In vivo validation of Magnetic Resonance Spin Tomography in Time-domain

Oscar van der Heide

Cover
Ollie

In vivo validation of Magnetic Resonance Spin Tomography in Time-domain
PhD Thesis, Utrecht University, the Netherlands

Manuscript

Layout:	O. van der Heide
Cover design:	O. van der Heide
Typesetting:	LaTeX
ISBN:	978-90-393-7639-3
Printing:	Ridderprint | `www.ridderprint.nl`

Funding
This work was funded by the Nederlandse Organisatie voor Wetenschappelijk Onderzoek, grant numbers 14125 and 17986.

In vivo validation of Magnetic Resonance Spin Tomography in Time-domain

In vivo validatie van kernspinresonantie-tomografie in het tijdsdomein

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op

dinsdag 6 februari 2024 des middags te 12.15 uur

door

Oscar van der Heide

geboren op 9 september 1987 te Amsterdam

Promotoren:
Prof. dr. C.A.T. van den Berg
Prof. dr. P.R. Luijten

Copromotor:
Dr. A. Sbrizzi

Beoordelingscommissie:
Prof. dr. R.H. Bisseling (voorzitter)
Prof. dr. R.M. Dijkhuizen
Prof. dr. D.W.J. Klomp
Prof. dr. M. Uecker
Dr. ir. E.P.A. Vonken

Contents

1

Introduction

Magnetic Resonance Imaging ("MRI") has established itself as one of the dominant imaging modalities in current clinical practice and biomedical research. Compared to other imaging modalities, MRI provides superior soft tissue contrast and, unlike X-ray and CT imaging, it does not expose the patient to ionizing radiation. Arguably though, the main strength of MRI is found in its versatility in generating different types of images, each providing different clinical insights. The versatility also continuous to make MRI an exciting field of research. To this day, new MRI imaging biomarkers are being discovered [187], new hardware [12] and software techniques [131] are being developed to improve image quality and non-diagnostic applications of MRI such as image-guided radiotherapy are being explored [40].

This thesis will focus on technological advancement of a relatively new quantitative MRI technique called Magnetic Resonance Spin Tomography in Time-domain ("MR-STAT") [141]. The main development goal was to design and implement data acquisition procedures as well as image reconstruction techniques for MR-STAT to allow the technique to be utilized in a clinical setting. In this general introduction, the image formation mechanisms in both qualitative and quantitative MRI are first described. Then, challenges associated with tradional quantitative MRI approaches are outlined and subsequently recent advancements in the field are discussed. Finally the MR-STAT technique is introduced.

1.1 Conventional Magnetic Resonance Imaging

1.1.1 MR Image Formation

The main image formation mechanism underlying all the different MRI techniques can be briefly explained as follows. Certain atomic nuclei within the human body, with hydrogen nuclei being the most abundant example, have a magnetic moment. In normal conditions, due to thermal motion, these magnetic moments cancel out on a macroscopic level. However, in MRI, a patient is placed within a strong, static magnetic field. The nuclei will interact with this external field in such a way that local ensembles of nuclei will have a net magnetization along the direction of the external field. By convention, in MRI the main magnetic field points in the z direction and the xy plane is referred to as the *transverse plane*. Local ensembles of nuclei are referred to as *spin isochromats* and can be thought of as three-dimensional magnetization vectors $\boldsymbol{M}(t) = (M_x(t), M_y(t), M_z(t))$, where t refers to time. By utilizing additional hardware components of the MR system such as *transmit coils* and *gradient coils*, the external magnetic field $\boldsymbol{B}(t)$ experienced by spin isochromats can be modified and motion of the spin isochromats can be induced. These spin dynamics are described by a set of phenomenological equations called the Bloch equations [19]:

$$\frac{d\boldsymbol{M}(t)}{dt} = \gamma\left(\boldsymbol{M}(t) \times \boldsymbol{B}(t)\right) - \left(\frac{M_x(t)}{T_2}, \frac{M_y(t)}{T_2}, \frac{M_z(t) - M_0}{T_1}\right). \tag{1.1.1}$$

Here γ is the *gyromagnetic ratio* of the nuclei, T_1 is the *longitudinal relaxation time* of the spin, T_2 the *transverse relaxation* and M_0 is the initial M_z component of the spin isochromat. In particular, if B_0 is the strength of the static magnetic field (in the z-direction) experienced by a spin isochromat, then the spin isochromat can be *excited* by transmitting electromagnetic waves at the *Larmor frequency* γB_0. The result of the excitation is that the spin isochromat starts precessing (with the same Larmor frequency) around a circle in the transverse plane: the plane orthogonal to the direction of the main magnetic field. During the precession, the time-varying transverse component of the spin isochromat itself generates electromagnetic waves that can be picked up by a receive coil placed in close vicinity of the patient. Such a receive coil, however, picks up signal from all excited spin isochromats within its receive field at once. To be able to form an image, *spatial localisation* of the signals coming from different spin isochromats is required. For this purpose, gradient coils are used to modify the magnetic field at different spatial locations. Typically there is one gradient coil for each spatial direction (x, y, z) which modifies the static magnetic field in that particular direction. With gradient coils, spin isochromats at different spatial locations can thus be made to precess at different frequencies, giving the spins phase offsets relative to each other in the transverse plane and allowing *spatial frequencies* within the excited object to be exposed and

measured. In MRI, the spatial frequency domain is typically referred to as *k-space*. The order in which k-space samples are acquired is described by the (time-dependent) *gradient trajectory* $\boldsymbol{k}(t)$, defined as the time-integral

$$\boldsymbol{k}(t) = \gamma \int_0^t \mathbf{G}(\mathbf{s})\mathbf{ds}, \tag{1.1.2}$$

with $\boldsymbol{G}(t)$ the vector of gradient waveforms in all spatial directions. If we let $m(\boldsymbol{r}, t)$ denote the transverse magnetization of a spin isochromat at spatial location $\boldsymbol{r}$, represented as a complex number, then the signal s picked up by a (homogeneous) receive coil at sampling time t can be expressed as the volume integral

$$s(\boldsymbol{k}(t)) = \int_V m(\boldsymbol{r}, t) e^{-2\pi i \boldsymbol{k}(t)\cdot\boldsymbol{r}} d\boldsymbol{r}, \tag{1.1.3}$$

where V is the excited volume. Note that Eq. 1.1.1 highlights a Fourier relationship between the measurable signal $s(t)$ and the spatial distribution of the transverse magnetization m. The image reconstruction procedure in conventional MRI approaches consists of applying a discrete Fourier transform to a discrete set of k-space samples denoted $\boldsymbol{s}$. For successful image reconstruction, however, several conditions imposed by the Fourier transform must be satisfied.

First, the Fourier transform assumes that m has no time-dependency. In reality, the spin isochromats in principle continuously evolve in time in accordance with the Bloch equations. To accommodate the Fourier-reconstruction, conventional MRI pulse sequences are designed in such a way that spin isochromats are brought into a *steady-state*. That is, during sampling times, the magnetization at each spatial location is - to good approximiation - independent of time. Steady-states are typically achieved by applying a repetitive sequence of radiofrequency excitations.

The second assumption is that the gradient coils used to modify the main magnetic field result in linear field variations in space. This assumption is encapsulated in the dot product $\boldsymbol{k}(t) \cdot \boldsymbol{r}$ within Eq. 1.1.1. Achieving gradient linearity over a sufficiently large volume is an important constraint that is taken into account in the design of gradient coils [75].

Third, successful image formation requires that the Nyquist sampling criterion is satisfied. Given a desired field-of-view ("FOV"), this criterion describes a maximum allowed k-space distance $\Delta\boldsymbol{k}$ between sample points to prevent *aliasing* artefacts in the image domain. The desired resolution in the image domain further specifies the highest spatial frequencies $\boldsymbol{k}_{max}$ that must be sampled.

Assuming a Nyquist-sampled k-space is acquired during a steady-state MR acquisition with linear gradient coils, a Fast Fourier Transform ("FFT") is typically applied to generate an image. Such an image only has clinical value if there is sufficient *contrast* between between healthy and non-healthy tissue. Contrast in MR images is realized by exploiting differences in nuclear magnetic properties of different tissues such as proton density and T_1 and T_2 relaxation

times. These nuclear magnetic tissue properties affect the signal emitted by spin isochromats. Through a choice of carefully timed RF excitations and gradient fields, images can be generated whose contrast is predominantly determined by differences in one of the nuclear magnetic properties at a time. The resulting contrast images are said to be *weighted* by the nuclear magnetic properties (e.g. T_1-weighted or T_2-weighted). The versatility in contrast generation mechanisms is one of the key strengths of MRI.

1.1.2 Conventional MRI Workflow

Certain tissue abnormalities may be hard to detect on one contrast while being clearly visible on another. Therefore, in a typical MRI exam, multiple contrasts are acquired one after another. For each individual contrast, a different pulse sequence is played out and - in principle - the requirements imposed by the FFT-based reconstruction are fulfilled. A graphical overview of this current clinical workflow is depicted in Fig. 1.1. Two fundamental issues can be identified with the conventional MRI workflow.

First, scan times are relatively long. For a single contrast, scan times are in the order of minutes. A full MRI exam during which several contrasts are acquired may take 30 minutes or longer. During the exam, patients are asked to lie still or perform breathholds to prevent motion artefacts in the images, adding to patient discomfort. The lengthy acquisition process also makes it challenging to perform dynamic imaging. The long scan times result from a combination of several factors. Given a desired field-of-view and spatial resolution, the Nyquist criterion dictates which and how many spatial frequencies must be sampled. The spatial frequencies, however, can in principle only be sampled sequentially (in contrast to say, a photo camera, that acquires voxel intensities in all voxels simulatenously). In between samples, the gradient coils must be activated to give spin isochromats different phase offsets. Furthermore, in order to achieve and maintain a steady-state for a specific contrast, RF excitations, other gradient waveforms and even waiting times may also be necessary in between sampling. Pulse sequences therefore typically consist of a base sequence block (with a duration that is referred to as *repetition time*) that is repeated multiple times with different gradient field encodings until a Nyquist-sampled k-space is acquired.

While the *acquisitions* are lengthy, the FFT-based image *reconstruction* procedure is typically very fast, allowing images to be generated at the scanner console almost instantly after having acquired the data. In general it can be said that in the current MRI paradigm the acquisiton procedure is designed in such a way to accommodate a fast and easy reconstruction.

Over time, several techniques have been introduced into clinical practice that relax the Nyquist sampling criterion at the expensive of more complicated reconstructions procedures. Notable examples include parallel imaging [135, 146] and compressed sensing [109]. With parallel imaging, k-spaces are

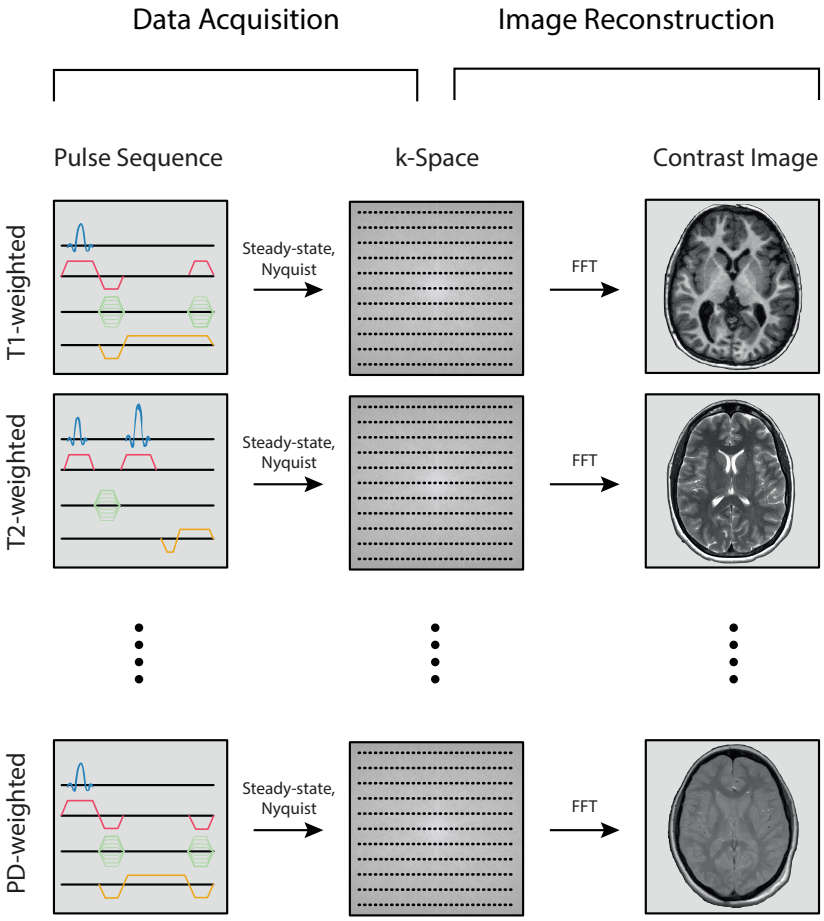

Figure 1.1: Graphical overview of current clinical MRI workflow. Multiple contrast images are acquired with different pulse sequences that are designed to emphasize differences in one nuclear magnetic tissue property at a time. Steady-states must be achieved and each k-space must in principle satisfy the Nyquist criterion. Contrast images are generated from the k-space using FFT-based reconstructions. Modern advancements in the form of parallel imaging and compressed sensing are ignored in this overview.

typically undersampled in a regular fashion. In the reconstruction, data from multiple receive coils is combined to compensate for the missing spatial frequency information. With compressed sensing, k-spaces are undersampled based on pseudo-random patterns. Regular FFT reconstructions would result in images with additional noise-like undersampling artefacts. By incorporating image priors into the reconstruction in the form of assumed sparsity in a suitably chosen domain (for example, the wavelet domain), the aliasing noise can be removed. These acceleration techniques have been further enhanced in recent years by incorporating data-driven artificial intelligence approaches [131]. With these acceleration strategies, the number of required repetitions of a base sequence block can be reduced. However, the duration of each sequence block (which may contain long waiting times) remains unchanged as its duration is dictated by the desired contrast.

A second fundamental issue with conventional MRI is that the images obtained by Fourier transform-based reconstructions only provide *qualitative* information of the imaged object. The voxel values in the reconstructed images are only *weighted* by biophysical tissue properties and provide no quantitative estimates of the properties themselves. Radiologists may diagnose a patient's condition based on certain regions of interest appearing hyper- or hypointense in, for example, T_1 or T_2-weighted scans without knowing the actual T_1 and T_2 values of the tissues.

Diagnosis in the current MRI paradigm is thus based on human interpretation rather than an objective assessment of quantitative values. The actual voxel values in qualitative MR images are, however, also influenced by many factors unrelated to biophysical tissue properties such as the pulse sequence used during scanning, hardware components of the scanner, scanner software releases and data processing pipelines. As such it is challenging with qualitative MRI to perform longitudinal- and multi-center patient studies [3, 99] as well as monitor disease progression [60] or treatment response [147].

1.2 Quantitative MRI

The issue of conventional MRI approaches resulting in qualitative images is adressed by quantitative MRI ("qMRI"). Rather than providing images that are weighted by certain biophysical tissue properties, it is possible with qMRI to provide maps with quantitative estimates of actual biophysical tissue properties. Traditional qMRI approaches follow a two-step procedure. In the first step, the *spatial localisation* step, several contrast images are acquired that - through a slight modification of the pulse sequences - are weighted differently by the tissue parameter of interest. For each voxel, multiple sample points are thus obtained. In a second step, the *parameter estimation* step, a signal model is fitted to the sample points in each voxel individually. The pulse sequences are typically designed in such a way that relatively simple analytic signal mod-

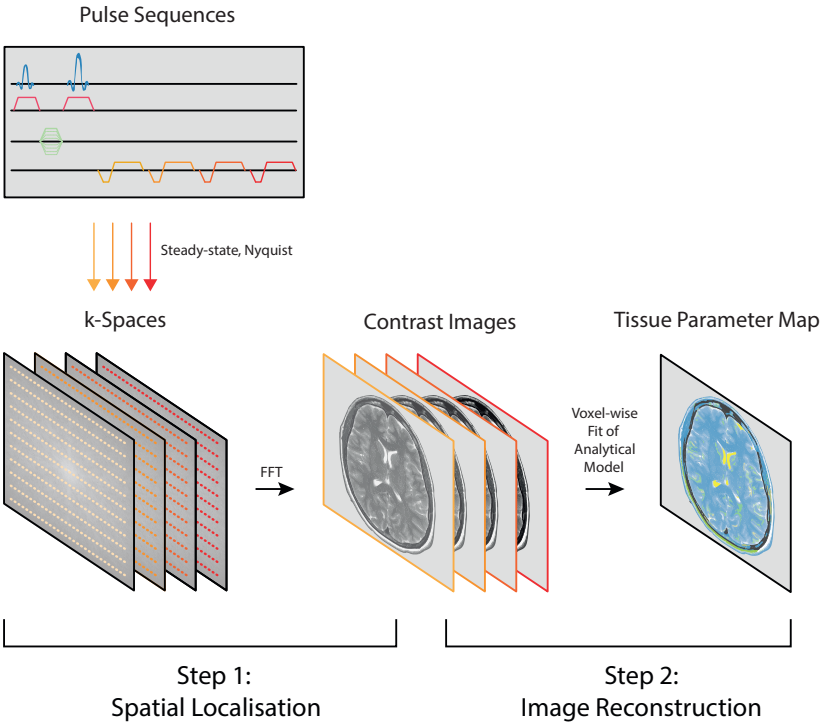

Figure 1.2: Graphical overview of the traditional, two-step qMRI approach. In the first step, the spatial localisation step, several contrast images are acquired, each weighted differently by the tissue parameter of interest. In the second step, the parameter estimation step, an analytical signal model is fitted to the signal in each voxel separately

els are available. Depending on whether the signal model admits a linear or non-linear dependency on the tissue parameter of interest, either direct inversion or iterative inversion techniques are employed. The traditional, two-step qMRI is visualized in Fig. 1.2.

As an example of the traditional qMRI approach, consider a spin-echo experiment. The transverse magnetization in each voxel is assumed to follow an exponential decay curve with the decay rate depending on the T_2 relaxation time of the tissue in that voxel. In order to arrive at a T_2 map, multiple qualitative images are acquired with different echo times. That is, the decay curve in each individual voxel is sampled at multiple (echo) time points. A T_2-dependent exponential decay model can then be fitted in each voxel separately in order to obtain a map of T_2 relaxation times.

Quantitative tissue parameter maps obtained with qMRI should in prin-

ciple be independent of MR hardware[1] and the exact pulse sequences used. Therefore, qMRI could play an important role in previously mentioned areas where the qualitative nature of conventional MRI proves to be challenging (longitudinal- and multi-center studies, disease progression, treatment response). In addition, qMRI could facilitate the formation of multi-center, standardized datasets [178, 62] which could be beneficial for further incorporating data-driven artificial intelligence techniques into radiological workflows.

Despite the above described promises, qMRI has only seen limited clinical adoption so far. A major hurdle in this respect is formed by the prolonged scan times of the traditional qMRI approach compared to conventional, qualitative MRI. For each quantitative tissue parameter map, multiple contrast images are acquired. And for each of these contrast images, the Nyquist sampling criterion must be satisfied. Each of the pulse sequences may contain waiting times to achieve the desired contrasts. The factors that result in qualitative MRI having lengthy acquisitions are thus amplified in qMRI.

Like in the case of conventional, qualitative MRI, it can be said that early qMRI approaches have been designed in such a way to facilitate computationally cheap and straightforward reconstructions. With computational resources having become cheaper and more accessible over time, in more recent years researchers have focussed their attention on qMRI methods that allow for more flexibility in the acquisition procedure at the expense of more computationally demanding reconstruction procedures. Prime examples of this paradigm shift in qMRI are MR Fingerprinting [110] and MR Spin Tomography in Time-domain [141].

1.3 MR Fingerprinting

MR Fingerprinting ("MRF") was introduced in 2013 by Ma et al. [110] as a qMRI technique that allows for the estimation of multiple quantitative tissue parameter maps using data from one short scan. With such a technique, the multiple qualitative scans that are typically performed during a conventional MRI exam (see Fig 1.1) could in principle be replaced by a single (MRF) scan. The initially proposed implementation of the MRF technique relies on the following ingredients. Pulse sequences are used that encode the information from multiple (non-linear) tissue parameters into the signal simulatenously as opposed to one parameter at a time in the traditional qMRI approach. Typical MRF scans continuously vary the flip angle, repetition time and/or echo time during the sequence to keep the spins in a *transient state*. Compared to steady-state sequences, such transient-state sequences can have better encoding capabilities [4]. Analytical signal models are in general not available for such transient-state sequences and numerical integration of the Bloch equations (i.e.

[1] Assuming the same field strength, since T_1 and T_2 are field-strength dependent parameters.

Bloch simulations) is typically required to describe the spin dynamics. Prior to scanning, a dictionary is assembled which consists of Bloch simulation results for many different combinations of tissue parameters (e.g. T_1 and T_2). Then, during the actual data acquisition, many snapshot images are acquired to sample the transient-state spin dynamics. Acquiring fully sampled k-spaces for each snapshot would result in scan times that are prohibitively long for clinical applications, just like in the traditional qMRI setting. Instead, each k-space is highly undersampled. Spatial localisation is performed through the application of (non-uniform) FFTs on each k-space. In each voxel, a time-series referred to as *fingerprint* is obtained. In a subsequent parameter estimation step, the fingerprint in each individual voxel is matched to the precomputed dictionary. However, due to the undersampling of the k-spaces, the fingerprints are corrupted by aliasing artefacts. Assuming *spatio-temporal incoherence* of the aliasing artefacts, and assuming that a pulse sequence with sufficient parameter encoding power is used, the dictionary matching procedure can still select the correct tissue parameters. The MRF workflow is visualized in Fig. 1.3. To this date, MRF is an active field of research and has been used with a wide variety of pulse sequences (e.g. gradient-balanced, gradient-spoiled and RF-spoiled) to estimate many different combinations of tissue parameters and scan parameters, including T_1, T_2, B_1 transmit imhomogeneity, ΔB_0, $T_1\rho$ [143], magnetization transfer [77, 181, 7] and apparent diffusion coefficient [1]. The MRF technique has been employed in various clinical trials [78].

The originally proposed MRF implementation can be considered a two-step approach with separate spatial localisation and parameter estimation steps. However, compared to the more traditional qMRI approaches it allows for shorter scan times. Scan times for MRF are typically in the order of minutes for entire 3D volumes. The reductions in scan time are made possible through a combination of several factors. Most importantly, rather than having acquisitions designed to (predominantly) encode only one tissue parameter type at a time, the pulse sequences are designed to encode multiple tissue parameter into the measured signal at once. The pulse sequences also typically do not have waiting times (e.g. long repetition times or inversion times) or dummy pulses to bring the spins into a steady-state are required. High undersampling factors within transient-state acqisitions are allowed by leveraging the favorable undersampling properties of non-Cartesian trajectories.

Two important limitations can be identified with the MRF technique that may limit its future potential as a definitive qMRI solution. First, MRF suffers from the *curse of dimensionality*. The computer memory required for storing the pre-computed dictionary scales expontially with the number of different tissue parameters that are estimated. For up to three parameters, dictionary sizes are in the order of gigabytes (depending on the resolution in the parameter directions) and are managable on modern desktop computers. For estimating four different tissue parameters simultaneously, dictionary sizes of hundreds of gigabytes have been reported. As the models used in qMRI become more complex,

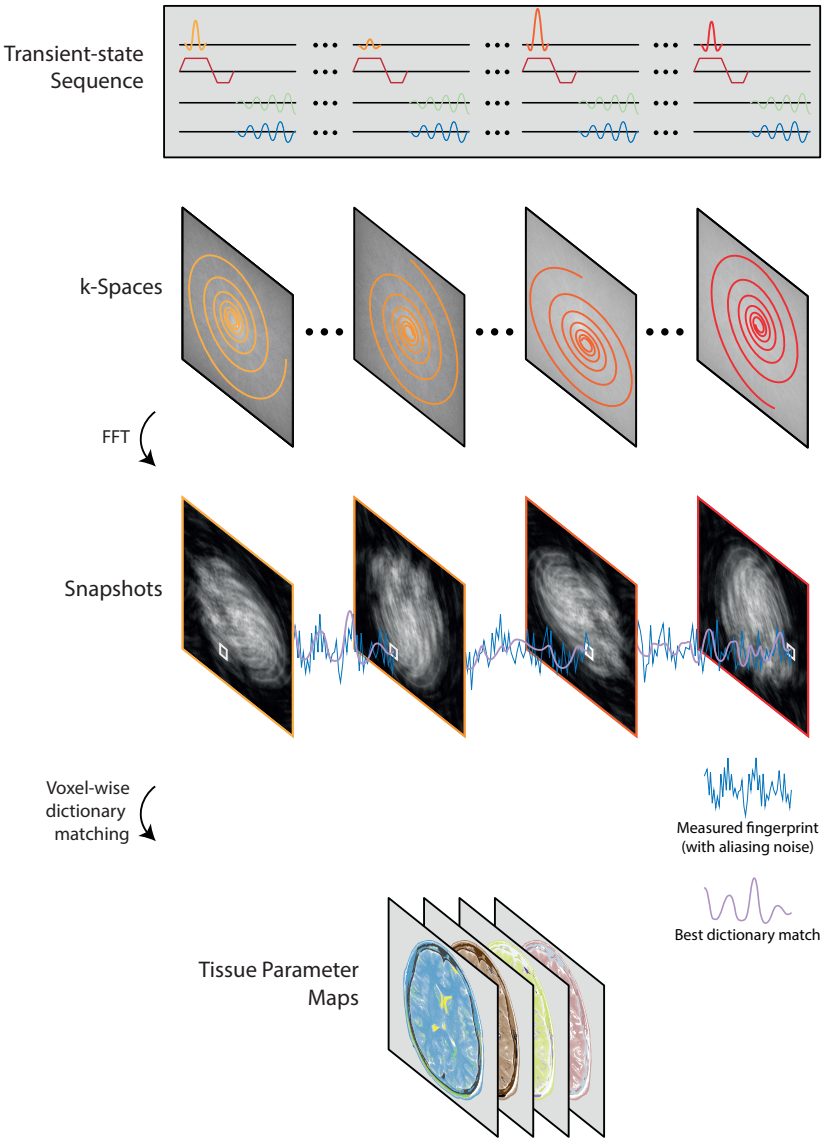

Figure 1.3: Graphical overview of the MRF approach to qMRI. A single transient-state sequence (in this example: a continuously varying flip angle sequence) is used that encodes information from multiple tissue parameters into the MR signal. Many snapshot images are acquired to sample the transient spin dynamics. Each snapshot image is highly undersampled (in this example: one spiral arm for each snapshot) and as a result the snapshots suffer from aliasing noise. On a voxel-per-voxel basis, the measured fingerprints are then matched against a pre-computed, Bloch equations-based dictionary. Assuming spatio-temporal incoherence of the aliasing noise, the dictionary matching procedure can succesfully return multiple tissue parameters estimates for each voxel. Note that the spatial localisation step ((non-uniform) FFT) and the parameter estimation step (dictionary matching) are separated like in traditional qMRI approaches.

with more parameters to be included in the signal models, the dictionary-based MRF approach may become infeasible due to memory requirements. Besides memory requirements, dictionary matching times may also become problematic. Dictionary matching times in the order of hours for estimating T_1 and T_2 maps for a 3D volume have been reported [35] and the matching times in principle also scale exponentially as more tissue parameter types are included in the reconstruction.

A second issue with the MRF approach is that, through the application of the (non-uniform) FFT on the undersampled k-spaces in the spatial localisation step, artificial aliasing noise is introduced that is not inherently part of the measured MR data. The dictionary matching procedure implicitly assumes that the aliasing noise in each voxel manifests itself as zero-mean Guassian noise. In that scenario, i.e. when spatio-temporal incoherence of the aliasing artefacts is achieved, the artificial noise is expected to result in reduced precision in the reconstructed tissue parameter maps only. However, when the aliasing noise is not incoherent, the accuracy of the reconstructed parameter maps may be negatively affected. This issue particularly manifests itself in case of Cartesian acquisitions [150]. Therefore the MRF technique is predominantly used in conjunction with non-Cartesian trajectories such as spiral and radial. At the same time, Cartesian acquisitions remain dominant in current clinical practice due to their robustness to hardware and modelling imperfections. In addition, radiologists are typically trained on contrast images generated using the conventional MRI approach with Cartesian acquisitions and are accustomed to the point spread function [137] associated with this type of acquisition. For clinical adoption of qMRI, a technique that is as fast as MRF yet works reliably with Cartesian gradient encoding schemes may be necessary.

1.4 MR-STAT

MR Spin Tomography in Time-domain ("MR-STAT") is a recently proposed technique that combines ideas from MRF together with *volumetric inversion*. With volumetric inversion, rather than following a two-step approach with separate spatial localisation and parameter estimation steps, there is only a single step during which spatial localisation and parameter estimation are performed jointly. In this single step, the tissue parameters in all voxels simultaneously are fitted against the measured k-space (or time-domain) data. No intermediate contrast images are formed in the process.

Volumetric inversion requires a forward model that takes quantitative tissue parameter *maps* as input and returns a simulated k-space signal based on a (pulse sequence dependent) physics model. The volumetric inversion technique was proposed by Block et al. in 2009 [20] in the context of pulse sequences for which analytical signal models are available and which typically encode one non-linear tissue parameter into the signal at once [189, 89, 175]. In MR-STAT,

a more generic forward model is used instead that, similar to MRF, allows arbitrary (transient-state) sequences to be utilized. More specifically, let N_v be the number of voxels within the desired field-of-view, $\boldsymbol{\alpha}_j$ the current tissue parameter estimates in voxel j, $\boldsymbol{r}_j$ the spatial coordinates of voxel j, $m(\boldsymbol{\alpha}_j, \boldsymbol{r}_j, t)$ the transverse magnetization in voxel j at timepoint t obtained through numerical integration of the Bloch equations and Δ_V the volume element for each voxel. The complex MR signal s at any arbitrary timepoint t is then modelled as the spatially discretized version of Eq. 1.1.1:

$$s(t) = \sum_{j=1}^{N_v} m(\boldsymbol{\alpha}_j, \boldsymbol{r}_j, t) e^{-2\pi i \boldsymbol{k}(t)\cdot \boldsymbol{r}_j} \Delta_V. \tag{1.4.1}$$

Now let N_t be the total number of sample points acquired during a pulse sequence and let $t_1, \ldots, t_{N_t}$ be the N_t sample times. The MR-STAT forward model is then defined as

$$\boldsymbol{s} := [s(t_1), \ldots, s(t_{N_t})] \in \mathbb{C}^{N_t}. \tag{1.4.2}$$

Next, let $\boldsymbol{d} \in \mathbb{C}^{N_t}$ be a vector of measured data obtained using an MR system. In MRF and other qMRI methods the measured data is organized into (undersampled) k-spaces in order to be Fourier-transformed. In MR-STAT, the measured data is not intended to be Fourier-transformed and instead the samples are simply stored in a vector in the order in which they were acquired. That is, the measured data is interpreted as *time-domain* data. Quantitative tissue parameter maps estimates are obtained by directly fitting 1.4.1 against the measured time-domain signal. Stated differently, if we let $\boldsymbol{\alpha}$ denote the quantitative parameter maps of interest (i.e. the concatenation of $\boldsymbol{\alpha}_j$ for all voxels j) then an MR-STAT reconstruction amounts to finding a numerical solution to the following inverse problem:

$$\text{argmin}_{\boldsymbol{\alpha}} \frac{1}{2} \|\boldsymbol{d} - \boldsymbol{s}(\boldsymbol{\alpha})\|_2^2. \tag{1.4.3}$$

Note that, rather than solving independent inversion problems in each voxel separately as in the two-step approach to qMRI, in MR-STAT a single large-scale non-linear inversion problem is solved for all voxels simultaneously. The MR-STAT workflow is summarized in Fig. 1.4. The main benefit of the one-step volumetric inversion approach employed in MR-STAT is that it naturally allows the sampling of spatial frequencies to be distributed over different contrasts [153]. This can be explained as follows. In the two-step approach, for each individual contrast image in the spatial localisation step, a Nyquist-sampled k-space must in principle be acquired (assuming no parallel imaging/compressed sensing techniques are employed). Spatial frequencies sampled for one contrast can not be used to fill in the k-space of another contrast without the FFT introducing image artefacts (e.g. blurring) that would propagate into

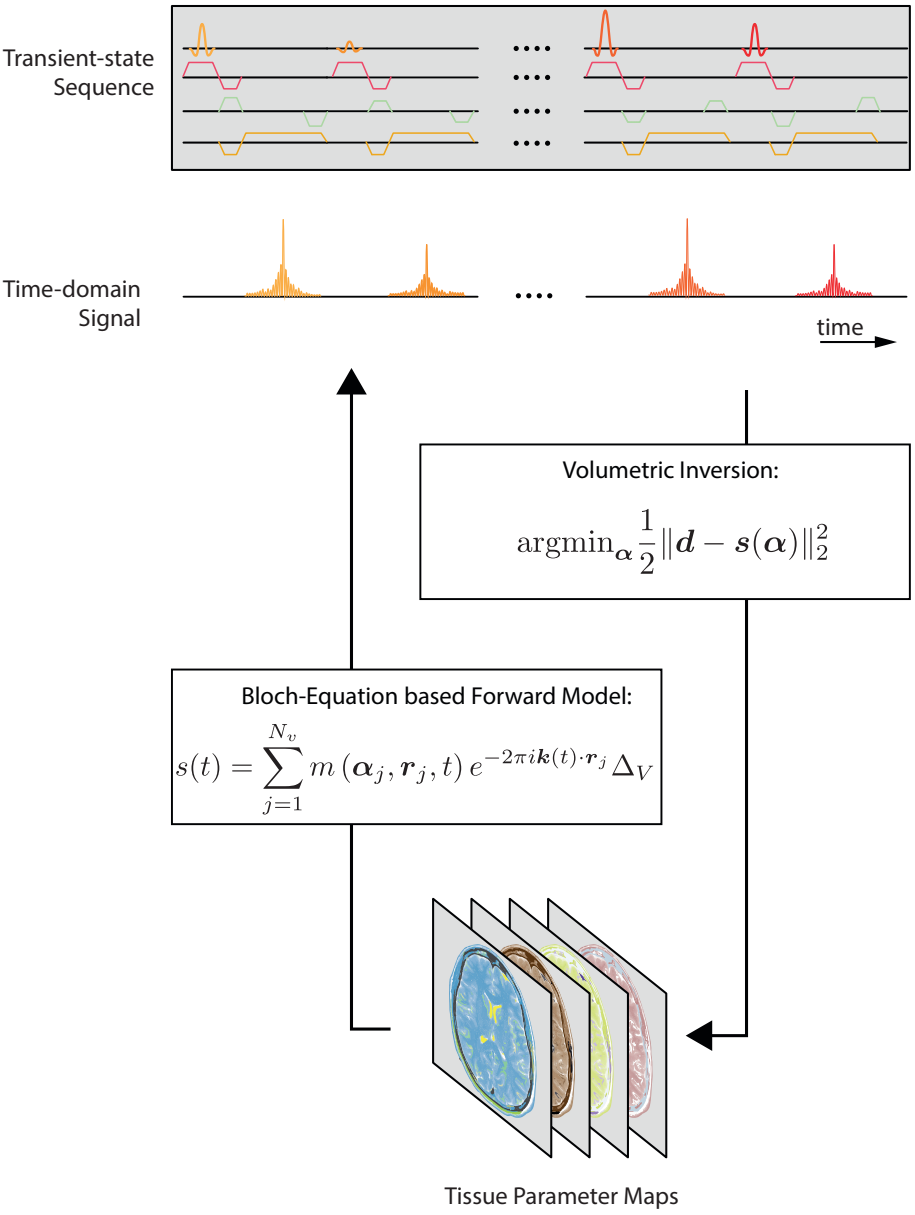

Figure 1.4: Graphical overview of the MR-STAT approach to qMRI. Like in MRF, a single transient-state sequence (in this example: a continuously varying flip angle sequence) is used that encodes information from multiple tissue parameters into the MR signal. The measured signal is not interpreted as k-space data but as time-domain data instead. A volumetric, Bloch equations-based signal model is fitted directly against the time-domain data to obtain estimates of multiple tissue parameter maps without forming intermediate contrast or snapshot images. The spatial localisation and parameter estimation steps are performed simultaneously.

the quantitative tissue parameter maps. On the other hand, in the volumetric inversion approach, both the time-varying contrast information and sampling pattern are intertwined in the forward model. Missing k-space information for one contrast is naturally complemented with k-space information from other contrasts. As such, even if a transient-state sequence with continuously changing contrast is utilized it may be possible to reconstruct tissue parameter maps free of aliasing artefacts despite each contrast being highly undersampled. Different from the MRF approach there is no artificial aliasing noise introduced into the reconstruction with MR-STAT and therefore, unlike with MRF, there is no incoherence requirement. This allows for even more freedom in the acquisition procedure compared to MRF. In particular, the MR-STAT technique is more suitable for Cartesian gradient encoding schemes, which demonstrate strongly coherent aliasing noise in the context of MRF [150].

In addition, since no pre-computed dictionary is used in MR-STAT, the curse of dimensionality that was identified with MRF (i.e. exponentially growing dictionary size as more types of tissue parameters are added) is elided.

Note that, whereas in the earlier proposed volumetric inversion techniques the FFT operator is used as part of the forward model, in MR-STAT the gradient encoding is modelled explicitly (through the $e^{-2\pi i \boldsymbol{k}(t)\cdot\boldsymbol{r}_j}$ term) rather than implictly (through the application of an FFT and an undersampling operator). A benefit of the MR-STAT approach in this respect is that spin dynamics *during* readouts, such as T_2, T_2^* and/or off-resonance induced rotation can be properly modelled. In addition, non-linear gradient fields could in principle be incorporated into the model (through modificiation of Eq. 1.1.1 and Eq. 1.1.1). Compared to other qMRI methods, the forward model in MR-STAT is more comprehensive and better describes the underlying biophysical processes during an MR examination.

A major challenge with the MR-STAT approach is the increased computational complexity of the reconstruction procedure. Due to the non-linear dependency of the signal model (i.e. Eq. 1.4.1) on the tissue parameters, there is no direct solution to the MR-STAT problem (i.e. Eq. 1.4.3). Iterative techniques must be employed instead. Such iterative techniques require an initial guess for the tissue parameter maps $\boldsymbol{\alpha}$ and then - based on certain update rules - modify the initial guess until the forward model $\boldsymbol{s}(\boldsymbol{\alpha})$ and measured data $\boldsymbol{d}$ are sufficiently close. At each iteration, relatively expensive Bloch simulations must be performed in each voxel in order to evaluate the forward model. Besides evaluating the forward model, the update rules typically require the gradient of the objective functon to be minimized (Eq. 1.4.3), which involves computing partial derivatives of the forward model with respect to the tissue parameters in all voxels (i.e. the *Jacobian matrix*). The gradient points in the direction in which the objective function increases fastest, so a gradient descent type algorithm may take a step in the direction of the negative gradient in an attempt to minimize the objective function. Convergence of such an algorithm can be improved by incorporating knowledge of the curvature of the

objective function. This knowledge is contained in the so called *Hessian matrix*, i.e. which contains second order partial derivatives of the forward model. Evaluating the forward model as well as the Jacobian and Hessian matrices is a challenge on its own due to the problem size of typical MR-STAT reconstructions. For 2D acquisitions, there can be tens of thousands of unknowns in the MR-STAT reconstruction problem. For 3D MR-STAT, problems the number of unknowns can run into the millions. The reconstruction procedure must be carefully designed to be able to deal with such large numbers of unknowns in terms of both computation time and computer memory. Another computational challenge is related to the fact that, rather than relying on the FFT in the forward operator for which highly optimized external libraries exist, the gradient encoding is modelled explicitly in MR-STAT. Custom software must therefore be designed and optimized to simulate the gradient encoding within the MR-STAT framework.

Whereas in conventional (q)MRI the acquisition procedure is designed to accommodate fast and easy image reconstructions, with MR-STAT maximum freedom is allowed in the acquisition at the expense of more challenging and lengthy reconstruction procedures. Even though in theory MR-STAT reconstructions can be performed off-line after a patient has already moved out of the MR scanner, relatively short reconstruction times remain important for acceptance of a new technique like MR-STAT into clinical workflows.

1.5 Thesis Outline

As stated in the opening section of this chapter, the main goal of the project that resulted in this thesis was to design and implement data acquisition procedures as well as image reconstruction techniques for MR-STAT to allow the technique to be utilized in a clinical demonstrator setting. Prior to starting the project, the theoretical framework for MR-STAT had already been developed and the feasibility was demonstrated in silico as well as experimentally on a pre-clinical MR system. The next step, the actual start of this project, involved designing a two-dimensional pulse sequence suitable for in vivo scanning, implementing the sequence on a clinical MR system and validating the sequence on healthy volunteers. The results of this work are reported in **Chapter 2**. The computational challenges associated with MR-STAT reconstructions are mitigated by utilizing a FFT-based decoupling technique in this work. With this technique, one-dimensional FFTs are applied along the readout direction of data from two-dimensional, Cartesian MR-STAT acquisitions. This allows the two-dimensional reconstruction problem to be reformulated as multiple independent one-dimensional reconstruction problems. For each of the individual subproblems, the variable project method is used which requires the Jacobian matrix to be stored in computer memory. Unlike in a two-dimensional setting, storing the model matrices is no problem in one-dimensional situations. In

addition, the decoupling allows for a straightforward parallel implementation of the reconstruction. While convenient on the one hand, the decoupling technique limits MR-STAT to two-dimensional Cartesian-based acquisitions and does not allow spin dynamics during readouts to be taken into account. A more generic reconstruction algorithm is therefore desired.

In **Chapter 3**, a Gauss-Newton based MR-STAT reconstruction algorithm is developed that - through a matrix-free approach - has a neglible computer memory footprint. As such, the technique can in principle be used to perform reconstructions for arbitrary, non-Cartesian MR-STAT sequences at clinically relevant resolutions. Although not demonstrated, it can in principle also be used in three-dimensional settings. The computationally demanding tasks of the reconstruction algorithm can be performed in parallel across different compute units. However, unlike with the FFT-based decoupling technique - communication between compute units is required during the reconstruction procedure and a parallel algorithm is more challenging to implement. In the work presented in Chapter 3, the algorithm was implemented on a high performance computing cluster with many CPUs and it was validated on two-dimensional Cartesian MR-STAT scans. Despite the use of many CPUs, reconstruction times are still in the order of hours per two-dimensional slice in this work.

One of the bottlenecks in the proposed MR-STAT reconstruction algorithm is formed by the need to scale the gradient of the objective function at each iteration with curvature information contained in the Gauss-Newton approximation to the Hessian matrix. In **Chapter 4**, it is recognized that the Hessian matrix - which was assumed to be too large to store into computer memory at first - admits a predictable structure. The actual structure depends on the pulse sequence used during the data acquisition. In the specific case of a pulse sequence comprising a Cartesian gradient trajectory and a smoothly varying flip angle it is demonstrated that the Hessian matrix admits a sparse structure. The entries of the (approximate) Hessian can be computed in parallel and stored in computer memory in this case. By adapting the MR-STAT reconstruction accordingly, reconstruction times in the order of fifteen minutes rather than several hours were achieved on a high performance computing cluster with many CPUs. The sparse Hessian approach can be interpreted as a generalization of the FFT-based decoupling approach. Unlike the decoupling approach, spin dynamics during readouts can be taken into account with this reconstruction technique. For non-Cartesian trajectories, the structure of the Hessian matrix is more complicated, less sparse and in general the technique may not be useful in non-Cartesian settings.

In many different fields, including the field of MRI, GPU hardware has been utilized to perform computations with better runtime performance compared to CPU-based implementations. In **Chapter 5**, a GPU-compatible MR-STAT reconstruction algorithm is introduced and validated. The Julia programming language is an indispensible tool in this context. It allows for relatively quick prototyping and great flexibility without compromising on runtime

performance. In addition, kernel functions that can be executed on GPU hardware can be written directly within Julia. Rather than a completely matrix-free Gauss-Newton approach as introduced in Chapter 3, a partially matrix-free algorithm is proposed in this work. With this algorithm, the entries of the model matrices which are the most expensive to compute are stored in GPU memory. The remaining cheaper entries are only computed at the time they are needed like in the matrix-free approach. The partially matrix-free MR-STAT reconstruction algorithm on the GPU allows for reconstruction times in the order of two minutes per slice. It also generalizes to non-Cartesian gradient trajectories. Bloch simulations form an important ingredient in the MR-STAT reconstructions but are also relevant to other qMRI techniques such as MRF. The Bloch simulations toolbox developed as part of this work is released as a standalone Julia package called BlochSimulators.jl. The package is open-source and freely available online.

In all chapters up to **Chapter 6**, all the MR-STAT validations were performed using Cartesian gradient trajectories. On the other hand, other qMRI techniques like MRF typically use non-Cartesian trajectories. Non-Cartesian trajectories may have benefits in terms of improved time-efficiency on the one hand, but are known to suffer from higher sensitivity to hardware imperfections. With MR-STAT, in principle both Cartesian and non-Cartesian trajectories can be used and the question arises which type of trajectory is more suitable in clinical settings. Therefore, in Chapter 6, Cartesian and radial MR-STAT are benchmarked in terms of time-efficiency and robustness.

Finally, in **Chapter 7**, the main findings of the individual chapters are discussed within the context of the aims of this research project. Challenges that need to be addressed are discussed and suggestions for future research directions are provided.

2

Fast quantitative MRI as a nonlinear tomography problem

Alessandro Sbrizzi
Oscar van der Heide
Martijn Cloos
Annette van der Toorn
Hans Hoogduin
Peter R. Luijten
Cornelis A. T. van den Berg

Magnetic Resonance Imaging, 2018, 46: 56-63.
https://doi.org/10.1016/j.mri.2017.10.015

Abstract

Quantitative Magnetic Resonance Imaging (MRI) is based on a two-steps approach: estimation of the magnetic moments distribution inside the body, followed by a voxel-by-voxel quantification of the human tissue properties. This splitting simplifies the computations but poses several constraints on the measurement process, limiting its efficiency. Here, we perform quantitative MRI as a one step process; signal localization and parameter quantification are simultaneously obtained by the solution of a large scale nonlinear inversion problem based on first-principles. As a consequence, the constraints on the measurement process can be relaxed and acquisition schemes that are time efficient and widely available in clinical MRI scanners can be employed. We show that the nonlinear tomography approach is applicable to MRI and returns human tissue maps from very short experiments.

2.1 Introduction

The possibility to store and process vast amounts of data at increasingly faster rates has boosted the application of numerical methods in physical sciences. Nowadays, solutions can be found to problems with hundred thousands or millions of unknowns [170, 90]. A representative example is seismic full waveform inversion [169]; the underlying process is based on a wave equation which is nonlinear in the spatially-dependent unknowns. The reconstruction over 2D or 3D regions of the Earth's interior is obtained by means of iterative algorithms. It is even possible to estimate multiple parameters simultaneously, such as wave velocity, density, anisotropy and attenuation.
Analogously to seismic waveform inversion, quantitative magnetic resonance imaging (qMRI) aims at reconstructing several parameters which characterize the internal structure of the human tissue; in particular, the proton density (ρ), the longitudinal (T_1) and transverse (T_2) relaxation rates, among others.
One important difference between tomographic techniques and state of the art qMRI lies in their methodology. Quantitative MRI is built upon a two step approach. Firstly, each local contribution to the volumetric signal is estimated (signal localization), returning spatial maps of the transverse magnetic moment; this is usually achieved by applying a multi dimensional inverse Fourier transform to the data. Subsequently, the tissue parameters quantification is carried out for each location separately. The second step (parameter estimation) is thus obtained from a series of magnetization images by fitting relatively simplistic signal models [112] or by searching over a dictionary of complex signal fingerprints [110, 36].
This separation leads to a simplified computational process but with significant costs. In order to satisfy the stringent criteria for Fourier encoding, one has to assume that the signal evolution during the read-out only reflects the intended gradient encoding. Long single-shot read-outs generally violate this condition, leading to image artifacts, e.g., geometrical distortion and intra-voxel dephasing. To avoid such artifacts, most clinical MR sequences have been designed to manipulate the nuclear spins into a reproducible state, which allows multiple measurements to be aggregated into one coherent frequency representation of the desired image (k-space). Consequently, MRI scans can be relatively time consuming when compared to CT or PET exams. Additionally, due to the overly simplifying assumptions in the Fourier encoding-based signal model, system imperfections such as off-resonances and radiofrequency field inhomogeneity are not easily taken into account.
MR Fingerprinting (MRF) [110] has shown a great potential to recover multiparametric maps from unprecedented short acquisitions allowing strong aliasing artifacts to exist in each of the individual images. The RF excitation and gradient acquisition schemes need to be designed properly to ensure incoherence between the signal and the undersampling artifacts which are interpreted

as zero-mean noise-like perturbations. Interleaved spiral [110] and radial [36] readout gradients are therefore preferred. These type of sequences are however, prone to gradient system imperfections such as eddy currents and thus require an additional sophisticated calibration of the hardware [156].
In this work, we pose quantitative MRI as a nonlinear tomographic problem by directly utilizing the fundamental relationship between the time-varying signal and the laws of physics that describe the experiment. Thereby, we unify the traditionally disjoined processes of signal localization and parameter estimation into one process. The macroscopic ensemble of magnetic spins in the body is treated as a large-scale nonlinear dynamical system, which is probed by superimposing a train of radiofrequency (RF) excitations and gradient fields. The tissue properties are obtained by inversion of the underlying large scale nonlinear model. We name this method MR-STAT, which stands for Magnetic Resonance Spin TomograAphy in Time-domain. We show that quantitative parameter maps can be accurately reconstructed by employing nonlinear optimization algorithms and parallel computing infrastructures which do not necessarily rely on the Fourier decoding step for spatial localization. The data collection process can thus be liberated from the standard sequence design constraints and very short acquisitions (order of seconds) provides sufficient data for correct reconstructions. Although the time-domain formulation would in principle accommodate any read-out strategy, we show that established, experimentally robust cartesian gradient acquisition schemes can also be employed; a step which should facilitate the translation of the technique to clinical MRI systems. Finally, MR-STAT is also able to estimate the precision of the reconstructed multi-parametric maps; another important step towards the clinical application of qMRI.

2.2 Theory

2.2.1 The coupled space-time signal model

The behavior of the space/time dependent magnetization vector, $\mathbf{m}(\mathbf{r},t)$ is determined by superimposed radiofrequency and gradient magnetic fields, respectively denoted by $b(t)$ and $\mathbf{G}(t)\cdot\mathbf{r}$. The response of the magnetic spins is also affected by the $T_1(\mathbf{r})$ and $T_2(\mathbf{r})$ relaxation rates, which carry diagnostic information. The relationship between all these quantities is given locally by the Bloch equation [82]:

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{m} = \gamma\mathbf{b}\times\mathbf{m} - \mathbf{q} \tag{2.2.1}$$

where

$$\mathbf{b} = \begin{pmatrix} \mathrm{Re}\{b\} \\ \mathrm{Im}\{b\} \\ \mathbf{G}\cdot\mathbf{r} \end{pmatrix},\ \mathbf{q} = \begin{pmatrix} \frac{m_x}{T_2} \\ \frac{m_y}{T_2} \\ \frac{m_z-1}{T_1} \end{pmatrix},\ \mathbf{m}(\mathbf{r},0) = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

and γ denotes the gyromagnetic ratio.
The signal, s, from a receiver coil is obtained from Faraday's law of induction [29]:

$$s(t) = \int_V \rho(\mathbf{r}) m(\mathbf{r}, t) \mathrm{d}\mathbf{r} \tag{2.2.2}$$

where ρ denotes the proton density of the tissue weighted by the spatially varying complex receive RF field B_1^-. m is the transverse component of $\mathbf{m}$ and V is the volume enclosing the spins which emit signal.
The first step in qMRI typically aims at reconstructing the spatially dependent magnetization state. This is achieved by designing the experiment such that the signal can be modeled as:

$$s(t) = s(\mathbf{k}(t)) = \int_V \rho(\mathbf{r}) m^*(\mathbf{r}) e^{-2\pi \imath \mathbf{k}(t)\cdot\mathbf{r}} \mathrm{d}\mathbf{r} \tag{2.2.3}$$

where m^* must be a time-independent state of the magnetization and $\mathbf{k}$ represents the accumulating effect of the gradient fields. Note that the system response is decoupled into a space-only dependent component $\rho\, m^*$ and a Fourier encoding term $\exp(-2\pi\imath\mathbf{k}(t)\cdot\mathbf{r})$ which is independent from tissue parameters. The unknown term is thus $\rho\, m^*$. If Fourier transform requirements are fulfilled by the experimental settings, Inverse Fourier transform can be applied to the data to reconstruct $\rho\, m^*$, obtaining thus a magnetization image. This decoupled approach typically leads to either long measurement times (m^* must be in the steady-states or in static equilibrium) or to large reconstruction artifacts if the Nyquist sampling criterion is not fulfilled [110]. In the subsequent step, model-fitting strategies based on the Bloch eq. (2.2.1) can be applied to each voxel separately to recover the tissue parameters on a local level. In the MR fingerprinting case [110], this is performed by an exhaustive search over a pre-computed dictionary of signals; a reconstruction strategy which although robust and straightforward, is undermined by the large dictionaries needed for high dimensional multi-parametric data. Furthermore, even a slight modification of a sequence requires an ad-hoc computation of the corresponding dictionary.
Instead of relying on the standard decoupled Fourier model, we reconsider the coupled space-time equation, Eq. (2.2.2), and solve it directly. Denoting by $d(t)$ the demodulated signal measured by the receiving coil of the MR scanner, the resulting tomographic approach is:

Find the system's parameters, $\boldsymbol{\alpha}$, that minimize

$$\int_\tau |s(\boldsymbol{\alpha}, t) - d(t)|^2 \mathrm{d}t$$

such that the Bloch equations (Eq. 2.2.1) and Faraday's law (Eq.)2.2.2) hold. (2.2.4)

In the equation, τ denotes the union of temporal acquisition intervals and $\boldsymbol{\alpha}$ represents the unknown parameters over the whole region. Note that the reconstruction acts on the signal in the time domain to directly derive the spatial distribution of the tissue's characteristics. In the MR-STAT framework, the link between temporal and spatial domain is still provided by the gradient fields, but now the k-space data set constitutes a non-trivial entanglement of spatial and spin-dynamic information.
During an MR-STAT experiment, the magnetization is thus no longer expected to be in steady-states or equilibrium conditions but is free to evolve. Since there are no particular requirements on the state of the system, the excitation/acquisition scheme can be designed to boost the time-efficiency and to minimize the impact of gradient hardware imperfections. In this work, we consider measurement schemes (sequences) where RF excitation pulses and acquisition intervals are contiguous, thus the repetition time T_R and echo-time T_E are kept as short as possible (see Fig. 2.1); there are no dead times and the data collection rate is thus maximized. We choose to employ a so-called gradient read-out scheme which is the standard acquisition modality due to its robustness with respect to hardware imperfections.

Since the reconstruction process no longer relies upon Fourier decoding, the

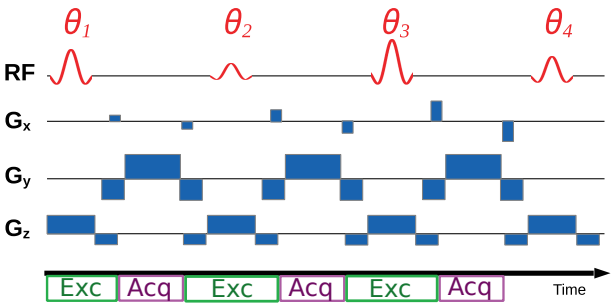

Figure 2.1: Fragment of an MR-STAT data acquisition sequence. The spatially selective RF pulse is scaled by the tip angles θ_j. G_x and G_y are encoding gradients. G_z is the slice selective gradient. Note that the excitation (**Exc**) and acquisition (**Acq**) intervals follow one another without interruption, that is, the fixed echo and repetition times are the shortest possible.

underlying physical model can be easily expanded to include system imperfections such as off-resonance frequency, $\omega(\mathbf{r})$, and transmit RF fields heterogeneity, $B_1^+(\mathbf{r})$. These quantities enter the reconstruction problem (2.2.4) through the vector of applied magnetic field $\mathbf{b}$ in the Bloch equation (2.2.1):

$$\mathbf{b} = \left(\mathrm{Re}\left\{B_1^+ b\right\}, \mathrm{Im}\left\{B_1^+ b\right\}, \mathbf{G}\cdot\mathbf{r} + \omega/\gamma\right)^T.$$

Consequently, the extended set of unknowns in the MR-STAT equation (2.2.4) is

$$\boldsymbol{\alpha} = (T_1, T_2, |\rho|, \angle\rho, |B_1^+|, \omega).$$

The MR-STAT reconstruction problem (Eq. (2.2.4)) can be solved by a generic purpose derivative-based nonlinear minimization algorithm upon the discretization of the spatial and temporal domains. See the Methods section for the implementation details.

Spatial encoding, identifiability and precision estimates

The encoding capability of the MR-STAT approach can be derived by standard techniques in inversion theory. In particular, the identifiability of a system's parameters [81] is reflected by the covariance matrix $\mathbf{C} \equiv \eta^2(\mathbf{D}^T\mathbf{D})^{-1}$ where $\mathbf{D}$ is the Jacobian matrix of the model with respect to the parameters $\boldsymbol{\alpha}$ and η is the noise variance.
To minimize noise amplifications, $\mathbf{C}$ should have a moderate condition number. This depends on both the acquisition length as well as the spatial resolution: for a fixed reconstruction grid, decreasing the sequence length leads to a more ill-conditioned matrix $\mathbf{C}$ and noise perturbations or model imperfections are thus amplified. In the extreme case that the sequence is too short, $\mathbf{C}$ becomes rank-deficient (infinitely large condition number) and the uniqueness of the solution is no longer guaranteed unless other regularization terms are introduced. This is analogous to reconstructions of undersampled k-space data in, for example, compressed sensing MRI [109, 50].
To illustrate this theoretical analysis with a concrete example, we consider a homogeneous object with properties:

$$(T_1, T_2, |\rho|, \angle\rho, |B_1^+|, \omega) = (0.833[\text{s}], 0.083[\text{s}], 1[\text{a.u.}], 0[\text{rad}], 1[\text{a.u.}], 0[\text{Hz}])$$

and construct $\mathbf{C}$ for varying spatial resolution and sequence length. The latter is expressed in terms of the number of readout lines in the sequence. The flip angles are randomly drawn from a normal distribution centered around 0 (see also the top of Fig. 2.2). The conditioning of the covariance matrix is reported in Fig. 2.3. As expected, the longer the sequence, the lower the noise amplification. The number of unknowns increases with the grid size, leading to a larger scale problem requiring more data (longer sequences) to be fully determined and to be robust to noise perturbations. When $\mathbf{C}$ has full rank, the MR-STAT problem is fully determined and the algorithm returns not only the parameter maps but also their spatially dependent standard deviations. The standard deviation of the n-th parameter is given by $\sigma_n \approx \sqrt{[\mathbf{C}]_{n,n}}$. Note the analogy between σ_n and the so-called geometry factor (g-factor) in parallel imaging [135].

