Journal Pre-proof

Model-Based Multi-Parameter Mapping

Yaël Balbastre, Mikael Brudfors, Michela Azzarito, Christian Lambert, Martina F. Callaghan, John Ashburner

PII: S1361-8415(21)00195-X
DOI: https://doi.org/10.1016/j.media.2021.102149
Reference: MEDIMA 102149

To appear in: *Medical Image Analysis*

Received date: 2 February 2021
Revised date: 7 May 2021
Accepted date: 24 June 2021

Please cite this article as: Yaël Balbastre, Mikael Brudfors, Michela Azzarito, Christian Lambert, Martina F. Callaghan, John Ashburner, Model-Based Multi-Parameter Mapping, *Medical Image Analysis* (2021), doi: https://doi.org/10.1016/j.media.2021.102149

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 Published by Elsevier B.V.

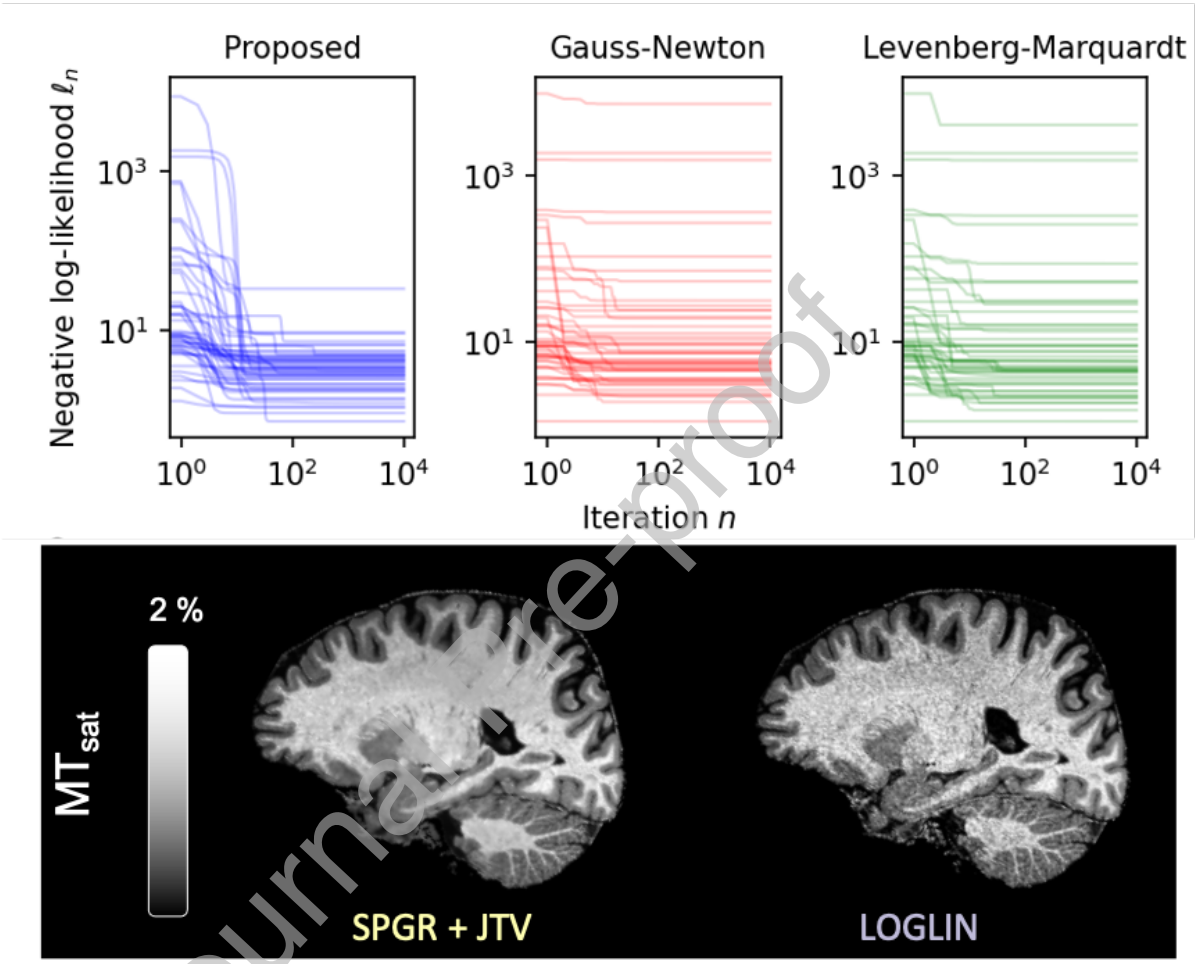
Journal Pre-proof

Highlights

- Model-based estimation of quantitative multi-parameter maps.
- Maximum-likelihood or Maximum a posteriori solutions.
- Embedded denoising using a joint total-variation prior.
- Stable second-order solver using a novel approximate Hessian

Journal Pre-proof

Graphical Abstract



Journal Pre-proof

Journal Pre-proof

Model-Based Multi-Parameter Mapping

Yaël Balbastre[a,b,*], Mikael Brudfors[a,c], Michela Azzarito[d], Christian Lambert[a], Martina F. Callaghan[a], John Ashburner[a]

[a] *Wellcome Center for Human Neuroimaging, Queen Square Institute of Neurology, University College London, London, UK*
[b] *Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, USA*
[c] *School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK*
[d] *Spinal Cord Injury Center Balgrist, University Hospital Zurich, University of Zurich, Switzerland*

Abstract

Quantitative MR imaging is increasingly favoured for its richer information content and standardised measures. However, computing quantitative parameter maps, such as those encoding longitudinal relaxation rate (R_1), apparent transverse relaxation rate (R_2^*) or magnetisation-transfer saturation ($\mathrm{MT}_{\mathrm{sat}}$), involves inverting a highly non-linear function. Many methods for deriving parameter maps assume perfect measurements and do not consider how noise is propagated through the estimation procedure, resulting in needlessly noisy maps. Instead, we propose a probabilistic generative (forward) model of the entire dataset, which is formulated and inverted to jointly recover (log) parameter maps with a well-defined probabilistic interpretation (e.g., maximum likelihood or maximum a posteriori). The second order optimisation we propose for model fitting achieves rapid and stable convergence thanks to a novel approximate Hessian. We demonstrate the utility of our flexible framework in the context of recovering more accurate maps from data acquired using the popular multi-parameter mapping protocol. We also show how to incorporate a joint total variation prior to further decrease the noise in the maps, noting that the probabilistic formulation allows the uncertainty on the recovered parameter maps to be estimated. Our implementation uses a PyTorch backend and benefits from GPU acceleration. It is available at https://github.com/balbasty/nitorch.

1. Introduction

The magnetic resonance imaging (MRI) signal is governed by a number of tissue-specific parameters. While many common MR sequences only aim to maximise the contrast between tissues of interest, the field of quantitative MRI (qMRI) is concerned with the extraction of the original parameters (Tofts, 2003). This interest stems from the fundamental relationship that exists between the magnetic parameters and the tissue microstructure: the longitudinal relaxation rate $R_1 = 1/T_1$ is sensitive to myelin content (Sigalovsky et al., 2006; Dick et al., 2012; Sereno et al., 2013; Lutti et al., 2014); the apparent transverse relaxation rate $R_2^* = 1/T_2^*$ can be used to probe iron content (Ordidge et al., 1994; Ogg et al., 1999; Hasan et al., 2012; Langkammer et al., 2010); the magnetization-transfer saturation ratio ($\mathrm{MT}_{\mathrm{sat}}$) is related to protons bound to macromolecules (in contrast to free water) and offers another metric to investigate myelination (Tofts et al., 2003; Helms et al., 2008b; Campbell et al., 2018). Additionally, quantitative MRI allows many of the scanner- and centre-specific effects to be factored out, making measures more comparable across sites (Tofts et al., 2006; Deoni et al., 2008b; Bauer et al., 2010; Weiskopf et al., 2013).

Multi-parameter mapping

Estimating quantitative parameters involves acquiring a collection of MR images while varying the sequence parameters (repetition time, echo time, flip angle) that are

*Corresponding author

Journal Pre-proof

known to interact with the quantitative parameters, so that the function that governs the weighted signal intensity can be inverted. This function emerges from the integration of ordinary differential equations (Bloch, 1946) and is therefore nonlinear and can be difficult to invert in closed-form. Therefore, early attempts at quantitative MRI targeted a single parameter at a time and used sequences with parameters carefully chosen so that other terms would disappear with a few algebraic manipulations (Gupta, 1977). However, as quantitative mapping moved from pure MR research to clinical applications, it became important to design sequences that allowed a maximum number of parameters to be estimated from a minimum number of acquisitions.

The multi-parameter mapping (MPM) protocol stems from this need: it uses multi-echo spoiled gradient echo (SPGR) acquisitions with variable flip angles and allows the quantification of R_1, R_2^*, $\mathrm{MT}_{\mathrm{sat}}$ and proton density (PD) at submillimetric resolution and in a clinically acceptable scan time (Weiskopf et al., 2013). In this context, a wider range of sequence parameters can be used, at the price of rational approximations of the signal (Helms et al., 2008b). Images acquired with a SPGR sequence are parameterised by three parameters (PD, R_1, R_2^*) plus an eventual fourth one if MT weighting is used. The ESTATICS forward model reparameterises the SPGR signal equation using one *intercept* per contrast (*i.e.*, weighted volumes extrapolated to the theoretical *zero* echo time) and a shared R_2^* decay (Weiskopf et al., 2014; Mohammadi et al., 2017). While ESTATICS provides a model-based estimation of R_2^*, the intercepts remain weighted images and another model must be inverted to recover the quantitative parameter maps: R_1, PD and $\mathrm{MT}_{\mathrm{sat}}$[1]. While this inversion has an analytical solution when all contrasts have the same repetition time (Wang et al., 1987; Liberman et al., 2014; Mohammadi et al., 2017), rational approximations must be used as soon as different repetition times are used (Helms et al., 2008a).

Non-linear fitting in MR relaxometry

Many quantitative mapping methods discard the probabilistic nature of the acquired signal, which randomly deviates from its theoretical value due to thermal and physiological noise. This noise propagates through every computational step and ends up in the maps with a non-trivial distribution (Polders et al., 2012; Cercignani and Alexander, 2006) and can even induce bias in the estimators (Sijbers et al., 1999; Chang et al., 2008). An alternative is to estimate well defined probabilistic quantities *conditioned* on the observed data, such as posterior distributions or maximum-likelihood estimators (Hurley et al., 2012; Tisdall and van der Kouwe, 2016) – although they can still be biased (Tisdall, 2017). The use of qMRI estimators that directly optimise a probability distribution have remained elusive, in part due to the non-linear and non-convex nature of the corresponding log-likelihood. Still, several groups have proposed to fit quantitative maps using iterative methods, either from magnitude images (Hurley et al., 2012; Tisdall, 2017) or directly from k-space measurements (Chen et al., 1998; Sumpf et al., 2011, 2014; Zhao et al., 2014). In order to solve the non-linear system, Tisdall (2017) used Brent's method, which does not require derivatives; Chen et al. (1998) used a Levenberg-Marquardt algorithm, Sumpf et al. (2011, 2014) used a conjugate gradient algorithm, while Zhao et al. (2014) used a quasi-Newton Broyden-Flecher-Goldfarb-Shannon (BFGS) optimiser, which is not ensured to converge in general (Mascarenhas, 2014) and is highly sensitive to initialisation. Scholand et al. (2020) pushed this idea further by modelling the entire Bloch equations and spatial encoding. In their work, the problem is solved using Gauss-Newton optimisation (which corresponds to Fisher's scoring of the Hessian). However, for this much more complex inverse-problem to be well-posed, spiral trajectories are needed so that each shot acquires as many spatial frequencies as possible. In multi-exponential T_2 mapping, the most common procedure involves non-negative least-squares fits (Whittall and MacKay, 1989), which can be very sensitive to noise and optimisation parameters (Wiggermann et al., 2020). The use of non-linear least-squares algorithms is also common and revolves around (constrained) Gauss-Newton or Levenberg-Marquardt (Milford et al., 2015; Gong et al., 2020).

Denoising

The push to higher resolutions and the use of parallel imaging reduces the signal-to-noise ratio (SNR) of the acquired images and consequently the precision of the com-

[1] This procedure neglects noise propagation and assumes that successive quantities are exact – even though they are not.

Journal Pre-proof

puted parameter maps. Papp et al. (2016) found a scan-rescan root mean squared error of about 7.5% for R_1 at 1mm, in the absence of inter-scan movement. Smoothing can be used to improve SNR, but at the cost of lower spatial specificity. In this context, providing mapping methods that are robust to high levels of noise, both in terms of bias and variance, is of major importance.

