Joint retrospective motion correction and reconstruction for brain MRI with a reference contrast

Gabrio Rizzuti[1,2], Alessandro Sbrizzi[2], and Tristan van Leeuwen[1,3]

[1]Utrecht University, Utrecht, 3584 CS, The Netherlands
[2]Universitair Medisch Centrum Utrecht, Utrecht, 3584 CX, The Netherlands
[3]Centrum Wiskunde & Informatica, Amsterdam, 1098 XG, The Netherlands

***Abstract*—We present a retrospective joint motion correction and reconstruction scheme for magnetic resonance imaging to reduce the imprint of subject motion on the reconstructed images. In multi-contrast imaging, reconstructions pertaining to distinct acquisition sequences (e.g., T1 or T2 weighted images) might not be equally affected by motion, due to the sequential nature of the data acquisition process or the specific sequence design. To avoid repeating the corrupted scan, we can leverage the uncorrupted reconstructions to post-process the contrasts that are most severely affected by motion, by assuming a shared underlying anatomy. Only rigid motion is considered here, but no further assumptions are required. Classical motion correction schemes are combined with weighted total-variation regularization, whose weight is defined by the structure of the well-resolved contrasts. We will particularly focus on brain imaging with conventional Cartesian sampling. We envision a practical workflow that can easily fit into the existing clinical practice without the need for changing the MRI acquisition protocols.**

***Index Terms*—Magnetic resonance imaging, motion correction, compressed sensing, structure-guided regularization**

I. Introduction

MAGNETIC resonance imaging (MRI) is a traditional medical imaging technique that utilizes non-ionizing magnetic fields to form detailed pictures of the human body. Compared to other imaging modalities (such as computed tomography, ultrasound, or photoacoustic imaging), MRI is especially apt at detecting and characterizing soft-tissue pathologies, but the data collection requires lengthier sessions (which last several minutes, as opposed to, e.g., several seconds for computed tomography). Since the timescale of patient motion is much smaller than the data acquisition process, MRI is highly sensitive to bulk motion. Moreover, the hardware and procedures involved in an MRI examination are an important cause for patient discomfort, and contribute sensibly to motion effects. Many overview papers discuss the impact of motion artifacts on MRI, here we will chiefly refer to [22, 53].

A. Motion corruption in MRI

Subject motion produces artifacts in the reconstructed images, which may be challenging for a correct radiologic interpretation and with important repercussions for the patient health. Depending on the severity of the artifacts, further scans may be needed, leading to considerable waste of time and financial resources for the provider, and extend the time and discomfort of the MR exam. It is estimated that almost 20% of MR examinations need scan repetition owing to motion artifacts, with potential hospital revenue forgone of approximately $115.000 per scanner per year, as per [2].

Preventive techniques for motion reduction, such as physical restraints or sedation, are sometime impractical. A different strategy altogether is to directly estimate the motion, and compensate for it in the reconstruction phase. Motion estimation can be aided by periodically acquiring additional data, known as MR navigators, that encode the spatial position of the object [18, 51] at the cost of longer acquisition or processing times. Alternatively, "self-navigating" sequences can be designed from which the position of the object can be determined directly from the data (often based on over-sampling the low-frequency region of the k-space, as in [48, 12, 52]). Motion correction can be applied prospectively [36], which means that the position of the object at a given time is used to dynamically adjust the scan parameters (such as the radiofrequency pulse or frequency and phase encoding) in order to keep the acquisition consistent throughout the MR session. In this paper, we will focus on retrospective motion correction, that is we seek to correct the data in a post-processing step, once the acquisition phase have been completed.

B. Prior art on retrospective motion correction

The technique proposed in the present work belongs to the class of retrospective motion correction schemes, where the removal of motion effects is applied entirely in the reconstruction phase. Retrospective techniques can be broadly characterized based on the level of prior information about the motion (direct of indirect) during the scan, ranging from exact knowledge to ignorance thereof ("blind" correction). Typically, the lack of information about the motion must be supplanted by reconstruction regularization priors, which assess the level of motion corruption with carefully designed cost functions. In a mixed approach, both priors can be exploited. Blind retrospective techniques are compelling in that they do not require additional hardware, scanner modifications, or markers. However, they typically rely on computationally expensive and hence time-consuming iterative reconstruction schemes, which is the primary obstacle to their integration

in the clinical practice. Despite the focus on rigid motion in this paper, it is important mentioning that one distinct advantage of retrospective correction over prospective methods is the freedom in choosing the type of motion considered (see, for example, [9, 34]). In [42], for example, non-rigid motion parameters and reconstruction volumes are jointly estimated in order to remove free-breathing motion artifacts. Even though this uses a more general motion model than what considered here, our proposed method (discussed in the next sections) can be applied there as well. We note, however, that generality comes at the cost of estimating the motion parameter unknowns on a higher dimensional space.

Examples of retrospective techniques for rigid motion based on extra navigator echoes can be found in [17, 39, 12, 49]. These methods rely on elementary properties of translations and rotations under the Fourier transform. Markers can further improve the estimation of rotational motion [27].

Blind retrospective correction schemes (also referred to as auto-focusing methods, [22]) attempt at estimating motion directly from the data by enforcing some degree of desired regularity on the reconstructed image. This class of methods are typically setup as a non-linear optimization problem, where the unknowns that describe the motion and the reconstruction image are jointly optimized. One of the earliest solution for rigid motion based on entropy regularization is due to [6, 5]. Other works that use entropy or gradient-entropy regularization can be found in [37, 31, 33]. Sparsity constraints are also a traditional choice in MRI, see for example [40]. The latter type of regularization is perhaps more widespread in the closely related field of dynamic MRI [44, 43]. While the goal of dynamic MRI is an accurate estimation of the motion evolution in time and retrospective motion correction rather aims at motion artifact removal, they are formally equivalent. Due to its objective, dynamic MRI is characterized by a non-standard sampling pattern and a less severely undersampled problem per time frame, and the techniques employed reflect this underlying assumption. Other hybrid approaches include prior information of various nature for further improvement, e.g. navigator data, as in [38, 14], or reference images, as in [4] or [32], as a calibration device for motion estimation. A more recent work by [16] employs a bi-level optimization scheme for rigid motion correction. It is characterized by a motion-resilient acquisition scheme based on distributed (e.g., spanning low- and high-frequency regions of the k-space for each shot) and pseudo-incoherent sampling order (DISORDER).

Lately, many data-driven approaches have been proposed for retrospective motion correction [46, 28, 25, 30, 21]. The work by [29] is closer in spirit to our contribution, as it leverages multi-contrast information for motion artifact removal. Assuming that enough data is available, retrospective techniques based on deep learning may potentially outperform the existing model-based methods, included the one presented in this work. However, hybrid strategies have been recently advanced that combine model-based algorithms with deep learning [1], in order to enhance robustness with respect to generalization and computational feasibility (hence, achieving the best of both worlds). We foresee that our work could also inform an analogous strategy.

C. Contributions

In this paper, we aim at retrospective methods. We will not consider special acquisition schemes designed for motion resilience (as DISORDER, previously discussed), since most clinical exams are still based on Cartesian acquisition. In Cartesian sampling, the acquisition follows standard filling patterns such as linear or low-to-high frequency ordering.

Our work consists in the combination of blind retrospective correction techniques with structural total-variation (TV) regularization. The abstract mathematical framework results in the joint optimization of the rigid motion parameters and reconstructed image, and follows the quite classical setup of, e.g., [33]. The major novelty of our approach is the adoption of the structure-guided TV regularization within the context of blind retrospective correction, which was originally proposed by [19] to tackle under-sampling for classical MRI (see also [13, 20]). Other versions of TV can also accommodate structural information [26]. This type of regularization is based on a weighted TV term that forces the gradient of the reconstructed image to align with the gradient of a reference (normalized) picture. The reference image ideally consists of a contrast devoid of motion artifacts, and possibly unregistered with respect to the target contrast. We also note that the reference image can be acquired with any arbitrary sequence, and interpolated on the grid of the target contrast. We assume that, in a multi-contrast setting, motion artifacts do not affect all the contrasts equally badly and well-resolved contrasts are available within the same MRI session for at least one contrast acquisition. The use of a reference for regularization is close in spirit to the proposal of [4], but very different in the way we exploit it (namely the reference and the target are not necessarily related to the same contrast, and the reference is used implicitly as a regularization term in an optimization setting). A novel aspect of our approach, compared to the existing literature, is resorting to a non-smooth variable projection scheme as optimization algorithm, recently proposed in [50]. Variable projection schemes results in better-conditioned problems and often benefit from faster convergence [23]. Lastly, as advocated in [47], we implement the TV regularization as a constraint rather than an additive penalty term. The goal is to simplify hyper-parameter tuning, which is typically complicated by non-trivial cross-dependency, and provide a practical workflow. A schematic overview of the method is depicted in Figure 1.

II. Method

In this section, we lay out the mathematical framework of joint motion correction and reconstruction with structural priors. We first present the algebraic formalism describing how the Fourier transform behaves under time-dependent rigid motions. We formulate a non-linear variational problem aimed at the joint optimization of motion parameters and the reconstructed image, under constraints that take into account prior knowledge obtained from well-resolved contrasts. The basic assumption, here, is that all contrasts are structurally similar

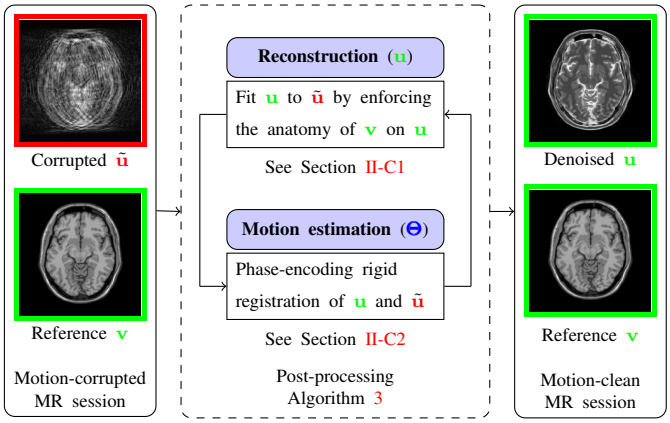


Fig. 1: Schematic overview of the joint retrospective method based on motion-free reference contrasts. After a multi-contrast MR session, a trained radiologist or technician select a contrast affected by motion artifacts (red) and a reference contrast not corrupted by motion (green). The corrupted contrast undergoes further post-processing, and, with the help of the reference, the motion artifacts may be removed. The technical description of this post-processing step is in Section II.

under a rigid transformation, e.g. the associated gradient vector fields are parallel to each other. We finally discuss sensible optimization strategies to tackle the problem.

A. Data model under rigid motion transformations

We start by describing the data corruption model caused by motion, and introduce the related notation. A target contrast image is represented by a square-integrable complex-valued function $u \in \mathbb{L}^2(\mathbb{R}^d;\mathbb{C})$, where $d = 2$ in 2D or $d = 3$ in 3D. The contrast u will undergo a time-dependent distortion u^t due to motion, here formalized as

$$u^t = u \circ (T^t)^{-1}, \tag{1}$$

by means of a differentiable coordinate transformation $T^t : \mathbb{R}^d \to \mathbb{R}^d$ (the symbol "$\circ$" indicates composition of functions). The contrasts u^t are labeled by a time-related index t. Under the rigid motion assumption, each transformation can be summarized by translation and rotation parameters encapsulated in a small vector $\boldsymbol{\theta}^t \in \mathbb{R}^{n_p}$, where $n_p = 3$ in 2D and $n_p = 6$ in 3D. We explicitly denote this dependency with $T^t = T(\boldsymbol{\theta}^t)$. For each t, we distinguish translation $\boldsymbol{\tau}^t$ and rotation $\boldsymbol{\varphi}^t$ parameters such that the related rigid transformation is given by:

$$T(\boldsymbol{\theta}^t)(\mathbf{x}) = R(\boldsymbol{\varphi}^t)\mathbf{x} + \boldsymbol{\tau}^t, \quad \boldsymbol{\theta}^t = (\boldsymbol{\tau}^t, \boldsymbol{\varphi}^t), \tag{2}$$

where R is a $d \times d$ matrix that corresponds to the rotation about the angles $\boldsymbol{\varphi}^t$ (with respect to the origin).