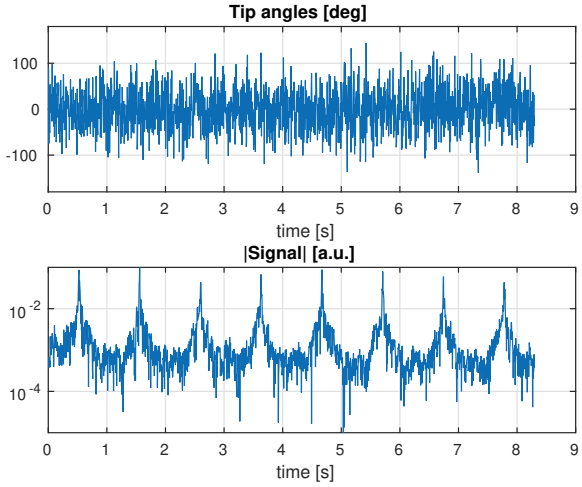

Figure 2.2: Tip angles and time-domain signal for the MR-STAT sequence applied to the in-silico simulated head experiment at 3T.

2.3 Methods

2.3.1 Implementation

For reasons that will soon become clear, we split the vector of unknowns in two parts, namely: $\boldsymbol{\alpha} = (\boldsymbol{\rho}, \boldsymbol{\beta})$ where $\boldsymbol{\beta}$ contains the spatial distribution of $(T_1, T_2, |B_1^+|, \omega)$. Given a demodulated dataset in the time domain, $d(t)$, the reconstructed parameter maps, $(\boldsymbol{\rho}^{\text{rec}}, \boldsymbol{\beta}^{\text{rec}})$, are obtained by solving the following nonlinear least squares problem, which is derived upon the discretization of Eq. (2.2.4):

$$(\boldsymbol{\rho}^{\text{rec}}, \boldsymbol{\beta}^{\text{rec}}) = \arg\min_{\boldsymbol{\rho},\boldsymbol{\beta}} \sum_{j=1}^{J} \left| d_j - \sum_{r=1}^{R} \rho_r m_{j,r}(T_{1,r}, T_{2,r}, |B_{1,r}^+|, \omega_r) \Delta_s \right|^2 \Delta_t,$$

$$\text{such that Eq. 2.2.1 holds.} \tag{2.3.1}$$

The first and second sum in the objective function approximate, respectively, the time and the volume integral from Eq. (2.2.4) and Eq. (2.2.2). J is the total number of acquired data samples, R is the number of spatial grid points, Δ_s and Δ_t are, respectively, the space and time discretization intervals. Using

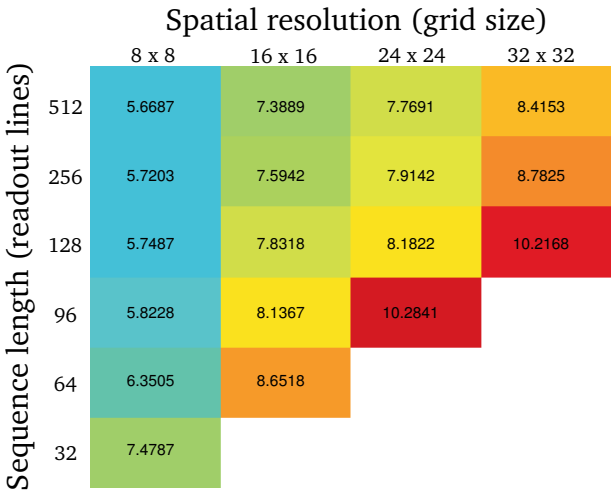

Figure 2.3: Noise amplification in the MR-STAT experiment when all six parameters $(T_1, T_2, |\rho|, \angle\rho, |B_1^+|, \omega)$ are reconstructed. The numerical values inside the table refer to the $\log_{10}$ of the condition number of the covariance matrix $\mathbf{C}$. Large values mean large noise amplification. The condition number is reported as a function of the experiment length (numbers of readout lines) and spatial resolution (grid sizes) for a small scale, homogeneous in-silico model. The number of samples per readout line is equal to the number of grid points along one dimension. An empty cell means that $\mathbf{C}$ is rank deficient (infinite condition number) and the problem can not be solved.

matrix-vector notation, Eq. (2.3.1) can be written as:

$$(\boldsymbol{\rho}^{\text{rec}}, \boldsymbol{\beta}^{\text{rec}}) = \arg\min_{\boldsymbol{\rho},\boldsymbol{\beta}} \|\mathbf{d} - \mathbf{M}(\boldsymbol{\beta})\boldsymbol{\rho}\|^2 \quad \text{such that Eq. 2.2.1 holds} \tag{2.3.2}$$

where the matrix $\mathbf{M}(\boldsymbol{\beta})$ is given by

$$[\mathbf{M}(\boldsymbol{\beta})]_{j,r} \equiv \Delta_s m_{j,r}(T_{1,r}, T_{2,r}, |B_{1,r}^+|, \omega_r).$$

Since the reconstruction problem is nonlinearly dependent on $\boldsymbol{\beta}$ and linearly dependent on $\boldsymbol{\rho}$, it can be solved by the variable projection method (VARPRO) [57]. Note that, if the vector $\boldsymbol{\beta}$ was a solution of Eq. (2.3.2), then the parameters $\boldsymbol{\rho}$ could be found by solving a *linear* least squares problem, whose solution is given by

$$\boldsymbol{\rho} = \mathbf{M}^\dagger(\boldsymbol{\beta})\mathbf{d} \tag{2.3.3}$$

where $\mathbf{M}^\dagger$ is the pseudo-inverse of $\mathbf{M}$. Substituting this back into Eq. (2.3.2) we obtain the reduced functional:

$$\boldsymbol{\beta}^* = \arg\min_{\boldsymbol{\beta}} \left\| \left[\mathbf{I} - \mathbf{M}(\boldsymbol{\beta})\mathbf{M}^\dagger(\boldsymbol{\beta})\right] \mathbf{d} \right\|^2. \tag{2.3.4}$$

Note that the linear parameter no longer plays a role in the equation.
VARPRO solves Eq. (2.3.2) by first solving the reduced nonlinear problem in Eq. (2.3.4). The optimal linear parameters are eventually found by substitution into Eq. (2.3.3): $\boldsymbol{\rho}^* = \mathbf{M}^\dagger(\boldsymbol{\beta}^*)\mathbf{d}$.
Solving Eq. (2.3.4) instead of Eq. (2.3.2) results in a faster and more robust convergence for non-convex problems. Additionally, initial guesses for $\boldsymbol{\rho}$ are unnecessary.
The largest computational burden for solving Eq. (2.3.4) is given by the calculation of the derivatives of the system matrix $\mathbf{M}$ with respect to the nonlinear variables, that is: $\mathrm{d}\mathbf{M}(\boldsymbol{\beta})/\mathrm{d}\boldsymbol{\beta}$. In this work, they are calculated by first order forward finite difference approximations. We point out that the VARPRO method has many applications and has even been used to solve different MR problems before [21, 74, 162].
The minimization problem is implemented in Matlab making use of the built-in trust region minimization algorithm and the VARPRO implementation given by [125]. The Bloch equation simulator is implemented in C [64] and was adapted to include slice profile response, off-resonance effects and B_1^+ inhomogeneities. The reconstruction is halted after 30 iterations or earlier if the maximum component of the gradient of the objective function is smaller than 10^{-6} (first order optimality measure).
Unless otherwise stated, the reconstruction algorithm is initialized with the following values:

$$(T_1, T_2, |B_1^+|, \omega)^{\text{start}} = (1.0[\text{s}], 0.1[\text{s}], 1.0[\text{a.u.}], 0.0[\text{Hz}]).$$

These values are uniform over the whole FOV. As explained in the Implementation subsection, the (complex) proton density variable need not be initialized since it is reconstructed by solving a standard linear least squares problem.

2.3.1.1 Computational complexity and parallelization

On the computation side, the MR-STAT reconstruction problem for a 2D or 3D geometry at realistic spatial resolution is extremely demanding. Since all parameter maps are reconstructed at once, the number of unknowns is vast. To illustrate: for a 2D acquisition of a $N_s \times N_s$ voxels grid, the number of unknowns is $N_s^2 \times 6$ since there are 6 parameters per voxel. Since $N_s \sim \mathcal{O}(10^2)$, the total number of unknowns is $\mathcal{O}(10^5)$. As a consequence, the number of data points should also be $\mathcal{O}(10^5)$. the response of the system also has to be calculated in the slice selective direction to correctly incorporate the effect of the slice profile. The reconstruction algorithm must calculate the response of the physical equations for $\mathcal{O}(10^5)$ voxels over $\mathcal{O}(10^5)$ time points.
For the second and third reconstruction tests in this work (see below), we parallelize the computations in the following way: suppose that we employ a Cartesian acquisition scheme with the read-out direction along the y-axis; in

this case, the signal, s_j, over the j-th read-out interval, τ_j, is given by

$$s_j(t) \propto \int_{X\times Y\times Z} m(\mathbf{r}, t_j) e^{\frac{t_j - t}{T_2}} e^{\imath(t-t_j)\omega(\mathbf{r})} e^{-\imath\gamma G_y (t-t_j) y} \mathrm{d}\mathbf{r}$$

where the 3D integration interval $X \times Y \times Z \subset \mathbb{R}^3$ contains all nuclear spins emitting a signal. Given that for this kind of sequence, the duration of the read-out τ_j is only one millisecond or less, we can neglect the T_2 decay and the dephasing due to ω. The signal equation becomes (we use the 1D k-space notation: $k_y \equiv \gamma/2\pi \int_{t_j}^{t} G_y(\tau)\mathrm{d}\tau$):

$$s_j(k_y) \approx \int_X \int_Y \int_Z m(x, y, z, t_j) e^{-i2\pi k_y y} \mathrm{d}x\mathrm{d}y\mathrm{d}z$$

and applying 1D Fourier Transform along the y direction, $\mathcal{F}^y$:

$$\mathcal{F}^y s_j(\tilde{y}) \approx \int_X \int_Z m(x, \tilde{y}, z, t_j) \mathrm{d}x\mathrm{d}z$$

$\mathcal{F}^y s_j(\tilde{y})$ represents the signal generated at time t_j by the nuclear spins located in the 2D interval $X \times Z$ at the y-coordinate given by $\tilde{y}$. The signal from spins with different y-coordinates does not contribute to $\mathcal{F}^y s_j(\tilde{y})$. In other words: the MR-STAT reconstruction problem can be decomposed into many independent subproblems, each one corresponding to a given coordinate $\tilde{y}_n$ with $n = 1, \ldots, N_s$. Parallelization is thus carried out by assigning each subproblem to a different computing core. The reconstruction time is defined as the longest runtime amongst all jobs.
The whole code is compiled as a Linux stand-alone executable and deployed to the High Performance Computing cluster of the UMC Utrecht by linking it to the corresponding Matlab run-time library.

2.3.2 Reconstructions

To demonstrate the design flexibility of MR-STAT, we employ several types of acquisition schemes: one where the tip angles are randomly drawn from a normal distribution (Fig. 2.2); one which follows a sinusoidal pattern where each lobe is weighted by a randomly chosen value (Fig. 2.4-Top) and one with piecewise constant excitations (Fig. 2.5-Top). For the latter RF-train, each constant tip angle section is preceded and followed by a half-angle pulse acting, respectively, as excitation and tip-back pulses. All the sequences start with a 180° inversion pulse. Each read-out interval is centered between excitations and all gradients are balanced, thus a single isochromat accurately represents the dynamics of a voxel.

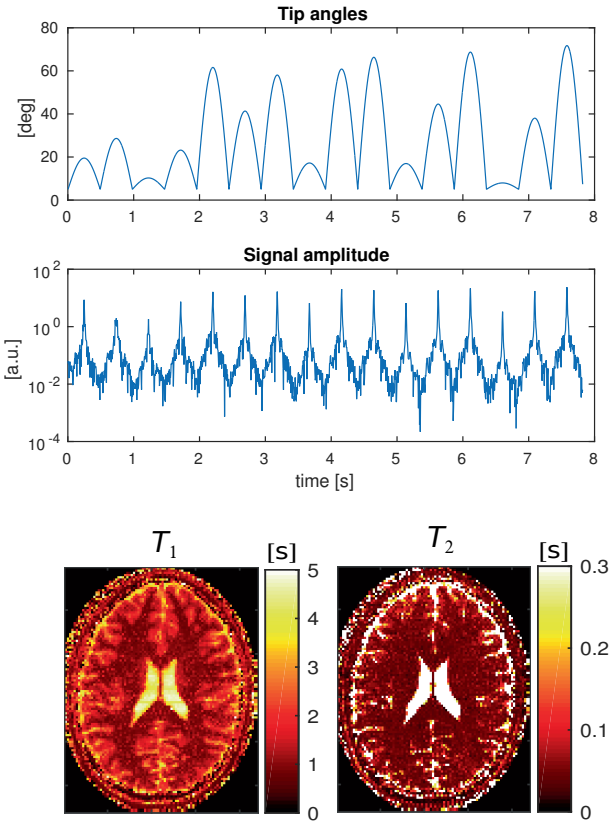

Figure 2.4: In-vivo experimental validation of MR-STAT for a sinusoidal RF train sweep. From top to bottom: the flip angle train, the recorded signal and the reconstructed parameter maps.

2.3.2.1 *In silica* low resolution reconstruction

A simple 2D object made of three homogeneous compartments is reconstructed on a 32×32 grid (See Fig. 2.6). The T_1 and T_2 rates for the three compartments A, B, and C correspond to cerebrospinal fluid (CSF), gray and white matter values, respectively. In this case, the off-resonance and transmit RF maps were set to $\omega = 0$ Hz and $B_1^+ = 1$ [a.u.], respectively. A random RF excitation train is applied analogously to the one shown in Fig. 2.2. Two-hundred and fifty-six RF pulses are interleaved with a 2D Cartesian read-out gradient scheme consisting of 32 phase encoding steps which are repeated 8 times. The resulting sequence duration is 1.2 seconds. Gaussian noise is superimposed to the time-domain signal such that $\|\text{noise}\|/\|\text{signal}\| = 0.01$.

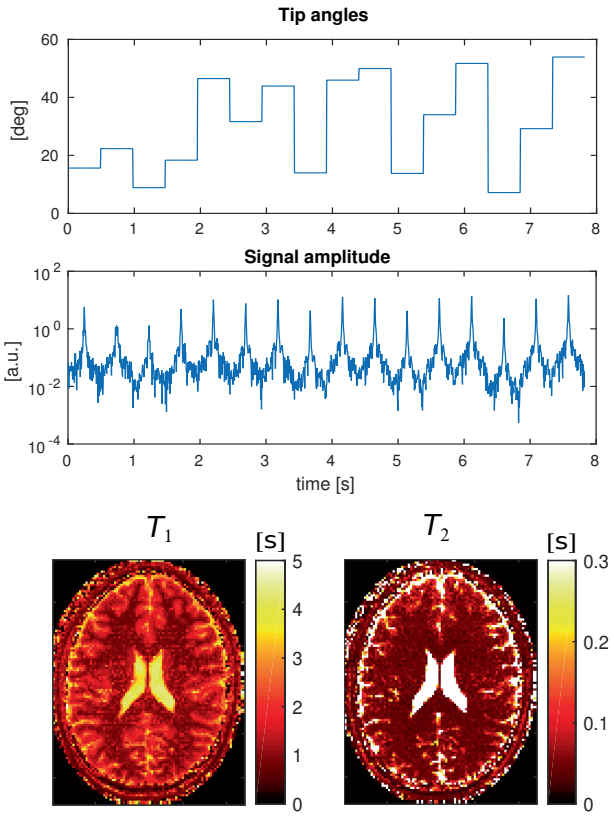

Figure 2.5: In-vivo experimental validation of MR-STAT for a piecewise constant tip angle excitation. From top to bottom: the flip angle train, the recorded signal and the reconstructed parameter maps.

2.3.2.2 *In-silica* high resolution reconstruction

The central slice of a numerical human brain model [9] is used to create a synthetic MR-STAT data set. The reconstructed in-plane resolution is 1mm×1mm which corresponds to a 216×216 voxels matrix. The tissue parameters for the biological components are given in Table 2.1. The amplitude and phase maps of the transmit RF field are obtained from a numerical electromagnetic simulation of a 3T headcoil driven in quadrature. Without loss of generality, a uniform receive sensitivity is assumed in this example. The off-resonance map is taken from [87] and is scaled to fit the range of $[-15, 15]$ Hz in the head (see the bottom of Fig. 2.7). For the acquisition, a Cartesian trajectory is used. The duration of each read out is 0.86 ms with a 4 μs dwell time per sample. The read out lines (k_y direction) cover the 2D k-space in ascending order, start-

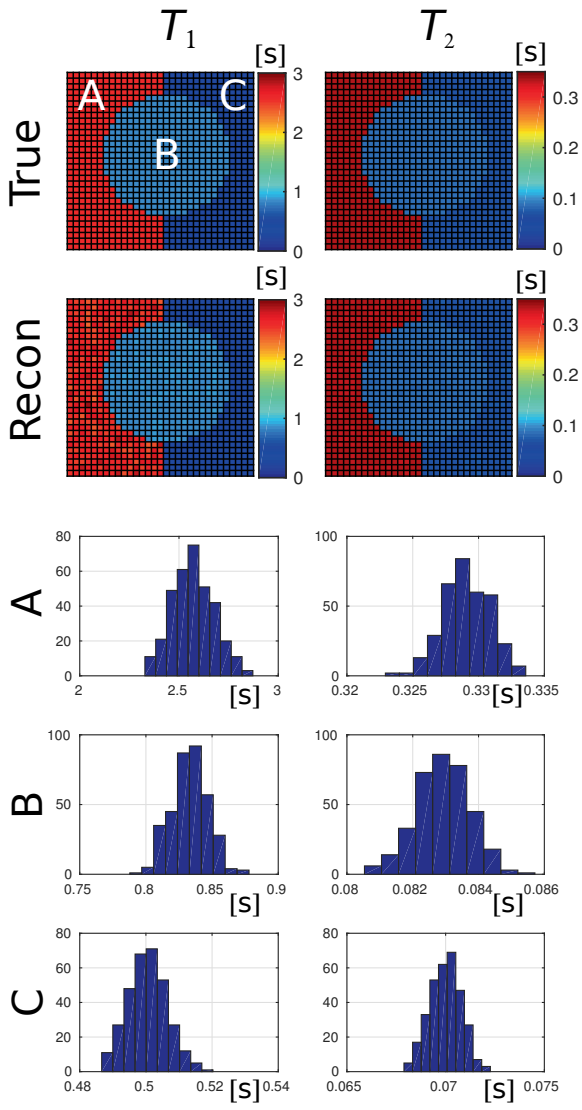

Figure 2.6: Precision estimate test. A simple 2D object (top row) undergoes a simulated MR-STAT acquisition and reconstruction. The reconstructed T_1 (left) and T_2 (right) maps are shown on the second row. The histogram plots report the distribution of the reconstructed values over each compartment A, B and C. The standard deviations of these distributions are reported in Table 2.2 and show great similarity with the estimated standard deviation values.

Table 2.1: True and Mean values of the reconstructed relaxation times per tissue type. The standard deviation of the reconstructed values for each tissue type is reported in brackets.

	T_1 [ms]			T_2 [ms]		
	true	recon	(std)	true	recon	(std)
CSF	2569	2565.7	(±38.9)	329	329.1	(±2.8)
Gray m.	833	833.4	(±18.9)	83	83.0	(±0.8)
White m.	500	500.9	(±12.2)	70	70.0	(±0.6)
Fat	350	352.2	(±8.9)	70	70.0	(±0.5)
Muscle	1000	1000.6	(±31.0)	47	47.0	(±0.6)
Skin	569	570.1	(±7.7)	329	328.3	(±4.0)
Blood	1700	1699.3	(±21.7)	300	299.6	(±2.5)
Dura	2000	2001.1	(±41.1)	280	279.2	(±5.2)

ing with the smallest negative values of k_x and repeating this pattern for the equivalent of 8 full k-space coverages. In total, 1728 lines are acquired in 8.3 seconds resulting in approximately $3.7 \cdot 10^5$ time data points. The random tip angles sequence is shown at the top of Fig. 2.2.
A Gaussian shaped RF pulse and a slice selective gradient waveform along the z axis are applied. The RF pulse is 1 ms long and is defined on a 0.1 ms dwell time step. The slice profile variation throughout the sequence is taken into account by discretizing the spatial domain in the slice-selective direction by 50 points and integrating the magnetization response for each point. This integration is applied to both the forward (signal simulation) and backward (reconstruction) steps. Gaussian noise is superimposed to the time-domain signal such that $\|\text{noise}\|_2/\|\text{signal}\|_2 = 0.01$. The resulting time-domain signal is shown at the bottom of Fig. 2.2.
The parameter ω is initialized by applying a median filter to the true off-resonance map. In experimental practice, this dataset could be generated with a fast B_0 calibration scan. The other parameters are initialized with the same values as reported in the Implementation subsection.

2.3.2.3 *In-vivo* experimental demonstration at 3.0 Tesla

Finally, MR-STAT is implemented on a 3T whole-body MR system (Philips-Ingenia). A single slice is acquired for a brain of a healthy volunteer with a 15 channel receive head-coil. Written informed consent from the volunteer participating in this experiment was obtained.
We employ two different sequences. The first RF train (Fig. 2.4, top) consists of 16 sinusoidal sweeps. Each lobe corresponds to a k-space filling and is randomly scaled to achieve maximum amplitude levels in the range $5^o \leq \theta \leq 75^o$.
The second RF train (Fig. 2.5, top) consists of piecewise constant flip angles, whose values are drawn from a uniform distribution in the range $[5^o, 60^o]$. Each

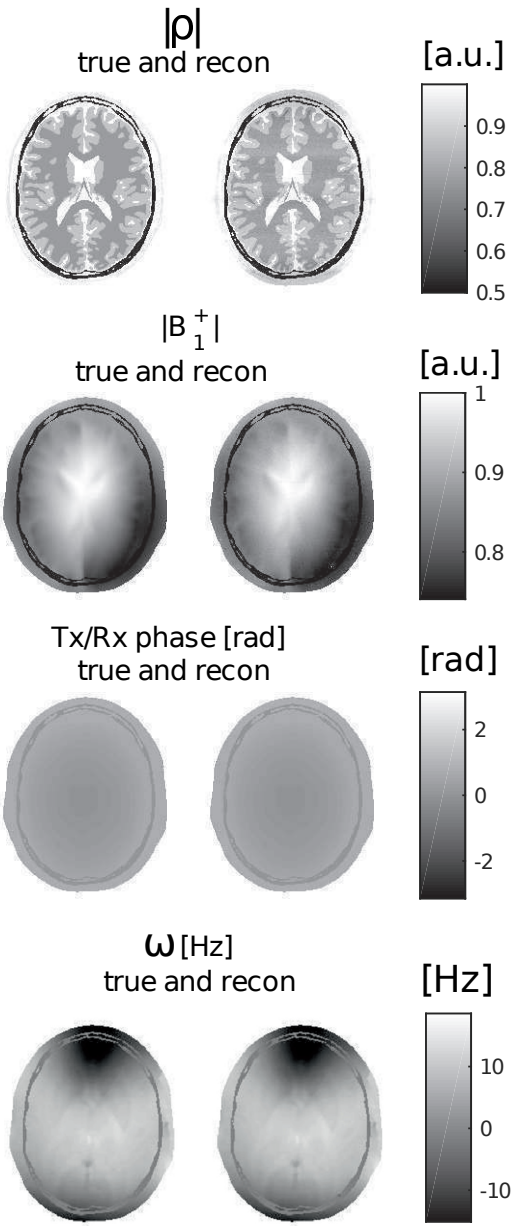

Figure 2.7: True and reconstructed maps of proton density, transceive phase, $|B_1^+|$ and ω.

of the 16 k-space fillings is thus characterized by the same tip angle excitation. In addition, a half-angle pre-pulse and a half angle tip-back pulse are applied, respectively, before and after each segment.
In both sequences, the excitation phases alternate between 0° and 180°. A Gaussian shaped RF pulse with duration 0.81 ms and a slice selective gradient are employed to achieve a 3mm slice thickness. The shortest possible values for T_E and T_R are chosen, namely $(T_E, T_R) = (2.78, 5.56)$ ms. The sequences are preceded by an adiabatic inversion pulse. The sequence parameters are converted into MATLAB format and imported in the reconstruction software. Analogously to the synthetic case, the slice profile variation across the sequence is included in the model by simulating the RF pulses on a 15 μs grid and taking 11 samples along the slice direction. As starting values for ω we choose 0 Hz everywhere.
The spatial resolution is 1.8×1.8 mm^2 and the scan time is 7.8 seconds. The measured signals are shown in Figures 2.4 and 2.5.
In these two tests, we reconstruct T_1 and T_2 values and we treat the other parameters as nuisance variables, that is, they are considered unknown but their estimation is not required to be precise.

2.4 Results

2.4.1 *In silica* low-resolution reconstruction

Fig. 2.6 illustrates the application of MR-STAT to the small scale reconstruction test. The distribution of reconstructed values from each compartment are reported in the histogram plots. The standard deviations as estimated from the covariance matrix $\mathbf{C}$ are averaged over each compartment and are reported in Table 2.2. In the same Table, also the true standard deviations obtained

	T_1		T_2	
Compartment	$\sqrt{[\mathbf{C}]_{n,n}}$	std of recon	$\sqrt{[\mathbf{C}]_{n,n}}$	std of recon
A (CSF)	112.8 [ms]	114.1 [ms]	2.0 [ms]	1.8 [ms]
B (Gray m.)	16.1 [ms]	14.2 [ms]	0.9 [ms]	0.8 [ms]
C (White m.)	6.6 [ms]	5.8 [ms]	0.9 [ms]	0.8 [ms]

Table 2.2: Precision estimation test. Estimated standard deviations per compartment as derived from the covariance matrix $\mathbf{C}$ versus the true standard deviation calculated after the reconstruction. The estimated precision levels are very close to the obtained ones.

from the reconstructed values are reported. These are calculated as

$$\sqrt{\frac{1}{K-1}\sum_{k=1}^{K} |T_i^{\text{recon}} - T_i^{\text{true}}|^2}, \quad i = 1, 2$$

where K is the number of voxels in a given compartment. From Table 1 it is clear that not only the T_1 and T_2 values are accurately reconstructed (as shown in Fig. 2.6), but also the estimated and truly obtained precision levels are very similar.
The convergence curve for the reconstruction algorithm is reported in Fig. 2.8 and displays the relative residual norm as a function of the iteration number, that is, the model-data misfit normalized on the norm of the data:

$$\text{relative residual norm} = \frac{\|\mathbf{d} - \mathbf{M}(\boldsymbol{\beta})\boldsymbol{\rho}\|}{\|\mathbf{d}\|}.$$

The data-model misfit eventually reaches the noise level after 5 iterations and

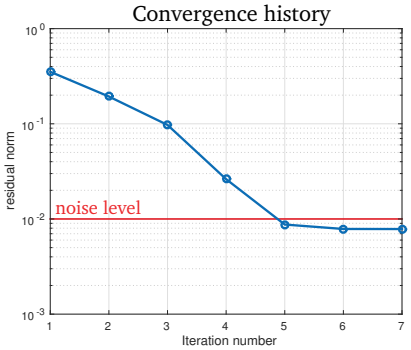

Figure 2.8: Convergence curve of the MR-STAT reconstruction algorithm for the precision estimation test (see also Fig. 2.6). The relative residual norm (data-model misfit normalized on the norm of the data) is reported as a function of the iteration number. Note that the algorithm eventually converges to the thermal noise level.

the algorithm halts soon afterwards.

2.4.2 *In-silica* high resolution reconstruction

Beside T_1, T_2 and ρ, also the transmit field profile and off-resonance map are reconstructed; they are displayed in Figures 2.9 and 2.7. They closely agree with the true values. In Table 2.1, the mean values and corresponding variations over each tissue type are reported and show high precision.
The root-mean-squared-errors (RMSE) for the B_1^+ and ω maps are also very small, namely:

$$\text{RMSE}(|B_1^+|) = 0.0043 \text{ [a.u.]}, \quad \text{RMSE}(\omega) = 0.12 \text{ [Hz]}.$$

The reconstruction time is about 90 minutes. The median number of performed iterations as calculated over all parallel reconstruction processes is 13.
The standard deviations estimated by MR-STAT for T_1 and T_2 are shown,

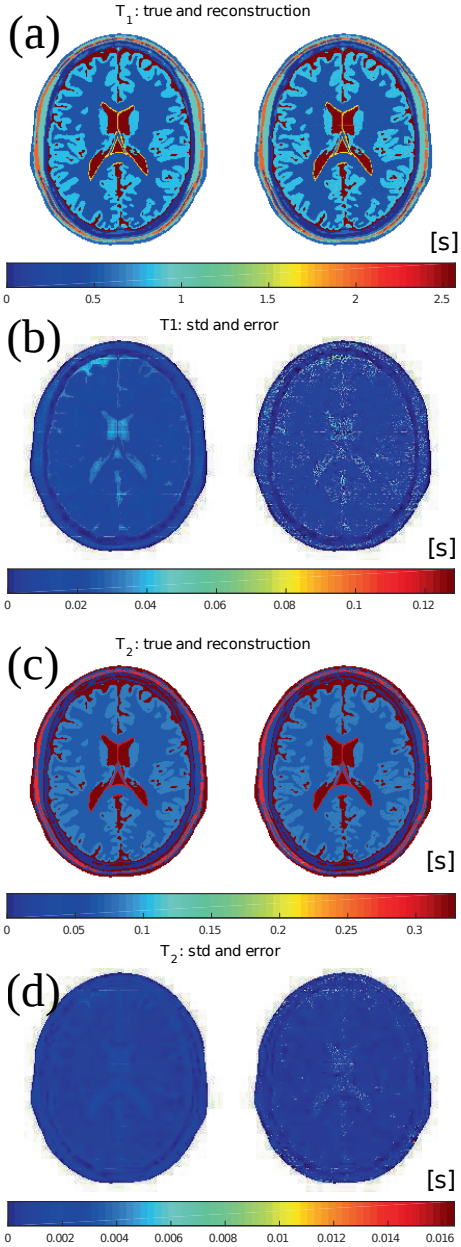

Figure 2.9: T_1 and T_2 maps for the synthetic MR-STAT acquisition and reconstruction. (a) and (c): true and reconstructed maps. (b) and (d): standard deviation maps estimated by MR-STAT and the error in the reconstructions ($|T_1^{\text{true}} - T_1^{\text{recon}}|$ and $|T_2^{\text{true}} - T_2^{\text{recon}}|$).

respectively, in Fig. 2.9(b) and Fig. 2.9(d). For comparison, the actual error maps, respectively defined as $|T_1 - T_1^{\text{recon}}|$ and $|T_2 - T_2^{\text{recon}}|$, are also reported and they show clear similarities.

2.4.3 *In-vivo* experimental demonstration at 3.0 Tesla

The obtained T_1 and T_2 maps are shown at the bottom of Fig. 2.4 and Fig. 2.5. The reconstruction algorithm was halted after 12 iterations since the solution did not significantly improve during the last few iterations. The computation time was about 12 minutes for both datasets.

2.5 Discussion

Traditional quantitative MR methods are typically performed in two steps; first a series of images is reconstructed, then the quantitative parameters are estimated from these images on a voxel-by-voxel basis. The recently introduced MRF method [110] works along similar lines, but shifts the focus away from the signal localization process and onto the temporal dynamics of the spin-system. Although MRF still adheres to the traditional two step procedure, it sacrifices accurate signal triangulation in favour of a high sampling rate. The resulting undersampling artifacts in each image are treated as a large, zero-mean, noise-like process, thus the signal model includes a substantial pseudo-stochastic component. MR-STAT relies instead on a fully deterministic strategy by employing a coupled space-time model that encapsulates the entire MR experiment. Consequently, the model accuracy is drastically enhanced and the brute-force exhaustive search is replaced by iterative minimization methods which exploit the structure of the underlying dynamics. The MR-STAT approach aims thus at a better utilization of the information carried by the data and to the elimination of the dictionary search, which is notoriously hindered by the curse of dimensionality. Another important benefit of taking this route is that it provides deep insights into the important aspect of error estimation. The availability of standard deviation maps is a valuable tool for quality monitoring; a fundamental aspect for the clinical application of quantitative MRI. It is important to realize that the gradient trajectory used in MR-STAT does not necessarily relate directly to the spatial resolution. The k-space in MR-STAT is not a spatial frequency domain, as is the case in standard MRI acquisition approaches. Although some demonstrations shown here still use a one dimensional Fourier transform along the read-out direction for parallelization, the MR-STAT formalism can, in principle, remove the explicit Fourier relationship between the time and image domain in its entirety. This will be beneficial in the case of non-cartesian trajectories such as radial and spiral or for non-linear gradient field systems [73]. As we move more and more along this direction, it may be better to think of trajectories in gradient space than in an actual

k-space. Inversion theory provides tools to generalize the concept of encoding capability for transient-states sequences when time and space dependence are implicitly entangled in the signal and results from Fourier theory are no longer applicable.
The primary cost of the MR-STAT approach is that all quantitative parameters must be estimated at once, which leads to a formidable inversion problem. We have however been able to reconstruct multi-parametric maps using a high performance computing facility within a reasonable computation time. The experimental design is more flexible since neither steady-states or static equilibrium conditions are needed nor the incoherence between undersampling artifacts and true signal; this allows for very short acquisitions (few seconds for a 2D slice) based upon experimentally reliable cartesian read-out schemes. In one of the experiments (see Fig. 2.5), we employed a step-wise flip-angle scheme combined with a standard bSSFP sequence, which is a widely available protocol on regular MR systems and does not require major adaptations on the acquisition. Also on the reconstruction side, flexibility is guaranteed by the inverse approach of MR-STAT; any changes made by the operator at the console during the exam can be easily accommodated in the reconstruction.
MR-STAT has been developed upon the philosophy that scanner time is much more expensive than computing time. We believe that this gap will keep growing in the future as computing power and algorithmic acceleration constantly increase. The current trends in bio-informatics and genomics show that local computing clusters or cloud computing on remote servers are becoming increasingly available in a hospital setting. The moderate investment in terms of the required computing infrastructure is highly profitable given the potential of MR-STAT for improving cost-effectiveness and patient comfort due to the reduced scan times and simpler workflows.
This study has focused on the computational and experimental proof-of-principle of MR-STAT. There is room left to study and optimize the accuracy, precision and speed of this framework. For instance, regularization techniques could be applied to reduce the noise amplification in the in-vivo measurements. Other techniques that could enhance MR-STAT are parallel imaging [146, 135, 61] and compressed sensing [109, 50]. The availability of multiple independent receivers and sparsity regularization terms can dramatically improve the triangulation of the signal origins thus greatly improving the conditioning of the comprehensive optimization problem. In general, optimum experiment design techniques [27, 184] could be applied to maximize the differentiation between signal evolutions and possibly enhance the rate of convergence while maintaining short acquisition times.
With this work, we intended to prove that quantitative MRI can be treated as a nonlinear tomographic problem and therefore large scale nonlinear optimization techniques can be successfully applied. We hope that that this manuscript will inspire researchers from other fields, to try and apply their experience and knowledge in the area of large scale inversion problems to the qMRI and

medical imaging in general.

2.6 Conclusion

A new framework for multi-parametric quantitative MRI, called MR-STAT, has been presented. Signal localization and parameter estimation are solved simultaneously by inverting a coupled space-time model from time domain data. This is obtained by established large scale nonlinear inversion techniques running on a high performance computing facility. The measurement efficiency is boosted by the elimination of dead times and traditional assumptions that inject artifacts into standard reconstruction approaches are circumvented. Moreover, this new formalism provides insights into the precision estimation of fast quantitative MRI.

Acknowledgments

Part of this work was funded by the Dutch Technology Foundation (NWO-STW), grant number 14125.
The authors are grateful to Dr. Tristan van Leeuwen, Prof. Jeannot Trampert and Dr. Ivan Vasconcelos for fruitful discussions and to Mrs Ying Lai Green for proofreading the manuscript.

3

High-resolution in vivo MR-STAT using a matrix-free and parallelized reconstruction algorithm

Oscar van der Heide
Alessandro Sbrizzi
Peter R. Luijten
Cornelis A. T. van den Berg

NMR in Biomedicine, 2020, 33.4: e4251.
https://doi.org/10.1002/nbm.4251

Abstract

MR-STAT is a recently proposed framework that allows the reconstruction of multiple quantitative parameter maps from a single short scan by performing spatial localisation and parameter estimation on the time domain data simultaneously, without relying on the FFT. To do this at high-resolution, specialized algorithms are required to solve the underlying large-scale non-linear optimisation problem. We propose a matrix-free and parallelized inexact Gauss-Newton based reconstruction algorithm for this purpose. The proposed algorithm is implemented on a high performance computing cluster and is demonstrated to be able to generate high-resolution ($1mm \times 1mm$ in-plane resolution) quantitative parameter maps in simulation, phantom and in-vivo brain experiments. Reconstructed T_1 and T_2 values for the gel phantoms are in agreement with results from gold standard measurements and for the in-vivo experiments the quantitative values show good agreement with literature values. In all experiments short pulse sequences with robust Cartesian sampling are used for which conventional MR Fingerprinting reconstructions are shown to fail.

3.1 Introduction

Conventional magnetic resonance imaging ("MRI") methods rely on the Fourier-Transform relationship between signal and local magnetization value for spatial encoding. Tissue differentiation is possible in the resulting qualitative images because different tissue types have distinct MR-related biophysical properties like T_1 and T_2 relaxation times. Quantitative MRI ("qMRI") methods aim to estimate MR-related biophysical properties like T_1 and T_2 relaxation times. Quantitative images could provide additional diagnostic value and are more suited for the purpose of multi-center studies and computer-aided diagnosis [159, 45]. The most straightforward and robust choices for T_1 and T_2 mapping sequences, i.e. single echo (inversion recovery) spin echo sequences have prohibitively long scan times. Over time, a multitude of alternative pulse sequences have been developed that reduce acquisition times to clinically acceptable levels [116, 108, 47, 46]. In recent years acquisition times have been reduced even further with advanced reconstruction techniques that include more information of the underlying physical processes in the reconstructions [13], add a-priori knowledge in the form of sparsity or low-rank constraints [189] and/or allow estimation of multiple parameter maps simultaneously [158, 144]. A prime example is MR Fingerprinting ("MRF", [110]). In MRF a transient-state pulse sequence with quasi-random components is used and many highly undersampled k-spaces are acquired in a single, short acquisition. The Fast Fourier Transform ("FFT") is applied on each k-space to generate many snapshot images. Then, on a voxel-per-voxel basis, the measured *fingerprints* are matched to a precomputed Bloch-equation based dictionary to obtain the quantitative parameters. Through this novel combination of transient-state sequences with a pattern recognition step, MRF has been able to drastically reduce qMRI acquisition times.

MR-STAT [141] is a recently proposed qMRI framework that, similarly to MRF, aims to estimate multiple parameter maps from a single short scan simultaneously. However, instead of performing FFTs in a separate step for spatial localisation of signal, parameter maps are fitted directly to the measured time-domain signal using a Bloch-equation based volumetric signal model. That is, a single large-scale non-linear optimisation problem is numerically solved in which the spatial localisation and parameter estimation are performed *simultaneously*. In addition, instead of using a dictionary matching procedure, in MR-STAT gradient-based iterative methods are used to solve the optimisation problem. Compared to MRF, the MR-STAT approach results in different trade-offs made in the reconstruction. Since the FFT is no longer explicitly used to transform back and forth between image space and frequency space, the spatial gradient encoding is entangled directly into the MR-STAT signal model. With this approach, data from different readouts of a transient state pulse sequence can be naturally combined into a single reconstruction process.

There is no reliance on dictionary compression [5] or compressed sensing [44] techniques to suppress aliasing artefacts. As will be demonstrated, MR-STAT allows for the reconstruction of high quality parameter maps from very short scans even when using standard and experimentally reliable Cartesian sampling strategies.

Solving the non-linear optimisation problem that results from using the volumetric signal model in MR-STAT does introduce new computational challenges. As will be discussed, the computational and memory requirements scale quadratically with the resolution and parallelizing the computations is non-trivial because the FFT is not used to spatially decouple the unknowns. In Sbrizzi et al. [141], to alleviate the computational challenges at high resolution, a 1D FFT along the readout direction was still employed to decouple the problem in one direction in space, resulting in many smaller and independent 1D subproblems to be solved. This hybrid approach only partly benefits from the above mentioned advantages of using a volumetric signal model, e.g., dynamical behaviour during readouts cannot be taken into account. Furthermore, it can only be used with Cartesian sampling strategies. Thirdly, if the technique is applied to 3D acquisitions, each of the resulting 2D subproblems will itself be a large-scale problem. Therefore, to unlock the full potential of MR-STAT, a specialized reconstruction algorithm is required that:

1. does not require storage of large model matrices (i.e. is *matrix-free*),

2. is suitable for a parallel computing implementation to reduce computation times,

3. is extensible to non-Cartesian sampling strategies.

In the current work we present a reconstruction algorithm for MR-STAT based on an inexact Gauss-Newton method (see [148] and Algorithm 7.2 in [120]) that satisfies the above requirements. For partial derivative computations that are needed in the reconstruction we propose to use exact algorithmic differentiation. With the new reconstruction algorithm we demonstrate the potential of the MR-STAT framework through simulation studies, phantom experiments and by reconstructing high-resolution in-vivo brain maps. Although in principle the reconstruction algorithm can be used with non-Cartesian sampling, in all experiments we will use Cartesian sampling patterns. The reason is that Cartesian sequences - which are used in the vast majority of clinical exams - are challenging to work with in the context of conventional MRF [150] whereas with MR-STAT parameter maps can be reconstructed successfully even when using very short acquisitions in the order of seconds per slice.

3.2 Theory

In this section we first review the MR-STAT framework as presented in Chapter 2. Then we discuss the computational challenges resulting from the large scale reconstruction problem and we propose techniques to deal with these challenges.

3.2.1 MR-STAT Framework

The time evolution of a single spin isochromat $\mathbf{m} = (m_x, m_y, m_z)$ with spatial coordinates $\mathbf{r} = (x, y, z)$ and tissue properties $\boldsymbol{\theta} = (T_1, T_2, \ldots)$ is governed by the Bloch equations.

Let $m = m_x + im_y$ be the transverse component of the magnetization in the rotating frame. The demodulated time-domain signal s is equal to the volume integral of the transverse magnetisation of all spins within the field of view V, weighted by their effective proton spin densities ρ. For the purpose of this work, ρ includes also the amplitude of the receive coil sensitivity and the transceive phase, thus ρ is a complex quantity, i.e. $\rho = \rho_x + i\rho_y$. In short:

$$s(t) = \int_V \rho(\boldsymbol{r}) m(\boldsymbol{r}, \boldsymbol{\theta}(\boldsymbol{r}), t) d\boldsymbol{r}. \tag{3.2.1}$$

After discretization of the field of view V into N_v voxels, each having volume Δ_V, equation 3.2.1 becomes

$$s(t) = \sum_{j=1}^{N_v} \rho_j m_j(\boldsymbol{\theta}_j, t) \Delta_V. \tag{3.2.2}$$

Here m_j is the magnetization in voxel j, which can be computed by numerical integration of the Bloch equations.

Let N_t be the total number of signal samples and let $t_1, \ldots, t_{N_t}$ denote the sampling times. Define the magnetization vector $\mathbf{m}_j$ in voxel j as

$$\mathbf{m}_j := (m_j(\boldsymbol{\theta}_j, t_1), \ldots, m_j(\boldsymbol{\theta}_j, t_{N_t})) \in \mathbb{C}_t^N \tag{3.2.3}$$

and the signal vector $\mathbf{s} \in \mathbb{C}^{N_t}$ as

$$\mathbf{s} = \sum_{j=1}^{N_v} \rho_j \mathbf{m}_j, \tag{3.2.4}$$

where Δ_V is assumed to be the same in all voxels and is absorbed into the proton density. Note that if we introduce the *magnetization matrix* $\mathbf{M} \in \mathbb{C}^{N_t \times N_v}$,

$$M_{i,j} := m_j(\boldsymbol{\theta}_j, t_i), \tag{3.2.5}$$

and proton density vector $\boldsymbol{\rho} \in \mathbb{C}^{N_v}$

$$\boldsymbol{\rho} = (\rho_1, \ldots, \rho_{N_v}), \tag{3.2.6}$$

then $\mathbf{s}$ can be written as

$$\mathbf{s} = \mathbf{M}\boldsymbol{\rho}. \tag{3.2.7}$$

Let N_p denote the number of distinct parameters per voxel (including real and imaginary parts of the proton density). Then s depends on $N := N_v \times N_p$ different parameters. All parameters are concatenated into a single vector $\boldsymbol{\alpha} \in \mathbb{R}^N$ in such a way that indices $\{j + kN_v \mid k = 0 \ldots, N_p - 1\}$ denote the parameters associated with voxel j.

Now, given a vector of measured time-domain samples $\mathbf{d} \in \mathbb{C}^{N_s}$, define the residual vector $\mathbf{r} \in \mathbb{C}^{N_t}$ as

$$\mathbf{r}(\boldsymbol{\alpha}) = \mathbf{d} - \mathbf{s}(\boldsymbol{\alpha}) \tag{3.2.8}$$

and define the non-linear least-squares objective function $f : \mathbb{R}^N \to \mathbb{R}$ as

$$f(\boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{r}(\boldsymbol{\alpha})\|_2^2 \tag{3.2.9}$$

The parameter maps $\boldsymbol{\alpha}^*$ are obtained by numerically solving

$$\boldsymbol{\alpha}^* = \operatorname{argmin}_{\boldsymbol{\alpha}} f(\boldsymbol{\alpha}), \tag{3.2.10}$$

subject to physical constraints represented by the Bloch equations and realistically attainable intervals for the parameters.

Computational Challenges

Note that (3.2.10) is a non-linear optimization problem that requires iterative algorithms to be solved. At each iteration, the signal $\mathbf{s} = \mathbf{M}\boldsymbol{\rho}$ needs to be computed and that requires the Bloch equations to be integrated for each voxel. In addition, the gradient of f (i.e. the vector of partial derivatives of f with respect to each of the parameters) needs to be computed. From the least-squares structure of the problem it follows that the gradient can be expressed as

$$\mathbf{g} = \nabla f = \operatorname{Re}\left\{\mathbf{J}^H\mathbf{r}\right\}, \tag{3.2.11}$$

where $\mathbf{J} \in \mathbb{C}^{N_t \times N_v}$ is the Jacobian matrix defined as

$$\mathbf{J}(\boldsymbol{\alpha}) := \left[\frac{\partial \mathbf{r}}{\partial \alpha_1} \cdots \frac{\partial \mathbf{r}}{\partial \alpha_N}\right], \tag{3.2.12}$$

$\mathbf{J}^H$ is the Hermitian transpose of $\mathbf{J}$ and $\operatorname{Re}\{\}$ is the real-part operator.

A gradient-descent type algorithm could be used to minimize (3.2.10) but it may result in poor convergence (see Chapter 3 of Nocedal and Wright [120]).

Table 3.1: On-disk sizes of MR-STAT matrices for $N_p = 4$ and $N_t = N = 4 \times N_v$ for an increasing number of voxels N_v. The memory sizes are computed as $2 \cdot N_v \cdot N_t \cdot 8$ bytes (**M**), $2 \cdot N \cdot N_t \cdot 8$ bytes (**J**) and $N^2 \cdot 8$ bytes (**H**) respectively. The factors of 2 come from the real and imaginary components and the factor of 8 represents the bytes necessary to store 64-bit floating point numbers.

Image size	**Voxels** (N_v)	**M**	**J**	**H**
64×64	$4,096$	1 GB	4 GB	2 GB
128×128	$16,384$	16 GB	64 GB	32 GB
256×256	$65,536$	256 GB	1,024 GB	512 GB
512×512	$262,144$	4,096 GB	16,384 GB	8,192 GB

Second-order methods (i.e. Newton methods) typically lead to better convergence. At each iteration, these methods require the inversion of a linear system involving (an approximation to) the Hessian matrix $\mathbf{H} \in \mathbb{R}^{N \times N}$, which includes curvature information and is defined as

$$\mathbf{H}(\boldsymbol{\alpha}) := \left[\frac{\partial^2 f}{\partial \alpha_i \partial \alpha_j} \right]_{i,j=1}^{N_p}. \tag{3.2.13}$$

A second-order MR-STAT reconstruction algorithm would follow the steps as outlined in Algorithm 3.1:

Algorithm 3.1 Minimize $f(\boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{d} - \mathbf{s}(\boldsymbol{\alpha})\|_2^2$

Require: Initial guess $\boldsymbol{\alpha}$
  while not converged **do**
    1. Compute residual: $\mathbf{r} = \mathbf{d} - \mathbf{s} = \mathbf{d} - \mathbf{M}\boldsymbol{\rho}$
    2. Compute gradient: $\mathbf{g} = \mathrm{Re}\left\{\mathbf{J}^H \mathbf{r}\right\}$
    3. Solve linear system: $\mathbf{H}\mathbf{p} = -\mathbf{g}$
    4. Update parameters: $\boldsymbol{\alpha} = \boldsymbol{\alpha} + \mathbf{p}$
  end while

Using Algorithm 3.1 for MR-STAT poses several practical challenges due to the large scale of the problem.

First of all, to estimate N parameters, the number of sample points N_t will in general be in the order of $N = N_p \times N_v$ as well. Assuming $N_t \approx N$, it follows that **M** will be of size $N_t \times N_v \approx (N_p N_v) \times N_v$ (complex entries) and **J** will be of size $2N_t \times N \approx (N_p N_v) \times (N_p N_v)$ (complex entries). Since **H** will be of size $(N_p N_v) \times (N_p N_v)$ as well, it follows that all three matrices scale with N_v^2. In Table 3.1, the required computer memory to store matrices of these sizes is reported for various values of N_v for the case $N_p = 4$. It can be seen that, even for 2D acquisitions, it will be infeasible to store these matrices in memory for clinically relevant resolutions.

Secondly, the actual time needed to compute the entries of $\mathbf{M}$, $\mathbf{J}$ and $\mathbf{H}$ scales with N_v^2 as well. When using a regular desktop computer the reconstruction times quickly become too long to make MR-STAT useful in clinical practice.

Fortunately, as will be detailed in the next section, Algorithm 3.1 only requires matrix-vector products with the matrices $\mathbf{M}, \mathbf{J}$ and (approximations to) $\mathbf{H}$. These matrix-vector product can be computed without having to store the full matrices in memory. Moreover, the computation of the matrix-vector products can be efficiently distributed among multiple computing cores on a high performance computing cluster, reducing the MR-STAT computation times to acceptable levels for off-line reconstructions.

3.2.2 Solution Strategies

3.2.2.1 Computing the time-domain signal s

In the first step of Algorithm 3.1 we need to compute $\mathbf{r} = \mathbf{d} - \mathbf{s}$ for the current estimate of the parameters $\boldsymbol{\alpha}$. Recall that

$$\mathbf{s} = \mathbf{M}\boldsymbol{\rho} = \sum_{j=1}^{N_v} \rho_j \mathbf{m}_j. \tag{3.2.14}$$

Since the time evolution of the magnetization in each voxel is assumed to be independent from other voxels, the $\mathbf{m}_j$ can be computed independently from each other. In particular, storage of the matrix $\mathbf{M}$ is not required for computing $\mathbf{s}$, see Algorithm 3.2.

Algorithm 3.2 Compute $\mathbf{s}(\boldsymbol{\alpha}) = \mathbf{M}\boldsymbol{\rho}$ (matrix-free, serial)

Initialize $\mathbf{s}$ = zeros(N_t,1)
for j $\leftarrow$ 1 to N_v **do**
1. Integrate Bloch equations in time to obtain $\mathbf{m}_j = [m_j(t_1), \ldots, m_j(t_{N_t})]^T$
2. Set $\mathbf{s} = \mathbf{s} + \rho_j \mathbf{m}_j$
end for
Return $\mathbf{s}$

Note that Algorithm 3.2 only requires the allocation of two vectors of length N_t, which is feasible on modern computing architectures for both 2D and 3D acquisitions. The computation of $\mathbf{s}$ can then be parallelized using N_c computing cores by following the procedure outlined in Algorithm 3.3 (see also [149, 102]).

The communication requirements for this parallelized algorithm can be summarized as follows:

- To distribute the parameters, the master process sends N/N_c parameters to each of the N_c slaves.

Algorithm 3.3 Compute $\mathbf{s}(\boldsymbol{\alpha}) = \mathbf{M}\boldsymbol{\rho}$ (matrix-free, parallel)

Require: Master process p_m, slave processes p_i for $i \in [1, \ldots, N_c]$.

1. p_m distributes $\boldsymbol{\alpha}$: each p_i receives the parameters associated with voxels $[(i-1) * N_v + 1, \ldots, i * N_v]$.
2. Each p_i uses Algorithm 3.2 to compute a "local" version of the signal $\tilde{\mathbf{s}}_i$.
3. Each p_i communicates $\tilde{\mathbf{s}}_i$ back to the p_m.
4. On p_m the signal $\mathbf{s}$ is computed as $\mathbf{s} = \sum_{i=1}^{N_c} \tilde{\mathbf{s}}_i$.

- To receive the local signals from the slaves, each slave sends a vector of length $N_t \approx N$ to the master process.

3.2.2.2 Computing the gradient g

To compute $\mathbf{g} = \nabla f$ for the current estimate of the parameters $\boldsymbol{\alpha}$, recall that

$$\mathbf{g} = \mathrm{Re}\left\{\mathbf{J}^H \mathbf{r}\right\}. \tag{3.2.15}$$

Since $\mathbf{J}$ is defined as

$$\mathbf{J} = \left[\frac{\partial \mathbf{r}}{\partial \alpha_1} \ldots \frac{\partial \mathbf{r}}{\partial \alpha_N}\right], \tag{3.2.16}$$

it follows that

$$\mathbf{g} = \begin{bmatrix} \mathrm{Re}\left\{\left\langle \frac{\partial \mathbf{r}}{\partial \alpha_1}, \mathbf{r}\right\rangle\right\} \\ \vdots \\ \mathrm{Re}\left\{\left\langle \frac{\partial \mathbf{r}}{\partial \alpha_N}, \mathbf{r}\right\rangle\right\} \end{bmatrix}. \tag{3.2.17}$$

To compute the $\frac{\partial \mathbf{r}}{\partial \alpha_i}$, again note that the magnetization in different voxels is assumed to evolve independently. Hence if α_i is a parameter associated with voxel j (i.e. $j = i \mod N_v$), it follows that

$$\frac{\partial \mathbf{r}}{\partial \alpha_i} = -\frac{\partial\left(\sum_{r=1}^{N_v} \rho_r \mathbf{m}_r\right)}{\partial \alpha_i} = -\frac{\partial\left(\rho_j \mathbf{m}_j\right)}{\partial \alpha_i}. \tag{3.2.18}$$

Using Algorithm 3.4 only requires storage of one vector of length N for the output and - in principle - one complex vector of length N_t to store the intermediate partial derivative vector. In practice we will compute the N_p partial derivatives for each voxel simultaneously so that N_p complex vectors of length N_t are stored simultaneously. For a high-resolution 2D scan this only requires limited memory (in the order of tens of megabytes).

Utilizing Algorithm 3.4, the computation of $\mathbf{g}$ can be performed in parallel as outlined in Algorithm 3.5. The parallelization schemes for both the signal and gradient computations are visualized in Figure 3.1.

Communication requirements for the parallel gradient computation can be summarized as follows:

Algorithm 3.4 Compute $\mathbf{g}(\boldsymbol{\alpha}) = \mathrm{Re}\left\{\mathbf{J}^H\mathbf{r}\right\}$ (matrix-free, serial)

Initialize $\mathbf{g}$ = zeros(N,1)
for j $\leftarrow$ 1 to N_v **do**
  for k $\leftarrow$ 1 to N_p **do**
    1. Set $i = j + (k-1)N_v$
    2. Compute partial derivative $\frac{\partial \mathbf{r}}{\partial \alpha_i} = -\frac{\partial(\rho_j \mathbf{m}_j)}{\partial \alpha_i}$
    3. Set $\mathbf{g}[i] = \mathrm{Re}\left\{\left\langle \frac{\partial \mathbf{r}}{\partial \alpha_i}, \mathbf{r} \right\rangle\right\}$
  end for
end for
Return $\mathbf{g}$

Algorithm 3.5 Compute $\mathbf{g}(\boldsymbol{\alpha}) = \mathrm{Re}\left\{\mathbf{J}^H\mathbf{r}\right\}$ (matrix-free, parallel)

Require: Master process p_M, slave processes p_i for $i \in [1, \ldots, N_c]$.
1. p_m distributes $\boldsymbol{\alpha}$: each p_i receives the parameters associated with voxels $[(i-1) * N_v + 1, \ldots, i * N_v]$.
2. p_m distributes $\mathbf{r}$ to each p_i.
3. Each p_i uses Algorithm 3.4 to compute a "local" gradient $\tilde{\mathbf{g}}_i$.
4. Each p_i communicates $\tilde{\mathbf{g}}_i$ back to the p_m.
5. On p_m the gradient $\mathbf{s}$ is computed by vertical concatenation of the $\tilde{\mathbf{g}}_i$.

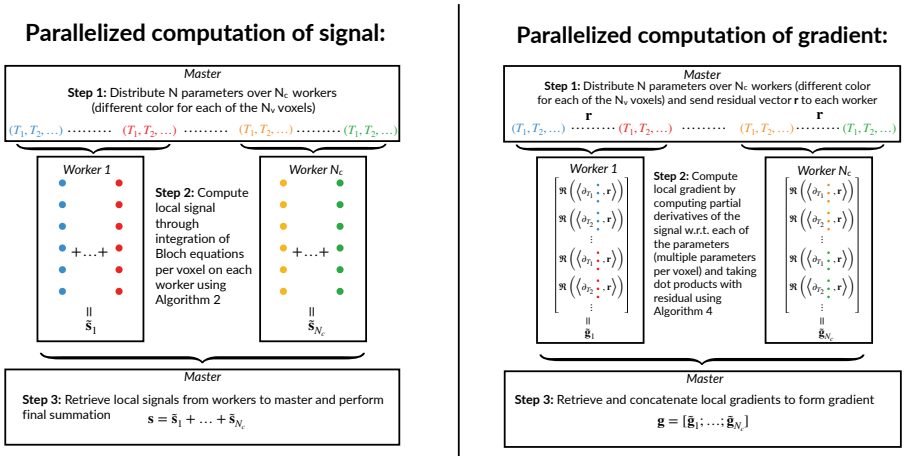

Figure 3.1: Visualization of the (matrix-free) algorithms to compute the signal [left] and the gradient [right] in a parallelized fashion.

- To distribute the parameters, the master process sends N/N_c parameters to each of the N_c slaves.
- To distribute the residual vector the master process sends a vector of length N_t to each slave.
- To receive the local gradients from the slaves, each slave sends a vector of length N/N_c to the master process.