Denoising methods aim to separate signal from noise. They take advantage of the fact that signal and noise have intrinsically different spatial profiles: thermal noise is spatially independent and often has a characteristic distribution while the signal is highly structured. Denoising methods originate from partial differential equations, adaptive filtering, variational optimisation or Markov random fields, and many connections exist between them. Two main families emerge:

1. Inversion of a generative model[2]:

$$\hat{Y} = \operatorname{argmin}_Y \mathcal{E}_l\left(X, \mathcal{F}(Y)\right) + \mathcal{E}_p\left(\mathcal{G}(Y)\right),$$

   where X is the observed data, Y is the unknown noise-free data, $\mathcal{F}$ is an arbitrary *forward* transformation (*e.g.*, spatial transformation, downsampling, smoothing) mapping from the reconstructed to the observed data and $\mathcal{G}$ is a linear transformation (*e.g.*, spatial gradients, Fourier transform, wavelet transform) that extracts features of interest from the reconstruction. $\mathcal{E}_l$ and $\mathcal{E}_p$ are two log-probability distributions that correspond to the likelihood of the parameters and their prior.
2. Application of an adaptive nonlocal filter:

$$\hat{Y}_i = \sum_{j \in \mathcal{N}_i} w\left(\mathcal{P}_i(X), \mathcal{P}_j(X)\right) X_j,$$

   where the reconstruction of a given voxel i is a weighted average over all observed voxels j in a given (possibly infinite) neighbourhood $\mathcal{N}_i$, with weights reflecting similarity between patches $\mathcal{P}$ centred about these voxels. Convolutional neural networks naturally extend this family by stacking many such filters and making their weights learnable rather than fixed.

[2] *Generative model* should be understood in the classical sense: a (probabilistic) function that mimics the true data-generating process.

For the first family of methods, it was found that the denoising effect is stronger when $\mathcal{E}_p$ is an absolute norm (or sum of), rather than a squared norm, because the solution is implicitly sparse in the feature domain (Bach, 2011). This family of methods includes total variation (TV) regularisation (Rudin et al., 1992) and wavelet soft-thresholding (Donoho, 1995). The second family also leverages sparsity in the form of redundancy in the spatial domain; that is, the dictionary of patches necessary to reconstruct the noise-free images is smaller than the actual number of patches in the image. Several such methods have been developed specifically for MRI, with the aim of finding an optimal, voxel-wise weighting based on the noise distribution (Coupe et al., 2008; Manjón et al., 2010; Coupé et al., 2012; Manjón et al., 2012).

Optimisation methods can naturally be interpreted as a maximum *a posteriori* (MAP) solution in a generative model, which eases their interpretation and extension. This feature is especially important in qMRI, where we possess a well-defined (nonlinear) forward function and log-likelihood, and wish to regularise a small number of maps. Joint total variation (JTV) (Sapiro and Ringach, 1996) is an extension of TV to multi-contrast images, where the absolute norm is defined across contrasts, introducing an implicit correlation between them. TV and JTV have been used before in MR reconstruction, *e.g.*, in compressed-sensing (Huang et al., 2012), quantitative susceptibility mapping (Liu et al., 2011) and super-resolution (Brudfors et al., 2018). Recently, Varadarajan et al. (2020) applied JTV to multi-echo SPGR data for B_0 distortion correction, with the joint prior defined across echoes. However, these problems are linear and (therefore) convex, while the MR signal equations are not. This non-convexity has limited the use of iterative methods in this context, as algorithms either have a slow convergence – in the case of first order methods – or are not ensured to converge – in the case of second order methods.

Non-linear MPM fitting

In Balbastre et al. (2020), we introduced a non-linear, spatially regularised version of ESTATICS that was fit using Gauss-Newton optimisation. However, since regularisation did not apply directly to the parameter maps but to these weighted intercepts, the denoising of the quantitative parameters was merely a by-product of the denoising of the intercepts. In this work, we instead pro-

Journal Pre-proof

pose a generic framework that directly finds MAP quantitative (log) parameters from all SPGR data at hand. To this end, we extend our optimisation framework to work directly with the SPGR signal equation, and jointly regularise the log-parameter maps. Working with log-parameters $\log \mathrm{PD}$, $\log R_1$, $\log R_2^*$, $\mathrm{logit}\,\mathrm{MT}_{\mathrm{sat}}$) is advantageous as their spatial gradients are independent from the choice of unit and representation (time, in s or ms, or rate, in s^{-1} or ms^{-1}), although it introduces an additional degree of nonlinearity. Another advantage is that posterior distributions are often better approximated by Laplace's method when a log basis is used (MacKay, 1998). We name this extended model SPGR+JTV when it uses regularisation and SPGR-ML when it does not. In Gauss-Newton, positive-definiteness of the Hessian is enforced by the use of Fisher's scoring, which discards terms that involve second derivatives of the SPGR signal function. In this new work, we found that a more stable algorithm can be derived by keeping the (absolute) diagonal of this second-order term. We illustrate the improved convergence on a toy problem, and experimentally show that it generalises to least-squares ESTATICS and SPGR. Additionally, our implementation uses a quadratic upper bound of the JTV functional, allowing the surrogate problem to be efficiently solved using second-order optimisation.

We validated our method on a unique dataset – five repeats of the MPM protocol acquired, within a single session, on a healthy subject – by leave-one-out cross-validation: maps were estimated on one repeat and used to predict the weighted images from the other repeats. Our methods (SPGR+JTV and SPGR-ML) were compared to algorithms that rely on different flavours of the ESTATICS model: LOGLIN uses a log-linear fit to recover the intercepts and decay (Weiskopf et al., 2014), whereas NONLIN, and NONLIN+JTV use a (regularised) non-linear fit (Balbastre et al., 2020). We also applied ESTATICS to echoes that had been previously denoised using the adaptive optimized nonlocal means (AONLM) method (Manjón et al., 2010), as a baseline for the denoising component of our method. Quantitative maps were then computed analytically from the intercepts obtained from all ESTATICS methods. SPGR+JTV consistently outperformed all baselines. Besides, we probed the effect of regularisation by reconstructing maps with a varying number of echoes, and explored the use of our generative model to estimate uncertainty about the inferred quantitative values.

2. Theory

2.1. Forward model: Spoiled Gradient Echo signal equation

The steady state signal intensity of a spoiled gradient echo acquired with flip angle α, repetition time TR and echo time TE is[3]

$$S = A \sin\alpha \frac{1 - \exp(-R_1 \mathrm{TR})}{1 - \cos\alpha \exp(-R_1 \mathrm{TR})} \exp(-R_2^* \mathrm{TE}) \,, \quad (1)$$

where A is proportional to the proton density (with a multiplicative factor related to the sensitivity of the receiver coil).

Multi-echo techniques acquire multiple echoes, allowing the R_2^* decay rate to be mapped, while variable flip-angle (VFA) techniques acquire images with multiple flip-angles to map the R_1 relaxation rate. Multi-parameter mapping protocols combine these two techniques in a single imaging session.

In addition, a similar sequence can be used to measure the magnetisation transfer saturation: in a two-pool model, protons are separated between the free pool – protons in free water that can diffuse freely, which cause the MR signal – and bound pool – non-acqueous protons, *e.g.*, in proteins and lipids of the myelin sheath encasing axons. Following selective saturation of the bound pool via an off-resonance pulse, the process of magnetisation transfer leads to partial saturation of the observable free pool. By acquiring images with and without this off-resonance pulse, the proportion of saturated protons can be measured, giving a surrogate measure of the bound pool. The resulting signal (which we write at TE=0 for brevity) can be modelled by adding an extra excitation pulse, with repetition time TR_2 and flip angle α_2 (Helms et al., 2008b):

$$S_{\mathrm{TE}=0} = A \sin\alpha \frac{1 - \cos\alpha_2 e^{-R_1 \mathrm{TR}} - (1 - \cos\alpha_2) e^{-R_1 \mathrm{TR}_2}}{1 - \cos\alpha \cos\alpha_2 e^{-R_1 \mathrm{TR}}} \,, \quad (2)$$

[3] A spoiled gradient echo acquisition is obtained by playing an excitation pulse that causes the bulk magnetization to rotate by an angle α (the flip angle), followed by a series of dephasing and rephasing gradients that create an echo train.

Journal Pre-proof

Here, α_2 is unknown (it depends on the degree of magnetisation transfer saturation) and, following Helms et al. (2008b), we name $\delta = 1 - \cos\alpha_2$ the MT saturation. Furthermore, we assume that $\mathrm{TR}_2 = 0$, and the signal equation simplifies to:

$$S_{\mathrm{TE}=0} = A \sin\alpha \frac{(1-\delta)\,(1-\exp(-R_1\mathrm{TR}))}{1-(1-\delta)\cos\alpha\exp(-R_1\mathrm{TR})}\,. \tag{3}$$

In practice, the MR environment is not perfect and multiple phenomena cause deviations from this theoretical model and introduce artefacts:

- The excitation field, B_1^+, is not perfectly homogeneous, causing the effective flip-angle to be location-dependent and to deviate from its nominal value. Multiple techniques allow this field to be measured (Jiru and Klose, 2006; Yarnykh, 2007; Sacolick et al., 2010).
- The receive field, B_1^-, suffers from the same issue, especially when high-density phased-array coils are used. This is particularly problematic when between-scan motion happens, as the receive field modulation then has a different value in the different contrasts (Papp et al., 2016).
- The main magnetic field, B_0, is not perfectly flat either, due to changes in magnetic susceptibility at air-tissue interfaces and imperfect shimming. As odd and even echoes are acquired with a readout of inverse polarity, between-echo distorsion is possible in regions where field inhomogeneity is high (Varadarajan et al., 2020).
- Even with the use of gradient and RF spoiling, the assumption that no transverse magnetisation persists across TRs is rarely met and can lead to additional dependencies, on both tissue properties and hardware, in the R_1 estimates (Preibisch and Deichmann, 2009; Corbin and Callaghan, 2021).

2.2. Noise model: Gaussian

The measured signal is hampered by thermal noise, which is uniform and Gaussian in each complex coil image and becomes Rician or non-central Chi distributed in the sum-of-square magnitude image. In the high SNR regime, this noise is approximately Gaussian, and we make that approximation in this work. This means that the fitting procedure – which consists of inferring the unknown parameters from noisy measurements – is effectively a non-linear least squares problem.

2.3. Optimisation: non-linear least squares

A non-linear least square problem is one of the form:

$$\min_{\boldsymbol{y}} \left\{ \ell(\boldsymbol{y}) := \frac{1}{2}\sum_i (f_i(\boldsymbol{y}) - x_i)^2 \right\}, \tag{4}$$

where all f_i are non-linear functions. In general, there is no closed-form solution (when there is a solution at all), and the problem can be non-convex. Therefore, carefully crafted iterative methods must be used. A method of choice is Gauss-Newton optimisation. Gauss-Newton emerges from Newton's method, that is, from a second order Taylor expansion of the objective function. Let us write the gradient of function f in $\boldsymbol{y}$ as $\boldsymbol{g}_f(\boldsymbol{y})$ and its Hessian as $\boldsymbol{H}_f(\boldsymbol{y})$. The Taylor expansion of equation (4) about $\boldsymbol{y}_0$ gives:

$$\begin{aligned}\ell(\boldsymbol{y}) \approx \ell(\boldsymbol{y}_0) &+ (\boldsymbol{y}-\boldsymbol{y}_0)^{\mathrm{T}}\,\boldsymbol{g}_\ell(\boldsymbol{y}_0) \\ &+ \frac{1}{2}(\boldsymbol{y}-\boldsymbol{y}_0)^{\mathrm{T}}\,\boldsymbol{H}_\ell(\boldsymbol{y}_0)\,(\boldsymbol{y}-\boldsymbol{y}_0)\end{aligned} \tag{5}$$

$$\boldsymbol{g}_\ell(\boldsymbol{y}_0) = \sum_i (f_i(\boldsymbol{y}_0) - x_i)\,\boldsymbol{g}_{f_i}(\boldsymbol{y}_0) \tag{6}$$

$$\boldsymbol{H}_\ell(\boldsymbol{y}_0) = \sum_i \boldsymbol{g}_{f_i}(\boldsymbol{y}_0)\boldsymbol{g}_{f_i}(\boldsymbol{y}_0)^{\mathrm{T}} + (f_i(\boldsymbol{y}_0) - x_i)\,\boldsymbol{H}_{f_i}(\boldsymbol{y}_0) \tag{7}$$

Under the assumption that $\boldsymbol{H}_\ell(\boldsymbol{y}_0)$ is positive-definite, Newton's method solves the surrogate quadratic problem by finding its stationary point:

$$\boldsymbol{y}_1 = \boldsymbol{y}_0 - \boldsymbol{H}_\ell(\boldsymbol{y}_0)^{-1}\boldsymbol{g}_\ell(\boldsymbol{y}_0)\,. \tag{8}$$

However, as previously stated, the problem is not necessarily convex and therefore the Hessian not necessarily positive-definite. Gauss-Newton circumvents this issue using Fisher's scoring, *i.e.*, by assuming – in the Hessian – that residuals are all zero:

$$\boldsymbol{P}_\ell(\boldsymbol{y}_0) = \sum_i \boldsymbol{g}_{f_i}(\boldsymbol{y}_0)\boldsymbol{g}_{f_i}(\boldsymbol{y}_0)^{\mathrm{T}}\,. \tag{9}$$

We denote this alternative Hessian $\boldsymbol{P}$ for *preconditioner*. The resulting quantity is (under mild conditions) positive-definite, and the surrogate quadratic problem now has

Journal Pre-proof

a minimum. When it converges, the Gauss-Newton iteration converges to a stationary point of the objective function. However, convergence is not guaranteed (Mascarenhas, 2014). In the following subsections, we introduce a sufficient condition for monotonic convergence in Newton-like methods and show simple examples related to MR relaxometry where Gauss-Newton does not converge monotonically or even diverges. We then introduce a new preconditioner that is shown (experimentally) to enforce the condition for monotonic convergence. This preconditioner will be used to optimise the non-linear SPGR least square problem.