The data acquisition process in MRI is mathematically equivalent to evaluating the Fourier transform $\mathcal{F}$ for some selected wave-number $\mathbf{k}$. We assume, now, that u^t is sampled at a particular subset of the k-space,

$$K^t = \{\mathbf{k}_1^t, \dots, \mathbf{k}_{n_k^t}^t\} \subset \mathbb{R}^d, \tag{3}$$

of cardinality $|K^t| = n_k^t$. The data at time t is then indicated by:

$$\mathbf{d}^t = \mathcal{S}^t \mathcal{F}[u^t], \tag{4}$$

where

$$\mathcal{S}^t \hat{u} = (\hat{u}(\mathbf{k}_1^t), \dots, \hat{u}(\mathbf{k}_{n_k^t}^t)) \tag{5}$$

is the (unbounded) sampling operator $\mathcal{S}^t : \mathbb{L}^2(\mathbb{R}^d;\mathbb{C}) \to \mathbb{C}^{n_k^t}$, whose domain is the set of continuous functions. Implicitly, we are assuming that no intra-set motion distortion occurs while measuring the data associated to K^t, for each t. For example, it is natural to consider K^t as the set of wave-numbers acquired during a readout line for Cartesian or radial acquisition schemes, which happens in few milliseconds. We now briefly review how the Fourier transform behave under rigid motions, that is:

$$\mathcal{F}[u_t](\mathbf{k}) = \exp(-\mathrm{i}\,\mathbf{k}\cdot\boldsymbol{\tau}^t)\mathcal{F}[u](R(\boldsymbol{\varphi}^t)^{\mathrm{T}}\mathbf{k}). \tag{6}$$

The symbol $^{\mathrm{T}}$ denotes transposition. In other words, translations in the x-space correspond to phase shifts in the k-space and rotations in the x-space (with respect to the origin) correspond to rotations in the k-space (with respect to the origin). Note that, in principle, (global) affine transformations also can be easily analyzed under the Fourier transform and accommodated in our model.

By means of a Cartesian grid $\{\mathbf{x}_1, \dots, \mathbf{x}_{n_x}\}$ spanning the rectangular field of view $\Omega \subset \mathbb{R}^d$, the target contrast u is discretized and represented by the vector $\mathbf{u} = (u_1, \dots, u_{n_x}) \in \mathbb{C}^{n_x}$, with components $u_j \approx u(\mathbf{x}_j)$. The Fourier transform can be efficiently approximated by the nonuniform discrete Fourier transform (NUDFT) discussed in [8, 7], which is readily available as a software library interface or natively for many programming languages[1]. Note that the NUDFT transform was made necessary by the rotation of the k-space in equation (6). We can now define the motion-induced data model:

$$\tilde{\mathbf{d}}^t = \mathcal{S}^t \tilde{\mathcal{F}}[\mathbf{u};\boldsymbol{\theta}^t], \tag{7}$$

where $\tilde{\mathcal{F}}$ is the perturbed Fourier transform:

$$\tilde{\mathcal{F}}[\mathbf{u};\boldsymbol{\theta}](\mathbf{k}) = \exp(-\mathrm{i}\,\mathbf{k}\cdot\boldsymbol{\tau})\mathcal{F}[u](R(\boldsymbol{\varphi})^{\mathrm{T}}\mathbf{k}), \quad \boldsymbol{\theta} = (\boldsymbol{\tau}, \boldsymbol{\varphi}). \tag{8}$$

Note that the NUDFT is an analytical function, and the selection operator $\mathcal{S}^t$ is then well-defined in equation (7).

For ease of exposition in the remainder of this paper, we stack all the equations (7), labeled by t, into a single expression. We obtain the aggregated data model

$$\tilde{\mathbf{D}} = \mathcal{S}\tilde{\mathcal{F}}[\mathbf{u};\boldsymbol{\Theta}], \tag{9}$$

where the rigid motion parameters are summarized into the unknown

$$\boldsymbol{\Theta} = (\boldsymbol{\theta}^1, \dots, \boldsymbol{\theta}^{n_t}). \tag{10}$$

The unknown $\boldsymbol{\Theta}$ is an array in $\mathbb{R}^{n_p \times n_t}$, where n_t denotes the number of time samples considered. The vector-valued perturbed Fourier transform $\tilde{\mathcal{F}}[\mathbf{u};\boldsymbol{\Theta}] \in \mathbb{L}^2(\mathbb{R}^d;\mathbb{C}^{n_t})$ is defined by

$$\tilde{\mathcal{F}}[\mathbf{u};\boldsymbol{\Theta}](\mathbf{k}) = (\tilde{\mathcal{F}}[\mathbf{u};\boldsymbol{\theta}^1](\mathbf{k}), \dots, \tilde{\mathcal{F}}[\mathbf{u};\boldsymbol{\theta}^{n_t}](\mathbf{k})), \tag{11}$$

[1] FINUFFT, https://finufft.readthedocs.io/

while the overall selection operator $\mathcal{S}$ is given by

$$\mathcal{S}(\hat{u}^1, \dots, \hat{u}^{n_t}) = (\mathcal{S}^1\hat{u}^1, \dots, \mathcal{S}^{n_t}\hat{u}^{n_t}). \tag{12}$$

B. The joint motion parameter estimation and reconstruction problem

Given the data perturbation model discussed in the previous section, we can now establish the optimization framework for jointly recovering the motion parameters $\boldsymbol{\Theta}$ and the reconstructed image $\mathbf{u}$. We aim at the following minimization problem:

$$\min_{\mathbf{u},\boldsymbol{\Theta}} \mathcal{J}(\mathbf{u}, \boldsymbol{\Theta}) = \mathcal{L}(\mathcal{S}\tilde{\mathcal{F}}[\mathbf{u}; \boldsymbol{\Theta}] - \mathbf{D}) + \lambda\mathcal{R}_u(\mathbf{u}) + \mu\mathcal{R}_\theta(\boldsymbol{\Theta}), \tag{13}$$

with given data $\mathbf{D}$, where λ and μ are appropriate weighting parameters. In this formulation, we are distinguishing a data misfit term, measured by a loss function $\mathcal{L}$, and additive regularization terms for the reconstruction unknowns, $\mathcal{R}_u$, and the rigid motion parameters, $\mathcal{R}_\theta$. The loss and regularization functions are assumed to be convex. The overall functional $\mathcal{J}$ is then also convex in $\mathbf{u}$, allowing efficient convex optimization tools. However, $\mathcal{J}$ is non-convex in $\boldsymbol{\Theta}$ and strongly multi-modal. The regularization strategy for $\mathcal{R}_\theta$ will generally aim to enforce smoothness of the motion parameters as a function of time, when appropriate. Note that the temporal evolution is defined by the phase-encoding ordering of the acquisition sequence, as described in Section II-A. We refer to Appendix A for some important considerations related to motion parameter regularization and phase wrapping.

1) The data misfit loss function: A conventional choice for the loss function is the least-squares norm:

$$\mathcal{L}(\mathbf{R}) = \frac{1}{2}\|\mathbf{R}\|_{2,2}^2 = \frac{1}{2}\sum_{k,t}|R_k^t|^2, \tag{14}$$

for a given array $\mathbf{R} \in \mathbb{C}^{n_k^1} \times \dots \times \mathbb{C}^{n_k^{n_t}}$. Other sensible options may involve mixed norms, such as $\|\mathbf{R}\|_{1,2} = \sum_t(\sum_k |R_k^t|^2)^{1/2}$, which is particularly compelling for outlier robustness when the motion causes signal loss due to spin de-phasing or undesired magnetization evolution for few time samples t [53]. Note that this noise mechanism cannot be accounted for in the data model in equation (7), and the mixed-norm loss ensures that the optimization will not be impaired by few out-of-distribution data samples [3]. In this paper, however, we will exclusively deal with synthetic examples devoid of this sort of problems. Additional complications involve the non-differentiability of $\|\cdot\|_{1,2}$.

2) Regularization for image reconstruction: Regularization for the reconstruction unknowns $\mathbf{u}$ is fundamental for avoiding data overfit. For example, when the wave-number samples are distinct and $n_x = \sum_t n_k^t$, the data model operator $\mathcal{S}\mathcal{F}[\,\cdot\,; \boldsymbol{\Theta}]$ in equation (9) is invertible and a trivial solution for the joint-recovery problem in equation (13) is $\mathbf{u} = (\mathcal{S}\mathcal{F}[\,\cdot\,; \boldsymbol{\Theta}])^{-1}\mathbf{D}$, with $\boldsymbol{\Theta} = 0$, when no regularization term $\mathcal{R}_u$ is applied to $\mathbf{u}$ and for any regularization term $\mathcal{R}_\theta$ that admits $\boldsymbol{\Theta} = 0$ as a global minimum. However, this result is undesirable since it will be heavily affected by motion distortions.

A classical regularization strategy for MRI is related to compressed sensing [35]. Compressed sensing aims at fast imaging by acquiring data corresponding to fewer k-space wave-numbers, compared to conventional Cartesian acquisition, but randomly located to avoid coherent noise such as aliasing. This under-determined problem is regularized by enforcing sparsity for $\mathbf{u}$ in some domain (e.g. via the wavelet transform or the gradient operator). Note that the problem of motion-corrected reconstruction here considered does not necessarily suffer from under-sampling (as the sampling is nominally dense), however some effective under-sampling might be induced by rotational motion, since spatial rotations correspond to rotation in k-space (see equation 8). For the same reasons, the motion may induce effective randomized acquisition even when we employ a regular Cartesian sampling.

In this work, we will focus on TV regularization:

$$\mathcal{R}_u(\mathbf{u}) = \|\nabla\mathbf{u}\|_{1,2} = \sum_i(\sum_l |\partial_l\mathbf{u}_i|^2)^{1/2}, \tag{15}$$

where the linear operator $\nabla : \mathbb{C}^{n_x} \to \mathbb{C}^{n_x \times d}$ stems from the discretization of the continuous gradient operator, which collects the partial derivatives ∂_l of $\mathbf{u}$ at each grid point $\mathbf{x}_i$. Weighted TV schemes were analyzed in [19] in the context of multi-contrast imaging. In particular, they proposed a structure-guided TV regularization defined by

$$\mathcal{R}_u(\mathbf{u}; \mathbf{v}) = \|P(\mathbf{v})\nabla\mathbf{u}\|_{1,2}, \quad P(\mathbf{v}) = I - \xi(\mathbf{v})\xi(\mathbf{v})^{\mathrm{H}}, \tag{16}$$

where I is the identity and $P(\mathbf{v})$ is the projection operator on the linear space that is orthogonal to the vector $\xi(\mathbf{v})$. The symbol $^{\mathrm{H}}$ indicates the adjoint operation. The vector $\xi(\mathbf{v})$ corresponds to the normalized gradient of a pre-selected contrast $\mathbf{v}$, e.g. $\xi(\mathbf{v}) \approx \nabla\mathbf{v}/|\nabla\mathbf{v}|$. To avoid singularities, $\xi(\mathbf{v})$ is rather defined as

$$\xi(\mathbf{v})_i = \frac{\nabla\mathbf{v}_i}{\sqrt{|\nabla\mathbf{v}_i|^2 + \eta^2}}, \tag{17}$$

where $\xi(\mathbf{v}) \in \mathbb{C}^{n_x \times d}$, for some constant $\eta > 0$. The effect of the regularization term in equation (16) is to promote a contrast $\mathbf{u}$ whose gradient aligns with the gradient of $\mathbf{v}$, hence enforcing the "structure" of $\mathbf{v}$ onto $\mathbf{u}$. Other types of weighted regularization were also compared in [19], but the choice in equation (16) was demonstrated to perform better.

Whenever a prior contrast $\mathbf{v}$ is available, the structural regularization in equation (17) defines a better-constrained problem than conventional TV in equation (15). Indeed, the joint reconstruction problem in equation (13) with conventional TV regularization — or any rigid-motion invariant regularization, for that matter — is fundamentally ambiguous, since any two solutions that differ by a global rigid transformation are equivalent (at least when no regularization for the motion parameters is applied). Therefore, a more rigorous definition for a solution of the problem (13) is by means of equivalence classes modulo a global rigid-motion registration. Structural TV, on the contrary, is not invariant for rigid motion transformations, and implicitly defines an absolute position for the target contrast based on the structural prior. We remark that $\mathbf{v}$ does not necessarily need to be pre-registered with the target contrast $\mathbf{u}$, since the optimization of the motion

parameters will ideally compensate for the prior and target contrast misalignment.

