Journal of Mathematical Imaging and Vision manuscript No.
(will be inserted by the editor)

Improved Susceptibility Artifact Correction of Echo Planar MRI using the Alternating Direction Method of Multipliers

Jan Macdonald · Lars Ruthotto

Received: 01 Feb 2016 / Accepted: 09 Aug 2017 (Final publication available at Springer via http://dx.doi.org/[insert DOI]).

Abstract We present an improved technique for susceptibility artifact correction in Echo Planar Imaging (EPI), a widely used ultra fast Magnetic Resonance Imaging (MRI) technique. Our method corrects geometric deformations and intensity modulations present in EPI images. We consider a tailored variational image registration problem incorporating a physical distortion model and aiming at minimizing the distance of two oppositely distorted images subject to invertibility constraints. We derive a novel face-staggered discretization of the variational problem that renders the discretized distance function and constraints separable. Motivated by the presence of a smoothness regularizer, which leads to global coupling, we apply the Alternating Direction Method of Multipliers (ADMM) to split the problem into simpler subproblems. We prove the convergence of ADMM for this non-convex optimization problem. We show the superiority of our scheme compared to two state-of-the-art methods both in terms of correction quality and time-to-solution for 13 high-resolution 3D imaging datasets.

Keywords Numerical Optimization · Image Registration · Echo Planar Imaging (EPI) · Magnetic Resonance Imaging (MRI) · Alternating Direction Method of Multipliers (ADMM)

Mathematics Subject Classification (2000) 65K10 · 92C55 · 94A08

The second author's work is supported in part by National Science Foundation (NSF) grant DMS 1522599.

Jan Macdonald
Technische Universität Berlin, Berlin, Germany
E-mail: macdonald@math.tu-berlin.de

Lars Ruthotto
Emory University, Atlanta, GA, USA
E-mail: lruthotto@emory.edu

1 Introduction

Echo-Planar-Imaging (EPI) is an ultra fast Magnetic Resonance Imaging (MRI) technique that is widely used in medical imaging applications [29]. For example, EPI is used in the neuroscience to accelerate the acquisition of Diffusion Tensor Images (DTI) [18] or intra-operatively to guide surgery [4]. While offering a substantial reduction of scan time, a drawback of EPI is its high sensitivity to small inhomogeneities of the local magnetic field. In practical applications the magnetic field is perturbed inevitably by susceptibility variations of the object being imaged. The strength of the inhomogeneity is correlated with the strength of the external magnetic field [21] and, thus, correcting for these artifacts becomes increasingly relevant for high-resolution MRI.

The distortions are mainly caused by susceptibility variations across different tissue types; see the physical distortion model derived in [3]. It was shown that the distortion consists of two components: a geometric displacement and a modulation of image intensities. It is important to note that for EPI-MRI the displacement is practically limited to a fixed and a priori known direction, the so-called *phase-encoding direction*. Distortions in the other directions (frequency encoding and slice selection direction) are negligible. It is this particular structure that we exploit in this work to obtain a partially separable objective function and an efficient optimization scheme. The intensity modulation is given by the Jacobian determinant of the geometric transformation and ensures mass-preservation under the assumption that the Jacobian determinant is strictly positive almost everywhere. As mentioned in [3] this property needs to be ensured by judicious choice of the measurement parameters.

In recent years many approaches for susceptibility artifact correction were proposed; see, e.g., [13] for an extensive overview. Most of them employ the physical distortion model in [3], and fall into one of two categories. One approach is to obtain a *field map*, which is an estimate of the field inhomogeneity, from a reference scan and apply the physical model for susceptibility artifacts [16,17]. Alternatively, the field map can be estimated using an additional EPI image with reversed phase-encoding gradients and thus opposite deformations. The estimation problem can be phrased as a nonlinear image registration problem as originally proposed in [3]. This approach, commonly referred to as *reversed gradient* method, is taken in the following.

There are several numerical implementations of the reversed gradient method, e.g., [15,26,24,13,27,1]. For a pair of images acquired with reversed phase-encoding gradients, the goal is to estimate the field inhomogeneity such that the resulting deformations render both images as similar as possible to one another. Recently, several studies have confirmed the quality of reversed gradient approaches, e.g., in quantitative MRI (qMRI) [14], perfusion weighted MRI [31], and Arterial Spin Labeling (ASL) [19].

Despite the increasing popularity of reversed gradient methods, relatively little attention has been paid to their efficient numerical implementation. Although the methods in [15,26,24,13,27,1] are all based on the same physical distortion model, they employ different discretizations and optimization strategies. Hessian-based minimization schemes are used in [26,24,13,27,1]. As to be expected, the computationally most expensive step in these iterative methods is computing the search direction, which requires approximately solving a large linear system. An often neglected aspect is the impact of the chosen discretization on the complexity of this step. As we show in this work, a careful numerical discretization, which is motivated by the physical distortion model, can be exploited to substantially reduce the computational cost of this step.