A sufficient condition for monotonic convergence

Let us focus on the simpler 1D case. Let $\ell : \mathbb{R} \to \mathbb{R}$ be a twice-differentiable non-linear function over the real line with a single local (and therefore global) minimum y^* and no local maximum[4]. This function does not need to be convex but, by construction, its first derivative $g : \mathbb{R} \to \mathbb{R}$ is negative everywhere on the left of its optimum and positive everywhere on its right. Let $h : \mathbb{R} \to \mathbb{R}_*^+$ be a strictly positive function that we will use as a preconditioner (*e.g.*, the absolute value of the curvature). Applying a Newton-like step at point y_n gives an update of the form $y_{n+1} = y_n - g(y_n)/h(y_n)$. Since h is positive, the step is ensured to be in the direction of the optimum; a sufficient condition for this step to improve the objective function ($\ell(y_{n+1}) < \ell(y_n)$) is for the updated point to fall *between* y_n and the optimum y^*. Finally, this condition can be formalised as

$$\frac{|g(y_n)|}{h(y_n)} \leqslant |y_n - y^*| \ . \tag{10}$$

While this condition seems difficult to enforce or check, note that it is always true at the optimum, where $g(y^*) = 0$. Let us write $r(y) = g(y)/h(y)$ and add the condition that h is differentiable; then condition (10) is enforced if the (absolute value of the) derivative of r is everywhere equal or below 1. On the right (positive) side, $r'(y) \leqslant 1$ on $[y^*, y_n]$ implies

$$|r(y_n)| = r(y_n) = \int_{y^*}^{y_n} r'(y)\,\mathrm{d}y \leqslant \int_{y^*}^{y_n} 1\,\mathrm{d}y = y_n - y^* = |y_n - y^*| \ .$$

[4]This proof could be easily extended to local-convergence when more than one minimum exists, but we focus on the simpler case for clarity

The same thing can be shown on the left (negative) side.

In the multivariate case, we want the update to stay in the same "quadrant" (that is, all coordinates fall between their original value and the optimum), which we formalise as:

$$\left|\boldsymbol{H}(\boldsymbol{y}_n)^{-1}\boldsymbol{g}(\boldsymbol{y}_n)\right| \leqslant \left|\boldsymbol{y}_n - \boldsymbol{y}^*\right| \ , \tag{11}$$

where $|\cdot|$ is the element-wise absolute value. However, we can apply any arbitrary rotation to the problem and this condition stays true. We can therefore relax it to:

$$\|\boldsymbol{H}(\boldsymbol{y}_n)^{-1}\boldsymbol{g}(\boldsymbol{y}_n)\| \leqslant \|\boldsymbol{y}_n - \boldsymbol{y}^*\| \ , \tag{12}$$

where $\|\cdot\|$ is the Euclidean norm.

Toy problem 1: exponential least square

Let us now discuss two toy problems that are closely related to MR relaxometry. The first one consists of finding the least square solution when the non-linear function is an exponential and there is a single data point:

$$\ell(y) = \frac{1}{2}\left(e^{-y} - x\right)^2 \tag{13}$$

The corresponding energy landscape, shown as a solid black line in Fig. 1, is non-convex on its left. When $x > 0$, there is an obvious analytical solution ($y^* = -\log(x)$), but we will use an iterative method to get to the same conclusion. Using the same notations as before, we have:

$$g_\ell(y) = -(e^{-y} - x)\,e^{-y} \ , \tag{14}$$

$$h_\ell(y) = e^{-2y} + (e^{-y} - x)\,e^{-y} \ , \tag{15}$$

$$p_\ell(y) = e^{-2y} \ . \tag{16}$$

If we compute the derivative of $r_\ell(y) = g_\ell(y)/p_\ell(y)$, we find

$$r'_\ell(y) = -xe^{y} \ , \tag{17}$$

whose absolute value is *not* bounded by 1. We therefore introduce an alternative pre-conditioner that uses the absolute value of the second term rather than discarding it:

$$\tilde{p}_\ell(y) = e^{-2y} + \left|e^{-y} - x\right|\,e^{-y} \ . \tag{18}$$

Although we do not prove it, we find experimentally that this preconditioner enforces condition (10). The corresponding quadratic approximations are plotted in red (Gauss-Newton) and blue (proposed preconditioner) in

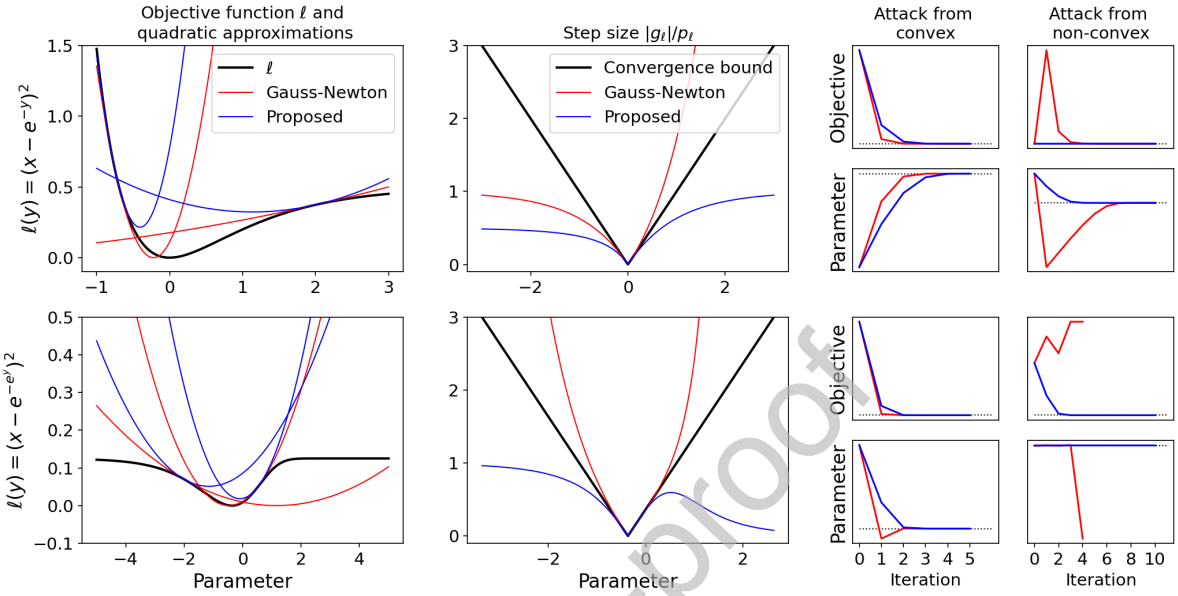


Figure 1: Qualitative analysis of two toy problems related to MR relaxometry. The first toy problem (first row) has an objective function of the form $(x - \exp(-y))^2$, while the other toy problem (second row) as one of the form $(x - \exp(-\exp(y)))^2$. In each case, the first column shows the objective function in black along with quadratic approximations corresponding to Gauss-Newton (red) and the proposed second order method (blue). The second column shows the magnitude of the gradient-hessian ratio. The black line corresponds to a purely quadratic objective function, whose optimum is found in one step ($|g(x)|/h(x) = |x - x^*|$). When the ratio is under this black line, monotonic convergence is ensured. The third column shows convergence of the Gauss-Newton (red) and proposed (blue) iterative schemes when attacking from either a convex or non-convex lobe, with the optimum shown in dashed black. In each case, convergence is shown both in terms of objective value per iteration and parameter value per iteration.

Fig. 1. Experimental convergence with both preconditioners is also shown. Conversely to Gauss-Newton, the proposed version is monotonic when attacking from the non-convex side and converges almost as fast as Gauss-Newton when attacking from the convex side.

Toy problem 2: nested exponential least square

Our second toy problem involves nested exponentials:

$$\ell(y) = \frac{1}{2}\left(x - e^{-e^y}\right)^2 . \tag{19}$$

The corresponding energy landscape is shown in Fig. 1: its curvature now tends towards zero on both sides of the real line. Its gradient, Hessian and preconditioners are:

$$g_\ell(y_0) = -e^{y_0 - e^{y_0}}\left(e^{-e^{y_0}} - x\right), \tag{20}$$

$$h_\ell(y_0) = \left(e^{y_0 - e^{y_0}}\right)^2 + \left(e^{2y_0} - e^{y_0}\right) e^{-e^{y_0}}\left(e^{y_0} - x\right), \tag{21}$$

$$p_\ell(y_0) = \left(e^{y_0 - e^{y_0}}\right)^2 , \tag{22}$$

$$\tilde{p}_\ell(y_0) = \left(e^{y_0 - e^{y_0}}\right)^2 + \left|\left(e^{2y_0} - e^{y_0}\right) e^{-e^{y_0}}\left(e^{y_0} - x\right)\right| . \tag{23}$$

Again, $|r'_\ell|$ is not bounded, and the Gauss-Newton preconditioner fails to enforce the convergence condition. Experimental convergence is shown in Fig. 1, where the proposed method is shown to monotonically converge while Gauss-Newton can diverge when starting in a non-convex section of the energy landscape.

Journal Pre-proof

Generalisation to the multivariate case

The toy problems investigated in the previous section are highly relevant to MR relaxometry: the first case is related to the exponential fit of the signal decay, where each data point is seen as the exponential of a linear combination of parameters, while the second case is an extension of the first, where positive parameters are encoded by their log. Given the apparent robustness of the proposed preconditioner on these toy problems, we propose two extensions to the general multivariate case by replacing the absolute value of the scalar Hessian in the one-dimensional case with either: (1) the absolute value of the diagonal of the Hessian, $\left|\boldsymbol{H}_{f_i}(\boldsymbol{y}_0)\right| \odot \boldsymbol{I}$; (2) the sum of the absolute value of the Hessian across columns, $\operatorname{diag}\left(\left|\boldsymbol{H}_{f_i}(\boldsymbol{y}_0)\right| \mathbf{1}\right)$. Only the second one is ensured to majorise the true Hessian (Chun and Fessler, 2018), but we show in our experiments that the first one yields a stable convergence in the MPM context. We therefore use the preconditioner:

$$\tilde{\boldsymbol{P}}_\ell(\boldsymbol{y}_0) = \sum_i \boldsymbol{g}_{f_i}(\boldsymbol{y}_0)\boldsymbol{g}_{f_i}(\boldsymbol{y}_0)^{\mathrm{T}} + \left|f_i(\boldsymbol{y}_0) - \boldsymbol{x}_i\right| \left|\boldsymbol{H}_{f_i}(\boldsymbol{y}_0)\right| \odot \boldsymbol{I}. \tag{24}$$

The impact of this loading is illustrated with a pair of two-dimensional toy problems that are closely related to the one-dimensional problems presented before. The first problem involves an objective function of the form $\ell(y,z) = \frac{1}{2}\left(x_0 - e^{z-t_0 y}\right)^2 + \frac{1}{2}\left(x_1 - e^{z-t_1 y}\right)^2$. In the second problem, the decay parameter is encoded by its log: $\ell(y,z) = \frac{1}{2}\left(x_0 - e^{z-t_0 e^y}\right)^2 + \frac{1}{2}\left(x_1 - e^{z-t_1 e^y}\right)^2$. Step sizes and optimisation trajectories are depicted in Fig. 2. When the Gauss-Newton preconditioner is used, a vast part of the parameter space has very large step sizes that lead to overshooting. When the proposed preconditioners are used, the entire parameter space (or most of it) lies below the monotonic convergence threshold.

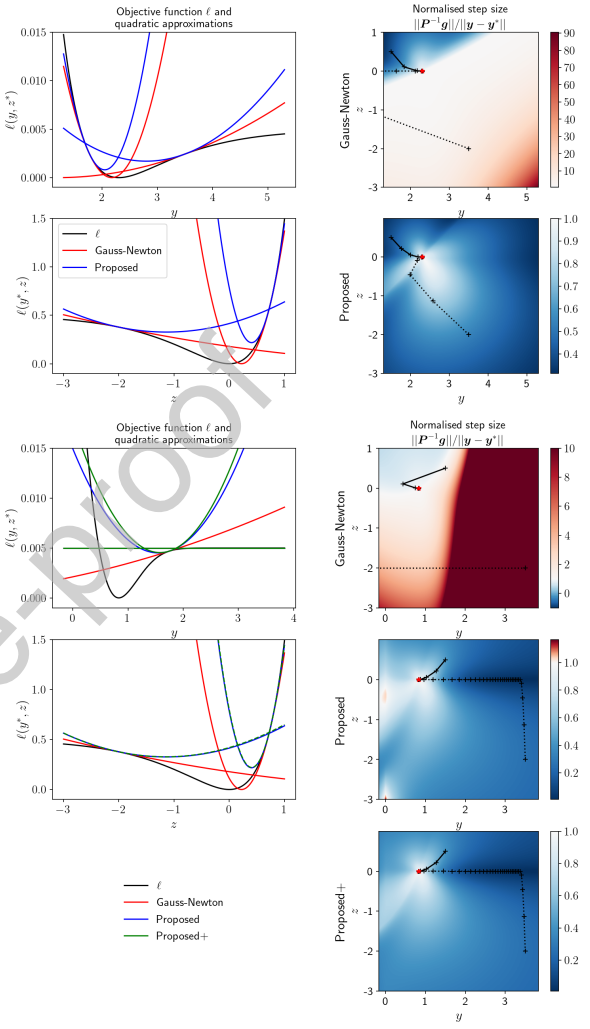


Figure 2: Qualitative analysis of two-dimensional toy problems related to MR relaxometry. The first toy problem (top half) has an objective function of the form $\ell(y,z) = \left(x_0 - e^{z-t_0 y}\right)^2 + \left(x_1 - e^{z-t_1 y}\right)^2$, while the other toy problem (bottom half) has one of the form $\ell(y,z) = \left(x_0 - e^{z-t_0 e^y}\right)^2 + \left(x_1 - e^{z-t_1 e^y}\right)^2$. In each case, the first column shows the objective function along each of the two dimensions (in black) along with quadratic approximations corresponding to Gauss-Newton (red) and the proposed second order methods (blue and green). The second column shows the normalised step size for each method, color-coded such that values below one are blue and values above one are red. The optimum is marked by a red star. Convergence paths from two random initial coordinates (solid and dashed curves) are overlaid. In the second toy problem, two different loading matrices are used: *Proposed* uses the diagonal of the absolute Hessian $\operatorname{diag}(|\boldsymbol{H}|)$ whereas *Proposed+* uses its sum across columns $|\boldsymbol{H}|\,\mathbf{1}$.