In this paper, we apply the structural regularization in equation (16) as a constraint rather than a penalty term. That means that we replace the variational problem in equation (13) with the constrained formulation

$$\min_{\mathbf{u}\in C(\mathbf{v}),\boldsymbol{\Theta}} \mathcal{J}(\mathbf{u},\boldsymbol{\Theta}) = \mathcal{L}(\mathcal{S}\tilde{\mathcal{F}}[\mathbf{u};\boldsymbol{\Theta}] - \mathbf{D}) + \mu\mathcal{R}_\theta(\boldsymbol{\Theta}), \quad (18)$$

where the constraint set is

$$C(\mathbf{v}) = \{\mathbf{u} : \|P(\mathbf{v})\nabla\mathbf{u}\|_{1,2} \leq \varepsilon\}, \quad (19)$$

for a prescribed noise level ε. Generally speaking, penalty and constrained formulations are equivalent for convex objectives in the following sense: given a noise level ε, there exists a unique weighting parameter λ such that the constrained and penalty formulations admit the same solution (and vice versa, [11]). We note, however, that the problem here considered is not convex. One important practical advantage of the constrained formulation (18) over the penalty formulation (17) is a more straightforward hyper-parameter choice. In [47], for example, it is argued that the effect of multiple hyper-parameters on the solution is highly non-trivial due to cross-talk, and the tuning phase results in a lengthy process. A constraint like in equation (18), instead, forces the solution to lie on a set with predefined regularity at each step of a chosen iterative method. The choice of the hyper-parameters involved in the constraint set, like ε, can be easily determined given the knowledge of the true solution of the problem, or calibrated based on a surrogate solution. Simply put, given a reconstruction $\mathbf{u}^*$ ideally close to the solution, one can set $\varepsilon = \|P(\mathbf{v})\nabla\mathbf{u}^*\|_{1,2}$ or a fraction thereof. Notably, this choice is independent of all the other hyper-parameters that appear in the problem (13) (such as the weighting μ for the regularization for the rigid motion unknowns).

C. Optimization strategy

We now discuss the optimization strategy to solve the joint motion estimation and reconstruction problem in equation (13). In the literature, it is common to employ an alternating minimization scheme, where the reconstruction and motion unknowns $\mathbf{u}$ and $\boldsymbol{\Theta}$ are updated in a leap-frog fashion. Here, we resort to a variable projection scheme, as discussed in the review of [23]. In other words, we introduce the reduced objective

$$\bar{\mathcal{J}}(\boldsymbol{\Theta}) = \min_{\mathbf{u}} \mathcal{J}(\mathbf{u},\boldsymbol{\Theta}), \quad (20)$$

function only of motion parameters $\boldsymbol{\Theta}$, where the inner variable $\mathbf{u}$ has been implicitly "projected-out" of the problem. Variable projection schemes are typically superior to joint optimization in terms of conditioning, i.e. the convergence to the solution is faster, but they usually hinge on closed-form expression for the inner variable $\mathbf{u}$. In practice, since this is not attainable due to computational costs, we ran a fixed number of iteration as a stopping criterion. This falls under the category of inexact non-smooth variable projection methods, as discussed in [50].

In the following, we briefly discuss how the inner problem in equation (20) is iteratively solved (Section II-C1), and which optimization algorithm is employed for $\bar{\mathcal{J}}$ (Section II-C2). All the computational step involved in the solution of the problem are detailed in Algorithm 3.

1) Solution of the reconstruction unknowns: In this section, we present a solution scheme for solving the minimization problem in equation (20). We start by remarking that the functional $\mathcal{J}$ (equation 13) is convex in $\mathbf{u}$ for any fixed $\boldsymbol{\Theta}$, so that fast convex optimization methods can be employed. We resort, in particular, to the general formulation of the fast iterative shrinkage threshold algorithm (FISTA, [10]), which is summarized in the reconstruction Algorithm 1. For practical

Algorithm 1 FISTA for image reconstruction

Input: $\mathbf{D}, \boldsymbol{\Theta}, \mathbf{u}_0, N, \alpha$ ▷ Data, motion par., init., iters, step-length
Output: $\mathbf{u}$
  $\mathbf{u} \leftarrow \mathbf{u}_0$
  for $n = 1, \dots, N$ **do**
    $\mathbf{g}_u \leftarrow \nabla_{\mathbf{u}}\mathcal{L}(\mathcal{S}\tilde{\mathcal{F}}[\mathbf{u};\boldsymbol{\Theta}] - \mathbf{D})$ ▷ Gradient computation
    $\mathbf{u} \leftarrow \text{proxy}_{\alpha\mathcal{R}_u}(\mathbf{u} - \alpha\,\mathbf{g}_u)$ ▷ See Algorithm 2
  end for

purposes, the reconstruction process will be inexact, which means that we control the solution accuracy indirectly by fixing the number of iterations N and updating the initial estimate $\mathbf{u}_0$ at each pass of Algorithm 1. The step length α is, ideally, the inverse of the Lipschitz constant of the objective $\mathbf{u} \mapsto \mathcal{L}(\mathcal{S}\tilde{\mathcal{F}}[\mathbf{u};\boldsymbol{\Theta}] - \mathbf{D})$. When $\mathcal{L}$ is the least-squares norm, α is the maximum singular value of $\mathcal{S}\tilde{\mathcal{F}}[\,\cdot\,;\boldsymbol{\Theta}]$ and can be estimated, for instance, by the power method.

An important detail of Algorithm 1 is the computation of the proximal operator of the regularization term $\mathcal{R}_u$:

$$\text{proxy}_{\mathcal{R}_u}(\mathbf{w}) = \arg\min_{\mathbf{u}} \frac{1}{2}\|\mathbf{u} - \mathbf{w}\|_2^2 + \mathcal{R}_u(\mathbf{u}).$$

For a general introduction on proximal operators, we refer to [45]. When $\mathcal{R}_u$ is the penalty TV regularization in equation (16), its proximal operator can be computed according to [19]. For the constrained formulation in equation (18), we exploit the saddle-point problem stemming from the dualization of $\mathcal{R}_u$, e.g.:

$$\max_{\mathbf{p}} \min_{\mathbf{u}} \frac{1}{2}\|\mathbf{u} - \mathbf{w}\|_2^2 + P(\mathbf{v})\nabla\mathbf{u}\cdot\mathbf{p} - \varepsilon\|\mathbf{p}\|_{\infty,2}\,.$$

The dual variable $\mathbf{p} \in \mathbb{C}^{n_x\times d}$ is a vector field with d components for each grid point. The mixed norm is given by $\|\mathbf{p}\|_{\infty,2} = \max_i(\sum_l \mathbf{p}_{il}^2)^{1/2}$. The inner minimization is solved by

$$\mathbf{u} = \mathbf{w} + \text{div}(P(\mathbf{v})^{\mathrm{H}}\mathbf{p}), \quad (21)$$

where $\text{div} = -\nabla^{\mathrm{T}}$ is the divergence operator. Everything boils down to a convex minimization problem in the dual variable $\mathbf{p}$,

$$\min_{\mathbf{p}} \frac{1}{2}\left\|\text{div}(P(\mathbf{v})^{\mathrm{H}}\mathbf{p}) + \mathbf{w}\right\|_2^2 + \varepsilon\|\mathbf{p}\|_{\infty,2}\,, \quad (22)$$

which can be iteratively solved with the FISTA, since the proximal operator for the norm $\|\cdot\|_{\infty,2}$ is known. Indeed, one can prove that

$$\text{proxy}_{\varepsilon\|\cdot\|_{\infty,2}}(\mathbf{p}) = \mathbf{p} - \Pi_{\|\cdot\|_{1,2}\leq\varepsilon}(\mathbf{p}), \quad (23)$$

where Π indicates the projection onto the $\|\cdot\|_{1,2}$-ball. Furthermore,

$$\Pi_{\|\cdot\|_{1,2}\leq\varepsilon}(\mathbf{p}) = \text{proxy}_{\lambda\|\cdot\|_{1,2}}(\mathbf{p}), \tag{24}$$

where λ satisfies

$$||\text{proxy}_{\lambda\|\cdot\|_{1,2}}(\mathbf{p})||_{1,2} = \varepsilon. \tag{25}$$

The threshold λ can be determined via root-finding algorithms. The resulting algorithm is collected in Algorithm 2. Note that the step-length can be chosen as $\beta \leq 1/8$.

Algorithm 2 Proximal operator for structural-TV projection

Input: $\mathbf{v}, \varepsilon, \mathbf{w}, N, \beta$ ▷ Prior, threshold, data, iters, step-length
Output: $\mathbf{u}$
  $\mathbf{p} \leftarrow 0$
  for $n = 1, \dots, N$ **do**
    $\mathbf{u} \leftarrow \mathbf{w} + \text{div}(P(\mathbf{v})^{\mathrm{T}}\mathbf{p})$ ▷ Primal variable
    $\mathbf{g}_p \leftarrow P(\mathbf{v})\nabla\mathbf{u}$ ▷ Gradient computation
    $\mathbf{p} \leftarrow \text{proxy}_{\beta\|\cdot\|_{\infty,2}}(\mathbf{p} + \beta\,\mathbf{g}_p)$ ▷ See equations (21)–(24)
  end for

2) Update of motion parameters: The optimization of the reduced objective in equation (20) is carried out by gradient-based methods. The gradient of $\bar{\mathcal{J}}$, owing to the nature of variable projection, is given by

$$\nabla_{\boldsymbol{\Theta}}\bar{\mathcal{J}} = \nabla_{\boldsymbol{\Theta}}\mathcal{J}(\bar{\mathbf{u}}, \boldsymbol{\Theta}), \tag{26}$$

where $\bar{\mathbf{u}} = \bar{\mathbf{u}}(\boldsymbol{\Theta})$ has been obtained from the reconstruction algorithm depicted in the previous section (see Algorithm 1). In other words, one can treat the variable $\mathbf{u}$ as fixed in the gradient calculation of $\bar{\mathcal{J}}$. We observe that since $\bar{\mathbf{u}}$ is the solution of a non-smooth problem, the method belongs to the unconventional class of non-smooth variable projection. The rigorous justification for (26) can be found in [50].

The update rule for the motion parameters $\boldsymbol{\Theta}$ will simply follow quasi-Newton methods [41]:

$$\boldsymbol{\Theta} \leftarrow \boldsymbol{\Theta} - \gamma H^{-1}\nabla_{\boldsymbol{\Theta}}\bar{\mathcal{J}}, \tag{27}$$

where H is the Hessian of the objective $\boldsymbol{\Theta} \mapsto \mathcal{J}(\bar{\mathbf{u}}, \boldsymbol{\Theta})$, with fixed $\bar{\mathbf{u}}$. We highlight the fact that the reduced objective $\bar{\mathcal{J}}$ is smooth but not necessarily twice differentiable due to the regularization term $\mathcal{R}_u$, so that its Hessian is ill-defined. However, the joint objective $\mathcal{J}$ is twice differentiable in $\boldsymbol{\Theta}$, hence we will use its Hessian merely as a pre-conditioner to accelerate the gradient descent. In our experimentation, the step-length is conveniently set at $\gamma = 1$, with no need for line searches. The overall joint optimization strategy is schematically reported in Algorithm 3. A block diagram of the method is reported in Figure 1.

3) Convergence of the joint scheme: The optimization strategy here delineated boils down to an alternating update of image reconstruction and motion-parameter unknowns, according to the inexact non-smooth variable projection method described in [50].

To the best of our knowledge, little is known for the convergence of alternating schemes to a global minimum for non-convex optimization problems as the one considered here,

Algorithm 3 Joint motion estimation and reconstruction

Input: $\mathbf{D}, \boldsymbol{\Theta}_0, N_u, N_\theta, \gamma$ ▷ Data, init., iters, step-length
Output: $\mathbf{u}, \boldsymbol{\Theta}$
  $\mathbf{u} \leftarrow 0$
  $\boldsymbol{\Theta} \leftarrow \boldsymbol{\Theta}_0$
  for $n = 1, \dots, N_\theta$ **do**
    Estimate α_u ▷ See Section II-C1
    $\mathbf{u} \leftarrow \text{image\_recon}(\mathbf{D}, \boldsymbol{\Theta}, \mathbf{u}, N_u, \alpha_u)$ ▷ See Algorithm 1
    $\mathbf{g}_\theta \leftarrow \nabla_{\boldsymbol{\Theta}}\{\boldsymbol{\Theta} \mapsto \mathcal{J}(\mathbf{u}, \boldsymbol{\Theta})\}$ ▷ Gradient computation
    $H_\theta \leftarrow \nabla^2_{\boldsymbol{\Theta}}\{\boldsymbol{\Theta} \mapsto \mathcal{J}(\mathbf{u}, \boldsymbol{\Theta})\}$ ▷ pseudo-Hessian computation
    $\boldsymbol{\Theta} \leftarrow \boldsymbol{\Theta} - \gamma H_\theta^{-1}\mathbf{g}_\theta$
  end for

and local minimum stagnation might occur. Note that many global minima exist for the problem in question, as discussed in the Appendix A (see equation 30), but these are entirely equivalent for our purposes since we are ultimately interested only in the motion-free reconstruction.

In order to avoid sub-optimal local minima, we advocate for a multiscale strategy, which recursively exploits the solution of the problem reformulated on a coarser grid and then refined on the original grid. This is a standard procedure in image registration, for example. Alternatively, one may solve the problem for small values of the regularization parameter ε in equation (19) and gradually relax this choice by continuation methods.

