3 to the corresponding ground truths of Figure 1. First row: Initial error of BLIP with a coarse dictionary. Middle row: error of BLIP with fine dictionary. Last row: Error of the proposed algorithm.

We depict the results in Figure 6 and the corresponding pointwise errors in Figure 7, using the same row system as in Figures 3 and 4, respectively. The result of the proposed algorithm again outperforms the refined BLIP algorithm, especially in the reconstruction of the density map, but not as significantly as in the noiseless case. In addition, our method consumes much less memory and requires much less CPU-time; see Table 1.

In Figure 8, we plot again the residual ratios again showing a linear rate of convergence.

5.5 Ideal data–fully sampled and no noise

We also discuss the results for the case of fully sampled and noise-free data. Here, we only need $L = 3$ data frames, which actually equals the number of unknown parameters. Thus, the resulting discrete system is non-singular. For both the BLIP and our algorithm, we execute 5 iterations. The
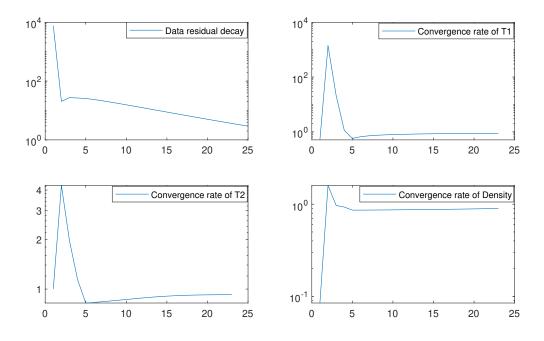
Figure 5: Experiment with undersampled data. From left to right and from above to bottom: Convergence of the data residual $\|Q(\mathbf{x}_n) - D\|$, convergence rates via plots of the iterate ratios $\frac{\|x_{n+1}-x_n\|_2}{\|x_n-x_{n-1}\|_2}$ for $x = T_1, T_2, \rho$, respectively.

regularization parameters were chosen as $\mu_n = 0$ for every $n \in \mathbb{N}$, $\lambda_0 = s^2 = 1$, and $\beta = 0$. Note that, as discussed earlier, this choice makes the L-M iteration equivalent to the Gauss-Newton method.

Here, we only show the error maps of the results in Figure 9. We observe that the Gauss-Newton algorithm essentially recovers the ground truth as expected, while the accuracy of BLIP is limited by the discretization mesh of the dictionary.

In contrast to the previous case, as we verify numerically in Figure 10, the convergence rate of the algorithm is superlinear.

5.6 Quantitative comparisons

In Table 1 we provide a summary of further qualitative comparisons for all of the previous tests. The index in our comparison is the cost in CPU-time as well as the error rates of each algorithm, with the latter defined as

$$\frac{\|x_{computed} - x_{gt}\|_2}{\|x_{gt}\|_2},$$

where $x = T_1, T_2$ or ρ. We provide comparisons with the initial value $\mathbf{x}_0$ and also with the results of the L-M algorithm without projection. Note that the CPU-time costs for the algorithm proposed here include the time needed for computing the initial value.

From the table we observe that the initial guess has been significantly improved by both algorithms, but in the end our proposed algorithm outperforms the refined BLIP in all of the indices. We can see from the table that the non-projected L-M algorithm provides comparable results as the projected L-M method in the cases the data with no additive noise. This is not surprising since the initialization
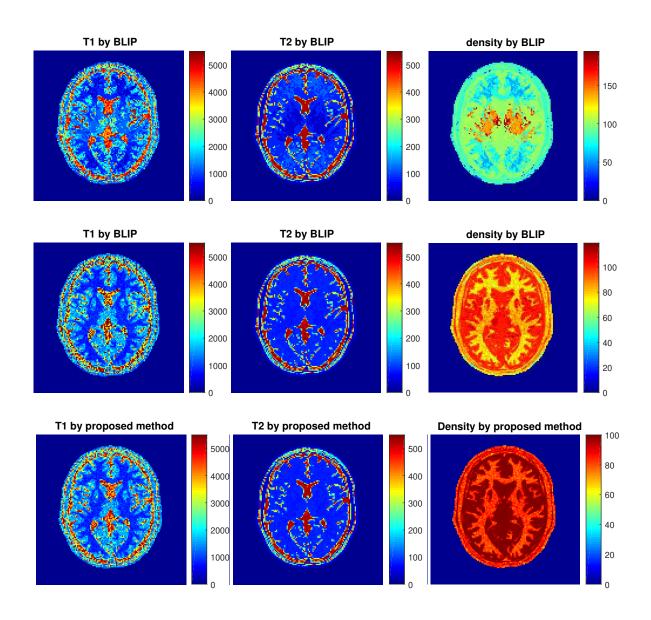
Figure 6: Experiment with undersampled and noisy data. First row: Initialization of our algorithm, computed by BLIP with a coarse dictionary. Middle row: Result by BLIP with fine dictionary. Last row: Solution by proposed algorithm.