2.4. Likelihood: spoiled gradient echo

The main contribution of this paper lies in the direct optimisation of A, R_1 R_2^* and δ. First, each parameter is encoded by its log (or logit in the case of δ). Let us switch from a physics to a mathematical notation, where scalars and vectors are written in lowercase while capital letters are reserved for matrices and tensors. We use a tilde to differentiate the log-parameters from their exponentiated

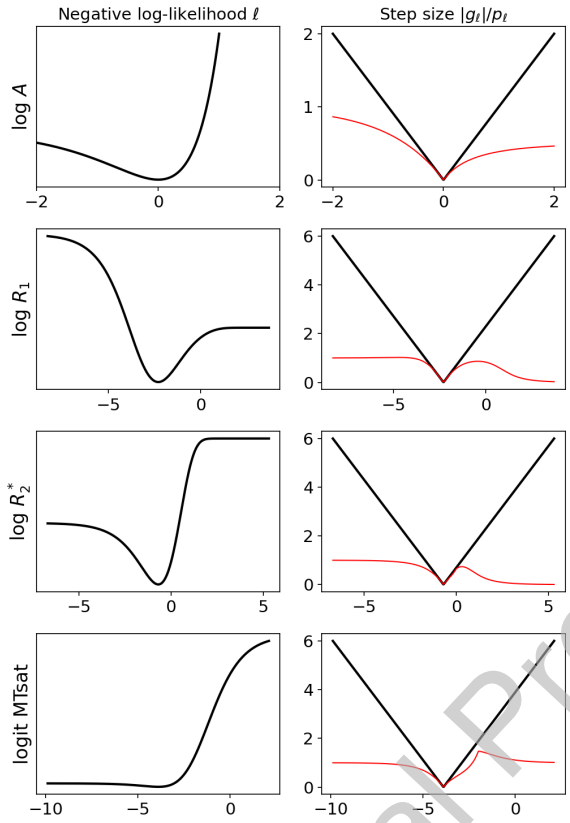


Figure 3: Negative log-likelihood and step size function of the nonlinear SPGR model. The first column shows the log-likelihood of a single observation with respect to each log-parameter, while keeping the others fixed. The second column shows the step size function (red) obtained with the proposed approximate Hessian, compared with the theoretical perfect step size function $f(x) = |x - x^*|$. As long as the actual step size is below this theoretical function, monotonic convergence is ensured.

version, so that in each voxel:

$$a = \exp(\tilde{a}), \qquad r_1 = \exp(\tilde{r}_1),$$
$$r_2 = \exp(\tilde{r}_2), \qquad \delta = \frac{1}{1 + \exp(-\tilde{\delta})}.$$

Let us assume that all images are defined on the same grid with N voxels. We write the set of log-parameters as $\boldsymbol{Y} = \left[\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{r}}_1, \tilde{\boldsymbol{r}}_2, \tilde{\boldsymbol{\delta}}\right] \in \mathbb{R}^{N\times 4}$ and the set of observed 3D images as $\boldsymbol{X} \in \mathbb{R}^{N\times I}$. Observed images come with a set of fixed parameters $\boldsymbol{\Theta} \in \mathbb{R}^{4\times I}$, where each quadruplet $\boldsymbol{\theta}_i = \left(\alpha_i, t_{ri}, t_{ei}, \sigma_i^2\right) \in \mathbb{R}^4$ contains the flip angle, repetition time, echo time and noise variance of an image. The likelihood term of the objective function involves a sum across images followed by a sum across voxels:

$$\ell(\boldsymbol{Y}) = \sum_i \sum_n \frac{1}{2\sigma_i^2} \left(x_{in} - s(\boldsymbol{y}_n \mid \boldsymbol{\theta}_i)\right)^2 , \tag{25}$$

where s is the intensity of a spoiled gradient echo[5]:

$$s(\tilde{a}, \tilde{r}_1, \tilde{r}_2, \tilde{\delta} \mid \alpha, t_r, t_e) = a \sin\alpha \frac{(1-\delta)(1-e^{-r_1 t_r})}{1-(1-\delta)\cos\alpha e^{-r_1 t_r}} e^{-r_2 t_e} . \tag{26}$$

For contrasts that do not include an MT pulse, $\delta = 0$ and gradients with respect to $\tilde{\delta}$ are zero. We recognise a nonlinear least squares problem, and propose to use the preconditioner from equation (24), which depends on the gradient and Hessian of s. Let us focus on a single voxel and drop the (i, n) subscripts for clarity. After differentiation, the gradient is:

$$\frac{\partial s}{\partial \tilde{a}} = s , \tag{27}$$
$$\frac{\partial s}{\partial \tilde{r}_1} = -t_r r_1 \frac{((1-\delta)\cos\alpha - 1)\, e^{-r_1 t_r}}{(1-(1-\delta)\cos\alpha\, e^{-r_1 t_r})\,(1-e^{-r_1 t_r})} s , \tag{28}$$
$$\frac{\partial s}{\partial \tilde{r}_2} = -t_e r_2 \, s , \tag{29}$$
$$\frac{\partial s}{\partial \tilde{\delta}} = \frac{\delta}{(1-\delta)\cos\alpha\, e^{-r_1 t_r} - 1} s , \tag{30}$$

[5] When the receive and transmit fields are known from *e.g.* field mapping, equation (26) is slighlty changed. The flip angle α takes a voxel-dependent value $\alpha_n = \alpha_{\text{nom}} b_n^+$, where the nominal flip angle is modulated by the transmit field efficiency; and the whole signal is modulated voxel-wise by the net receive field b_n^-.

Journal Pre-proof

and the Hessian is:

$$\frac{\partial^2 s}{\partial \tilde{a}^2} = \frac{\partial s}{\partial \tilde{a}} , \tag{31}$$

$$\frac{\partial^2 s}{\partial \tilde{r}_1^2} = \left(1 - t_r r_1 \frac{1 + (1-\delta)\cos\alpha\, e^{-r_1 t_r}}{1 - (1-\delta)\cos\alpha\, e^{-r_1 t_r}}\right) \frac{\partial s}{\partial \tilde{r}_1} , \tag{32}$$

$$\frac{\partial^2 s}{\partial \tilde{r}_2^2} = (1 - t_e r_2)\ \frac{\partial s}{\partial \tilde{r}_2} , \tag{33}$$

$$\frac{\partial^2 s}{\partial \tilde{\delta}^2} = \left(\frac{2(1-\delta)(1-\cos\alpha\, e^{-r_1 t_r})}{1 - (1-\delta)\cos\alpha\, e^{-r_1 t_r}} - 1\right) \frac{\partial s}{\partial \tilde{\delta}} . \tag{34}$$

The gradient $\boldsymbol{g}_\ell$ and preconditioner $\tilde{\boldsymbol{P}}_\ell$ of the least-square objective function are then computed as in equations (6) and (24) and summed across echo times and contrasts. The log-likelihood of one observation with respect to each parameter – keeping the others fixed – is shown in Fig. 3, along with the step size function, which is consistently under the monotonic convergence bound. Although this figure only shows the step size function for one given set of parameters, the convergence property is consistently enforced when sampling random parameters and observed values.

Spatial projection

In practice, it is common for people to move their head between scans, making inter-scan registration a mandatory initial step. While reslicing all scans in the same space is a possibility, we prefer to keep the raw data untouched and integrate the spatial projection from the reconstruction space to the different acquisition spaces in the forward model, modifying slightly the objective function. Let us write all log parameter maps as $\boldsymbol{y} = \mathrm{vec}(\boldsymbol{Y}) \in \mathbb{R}^{NK}$, where N is the number of voxels in reconstruction space and K the number of maps, and $\boldsymbol{\Psi} \in \mathbb{R}^{MK \times NK}$ a matrix that encodes linear reslicing from reconstruction space to one acquisition space[6]. The new objective function is $\ell_\psi(\boldsymbol{y}) = \ell(\boldsymbol{\Psi}\boldsymbol{y})$. Application of the chain rule yields:

$$\frac{\partial \ell_\psi}{\partial \boldsymbol{y}} = \boldsymbol{\Psi}^{\mathrm{T}} \left(\frac{\partial \ell}{\partial \boldsymbol{y}} (\boldsymbol{\Psi}\boldsymbol{y})\right) , \tag{35}$$

$$\frac{\partial^2 \ell_\psi}{\partial \boldsymbol{y} \partial \boldsymbol{y}^{\mathrm{T}}} = \boldsymbol{\Psi}^{\mathrm{T}} \left(\frac{\partial^2 \ell}{\partial \boldsymbol{y} \partial \boldsymbol{y}^{\mathrm{T}}} (\boldsymbol{\Psi}\boldsymbol{y})\right) \boldsymbol{\Psi} . \tag{36}$$

However, this new Hessian is much less sparse: it cannot be seen as block diagonal anymore. It was shown in Ashburner et al. (2018) that this matrix can be majorised (in the Löwner order sense) by a block-diagonal matrix:

$$\frac{\partial^2 \ell_\psi}{\partial \boldsymbol{y} \partial \boldsymbol{y}^{\mathrm{T}}} \preceq \mathrm{diag}\left(\boldsymbol{\Psi}^{\mathrm{T}} \left(\frac{\partial^2 \ell}{\partial \boldsymbol{y} \partial \boldsymbol{y}^{\mathrm{T}}} (\boldsymbol{\Psi}\boldsymbol{y})\right)\right) . \tag{37}$$

In other words, for each contrast, the log parameter maps are resliced (*pulled*) to the acquisition space, where the gradient and Hessian of the least square function are computed and sent back to recon space using the adjoint of the reslicing operation (*pushed*).

2.5. Prior: joint total variation

Multiple types of spatial regularisation exist, with different properties. Here, we focus on joint total-variation, equivalent to an $\ell_{1,2}$ norm over spatial gradients, which benefits from the same sparsity-inducing properties as other ℓ_1 norms, except that sparsification is shared across contrasts. This property is particularly interesting when the same object is seen through different contrasts, as is the case in the context of MPMs, since edges are shared across parameter maps. Again, let $\boldsymbol{Y} = \left[\tilde{\boldsymbol{a}}, \tilde{\boldsymbol{r}}_1, \tilde{\boldsymbol{r}}_2, \tilde{\boldsymbol{\delta}}\right] \in \mathbb{R}^{N \times 4}$ be the set of all parameter maps, the regulariser has the form:

$$\mathrm{JTV}(\boldsymbol{Y}) = \sum_{n=1}^{N} \sqrt{\sum_{k=1}^{4} \lambda_k \left(\boldsymbol{G}_n \boldsymbol{y}_k\right)^{\mathrm{T}} \left(\boldsymbol{G}_n \boldsymbol{y}_k\right)} , \tag{38}$$

where $\boldsymbol{G}_n$ extracts all 6 forward and backward finite-differences (in 3D) about the n-th voxel and λ_k is a parameter-specific regularisation factor.

This regulariser is convex but non-smooth, complicating its use in optimisation. However, several methods have been developed over the years to solve this problem; *e.g.*, ISTA (Daubechies et al., 2004), FISTA (Beck and Teboulle, 2009), ADMM (Boyd et al., 2011) and variations thereof. In this paper, we use an iteratively

[6] Since the same transformation is applied to all maps, $\boldsymbol{\Psi}$ really is the Kronecker product of an $M \times N$ reslicing matrix and a $K \times K$ identity matrix.