Regarding the local convergence of the scheme, some formal guarantees are discussed in [50], Theorem 2.5, in the context of non-smooth variable projection. The theorem states that some degree of accuracy must be achieved for the solution of the inner problem in equation (20) to achieve convergence. However, in the Algorithm 3, we adopt a more practical approach based on a fixed number of iterations for the solution of the inner problem (Algorithm 1), which was effective for all the numerical experiments presented here.

III. Numerical Experiments

In this section, we compare 2D reconstruction results obtained with different methods, namely:

- *conventional*: $\bar{\mathbf{u}} = (\mathcal{S}\mathcal{F}[\cdot\,;\boldsymbol{\Theta}])^{\mathrm{H}}\mathbf{D}$ with fixed motion parameters $\boldsymbol{\Theta} = 0$, which corresponds to a basic inverse Fourier transform;
- *TV*: $\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \mathcal{J}(\mathbf{u}, \boldsymbol{\Theta})$ in equation (13) with fixed motion parameters $\boldsymbol{\Theta} = 0$ and standard TV regularization;
- *sTV*: $\bar{\mathbf{u}} = \arg\min_{\mathbf{u}} \mathcal{J}(\mathbf{u}, \boldsymbol{\Theta})$ in equation (13) with fixed motion parameters $\boldsymbol{\Theta} = 0$ and structure-guided TV regularization (Section II-B2);
- *joint-TV*: $\bar{\mathbf{u}}, \bar{\boldsymbol{\Theta}} = \arg\min_{\mathbf{u},\boldsymbol{\Theta}} \mathcal{J}(\mathbf{u}, \boldsymbol{\Theta})$ in equation (13) with standard TV regularization;
- *joint-sTV*: $\bar{\mathbf{u}}, \bar{\boldsymbol{\Theta}} = \arg\min_{\mathbf{u},\boldsymbol{\Theta}} \mathcal{J}(\mathbf{u}, \boldsymbol{\Theta})$ in equation (13) with structure-guided TV regularization (Section II-B2).

We analyze the performance with respect to synthetic (Section III-A) and in-vivo contrasts (Section III-B). The methods are tested on different simulated motion types, that is:

- sudden motion, where the phantom assumes two distinct positions during the measurements (the change in position happens roughly halfway through the scan, see Figures 11 or 14 in Appendix B for a depiction of the true motion);
- periodic motion: the phantom is continuously moving in periodic fashion (Figures 12 or 15 in Appendix B, the corresponding period is a fifth of the total scan time);
- random smooth motion obtained from a Gaussian process with squared exponential covariance function (Figures 13 or 16 in Appendix B, the corresponding covariance function is defined as $\kappa(t,t') = \exp(-|t-t'|^2/(2\sigma^2)$ with $\sigma = 4$ for all the motion parameters. The realizations of the Gaussian process for the translational parameter are multiplied by the factor 2, while the realization for the angle of rotation is scaled by $2\pi/180$).

In this set of experiments, no out-of-plane motion is assumed. The sampling scheme considered here is dense Cartesian, which consists of time-ordered readout segments

$$K^t = \{(k_x^t, k_y) : k_y = j\Delta k, j = -n_y/2, \dots, n_y/2\}.$$

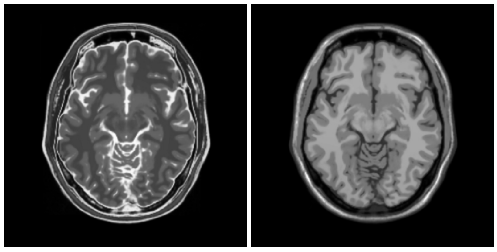
Hence, the readout is performed along the k_y direction, and the phase-encoding along the k_x direction, e.g. $k_x^t = t\Delta k, t = -n_x/2, \dots, n_x/2$, from negative to positive wave-numbers (which is aligned to the horizontal axis of the reported images, or "left-right" in radiological convention). The k-space spacing Δk is chosen accordingly to the field of view. The motion-corrupted data have been obtained synthetically from the data model described in Section II-A, equation (9). To reduce the "inverse crime", we add artificial Gaussian white noise corresponding to SNR (signal-to-noise ratio) equal to 70 dB. Beside a qualitative comparison, we assess the results according to the following performance metrics: PSNR (peak signal-to-noise ratio, measured in dB) and SSIM (structural similarity). Due to the reconstruction ambiguity related to bulk rigid motion, mentioned in Section A and equation (30), we compute these quantities after rigid registration with the ground truth.

The experiments are carried on a laptop with the following processor specification: Intel Core i7-10750H CPU @ 2.60GHz x 12. The method has been implemented in the programming language Julia. Computational times are around 3 minutes per experiment. Note that we are not resorting to GPUs. The code will be publicly available upon acceptance of this manuscript.

A. BrainWeb

This numerical experiment is based on the synthetic phantom BrainWeb[2] [15]. We select T2- and T1-weighted 2D axial slices, respectively, as ground truth and structural prior images (see Figure 2). Note that, for this example, the ground truth and the prior images are perfectly registered. The reconstruction results are collected in Figures 3, 4, and 5, respectively, for sudden, periodic, and random smooth motions. For the joint reconstructions, we also compare the estimated motion parameters in Appendix B, Figures 11, 12, and 13.

[2]https://brainweb.bic.mni.mcgill.ca/



(a) T2-weighted (b) T1-weighted

Fig. 2: BrainWeb contrast phantom: (a) ground-truth, (b) structural prior

B. In-vivo data with simulated motion

In the second experiment, we consider ground truth and prior contrasts selected from in-house in-vivo data (see Figure 6). Note that, for the following tests, we keep the same resolution (224×224, 1mm^2) and field-of-view of the original data. As for the BrainWeb phantom, we choose T2- and T1-weighted contrasts. We aim at assessing the effect of the prior on the reconstruction in the realistic case where it is not registered with the target contrast and was originally acquired on a different field of view before being interpolated on the target contrast grid. More specifically, the structural prior is slightly mis-registered with the ground truth by a bulk rigid motion corresponding to $\tau_x = 0.68$ pixels, $\tau_y = -0.52$ pixels, $\varphi = 0.56$ degrees (according to the model in equation 2). The results are summarized in Figures 7, 8, and 9, for sudden, periodic, and random smooth motions, respectively. The estimated motion parameters for the joint reconstructions are compared in Appendix B, Figures 11, 12, and 13, respectively.

IV. Discussion

The numerical experiments presented in the previous section demonstrate the need for motion estimation schemes to effectively remove the motion imprint from the reconstruction results. Conventional methods based only on regularization of the reconstruction are fundamentally defective when moderate or severe motion occurs (e.g. for non-cooperative patients). Total-variation regularization, even comprising structure-informed weights, do not sensibly improve the results compared to the basic Fourier inversion according to qualitative inspection and performance metrics based on least-squares error or structural similarity.

On the contrary, as widely reported in the literature and our own experimentation, joint motion estimation and reconstruction methods provide a reconstruction quality that is superior by several dB's in PSNR to conventional, TV, or sTV results. The typical ghosting and blurring artifacts caused by motion are greatly reduced or even removed for both synthetic and realistic brain phantoms. These findings are consistent over a wide range of motion types.

Our proposal consists in the combination of joint motion and estimation and reconstruction with structure-guided regularization, originally advanced by [19] for compressed sensing MRI. Motion artifacts do not affect all the contrasts equally badly and well-resolved contrasts are often available

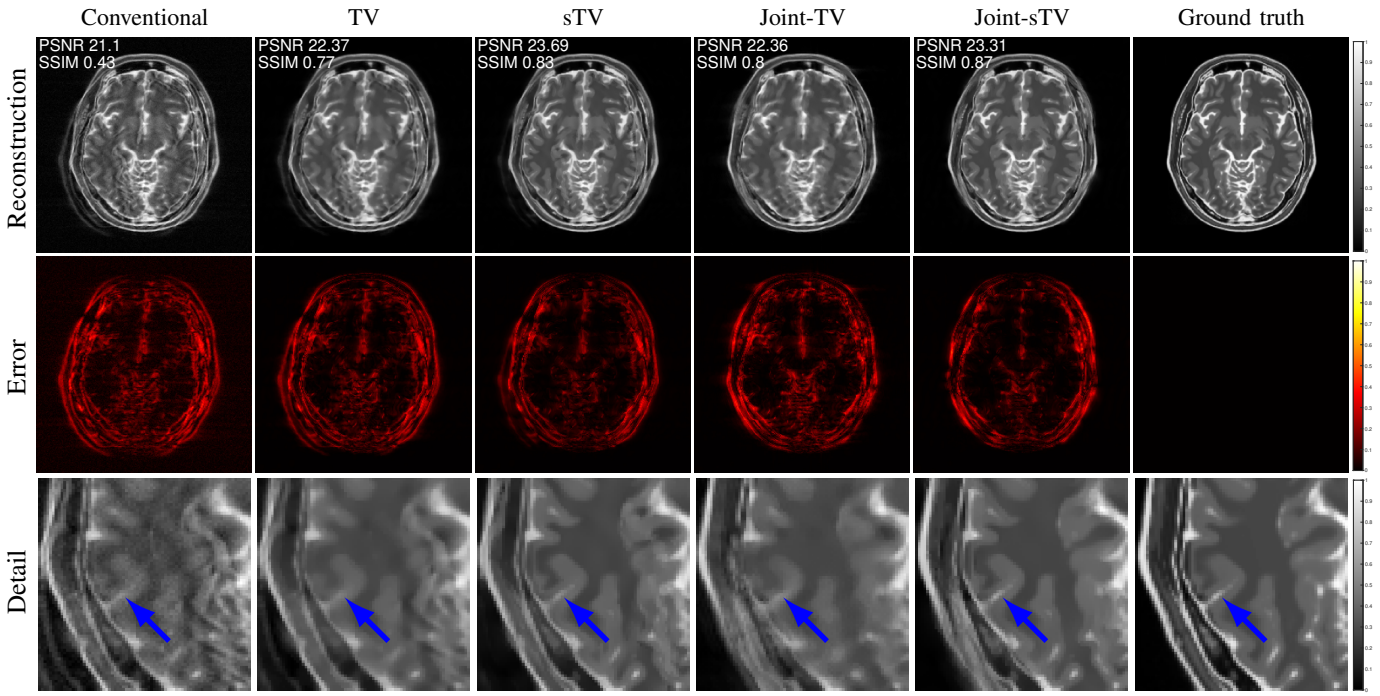


Fig. 3: Reconstruction results for sudden motion (BrainWeb phantom). We compare the motion-corrupted image (conventional) with several reconstruction methods that ignore motion-parameter estimation (TV and sTV) and that estimate motion parameter unknowns (Joint-TV and Joint-sTV). The proposed method is Joint-sTV. TV refers to the use of conventional TV regularization (equation 15), while sTV employs reference-guided TV regularization (equation 16).

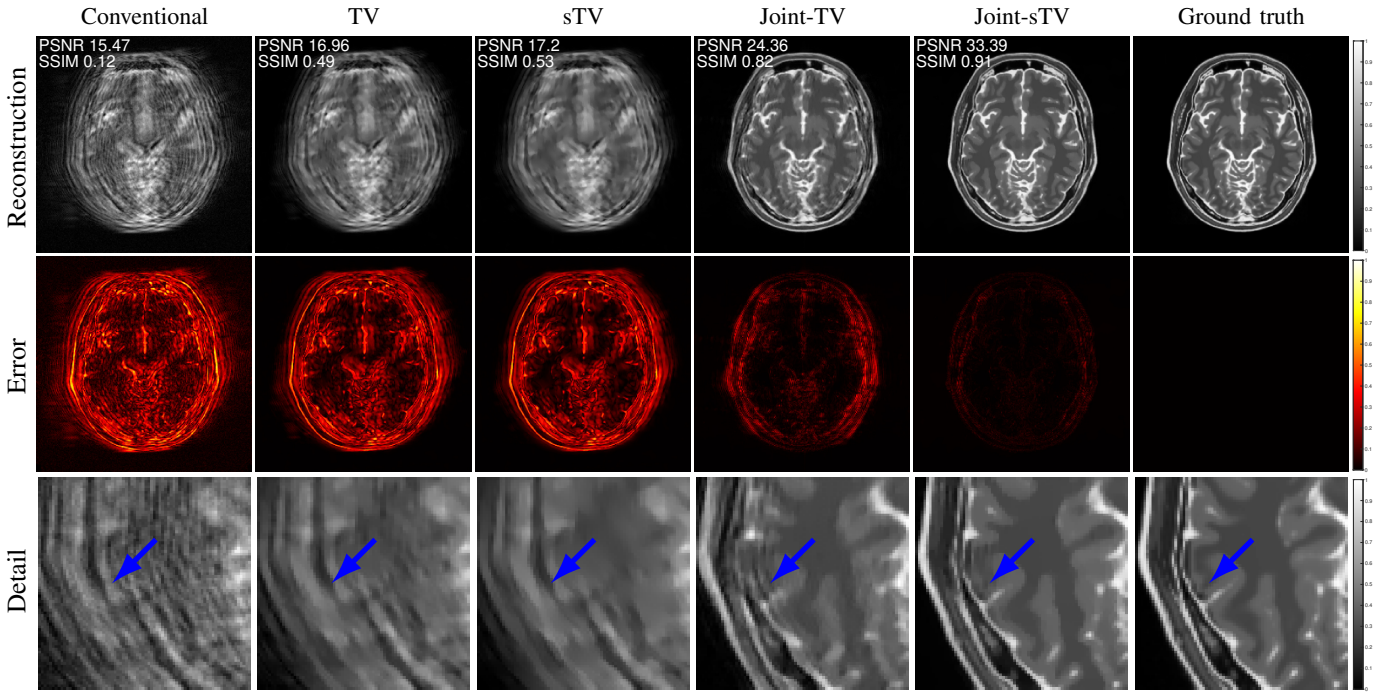


Fig. 4: Reconstruction results for periodic motion (BrainWeb phantom). We compare the motion-corrupted image (conventional) with several reconstruction methods that ignore motion-parameter estimation (TV and sTV) and that estimate motion parameter unknowns (Joint-TV and Joint-sTV). The proposed method is Joint-sTV. TV refers to the use of conventional TV regularization (equation 15), while sTV employs reference-guided TV regularization (equation 16).

within the same MRI session. For brain imaging, in particular, we can reasonably assume that all the contrasts share the same anatomy (i.e. structure) and focus on rigid motion. The numerical results presented in the previous section validate the

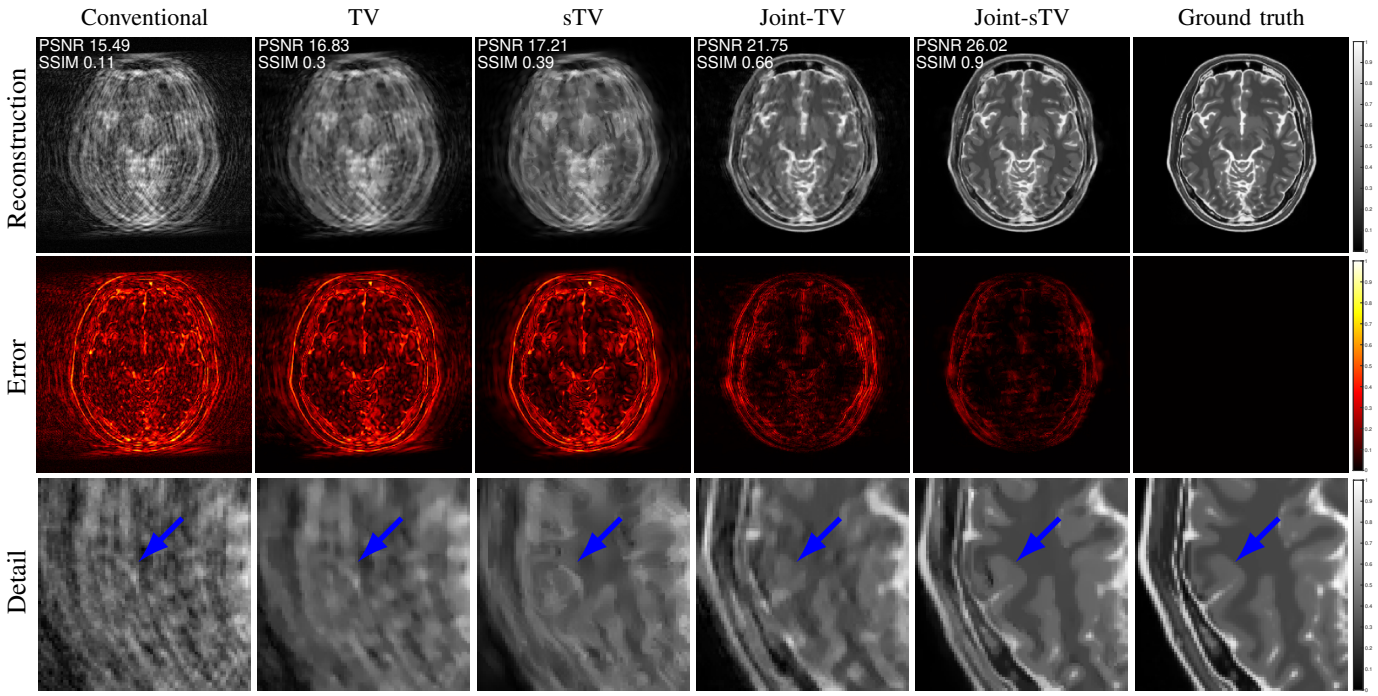
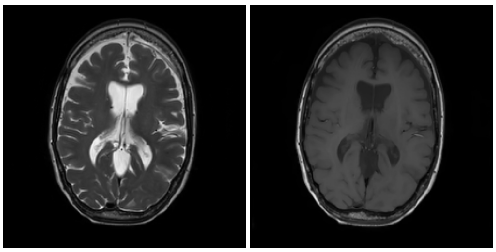


Fig. 5: Reconstruction results for random smooth motion (BrainWeb phantom). We compare the motion-corrupted image (conventional) with several reconstruction methods that ignore motion-parameter estimation (TV and sTV) and that estimate motion parameter unknowns (Joint-TV and Joint-sTV). The proposed method is Joint-sTV. TV refers to the use of conventional TV regularization (equation 15), while sTV employs reference-guided TV regularization (equation 16).



(a) T2-weighted (b) T1-weighted

Fig. 6: In-vivo contrast data: (a) ground-truth, (b) structural prior

use of contrast prior for both synthetic and realistic scenarios. In particular, we are able to consistently improve on the plain joint reconstruction results (not using structure-guided regularization) by several dB's in PSNR, with higher resolution and less motion defects. We note that the prior and the target contrasts do not need to be co-registered.

Retrospective motion correction schemes are notoriously demanding in terms of computational resources, especially in view of 3D applications. The method proposed in this work is no exception, and will likely require a GPU implementation to be effectively adopted in the clinical setting. The current CPU implementation requires about 3 minutes for each of the experiments discussed in this paper. For 3D, a 2D multi-slice reconstruction approach can also be a sensible choice. Note that, contrary to many retrospective techniques, we implemented rigid motions directly in the k-space by means of the NUDFT. Direct interpolation in the k-space allows to reduce the complexity of the image-to-data forward operator to $\mathcal{O}(n_x \log n_x)$ (in big-$\mathcal{O}$ notation), while rigid motion via interpolation in the spatial domain corresponds to $\mathcal{O}(n_t n_x \log n_x)$ (since a Fourier transform evaluation is required for each time sample). A minor drawback of this approach is that for parallel imaging the complexity increases to $\mathcal{O}(n_c n_x \log n_x)$, where n_c is the number of coil channels. This will still compare favorably to $\mathcal{O}(n_t n_x \log n_x)$ when the number of channels remains limited. In practice, it is more convenient to rely on classical pre-processing steps that either perform a preliminary (corrupted) reconstruction via parallel imaging or interpolate multi-coil data on the full k-space (e.g. via GRAPPA-like methods, [24]). We also note that the choice of NUDFT is tied to the rigid motion assumption, and more complex models will have to resort to classical DFT. Further improvements from a computational point of view can be achieved by simplifying the objective functional in equation (13) from $\mathcal{L}(\mathcal{S}\tilde{\mathcal{F}}[\mathbf{u};\boldsymbol{\Theta}] - \mathbf{D})$ to $\mathcal{L}(\mathbf{u} - (\mathcal{S}\tilde{\mathcal{F}}[\cdot;\boldsymbol{\Theta}])^{\mathrm{H}}\mathbf{D})$, as advocated in [33], since the reconstruction step for $\mathbf{u}$ boils down to a single TV proximal or projection evaluation. Alas, per [33], this might limit the magnitude of the motion that can be reasonably resolved.

An important issue is related to the bias introduced by the structural prior when the basic assumption of shared anatomy fails, with strong implications for image quality. Beside the situation where prior and target contrasts are strongly mis-registered, which can be in principle solved by a more careful choice and pre-processing, this issue might arise when the contrast prior and target come from different MRI sessions

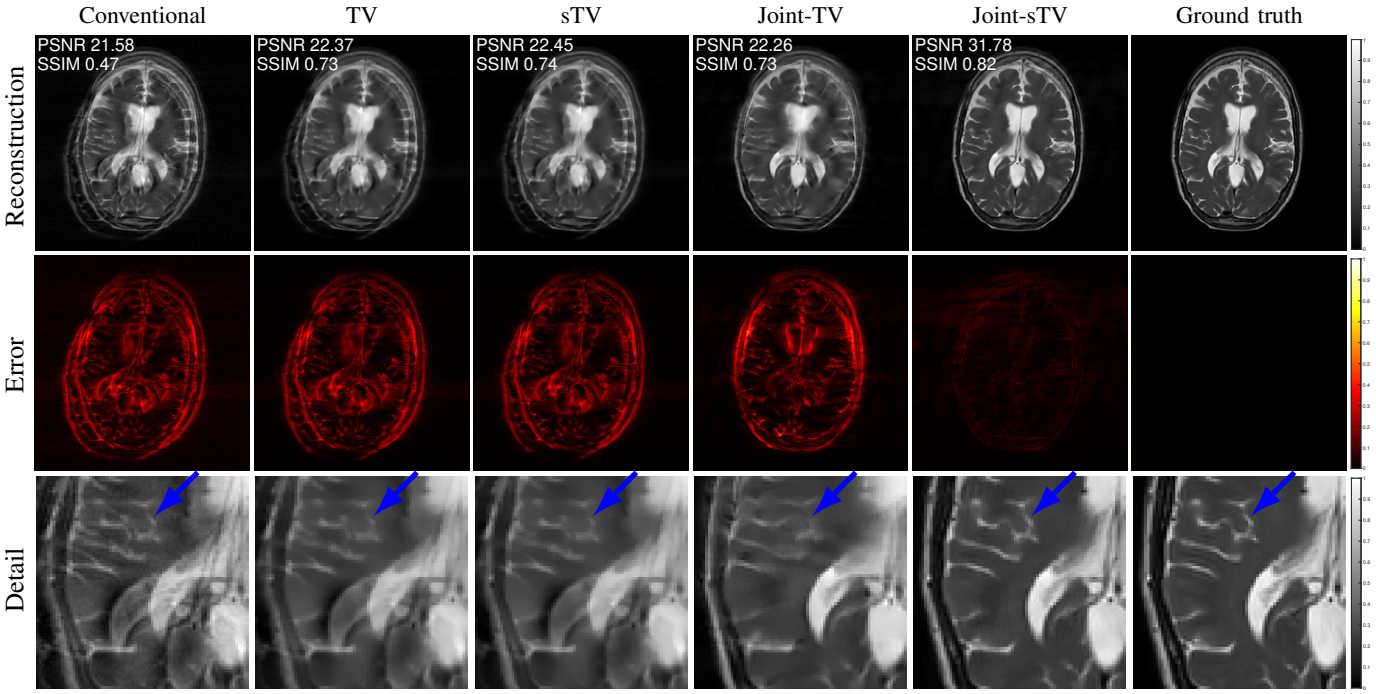


Fig. 7: Reconstruction results for sudden motion (in-vivo data with simulated motion experiment). We compare the motion-corrupted image (conventional) with several reconstruction methods that ignore motion-parameter estimation (TV and sTV) and that estimate motion parameter unknowns (Joint-TV and Joint-sTV). The proposed method is Joint-sTV. TV refers to the use of conventional TV regularization (equation 15), while sTV employs reference-guided TV regularization (equation 16).