Note that for both algorithms 3.3 and 3.5, the communication requirements scale linearly with the number of parameters N for a fixed number of cores N_c. Since $N = N_v \times N_p$, it follows that the communication requirements scale linearly with N_v as well. As discussed in the previous section the computational requirements scale quadratically with N_v. Therefore we hypothesize that, as long as $N_c \ll N_v$, the communication overhead is negligible compared to reduction in computation times achieved by dividing the computation load over N_c computing cores. That is, we expect the total computation time to decrease linearly with the number of cores available under this assumption. This hypothesis is confirmed in Figure 3.2 in the results section.

3.2.2.3 Incorporating Curvature Information

Given the ability to compute the gradient $\mathbf{g}$ using the matrix-free, parallelized algorithm from the previous subsection, in principle the so called limited-memory Broyden–Fletcher–Goldfarb–Shanno ("L-BFGS", [119]) method can be applied to obtain the update step $\mathbf{p}$ at each iteration. The L-BFGS method approximates the inverse of the Hessian matrix using a limited number of gradient vectors from previous iterations. However, in practice the L-BFGS method was observed to result in poor convergence.

Alternatively, since we are dealing with a least-squares problem, a Gauss-Newton method might be used in which the Hessian matrix $\mathbf{H}$ in Algorithm 3.1 is approximated by $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ and

$$\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}\mathbf{p} = -\mathbf{g} \tag{3.2.19}$$

is solved to obtain update steps $\mathbf{p}$. Note that the matrix $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ is of the same size as the Hessian matrix itself and thus, in principle, cannot be stored into computer memory. If, however, we use iterative techniques (e.g. a Conjugate Gradient method) to solve the linear system $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}\mathbf{p} = -\mathrm{Re}\left\{\mathbf{g}\right\}$, we only need matrix-vector products with $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$. In the previous subsection it was outlined how matrix-vector products of the form $\mathbf{J}^H\mathbf{v}$ may be computed in a matrix-free, parallelized fashion. Similar techniques can be applied to matrix-vector products of the form $\mathbf{Jv}$. Hence matrix-vector products of the form $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}\mathbf{v}$ can be computed in a matrix-free, parallelized fashion by first computing $\mathbf{y} = \mathbf{Jv}$ and subsequently computing $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{y}\right\}$. With this technique, the linear system in equation (3.2.19) can be solved numerically even

for large scale problems. In practice it will not be necessary to solve equation (3.2.19) to high precision and the number of iterations in this *inner loop* can be limited, resulting in an *inexact* Gauss-Newton method (see [148] and Algorithm 7.2 in [120]) as outlined in Algorithm 3.6

Algorithm 3.6 (Inexact Gauss-Newton MR-STAT) Minimize $f(\boldsymbol{\alpha}) = \frac{1}{2}\|\mathbf{d} - \mathbf{s}(\boldsymbol{\alpha})\|_2^2$

Require: Initial guess $\boldsymbol{\alpha_0}$, initial trust radius Δ_0

while not converged **do**
  Compute $\mathbf{s}$ (matrix-free, parallel)
  Set $\mathbf{r} = \mathbf{d} - \mathbf{s}$
  Compute $\mathbf{g} = \mathbf{J}^H\mathbf{r}$ (matrix-free, parallel)
  Solve with CG iterations (inner GN loop):
    $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}\mathbf{p} = -\mathbf{g}$ (matrix-free, parallel)
  Set $\boldsymbol{\alpha} = \boldsymbol{\alpha} + \mathbf{p}$
end while

3.3 Methods

The matrix-free, parallelized MR-STAT reconstruction algorithm was tested on both simulated and experimentally acquired data.

3.3.1 Pulse Sequence

In all test cases, a transient-state 2D balanced gradient-echo pulse sequence similar to the pulse sequence in Sbrizzi et al [141] was used. Throughout the whole sequence the TR was fixed and TE was set to TR/2. A linear, Cartesian sampling strategy was employed together with time-varying flip angles that change according to a smoothly varying pattern. We refer to Supplementary Material S1 for more details on the sampling trajectory and flip angle pattern. The phase of the RF pulse alternated between 0 and 180 degrees. Changing the flip angles prevents the spins from reaching a steady-state and by following a smoothly varying pattern the spin-echo behaviour of bSSFP sequences [142] is preserved to a large extent. This spin-echo like behaviour is needed for proper T_2 estimation and at the same time it also effectively eliminates sensitivity to ΔB_0 within a certain passband of off-resonances [6]. An added benefit of the smoothly changing flip angle train is the improved convexity of the minimization landscape [140].

Each RF pulse has a Gaussian envelope and at the start of the pulse sequence a non-selective inversion pulse is played out for enhanced T_1 encoding.

The pulse sequence was implemented on a 1.5 T clinical MR system (Ingenia, Philips Healthcare, Best, The Netherlands).

3.3.2 Reconstructions

All reconstruction code was written in the open-source Julia programming language [15]. To compute the MR-signal for a given set of parameters, an optimized Bloch-equation solver was implemented which also takes into account the slice profile [165]. To compute exact partial derivatives algorithmic differentiation in forward mode [179] was implemented. We refer to the Supplementary Material S2 for more details.

The inexact Gauss-Newton method was implemented using a trust-region framework (following [148] and Algorithm 7.2 in [120]). In order to facilitate bound constraints on the parameters, reflection at feasible boundaries was incorporated [38]. For the L-BFGS method, an implementation from the Optim.jl package [86] was used. The reconstruction algorithm was implemented on a high performance computing cluster which consists of multiple Intel Xeon Gold 6148 nodes with 40 cores each, on which the CentOS Linux 7 (Core) operating system is installed.

For all experiments, T_1, T_2 and ρ (complex) maps are reconstructed. For the data obtained with clinical MR systems we also reconstruct $|B_1^+|$ to take into account transmit field inhomogeneities. The off-resonance ΔB_0 was set to zero and thus it was not reconstructed because of the flat spectral response of the balanced sequence within the passband. The non-linear parameters were initialized as follows: $T_1 = 1000$ ms, $T_2 = 100$ ms, $|B_1^+| = 1$ a.u. and $\Delta B_0 = 0$ Hz. In previous work [141] the Variable Projection method ("VARPRO", [57]) was utilized to separate the linear parameters (i.e. proton density) from the non-linear parameters. The VARPRO method in principle requires computing (through SVD or QR decomposition) and storing an orthogonal basis for the matrix $\mathbf{M}$. For the matrix sizes in the current work that would be computationally infeasible and it is non-trivial to extend the VARPRO technique to matrix-free methods. Therefore, in the current work we treat the proton densities as non-linear parameters. We only make use of the linearity to provide an initial guess for the proton densities. That is, given the initial guess for the non-linear parameters, the (complex) proton density was initialized as the least squares solution to the linear system $\mathbf{M}(\boldsymbol{\alpha}_0)\boldsymbol{\rho} = \mathbf{d}$ obtained using a linear solver (LSQR). Based on the resulting initial guess for the proton density, a mask was drawn to exclude regions with no significant proton density from subsequent simulations.

In all reconstructions, logarithmic scaling is applied to both T_1 and T_2 parameters. The variable substitution brings both variables in a similar range and it thus improves convergence of the algorithm.

3.3.3 Numerical Brain Simulation

Signal from a numerical brain phantom [9] with a field-of-view of 192 mm × 192 mm and voxel size of 1 mm × 1 mm was simulated using the transient-state pulse sequence. A total number of $8 \times 192 = 1536$ readouts were simulated (each phase encoding line was acquired eight times but note that for each readout line the flip angle and thus state of the spins is different) with a TR of 7.88 ms and a TE of 3.94 ms. The total sequence duration was 12.1 s.

Reconstructions were performed using 64 cores. The number of outer and inner iterations for the inexact Gauss-Newton method were limited to 40 and 20, respectively.

For comparison purposes, we also performed MRF reconstructions on signal from the numerical brain phantom using the Cartesian trajectory, as well as signal from radial and spiral trajectories for which MRF is known to work well. In all three cases the same flip angle train, TE and TR were used. For the radial case, k_{max} was extended by a factor of $\sqrt{2}$ and each readout the spoke was rotated by the golden angle. For the spiral acquisition a variable density spiral was generated [98, 84] that would require 24 interleaves to fully sample the inner region of k-space and 48 interleaves for the outer region of k-space. The spiral was rotated by the golden angle each readout. Data $\mathbf{d}^{MRF}$ was then simulated by applying a forward operator, consisting of the (non-uniform) FFT [51] and an undersampling operator, on fingerprints simulated using the numerical brain phantom. To perform the MRF reconstructions, a dictionary consisting of 100 T_1 values from 0.1 s to 5.0 s in increments of 4 % and 100 T_2 values from 0.01 s to 2.0 s in increments of 5.5 % was generated. The T_1 and T_2 values of the phantom were not used in generating the dictionary. The dictionary was compressed in the time direction to rank 5 [114] using the SVD. For all trajectories (linear) low-rank forward operators $\boldsymbol{E}^{MRF}$ were formed that consisted of the expansion of low-rank coefficients to the full time-series, a nuFFT operator, and a sampling operator compression [5]. Low-rank snapshot images $\mathbf{x}^{MRF}$ were reconstructed from the undersampled data $\mathbf{d}^{MRF}$ by solving the linear system

$$\mathbf{x}^{MRF} = \text{argmin}_{\mathbf{x}} \|\mathbf{d}^{MRF} - \boldsymbol{E}^{MRF}\mathbf{x}\|_2^2 \tag{3.3.1}$$

with LSQR (similar to e.g. [191] and low-rank inversion in [5]). Finally, dictionary matching with the compressed dictionary was performed on $\mathbf{x}^{MRF}$ to obtain the parameter estimates.

To further study the effect of noise on the MR-STAT reconstruction algorithm, additional reconstructions were performed where complex Gaussian noise was added to the simulated signal such that $\|\text{signal}\|_2/\|\text{noise}\|_2 = 50, 25$ and 10.

3.3.4 Gel Phantom Experiment

Signal from a 2D transverse slice of six gadolinium-doped gel phantoms (TO5, Eurospin II test system, Scotland) was collected on the 1.5 T MR system using the manufacturer's thirteen-channel receive headcoil. In total $8 \times 96 = 768$ readout lines were acquired with a spatial resolution of 1 mm $\times$ 1 mm $\times$ 5 mm and a field-of-view of 96 mm $\times$ 96 mm. The TR and TE were 7.4 ms and 3.7 ms, respectively, resulting in a total acquisition time of 5.7 s. For reproducibility purposes the MR-STAT scan was repeated four times with full relaxation in between the different scans.

Parameters that describe the pulse sequence were exported from the scanner and subsequently loaded into Matlab [113]. The measured signals from different receive channels were compressed into a single signal by applying the principal component analysis and choosing the principle mode [31].

Reconstructions of the parameter maps were performed using the inexact Gauss-Newton method on the computing cluster using 32 cores. The number of inner iterations was limited to fifteen whereas the number of outer iterations was limited to ten.

To assess correctness of the T_1 and T_2 maps reconstructed with MR-STAT, data was also acquired using gold standard methods in the form of an inversion-recovery single spin-echo protocol with inversion times of [50, 100, 150, 350, 550, 850, 1250] ms for T_1 mapping and a single echo spin-echo protocol with echo times of [8, 28, 48, 88, 138, 188] ms for T_2 mapping.

3.3.5 In-vivo experiments

Using the 1.5 T clinical MR system we also acquired signal from 2D transverse slices of the brain in three healthy volunteers. Each volunteer gave written informed consent. For each acquisition, a total of $8 \times 192 = 1536$ readout lines were acquired with acquisition parameters as reported in Table 3.2. After masking, approximately 25,000 voxels remain for which quantitative parameters are estimated. The MR-STAT reconstructions were performed with 64 cores using the reconstruction settings as for the gel phantom experiment.

To demonstrate the effect of accelerated acquisitions, we also performed reconstructions using time-domain data corresponding to the first 896 TRs and the first 448 TRs from one of the subjects. The corresponding acquisition times were 6.8 s and 3.4 s respectively.

One of the in-vivo brain datasets was also used to test the effectiveness of the parallelization scheme. Individual matrix-vector products of the form $\mathbf{Mv}$ and $\mathbf{J}^H\mathbf{v}$ were computed and timed for $5, 10, 20, 40, 60, 80, 100, 120$ and 240 cores respectively.

Table 3.2: Acquisition parameters for in-vivo MR-STAT brain scans.

Acquisition parameter	**Subjects 1 and 2**	**Subject 3**
Field strength	1.5 T	1.5 T
In-plane resolution	1 mm × 1 mm	1 mm × 1 mm
Field-of-view	224 mm × 224 mm	224 mm × 224 mm
Slice thickness	5 mm	3 mm
TR	7.6 ms	7.9 ms
TE	3.8 ms	3.95 ms
Readout bandwidth	85.6 kHz	85.6 kHz
Pulse duration	0.76 ms	0.81 ms
Scan time	13.6 s	14.15 s

3.4 Results

3.4.1 Parallelization

In Figure 3.2 the time required to compute matrix-vector products of the form $\mathbf{Mv}$ and $\mathbf{J}^H\mathbf{v}$ for one of the in-vivo datasets is shown for an increasing number of computing cores N_c. Initially we observe a linear decrease in computation times, however this linear decrease flattens beyond approximately 64 cores. This effect can be explained by the increase in communication overhead when using more cores and increased competition between cores for shared resources like memory bandwidth and cache memory. Although the linear decrease flattens beyond 64 cores, a decrease in computation times is still observed even when going towards 240 cores. Because for MR-STAT reconstruction times are dominated by the computation of these matrix-vector products, the reconstruction times can thus be effectively reduced by the proposed parallel implementation.

3.4.2 Numerical Brain Phantom

The T_1, T_2 and proton density maps reconstructed using (Cartesian) MR-STAT and Cartesian, radial and spiral MRF are shown in Figure 3.3 as well as the corresponding absolute relative error maps. It can be seen that the parameter maps reconstructed with either MR-STAT or spiral MRF are in excellent agreement with the ground truth. The radial MRF reconstructions show stronger residual streaking artefacts but in general the estimated parameter values are close to the ground truth. For the Cartesian case the MRF reconstruction is unable to cope with the high undersampling (factor 192), resulting in severely biased parameter maps.

To quantify the quality of the reconstructions, normalized root mean square errors ("NRMSE"), high-frequency error norms ("HFEN", [136], with standard

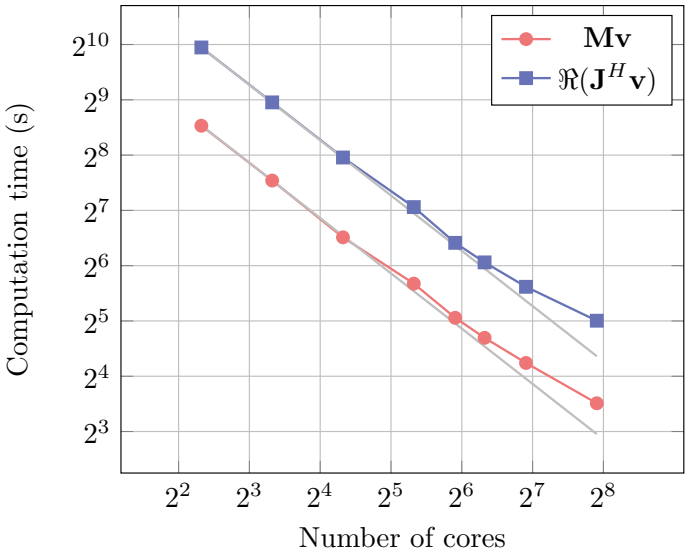

Figure 3.2: Time needed to compute matrix-vector products of the form $\mathbf{Mv}$ and $\mathbf{J}^H\mathbf{v}$ for different numbers of cores N_c used on a high performance computing cluster.

deviation of 1.5 pixels) and mean absolute relative errors ("MAPE") were computed and are reported in Table 3.3. It can be seen that the MR-STAT reconstruction results in the lowest NRSME and MAPE for all three parameters. The HFEN for the radial and spiral MRF and Cartesian MR-STAT reconstructions are similar.

In Figure 3.4 convergence curves for MR-STAT with the inexact Gauss-Newton method for different SNR levels (50, 25 and 10) are shown as well as mean absolute percentage errors per iteration for T_1, T_2 and proton density. For the higher SNR cases the error values stabilize and the method converges after relatively few, e.g. ten, outer iterations. It can be seen that for the lowest SNR case, overfitting occurs after around six iterations. Based on these observations the number of outer iterations for the in-vivo case was chosen to be ten.

3.4.3 Gel Phantoms

In Figure 3.5, reconstructed T_1 and T_2 maps for the gel phantoms are shown and the mean T_1 and T_2 values per tube are compared to the gold standard measurements. It can be seen that the mean values are in excellent agreement. The mean values reported for the different repetitions of the MR-STAT scans are also in good agreement with each other (i.e. within standard deviations). In general the standard deviations for the reconstructed T_2 values is higher

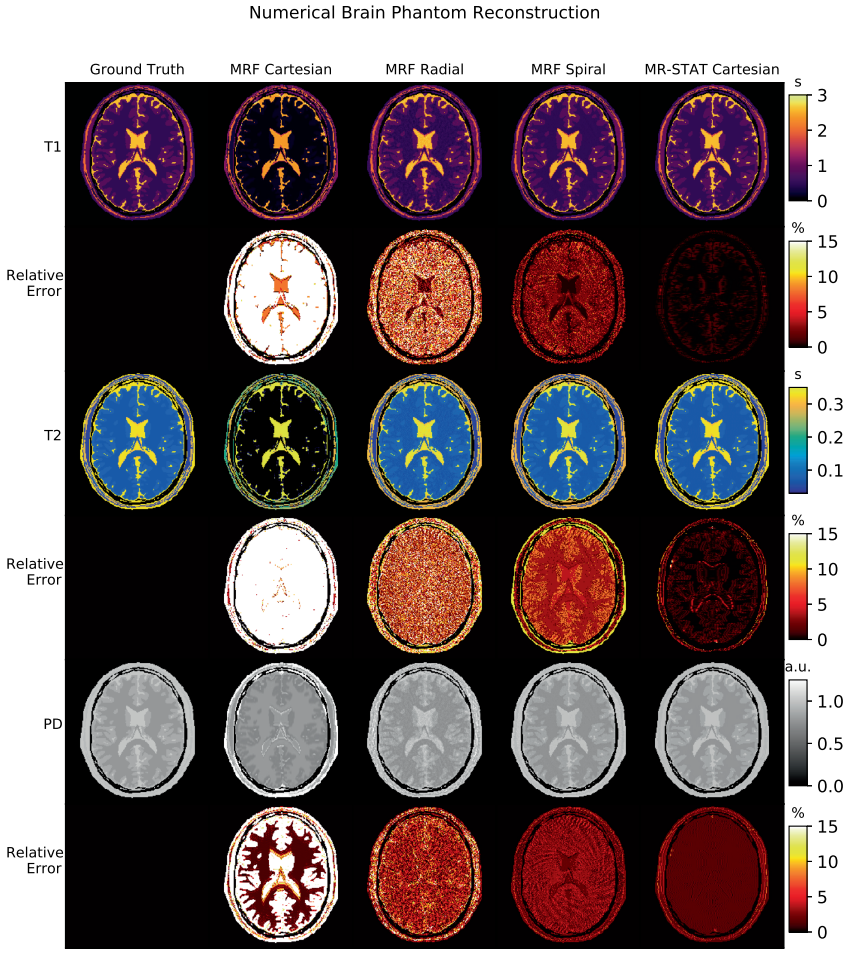

Figure 3.3: [First column] Ground truth T_1, T_2 and proton density maps for the numerical brain phantom. [Second, third and fourth columns] Reconstructed parameter maps and relative error maps for MRF with linear Cartesian, golden angle radial and golden angle spiral trajectories, respectively. [Fifth column] Reconstructed parameter maps and relative error maps for MR-STAT using a linear, Cartesian sampling trajectory. The MRF spiral and MR-STAT reconstructions both show excellent agreement with the ground truth values. The radial MRF reconstructions show residual aliasing artefacts and the Cartesian MRF reconstruction is heavily biased.

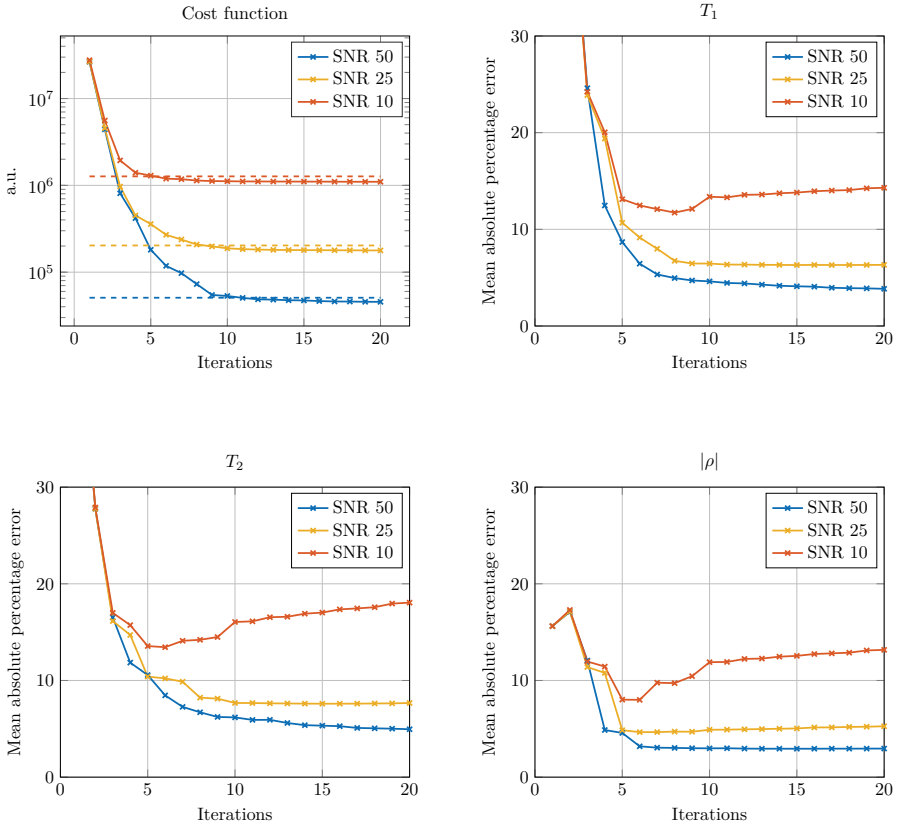

Figure 3.4: [Top Left] Convergence curves for the inexact Gauss-Newton MR-STAT method applied to data generated from the numerical brain phantom with different noise levels (SNR 50, 25, 10). In all cases the value of the cost function converges to the value expected based on the noise level. [Top right and bottom row] Mean absolute percentage errors for T_1, T_2 and proton density (magnitude) maps per iteration of the inexact Gauss-Newton method for different noise levels.

Table 3.3: Three different error metrics (NRMSE, HFEN, MAPE) computed for the MRF (Cartesian, Radial and Spiral) and MR-STAT (Cartesian) reconstructions on the numerical brain phantom. No noise was added to the data in these reconstructions. The MR-STAT reconstructions result in the lowest errors because the reconstructions do not suffer from undersampling artefacts and because there are no discretization errors due to a finite dictionary.

Parameter	Metric	Units	MRF Cartesian	MRF Radial	MRF Spiral	MR-STAT Cartesian
T_1	NRMSE	[a.u.]	0.2302	0.0432	0.0110	0.0025
	MAPE	[%]	65.6	8.5	2.5	0.4
	HFEN	[a.u.]	18.1	15.6	15.7	15.7
T_2	NRMSE	[a.u.]	0.2486	0.0756	0.0492	0.0048
	MAPE	[%]	56.1	8.1	4.9	0.9
	HFEN	[a.u.]	3.2	2.5	2.4	2.5
ρ	NRMSE	[a.u.]	1.0979	0.2626	0.1129	0.0830
	MAPE	[%]	15.8	4.8	2.7	1.8
	HFEN	[a.u.]	8.5	5.6	5.5	5.4

than for T_1 values, indicating a much stronger encoding of T_1 information into the signal which can be explained by the inversion pulse at the start of the sequence.

To reconstruct the parameter maps, only five iterations of the reconstruction algorithm were needed and the total reconstruction time was approximately nine minutes using 32 computing cores. In Figure 3.6 a logarithmic plot of the measured signal magnitude and the residual vector after the fifth iteration are displayed for one of the MR-STAT repetitions. Histograms of the measured noise and the residual vectors are also shown. It can be seen that the residual vector follows a zero-mean Gaussian distribution with standard deviation similar to the noise, indicating that the model used in MR-STAT is able to adequately describe the measured time-domain signal.

3.4.4 High-resolution 2D brain scan

In Figure 3.7, the reconstructed T_1, T_2 and proton density (magnitude) maps for the in-vivo brain scans performed on the three volunteers are shown. The maps show clear contrast between white matter, gray matter and cerebrospinal fluid ("CSF"). The maps corresponding to subject 3 appear noisier compared to the maps corresponding to subjects 1 and 2, which can be explained by the differences in slice thickness in the acquisition (3 mm vs 5 mm). Mean T_1 and T_2 values and standard deviations in regions of white- and gray matter are reported in Table 3.4. The mean values are generally in good agreement with values found in literature for 1.5 T experiments [41, 46, 28] although we do observe an underestimation compared to some other studies, especially in white matter. We expect the underestimation is related to magnetization transfer that is known to affect the signal of balanced gradient-echo sequences (in a way that depends on the used TR and RF pulse duration) [17, 127]. The

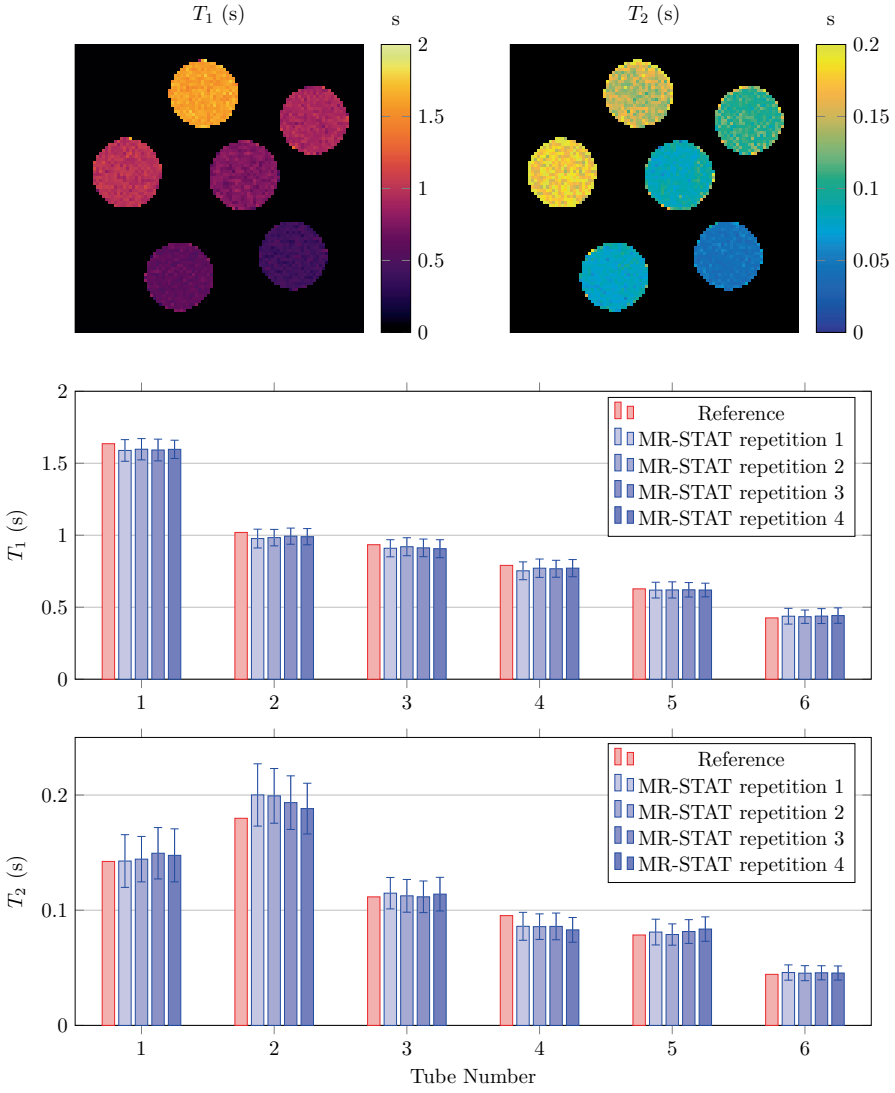

Figure 3.5: [Top row] T_1 and T_2 maps reconstructed with MR-STAT from gel-phantom data. [Middle and bottom rows] Comparison of mean T_1 and T_2 values obtained with MR-STAT and gold standard methods for each of the six gel phantom tubes. For MR-STAT the acquisition has been repeated four times.

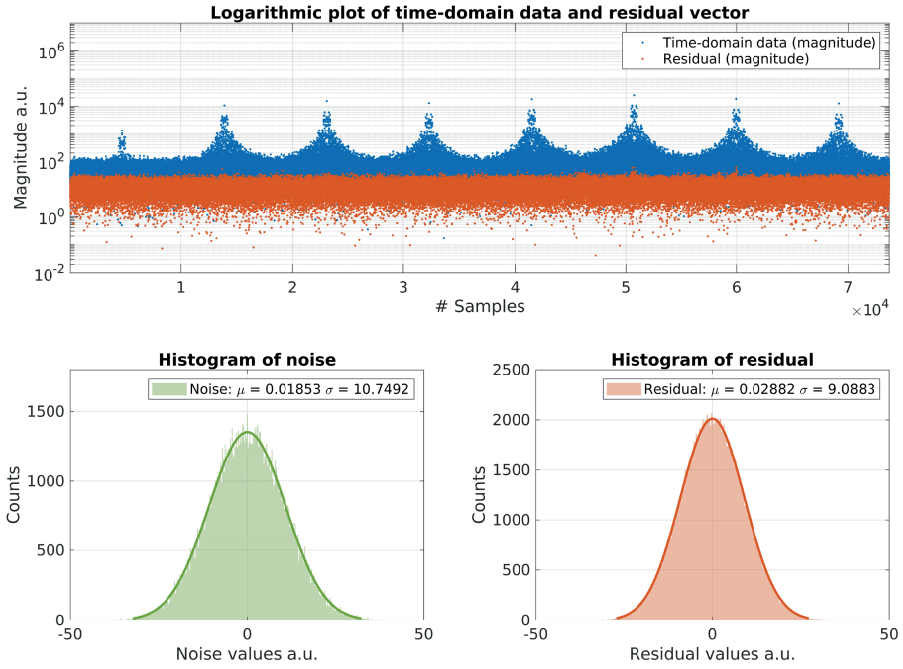

Figure 3.6: [Top row] Logarithmic plot of the magnitude of the measured time domain data obtained from the gel phantoms and the magnitude of the residual vector entries after the fifth iteration of the inexact Gauss-Newton method. [Bottom left] Histogram of noise values (real and imaginary values concatenated). The noise was measured using the receive channels right before the actual acquisition and it was subjected to the same pre-processing steps as the data used in the reconstruction (e.g. compression to a single channel using SVD). [Bottom right] Histogram of the residual vector entries (real and imaginary values concatenated) after the fifth iteration of the inexact Gauss-Newton method.

Table 3.4: Mean T_1 and T_2 values and standard deviation in white- and gray matter regions for each of the three in-vivo brain scans.

Tissue type	**Subject**	T_1	T_2
Frontal white matter	1	505 ± 48 ms	53.3 ± 4.0 ms
	2	542 ± 48 ms	57.4 ± 3.8 ms
	3	519 ± 54 ms	56.1 ± 4.3 ms
Putamen (gray matter)	1	874 ± 64 ms	74.8 ± 4.4 ms
	2	956 ± 66 ms	80.2 ± 4.5 ms
	3	895 ± 107 ms	78.4 ± 7.0 ms

reconstruction time for each slice was approximately five hours using 64 cores.

In Figure 3.8 we show T_1, T_2 and proton density (magnitude) for the same 2D brain slice but reconstructed using data corresponding to, respectively, 13.6 s, 7.8 s and 3.4 s acquisitions. It can be seen that the maps corresponding to the 6.8 s acquisition are comparable to the maps corresponding to the 13.6 s acquisition except that more noise is present. Depending on the application it might be more beneficial to repeat such a shorter sequence multiple times for noise averaging instead of scanning with the longer sequence. An added benefit of a shorter sequence duration is that the Bloch simulations are faster and thus reconstruction times are reduced by approximately the same factor with which the scantime is reduced. For the 3.4 s acquisition the MR-STAT problem (Eq. 3.2.10) is underdetermined in the sense that the number of datapoints is less than the number of unknowns in the problem. As can be seen in the reconstructed maps, this mostly results in biases in the CSF values. Note that in none of the reconstructions parallel imaging or compressed sensing techniques were utilized.

3.5 Discussion & Conclusion

MR-STAT is a framework for obtaining multiple quantitative parameter maps by fitting directly to measured time-domain data obtained from one short scan. Rather than relying on the FFT for spatial localisation of signal in a separate step, the spatial localisation and parameter estimation are performed simultaneously by iteratively solving a single non-linear optimization problem using a signal model that explicitly includes the spatial encoding gradients. The inherent large scale of the problem brings along new challenges in terms of computer memory requirements and computation times that make it difficult to perform MR-STAT reconstructions at high resolutions. To address these issues, we have presented a parallel and matrix-free reconstruction algorithm for MR-STAT and demonstrated that it can be used to generate high-resolution quantitative parameter maps.

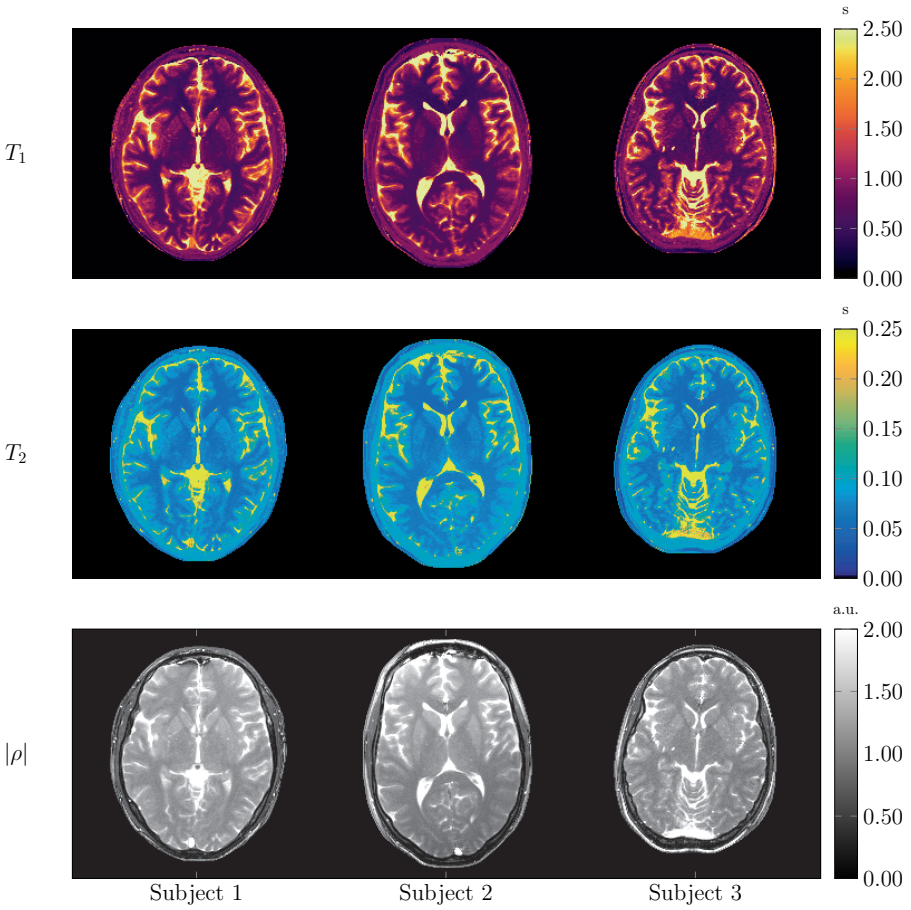

Figure 3.7: T_1, T_2 and proton density (magnitude) maps reconstructed with MR-STAT from in-vivo brain data obtained at 1.5 T (Philips, Ingenia) from multiple healthy volunteers. The in-plane resolution was 1×1 mm^2 for all three subjects. For subjects 1 and 2 the acquisition time was 13.6 s and the slice thickness was 5 mm. For subject 3 the acquisition time was 14.15 s and the slice thickness was 3 mm.

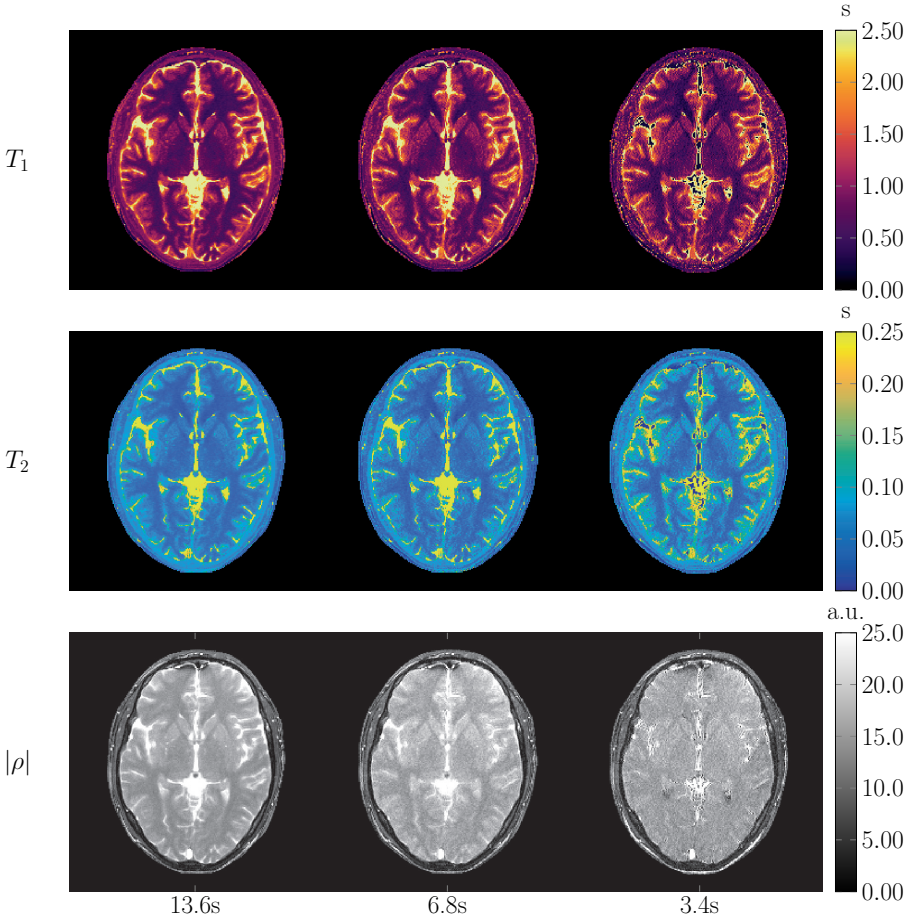

Figure 3.8: In-vivo T_1, T_2 and proton density (magnitude) maps at $1mm \times 1mm$ in-plane resolution reconstructed with MR-STAT based on acquisitions of, respectively, 13.6 s, 6.8 s and 3.4 s on a 1.5 T MR system (Philips, Ingenia).

All MR-STAT experiments in the current work have been performed with linear, Cartesian sampling strategies. This sampling strategy offers important advantages in the form of robustness to hardware imperfections (e.g. eddy currents, especially for gradient-balanced sequences [18, 30]), less susceptibility to ΔB_0 related blurring artefacts [14] and direct availability on clinical MR systems. Within the conventional MRF framework it is more challenging to work with Cartesian sampling strategies, as demonstrated using the simulation experiments. Studies that perform Cartesian MRF [93, 32] therefore typically acquire multiple readout lines per snapshot, resulting in much longer acquisition times compared to non-Cartesian MRF acquisitions. A formal explanation of why Cartesian acquisitions are less suitable for MRF is reported in Stolk et al. (2019) [150]. More advanced iterative MRF reconstructions [192, 191, 5, 44] might perform better with Cartesian sampling than the currently used MRF reconstructions (low-rank inversion followed by low-rank dictionary matching) and an in-depth comparison will be the subject of further studies. It should also be noted that neither the MR-STAT framework nor the currently proposed reconstruction algorithm are restricted to Cartesian sampling and further research is also aimed at incorporating non-Cartesian trajectories into MR-STAT.

An additional benefit of the volumetric signal model used in MR-STAT over FFT-based methods is that dynamic behaviour during the readouts (e.g. T_2-decay and ΔB_0 induced phase accumulation) is taken into account. This may especially be beneficial for improving reconstructions based on acquisitions with long readouts (e.g. spiral readouts).

MR-STAT reconstructions are performed by solving a non-linear optimization problem using gradient-based iterative methods. No pre-computed dictionary is used. Compared to dictionary-matching approaches there are no discretization errors and the reconstruction procedure is also flexible with respect to changes in sequence parameters (e.g. no rebuilding of a dictionary required when scan settings change). A downside of using iterative reconstruction algorithms to solve non-linear optimization problems is the risk of landing in a local minimum. In practice, with the currently used pulse sequence with smoothly changing flip angles and initial guess of the parameters, we have not encountered issues with local minima [140].

Whereas with MRF the addition of new parameters results in an exponential increase in dictionary size (and thus also an exponential increase in dictionary generation and matching time), with MR-STAT additional parameters can be added at a quadratic increase in computation time. The quadratic increase can be explained as follows. The total number of parameters to be reconstructed N increases linearly with the number of parameters per voxel ($N = N_p N_v$). Since the minimum number of time points N_t that needs to be acquired - and thus simulated - is in the order of N, the computation time per Bloch simulation increases linearly as well. In addition, the number of partial derivative computations that needs to be performed per voxel also increases linearly. That is, both the number of rows and columns of the Jacobian $\mathbf{J}$ increase linearly,

resulting in approximately a quadratic increase in computation time. In this respect we do note that, although currently ΔB_0 maps are not reconstructed (because the employed bSSFP sequence used in this work is designed not to be sensitive to ΔB_0 within the "passband"), it is part of all our Bloch simulations and partial derivative computations. In addition, for the MR-STAT experiments described in the manuscript we used pulse sequences such that $N_t \approx 2N$ so that the problem remains overdetermined when an additional parameter is reconstructed. Therefore, assuming a pulse sequence is used that has sufficient ΔB_0 encoding [144, 171], we do not expect to see an increase in computation times when reconstructing ΔB_0 as an additional parameter.

For the phantom experiment we observed that the noise level was reached for the residual. However, this was not observed for the in-vivo case as certain effects are still accounted for in the model. Examples include patient motion, blood flow, magnetization transfer and diffusion effects.

A limitation of the proposed method is that at this moment reconstruction times are still long for high-resolution scans, especially when compared to the dictionary matching procedures used in MRF. Even when employing a high performance computing cluster, reconstruction times are in the order of hours for a single 2D brain slice. Although possible from a memory point-of-view, 3D reconstructions will take too long for practical purposes with the current reconstruction setup. The main bottleneck in the reconstructions is formed by the partial derivative computations needed to solve equation (3.2.19). Further research is aimed at performing these computations on GPU architectures [183, 96], reducing the computational effort through algorithmic improvements [68] and through the use of surrogate models [186]. Together with (cloud) computing resources becoming cheaper and more accessible over time, we believe it is possible to accelerate the computations to an extent that MR-STAT becomes applicable in clinical settings.

Further research is also aimed at reduction of acquisiton time and improving precision and accuracy of the MR-STAT parameter maps by incorporating parallel imaging [65], compressed sensing and through sequence optimization.

The main aim of the MR-STAT project is to explore possibilities to achieve very short acquisition times beyond what is possible with FFT-based frameworks. Although the MR-STAT framework in principle allows for much flexibility in the data acquisition process (e.g. non-Cartesian acquisitions), in the current work we have opted for Cartesian sampling patterns because of their robustness to hardware imperfections and because they clearly exemplify the benefits of skipping the FFT step (i.e. no introduction of artificial aliasing noise through application of the FFT on undersampled k-spaces). An additional benefit is the direct availability of such sequences on clinical MR systems. In the current work we used constantly varying flip angle trains, however, as shown in Supplementary Material S3, MR-STAT could even be used with Cartesian bSSFP sequences with a fixed flip angle per k-space that require little to no pulse programming for their implementation.

Acknowledgements

This work was funded by the Dutch Technology Foundation, grant #14125.

3.6 Supplementary Materials

Supplementary Material S1: MR-STAT pulse sequence

For the simulation, gel phantom and in-vivo MR-STAT experiments discussed in the main text a pulse sequence was used with linear, Cartesian sampling and flip angles that change each TR in accordance with a smoothly varying pattern shown in Fig 3.9. The flip angle pattern starts at zero and then consists of connected sine-squared waves with randomly generated peak amplitudes. The sine-squared waves are aligned such that the central line of k-space is sampled at the peaks.

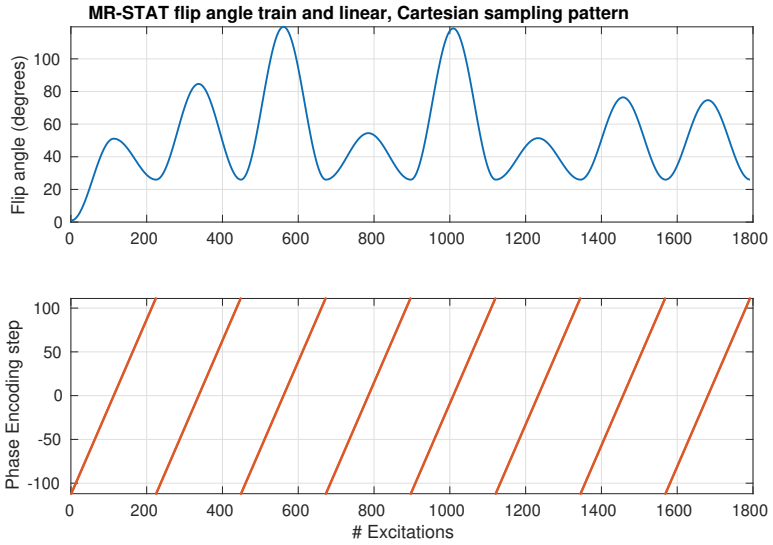

Figure 3.9: Overview of the flip angle train and the Cartesian gradient-encoding order used in the 2D (transient-state) MR-STAT pulse sequence.

Supplementary Material S2: Algorithmic Differentiation of Bloch equation solver

The Bloch equations define a system of (linear, inhomogeneous) ordinary differential equations with time-dependent coefficients (e.g. RF and gradient wave-

forms). To numerically integrate the Bloch equations, the time domain is discretized into time intervals during which the RF and gradient waveforms are assumed to be constant. Given the magnetization $\mathbf{m}_t = [m_{t,x}, m_{t,y}, m_{t,z}]^T$ in a voxel with coordinates $\mathbf{r} = [x, y, z]^T$ at time t, to find (an approximation to) the magnetization at time $t + \Delta t$ we first apply a rotation induced by the (complex) RF pulse and gradients $\mathbf{GR} = (GR_x, GR_y, GR_z)$ that are present during the time interval $[t, t + \Delta t]$ to obtain the rotated magnetization $\mathbf{m}_{rot}$. Afterwards we apply T_1 and T_2 induced decay and regrowth to $\mathbf{m}_{rot}$ to obtain $\mathbf{m}_{t+\Delta t}$. We refer to [165] for more details on the discretization.

To apply the rotation during each time interval, we first compute a rotation vector $\mathbf{a}$ as

$$\mathbf{a} = \gamma \Delta t \begin{bmatrix} -|B_1^+| \mathrm{Re}\{RF\} \\ |B_1^+| \mathrm{Im}\{RF\} \\ \mathbf{GR} \cdot \mathbf{r} + \Delta B_0 / \gamma \end{bmatrix}. \tag{3.6.1}$$

Now define $\mathbf{k} = \mathbf{a}/\|\mathbf{a}\|_2$ (i.e. $\mathbf{k}$ is a unit vector pointing in the same direction as $\mathbf{a}$) and $\theta = \|\mathbf{a}\|_2$. By using Rodrigues' rotation formula we can compute $\mathbf{m}_{rot}$ as:

$$\mathbf{m}_{rot} = \cos(\theta)\mathbf{m}_t + \sin(\theta)(\mathbf{k} \times \mathbf{m}_t) + (1 - \cos(\theta))(\mathbf{k} \cdot \mathbf{m}_t)\mathbf{k}. \tag{3.6.2}$$

To apply the decay and regrowth we first compute $E_1 := \exp(-\Delta t/T_1)$ and $E_2 := \exp(-\Delta t/T_2)$. Then, the magnetization at time $t + \Delta t$ is obtained as follows:

$$\mathbf{m}_{t+\Delta t} = \mathbf{m}_{rot} \otimes [E_2, E_2, E_1]^T + [0, 0, 1 - E_1]^T, \tag{3.6.3}$$

where $\otimes$ denotes the Hadamard product. The simulation is initialized with the magnetization in equilibrium position along the $z-$axis, i.e. $\mathbf{m}_0 = [0, 0, 1]^T$.

Given the update formula for the magnetization, we can apply algorithmic differentiation to obtain update formulas for the partial derivatives. We illustrate this procedure for partial derivatives with respect to T_1. First of all, note that from Eq. 3.6.2 we can derive that

$$\begin{aligned} \frac{\partial \mathbf{m}_{rot}}{\partial T_1} = &\frac{\partial}{\partial T_1}\left(\cos(\theta)\mathbf{m}_t\right) + \frac{\partial}{\partial T_1}\left(\sin(\theta)(\mathbf{k} \times \mathbf{m}_t)\right) \\ &+ \frac{\partial}{\partial T_1}\left((1 - \cos(\theta))(\mathbf{k} \cdot \mathbf{m}_t)\mathbf{k}\right). \end{aligned} \tag{3.6.4}$$

Since neither the rotation axis $\mathbf{k}$ nor the rotation angle θ depend on T_1, it

follows that

$$\begin{aligned}\frac{\partial \mathbf{m}_{rot}}{\partial T_1} &= \cos(\theta)\frac{\partial \mathbf{m}_t}{\partial T_1} + \sin(\theta)\frac{\partial(\mathbf{k}\times\mathbf{m}_t)}{\partial T_1} \\ &\quad +(1-\cos(\theta))\frac{\partial(\mathbf{k}\cdot\mathbf{m}_t)}{\partial T_1}\mathbf{k} \\ &= \cos(\theta)\frac{\partial \mathbf{m}_t}{\partial T_1} + \sin(\theta)\left(\mathbf{k}\times\frac{\partial \mathbf{m}_t}{\partial T_1}\right) \\ &\quad +(1-\cos(\theta))\left(\mathbf{k}\cdot\frac{\partial \mathbf{m}_t}{\partial T_1}\right)\mathbf{k}. \end{aligned} \tag{3.6.5}$$

Next, using Eq. 3.6.3, we find that

$$\begin{aligned}\frac{\partial \mathbf{m}_{t+1}}{\partial T_1} &= \frac{\partial \mathbf{m}_{rot}}{\partial T_1}\otimes[E_2,E_2,E_1]^T + \mathbf{m}_{rot}\otimes\frac{\partial[E_2,E_2,E_1]^T}{\partial T_1} \\ &\quad +\frac{\partial[0,0,1-E_1]^T}{\partial T_1} \\ &= \frac{\partial \mathbf{m}_{rot}}{\partial T_1}\otimes[E_2,E_2,E_1]^T + \mathbf{m}_{rot}\otimes[0,0,\frac{\Delta t}{T_1^2}E_1]^T \\ &\quad +[0,0,-\frac{\Delta t}{T_1^2}E_1]^T. \end{aligned} \tag{3.6.6}$$

By combining Eq. 3.6.5 and Eq. 3.6.6 we have a recursive update formula for the partial derivatives. Since $\mathbf{m}_0 = [0,0,1]^T$ is independent of T_1, the partial derivatives are initialized with $\frac{\partial \mathbf{m}_0}{\partial T_1} = [0,0,0]^T$. The procedure to find partial derivatives with respect to the other parameters is similar.

Although toolboxes exist that can automate the process of computing the derivatives, we have opted for a manual implementation to optimize for performance. Compared to the finite-difference method utilized in Sbrizzi et al. (2018) [141] algorithmic differentiation does not require the (non-trivial) choice of step sizes. And whereas the finite difference method approximates derivatives, with algorithmic differentiation the derivatives are exact. On top of that, our implementation of the algorithmic differentiation runs faster than the finite difference method by approximately a factor of two. This speedup can be partially explained by the fact that the relatively expensive $\sin, \cos$ (for rotations) and exp (for decay and regrowth) terms that need to be computed at each time step of the numerical integration can be efficiently reused for computing the partial derivatives.

Supplementary Material S3: MR-STAT with Cartesian sampling and constant flip angle per k-space

To demonstrate the flexibility of the MR-STAT framework we also acquired in-vivo brain data using a pulse sequence that has a constant flip angle per k-space (but different flip angle between the k-spaces). The data was acquired on

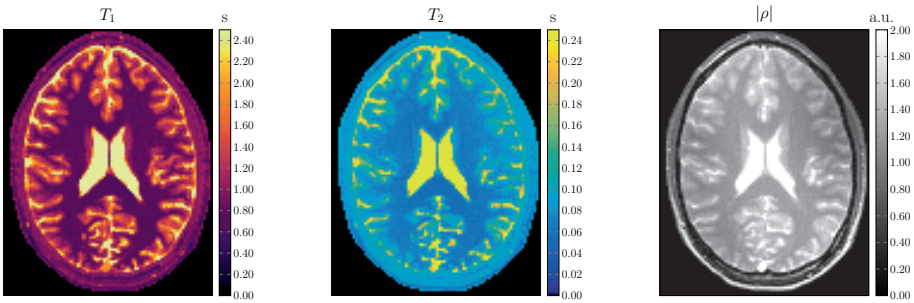

Figure 3.10: Reconstructed T_1, T_2 and proton density (magnitude) maps for a 2D slice of an in-vivo brain of a healthy volunteer at 1.8 mm $\times$ 1.8 mm in-plane resolution using a the pulse sequence with constant flip angle per k-space on a 3 T (Philips, Ingenia) MR system.

a 3 T clinical MR-system with the following acquisition parameters: in-plane resolution 1.8 mm $\times$ 1.8 mm, field-of-view 201.5 mm $\times$ 158 mm, TR 5.6 ms, TE 2.8 ms, slice thickness 3 mm, readout bandwidth 77.2 kHz, number of k-spaces 16, total scan time 7.82 s. The inexact Gauss-Newton method with ten outer iterations and fifteen inner iterations was used for the reconstruction. The total reconstruction time was approximately 30 minutes using 64 cores on the computing cluster. The reconstructed parameter maps and the flip angle train are shown in Fig. 3.10 and Fig. 3.11 respectively. The results demonstrate that excellent quality parameter maps can be obtained in a short time using a pulse sequence that is readily available even on older clinical MR-systems and requires little to no pulse programming.

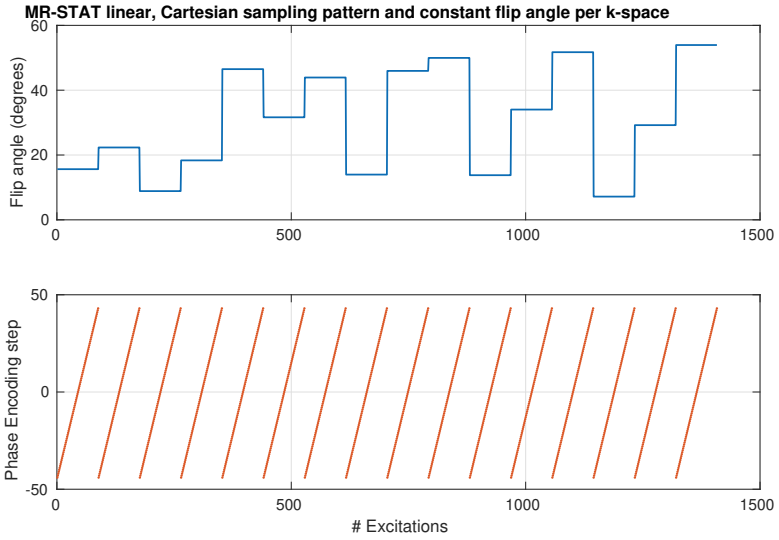

Figure 3.11: Visualization of the pulse sequence with linear, Cartesian sampling and constant flip angle per k-space.

4

Accelerated MR-STAT reconstructions using sparse Hessian approximations

Oscar van der Heide
Alessandro Sbrizzi
Cornelis A. T. van den Berg

IEEE Transactions on Medical Imaging, 2020, 39.11: 3737-3748.
https://doi.org/10.1109/TMI.2020.3003893

Abstract

MR-STAT is a quantitative magnetic resonance imaging framework for obtaining multi-parametric quantitative tissue parameter maps using data from single short scans. A large-scale optimization problem is solved in which spatial localization of signal and estimation of tissue parameters are performed simultaneously by directly fitting a Bloch-based volumetric signal model to measured time-domain data. In previous work, a highly parallelized, matrix-free Gauss-Newton reconstruction algorithm was presented that can solve the large-scale optimization problem for high-resolution scans. The main computational bottleneck in this matrix-free method is solving a linear system involving (an approximation to) the Hessian matrix at each iteration. In the current work, we analyze the structure of the Hessian matrix in relation to the dynamics of the spin system and derive conditions under which the (approximate) Hessian admits a sparse structure. In the case of Cartesian sampling patterns with smooth RF trains we demonstrate how exploiting this sparsity can reduce MR-STAT reconstruction times by approximately an order of magnitude.

4.1 Introduction

Magnetic Resonance Imaging ("MRI") is an imaging modality in which typically the Fast Fourier Transform ("FFT") is used to reconstruct qualitative contrast images from the raw k-space data obtained with an MR system. Tissue types can be differentiated in these images because each tissue is weighted differently under the influence of its MR-related biophysical properties like T_1 and T_2 relaxation times. Quantitative MRI ("qMRI") methods aim to provide estimates of these biophysical tissue properties. Compared to qualitative MR images, qMRI can provide additional diagnostic information, and by removing scanner and sequence variability they are more suited for multi-center studies and computer-aided diagnosis [159, 45].

Despite these advantages over qualitative MR imaging, qMRI has only found limited use in clinical practice so far, mainly because of increased scan times compared to qualitative imaging. In traditional qMRI, series of steady-state acquisitions are performed in which quantitative tissue parameters are encoded individually into multiple contrast images (*spatial localisation step*). Relatively simple analytical models are subsequently fitted on a voxel-per-voxel basis to yield estimates of the quantitative tissue parameters [116, 108, 46] (*parameter estimation step*). The total scan time can be lenghty as the acquisition for each contrast image has to satisfy the steady-state and Nyquist sampling conditions as set forward by the FFT.