Journal Pre-proof

reweighted least squares (IRLS) scheme (Daubechies et al., 2010; Bach, 2011), which relies on the bound:

$$\sqrt{\sum_k \lambda_k \left(\boldsymbol{G}_n \boldsymbol{y}_k\right)^{\mathrm{T}} \left(\boldsymbol{G}_n \boldsymbol{y}_k\right)} = \min_{w_n > 0} \left\{ \frac{1}{2w_n} + \frac{w_n}{2} \sum_k \lambda_k \left(\boldsymbol{G}_n \boldsymbol{y}_k\right)^{\mathrm{T}} \left(\boldsymbol{G}_n \boldsymbol{y}_k\right) \right\} . \quad (39)$$

When the weight map $\boldsymbol{w}$ is fixed, this bound can be seen as a Tikhonov (*i.e.*, ℓ_2) prior with nonstationary regularisation, which is a quadratic prior that factorises across parameters. Therefore, the between-parameters correlations induced by the JTV prior are entirely captured by the weights. Conversely, when the parameter maps are fixed, the weights can be updated in closed-form:

$$w_n = \left(\sum_k \lambda_k \left(\boldsymbol{G}_n \boldsymbol{y}_k\right)^{\mathrm{T}} \left(\boldsymbol{G}_n \boldsymbol{y}_k\right) \right)^{-\frac{1}{2}} . \quad (40)$$

Let us now vectorise all log-parameter maps: $\boldsymbol{y} = \mathrm{vec}\left(\boldsymbol{Y}\right)$. The quadratic term in equation (39) – when summed over all voxels – can be written as $\boldsymbol{y}^{\mathrm{T}} \boldsymbol{L} \boldsymbol{y}$ with $\boldsymbol{L} = \boldsymbol{G}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{G} \otimes \mathrm{diag}\left(\boldsymbol{\lambda}\right)$, where $\boldsymbol{G} \in \mathbb{R}^{6N \times N}$ extracts all 6 forward and backward finite-differences about all voxels and $\boldsymbol{W} = \mathrm{diag}\left(\boldsymbol{w}\right) \otimes \boldsymbol{I}_6$. Note further that $\boldsymbol{G}^{\mathrm{T}} \boldsymbol{G}$ is Toeplitz and can be implemented as a convolution by a small kernel (Ashburner, 2007). Similarly, $\mathrm{diag}\left(\boldsymbol{G}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{G}\right)$ can be obtained by convolving the weight map $\boldsymbol{w}$ with a small "cross" kernel.

2.6. Optimisation: regularised spoiled gradient echo

Optimisation follows an IRLS scheme, where each iteration consists of alternately solving for the parameter maps $\boldsymbol{y}$ and the JTV weights $\boldsymbol{w}$. At each iteration, the objective function in $\boldsymbol{y}$ is of the form:

$$\ell(\boldsymbol{y}) = \sum_{c,e} \frac{1}{2\sigma_c^2} \| \boldsymbol{x}_{ce} - s_{ce}\left(\boldsymbol{y}\right) \|^2 + \frac{1}{2} \boldsymbol{y}^{\mathrm{T}} \boldsymbol{L} \boldsymbol{y} , \quad (41)$$

where c and e are contrasts and echoes. This function is non-linear and solved using the proposed second order method, from which we get a gradient $\boldsymbol{g} \in \mathbb{R}^{4N}$ and an approximate Hessian $\boldsymbol{H} \in \mathbb{R}^{4N \times 4N}$. This matrix is actually block diagonal as it consists of N small 4×4 matrices. Performing the Newton update step involves solving the linear system

$$\left(\boldsymbol{H} + \boldsymbol{L}\right)^{-1} \left(\boldsymbol{g} + \boldsymbol{L} \boldsymbol{y}\right) , \quad (42)$$

which is done using the preconditioned conjugate gradient method. The preconditioner that we use is

$$\boldsymbol{H} + \left(\boldsymbol{G}^{\mathrm{T}} \boldsymbol{W} \boldsymbol{G} \odot \boldsymbol{I} \right) \otimes \mathrm{diag}\left(\boldsymbol{\lambda}\right) ,$$

2.7. Implementation

In practice, 10 IRLS iterations were used, with 5 inner Newton iterations and 32 conjugate gradient iterations per Newton step. A tolerance threshold on the gain between two subsequent iterations was set for early stopping (IRLS: 10^{-5}, Newton: 10^{-5}, CG: 10^{-3}). The variance of the thermal noise was estimated by fitting a two-class Rician mixture to each acquired volume (Ashburner and Ridgway, 2013): the class with lowest mean was considered as the background. Variances were then averaged across echo times using the geometric mean. Initial (constant) parameter maps were estimated using the mean of the other (non-background) class in each volume.

This algorithm, along with baseline methods, is implemented in NiTorch, a Python library with a PyTorch backend dedicated to neuroimaging. It is publicly available at https://github.com/balbasty/nitorch. The processing of a 0.8 mm MPM dataset – from which A, R_1, R_2^* and $\mathrm{MT}_{\mathrm{sat}}$ can be estimated – fits in 12GB of memory and takes about 10 minutes on a Nvidia Titan V GPU for SPGR+JTV (5 minutes for NONLIN+JTV, $<$ 1 minute for LOGLIN).

3. Experiments

3.1. Convergence analysis

This experiment was based on simulated values. One thousand single-voxel datasets were simulated, each with their own relaxation parameters, TRs, TEs and flip angles. Each voxel had three contrasts (i.e., three different TRs and flip angles) – including one with an MT pulse – and five echo times per contrast. All log parameters ($\log A$, $\log R_1$, $\log R_2^*$, $\mathrm{logit}\,\mathrm{MT}_{\mathrm{sat}}$, log TR, log TE) were uniformly sampled in $[-5, 5]$, while flip angles were uniformly sampled in $[0, \frac{\pi}{4}]$ and the noise variance was fixed at $\sigma^2 = 1$. The (unregularised) SPGR model was then fitted using 10,000 iterations of the proposed second-order optimisation, and the negative log-likelihood was computed after each iteration. Two baseline algorithms were used as well: Gauss-Newton with a backtracking line-search (at each step, the descent direction was modulated

Journal Pre-proof

by an Armijo factor) and Levenberg-Marquardt (LM). Following Press et al. (2007), updates were only accepted if they improved the objective-function. On failure, the GN Armijo factor (respectively the LM damping parameter) was divided (resp. multiplied) by 10, while it was multiplied (resp. divided) by 10 on success. One Armijo or damping factor was defined in each virtual voxel.

A similar experiment was conducted in order to compare the use of a log, rate or time representation during optimisation. Two alternative optimisation schemes were derived such that either (A, R_1, R_2^*, $\mathrm{MT}_{\mathrm{sat}}$) or ($A$, T_1, T_2^*, $\mathrm{MT}_{\mathrm{sat}}$) are optimised instead of ($\log A$, $\log R_1$, $\log R_2^*$, $\mathrm{logit}\,\mathrm{MT}_{\mathrm{sat}}$). These two representations required an even more stable Hessian, were $|\boldsymbol{H}|\,\mathbf{1}$ is used to load the diagonal instead of diag ($|\boldsymbol{H}|$).

3.2. Datasets

Two dataset were used in this study:

1. The first dataset used in this study is a single-subject, single-session reproducibility dataset: a single participant was scanned five times in a single session with a 0.8 mm MPM protocol. This protocol consists of three multi-echo gradient echo contrasts (flip angle: 21/6/6 deg, number of echoes: 8/8/6 with 2.3 ms echo spacing, off-resonance pulse: No/No/Yes, TR: 25 ms, resolution: 0.8 mm, FOV: $256 \times 224 \times 179$ mm^3, GRAPPA acceleration: 2×2 with 40 reference lines). Calibration data were acquired to additionally account for contrast-specific receive field (Papp et al., 2016) and participant-specific transmit field inohomogeneities (Lutti et al., 2010) using the protocols described in Callaghan et al. (2019). Signal-to-noise ratios were estimated by fitting a two-class Rician mixture model to the image histograms. SNRs in the first echo were 29 (PDw), 25 (T1w) and 28 (MTw). SNRs in the last echo were 16 (PDw), 15 (T1w) and 16 (MTw).
2. The second dataset is the demonstration dataset from the hMRI toolbox (Callaghan et al., 2019), which consists of the same 0.8 mm MPM protocol acquired in a single subject, with motion between the MT-weighted scan and the PD- and T1-weighted scans.

3.3. Algorithms

We compared two variants of our method (ML and MAP) to methods based on a two-step process described in following Tabelow et al. (2019) (ESTATICS to estimate R_2^* and $T_E = 0$ intercepts, followed by an analytical fit of R_1, A and $\mathrm{MT}_{\mathrm{sat}}$ from these intercepts). These baseline methods use either a log-linear fit or a non-linear fit of ESTATICS, and use different denoising methods (preprocessing with a non-local means filter or MAP with a JTV prior):

- LOGLIN performs a log-linear fit of R_2^* from which weighted images extrapolated to TE=0 are obtained. This is equivalent to ESTATICS (Weiskopf et al., 2013), except that the spatial projection is integrated in the model. The Hessian matrix that is used is a (slight) majoriser of the true Hessian, so three Newton-like iterations are needed to reach the optimum (compared to a single one, had the true Hessian been used). Since all TRs are identical, A, R_1 and $\mathrm{MT}_{\mathrm{sat}}$ are then computed analytically according to Mohammadi et al. (2017), without requiring rational approximations.
- LOGLIN+AONLM applies an anisotropic non-local mean filter (Manjón et al., 2010) to all individual echoes before processing them with LOGLIN.
- NONLIN performs a non-linear fit or R_2^* instead of a log-linear fit, but does not use any regularisation. This method was not used in Balbastre et al. (2020), because Gauss-Newton was not stable enough without regularisation. The new Hessian used in this work makes its optimisation possible.
- NONLIN+JTV adds JTV-regularisation to NONLIN (Balbastre et al., 2020).
- SPGR-ML performs a non-linear maximum-likelihood fit of the log-parameters $\log R_1$, $\log R_2^*$, $\log A$ and $\mathrm{logit}\,\mathrm{MT}_{\mathrm{sat}}$.
- SPGR+JTV adds JTV regularisation to SPGR-ML, and effectively performs a non-linear maximum *a posteriori* fit.

For each dataset, as a preprocessing step, all contrasts across all repeats (and their $B_1^{+/-}$ maps) were registered towards the PD-weighted series of the first repeat with SPM12. This co-registration step only modified the orientation matrices in the header of the NIfTI files and did not involve any reslicing.

Journal Pre-proof

3.4. Cross-validation

This experiment was based on the single-session dataset. First, one of the repeats was used to estimate optimal regularisation factors for LOGLIN+AONLM, NONLIN+JTV, and SPGR+JTV by cross-validation across echoes. Five folds were constructed, with the first six echo times randomly permuted and split between a training set – used to estimate parameter maps – and a testing set – used to evaluate predictions. In each fold, the training echoes were used to estimate parameters with multiple regularisation values; these parameters were then used to simulate the left-out echoes and compared against the true (acquired) volumes. The mean-squared error (MSE) between the true and predicted echoes was computed and z-normalised across folds, contrast and echo times. Finally, the median z-normalised MSE was used as a metric to select the most adequate regularisation factor.

The remaining four repeats were used to compare all (optimised) methods by cross-validation across repeats. Parameter maps were computed for each repeat and used to predict individual echoes from all other repeats.

3.5. Echo decimation

This experiment was based on the reproducibility dataset. First, reference parameter maps were obtained by computing the maximum-likelihood solution (SPGR-ML) of combined data from all four test runs. Parameter maps of one of the runs were then reconstructed by either SPGR+JTV or SPGR-ML using the first 2, 4 or 6 echoes. The root mean squared error (RMSE) – within a brain ROI – was computed between each map and its reference.

3.6. Uncertainty mapping

The inverse of the Hessian of the objective function at the optimum can be used, under the Laplace approximation, to estimate the posterior uncertainty of the log-parameters (MacKay, 2003). Since it is easier to invert, we rather use the inverse of the block-diagonal preconditioner and extract its diagonal:

$$\operatorname{cov}\left[\boldsymbol{y} \mid \mathcal{X}\right] \approx \left(\boldsymbol{H} + \operatorname{diag}\left(\boldsymbol{G}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{G}\right) \otimes \boldsymbol{\lambda}\right)^{-1} , \tag{43}$$

where $\mathcal{X}$ contains all observed volumes used to estimate the maps. Note that under the Laplace approximation, the posterior of the log-parameter maps is Normal-distributed, making the parameters themselves Log-Normal-distributed. If the parameter maps themselves are of interest, it might be more interesting to return their expected value and variance, which in the case of R_1 gives:

$$\mathbb{E}\left[r_1\right] = \sqrt{\exp\left(\tilde{\sigma}^2\right)}\, \exp\left(\tilde{r}_1\right) , \tag{44}$$

$$\mathbb{V}\left[r_1\right] = \left(\exp\left(\tilde{\sigma}^2\right) - 1\right) \exp\left(\tilde{\sigma}^2\right) \exp\left(\tilde{r}_1\right)^2 . \tag{45}$$

Conversely, if $T_1 = 1/R_1$ is of interest:

$$\mathbb{E}\left[t_1\right] = \sqrt{\exp\left(\tilde{\sigma}^2\right)}\, \exp\left(-\tilde{r}_1\right) , \tag{46}$$

$$\mathbb{V}\left[t_1\right] = \left(\exp\left(\tilde{\sigma}^2\right) - 1\right) \exp\left(\tilde{\sigma}^2\right) \exp\left(-\tilde{r}_1\right)^2 . \tag{47}$$

In this experiment, we used three echoes from the hMRI demonstration dataset to reconstruct maps using SPGR+JTV and SPGR-ML and computed the posterior uncertainty according to the above formula. This allowed us to map the posterior mean and standard deviation of log-R_1, R_1 and T_1. The ratio between the expected value of R_1 under the Laplace approximation and under a point estimate is given by $\sqrt{\exp\left(\tilde{\sigma}^2\right)}$.