In this paper, we present a novel fast and scalable numerical optimization scheme that considerably accelerates reversed gradient-based susceptibility artifact correction. Similar to [26,24,13], we consider a variational formulation consisting of a distance functional and a smoothness regularization functional. We use a discretize-then-optimize paradigm and follow the general guidelines described in [20]. In contrast to existing works, we derive a face-staggered discretization that exploits the fact that, in EPI-MRI, displacements are practically limited along the phase encoding direction. We show that this discretization leads to a separable structure of the discrete distance function, whereas the smoothness regularizer yields global coupling but has exploitable structure. We derive a completely parallelizable algorithm that aims at minimizing the non-convex objective function using the framework of the Alternating Direction Method of Multipliers (ADMM) [2]. We prove that ADMM converges to a stationary point of the Lagrangian associated with the non-convex objective function. The complexity of our method is essentially linear and the steps of the ADMM method can be separated into several small-dimensional subproblems that can be solved efficiently and in parallel.

The paper is organized as follows. Section 2 introduces the forward and inverse problem of susceptibility artifact correction. Section 3 describes the discretization using a face-staggered grid for the displacement. Section 4 describes and proves the convergence of the optimization method. Section 5 outlines the potential of our method using high-resolution imaging data. Finally, Section 6 concludes the paper.

2 Variational Formulation

In this section, we briefly review the physical distortion model derived in [3] and the variational formulation of EPI susceptibility artifact correction also used in [26, 24]. For clarity of presentation, we limit the discussion to the three-dimensional case, which is most relevant for our applications.

Let us first derive the forward problem. Let $\Omega \subset \mathbb{R}^3$ be a rectangular domain of interest, let $v \in \mathbb{R}^3$ denote the phase-encoding direction, and let the magnitude of the field inhomogeneity at a point $x \in \Omega$ be $b(x)$, where in the forward problem $b : \Omega \to \mathbb{R}$ is assumed to be known. As derived in [3], the distorted measurement $\mathcal{I}_{+v} : \mathbb{R}^3 \to \mathbb{R}$ and the undistorted image $\mathcal{I} : \mathbb{R}^3 \to \mathbb{R}$ satisfy

$$\begin{aligned}\mathcal{I}(x) &= \mathcal{I}_{+v}(x + b(x)v) \cdot \det \nabla(x + b(x)v) \\ &= \mathcal{I}_{+v}(x + b(x)v) \cdot (1 + v^\top \nabla b(x)),\end{aligned} \tag{1}$$

where $\det \nabla(x + b(x)v)$ denotes the Jacobian determinant of the transformation $x + b(x)v$. Following [20] we assume that the images are continuously differentiable and compactly supported in Ω. Note that the Jacobian determinant simplifies to a directional derivative since displacements are limited along one direction. Similarly, let $\mathcal{I}_{-v}$ denote a second image acquired with reversed phase-encoding direction $-v$ but otherwise unchanged imaging parameters. Using the physical distortion model (1), we have

$$\begin{aligned}\mathcal{I}(x) &= \mathcal{I}_{+v}(x + b(x)v) \cdot (1 + v^\top \nabla b(x)) \\ &= \mathcal{I}_{-v}(x - b(x)v) \cdot (1 - v^\top \nabla b(x)).\end{aligned} \tag{2}$$

In the inverse problem, both the inhomogeneity b and the undistorted image $\mathcal{I}$ are unknown. However, given two images $\mathcal{I}_{+v}$ and $\mathcal{I}_{-v}$ acquired with opposite phase-encoding directions $+v$ and $-v$, respectively, the goal is to estimate b such that the second equality in (2) holds approximately. As common we relax the equality in (2) to an $\mathcal{L}^2$-distance term

$$\mathcal{D}(b) = \frac{1}{2}\int_\Omega \left(\mathcal{T}_{+v}(b) - \mathcal{T}_{-v}(b)\right)^2 \,\mathrm{d}x \tag{3}$$

where we denote transformed images by

$$\mathcal{T}_{+v}(b) = \mathcal{I}_{+v}(x + b(x)v) \cdot (1 + v^\top \nabla b(x))$$

and

$$\mathcal{T}_{-v}(b) = \mathcal{I}_{-v}(x - b(x)v) \cdot (1 - v^\top \nabla b(x)).$$

Minimizing the distance term alone is an ill-posed problem and thus regularization is added; see [26,24,13]. In [21], methods for computing the field inhomogeneity b from susceptibility properties of the brain are developed. In this work it is shown that the field inhomogeneity should be in the Sobolev space $H^1(\Omega)$ and thus, we consider the smoothness regularizer

$$\mathcal{S}(b) = \frac{1}{2}\int_\Omega \|\nabla b(x)\|^2 \,\mathrm{d}x. \tag{4}$$

It is important to note that the physical distortion model (1) only holds if the Jacobian determinants for both phase-encoding directions are strictly positive for almost all $x \in \Omega$; see [3]. Therefore, as firstly suggested in [26,24], we impose a constraint on the Jacobian determinant

$$-1 \le v^\top \nabla b(x) \le 1 \quad \text{for almost all} \quad x \in \Omega. \tag{5}$$

As also discussed in [25, Sec. 3], we restrict the set of feasible field inhomogeneities to a closed ball $\mathcal{B}$ with respect to the $\mathcal{L}^\infty$-norm whose radius depends only on the diameter of Ω. To this end, note that both the distance and the regularization functional vanish for any large enough constant field inhomogeneity b, due to the fact that $\mathcal{I}_{+v}$ and $\mathcal{I}_{-v}$ are supported within the bounded set Ω. However, these global minimizers of $\mathcal{D}(b)$ and $\mathcal{S}(b)$ would be implausible in practical applications. This leads to the variational problem

$$\begin{aligned} \min_{b\in\mathcal{B}} \quad & \mathcal{J}(b) = \mathcal{D}(b) + \alpha\mathcal{S}(b) \\ \text{subject to} \quad & -1 \le v^\top \nabla b(x) \le 1, \quad \forall x \in \Omega, \end{aligned} \tag{6}$$

where the regularization parameter $\alpha > 0$ balances between minimizing the distance and the regularity of the solution. There is no general rule for choosing an "optimal" regularization parameter especially for nonlinear inverse problems, however, several criteria such as generalized cross validation [9,10], L-curve [11], or discrepancy principle [32] are commonly used. In this paper, we assume that α is chosen by the user. Following the guidelines in [20] we first discretize the variational problem, see Section 3, and then discuss a numerical method for solving the discrete optimization problem in Section 4.

3 Discretization

In this section, we derive a novel face-staggered discretization of the variational problem (6) that renders the discrete version of the distance functional (3) *separable* and, thus, amendable for efficient numerical optimization.

Our notation follows the general guidelines in [20]. For ease of presentation, we consider a rectangular domain $\Omega = (0,1)\times(0,1)\times(0,1) \subset \mathbb{R}^3$ and assume that the phase-encoding direction and thus the direction of the distortion is aligned with the first coordinate axis. In other words, we assume $v = e_1$, where e_1 is the first unit vector. In our experience, this is not a practical limitation since image data can be rearranged accordingly. To simplify our notation, we assume that Ω is divided into m^3 voxels with edge length $h = 1/m$ in all three coordinate directions although our implementation supports arbitrary numbers of voxels and anisotropic voxel sizes. The images $\mathcal{I}_{+v}$ and $\mathcal{I}_{-v}$ are assumed to be compactly supported and continuously differentiable functions. In practice, a continuous image model is built from discrete data by using interpolation; see [20, Sec. 3] for details.

To obtain a separable structure of the discrete distance term, we discretize the field inhomogeneity, b, by a vector $\mathbf{b} \in \mathbb{R}^{(m+1)m^2}$ on the x_1-faces of a regular grid as visualized in Figure 1. For brevity, we denote the number of x_1-faces by $n = (m+1)m^2$. Clearly, the elements of the vector $\mathbf{b}$ can be accessed using linear indices or sub indices

$$\mathbf{b}_{ijk} = b\left((i-1)h, (j-0.5)h, (k-0.5)h\right),$$

for $i = 1,\dots,m+1$ and $j,k = 1,\dots,m$. The restriction of the field inhomogeneities b to the $\mathcal{L}^\infty$-ball $\mathcal{B}$ is discretized by restricting $\mathbf{b}$ to a symmetric closed box $\Sigma \subset \mathbb{R}^n$ with edge length $2\cdot\mathrm{diam}(\Omega)$ around the origin. In our numerical experiments, it was not necessary to enforce this constraint. The distance functional (3) is approximated by a midpoint rule. To this end, the geometric transformation and the intensity modulation in (1) are approximated in the cell-centers by simple averaging and short finite differences, respectively. To compute the field inhomogeneity in the cell-

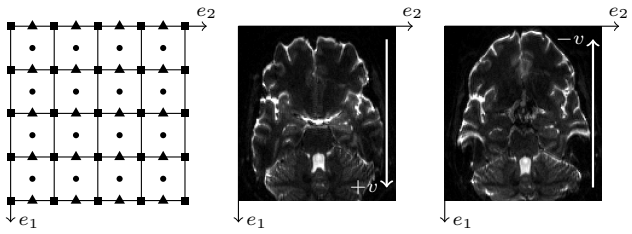

Fig. 1 Left: Cell-centered grid (circles), nodal grid (squares), and e_1-staggered grid (triangles) on a 2-dimensional box-domain with 4×4 pixels. Middle and right: Example of a reversed-gradient EPI-MRI pair provided by the Human Connectome Project [30]; see also Section 5.

centers given an e_1-staggered discretization, we use the averaging matrix

$$\mathbf{A}_1 = \mathbf{I}\left(m^2\right) \otimes \widetilde{\mathbf{A}}_1,$$

with

$$\widetilde{\mathbf{A}}_1 = \frac{1}{2}\begin{pmatrix} 1 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & 1 & 1 \end{pmatrix} \in \mathbb{R}^{m\times(m+1)},$$

where $\otimes$ denotes the Kronecker product of two matrices and $\mathbf{I}(k)$ denotes the identity matrix of size $k \times k$. From that we obtain the discretized displacement in e_1-direction using

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 \\ \mathbf{0} \\ \mathbf{0} \end{pmatrix} \in \mathbb{R}^{3m^3\times n}, \tag{7}$$

where $\mathbf{0} \in \mathbb{R}^{m^3\times n}$ is a matrix of all zeroes.

Similarly, we discretize the first partial differential operator using short finite differences matrices

$$\mathbf{D}_1 = \mathbf{I}\left(m^2\right) \otimes \widetilde{\mathbf{D}}(m+1, h), \tag{8}$$

where

$$\widetilde{\mathbf{D}}(m,h) = \frac{1}{h}\begin{pmatrix} -1 & 1 & & 0 \\ & \ddots & \ddots & \\ 0 & & -1 & 1 \end{pmatrix} \in \mathbb{R}^{(m-1)\times m}.$$

Combining (7) and (8) the distance functional (3) is approximated by

$$D(\mathbf{b}) = \frac{h^3}{2}\big\|\mathcal{I}_{+v}(\mathbf{x}+\mathbf{A}\mathbf{b}) \odot (\mathbf{e}+\mathbf{D}_1\mathbf{b}) - \mathcal{I}_{-v}(\mathbf{x}-\mathbf{A}\mathbf{b}) \odot (\mathbf{e}-\mathbf{D}_1\mathbf{b})\big\|^2, \tag{9}$$

where $\mathbf{x}$ are the cell-centers of a uniform grid, $\mathcal{I}_{\pm v}(\mathbf{x} \pm \mathbf{A}\mathbf{b})$ denote vectors containing the image intensities at the shifted grid points, $\odot$ denotes the component-wise Hadamard product, and $\mathbf{e} \in \mathbb{R}^{m^3}$ is a vector of all ones.

Similarly, we approximate the regularization functional (4) by

$$S(\mathbf{b}) = \frac{h^3}{2}\left(\|\mathbf{D}_1\mathbf{b}\|^2 + \|\mathbf{D}_2\mathbf{b}\|^2 + \|\mathbf{D}_3\mathbf{b}\|^2\right), \tag{10}$$

where the discrete partial derivative operators are

$$\mathbf{D}_2 = \mathbf{I}(m) \otimes \widetilde{\mathbf{D}}(m,h) \otimes \mathbf{I}(m+1) \tag{11}$$

and

$$\mathbf{D}_3 = \widetilde{\mathbf{D}}(m,h) \otimes \mathbf{I}(m\cdot(m+1)). \tag{12}$$

Using (8) to discretize the constraint and combining (9) and (10), we obtain the finite-dimensional optimization problem

$$\begin{aligned} \min_{\mathbf{b}\in\Sigma} \quad & J(\mathbf{b}) = D(\mathbf{b}) + \alpha S(\mathbf{b}) \\ \text{subject to} \quad & -1 \le \mathbf{D}_1\mathbf{b} \le 1. \end{aligned} \tag{13}$$

All components of (13) are smooth, the regularizer is a convex quadratic, the constraints are convex, however, the distance function is in general non-convex. The non-convexity is addressed using a multilevel strategy; see Section 4.3. We exploit the separable structure of the distance term and the constraints in (13) and derive an efficient implementation of a Sequential Quadratic Programming (SQP) method within the framework of ADMM. The approach is described in more detail in Section 4.

In preparation for efficient, Hessian-based optimization, we quickly derive the gradients and approximate Hessians of the discretized distance term (9) and regularizer term (10). Let us denote the residual of the distance term (9) by

$$\mathbf{r}(\mathbf{b}) = \mathcal{I}_{+v}(\mathbf{x}+\mathbf{A}\mathbf{b}) \odot (\mathbf{e}+\mathbf{D}_1\mathbf{b}) - \mathcal{I}_{-v}(\mathbf{x}-\mathbf{A}\mathbf{b}) \odot (\mathbf{e}-\mathbf{D}_1\mathbf{b}).$$

Then, denoting the Jacobian matrix of the residual for a fixed $\mathbf{b}$ by $\mathbf{J}_\mathbf{r}(\mathbf{b}) \in \mathbb{R}^{m^3\times n}$, we obtain the gradient and Gauss-Newton approximation to the Hessian as

$$\nabla D(\mathbf{b}) = h^3\ \mathbf{J}_\mathbf{r}(\mathbf{b})^\top \mathbf{r}(\mathbf{b})$$

and

$$\nabla^2 D(\mathbf{b}) \approx \mathbf{H}_D(\mathbf{b}) = h^3\ \mathbf{J}_\mathbf{r}(\mathbf{b})^\top \mathbf{J}_\mathbf{r}(\mathbf{b}).$$

The Jacobian of the residual, is given by

$$\begin{aligned} \mathbf{J}_\mathbf{r}(\mathbf{b}) = {} & \operatorname{diag}\left(\partial_1\mathcal{I}_{+v}(\mathbf{x}+\mathbf{A}\mathbf{b}) \odot (\mathbf{e}+\mathbf{D}_1\mathbf{b})\right)\mathbf{A}_1 \\ & + \operatorname{diag}\left(\mathcal{I}_{+v}(\mathbf{x}+\mathbf{A}\mathbf{b})\right)\mathbf{D}_1 \\ & + \operatorname{diag}\left(\partial_1\mathcal{I}_{-v}(\mathbf{x}-\mathbf{A}\mathbf{b}) \odot (\mathbf{e}-\mathbf{D}_1\mathbf{b})\right)\mathbf{A}_1 \\ & + \operatorname{diag}\left(\mathcal{I}_{-v}(\mathbf{x}-\mathbf{A}\mathbf{b})\right)\mathbf{D}_1, \end{aligned}$$

where $\partial_1\mathcal{I}_{\pm v}(\mathbf{x}\pm\mathbf{A}\mathbf{b})$ denote vectors of the first partial derivatives of the images evaluated at the shifted grid

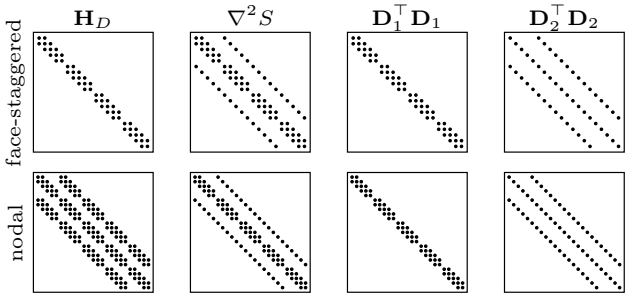

Fig. 2 Sparsity patterns of the (approximated) Hessians of the distance function, smoothness regularizer, and its first and second term (from left to right) for the e_1-staggered discretization (top row) and the nodal discretization used in [26, 24] (bottom row) for the 4×4 example shown in Figure 1. Note that the Hessian of the distance term and the first term of the regularizer are block diagonal and therefore separable with respect to image columns for the face-staggered discretization. Coupling is only introduced by the second term of the Hessian of the smoothness regularizer.

points and $\mathrm{diag}(\mathbf{v}) \in \mathbb{R}^{n\times n}$ is a diagonal matrix with diagonal entries given by the vector $\mathbf{v} \in \mathbb{R}^n$.

Unlike the distance function, the regularizer introduces coupling along all dimensions. The Hessian and the gradient of the regularization function (10) are

$$\nabla^2 S = h^3 \left(\mathbf{D}_1^\top \mathbf{D}_1 + \mathbf{D}_2^\top \mathbf{D}_2 + \mathbf{D}_3^\top \mathbf{D}_3\right)$$

and

$$\nabla S(\mathbf{b}) = \nabla^2 S\ \mathbf{b}.$$

In view of efficient numerical methods, we note that $\nabla^2 S$ is of Block-Toeplitz-Toeplitz-Block (BTTB) structure; see, e.g., [12]. Combining the above derivations, the gradient and approximated Hessian of the discrete objective functions are

$$\nabla J(\mathbf{b}) = \nabla D(\mathbf{b}) + \alpha \nabla S(\mathbf{b}) \tag{14}$$

and

$$\mathbf{H}_J(\mathbf{b}) = \mathbf{H}_D(\mathbf{b}) + \alpha \nabla^2 S. \tag{15}$$

A key idea of the optimization method presented in the following section is to exploit the separability of the distance function that can also be seen in the sparsity structure of the Hessian matrix. Due to the choice of the average operator $\mathbf{A}$ and the short finite difference operator $\mathbf{D}_1$, the Hessian of the distance function has a block-diagonal structure with tridiagonal blocks of size $(m+1)\times(m+1)$; see also Figure 2. Thus, minimizing the distance term would decouple into several smaller optimization problems. The Hessian of the regularizer, $\nabla^2 S$, is a discrete version of the negative Laplacian on Ω with homogeneous Neumann boundary conditions and has a banded structure. The term $\mathbf{D}_1^\top \mathbf{D}_1$ has the same block-diagonal structure with tridiagonal blocks of size $(m+1)\times(m+1)$ as $\nabla^2 D$. So the coupling introduced by the regularizer comes from the terms $\mathbf{D}_2^\top \mathbf{D}_2$ and $\mathbf{D}_3^\top \mathbf{D}_3$ only. Figure 2 shows a comparison of the proposed face-staggered discretization and the nodal discretization used in [26, 24].

4 Numerical Optimization

In this section, we propose an efficient iterative method for solving the discretized constrained non-convex optimization problem (13) using the Alternating Direction Method of Multipliers (ADMM). Our method exploits the separability of the discrete distance function achieved by the face-staggered discretization while enforcing the smoothness of the regularizer. We obtain a scalable method in the sense that its complexity grows *essentially linearly* with the number of unknowns in the discrete optimization problem. In section 4.1 we use ADMM to decouple the optimization into two subproblems that can be solved efficiently and in parallel. Due to the non-convexity of the objective function $J(\mathbf{b})$ traditional convergence results for ADMM such as in [2] do not hold here. Therefore, we prove convergence of ADMM for our problem (13) in Section 4.2. Finally, in Section 4.3 we describe the coarse-to-fine multilevel strategy used in our experiments. Throughout this section, we denote the iteration counter by superscripts.

4.1 ADMM

We aim at exploiting the partial separability of the discrete distance function (9) by applying the Alternating Direction Method of Multipliers (ADMM). Originally developed in the mid 1970s, ADMM recently received a lot of attention in many data science and imaging applications; see, e.g., [7] and the recent surveys [2,5,8]. The key idea in our case is to split the terms of the objective function (13) into parts that are separable with respect to image columns and terms that couple across image columns. To this end, let $\mathbf{z} \in \mathbb{R}^n$ be a new artificial variable and split the objective function J into

$$f(\mathbf{b}) = D(\mathbf{b}) + \frac{\alpha h^3}{2} \|\mathbf{D}_1 \mathbf{b}\|^2 \tag{16}$$

and

$$g(\mathbf{z}) = \frac{\alpha h^3}{2} \left(\|\mathbf{D}_2 \mathbf{z}\|^2 + \|\mathbf{D}_3 \mathbf{z}\|^2\right). \tag{17}$$

Then problem (13) is equivalent to

$$\min_{\mathbf{b},\mathbf{z}\in\mathbb{R}^n} \iota_C(\mathbf{b}) + f(\mathbf{b}) + g(\mathbf{z}) \quad \text{subject to} \quad \mathbf{b} = \mathbf{z}, \tag{18}$$

where we encode the linear inequality and box constraints into the objective function, by using the indicator function $\iota_C(\mathbf{b})$ of the compact and convex set $C = \{\, \mathbf{b} \in \Sigma \,:\, -1 \leq \mathbf{D}_1\mathbf{b} \leq 1 \,\}$, taking the value 0, whenever $\mathbf{b} \in C$, and ∞ otherwise. For an augmentation parameter $\rho > 0$, whose choice is discussed below, we aim at finding a stationary point of the augmented Lagrangian of (18)

$$\mathcal{L}_\rho(\mathbf{b}, \mathbf{z}, \mathbf{y}) = \iota_C(\mathbf{b}) + f(\mathbf{b}) + g(\mathbf{z}) + \mathbf{y}^\top(\mathbf{b} - \mathbf{z}) + \frac{\rho h^3}{2} \|\mathbf{b} - \mathbf{z}\|^2, \tag{19}$$

where $\mathbf{y} \in \mathbb{R}^n$ is the Lagrange multiplier associated with the equality constraint in (18). The idea of ADMM is to update $\mathbf{b}$, $\mathbf{z}$, and $\mathbf{y}$ in an alternating fashion

$$\mathbf{b}^{k+1} = \operatorname*{argmin}_{\mathbf{b}\in\mathbb{R}^n} \mathcal{L}_\rho(\mathbf{b}, \mathbf{z}^k, \mathbf{y}^k), \tag{20}$$

$$\mathbf{z}^{k+1} = \operatorname*{argmin}_{\mathbf{z}\in\mathbb{R}^n} \mathcal{L}_\rho(\mathbf{b}^{k+1}, \mathbf{z}, \mathbf{y}^k), \tag{21}$$

$$\mathbf{y}^{k+1} = \mathbf{y}^k + \rho h^3(\mathbf{b}^{k+1} - \mathbf{z}^{k+1}). \tag{22}$$

Introducing the scaled dual variable $\mathbf{u} = \mathbf{y}/(\rho h^3)$, we obtain the following iteration

$$\mathbf{b}^{k+1} = \operatorname*{argmin}_{\mathbf{b}\in\mathbb{R}^n} \iota_C(\mathbf{b}) + f(\mathbf{b}) + \frac{\rho h^3}{2} \left\|\mathbf{b} - \mathbf{z}^k + \mathbf{u}^k\right\|^2, \tag{23}$$

$$\mathbf{z}^{k+1} = \operatorname*{argmin}_{\mathbf{z}\in\mathbb{R}^n} g(\mathbf{z}) + \frac{\rho h^3}{2} \left\|\mathbf{b}^{k+1} - \mathbf{z} + \mathbf{u}^k\right\|^2, \tag{24}$$

$$\mathbf{u}^{k+1} = \mathbf{u}^k + \mathbf{b}^{k+1} - \mathbf{z}^{k+1}. \tag{25}$$

The first subproblem, updating $\mathbf{b}$, is a non-convex constrained optimization problem and can be solved approximately using Sequential Quadratic Programming (SQP) as described below. For the discretization derived in the previous section, this problem is separable with respect to the columns in the image and thus can be further broken down into m^2 separate steps. In each iteration of the SQP method, we form a quadratic approximation of the objective function and solve the resulting quadratic program (QP) using the active set method with Schur complement solver described in [22, Ch. 16]. The general form of the QP is

$$\min_{\mathbf{x}} \frac{1}{2}\mathbf{x}^\top\mathbf{G}\mathbf{x} + \mathbf{c}^\top\mathbf{x} \quad \text{subject to} \quad \mathbf{A}\mathbf{x} \geq \mathbf{d}, \tag{26}$$

where, in our case, $\mathbf{G}$ is the respective Hessian approximation, $\mathbf{c}$ is the gradient, $\mathbf{D}_1$ and $-\mathbf{D}_1$ are stacked into $\mathbf{A}$, and $\mathbf{d}$ is a vector of all negative ones. Let $\mathbf{x}^k$ be the current iterate in this QP and $I \subset \{1, 2, \ldots, n\}$ be a subset of component-indices describing the active set. Further, let $\mathbf{A}_I$ be the matrix containing the rows of $\mathbf{A}$ associated with active constraints and let $\mathbf{d}_I$ be the corresponding right-hand-side. Then we obtain an update direction $\mathbf{p}$ for $\mathbf{x}^k$ and the Lagrange multiplier $\boldsymbol{\lambda}$ of the active constraints by solving

$$\begin{pmatrix} \mathbf{G} & \mathbf{A}_I^\top \\ \mathbf{A}_I & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{p} \\ -\boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} -\mathbf{c} - \mathbf{G}\mathbf{x}^k \\ \mathbf{d}_I - \mathbf{A}_I\mathbf{x}^k \end{pmatrix} =: \begin{pmatrix} \mathbf{g} \\ \mathbf{h} \end{pmatrix} \tag{27}$$

using the Schur complement, i.e., by setting

$$\boldsymbol{\lambda} = -\left(\mathbf{A}_I\mathbf{G}^{-1}\mathbf{A}_I^\top\right)^{-1}\left(\mathbf{A}_I\mathbf{G}^{-1}\mathbf{g} - \mathbf{h}\right)$$

and

$$\mathbf{p} = \mathbf{G}^{-1}\left(\mathbf{A}_I^\top\boldsymbol{\lambda} + \mathbf{g}\right).$$