Over time, advanced MR reconstruction techniques like parallel imaging [135, 61] and compressed sensing [109] have been developed that can relax the sampling conditions per contrast image and thus reduce acquisition times. For qMRI specifically, methods have been developed that also incorporate sparsity or low-rankness in the contrast dimension [50, 130, 79, 155] in the sense that the temporal spin dynamics are represented as a linear combination of relatively few basis vectors. This approach can improve the conditioning of the spatial localisation step, allowing for further reductions in acquisition times, but it does require careful tuning of the number of basis vectors used in the reconstruction. Using too few basis vectors can bias the estimated quantitative tissue parameters whereas using too many basis vectors will result in an underdetermined problem.

Exploring a different strategy, volumetric model-based reconstruction techniques have been proposed that combine the spatial localisation step and parameter estimation step into one [20]. Instead of generating contrast images first, tissue parameter estimates are obtained for all voxels simultaneously by directly fitting analytical signal models to the measured k-space (or time-domain) data. These signal models are non-linear and naturally provide sparse representations of the temporal spin dynamics. Moreover, the spatial encodings may be naturally distributed over the different contrasts [153], allowing for higher undersampling factors. Acquisition times can thus be further reduced com-

pared to the traditional approaches that separate the spatial localisation and parameter estimation steps. However, because of the non-linear nature of the signal models used, iterative reconstruction techniques need to be employed that require the computation of the gradient of the volumetric signal model with respect to the tissue parameters. This significantly increases reconstruction times. The non-linear volumetric inversion approach has been succesfully applied in the context of e.g. T1-mapping [175], T2-mapping [189] and diffusion tensor mapping [89].

Another important advancement in the field of qMRI has been the development of MR Fingerprinting ("MRF") [110]. In the original MRF paper it was proposed to use a transient-state pulse sequence with randomized flip angles, TR's and TE's to encode multiple quantitative tissue parameters into the signal simultaneously. Many highly-undersampled contrast images are reconstructed with the (non-uniform) FFT. Subsequently, on a voxel-per-voxel basis, the measured fingerprints are matched to a pre-computed dictionary of fingerprints to generate the tissue parameter estimates. Because analytical signal models are not available for the transient-state MRF sequences, numerical integration of the Bloch equations is required to form the dictionary entries. It was demonstrated that the dictionary matching procedure is resilient against the spatio-temporal incoherent undersampling artefacts that are present in the reconstructed contrast images. With MRF a method was thus introduced that can generate multiple quantitative tissue parameter maps based on a single, short acquisition.

In recent years more advanced MRF reconstruction methods have emerged that improve upon the initially proposed dictionary matching procedure. In Zhao et al. (2016) [192] the MRF parameter reconstruction problem was formulated as a maximum likelihood estimation problem. A novel reconstruction algorithm based on the alternating direction method of multipliers ("ADMM") was proposed that iterates between reconstructing the time frames and performing the dictionary matching to reduce the effects of undersampling artefacts. In Asslander et al. (2017) [6], it was further recognized that - due to the high undersampling factors applied - the contrast image reconstruction problem is in principle highly-underdetermined. By incorporating dictionary compression techniques (originally proposed in McGivney et al. (2014) [114]) into the ADMM method the conditioning of the reconstruction problem was improved and at the same time reconstruction times were reduced. In a similar fashion, methods have been proposed that instead rely on compressed sensing [44] or matrix completion techniques [49]. A common theme in the original and more advanced MRF methods is to separate the steps of the FFT-based spatial localisation of the measured signal and the subsequent parameter estimation like in the tradional qMRI methods.

MR Spin Tomography in Time-Domain ("MR-STAT") [141] is a recently proposed qMRI framework that combines the ideas of volumetric non-linear inversion on the time-domain data with using general transient-state sequences

that encode multiple tissue parameters into the signal simultaneously. Compared to MRF, the (typically) underdetermined problem of reconstructing the intermediate contrast images is avoided. Also, no precomputed dictionary is used thats grows exponentially in size with the addition of new parameters (like B_1^+, magnetization transfer or ΔB_0) and that can potentially bias the parameter estimates. Compared to the previously mentioned volumetric non-linear inversion methods a more general signal model is used in which the Bloch equations are integrated numerically, thus allowing for great flexibility in the acquisition process like in MRF. Moreover, in MR-STAT the FFT is no longer used for transforming between image space and k-space. Instead, the effects of the spatial encoding gradients are modelled explicitly. This allows for the inclusion of dynamic spin behaviour during readouts (e.g. T2 decay or B0 dephasing) like in time-resolved approaches [121, 154, 122]. For the acquisitions used in the current work short readout are used for which these effects are assumed to be have negligible impact.

The downside of using such a comprehensive, volumetric forward model in MR-STAT is that it gives rise to a computationally challenging large-scale non-linear optimization problem for which dedicated reconstruction algorithms are required. In previous work [69], a parallelized, matrix-free Gauss-Newton reconstruction algorithm was presented that can solve the large-scale optimization problem also for high-resolution scans where the number of unknowns is in the order of 10^5 (2D) to 10^6 (3D). It was demonstrated that with the MR-STAT approach high quality tissue parameter maps can be generated from short acquisitions even when using Cartesian readout strategies. However, even when doing reconstructions on a high performance computing cluster using many computing cores, reconstruction times for the matrix-free method are still in the order of hours for a single 2D slice. The main computational bottleneck in the previously proposed matrix-free method is solving a linear system involving the Gauss-Newton approximation of the Hessian matrix at each iteration to produce update steps in parameter space.

In the current work we study the structure of the Hessian in relation to the dynamics of the physical process (governed by the Bloch equations) and the sampling pattern used in the data acquisition. We are able to derive conditions under which the Hessian admits a sparse structure, which allows for

1. storage of the Hessian at each outer iteration of the reconstruction; and consequently

2. rapid computation of the update directions in parameter space using sparse matrix-vector multiplications embedded in the iterative solver (Conjugate Gradient iterations).

In case of Cartesian sampling patterns combined with smooth RF trains we demonstrate through in-silico, phantom and in-vivo experiments that MR-STAT reconstruction times can be reduced by an order of magnitude with the

sparse Hessian technique without compromising on quality of the reconstructed parameter maps.

4.2 Theory

In this section we first review the MR-STAT framework as presented in Sbrizzi et al. [141] and summarize how some of the computational challenges in the reconstrucion process were previously adressed. Then we derive conditions under which the Hessian matrix as well as the Gauss-Newton approximation to the Hessian matrix admit a sparse structure.

Scalar quantities (both real and complex) will be denoted in lower case, vector quantities in bold lower case letters and matrices in bold upper case letters.

4.2.1 MR-STAT Framework

The time evolution of a single spin isochromat $\mathbf{m} = (m_x, m_y, m_z)$ with spatial coordinates $\mathbf{r} = (x, y, z)$ is governed by the Bloch equations. This time evolution depends on the pulse sequence used (e.g. the RF excitation pulses and the spatial gradients acting on the spin isochromat) and also on its MR-relevant biophysical tissue properties $\boldsymbol{\theta} = (T_1, T_2, \ldots)$. Let the gradient trajectory associated with the pulse sequence be denoted by $\mathbf{k}(t)$ and let $m = m_x + im_y$ be the transverse magnetization component of a spin isochromat in the rotating frame. The demodulated time-domain signal s obtained with an MR system is modeled as the volume integral of the transverse magnetization of all excited spin isochromats within the field of view V of the receive coil. That is,

$$s(t) = \int_V m(\boldsymbol{\theta}(\boldsymbol{r}), t) e^{-2\pi i \mathbf{k}(t)\cdot\mathbf{r}} d\boldsymbol{r}. \tag{4.2.1}$$

In Eq. (4.2.1), the effects of spatial encoding gradients are included in the exponential term. Other space/time-dependent behavior of the magnetization, like rotations due to RF pulses, T_1-regrowth and T_2-decay, is included in the m term. Which specific effects are taken into account in the m term in practice will depend on the type of pulse sequence used in combination with the choice of which tissue parameters one aims to reconstruct. We also note that in Eq. (4.2.1) it is possible to include dynamic behaviour during readouts, e.g. T_2-decay and ΔB_0 dephasing, although in the current work we will focus on sequences with short readouts for which these effects are assumed to have negligble impact on the measureable signal. In a Nyquist-sampled steady-state sequence, and ignoring transient behaviour during readouts, the transverse magnetization m loses its time-dependency and the FFT can be used to recover qualitative images. In the more general case of a transient-state sequence (as proposed in [110, 140]) the FFT can in principle no longer be directly used to

transform between measured time-domain (or k-space) data and image space. Therefore, in the transient-state case, we proceed as follows with MR-STAT. First, we perform spatial discretization on eq. (4.2.1) to obtain

$$s(t) = \sum_{j=1}^{N_v} m(\boldsymbol{\theta}_j, t) e^{-2\pi i \mathbf{k}(t)\cdot \mathbf{r}_j} \Delta_V, \tag{4.2.2}$$

where m_j is the magnetization in voxel j and Δ_V is the volume element for each voxel. Now let N_t be the total number of samples acquired with the MR system's receiver and let $t_1, \ldots, t_{N_t}$ denote the sampling times. If we define the magnetization vector $\mathbf{m}_j$ in voxel j as

$$\mathbf{m}_j := [m(\boldsymbol{\theta}_j, t_1), \ldots, m(\boldsymbol{\theta}_j, t_{N_t})] \in \mathbb{C}^{N_t} \tag{4.2.3}$$

and the gradient encoding vector for voxel j as

$$\mathbf{GR}_j := \left[e^{-2\pi i \mathbf{k}(t_1)\cdot \mathbf{r}_j}, \ldots, e^{-2\pi i \mathbf{k}(t_{N_t})\cdot \mathbf{r}_j} \right] \in \mathbb{C}^{N_t} \tag{4.2.4}$$

then the discretized signal vector $\mathbf{s} \in \mathbb{C}^{N_t}$ can be defined as

$$\mathbf{s} = \sum_{j=1}^{N_v} \mathbf{m}_j \odot \mathbf{GR}_j. \tag{4.2.5}$$

Here $\odot$ denotes the Hadamard product. Let N_p denote the number of distinct parameters per voxel (including real and imaginary parts of the proton density). Then $\mathbf{s}$ depends on $N := N_v \times N_p$ different parameters. We concatenate all parameters into a single vector $\boldsymbol{\alpha} \in \mathbb{R}^N$ in such a way that parameters $\{\alpha_{j+kN_v} \mid k = 0 \ldots, N_p - 1\}$ are the parameters associated with voxel j.

Given a vector of measured time-domain samples $\mathbf{d} \in \mathbb{C}^{N_s}$, define the residual vector $\mathbf{r} \in \mathbb{C}^{N_t}$ as

$$\mathbf{r}(\boldsymbol{\alpha}) = \mathbf{d} - \mathbf{s}(\boldsymbol{\alpha}) \tag{4.2.6}$$

and define the least squares objective function $f : \mathbb{R}^N \to \mathbb{R}$ as

$$f(\boldsymbol{\alpha}) = \frac{1}{2} \|\mathbf{r}(\boldsymbol{\alpha})\|_2^2 \tag{4.2.7}$$

The parameter maps $\boldsymbol{\alpha}^*$ are obtained by numerically solving

$$\boldsymbol{\alpha}^* = \operatorname{argmin}_{\boldsymbol{\alpha}} f(\boldsymbol{\alpha}), \tag{4.2.8}$$

subject to physical constraints represented by the Bloch equations and realistically attainable intervals for the parameters. We point out that the only stochastic component in the model is the thermal noise, which is Gaussian distributed with zero mean. As a consequence, the present formulation is equivalent to a Maximum Likelihood estimation of $\boldsymbol{\alpha}^*$.

4.2.2 Inexact Gauss-Newton Method

Note that Eq. (4.2.8) is a non-linear optimization problem. Such problems are typically solved using iterative procedures such as Newton's method or variants thereof. Newton methods start with an initial guess $\boldsymbol{\alpha}$ and then obtain update steps $\mathbf{p}$ in parameter space at each *outer iteration* of the reconstruction algorithm by solving the linear system

$$\mathbf{H}(\boldsymbol{\alpha})\mathbf{p} = -\mathbf{g}(\boldsymbol{\alpha}). \tag{4.2.9}$$

Here $\mathbf{g} \in \mathbb{R}^N$ is the gradient of the objective function with respect to all the tissue parameters and $\mathbf{H} \in \mathbb{R}^{N\times N}$ is the Hessian matrix which is defined as

$$\mathbf{H}(\boldsymbol{\alpha}) := \left[\frac{\partial^2 f}{\partial\alpha_i \partial\alpha_j}\right]_{i,j=1}^{N}. \tag{4.2.10}$$

A difficulty with directly applying a Newton method in the context of MR-STAT is the inherent large scale of the problem. Even for 2D problems the number of parameters N can be in the order of 10^6. Explicitly forming the Hessian matrix or its inverse is therefore infeasible on present-day computer architectures. A commonly used technique in *least squares* problem is to approximate the Hessian matrix with the Gauss-Newton matrix $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ [120], where $\mathbf{J} \in \mathbb{C}^{N_t\times N}$ is the Jacobian matrix defined as

$$\mathbf{J}(\boldsymbol{\alpha}) := \left[\frac{\partial\mathbf{r}}{\partial\alpha_1} \cdots \frac{\partial\mathbf{r}}{\partial\alpha_N}\right], \tag{4.2.11}$$

$\mathbf{J}^H$ is the Hermitian transpose of $\mathbf{J}$ and $\mathrm{Re}\{\ldots\}$ is the real-part operator. Even though the matrices $\mathbf{J}$ and $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ are also typically too large to be stored in computer memory for MR-STAT problems, in previous work [69] it was shown how matrix-vector products of the form $\mathbf{Jv}$ and $\mathbf{J}^H\mathbf{v}$ can be computed without having to explicitly store $\mathbf{J}$ in memory. Given the ability to compute these matrix-vector products, the linear system in Eq. (4.2.9) (with $\mathbf{H}$ replaced by $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$) can then be solved in an iterative fashion using conjugate-gradient based techniques. Rather than solving Eq. (4.2.9) to arbitrary precision, the number of iterations performed in this *inner loop* is limited, resulting in a so called *inexact*, matrix-free Gauss-Newton method. A pseudo-algorithm for the matrix-free Gauss-Newton method is shown in Alg. 4.1.

As demonstrated in Chapter 3, the matrix-free Gauss-Newton MR-STAT method can be used to reconstruct high resolution parameter maps. However, a downside of the method is that the columns of $\mathbf{J}$ need to be re-computed for each matrix-vector multiplication in the inner loop (because the size of the problem prevents $\mathbf{J}$ from being stored in computer memory). In other words, at each iteration of the inner loop, partial derivatives of the signal with respect to each of the N parameters are computed. With approximately ten inner iterations needed per outer iteration [69], these computations form the computational bottleneck of the matrix-free MR-STAT reconstruction algorithm.

Algorithm 4.1 Matrix-free Gauss-Newton MR-STAT

Minimize $\frac{1}{2}\|\mathbf{d} - \mathbf{s}(\boldsymbol{\alpha})\|_2^2$
Require: Initial guess $\boldsymbol{\alpha}$
("Outer Loop")
while !converged **do**
1. Compute residual: $\mathbf{r} = \mathbf{d} - \mathbf{s}$
2. Compute gradient: $\mathbf{g} = \mathrm{Re}\left\{\mathbf{J}^H\mathbf{r}\right\}$
3. ("Inner loop") Iteratively solve linear system in matrix-free fashion: $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}\mathbf{p} = -\mathbf{g}$
4. Update parameters: $\boldsymbol{\alpha} = \boldsymbol{\alpha} + \mathbf{p}$

end while

4.2.3 Analysis of the sparsity pattern of the Hessian matrix in MR-STAT reconstructions

In this work we will derive conditions under which both the Hessian matrix $\mathbf{H}$ and the approximation $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ admit a sparse structure. The sparse structure allows for $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ to be computed and stored at each outer iteration of the reconstruction. The computational effort of subsequent multiplications with the sparse $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ is then negligible and the linear system in Eq. 4.2.9 can be solved rapidly by an iterative, conjugate gradient based algorithm to obtain $\mathbf{p}$. Since in practice approximately ten inner iterations are performed per outer iteration, we expect to see a decrease in MR-STAT reconstruction times of approximately an order of magnitude. A pseudo-algorithm for this sparse Gauss-Newton reconstruction algorithm is provided in Alg. 4.2.

Algorithm 4.2 Sparse Gauss-Newton MR-STAT

Minimize $\frac{1}{2}\|\mathbf{d} - \mathbf{s}(\boldsymbol{\alpha})\|_2^2$
Require: Initial guess $\boldsymbol{\alpha}$
("Outer loop")
while !converged **do**
1. Compute residual: $\mathbf{r} = \mathbf{d} - \mathbf{s}$
2. Compute gradient: $\mathbf{g} = \mathrm{Re}\left\{\mathbf{J}^H\mathbf{r}\right\}$
3. Compute and store sparse approximation: $\tilde{\mathbf{H}} \approx \mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$
4. ("Inner loop") Iteratively solve the linear system: $\tilde{\mathbf{H}}\mathbf{p} = -\mathbf{g}$
5. Update parameters: $\boldsymbol{\alpha} = \boldsymbol{\alpha} + \mathbf{p}$

end while

In the next subsections we first provide an intuitive argument for sparsity patterns of the Hessian followed by formal derivation afterwards.

4.2.3.1 Intuition for sparsity pattern

Intuitively, the sparsity of the Hessian matrix $\mathbf{H}$ may be understood as follows. Assume for simplicity that per voxel we are only estimating T_1 and T_2 values. Let $\mathbf{T_1}$ and $\mathbf{T_2}$ be the arrays obtained by concatenating the T_1 and T_2 values of each voxel respectively. Then $\mathbf{H}$ admits the following 2×2 block structure:

$$\mathbf{H} = \left(\begin{array}{c|c} \left(\frac{\partial^2 f}{\partial \boldsymbol{T_1}^2} \right) & \left(\frac{\partial^2 f}{\partial \boldsymbol{T_1} \partial \boldsymbol{T_2}} \right) \\ \hline \left(\frac{\partial^2 f}{\partial \boldsymbol{T_2} \partial \boldsymbol{T_1}} \right) & \left(\frac{\partial^2 f}{\partial \boldsymbol{T_2}^2} \right) \end{array} \right).$$

In the more general case of N_p parameters per voxel, the matrix $\mathbf{H}$ will consist of N_p^2 blocks (of which $N_p(N_p + 1)/2$ are unique due to symmetry of the matrix). The off-diagonal elements of each block provide estimates of covariances between parameters associated with different voxels.

In case of a conventional quantitative MR experiment with Nyquist-sampled, steady-state sequences, it is common to minimize an objective function for each voxel individually. This is possible because the spatial encoding gradients eliminate dependencies between voxels, i.e. no aliasing between signal from different voxels occurs. Spatial decoupling is performed through application of the FFT. On the other hand, if we applied an MR-STAT reconstruction on such Nyquist sampled data series (e.g. obtain the parameter maps for all voxels simultaneously by solving Eq. (4.2.8)), then the spatial decoupling would manifest itself in each block of the Hessian consisting of only a main diagonal.

In MR Fingerprinting [110] and MR-STAT, transient-state sequences are utilized that simultaneously encode multiple parameters into the signal. For 2D problems it is common though that throughout the pulse sequence each spatial frequency is sampled multiple times ($\approx$ 1000 spirals, radial spokes or Cartesian readout lines [110, 6, 140]). Over the acquisition as a whole it can therefore be said that such sequences are still "Nyquist-sampled". However, the transienst-state nature of the sequences prevents complete decoupling of parameters associated with different voxels. In an FFT framework, the transient states can be interpreted as time dependent (or spatial frequency dependent) perturbations of a steady-state signal. Thus, when applying the FFT, these perturbations result in convolutions with point spread functions in image space. Signals and thus parameters associated with different voxels become coupled despite the Nyquist sampling. Translated to the Hessian matrix as encountered in the MR-STAT framework, we can no longer expect the blocks to be diagonal matrices. However, if the transient-state transitions are sufficiently smooth the convolution kernels will be spatially confined and then we only expect dependencies between parameters associated with closely neighboring voxels. For the corresponding Hessian matrices we then expect each block to have a sparse, banded structure and by iteratively inverting it the point spread function is essentially refocused [52].

4.2.3.2 Formal derivation of sparsity pattern

For a formal derivation of the sparsity pattern, first recall that the magnetization in a voxel j can expressed as the component-wise product of $\mathbf{m}_j$ (which depends on the RF excitation pulses and the parameters in voxel j, e.g. $T_1, T_2, \ldots$, see eq. (4.2.3)) and the vector $\mathbf{GR}_j$ containing phase terms from spatial encoding gradients (see eq. (4.2.4)).

In order to simplify notation, assume each voxel to only have one associated parameter so that the same index can be used to denote a voxel j and its corresponding parameter α_j. Since $\mathbf{GR}_j$ is independent of the tissue parameters, it follows that

$$\frac{\partial \mathbf{s}}{\partial \alpha_j} = \frac{\partial \left(\mathbf{m}_j \odot \mathbf{GR}_j \right)}{\partial \alpha_j} = \frac{\partial \mathbf{m}_j}{\partial \alpha_j} \odot \mathbf{GR}_j. \tag{4.2.12}$$

From the least-squares structure of problem (4.2.8) it follows that

$$\begin{aligned} \frac{\partial f}{\partial \alpha_j} &= -\mathrm{Re}\left\{ \left\langle \frac{\partial \mathbf{s}}{\partial \alpha_j}, \mathbf{r} \right\rangle \right\} \\ &= -\mathrm{Re}\left\{ \left\langle \frac{\partial \mathbf{m}_j}{\partial \alpha_j} \odot \mathbf{GR}_j, \mathbf{r} \right\rangle \right\}. \end{aligned} \tag{4.2.13}$$

The Hessian entries can then be computed as

$$\begin{aligned} \mathbf{H}_{i,j} &= \frac{\partial^2 f}{\partial \alpha_i \partial \alpha_j} \\ &= -\mathrm{Re}\left\{ \left\langle \frac{\partial^2 \mathbf{m}_j}{\partial \alpha_i \partial \alpha_j} \odot \mathbf{GR}_j, \mathbf{r} \right\rangle \right\} \\ &\quad +\mathrm{Re}\left\{ \left\langle \frac{\partial \mathbf{m}_i}{\partial \alpha_i} \odot \mathbf{GR}_i, \frac{\partial \mathbf{m}_j}{\partial \alpha_j} \odot \mathbf{GR}_j \right\rangle \right\}. \end{aligned} \tag{4.2.14}$$

However, because the transverse magnetization in voxel i is independent of the parameters associated with voxel j, the term $\frac{\partial^2 \mathbf{m}_j}{\partial \alpha_i \partial \alpha_j}$ is zero if $i \neq j$. Therefore,

$$\mathbf{H}_{i,j} = \mathrm{Re}\left\{ \left\langle \frac{\partial \mathbf{m}_i}{\partial \alpha_i} \odot \mathbf{GR}_i, \frac{\partial \mathbf{m}_j}{\partial \alpha_j} \odot \mathbf{GR}_j \right\rangle \right\}, \quad i \neq j. \tag{4.2.15}$$

On the other hand,

$$(\mathbf{J}^H \mathbf{J})_{i,j} = \left\langle \frac{\partial \mathbf{m}_i}{\partial \alpha_i} \odot \mathbf{GR}_i, \frac{\partial \mathbf{m}_j}{\partial \alpha_j} \odot \mathbf{GR}_j \right\rangle. \tag{4.2.16}$$

It can thus be observed that the Gauss-Newton approximation $\mathrm{Re}\left\{\mathbf{J}^H \mathbf{J}\right\}$ to the Hessian matrix is *exact* for the off-diagonal terms of each block. As a result both matrices will have the same sparsity pattern and in what follows we will only focus on the matrix $\mathrm{Re}\left\{\mathbf{J}^H \mathbf{J}\right\}$ instead of $\mathbf{H}$.

Suppose that we want to reconstruct N_p 2D parameter maps, each consisting of $N_v = N_x \times N_y$ voxels. A single full Nyquist sampled k-space would consist of N_v spatial frequencies in that scenario. Now we make the requirement that a sampling strategy is used where each spatial frequency is sampled exactly $N_k \geq 1$ times (Requirement 1). We anctipate here that in practice this will not be a restrictive requirement.

Assuming that Requirement 1 is satisfied, we sort and reshape $\mathbf{m}_j$ and $\mathbf{GR}_j$ into arrays of size (N_x, N_y, N_k) which are consistent with the k-space sampling structure. In other words, data is reorganized as standard spatial frequency arrays for each k-space acquisition.[1] Using Matlab notation, define $\mathbf{m}_j^{(k)} := \mathbf{m}_j(:,:,k)$ and $\mathbf{GR}_j^{(k)} := \mathbf{GR}_j(:,:,k)$. Then Eq. (4.2.16) can be rewritten as

$$\left(\mathbf{J}^H\mathbf{J}\right)_{i,j} = \sum_{k=1}^{N_k} \left\langle \frac{\partial \mathbf{m}_i^{(k)}}{\partial \alpha_i} \odot \mathbf{GR}_i^{(k)}, \frac{\partial \mathbf{m}_j^{(k)}}{\partial \alpha_j} \odot \mathbf{GR}_j^{(k)} \right\rangle_F , \tag{4.2.17}$$

where $\langle .,. \rangle_F$ is the Frobenius inner product.

Using Parseval's Theorem and the Convolution Theorem, with $\mathcal{F}$ being the two-dimensional DFT, the time-domain quantities in the above expression can be transformed into spatial domain quantities as follows:

$$\left(\mathbf{J}^H\mathbf{J}\right)_{i,j} = \tag{4.2.18}$$

$$\sum_k \left\langle \mathcal{F}\left(\frac{\partial \mathbf{m}_i^{(k)}}{\partial \alpha_i} \odot \mathbf{GR}_i^{(k)} \right), \mathcal{F}\left(\frac{\partial \mathbf{m}_j^{(k)}}{\partial \alpha_j} \odot \mathbf{GR}_j^{(k)} \right) \right\rangle_F = \tag{4.2.19}$$

$$\sum_k \left\langle \mathcal{F}\left(\frac{\partial \mathbf{m}_i^{(k)}}{\partial \alpha_i} \right) \otimes \mathcal{F}\left(\mathbf{GR}_i^{(k)} \right), \mathcal{F}\left(\frac{\partial \mathbf{m}_j^{(k)}}{\partial \alpha_j} \right) \otimes \mathcal{F}\left(\mathbf{GR}_j^{(k)} \right) \right\rangle_F . \tag{4.2.20}$$

Now, because of Requirement 1 and the sorting of $\mathbf{GR}_j$, it follows that $\mathcal{F}\left(\mathbf{GR}_j^{(k)}\right) = \delta_j$, where δ_j is a delta-peak at the position of voxel j. The above equation can thus be rewritten as

$$\left(\mathbf{J}^H\mathbf{J}\right)_{i,j} = \sum_k \left\langle \mathcal{F}\left(\frac{\partial \mathbf{m}_i^{(k)}}{\partial \alpha_i} \right) \otimes \delta_i, \mathcal{F}\left(\frac{\partial \mathbf{m}_j^{(k)}}{\partial \alpha_j} \right) \otimes \delta_j \right\rangle_F . \tag{4.2.21}$$

Next, by using Leibniz' integral rule it follows that

$$\mathcal{F}\left(\frac{\partial \mathbf{m}_i^{(k)}}{\partial \alpha_i} \right) = \frac{\partial}{\partial \alpha_i} \mathcal{F}\left(\mathbf{m}_i^{(k)} \right) \tag{4.2.22}$$

[1] For simplicity we ignore possible readout oversampling. Also, for non-Cartesian sequences, we assume the data is still acquired on (or interpolated to) Cartesian coordinates.

and therefore

$$\left(\mathbf{J}^H\mathbf{J}\right)_{i,j} = \sum_k \left\langle \frac{\partial}{\partial \alpha_i}\mathcal{F}\left(\mathbf{m}_i^{(k)}\right) \otimes \delta_i, \frac{\partial}{\partial \alpha_j}\mathcal{F}\left(\mathbf{m}_j^{(k)}\right) \otimes \delta_j \right\rangle_F . \tag{4.2.23}$$

If the pulse sequence used in the acquisition is designed in such a way that each of the *magnetization matrices* $\mathbf{m}_j^{(k)}$ is smooth in both spatial directions (Requirement 2, the related sequence design considerations will be discussed in Section 4.2.4), then the point spread functions $\mathcal{F}(\mathbf{m}_j^{(k)})$ will have (to good approximation) a compact support. Each support $\operatorname{supp}(\mathbf{m}_j^{(k)})$ in principle depends on the tissue parameters $\boldsymbol{\theta}_j$. Let Θ be the subset of $\mathbb{R}^{N_p}$ of physically viable tissue parameters and define the *maximum support* χ as

$$\chi := \bigcup_{k=1}^{N_k} \bigcup_{\boldsymbol{\theta}\in\Theta} \operatorname{supp}\left(\mathcal{F}\left(\mathbf{m}^{(k)}(\boldsymbol{\theta})\right)\right). \tag{4.2.24}$$

Outside of χ the point spread functions will vanish for all potential tissue parameters and therefore the partial derivative terms $\frac{\partial}{\partial \alpha_i}\mathcal{F}\left(\mathbf{m}_i^{(k)}\right)$ will vanish outside of χ as well. We can then conclude that the right-hand term of (4.2.23) will be zero for voxels i and j that are separated in space by a distance more than $\operatorname{diam}(\chi)$, where

$$\operatorname{diam}(A) := \sup\{\|x - y\| \mid x, y \in A\}. \tag{4.2.25}$$

A sparsity pattern thus emerges for the matrix $\operatorname{Re}\{(\}\mathbf{J}^H\mathbf{J}$). In the more general case of N_p parameters per voxel the above arguments can be repeated for each of the N_p^2 blocks separately.

To conclude the formal derivation, if a pulse sequence is used where

1. each of the N_v spatial frequencies is sampled exactly N_k times, and
2. each of the magnetization matrices $\mathbf{m}_j^{(k)}$ is smooth,

then each of the N_p^2 blocks of the matrix $\operatorname{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ admits a sparse structure with off-diagonal elements corresponding to closely neighbouring voxels only.

4.2.4 Pulse sequence design considerations

In this section we discuss the design of pulse sequences that result in smooth $\mathbf{m}_j^{(k)}$ and thus sparse $\operatorname{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$.

4.2.4.1 Cartesian sampling

For Cartesian sampling patterns with short readouts, the spin dynamics during the readouts (e.g. T_2 decay and ΔB_0 induced dephasing) are assumed to

have negligble impact on the measurable magnetization. As a result, the magnetization matrices $\mathbf{m}_j^{(k)}$ will be (practically) constant, and thus smooth, in the readout direction and the matrix $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ is already expected to have a sparsity pattern. More precisely, if we assume the parameters in the MR-STAT optimization problem are ordered by stepping into the phase-encoding direction first, the sparsity pattern will be such that each block of $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ will - at most - be a block diagonal matrix with blocks of size $N_y \times N_y$. See Supplementary Fig. 4.9 for an example. In this "worst case scenario" for Cartesian sampling, where we only assume smoothness of $\mathbf{m}_j^{(k)}$ in the readout direction, storage requirements and computational requirements for $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ are already reduced by a factor of N_x.

For increased sparsity of $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$, smoothness of $\mathbf{m}_j^{(k)}$ in the phase encoding direction must be achieved. To this end, we note that if a sequence with smoothly varying flip angles is used, then the magnetization vector $\mathbf{m}_j$ will also be smooth [140]. If, in addition, a linear Cartesian sampling trajectory is chosen, then after reordering and reshaping $\mathbf{m}_j$, each of the $\mathbf{m}_j^{(k)}$ will be smooth as well. In this scenario, there will only be coupling between closely neighboring voxels in the phase encoding direction. Instead of the blocks of $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$ being block-diagonal matrices (indicating coupling between all voxels in the phase encoding direction), we expect them to be banded matrices (i.e. matrices whose only non-zero entries lie on the main diagonal and other diagonals on either side) with very moderate bandwidths. To verify the sparsity pattern for such a pulse sequence (linear Cartesian with smoothly varying flip angles), in Fig. 4.1 a logarithmic plot of the magnitude of $\mathbf{J}^H\mathbf{J}$ is shown for a small, square numerical phantom of size $N_x = N_y = 16$. This numerical phantom is small enough so that $\mathbf{J}^H\mathbf{J}$ can be computed and stored in its entirety. It can be seen that $\mathbf{J}^H\mathbf{J}$ admits a block structure (5×5 blocks, one for each pair of distinct parameters: $T_1, T_2, |B_1|^+, \mathrm{Re}\left\{\rho\right\}, \mathrm{Im}\left\{\rho\right\}$) and that each block is, to good approximation, a very sparse, banded matrix. To capture the relevant information that is present in $\mathrm{Re}\left\{\mathbf{J}^H\mathbf{J}\right\}$, only a tiny fraction of its N^2 components need to be computed and stored.

If instead a very irregular flip angle pattern is used, or if the ordering of phase encoding lines is randomized, the magnetization matrices $\mathbf{m}_j^{(k)}$ will no longer be smooth and there will be coupling along the entire phase encoding direction. This can be seen in Supplementary Fig. 4.7 (irregular flip angle train) and Supplementary Fig. 4.8 (random ordering of phase encoding lines), where the blocks of $\mathbf{J}^H\mathbf{J}$ are block diagonal matrices with blocks of size $N_y \times N_y$.

In Supplementary Fig. 4.9 a logarithmic plot of the magnitude of $\mathbf{J}^H\mathbf{J}$ is shown for a situation where the sampling condition that each spatial frequency is sampled exactly $N_k \geq 1$ times is not fulfilled. In this scenario the central phase encoding line is sampled much more often than the other phase encoding lines, resulting in coupling along the entire phase encoding direction.

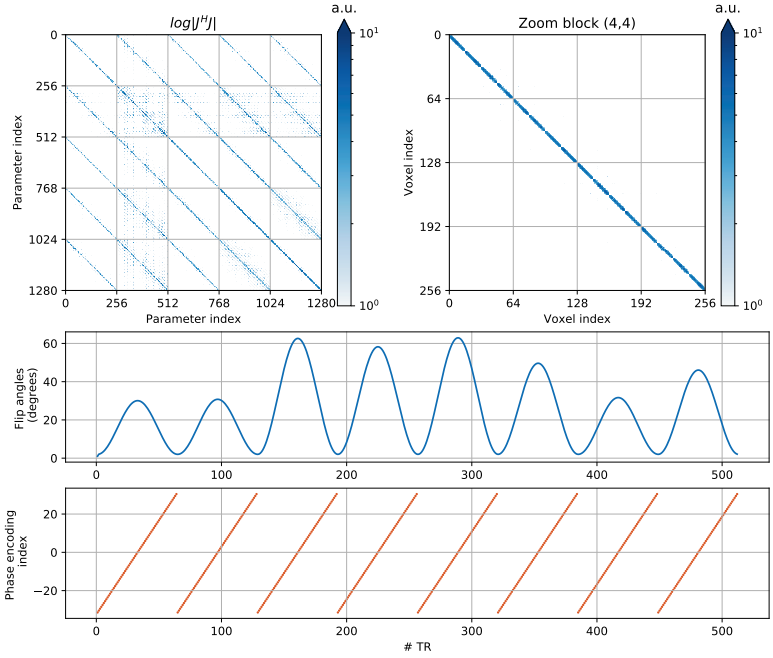

Figure 4.1: [a] Logplot of magnitude of $\mathbf{J}^H\mathbf{J}$ at a certain set of randomly generated parameters $(T_1, T_2, |B_1^+|, \mathrm{Re}\{\rho\}, \mathrm{Im}\{\rho\}$ in each voxel of a square matrix of size $N_x = N_y = 16$). [b] A zoom of one of the blocks of the matrix in [a]. [c-d] A pulse sequence with smoothly varying flip angles [c] and linear Cartesian sampling was used where each spatial frequency was sampled $N_k = 8$ times [d]. The matrix $\mathbf{J}^H\mathbf{J}$ in [a] is seen to consist of 25 blocks (of which 15 are unique due to symmetry) associated with the different combinations of parameter pairs. Each block contains, approximately, a main diagonal and a limited number of off-diagonals, indicating that there is only coupling between close neighbours in the phase encoding direction.

An example with a gradient-spoiled (i.e. "FISP") sequence instead of a gradient-balanced sequence is shown in Supplementary Fig. 4.10. In Supplementary Fig. 4.11 a (gradient-spoiled) sequence is used with random TRs and TEs throughout the sequence.

4.2.4.2 Non-Cartesian sampling

For non-Cartesian sampling patterns like radial or spiral the point spread functions $\mathcal{F}\left(\mathbf{m}_i^{(k)}\right)$ will propagate into all spatial directions. Assuming smooth $\mathbf{m}_j^{(k)}$ can be achieved, the blocks of $\mathbf{J}^H\mathbf{J}$ will not just have a band around the main diagonal like in the Cartesian case but there will be multiple bands. Depending on the pulse sequence used (e.g. smoothness of the flip angle train), it may be possible to compute and store the additional bands although the computational effort and memory requirements will be larger. Experiments with non-Cartesian sampling are outside the scope of the current work since MR-STAT, unlike MR Fingerprinting, is usually implemented with Cartesian readouts.

4.2.5 Determining sparse Hessian bandwidth

To determine an appropriate number of off-diagonals ("bandwidth") to be computed for each of the $N_p(N_p+1)/2$ unique blocks of $\mathbf{J}^H\mathbf{J}$ for a pulse sequence with smoothly varying flip angles and linear Cartesian sampling, we proceed as follows. Assume x and y correspond to the readout- and phase encoding directions, respectively. The magnetization response in a voxel with coordinates $x = y = 0$ is simulated using different combinations of physically realistic quantitative tissue parameters. After each simulation the resulting magnetization vector is reshaped into magnetization matrices ($\mathbf{m}_j^{(k)}$) and the 2D FFT is applied on each to generate point spread functions. See Fig. 4.2a for an example. To determine an appropriate width (in voxels) of the point spread function in the phase encoding direction, a threshold value is needed. The maximum magnitude of the point spread functions in the readout direction (i.e. the line $\{(x,y)|y=0\} \setminus (x=0, y=0)$) over all parameter combinations is taken as a threshold value. Given this threshold, for each of the point spread functions the width is determined as the number of voxels along the line $\{(x,y)|x=0\}$ for which the magnitude of the point spread function is above the threshold, see e.g. Fig. 4.2b for an example. Based on Eq. (4.2.23), the maximum kernel width over all simulated parameter pairs is expected to be a good estimate for the number of off-diagonals to be computed for the sparse Hessian approximation. Since only a few combinations of parameters are simulated for a single voxel, the time needed to compute the bandwidth is negligible compared to the time needed for the MR-STAT reconstructions and can be easily performed at the start of the reconstruction (or the bandwidth could already be known from previous reconstructions with similar sequence parameters).

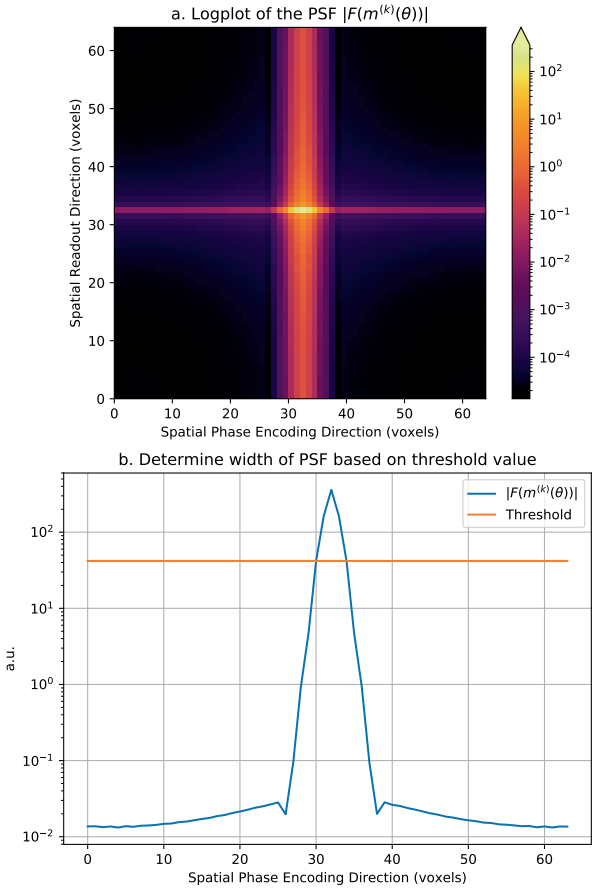

Figure 4.2: [a] Logarithmic plot of the magnitude of one of the point spread functions for a sequence with linear, Cartesian sampling and smoothly varying flip angles from Fig. 4.1. The quantitative parameter values used for the simulation were $T_1 = 0.8$ s, $T_2 = 0.05$s and $|B_1^+| = 1.2$. In the readout direction the point spread function is non-zero because of T_2-decay during the readouts whereas in the phase encoding direction it is non-zero because of the varying flip angles. [b] Line plot of the central, horizontal line of the point spread function shown in the top plot. The threshold value is determined as the maximum magnitude of the point spread functions in the readout direction (i.e. the lines $\{(x, y)|y = 0\} \setminus (x = 0, y = 0)$) , over a range of physically feasible quantitative tissue parameter combinations. The number of voxels in the phase encoding direction for which the magnitude of the point spread function is above the threshold is determined to be the width of this point spread function. In this example the width is five.

4.2.6 Distributed and memory-efficient computation of $\mathbf{J}^H\mathbf{J}$

Although general methods exist to efficiently compute the entries of a Hessian with a given sparsity structure [133, 39, 48], in the current work a custom algorithm is used to compute $\mathbf{J}^H\mathbf{J}$ for the case of sequences with linear, Cartesian sampling and smoothly varying flip angles. The algorithm is described in the Supplementary Material and utilizes the Gauss-Newton approximation to the Hessian and the fact that columns of the Jacobian matrix (associated with different voxels) can be computed independently from one another. The algorithm is memory-efficient in the sense that only few columns of the Jacobian matrix need to be stored in memory at once. In addition, the computational load can also be distributed efficiently among multiple CPUs.

4.3 Methods

The matrix-free MR-STAT (state-of-the-art) method (Alg. 4.1) and the proposed sparse MR-STAT method (Alg. 4.2) are compared on synthetically generated data, on data from gel phantoms and on in-vivo acquired brain data.

4.3.1 Acquisition scheme

In all experiments, a 2D balanced gradient-echo pulse sequence was used with smoothly varying flip angles and linear, Cartesian sampling similar to the sequence shown in Fig. 4.1. Note that the sequence is designed such that high flip angles (resulting in high signal) occur when the lower spatial frequencies are sampled. The peak values of the flip angle trains were generated randomly. To perform the gel phantom and in-vivo measurements the sequence was implemented on a 1.5 clinical MR system (Ingenia, Philips Healthcare, Best, The Netherlands). Data was acquired using the manufacturer's thirteen channel receive headcoil.

4.3.2 In-silico

For the synthetic experiment, the transienst-state sequence was used to simulate data with a (simulated) scan time of 7.8 seconds from a numerical brain phantom [9]. Noise was added to the simulated signal such that

$$\|\text{signal}\|_2/\|\text{noise}\|_2 = 50.$$

For the sparse method an optimal bandwidth of five was derived using the technique described in Section 4.2.5. To test the influence of the bandwidth on the convergence, reconstructions were performed using bandwidths of one, three, five (the derived optimal bandwidth for this sequence) and 55. For the

reconstruction with bandwidth one the number of outer iterations was increased to 100 because 25 was observed not to be enough to converge in this case.

4.3.3 Gel phantoms

Six gadolinium-doped gel phantoms (TO5, Eurospin II test system, Scotland) were scanned on the clinical 1.5 T MR system using the manufacturer's thirteen-channel receive headcoil. The field-of-view was set to 96 mm × 96 mm with a spatial resolution of 1 mm × 1 mm × 5 mm. In total 768 readout lines were acquired with a TR and TE of 7.4 ms and 3.7 ms, respectively, and such that each spatial frequency was sampled eight times. The readout bandwidth was 85.6 kHz and the total scan time was 5.7 s.

For comparison purposes, data was also acquired using gold standard T_1 and T_2 mapping methods. For T_1 mapping, an inversion-recovery single spin-echo experiment with inversion times of [50, 100, 150, 350, 550, 850, 1250] ms was performed. For T_2 mapping a single echo spin-echo experiments with echo times of [8, 28, 48, 88, 138, 188] ms was performed.

4.3.4 In-vivo

After having obtained written informed consent, signal from a 2D transverse slice of the brain of three healthy volunteers was acquired on the clinical 1.5T MR system using the vendors receive headcoil. For the first two volunteers, the field-of-view was set to 224 mm × 224 mm with a spatial resolution 1 mm × 1 mm × 5 mm. In total 1120 readout lines were acquired with a TR and TE of 7.6 ms and 3.8 ms, respectively, and such that each spatial frequency was sampled five times. The readout bandwidth was 85.6 kHz and the total scan time was 8.5 s. For the third volunteer, the spatial resolution was 1 mm × 1 mm × 3 mm, the TR and TE were 7.9 ms and 3.95 ms and again 1120 readout lines were acquired in 8.85 s with a readout bandwidth of 85.6 kHz.

4.3.5 Reconstructions

For all three experiments we reconstructed T_1, T_2 and proton density maps using both the matrix-free and sparse GN-method. Neither transmit field inhomogeneity nor off-resonance are taken into account in the reconstructions. All reconstructions were run on a local high performance computing cluster. The cluster consists of multiple Intel Xeon Gold 6148 nodes with 40 cores each and runs on the CentOS Linux 7 (Core) operating system. The reconstruction algorithms and the Bloch equation simulator were written in the open-source Julia programming language [15]. The number of inner and outer iterations for the GN-method were both limited to twenty. Prior to the reconstructions, signals from the different receive channels were compressed into a single virtual

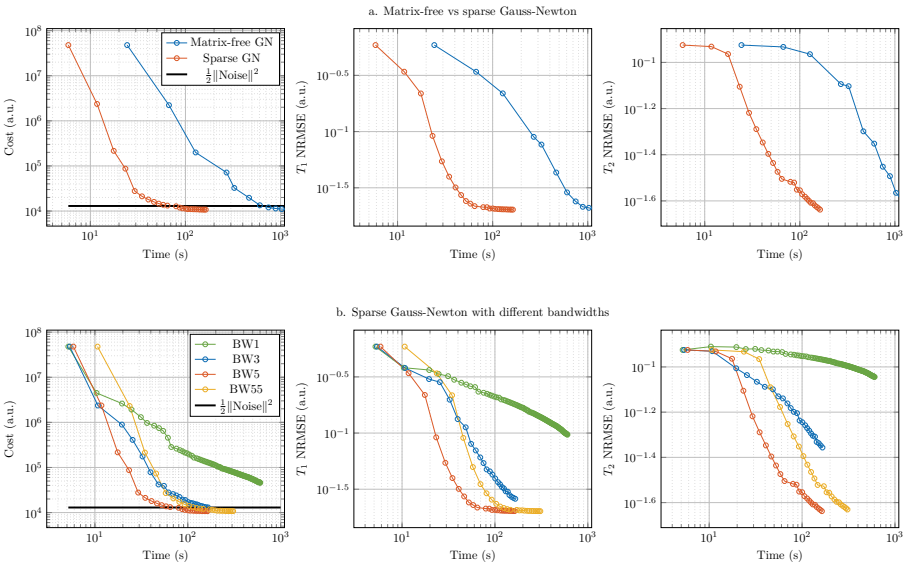

Figure 4.3: Convergence curves in terms of both the cost function and normalized root mean squared errors ("NRMSE") of the reconstructed T_1 and T_2 maps are shown for data from a numerical brain phantom. [a] In the first row, the matrix-free GN-method and the sparse GN-method are compared. The expected cost upon convergence is plotted as a black line in the left column. The sparse GN-method (with the derived optimal bandwidth of five) is seen to be approximately an order of magnitude faster. [b] In the second row, convergence curves for the sparse GN-method with different bandwidths are shown. The derived optimal bandwidth of five shows the best performance. A bandwidth of one results in poor convergence, even slower than the matrix-free method. Having a high bandwidth (e.g. 55) does not result in better steps in parameter space, it only increases computation time compared to the derived optimal bandwidth.

receive channel by performing a singular value decomposition and selecting the dominant singular vector [31].

4.4 Results

In-silico In Fig. 4.3, convergence curves in terms of both the cost function and the normalized root mean square errors ("NRSME") are shown. In the first row, the matrix-free and the sparse method with the derived optimal bandwidth of five are compared. Upon convergence of the parameter maps towards the ground truth values, the residual vector $\mathbf{r}$ is indistinguishable from noise and the achieved value of the cost function is expected to be similar to $\frac{1}{2}\|\text{noise}\|^2$. This value is indicated in the plots in the first column by a horizontal black line.

In Fig. 4.3a it can be seen that both the matrix-free method and the sparse method (with the derived optimal bandwidth of five) converge to the noise level. The sparse method outperforms the matrix-free method by approximately an order of magnitude.

Convergence curves for different bandwidths used in the sparse method are shown in the second row of Fig. 4.3. The sparse method with the derived optimal bandwidth of five shows the best performance. The sparse method with a bandwidth of one shows poor convergence because important inter-voxel correlations that arise from the transienst-state nature of the pulse sequence are ignored. The sparse method with a high bandwidth of 55 also converges but requires more computational effort because time is spent on non-relevant inter-voxel correlations.

Gel phantoms Reconstructed T_1 and T_2 maps for the gel phantoms using the matrix-free and the sparse reconstruction methods as well as the gold standard methods are displayed in Fig. 4.4. The mean values and standard deviations per tube are also reported. It can be seen that the mean values for both methods agree with the gold standard measurements. The standard deviations for both methods are also similar. The reconstructions were performed using 32 CPUs. The reconstruction time for the matrix-free method was nine minutes whereas for the sparse method it was one minute.

In-vivo In Fig. 4.5 in-vivo T_1, T_2 and proton density maps that are reconstructed using both the matrix-free and the sparse MR-STAT reconstruction algorithms are shown. Histograms for all maps are also displayed. From the maps and the histograms for the three volunteers it can be seen that, for practical purposes, the matrix-free and sparse methods converge to the same solution. For both methods we used 96 cores to perform the reconstructions. For the matrix-free method the reconstruction times were approximately 180 minutes per dataset whereas for the sparse method the reconstruction times were approximately sixteen minutes per dataset.

4.5 Discussion

Reconstructing parameter maps with MR-STAT requires dedicated algorithms that can deal with the inherent large scale of the non-linear optimization problem that needs to be solved. In the current work we have derived conditions under which the Hessian, as well as the Gauss-Newton approximation to the Hessian, admit a sparse structure. After having computed the sparse Hessian, the linear system in Eq. (4.2.9) can be solved rapidly in order to obtain steps $\mathbf{p}$ in parameter space. Compared to the previously used matrix-free method the reconstruction times can be reduced by approximately an order of mag-

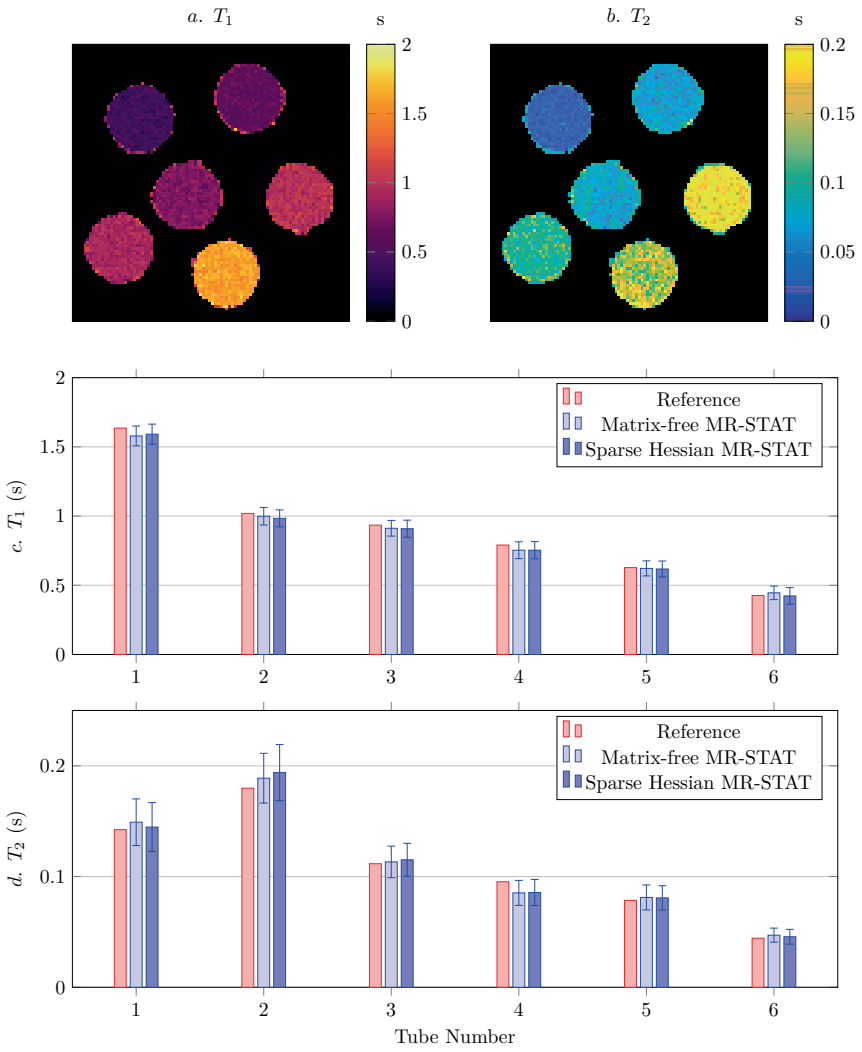

Figure 4.4: [a,b] The T_1 and T_2 maps reconstructed from the gel phantom data using the sparse MR-STAT method are shown. [c,d] Mean values and standard deviations are comparable to the values obtained with the matrix-free MR-STAT method. Reconstructions were performed using 32 cores. The reconstruction time for the matrix-free method was nine minutes whereas for the sparse method it was one minute. The mean values obtained per tube also show excellent agreement with the gold standard reference methods.

nitude without compromising on image quality, as has been demonstrated on simulation data, gel phantom data and in-vivo brain data.

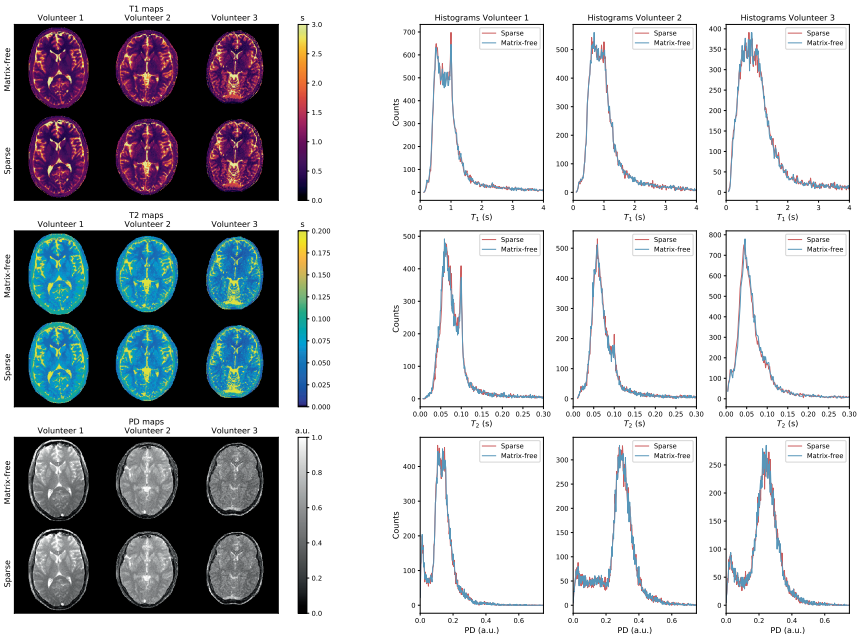

Figure 4.5: The T_1, T_2 and proton density maps as well as histograms reconstructed from the in-vivo brain data sets using both the sparse and matrix-free MR-STAT methods are shown. The maps and histograms show excellent agreement between the sparse and matrix-free method in all three volunteers, demonstrating that the sparse method can be used to accelerate MR-STAT reconstruction by an order of magnitude. Note that for volunteers 1 and 2, a slice thickness of 5 mm was use whereas for volunteer 3 a slice thickness of 3 mm was used in the acquisition.

The sparse Hessian method has only been verified in-vivo using gradient-balanced sequences in the current work. However, the conditions under which the Hessian is sparse were derived in a more general setting and , as shown in Supplementary Figures 4.10 and 4.11, for spoiled sequences similar sparsity patterns can be obtained as for gradient-balanced sequences. It is therefore expected that the sparse method will yield similar results in terms of acceleration for in-vivo measurements with gradient-spoiled sequences.

The derived conditions under which the Hessian is sparse are in principle not restrictive when performing actual MR measurements. The requirement of smooth magnetization matrices, which can be accomplished with smoothly varying flip angles and linear Cartesian sampling, are also important for mitigating the effects of eddy currents in gradient-balanced sequences [18, 30]. In addition, smoothly varying flip angles have also been shown to arise naturally from optimal sequence design experiments [8, 140]. The requirement to sample

each of the N_v spatial frequencies $N_k \geq 1$ times is in general also easily satisfied because, in order to make the MR-STAT problem (4.2.8) well-posed, at least $N_p \cdot N_v$ sample points need to be acquired so that N_k can be set to N_p.

In previous work [141, 69] linear, Cartesian sampling strategies were also used and for high resolution imaging a 1D FFT was applied along the readout direction prior to reconstruction to decouple the MR-STAT problem into multiple independent subproblems. The proposed sparse Hessian technique can be seen as a generalization of this approach. Rather than explicitly decoupling through application of the FFT - and introducing model inaccuracies (e.g. T_2 decay and ΔB_0 effects during readout are no longer modelled), the decoupling in the readout direction happens implicitly through assumptions on the sparsity pattern of the Hessian matrix. In addition, smoothness of the magnetization response - resulting in a limited bandwidth in the phase encoding direction and thus further sparsity - was not used in previous work.

Even though in the sparse method an approximation of $\mathrm{Re}\left\{J^H J\right\}$ is used (which itself approximates the Hessian matrix), the optimization is observed to converge to the same parameter maps. In fact, the sparse method can potentially lead to better update steps because iterations in the inner loop are computationally cheap and need not be limited, allowing it to take into account the curvature of the objective function in a more efficient manner. In the presence of noise we empirically observed it to be beneficial to limit the number of inner iterations in order to prevent overfitting. Adding regularization to the reconstruction problem in the form of (joint) sparsity constraints on the parameter maps [190, 129] might make the reconstruction procedure more robust and is the subject of further studies.

Note that whereas the matrix $\mathrm{Re}\left\{\mathbf{J}^H \mathbf{J}\right\}$ is positive-definite by construction, the sparse approximation is not guaranteed to be positive definite. The Steihaug algorithm [148] that is used in Alg 4.2 to determine steps in parameter space is able to deal with non-positive definite (approximate) Hessians by terminating when a direction of negative curvature is encountered. In practice no issues with non-positive definite matrices were encountered.

In the present work we restricted ourselves to sequences with short readouts for which the spin dynamics during readouts are assumed to have negligble impact on the measureable signal. Exploring the benefits of having a forward model which includes these effects (similar to e.g. time-segmented approaches [121, 154, 122]) for MR-STAT sequences with longer (e.g. EPI or spiral) readouts is the subject of further studies.

Although the conditions that result in a sparse Hessian were derived in the context of general transient-state sequences for which no analytical models are available, the outcomes apply equally well if such analytical models are available. The ideas presented in this work could therefore also be applied to accelerate other volumetric non-linear inversion methods that rely on such analytical models [20, 153, 189, 76, 175]. In this respect we note that the magnetization responses for sequences for which analytical models are available

(e.g. multi-echo spin-echo, inversion-recovery bSSFP) are typically smooth. We also note that in some of these works [20, 128], the authors propose first-order methods to solve the optimization problem while explicitly noting that scale-invariant methods like the Gauss-Newton method might be preferable but come with the disadvantage of having to compute the (approximate) Hessian. By relying on the ideas presented here it might become tractable to compute the Hessian, making the reconstruction procedures more robust as well.

A major challenge with the MR-STAT approach to qMRI is the long durations of the reconstructions. With the proposed sparse (approximate) Hessian technique we have been able to significantly reduce reconstruction times. The current computational bottleneck in the reconstruction algorithm is formed by the need to perform Bloch simulations, as well as partial derivative computations, at each outer iteration of the reconstruction algorithm to compute the signal $\mathbf{s}$ and the columns of the Jacobian $\mathbf{J}$, respectively. Further research is aimed at using surrogate models (see e.g. [37, 166]) for both these types of computations in order to reduce reconstruction times so that the potential benefits of MR-STAT in clinical practice can be explored.

4.6 Conclusion

We have derived conditions under which the Hessian matrices in the MR-STAT optimization admit a sparse structure. For pulse sequences with linear, Cartesian sampling a distributed and memory-efficient algorithm was provided to explicitely compute the sparse Hessians. Through in-silica, phantom and in-vivo experiments we have demonstrated that by exploiting the sparsity the MR-STAT reconstruction times can be reduced by approximately an order of magnitude, bringing the MR-STAT method closer to clinical practice.

Acknowledgments

This research is funded by the Netherlands Organisation for Scientific Research, domain Applied and Engineering Sciences, Grant #14125.

4.7 Supplementary Materials

Supplementary Material S1: Sparsity patterns

In Section 2.3 of the main manuscript conditions are derived under which the Hessian matrix in the MR-STAT problem is expected to have a sparse structure. In Section 2.4 pulse sequence design considerations are discussed for which the sparsity conditions are satisfied. For gradient-balanced (i.e. "bSSFP" or "trueFISP" based) pulse sequences with smoothly varying flip angles and linear,

Cartesian sampling a strong sparsity pattern is obtained as shown in Figure 1. In Supplementary Figure 4.6 a flip angle train is used that is less smooth and based on the analysis in the main manuscript this will result in more coupling along the phase encoding direction and thus less sparsity in the approximate Hessian. This effect can indeed by observed in the figure by the broadening of the diagonal bands.

If instead an extremely irregular flip angle pattern is used, or if the ordering of phase encoding lines is randomized, the magnetization matrices $\mathbf{m}_j^{(k)}$ will no longer be smooth at all and there will be coupling along the entire phase encoding direction. This can be seen in Supplementary Figure 4.7 (irregular flip angle train) and Supplementary Figure 4.8 (random ordering of phase encoding lines), where the blocks of $\mathbf{J}^H\mathbf{J}$ are block diagonal matrices with blocks of size $N_y \times N_y$.

In Supplementary Figure 4.9 a logarithmic plot of the magnitude of $\mathbf{J}^H\mathbf{J}$ is shown for a situation where we break the condition that each spatial frequency is sampled exactly $N_k \geq 1$ times. In this scenario the central phase encoding line is sampled many more times than the other phase encoding line, resulting in coupling along the entire phase encoding direction.

In Supplementary Figure 4.10 results are shown for a gradient-spoiled (i.e. "FISP") sequence with smoothly varying flip angles. In this sequence, a fixed TR of 8.8 ms and TE of 4.4 ms are used. If instead the TR and TE are randomized throughout the sequence (with mean, maximum and minimum TR of 16 ms, 8.8ms and 23.8 ms, respectively, and mean, maximum and minimum TE of 6.9 ms, 4.4 ms and 9.3 ms, respectively) the magnetization response remains smooth and has little effect on the sparsity pattern, as can be seen in Supplementary Figure 4.11.

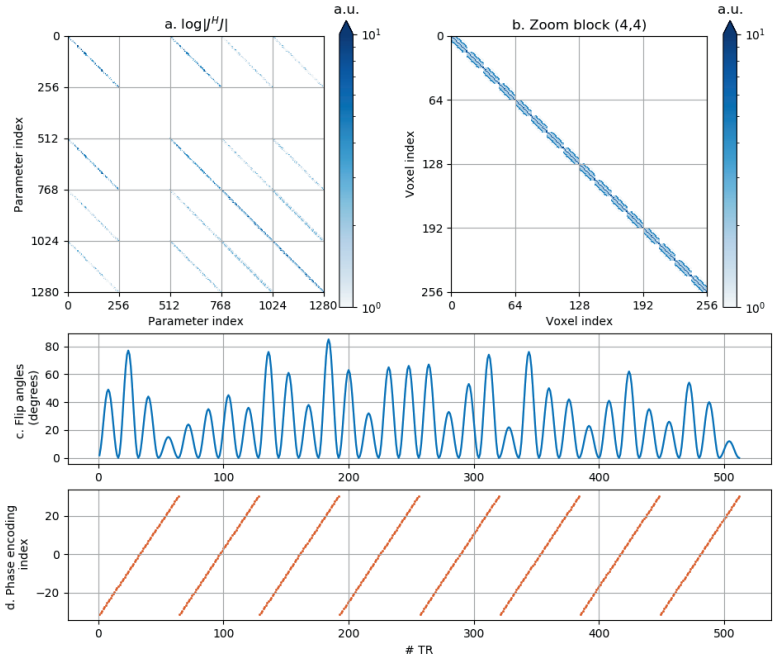

Figure 4.6: Logplot of magnitude of $\mathbf{J}^H\mathbf{J}$ in a similar scenario as in Fig. 1 of the main manuscript, except that the flip angle train is made less smooth. Each spatial frequency is still sampled $N_k = 8$ times. The diagonal block structure of each of the 25 blocks of $\mathbf{J}^H\mathbf{J}$ indicates coupling between more voxels in the phase encoding direction compared to the situation of Fig. 1 of main manuscript.

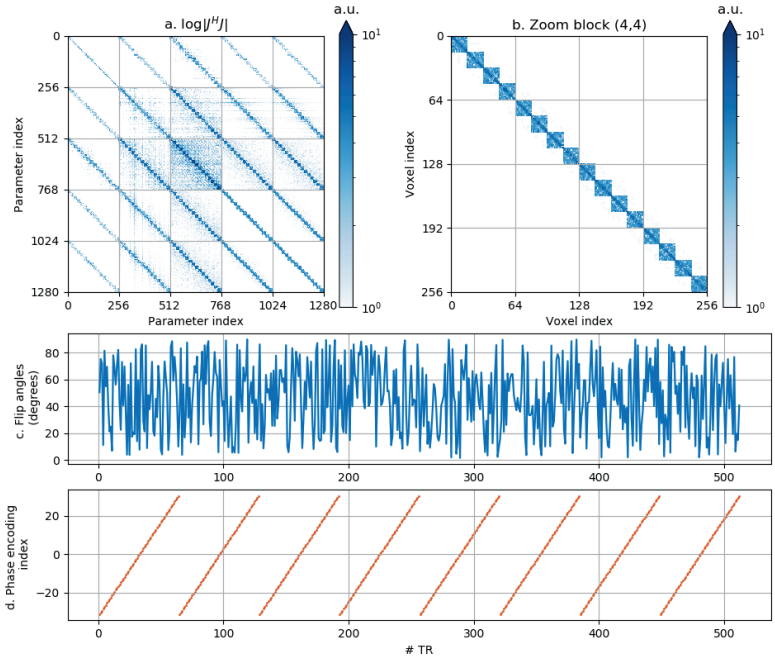

Figure 4.7: Logplot of magnitude of $\mathbf{J}^H\mathbf{J}$ in a similar scenario as in Fig. 1 of the main manuscript, except that a very irregular flip angle pattern is used. The diagonal block structure of each of the 25 blocks of $\mathbf{J}^H\mathbf{J}$ indicates coupling between (almost) all voxels in the phase encoding direction. Compared to the situation of Fig. 1 of the main manuscript, the matrix is much less sparse.

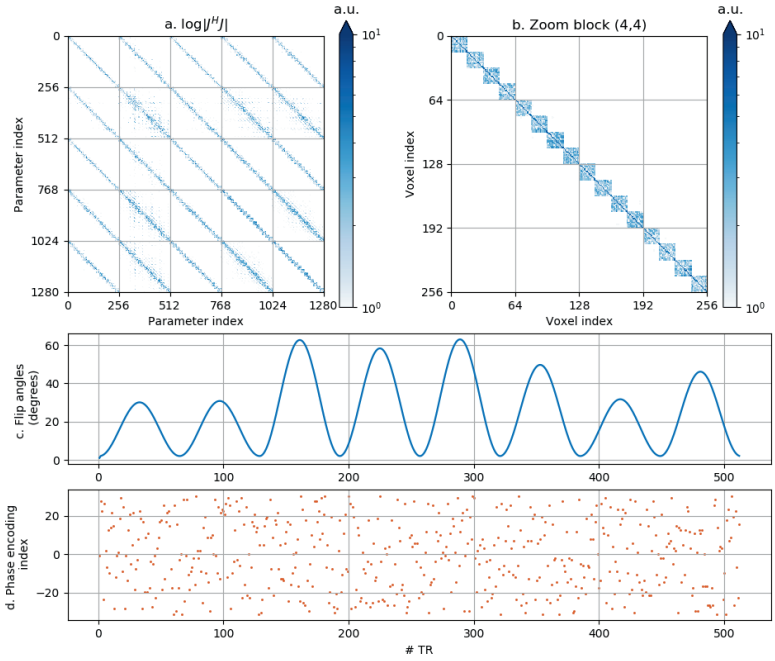

Figure 4.8: Logplot of magnitude of $\mathbf{J}^H\mathbf{J}$ in a similar scenario as in Fig. 1 of the main manuscript, except that a random ordering of the phase encoding lines is chosen. Each spatial frequency is still sampled $N_k = 8$ times. The diagonal block structure of each of the 25 blocks of $\mathbf{J}^H\mathbf{J}$ indicate coupling between (almost) all voxels in the phase encoding direction. Compared to the situation of Fig. 1 of main manuscript, the matrix is much less sparse.

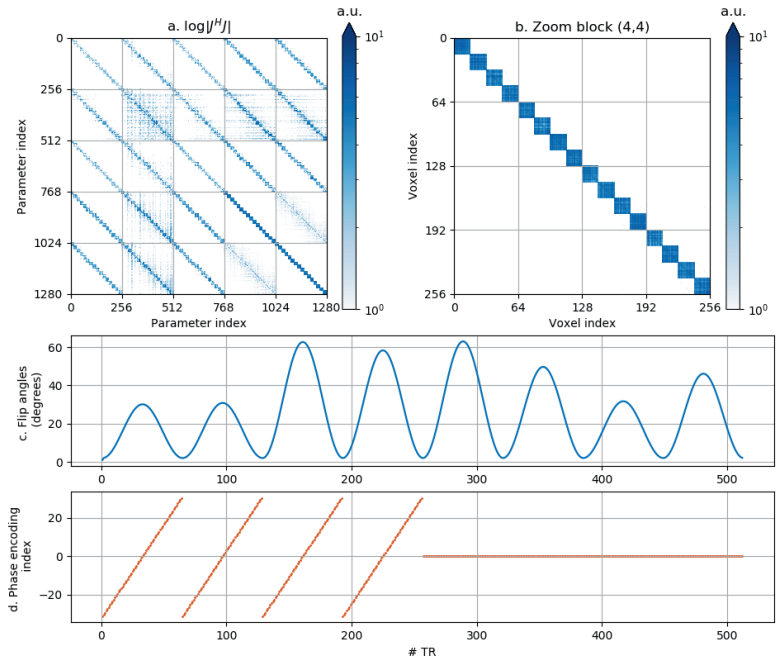

Figure 4.9: Logplot of magnitude of $\mathbf{J}^H\mathbf{J}$ in a similar scenario as in Fig. 1 of the main manuscript, except that the central phase encoding line was sampled many more times than the other phase encoding lines. As a result there is coupling between (almost) all voxels in the phase encoding direction, as indicated by each block $\mathbf{J}^H\mathbf{J}$ being a block diagonal matrix with blocks of size $N_y \times N_y$. Compared to the situation of Fig. 1 of the main manuscript, the matrix is much less sparse.

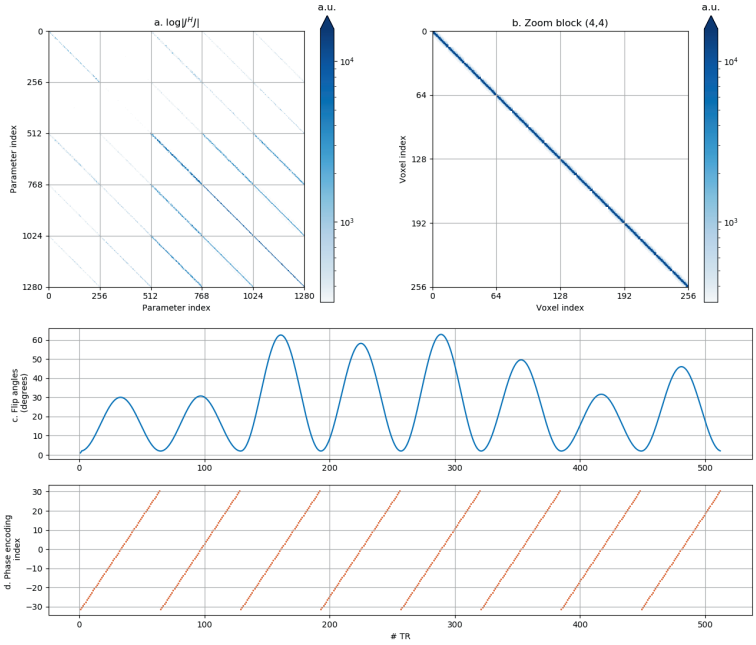

Figure 4.10: Logplot of magnitude of $\mathbf{J}^H\mathbf{J}$ in a similar scenario as in Fig. 1 of the main manuscript, except that a gradient-spoiled (i.e. "FISP") sequence is used. Similar to the case with a gradient-balanced (i.e. "bSSFP" or "trueFISP") sequence, there is a high degree of sparsity in the phase encoding direction.

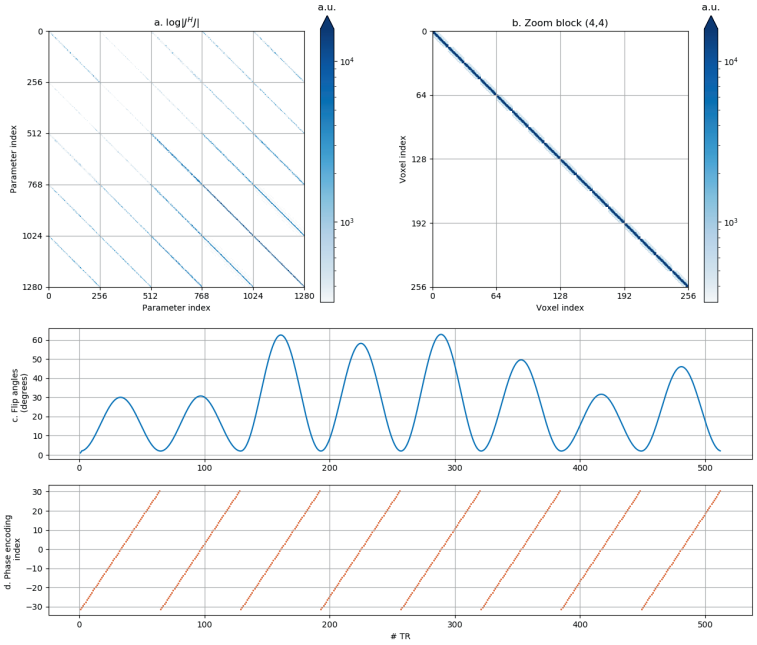

Figure 4.11: Logplot of magnitude of $\mathbf{J}^H\mathbf{J}$ in a similar scenario as in Supplementary Figure 4.10, except that the TR and TE are randomized throughout the pulse sequence. As can be seen from the figure, the randomized TR and TE have little effect on the sparsity pattern.

Supplementary Material S2: Memory-efficient and distributed computation of sparse $\mathbf{J}^H\mathbf{J}$

Knowing the sparsity structure for $\mathbf{J}^H\mathbf{J}$ for a sequence with linear, Cartesian sampling and smoothly varying flip angles, we still need a procedure to compute its non-zero entries. In this respect, note that the Jacobian matrix $\mathbf{J}$ will generally be too large to fit into computer memory [69]. With the help of Figure 4.12 we now outline an algorithm to compute the non-zero entries of $\mathbf{J}^H\mathbf{J}$ in a memory-efficient manner. Assume the appropriate bandwidth has been determined to be BW. In the illustration $BW = 9$ and only two parameters per voxel, namely T_1 and T_2, are considered. Assume that the gray entries of $\mathbf{J}$ and $\mathbf{J}^H\mathbf{J}$ have already been computed and that currently the green and blue colums of $\mathbf{J}$ are stored in memory. With these columns, the green and blue entries of $\mathbf{J}^H\mathbf{J}$ can be computed by taking pair-wise inner products of these columns. Now the blue columns can be dropped from memory and the red columns of $\mathbf{J}$ can be computed and stored. Subsequently, the red entries of $\mathbf{J}^H\mathbf{J}$ can be computed and stored. Repeat this process until the entire sparse $\mathbf{J}^H\mathbf{J}$ has been stored and computed. This procedure only requires $BW \times N_p$ columns of $\mathbf{J}$ to be stored into memory at the same time which, for moderate vales of BW, will be feasible on modern computer architectures.

Also note that, since the columns of $\mathbf{J}$ associated with different voxels can be computed independently from columns associated with other voxels, the computation of $\mathbf{J}^H\mathbf{J}$ can be performed parallel by distributing the voxels over the available number of computing cores.

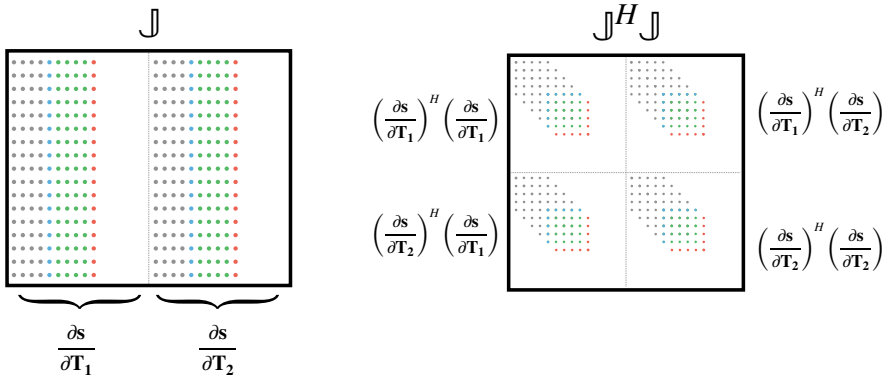

Figure 4.12: Illustration of how to compute the sparse $\mathbf{J}^H\mathbf{J}$ for a linear, Cartesian sequence with smoothly varying flip angles in a memory-efficient manner. In this figure we only consider two parameters per voxel, T_1 and T_2, and the bandwidth is set to nine. Assume that the gray entries of $\mathbf{J}$ and $\mathbf{J}^H\mathbf{J}$ have already been computed and that currently the green and blue colums of $\mathbf{J}$ are stored in memory. With the columns, the green and blue entries of $\mathbf{J}^H\mathbf{J}$ can be computed by taking pair-wise dot products of these columns. Now drop the blue columns from memory and compute and store the red columns of $\mathbf{J}$. Then the red entries of $\mathbf{J}^H\mathbf{J}$ can be computed and stored. Repeat this process until the entire sparse $\mathbf{J}^H\mathbf{J}$ has been stored and computed. Note that, since the columns of $\mathbf{J}$ associated with different voxels can be computed independently from columns associated with other voxels, the computation of $\mathbf{J}^H\mathbf{J}$ can be performed parallel by distributing the voxels over the available number of computing cores.

Supplementary Material S3: Multi-channel receive arrays

In the derivation in the main manuscript it was assumed that data is received by only a single receive channel. However, in most MRI acquisitions, data is received with multiple receive channels. Let $N_c \geq 1$ be the number of receive channels and define $d_i, i \in 1, \ldots, N_c$ to be the data from receiver i. Assuming the coil sensitivities are known (from either a pre-scan or by using a self-calibration method), let s_i be the signal model for receiver i. For each channel a residual vector is defined as $r_i := d_i - s_i$ and the multi-coil equivalent of the objective function becomes

$$f(\alpha) = \sum_{i=1}^{N_c} \|r_i(\alpha)\|_2^2. \tag{4.7.1}$$

The corresponding Gauss-Newton approximation to the Hessian matrix is then given by

$$\operatorname{Re}\left\{J^H J\right\} = \sum_{i=1}^{N_c} \operatorname{Re}\left\{J_i^H J_i\right\}, \tag{4.7.2}$$

where

$$J_i := \frac{\partial r_i}{\partial \alpha}. \tag{4.7.3}$$

The analysis from the main manuscript can be applied to each of the $J_i^H J_i$ terms separately. Assuming the $J_i^H J_i$ to admit a sparsity pattern, from Eq. (4.7.2) it follows that the sparsity pattern carries over to $\operatorname{Re}\left\{J^H J\right\}$.

We also note that because the coil sensitivities only enter as linear scaling factors in the Jacobian matrices compared to the single-coil case, the computational overhead of computing the sparse Hessian for the multi-coil case is very limited.

5

GPU-accelerated Bloch simulations and MR-STAT reconstructions using the Julia programming language

Oscar van der Heide
Cornelis A. T. van den Berg
Alessandro Sbrizzi

Submitted to IEEE Transactions on Medical Imaging

Abstract

MR-STAT is a relatively new multi-parametric quantitative MRI technique. Quantitative paramter maps are obtained by solving a computationally demanding large-scale non-linear optimization problem. In this work, we aim to reduce MR-STAT reconstruction times by leveraging GPU hardware using the Julia programming language. The computationally demanding tasks of a previously proposed MR-STAT reconstruction algorithm are identified. One of these tasks consists of performing many Bloch simulations. A Bloch simulation framework is developed and released as a standalone package called BlochSimulators.jl. The package is benchmarked in terms of runtime performance against other Bloch simulation toolboxes developed in C, CUDA or Julia and demonstrates superior performance. Furthermore, a GPU-compatible (partially matrix-free) modification of the MR-STAT reconstruction algorithm is presented that utilizes BlochSimulators.jl. The proposed algorithm results in MR-STAT reconstruction times in the order of one minute per 2D slice.

5.1 Introduction

Magnetic resonance imaging ("MRI") is a medical imaging modality that is used in clinical practice to visualize anatomical structures. Compared to other imaging modalities such as CT and PET, the main advantage of MRI is the superior soft tissue contrast. The MR scanner utilizes radiofrequency waves together with rapidly switching magnetic gradient fields to be able to sample spatial frequency ("k-space") information. The inverse Fast Fourier Transform ("FFT") is then typically used to generate diagnostic contrast images. While the direct FFT approach results in very fast reconstructions, the Nyquist-sampling criterion must be satisfied for the FFT to be succesful, which, due to the sequential manner of sampling spatial frequencies, results in relatively lenghty data acquisition procedures compared to other imaging modalities. For quantitative MRI ("qMRI"), where the goal is to provide quantitative estimates of certain physical tissue properties, the same field-of-view is typically imaged multiple times with different scan settings to generate multiple contrasts, followed by a voxelwise fit of an analytical signal model. The issue of long scan times is thus amplified for these traditional qMRI approaches and consequently qMRI has only seen limited clinical adoption.

In recent years, new transient-state qMRI techniques like MR Fingerprinting [110] and MR-STAT [141] have emerged that drastically reduce qMRI acquisition times. In MR-STAT, the qMRI problem is posed as a large-scale non-linear inversion problem. Multiple quantitative tissue parameter maps (e.g. T_1, T_2) are generated from a single-short scan by fitting a Bloch-equation based forward model to the measured, transient-state k-space samples. A major challenge with this technique is managing the reconstruction times and memory requirements of the reconstruction algorithm. In previous work, MR-STAT reconstruction times have been accelerated by utilizing CPU-parallelization [69], surrogate modelling [106] as well as algorithmic techniques that rely on the assumption of Cartesian-based gradient trajectories ([66, 106]).

Since GPU hardware has been demonstrated to result in runtime performance gains in many different areas of research, including the field of MRI reconstructions [152, 174], it could be beneficial to implement (the computationally demanding parts) of the MR-STAT reconstruction algorithm on GPU hardware. However, implementing such an algorithm in a low-level language like CUDA is challenging and time consuming, especially for MRI researchers whose main area expertise typically does not include low-level programming.

In this work, we will instead use the Julia language programming language [15] to implement a GPU-accelerated adaptation of the MR-STAT reconstruction algorithm. Julia is a relatively new programming language that is free and open source, and it is designed specifically for scientific programming purposes. Its so-called multiple dispatch type system together with its Just-In-Time ("JIT") compiler allows for quick and easy prototyping on the one hand

(similar to dynamically typed, interactive languages like Matlab and Python) while being able to achieve the same runtime performance as low-level, compiled languages such as C, C++ and Fortran [83]. Together with the CUDA.jl package [**?**], one can use Julia to write functions that compile down to native CUDA kernels to be executed on GPU hardware. We will make use of these language features to implement custom kernels for the computationally demanding tasks of the MR-STAT reconstruction algorithm. The relevant implementation considerations will be described and the runtime performance will be benchmarked against previously presented MR-STAT reconstruction implementations.

One of the important and computationally demanding parts of the MR-STAT reconstructions is formed by the need to perform Bloch simulations at each iteration of the iterative reconstruction algorithm. Bloch simulations are also important for other quantitative, transient-state MR reconstruction techniques such as MR Fingerprinting and several Bloch simulation software tools written in a range of programming languages are already available online [64, 149, 102, 172, 33]. Instead of relying on these pre-existing toolboxes, a new Julia Bloch simulation toolbox was developed as part of this work. This toolbox is highly flexible in the sense that it allows users to implement pulse sequences that can then be executed on multiple hardware architectures (single CPU, multi-threaded CPU, distributed CPU as well as GPU) using multiple numbers types (e.g. single- and double precision) and without code repetition. State-of-the-art runtime performance is achieved, as will be demonstrated through benchmarks against existing MR physics and fast surrogate-modelling based toolboxes. Our Bloch simulation toolbox is released as a stand-alone Julia package called BlochSimulators.jl, available at `https://github.com/oscarvanderheide/BlochSimulators.jl`.

5.2 Theory

In this section we first provide a high-level overview of the MR-STAT reconstruction algorithm presented in [69] and we list the computational bottlenecks that can benefit from GPU acceleration.

Let $\mathbf{d} \in \mathbb{C}^{N_t}$ be the vector of time-domain samples measured at N_t sampling times $t_1, \ldots, t_{N_t}$ during an MR experiment. The time-dependent forward model s (after spatial discretization) that is used to synthesize time-domain samples is given by

$$\mathbf{s}(t) = \sum_{j=1}^{N_v} m(\boldsymbol{\theta}_j, t) e^{-2\pi i \boldsymbol{k}(t)\cdot \boldsymbol{r}_j} \Delta_V, \tag{5.2.1}$$

where N_v is the number of voxels within the field-of-view, Δ_V is the volume element for each voxel, $\boldsymbol{r}_j$ is the vector of spatial coordinates for the voxel associated with index j, $\boldsymbol{\theta}_j$ is the vector of MR-relevant biophysical tissue

properties (e.g. $T_1, T_2, \rho, \ldots$) for the voxel associated with index j, $\boldsymbol{k}(t)$ is the k-space trajectory and m is the complex transverse magnetization whose time-varying behavior is obtained from Bloch simulations. Define the vector $\boldsymbol{s}$ as

$$\boldsymbol{s} := [s(t_1), \ldots, s(t_{N_t})] \in \mathbb{C}^{N_t}.$$

Note that $\boldsymbol{s}$ depends on the tissue parameters $\boldsymbol{\theta}_j$ for all voxels j. Concatenate all tissue parameters into a single vector $\boldsymbol{\alpha}$. The parameter maps (contained in $\boldsymbol{\alpha}$) are then obtained by numerically solving the inverse problem

$$\boldsymbol{\alpha}^* = \operatorname{argmin}_{\boldsymbol{\alpha}} \frac{1}{2} \|\boldsymbol{d} - \boldsymbol{s}(\boldsymbol{\alpha})\|_2^2 . \tag{5.2.2}$$

using the Gauss-Newton iterative algorithm as outlined in Alg. 5.1.

Algorithm 5.1 Minimize $\boldsymbol{\alpha} \to \frac{1}{2}\|\mathbf{d} - \mathbf{s}(\boldsymbol{\alpha})\|_2^2$

Require: Initial guess $\boldsymbol{\alpha}_0$
while not converged **do**
1. Evaluate forward model $\mathbf{s}$ at $\boldsymbol{\alpha}$
2. Compute residual: $\mathbf{r} = \mathbf{d} - \mathbf{s}$
3. Compute gradient: $\mathbf{g} = \operatorname{Re}\left(\mathbf{J}^H \mathbf{r}\right)$, where $\mathbf{J} := -\frac{\partial \boldsymbol{s}}{\partial \boldsymbol{\alpha}}$
4. Solve linear system: $\operatorname{Re}\left(\mathbf{J}^H \mathbf{J}\right) \mathbf{p} = -\mathbf{g}$
5. Update parameters: $\boldsymbol{\alpha} = \boldsymbol{\alpha} + \mathbf{p}$

end while

The computationally demanding steps of this iterative algorithm are Steps 1, 3 and 4. In Step 1, Bloch simulations must be performed in each voxel to be able to evaluate the forward model $\mathbf{s}$. Even for 2D reconstructions, there can be approximately 10^5 voxels with approximately 10^6 time-points per voxel. In Step 3, the gradient of the objective function is computed, which requires a matrix-vector multiplication with the Jacobian matrix $\mathbf{J}$ whose columns are given by partial derivatives of the forward model $\mathbf{s}$ with respect to the tissue parameters in all voxels α. Computing partial derivatives of the forward model is typically even more computationally demanding than the Bloch simulations themselves. In Step 4, a linear system involving the Gauss-Newton matrix $\mathbf{J}^H\mathbf{J}$ is numerically solved. However, as described in [69], directly inverting $\mathbf{J}^H\mathbf{J}$ is infeasible for the typical MR-STAT problem sizes. Instead, it was proposed to numerically solve the linear system by an iterative solver for linear least squares problems (LSQR [126]). This solver requires repeated multiplications with $\mathbf{J}$ and $\mathbf{J}^H$. The number of entries of the matrix $\mathbf{J}$ is typically in the order of $10^6 \times 10^6$ for 2D problems and is therefore too large to store in computer memory. A matrix-free algorithm was proposed to be able to compute the matrix-vector multiplications required for steps 3 and 4 and the feasibility of this approach was demonstrated on a CPU computing cluster.

The main aim of this work will be to accelerate MR-STAT reconstructions by using Julia to perform each of the following computationally demanding tasks on GPU hardware:

(A) Bloch simulations (Step 1);

(B) Partial derivatives of the forward model $\mathbf{s}$ with respect to α (Step 3); and

(C) Matrix-vector products with $\mathbf{J}$ and $\mathbf{J}^H$ (Step 4).

5.2.1 Bloch Simulations

5.2.1.1 Numerical integration of Bloch equations

In MRI, individual spin isochromats are modeled as 3D vectors $\boldsymbol{M} = (M_x, M_y, M_z)$ whose dynamics are described by the Bloch equations [82]

$$\frac{d\mathbf{M}}{dt} = (\gamma \boldsymbol{B}(t) + \mathbf{D})\, \boldsymbol{M}(t) + \frac{\mathbf{M}_0}{T_1}. \tag{5.2.3}$$

Here γ is the nuclear magnetic ratio, $\boldsymbol{B}(t)$ is a 3×3 matrix containing the pulse sequence-dependent magnetic field at each timepoint,

$$\mathbf{D} = \operatorname{diag}(-1/T_2, -1/T_2, -1/T_1) \tag{5.2.4}$$

contains the longitudinal (T_1) and transversal (T_2) relaxation times of the spin isochromat, and $\mathbf{M}_0 = (0, 0, M_{z,0})$ is its equilibrium magnetization.

In Bloch simulations, the ordinary differential equation ("ODE") with time-varying coefficients from Eq. 5.2.3 is integrated in time for a specific pulse sequence. In general, no analytic solutions exist. However, in specific scenario's analytic solutions do exist: when relaxation is ignored and the magnetic field $\mathbf{B}(t)$ is assumed constant during a timestep Δt, an analytic solution exists in the form of multiplication with the 3D rotation matrix $e^{-\gamma \Delta t \mathbf{B}}$. Alternatively, when $\mathbf{B}(t)$ is ignored during a timestep Δt, an analytic solution is available in the form of multiplication with the (diagonal) decay matrix $e^{\Delta t \mathbf{D}}$ and addition of the (longitudinal) regrowth vector $(1 - e^{-1/T_1})\mathbf{M}_0$. When performing Bloch simulations, instead of relying on general purpose ODE solvers, a commonly used method is to manually discretize the pulse sequence in time and for each time-step applying a splitting method where first the 3D rotation is applied, followed by decay and regrowth operations. That is, given $\mathbf{M}$ at time t, the magnetization at time $t + \Delta t$ is computed as

$$\boldsymbol{M}(t + \Delta t) = e^{\Delta t \mathbf{D}} e^{-\gamma \Delta t \mathbf{B}(\mathbf{t})} \mathbf{M}(t) + (1 - e^{-1/T_1})\mathbf{M}_0. \tag{5.2.5}$$

We refer to [167] for more details on the splitting approach.

In the presence of spoiling gradients, it may be inefficient to model the MR signal using individual spin isochromats and the extended phase graph ("EPG")

model [71, 72] may be more appropriate. In the EPG model, instead of keeping track of individual spin isochromats, one tracks the dynamics of so called configuration states. Like in the case of the isochromat model, Bloch simulations can be performed by repeatedly applying rotation, decay and regrowth operations but in addition, a spoiling operator is required that shifts configuration states (see [177] for a review of the EPG framework).

5.2.1.2 BlochSimulators.jl

In this subsection we provide a high level overview of the design of BlochSimulators.jl - our Julia implementation of a Bloch simulation framework with a focus on speed and flexibility.

For both the isochromat and EPG models, functions are written that implement the above mentioned `rotation, decay, regrowth` and `spoiling` (EPG only) operations. The functions are designed to be both *type-stable* and *non-allocating*, and form the core of the BlochSimulators.jl package. Type-stability allows Julia's JIT compiler to generate efficient machine code without having to hard-code - for example - whether single or double-precision number types are used. This is useful because, for example, on GPU one might want to perform computations using 32-bit (complex) floating point numbers for performance reasons. On the other hand, when precision is more important than runtime performance, one might want to use 64-bit floating point numbers instead.

The operator functions are non-allocating in the sense that relatively expensive heap allocations are avoided during their execution. This is in general beneficial for runtime performance. An important ingredient for achieving non-allocating code is the StaticArrays package which introduces arrays whose sizes are known at compile time and thus allows them to be stack-allocated. In BlochSimulators.jl, the isochromats and configuration state matrices are stored using `StaticArrays`. In addition, on the GPU, the configuration state matrices are stored in shared memory instead to prevent L1-cache eviction during simulations. This was observed to result in significantly faster GPU simulations. For the isochromat-based model no difference was observed when using shared memory, likely because the isochromat vectors are small enough that they already remain in L1 cache during the entire simulation.

Simulators for entire pulse sequences can in principle be assembled by combining the individual operator functions. Because the individual operator functions are type-stable and non-allocating, a proper combination of these functions should result in type-stable and non-allocating sequence simulations as well. In BlochSimulators.jl, we follow a convention where computing the magnetization at echo times without taking into account gradient encoding is separated from computing the magnetization at other readout times (with gradient encoding) and performing the volume integral in Eq. 5.2.1. The reason for this separation is twofold. First of all, for some applications such as computing MRF dictionaries, only the magnetization at echo times without gradient en-

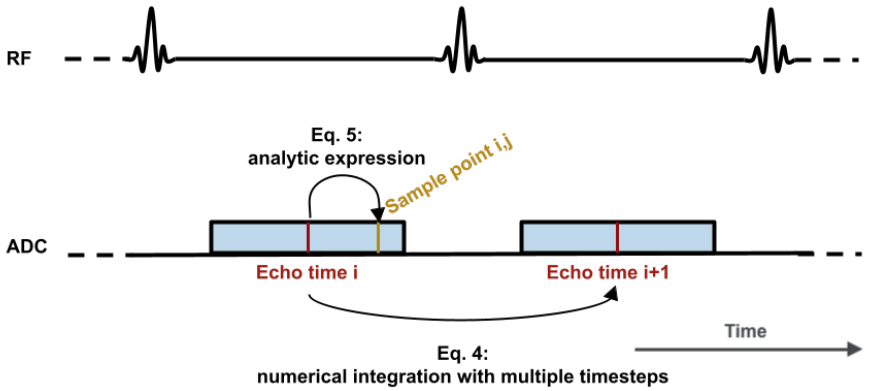

Figure 5.1: In BlochSimulators.jl a convention is followed where at first only the magnetization at the different echo times is computed. Eq. 5.2.5 is used to propagate the magnetization state from the i-th echo time to the $(i+1)$-th echo time. Typically there are RF excitations (indicated by the waveforms on the RF line) in between echo times which may require multiple timesteps to be properly simulated. Given the magnetization at the ith echo time, the magnetization at the j-th sample point of that readout (the blue bar on the ADC line) is computed analytically using Eq. 5.2.6.

coding is needed. Secondly, when performing 2D scans, slice profile correction mechanisms typically involve a summation over multiple simulations per voxel using either different locations of isochromats in the slice-select direction or different effective RF flip angles [97]. After the net transverse magnetization (i.e. after summation) in a voxel at some echo time t^* has been computed, a simplification of Eq. 5.2.5 allows the net transverse magnetization at another time t during that readout to be computed analytically as

$$\mathbf{M}(t) = \mathbf{M}(t^*)e^{-\frac{t-t*}{T_2}}e^{-2\pi i(t-t^*)\Delta B_0}e^{-2\pi i\boldsymbol{k}(t)\cdot\mathbf{r}}, \tag{5.2.6}$$

where $\mathbf{r}$ is the spatial position of the voxel, $\mathbf{k}$ is the k-space trajectory and ΔB_0 is the local off-resonance. The net magnetization in a voxel at other readout times can thus be computed *after* the summation necessary for the slice profile correction, therefore reducing the total number of computations. The two different timescales (i.e. between echo times and between sample points within a readout) are illustrated in Fig. 5.1.

In order to compute the magnetization at echo times for a particular sequence, a new custom Julia "struct" needs to be introduced first. The struct should be a subtype of either `IsochromatSimulator` or `EPGSimulator` to indicate which model will be used for simulations. The struct must have fields containing the necessary information to describe the pulse sequence. For example, in a variable-flip angle MR-STAT sequence, the sequence struct should

hold - among other things - the list of different flip angles for each T_R. In a multi-echo experiment, the struct should contain information on how many echos there are and what their respective echo times are. Given the sequence struct, a new `simulate!` method must be added which uses the fields of the struct, together with the above mentioned core operators, to implement the actual pulse sequence. Example sequence implementations are provided as part of the source code of BlochSimulators.jl (stored in `src/sequences`) but users of the package are able to assemble their own as well.

To perform Bloch simulations with a given `sequence`, the user must also provide a `parameters` array which contains tissue parameters to be used as inputs for the simulations (e.g. different combinations of T_1 and T_2 values, see `src/tissueparameters.jl` for currently supported parameter combinations). In addition, the user must specifiy the computational `resource` to be used for the simulations. BlochSimulators.jl supports single CPU (`CPU1()`), multi-threaded CPU (`CPUThreads()`), distributed CPU (`CPUProcesses`) and GPU modes (`CUDALibs()`). A function call of the form

```
output = simulate(resource, sequence,
  parameters)
```

will result in an array with the magnetization at echo times for each combination of input parameters, which can be used as a dictionary for MR Fingerprinting.

In order to evaluate the forward model Eq. 5.2.1, which involves a volume integral of the magnetization at all readout times, a custom trajectory struct (subtype of `AbstractTrajectory`) must be introduced with fields that can describe the gradient trajectory. For each such struct, a method must be added to the function `to_readout_point` which implements Eq. 5.2.6 for that particular trajectory. Examples are provided for Cartesian and radial gradient trajectories. In these examples, rather than storing the k-space coordinates for all sample points, we use a compact representation where for each readout only the starting point in k-space and the step in k-space per sample point is stored. This compact representation of the trajectory reduces memory access during runtime to allow for an increased compute throughput.

Besides a `trajectory`, the user must also provide coil sensitivity profiles (`coil_sensitivities`). The forward model Eq. 5.2.1 is then evaluated with a function call of the form

```
output = simulate(resource, sequence,
  parameters, trajectory,
  coil_sensitivities)
```

A graphical overview of the code structure underlying BlochSimulators.jl is shown at the online repository (`https://github.com/oscarvanderheide/BlochSimulators.jl`).

5.2.2 Partial derivatives of the forward model

To compute partial derivatives of the forward model with respect to the tissue parameters $\boldsymbol{\alpha}$, we use the finite difference technique [140] to compute the partial derivatives at echo times first. The finite difference technique directly translates the GPU acceleration from BlochSimulators.jl to the partial derivative computations. Given the partial derivatives at echo times, the partial derivatives at other readout times are computed by manually differentiating the analytical expression from Eq. 5.2.6 and, again, custom CUDA kernels are written in Julia to execute these partial computations on GPU hardware.

5.2.3 Matrix-vector products with J and $\mathbf{J}^H$

Computing the gradient in step 3 of Alg. 5.1 involves a matrix-vector multiplication with $\mathbf{J}^H$. Numerically solving the linear system in step 4 of Alg. 5.1 also involves repeated multiplications with $\mathbf{J}$ and $\mathbf{J}^H$. In previous work [69], it was argued that storing $\mathbf{J}$ requires more computer memory than typically available and a matrix-free implementation was proposed that computes the matrix-vector products without having to store $\mathbf{J}$ into computer memory. However, this method requires recomputation of the entries of $\mathbf{J}$ for each matrix-vector multiplication. While memory-efficient, the method involves many redundant computations. In this work, we propose a *partially matrix-free* implementation instead, where at first only the partial derivatives of the transverse magnetization at echo times are computed and stored in computer memory. For 2D MR-STAT reconstructions at clinically relevant resolutions it should be possible to store these arrays on modern GPU cards, see Supplementary Material 5.6.

Having stored arrays with partial derivatives at echo times in memory, matrix-vector multiplications with $\mathbf{J}$ and $\mathbf{J}^H$ can be then performed in parallel by following slight modifications of the algorithms presented in [69]: instead of recomputing the partial derivatives at all readout times, the processes read in the partial derivatives at echo times and then use Eq. 5.2.6 (differentiated w.r.t. the tissue parameters at hand) to compute the values at the remaining readout times. Pseudo-code for these algorithms is presented in Alg. 5.2 and Alg. 5.3. In this pseudo-code, we assume for simplicity that only a single tissue parameter per voxel is considered and that no coilmaps are used in the reconstruction.

CUDA kernels are written in Julia to perform the partially matrix-free evaluation of Eq. 5.2.1 and multiplications with $\mathbf{J}$ (Alg. 5.2) and $\mathbf{J}^H$ (Alg. 5.3) respectively.

With tasks A, B and C being made GPU compatible (with the help of BlochSimulators.jl), we arrive at a GPU-compatible, partially matrix-free MR-STAT reconstruction algorithm.

Algorithm 5.2 Partially matrix-free, parallelized algorithm for computing matrix-vector products $\mathbf{Jx}$.

Assumptions: Let $\{nr, ns, nv\}$ be the total number of {readouts, samples per readout, voxels}.
Input:
- Matrix with partial derivatives at echo times $\tilde{\mathbf{J}} \in \mathbb{C}^{nr \times nv}$ computed with a `sequence` struct
- A `trajectory` struct describing the gradient trajectory
- Input vector $\mathbf{x} \in \mathbb{R}^{nv}$
Output:
- Matrix-vector product $\mathbf{Jx}$ stored in $output \in \mathbb{C}^{nr \times ns}$

Algorithm Kernel:
Let $1 \leq i \leq (ns \times nr)$ be the current process' global index.
$r, s = \text{divrem}(i, nr)$ # Readout and sample indices
for $v = 1, \ldots, nv$ **do**
  $J_{r,v} = \tilde{\mathbf{J}}(r, v)$ # Partial derivative at r-th readout
  # Compute partial derivative at s-th sample point
  $J_{s,v} = \partial\text{to\_sample\_point}(J_{r,v}, trajectory, s, \ldots)$
  # Multiply with input vector and accumulate result
  $output(i)$ += $J_{s,v} * \mathbf{x}(v)$
end for

5.3 Methods

In this section we will benchmark BlochSimulators.jl in terms of runtime performance against several existing and online available Bloch simulation packages. Benchmarks will be performed on both CPU and GPU hardware, using both the isochromat and EPG model. Secondly, we benchmark the runtime of the partially matrix-free MR-STAT reconstruction algorithm on GPU against previously presented MR-STAT reconstruction techniques. All CPU-based experiments are performed on an Intel(R) Xeon(R) W-2245. The GPU-based experiments are performed on an NVIDIA RTX A5000 card.

Algorithm 5.3 Pseudo-code to compute $\mathbf{J}^H\mathbf{y}$ in partially matrix-free, parallelized fashion.

Assumptions: Let $\{nr, ns, nv\}$ be the total number of {readouts, samples per readout, voxels}.
Input:
- Matrix with partial derivatives at echo times $\tilde{\mathbf{J}} \in \mathbb{C}^{nr \times nv}$ computed with a `sequence` struct
- A `trajectory` struct describing the gradient trajectory
- Input vector $\mathbf{y} \in \mathbb{C}^{nr \times ns}$
Output:
- Matrix-vector product $\mathbf{J}^H\mathbf{y}$ stored in $output \in \mathbb{C}^{nv}$
Algorithm Kernel:
Let $1 \leq v \leq nv$ be the current process' global index.
$tmp = 0$ # Initialize accumulator
for $r = 1, \ldots, nr$ **do**
  $J_{r,v} = \tilde{\mathbf{J}}(r, v)$ #Partial derivative at r-th readout
  for $s = 1, \ldots, ns$ **do**
    # Compute partial derivative at s-th sample point
    $J_{s,v} = \partial\text{to\_sample\_point}(J_{r,v}, trajectory, s, \ldots)$
    # Multiply with input vector and accumulate result
    $tmp \mathrel{+}= \text{conj}\,(J_{s,v}) * \mathbf{x}(v)$
  end for
end for
$output(v) = tmp$

5.3.1 Performance Benchmarks: BlochSimulators.jl

First, we compare BlochSimulators.jl against an isochromat-based Bloch simulator written in the C programming language written by Brain Hargreaves (URL). This Bloch simulator takes as input vectors with RF excitation and gradient waveforms and returns the x, y and z components of a spin isochro-

mat for each simulated time interval. One might use such an isochromat simulator to simulate gradient-balanced pulse sequences. In BlochSimulators.jl, a sequence simulator called "Generic3D" is implemented that takes similar input and produces the same output. For benchmarking purposes, simulations are performed to generate dictionaries for 1000 up to 10,000 different combinations of T_1 and T_2 values for a sequence with 1120 readouts. The simulations are performed on a single CPU using both single- and double precision floating point numbers. More details of the simulation setup are described in Supplementary Material 5.6 and the variable flip angle train used in the simulations is depicted in Supplementary Fig. 5.6. For the Generic3D simulator we generate dictionaries on the GPU as well using 10,000 up to 350,000 combinations of T_1 and T_2 values. The maximum number of 350,000 was chosen to be in the order of a typical MR Fingerprinting dictionary size.

Most pulse sequences are repetitive by nature. For example, in most conventional MR sequences, only the gradient encoding is different between different repetitions of the base sequence block. For transient-state sequences as used in MR Fingerprinting or MR-STAT the flip angle typically does change each TR but the nominal RF excitation waveform is fixed throughout the sequence. Whereas it would be possible to perform simulations for such a (transient-state) sequence with the simulator from Hargreaves or the Generic3D simulator from BlochSimulators.jl, runtime memory access could be reduced by incorporating knowledge of the repetitive nature of the sequence into the simulation code. For example, instead of reading in from memory at each timepoint of the simulation the actual RF field during that timepoint, the nominal RF waveform could be loaded once up front and scaled with the desired flip angle each TR instead. We take this approach with the "pSSFP" (pseudo steady-state free precession [6]) sequence that is provided as an example in BlochSimulators.jl. For this particular sequence implementation, the TR and TE remain fixed throughout the sequence but the flip angle is allowed to change each TR. The runtime for this sequence-specific simulator was benchmarked using a similar setup as used for the Hargreaves simulator and the Generic3D simulator.

For a third benchmark, we perform simulations for a FISP-type sequence with time-varying flip angles [85] based on the EPG model (additional sequence simulation details are provided in Supplementary Material 5.6). In BlochSimulators.jl, a sequence simulator "FISP" was implemented for this purpose. We compare this against the FISP-type simulator from SnapMRF [172], an MR Fingerprinting dictionary simulation framework that runs on GPU hardware and is written in native CUDA, available at `https://github.com/dongwang881107/snapMRF`.

It has been demonstrated that neural networks can be trained to act as a surrogate model for Bloch simulations with high accuracy and fast runtime performance [186]. For the fourth benchmark, we compare the FISP simulator from BlochSimulators.jl against against the recurrent neural network proposed ("RNN-EPG") in [104] that is trained to perform simulations for a FISP se-

quence. We note that EPG-RNN was trained on sequences for which the RF excitations always have the same phase. The calculation of the response of configuration states to an RF excitation then simplifies and the magnetization can also be described using real numbers instead of complex numbers. Using type-stability of BlochSimulators.jl's kernel function, together with Julia's multiple dispatch system, we could use the same FISP sequence implementation for both scenarios (RF excitations with constant and time-varying phases).

For the final benchmark, we measure runtimes for evaluating the forward model Eq. 5.2.1 using both BlochSimulators.jl and KomaMRI.jl. KomaMRI.jl is another Julia Bloch simulation package that allows for evaluation of the forward model Eq. 5.2.1 for arbitrary sequences and it supports GPU hardware as well. Both toolboxes were used to simulate the signal from a gradient-balanced sequence with golden angle radial readouts. We performed signal simulations for scenarios with 10,000 up to 350,000 voxels using single precision computations on GPU hardware. Additional sequence details are provided in Supplementary Material 5.6.

5.3.2 Performance Benchmarks: MR-STAT Reconstructions on GPU

The Julia GPU implementation of the partially matrix-free MR-STAT reconstruction algorithm is benchmarked in terms of runtime performance against the matrix-free distributed CPU implementation from [69], a sparse Hessian distributed CPU implementation from [66] and a neural-network and ADMM implementation from [106]. In all these previous reports, the same 2D in-vivo brain dataset obtained using a clinical 1.5T MR System (Ingenia, Philips Healthcare, Best, The Netherlands) was utilized. For this dataset, 1120 Cartesian readouts were acquired with gradient-balanced, variable-flip angle sequence with TE/TR = 3.8 ms / 7.6 ms and a total scan time of 8.5 s. The flip angle train and phase encoding order are depicted in Supplementary Fig. 5.6. The field-of-view was set to 224 mm × 224 mm × 5 mm and the reconstructed resolution was 1 mm × 1 mm × 5 mm. The partially matrix-free GPU implementation will be used on this same dataset, using ten outer iterations of Alg. 5.1 and a maximum of twenty inner iterations (Step 4 of Alg. 5.1) to reconstruct T_1, T_2 and proton density (complex) maps.

5.4 Results

In Fig. 5.2, the CPU runtimes for Hargreaves' simulator and the Generic3D and pSSFP simulators from BlochSimulators.jl that all use the isochromat model are displayed. It can be observed that Generic3D simulator outperforms Hargreaves' simulator by approximately 33% while both take the same input, return the same output, and at runtime essentially perform the same type of calcu-

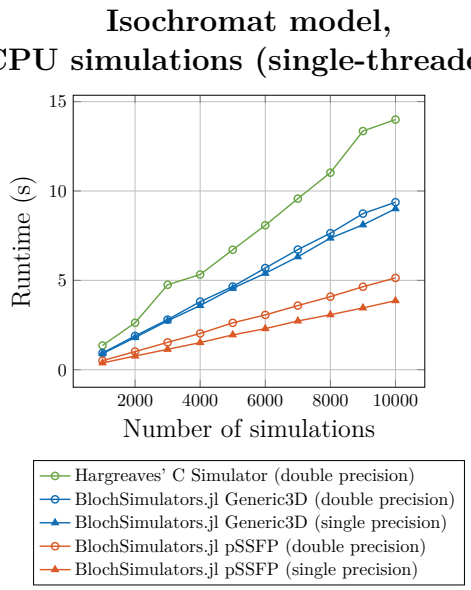

Figure 5.2: Runtime performance comparison of Hargreaves' C simulator (double precision only) and the Generic3D and pSSFP simulators from BlochSimulators.jl (single and double precision) on CPU.

lations. This demonstrates the effectivenes of Julia's Just-In-Time compiler to generate efficient machine code in this scenario. Secondly, we observe that the pSSFP simulator is approximately 50% faster than the Generic3D simulator. That is, making use of the repetitive nature of a pulse sequence to reduce runtime memory access (e.g. not having to load the same nominal RF waveform for each TR) significantly increases the compute throughput. In BlochSimulators.jl it is relatively easy to use its kernel functions to assemble a simulator corresponding to a specific pulse sequence. On a CPU, there is only a small performance gain when simulations are run using single precision compared to double precision. Note that the Hargreaves' simulator is hardcoded to double precision and therefore no single precision simulations were performed.

Whereas the Hargreaves' simulator is not GPU compatible, the design of BlochSimulators.jl allows the Generic3D and pSSFP sequence implementations to be directly executed on the GPU. In Fig. 5.3, the runtime performance results for the GPU calculations are shown (on a logarithmic scale). On the GPU, as expected, we observe more significant differences between single and double precision. The single precision Generic3D simulator is approximately six times faster compared to its double precision counterpart. For the pSSFP simulator, the difference is even larger with the single precision one being fifteen times faster. Furthermore we observe that the single precision pSSFP simulator

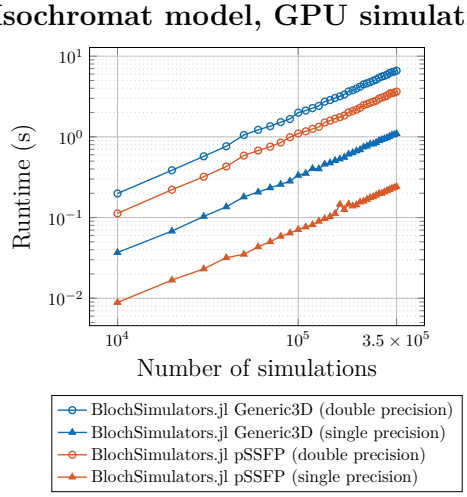

Figure 5.3: Runtime performance comparison of the Generic3D and pSSFP simulators from BlochSimulators.jl (single and double precision), this time on a GPU. The runtime performance benefit of using single precision floating point numbers is more pronounced on GPU hardware.

is approximately five times faster than the single precision Generic3D simulator. The former performs simulations in 350,000 voxels in 0.22 s. Extrapolating the CPU runtimes for the single precision pSSFP simulator to 350,000 simulations would result in a runtime of approximately 140 s. That is, the GPU results in a 600x speedup in this scenario.

The results from the EPG simulation benchmarks are displayed in Fig. 5.4. For the EPG simulations, we only performed simulations on the GPU at single precision and we compare the FISP simulator from BlochSimulators.jl, SnapMRF [172] and RNN-EPG [104]. Even though SnapMRF is written in native CUDA, the FISP implementation from our BlochSimulators.jl (annotated with "complex" in Fig 5.4 to indicate that the RF excitations can have varying phases) is more than 50 times faster. This difference may be explained by different design choices in the implementation such as using shared memory to store the configuration states, and further exemplifies the ability of Julia's compiler to generate efficient machine code. Furthermore, we observe that in the case where RF excitations are assumed to have constant phase (annoted with "real" in Fig. 5.4), the FISP implementation from BlochSimulators.jl is approximately three times as fast as RNN-EPG. However, RNN-EPG was trained to also return partial derivatives with respect to T_1 and T_2. If we were to use a finite difference approach to compute these partial derivatives with BlochSimulators.jl, the runtimes would be approximately the same.

In Fig. 5.5 the runtimes for evaluating the forward model Eq. 5.2.1 using

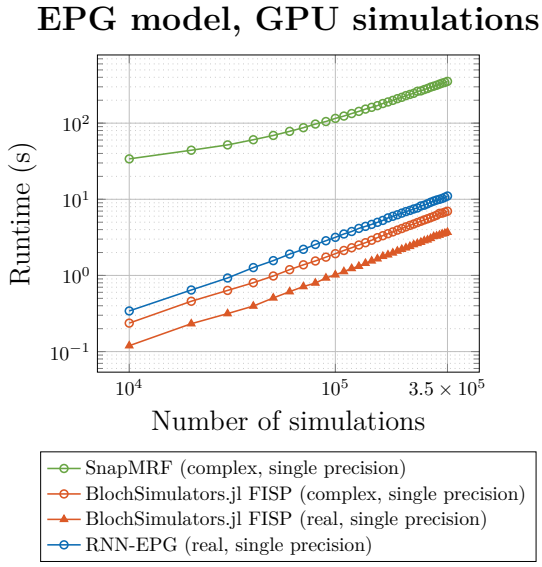

Figure 5.4: Runtime performance comparison of a variable flip-angle FISP-type sequence. The EPG model is used for signal simulations using an implementation from BlochSimulators.jl, the native CUDA implementation from SnapMRF [172] and RNN-EPG [104].

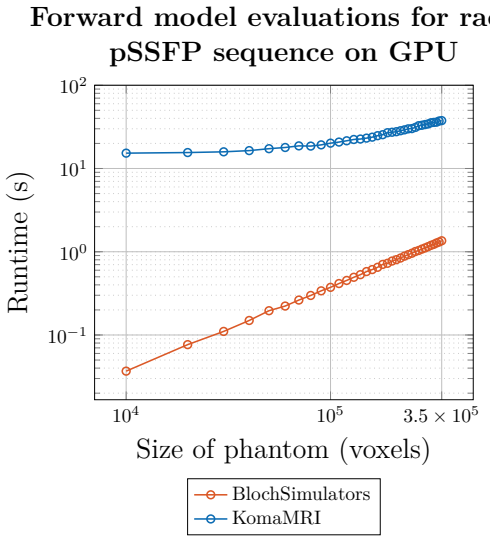

Figure 5.5: Runtimes for evaluating the forward model Eq. 5.2.1 using both KomaMRI.jl and BlochSimulators.jl on a gradient-balanced sequence with golden angle radial readouts.

both KomaMRI.jl [33] and BlochSimulators.jl are displayed. We observe that BlochSimulators.jl is faster, with the speedup factor depending on the number of voxels used in the simulations. For a typical 2D phantom size of $256^2 = 65,536$ voxels BlochSimulators.jl is approximately 75 times faster whereas for 350,000 voxels it is approximately 30 times faster.

In Table 5.1, the reconstruction times for the different MR-STAT reconstruction techniques on the same underlying dataset are reported. For the proposed partially matrix-free GPU case, the reconstruction time was 68 seconds, almost twice as fast as the previous state-of-the-art MR-STAT reconstruction technique from [106]. The T_1, T_2 and proton density maps reconstructed with the partially matrix-free MR-STAT algorithm on GPU are shown in Supplementary Fig. 5.7. Regarding the accuracy of the reconstructed maps, we note that the reconstruction procedure computes the same update steps (see Alg. 5.1) as the matrix-free [69] approach, and therefore it results in the same quantitative T_1, T_2 and proton density. In Table 5.2 we further outline the computationally demanding tasks of the reconstruction algorithm and what percentage of the reconstruction time is spent on these tasks. It can be observed that matrix-vector multiplications with $\mathbf{J}$ account for approximately 65% of the total reconstruction time, thus forming the main computational bottleneck.

Table 5.1: 2D in-vivo MR-STAT reconstruction times

Method	Hardware	Reconstruction Time
Matrix-free [69]	96 CPU cores	193 minutes
Sparse-Hessian [66]	96 CPU cores	16 minutes
ADMM + neural network [106]	8 CPU cores	2 minutes
Partially matrix-free [this work]	1 GPU	68 seconds

Table 5.2: Breakdown of 2D in-vivo MR-STAT reconstruction times using the partially matrix-free reconstruction algorithm on a GPU

Task	Time	% of total
Forward model evaluation	6.22 s	9.1 %
Partial derivatives at echo times	2.84 s	4.2 %
Matrix-vector products with $\mathbf{J}$	44.1 s	64.9 %
Matrix-vector products with $\mathbf{J}^H$	9.1 s	13.5 %

5.5 Discussion

Managing reconstruction times is one of the main challenges of the MR-STAT technique. Previously presented acceleration techniques focussed on CPU-parallelization [69], surrogate modelling [106] or algorithmic techniques that rely on the assumption of Cartesian-based gradient trajectories ([66, 106]). In this work we explored a different acceleration direction: GPU architectures. The Julia language was an essential tool in exploring this direction. Instead of having to work with a low-level, labor intensive programming language like CUDA, kernel functions for the computationally demanding parts of the reconstruction algorithm could be written directly in Julia, making use of many of its high-level language features, thus allowing for relatively quick prototyping.

The standalone software package BlochSimulators.jl that was developed as part of this work can be used to perform Bloch simulations using both the isochromat model and the EPG model with runtime performance that is comparable to or better than toolboxes developed in statically typed, compiled languages (C, CUDA). At the same time, BlochSimulators.jl allows for high flexibility: sequence simulators can be assembled that can take into account the repetitive nature that most pulse sequence exhibit for better runtime performance. Besides using BlochSimulators.jl for forward model evaluations in the context of MR-STAT, the package could be used, for example, for online generation of MR Fingerprinting dictionaries.

Although not demonstrated in the benchmarks, BlochSimulators.jl also supports multi-threaded as well as distributed CPU computations. Besides gener-

alizing to different hardware architectures, it also generalizes to different number types. This property is convenient because it allows one, for example, to use single precision floating point numbers on GPU without having to write additional code. The type-stable design of the kernel functions also allows one to use more exotic number types in the simulations. As an example, stochastic variables from the Measurements.jl package could be used to propagate uncertainties in certain parameters (flip angles for example) throughout the simulations automatically.

We observed that the EPG-based FISP sequence simulator from BlochSimulators.jl has similar runtime performance as the recurrent neural network RNN-EPG. At the same time, it does not require a training phase and there is no loss of accuracy that is inherent to trained networks. However, we still believe neural networks can play an important role in MR-STAT reconstructions. The RNN-EPG network was trained to be able to predict the signal for a wide range of varying flip angle trains. For scenario's where the fixed flip angle trains are fixed, a different network architecture (e.g. [186]) with better runtime performance may be more suitable and have better performance. BlochSimulators.jl may be a relevant tool in training such networks.

At the moment, the kernel functions in BlochSimulators.jl only recognize $T_1, T_2, B_1^+, \Delta B_0$ and spatial coordinates as inputs. In the future we plan to add support for additional contrast mechanisms such as diffusion using the Bloch-Torrey equations [160] and magnetization transfer [10].

Both our BlochSimulators.jl and the KomaMRI.jl [33] package use the `CUDA.jl` package to achieve GPU support. However, both packages follow a different implementation approach. In BlochSimulators.jl, signal simulations are performed with only two kernel calls: one kernel call for simulating the magnetization at echo times and a second kernel call for incorporating the gradient encoding and volume integration. On the other hand, in KomaMRI.jl, GPU support is provided by means of performing individual operations on CuArrays, resulting in many kernel launches per simulation. The overhead associated with launching many kernels may explain the difference in runtime performance observed in the benchmarks in this work. Another factor that may account for the runtime difference is that KomaMRI.jl uses a generic sequence format (similar to e.g. the Hargreaves simulator and Generic3D from BlochSimulators.jl) rather than a sequence specific format (such as pSSFP from BlochSimulators.jl).

By utilizing BlochSimulators.jl together with additional custom kernel functions, we could implement the partially matrix-free MR-STAT reconstruction algorithm on GPU hardware. We were able to achieve MR-STAT reconstruction times that are much faster compared to the MR-STAT reconstructions performed on a CPU cluster (68 seconds vs 193 and 16 minutes, respectively) and similar to the ADMM MR-STAT reconstruction technique [106]. The ADMM method, however, assumes a Cartesian-based gradient encoding scheme. The proposed (GPU accelerated) partially matrix-free method does not rely on this

assumption and could also work for non-Cartesian MR-STAT [67].

The currently reported reconstruction time of 68 seconds for the 2D in-vivo brain scan is specific for the acquisition and reconstruction setup that was used in this work. Increasing the number of samples during the acquisition, increasing the number of simulated isochromats per voxel (for the purpose of computing a more accurate slice profile compensation), or increasing the number of receive coils used in the reconstruction results in increased computation times. The same holds true also for the previous MR-STAT reconstruction techniques that were used in the benchmark.

Despite the speedup achieved in this work, reconstruction times are still too long for online reconstruction in clinical practice. Several options may be considered to further reduce reconstruction times. First, note that currently the computational bottleneck is formed by the matrix-vector products with the Jacobian $\mathbf{J}$ (see Table 5.2). In the current GPU implementation of this matrix-vector product, threads read from and write to global GPU memory only. Potential acceleration could be achieved by utilizing the GPU memory layout in a more optimized fashion, for example by performing warp and/or block reductions prior to writing to global memory.

Besides improving the implementation of the current algorithm, if Cartesian-based gradient encoding schemes are used, it may be possible to adapt the Sparse Hessian technique and/or the ADMM splitting technique to GPU hardware for further acceleration. Further, as mentioned above, despite the pSSFP implementation from BlochSimulators.jl showing comparable performance to RNN-EPG, other, potentially faster, neural network architectures could be explored for accelerating Bloch simulations and they could then be utilized to speed up MR-STAT reconstructions as well.

As outlined in Supplementary Material 5.6, for 2D MR-STAT scans, we can typically store the transverse magnetization at echo times and partial derivatives in GPU memory. For 3D MR-STAT acquisitions [105] it may not be possible to do so on a single GPU card due to memory constraints: there are many more voxels and potentially also many more echo times (depending on the actual pulse sequence). It may still be beneficial though to store the partial derivatives at *some* of the echo times (i.e. checkpointing [34]), and compute others only when needed during evaluation of the matrix-vector products with $\mathbf{J}$ and $\mathbf{J}^H$. We also note that in 3D cases, the Bloch simulations per individual voxel are computationally cheaper because in principle it is no longer necessary to perform multiple simulations per voxel to account for slice profile effects. Alternatively, if available, a cluster with multiple GPUs could potentially be used to store the transverse magnetization and partial derivatives at more/all echo times accross different devices. Subsequent evaluation of the forward model and matrix-vector multiplications with $\mathbf{J}$ and $\mathbf{J}^H$ would require communication between the different devices.

By combining the above mentioned acceleration strategies, we believe clinically acceptable reconstruction times for multi-slice or 3D MR-STAT scans are

within reach.

Acknowledgements

This research has been financed by the Dutch Technology Foundation under grant #17986. The authors are thankful for fruitful discussions with Christian Hundt, Oliver Kutter and Boris Bonev from NVIDIA.

5.6 Supplementary Materials

Supplementary Material S1: Partially matrix-free 2D MR-STAT memory estimate

Assume a 2D MR-STAT scan with a field-of-view consisting of 256×256 voxels and 5×256 readouts is performed. To store the magnetization at echo times, together with partial derivatives with respect to T_1 and T_2, using single precision (complex) floating point numbers ($2 * 4$ bytes per complex number) would require a total of

$$(256 * 256) * (5 * 256) * 3 * (2 * 4) * 1024^{-3} = 1.875\,\text{GB}.$$

Modern GPU cards typically have more than ten GB of global memory available, meaning that there is some flexiblity to increase the resolution, sequence length or number of parameters while still being able to store the magnetization at echo times (together with the partial derivatives) in global memory.

Supplementary Material S2: Isochromat simulation details

For the dictionary simulations based on the isochromat model, a gradient-balanced sequence consisting of 1120 excitations and readouts was used. A Gaussian RF waveform was used for the excitation and it was time-discretized into 25 piecewise constant parts. For each readout, only the magnetization at echo time was computed. The flip angle was different for each excitation. The flip angle pattern is depicted in Supplementary Fig. 5.6. The TE and TR were fixed at 8.0 ms and 4.0 ms. No slice profile correction was performed in the simulations. The T_1 and T_2 values used as input for the simulation are randomly selected from 1000 to 5000 ms and 10 to 1000 ms, respectively.

Supplementary Material S3: Extended Phase Graph simulation details

For the dictionary simulations based on the EPG model, a gradient-spoiled sequence consisting of 1120 excitations and readouts was used. For each readout,

only the magnetization at echo time was computed. The flip angle was different for each excitation. The flip angle pattern is depicted in Supplementary Fig. 5.6. The TE and TR were fixed at 8.0 ms and 4.0 ms. To account for slice profile effects, a partitioned EPG method [97] was used. Multiple independent simulations were performed per dictionary entry with each simulation using a different RF excitation scaling factor to mimic the effect of different positions along the slice direction. The individual simulations were then summed together. For the FISP implementation from BlochSimulators.jl, this slice profile correction method was part of the sequence implementation and 35 slice positions were used in the simulations. For SnapMRF, no slice profile correction mechanism is provided out-of-the-box. For a fair comparison, we therefore performed the simulations without corrections and multiplied the runtimes by 35.

Supplementary Material S4: Forward model simulation details

For the benchmark in which the forward model Eq. 5.2.1 is evaluated, a gradient-balanced sequence consisting of 500 TRs was used together with a golden angle radial readout trajectory. Each readout had 224 samples. The TR and TE remained fixed at 8.0 and 4.0, respectively. The nominal RF waveform was discretized using 25 timesteps. No slice profile correction mechanism was applied. To evaluate the forward model within BlochSimulators.jl we used the pSSFP sequence simulator together with the radial trajectory implementation (see `src/trajectories/radial.jl`). For KomaMRI.jl, we modified its MR Fingerprinting example script (see `https://github.com/cncastillo/KomaMRI.jl/tree/master/examples/3.koma_paper/mrf/MRFsimulation.jl`) to achieve the same sequence setup.

Supplementary Material S5: In-vivo parameter maps

The MR-STAT in-vivo acquisition used a smoothly varying, sinusoidal flip angle pattern with the peaks aligned with the sampling of the center of k-space. The flip angle train and the (Cartesian) phase encoding order is depicted in Supplementary Fig. 1. The quantitative T_1, T_2 and proton density maps reconstructed from this data using the partially matrix-free MR-STAT algorithm on GPU are shown in Supplementary Fig. 2.

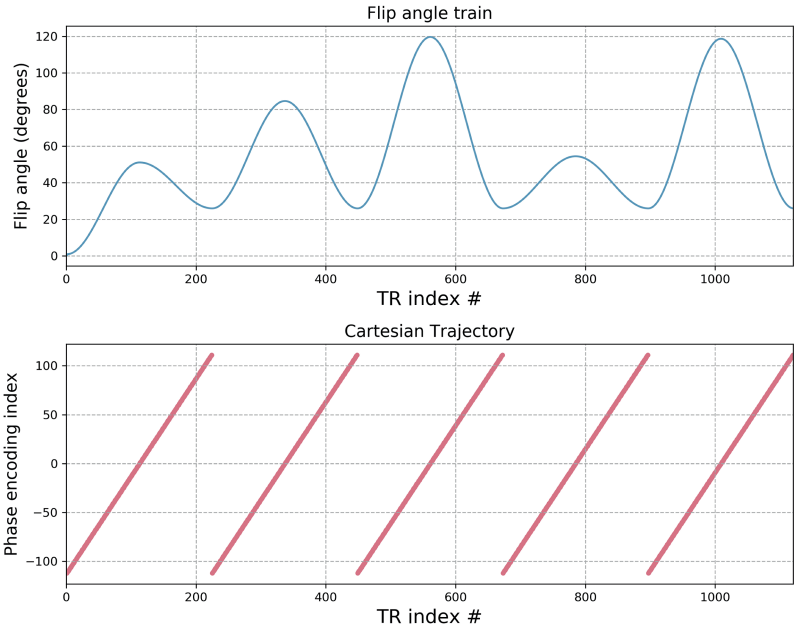

Figure 5.6: The time-varying flip angle train [top] and phase encoding indices for the Cartesian trajectory [bottom] used in the in-vivo MR-STAT data acquisition as well as the dictionary/signal simulations.

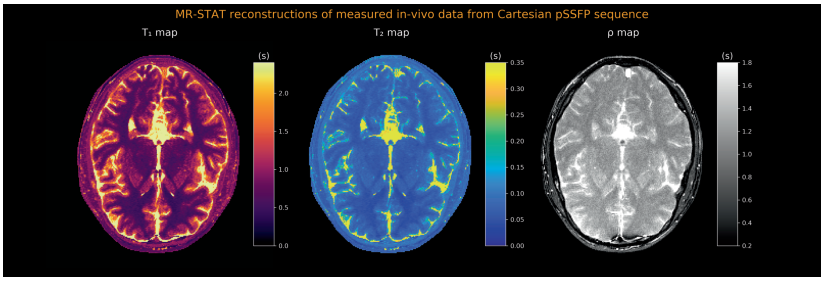

Figure 5.7: In-vivo T_1, T_2 and proton density maps reconstructed with the partially matrix-free MR-STAT algorithm on GPU in 68 seconds. For the data acquisition, a gradient-balanced sequence with variable flip angle train and Cartesian readouts (see Supplementary Fig. 5.7) was used.

6

Cartesian vs radial MR-STAT: An efficiency and robustness study

Oscar van der Heide
Alessandro Sbrizzi
Cornelis A. T. van den Berg

Magnetic Resonance Imaging, 2023, 99: 7-19.
https://doi.org/10.1016/j.mri.2023.01.017

Abstract

MR Spin TomogrAphy in Time-domain ("MR-STAT") is a quantitative MR technique in which multiple quantitative parameters are estimated from a single short scan by solving a large-scale nonlinear optimization problem. In this work we extended the MR-STAT framework to non-Cartesian gradient trajectories. Cartesian MR-STAT and radial MR-STAT were compared in terms of time-efficiency and robustness in simulations, gel phantom measurements and in-vivo measurements.

In simulations, we observed that both Cartesian and radial MR-STAT are highly robust against undersampling. Radial MR-STAT does have a lower spatial encoding power because the outer corners of k-space are never sampled. However, especially in T2, this is compensated by a higher dynamic encoding power that comes from sampling the k-space center with each readout. In gel phantom measurements, Cartesian MR-STAT was observed to be robust against overfitting whereas radial MR-STAT suffered from high-frequency artefacts in the parameter maps at later iterations. These artefacts are hypothesized to be related to hardware imperfections and were (partially) suppressed with image filters. The time-efficiencies were higher for Cartesian MR-STAT in all vials. In-vivo, the radial reconstruction again suffered from overfitting artefacts. The robustness of Cartesian MR-STAT over the entire range of experiments may make it preferable in a clinical setting, despite radial MR-STAT resulting in a higher T1 time-efficiency in white matter.

6.1 Introduction

Quantitative magnetic resonance imaging ("qMRI") techniques aim to provide estimates of MR-related tissue properties like T_1 and T_2. From these tissue property maps, image contrasts can be synthesized retrospectively using signal equations for different MR sequences [16, 22]. An additional advantage of qMRI over regular, qualitative MRI is that it removes scanner- and sequence induced variability in the images, which is beneficial for multi-center studies and computer aided diagnosis [45, 159]. Despite its advantages, clinical adoption of qMRI is currently limited because most conventional qMRI techniques require prohibitively long scan times and only produce one tissue parameter map at a time.

MR Fingerprinting ("MRF") is a recent multi-parametric qMRI technique which drastically reduces scan times compared to conventional qMRI methods [110]. In MRF, transient-state sequences with randomized components (e.g. flip angles, TR, TE) are used to generate "fingerprints" in each voxel. These fingerprints are measured with a highly undersampled acquisition and are subsequently matched to a pre-computed dictionary. Assuming spatio-temporal incoherence of the undersampling artefacts, the dictionary matching procedure can successfully select the correct tissue parameters [110]. The success of MRF approach has sparked research interest in highly accelerated, multi-parametric qMRI techniques [59, 173].

In order to achieve spatio-temporal incoherence in MRF, it is beneficial to use non-Cartesian acquisition strategies [150]. In the original MRF work and most subsequent MRF studies, variably density spirals are used in the acquisition [110, 85]. Radial acquisitions have also been used extensively [6, 43, 36]. Only few studies have been reported that use Cartesian (spin-warp) imaging in the context of MRF and the studies that do use such acquisitions report much longer scan times compared to the non-Cartesian MFR studies [32, 49, 93].

In a different direction from MRF, qMRI has seen advancements in the form of non-linear volumetric inversion methods that directly estimate tissue parameters from measured k-space (or "time-domain") data [20]. A benefit of using such a time-domain signal model is that the sampling of spatial frequencies that are required by the Nyquist criterion can be distributed among different contrasts [153]. That is, the *spatial encoding* and the *dynamic encoding* (i.e. T_1 and T_2 encoding) are intertwined. These techniques have initially been proposed in the context of pulse sequences for which analytical signal models are available [89, 175, 190].

Magnetic Resonance Spin Tomography in Time-domain ("MR-STAT") is a recently proposed multi-parametric qMRI framework [141] that combines the idea of non-linear, volumetric inversion together with the use of generic, transient-state sequences for performing multi-parametric qMRI. For transient-

state sequences, each contrast is usually highly undersampled (typically only one readout per contrast is acquired in MR-STAT), yet by applying the non-linear volumetric inversion it is still possible to reconstruct high-quality quantitative parameters maps from the data [69]. The MR-STAT reconstruction procedure relies on a forward model that relates multiple quantitative tissue parameters maps directly to the measured time-domain (k-space) signal. Evaluation of the forward model involves numerical integration of the Bloch equations as, in general, no analytic signal models are available for the type of time-varying flip angle sequences used in MR-STAT. The forward model gives rise to a large-scale non-linear inversion problem which is solved through iterative algorithms that require computing partial derivatives of the forward model with respect to all tissue parameters. In addition, instead of using the (non-uniform) Fast Fourier Transform ("FFT") for transforming back and forth between image space and spatial frequency space, the gradient encoding is taken into account explictly within the Bloch simulations. In other words, the spatial and dynamic encoding are simultaneously modeled. The MR-STAT reconstruction is computationally challenging and requires dedicated algorithms to manage computation times as well as computer memory requirements [69, 103, 66].

The feasibility of the MR-STAT approach has been demonstrated in previous work with acquisitions that rely on Cartesian gradient trajectories. High-resolution (1 mm in-plane) 2D quantitative T_1, T_2 and proton density maps were reconstructed from in-vivo brain data acquired in scan times comparable to what is reported in MRF studies using non-Cartesian acquisitions [69, 66]. The theory behind the MR-STAT concept, however, is not restricted to Cartesian acquisitions. To demonstrate the generic nature of MR-STAT, in the current work we have extended the framework to non-Cartesian acquisitions. This allows us to compare Cartesian and non-Cartesian trajectories within MR-STAT.

The main aim of this chapter will be to perform an empirical comparison of Cartesian and radial MR-STAT reconstructions performed on both simulated and measured data. The focus will lie on the efficiency and robustness of the two different encoding strategies. Efficiency is interpreted in this context as $T_{1,2}$-*to-noise* ratio per square root of scan time [42]. The efficiency is assumed to capture both the spatial and dynamic encoding capabilities of an acquisition. Robustness of the iterative reconstruction procedures against model inaccuracies (e.g. hardware imperfections) and reconstruction parameters (e.g. the number of iterations used) will also be studied since they can have a significant impact on efficiency as well.

For conventional, qualitative MR imaging, it is known that, for example, radial trajectories have a lower SNR efficiency due to the non-uniform sampling density [163]. That is, in terms of *spatial* encoding, radial is less efficient than Cartesian. However, we expect non-Cartesian MR-STAT acquisitions to potentially have higher *dynamic* encoding capabilities based on the following reasoning. The transient-state nature of the MR-STAT acquisition implies that

the underlying image contrast is constantly changing. Image contrast information is mostly contained in the central region of k-space. With non-Cartesian acquisitions like radial and spiral, the center of k-space can be sampled with each readout and thus each readout can provide information about the changing contrast. On the other hand, with a Cartesian acquisition, readouts that sample the outer parts of k-space are expected to provide relatively little information on the underlying contrast changes and thus provide less dynamic encoding power compared to the central k-space lines. Because in MR-STAT both spatial and dynamic encoding is required, it is a priori unclear which trajectory type will result in a higher efficiency.

The choice to consider radial acquisitions in this work - as opposed to other non-Cartesian trajectories - is motivated by the fact that most acquisition parameters like TR, TE, total number of readouts, number of samples per readout, readout bandwidth and total scan time can be kept similar for both. With, for example, spiral trajectories, one typically acquires fewer readouts with longer TRs and more samples are acquired per readout depending on the design of the spirals. A comparison of Cartesian against other non-Cartesian acquisitions thus requires many design choices that may influence the outcomes.

To study the efficiency and robustness question in the context of Cartesian and radial MR-STAT, we proceed as follows. Assuming that a pulse sequence is used that has sufficient T_1 and T_2 encoding power, and assuming that the MR-STAT reconstructions have succesfully converged, errors in the final parameters maps are caused by either 1) thermal noise on the data, 2) undersampling artefacts and/or 3) inaccuracies in the forward model (e.g. hardware imperfections, partial volume effects or unmodelled biophysical phenomena). Since the model-based MR-STAT reconstruction is expected to be highly robust against undersampling, we expect errors from undersampling to be minimal for both the Cartesian and radial cases. To verify whether this is the case, we will first perform numerical simulations without thermal noise or model imperfections such that undersampling artefacts are the only expected error source. Afterwards, we will study the efficiency and robustness in the presence of thermal noise by adding (complex) Gaussian noise to the simulated data. We then proceed by performing gel phantom and in-vivo measurements where also model imperfections are expected to be present.

6.2 Methods

6.2.1 MR-STAT

In MR-STAT, the forward model for the measurable time-domain signal s after spatial discretization is given by

$$s(t) = \sum_{j=1}^{N_v} m(\boldsymbol{\theta}_j, t) e^{-2\pi i \boldsymbol{k}(t)\cdot \boldsymbol{r}_j} \Delta_V. \tag{6.2.1}$$

Here N_v is the number of voxels within the field-of-view, Δ_V is the volume element for each voxel, $\boldsymbol{r}_j$ is the vector of spatial coordinates for the voxel associated with index j, $\boldsymbol{\theta}_j$ is the vector of MR-relevant biophysical tissue properties (e.g. $T_1, T_2, \rho, \ldots$) for the voxel associated with index j, $\boldsymbol{k}(t)$ is the k-space trajectory and m is the complex transverse magnetization whose time-varying behavior is modeled by the Bloch equations. For simplicity, we do not include receive coils in the forward model in here but it should be noted that within the actual MR-STAT reconstructions data from multiple coils is taken into account [69]. Let $t_1, \ldots, t_{N_t}$ denote the sampling times with N_t the total number of samples and define the vector of time-domain samples $\boldsymbol{s}$

$$\boldsymbol{s} := [s(t_1), \ldots, s(t_{N_t})] \in \mathbb{C}^{N_t}.$$

Note that $\boldsymbol{s}$ depends on the tissue parameters $\boldsymbol{\theta}_j$ for all voxels j. All tissue parameters are concatenated into a single vector $\boldsymbol{\alpha}$. The forward model (Eq. 6.2.1) then gives rise to a large-scale non-linear inverse problem

$$\boldsymbol{\alpha}^* = \operatorname{argmin}_{\boldsymbol{\alpha}} \frac{1}{2} \left\| \boldsymbol{d} - \boldsymbol{s}(\boldsymbol{\alpha}) \right\|_2^2. \tag{6.2.2}$$

This inverse problem is numerically solved using a Gauss-Newton method that requires the computation of partial derivatives of the objective function at each iteration [69].

Instead of using the (non-uniform) FFT to evaluate an approximation to the forward model (Eq. 6.2.1) as is common in other frameworks (e.g. MRF), in MR-STAT the forward model (Eq. 6.2.1) is evaluated directly. Both Cartesian and non-Cartesian gradient trajectories $\boldsymbol{k}(t)$ can be inserted into the forward model. In the case of non-Cartesian trajectories, no gridding or density compensation are required as opposed to typical non-uniform FFT implementations. Also note that, in principle, spin dynamics during readouts (e.g. $T_2^{(*)}-$decay or off-resonance induced rotations) can be modeled. These effects are expected to be neglible for acquisitions with short readouts and therefore exploring potential benefits of including these dynamics in the forward model is outside the scope of the current work.

6.2.2 Acquisition

6.2.2.1 Numerical brain phantom simulations

To compare the theoretical efficiency of both Cartesian and radial acquisitions in the context of MR-STAT, we first perform a simulation study for which ground truth parameter values are available. A 2D numerical brain phantom [9] was generated consisting of several compartments with different combinations of T_1, T_2 and proton density values. The field-of-view of the phantom was set to 224 mm x 224 mm with an in-plane resolution of 1 mm x 1 mm, resulting in a matrix of 224×224 voxels. In a conventional (qualitative) MR setting, a minimum of 224 phase encoding lines would need to be acquired for the Cartesian case to satisfy the Nyquist criterion. In a radial setting, a minimum number of $\pi/2 \times 224$ would need to be acquired [29] to satisfy the Nyquist criterion.

For the MR-STAT acquisitions, a gradient-spoiled sequence consisting of 1792 ($= 8 \times 224$) TR's was employed with a varying flip angle train such that the flip angle at the n-th TR was given by $35 \times 1 - \cos(2n/280)$ (resulting in a sinusoidal pattern of flip angles between 0 and 70 degrees, see Fig. 6.1a). For the Cartesian case, linear ordering of the 224 different phase encoding lines was chosen and the pattern is repeated eight times (Fig. 6.1b middle row). For the radial trajectory the first readout was identical to the Cartesian $k_y = 0$ line, and subsequent readout lines were obtained by rotating the previous readout line with the golden angle, resulting in 1792 different radial angles with a very dense cumulative sampling of k-space. The TR and TE remained fixed throughout the sequence with values of 8.8 ms and 4 ms, respectively. For each of the 1792 readouts a total number of 448 samples per readout (corresponding to a readout oversampling factor of two) were simulated. The total simulated acquisition time was 15.8 for both the Cartesian and the radial case. The k-space coverage for both acquisitions is depicted in Fig. 6.1c and Fig. 6.1d, respectively.

The magnetization response in each voxel was simulated using the Extended Phase Graph ("EPG") method [177]. In these simulations, a Gaussian-shaped RF excitation waveform was used and the corresponding slice profile was modelled by partitioning each voxel into 35 compartments that experience different effective flip angles (i.e. the partitioned EPG method [23, 26]). The scaling factors to determine the effective flip angles were obtained from simulating the magnetization response at different z-locations for all the different RF-excitation waveforms corresponding to different flip angles. Using the MR-STAT forward model (Eq. 6.2.1), time-domain data was simulated for both Cartesian and radial readout trajectories. The simulation code was written in the Julia programming language [4] and the simulations were performed on an NVIDIA GeForce RTX A5000 graphics card.

To study the efficiency and robustness in the presence of thermal noise, we

corrupted the noiseless datasets with random noise sampled from a complex Gaussian distribution. The noise was generated such that the signal-to-noise ratio in decibels (SNR_{DB}) was 15.36 DB for the Cartesian case. The SNR_{DB} is computed as

$$\mathrm{SNR}_{DB} = 10 \log_{10} \left(\frac{RMS(\mathrm{signal})^2}{RMS(\mathrm{noise})^2} \right),$$

where $\mathrm{RMS}(x)$ is the root mean square of a vector x. The same noise vector was added to the radial dataset.

6.2.2.2 Gel phantom measurements

Twelve vials containing gadolinium-doped gel with varying T_1 and T_2 values (TO5, Eurospin II test system, Scotland) were scanned using a 3T Philips Ingenia Elition X MR System (DDAS spectrometer, software release 5.6) with the vendors 16-channel receive headcoil. Data from a single 2D slice was acquired using Cartesian and radial sequences similar to the ones used in the simulation study. In both cases the in-plane resolution was 1 mm x 1 mm, the field-of-view was 224 mm x 224 mm and slice thickness was 5 mm. A total number of 1792 readouts were acquired with 448 samples per readout (factor two readout oversampling) with a readout bandwith of 85.6 kHz. For both acquisitions, the TE and TR were set to their shortest possible values. For the Cartesian acquisition the TE was 3.8 ms and the TR was 7.5 ms, resulting in a scan time of 13.4 s. For the radial acquisition, the TE and TR were slightly different at 4 ms and 8.5 ms, respectively, resulting in a total scan time of 15.2 s.

In addition to the MR-STAT scans, inversion-recovery single spin-echo T_1 measurements as well as single echo spin-echo T_2 measurements were performed to obtain reference parameter values for the vials. For the T_1 mapping measurement, inversion times of $[50, 100, 150, 350, 550, 850, 1250]$ ms were chosen and for the T_2 mapping measurement echo times of $[8, 28, 48, 88, 138, 188]$ ms were chosen. The T_1 and T_2 values were obtained pixel-wise by fitting (non-linear) exponential regrowth (T_1) and exponential decay (T_2) curves to the measurements using the variable projection method [58].

6.2.2.3 In-vivo measurements

In-vivo measurements were performed on a healthy volunteer after having obtained written informed consent. The same scanner hardware acquisition settings were used as for the gel phantom measurements, except that for the Cartesian acquisition the TR, TE and total scan time were 7.7 ms, 4.0 ms and 13.8 s, respectively. For the radial acquisition the TR, TE and total scan time were 8.6 ms, 4.0 ms and 15.4 s, respectively.

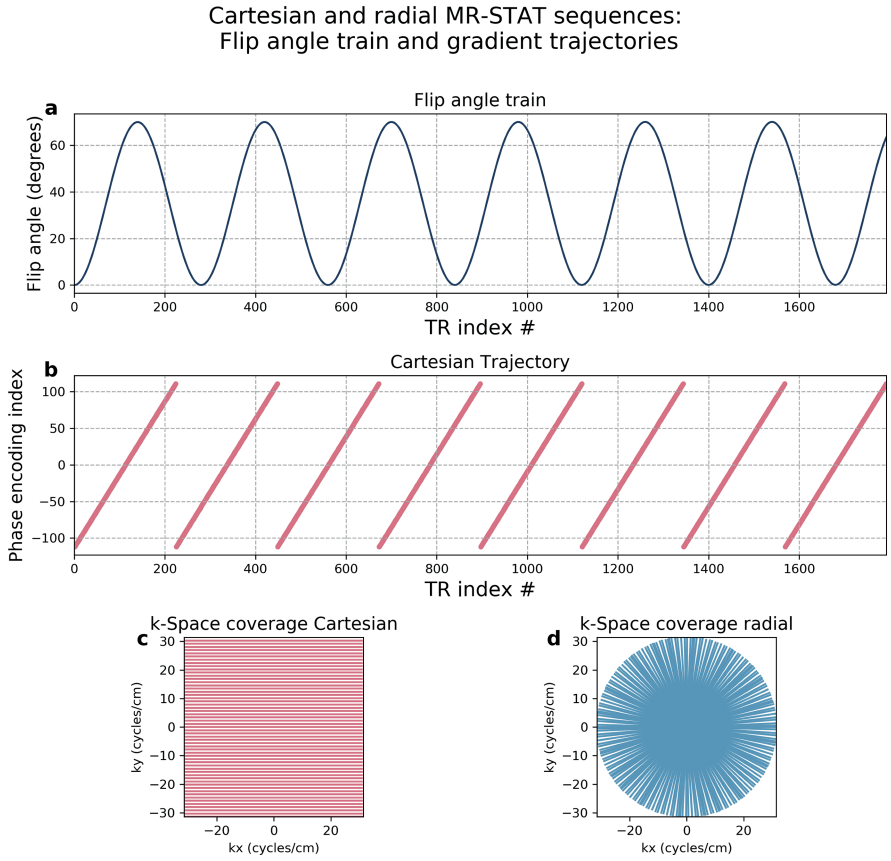

Figure 6.1: **[a]** Flip angle train used for both the Cartesian and radial acquisitions. **[b]** The phase encoding order for the Cartesian acquisitions. **[c** and **d]** K-space coverage for the Cartesian and radial acquisitions, respectively. Note that the Cartesian sampling scheme is repeated eight times.

6.2.3 MR-STAT Reconstructions

To reconstruct quantitative parameter maps from the data, the matrix-free Gauss-Newton MR-STAT method proposed in van der Heide et. al. [69] was used to reconstruct quantitative parameter maps from the data. Within this method, the magnetization response in each voxel was simulated using the EPG method with the same slice profile correction technique and computer hardware as described in Subsection 6.2.2.1. Partial derivatives of the magnetization response were computed using finite differences. The initial parameter estimates for T_1 and T_2 were set to 1.0 s and 0.1 s, respectively. The proton density was initialized by inserting the initial T_1 and T_2 values into Eq. 6.2.2, which then reduces to a linear problem that can be solved using the LSQR algorithm [126]. Given the initial proton density, a spatial mask was generated by selecting the voxels for which the magnitude of the initial proton density was higher than 10 % of the maximum magnitude. The maximum number of outer iterations of the iterative reconstruction algorithm was set to twenty and the number of Conjugate Gradient iterations within each outer iteration was set to twenty as well.

Raw data from the scanner was exported using ReconFrame (Gyrotools, Switzerland). In all reconstructions on measured data, the singular value decomposition was applied prior to the reconstruction to generate virtual coil data [31]. The number of virtual coils was chosen such that approximately 85% of the total energy was captured (i.e. the sum of the singular values was approximately 0.85). Coil sensitivity maps were estimated from the measured data using the ESPIRiT algorithm [164]. To correct for B_1^+ inhomogeneities that may be present at 3T, we measured a B_1^+ map separately using the dual angle method [151]. This B_1^+ map was then used in the reconstruction model to scale the effective RF induced flip angles in each voxel.

For the radial acquisitions, the outer corners of k-space are never sampled, as can be seen from Fig. 6.1d. In conventional non-uniform FFT based MR reconstructions, this is not necessarily problematic since the gridding procedure essentially zero-fills these non-sampled k-space regions. However, for iterative reconstruction techniques like MR-STAT, these non-sampled regions can cause strong noise amplification [134]. We therefore filter the reconstructed quantitative parameter maps by applying a circular symmetric window function in the Fourier domain as after the MR-STAT reconstruction. For the gel phantom data, the arctan-based filter proposed by Pruessmann et al. (2001) [134] was used with a cutoff of value set to $k_{x,\max}$ and $\beta = 100$. For the in-vivo data, we used the Hann filter defined as

$$\boldsymbol{k} \rightarrow \cos\left(\frac{\sqrt{k_x^2 + k_y^2}}{|k_{\max}|}\right)^2$$

where $\boldsymbol{k}$ is the k-space coordinate. The same filters will be applied to parameter

maps obtained from Cartesian and radial data, respectively.

6.2.4 Efficiency assessment

The time-efficiency of the acquisitions will be computed as

$$\frac{T_nNR}{\sqrt{T_{scan}}}, \quad n = 1, 2,$$

where T_nNR is the T_n-*to-noise* ratio and T is the scan time [42]. For the numerical brain phantom, we compute the T_nNR ($n = 1, 2$) for each tissue type separately by dividing the mean T_n value by the standard deviation of the T_n value for that tissue type. For the gel phantom measurements, values for T_nNR are obtained in a similar fashion by computing the mean values and standard deviations in manually drawn regions-of-interest in each vial. For the in-vivo measurements, gray- and white matter segmentation is performed using the T_1 maps as input to FSL Fast [188] and the T_nNR values are computed in these regions.

Because the MR-STAT reconstructions follow an iterative procedure, one question that needs to be addressed is which iterations will be used for analyzing the efficiency. For this purpose, whenever ground truth parameter maps are available, the root mean squared relative errors ("RMSRE") is utilized. If $\boldsymbol{\alpha}^{gt}$ denotes the vector of ground truth parameter values, the RMRSE value for the current estimates of the tissue parameter $\boldsymbol{\alpha}$ is computed as

$$RMRSE(\boldsymbol{\alpha}) = \left(\frac{1}{N_p} \sum_{j=1}^{N_p} \left(\frac{\alpha_j - \alpha_j^{gt}}{\alpha_j} \right) \right)$$

The iterations with minimum RMSRE value are assumed to strike a balance between bias and precision. To assess the robustness of the reconstructions against overfitting, the efficiencies will also be computed at the final iterations in each case. Note that in the in-vivo case, no ground truth parameter values are available so the RMSRE cannot be computed and therefore only the final iterations are considered.

6.3 Results

6.3.1 Numerical brain phantom simulations

6.3.1.1 Noiseless dataset

In Fig. 6.2 the T_1 and T_2 maps reconstructed from the noiseless numerical brainweb phantom datasets are displayed, as well as the relative error maps.

The RMRSE values for each iteration are shown in Fig. 6.3. In the Cartesian case, the relative errors and RMSRE value at iteration twenty are neglibly small. The parameter maps are reconstructed without apparent aliasing artefacts despite each contrast (i.e. each TR) being sampled with only one Cartesian readout line. The Cartesian acquisition thus has sufficient spatial and dynamic encoding power.

In the radial case we do observe non-neglible errors in the parameter maps and the RMSRE value is orders of magnitude higher compared to the Cartesian case. We argue that the higher errors are not resolved by acquiring more spokes per contrast or running more iterations in the reconstruction procedure. Instead, this reduced *spatial* encoding power of the radial acquisition is the result of not sampling the outer corners of k-space (see Fig. 6.1d). This makes it impossible for the reconstruction procedure to properly resolve the finer structures in the parameter maps.

In Supplementary Material S1 we adjusted the radial k-space trajectory to also sample the outer corners to improve the spatial encoding. In that situation we indeed observe that the parameter maps are properly reconstructed with neglibly low relative errors and RMSRE values. The errors - although neglible - are still higher compared to the Cartesian case, but this is expected because, unlike in the Cartesian case, the image grid does not exactly match with the sampling grid.

6.3.1.2 Noisy dataset

In Fig. 6.4a the cost ($\boldsymbol{\alpha} \mapsto \frac{1}{2} \left\| \boldsymbol{d} - \boldsymbol{s}(\boldsymbol{\alpha}) \right\|_2^2$) is plotted for each iteration of the MR-STAT reconstructions on the noisy numerical brain phantom datasets. Since the model used to simulate the data is equal to the model used in the reconstruction, upon convergence, the cost function is expected to be similar to $\frac{1}{2} \left\| \boldsymbol{\eta} \right\|_2^2$, where $\boldsymbol{\eta}$ denotes the vector of complex noise that was added to both datasets. The noise level $\frac{1}{2} \left\| \boldsymbol{\eta} \right\|_2^2$ is plotted as a horizontal line and it can be seen that for both reconstructions, the noise level is indeed reached. In Fig. 6.4b the RMRSE values per iteration of the MR-STAT reconstruction algorithm are shown. No filtering was applied to these reconstructions. Iterations three and eight results in the lowest RMSRE values for the Cartesian and radial cases, respectively. At these *optimal* iterations, the RMSRE value for radial is lower than for Cartesian. After the optimal iterations, the RMSRE values go up slightly whereas the cost functions decrease, suggesting that the reconstructions may be susceptible to overfitting artefacts.

The reconstructed parameter maps and relative error maps at the optimal iterations are shown in Fig. 6.5 together with the parameter maps at the final iterations. Like in the noiseless case, no apparent undersampling artefacts are observed. Although the increase of the RMSRE value suggests that the reconstructions may be susceptible to overfitting, it is visually difficult to observe for both the Cartesian and the radial case. That is, the reconstructions appear

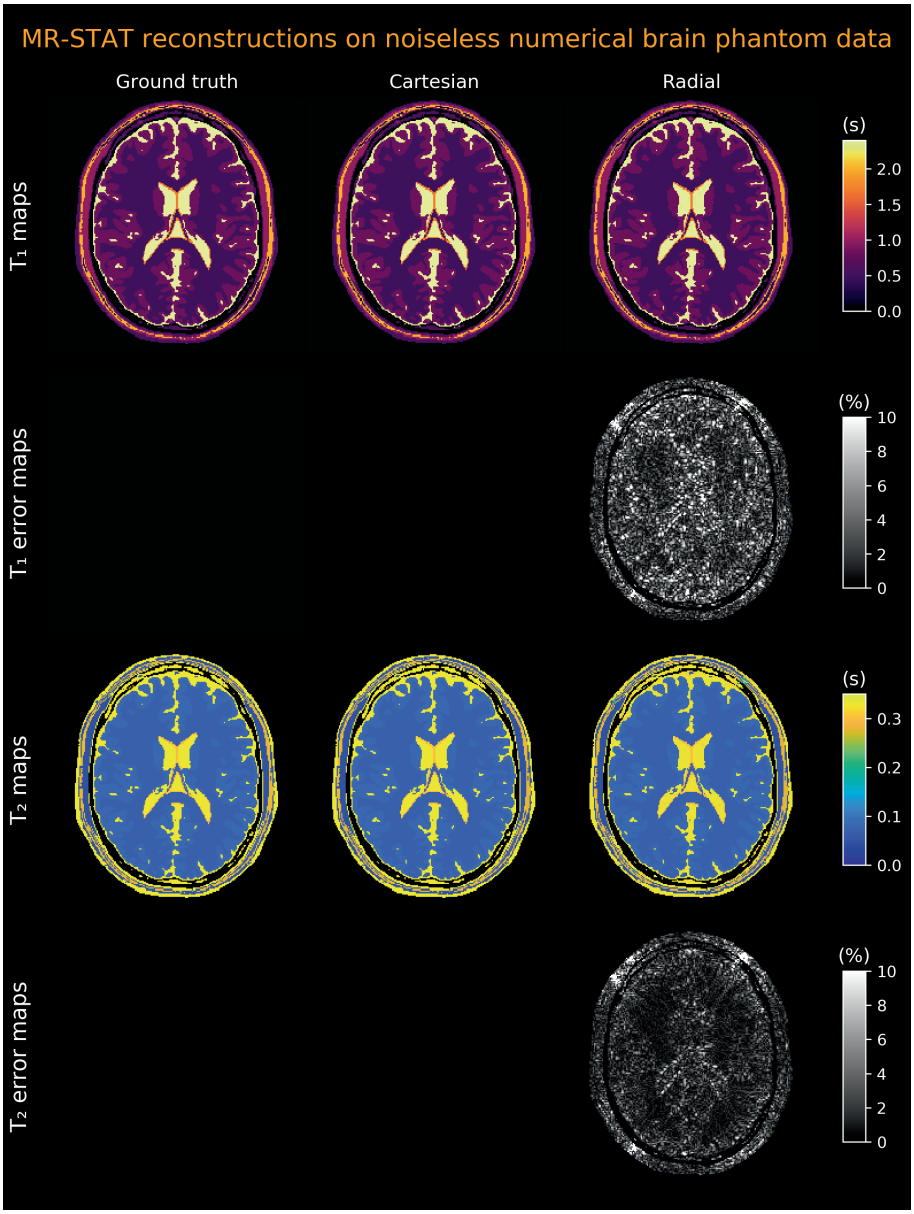

Figure 6.2: The T_1 and T_2 parameter maps reconstructed from noiseless simulated brain data. The first column shows the ground truth T_1 and T_2 maps. The second column shows the parameter maps and relative error maps for the Cartesian case after twenty iterations. The third column shows the parameter maps and relative error maps at iteration twenty for the radial case.

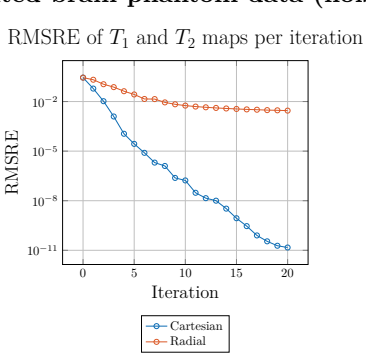

Figure 6.3: RMRSE values (for T_1 and T_2 combined) per outer iteration. For the radial case the RMRSE values are higher because the outer corners of k-space are never sampled and therefore finer structures in the parameter maps cannot be properly resolved (i.e. reduced spatial encoding)

to be robust against simulated thermal noise.

The mean T_1 and T_2 values as well as standard deviations per compartment at the optimal iterations as well as the final iterations are shown in Fig. 6.6a-b. It can be observed that the mean tissue parameter values per compartment are in good agreement with the ground truth values (well within one standard deviation). The efficiencies per compartment are displayed in Fig. 6.6c-d. We see that at the final iterations, the radial acquisition results in higher efficiencies (approximately 25%) for T_2 in six out of the seven distinct tissues despite the radial reconstruction suffering from errors caused by not sampling the outer corners of k-space. For T_1, the Cartesian acquisition is more efficient for four out of the seven tissues, but in gray- and white matter (arguably the most relevent tissues for most brain imaging applications) the radial acquisition is more efficient.

6.3.2 Gel phantom measurements

To assess the efficiency and robustness of the MR-STAT reconstructions on the Cartesian and radial data obtained from gel phantoms, we follow a similar procedure as in the numerical brain phantom study. In Fig. 6.7a the RMRSE values for each iteration are shown. The RMSRE values for the Cartesian case follow a similar curve as in the noisy numerical brain phantom case and iteration two is observed to result in the lowest RMSRE value. For the radial case, the situation is different compared to the noisy numerical brain phantom case. Without filtering the parameter maps in a post-processing step, the lowest RMSRE occurs at iteration three but at further iterations the RMSRE

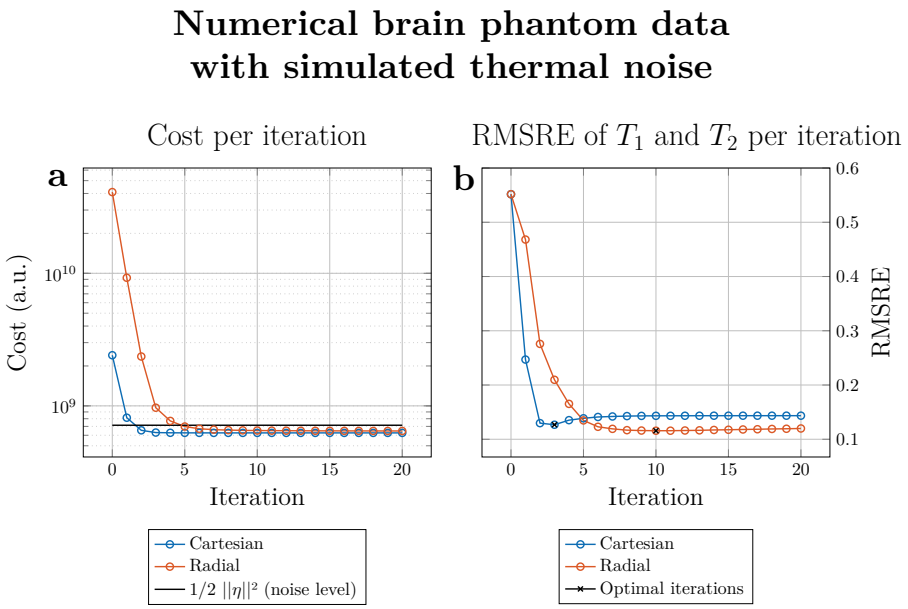

Figure 6.4: Reconstruction results for the noisy numerical brain phantom dataset. **a**: Cost as function of the number of outer iterations. **b**: RMRSE values (for T_1 and T_2 combined) per outer iteration.

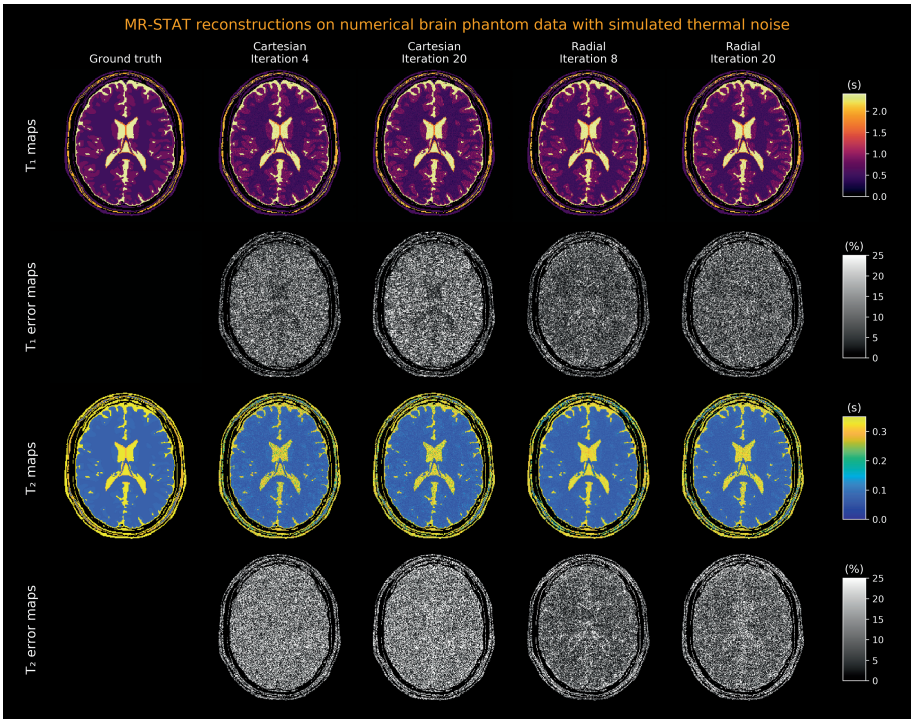

Figure 6.5: The T_1 and T_2 parameter maps reconstructed from noisy simulated brain data. The first column shows the ground truth T_1 and T_2 maps. The second and third columns show the parameter maps and relative error maps for the Cartesian case for the iteration with the lowest RMSRE value (iteration three) and the final iteration (iteration twenty), respectively. The fourth and fifth columns show the parameter maps and relative error maps for the radial case.

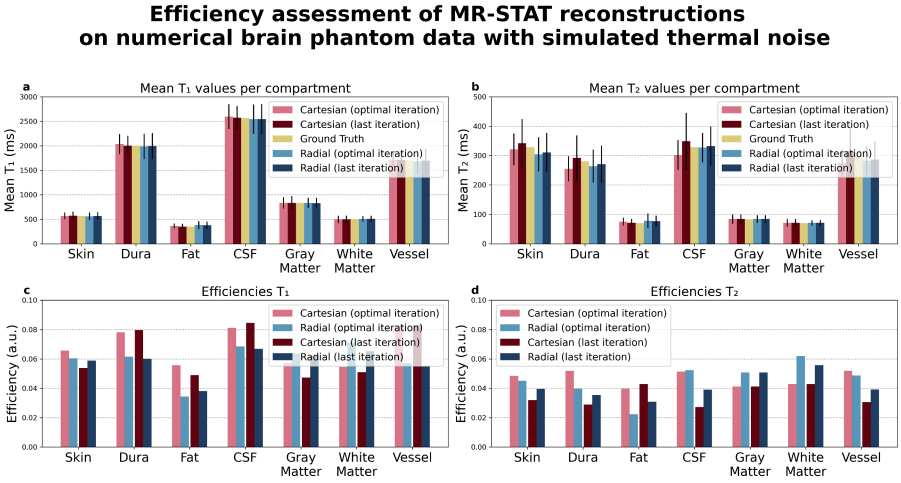

Figure 6.6: **a-b**: Mean T_1 and T_2 values and standard deviations per tissue type of the numerical brain phantom reconstructions. Ground truth parameter values are displayed using the yellow bars. **c-d**: Efficiencies for T_1 and T_2 per tissue type. It can be observed that overall the radial case results in higher efficiencies.

values rapidly increases, indicating that the radial case is highly susceptible to overfitting.

In Fig. 6.8 the T_1 and T_2 parameter maps for the MR-STAT reconstructions of the gel phantom data are shown. In columns one and two the maps from the Cartesian acquisition are shown at iterations two and twenty, respectively. Some mild overfitting artefacts (e.g. high frequency noise) can be observed in the maps corresponding to iteration twenty when compared to iteration two. On the other hand, when comparing iterations three and twenty for the radial acquisition (columns four and five, respectively), the appearance of strong overfitting artefacts can indeed be observed. Therefore, the application of a window function to the parameter maps in Fourier domain is deemed necessary. In this case, an arctan-based filter was chosen [134]. The resulting RMSRE curves are displayed in Fig. 6.7b and in Fig. 6.8, columns three and six, the filtered parameter maps are displayed as well. We observe that the RMSRE values for the radial case stabilize to values comparable to the RMSRE values and the high frequency noise in the parameter maps is greatly reduced. For the Cartesian case, this particular choice of filter has no significant impact on the RMSRE values or the parameter maps. We also visually observe that for both the Cartesian and radial cases, at the optimal iterations, the T_2 of the vial with the highest T_2 value has not fully converged yet.

The mean T_1 and T_2 values as well as standard deviations per vial at the iterations with lowest RMSRE values as well as the filtered versions of the

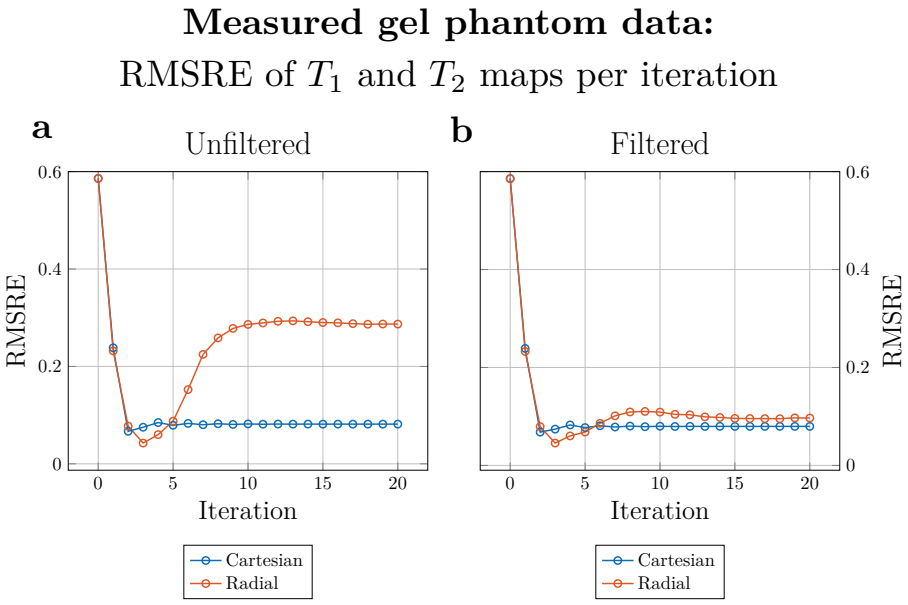

Figure 6.7: Reconstruction results for the gel phantom measurements. **a**: RMRSE values (for T_1 and T_2 combined) per outer iteration. **b**: RMRSE values after applying an arctan-based filter to the parameter maps in a post-processing step.

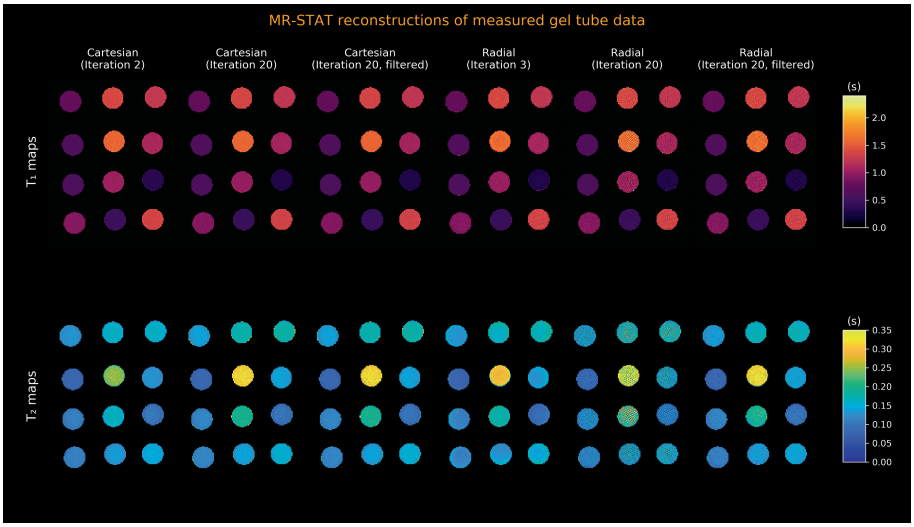

Figure 6.8: The T_1 and T_2 parameter maps reconstructed from measured gel phantom data. The first two columns show the Cartesian reconstrutions at the optimal iteration (two) and the final iteration (twenty). In columns four and five the corresponding radial reconstructions are shown. It can be observed that - unlike in the simulation study - the radial reconstruction is highly susceptible to overfitting artefacts and applying a filter is deemed necessary. The filtered maps are shown in columns three (Cartesian) and six (radial). For the Cartesian case, the effect of the filter is difficult to observe visually. For the radial case, it significantly reduces the high-frequency noise in the parameter maps.