4. Results

4.1. Convergence

The negative log-likelihood and gain after each fitting iteration in 50 randomly selected voxels are shown in Fig. 4. These plots show that all voxels converge monotonically with our approach (the gain is never negative) despite the very wide range of parameters (*e.g.*, SNR, observed relaxation times) covered by the simulations. Conversely, Gauss-Newton and Levenberg-Marquardt fail to converge to the optimum in a large number of voxels, despite the fact that this simulation context is very favourable to them: they are allowed to define one regularisation parameter per voxel, whereas in a spatially regularised case, voxels would not be independent and would have to share the same parameter.

Note that under this setup (3 contrasts, 5 echoes per contrasts), the vast majority of these maximum-likelihood fits yield optimal negative log-likelihood that are lower than the expected negative log-likelihood of the true parameters. This shows that ML likely over-fits the data and is strong evidence that regularisation is needed.

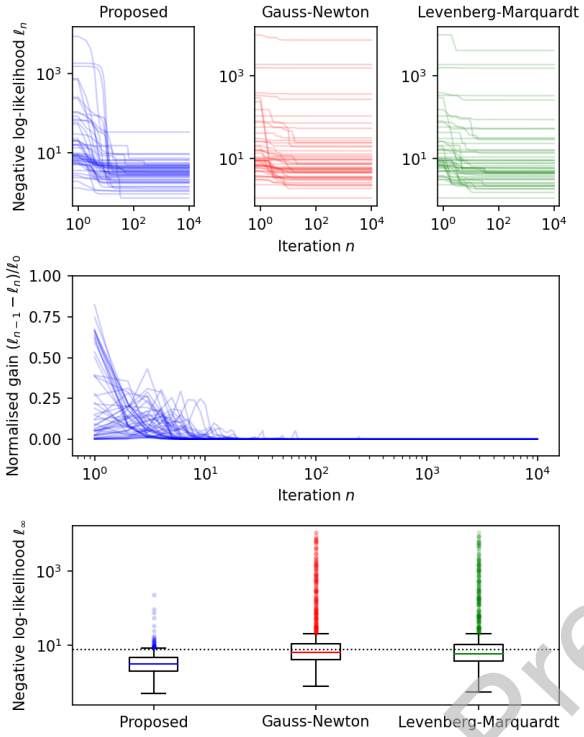


Figure 4: Optimisation of 1,000 simulated voxels for 10,000 iterations using the proposed second order method (left), Gauss-Newton (middle) and Levenberg-Marquardt (right). The top graph shows the negative log-likelihood per iteration for 50 random samples, while the middle graph shows the normalised gain per iteration for these same 50 samples. Gain is only shown for the proposed method. Note that if an iteration was non-monotonic, its gain would be negative, which never happens in our simulations. The bottom graph shows the distribution of the final log-likelihood of all 1,000 samples. The dashed line represents the expected log-likelihood of the true parameters. Log-likelihood and iterations are shown in log scale.

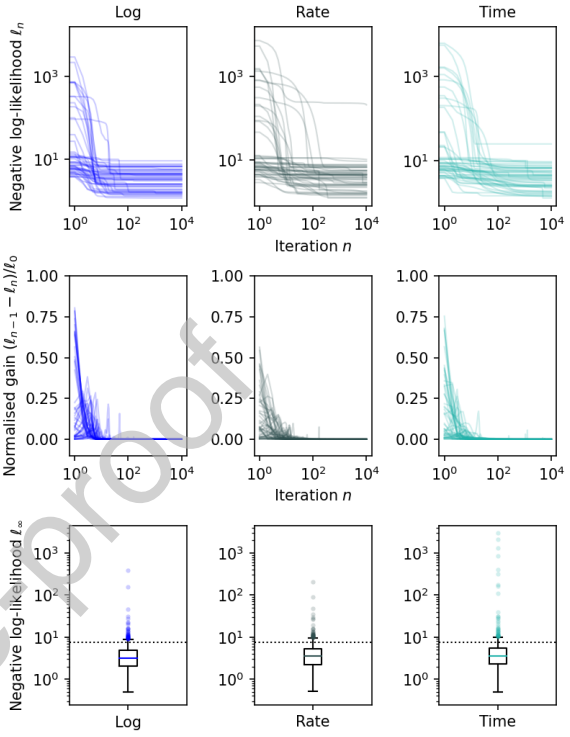


Figure 5: Optimisation of 1,000 simulated voxels for 10,000 iterations using the proposed second order method. The optimisation is either performed in the basis of the log-parameters (left), the rate representation (middle) or the time representation (right). The top graph shows the negative log-likelihood per iteration for 50 random samples, while the middle graph shows the normalised gain per iteration for these same 50 samples. The bottom graph shows the distribution of the final log-likelihood of all 1,000 samples. The dashed line represents the expected log-likelihood of the true parameters. Log-likelihood and iterations are shown in log scale.

Fig. 5 shows results for the same experiment, this time comparing different optimisation basis (log, rate or time). While all methods converge to the same optimum, the rate of convergence is clearly higher when log-parameters are optimised.

4.2. Hyper-parameter selection

SPGR+JTV used the same regularisation factor (λ) for all contrasts. Values $\{1, 5, 10, 15, 20\}$ were tried and the smallest MSE was obtained with $\lambda = 10$. Because they have different dynamic scales, NONLIN+JTV used a different factor for the R_2^* map (λ_r) and for the log-intercepts (λ_i). We tried values $\{0.001, 0.01, 0.1, 1.0, 10\}$ and $\{100, 500, 1000, 5000, 10000\}$ respectively and found an optimal combination with $\lambda_r = 0.1$ and $\lambda_i = 1000$. For LOGLIN+AONLM, we used the optimal value $\beta = 0.2$ found in Balbastre et al. (2020).

4.3. Methods cross-validation

Quantitative maps obtained with all methods are displayed in Fig. 6, along with MSE. As expected, directly

fitting the parameters is beneficial over the use of sequential parameter estimation. SPGR+JTV (median MSE = 1459) outperform non-regularised approaches such as SPGR-ML (median MSE = 1706). Methods that make use of rational approximations all trail behind, with the regularised NONLIN+JTV (median 1716) again improving over the non-regularised NONLIN (median 2467). The log-linear version trails far behind (median MSE = 2873), with AONLM preprocessing only slighlty improving it (median MSE = 2799).

4.4. Echo decimation

Fig. 7 displays all four parameter maps (R_1, R_2^*, PD, $\mathrm{MT_{sat}}$) reconstructed using a growing number of echoes (2, 4, 6) with a regularised algorithm (SPGR+JTV) and a non-regularised one (SPGR-ML). The RMSE between all maps and a reference map (the combined ML solution of all 4 runs) was computed and the regularised maps show a much lower RMSE than the non-regularised ones, even when fewer echoes are used. Noise amplification in the center of the brain appears clearly in the SPGR-ML maps when only two echoes are used, whereas SPGR+JTV manages to preserve anatomical details in this region, despite some additional smoothing. Note that in our implementation, the noise variance and regularisation are two different parameters, even though they share a single degree of freedom in the optimisation. The noise variance is estimated on a volume-by-volume basis while the regularisation factor is kept *a priori* fixed. Consequently, the balance between the likelihood and the prior appears clearly in the maps displayed here, as more smoothing happens when less information (*e.g.*, fewer echoes) is present in the data.

4.5. Uncertainty mapping

The capacity to estimate uncertainty is exemplified via R_1: posterior expected R_1 and T_1 maps and their uncertainties are shown in Fig. 8. Uncertainty after a SPGR-ML fit is mostly driven by intensities: white matter has a higher uncertainty because its MR signal, and therefore SNR, is generally lower than the grey matter. Conversely, uncertainty after a SPGR+JTV fit is mostly driven by edges, since less local signal averaging is possible in their vicinity.

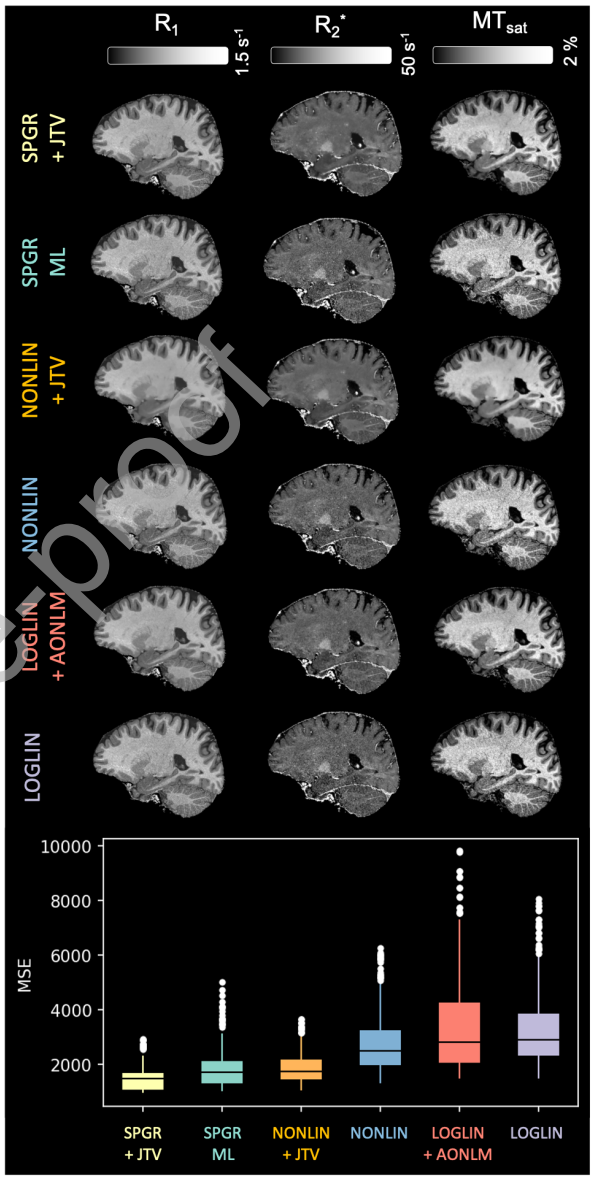


Figure 6: Top: R_1, R_2^* and $\mathrm{MT_{sat}}$ maps obtained with all methods. Bottom: z-normalised mean squared error between (unseen) predicted and acquired gradient-echo images. MSE was computed across all voxels in a single volume. Lower is better.

Journal Pre-proof

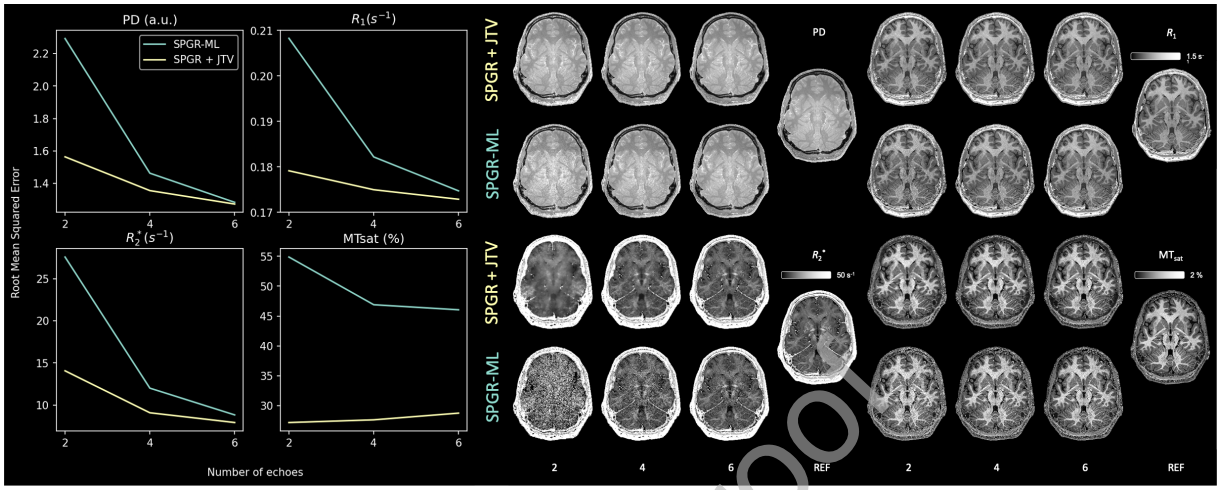


Figure 7: Decimation of the number of echoes. Maps were computed using the first 2, 4 or 6 echoes. Graphs on the left display the mean squared error (lower is better) relative to the reference maps (SPGR-ML with four repeats of 8 echoes) as a function of the number of echoes used. The corresponding parameter maps are displayed on the right.

5. Discussion & Conclusion

In this paper, we introduced an optimisation-based framework for multi-parameter mapping that directly fits the signal equations to all available SPGR data, and can work with (or without) a wide range of regularisers. The key innovation of our method is a second-order optimiser that makes use of a bespoke approximate Hessian that shows excellent experimental convergence. We showed on toy problems that our update steps have theoretical properties that ensure monotonic convergence even when the quadratic approximation is not a majoriser of the objective function. Although we did not offer analytical proof that this monotonic convergence extends to the multivariate case, our solution is at least as good as Gauss-Newton, which is known to converge to the optimum when it converges. Our optimiser therefore benefits from the same property and, experimentally, was never found to diverge.

We also introduced a novel encoding of the relaxation parameters, based on their log, that makes regularisation insensitive to the choice of unit. This encoding has the additional advantage that parameters live on the entire real line, making the optimisation problem unconstrained.