Table 1: Quantitative summary of the results: computational times and error rates

	Full data				$1/8$ sampled data				$1/4$ sampled and noisy data			
	time (s)	error rate $\frac{\|e\|_2}{\|x^*\|_2}$			time (s)	error rate $\frac{\|e\|_2}{\|x^*\|_2}$			time (s)	error rate $\frac{\|e\|_2}{\|x^*\|_2}$		
		T_1	T_2	ρ		T_1	T_2	ρ		T_1	T_2	ρ
Initial	1.20	0.036	0.009	0.008	15.26	0.472	0.010	0.003	13.67	0.148	0.088	0.188
BLIP	78.94	0.005	0.002	0.003	964.04	0.072	0.002	0.001	1073.86	0.078	0.019	0.028
L-M	8.41	$1.6*10^{-13}$	$2.6*10^{-15}$	$6.1*10^{-16}$	489.82	0.009	0.001	0.0002	493.23	0.072	0.014	0.020
Proposed	8.47	$1.6*10^{-13}$	$2.4*10^{-15}$	$5.6*10^{-16}$	494.56	0.015	0.002	0.0002	495.77	0.070	0.011	0.009

produces values well located in the interior of the feasible set. As a consequence the non-projected L-M iterations can almost be reside in the feasible domain. Hence, the projection appears unnecessary
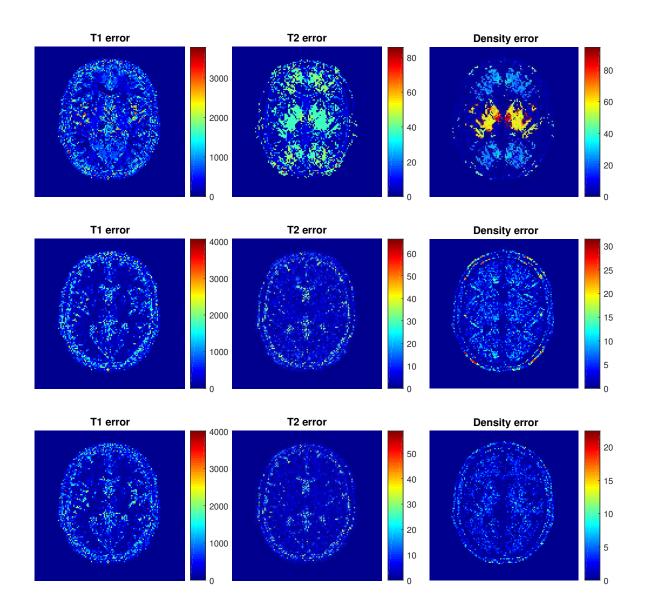
Figure 7: Experiment with undersampled and noisy data. Pointwise distance of the solutions of Figure 6, to the corresponding ground truths of Figure 1. First row: Initial error of BLIP with a coarse dictionary. Middle row: error of BLIP with fine dictionary. Last row: Error of the proposed algorithm.

in the noise-free case.