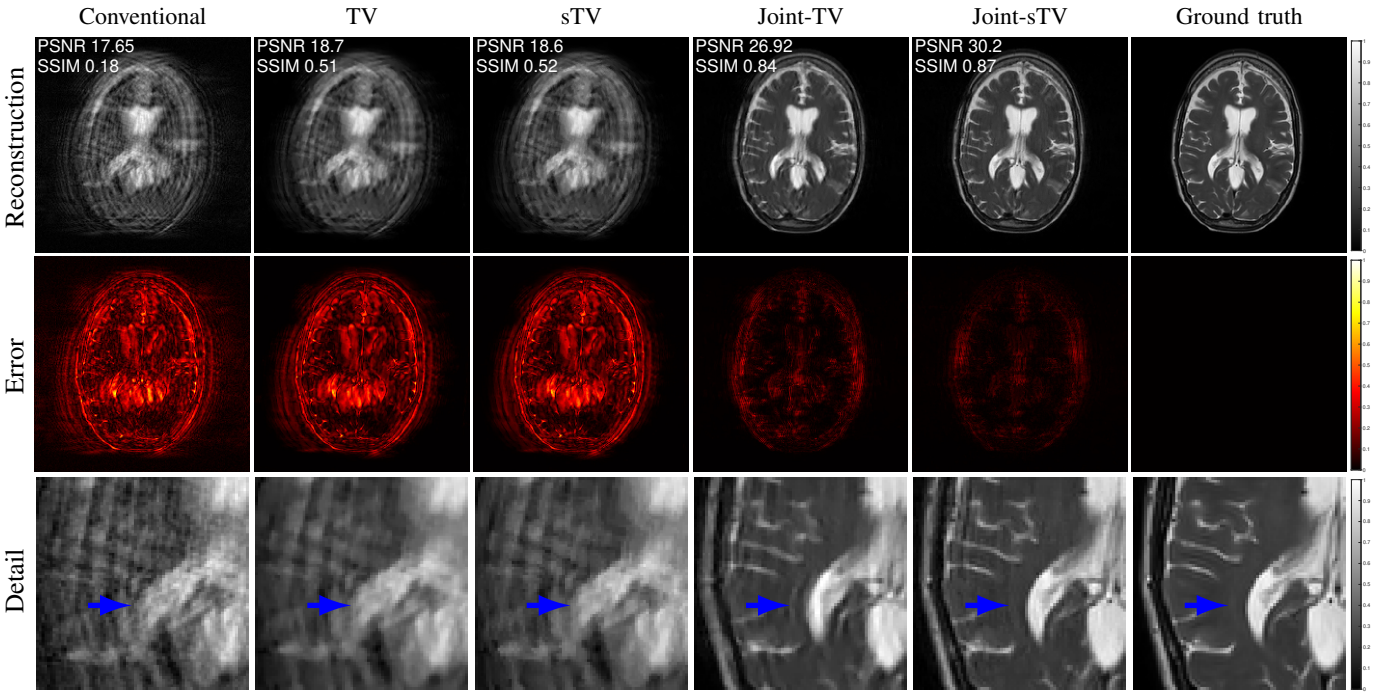


Fig. 8: Reconstruction results for periodic motion (in-vivo data with simulated motion experiment). We compare the motion-corrupted image (conventional) with several reconstruction methods that ignore motion-parameter estimation (TV and sTV) and that estimate motion parameter unknowns (Joint-TV and Joint-sTV). The proposed method is Joint-sTV. TV refers to the use of conventional TV regularization (equation 15), while sTV employs reference-guided TV regularization (equation 16).

and the patient is affected by an evolving pathology. When structural changes involve a limited region, the adverse bias can be easily mitigated by selecting the affected region and devising a mask for the regularization weight that defines the

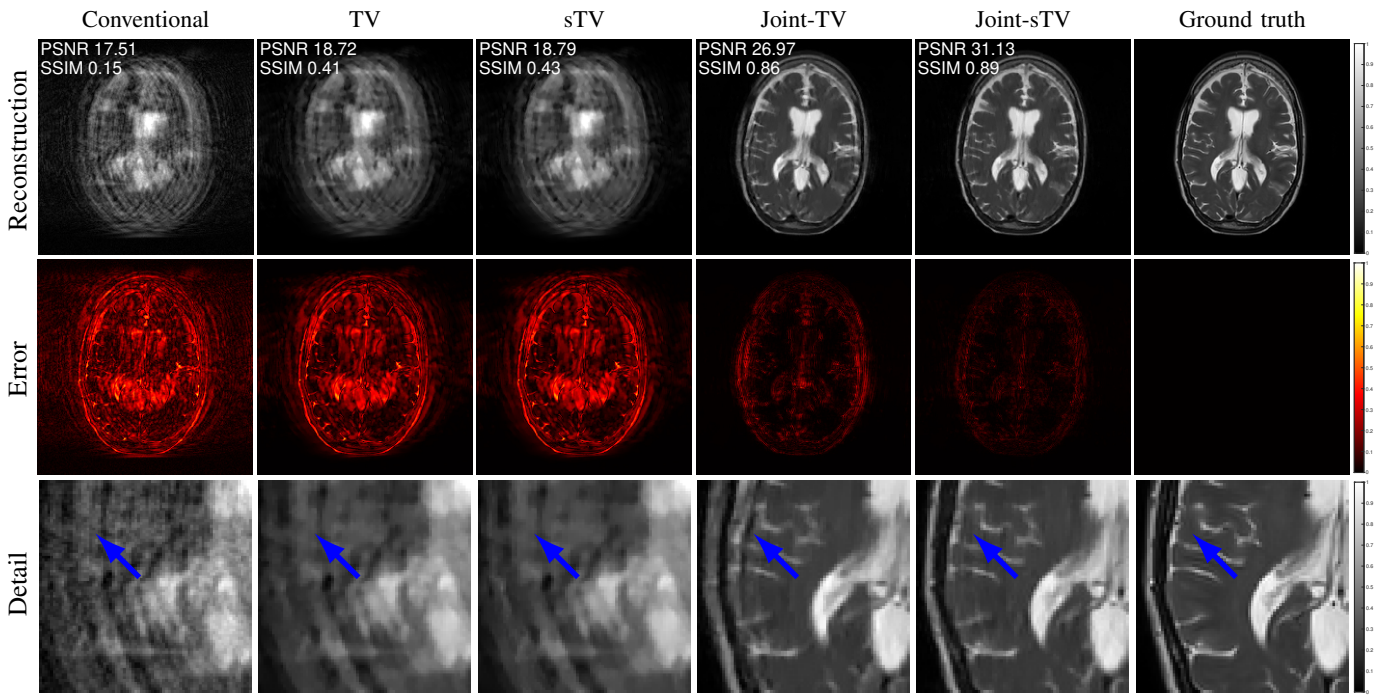


Fig. 9: Reconstruction results for random smooth motion (in-vivo data with simulated motion experiment). We compare the motion-corrupted image (conventional) with several reconstruction methods that ignore motion-parameter estimation (TV and sTV) and that estimate motion parameter unknowns (Joint-TV and Joint-sTV). The proposed method is Joint-sTV. TV refers to the use of conventional TV regularization (equation 15), while sTV employs reference-guided TV regularization (equation 16).

structure-guided TV. The effect of this bias and remedies will be investigated in the future for synthetic and in-vivo data.

The primary aim of this work is to provide a solution for reducing motion artifacts in MRI for relevant and practical clinical scenarios. For these reasons, we focused on clinically used Cartesian acquisition in a retrospective fashion. Note that arbitrary acquisition sequences can also be trivially incorporated in our scheme, and motion-resilient sampling (such as [16]) can greatly benefit from structure-guided regularization. This paper lays the foundation for such goals, but many important practical details (such as different resolution or field-of-view for different scans, multi-slice 2D vs 3D implementation, etc.) are left to future work.

The clinical implementation may be challenged by the different reconstruction geometries and field of view of structure-guided priors and motion-corrupted images, but this can be trivially address by interpolation and proper masking. We note that a motion-free structure-guided prior could have been acquired in a past MR session or by other imaging modalities. In this case, some extra care is required in using this prior information when evolving pathologies are involved, as the structural identity between the prior and the motion-corrupted image underlying our proposal will be violated. However, when the region affected by the pathology is known, we can limit its adverse effect by a simple masking procedure. We emphasize that the 2D implementation of the motion correction scheme discussed here is not designed to handle out-of-plane motion, and a 3D extension is likely required for a realistic scenario. Lastly, potential issues are related to the gap arising from the idealized rigid motion model and the actual non-rigid motion. Since we are focusing on brain imaging, the relevant literature on motion correction suggests that these issues may be minor.

V. Conclusions

We demonstrated that blind retrospective motion correction methods have the potential to reduce the imprint of motion artifacts on the MRI reconstruction. When prior information about the target structure is also included, we can achieve further improvements. For brain imaging, where motion can be effectively limited to time-dependent rigid transformations, such priors can be obtained from other contrasts, within the same MRI session, that are less affected by motion. In principle, a trained MRI technician can select well-resolved contrasts as a candidate for structure-guided prior and post-process the target contrast, without the need for repeating the corrupted scan. Such workflow fits naturally in a clinical setting. Whether the proposed methodology can be applied to realistic 3D imaging is subject to future in-vivo studies.

Data and software

The Julia implementation of the method described in this paper consists of the following packages: FastSolversForWeightedTV.jl for the structure-guided total variation utilities, and UtilitiesForMRI.jl for the motion-perturbed Fourier transform and other MRI-specific utilities.

ACKNOWLEDGMENTS

This publication is part of the project "Reducing re-scans in clinical MRI exams" (with project number 104022007 of the research program "IMDI, Technologie voor bemensbare zorg: Doorbraakprojecten") which is financed by The Netherlands Organization for Health Research and Development (ZonMW). The project is also supported by Philips Medical Systems Netherlands BV.

REFERENCES

[1] Jonas Adler. *Learned Iterative Reconstruction*, pages 1–22. Springer International Publishing, Cham, 2021.

[2] Jalal B. Andre, Brian W. Bresnahan, Mahmud Mossa-Basha, Michael N. Hoff, C. Patrick Smith, Yoshimi Anzai, and Wendy A. Cohen. Toward quantifying the prevalence, severity, and cost associated with patient motion during clinical MR examinations. *Journal of the American College of Radiology*, 12(7):689–695, jul 2015.

[3] Aleksandr Aravkin, Tristan Van Leeuwen, and Felix Herrmann. Robust full-waveform inversion using the Student's t-distribution. In *SEG Technical Program Expanded Abstracts 2011*, pages 2669–2673. Society of Exploration Geophysicists, 2011.

[4] David Atkinson and Derek LG Hill. Automatic motion correction using prior knowledge. In *Proceedings of the 9th Annual Meeting ISMRM*, page 747, 2001.

[5] David Atkinson, Derek L.G. Hill, Peter N.R. Stoyle, Paul E. Summers, Stuart Clare, Richard Bowtell, and Stephen F. Keevil. Automatic compensation of motion artifacts in MRI. *Magnetic Resonance in Medicine*, 41(1):163–170, 1999.

[6] David Atkinson, Derek L.G. Hill, Peter N.R. Stoyle, Paul E. Summers, and Stephen F. Keevil. Automatic correction of motion artifacts in magnetic resonance images using an entropy focus criterion. *IEEE Transactions on Medical Imaging*, 16(6):903–910, 1997.

[7] Alex H. Barnett. Aliasing error of the exp(β1-z2) kernel in the nonuniform fast Fourier transform. *Applied and Computational Harmonic Analysis*, 51(2):1–16, 2021.

[8] Alexander H. Barnett, Jeremy Magland, and Ludvig A.F. Klinteberg. A parallel nonuniform fast Fourier transform library based on an "exponential of semicircle" kernel. *SIAM Journal on Scientific Computing*, 41(5):C479–C504, 2019.

[9] P. G. Batchelor, D. Atkinson, P. Irarrazaval, D. L.G. Hill, J. Hajnal, and D. Larkman. Matrix description of general motion correction applied to multishot images. *Magnetic Resonance in Medicine*, 54(5):1273–1280, nov 2005.

[10] Amir Beck and Marc Teboulle. A Fast Iterative Shrinkage-Thresholding Algorithm. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[11] Åke Björck. *Numerical methods for least squares problems*. SIAM, 1996.

[12] Candice A Bookwalter, Mark A Griswold, and Duerk Jeffrey L. Multiple Overlapping k-Space Junctions for Investigating Translating Objects (MOJITO). *IEEE Transactions on Medical Imaging*, 29(2):339–349, 2010.

[13] Leon Bungert and Matthias J Ehrhardt. Robust Image Reconstruction With Misaligned Structural Information. *IEEE Access*, 8:222944–222955, 2020.

[14] Joseph Y. Cheng, Marcus T. Alley, Charles H. Cunningham, Shreyas S. Vasanawala, John M. Pauly, and Michael Lustig. Nonrigid motion correction in 3D using autofocusing with localized linear translations. *Magnetic Resonance in Medicine*, 68(6):1785–1797, 2012.

[15] Chris A Cocosco, Vasken Kollokian, Remi K-S Kwan, G Bruce Pike, and Alan C Evans. Brainweb: Online interface to a 3D MRI simulated brain database. In *NeuroImage*. Citeseer, 1997.

[16] Lucilio Cordero-Grande, Giulio Ferrazzi, Rui Pedro A.G. Teixeira, Jonathan O'Muircheartaigh, Anthony N. Price, and Joseph V. Hajnal. Motion-corrected MRI with DISORDER: Distributed and incoherent sample orders for reconstruction deblurring using encoding redundancy. *Magnetic Resonance in Medicine*, 84(2):713–726, 2020.

[17] R. L. Ehman and J. P. Felmlee. Adaptive technique for high-definition MR imaging of moving structures. *Radiology*, 173(1):255–263, 1989.

[18] Richard L Ehman and Joel P Felmlee. Adaptive technique for high-definition MR imaging of moving structures. *Radiology*, 173(1):255–263, 1989.

[19] Matthias J. Ehrhardt and Marta M. Betcke. Multicontrast MRI reconstruction with structure-guided total variation. *SIAM Journal on Imaging Sciences*, 9(3):1084–1106, 2016.

[20] Matthias J. Ehrhardt, Kris Thielemans, Luis Pizarro, David Atkinson, Sébastien Ourselin, Brian F. Hutton, and Simon R. Arridge. Joint reconstruction of PET-MRI by exploiting structural similarity. *Inverse Problems*, 31(1), 2015.

[21] Mina Ghaffari, Kamlesh Pawar, and Ruth Oliver. Brain MRI motion artifact reduction using 3D conditional generative adversarial networks on simulated motion. In *2021 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–7, 2021.

[22] Frank Godenschweger, Urte Kägebein, Daniel Stucht, Uten Yarach, Alessandro Sciarra, Renat Yakupov, Falk Lüsebrink, Peter Schulze, and Oliver Speck. Motion correction in MRI of the brain. *Physics in Medicine & Biology*, 61(5):R32–R56, 2016.

[23] Gene Golub and Victor Pereyra. Separable nonlinear least squares: The variable projection method and its applications. *Inverse Problems*, 19(2), 2003.

[24] Mark A Griswold, Peter M Jakob, Robin M Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002.

[25] Melissa W. Haskell, Stephen F. Cauley, Berkin Bilgic, Julian Hossbach, Daniel N. Splitthoff, Josef Pfeuffer, Kawin Setsompop, and Lawrence L. Wald. Network Accelerated Motion Estimation and Reduction (NAMER): Convolutional neural network guided retrospective motion correction using a separable motion model. *Magnetic Resonance in Medicine*, 82(4):1452–1461, 2019.

[26] Florian Knoll, Kristian Bredies, Thomas Pock, and Rudolf Stollberger. Second order total generalized variation (TGV) for MRI. *Magnetic Resonance in Medicine*, 65(2):480–491, 2011.

[27] Hope W Korin, Joel P Felmlee, Stephen J Riederer, and Richard L Ehman. Spatial-frequency-tuned markers and adaptive correction for rotational motion. *Magnetic resonance in medicine*, 33(5):663–669, 1995.

[28] Thomas Küstner, Karim Armanious, Jiahuan Yang, Bin Yang, Fritz Schick, and Sergios Gatidis. Retrospective correction of motion-affected MR images using deep learning frameworks. *Magnetic Resonance in Medicine*, 82(4):1527–1540, 2019.

[29] Jongyeon Lee, Byungjai Kim, and HyunWook Park. MC2-Net: motion correction network for multi-contrast brain MRI. *Magnetic Resonance in Medicine*, 86(2):1077–1092, 2021.

[30] Seul Lee, Soozy Jung, Kyu-Jin Jung, and Dong-Hyun Kim. Deep Learning in MR Motion Correction: a Brief Review and a New Motion Simulation Tool (view2Dmotion). *Investigative Magnetic Resonance Imaging*, 24(4):196–206, 2020.

[31] Wei Lin and Hee Kwon Song. Improved optimization strategies for autofocusing motion compensation in MRI via the analysis of image metric maps. *Magnetic Resonance Imaging*, 24(6):751–760, 2006.

[32] Wei Lin and Hee Kwon Song. Extrapolation and correlation (EXTRACT): A new method for motion compensation in MRI. *IEEE Transactions on Medical Imaging*, 28(1):82–93, 2009.

[33] Alexander Loktyushin, Hannes Nickisch, Rolf Pohmann, and Bernhard Schölkopf. Blind retrospective motion correction of MR images. *Magnetic Resonance in Medicine*, 70(6):1608–1618, 2013.

[34] Alexander Loktyushin, Hannes Nickisch, Rolf Pohmann, and Bernhard Schölkopf. Blind multirigid retrospective motion correction of MR images. *Magnetic Resonance in Medicine*, 73(4):1457–1468, 2015.

[35] Michael Lustig, David Donoho, and John M. Pauly. Sparse MRI: The application of compressed sensing for rapid MR imaging. *Magnetic Resonance in Medicine*, 58(6):1182–1195, 2007.

[36] Julian Maclaren, Michael Herbst, Oliver Speck, and Maxim Zaitsev. Prospective motion correction in brain imaging: a review. *Magnetic resonance in medicine*, 69(3):621–636, 2013.

[37] Armando Manduca, Kiaran P. McGee, E. Brian Welch, Joel P. Felmlee, Roger C. Grimm, and Richard L. Ehman. Autocorrection in MR Imaging: Adaptive Motion Correction without Navigator Echoes. *Radiology*, 215(3):904–909, jun 2000.

[38] Kiaran P. McGee, Joel P. Felmlee, Armando Manduca, Stephen J. Riederer, and Richard L. Ehman. Rapid autocorrection using prescan navigator echoes. *Magnetic Resonance in Medicine*, 43(4):583–588, 2000.

[39] Jason Mendes, Eugene Kholmovski, and Dennis L Parker. Rigid-Body Motion Correction with Self-Navigation MRI. *Magnetic Resonance in Medicine*, 61(3):739–747, 2009.

[40] Anita Möller, Marco Maaß, and Alfred Mertins. Blind sparse motion mri with linear subpixel interpolation. In *Bildverarbeitung für die Medizin 2015*, pages 510–515. Springer, 2015.

[41] Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[42] Freddy Odille, Pierre André Vuissoz, Pierre Yves Marie, and Jacques Felblinger. Generalized reconstruction by inversion of coupled systems (GRICS) applied to free-breathing MRI. *Magnetic Resonance in Medicine*, 60(1):146–157, 2008.

[43] Frank Ong and Michael Lustig. Joint Non-Rigid Motion Estimation and Image Reconstruction via Sparse Blind Deconvolution. In *Proceedings of the 25th Annual Meeting ISMRM*, page 3937, 2017.

[44] Ricardo Otazo, Emmanuel Candès, and Daniel K. Sodickson. Low-rank plus sparse matrix decomposition for accelerated dynamic MRI with separation of background and dynamic components. *Magnetic Resonance in Medicine*, 73(3):1125–1136, 2015.

[45] Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1(3):127–239, 2014.

[46] Kamlesh Pawar, Zhaolin Chen, N Jon Shah, and Gary F Egan. Motion Correction in MRI using Deep Convolutional Neural Network. In *Proceedings of the 27th Annual Meeting ISMRM*, page 1174, 2018.

[47] Bas Peters, Brendan R. Smithyman, and Felix J. Herrmann. Projection methods and applications for seismic nonlinear inverse problems with multiple constraints. *Geophysics*, 84(2):R251–R269, 02 2018.

[48] James G Pipe. Motion correction with PROPELLER MRI: application to head motion and free-breathing cardiac imaging. *Magnetic Resonance in Medicine*, 42(5):963–969, 1999.

[49] Ghislain Vaillant, Claudia Prieto, Christoph Kolbitsch, Graeme Penney, and Tobias Schaeffter. Retrospective rigid motion correction in k-space for segmented radial MRI. *IEEE Transactions on Medical Imaging*, 33(1):1–10, 2014.

[50] Tristan van Leeuwen and Aleksandr Y. Aravkin. Variable Projection for Non-Smooth Problems. *SIAM Journal on Scientific Computing*, 9:S249–S268, 2021.

[51] Edward Brian Welch, Armando Manduca, Roger C Grimm, Heidi A Ward, and Clifford R Jack Jr. Spherical navigator echoes for full 3D rigid body motion measurement in MRI. *Magnetic Resonance in Medicine*, 47(1):32–41, 2002.

[52] Edward Brian Welch, Phillip J Rossman, Joel P Felmlee, and Armando Manduca. Self-navigated motion correction using moments of spatial projections in radial MRI. *Magnetic Resonance in Medicine*, 52(2):337–345, 2004.

[53] Maxim Zaitsev, Julian Maclaren, and Michael Herbst. Motion artifacts in MRI: A complex problem with many partial solutions. *Journal of Magnetic Resonance Imaging*, 42(4):887–901, oct 2015.

APPENDIX A
PHASE WRAPPING AND REGULARIZATION FOR MOTION PARAMETER ESTIMATION

In blind motion correction, the a priori knowledge about the motion evolution in time is limited by definition. However, one might reasonably assume that these variations occur smoothly in time, and choose a regularization term according to

$$\mathcal{R}_\Theta(\boldsymbol{\Theta}) = \frac{\lambda}{2}^2 \|D\,\boldsymbol{\Theta}\|_{2,2}^2 . \tag{28}$$

The least-squares norm for motion unknowns is $\|\boldsymbol{\Theta}\|_{2,2}^2 = \sum_{t,p} |\boldsymbol{\theta}_p^t|^2$, and the derivative operator is, for instance,

$$(D\,\boldsymbol{\Theta})^t = \frac{1}{\Delta t}(\boldsymbol{\theta}^{t+1} - \boldsymbol{\theta}^t),$$

for a given time spacing Δt. Alternatively, one can choose higher order derivative operators, or restrict $\boldsymbol{\Theta} \in \mathbb{R}^{n_p \times n_t}$ to a lower-dimensional space by means of interpolation, e.g. by the change of variable

$$\boldsymbol{\Theta} = \mathcal{I}\,\tilde{\boldsymbol{\Theta}}, \quad \tilde{\boldsymbol{\Theta}} \in \mathbb{R}^{n_p \times \tilde{n_t}}, \tag{29}$$

where $\mathcal{I}$ is an interpolation operator. Ideally, $\tilde{n_t} \ll n_t$. Even when dealing with motion artifacts caused by smooth motion, the naive choices in equations (28) and (29) are, however, subject to phase wrapping and should be adapted accordingly. In the following, we will detail the issue.

The definition of the perturbed data model in equation (9), for a fixed image $\mathbf{u}$, reveals that the same data can be obtained with different rigid motion parameters, as phase-shift and rotation are periodic functions. The respective periods depend on the type of acquisition scheme. In general, for any $\boldsymbol{\theta}^t = (\boldsymbol{\tau}^t, \varphi^t)$ we have that

$$\tilde{\mathcal{F}}[\mathbf{u}; \boldsymbol{\theta}^t + \Delta\boldsymbol{\theta}^t] = \tilde{\mathcal{F}}[\mathbf{u}; \boldsymbol{\theta}^t], \tag{30}$$

where $\Delta\boldsymbol{\theta}^t = (\Delta\boldsymbol{\tau}^t, \Delta\varphi^t)$ satisfies the following (time-dependent) conditions:

$$\mathbf{k} \cdot \Delta\boldsymbol{\tau}^t \in \mathbb{Z}, \quad \forall\, \mathbf{k} \in K^t, \quad \text{and} \quad (2\pi)^{-1}\Delta\varphi^t \in \mathbb{Z}. \tag{31}$$

Here, $\mathbb{Z}$ is the set of relative integers. It follows that infinite different values for $\boldsymbol{\Theta}$ are equivalent solutions for our problem when no prior information is enforced over $\boldsymbol{\Theta}$. Note, however, that the periodic behavior of the rotation angle φ^t in equation (31) can be neglected, since rotations of this magnitude are unlikely in practice, and will not be addressed in this work.

As a concrete example, we might consider 2D Cartesian sampling, where each time segment K^t is defined by

$$K^t = \{(k_x^t, k_y) : k_y = j\Delta k, j = -n_y/2, \ldots, n_y/2\}. \tag{32}$$

According to this notation, the readout is performed along the k_y direction, and the phase-encoding coordinate is

$$k_x^t = t\Delta k, \quad t = -n_x/2, \ldots, n_x/2. \tag{33}$$

The k-space spacing is determined by $\Delta k = 1/L$, where L is the size of the region of interest $|\Omega| = L^2$. The solution of the periodic conditions in equation (31) is:

$$\Delta\boldsymbol{\tau}^t = \left(\frac{n}{k_x^t}, \frac{m}{\Delta k}\right), \quad n, m \in \mathbb{Z}. \tag{34}$$

The horizontal period $T_x^t = 1/k_x^t$ depends on the phase encoding value, and decreases as a function of k_x^t, which makes the phase-wrapping ambiguity more severe for high frequencies. On the other hand, the vertical period $T_y^t = 1/\Delta k$ is constant in time and is large enough (comparable to the imaging object size) that its effect can be safely ignored (similarly to the rotation angle). To illustrate the consequences of phase wrapping for Cartesian sampling, we setup the following experiment: we assume that the ideal reconstruction $\mathbf{u}$ (corresponding to the ground truth) is available, and we solve the optimization problem in equation (13) by estimating the motion parameter $\boldsymbol{\Theta}$ with initialization $\boldsymbol{\Theta} = 0$ (corresponding to no motion). No noise is artificially added to the simulated data. The estimated parameters are depicted in Figure 10. The

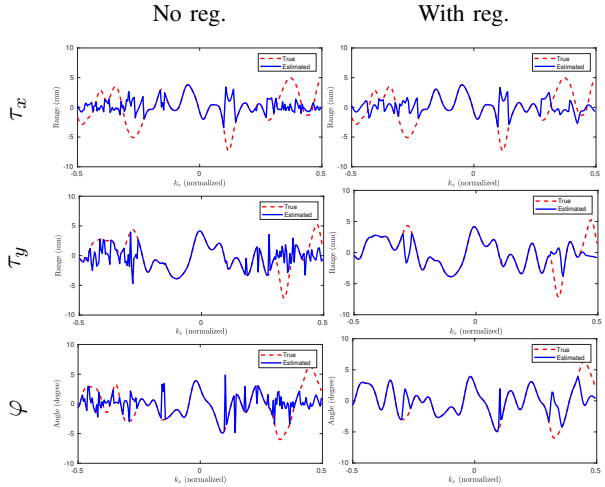


Fig. 10: Estimated rigid motion parameters $\boldsymbol{\Theta}$ for the 2D Cartesian sampling scheme under ideal conditions: we assume perfect reconstruction $\mathbf{u}$ and noiseless data. The horizontal axis represents the phase encoding wave-numbers k_x ordered according to their temporal sequence (from negative to positive wave-numbers). Due to phase wrapping and poor initial estimate, the result will not match the known solution. Enforcing smoothness improves the fit (mostly for τ_y and φ), but does not resolve the phase wrapping effects on τ_x and eventually leads to poorer data misfit.

solution, especially τ_x, presents sharp discontinuities across time. This phenomenon happens whenever it has converged to different values in contiguous time segments that differ by a multiple of the period corresponding to the transition. These results clearly indicate that enforcing smoothness for all the rigid motion parameters does not adequately address the phase wrapping issue, and an aggressive regularization strategy might instead deteriorate the data fit.

A notable alternative to Cartesian sampling is radial sampling, which corresponds, in 2D, to:

$$K^t = \{j\,R_{\alpha^t}(\Delta k, 0) : j = -n/2, \ldots, n/2\}, \tag{35}$$

where α^t denotes the angle of the line intersecting the k-space origin:

$$\alpha^t = t\Delta\alpha, \quad t = 1, 2, \ldots, \pi/\Delta\alpha. \tag{36}$$

The phase wrapping ambiguities defined by the conditions in equation (31) are explicitly given by

$$\Delta\boldsymbol{\tau}^t = R_{\alpha^t}\left(\frac{n}{\Delta k}, c\right), \quad n \in \mathbb{Z}, c \in \mathbb{R}. \tag{37}$$

While the component of the translation parameters perpendicular to K^t, here parameterized by c, cannot be fundamentally resolved, the parallel component of $\Delta\boldsymbol{\tau}^t$ has a long period $T_{\parallel} = 1/\Delta k$ and will not be affected by phase wrapping as much as for Cartesian sampling. In this work, however, we focus on scenarios more relevant for the clinical practice.

Cartesian schemes can be modified to include jittered sampling, as employed in compressed sensing. While, in principle, combining these conventional schemes with jittered sampling removes the periodic behavior of $\boldsymbol{\Theta}$ (since the conditions in equation (31) are only satisfied for $\Delta\boldsymbol{\tau}^t = 0$), the problem will still admit many local minima unless one adopts impractical randomization strategies.

In conclusion, we remark that the estimation of the rigid motion parameters is purely instrumental in improving the image quality of $\mathbf{u}$, so that the absolute values of $\boldsymbol{\Theta}$ are of relative importance. Mathematically, we are only interested in the equivalence class of $\boldsymbol{\Theta}$ modulo the respective period. Even though this ambiguity is, by itself, inconsequential for our scopes, it gives rise to some challenges for conventional regularization schemes that seek to enforce regularity over $\boldsymbol{\Theta}$ across time (as depicted by the experiment in Figure 10). A practical strategy, for 2D Cartesian sampling, is to avoid regularization for τ_x, which is the rigid motion parameter most affected by phase wrapping, and enforce regularity only for τ_y and φ. We will adopt this approach in our numerical experimentation.

Appendix B
Estimated motion parameters in joint reconstruction

The joint schemes here discussed produce estimated motion parameters as a function of "time", which for Cartesian sampling is encoded by the horizontal wave-number k_x (here conventionally aligned with the phase-encoding direction). As argued in Appendix A, the quality of these estimations is difficult to assess due to phase wrapping, which makes the comparison with the ground truth motion quite deceiving (that is, unfairly negative). As predicted theoretically, this is particularly pronounced for horizontal translation τ_x, but vertical translation τ_y and rotation angle φ are also affected because of cross-talk effects. One fair conclusion stemming from Figures 11, 12, 13, 14, 15, 16 is that the high-frequency components of the reconstruction are relatively more difficult to retrieve than the low-resolution components. At any rate, including structural priors in the joint scheme clearly widens the spectrum interval where the ground truth and the estimated motion parameters match.

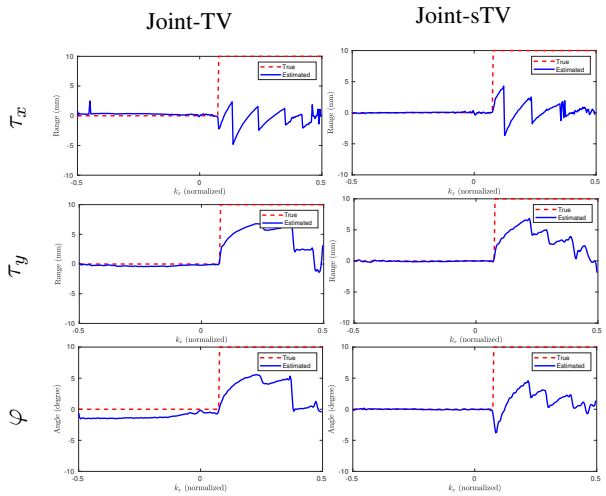


Fig. 11: Estimated motion parameters for sudden motion (BrainWeb phantom). The horizontal axis represents the phase encoding wave-numbers k_x ordered according to their temporal sequence (from negative to positive wave-numbers).

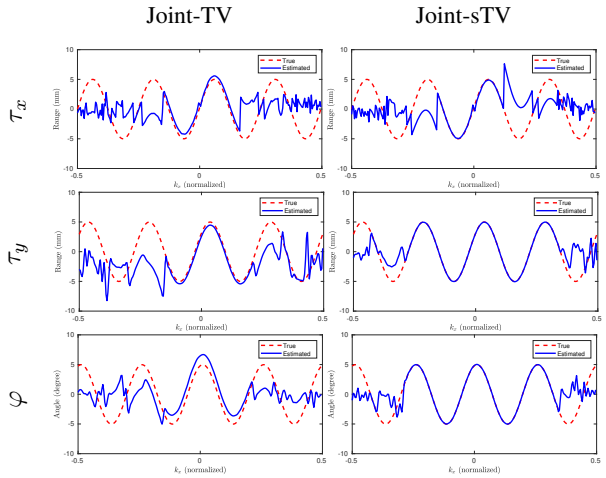


Fig. 12: Estimated motion parameters for periodic motion (BrainWeb phantom). The horizontal axis represents the phase encoding wave-numbers k_x ordered according to their temporal sequence (from negative to positive wave-numbers).

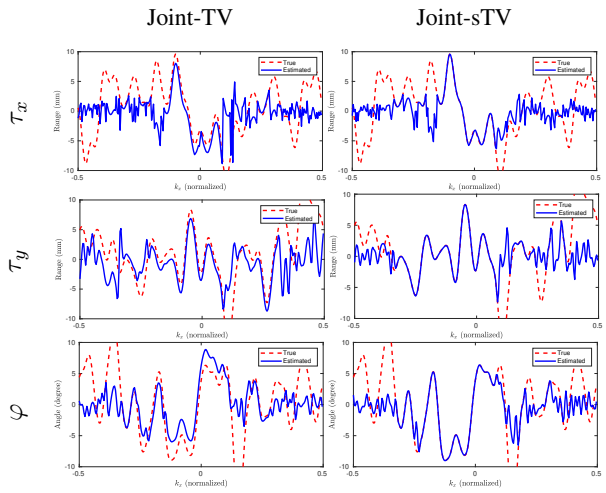


Fig. 13: Estimated motion parameters for random motion (BrainWeb phantom). The horizontal axis represents the phase encoding wave-numbers k_x ordered according to their temporal sequence (from negative to positive wave-numbers).

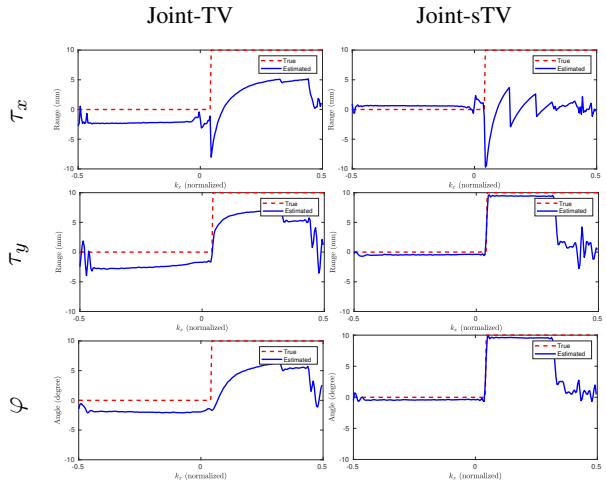


Fig. 14: Estimated motion parameters for sudden motion (in-vivo data with simulated motion experiment). The horizontal axis represents the phase encoding wave-numbers k_x ordered according to their temporal sequence (from negative to positive wave-numbers).

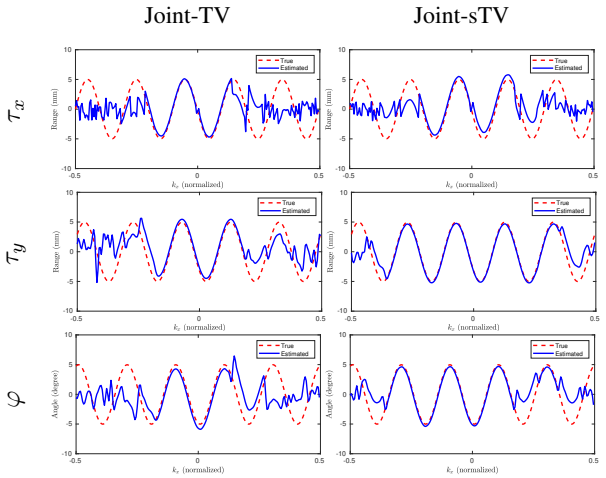


Fig. 15: Estimated motion parameters for periodic motion (in-vivo data with simulated motion experiment). The horizontal axis represents the phase encoding wave-numbers k_x ordered according to their temporal sequence (from negative to positive wave-numbers).

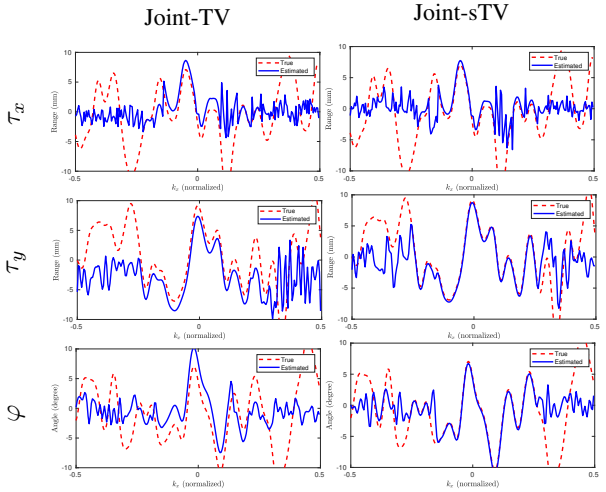


Fig. 16: Estimated motion parameters for random motion (in-vivo data with simulated motion experiment). The horizontal axis represents the phase encoding wave-numbers k_x ordered according to their temporal sequence (from negative to positive wave-numbers).