We evaluated the use of a joint total variation regulariser, an edge-preserving prior that introduces implicit correlations between channels, increasing its denoising power over parameter-wise total variation. We show that this prior yields the best parameter inference, under the metric that inferred parameters best predict unseen data. Although this paper focuses on relatively uninformative spatial regularisers, our method opens the door to the use of informative learned priors. One possibility, inspired by Brudfors et al. (2019, 2020), is to include a multivariate Gaussian mixture model of the log-parameter maps, with learned tissue parameters (means, covariances, tissue probability maps). Alternatively, priors based on learned dictionaries of patches could be used (Dalca et al., 2017).

Although this was not the topic of this paper, it is important to investigate some properties of the maximum-likelihood estimator. While our second-order scheme may output an approximate posterior distribution under the Laplace approximation, it is centered about the mode of the posterior, which is certainly not aligned with the posterior expected value. Different corrections could be devised depending on the map of interest (*e.g.*, R_1, T_1 or $\pm \log R_1$).

The Laplace approximation allowed posterior uncertainty to be estimated. These uncertainty maps are far from perfect: they only contain aleatory uncertainty (*i.e.*,

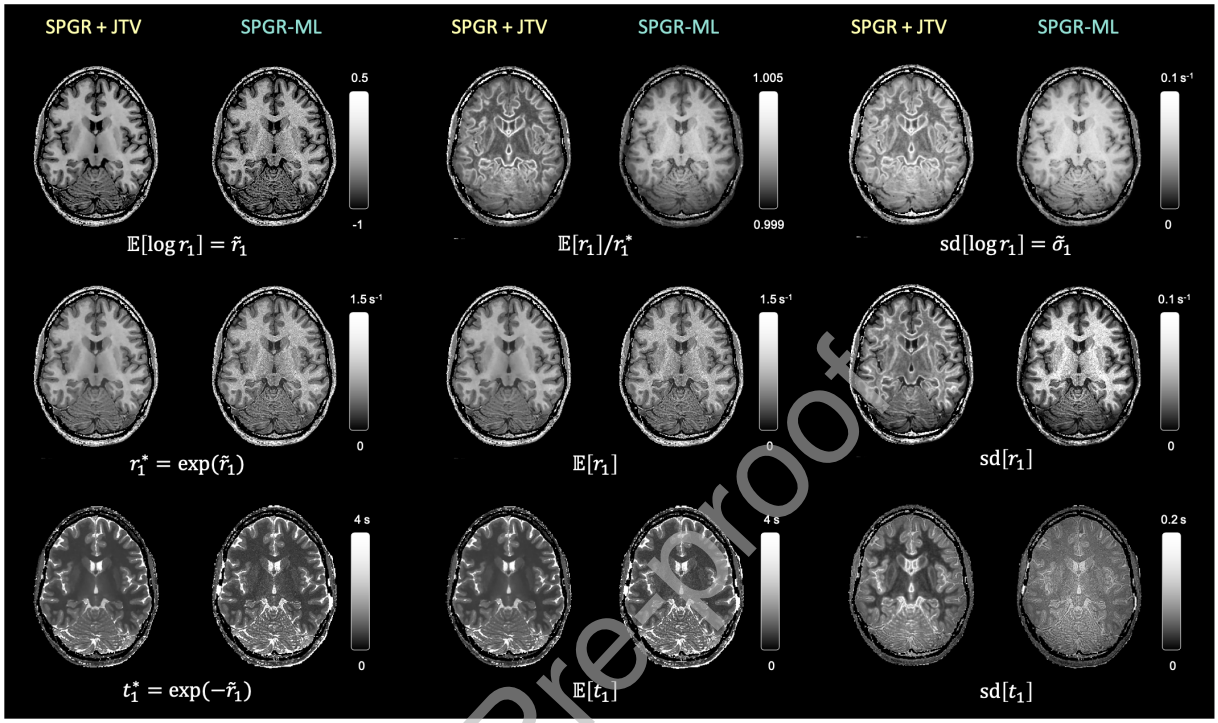


Figure 8: Uncertainty mapping and propagation. Log parameter maps ($\mathbb{E}[\log r_1] = \tilde{r}_1$) and their uncertainty ($\mathrm{sd}[\log r_1]$) were estimated using SPGR+JTV, which includes an edge-preserving prior and SPGR-ML, which has no (or infinitely uninformative) prior. The expected value and standard deviation of r_1 and t_1 under a Laplace posterior ($\mathbb{E}[r_1]$) or a peaked posterior ($\exp(\tilde{r}_1)$) are shown.

uncertainty that stems from the probabilistic nature of the variables in play, that is known *and* can be modelled), and are based on the assumption that the $B_1^{+/-}$ fields and thermal noise variance are known, thereby discarding any uncertainty about their values. Furthermore, it relies on a Gaussian assumption that could be very wrong. Nonetheless, all models are wrong but some are useful, and – while flawed – these uncertainty maps are better than blind confidence. They could be used to *e.g.* down-weight low-quality data points in downstream regression or classification tasks.

Finally, transmit and receive field maps are assumed pre-computed and fixed in this work and spoiling is assumed perfect. However, it has been shown that key sources of variance in T_1 mapping come from imperfect spoiling and errors in the estimate of the B_1^+ field (Stikov et al., 2015). In theory, transmit field map estimation could be integrated in the optimisation framework as long as the corresponding imaging data is included as well (Hurley et al., 2012). However, B_0 and B_1^+ mapping protocols based on phase differences would be difficult to fit using second-order methods because of the circular support of phase values (Fessler and Noll, 2004). On the other hand, imperfect spoiling is challenging to integrate in the generative model because, in its current form, it is based on full numerical Bloch simulations and *a posteriori* transformation of *apparent* T_1 maps under a polynomial whose coefficients depend on the B_1^+ (Preibisch and Deichmann, 2009). Of course, those *post hoc* corrections can still be applied to the parameters estimated here.

Some remaining approximations could also be removed: (1) our Gaussian noise model could be replaced by a more accurate non-central Chi or Rice model that can be fitted using a majorisation-minimisation frame-

Journal Pre-proof

work similar to the one used for optimising the JTV functional in this work (Varadarajan and Haldar, 2015); (2) the true TR of the off-resonance pulse could be used instead of being set to zero (Mohammadi et al., 2017), which may remove some residual bias in the R_1 and MT_{sat} maps. Furthermore, the generative model could be extended to more complex biophysical models, such as multi-compartment models (Deoni et al., 2008a), and to additional sequences, such as spin echo or inversion recovery. Diffusion data could also be integrated and modelled (Jian et al., 2007), as joint diffusion-relaxometry datasets are increasingly being used for accurate microstructure modelling (Hutter et al., 2018). Note that the diffusion signal can also be described using exponential-linear models, which are likely to benefit from our improved Hessian.

On the computational side, this method would benefit from any progress made on TV optimisation. Our implementation uses a relatively simple conjugate-gradient optimiser with Jacobi preconditioning, and it is certain that using preconditioners better tailored to TV would improve its convergence. Additionally, although we experimentally show the stability of our approximate Hessian, we did not provide a proof of convergence in this work. A formal proof that builds on our analysis would surely help move the field forward.

In conclusion, we have presented a flexible probabilistic generative modelling framework, which capitalises on a novel approximate Hessian to enable the robust joint estimation of multiple quantitative MRI parameters. The framework can be adapted to incorporate spatial projections, denoising schemes and additional priors. With this framework we have demonstrated enhanced performance relative to the consecutive estimation techniques commonly used in multi-parameter mapping, with additional benefits such as uncertainty mapping.

Acknowledgements

YB, MFC and JA were funded by the MRC and Spinal Research Charity through the ERA-NET Neuron joint call (MR/R000050/1). YB was partially supported by the National Institutes of Health under award numbers U01MH117023, R01AG064027 and P41EB030006. CL is supported by an MRC Clinician Scientist award (MR/R006504/1). The Wellcome Centre for Human Neuroimaging is supported by core funding from the Wellcome [203147/Z/16/Z].

References

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. NeuroImage 38, 95–113.

Ashburner, J., Brudfors, M., Bronik, K., Balbastre, Y., 2018. An Algorithm for Learning Shape and Appearance Models without Annotations. NeuroImage 55, 197–215.

Ashburner, J., Ridgway, G.R., 2013. Symmetric diffeomorphic modeling of longitudinal structural mri. Frontiers in neuroscience 6, 197.

Bach, F., 2011. Optimization with Sparsity-Inducing Penalties. FNT in Machine Learning 4, 1–106.

Balbastre, Y., Brudfors, M., Azzarito, M., Lambert, C., Callaghan, M.F., Ashburner, J., 2020. Joint Total Variation ESTATICS for Robust Multi-parameter Mapping, in: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (Eds.), Medical Image Computing and Computer Assisted Intervention MICCAI 2020, Springer International Publishing, Cham. pp. 53–63.

Bauer, C.M., Jara, H., Killiany, R., 2010. Whole brain quantitative T2 MRI across multiple scanners with dual echo FSE: Applications to AD, MCI, and normal aging. NeuroImage 52, 508–514.

Beck, A., Teboulle, M., 2009. Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems. IEEE Transactions on Image Processing 18, 2419–2434. Conference Name: IEEE Transactions on Image Processing.

Bloch, F., 1946. Nuclear induction. Physical Review 70, 460–474. 07663.

Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J., 2011. Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Foundations and Trends in Machine Learning 3, 1–122.

Journal Pre-proof

Brudfors, M., Ashburner, J., Nachev, P., Balbastre, Y., 2019. Empirical Bayesian Mixture Models for Medical Image Translation, in: Burgos, N., Gooya, A., Svoboda, D. (Eds.), Simulation and Synthesis in Medical Imaging, Springer International Publishing, Cham. pp. 1–12.

Brudfors, M., Balbastre, Y., Flandin, G., Nachev, P., Ashburner, J., 2020. Flexible Bayesian Modelling for Nonlinear Image Registration, in: Martel, A.L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M.A., Zhou, S.K., Racoceanu, D., Joskowicz, L. (Eds.), Medical Image Computing and Computer Assisted Intervention MICCAI 2020, Springer International Publishing, Cham. pp. 253–263.

Brudfors, M., Balbastre, Y., Nachev, P., Ashburner, J., 2018. MRI Super-Resolution using Multi-Channel Total Variation, in: 22nd Conference on Medical Image Understanding and Analysis, Southampton, UK. 00000.

Callaghan, M.F., Lutti, A., Ashburner, J., Balteau, E., Corbin, N., Draganski, B., Helms, G., Kherif, F., Leutritz, T., Mohammadi, S., Phillips, C., Reimer, E., Ruthotto, L., Seif, M., Tabelow, K., Ziegler, G., Weiskopf, N., 2019. Example dataset for the hMRI toolbox. Data in Brief 25, 104132.

Campbell, J.S.W., Leppert, I.R., Narayanan, S., Boudreau, M., Duval, T., Cohen-Adad, J., Pike, G.B., Stikov, N., 2018. Promise and pitfalls of g-ratio estimation with MRI. NeuroImage 182, 80–96.

Cercignani, M., Alexander, D.C., 2006. Optimal acquisition schemes for in vivo quantitative magnetization transfer MRI. Magnetic Resonance in Medicine 56, 803–810.

Chang, L.C., Koay, C.G., Basser, P.J., Pierpaoli, C., 2008. Linear least-squares method for unbiased estimation of T1 from SPGR signals. Magnetic Resonance in Medicine 60, 496–501.

Chen, Z., Prato, F.S., McKenzie, C., 1998. T1 fast acquisition relaxation mapping (T1/-FARM): an optimized reconstruction. IEEE Transactions on Medical Imaging 17, 155–160. Conference Name: IEEE Transactions on Medical Imaging.

Chun, I.Y., Fessler, J.A., 2018. Convolutional Dictionary Learning: Acceleration and Convergence. IEEE Transactions on Image Processing 27, 1697–1712.

Corbin, N., Callaghan, M.F., 2021. Imperfect spoiling in variable flip angle T1 mapping at 7T: Quantifying and minimizing impact. Magnetic Resonance in Medicine .

Coupé, P., Manjón, J., Robles, M., Collins, D., 2012. Adaptive multiresolution non-local means filter for three-dimensional magnetic resonance image denoising. IET Image Process. 6, 558–568.

Coupe, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C., 2008. An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images. IEEE T. Med. Imaging 27, 425–441.

Dalca, A.V., Bouman, K.L., Freeman, W.T., Rost, N.S., Sabuncu, M.R., Golland, P., 2017. Population Based Image Imputation, in: Niethammer, M., Styner, M., Aylward, S., Zhu, H., Oguz, I., Yap, P.T., Shen, D. (Eds.), Information Processing in Medical Imaging, Springer International Publishing, Cham. pp. 659–671.

Daubechies, I., Defrise, M., Mol, C.D., 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. Communications on Pure and Applied Mathematics 57, 1413–1457.

Daubechies, I., DeVore, R., Fornasier, M., Güntürk, C.S., 2010. Iteratively reweighted least squares minimization for sparse recovery. Commun. Pure Appl. Math. 63, 1–38.

Deoni, S.C.L., Rutt, B.K., Arun, T., Pierpaoli, C., Jones, D.K., 2008a. Gleaning multicomponent T1 and T2 information from steady-state imaging data. Magnetic Resonance in Medicine 60, 1372–1387.

Deoni, S.C.L., Williams, S.C.R., Jezzard, P., Suckling, J., Murphy, D.G.M., Jones, D.K., 2008b. Standardized structural magnetic resonance imaging in multicentre studies using quantitative T1 and T2 imaging at 1.5 T. NeuroImage 40, 662–671.