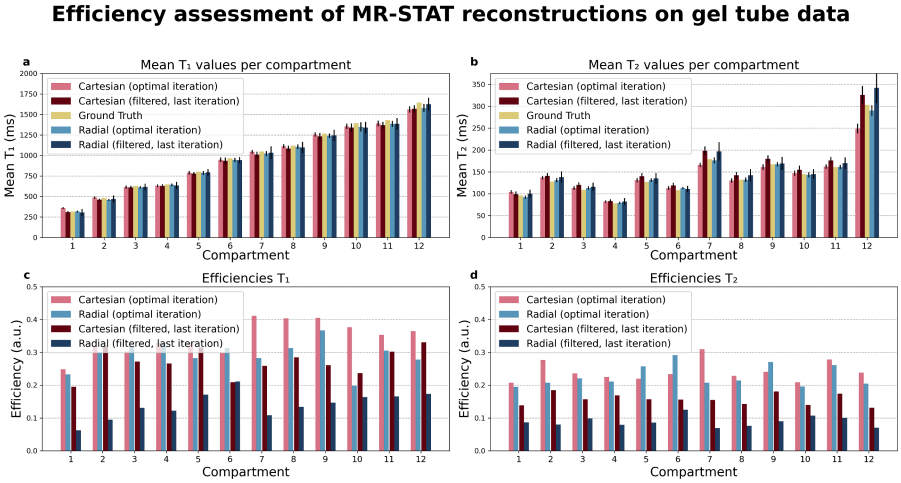

Figure 6.9: Reconstruction results for the gel phantom measurements. **a-b:** Mean T_1 and T_2 values as well as standard deviations per vial of the gel phantom for the outer iterations with the lowest RMRSE values (two for Cartesian, three for radial) and the final iterations. **c-d:** Efficiencies per vial are shown. It can be observed that overall the radial case results in higher efficiencies at the optimal iterations while the Cartesian case results in higher efficiencies at the latest iteration.

final iterations are shown in Fig. 6.9. The T_1 values obtained from both the Cartesian and the radial reconstructions are in excellent agreement with the reference values. For T_2 we observe good agreement but the deviations from the reference values are larger compared to T_1. This may be explained by potential errors in the B_1^+ maps used in the reconstructions that mostly have an impact on estimated T_2 values [123]. As observed before from Fig. 6.8, the vials with the highest T_2 values have not converged yet at the optimal iterations. The efficiencies per vial are displayed in 6.9c-d. At the optimal iterations, the radial case results in higher efficiencies in all vials(except for T_1 in vial number ten). However, at the (filtered) final iterations, a significant drop in efficiencies is observed for the radial case and the Cartesian case results in higher efficiencies in all vials (except for T_1 in vial number six).

6.3.3 In-vivo measurements

With no ground truth measurements available, the RMSRE values cannot be computed and an optimal iteration cannot be selected. Therefore we only consider the final iterations (i.e. iteration twenty) for both the Cartesian and radial cases. In Fig. 6.10, the T_1 and T_2 maps at these final iterations are shown. Like in the case for the gel phantoms, the radial reconstruction suffers

		Cartesian		Radial		Literature
		Mean ± Std (ms)	**Efficiency (a.u.)**	**Mean ± Std (ms)**	**Efficiency (a.u.)**	**Mean (ms)**
White Matter	**T1**	898.4 ± 78.0	0.098	881.3 ± 49.3	0.144	954
	T2	32.2 ± 5.0	0.055	31.4 ± 3.9	0.065	38.7
Gray Matter	**T1**	1471.4 ± 137.7	0.091	1398.1 ± 128.8	0.088	1372
	T2	57.7 ± 14.7	0.033	45.3 ± 12.2	0.03	52.7

Table 6.1: Mean values, standard deviations and efficiencies for in-vivo T_1 and T_2 in gray- and white matter regions. Literature values [95] are reported in the last column.

from high frequency noise and windowing is deemed necessary. For this dataset, a Hann filter was utilized and the filtered maps are shown in Fig. 6.10 as well. Although the filtering greatly improves the quality of the radial parameter maps, some high frequency noise is still present after filtering (mostly in or around cerebrospinal fluid ("CSF") regions) whereas no high-frequency artefacts are observed to be present in the filtered Cartesian maps.

In the Cartesian reconstruction, the CSF appears to suffer from flow-induced artefacts that are known to be present in gradient-spoiled sequences as used in this work [100]. For the radial case these artefacts are much less severe, likely because the radial sampling pattern effectively results in the flow effects being averaged out over the whole duration of the acquisition.

Mean T_1 and T_2 values in gray- and white matter for the filtered maps as well as standard deviations and efficiencies are displayed in Table 6.1. Recent literature values [95] are also reported. For T_1, the Cartesian and radial reconstructions result in similar mean values in gray- and white matter and these values are in line with literature values. For T_2 the mean values for white matter between Cartesian and radial agree but in gray matter there is a larger discrepancy between the Cartesian and radial cases. Segmenting the gray matter is more challenging and segmentation errors may partly explain the larger discrepancy and larger standard deviations.

In terms of efficiency, the radial reconstruction results in higher efficiencies in white matter (0.098 vs 0.144 for T_1, 0.055 vs 0.065 for T_2) whereas for gray matter the efficiencies are similar to their Cartesian counterparts (0.091 vs 0.088 for T_1, 0.033 vs 0.03 for T_2).

6.4 Discussion

In this work we have extended the MR-STAT framework with non-Cartesian gradient trajectories with the main purpose of comparing Cartesian and radial MR-STAT in terms of robustness and time-efficiency. Because MR-STAT uses a model-based iterative reconstruction in which spatial and dynamic encoding are coupled, providing predictions upfront on which acquisition type will result in higher efficiencies is challenging. We therefore performed an empirical study

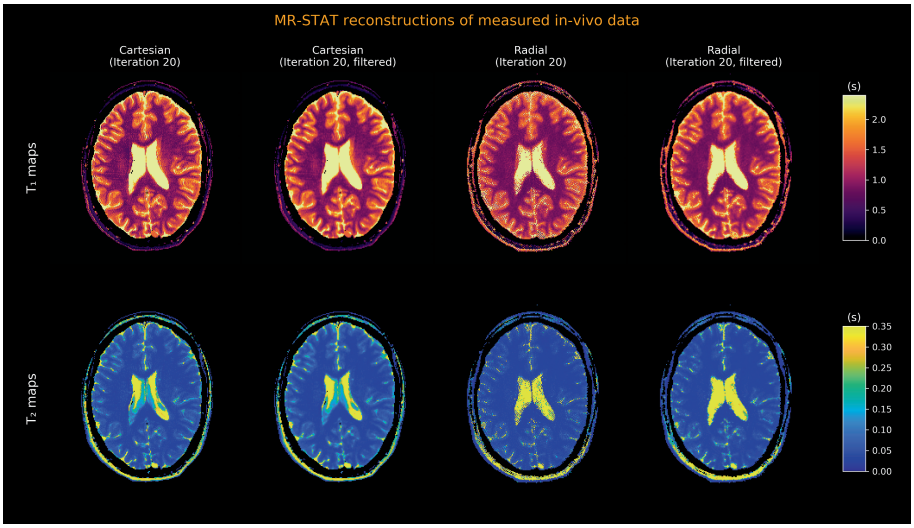

Figure 6.10: Reconstruction results for the in-vivo measurements. The Cartesian reconstructions at their final iterations are shown in the first (unfiltered) and second (Hann-filtered) columns. The radial reconstructions at their final iterations are shown in the third (unfiltered) and fourth (Hann-filtered) columns.

based on simulations and experiments instead.

In general, errors in the reconstructed parameter maps (which directly influence the efficiency) are the result of thermal noise on the measured data, undersampling and/or imperfections in the signal model used in the reconstruction (due to e.g. hardware imperfections). The parameter maps reconstructed from the Cartesian noiseless numerical brain phantom dataset did not suffer from any undersampling (i.e. aliasing) artefacts despite the high undersampling factor used (one readout per contrast). Also in the radial case no streaking artefacts were observed in the reconstructed maps. Both Cartesian and radial MR-STAT are therefore robust against undersampling. However, in the radial case the outer k-space corners were never sampled and therefore the smaller, more narrow structures and tissue boundaries could not be properly resolved. As such, radial MR-STAT has a lower spatial encoding efficiency compared to Cartesian MR-STAT. For the numerical brain phantom datasets corrupted with simulated thermal noise, we observed that the radial reconstruction resulted in higher T_2 time-efficiencies (by approximately 25%) compared to the Cartesian reconstruction in most tissue types. We hypothesize these higher efficiencies result from the fact that with the radial acquisition the center of k-space is sampled with each readout, and as such it has higher dynamic encoding power which compensates for the reduced spatial encoding that was observed in the noiseless numerical phantom reconstruction. For T_1 the radial reconstruction

only results in higher efficiencies in gray- and white matter (and only slightly in skin). The difference between T_1 and T_2 may be explained by the presence of the inversion prepulse. This prepulse adds strong dynamic T_1 encoding to both acquisitions to the point where the intrinsic higher dynamic encoding of the radial acquisition may provide little to no benefit. On the other hand, the reduced spatial encoding for radial is still present, resulting in lower efficiencies in finer structures for the radial acquisition.

In the gel phantom measurements, we observed that - while the Cartesian reconstruction procedure results in an RMSRE curve that is similar to the noisy brain simulation case - for the radial case the situation is different because the reconstruction procedure is highly susceptible to overfitting. In terms of sources of errors in the parameter maps, the main difference between the numerical simulation study and the gel phantom study is the presence of potential model inaccuracies in the latter. For radial (or more general: non-Cartesian) acquisitions it is known that they are more sensitive to hardware imperfections (e.g. gradient delays) as compared to Cartesian acquisitions [145]. We thus hypothesize that the differences between the simulation results and the gel phantom results come from inaccuracies in the forward model due to hardware imperfections and that these inaccuracies result in high-frequency artefacts in the parameter maps at the later iterations of the reconstruction procedure. In that scenario, the benefits of higher dynamic encoding for radial that was observed in the noisy numerical brain phantom case is overshadowed by the presence of these artefacts. We also observed that at the "optimal" iterations, the vials with the highest T_2 had not yet fully converged. An early stopping strategy thus poses a risk of introducing a signficant bias in the parameter maps. As an alternative strategy to reduce the impact of the high-frequency artefacts on the radial reconstructions we applied a k-space filter to the parameter maps in a post-processing step (similar to Pruessmann et al. (2001) [134]). Even after the filtering step, the Cartesian acquisition remained more efficient.

In both the noisy numerical brain phantom and gel phantom experiments the cost and RMSRE values had stabilized at iteration twenty. We therefore assumed in this work that twenty iterations would be sufficient for convergence in the in-vivo reconstructions as well. At these final iterations, the application of a (Hann) filter was again considered necessary for the radial reconstruction. For the filtered reconstructions, the radial case resulted in higher efficiencies in white matter and comparable efficiencies in gray matter. However, some of the high spatial-frequency artefacts remained visible in the radial parameter maps even after filtering, which may hamper the clinical acceptability. We expect the impact of these artefacts to be even stronger on older MR systems with less performant gradient systems and less advanced eddy-current compensation mechanisms. The necessity of the filtering also complicates the reconstruction procedure in the sense that it involves more tweaking parameters (e.g. filter type and strength). The Cartesian reconstruction, like in the simulations and

gel phantom experiments, did not display severe high-frequency artefacts. Filtering was not considered necessary but could still be desirable for visualisation of the parameter maps.

Over the entire range of experiments performed in this work, Cartesian MR-STAT was able to produce tissue parameter maps without apparent aliasing or overfitting artefacts. The reconstruction procedure was robust against the number of iterations in the reconstruction in all cases, displayed similar behaviour in both simulations and measurements and required less tweaking parameters compared to radial MR-STAT. The combination of these factors may make Cartesian MR-STAT preferable over radial MR-STAT in a clinical setting despite the lower efficiency in white matter.

An additional downside of the radial acquisitions that is not taken into account in the current analysis is the effects of off-resonances. Whereas with Cartesian acquisitions the presence of off-resonances causes shifts in the image, with radial acquisitions, because the readout direction is different each TR, off-resonances will have a blurring effect on the reconstructed images. It should be noted though that with MR-STAT the off-resonance effects during readouts can be included in the forward model. As such, it should be possible to correct for this blurring effect by providing a separately measured off-resonance map to the reconstruction algorithm. Alternatively, a sequence could be designed that allows for the joint estimation of off-resonance maps with the T_1, T_2 and proton-density maps [94].

In the current work, we did not use a density compensation function for the radial reconstruction to compensate for the fact that the lower spatial frequencies are over-represented. Using such a density compensation, for example the Ram-Lak filter, as a pre-conditioner may result in faster convergence. We expect the (unfiltered) final parameter maps to suffer from the same high-frequency errors as in the reconstructions without density compensation. Reducing the number of iterations may aid in lowering reconstruction times for radial MR-STAT.

Instead of applying post-processing k-space filters, a more principled approach to suppress high frequency artefacts during reconstructions would be to add a spatial regularization term to the parameter maps in the MR-STAT objective function. This would, however, add additional complexity into the reconstruction procedure (e.g. choice of regularization function(s) and parameter(s), potential non-differentiability of the regularization term). At the same time we expect the outcomes to be the same: for radial, regularization would be necessary to stabilize the reconstructions whereas for Cartesian it is not necessary (but it may still be used to reduce noise in the reconstructed parameter maps).

One important limitation of this work is that we only considered a single RF train type in all acquisitions. The RF train was designed such that its local maxima were incoherent with respect to the sampling of the k-space center for the Cartesian acquisitions [54]. No optimization schemes were applied. For the

radial case, since all readouts cover equivalent portions of k-space, it may not be necessary to take the spatial encoding into account such that the optimization can focus purely on enhancing the dynamic encoding power [8, 140]. In the Cartesian case however, different readouts cover distinct parts of k-space and the optimization will have to strike a balance between spatial and dynamic encoding power. Performing such an optimization is non-trivial and is outside the scope of this work [55]. Since optimization schemes for Cartesian and radial can result in different RF trains for both, the efficiencies and conclusion drawn in the current work may be influenced when considering such optimized sequences.

Another limitation of this work is that we only considered 2D acquisitions in this work. For 3D acquisitions, we expect an amplification of the dynamic encoding benefits of radial over Cartesian. Hardware imperfections in radial acquisitions remain an issue also at 3D and future research will be aimed at better understanding and mitigating the impact of these imperfections on the radial reconstructions, for example by utilizing gradient impulse response functions [168] to correct the gradient trajectory in the forward model (Eq. 6.2.1).

6.5 Conclusion

We extended the MR-STAT framework to non-Cartesian acquisitions and compared Cartesian and radial MR-STAT in terms of robustness and time-efficiency. While radial MR-STAT resulted in higher T_2 efficiencies in numerical simulations, in the gel phantom experiment the efficiencies were lower compared to Cartesian MR-STAT and we argue this is due to increased sensitivity to hardware imperfections. In clinical practice, the robustness and reliability of Cartesian MR-STAT may be preferred, especially on older MR systems where the impact of hardware imperfections on the radial reconstructions are expected to be more severe. With this work we thus would like emphasize that Cartesian acquisitions are still highly relevant in the field of multi-parametric qMRI.

Acknowledgements

The authors would like to thank Tom Bruijnen and Mariya Doneva for fruitful discussions regarding the experimental setup and interpretation of the results and Jordi Kleinloog for assisting with image segmentation. This research has been financed by the Dutch Technology Foundation under grant #17986.

6.6 Supplementary Materials

Supplementary Material S1: Extending radial readouts

In Fig. 6.2 of the main manuscript the MR-STAT reconstructions performed on noiseless numerical brain phantom data are shown. The Cartesian reconstruction results in errors in the parameter maps that are practically zero. The radial reconstruction has some noise-like residual errors in the parameter maps. Here we show that these errors are the result of the outer corners of k-space never being sampled for the radial acquisition.

In columns two and three of Fig. 6.11, we show the same reconstructions as shown in Fig. 6.2 of the main manuscript but we also display the spatial frequency content (magnitude) of the error maps. For the radial reconstruction, the errors occur exactly at the corners of k-space that were never sampled. We then modified the radial trajectory to also cover these corners during the acquisition. This was achieved by increasing the number of samples per spoke by a factor of 2 while keeping the Δk per sample fixed. Data was simulated with this modified trajectory and reconstructions performed on this data are shown in the fourth column of Fig. 6.11. The errors in both T_1 and T_2 are practically zero for this modified radial case.

The (original) radial acquisition thus comes with an inherent loss of resolution as compared to the Cartesian acquisition. As demonstrated in the main manuscript, in the presence of thermal noise, this reduced *spatial* encoding power of radial is partly compensated for by the higher *dynamic* encoding power.

The RMSRE curves for the three reconstructions are shown in Fig. 6.12. While the modified radial case results in RMSRE values that are almost five orders of magnitude lower than the original radial case, it is still orders of magnitude higher than the Cartesian case. We hypothesize the higher errors (as compared to the Cartesian case) are the result of discretization errors sampling an object (i.e. parameter maps) defined on a Cartesian grid using a non-Cartesian trajectory and can be ignored for practical purposes.

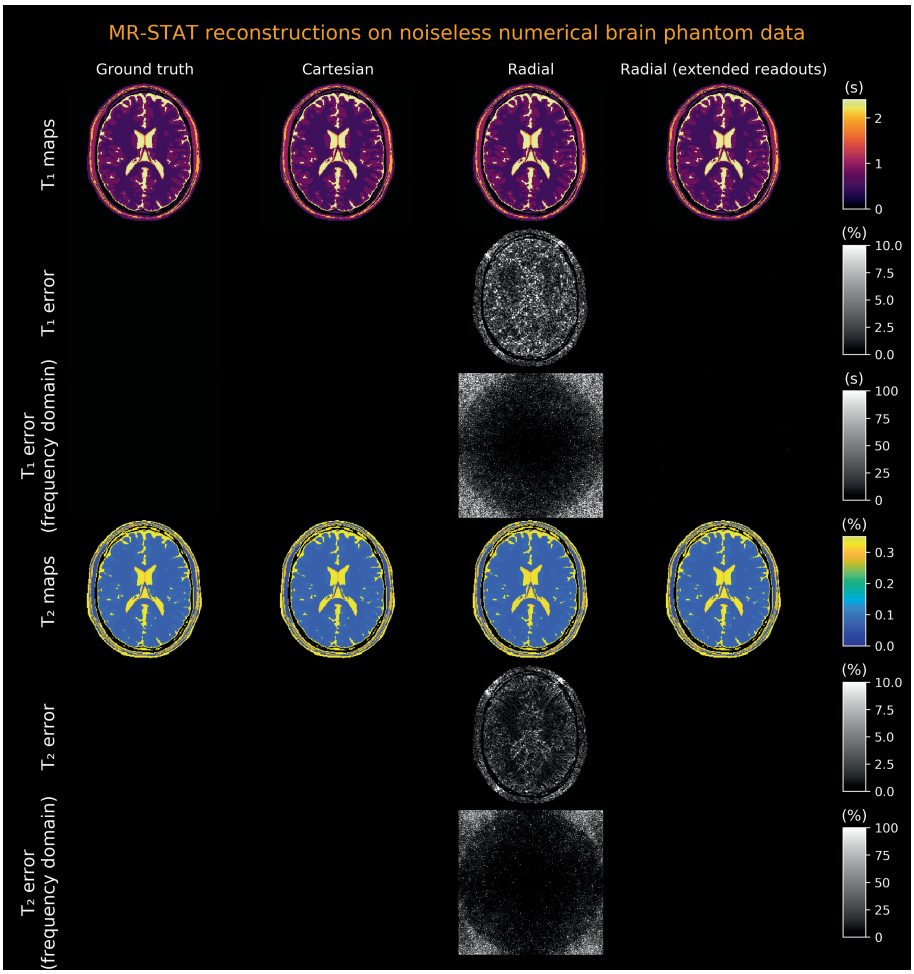

Figure 6.11: The T_1 and T_2 parameter maps reconstructed from noiseless simulated brain data. The first column shows the ground truth T_1 and T_2 maps. The second column shows the parameter maps and relative error maps for the Cartesian case after twenty iterations (also in frequency domain). The third and fourth columns show the parameter maps and relative error maps at iteration twenty for the original and modified radial trajectory, respectively.

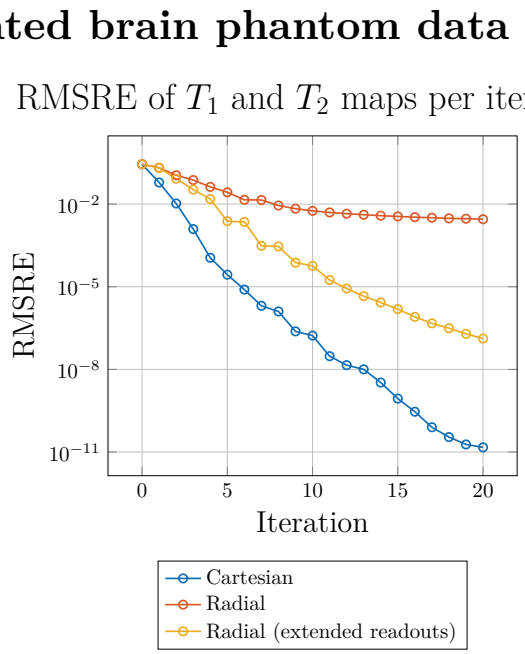

Figure 6.12: RMRSE values (for T_1 and T_2 combined) per outer iteration. The extended radial case results in significantly lower RMSRE values as compared to the original radial case.

7

Summary and Outlook

In this chapter the contributions from the individual chapters are contextualized and discussed. Current technical challenges with potential solution strategies are outlined and directions for future research are provided.

7.1 Contextual Summary

The overal goal of this thesis was to advance the MR-STAT framework to the point where it could be utilized in a clinical demonstrator setting to reconstruct T_1, T_2 and proton density maps. The project focussed equally the data acquisition as well as the image reconstruction parts of MR-STAT.

7.1.1 Data Acquisition

On the data acquisition side, pulse sequences with short scan times had to be designed and implemented that could result in high-resolution in-vivo quantitative tissue parameter maps. Initially, in Chapter 2, a Cartesian gradient-balanced sequence with flip angles drawn from a normal distribution was proposed. While this sequence resulted in succesful MR-STAT reconstructions in silico, reconstructions on phantom and in vivo data would typically be corrupted by image artefacts. Based on several works that study (hardware induced) model imperfections for gradient-balanced sequences [142, 18, 6], it was realized that for completely random flip angles, intra-voxel dephasing may occur and the spin-echo nature of balanced sequences is lost. This spin-echo nature, however, was an important assumption in the single-isochromat based

forward model used in the reconstructions. The model could therefore not accurately describe the data and as a result the reconstructed parameter maps would be inaccurate. To resolve this issue, a transition to smoothly varying flip angles was made that, together with a linear Cartesian sampling of k-space, do preserve the spin-echo behaviour. The results presented in Chapters 2-5 are all based on this type of sequence. The linear Cartesian sampling pattern is known to be robust to scanner hardware imperfections and - being the workhorse in current clinical practice - is directly available on the installed base of MR systems.

While gradient-balanced sequences are superior in terms of signal-to-noise ratio, a known fundamental issue with such sequences is their sensitivity to off-resonances. At lower field strengths (e.g. 1.5 Tesla and lower) and with short repetition times, these issues are manageable but at higher field strengths they result in severe banding artefacts. In order to facilitate a demonstrator patient study at 3 Tesla [88], a switch was made during the project to gradient-spoiled sequences which have little sensitivity to off-resonances but come with reduced signal-to-noise ratio.

In an attempt to improve the signal-to-noise ratio, the MR-STAT framework was extended to non-Cartesian acquisition and experiments were performed using golden angle radial trajectories. The motivation for using radial trajectories rather than, for example, spiral trajectories was to allow for a relatively straightforward comparison against Cartesian trajectories. This comparison was performed in Chapter 6. While in simulations the radial trajectory was shown to be more efficient, in actual measurements it came with reliability and robustness challenges that were attributed to a higher sensitivity to hardware imperfections.

7.1.2 Image Reconstruction

On the image reconstruction side, initially in Chapter 2 a one-dimensional FFT was applied along the readout direction to decouple the MR-STAT reconstruction problem into multiple one-dimensional MR-STAT problems. For each of these individual problems, the variable projection method was used. For the Bloch simulations that are required to evaluate the forward model (and for the computation of partial derivatives of the forward model using finite differences) an online available Bloch simulation toolbox written in the C programming language was used. The reconstruction problems were solved in parallel using multiple CPUs on a high performance computing cluster.

An issue with the reconstruction technique presented in Chapter 2 is that the variable projection method requires model matrices to be stored in computer memory. For 1D problems this is feasible but for 2D or 3D problems it is not. Because the decoupling strategy relies on the assumption of Cartesian readout trajectories, this reconstruction setup would limit MR-STAT to Cartesian only even though the theoretical framework is also valid for non-Cartesian

trajectories. In addition, for 3D Cartesian acquisitions, the decoupling strategy would result in multiple 2D problems for which the variable projection method could not be used. A more generic MR-STAT algorithm that could be used for arbitrary gradient trajectories was therefore desired.

In Chapter 3 a matrix-free Gauss-Newton based algorithm was developed that, in principle, can be used for arbitrary sequences because it does not require storage of model matrices. The matrix-free (and thus memory-efficient) nature of the algorithm comes at the price of increased computational costs: entries of the relevant model matrices that are needed multiple times during the reconstruction are re-computed multiple times because they are no longer stored. Fortunately, the computations can be performed in parallel to reduce reconstruction times. Unlike the FFT-based decoupling strategy, communication between compute units is required during the reconstruction procedure. In order to implement this algorithm on a high performance computing cluster, a programming language was needed that could facilitate the communication between compute units without requiring labor intensive, low-level programming. The relatively new, free and open-source Julia programming language satisfied these requirements and was thus used to implement the parallel, matrix-free Gauss-Newton algorithm on a high performance computing (CPU) cluster. While the algorithm allowed for the reconstruction of high-resolution (i.e. 1 mm $\times$ 1 mm in-plane resolution) quantitative tissue parameters, reconstruction times were in the order of hours for a single slice even when approximately 100 CPUs were utilized.

In the search for techniques to reduce reconstruction times, it was realized in Chapter 4 that the Gauss-Newton matrix used in the MR-STAT reconstruction technique proposed in Chapter 3 admits a sequence-dependent structure. Predictions were made based on a theoretical foundation that, for sequences with Cartesian trajectories and smoothly varying flip angles, the Gauss-Newton matrix admits (to good approximation) a sparse structure. This insight allows the non-zero entries to be computed (in parallel) and stored in computer memory. Subsequent matrix-vector multiplications with the sparse matrix are then significantly faster when compared to the matrix-free approach. The theoretical predictions were validated in simulations as well as phantom and in vivo experiments. Reconstruction times were reduced by an order of magnitude, resulting in reconstruction times of approximately fifteen minutes per slice while still using almost 100 CPUs. The technique developed in Chapter 4 can be interpreted as a generalization of the FFT-based decoupling technique used in Chapter 5. Unfortunately, for non-Cartesian trajectories, the technique would be more much challenging to implement and may not result in significant reductions in reconstruction times due to the less sparse matrix structure.

Prompted by the availability of GPU hardware, as well as having observed significant computational speedups achieved with GPUs in other research fields (including MRI), the decision was made to work towards a GPU compatible MR-STAT reconstruction algorithm in Chapter 5. The Julia programming lan-

guage turned out to be a valuable tool here as well since it allows one to write custom CUDA kernels in a high-level manner that can then be executed on NVIDIA GPU hardware. Kernel functions were written for the computationally demanding tasks of the MR-STAT reconstruction algorithm proposed in Chapter 3.

In addition, it was recognized that, given the magnetization (and partial derivatives) in a voxel at some echo time, the magnetization at other sample points during the same readout are relatively cheap to compute. At the same time, for 2D reconstructions, storing the magnetization (and partial derivatives) at echo times only is feasible on modern GPU cards. While the matrix-free Gauss-Newton method proposed in Chapter 3 requires all entries of the Jacobian matrix to be recomputed multiple times within one outer iteration of the MR-STAT reconstruction algorithm, with the *partially* matrix-free approach proposed in Chapter 5, the expensive entries of the Jacobian are computed once and then stored into GPU memory. Only the entries of the Jacobian that are relatively cheap to compute are recomputed multiple times within one outer iteration. With this technique, 2D MR-STAT reconstructions that took approximately three hours in Chapter 3 and approximately fifteen minutes in Chapter 4 could be performed in 68 seconds on a modern GPU card. An additional advantage of this approach is that - unlike the sparse Hessian based acceleration technique proposed in 4 - the technique can also be utilized for non-Cartesian acquisitions and was essential in performing the radial MR-STAT reconstructions in Chapter 6.

An important part of the MR-STAT reconstructions is formed by the need to perform Bloch simulations (using either the isochromat or the EPG model) and to evaluate the signal model 1.1.1. These simulations, however, could also be relevant to other (q)MRI techniques and therefore the developed Julia code to perform these simulations was released as a separate, free and open-source software package entitled BlochSimulators.jl. In Chapter 5 the toolbox is demonstrated to have state-of-the-art runtime performance on both CPU and GPU architectures.

Based on the work presented in this thesis, whole-brain multi-2D MR-STAT acquisitions of five minutes are possible from which T_1, T_2 and proton density maps can be estimated with reconstruction times in the order of two minutes per slice. A first clinical demonstrator study on 30 patients with various diseases has been concluded [88]. A second clinical demonstrator study on a large group of patients suffering from Parkinson's disease is being conducted at the time of this writing [24].

7.2 Discussion

Going forward with MR-STAT, several challenges will need to be addressed prior to clinical adoption. These challenges are discussed below together with

suggestions for possible mitigation strategies and other future research directions leverage the unique features of the MR-STAT framework.

7.2.1 Volumetric coverage

This thesis focussed entirely on single-slice, 2D acquisitions. Volumetric coverage can be achieved in a straightforward manner by performing multiple 2D acquisitions in a serial manner [88]. However, such an approach is likely to be suboptimal in terms of scan efficiency and makes it difficult to achieve an isotropic resolution. Two potentially better approaches should be explored: interleaved multi-slice and/or 3D acquisitions.

With interleaved multi-slice type acquisitions, the data acquisition for a slice is interrupted to excite and acquire data from other slices. The interruption time would allow for T_1 regrowth of the longitudinal spin components, resulting in higher signal (and thus higher scan efficiency) without increasing the overal scan time. Such an acquisition involves many design choices, e.g. the number of interruptions and the interruption times are involved, that influence the scan efficiency. Sequence optimization tools, such as the BLAKJac framework [55] that was designed specifically for optimizing MR-STAT sequences, are essential in this context.

Alternatively, 3D MR-STAT acquisitions could be utilized, especially when clinical applications demand isotropic voxels. A major challenge with 3D acquisitions is that memory requirements and reconstruction times are even more difficult to manage compared to 2D acquisitions. In exploratory work by Liu et al. (2023) [107], a *stationary* [2] MR-STAT sequence was developed that repeats the same 2D MR-STAT acquisition for different phase encoding steps in the k_z direction. This method allows for decoupling (with the help of parallel imaging if undersampling in the k_z direction is applied) of the 3D reconstruction problem into multiple 2D reconstruction problems. Compared to multi-2D acquisitions, these type of 3D acquisitions were observed to have higher scan efficiencies [107]. From a conceptual point-of-view it would also be interesting to develop (non-decoupled) 3D MR-STAT reconstruction techniques since the decoupling strategies put constraints on the acquisition strategy. The benefits of non-Cartesian MR-STAT imaging that were explored in Chapter 6 could also be more pronounced for 3D acquisitions (although the challenges in terms of robustness remain). Like in the interleaved multi-slice case, 3D MR-STAT acquisitions require many design choices and potential pitfalls that need to be thoroughly investigated with the aid of sequence optimization tools [55].

7.2.2 Reconstruction times

Using the reconstruction technique presented in Chapter 5, 2D MR-STAT reconstructions can be performed in the order of minutes on GPU hardware,

with the actual reconstruction time depending on the desired resolution, sequence length, number of receive coils and simulation parameters (e.g. number of discretization steps for RF excitation waveforms or the maximum allowed configuration state order in EPG simulations). Reconstruction times for 3D volumes (acquired using the multi-2D approach) are currently in the order of hours. While acceptable for research purposes or clinical demonstrators, such reconstruction times would not be sufficient if MR-STAT were to be adopted on a large scale in clinical protocols. Reconstruction times would ideally be in the order of a minute for a 3D volume as this would also allow direct feedback of scan results to radiologic technologists. Several strategies could be considered to work towards this goal.

First of all, a GPU implementation of the sparse Hessian technique introduced in Chapter 4 could significantly reduce reconstruction times for Cartesian MR-STAT sequences. The design of a kernel function for computing and storing the sparse Hessian approximation is non-trivial and involves many design choices that could impact the performance of such a kernel. Assuming that a sparse Hessian approximation is stored in GPU memory, existing libraries to perform sparse matrix-vector multiplications on GPU hardware [118] could be leveraged.

For non-Cartesian MR-STAT sequences for which the sparse Hessian technique is less suitable, speedups could be achieved by improving the existing kernel functions for the partially-matrix free technique from Chapter 5. As demonstrated in that chapter, the bottleneck in the partially matrix-free MR-STAT reconstructions is found in the repeated matrix-vector multiplications with the Jacobian matrix $\mathbf{J}$ and its conjugate transpose. The kernels that implement these multiplications involve many design choices (e.g. letting threads loop over either columns or row of the matrix) and have not been thoroughly investigated. Rather than manually searching for optimal design choices in this respect, tools to automate this process could be utilized [180]. An interesting question in this respect is whether the Julia language - which was used in this work to develop the kernel functions - remains a viable option or whether significant speedups can be achieved by using the CUDA language instead.

Rather than performing reconstructions on a single GPU card, MR-STAT reconstructions could also be performed on (local or cloud-based) GPU clusters consisting of multiple GPU cards. For multi-2D MR-STAT acquisitions, reconstructing different slices is an *embarrassingly parallel* problem and each available GPU could be assigned to a different subset of the slices. For 3D MR-STAT acquisitions, communication between different GPU cards during the reconstruction would be required and tools to facilitate this communication need to be explored [70]. The multi-GPU approach may be necessary to utilize the partially matrix-free approach in the 3D scenario because the memory requirements for this algorithm are expected to exceed what is available on a single GPU card at the time of this writing.

On an algorithmic level, a stochastic modification of the inexact Gauss-

Newton method proposed in Chapter 3 could be explored. In the inexact Gauss-Newton method, a linear system of equations involving the full Jacobian matrix $\mathbf{J}$ is numerically solved for each outer iteration of the iterative algorithm to obtain update steps in parameter space. The Jacobian matrix $\mathbf{J}$ contains partial derivatives of the forward model with respect to the tissue parameters of interest at all sample times. Instead of using *all* sample points in the linear system of equations, a stochastic approach that uses only subsets of the sample points [161, 176] could accelerate the computation of the update steps. Predicting potential speedup up front is difficult. Reducing the number of sample points linearly reduces the computation time but at the same time may require additional iterations for convergence. An optimal balance must thus be found.

In a different direction, speedups may be possible by incorporating surrogate models into the reconstruction. In Chapter 5, the performance of the EPG simulator used in MR-STAT is on-par with the performance of a recurrent neural network-based surrogate model [104]. However, the recurrent neural network was chosen as a balance between performance on the one hand and flexibility with respect to sequence parameters on the other. If, for example in a clinical setting, a fixed sequence (e.g. fixed flip angles at a fixed resolution) is used, a different network architecture may be utilized that sacrifices flexibility to gain additional runtime performance [80].

Besides surrogate modeling, neural networks could potentially be used to learn the iterative steps of the MR-STAT reconstruction. Such *loop-unrolling* techniques [117] have been used successfully in the context of compressed sensing MRI to significantly reduce reconstruction times [139].

A proper combination of the proposed strategies may make it possible to achieve one-minute MR-STAT reconstructions for 3D volumes.

7.2.3 Accuracy of tissue parameter estimates

In Chapters 2,3,4 and 6 the T_1 and T_2 maps obtained with MR-STAT were shown to be accurate for in silico and phantom experiments. For the phantom experiments, accuracy was assessed by comparing against results from gold standard techniques (single-echo spin-echo inversion recovery with different inversion times for T_1 mapping, single-echo spin-echo with different echo times for T_2 mapping). Unfortunately, the gold standard techniques have scan times that are prohibitively long for being performed in vivo. Assessing accuracy by comparing against other qMRI techniques that have shorter scan times is not straightforward. For example, for brain imaging, there is a wide range of reported T_1 and T_2 values in gray- and white matter [23] and there is no concensus on what the "true" values are.

The wide variety of reported in vivo relaxation times can be partly attributed to the use of incomplete signal models in qMRI. There are many biophysical processes (besides transverse/longitudinal relaxation and proton den-

sity) that have an impact on the MR signal and can bias the relaxation times if not accounted for during the reconstruction procedure. Examples include diffusion, flow and magnetization transfer. In addition, system imperfections such as inhomogeneous B_1^+ and B_0 fields can also bias the reconstructed relaxation times. Unfortunately, different qMRI techniques have different sensitivities to each of these confounding factors. As a result, T_1 estimates obtained with one technique may be biased by one confounder while the T_1 estimates obtained with another technique may be biased by another. Both techniques may thus return different T_1 estimates. Several approaches exist to deal with confounding factors in qMRI.

The first approach is to design sequences that are insensitive to the confounding factors such that they can be ignored during the reconstruction procedure without adversely affecting the relaxation time estimates. As an example, single-echo FISP-type [124] sequences as used in Chapter 6 are known to be largely insensitive to B_0 inhomogeneities. For such sequences, it is therefore not necessary to include potential B_0 inhomogeneities into the signal model used in the qMRI reconstruction. It is, however, highly questionable whether sequences can be designed that are sufficiently insensitive to *all* potential confounders.

As a second option, separate measurements could be performed to retrieve confounding parameters so that they can subsequently be included in the MR-STAT forward model. This is the approach currently taken for B_1^+ in Chapter 6. The obvious downside is an increase in acquisition time and a fallback to the more traditional qMRI approach of estimating individual tissue parameter types one at a time.

A third option would be to design sequences that - on purpose - have high sensitivity to the confounders such that they can be reconstructed from the data as well. For example, in the field of MRF, sequences have been proposed that allow for reconstruction B_1^+, off-resonance, diffusion, magnetization transfer, flow, perfusion. A challenge with this latter approach is that the acquisition times typically become longer as more parameters are encoded into the signal. At the same time, reconstruction procedures become more complex as well. With the dictionary-based MRF approach to multi-parametric qMRI, the size of the dictionary in principle scales exponentially with the number of different tissue parameter types and one quickly runs into both memory and matching time issues as more tissue parameter types are included. For MR-STAT, the memory requirements (for storing part of the Jacobian matrix in the partially matrix-free algorithm from Chapter 5) and computation time (for computing partial derivatives of transverse magnetization) in principle scale linearly as more parameters are included. The MR-STAT approach to qMRI may therefore be more future-proof in this respect. A challenge with the MR-STAT approach is that - as more parameter types are included - the risk of encountering local minima during the iterative reconstruction procedure may increase.

7.2.4 Spin dynamics during readouts

In MR-STAT, a forward model (Eq. 1.4.1) is used that describes time-domain sample points by explicitly taking into account the spatial encoding gradients. A benefit of this approach over FFT-based techniques is that spin dynamics *during* readouts, such as T_2, T_2^* or off-resonance effects, can be considered. However, in this thesis, this benefit has not been explored. Only sequences with relatively short readouts (in the order of a few milliseconds) were tested for which the spin dynamics during readouts can typically be ignored. For sequences with longer readouts, such as spiral or echo planar imaging ("EPI") acquisitions [111], spin dynamics during readouts are known to result in image artefacts (e.g. blurring or geometric distortions) in FFT-based reconstructions. With MR-STAT it should be possible to reconstruct quantitative maps that do not suffer from such artefacts and future research could be dedicated to exploring this unique feature of MR-STAT.

7.2.5 Continuous sampling

In a typical MRI pulse sequence, data is only acquired during the constant part of the readout gradient. In theory, however, data can be sampled continuously in between RF excitations, including the ramp up and ramp down moments of the spatial encoding gradients. Regular Cartesian sequences become hybrid Cartesian/radial sequences with such an approach [25]. In conventional (non-uniform) FFT-based reconstructions, the additional data points can be included by gridding prior to applying an FFT [182, 26]. Data points that are sampled multiple times during a continuous sampling window are essentially averaged during the gridding step without taking into account the spin dynamics during the sampling window. In MR-STAT, no gridding is required and the additional data points can be included naturally into the reconstruction procedure. This continuous sampling technique could be leveraged to increase the signal-to-noise ratio for MR-STAT without increasing scan times. For this approach to be successfull precise knowledge of the actual gradient waveforms is required. Tools such as gradient impulse response functions could be essential in this respect [168].

7.2.6 Convergence and regularization

The MR-STAT reconstruction algorithms presented in this work are all iterative in nature due to the use of a non-linear forward model. Proper convergence criteria for halting the iterative reconstructions have not been investigated. Early stopping may result in heavily biased parameter estimates whereas stopping too late increases reconstruction times and can result in overfitting artefacts. Commonly used stopping criteria involve the vector norm of the gradient of the objective function [120], the vector norm of the step in parameter space or

the change in objective function.

A topic closely related to stopping criteria is regularization. Many (q)MRI techniques incorporate regularization terms as they allow prior information to be included to improve image quality [101, 109, 131]. While technically possible [106], in this thesis no explicit regularization terms have been included in the MR-STAT reconstructions. A more thorough investigation of different (joint) regularization strategies for MR-STAT is recommended since regularization can improve image quality as well as prevent overfitting artefacts.

7.2.7 Uncertainty Estimation

With the MR-STAT approach to qMRI it is in theory possible to estimate uncertainties associated with the reconstructed tissue parameter maps. If η denotes the noise on the measured MR data and $\mathbf{J}$ the Jacobian matrix of the MR-STAT problem evaluated at the reconstructed tissue parameters, then the uncertainty associated with the i-th reconstructed parameter is estimated as $\sqrt{\text{diag}(\text{Re}\left\{\mathbf{J^H J}\right\}^{-1})_i}/\eta$ (see Chapter 2). Such uncertainty maps may prove to be useful in the clinical assessment of quantitative tissue parameter maps. For small numerical phantoms, or when using the 1D FFT-based decoupling strategy, the covariance matrix $(\text{Re}\left\{\mathbf{J^H J}\right\})^{-1}$, and therefore the uncertainty maps, can be computed in a straightforward manner. However, when no decoupling strategy is employed (such as in Chapters 3-6), directly computing the covariance matrix is no longer feasible from both a memory and computation time point of view. Because matrix-vector products with $\text{Re}\left\{\mathbf{J^H J}\right\}$ can be computed (either in a matrix-free fashion as proposed in Chapter 3, a partially matrix-free fashion as proposed in Chapter 5, or by using the sparse approximation in the case of Cartesian acquisitions as proposed in Chapter 4), the columns of $(\text{Re}\left\{\mathbf{J^H J}\right\})^{-1}$ can in principle be estimated one at a time by solving linear systems of the form

$$\text{Re}\left\{\mathbf{J^H J}\right\}\mathbf{x} = \mathbf{e_i}, \tag{7.2.1}$$

where $\mathbf{e_i}$ is the i-th standard basis vector. When $\text{Re}\left\{\mathbf{J^H J}\right\}$ can be approximated by a sparse matrix, solving Eq. **??** for each reconstructed may be computationally feasible. In the more general case, however, the computation times will be prohibitively long. Further research into alternative strategies for estimating the diagonal of the covariance matrix [11, 53] is recommended as uncertainty maps may be important for clinical acceptance of MR-STAT.

7.2.8 MRF vs MR-STAT

Reconstructions in the original MRF implementation proposed by Ma et al. (2013) [110] and described in Chapter 1 consist of applying (non-uniform) FFTs

on many highly undersampled snapshot images followed by a voxel-wise dictionary matching procedure. Matching times reported for these original MRF reconstructions [110, 84] are typically much lower compared to MR-STAT reconstruction times reported in this thesis. The matching, which is in essence a discretized global parameter search, also avoids potential issues with local minima. However, as described in Chapter 1, MRF introduces artificial aliasing noise into the reconstructions. In order to reduce the effects of aliasing artifacts or to reduce acquisition times, iterative reconstruction techniques have been proposed for MRF [132, 192]. These newer, iterative MRF techniques typically do not treat the reconstruction problem as a one-step non-linear inversion problem like in MR-STAT but maintain a two-step approach [4]. In the spatial localisation step, rather than direct (non-uniform) FFT reconstructions, a linear imaging operator is inverted iteratively to obtain snapshot images from the measured data. The subsequent parameter estimation step is then performed on a voxel-per-voxel basis through dictionary matching. The reconstructions may alternate between the two different steps [192, 5, 191].

Compared to the initially proposed MRF reconstruction method the iterative MRF reconstructions are computationally more demanding and reintroduce potential issues with local minima. Also note that the linear system in the spatial localisation step is in principle highly underdetermined: since each snapshot is highly undersampled, there are many more unknowns than equations. To deal with this latter issue, techniques have been proposed that reconstruct the snapshots in a low-rank (time-compressed) basis instead [5, 191, 115]. That way the number of unknowns in the linear spatial localisation problem can be reduced significantly. The singular value decomposition is commonly used as a tool to extract a low-rank basis from the MRF dictionary [114]. However, as more different types of tissue parameters may be needed in qMRI signal models in the future, forming and compressing dictionaries becomes (exponentially) more challenging. In addition, a higher rank (i.e. more singular vectors) may be required to capture enough information about the underlying spin dynamics, to the point where the spatial localisation problem becomes highly underdetermined again.

Despite technical differences in the reconstruction techniques, each with its own set of advantages and disadvantages, it can be said that the distinction between MRF and MR-STAT has decreased over time. However, by using the one-step, non-linear inversion approach, the MR-STAT technique is arguably more future-proof than the two-step approach-based MRF technique. At the same time the forward model used in MR-STAT opens up research directions (e.g. non-linear encoding fields) not permitted by techniques that rely on the (non-uniform) FFT in their reconstructions.

7.3 Future Perspectives

Current clinical MRI workflows consist of multiple scans, one for each desired contrast image. A fast, multi-parametric qMRI technique may replace all those scans with just a single scan. The reduced scan times can result in lowered costs and increased patient throughput for hospitals. At the same time, the obtained quantitative images are richer in information than their qualitative counterparts. With MR-STAT, a technique has been developed that may deliver on this promise of multi-parametric qMRI. The MR-STAT framework uses a forward model that is more comprehensive compared to other qMRI methods to allow for maximum freedom in the acquisition procedure. This freedom has been utilized to design multi-parametric Cartesian-based acquisitions with scan times that are sufficiently short for incorporation into clinical routines.

The freedom in the MR-STAT acquisitions come at the price of more challenging reconstructions. A large part of this thesis was devoted to managing MR-STAT reconstructions in terms of computation times and computer memory requirements. With the suggestions provided in this chapter, together with the fact that, in general, computing capabilities become cheaper and more accessible over time, the reconstruction times may no longer be a limiting factor in the near future.

Arguably the most important limiting factor when it comes to clinical acceptance of MR-STAT is the accuracy issue described in subsection 7.2.3. This issue, however, is not specific to MR-STAT but in principle affects all qMRI methods. The signal models currently used in qMRI are too simplistic and do not cover all of the biophysical processes that affect the measured MR signal. Different qMRI methods may therefore have different sensititivites to confounding factors and thus result in different tissue parameter estimates. It is an open question in the field of fast, multi-parametric qMRI on what type of acquisition should be used and which parameters could or should be estimated jointly from the measured data. Without consensus on this matter, clinical interpretation of estimated tissue parameters is challenging [60]. Even though quantitative tissue parameter maps are already being explored as potential clinical biomarkers [56, 92], joint research efforts by the entire qMRI community may be required on this particular topic before qMRI can truly deliver on its promised benefits.

In the meantime, however, fast multiparametric qMRI techniques such as MR-STAT could still see adoption in clinical workflows through so called *synthetic MRI* methods [63]. With synthetic MRI, Bloch-equation based signal models are used to generate contrast images from quantitative tissue parameter maps. By using MR-STAT as the backbone for a synthetic MRI engine, the contrast images that are currently acquired in clinical workflows can be generated with reduced acquisition times. Succesful synthetisation does require signal models that accurately describe the clinically used contrast sequences. For some contrasts analytical signal models have been used to generate synthetic

images with comparable image quality and diagnostic value as the conventionally acquired contrast images [157, 88, 91]. However, for other contrasts, the analytical signal models are currently lacking and more advanced, data-driven techniques may be required [138, 185].

Bibliography

[1] Afzali, M., Mueller, L., Sakaie, K., Hu, S., Chen, Y., Szczepankiewicz, F., Griswold, M. A., Jones, D. K., and Ma, D. (2022). MR Fingerprinting with b-Tensor Encoding for Simultaneous Quantification of Relaxation and Diffusion in a Single Scan. *Magnetic Resonance in Medicine*, 88(5):2043–2057.

[2] Amthor, T., Doneva, M., Koken, P., Sommer, K., Meineke, J., and Börnert, P. (2017). Magnetic resonance fingerprinting with short relaxation intervals. *Magnetic Resonance Imaging*, 41:22–28.

[3] Ashton, E. (2010). Quantitative MR in multi-center clinical trials. *Journal of Magnetic Resonance Imaging*, 31(2):279–288.

[4] Assländer, J. (2021). A perspective on MR fingerprinting. *Journal of Magnetic Resonance Imaging*, 53(3):676–685.

[5] Assländer, J., Cloos, M. A., Knoll, F., Sodickson, D. K., Hennig, J., and Lattanzi, R. (2018). Low rank alternating direction method of multipliers reconstruction for MR fingerprinting. *Magnetic Resonance in Medicine*, 79(1):83–96.

[6] Assländer, J., Glaser, S. J., and Hennig, J. (2017). Pseudo Steady-State Free Precession for MR-Fingerprinting. *Magnetic Resonance in Medicine*, 77(3):1151–1161.

[7] Assländer, J., Gultekin, C., Mao, A., Zhang, X., Duchemin, Q., Liu, K., Charlson, R. W., Shepherd, T., Fernandez-Granda, C., and Flassbeck, S. (2022). Rapid quantitative magnetization transfer imaging: utilizing the hybrid state and the generalized Bloch model. *arXiv preprint arXiv:2207.08259*.

[8] Assländer, J., Lattanzi, R., Sodickson, D. K., and Cloos, M. A. (2019). Optimized quantification of spin relaxation times in the hybrid state. *Magnetic Resonance in Medicine*.

[9] Aubert-Broche, B., Evans, A. C., and Collins, L. (2006). A new improved version of the realistic digital brain phantom. *NeuroImage*, 32(1):138–145.

[10] Balaban, R. S. and Ceckler, T. (1992). Magnetization transfer contrast in magnetic resonance imaging. *Magnetic resonance quarterly*, 8(2):116–137.

[11] Balmand, S. and Dalalyan, A. S. (2016). On estimation of the diagonal elements of a sparse precision matrix. `https://arxiv.org/pdf/1504.04696.pdf`. [Online; accessed 12-June-2023].

[12] Barisano, G., Sepehrband, F., Ma, S., Jann, K., Cabeen, R., Wang, D. J., Toga, A. W., and Law, M. (2019). Clinical 7T MRI: Are we there yet? A review about magnetic resonance imaging at ultra-high field. *The British journal of radiology*, 92(1094):20180492.

[13] Ben-Eliezer, N., Sodickson, D. K., and Block, K. T. (2015). Rapid and accurate T2mapping from multi-spin-echo data using bloch-simulation-based reconstruction. *Magnetic Resonance in Medicine*, 73(2):809–817.

[14] Bernstein, M. A., King, K. F., and Zhou, X. J. (2004). *Handbook of MRI Pulse Sequences*. Elsevier.

[15] Bezanson, J., Edelman, A., Karpinski, S., and Shah, V. B. (2017). Julia: A fresh approach to numerical computing. *SIAM review*, 59(1):65–98.

[16] Bielke, G., Meves, M., Meindl, S., Brückner, A., Rinck, P., von Seelen, W., and Pfannenstiel, P. (1984). A systematic approach to optimization of pulse sequences in NMR-imaging by computer simulations. In *Technology of Nuclear Magnetic Resonance*. The Society of Nuclear Medicine.

[17] Bieri, O., Mamisch, T. C., Trattnig, S., and Scheffler, K. (2008). Steady state free precession magnetization transfer imaging. *Magnetic Resonance in Medicine*, 60(5):1261–1266.

[18] Bieri, O., Markl, M., and Scheffler, K. (2005). Analysis and compensation of eddy currents in balanced SSFP. *Magnetic Resonance in Medicine*, 54(1):129–137.

[19] Bloch, F. (1946). Nuclear induction. *Physical review*, 70(7-8):460.

[20] Block, K. T., Uecker, M., and Frahm, J. (2009). Model-based iterative reconstruction for radial fast spin-echo MRI. *IEEE Transactions on Medical Imaging*, 28(11):1759–1769.

[21] Boada, F., Liang, Z.-P., and Haacke, E. M. (1998). Improved parametric reconstruction using variable projection optimization. *Inverse problems*, 14(1):19.

[22] Bobman, S. A., Riederer, S. J., Lee, J. N., Suddarth, S. A., Wang, H. Z., Drayer, B. P., and MacFall, J. R. (1985). Cerebral magnetic resonance image synthesis. *American Journal of Neuroradiology*, 6(2):265–269.

[23] Bojorquez, J. Z., Bricq, S., Acquitter, C., Brunotte, F., Walker, P. M., and Lalande, A. (2017). What are normal relaxation times of tissues at 3 T? *Magnetic resonance imaging*, 35:69–80.

[24] Booij, J., de Bie, R. M., Beudel, M., van de Giessen, E., and Wallert, E. D. (2021). Diagnostic accuracy of MRI for neurodegenerative parkinsonism. `https://www.ccmo.nl/binaries/ccmo/documenten/jaarverslagen/2023/05/09/jaarverslag-metc-amsterdam-umc-2022/TC-09B+Jaarverslag+2022+METC+Amsterdam+UMC.pdf`. Protocol ID NL79240.018.21.

[25] Bookwalter, C. A., Griswold, M. A., and Duerk, J. L. (2009). Multiple overlapping k-space junctions for investigating translating objects (MOJITO). *IEEE Transactions on Medical Imaging*, 29(2):339–349.

[26] Bookwalter, C. A., Griswold, M. A., Sunshine, J. L., and Duerk, J. L. (2007). Analysis of signal-to-noise behavior in Cartesian continuous sampling sequences: Predictions and experimental validation of opportunities for improved image SNR. *Magnetic Resonance in Medicine*, 58(4):819–824.

[27] Box, G. E. P. and Lucas, H. (1959). Design of experiments in non-linear situations. *Biometrika*, 46(1/2):77–90.

[28] Breger, R. K., Rimm, A. A., Fischer, M. E., Papke, R. A., and Haughton, V. M. (1989). T1 and T2 measurements on a 1.5-T commercial MR imager. *Radiology*, 171(1):273–276.

[29] Brown, R. W., Cheng, Y.-C. N., Haacke, E. M., Thompson, M. R., and Venkatesan, R. (2014). *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons.

[30] Bruijnen, T., Stemkens, B., Berg, C. A. T., and Tijssen, R. H. N. (2019). Prospective GIRF-based RF phase cycling to reduce eddy current-induced steady-state disruption in bSSFP imaging. *Magnetic Resonance in Medicine*, page mrm.28097.

[31] Buehrer, M., Pruessmann, K. P., Boesiger, P., and Kozerke, S. (2007). Array compression for MRI with large coil arrays. *Magnetic Resonance in Medicine*, 57(6):1131–1139.

[32] Buonincontri, G. and Sawiak, S. J. (2016). MR fingerprinting with simultaneous B1 estimation. *Magnetic Resonance in Medicine*, 76(4):1127–1135.

[33] Castillo-Passi, C., Coronado, R., Varela-Mattatall, G., Alberola-López, C., Botnar, R., and Irarrazaval, P. (2023). KomaMRI.jl: An Open-Source Framework for General MRI Simulations with GPU Acceleration. *Magnetic Resonance in Medicine*, 90(1):29–342.

[34] Charpentier, I. (2001). Checkpointing schemes for adjoint codes: Application to the meteorological model Meso-NH. *SIAM Journal on Scientific Computing*, 22(6):2135–2151.

[35] Choi, J. Y., Hu, S., Su, T.-Y., Murakami, H., Tang, Y., Blümcke, I., Najm, I., Sakaie, K., Jones, S., Griswold, M., et al. (2023). Normative quantitative relaxation atlases for characterization of cortical regions using magnetic resonance fingerprinting. *Cerebral Cortex*, 33(7):3562–3574.

[36] Cloos, M. A., Knoll, F., Zhao, T., Block, K. T., Bruno, M., Wiggins, G. C., and Sodickson, D. K. (2016). Multiparametric imaging with heterogeneous radiofrequency fields. *Nature communications*, 7.

[37] Cohen, O., Zhu, B., and Rosen, M. S. (2018). MR fingerprinting Deep RecOnstruction NEtwork (DRONE). *Magnetic Resonance in Medicine*, 80(3):885–894.

[38] Coleman, T. F. and Li, Y. (2005). An Interior Trust Region Approach for Nonlinear Minimization Subject to Bounds. *SIAM Journal on Optimization*, 6(2):418–445.

[39] Coleman, T. F. and Moré, J. J. (1984). Estimation of sparse hessian matrices and graph coloring problems. *Mathematical Programming*, 28(3):243–270.

[40] Corradini, S., Alongi, F., Andratschke, N., Belka, C., Boldrini, L., Cellini, F., Debus, J., Guckenberger, M., Hörner-Rieber, J., Lagerwaard, F., et al. (2019). MR-guidance in clinical reality: current treatment challenges and future perspectives. *Radiation Oncology*, 14(1):1–12.

[41] Coxon, R., Brookes, M. J., Francis, S. T., Bowtell, R. W., Morris, P. E., Wright, P. J., Gowland, P. A., Mougin, O. E., Totman, J. J., Peters, A. M., and Clemence, M. (2008). Water proton T 1 measurements in brain tissue at 7, 3, and 1.5T using IR-EPI, IR-TSE, and MPRAGE: results and optimization. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 21(1-2):121–130.

[42] Crawley, A. P. and Henkelman, R. M. (1988). A comparison of one-shot and recovery methods in T1 imaging. *Magnetic Resonance in Medicine*, 7(1):23–34.

[43] Cruz, G., Schneider, T., Bruijnen, T., Gaspar, A. S., Botnar, R. M., and Prieto, C. (2018). Accelerated Magnetic Resonance Fingerprinting using soft-weighted key-hole (MRF-SOHO). *PloS one*, 13(8):e0201808.

[44] Davies, M., Puy, G., Vandergheynst, P., and Wiaux, Y. (2014). A Compressed Sensing Framework for Magnetic Resonance Fingerprinting. *SIAM Journal on Imaging Sciences*, 7(4):2623–2656.

[45] Deoni, S. C. (2010). Quantitative relaxometry of the brain. *Topics in Magnetic Resonance Imaging*, 21(2):101–113.

[46] Deoni, S. C., Peters, T. M., and Rutt, B. K. (2005). High-resolution T1 and T2 mapping of the brain in a clinically acceptable time with DESPOT1 and DESPOT2. *Magnetic Resonance in Medicine*, 53(1):237–241.

[47] Deoni, S. C., Rutt, B. K., and Peters, T. M. (2003). Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state. *Magnetic Resonance in Medicine*, 49(3):515–526.

[48] Dixon, L. C. W. and Price, R. C. (1989). Truncated Newton method for sparse unconstrained optimization using automatic differentiation. *Journal of Optimization Theory and Applications*, 60(2):261–275.

[49] Doneva, M., Amthor, T., Koken, P., Sommer, K., and Börnert, P. (2017). Matrix completion-based reconstruction for undersampled magnetic resonance fingerprinting data. *Magnetic Resonance Imaging*.

[50] Doneva, M., Börnert, P., Eggers, H., Stehning, C., Sénégas, J., and Mertins, A. (2010). Compressed sensing reconstruction for magnetic resonance parameter mapping. *Magnetic Resonance in Medicine*, 64(4):1114–1120.

[51] Fessler, J. and Sutton, B. (2003). Nonuniform Fast Fourier Transforms using min-max interpolation. *IEEE Transactions on Signal Processing*, 51(2):560–574.

[52] Fichtner, A. and Trampert, J. (2011). Resolution analysis in full waveform inversion. *Geophysical Journal International*, 187(3):1604–1624.

[53] Fika, P. and Mitrouli, M. (2019). Fast estimates for the diagonal of the inverse of large scale matrices appearing in applications. *Journal of Computational and Applied Mathematics*, 355:91–105.

[54] Fuderer, M., van der Heide, O., Liu, H., van den Berg, C. A. T., and Sbrizzi, A. (2022). Non-steady-state sequences for multi-parametric MRI need to be evaluated in the context of gradient-encoding. In *Proc. Intl. Soc. Mag. Reson. Med.*

[55] Fuderer, M., van der Heide, O., Liu, H., van den Berg, C. A. T., and Sbrizzi, A. (2023). Efficient performance analysis and optimization of transient-state sequences for multiparametric magnetic resonance imaging. *NMR in Biomedicine*, 36(3):e4864.

[56] Gao, W., Yang, Q., Li, X., Chen, X., Wei, X., Diao, Y., Zhang, Y., Chen, C., Guo, B., Wang, Y., et al. (2022). Synthetic MRI with quantitative mappings for identifying receptor status, proliferation rate, and molecular subtypes of breast cancer. *European Journal of Radiology*, 148:110168.

[57] Golub, G. and Pereyra, V. (2003). Separable nonlinear least squares: the variable projection method and its applications. *Inverse Problems*, 19(2):R1–R26.

[58] Golub, G. H. and Pereyra, V. (1973). The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate. *SIAM Journal on numerical analysis*, 10(2):413–432.

[59] Gómez, P. A., Cencini, M., Golbabaee, M., Schulte, R. F., Pirkl, C., Horvath, I., Fallo, G., Peretti, L., Tosetti, M., Menze, B. H., et al. (2020). Rapid three-dimensional multiparametric MRI with quantitative transient-state imaging. *Scientific reports*, 10(1):1–17.

[60] Granziera, C., Wuerfel, J., Barkhof, F., Calabrese, M., De Stefano, N., Enzinger, C., Evangelou, N., Filippi, M., Geurts, J. J., Reich, D. S., et al. (2021). Quantitative magnetic resonance imaging towards clinical application in multiple sclerosis. *Brain*, 144(5):1296–1311.

[61] Griswold, M. A., Jakob, P. M., Heidemann, R. M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., and Haase, A. (2002). Generalized Autocalibrating Partially Parallel Acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210.

[62] Gurney-Champion, O. J., Mahmood, F., van Schie, M., Julian, R., George, B., Philippens, M. E., van der Heide, U. A., Thorwarth, D., and Redalen, K. R. (2020). Quantitative imaging for radiotherapy purposes. *Radiotherapy and Oncology*, 146:66–75.

[63] Hagiwara, A., Warntjes, M., Hori, M., Andica, C., Nakazawa, M., Kumamaru, K. K., Abe, O., and Aoki, S. (2017). SyMRI of the brain: rapid quantification of relaxation rates and proton density, with synthetic MRI, automatic brain segmentation, and myelin measurement. *Investigative radiology*, 52(10):647.

[64] Hargreaves, B. (2005). Bloch Equation Simulator. `http://mrsrl.stanford.edu/~brian/blochsim/`. [Online; accessed 12-June-2023].

[65] van der Heide, O., Eijbersen, M. A., Sbrizzi, A., Luijten, P. R., and van den Berg, C. A. T. (2019a). Enhanced MR-STAT by a multi-coil reconstruction framework. In *Proc. Intl. Soc. Mag. Reson. Med.*, page 2414.

[66] van der Heide, O., Sbrizzi, A., and van den Berg, C. A. T. (2020a). Accelerated MR-STAT reconstructions using sparse Hessian approximations. *IEEE Transactions on Medical Imaging*.

[67] van der Heide, O., Sbrizzi, A., and van den Berg, C. A. T. (2023). Cartesian vs radial MR-STAT: An efficiency and robustness study. *Magnetic Resonance Imaging*, 99:7–19.

[68] van der Heide, O., Sbrizzi, A., Luijten, P. R., and van den Berg, C. A. T. (2019b). Sparse MR-STAT: Order of magnitude acceleration in reconstruction times. In *Proc. Intl. Soc. Mag. Reson. Med.*, page 4538.

[69] van der Heide, O., Sbrizzi, A., Luijten, P. R., and van den Berg, C. A. T. (2020b). High resolution in-vivo MR-STAT using a matrix-free and parallelized reconstruction algorithm. *NMR in Biomedicine*, 33(4):e4251.

[70] Heldens, S., Hijma, P., Van Werkhoven, B., Maassen, J., and Van Nieuwpoort, R. V. (2022). Lightning: Scaling the GPU Programming Model Beyond a Single GPU. In *2022 IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, pages 492–503.

[71] Hennig, J. (1991a). Echoes—how to generate, recognize, use or avoid them in MR-imaging sequences. Part I: Fundamental and not so fundamental properties of spin echoes. *Concepts in Magnetic Resonance*, 3(3):125–143.

[72] Hennig, J. (1991b). Echoes—how to generate, recognize, use or avoid them in MR-imaging sequences. Part II: Echoes in imaging sequences. *Concepts in Magnetic Resonance*, 3(4):179–192.

[73] Hennig, J., Welz, A. M., Schultz, G., Korvink, J., Liu, Z., Speck, O., and Zaitsev, M. (2008). Parallel imaging in non-bijective, curvilinear magnetic field gradients: a concept study. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 21(1):5–14.

[74] Hernando, D., Haldar, J., Sutton, B., Ma, J., Kellman, P., and Liang, Z.-P. (2008). Joint estimation of water/fat images and field inhomogeneity map. *Magnetic Resonance in Medicine*, 59(3):571–580.

[75] Hidalgo-Tobon, S. S. (2010). Theory of gradient coil design methods for magnetic resonance imaging. *Concepts in Magnetic Resonance Part A*, 36(4):223–242.

[76] Hilbert, T., Sumpf, T. J., Weiland, E., Frahm, J., Thiran, J.-P., Meuli, R., Kober, T., and Krueger, G. (2018). Accelerated T2 mapping combining parallel MRI and model-based reconstruction: GRAPPATINI. *Journal of Magnetic Resonance Imaging*, 48(2):359–368.

[77] Hilbert, T., Xia, D., Block, K. T., Yu, Z., Lattanzi, R., Sodickson, D. K., Kober, T., and Cloos, M. A. (2020). Magnetization transfer in magnetic resonance fingerprinting. *Magnetic Resonance in Medicine*, 84(1):128–141.

[78] Hsieh, J. J. and Svalbe, I. (2020). Magnetic resonance fingerprinting: from evolution to clinical applications. *Journal of Medical Radiation Sciences*, 67(4):333–344.

[79] Huang, C., Graff, C. G., Clarkson, E. W., Bilgin, A., and Altbach, M. I. (2012). T2 mapping from highly undersampled data by reconstruction of principal component coefficient maps using compressed sensing. *Magnetic Resonance in Medicine*, 67(5):1355–1366.

[80] Huang, H., Yang, Q., Wang, J., Zhang, P., Cai, S., and Cai, C. (2023). High-efficient Bloch simulation of magnetic resonance imaging sequences based on deep learning. *Physics in Medicine & Biology*, 68(8):085002.

[81] Jacquez, J. A. and Greif, P. (1985). Numerical parameter identifiability and estimability: Integrating identifiability, estimability, and optimal sampling design. *Mathematical Biosciences*, 77(1-2):201–227.

[82] Jaynes, E. T. (1955). Matrix treatment of nuclear induction. *Physical Review*, 98(4):1099–1105.

[83] Jeff Bezanson, S. K. (2022). Julia Micro-Benchmarks.

[84] Jiang, Y., Ma, D., Seiberlich, N., Gulani, V., and Griswold, M. A. (2015a). MR fingerprinting using fast imaging with steady state precession (FISP) with spiral readout. *Magnetic Resonance in Medicine*, 74(6):1621–1631.

[85] Jiang, Y., Ma, D., Seiberlich, N., Gulani, V., and Griswold, M. A. (2015b). MR fingerprinting using fast imaging with steady state precession (FISP) with spiral readout. *Magnetic Resonance in Medicine*, 74(6):1621–1631.

[86] K Mogensen, P. and N Riseth, A. (2018). Optim: A mathematical optimization package for Julia. *Journal of Open Source Software*, 3(24):615.

[87] Kikinis, R. (2012). Website of the National Alliance for Medical Image Computing. `http://wiki.na-mic.org/Wiki/index.php/Projects:QuantitativeSusceptibilityMapping`. [Online; accessed 12-June-2023].

[88] Kleinloog, J. P., Mandija, S., D'Agata, F., Liu, H., van der Heide, O., Koktas, B., Dankbaar, J. W., Keil, V. C., Vonken, E.-J., Jacobs, S. M., et al. (2022). Synthetic MRI with Magnetic Resonance Spin TomogrAphy in Time-Domain (MR-STAT): Results from a Prospective Cross-Sectional Clinical Trial. *Journal of Magnetic Resonance Imaging*.

[89] Knoll, F., Raya, J. G., Halloran, R. O., Baete, S., Sigmund, E., Bammer, R., Block, T., Otazo, R., and Sodickson, D. K. (2015). A model-based reconstruction for undersampled radial spin-echo DTI with variational penalties on the diffusion tensor. *NMR in Biomedicine*, 28(3):353–366.

[90] Komodakis, N. and Pesquet, J.-C. (2015). Playing with duality: An overview of recent primal-dual approaches for solving large-scale optimization problems. *IEEE Signal Processing Magazine*, 32(6):31–54.

[91] Konar, A. S., Paudyal, R., Shah, A. D., Fung, M., Banerjee, S., Dave, A., Lee, N., Hatzoglou, V., and Shukla-Dave, A. (2022a). Qualitative and Quantitative Performance of Magnetic Resonance Image Compilation (MAGiC) Method: An Exploratory Analysis for Head and Neck Imaging. *Cancers*, 14(15):3624.

[92] Konar, A. S., Shah, A. D., Paudyal, R., Fung, M., Banerjee, S., Dave, A., Hatzoglou, V., and Shukla-Dave, A. (2022b). Quantitative Synthetic Magnetic Resonance Imaging for Brain Metastases: A Feasibility Study. *Cancers*, 14(11):2651.

[93] Koolstra, K., Beenakker, J. M., Koken, P., Webb, A., and Börnert, P. (2019). Cartesian MR fingerprinting in the eye at 7T using compressed sensing and matrix completion-based reconstructions. *Magnetic Resonance in Medicine*, 81(4):2551–2565.

[94] Körzdörfer, G., Jiang, Y., Speier, P., Pang, J., Ma, D., Pfeuffer, J., Hensel, B., Gulani, V., Griswold, M., and Nittka, M. (2019a). Magnetic resonance field fingerprinting. *Magnetic Resonance in Medicine*, 81(4):2347–2359.

[95] Körzdörfer, G., Kirsch, R., Liu, K., Pfeuffer, J., Hensel, B., Jiang, Y., Ma, D., Gratz, M., Bär, P., Bogner, W., et al. (2019b). Reproducibility and repeatability of MR fingerprinting relaxometry in the human brain. *Radiology*, 292(2):429–437.

[96] Kose, R. and Kose, K. (2017). BlochSolver: A GPU-optimized fast 3D MRI simulator for experimentally compatible pulse sequences. *Journal of Magnetic Resonance*, 281:51–65.

[97] Lebel, R. M. and Wilman, A. H. (2010). Transverse relaxometry with stimulated echo compensation. *Magnetic Resonance in Medicine*, 64(4):1005–1014.

[98] Lee, J. H., Hargreaves, B. A., Hu, B. S., and Nishimura, D. G. (2003). Fast 3D imaging using variable-density spiral trajectories with applications to limb perfusion. *Magnetic Resonance in Medicine*, 50(6):1276–1285.

[99] Li, X. T. and Huang, R. Y. (2020). Standardization of imaging methods for machine learning in neuro-oncology. *Neuro-oncology Advances*, 2(Supplement_4):iv49–iv55.

[100] Lisanti, C., Carlin, C., Banks, K. P., and Wang, D. (2007). Normal MRI appearance and motion-related phenomena of CSF. *American Journal of Roentgenology*, 188(3):716–725.

[101] Liu, B., King, K., Steckner, M., Xie, J., Sheng, J., and Ying, L. (2009). Regularized sensitivity encoding (SENSE) reconstruction using Bregman iterations. *Magnetic Resonance in Medicine*, 61(1):145–152.

[102] Liu, F., Velikina, J. V., Block, W. F., Kijowski, R., and Samsonov, A. A. (2017). Fast Realistic MRI Simulations Based on Generalized Multi-Pool Exchange Tissue Model. *IEEE Transactions on Medical Imaging*, 36(2):527–537.

[103] Liu, H., van der Heide, O., van den Berg, C. A. T., and Sbrizzi, A. (2020). Accelerated MR-STAT Algorithm: Achieving 10-minute High-Resolution Reconstructions on a Desktop PC. In *Proc. Intl. Soc. Mag. Reson. Med.*

[104] Liu, H., van der Heide, O., van den Berg, C. A. T., and Sbrizzi, A. (2021). Fast and accurate modeling of transient-state, gradient-spoiled sequences by recurrent neural networks. *NMR in Biomedicine*, 34(7):e4527.

[105] Liu, H., van der Heide, O., Fuderer, M., van den Berg, C. A. T., and Sbrizzi, A. (2022a). 3D MR-STAT: Towards a fast multi-parametric protocol with increased snr. In *Proc Intl Soc Mag Reson Med*, volume 30, page 1348.

[106] Liu, H., van der Heide, O., Mandija, S., van den Berg, C. A. T., and Sbrizzi, A. (2022b). Acceleration strategies for MR-STAT: Achieving high-resolution reconstructions on a desktop pc within 3 minutes. *IEEE Transactions on Medical Imaging*, 41(10):2681–2692.

[107] Liu, H., van der Heide, O., Versteeg, E., Froeling, M., Fuderer, M., Xu, F., van den Berg, C. A. T., and Sbrizzi, A. (In-press). A three-dimensional MR-STAT protocol for high-resolution multi-parametric quantitative MRI. *NMR in Biomedicine*, page e5050.

[108] Look, D. C. and Locker, D. R. (1970). Time saving in measurement of NMR and EPR relaxation times. *Review of Scientific Instruments*, 41(2):250–251.

[109] Lustig, M., Donoho, D., and Pauly, J. M. (2007). Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*.

[110] Ma, D., Gulani, V., Seiberlich, N., Liu, K., Sunshine, J. L., Duerk, J. L., and Griswold, M. A. (2013). Magnetic Resonance Fingerprinting. *Nature*, 495:187–192.

[111] Mansfield, P. (1977). Multi-planar image formation using NMR spin echoes. *Journal of Physics C: Solid State Physics*, 10(3):L55.

[112] Margaret Cheng, H.-L., Stikov, N., Ghugre, N. R., and Wright, G. A. (2012). Practical medical applications of quantitative MR relaxometry. *Journal of Magnetic Resonance Imaging*, 36(4):805–824.

[113] MATLAB (2015). *version 8.5.0 (R2015a)*. The MathWorks Inc., Natick, Massachusetts.

[114] McGivney, D. F., Pierre, E., Ma, D., Jiang, Y., Saybasili, H., Gulani, V., and Griswold, M. A. (2014). SVD Compression for Magnetic Resonance Fingerprinting in the Time Domain. *IEEE Transactions on Medical Imaging*, 33(12):2311–2322.

[115] McGivney, Debra F and Boyaciouglu, Rasim and Jiang, Yun and Poorman, Megan E and Seiberlich, Nicole and Gulani, Vikas and Keenan, Kathryn E and Griswold, Mark A and Ma, Dan (2020). Magnetic resonance fingerprinting review part 2: Technique and directions. *Journal of Magnetic Resonance Imaging*, 51(4):993–1007.

[116] Meiboom, S. and Gill, D. (1958). Modified spin-echo method for measuring nuclear relaxation times. *Review of Scientific Instruments*, 29(8):688–691.

[117] Monga, V., Li, Y., and Eldar, Y. C. (2021). Algorithm unrolling: Interpretable, efficient deep learning for signal and image processing. *IEEE Signal Processing Magazine*, 38(2):18–44.

[118] Naumov, M., Chien, L., Vandermersch, P., and Kapasi, U. (2010). Cusparse library. In *GPU Technology Conference*.

[119] Nocedal, J. (2006). Updating Quasi-Newton Matrices with Limited Storage. *Mathematics of Computation*, 35(151):773.

[120] Nocedal, J. and Wright, S. (2006). *Numerical Optimization 2nd Ed.* Springer.

[121] Noll, D. C., Meyer, C. H., Pauly, J. M., Nishimura, D. G., and Macovski, A. (1991). A Homogeneity Correction Method for Magnetic Resonance Imaging with Time-Variying Gradients. *IEEE Transactions on Medical Imaging*.

[122] Olafsson, V. T., Noll, D. C., and Fessier, J. A. (2008). Fast joint reconstruction of dynamic R2* and field maps in functional MRI. *IEEE Transactions on Medical Imaging*.

[123] Olsson, H., Andersen, M., and Helms, G. (2020). Reducing bias in DREAM flip angle mapping in human brain at 7T by multiple preparation flip angles. *Magnetic Resonance Imaging*, 72:71–77.

[124] Oppelt, A., Graumann, R., Barfuss, H., Fischer, H., Hartl, W., Schajor, W., et al. (1986). FISP—a new fast MRI sequence. *Electromedica*, 54(1):15–18.

[125] O'leary, D. P. and Rust, B. W. (2013). Variable projection for nonlinear least squares problems. *Computational Optimization and Applications*, pages 1–15.

[126] Paige, C. C. and Saunders, M. A. (1982). LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software (TOMS)*, 8(1):43–71.

[127] Pedro Teixeira, R. A., Malik, S. J., and Hajnal, J. V. (2018). Magnetic Resonance in Medicine Fast quantitative MRI using controlled saturation magnetization transfer. *Magnetic Resonance in Medicine*, 81(2):907–920.

[128] Peng, X., Liu, X., Zheng, H., and Liang, D. (2014). Exploiting parameter sparsity in model-based reconstruction to accelerate proton density and T2 mapping. *Medical engineering & physics*, 36(11):1428–1435.

[129] Peng, X., Ying, L., Liu, Y., Yuan, J., Liu, X., and Liang, D. (2016). Accelerated exponential parameterization of T2 relaxation with model-driven low rank and sparsity priors (MORASA). *Magnetic Resonance in Medicine*, 76(6):1865–1878.

[130] Petzschner, F. H., Ponce, I. P., Blaimer, M., Jakob, P. M., and Breuer, F. A. (2011). Fast MR parameter mapping using k-t principal component analysis. *Magnetic Resonance in Medicine*, 66(3):706–716.

[131] Pezzotti, N., Yousefi, S., Elmahdy, M. S., Van Gemert, J. H. F., Schuelke, C., Doneva, M., Nielsen, T., Kastryulin, S., Lelieveldt, B. P., Van Osch, M. J., et al. (2020). An adaptive intelligence algorithm for undersampled knee MRI reconstruction. *IEEE Access*, 8:204825–204838.

[132] Pierre, E. Y., Ma, D., Chen, Y., Badve, C., and Griswold, M. A. (2016). Multiscale reconstruction for mr fingerprinting. *Magnetic Resonance in Medicine*, 75(6):2481–2492.

[133] Powell, M. J. D. and Toint, P. L. (1979). On the Estimation of Sparse Hessian Matrices. *SIAM Journal on Numerical Analysis*.

[134] Pruessmann, K. P., Weiger, M., Börnert, P., and Boesiger, P. (2001). Advances in sensitivity encoding with arbitrary k-space trajectories. *Magnetic Resonance in Medicine*, 46(4):638–651.

[135] Pruessmann, K. P., Weiger, M., Scheidegger, M. B., and Boesiger, P. (1999). SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*.

[136] Ravishankar, S. and Bresler, Y. (2011). MR Image Reconstruction From Highly Undersampled k-Space Data by Dictionary Learning. *IEEE Transactions on Medical Imaging*, 30(5):1028–1041.

[137] Robson, M. D., Gore, J. C., and Constable, R. T. (1997). Measurement of the point spread function in MRI using constant time imaging. *Magnetic Resonance in Medicine*, 38(5):733–740.

[138] Ryu, K., Nam, Y., Gho, S.-M., Jang, J., Lee, H.-J., Cha, J., Baek, H. J., Park, J., and Kim, D.-H. (2019). Data-driven synthetic MRI FLAIR artifact correction via deep neural network. *Journal of Magnetic Resonance Imaging*, 50(5):1413–1423.

[139] Sandino, C. M., Cheng, J. Y., Chen, F., Mardani, M., Pauly, J. M., and Vasanawala, S. S. (2020). Compressed sensing: From research to clinical practice with deep neural networks: Shortening scan times for magnetic resonance imaging. *IEEE signal processing magazine*, 37(1):117–127.

[140] Sbrizzi, A., Bruijnen, T., van der Heide, O., Luijten, P., and van den Berg, C. A. T. (2017). Dictionary-free MR Fingerprinting reconstruction of balanced-GRE sequences.

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

[142] Scheffler, K. and Hennig, J. (2003). Is TrueFISP a gradient-echo or a spin-echo sequence? *Magnetic Resonance in Medicine*, 49(2):395–7.

[143] Sharafi, A., Zibetti, M. V., Chang, G., Cloos, M., and Regatte, R. R. (2022). 3D magnetic resonance fingerprinting for rapid simultaneous T1, T2, and T1ρ volumetric mapping of human articular cartilage at 3T. *NMR in Biomedicine*, 35(12):e4800.

[144] Shcherbakova, Y., van den Berg, C. A. T., Moonen, C. T., and Bartels, L. W. (2018). PLANET: An ellipse fitting approach for simultaneous T 1 and T 2 mapping using phase-cycled balanced steady-state free precession. *Magnetic Resonance in Medicine*, 79(2):711–722.

[145] Smith, T. B. (2010). MRI artifacts and correction strategies. *Imaging in Medicine*, 2(4):445.

[146] Sodickson, D. K. and Manning, W. J. (1997). Simultaneous acquisition of spatial harmonics (SMASH): fast imaging with radiofrequency coil arrays. *Magnetic Resonance in Medicine*, 38(4):591–603.

[147] Spinnato, P., Kind, M., Le Loarer, F., Bianchi, G., Colangeli, M., Sambri, A., Ponti, F., van Langevelde, K., and Crombé, A. (2022). Soft tissue sarcomas: the role of quantitative MRI in treatment response evaluation. *Academic Radiology*, 29(7):1065–1084.

[148] Steihaug, T. (2005). The Conjugate Gradient Method and Trust Regions in Large Scale Optimization. *SIAM Journal on Numerical Analysis*, 20(3):626–637.

[149] Stöcker, T., Vahedipour, K., Pflugfelder, D., and Shah, N. J. (2010). High-performance computing MRI simulations. *Magnetic Resonance in Medicine*, 64(1):186–193.

[150] Stolk, C. C. and Sbrizzi, A. (2019). Understanding the combined effect of k-space undersampling and transient states excitation in MR Fingerprinting reconstructions. *IEEE Transactions on Medical Imaging*.

[151] Stollberger, R. and Wach, P. (1996). Imaging of the active B1 field in vivo. *Magnetic Resonance in Medicine*, 35(2):246–251.

[152] Stone, S. S., Haldar, J. P., Tsao, S. C., Hwu, W.-m. W., Liang, Z.-P., and Sutton, B. P. (2008). Accelerating advanced MRI reconstructions on GPUs. In *Proceedings of the 5th conference on Computing frontiers*, pages 261–272.

[153] Sumpf, T. J., Uecker, M., Boretius, S., and Frahm, J. (2011). Model-based nonlinear inverse reconstruction for T2 mapping using highly undersampled spin-echo MRI. *Journal of Magnetic Resonance Imaging*, 34(2):420–428.

[154] Sutton, B. P., Noll, D. C., and Fessler, J. A. (2003). Fast, iterative image reconstruction for mri in the presence of field inhomogeneities. *IEEE Transactions on Medical Imaging*, 22(2):178–188.

[155] Tamir, J. I., Uecker, M., Chen, W., Lai, P., Alley, M. T., Vasanawala, S. S., and Lustig, M. (2017). T2 shuffling: sharp, multicontrast, volumetric fast spin-echo imaging. *Magnetic Resonance in Medicine*, 77(1):180–195.

[156] Tan, H. and Meyer, C. H. (2009). Estimation of k-space trajectories in spiral MRI. *Magnetic Resonance in Medicine*, 61(6):1396–1404.

[157] Tanenbaum, L. N., Tsiouris, A. J., Johnson, A. N., Naidich, T. P., DeLano, M. C., Melhem, E. R., Quarterman, P., Parameswaran, S., Shankaranarayanan, A., Goyen, M., et al. (2017). Synthetic MRI for clinical neuroimaging: results of the magnetic resonance image compilation (MAGiC) prospective, multicenter, multireader trial. *American Journal of Neuroradiology*, 38(6):1103–1110.

[158] Teixeira, R. P. A., Malik, S. J., and Hajnal, J. V. (2018). Joint system relaxometry (JSR) and Crámer-Rao lower bound optimization of sequence parameters: A framework for enhanced precision of DESPOT T1 and T2 estimation. *Magnetic Resonance in Medicine*, 79(1):234–245.

[159] Tofts, P. (2003). *Quantitative MRI of the Brain*. John Wiley & Sons, Ltd, Chichester, UK.

[160] Torrey, H. C. (1956). Bloch equations with diffusion terms. *Physical review*, 104(3):563.

[161] Tran-Dinh, Q., Pham, N., and Nguyen, L. (2020). Stochastic Gauss-Newton algorithms for nonconvex compositional optimization. In *International Conference on Machine Learning*, pages 9572–9582. PMLR.

[162] Trzasko, J. D., Mostardi, P. M., Riederer, S. J., and Manduca, A. (2013). Estimating T1 from multichannel variable flip angle SPGR sequences. *Magnetic Resonance in Medicine*, 69(6):1787–1794.

[163] Tsai, C.-M. and Nishimura, D. G. (2000). Reduced aliasing artifacts using variable-density k-space sampling trajectories. *Magnetic Resonance in Medicine*, 43(3):452–458.

[164] Uecker, M., Lai, P., Murphy, M. J., Virtue, P., Elad, M., Pauly, J. M., Vasanawala, S. S., and Lustig, M. (2014). ESPIRiT - An eigenvalue approach to autocalibrating parallel MRI: Where SENSE meets GRAPPA. *Magnetic Resonance in Medicine*, 71(3):990–1001.

[165] van Valenberg, W. (2015). MR pulse design through optimal control and model order reduction of the Bloch equations. Master Thesis Utrecht University.

[166] van Valenberg, W., Klein, S., Vos, F. M., Koolstra, K., van Vliet, L. J., and Poot, D. H. (2019). An Efficient Method for Multi-Parameter Mapping in Quantitative MRI using B-Spline Interpolation. *IEEE Transactions on Medical Imaging*, pages 1–1.

[167] Valenberg, W. v. (2015). Radiofrequency pulse design through optimal control and model order reduction of the Bloch equation. Master's thesis, Utrecht University.

[168] Vannesjo, S. J., Haeberlin, M., Kasper, L., Pavan, M., Wilm, B. J., Barmet, C., and Pruessmann, K. P. (2013). Gradient system characterization by impulse response measurements with a dynamic field camera. *Magnetic Resonance in Medicine*, 69(2):583–593.

[169] Virieux, J. and Operto, S. (2009). An overview of full-waveform inversion in exploration geophysics. *Geophysics*, 74(6):WCC1–WCC26.

[170] Wächter, A. and Biegler, L. T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical programming*, 106(1):25–57.

[171] Wang, C. Y., Coppo, S., Mehta, B. B., Seiberlich, N., Yu, X., and Griswold, M. A. (2019). Magnetic resonance fingerprinting with quadratic RF phase for measurement of T2* simultaneously with δf , T1 , and T2. *Magnetic Resonance in Medicine*, 81(3):1849–1862.

[172] Wang, D., Ostenson, J., and Smith, D. S. (2020). snapMRF: GPU-accelerated magnetic resonance fingerprinting dictionary generation and matching using extended phase graphs. *Magnetic resonance imaging*, 66:248–256.

[173] Wang, F., Dong, Z., Reese, T. G., Rosen, B., Wald, L. L., and Setsompop, K. (2022). 3D Echo Planar Time-resolved Imaging (3D-EPTI) for ultrafast multi-parametric quantitative MRI. *NeuroImage*, 250:118963.

[174] Wang, H., Peng, H., Chang, Y., and Liang, D. (2018a). A survey of GPU-based acceleration techniques in MRI reconstructions. *Quantitative imaging in medicine and surgery*, 8(2):196.

[175] Wang, X., Roeloffs, V., Klosowski, J., Tan, Z., Voit, D., Uecker, M., and Frahm, J. (2018b). Model-based T1 mapping with sparsity constraints using single-shot inversion-recovery radial FLASH. *Magnetic Resonance in Medicine*, 79(2):730–740.

[176] Wang, Z. and Wen, B. (2022). Stochastic Gauss–Newton algorithm with STORM estimators for nonconvex composite optimization. *Journal of Applied Mathematics and Computing*, pages 1–23.

[177] Weigel, M. (2015). Extended phase graphs: dephasing, RF pulses, and echoes-pure and simple. *Journal of Magnetic Resonance Imaging*, 41(2):266–295.

[178] Weiskopf, N., Suckling, J., Williams, G., Correia, M. M., Inkster, B., Tait, R., Ooi, C., Bullmore, E. T., and Lutti, A. (2013). Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center validation. *Frontiers in neuroscience*, 7:95.

[179] Wengert, R. E. (2002). A simple automatic derivative evaluation program. *Communications of the ACM*, 7(8):463–464.

[180] van Werkhoven, B. (2019). Kernel Tuner: A search-optimizing GPU code auto-tuner. *Future Generation Computer Systems*, 90:347–358.

[181] West, D. J., Cruz, G., Teixeira, R. P., Schneider, T., Tournier, J.-D., Hajnal, J. V., Prieto, C., and Malik, S. J. (2022). An MR fingerprinting approach for quantitative inhomogeneous magnetization transfer imaging. *Magnetic Resonance in Medicine*, 87(1):220–235.

[182] Winkelmann, S., Schaeffter, T., Eggers, H., and Doessel, O. (2005). SNR enhancement in radial SSFP imaging using partial k-space averaging. *IEEE Transactions on Medical Imaging*, 24(2):254–262.

[183] Xanthis, C. G., Venetis, I. E., Chalkias, A. V., and Aletras, A. H. (2014). MRISIMUL: A GPU-based parallel approach to MRI simulations. *IEEE Transactions on Medical Imaging*, 33(3):607–617.

[184] Xiao, Z. and Vien, A. (2004). Experimental designs for precise parameter estimation for non-linear models. *Minerals Engineering*, 17(3):431–436.

[185] Xu, F., Mandija, S., Kleinloog, J. P. D., Liu, H., van der Heide, O., van der Kolk, A., van den Berg, C. A. T., and Sbrizzi, A. (2023). Fast and Accurate Synthetization of FLAIR images by MR-STAT/synth. In *Proceedings 15th Annual Meeting Benelux Chapter of International Society of Magnetic Resonance in Medicine*.

[186] Yang, M., Jiang, Y., Ma, D., Mehta, B., and Griswold, M. (2018). Game of Learning Bloch Equation Simulations for MR Fingerprinting. ISMRM.

[187] Young, P. N., Estarellas, M., Coomans, E., Srikrishna, M., Beaumont, H., Maass, A., Venkataraman, A. V., Lissaman, R., Jiménez, D., Betts, M. J., et al. (2020). Imaging biomarkers in neurodegeneration: current and future practices. *Alzheimer's research & therapy*, 12(1):1–17.

[188] Zhang, Y., Brady, M., and Smith, S. (2001). Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Transactions on Medical Imaging*, 20(1):45–57.

[189] Zhao, B., Lam, F., and Liang, Z. P. (2014). Model-based MR parameter mapping with sparsity constraints: Parameter estimation and performance bounds. *IEEE Transactions on Medical Imaging*, 33(9):1832–1844.

[190] Zhao, B., Lu, W., Hitchens, T. K., Lam, F., Ho, C., and Liang, Z. P. (2015). Accelerated MR parameter mapping with low-rank and sparsity constraints. *Magnetic Resonance in Medicine*, 74(2):489–498.

[191] Zhao, B., Setsompop, K., Adalsteinsson, E., Gagoski, B., Ye, H., Ma, D., Jiang, Y., Ellen Grant, P., Griswold, M. A., and Wald, L. L. (2018). Improved magnetic resonance fingerprinting reconstruction with low-rank and subspace modeling. *Magnetic Resonance in Medicine*.

[192] Zhao, B., Setsompop, K., Ye, H., Cauley, S. F., and Wald, L. L. (2016). Maximum Likelihood Reconstruction for Magnetic Resonance Fingerprinting. *IEEE Transactions on Medical Imaging*, 35(8):1812–1823.

Nederlandse Samenvatting

Magnetic Resonance Imaging ("MRI") is vandaag de dag één van de belangrijkste medische beeldvormingstechnieken. Door een combinatie van sterke magneetvelden en radiogolven kan de binnenkant van het lichaam bekeken worden zonder dat patiënten daarbij worden blootgesteld aan ioniserende straling. Door de scanner op verschillende manieren aan te sturen kan een diversiteit aan contrasten gegenereerd worden. De contrastbeelden belichten verschillen in weefseleigenschappen maar voorzien in principe niet in een daadwerkelijke kwantificatie van deze weefseleigenschappen. Kwantitatieve beelden bevatten meer informatie dan de conventionele contrastbeelden en kunnen longitudinale alsmede grootschalige, multi-center patiëntenonderzoeken faciliteren. Bestaande methoden om weefseleigenschappen te kwantificeren volgen een tweestaps procedure waarbij per weefseleigenschap eerst meerdere contrastbeelden worden gegenereerd, gevolgd door het fitten van relatief eenvoudige signaalmodellen in iedere voxel afzonderlijk. Hoewel de reconstructies doorgaans snel uitgevoerd kunnen worden zorgen de lange acquisitietijden ervoor dat dergelijke kwantitatieve MRI methoden slechts zeer beperkt worden toegepast in de klinische praktijk.

Magnetic Resonance Spin Tomography in Time-domain (MR-STAT) is een nieuwe techniek binnen de MRI waarmee meerdere weefseleigenschappen kunnen worden gekwantifieerd op basis van één scan van enkele minuten. Anders dan bij de twee-staps methoden worden geen contrastbeelden gegenereerd maar worden de weefseleigenscahppen gelijktijdig spatieel gelocaliseerd en gekwantificeerd. Het reconstructieproces binnen MR-STAT bestaat uit het oplossen van een grootschalig, niet-linear optimalisatieprobleem in alle voxels tegelijkertijd en brengt rekenkundige uitdagingen met zich mee. In dit proefschrift wordt de MR-STAT techniek verder ontwikkeld op zowel het gebied van dataacquisitie als beeldreconstructie om een klinische demonstratie van de techniek in praktische zin mogelijk te maken.

In Hoofdstuk 2 wordt de theorie beschreven die ten grondslag ligt aan MR-

STAT. Middels in silico en lage-resolutie in vivo experimenten wordt de potentie van de techniek aangetoond. Voor de acquisitie wordt gebruik gemaakt van Cartesische uitleestrajecten in combinatie met variabele fliphoeken. Voor de reconstructies in dit hoofdstuk wordt gebruik gemaakt van een ontkoppelingsmethode gebaseerd op een ééndimensionale Fourier transformatie. Voor ieder van de resulterende deelproblemen wordt het variabelen projectie-algoritme gebruikt. Deze hybride reconstructiemethodiek beperkt MR-STAT echter tot tweedimensionale, Cartesische acquisities.

In Hoofdstuk 3 wordt een meer generieke reconstructiemethodiek voor MR-STAT geintroduceerd welke gebaseerd is op een Gauss-Newton algoritme. Er wordt niet langer gebruik gemaakt van een ontkoppelingsmechanisme waardoor de reconstructiemethodiek in theorie ook generaliseert naar niet-Cartesisch acquisities. Een hiermee verbonden nadeel is dat zonder ontkoppeling de in de Gauss-Newton methode benodigde modelmatrices dusdanig groot worden dat opslag in computergeheugen praktisch onmogelijk is. Opslag van deze matrices is echter niet strikt noodzakelijk: voor het algoritme is het voldoende om matrix-vector producten met deze matrices uit te kunnen rekenen. Een zogeheten matrixvrije implementatie wordt voorgesteld waarbij het tevens mogelijk is om de rekentaken in parallel uit te voeren. Communicatie tussen de processen is hierbij wel vereist. Het parallelle, matrixvrije algoritme wordt geimplementeerd op een CPU-rekencluster en succesvol toegepast op data uit hoge-resolutie, tweedimensionale MR-STAT acquisities. Door de matrixvrije implementatie zou de voorgestelde reconstructiemethodiek in theorie ook kunnen werken voor driedimensionale acquisities. De reconstructietijden vormen echter nog wel een belemmerende factor: met de voorgestelde methode duren zelfs tweedimensionale reconstructies al enkele uren op het rekencluster.

In de zoektocht naar technieken om de reconstructietijden omlaag te brengen wordt in Hoofdstuk 4 aangetoond dat de zogeheten Gauss-Newton matrix die wordt gebruikt in de MR-STAT-reconstructietechniek uit Hoofdstuk 3 een sequentie-afhankelijke structuur heeft. Op basis van een theorie wordt beargumenteerd dat de Gauss-Newton-matrix voor sequenties met Cartesische uitleestrajecten en geleidelijk variërende fliphoeken een (bij goede benadering) ijle structuur heeft. Dit inzicht maakt het mogelijk om de niet-nul elementen (in parallel) te berekenen en op te slaan in het computergeheugen. Met de ijle benadering kunnen matrix-vectorvermenigvuldigingen aanzienlijk sneller uitgevoerd worden dan via de matrixvrije aanpak. De theoretische voorspellingen worden gevalideerd in simulaties, alsmede fantoom- en in vivo experimenten. De reconstructietijden voor een tweedimensionale acquisitie worden hiermee gereduceerd van enkele uren tot ongeveer vijftien minuten op een CPU-rekencluster. De techniek ontwikkeld in dit hoofdstuk kan worden geïnterpreteerd als een generalisatie van de op de ééndimensionale Fourier-transformatie gebaseerde ontkoppelingsmethode die wordt gebruikt in Hoofdstuk 2. De voorgestelde techniek is voor niet-Cartesiaanse trajecten echter moeilijk te implementeren en leidt mogelijk niet tot significante verminderingen in reconstructietijden van-

wege de minder ijle matrixstructuur.

Geïnspireerd door de toegenomen beschikbaarheid van videokaarten en door de observatie van significante rekenversnellingen met videokaarten in andere onderzoeksgebieden (waaronder MRI), is besloten om in Hoofdstuk 5 te werken aan een GPU-implementatie van het MR-STAT reconstructie-algoritme. De relatief nieuwe Julia-programmeertaal blijkt hierin een waardevol instrument te zijn. Met deze hoog-niveau programmeertaal is het namelijk mogelijk om code te schrijven welke direct gecompileerd kan worden naar instructies welke op videokaarten uitgevoerd kunnen worden zonder in te leveren op rekensnelheid. De rekenkundig intensieve taken van het reconstructie-algoritme uit Hoofdstuk 3 worden middels Julia uitvoerbaar gemaakt op videokaarten. Tevens wordt een aanpassing van het reconstructie-algoritme voorgesteld. In plaats van op een volledig matrixvrije manier te werk te gaan wordt een deel van de modelmatrices (namelijk die elementen die het duurst zijn om uit te rekenen) wel degelijk opgeslagen en een ander deel (de elementen die relatief snel uitgerekend kunnen worden) wordt zoals in de matrixvrije methode behandeld. Met deze gedeeltelijk matrixvrije Gauss-Newton reconstructietechniek blijkt het mogelijk te zijn om tweedimensionale MR-STAT reconstructies in ongeveer een minuut uit te voeren op één videokaart. Een bijkomend voordeel van deze aanpak is dat de methode - in tegenstelling tot de ijle benaderingsmethode uit Hoofdstuk 4 - ook toegepast kan worden ingeval van niet-Cartesische acquisities. Een belangrijk onderdeel van de MR-STAT reconstructies bestaat uit het uitvoeren van zogeheten Bloch simulaties. Bloch simulaties zijn ook voor andere (kwantitatieve) MRI technieken relevant. De Julia code om Bloch simulaties uit te voeren (al dan niet op videokaarten) is als zelfstandig pakket genaamd BlochSimulators.jl beschikbaar gesteld op het internet. In Hoofdstuk 5 wordt aangetoond dat dit softwarepakket beter presteert in termen van rekentijden dan andere beschikbare softwarepakketten om Bloch simulaties uit te voeren.

Gebruikmakend van de in Hoofdstuk 5 ontwikkelde reconstructmethode wordt in Hoofdstuk 6 een verkenning gedaan van niet-Cartesische uitleesstrategiën voor MR-STAT. Meer specifiek wordt er een vergelijking tussen Cartesische en radiale MR-STAT gedaan in termen van efficiëntie en robuustheid. Anders dan bij conventionele, niet-kwantitatieve MRI is het niet op voorhand te voorspellen welke van de twee acquisitiemethoden tot een hogere efficiëntie leidt. In simulatie-experimenten resulteert de radiale uitleesstrategie in een hogere efficiëntie. Bij fantoom- en in vivo experimenten zijn de conclusies minder eenduidig. Dit komt doordat de radiale uitleesstrategie gepaard gaat met betrouwbaarheids- en robuustheidsuitdagingen die worden toegeschreven aan een hogere gevoeligheid voor onvolkomenheden in scanner-hardware. Voor klinische toepassingen kan de Cartesische uitleesstrategie daarom mogelijk alsnog de voorkeur genieten.

Op basis van het werk dat in dit proefschrift wordt gepresenteerd, zijn volledige multi-tweedimensionale MR-STAT-acquisities van vijf minuten mogelijk waaruit kwantitatieve afbeeldingen van T_1, T_2 en proton dichtheid kunnen wor-

den afgeschat met reconstructietijden van enkele minuten per plak. Een eerste klinische demonstratie op 30 patiënten met verschillende ziekteverschijnselen is succesvol afgerond. Op het moment van dit schrijven wordt een tweede klinische demonstratie uitgevoerd bij een grote groep patiënten die lijden aan de ziekte van Parkinson.

Publications

Journal Publications

[1] A. Sbrizzi, **O. van der Heide**, M. Cloos, A. van der Toorn, H. Hoogduin, P.R. Luijten, and C.A.T. van den Berg, Fast quantitative MRI as a nonlinear tomography problem, *Magnetic resonance imaging*, vol. 46, pp. 56–63, 2018.

[2] van der Heide, Oscar and Sbrizzi, Alessandro and Luijten, Peter R. and van den Berg, Cornelis A T, High resolution in-vivo MR-STAT using a matrix-free and parallelized reconstruction algorithm, *NMR in Biomedicine*, vol. 33, p. e4251, apr 2020.

[3] **O. van der Heide**, A. Sbrizzi, and C.A. van den Berg, Accelerated MR-STAT reconstructions using sparse Hessian approximations, *IEEE Transactions on Medical Imaging*, vol. 39, no. 11, pp. 3737–3748, 2020.

[4] T. Bruijnen, **O. van der Heide**, M.P.W. Intven, S. Mook, J.J.W. Lagendijk, C.A.T. van den Berg, and R.H.N. Tijssen, Technical feasibility of magnetic resonance fingerprinting on a 1.5 T MRI-linac, *Physics in Medicine & Biology*, vol. 65, no. 22, p. 22NT01, 2020.

[5] H. Liu, **O. van der Heide**, C.A.T. van den Berg, and A. Sbrizzi, Fast and accurate modeling of transient-state, gradient-spoiled sequences by recurrent neural networks, *NMR in Biomedicine*, vol. 34, no. 7, p. e4527, 2021.

[6] H. Liu, **O. van der Heide**, S. Mandija, C.A.T. van den Berg, and A. Sbrizzi, Acceleration strategies for MR-STAT: Achieving high-resolution reconstructions on a desktop pc within 3 minutes, *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2681–2692, 2022.

[7] O. Akdag, S. Mandija, A.L.H.M.W. van Lier, P.T.S. Borman, T. Schakel, E. Alberts, **O. van der Heide**, R.J. Hassink, J.J.C. Verhoeff, F.A.A.M. Hoesein, and others, Feasibility of cardiac-synchronized quantitative T1 and T2 mapping on a hybrid 1.5 Tesla magnetic resonance imaging and linear accelerator system, *Physics and Imaging in Radiation Oncology*, vol. 21, pp. 153–159, 2022.

[8] M. Fuderer, **O. van der Heide**, H. Liu, C.A.T. van den Berg, and A. Sbrizzi, Efficient performance analysis and optimization of transient-state sequences for multiparametric magnetic resonance imaging, *NMR in Biomedicine*, vol. 36, no. 3, p. e4864, 2023.

[9] J.P. Kleinloog, S. Mandija, F. D'Agata, H. Liu, **O. van der Heide**, B. Koktas, J.W. Dankbaar, V.C. Keil, E.J. Vonken, S.M. Jacobs, and others, Synthetic MRI with Magnetic Resonance Spin TomogrAphy in Time-Domain (MR-STAT): Results from a Prospective Cross-Sectional Clinical Trial, *Journal of Magnetic Resonance Imaging*, vol. 57, no. 5, pp. 1451–1461, 2023.

[10] **O. van der Heide**, A. Sbrizzi, and C.A.T. van den Berg, Cartesian vs radial MR-STAT: An efficiency and robustness study, *Magnetic Resonance Imaging*, vol. 99, pp. 7–19, 2023.

[11] S. Mandija, S.M. Jacobs, J.P.D. Kleinloog, H. Liu, **O. van der Heide**, A.G. Kolk, A. Sbrizzi, and C.A.T. van den Berg, Water content-based electrical properties tomography: results from a retrospective clinical study, *in submission*, 2023.

[12] **O. van der Heide**, C.A.T. van den Berg, and A. Sbrizzi, GPU-accelerated Bloch simulations and MR-STAT reconstructions using the Julia programming language, *in submission*, 2023.

[13] H. Liu, **O. van der Heide**, E. Versteeg, M. Froeling, M. Fuderer, F. Xu, C.A.T. van den Berg, and A. Sbrizzi, A three-dimensional MR-STAT protocol for high-resolution multi-parametric quantitative MRI, *NMR in Biomedicine*, p. e5050, in-press.

[14] F. Xu, S. Mandija, J.P.D. Kleinloog, H. Liu, **O. van der Heide**, A. van der Kolk, J.W. Dankbaar, C.A. van den Berg, and A. Sbrizzi, Improving the lesion appearance on FLAIR images synthetized from quantitative MRI: a fast, hybrid approach, *in submission*, 2023.

[15] M. Fuderer, H. Liu, **O. van der Heide**, C.A. van den Berg, and A. Sbrizzi, RF phase modulation improves quantitative transient state sequences under constrained conditions, *in submission*, 2023.

[16] E. Versteeg, H. Liu, **O. van der Heide**, M. Fuderer, C.A. van den Berg, and A. Sbrizzi, High SNR full brain relaxometry at 7T by accelerated MR-STAT, *in submission*, 2023.

Conference Proceedings

[1] **O. van der Heide**, A. Sbrizzi, A. Kruseman, M. Cloos, P.R. Luijten, and C.A.T. van den Berg, In-vivo Validation of MR-STAT: Simultaneous Signal Localization and Quantification of Tissue Parameters on a 3T Clinical MR-System, in *Proc Intl Soc Mag Reson Med*, p. 1790, 2017.

[2] **O. van der Heide**, A. Sbrizzi, P.R. Luijten, and C.A.T. van den Berg, High-resolution in-vivo multi-parametric MRI using MR-STAT with a highly parallelized, limited-memory reconstruction algorithm, in *Proc Intl Soc Mag Reson Med*, p. 226, 2018.

[3] **O. van der Heide**, M.A. Eijbersen, C.A.T. van den Berg, P.R. Luijten, and A. Sbrizzi, Enhanced MR-STAT by a multi-coil reconstruction framework, in *Proc Intl Soc Mag Reson Med*, p. 2414, 2019.

[4] **O. van der Heide**, A. Sbrizzi, and C.A.T. van den Berg, Sparse MR-STAT: Order of magnitude acceleration in reconstruction times, in *Proc Intl Soc Mag Reson Med*, p. 2414, 2019.

[5] **O. van der Heide**, A. Sbrizzi, T. Bruijnen, and C.A.T. van den Berg, Extension of MR-STAT to non-Cartesian and gradient-spoiled sequences, in *Proc Intl Soc Mag Reson Med*, vol. 886, 2020.

[6] M. Eijbersen, H. Liu, **O. van der Heide**, C.A.T. van den Berg, and A. Sbrizzi, Multi-parametric quantification of multiple spectral components by an extended MR-STAT framework, in *Proc Intl Soc Mag Reson Med*, p. 3228, 2020.

[7] S. Mandija, F. D'Agata, H. Liu, **O. van der Heide**, B. Koktas, C.A.T. van den Berg, J. Hendrikse, A. van der Kolk, and A. Sbrizzi, A five-minute multi-parametric high-resolution whole-brain MR-STAT exam: first results from a clinical trial, in *Proc Intl Soc Mag Reson Med*, vol. 558, 2020.

[8] H. Liu, **O. van der Heide**, C.A.T. van den Berg, and A. Sbrizzi, Accelerated MR-STAT Algorithm: Achieving 10-minute High-Resolution Reconstructions on a Desktop PC, in *Proc Intl Soc Mag Reson Med*, vol. 3477, 2020.

[9] **O. van der Heide**, A. Sbrizzi, and C.A.T. van den Berg, Faster Bloch simulations and MR-STAT reconstructions on GPU using the Julia programming language, in *Proc Intl Soc Mag Reson Med*, p. 3063, 2021.

[10] H. Liu, **O. van der Heide**, C.A.T. van den Berg, and A. Sbrizzi, Fast and Accurate Modeling of Transient-state Sequences by Recurrent Neural Networks, in *Proc Intl Soc Mag Reson Med*, p. 329, 2021.

[11] H. Liu, T. Bruijnen, M. van Haandel, **O. van der Heide**, M. Fuderer, C.A.T. van den Berg, and A. Sbrizzi, Increasing the T2 sensitivity of MR-STAT sequences by small quadratic RF phase increments, in *Proc Intl Soc Mag Reson Med*, p. 625, 2022.

[12] M. Fuderer, **O. van der Heide**, H. Liu, C.A.T. van den Berg, and A. Sbrizzi, Non-steady-state sequences for multi-parametric MRI need to be evaluated in the context of gradient-encoding, in *Proc Intl Soc Mag Reson Med*, p. 2786, 2022.

[13] J.P.D. Kleinloog, S. Mandija, F. D'Agata, **O. van der Heide**, B. Koktas, S.M. Jacobs, C.A.T. van den Berg, J. Hendrikse, A.G. van der Kolk, and A. Sbrizzi, Synthetic MRI with MR-STAT: results from a clinical trial, in *Proc Intl Soc Mag Reson Med*, p. 597, 2022.

[14] H. Liu, **O. van der Heide**, M. Fuderer, C.A.T. van den Berg, and A. Sbrizzi, 3D MR-STAT: Towards a fast multi-parametric protocol with increased snr, in *Proc Intl Soc Mag Reson Med*, p. 1348, 2022.

[15] M. Fuderer, H. Liu, **O. van der Heide**, C.A.T. van den Berg, and A. Sbrizzi, Multi-parametric quantitative MRI without inversion pulses by optimized RF phase modulation, in *Proc Intl Soc Mag Reson Med*, p. 2176, 2023.

[16] M. Fuderer, **O. van der Heide**, C.A.T. van den Berg, and A. Sbrizzi, The influence of diffusion in fast multi-parametric relaxometry, in *Proc Intl Soc Mag Reson Med*, p. 2187, 2023.

[17] **O. van der Heide**, M. Doneva, P. Koken, J. Meineke, M. Fuderer, C.A.T. van den Berg, and A. Sbrizzi, Cartesian MR-STAT vs spiral MR Fingerprinting: a comparison, in *Proc Intl Soc Mag Reson Med*, p. 2197, 2023.

[18] E. Versteeg, H. Liu, **O. van der Heide**, M. Fuderer, C.A.T. van den Berg, and A. Sbrizzi, Full brain relaxometry at 7T in 3 minutes by accelerated MR-STAT using a low-SAR flip angle train, in *Proc Intl Soc Mag Reson Med*, p. 271, 2023.

[19] H. Liu, **O. van der Heide**, E. Versteeg, M. Fuderer, F. Xu, M. Froeling, C.A.T. van den Berg, and A. Sbrizzi, High-resolution three-dimensional MR-STAT for musculoskeletal applications, in *Proc Intl Soc Mag Reson Med*, p. 672, 2023.

[20] F. Xu, H. Liu, S. Mandija, **O. van der Heide**, E. Versteeg, M. Fuderer, C.A.T. van den Berg, and A. Sbrizzi, MR-STAT for fast contrast agent uptake quantification, in *Proc Intl Soc Mag Reson Med*, p. 1106, 2023.

[21] S. Mandija, S.M. Jacobs, J.P.D. Kleinloog, H. Liu, **O. van der Heide**, A.G. Kolk, A. Sbrizzi, and C.A.T. van den Berg, MR-STAT for fast contrast agent uptake quantification, in *Proc Intl Soc Mag Reson Med*, p. 5166, 2023.

Patents

[1] **O. van der Heide**, A. Sbrizzi, and C.A.T. van den Berg, Parameter map determination for time domain magnetic resonance, May 26 2022. Pending patent applications US17/602,259US, EP20719593.4 and EP2020/059997.

Dankwoord

Nico. **Alessandro**. Toen ik na mijn afstuderen op zoek ging naar een baan leek het me leuk om samen te werken met met mensen met verschillende achtergronden. Tegelijkertijd vond ik het ook erg belangrijk - en dat vind ik nog steeds - om met de fiets naar mijn werk te kunnen. Destijds wonend in de Bilt kwam ik al snel terecht bij de vacaturebank van het UMC Utrecht. Hier stond een vacature over een nieuwe MRI techniek en het had iets met wiskunde en programmeren te maken. Het klonk interessant maar het was wel een promotieplek en daar was ik niet specifiek naar op zoek. Dat leek me namelijk nogal eenzaam en te theoretisch. Ook had ik geen kennis van MRI dus ik had sterke twijfels of reageren op de vacature wel zin zou hebben. Ik besloot het toch te doen (nog bedankt trouwens dat ik geen motivatiebrief hoefde te schrijven) en twee gesprekken verder kon ik starten aan het MR-STAT project. Nu, aan het einde van het promotietraject, kan ik zeggen dat ik geen enkel moment spijt heb gehad van de keuze om te gaan promoveren. En dat komt voor een heel groot deel door hoe jullie de begeleiding hebben aangepakt. In de beginperiode heb ik uitgebreid de tijd gekregen om mijn gebrek aan MRI kennis weg te werken. Wekelijkse groepsbespreking waarin we artikelen over specifieke MRI sequenties en technieken bestudeerden zijn daarin zeer waardevol geweest. Later heb ik van jullie de ruimte en het vertrouwen gekregen om grotendeels zelf te bepalen aan welke onderdelen van het MR-STAT project ik wilde werken. Daar waar ik om sturing vroeg heb ik die gekregen. Jullie hebben gezorgd voor een prettige balans tussen fundamentele vraagstukken onderzoeken en pragmatisch handelen om patiënten te kunnen scannen. Door jullie voortdurende inzet is het MR-STAT project gaandeweg steeds groter geworden en is het allesbehalve een eenzaam traject geweest. Bovenal zijn jullie ook gewoon een vrolijk duo. Allebei totaal verschillend, maar perfect op elkaar ingespeeld. Bedankt.

Els. **Erik**. Bedankt voor alle jaren aan onvoorwaardelijke steun. Het maakt niet uit wat en wanneer ik iets vraag, jullie schieten altijd te hulp. Gaandeweg

ben ik me steeds meer gaan realiseren hoe bijzonder dit is en dat mag ook best eens gezegd worden.

Femke, **Fenna**, **Anna** en Saartje. Met de verdediging van dit proefschrift eindigt niet alleen mijn promotietraject, er eindigt tegelijkertijd ook een turbulente periode van verhuizingen, gebroken nachten en kinderopvangpuzzels. We hebben gelukkig weer een thuis gevonden in Utrecht waar we opgelucht adem kunnen halen en met uitzicht op veel moois.

Jurjen. **Tjebbe**. Bedankt dat jullie aan mijn zijde willen staan tijdens deze - stiekem toch best spannende - levensgebeurtenis.

Verder wil ik alle leden van de promotiecommissie bedanken voor de tijd en moeite die het gekost heeft om dit proefschrift te beoordelen.

Tot slot wil ik mijn dank uitspreken aan iedereen met wie ik de afgelopen jaren heb samengewerkt, ideeën heb uitgewisseld, op reis ben geweest, heb getafelvoetbald, koffie gezet en gedronken heb, aan wie ik Julia heb opgedrongen en met wie ik heb genoten van de maaltijden in restaurant de Brink. Ik hoop dat degenen voor wie deze boodschap bedoeld is zich erkend voelen en dat de waardering wederzijds is.

Curriculum vitae

Oscar van der Heide was born in 1987 Amsterdam and moved to Wilnis at the age of three. After graduating from Veenlanden College Mijdrecht in 2006 he started studying Law at Utrecht University. He completed his Bachelor's degree in Tax Law in 2009 and his Master's degree in International Tax Law in 2011. In the meantime, in 2010, he started studying Mathematics at Utrecht University. After obtaining his Bachelor's degree in 2013 he continued with his Master's in Industrial and Applied Mathematics at the Technical University Eindhoven. For his Master's thesis project he analysed a diffusion-reaction system to model the formation of flexible solar at TNO/Solliance. Shortly after graduating he started his Phd at the University Medical Center Utrecht on the development of the MR-STAT technique. The findings of this research are outlined in the presented thesis.

Oscar will continue working at the University Medical Center Utrecht as a postdoctoral researcher.