Now three cases can occur: First, if $\mathbf{p} = \mathbf{0}$ and $\boldsymbol{\lambda} \geq \mathbf{0}$ component-wise, then $\mathbf{x}^k$ is the global solution to (26). Second, if $\mathbf{p} = \mathbf{0}$ but some components of $\boldsymbol{\lambda}$ are negative, we remove up to one constraint with negative Lagrange multiplier per block in $\mathbf{A}$ from the active set and repeat the above computations. Third, if $\mathbf{p} \neq \mathbf{0}$, compute a step length λ^k such that $\mathbf{x}^k + \lambda^k\mathbf{p}$ is feasible but at least one additional constraint has become active. Then all of these constraints are added to the active set. Linear independence of the active constraints is guaranteed since the partial derivative at an active voxel can be either 1 or -1. Note that the Schur complement involves $\mathbf{G}^{-1}$ and thus an efficient method to solve linear systems with matrix $\mathbf{G}$ is needed. In our application, the chosen discretization allows for fast inversions.

More precisely, for the update of $\mathbf{b}$, we have

$$\mathbf{c} = \nabla D(\mathbf{b}^k) + \alpha h^3\mathbf{D}_1\mathbf{b}^k + \rho h^3(\mathbf{b}^k - \mathbf{z}^k + \mathbf{u}^k)$$

and

$$\mathbf{G} = \mathbf{H}_D\left(\mathbf{b}^k\right) + \alpha h^3\mathbf{D}_1^\top\mathbf{D}_1 + \rho h^3\mathbf{I}\,(n),$$

which has the before mentioned block-diagonal structure with tridiagonal blocks of size $m + 1$. Therefore, linear systems involving $\mathbf{G}$ can be solved in parallel by a direct method with a linear complexity of $\mathcal{O}\,(n)$.

The second subproblem, updating $\mathbf{z}$, is an unconstrained strictly convex quadratic optimization problem with a structured, symmetric positive definite Hessian

$$\widetilde{\mathbf{G}} = \alpha h^3\left(\mathbf{D}_2^\top\mathbf{D}_2 + \mathbf{D}_3^\top\mathbf{D}_3\right) + \rho h^3\mathbf{I}\,(n)$$

and has the closed form solution

$$\mathbf{z}^{k+1} = \widetilde{\mathbf{G}}^{-1}\left(\rho h^3(\mathbf{b}^{k+1} + \mathbf{u}^k)\right).$$

The matrix $\widetilde{\mathbf{G}}$ can be reordered into a block-diagonal matrix with $m + 1$ blocks, whose blocks are matrices of size $m^2 \times m^2$. Each block is a discretization of the negative two-dimensional Laplacian with homogeneous Neumann boundary conditions on a regular mesh and

thus is a Block-Toeplitz-Toeplitz-Block (BTTB) matrix. Hence, $\widetilde{\mathbf{G}}$ can be diagonalized using two-dimensional Discrete Cosine Transforms (DCT); see [12, Ch. 4]. More specifically, denote by $\mathbf{C}(m) \in \mathbb{R}^{m\times m}$ the one-dimensional DCT of size m. Then we have

$$\widetilde{\mathbf{D}}(m,h)^\top \widetilde{\mathbf{D}}(m,h) = \mathbf{C}(m)^\top \boldsymbol{\Lambda}(m,h)\mathbf{C}(m) \tag{28}$$

for some diagonal matrix $\boldsymbol{\Lambda}(m,h)$. Combining (11), (12) and (28), we immediately obtain

$$\mathbf{D}_2^\top \mathbf{D}_2 + \mathbf{D}_3^\top \mathbf{D}_3 = \mathbf{C}^\top \boldsymbol{\Lambda} \mathbf{C},$$

where

$$\mathbf{C} = \mathbf{C}(m) \otimes \mathbf{C}(m) \otimes \mathbf{I}(m+1)$$

is the two-dimensional DCT along e_2- and e_3-direction, and

$$\begin{aligned}\boldsymbol{\Lambda} = {} & \mathbf{I}(m) \otimes \boldsymbol{\Lambda}(m,h) \otimes \mathbf{I}(m+1) \\ & + \boldsymbol{\Lambda}(m,h) \otimes \mathbf{I}(m) \otimes \mathbf{I}(m+1)\end{aligned}$$

is a diagonal matrix that only needs to be computed once for every fixed discretization (m,h). Finally, since $\mathbf{C}$ is orthogonal, we obtain

$$\widetilde{\mathbf{G}} = \mathbf{C}^\top \left(\alpha h^3 \boldsymbol{\Lambda} + \rho h^3 \mathbf{I}(n)\right) \mathbf{C}.$$

Thus, the second ADMM step, updating $\mathbf{z}$, requires $m+1$ two-dimensional DCTs, which is of complexity $\mathcal{O}\left((m+1)(m^2 \log m^2)\right)$, followed by a diagonal solve, and finally $m+1$ inverse two-dimensional DCTs. Again this is a direct solve and does not require the use of iterative methods.