Dick, F., Tierney, A.T., Lutti, A., Josephs, O., Sereno, M.I., Weiskopf, N., 2012. In Vivo Functional and Myeloarchitectonic Mapping of Human Primary Auditory Areas. J. Neurosci. 32, 16095–16105.

Journal Pre-proof

Donoho, D., 1995. De-noising by soft-thresholding. IEEE T. Inform. Theory 41, 613–627.

Fessler, J.A., Noll, D.C., 2004. Iterative image reconstruction in MRI with separate magnitude and phase regularization, in: 2004 2nd IEEE International Symposium on Biomedical Imaging: Nano to Macro (IEEE Cat No. 04EX821), pp. 209–212 Vol. 1.

Gong, T., Tong, Q., He, H., Sun, Y., Zhong, J., Zhang, H., 2020. MTE-NODDI: Multi-TE NODDI for disentangling non-T2-weighted signal fractions from compartment-specific T2 relaxation times. NeuroImage 217, 116906.

Gupta, R.K., 1977. A new look at the method of variable nutation angle for the measurement of spin-lattice relaxation times using fourier transform NMR. Journal of Magnetic Resonance (1969) 25, 231–235.

Hasan, K.M., Walimuni, I.S., Kramer, L.A., Narayana, P.A., 2012. Human brain iron mapping using atlas-based T2 relaxometry. Magn. Reson. Med. 67, 731–739.

Helms, G., Dathe, H., Dechent, P., 2008a. Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation. Magn. Reson. Med. 59, 667–672.

Helms, G., Dathe, H., Kallenberg, K., Dechent, P., 2008b. High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and T1 relaxation obtained from 3D FLASH MRI. Magn. Reson. Med. 60, 1396–1407.

Huang, J., Chen, C., Axel, L., 2012. Fast Multi-contrast MRI Reconstruction, in: Ayache, N., Delingette, H., Golland, P., Mori, K. (Eds.), MICCAI 2012, Springer, Berlin, Heidelberg. pp. 281–288.

Hurley, S.A., Yarnykh, V.L., Johnson, K.M., Field, A.S., Alexander, A.L., Samsonov, A.A., 2012. Simultaneous variable flip angleactual flip angle imaging method for improved accuracy and precision of three-dimensional T1 and B1 measurements. Magnetic Resonance in Medicine 68, 54–64.

Hutter, J., Slator, P.J., Christiaens, D., Teixeira, R.P.A.G., Roberts, T., Jackson, L., Price, A.N., Malik, S., Hajnal, J.V., 2018. Integrated and efficient diffusion-relaxometry using ZEBRA. Scientific Reports 8, 15138. Number: 1 Publisher: Nature Publishing Group.

Jian, B., Vemuri, B.C., zarslan, E., Carney, P.R., Mareci, T.H., 2007. A novel tensor distribution model for the diffusion-weighted MR signal. NeuroImage 37, 164–176.

Jiru, F., Klose, U., 2006. Fast 3D radiofrequency field mapping using echo-planar imaging. Magnetic Resonance in Medicine 56, 1375–1379.

Langkammer, C., Krebs, N., Goessler, W., Scheurer, E., Ebner, F., Yen, K., Fazekas, F., Ropele, S., 2010. Quantitative MR Imaging of Brain Iron: A Postmortem Validation Study. Radiology 257, 455–462.

Liberman, G., Louzoun, Y., Bashat, D.B., 2014. T1 Mapping using variable flip angle SPGR data with flip angle correction. Journal of Magnetic Resonance Imaging 40, 171–180.

Liu, T., Liu, J., de Rochefort, L., Spincemaille, P., Khalidov, I., Ledoux, J.R., Wang, Y., 2011. Morphology enabled dipole inversion (MEDI) from a single-angle acquisition: Comparison with COSMOS in human brain imaging. Magn. Reson. Med. 66, 777–783.

Lutti, A., Dick, F., Sereno, M.I., Weiskopf, N., 2014. Using high-resolution quantitative mapping of R1 as an index of cortical myelination. NeuroImage 93, 176–188.

Lutti, A., Hutton, C., Finsterbusch, J., Helms, G., Weiskopf, N., 2010. Optimization and validation of methods for mapping of the radiofrequency transmit field at 3T. Magn. Reson. Med. 64, 229–238.

MacKay, D.J., 1998. Choice of Basis for Laplace Approximation. Machine Learning 33, 77–86.

MacKay, D.J.C., 2003. Information Theory, Inference and Learning Algorithms. Sixth printing 2007 edition ed., Cambridge University Press, Cambridge, UK ; New York.

Journal Pre-proof

Manjón, J.V., Coupé, P., Buades, A., Louis Collins, D., Robles, M., 2012. New methods for MRI denoising based on sparseness and self-similarity. Med. Image Anal. 16, 18–27.

Manjón, J.V., Coupé, P., Martí-Bonmatí, L., Collins, D.L., Robles, M., 2010. Adaptive non-local means denoising of MR images with spatially varying noise levels. J. Magn. Reson. Imaging 31, 192–203.

Mascarenhas, W.F., 2014. The divergence of the BFGS and Gauss Newton methods. Mathematical Programming 147, 253–276.

Milford, D., Rosbach, N., Bendszus, M., Heiland, S., 2015. Mono-Exponential Fitting in T2-Relaxometry: Relevance of Offset and First Echo. PLOS ONE 10, e0145255. Publisher: Public Library of Science.

Mohammadi, S., D'Alonzo, C., Ruthotto, L., Polzehl, J., Ellerbrock, I., Callaghan, M.F., Weiskopf, N., Tabelow, K., 2017. Simultaneous adaptive smoothing of relaxometry and quantitative magnetization transfer mapping. Weierstrass Institute for Applied Analysis and Stochastics: Preprint 2432 doi:10.20347/WIAS.PREPRINT.2432.

Ogg, R.J., Langston, J.W., Haacke, E.M., Steen, R.G., Taylor, J.S., 1999. The correlation between phase shifts in gradient-echo MR images and regional brain iron concentration. Magn. Reson. Imaging 17, 1141–1148.

Ordidge, R.J., Gorell, J.M., Deniau, J.C., Knight, R.A., Helpern, J.A., 1994. Assessment of relative brain iron concentrations using T2-weighted and T2*-weighted MRI at 3 Tesla. Magn. Reson. Med. 32, 335–341.

Papp, D., Callaghan, M.F., Meyer, H., Buckley, C., Weiskopf, N., 2016. Correction of inter-scan motion artifacts in quantitative R1 mapping by accounting for receive coil sensitivity effects. Magn. Reson. Med. 76, 1478–1485.

Polders, D.L., Leemans, A., Luijten, P.R., Hoogduin, H., 2012. Uncertainty estimations for quantitative in vivo MRI T1 mapping. Journal of Magnetic Resonance 224, 53–60.

Preibisch, C., Deichmann, R., 2009. Influence of RF spoiling on the stability and accuracy of T1 mapping based on spoiled FLASH with varying flip angles. Magnetic Resonance in Medicine 61, 125–135.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P., 2007. Numerical Recipes 3rd Edition: The Art of Scientific Computing. 3 edition ed., Cambridge University Press, Cambridge, UK ; New York.

Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. Physica D: Nonlinear Phenomena 60, 259–268.

Sacolick, L.I., Wiesinger, F., Hancu, I., Vogel, M.W., 2010. B1 mapping by Bloch-Siegert shift. Magnetic Resonance in Medicine 63, 1315–1322.

Sapiro, G., Ringach, D.L., 1996. Anisotropic diffusion of multivalued images with applications to color filtering. IEEE T. Image Process. 5, 1582–1586.

Scholand, N., Wang, X., Rosenzweig, S., Holme, H.C.M., Uecker, M., 2020. Generic Quantitative MRI using Model-Based Reconstruction with the Bloch Equations, in: Proc. Intl. Soc. Mag. Reson. Med.

Sereno, M.I., Lutti, A., Weiskopf, N., Dick, F., 2013. Mapping the Human Cortical Surface by Combining Quantitative T1 with Retinotopy. Cereb. Cortex 23, 2261–2268.

Sigalovsky, I.S., Fischl, B., Melcher, J.R., 2006. Mapping an intrinsic MR property of gray matter in auditory cortex of living humans: A possible marker for primary cortex and hemispheric differences. NeuroImage 32, 1524–1537.

Sijbers, J., Dekker, A.J.d., Raman, E., Dyck, D.V., 1999. Parameter estimation from magnitude MR images. International Journal of Imaging Systems and Technology 10, 109–114.

Stikov, N., Boudreau, M., Levesque, I.R., Tardif, C.L., Barral, J.K., Pike, G.B., 2015. On the accuracy of T1 mapping: Searching for common ground. Magnetic Resonance in Medicine 73, 514–522.

Journal Pre-proof

Sumpf, T.J., Petrovic, A., Uecker, M., Knoll, F., Frahm, J., 2014. Fast T2 Mapping With Improved Accuracy Using Undersampled Spin-Echo MRI and Model-Based Reconstructions With a Generating Function. IEEE Transactions on Medical Imaging 33, 2213–2222. Conference Name: IEEE Transactions on Medical Imaging.

Sumpf, T.J., Uecker, M., Boretius, S., Frahm, J., 2011. Model-based nonlinear inverse reconstruction for T2 mapping using highly undersampled spin-echo MRI. Journal of Magnetic Resonance Imaging 34, 420–428.

Tabelow, K., Balteau, E., Ashburner, J., Callaghan, M.F., Draganski, B., Helms, G., Kherif, F., Leutritz, T., Lutti, A., Phillips, C., Reimer, E., Ruthotto, L., Seif, M., Weiskopf, N., Ziegler, G., Mohammadi, S., 2019. hMRI – A toolbox for quantitative MRI in neuroscience and clinical research. NeuroImage 194, 191–210.

Tisdall, M.D., 2017. Bias and SNR of T1 estimates derived from joint fitting of actual flip-angle and FLASH imaging data with variable flip angles. Proc. Intl. Soc. Mag. Reson. Med. 25, 1445.

Tisdall, M.D., van der Kouwe, A.J.W., 2016. Efficient algorithm for maximum likelihood estimate and confidence intervals of T1 from multi-flip, multi-echo FLASH, in: Proc. Intl. Soc. Mag. Reson. Med.

Tofts, P.S., 2003. Quantitative MRI of the Brain. 1 ed., John Wiley & Sons, Ltd.

Tofts, P.S., Steens, S.C.A., van Buchem, M.A., 2003. MT: Magnetization Transfer, in: Quantitative MRI of the Brain. John Wiley & Sons, Ltd, pp. 257–298.

Tofts, P.S., Steens, S.C.A., Cercignani, M., Admiraal-Behloul, F., Hofman, P.A.M., van Osch, M.J.P., Teeuwisse, W.M., Tozer, D.J., van Waesberghe, J.H.T.M., Yeung, R., Barker, G.J., van Buchem, M.A., 2006. Sources of variation in multi-centre brain MTR histogram studies: Body-coil transmission eliminates inter-centre differences. Magn. Reson. Mater. Phy. 19, 209–222.

Varadarajan, D., Frost, R., van der Kouwe, A., Morgan, L., Diamond, B., Boyd, E., Fogarty, M., Stevens, A., Fischl, B., Polimeni, J.R., 2020. Edge-preserving B0 inhomogeneity distortion correction for high-resolution multi-echo ex vivo MRI at 7T, in: Proc. Intl. Soc. Mag. Reson. Med.

Varadarajan, D., Haldar, J.P., 2015. A Majorize-Minimize Framework for Rician and Non-Central Chi MR Images. IEEE transactions on medical imaging 34, 2191–2202.

Wang, H.Z., Riederer, S.J., Lee, J.N., 1987. Optimizing the precision in T1 relaxation estimation using limited flip angles. Magnetic Resonance in Medicine 5, 399–416.

Weiskopf, N., Callaghan, M.F., Josephs, O., Lutti, A., Mohammadi, S., 2014. Estimating the apparent transverse relaxation time (R2*) from images with different contrasts (ESTATICS) reduces motion artifacts. Front. Neurosci. 8.

Weiskopf, N., Suckling, J., Williams, G., Correia, M.M., Inkster, B., Tait, R., Ooi, C., Bullmore, E.T., Lutti, A., 2013. Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: A multi-center validation. Front. Neurosci. 7.

Whittall, K.P., MacKay, A.L., 1989. Quantitative interpretation of NMR relaxation data. Journal of Magnetic Resonance (1969) 84, 134–152.

Wiggermann, V., Vavasour, I.M., Kolind, S.H., MacKay, A.L., Helms, G., Rauscher, A., 2020. Non-negative least squares computation for in vivo myelin mapping using simulated multi-echo spin-echo T2 decay data. NMR in Biomedicine 33, e4277.

Yarnykh, V.L., 2007. Actual flip-angle imaging in the pulsed steady state: A method for rapid three-dimensional mapping of the transmitted radiofrequency field. Magnetic Resonance in Medicine 57, 192–200.

Zhao, B., Lam, F., Liang, Z., 2014. Model-Based MR Parameter Mapping With Sparsity Constraints: Parameter Estimation and Performance Bounds. IEEE Transactions on Medical Imaging 33, 1832–1844. Conference Name: IEEE Transactions on Medical Imaging.

Journal Pre-proof

Journal Pre-proof

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof