TU Delft

Delft University of Technology

A unified Maximum Likelihood framework for simultaneous motion and T1 estimation in quantitative MR T1 mapping

Ramos-Llorden, Gabriel; Den Dekker, Arnold J.; Van Steenkiste, Gwendolyn Van; Jeurissen, Ben; Vanhevel, Floris; Audekerke, Johan Van; Verhoye, Marleen; Sijbers, Jan

DOI
10.1109/TMI.2016.2611653

Publication date
2017

Document Version
Accepted author manuscript

Published in
IEEE Transactions on Medical Imaging

Citation (APA)
Ramos-Llorden, G., Den Dekker, A. J., Van Steenkiste, G. V., Jeurissen, B., Vanhevel, F., Audekerke, J. V., Verhoye, M., & Sijbers, J. (2017). A unified Maximum Likelihood framework for simultaneous motion and T1 estimation in quantitative MR T1 mapping. *IEEE Transactions on Medical Imaging*, *36*(2), 433-446. https://doi.org/10.1109/TMI.2016.2611653

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.
For technical reasons the number of authors shown on this cover page is limited to a maximum of 10.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

A unified Maximum Likelihood framework for simultaneous motion and T_1 estimation in quantitative MR T_1 mapping

Gabriel Ramos-Llordén*, *Member, IEEE,* Arnold J. den Dekker, Gwendolyn Van Steenkiste, Ben Jeurissen, Floris Vanhevel, Johan Van Audekerke, Marleen Verhoye and Jan Sijbers

***Abstract*—In quantitative MR T_1 mapping, the spin-lattice relaxation time T_1 of tissues is estimated from a series of T_1-weighted images. As the T_1 estimation is a voxel-wise estimation procedure, correct spatial alignment of the T_1-weighted images is crucial. Conventionally, the T_1-weighted images are first registered based on a general-purpose registration metric, after which the T_1 map is estimated. However, as demonstrated in this paper, such a two-step approach leads to a bias in the final T_1 map. In our work, instead of considering motion correction as a preprocessing step, we recover the motion-free T_1 map using a unified estimation approach. In particular, we propose a unified framework where the motion parameters and the T_1 map are simultaneously estimated with a Maximum Likelihood (ML) estimator. With our framework, the relaxation model, the motion model as well as the data statistics are jointly incorporated to provide substantially more accurate motion and T_1 parameter estimates. Experiments with realistic Monte Carlo simulations show that the proposed unified ML framework outperforms the conventional two-step approach as well as state-of-the-art model-based approaches, in terms of both motion and T_1 map accuracy and mean-square error. Furthermore, the proposed method was additionally validated in a controlled experiment with real T_1-weighted data and with two in vivo human brain T_1-weighted data sets, showing its applicability in real-life scenarios.**

***Index Terms*—T_1 mapping, Maximum Likelihood, motion correction, dynamic MRI, registration**

I. Introduction

Quantitative T_1 mapping is a Magnetic Resonance Imaging (MRI) technique in which the spin-lattice relaxation time T_1 of tissues is measured [1]. Because T_1 depends on biophysical properties, it is used as biomarker in a broad range of diseases, such as multiple sclerosis [2], epilepsy [3] and Alzheimer's disease [4], as well as in the measurement of perfusion [5] and blood flow [6]. Hence, its accurate and precise estimation is of uttermost importance [1], [7]. In order to quantify T_1, a set of T_1-weighted images with different sequence settings needs to be acquired [1], [8], [9]. From this set, a spatial map of T_1 values can be calculated by fitting a known relaxation model at every voxel. Evidently, to obtain a meaningful T_1 map, spatial correspondence between the images in the acquired series is crucial [10]. However, due to patient motion and/or apparent spatial shifts introduced by the scanner (e.g., scanner drift [11]), T_1-weighted images are often misaligned.

To deal with this problem, T_1-weighted images are commonly spatially registered prior to the estimation of the T_1 map [12], [13]. This is often done by choosing one T_1-weighted image as a target and subsequently registering the remaining T_1-weighted images to this target image by using a similarity measure such as Mutual Information (MI) [14], [15].

Such an approach, however, suffers from inherent problems. First, the specific relation between the intensity value as a function of time of the (aligned) voxels is ignored. Second, the registration is not driven by a global optimization criterion that considers all T_1-weighted images simultaneously. Even more problematic is the fact that current motion correction is a *preprocessing* step prior to the estimation of the T_1 values. Such a two-step processing pipeline lacks a feedback mechanism between the image registration and the T_1 map estimation step. As a result, registration errors will propagate to the estimation step, leading to biased estimates [16].

Recently, progress in registration of T_1-weighted images was made by the introduction of model-based approaches. Such techniques integrate the signal model connecting the series of images (such as a T_1 relaxation model) into the registration step. State-of-the-art model-based methods have shown to outperform the conventional two-step approach in terms of accuracy, for example, in myocardial T_1 mapping [17], [18]. Unfortunately, they all come with serious limitations for precise and accurate T_1 mapping, mainly because different criteria for registration and estimation are heuristically combined [17], [18]. Since they do not constitute a truly unified framework, the output of the algorithms cannot be related to the optimal value of a given global information-based criterion. As a consequence, it is doubtful whether all

This work was supported by the Research Foundation - Flanders (FWO, Belgium) through project funding G037813N, and by the Belgian Science Policy Office through the Inter-university Attraction Poles Program (P7/11). The work of B. Jeurissen was supported by a postdoctoral grant given by the Research Foundation - Flanders (FWO, Belgium). *Asterisk indicates corresponding author.*

This paper has supplementary downloadable material available at http://ieeexplore.ieee.org, provided by the authors. This includes four animated GIFs showing real and simulated motion of several T_1-weighted images data sets, and an additional document with further technical content of the work presented here. This material is 2 MB in size.

*G. Ramos-Llordén, A. J. den Dekker, G. Van Steenkiste, B. Jeurissen and J. Sijbers are with the iMinds-Vision Lab, Department of Physics, University of Antwerp, 2610 Antwerp, Belgium (e-mail:Gabriel.Ramos-Llorden@uantwerpen.be).

A. J. den Dekker is also with the Delft Center for Systems and Control, Delft University of Technology, 2628 CD Delft, The Netherlands.

F. Vanhevel is with Department of Radiology, University of Antwerp, Antwerp University Hospital, 2650 Antwerp, Belgium.

J. Van Audekerke and M. Verhoye are with the Bio-Imaging Lab, Department of Biomedical Sciences, University of Antwerp, 2610 Antwerp, Belgium.

Copyright (c) 2010 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

the information gathered in the series of T_1-weighted images, including the data statistics [7], [19], is optimally exploited.

In our work, we propose an integrated model-based image registration and T_1 estimation approach, where the motion parameters and T_1 map are jointly estimated using a unified global information criterion, more specifically, the maximum likelihood (ML) criterion [19]–[21]. By combining models of T_1 relaxation, motion, and noise into one statistical model of the T_1-weighted images, we are able to restore the original motion-free T_1 map using a joint ML estimator. The unified ML framework allows to account for the statistical noise model, the relaxation model and the motion model simultaneously, exploiting, in addition to the temporal information, knowledge on data statistics. The large-scale ML optimization problem is solved by alternating between the estimation of motion and relaxation parameters in an efficient and robust manner, making use of block coordinate descent [22] and Majorize-Minimize (MM) algorithms [23]. Exact convergence properties of the algorithm are presented, demonstrating that the proposed iterative procedure leads to the ML estimates in a computationally efficient way.

We thoroughly validate the proposed joint maximum likelihood estimator (MLE) with realistic Monte Carlo (MC) simulations and compare it with the conventional two-step approach as well as the newest state-of-the-art model-based approach of Hallack [18]. We show that substantially more accurate T_1 maps as well as motion parameters can be obtained with our proposed joint MLE. Additionally, the T_1 maps estimated with the joint MLE are superior in terms of the root-mean-square error (RMSE). Apart from simulation experiments, we also quantitatively evaluate the performance of the joint MLE in a controlled experiment involving real T_1-weighted data. Further, we validate it with two in vivo human brain T_1-weighted data sets corrupted by patient motion, showing its applicability in real-life scenarios.

The remainder of the paper is organized as follows. In Section II, the image model used to construct the joint MLE is presented. Section III is devoted to the joint MLE algorithm. Section IV describes the experiments of which the results are presented in Section V, which is followed by a discussion in Section VI. Finally, conclusions are drawn in Section VII.

II. Theory

The derivation of the joint MLE requires a parametric statistical model of the images. This section is devoted to the derivation of such a model, which comprises a relaxation signal model, a motion model, and a statistical noise model.

A. Relaxation signal model

In the absence of noise, the evolution of the magnitude MRI signal in each voxel of a series of N T_1-weighted images can be described by a parametric model $\{f_n(\boldsymbol{\kappa}, T_1)\}_{n=1}^{N}$, where $\boldsymbol{\kappa}$ denotes a vector of nuisance parameters. The exact expression for this T_1-relaxation model depends on the pulse sequence that is used. In this work, we will use the Inversion Recovery (IR) sequence, being the gold standard for T_1-mapping [1]. Note that signal models corresponding with other sequences, such as SPoiled Gradient Recalled echo (SPGR) [12] or MOdified Look-Locker Inversion recovery (MOLLI) [24], can be accommodated within our framework as well. For the IR sequence, a common magnitude relaxation model is given by [25]:

$$f_n(a, b, T_1) = |a + be^{-\frac{TI_n}{T_1}}|, \quad \text{with} \quad n = 1, ..., N, \qquad (1)$$

where $\{TI_n\}_{n=1}^{N}$ are the inversion times. The other parameters a and b are related to, among others, the repetition time (TR), radio frequency (RF) pulse angles and the tissue-dependent proton density. The mathematical expressions that relate a and b to these quantities for the case of gradient-recalled-echo (GRE) IR and spin-echo (SE) IR can be found in [25].

To model the noiseless T_1-weighted images, we use a vector notation for the spatially varying parameters T_1, a and b. Let $\boldsymbol{r} = (x, y, z)^T$ be a vector in the Cartesian coordinate system in which they are defined. Then, a 3D spatial T_1 map of M voxels can be defined as a column vector, $\boldsymbol{T_1} \in \mathbb{R}^{M\times 1}$, where $[\boldsymbol{T_1}]_m$ represents T_1 defined at the spatial point $\boldsymbol{r}_m$, indexed by voxel m. Similarly, we define $\boldsymbol{a} \in \mathbb{R}^{M\times 1}$ and $\boldsymbol{b} \in \mathbb{R}^{M\times 1}$ as the parameter maps of a and b. For ease of readability and to alleviate the notation, we introduce the parameter vector $\boldsymbol{\kappa} = (\boldsymbol{a}^T, \boldsymbol{b}^T)^T \in \mathbb{R}^{2M\times 1}$. The relaxation model for the noiseless n-th T_1-weighted image is then given by

$$\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}) = |\boldsymbol{a} + \boldsymbol{b} \odot e^{-\frac{TI_n}{\boldsymbol{T_1}}}|, \qquad (2)$$

where $\odot$ and $|\cdot|$ denote the point-wise multiplication and point-wise modulus operator, respectively.

B. Motion model

In what follows, we will restrict the motion model of the unified ML framework to inter-image motion, that is, motion between the 3D T_1-weighted images, as in [18]. In the discussion section (Section VI), we further elaborate on extensions of the unified ML framework in which intra-image motion is incorporated, in particular, motion between the slices of a multi-slice T_1-weighted image.

The effect of inter-image motion is modeled by assuming that $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$ is observed in a different Cartesian coordinate system $\boldsymbol{r}^n$ for each acquisition $n = 1, ..., N$. In this work, we illustrate the joint MLE with rigid motion. Hence, the spatial point $\boldsymbol{r}_m^n$, with $m = 1, ..., M$, is related to the reference-system point $\boldsymbol{r}_m$, through a rigid transformation matrix, $\boldsymbol{M}_{\boldsymbol{\theta}_n} \in \mathbb{R}^{4\times 4}$ (in homogeneous coordinates), parameterized by

$$\boldsymbol{\theta}_n = (t_{xn}, t_{y_n}, t_{zn}, \alpha_n, \beta_n, \gamma_n)^T \quad , \qquad (3)$$

with t_{xn}, t_{y_n}, t_{zn} the translation parameters and α_n, β_n, γ_n are the Euler angles of the three elementary rotation matrices around axis x, y and z, respectively [26]. In our work, the reference system $\boldsymbol{r}$ is defined similarly as the intrinsic coordinate system which MATLAB uses to represent 3D images. That is, axis x points in the direction of increasing column index while y points in the direction of increasing row index. Finally, the axis z is aligned with the direction of increasing index of the third dimension. The origin of this coordinate

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

system is the center of the 3D image. Furthermore, in multi-slice acquisitions the axis z is aligned with the slice-encoding direction.

The noiseless T_1-weighted image observed at $\boldsymbol{r}^n$ can be modeled as the output of a linear operator that performs rigid motion, $\mathcal{H}_{\boldsymbol{\theta}_n}\{\cdot\}$, and whose input is the unobserved $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$. Because $\mathcal{H}_{\boldsymbol{\theta}_n}\{\cdot\}$ is linear, the input-output relation can be concisely written in matrix form as:

$$\tilde{\boldsymbol{f}}_n(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1}) = \boldsymbol{H}_{\boldsymbol{\theta}_n} \boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}), \tag{4}$$

where $\tilde{\boldsymbol{f}}_n(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1})$ is the motion-corrupted noiseless T_1-weighted image acquired at TI_n and $\boldsymbol{H}_{\boldsymbol{\theta}_n} \in \mathbb{R}^{M\times M}$ is the matrix representation of the linear motion operator $\mathcal{H}_{\boldsymbol{\theta}_n}\{\cdot\}$.

To design $\boldsymbol{H}_{\boldsymbol{\theta}_n} \in \mathbb{R}^{M\times M}$, we use the method proposed in [27], where it was demonstrated that each of the rotation matrices of $\boldsymbol{M}_{\boldsymbol{\theta}_n}$ can be decomposed as the product of three shear matrices. Each of the shearings is implemented very efficiently with Fast Fourier Transforms (FFT). Translation is implemented using an FFT as well.

With the FFT approach, the motion operator $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ can be shown to be unitary, which means that its inverse is given by $\boldsymbol{H}_{\boldsymbol{\theta}_n}^H$, where the superscript H denotes the Hermitian conjugate. Hence, the motion operator $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ is reversible, i.e., when applied to an image, this image can be retrieved by applying $\boldsymbol{H}_{\boldsymbol{\theta}_n}^H$ to the output of this operation. The unitarity property of the motion operator will turn out to be useful in the derivation of the joint MLE algorithm. Details of the exact analytical expression of $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ and the proof of the unitarity property are provided in Section I of the additional document which is included as part of the downloadable supplementary material which accompanies this paper.

C. Statistical noise model

In practice, acquired T_1-weighted images are inherently disturbed by noise. A typical data distribution for (single-coil) magnitude T_1-weighted images is the Rice distribution [28]:

$$p_s(s|\mu, \sigma) = \frac{s}{\sigma^2} e^{\frac{-(s^2+\mu^2)}{2\sigma^2}} I_0\left(\frac{s\mu}{\sigma^2}\right) u(s), \tag{5}$$

with μ the noiseless magnitude signal in a voxel, s the noise disturbed signal, $I_0(\cdot)$ the zeroth order modified Bessel function of the first kind, and σ the standard deviation of the Gaussian noise disturbing the underlying complex data [21]. The unit step function $u(\cdot)$ is used to indicate that the expression for the probability density function (PDF) of s is valid for nonnegative values of s only. Note that for high signal-to-noise ratio $\frac{\mu}{\sigma} > 3$, the Rician PDF becomes quasi Gaussian [29]. If multiple - instead of just one - receiving coils are used to acquire the data and the k-space is fully sampled (by each coil), the magnitude image that is reconstructed using the Sum of Squares (SoS) method obeys a noncentral chi (nc-χ) distribution, being the natural extension of the Rician distribution for the single-coil case [30]. When parallel MRI techniques that undersample the k-space to decrease the acquisition time are performed, such as SENSE or GRAPPA, other distributions may apply [31]. For a recent review on data distributions in MRI, the reader is referred to [21]. In this work, we will illustrate the proposed joint MLE by deriving it for the case of independent Rician distributed voxels, with different noise standard deviation σ for each voxel m and for each acquisition n. This is an accurate noise model for magnitude images that are reconstructed with SENSE [31]. It is also a valid noise model for magnitude images that are reconstructed with GRAPPA jointly with a spatial-matched-filter (SMF) data combination [32]. If, instead of SMF, SoS is used in combination of GRAPPA, the data distribution can be well approximated at high SNR by a Gaussian distribution with a spatially variant variance [33]. The derivation of the joint MLE for Gaussian distributed data will be covered in subsection III-E.

III. Joint MLE

Let $\boldsymbol{s}_n \in \mathbb{R}^{M\times 1}$, with $n = 1, ..., N$, denote an actual, noisy T_1-weighted image acquired at inversion time TI_n. Assuming Rician distributed data, it follows from Eq. (5) and the motion-corrupted noiseless T_1-weighted model Eq. (4) that the PDF of the voxels $[\boldsymbol{s}_n]_m$, $m = 1, \ldots, M$, of this image is given by

$$p_{[\boldsymbol{s}_n]_m}([\boldsymbol{s}_n]_m|[\tilde{\boldsymbol{f}}_n]_m, [\boldsymbol{\sigma}_n]_m) = \frac{[\boldsymbol{s}_n]_m}{[\boldsymbol{\sigma}_n]_m^2} e^{\frac{-([\boldsymbol{s}_n]_m^2+[\tilde{\boldsymbol{f}}_n]_m^2)}{2[\boldsymbol{\sigma}_n]_m^2}} \times I_0\left(\frac{[\boldsymbol{s}_n]_m[\tilde{\boldsymbol{f}}_n]_m}{[\boldsymbol{\sigma}_n]_m^2}\right) u([\boldsymbol{s}_n]_m). \tag{6}$$

Furthermore, if all voxels are assumed to be independent, the joint PDF of the voxels constituting the image $\boldsymbol{s}_n$ is given by the product of the PDFs of the individual voxels, i.e., $p_{\boldsymbol{s}_n}(\boldsymbol{s}_n|\tilde{\boldsymbol{f}}_n, \boldsymbol{\sigma}_n) = \prod_{m=1}^{M} p_{[\boldsymbol{s}_n]_m}([\boldsymbol{s}_n]_m|[\tilde{\boldsymbol{f}}_n]_m, [\boldsymbol{\sigma}_n]_m)$. Similarly, the joint PDF of the supposedly independent voxels of a set of N T_1-weighted images $\{\boldsymbol{s}_n\}_{n=1}^N$ is given by

$$p_{\boldsymbol{s}}(\boldsymbol{s}|\tilde{\boldsymbol{f}}, \boldsymbol{\sigma}) = \prod_{n=1}^{N} p_{\boldsymbol{s}_n}(\boldsymbol{s}_n|\tilde{\boldsymbol{f}}_n, \boldsymbol{\sigma}_n) \tag{7}$$

with $\boldsymbol{s} = (\boldsymbol{s}_1^T, \ldots, \boldsymbol{s}_N^T)^T$, $\tilde{\boldsymbol{f}} = (\tilde{\boldsymbol{f}}_1^T, \ldots, \tilde{\boldsymbol{f}}_N^T)^T$ and $\boldsymbol{\sigma} = (\boldsymbol{\sigma}_1^T, \ldots, \boldsymbol{\sigma}_N^T)^T$. Note that this joint PDF depends on the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T, \boldsymbol{\theta}_2^T, ..., \boldsymbol{\theta}_N^T)^T$, $\boldsymbol{\kappa}$ and $\boldsymbol{T_1}$ via $\tilde{\boldsymbol{f}}$ and can hence be written as $p_{\boldsymbol{s}}(\boldsymbol{s}|\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T_1}, \boldsymbol{\sigma})$. To construct the MLE of these parameters, the *likelihood function* must be derived. The likelihood function is obtained from the joint PDF, Eq. (7), by replacing the independent variables $\boldsymbol{s}$ by the actual acquired voxel intensity values - that is, by numbers - and the supposedly fixed, exact parameters $\boldsymbol{\theta}, \boldsymbol{\kappa}$ and $\boldsymbol{T_1}$ by independent variables. The likelihood function is, therefore, a function of the parameters considered as independent variables and is parametric in the acquired voxel intensities, from now on called *observations* [7]. To express this, the likelihood function is written as $L(\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s})$. Strictly speaking, the likelihood function also depends on $\boldsymbol{\sigma}$. However, in our work, we assume that $\boldsymbol{\sigma}$ can be estimated prior to the construction of the joint MLE using tailored noise estimation techniques [33], [34]. Hence, we omit the explicit σ-dependence in the notation.

To simplify the notation, let us define the parameter vector $\boldsymbol{\tau} = (\boldsymbol{\theta}^T, \boldsymbol{\kappa}^T, \boldsymbol{T_1}^T)^T$. The joint MLE $\hat{\boldsymbol{\tau}}_{\text{ML}}$ of $\boldsymbol{\tau}$ from the observations $\boldsymbol{s}$ is that value of $\boldsymbol{\tau}$ that maximizes the likelihood

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

function $L(\boldsymbol{\tau}|\boldsymbol{s})$, or equivalently, minimizes the so-called negative log-likelihood function $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\tau}|\boldsymbol{s}) \triangleq -\log L(\boldsymbol{\tau}|\boldsymbol{s})$ with respect to $\boldsymbol{\tau}$, i.e.,

$$\hat{\boldsymbol{\tau}}_{\text{ML}} = \arg\min_{\boldsymbol{\tau}} \mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\tau}|\boldsymbol{s}). \tag{8}$$

It follows from Eq. (7) that $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\tau}|\boldsymbol{s})$ can be written as

$$\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\tau}|\boldsymbol{s}) = \sum_{n=1}^{N} \mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s}_n), \tag{9}$$

with $\mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s}_n) = -\log p_{\boldsymbol{s}_n}(\boldsymbol{s}_n|\tilde{\boldsymbol{f}}_n, \boldsymbol{\sigma}_n)$.

As the joint MLE can not be found analytically, one has to resort to numerical optimization algorithms. In order to solve this very-large-scale optimization problem, a cyclic block-coordinate descent (cBCD) method was used [22]. cBCD methods work by iteratively minimizing the cost function $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s})$ with respect to a subset of the optimization variables, holding the remaining variables fixed [35], where in each iteration, the roles of the optimization and fixed variables are reversed [35]. The utility of the cBCD algorithm relies on a smart selection of the subset of optimization variables. In the case of our joint MLE, this subset is chosen to contain the motion parameters or the relaxation parameters. In this way, the very-large-scale optimization problem is separated into more easily solvable problems.

Indeed, alternating between the motion estimation problem and the relaxation estimation problem, the joint MLE is found in an efficient way. Moreover, the cBCD method assures that $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s})$ decreases at every iteration [22]. Therefore, convergence to at least a local minimum is guaranteed [36]. In summary, the cBCD-based joint MLE is obtained by the following iterative recursive procedure:

$$\hat{\boldsymbol{\theta}}^{(t+1)} = \arg\min_{\boldsymbol{\theta}} \sum_{n=1}^{N} \mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n, \hat{\boldsymbol{\kappa}}^{(t)}, \hat{\boldsymbol{T}_1}^{(t)}|\boldsymbol{s}_n), \tag{P.1}$$

$$\hat{\boldsymbol{\kappa}}^{(t+1)}, \hat{\boldsymbol{T}_1}^{(t+1)} = \arg\min_{\boldsymbol{\kappa}, \boldsymbol{T_1}} \sum_{n=1}^{N} \mathcal{L}_{\boldsymbol{s}_n}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s}_n), \tag{P.2}$$

with $\hat{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}_{\text{ini}}$, $\hat{\boldsymbol{\kappa}}^{(0)} = \boldsymbol{\kappa}_{\text{ini}}$ and $\hat{\boldsymbol{T}_1}^{(0)} = \boldsymbol{T_1}_{\text{ini}}$ the initial values of the parameters $\boldsymbol{\theta}$, $\boldsymbol{\kappa}$ and $\boldsymbol{T}_1$, respectively. This procedure is terminated when the number of iterations exceeds $t_{\max}$ or the relative decrease $\mathcal{E}^{(t)}$ of $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T}_1|\boldsymbol{s})$ between two consecutive iterations is below a fixed tolerance, $\mathcal{E}_{\min}$.

A detailed description of the problems (P.1) and (P.2) is provided in subsections III-A and III-B, respectively. Furthermore, a pseudo-code of the joint MLE algorithm is shown in subsection III-C, whereas its implementation is described in subsection III-D.

A. Problem 1 (P.1): estimation of the motion parameters

The motion estimation problem adopts a particularly simple structure when the relaxation parameters are fixed. If no dependence of $\{\boldsymbol{\theta}_n\}_{n=1}^N$ through index n is assumed, as is done here, the minimization can be decoupled into N optimization problems, which can be implemented very efficiently by parallel operations. The parameters $\{\boldsymbol{\theta}_n\}_{n=1}^N$ enter the linear motion operator in a non-trivial way, which renders the analytical calculation of the derivatives infeasible. Fortunately, the low dimensionality of the N minimization problems, involving only six variables each, allows us to use a derivative-free optimization method. In our approach, simulated annealing (SA) minimization is performed [37], which is known for its ability to avoid being trapped in local minima and its robustness to functions with complex structure [37]–[40]. Each time, motion estimates from the previous iteration t are used as initial guesses.

B. Problem 2 (P.2): estimation of the relaxation parameters

In contrast to P.1, the relaxation parameter estimation problem is a very-large-scale minimization problem. Derivative-free algorithms such as SA are therefore impractical to use. Quasi-Newton algorithms, on the other hand, may produce critical memory storage problems, due to the large dimensionality of the Hessian matrix approximation. Furthermore, line searches dramatically slow down the algorithm [41].

The optimization method of choice for solving this minimization problem is the use of an MM framework [42], which yields a voxel-wise independent algorithm, allowing a computationally efficient implementation. Here, we sketch the basics of the MM algorithm applied to the estimation problem at hand. The reader is referred to [42] for further details.

Let $J(\boldsymbol{\kappa}, \boldsymbol{T_1}) = \sum_{n=1}^{N} \mathcal{L}_{\boldsymbol{s}_n}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s}_n)$ be the cost function of P.2 that we seek to minimize w.r.t. $\boldsymbol{\kappa}$ and $\boldsymbol{T_1}$. MM algorithms are defined through the following recursive minimization problem:

$$\boldsymbol{\kappa}^{k+1}, \boldsymbol{T_1}^{k+1} = \arg\min_{\boldsymbol{\kappa}, \boldsymbol{T_1}} G(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k), \tag{10}$$

where $G(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ is a new user-designed cost function. It can be demonstrated that the sequence of iterates $\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k$ obtained from Eq. (10) converges to a local minimum of $J(\boldsymbol{\kappa}, \boldsymbol{T_1})$ if $G(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ is what is called in the optimization literature a *surrogate function* of $J(\boldsymbol{\kappa}, \boldsymbol{T_1})$. The properties that characterize a surrogate function are 1) $J(\boldsymbol{\kappa}, \boldsymbol{T_1}) \leq G(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) \quad \forall \boldsymbol{\kappa}, \boldsymbol{T_1}$ and 2) $J(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = G(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$.

Obviously, to really benefit from the MM algorithm, the surrogate function $G(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ should be easier to minimize than the original cost function $J(\boldsymbol{\kappa}, \boldsymbol{T_1})$. A key result is presented by Varadarajan and Haldar [23], showing that the following function is a valid surrogate function of $-\log p_{[\boldsymbol{s}_n]_m}([\boldsymbol{s}_n]_m|[\tilde{\boldsymbol{f}}_n]_m, [\boldsymbol{\sigma}_n]_m)$:

$$\frac{1}{2[\boldsymbol{\sigma}_n]_m^2}\left([\tilde{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1})]_m - [\breve{\boldsymbol{s}}_n^k]_m\right)^2 + C_n^m(k) \tag{11}$$

with $C_n^m(k)$ a constant independent of $\boldsymbol{\kappa}$ and $\boldsymbol{T_1}$,

$$\tilde{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1}) = \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}} \boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}), \tag{12}$$

and

$$[\breve{\boldsymbol{s}}_n^k]_m = [\boldsymbol{s}_n]_m \frac{I_1\left(\frac{[\boldsymbol{s}_n]_m [\tilde{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)]_m}{[\boldsymbol{\sigma}_n]_m^2}\right)}{I_0\left(\frac{[\boldsymbol{s}_n]_m [\tilde{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)]_m}{[\boldsymbol{\sigma}_n]_m^2}\right)}. \tag{13}$$

with $I_1(\cdot)$ the first order modified Bessel function of the first kind.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

Note that $\tilde{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1})$ describes the motion-corrupted synthetic T_1-weighted image, whereas $\breve{\boldsymbol{s}}_n^k$, from now on called the Bessel image, is the actual acquired image $\boldsymbol{s}_n$ corrected with a Bessel correction factor. A surrogate function for $\mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s}_n)$ is now obtained by summing Eq. (11) over m, i.e.,

$$G_n(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = ||\boldsymbol{W}_n^{1/2}\big(\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}) - \breve{\boldsymbol{s}}_n^k\big)||_2^2 + C_n(k) \tag{14}$$

with $\boldsymbol{W}_n = \text{diag}\{\frac{1}{2\boldsymbol{\sigma}_n^2}\}$. By summing $G_n(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ over n, we would obtain a global surrogate function for $J(\boldsymbol{\kappa}, \boldsymbol{T_1})$. At this point, the main benefit of applying the MM framework is that the relaxation estimation problem has now been transformed into a collection of weighted non-linear least squares (NLLS) problems, avoiding complicated minimization with Bessel functions. However, we still can further simplify the problem and convert it into a fully separable (voxel-wise independent) NLLS problem. To that end, we apply another surrogate function $G_n^*(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ to each $G_n(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$. It is easy to demonstrate that if $G_n^*(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ is a surrogate function for $G_n(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$, it is also a valid surrogate function for $\mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s}_n)$. Therefore, we finally define $G(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ as

$$G(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = \sum_{n=1}^{N} G_n^*(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k). \tag{15}$$

The choice for $G_n^*(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ is a separable quadratic surrogate (SQS) function [43], which, when applied to our problem at hand, takes the expression,

$$G_n^*(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = ||\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}) - \boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)||_2^2 + C_n^*(k), \tag{16}$$

with

$$\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = \boldsymbol{f}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) + \sigma^* \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H \boldsymbol{W}_n\big(\breve{\boldsymbol{s}}_n^k - \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}\boldsymbol{f}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)\big) \tag{17}$$

and $\sigma^* = 2\min_{n,m}[\boldsymbol{\sigma}_n]_m^2$. The complete derivation of $G_n^*(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ can be found in Section II of the additional document which is included as part of the downloadable supplementary material. With $G_n^*(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$, minimization of Eq. (15) is nothing more than fitting the relaxation model $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$ to the "residual" images $\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ with $n = 1, ..., N$ in a least squares sense. Therefore, it is a completely separable optimization problem and hence it can be implemented in parallel for every voxel m. This is the main distinct characteristic of the joint MLE that we present in this work, which makes it an efficient method to be used in practice. Once the model-fitting is performed, the new iterate serves to update the "residual" images. This process is repeated until $k > k_{\text{max}}$ or $\mathcal{EJ}^{(k)} < \mathcal{EJ}_{\text{min}}$ where $\mathcal{EJ}^{(k)}$ is the analogous of $\mathcal{E}^{(t)}$ for $J(\boldsymbol{\kappa}, \boldsymbol{T_1})$. The final iterate yields the new $\hat{\boldsymbol{\kappa}}^{(t+1)}$ and $\hat{\boldsymbol{T}}_{\boldsymbol{1}}^{(t+1)}$, which are then used as input in the motion estimation problem P.1.

C. Initialisation

Although convergence to a local minimum is guaranteed, convergence to the global minimum, which corresponds with the MLE estimate, cannot be proved, since $\mathcal{L}_{\boldsymbol{s}}(\boldsymbol{\theta}, \boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{s})$ is non-convex. To increase the chances of finding the global minimum, providing good initial values is crucial. In our approach, initial values were obtained using the conventional approach (CA), which consists of image registration prior to voxel-wise relaxation model fitting. A sequential estimation to initialize unified motion model-based approaches was also used in [44], with very good results, and we found it a robust method to initialize our joint MLE. In particular, firstly, the initial motion parameters $\boldsymbol{\theta}_{\text{ini}}$ were obtained by registering the set of T_1-weighted images based on maximization of MI between the images [45], [46]. All images were pairwise registered to the reference system $\boldsymbol{r}$ with a pyramidal multi resolution scheme of three levels. The number of iterations of the internal optimization algorithm was set to a very high value (> 900) to ensure convergence of the motion parameter estimation. Next, the relaxation parameters $\boldsymbol{\kappa}_{\text{ini}}$ and $\boldsymbol{T_1}_{\text{ini}}$ were voxel-wise estimated from the registered images using the MLE based on Rician distributed data [20]. To compute the MLE, a Broyden-Fletcher-Goldfarb-Shanno (BFGS) quasi-newton algorithm [41] was used with exact analytical derivatives. The spatially variant standard deviation, which is required for the MLE, was estimated with the method of [33]. By using the CA as initialization, we have invariably found that the estimated T_1 maps, $\boldsymbol{T_1}_{\text{ML}}$, are superior in terms of accuracy and rMSE compared to those obtained with the CA.

Furthermore, simulation results have shown that the joint MLE is stable and robust to occasional inaccuracies in the CA-based initial motion parameters.

Pseudo-code of the joint MLE algorithm is presented in Fig. 1 and an illustrative flow-chart is shown in Fig. 2. In practice, the joint MLE requires as input, apart from the initial motion and relaxation parameters, an estimate of the spatially variant standard deviation $\boldsymbol{\sigma}$, $\hat{\boldsymbol{\sigma}}$. Such an estimate is obtained with the method of [33] directly applied on the acquired images $\{\boldsymbol{s}_n\}_{n=1}^N$.

D. Parameters selection, code implementation and computational cost

The proposed joint MLE was implemented in MATLAB and run on a computer with an Intel i7-4770K processor consisting of four cores at 3.5 GHz. The machine had 32 GB of RAM. The SA algorithm of P.1 was implemented using the MATLAB routine `simulannealbnd` with the default parameters. The NLLS fitting of P.2 was performed by the MATLAB routine `lsqnonlin` using the Levenberg-Marquardt (LM) [41] method, also with the default parameters. The tolerance criteria and the maximum number of iterations to halt the algorithm were chosen to be $\mathcal{EJ}_{\text{min}} = 10^{-2}$ and $\mathcal{E}_{\text{min}} = 10^{-3}$, and $t_{\text{max}} = k_{\text{max}} = 10$, respectively. To exploit the highly parallelizable structure of the joint MLE, MATLAB parallel computing tools were used to estimate $\boldsymbol{\theta}_n$ for each value of n separately. Similarly, the voxel-wise NLLS relaxation model fitting was performed in a parallel

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

initialize $t = 0$, $\boldsymbol{\sigma} = \hat{\boldsymbol{\sigma}}$, $\hat{\boldsymbol{\theta}}^{(0)} = \boldsymbol{\theta}_{\text{ini}}$, $\hat{\boldsymbol{T}_1}^{(0)} = \boldsymbol{T_1}_{\text{ini}}$ and $\hat{\boldsymbol{\kappa}}^{(0)} = \boldsymbol{\kappa}_{\text{ini}}$
while $\mathcal{E}^{(t)} \geq \mathcal{E}_{\min}$ and $t < t_{\max}$ **do**
  Solve (P.1) to get $\hat{\boldsymbol{\theta}}^{(t+1)}$
  Set $k = 0$ (P.2 begins)
  $\boldsymbol{\kappa}^k = \boldsymbol{\kappa}^{(t)}$ and $\boldsymbol{T_1}^k = \boldsymbol{T_1}^{(t)}$
  while $\mathcal{EJ}^{(k)} \geq \mathcal{EJ}_{\min}$ and $k < k_{\max}$ **do**
    Calculate $\tilde{\boldsymbol{f}}_n^{(t+1)}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ with Eq. (12)
    Calculate $\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ with Eq. (17)
    Voxel-wise NLLS fitting of $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$ to $\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ so as to get $\boldsymbol{\kappa}^{k+1}$ and $\boldsymbol{T_1}^{k+1}$
    Calculate $\mathcal{EJ}^{(k)}$ and set $k = k + 1$
  end while
  Set $\hat{\boldsymbol{\kappa}}^{(t+1)} = \boldsymbol{\kappa}^k$ and $\hat{\boldsymbol{T_1}}^{(t+1)} = \boldsymbol{T_1}^k$ (P.2 ends)
  Calculate $\mathcal{E}^{(t)}$ and set $t = t + 1$
end while
$\boldsymbol{\theta}_{\text{ML}} = \hat{\boldsymbol{\theta}}^{(t)}$ and $\boldsymbol{T_1}_{\text{ML}} = \hat{\boldsymbol{T_1}}^{(t)}$

Fig. 1: Pseudo-code of the joint MLE algorithm.

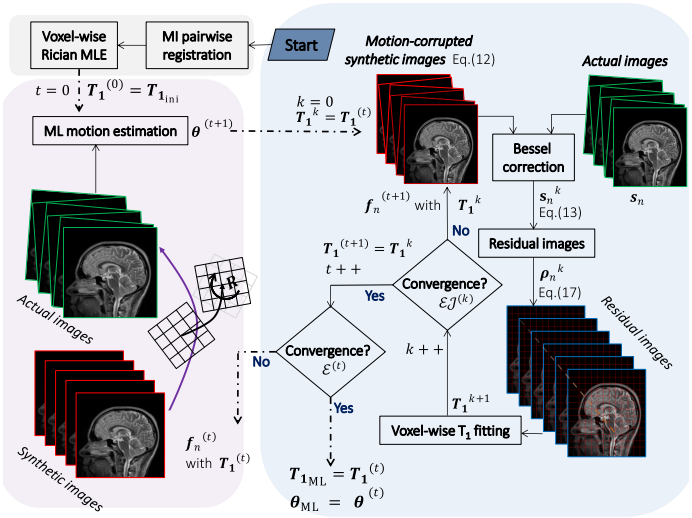

Fig. 2: Flow-chart of the joint MLE algorithm.

manner by dividing the spatial grid into 8 non-overlapping 3D blocks. Finally, a mask was used to avoid calculation of the relaxation parameters in background areas, hence speeding up the implementation.

The computational time per iteration of the joint MLE algorithm is dominated by the voxel-wise T_1 fitting, which depends linearly on the number of voxels M, depending in turn on the Field-of-View (FOV) and the voxel size. With relatively little effort to optimize our code, and using the MATLAB parallel tools mentioned earlier, the voxel-wise T_1 fitting took approximately 8 min to process a series of $N = 8$ T_1-weighted images with $M \approx 10^5$ voxels. Overall, with the tolerance criterion described above, the average number of total iterations (external plus internal) were roughly 15, providing precise and accurate $\boldsymbol{\theta}_{\text{ML}}$ and $\boldsymbol{T_1}_{\text{ML}}$ in an average time of 2.2 hours. Note that migration of the MATLAB code to C++ would produce a much faster implementation, especially if multi-threading is used for the highly parallelizable relaxation estimation problem [44].

E. Gaussian approximation for GRAPPA+SoS

When GRAPPA reconstructed data is combined with SoS, the statistical distribution of the composite magnitude image can be well approximated with a non-stationary nc-χ distribution, where both the variance and the (effective) degrees-of-freedom parameter, L_{eff}, are spatially non-stationary (i.e., vary from voxel to voxel) [31]. Since the MM framework was originally developed for the nc-χ distribution [23], the application of the proposed joint MLE is straightforward provided an estimate of L_{eff} for every voxel is available. Unfortunately, practical estimators of spatial maps of L_{eff} are, to the authors' knowledge, not yet available in the literature. Nevertheless, for high SNR, a Gaussian distribution with spatially variant σ has been proved to be an accurate model in replacement of the nc-χ model [33]. In this case, the joint MLE is even simpler than it was for the Rician case. Indeed, it can easily be shown that

$$\mathcal{L}_{\boldsymbol{s}_n}(\boldsymbol{\theta}_n, \hat{\boldsymbol{\kappa}}^{(t)}, \hat{\boldsymbol{T_1}}^{(t)} | \boldsymbol{s}_n) = ||\boldsymbol{W}_n^{1/2}\big(\boldsymbol{H}_{\boldsymbol{\theta}_n}\boldsymbol{f}_n(\hat{\boldsymbol{\kappa}}^{(t)}, \hat{\boldsymbol{T_1}}^{(t)}) - \boldsymbol{s}_n\big)||_2^2 \quad (18)$$

and

$$\mathcal{L}_{\boldsymbol{s}_n}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1} | \boldsymbol{s}_n) = ||\boldsymbol{W}_n^{1/2}\big(\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}) - \boldsymbol{s}_n\big)||_2^2. \quad (19)$$

The same SA optimization algorithm as before can be used to minimize Eq. (18) for solving the motion estimation problem (P.1). To simplify the minimization of Eq. (19) for solving the relaxation parameter estimation problem (P.2), we can apply directly the SQS function on $\mathcal{L}_{\boldsymbol{s}_n}(\hat{\boldsymbol{\theta}}_n^{(t+1)}, \boldsymbol{\kappa}, \boldsymbol{T_1} | \boldsymbol{s}_n)$, avoiding the Bessel correction step. Indeed, the relaxation parameter estimation problem is again a NLLS fitting of $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$ to a different $\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$ where just $\breve{\boldsymbol{s}}_n^k$ in Eq. (17) has to be replaced by $\boldsymbol{s}_n$, the actual acquired images.

IV. Experiments

The proposed joint motion and T_1 MLE was validated using both simulated and real data. Moreover, its performance was compared to that of the CA with MI-based registration [46], and a recently proposed model-based approach of Hallack et al. [18].

A. Simulated T_1-weighted data

A set of 3D IR-SE T_1-weighted images $\{\boldsymbol{s}_n\}_{n=1}^N$ affected by inter-image motion (as in Eq. (4)) and noise was simulated from ground truth T_1 and proton density maps. The ground truth T_1 map was created from the BrainWeb anatomical model, using reported T_1 values in human brain tissue at 3T [47], [48]. For the three main brain tissues, white matter, grey matter and cerebrospinal fluid (CSF), the reference values were 838 ms, 1607 ms, and 4300 ms, respectively. The ground truth proton density map was created in a similar fashion. The size of both 3D maps was $111 \times 93 \times 71$ with an isotropic voxel size of 1.5 mm. From these maps, a set of IR-SE T_1-weighted images was simulated based on [25] with TR/TE = 10000 /14 ms, and $N = 8$ logarithmically equidistant inversion times

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

$\{TI_n\}_{n=1}^N$ between 20 ms and 8000 ms. The three consecutive RF pulse angles were set to 180°, 90° and 180°. In the next step, we randomly generated ground truth motion parameters $\{\boldsymbol{\theta}_n\}_{n=1}^N$. Each of the six rigid motion parameters followed an independent Gaussian Random Walk (RW) [44] along the temporal dimension n.

More precisely, the motion parameters were generated as

$$\boldsymbol{\theta}_n = \boldsymbol{c} + \boldsymbol{\theta}_{n-1} + \boldsymbol{w}_n, \tag{20}$$

where $\boldsymbol{c} \in \mathbb{R}^{6\times 1}$ denotes the motion drift and $\boldsymbol{w}_n \in \mathbb{R}^{6\times 1}$ a vector valued, zero mean, Gaussian random variable with covariance matrix $\boldsymbol{\Sigma} = \sigma_{\text{RW}}^2 \boldsymbol{I}$, with σ_{RW} the standard deviation of each of the elements of $\boldsymbol{w}_n$ and $\boldsymbol{I}$ the 6×6 identity matrix. The reference system $\boldsymbol{r}$ was chosen to be $\boldsymbol{r}_1$, hence $\boldsymbol{\theta}_1 = \boldsymbol{0}$. Finally, to account for noise, Rician distributed images $\{\boldsymbol{s}_n\}_{n=1}^N$ were simulated [49] with spatially variant noise maps. Synthetic spatially variant noise maps were generated based on a realistic noise pattern that was presented in [33]. This pattern was derived from a real parallel MRI acquisition [50].

The proposed joint motion and T_1 MLE was compared to that of the CA with MI-based registration [46] (the initialization technique for the joint MLE) and a recently proposed inter-image model-based approach of Hallack et al. [18]. The MI-based rigid registration step of the CA was implemented using the first image of the series as a reference. Details of the implementation were already given in subsection III-C. The remaining MI registration parameters were set to those provided in the MATLAB built-in code. Hallack's method was implemented by following the guidelines provided in [18]. Just like the joint MLE, it was initialized with the CA. The parameters $\boldsymbol{\kappa}$ and $\boldsymbol{T_1}$ were estimated with the LM algorithm. Hallack's algorithm was stopped when either the decrease of the cost function between iterations was below $\mathcal{E}_{\text{min}}$, or the number of iterations exceeded t_{max}.

Two types of simulation experiments were performed:

1) Exp.1: Performance as a function of SNR: In a first set of experiments (Exp.1), the performance of the joint MLE as a function of the SNR of the T_1-weighted image data set was tested. To that end, motion parameters $\{\boldsymbol{\theta}_n\}_{n=1}^N$ were generated with $\sigma_{\text{RW}} = 0.4$ mm/degree and no drift. After fixing the motion parameters, T_1-weighted image data sets with SNR values between 20 and 100 were simulated, where the SNR is defined as the spatial mean of the ratio of the reference noise-free T_1-weighted image and the standard deviation noise map of this reference image. For each SNR, $N_{\text{MC}} = 15$ MC simulations were generated.

2) Exp.2: Performance as a function of the type of motion: In a second set of experiments (Exp.2), the performance of the joint MLE was evaluated for different types of motion (and fixed SNR $= 40$). For completeness, we also included a case without motion.

a) *Low Abrupt motion (LA-m).* The motion parameters were generated as in Exp. 1, i.e., without drift/tendency and with $\sigma_{\text{RW}} = 0.4$ mm/degree.

b) *High Abrupt motion (HA-m).* The motion parameters were generated without drift/tendency and with $\sigma_{\text{RW}} = 1.5$ mm/degree.

c) *Rotational motion (R-m).* The motion parameters were generated with drift parameter vector $\boldsymbol{c} = (0, 0, 0, 0.5, 0.5, 0.5)^T$ and $\sigma_{\text{RW}} = 0.4$ mm/degree. Note that it follows from Eq. (20) and Eq. (3) that in this scenario only the rotation parameters are affected by drift.

d) *Translational motion (T-m).* The motion parameters were generated with $\boldsymbol{c} = (0.5, 0.5, 0.5, 0, 0, 0)^T$ and $\sigma_{\text{RW}} = 0.4$ mm/degree.

e) *No motion (No-m).* No motion was applied, i.e., $\tilde{\boldsymbol{f}}_n(\boldsymbol{\theta}_n, \boldsymbol{\kappa}, \boldsymbol{T_1}) = \boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$.