Finally, we would like to verify the fact that larger frames sequences can help to get more accurate estimations; compare the discussion at the end of Section 4.5. For the results shown in Table 2 we performed a set of experiments for data frames of increasing length L. All data were $1/4$ sub-sampled and corrupted by additive i.d.d. Gaussian noise as described before with variance 1 and mean 0. This gives a total signal noise ratio $SNR = 15$. We then ran our L-M algorithm for 20 iterations always using the same initial value $\mathbf{x}_0$, which was generated by using BLIP with 160 frames and the coarse dictionary as described in Section 5.4. The parameter had values $\lambda_n = \mu_n = \lambda_0 \left\| Q\mathbf{x}_n - D^\delta \right\|_2$ for all $n \in \mathbb{N}$, and $\lambda_0 = 10^{-8}$.

The results in Table 2 clearly indicate that an increasing number of sequences improves the accuracy of the estimated parameters. There we have also computed the errors of the standard L-M method, i.e., with no projection. It can be observed (also in Table 1 for the noisy case) that the projected L-M
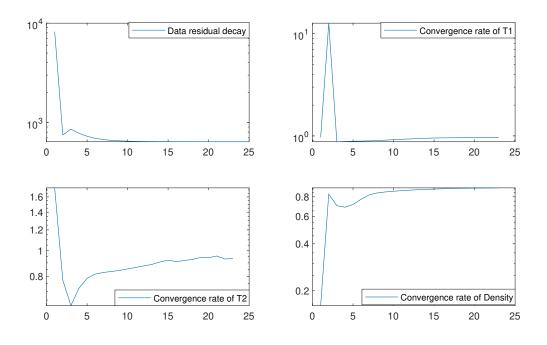
Figure 8: Experiment with noisy data. From left to right and from above to bottom: Convergence of the data residual $\|Q(\mathbf{x}_n) - D^\delta\|$, convergence rates via plots of the iterate ratios $\frac{\|x_{n+1}-x_n\|_2}{\|x_n-x_{n-1}\|_2}$ for $x = T_1, T_2, \rho$ respectively.

method outperforms the standard L-M method of no projection. Note that the CPU-time costs that we show here do not include initialization as the latter was the same in each case.

Table 2: The length of data sequences and its influence on the solution accuracy

	Standard L-M method				Projected L-M method (proposed)			
	ER T_1	ER T_2	ER ρ	time (s)	ER T_1	ER T_2	ER ρ	time (s)
$L = 5$	0.2267	0.4923	0.1682	36.66	0.1743	0.2028	0.0424	36.40
$L = 10$	0.1818	0.0805	0.0757	59.69	0.1699	0.0348	0.0275	59.33
$L = 20$	0.0542	0.0182	0.0317	104.56	0.0290	0.0072	0.0099	104.65
$L = 40$	0.0413	0.0131	0.0276	193.65	0.0211	0.0051	0.0090	194.28
$L = 80$	0.0268	0.0117	0.0268	375.90	0.0121	0.0043	0.0087	374.41
$L = 160$	0.0193	0.0112	0.0266	736.77	0.0078	0.0041	0.0085	737.45

6 Concluding remarks

In this paper, we analysed MRF from the perspective of inverse problems, and we were able to provide some mathematical insights in order to better understand the functionalities of the method. Subsequently, we have proposed a novel model for quantitative MRI which is in accordance with the standard routine of the MRI experiment setting. The model is dictionary-free and incorporates the physical
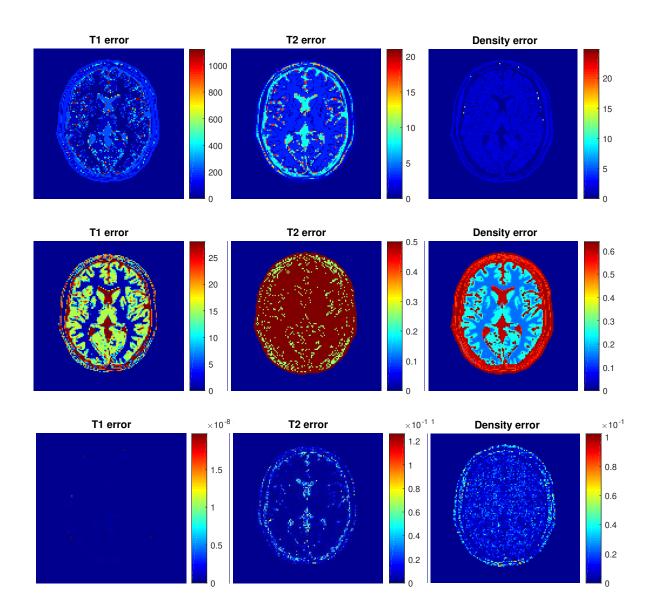
Figure 9: Fully sampled data. Pointwise distance of the solutions of BLIP algorithm and Newton algorithm to the corresponding ground truths. First row: Initial error of BLIP with a coarse dictionary. Middle row: error of BLIP with fine dictionary. Last row: Error of the proposed algorithm.

setting of MRI into one single non-linear equation. We have proposed a robust algorithm that was shown to be capable of estimating the tissue parameters with high precision. In contrast to the original MRF method and its variant BLIP, it does not rely on refining a dictionary to improve the accuracy. Even though our new algorithm is based on a specific Bloch dynamics referred to as IR-bSSFP, this constitutes by no means a limitation for the method. Rather, other type of discrete dynamics or approximations to Bloch equations can be fitted to this approach as well. Furthermore, we have considered the relaxation parameters T_1, T_2 and the proton density ρ as unknowns in the present paper, but as long as parameters can be related to the Bloch dynamics, there would be no difficulty to extending the algorithm to incorporate the further parameters into the current framework.
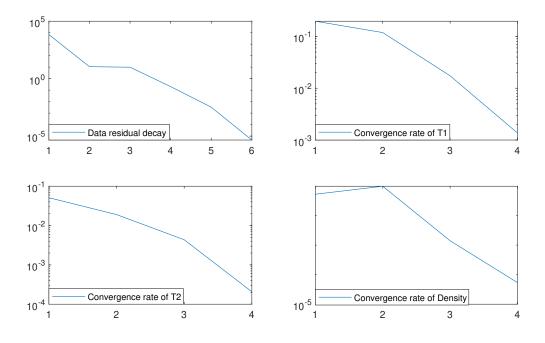
Figure 10: Fully sampled data and superlinear convergence of the Newton method. From left to right and from above to bottom: Convergence of the data residual $\|Q(\mathbf{x}_n) - D\|_2$, convergence rates via plots of the iterate ratios $\frac{\|x_{n+1}-x_n\|_2}{\|x_n-x_{n-1}\|_2}$ for $x = T_1, T_2, \rho$ respectively.

Acknowledgement

Appendix: Solutions of Bloch equations with different cases

Here we briefly review several simplified cases towards the solutions of the Bloch equations, which are helpful in order to understand the simulations based on discrete dynamics. More detailed descriptions can be found in [20, 25]. Note that here we omit the position dependence in Bloch equations.

Only main field with no relaxation

The Bloch equations, in a setting which only takes into account the main magnetic field and with no relaxation, represent an autonomous dynamical system, that is

$$\frac{\partial m}{\partial t} = m \times \gamma B_0.$$

The solution in this case is

$$m(t) = P_{\omega_0}(t)m(0),$$

where

$$P_{\omega_0}(t) = \begin{pmatrix} \cos(\omega_0 t) & \sin(\omega_0 t) & 0 \\ -\sin(\omega_0 t) & \cos(\omega_0 t) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \text{and } \omega_0 = \gamma \left| B_0 \right|.$$

It can be interpreted in a way that the magnetization precesses about the main magnetic field at a frequency ω_0, called Larmor frequency.

Main field with relaxation

This is the case of Bloch equations (1) with $B(t, r) = B_0$, meaning that radio frequencies and gradient fields are not considered here. After some change of variable and further calculations, the solution turns out to be

$$m(t) = P_{\omega_0}(t)E(t)m(0) + (1 - e^{-\frac{t}{T_1}})m_e,$$

where

$$E(t) = \begin{pmatrix} e^{-\frac{t}{T_2}} & 0 & 0 \\ 0 & e^{-\frac{t}{T_2}} & 0 \\ 0 & 0 & e^{-\frac{t}{T_1}} \end{pmatrix}.$$

Note that the matrices $P_{\omega_0}(t)$ and $E(t)$ are commutable.

With perturbations and without relaxation

By perturbation we mean that there is a B_1 field which rotates at the Larmor frequency, and it is always orthogonal to the main field, such that $\langle B_0, B_1 \rangle = 0$. This models the excitation of radio pulses in the MRI machine. By convention, the direction of the B_1 field can be defined to be along the x-axis in space. Since in reality, the excitation pulse only lasts for a very short length of time in comparison with T_1 and T_2, we can ignore the relaxation terms. The solution of (1) in the case of no relaxation terms but with perturbation is

$$m(t) = P_{\omega_0}(t)R_x(\alpha(t))m(0),$$

where $\alpha(t) := \gamma \int_0^t |B_1(s)| \, ds$ is the flip angle, and

$$R_x(t) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha(t)) & \sin(\alpha(t)) \\ 0 & -\sin(\alpha(t)) & \cos(\alpha(t)) \end{pmatrix}.$$

With perturbations and relaxation

Finally we are able to simulate the solutions of (1) in the case where both the perturbations of the main field and relaxation terms are taken into account. This is based on the assumption that the excitation pulse is turned on at the time period $(0, t_0)$, where t_0 is a very small number in comparison to the relaxation parameters. Therefore, we can estimate the solution of (1) with the following formula:

$$m(t) = P_{\omega_0}(t)E(t)R_x(\alpha(t_0))m(0) + (1 - e^{-\frac{t}{T_1}})m_e.$$

The main tool in all of the above calculations is to change variables to a rotating frame of reference in order to match the Larmor precession.

References

[1] Brainweb: Simulated brain database. *http://www.bic.mni.mcgill.ca/brainweb/*.

[2] J. Assländer, M.A. Cloos, F. Knoll, D.K. Sodickson, J. Hennig, and R. Lattanzi. Low rank alternating direction method of multipliers reconstruction for MR fingerprinting. *Magnetic Resonance Medicine*, 79:83–96, 2018.

[3] A. Bensoussan, G. Da Prato, M.C. Delfour, and S.K. Mitter. *Representation and Control of Infinite Dimensional Systems*. Systems and Control Foundations and Applications. Birkhauser, second edition, 2007.

[4] F. Bloch. Nuclear induction. *Physical Review*, 70:460–473, 1946.

[5] 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 Transactions on Medical Imaging*, 17(3):463–468, 1998.

[6] M. Davies, G. Puy, P. Vandergheynst, and Y. Wiaux. A compressed sensing framework for magnetic resonance fingerprinting. *SIAM Journal on Imaging Sciences*, 7(4):2623–2656, 2014.

[7] P. Deuflhard. *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*, volume 35 of *Springer series in computational mathematics*. Springer, first edition, 2004.

[8] L.S. Evans. *Partial Differential Equations*, volume 19 of *Graduate studies in mathematics*. American Mathematical Society, second edition, 2010.

[9] J. Fan and Y. Yuan. On the quadratic convergence of the Levenberg-Marquardt method without nonsingularity assumption. *Computing*, 74:23–39, 2005.

[10] K. Ferentinos. On Tcebycheff's type inequalities. *Trabajos de Estadistica y de Investigacion Operativa*, 33(1):125–132, 1982.

[11] W. Guo, Guohui Song, and Yue Zhang. PCM-TV-TFV: A novel two-stage framework for image reconstruction from Fourier data. *SIAM Journal on Imaging Sciences*, 10(4):2250–2274, 2017.

[12] M. Hanke. A regularizing Levenberg-Marquardt scheme, with applications to inverse groundwater filtration problems. *Inverse Problems*, 13:79–95, 1997.

[13] M. Hintermüller, C.N. Rautenberg, T. Wu, and A. Langer. Optimal selection of the regularization function in a weighted total variation model. Part II: Algorithm, its analysis and numerical tests. *Journal of Mathematical Imaging and Vision*, 59(3):515–533, 2017.

[14] C. Huang, C.G. Graff, E.W. Clarkson, A. Bilgin, and M.I. Altbach. T2 mapping from highly undersampled data by reconstruction of principal component coefficient maps using compressed sensing. *Magnetic Resonance in Medicine*, 67:1355–1366, 2012.

[15] K. Ito and K. Kunisch. *Lagrange Multiplier Approach to Variational Problems and Applications*, volume 15 of *Advances in Design and Control*. Society for Industrial and Applied Mathematics, 2008.

[16] B. Kaltenbacher, A. Neubauer, and O. Scherzer. *Iterative Regularization Methods for Nonlinear Ill-Posed Problems*, volume 6 of *Radon Series on Computational and Applied Mathematics*. De Gruyter, 2008.

[17] C. Kanzow, N. Yamashita, and M. Fukushima. Levenberg-Marquadt methods with strong local convergence properties for solving nonlinear equations with convex constraints. *Journal of Computational and Applied Mathematics*, 172:375–397, 2004.

[18] C. Karmonik, A. Malaty, M. Bikram, P. Schmitt, S. Partovi, and D.J. Shah. Fast in vivo quantification of T1 and T2 MRI relaxation times in the myocardium based on Inversion Recovery SSFP with in vitro validation post Gd-based contrast administration. *Cardiovascular Diagnosis & Therapy*, 4(2):88–95, 2014.

[19] F. Knoll, K. Bredies, T. Pock, and R. Stollberger. Second order total generalized variation (TGV) for MRI. *Magnetic Resonance in Medicine*, 65(2):480–491, 2011.

[20] V. Kuperman. *Magnetic Resonance Imaging: physical principles and applications*. Electromagnetism. Academic Press, first edition, 2000.

[21] H. Larsson, J. Frederiksen, J. Petersen, I. Nordenbo, A.and Zeeberg, O. Henriksen, and J. Olesen. Assessment of demyelination, edema, and gliosis by in vivo determination of T1 and T2 in the brain of patients with acute attack of multiple sclerosis. *Magnetic Resonance in Medicine*, 11:337–348, 1989.

[22] M. Lustig, D.L. Donoho, J.M. Santos, and J.M. Pauly. Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25:72–82, 2008.

[23] D. Ma, V. Gulani, N. Seiberlich, K. Liu, J. Sunshine, J.L. Duerk, and M.A. Griswold. Magnetic resonance fingerprinting. *Nature*, 495(187):187–193, 2013.

[24] G.C. McKinnon. Ultrafast interleaved gradient-echo-planar imaging on a standard scanner. *Magnetic Resonance in Medicine*, 30:609–616, 1993.

[25] D. Nishimura. *Principles of Magnetic Resonance Imaging*. Stanford University, second edition, 2010.

[26] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill Series in Electrical Engineering. McGraw Hill Higher Education, third edition, 1991.

[27] A. Sbrizzi, T. Bruijnen, O. van der Heide, P. Luijten, and C.A.T. van den Berg. Dictionary-free MR fingerprinting reconstruction of balanced-GRE sequences. *arXiv preprint*, 2017. `https://arxiv.org/abs/1711.08905v1`.

[28] A. Sbrizzi, O. van der Heide, M. Cloos, A. van der Toorn, H. Hoogduin, P.R. Luijten, and C.A.T. van der Berg. Fast quantitative MRI as a nonlinear tomography problem. *Magnetic Resonance Imaging*, 46:56–63, 2018.

[29] K. Scheffler. A pictorial description of steady-states in rapid magnetic resonance imaging. *Concepts in Magnetic Resonance*, 11(5):187–193, 1999.

[30] G. Teschl. *Ordinary Differential Equations and Dynamical Systems*, volume 140 of *Graduate Studies in Mathematics*. American Mathematical Society, first edition, 2012.

[31] G.A. Wright. Magnetic resonance imaging. *IEEE Signal Processing Magazine*, 14:56–66, 1997.

[32] N. Yamashita and M. Fukushima. On the rate of convergence of the Levenberg-Marquardt method. In G. Alefeld and X. Chen, editors, *Topics in Numerical Analysis: With Special Emphasis on Nonlinear Problems*, Computing Supplementa book series, pages 239–249. Springer, 2001.