For all types of motion, extreme values as well as mean values for each of the six rigid parameters along the temporal dimension n are shown in the second column of Table I. For each type of motion, $N_{\text{MC}} = 15$ simulations, i.e., noisy T_1-weighted image data sets, were generated (with fixed motion parameters and SNR$= 40$).

To assess the ability of each method to estimate the T_1 map, the following performance measures were used:

(a) *Relative bias*. The bias quantifies the *accuracy* of the estimator. For each voxel, the relative sample bias was calculated as $(\bar{\hat{T}}_1 - T_1)/T_1$, where $\bar{\hat{T}}_1$ is the sample mean of the N_{MC} estimates $\hat{T}_1$ and T_1 is the true value. A measure of the overall accuracy of the T_1 map was obtained by calculating the spatial mean of the absolute value of the relative sample bias, using a brain mask to avoid the skull.

b) *Relative standard deviation*. The standard deviation quantifies the *precision* of the estimator. For each voxel the relative sample standard deviation was calculated as $\text{std}(\hat{T}_1)/T_1$, and an overall precision measure was obtained by taking the spatial mean of these relative sample standard deviations, using the same brain mask.

c) *Relative root-mean-square error (relative RMSE).* The RMSE is a measure that incorporates both accuracy and precision. For each voxel, the relative sample RMSE was calculated as $\sqrt{\overline{(\hat{T}_1 - T_1)^2}}/T_1$. An overall RMSE measure was obtained by calculating the spatial mean of these relative sample RMSE values, again within the same brain mask.

To assess the ability of each method to estimate motion, the following performance measures were used:

d) *Relative motion error*, defined as

$$||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}||_2 / ||\boldsymbol{\theta}||_2. \tag{21}$$

e) *Motion component relative bias*, defined as

$$\frac{1}{N}\sum_{n=1}^{N} |(\overline{[\hat{\boldsymbol{\theta}}_n]_j} - [\boldsymbol{\theta}_n]_j)/[\boldsymbol{\theta}_n]_j|, \tag{22}$$

with $[\boldsymbol{\theta}_n]_j$ the jth component of $\boldsymbol{\theta}_n$ and $\overline{[\hat{\boldsymbol{\theta}}_n]_j}$ the sample mean of the N_{MC} estimates $[\hat{\boldsymbol{\theta}}_n]_j$.

B. Ground-truth based real experiment

In order to assess the performance of the joint MLE with an actual T_1-weighted data set corrupted by motion, we performed a controlled experiment. The experiment comprised

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

the acquisition of two data sets. Firstly, we acquired an IR T_1-weighted data set of a (static) watermelon. In a second step, we deliberately introduced motion between the acquisition of each of the acquired 2D multi-slice T_1-weighted images. In particular, we manually translated and rotated the watermelon after the complete acquisition for a fixed TI. From this data, estimated T_1 maps were obtained with the CA, Hallack's method and the joint MLE. We then quantitatively compared these T_1 maps to the estimated T_1 map of the first dataset, which was unaffected by motion and hence can be considered as a reasonable ground-truth.

Both IR T_1-weighted data sets were acquired with a 3T MRI scanner (MAGNETOM Prisma$^{\text{fit}}$, Siemens) using a 32-channel head coil. The IR T_1-weighted data sets comprised $N = 8$ T_1-weighted multi-slice images whose inversion times were logarithmically spaced between 300 and 6000 ms. For each inversion time, we acquired a 2D multi-slice image with a 2D interleaved multi-slice IR Turbo Spin Echo (TSE) sequence [51]. The Echo Train Length (ETL) was 10 and TR/TE $= 7920/8.8$ms. Each multi-slice image was acquired within approximately 3 min. The total scan time was about 24 minutes. The acquisition plane was axial and the acquisition matrix was $256 \times 256 \times 40$ with an anisotropic voxel size of $1 \times 1 \times 4$ mm^3 and no slice gap. Magnitude data were reconstructed using the SENSE method (acceleration factor of 2) [52]. The coordinate system of the image with the highest inversion time was chosen to be the reference system $\boldsymbol{r}$. An estimated SNR of the reference image, as defined in subsection IV-A1, was found to be 26. The estimation of the SNR was performed with the Expectation-Maximization-based method of De Vore [53] adapted for local estimation [33] to work with one single image (3×3 neighborhoods were used). The ground-truth T_1 map was estimated with a voxel-wise MLE based on Rician data [20], where the noise standard deviation map was estimated in a preprocessing step using the method described in [33].

C. In vivo T_1-weighted data

We validated the joint MLE with two in vivo human brain data sets suffering from involuntary patient motion.

1) In vivo axial human brain data: An IR T_1-weighted data set of a healthy 26-year old male volunteer was acquired with a 3T MRI scanner (MAGNETOM Skyra, Siemens) using a 20-channel head coil. For each inversion time, we acquired a 2D multi-slice image with an interleaved 2D multi-slice IR TSE sequence [51], [54]. The sequence parameters were: ETL $= 4$ and TR/TE = 8040/18 ms. Each multi-slice image was acquired within 2.5 min approximately. The IR T_1-weighted data set comprised $N = 7$ T_1-weighted multi-slice images whose inversion times were logarithmically spaced between 50 and 3200 ms. The total scan time was about 19 minutes. The acquisition plane was axial and the acquisition matrix was $128 \times 128 \times 25$ with an anisotropic voxel size of $1.9 \times 1.9 \times 6$ mm^3 and a slice gap of 10%. The SENSE method was employed to reconstruct the magnitude data with an acceleration factor of 3. Noise maps were obtained with the method of [33]. We estimated an SNR of 24.3 with the method of [53]. In this case, the reference image was the one with lowest inversion time.

2) In vivo sagittal human brain data: An IR T_1-weighted data set of a healthy 28-year old male volunteer was acquired with a 3T MRI scanner (MAGNETOM Prisma$^{\text{fit}}$, Siemens) using a 32-channel head coil. As in the previous in vivo experiment, we acquired for each inversion time a 2D multi-slice image with an interleaved 2D multi-slice IR TSE sequence [51]. The sequence parameters were: ETL = 10 and TR/TE = 5000/4.8 ms. Each multi-slice image was acquired within 2 min approximately. The IR T_1-weighted data set comprised $N = 14$ T_1-weighted images whose inversion times were logarithmically spaced between 100 and 3000 ms, giving a total acquisition time of 28 min. The acquisition plane was sagittal and the acquisition matrix was $256 \times 256 \times 40$ with an anisotropic voxel size of $1 \times 1 \times 4$ mm^3 and no slice gap. Magnitude data were reconstructed with the GRAPPA method with SoS reconstruction (acceleration factor of 3) [55]. The image with the lowest inversion time was chosen as a reference. We estimated an SNR of 55 with a locally adapted ML estimator (3×3 neighborhoods) assuming a nc-χ distribution. Due to the high SNR, we relied on results of [33] and used the version of the joint MLE algorithm adapted for spatially variant Gaussian noise (subsection III-E). Noise maps were obtained with the method of [33]. The CA was implemented with a Gaussian MLE where the noise standard deviation was estimated with the method described in [33].

V. Results

A. Simulated T_1-weighted data

1) Exp.1: Performance as a function of SNR:

Overall relative T_1 bias, standard deviation and RMSE results are shown in Fig.3(a-c). For the whole range of SNR, the joint MLE allows a much more accurate estimation of the T_1 map than Hallack's method and especially than the CA.

In terms of precision, CA obtains the best result, followed by the joint MLE and Hallack's method. However, in terms of the overall RMSE, the joint MLE performs best for all values of the SNR. Furthermore, the box-plot shown in Fig. 3(d) demonstrates the superiority of the proposed joint MLE in terms of motion estimation.

To complement the results, maps of the absolute value of the relative sample bias for the three methods are shown in Fig. 4(b-d), along with the simulated ground truth in Fig. 4(a). A close look at the bias maps corroborates the poor performance of the CA compared to the joint MLE. It is also clearly seen that the bias map of Hallack's method presents much higher values than that of the joint MLE, especially in white/grey matter surroundings.

2) Exp.2: Performance as a function of the type of motion: Bar charts representing the overall T_1 accuracy, precision and rMSE for the four cases of motion and the no-motion scenario are shown in Fig. 5.

In light of these results, it can be concluded that the joint MLE yields the most accurate T_1 maps in all the four considered motion scenarios, followed by Hallack's method. Furthermore, the performance of all methods seems to be

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

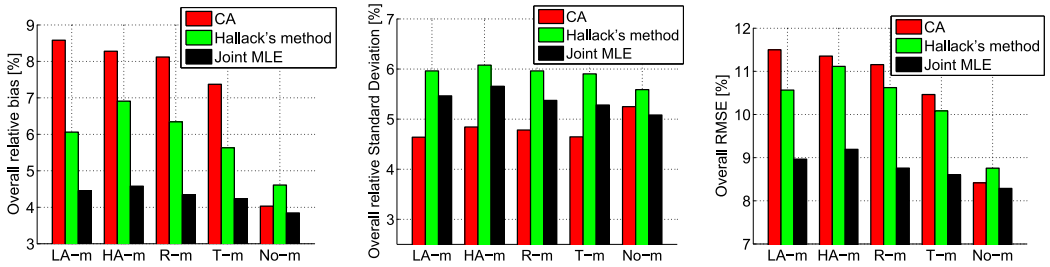

Fig. 5: Results of Exp.2: (a) relative T_1 bias, (b) relative T_1 standard deviation and (c) relative T_1 RMSE.

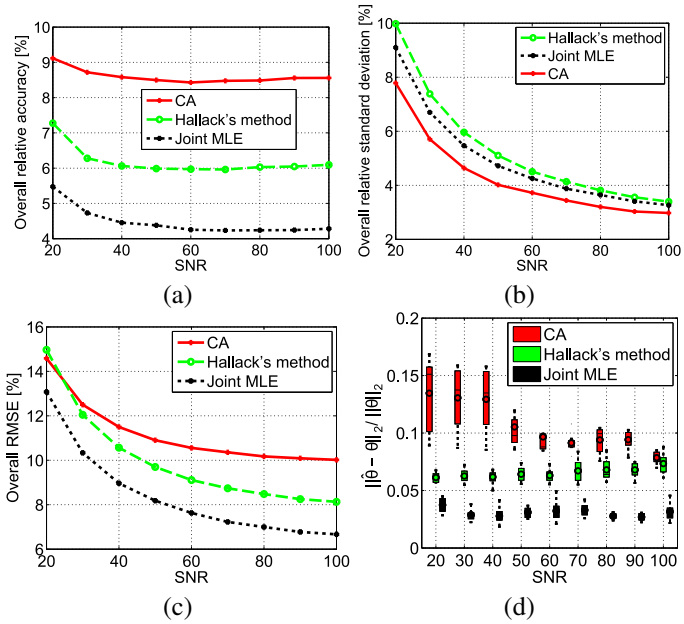

Fig. 3: Results of Exp.1: (a) relative T_1 bias, (b) relative T_1 standard deviation, (c) relative T_1 RMSE and (d) box plots of the relative motion error, as a function of the SNR.

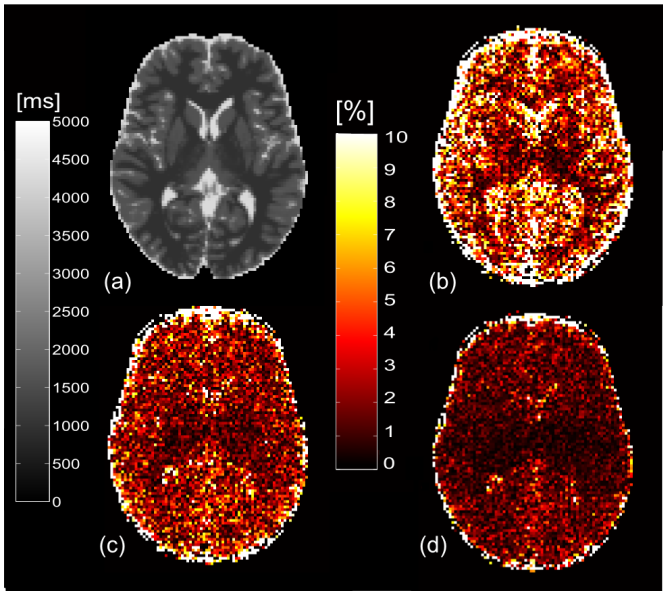

Fig. 4: Results of Exp.1: (a) Mid-axial slice of the 3D ground truth T_1 map, and maps of the absolute value of the relative sample bias [in %] for (b) CA, (c) Hallack's method and (d) joint MLE, for SNR = 40.

fairly insensitive to the type of motion considered. Even though the highest precision was consistently obtained with the CA, its overall relative RMSE is much higher compared to Hallack's method and especially to the joint MLE, which again produces the best T_1 maps in RMSE sense. The case of no-motion is particularly interesting. In such a scenario, the CA performance drastically improves in terms of accuracy and RMSE, though its precision decreases. In the absence of motion, the error propagation of the CA approach is negligible, and hence, the overall relative accuracy of the CA approaches the one obtained with the joint MLE, which is yet the highest. The decrease in precision of the CA can be understood as follows. In the absence of motion, the interpolation (and hence smoothing) effects that are inherent to the registration step of the conventional two-step approach become marginal, hence not contributing to a reduction in the variability of the estimates.

The motion component relative bias for each of the six components are reported in Table I. For the no-motion case, the motion component relative bias is not well-defined (division by zero). Instead, we report the motion component absolute bias.

The best results are highlighted in shaded gray. In 27 of 30 cases, the joint MLE achieved the highest accuracy in both the translation and rotation parameters. Sometimes, this improvement is even more than 5-fold compared to the CA. In general, Hallack's method provides more accurate T_1 estimates than CA, which is in agreement with previously reported results [10], [18]. Nonetheless, further substantial improvement can be obtained if the joint MLE is used. To illustrate the quality of the motion estimation, we have shown graphs, as a function of n, of the ground-truth and estimated motion parameters for one of the rotational motion (R-m) simulations in the additional document (Fig.1) of the downloadable supplementary material.

It is also important to notice that the occasionally poor CA-based motion initialization does not prevent the joint MLE from producing the most accurate motion estimates. This highlights another feature of the joint MLE: it is fairly robust to scenarios where the CA-based motion initialization is relatively poor.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

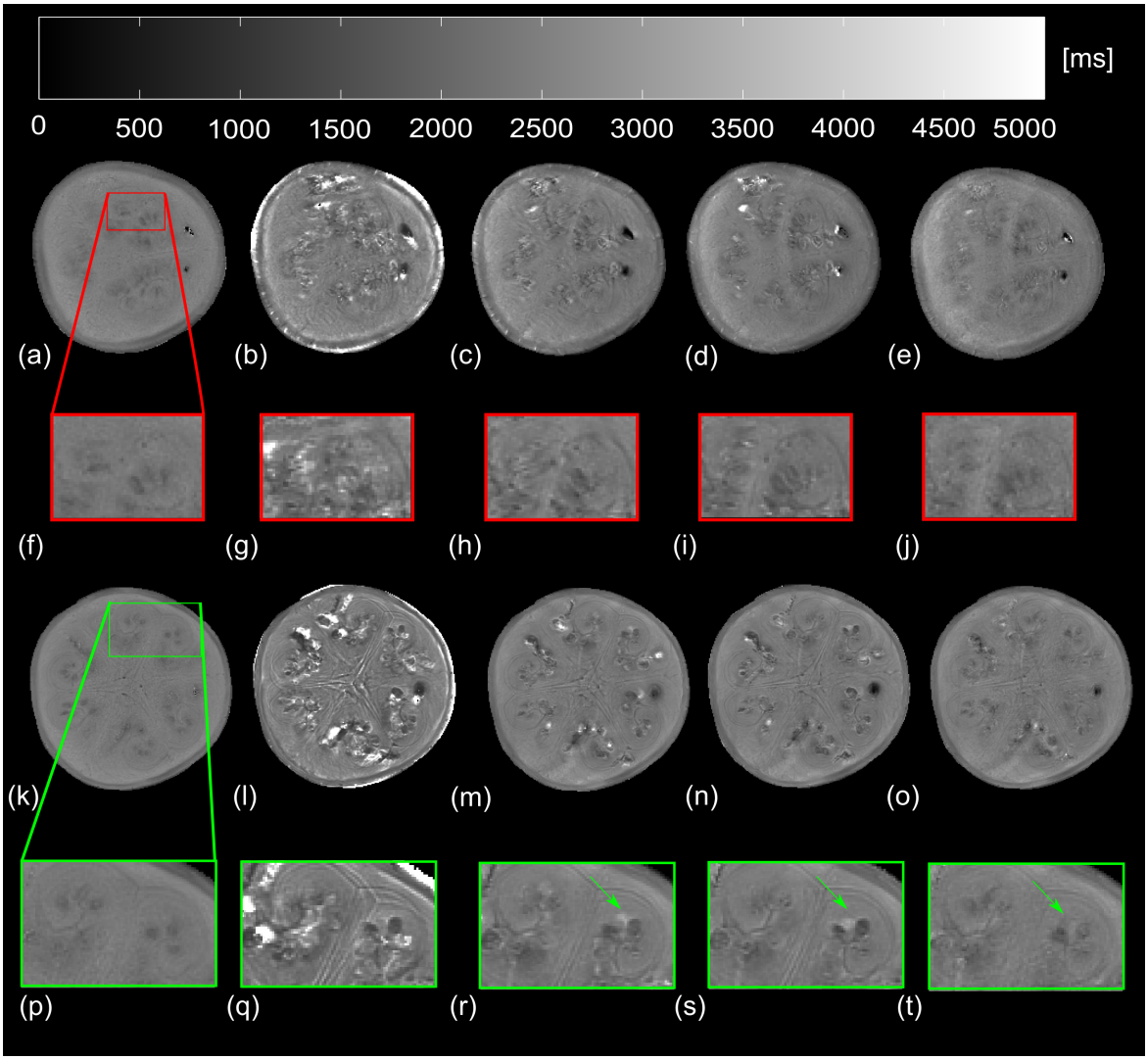

Fig. 6: Two axial slices of a watermelon T_1 map. Top-axial slice: (a) ground-truth (no motion), (b) without motion correction, (c) CA, (d) Hallack's method, (e) joint MLE. Magnified regions are shown in (f-j). Mid-axial slice: (k) ground-truth (no motion), (l) without motion correction, (m) CA, (n) Hallack's method, (o) joint MLE. Magnified regions are shown in (p-t).

B. Ground-truth based real experiment

A top-axial and a mid-axial slice of the estimated T_1 maps for the CA, Hallack's method and the joint MLE are shown in Fig. 6(c-e) and Fig. 6(m-o), whereas the ground truth T_1 map and the T_1 map estimated without motion correction are displayed in Fig. 6(a-b) for the top-axial slice and in Fig. 6(k-l) for the mid-axial slice. From this experiment, it can be observed that more detailed T_1 maps can be obtained with the joint MLE in comparison to Hallack's method and especially to the CA. Aside from the presence of a large number of outliers in the T_1 maps obtained with the CA and Hallack's method, which are drastically reduced with the joint MLE, fine structural details seem better preserved with our proposed method. This observation is confirmed by inspecting the magnified regions. The heterogeneity of the T_1 values in those regions, as noticed from the ground-truth T_1 map, is better maintained with the joint MLE. See for instance the delineation of low T_1 value structures in Fig. 6(f-j). Note as well that artifacts in the T_1 maps, as shown in Fig. 6(p-t) (green arrow), are considerably mitigated with the joint MLE. Quantitative validation of the estimated T_1 maps was based on spatial maps of the absolute value of the relative errors [%](Fig. 7).

In accordance to our previous discussion, the spatial distributions of the relative errors further indicate the good performance of the joint MLE in comparison to competing methods. It is manifestly clear that the error maps of Hallack's method and the CA present much higher values than that of the joint MLE. To complement the quantitative analysis, we calculated an overall relative error, within a mask neglecting the background, in a similar fashion as done in Exp.1. Numerical results are in agreement with the observation made from the spatial maps. We found that the joint MLE produced the T_1 map with the lowest overall relative error. Indeed, the overall relative error for the without motion correction case, the CA, Hallack's method and the joint MLE was 87%, 20.2%, 19.7% and 15.1%, respectively. To further complement the

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

Type	Motion (max / mean)	CA	Hallack	Joint MLE
LA-m	t_x (2.5 / 0.98 mm)	8.4 %	4 %	**1.1 %**
	t_y (1.4 / 0.88 mm)	1.4 %	2.1 %	**0.4 %**
	t_z (1 / 0.38 mm)	33.5 %	30.9 %	**5.3 %**
	α (0.2 / 0.003 degree)	63.5 %	20.6 %	**9.5 %**
	β (-0.2 / -0.06 degree)	90 %	43.8 %	**19.1 %**
	γ (0.57 / 0.26 degree)	75.4 %	57.5 %	**6.8 %**
HA-m	t_x (-9.4 / -4.4 mm)	2.3 %	3.9 %	**0.1 %**
	t_y (-2.3 / 0.08 mm)	30 %	12.9 %	**2.4 %**
	t_z (4.6 / 2 mm)	27 %	19.8 %	**13.5 %**
	α (-2.1 / -0.51 degree)	**4.9 %**	10.5 %	6.4 %
	β (3.7 / 1.9 degree)	1.4 %	2 %	**0.9 %**
	γ (1.3 / 0.5 degree)	**4.9 %**	18.2 %	19.6 %
R-m	t_x (-0.8 / -0.4 mm)	5.5 %	3.4 %	**1.2 %**
	t_y (1.1 / 0.8 mm)	2.3%	3.1 %	**1.1 %**
	t_z (0.3 / 0.02 mm)	466 %	166 %	**29.8 %**
	α (3.2 / 1.5 degree)	2.1 %	1.3 %	**0.7 %**
	β (3.1 / 1.5 degree)	1.6 %	**0.9 %**	1.1 %
	γ (1.9 / 1.1 degree)	2.9 %	5 %	**0.13 %**
T-m	t_x (2.9 / 1.4 mm)	1.3 %	1.9 %	**0.8 %**
	t_y (2.8 / 1.9 mm)	1.2 %	3 %	**1.1 %**
	t_z (2.1 / 1.2 mm)	2 %	2.6 %	**0.8 %**
	α (-0.3 / -0.2 degree)	26.7 %	64.7 %	**20.7 %**
	β (-0.8 / -0.3 degree)	121.15 %	17.2 %	**4.7 %**
	γ (-0.5 / -0.2 degree)	91.1 %	27.1 %	**21.7 %**
No-m	t_x (0 / 0 mm)	0.012	0.032	**0.009**
	t_y (0 / 0 mm)	0.112	0.077	**0.03**
	t_z (0 / 0 mm)	0.078	0.159	**0.086**
	α (0 / 0 degree)	0.0281	0.0133	**0.01**
	β (0 / 0 degree)	0.083	0.057	**0.019**
	γ (0 / 0 degree)	0.0201	0.0183	**0.011**

TABLE I: Results of Exp.2: for four types of motion (column 1), the maximum and mean values of the motion parameters (column 2), and the motion component relative bias for each of the six components for CA, Hallack's and the joint MLE method (columns 3-5) are shown. For the no-motion case, the motion component absolute bias is reported instead of the motion component relative bias since the latter metric is not well-defined for parameters that are equal to zero.

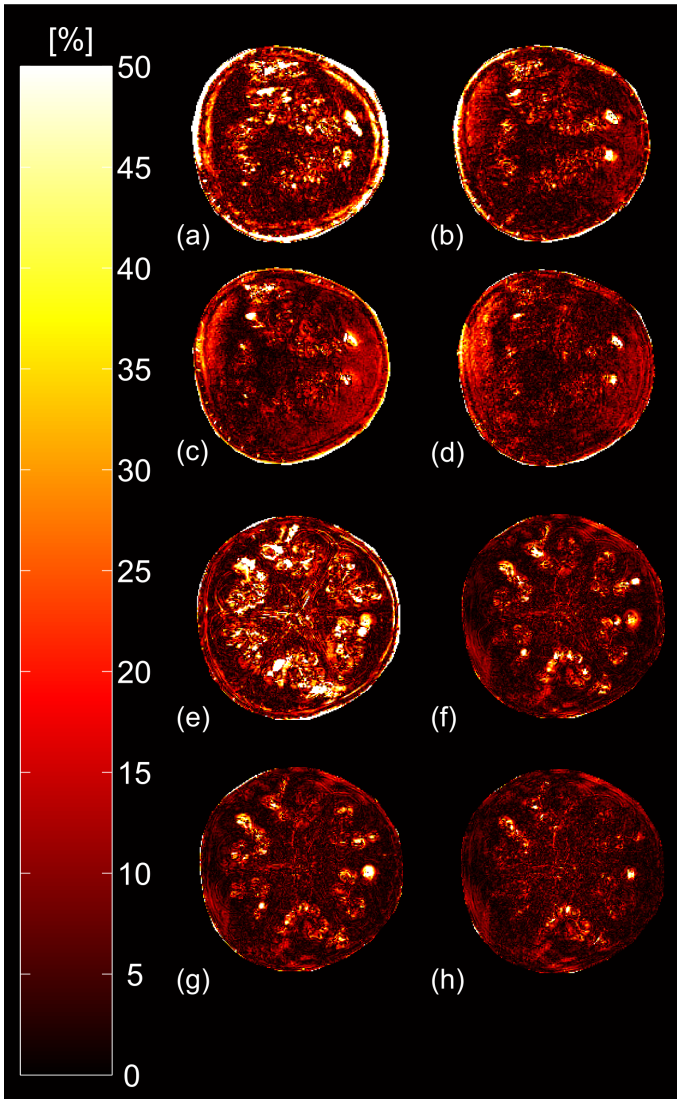

Fig. 7: Maps of the absolute value of relative error with respect to the ground-truth T_1 map. Top-axial slice: (a) without motion correction, (b) CA, (c) Hallack's method, (d) joint MLE. Mid-axial slice: (e) without motion correction, (f) CA, (g) Hallack's method, (h) joint MLE.

quantitative analysis, graphs of the motion parameter estimates for the three methods are shown in Fig. 2 of the additional document which is included as part of the downloadable supplementary material which accompanies this paper.

C. *In vivo T_1-weighted data*

1) In vivo axial human brain data: A mid-axial and a top-axial slice of the estimated T_1 map for the three methods are shown in Fig. 8(b-d) and Fig. 8(f-h), respectively. The estimated T_1 maps without motion correction are shown in Fig. 8(a) and Fig. 8(e). The presence of outliers in the T_1 map when motion correction is not applied is not completely avoided with the CA. Indeed, it can be clearly observed that outliers are still present, in particular at the interfaces between ventricles and white matter (green arrow) for the mid-axial slice, and in the interfaces between white and grey matter for the top-axial slice Fig. 8(f). The T_1 map produced by Hallack's method seems free from outliers in the mid-axial but not in top-axial slice Fig. 8(g) (green arrow). The joint MLE provides T_1 maps which does not suffer from this issue. The motion parameter estimates obtained with the three methods are shown in the additional document (Fig. 3) which is included as part of the supplementary material which accompanies this paper.

2) In vivo sagittal human brain data: Two mid-sagittal slice of the estimated T_1 map for the three methods are shown in Fig. 9(b-d) and Fig. 9(j-l), respectively. In Fig. 9(a) and Fig. 9(i) estimated T_1 maps for the no motion correction case are presented.

In this case, the estimated T_1 map with the CA is almost free from outliers, which are widespread when no motion correction is accomplished, see magnified regions in Fig. 9(e-f) and Fig. 9(m-n). However, the CA sacrifices the final resolution of the T_1 map. It is clear that in the T_1 map obtained with Hallack's method and especially in that obtained with the joint MLE, structure details are better defined and contours better delineated, particularly in the interfaces between white/grey matter and CSF. Visual differences between the T_1 maps provided by Hallack's method and the joint MLE can be discerned as well. For example, the CSF infiltrating the cortical

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

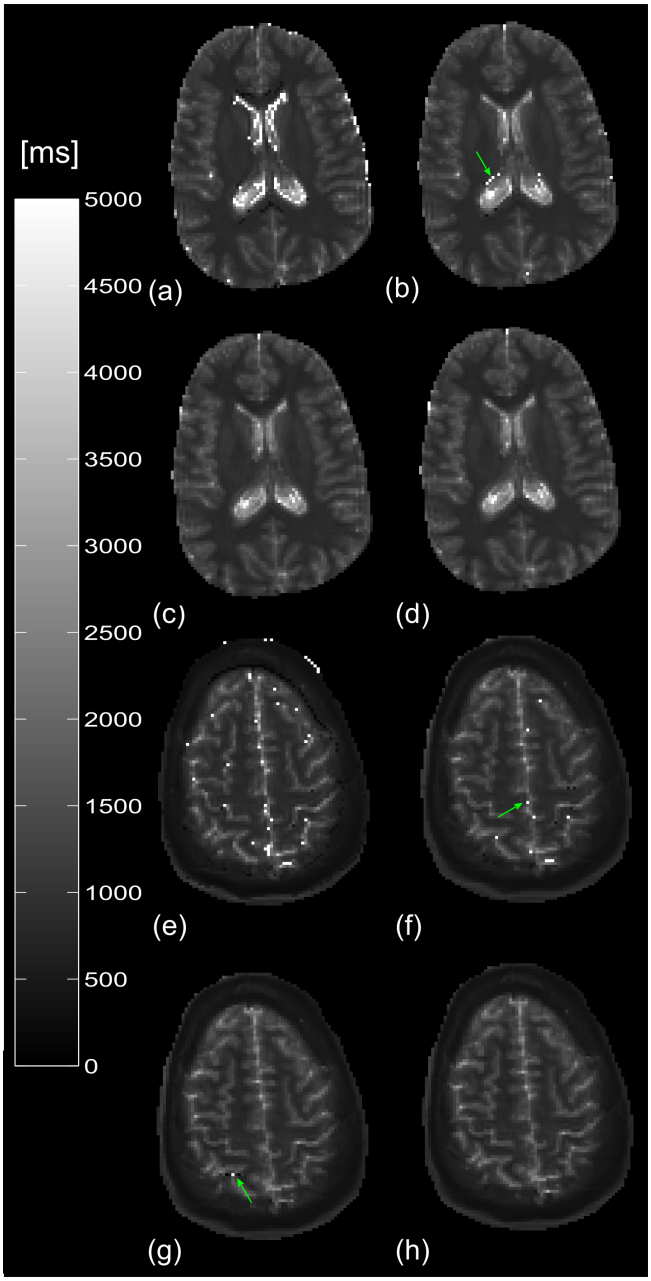

Fig. 8: Two axial slices of a whole human brain T_1 map. Mid-axial slice: (a) without motion correction, and corrected with (b) CA, (c) Hallack's method, (d) joint MLE. Top-axial slice: (e) without motion correction, and corrected with (f) CA, (g) Hallack's method, (h) joint MLE.

folds (see Fig. 9(g-h)) seems better defined with the joint MLE. In addition, yet a (reduced) number of outliers can be detected in the T_1 map estimated with the Hallack's method. See for example, Fig. 9(o). Such outliers, as in the in vivo axial experiment, are not present provided the joint MLE is applied. In Fig. 4 of the additional document, which is part of the downloadable supplementary material, the motion parameter estimates of this experiment are shown.

VI. Discussion

We presented a unified model-based approach for simultaneous motion correction and T_1 mapping that jointly estimates the motion parameters and the T_1 map using a maximum likelihood estimator (MLE). The proposed joint MLE possesses optimal statistical properties, which are shared by neither the conventional two-step approach (image registration prior to T_1 estimation) nor other heuristic integrated model-based methods.

Using realistic MC simulation experiments, it was shown that the proposed joint MLE outperforms existing T_1 mapping methods in terms of both accuracy and RMSE, next to providing more accurate motion parameter estimates. Results of the controlled experiment based on ground-truth real data are in line with the findings of MC simulations. We have shown that detailed and meaningful T_1 map can be recovered with the joint MLE under the influence of manually induced, severe motion. Quantitative comparison against a ground-truth T_1 map demonstrates the superior quality in T_1 map restoration compared to CA and Hallack's method.

Furthermore, the optimal unified ML framework has been validated with in vivo human brain data experiments, suffering form involuntary motion. From these experiments, it has become evident that motion correction is indispensable in T_1 mapping, even when subject motion is relatively small. Interestingly, yet recognizing the limitation of visual assessment in quantitative MRI, some of the rigorously derived statistical conclusions from the MC simulations can be noticed in the two whole brain human data sets. For instance, the arguably poorer motion estimation performance of the CA compared to model-based registration approaches, already reported by [10], [18] and confirmed by our MC results (see Fig. 3(d) and Table I), may be the cause of the presence of outliers in the estimated T_1 map and the loss of fine details. The presence of small number of outliers in the estimated T_1 map with Hallack's method can be attributed to its non-optimal/heuristic design. While improvement in motion estimation compared to CA has been demonstrated, optimality in terms of T_1 estimation cannot be theoretically and empirically guaranteed. In contrast, our ML framework combines, in a single integrated approach, the benefits of model-based registration techniques with optimal T_1 map restoration based on statistical theory, where the noise statistics are properly accounted for.

On top of that, our careful algorithm design avoids heavy computational burden. Note that the voxel-wise T_1 fitting task, which contributes most to the computational cost of the proposed joint MLE, is often also included in the iterative loop of other model-based integrated methods. Consequently, the computation time per iteration is comparable.

The proposed ML framework can be extended in different ways without compromising its optimal statistical properties. Extension to other MRI sequences or modalities is straightforward by substituting a properly modified parametric signal model for Eq. (1). Such potential extensions include T_2 and T_2^* mapping [29]. Moreover, multi-component T_2 mapping would benefit as well from the proposed ML framework [56]. Furthermore, it is worthwhile mentioning that when the joint

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

MLE proposed in this work is applied to the particular case of T_1 mapping using the spoiled gradient recalled echo (SPGR) sequence, its computational efficiency can even be further improved by using the recently proposed fast non-linear least squares T_1 estimator NOVIFAST [57] for solving the voxel-wise NLLS problems in P.2.

Extensions towards the inclusion of different types of motion, e.g., non-rigid motion, are also possible but require further study, which is considered future work. In this work, we have assumed a motion model which accounts for inter-image motion, that is, motion between each of the 3D T_1-weighted images. Although this model left aside intra-motion effects, which is known to affect the k-space reconstruction process, we have not observed any derived ghosting artifacts in the in vivo reconstructed T_1-weighted images. It should be noted that, to deal with such kind of motion, navigators [58] or advanced k-space reconstruction methods with motion correction [59] can be applied in conjunction with the joint MLE. However, though intra-image motion may be alleviated with such techniques, image registration, that is, inter-image motion correction, will remain necessary, which further emphasizes the relevance of the unified ML framework. Finally, it is relatively straightforward to extend the ML framework to cope with inter-slice motion, that is, motion occurring between the acquisition of 2D slices of a T_1-weighted dataset. An outlook to such an extension, which is especially relevant for image acquisition methods that acquire 3D volumes slice by slice sequentially, such as Echo Planar Imaging (EPI) sequences [44], is given in the additional document (Section IV) which is part of the downloadable supplementary material.

VII. Conclusion

In quantitative MR T_1 mapping, it is common practice to register the T_1-weighted images prior to T_1 map estimation. However, as demonstrated in this paper, this conventional two-step approach lacks high accuracy motion estimation and leads to biased T_1 estimates. Hence, we have proposed a rigorous unified framework for simultaneous motion and T_1 estimation using a Maximum Likelihood (ML) estimator. It has been demonstrated that the proposed joint MLE outperforms the conventional approach as well as a recently proposed model-based method [18] in terms of motion and T_1 estimation accuracy and RMSE. Our ML framework, which uses an efficient algorithm, has been validated in a controlled experiment with real T_1-weighted data and also with two in vivo human brain data sets. We believe that the unified ML framework possesses serious advantages over the conventional approach to replace it in clinical scenarios where precise and accurate T_1 estimates are the ultimate goal.

Acknowledgments

The authors would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the quality of the paper. The authors would also like to thank Santiago Aja-Fernández and Sayuan Liang for interesting discussions on noise estimation and acquisition settings, respectively.

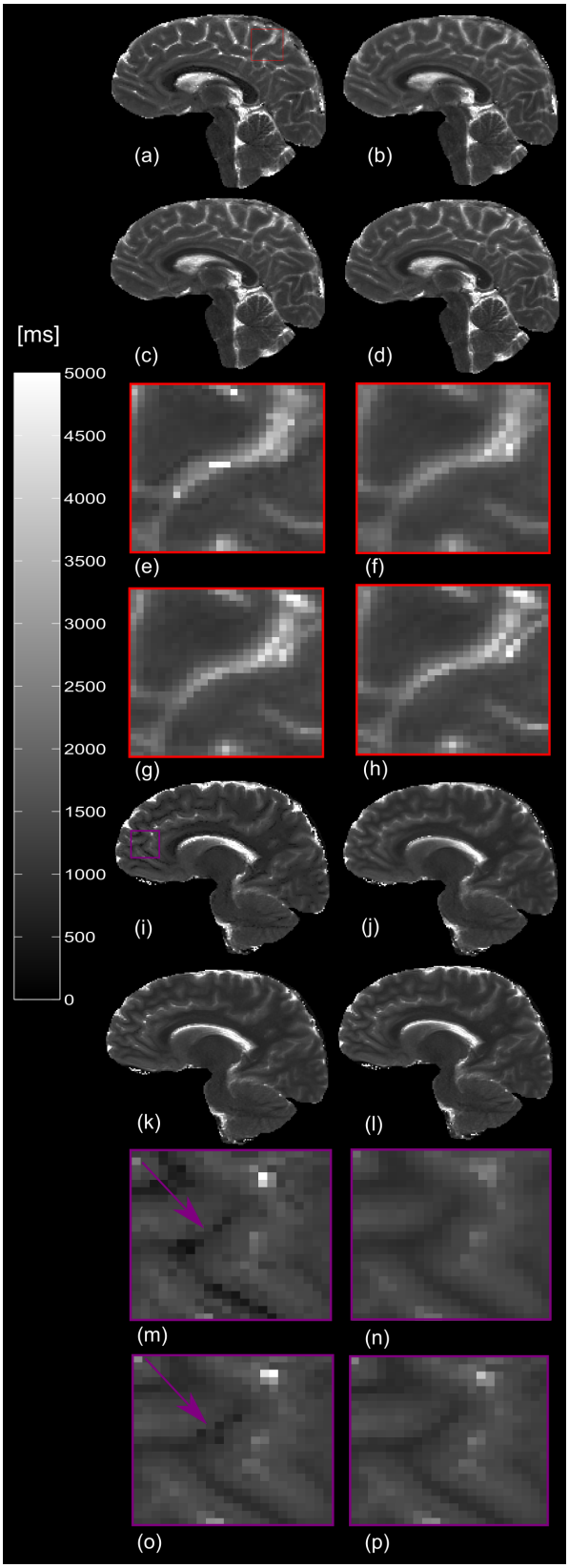

Fig. 9: Two mid-sagittal slices of a whole human brain T_1 map. First mid-sagittal slice: (a) without motion correction, (b) CA, (c) Hallack's method, (d) joint MLE. Magnified regions are shown in (e), (f), (g) and (h), respectively. Second mid-sagittal slice: (i) without motion correction, (j) CA, (k) Hallack's method, (l) joint MLE. Magnified regions are shown in (m), (n), (o) and (p), respectively.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

REFERENCES

[1] P. Tofts, *Quantitative MRI of the brain: measuring changes caused by disease*. Chichester, England: John Wiley & Sons, 2004.

[2] H. B. W. Larsson *et al.*, "Assessment of demyelination, edema, and gliosis by in vivo determination of T1 and T2 in the brain of patients with acute attack of multiple sclerosis," *Magn. Reson. Med.*, vol. 11, pp. 337–348, Sep 1989.

[3] P. Conlon, M. R. Trimble, D. Rogers, and C. Callicott, "Magnetic resonance imaging in epilepsy: a controlled study," *Epilepsy Res.*, vol. 2, pp. 37–43, Jan 1988.

[4] T. Erkinjuntti *et al.*, "Do white matter changes on MRI and CT differentiate vascular dementia from Alzheimer's disease?," *J. Neurol. Neurosurg. Psychiatry*, vol. 50, pp. 37–42, Jan 1987.

[5] J. Detre, J. Leigh, D. Williams, and A. Koretsky, "Perfusion imaging," *Magn. Reson. Med.*, vol. 23, pp. 37–45, Jan 1992.

[6] H.-L. M. Cheng, "T1 measurement of flowing blood and arterial input function determination for quantitative 3D T1-weighted DCE-MRI," *J. Magn. Reson. Imaging*, vol. 25, pp. 1073–1078, May 2007.

[7] A. van den Bos, *Parameter Estimation for Scientists and Engineers*. Hoboken, New Yersey, USA: John Wiley & Sons, 2007.

[8] J. D. Trzasko, P. M. Mostardi, S. J. Riederer, and A. Manduca, "Estimating T1 from multichannel variable flip angle SPGR sequences," *Magn. Reson. Med.*, vol. 69, pp. 1787–1794, Jun 2013.

[9] G. Van Steenkiste *et al.*, "Super-resolution T1 estimation: Quantitative high resolution T1 mapping from a set of low resolution T1-weighted images with different slice orientations." *Magn. Reson. Med.*, doi:10.1002/mrm.26262, 2016.

[10] W. Huizinga *et al.*, "PCA-based groupwise image registration for quantitative MRI," *Med. Image Anal.*, vol. 29, pp. 65 – 78, Apr 2016.

[11] B. U. Foerster, D. Tomasi, and E. C. Caparelli, "Magnetic field shift due to mechanical vibration in functional magnetic resonance imaging," *Magn. Reson. Med.*, vol. 54, no. 5, pp. 1261–1267, 2005.

[12] S. C. Deoni, B. K. Rutt, and T. M. Peters, "Rapid combined T1 and T2 mapping using gradient recalled acquisition in the steady state," *Magn. Reson. Med.*, vol. 49, pp. 515–526, Mar 2003.

[13] J. B. M. Warntjes, O. Dahlqvist, and P. Lundberg, "Novel method for rapid, simultaneous T1, T2*, and proton density quantification," *Magn. Reson. Med.*, vol. 57, pp. 528–537, Mar 2007.

[14] U. Studler *et al.*, "Impact of motion on T1 mapping acquired with inversion recovery fast spin echo and rapid spoiled gradient recalled-echo pulse sequences for delayed gadolinium-enhanced MRI of cartilage (dGEMRIC) in volunteers," *J. Magn. Reson. Imaging*, vol. 32, pp. 394–398, Aug 2010.

[15] E. E. Bron *et al.*, "Image registration improves human knee cartilage T1 mapping with delayed gadolinium-enhanced MRI of cartilage (dGEMRIC)," *Eur. Radiol.*, vol. 23, no. 1, pp. 246–252, 2013.

[16] G. Ramos-Llordén *et al.*, "Simultaneous motion correction and T1 estimation in quantitative T1 mapping: An ML restoration approach," in *Proc. IEEE ICIP*, pp. 3160–3164, 2015.

[17] H. Xue *et al.*, "Phase-sensitive inversion recovery for myocardial T1 mapping with motion correction and parametric fitting," *Magn. Reson. Med.*, vol. 69, pp. 1408–1420, May 2013.

[18] A. Hallack, M. A. Chappel, M. Gooding, and J. Schnabel, "A new similarity metric for groupwise registration of variable flip angle sequences for improved T_{10} estimation in DCE-MRI," in *Biomedical Image Registration*, vol. 8545 of *Lect. Notes Comput. Sc.*, pp. 154–163, 2014.

[19] J. Sijbers, A. J. den Dekker, P. Scheunders, and D. Van Dyck, "Maximum-likelihood estimation of Rician distribution parameters," *IEEE Trans. Med. Imaging*, vol. 17, pp. 357–361, Jun 1998.

[20] J. Sijbers, A. J. den Dekker, E. Raman, and D. Van Dyck, "Parameter estimation from magnitude MR images," *Int. J. Imag. Syst. Tech.*, vol. 10, pp. 109–114, Mar 1999.

[21] A. J. den Dekker and J. Sijbers, "Data distributions in magnetic resonance images: A review," *Phys. Medica*, vol. 30, pp. 725–741, Nov 2014.

[22] J. A. Fessler and D. Kim, "Axial block coordinate descent (ABCD) algorithm for X-ray CT image reconstruction," in *Proc. Intl. Mtg. on Fully 3D image Recon. in Rad. and Nuc. Med.*, pp. 262–5, 2011.

[23] D. Varadarajan and J. P. Haldar, "A majorize-minimize framework for Rician and non-central chi MR images," *IEEE Trans. Med. Imag*, vol. 34, pp. 2191–2202, Oct 2015.

[24] D. R. Messroghli *et al.*, "Modified look-locker inversion recovery (MOLLI) for high-resolution T1 mapping of the heart," *Magn. Reson. Med.*, vol. 52, pp. 141–146, Jul 2004.

[25] J. K. Barral *et al.*, "A robust methodology for in vivo T1 mapping," *Magn. Reson. Med.*, vol. 64, pp. 1057–1067, Oct 2010.

[26] H. Goldstein *et al.*, *Classical Mechanics*. Harlow, Essex, England: Pearson Higher Ed., 2014.

[27] K. G. Larkin, M. A. Oldfield, and H. Klemm, "Fast Fourier method for the accurate rotation of sampled images," *Opt. Commun.*, vol. 139, pp. 99 – 106, Jun 1997.

[28] H. Gudbjartsson and S. Patz, "The Rician distribution of noisy MRI data," *Magn. Reson. Med.*, vol. 34, pp. 910–914, Feb 1995.

[29] J. Sijbers, A. J. den Dekker, M. Verhoye, E. R. Raman, and D. Van Dyck, "Optimal estimation of T2 maps from magnitude MR images," in *Proc. SPIE'98: Medical Imaging*, vol. 3338, pp. 384–390, 1998.

[30] C. Constantinides, E. Atalar, and E. McVeigh, "Signal-to-noise measurements in magnitude images from NMR phased arrays," *Magn. Reson. Med.*, vol. 38, pp. 852–857, Nov 1997.

[31] S. Aja-Fernández and A. Tristán-Vega, "A review on statistical noise models for magnetic resonance imaging," tech. rep., Univ. Valladolid, Valladolid, Spain. Tech. Rep. LPI2013-01, Jun 2013.

[32] O. Dietrich *et al.*, "Influence of multichannel combination, parallel imaging and other reconstruction techniques on MRI noise characteristics," *Magn. Reson. Imaging*, vol. 26, pp. 754 – 762, Jul 2008.

[33] S. Aja-Fernández, T. Pieciak, and G. Vegas-Sánchez-Ferrero, "Spatially variant noise estimation in MRI: A homomorphic approach," *Med. Image Anal.*, vol. 20, no. 1, pp. 184 – 197, 2015.

[34] B. Goossens, A. Pizurica, and W. Philips, "Wavelet domain image denoising for non-stationary noise and signal-dependent noise," in *Proc. IEEE ICIP*, pp. 1425–1428, Oct 2006.

[35] M. Hong, M. Razaviyayn, Z.-Q. Luo, and J.-S. Pang, "A unified algorithmic framework for block-structured optimization involving big data: With applications in machine learning and signal processing," *IEEE Signal Process. Mag.*, vol. 33, pp. 57–77, Jan 2016.

[36] J. Fan, M. Farmen, and I. Gijbels, "Local maximum likelihood estimation and inference," *J. R. Stat. Soc. Series B. Stat. Methodol.*, vol. 60, pp. 591–608, Jan 1998.

[37] W. H. Press *et al.*, *Numerical Recipes in C++: the art of scientific computing*. New York, NJ, USA: Cambridge Univ. Press, 2007.

[38] R. Eglese, "Simulated annealing: a tool for operational research," *European journal of operational research*, vol. 46, no. 3, pp. 271–281, 1990.

[39] W. L. Goffe, G. D. Ferrier, and J. Rogers, "Global optimization of statistical functions with simulated annealing," *J. Econom.*, vol. 60, no. 1-2, pp. 65–99, 1994.

[40] D. Bertsimas and O. Nohadani, "Robust optimization with simulated annealing," *J. Global Optim.*, vol. 48, no. 2, pp. 323–334, 2010.

[41] J. Nocedal and S. Wright, *Numerical optimization*. New York, NJ, USA: Springer-Verlag, 2006.

[42] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Am. Stat.*, vol. 58, pp. 30–37, Jan 2004.

[43] M. J. Muckley, D. C. Noll, and J. A. Fessler, "Fast parallel MR image reconstruction via B1-based, adaptive restart, iterative soft thresholding algorithms (BARISTA)," *IEEE Trans. Med. Imag*, vol. 34, pp. 578–588, Oct 2015.

[44] M. Fogtmann *et al.*, "A unified approach to diffusion direction sensitive slice registration and 3-D DTI reconstruction from moving fetal brain anatomy," *IEEE Trans. Med. Imag*, vol. 33, pp. 272–289, Feb 2014.

[45] D. Mattes *et al.*, "PET-CT image registration in the chest using free-form deformations," *IEEE Trans. Med. Imag*, vol. 22, pp. 120–128, Jan 2003.

[46] J. P. W. Pluim, J. B. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE Trans. Med. Imag*, vol. 22, pp. 986–1004, Aug 2003.

[47] G. E. Gold *et al.*, "Musculoskeletal MRI at 3.0 T: relaxation times and image contrast," *Am. J. Roentgenol.*, vol. 183, pp. 343–351, Aug 2004.

[48] P. J. Wright *et al.*, "Water proton T1 measurements in brain tissue at 7, 3, and 1.5T using IR-EPI, IR-TSE, and MPRAGE: results and optimization," *Magn. Reson. Mater. Phy.*, vol. 21, pp. 121–130, Mar 2008.

[49] J. Sijbers and A. J. den Dekker, "Maximum likelihood estimation of signal amplitude and noise variance from MR data," *Magn. Reson. Med.*, vol. 51, pp. 586–594, Mar 2004.

[50] S. Aja-Fernández, G. Vegas-Sánchez-Ferrero, and A. Tristán-Vega, "Noise estimation in parallel MRI: GRAPPA and SENSE," *Magn. Reson. Imaging*, vol. 32, no. 3, pp. 281–290, 2014.

[51] M. Bernstein, K. King, and X. Zhou, *Handbook of MRI Pulse Sequences*. Burlington, MA, USA: Elsevier Inc, 2004.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

This is the author's version of an article that has been published in this journal. Changes were made to this version by the publisher prior to publication.
The final version of record is available at http://dx.doi.org/10.1109/TMI.2016.2611653

[52] K. P. Pruessmann, M. Weiger, B. M. Scheidegger, and P. Boesiger, "SENSE: sensitivity encoding for fast MRI," *Magn. Reson. Med.*, vol. 42, pp. 952–962, Nov 1999.
[53] M. D. DeVore, A. D. Lanterman, and J. A. O'Sullivan, "ATR performance of a Rician model for SAR images," in *AeroSense 2000*, pp. 34–45, International Society for Optics and Photonics, 2000.
[54] H. W. Park, M. H. Cho, and Z. H. Cho, "Time-multiplexed multislice inversion-recovery techniques for NMR imaging," *Magnetic Resonance in Medicine*, vol. 2, no. 6, pp. 534–539, 1985.
[55] M. A. Griswold *et al.*, "Generalized autocalibrating partially parallel acquisitions (GRAPPA)," *Magn. Reson. Med.*, vol. 47, pp. 1202–1210, Jun 2002.
[56] M. Björk, D. Zachariah, J. Kullberg, and P. Stoica, "A multicomponent T2 relaxometry algorithm for myelin water imaging of the brain," *Magn. Reson. Med.*, vol. 75, no. 1, pp. 390–402, 2016.
[57] G. Ramos-Llordén, A. J. den Dekker, M. Björk, M. Verhoye, and J. Sijbers, "NOVIFAST: A fast non-linear least squares method for accurate and precise estimation of T1 from SPGR signals," *Proc. Intl. Soc. Mag. Reson. Med.*, vol. 24, p. 2820, 2016.
[58] E. B. Welch, A. Manduca, R. C. Grimm, H. A. Ward, and C. R. Jack Jr, "Spherical navigator echoes for full 3D rigid body motion measurement in MRI," *Magn. Reson. Med.*, vol. 47, no. 1, pp. 32–41, 2002.
[59] L. Cordero-Grande, R. P. A. G. Teixeira, E. J. Hughes, J. Hutter, A. N. Price, and J. Hajnal, "Sensitivity encoding for aligned multishot magnetic resonance reconstruction," *IEEE Trans. Comput. Imaging*, vol. 2, no. 3, pp. 266–280, 2016.

Copyright (c) 2016 IEEE. Personal use is permitted. For any other purposes, permission must be obtained from the IEEE by emailing pubs-permissions@ieee.org.

Additional document for "A unified Maximum Likelihood framework for simultaneous motion and T_1 estimation in quantitative MR T_1 mapping"

Gabriel Ramos-Llordén, *Member, IEEE,* Arnold J. den Dekker, Gwendolyn Van Steenkiste, Ben Jeurissen, Floris Vanhevel, Johan Van Audekerke, Marleen Verhoye and Jan Sijbers

***Abstract*—In Section I of this additional document we provide details on the motion operator $\boldsymbol{H}_{\boldsymbol{\theta}_n}$. In Section II, we prove Eqs.(16-17) of the main body of the paper, which gives a complete description of the MM framework used in P.2. In Section III, we show graphs of estimated motion parameters for the in vivo experiments and one of the simulation experiments described in the main body of the paper. Finally, Section IV discusses an extension of the joint MLE to account for intra-image motion.**

I. Details on motion operator $\boldsymbol{H}_{\boldsymbol{\theta}_n}$

In this section, an explicit expression of the motion operator $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ is derived. Furthermore, we sketch the proof for its unitarity property, i.e., $\boldsymbol{H}_{\boldsymbol{\theta}_n}^H \boldsymbol{H}_{\boldsymbol{\theta}_n} = \boldsymbol{H}_{\boldsymbol{\theta}_n} \boldsymbol{H}_{\boldsymbol{\theta}_n}^H = \boldsymbol{I}$, with $\boldsymbol{I}$ the identity matrix. Specific details can be found in [1], [2] and especially in [3].

Let $\boldsymbol{r}_m^n$ be a spatial point related to the reference-system point $\boldsymbol{r}_m$ through a rigid transformation matrix $\boldsymbol{M}_{\boldsymbol{\theta}_n} \in \mathbb{R}^{4\times4}$:

$$\begin{pmatrix} \boldsymbol{r}_m^n \\ 1 \end{pmatrix} = \boldsymbol{M}_{\boldsymbol{\theta}_n} \begin{pmatrix} \boldsymbol{r}_m \\ 1 \end{pmatrix}. \tag{S1}$$

The rigid transformation matrix $\boldsymbol{M}_{\boldsymbol{\theta}_n} \in \mathbb{R}^{4\times4}$, which includes 3D rotation and translation, can then be written as [4]:

$$\boldsymbol{M}_{\boldsymbol{\theta}_n} = \begin{pmatrix} \boldsymbol{R}(\alpha_n, \beta_n, \gamma_n) & \boldsymbol{t}_n \\ \boldsymbol{0}^T & 1 \end{pmatrix} \quad , \tag{S2}$$

with $\boldsymbol{t}_n = (t_{xn}, t_{y_n}, t_{zn})^T$ a vector of translation parameters, $\boldsymbol{0}^T$ a 1×3 zero vector, and $\boldsymbol{R}(\alpha_n, \beta_n, \gamma_n) \in \mathbb{R}^{3\times3}$ the product of three elementary rotation matrices $\big(\boldsymbol{R}_x(\alpha_n)$, $\boldsymbol{R}_y(\beta_n)$ and $\boldsymbol{R}_z(\gamma_n)\big)$ describing rotations around the x, y and z axis, with angles α_n, β_n and γ_n, respectively. With such parametrization, we get

$$\boldsymbol{r}_m^n = \boldsymbol{R}_x(\alpha_n)\boldsymbol{R}_y(\beta_n)\boldsymbol{R}_z(\gamma_n)\boldsymbol{r}_m + \boldsymbol{t}_n \quad . \tag{S3}$$

Let $f(\cdot)$ be spatially-continuous function (a relaxation model in our problem). Then, to calculate $f(\boldsymbol{r}_m^n)$, the following spatial transformations on $f(\cdot)$ are consecutively applied.

1) 3D translation: $f_{\text{T}}(\boldsymbol{r}) = f(\boldsymbol{r} + \boldsymbol{t}_n)$
2) Rotation around x axis $f_{\text{Rot-x}}(\boldsymbol{r}) = f_{\text{T}}(\boldsymbol{R}_x\boldsymbol{r})$
3) Rotation around y axis $f_{\text{Rot-y}}(\boldsymbol{r}) = f_{\text{Rot-x}}(\boldsymbol{R}_y\boldsymbol{r})$
4) Rotation around z axis $f_{\text{Rot-z}}(\boldsymbol{r}) = f_{\text{Rot-y}}(\boldsymbol{R}_z\boldsymbol{r})$

Indeed, by evaluating $f_{\text{Rot-z}}(\cdot)$ at $\boldsymbol{r}_m$, we get $f(\boldsymbol{r}_m^n)$.

In a discrete domain, each of the previous four operations is represented by linear operators, hence matrices, that we denote as $\boldsymbol{H}_{\text{T}}$, $\boldsymbol{H}_{\text{Rot-x}}$, $\boldsymbol{H}_{\text{Rot-y}}$, and $\boldsymbol{H}_{\text{Rot-z}}$, respectively. Note that we have omitted the dependence on the motion parameter for the sake of notational convenience.

As a consequence, the motion operator, $\boldsymbol{H}_{\boldsymbol{\theta}_n}$, can be written as

$$\boldsymbol{H}_{\boldsymbol{\theta}_n} = \boldsymbol{H}_{\text{Rot-z}}\boldsymbol{H}_{\text{Rot-y}}\boldsymbol{H}_{\text{Rot-x}}\boldsymbol{H}_{\text{T}} \tag{S4}$$

and its Hermitian transpose as

$$\boldsymbol{H}_{\boldsymbol{\theta}_n}^H = \boldsymbol{H}_{\text{T}}^H\boldsymbol{H}_{\text{Rot-x}}^H\boldsymbol{H}_{\text{Rot-y}}^H\boldsymbol{H}_{\text{Rot-z}}^H. \tag{S5}$$

It is clear that if $\boldsymbol{H_T}$, $\boldsymbol{H}_{\text{Rot-x}}$, $\boldsymbol{H}_{\text{Rot-y}}$ and $\boldsymbol{H}_{\text{Rot-z}}$ are unitary, $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ is unitary as well.

Sketch of Proof 1: $\boldsymbol{H}_T$ is unitary

The translation operator $\boldsymbol{H}_{\text{T}}$ consists of 1) a 3D FFT, 2) a voxel-wise multiplication with a purely complex exponential whose phase depends linearly on the translation parameters, and 3) an inverse 3D FFT [3]. By noting that the multidimensional FFT is a unitary operator [2], the translation operator can be succinctly written as

$$\boldsymbol{H}_{\text{T}} = \boldsymbol{F}_{\text{3D}}^H\boldsymbol{\Delta}\boldsymbol{F}_{\text{3D}} \quad , \tag{S6}$$

where $\boldsymbol{F}_{\text{3D}}$ is the 3D unitary Discrete Fourier Transform (DFT) matrix and $\boldsymbol{\Delta}$ is a diagonal matrix whose entries are purely complex exponentials.

It is known that $\boldsymbol{\Delta}$ is a unitary matrix if and only if the modulus of each diagonal entries is one. Since this is always true for purely complex exponentials, it demonstrates that $\boldsymbol{H}_{\text{T}}$ is unitary.

Sketch of Proof 2: $\boldsymbol{H}_{\text{Rot-x}}$, $\boldsymbol{H}_{\text{Rot-y}}$, and $\boldsymbol{H}_{\text{Rot-z}}$ are unitary.

For brevity, we present the proof only for $\boldsymbol{H}_{\text{Rot-x}}$. The proof for $\boldsymbol{H}_{\text{Rot-y}}$ and $\boldsymbol{H}_{\text{Rot-z}}$ is completely similar.

Because $\boldsymbol{R}_x(\alpha_n)$ can be decomposed as the product of three one-dimensional shear matrices [3], it is possible to write

$$\boldsymbol{H}_{\text{Rot-x}} = \boldsymbol{S_x}\boldsymbol{S_y}\boldsymbol{S_x} \quad , \tag{S7}$$

where $\boldsymbol{S_x}$ and $\boldsymbol{S_y}$ are fractional delay filters [1], which model the shearings in the x and y dimension, respectively. Note that these filters can be implemented efficiently with FFT [3]. If both $\boldsymbol{S_x}$ and $\boldsymbol{S_y}$ are unitary, $\boldsymbol{H}_{\text{Rot-x}}$ is unitary as well. Indeed, $\boldsymbol{S_x}$ has essentially the same diagonal expression as Eq. (S6), where the role of the 3D DFT matrices is fulfilled by a (unitary) Fourier matrix which applies an FFT only along the x direction. The phase of the complex exponential

in the diagonal matrix now depends linearly on the shearing parameter [3], which is a real value. Therefore, the associated diagonal matrix is unitary. The unitarity property of $\boldsymbol{S_x}$ follows immediately.

The proof for $\boldsymbol{S_y}$ is equivalent, with the exception that the unitary Fourier matrix now represents an FFT along the y direction. We can prove then that $\boldsymbol{S_y}$ is unitary and thus $\boldsymbol{H}_{\text{Rot-x}}$ is unitary.

As already mentioned, the proof for $\boldsymbol{H}_{\text{Rot-y}}$ and $\boldsymbol{H}_{\text{Rot-z}}$ are analogous. Combining Proof 1 and Proof 2, the unitary property of $\boldsymbol{H}_{\boldsymbol{\theta}_n}$ is demonstrated.

II. Separable Quadratic Surrogate (SQS) function derivation for the joint MLE

In order to get the final version of the joint MLE algorithm, a necessary step was to obtain a surrogate function for

$$G_n(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = ||\boldsymbol{W}_n^{1/2}\big(\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}) - \breve{\boldsymbol{s}}_n^k\big)||_2^2 + C_n(k) \tag{S8}$$

with $\boldsymbol{W}_n = \text{diag}\{\frac{1}{2\boldsymbol{\sigma}_n^2}\}$.

The choice we made in this work was a SQS function [5], that when applied to Eq.(S8), yields Eq.(16-17). Here, we present the proof of these equations. To that end, we build on results presented in [5]. In that work, a SQS function was applied to a generic quadratic form $\frac{1}{2}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2$. Such SQS function had the following expression:

$$\frac{1}{2}||\boldsymbol{x} - \big(\boldsymbol{x}^k - \boldsymbol{D_f^{-1}}\boldsymbol{A}^H\big(\boldsymbol{A}\boldsymbol{x}^k - \boldsymbol{y}\big)\big)||_{\boldsymbol{D_f}}^2 + \xi, \tag{S9}$$

with ξ a constant independent of $\boldsymbol{x}$ and where the matrix $\boldsymbol{D_f}$ is defined in such way that it satisfies $\boldsymbol{D_f} \succeq \boldsymbol{A}^H\boldsymbol{A}$, that is, $\boldsymbol{D_f} - \boldsymbol{A}^H\boldsymbol{A}$ is a positive-semidefinite matrix.

We can easily identify the terms of the quadratic form at hand, i.e., $G_n(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$, with the terms of $\frac{1}{2}||\boldsymbol{y} - \boldsymbol{A}\boldsymbol{x}||_2^2$, and hence easily define our SQS function, $G_n^*(\boldsymbol{\kappa}, \boldsymbol{T_1}|, \boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)$, as

$$G_n^*(\boldsymbol{\kappa}, \boldsymbol{T_1}|\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = ||\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1}) - \boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)||_{\boldsymbol{D_f}}^2 + C_n^*(k), \tag{S10}$$

with

$$\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = \boldsymbol{f}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) - \boldsymbol{D_f^{-1}}\boldsymbol{A}^H(\boldsymbol{A}\boldsymbol{f}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) - \boldsymbol{y}), \tag{S11}$$

$C_n^*(k)$ a constant independent of $\boldsymbol{\kappa}$ and $\boldsymbol{T_1}$, and where $\boldsymbol{A} = \boldsymbol{W}_n^{1/2}\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}$ and $\boldsymbol{y} = \boldsymbol{W}_n^{1/2}\breve{\boldsymbol{s}}_n^k$.

After some algebra, we obtain

$$\boldsymbol{\rho}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) = \boldsymbol{f}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k) + \boldsymbol{D_f^{-1}}\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H\boldsymbol{W}_n\big(\breve{\boldsymbol{s}}_n^k - \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}\boldsymbol{f}_n(\boldsymbol{\kappa}^k, \boldsymbol{T_1}^k)\big). \tag{S12}$$

Before giving an expression for $\boldsymbol{D_f}$ satisfying $\boldsymbol{D_f} \succeq \boldsymbol{A}^H\boldsymbol{A}$, first we recognize that

$$\boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H\boldsymbol{W}_n^{1/2^H}\boldsymbol{W}_n^{1/2}\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}} = \boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H\boldsymbol{W}_n\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}. \tag{S13}$$

Furthermore, it is easy to show that the diagonal matrix $\boldsymbol{W}_n$ fulfills $\boldsymbol{W}_n \preceq (\sigma_n^*)^{-1}\boldsymbol{I}$ with $(\sigma_n^*)^{-1}$ being the maximum value along its diagonal, which is,

$$(\sigma_n^*)^{-1} \triangleq \frac{1}{2\min\limits_m [\boldsymbol{\sigma}_n]_m^2}. \tag{S14}$$

Thus, if $\boldsymbol{W}_n \preceq (\sigma_n^*)^{-1}\boldsymbol{I}$, it follows that $\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H\boldsymbol{W}_n\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}} \preceq (\sigma_n^*)^{-1}\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}}^H\boldsymbol{H}_{\hat{\boldsymbol{\theta}}_n^{(t+1)}} = (\sigma_n^*)^{-1}\boldsymbol{I}$, since the motion operator is unitary. Therefore, by defining $\boldsymbol{D_f}$ as $\boldsymbol{D_f} \triangleq (\sigma_n^*)^{-1}\boldsymbol{I}$, $\boldsymbol{D_f} \succeq \boldsymbol{A}^H\boldsymbol{A}$ holds.

Note that previous $\boldsymbol{D_f}$ definition depends on n and hence the NNLS problem (Eq.S9) is weighted differently along dimension n. To provide an unweighted NNLS problem, that is, the version we have presented in the main body of the paper, we set $\sigma^* = 2\min\limits_{n,m}[\boldsymbol{\sigma}_n]_m^2$. Clearly $\boldsymbol{W}_n \preceq (\sigma^*)^{-1}\boldsymbol{I}$ for all n. Hence, we redefine $\boldsymbol{D_f}$ as $\boldsymbol{D_f} \triangleq (\sigma^*)^{-1}\boldsymbol{I}$, and trivially we get $\boldsymbol{D_f} \succeq \boldsymbol{A}^H\boldsymbol{A}$ as desired. By substituting $\boldsymbol{D_f}$ into Eq.(S11), we arrive at the final expressions which are shown in the main body of the paper.

III. Graphs of motion

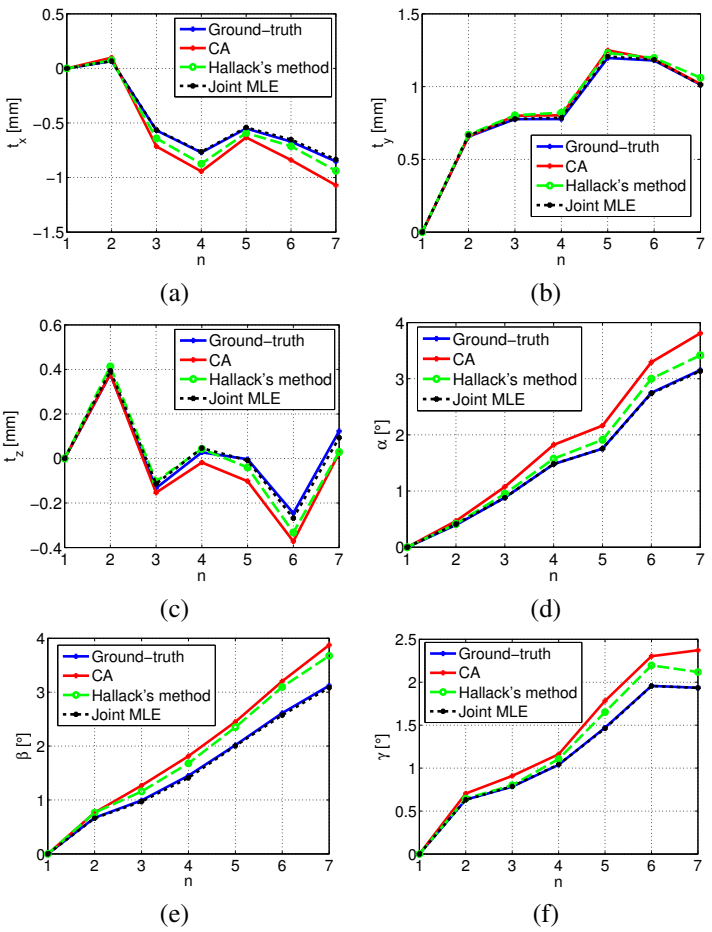

Fig. 1: Graphs of the ground-truth and estimated motion parameters for one realization of the simulation experiment with rotational motion: (a) t_x, (b) t_y, (c) t_z, (d) α, (e) β, (f) γ.

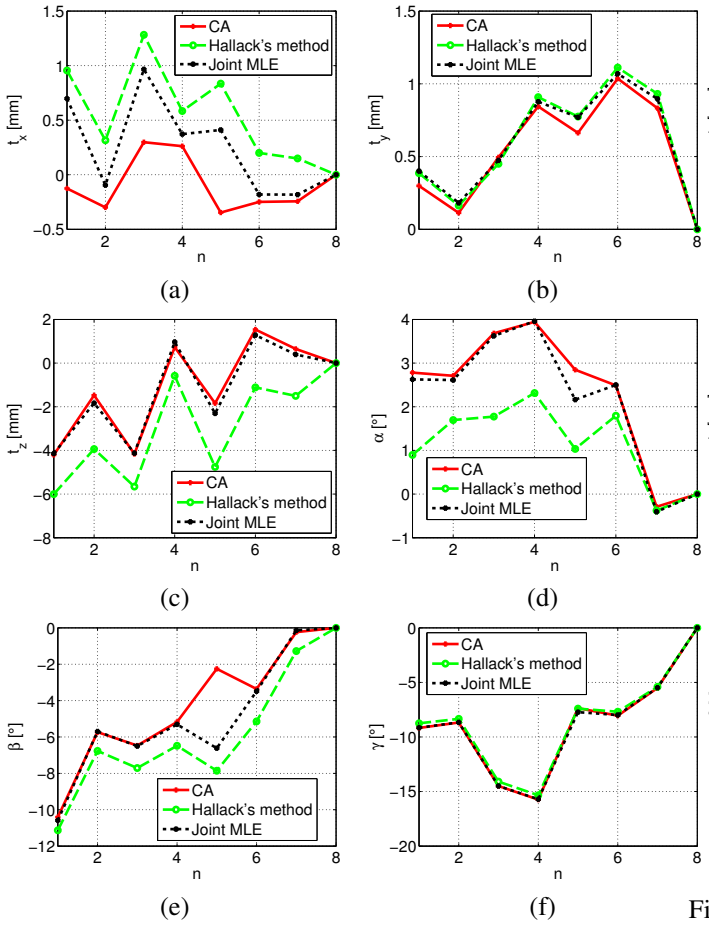

Fig. 2: Graphs of the estimated motion parameters for the ground-truth watermelon experiment: (a) t_x, (b) t_y, (c) t_z, (d) α, (e) β, (f) γ.

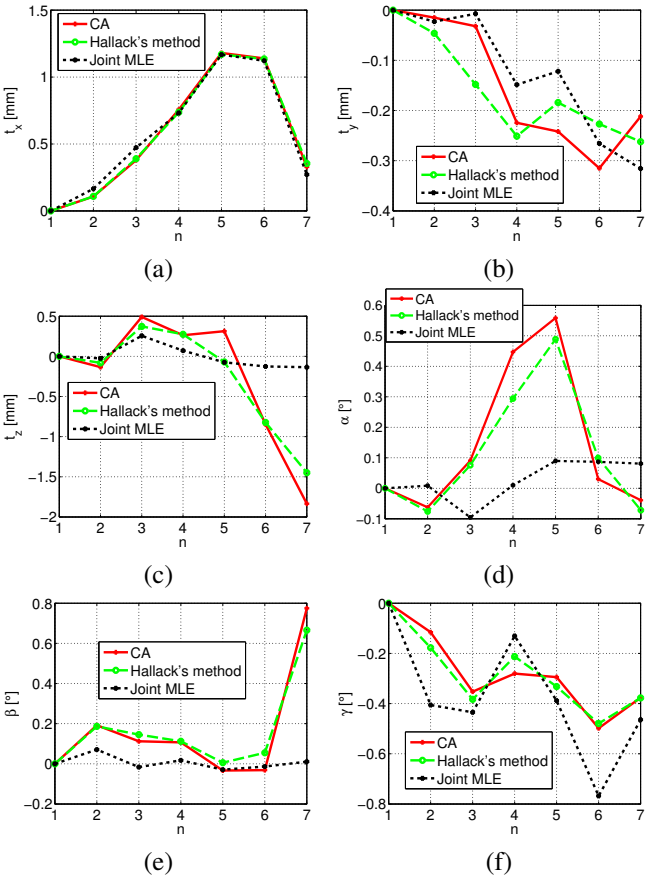

Fig. 3: Graphs of the estimated motion parameters for the in vivo axial human brain data experiment: (a) t_x, (b) t_y, (c) t_z, (d) α, (e) β, (f) γ.

IV. Extension of the joint MLE to account for intra-image motion

As mentioned in the discussion section of the main body of the paper, the joint MLE can be extended to include intra-image motion, in particular, motion between the acquisition of the different slices of a multi-slice image. A brief outlook to such an extension is given here. The implementation of the extended algorithm should take into account the following considerations. First, given an inversion time TI_n, the z-th noiseless and motion-corrupted 2D slice T_1-weighted $\tilde{\boldsymbol{f}}_{n,z}(\boldsymbol{\theta}_{n,z}, \boldsymbol{\kappa}, \boldsymbol{T_1})$ is related to the unobserved 3D image, $\boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$, through the motion parameters $\boldsymbol{\theta}_{n,z}$. Note that the number of motion parameters scales with $M_z \times N$, where M_z is the number of slices. Second, the mapping between a 2D slice T_1-weighted image and the noiseless unobserved 3D image also requires a slice-selective profile filter, which can be included as a matrix $\boldsymbol{\delta}_z$ [6], just after the motion operator, that is, $\tilde{\boldsymbol{f}}_{n,z}(\boldsymbol{\theta}_{n,z}, \boldsymbol{\kappa}, \boldsymbol{T_1}) = \boldsymbol{\delta}_z \boldsymbol{H}_{\boldsymbol{\theta}_{n,z}} \boldsymbol{f}_n(\boldsymbol{\kappa}, \boldsymbol{T_1})$. The final details of the derivation of the MM algorithm are beyond the scope of this paper and therefore not presented here.

References

[1] L. Condat, D. Van De Ville, and B. Forster-Heinlein, "Reversible, fast, and high-quality grid conversions," *IEEE Trans. Image Process.*, vol. 17, pp. 679–693, May 2008.

[2] A. K. Jain, *Fundamentals of digital image processing*. Englewood Cliffs, New Yersey, USA: Prentice-Hall, Inc., 1989.

[3] K. G. Larkin, M. A. Oldfield, and H. Klemm, "Fast Fourier method for the accurate rotation of sampled images," *Opt. Commun.*, vol. 139, pp. 99 – 106, Jun 1997.

[4] H. Goldstein *et al.*, *Classical Mechanics*. Harlow, Essex, England: Pearson Higher Ed., 2014.

[5] M. J. Muckley, D. C. Noll, and J. A. Fessler, "Fast parallel MR image reconstruction via B1-based, adaptive restart, iterative soft thresholding algorithms (BARISTA)," *IEEE Trans. Med. Imag*, vol. 34, pp. 578–588, Oct 2015.

[6] A. Gholipour, J. A. Estroff, and S. K. Warfield, "Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain mri," *IEEE Trans. Med. Imag*, vol. 29, no. 10, pp. 1739–1758, 2010.

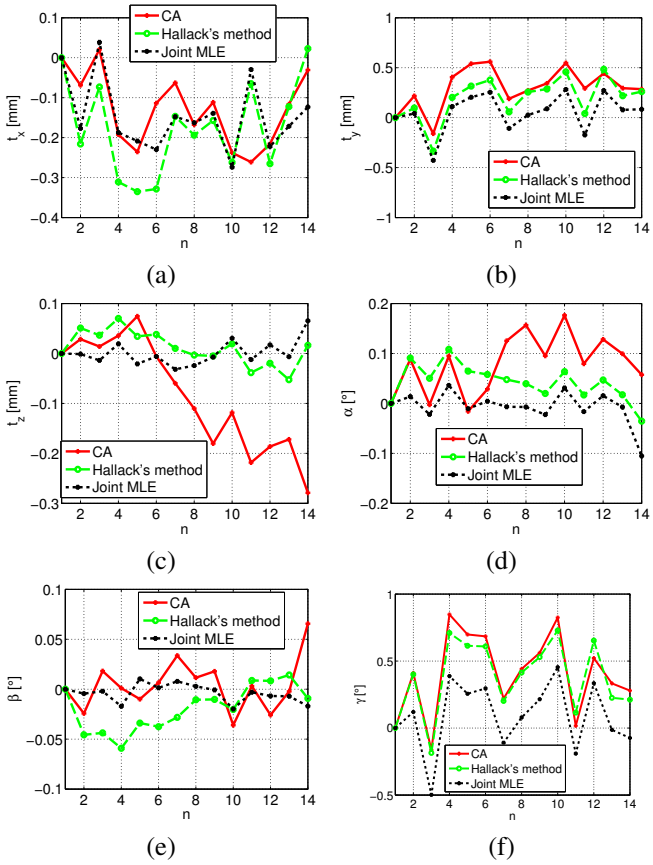

Fig. 4: Graphs of the estimated motion parameters for in vivo sagittal human brain data experiment. (a) t_x, (b) t_y, (c) t_z, (d) α, (e) β, (f) γ.

View publication stats