For the ADMM algorithm (23) to (25), we use the stopping criteria proposed in [2, Sec. 3], which are also justified by our convergence analysis in Section 4.2. We stop when the norm of the primal and dual residual satisfy

$$\left\|\mathbf{b}^{k+1} - \mathbf{z}^{k+1}\right\| \le \epsilon_{\text{pri}} \tag{29}$$

and

$$\rho h^3 \left\|\mathbf{z}^k - \mathbf{z}^{k+1}\right\| \le \epsilon_{\text{dual}}, \tag{30}$$

where ϵ_{pri} and ϵ_{dual} are computed exactly as (3.12) in [2] using a combination of an absolute and a relative tolerance

$$\begin{aligned}\epsilon_{\text{pri}} &= \sqrt{n}\,\epsilon_{\text{abs}} + \epsilon_{\text{rel}} \max\left\{\left\|\mathbf{b}^k\right\|, \left\|\mathbf{z}^k\right\|\right\} \\ \epsilon_{\text{dual}} &= \sqrt{n}\,\epsilon_{\text{abs}} + \epsilon_{\text{rel}} \rho h^3 \left\|\mathbf{u}^k\right\|.\end{aligned}$$

In our numerical experiments in Section 5, we choose $\epsilon_{\text{abs}} = \epsilon_{\text{rel}} = 2 \cdot 10^{-1}$.

4.2 Convergence of ADMM

It is important to stress that the first subproblem in our ADMM algorithm, (23), is non-convex and thus the traditional convergence results for ADMM do not hold. However, ADMM can be considered a local optimization method and has been successfully applied to non-convex problems in other applications; see [2, Ch.9] for some examples. Recently, convergence results have been established under some modest conditions on the functions involved; see, e.g., [33]. In the following, we show convergence of ADMM for the specific problem at hand. Using the smoothness of our problem we obtain a simplified, but less general, convergence proof as compared to [33]. We first recall that the functions f and g in (16) and (17) are twice continuously differentiable by construction. Further, ∇f is Lipschitz continuous over C and ∇g is Lipschitz continuous over $\mathbb{R}^n$. We denote the corresponding Lipschitz constants by L_f and L_g respectively. The relation between these Lipschitz constants and the augmentation parameter, ρ, is crucial in the convergence analysis. Throughout this section we use both the unscaled ADMM formulation (20) to (22) as well as the equivalent scaled formulation (23) to (25), whichever is more convenient. Recall that $\mathbf{u}^k = \mathbf{y}^k/(\rho h^3)$ is the scaled Lagrange multiplier. The main result of this section is the following.

Theorem 1. *For $\rho > \frac{2}{h^3}\max\{L_f, L_g\}$, the sequence of iterates $\{\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k\}_k$ generated by the ADMM algorithm (20) to (22) converges subsequentially and each limit point $\{\mathbf{b}^*, \mathbf{z}^*, \mathbf{y}^*\}$ is a stationary point of the augmented Lagrangian $\mathcal{L}_\rho$ in (19).*

As in [33], our proof is based on the following three properties.

Theorem 2. *For $\rho > \frac{2}{h^3}\max\{L_f, L_g\}$, the sequence of iterates $\{\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k\}_k$ generated by the ADMM algorithm (20) to (22) has the following properties:*

(P1) The iterates $\{\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k\}_k$ are well-defined, bounded, and $\{\mathcal{L}_\rho(\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k)\}_k$ is bounded below.

(P2) The value of the Lagrangian decreases sufficiently fast, meaning there is a constant $C_1 > 0$ such that

$$\begin{aligned}&\mathcal{L}_\rho(\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k) - \mathcal{L}_\rho(\mathbf{b}^{k+1}, \mathbf{z}^{k+1}, \mathbf{y}^{k+1}) \\ &\quad \ge C_1 \left(\left\|\mathbf{b}^k - \mathbf{b}^{k+1}\right\|^2 + \left\|\mathbf{z}^k - \mathbf{z}^{k+1}\right\|^2\right).\end{aligned}$$

(P3) There exists a constant $C_2 > 0$ and subgradients $\mathbf{d}^{k+1} \in \partial\mathcal{L}_\rho(\mathbf{b}^{k+1}, \mathbf{z}^{k+1}, \mathbf{y}^{k+1})$ such that

$$\left\|\mathbf{d}^{k+1}\right\| \le C_2 \left(\left\|\mathbf{b}^k - \mathbf{b}^{k+1}\right\| + \left\|\mathbf{z}^k - \mathbf{z}^{k+1}\right\|\right).$$

We split the proof of Theorem 2 into several lemmas.

Lemma 3. *For $\rho > 0$ the subproblems (23) and (24) have at least one solution for all $k \in \mathbb{N}$.*

Proof. For arbitrary but fixed k, the first subproblem, the update $\mathbf{b}^k \to \mathbf{b}^{k+1}$, consists of minimizing the function

$$f(\mathbf{b}) + \frac{\rho h^3}{2} \left\| \mathbf{b} - \mathbf{z}^k + \mathbf{u}^k \right\|^2$$

over the compact and convex set C. On this set f is twice continuously differentiable, so the problem is well-defined, i.e., there exists a global solution. The second subproblem, the update $\mathbf{z}^k \to \mathbf{z}^{k+1}$, consists of minimizing

$$g(\mathbf{z}) + \frac{\rho h^3}{2} \left\| \mathbf{b}^{k+1} - \mathbf{z} + \mathbf{u}^k \right\|^2$$

over $\mathbf{z} \in \mathbb{R}^n$. This problem is well defined, since g is a convex quadratic function and $\rho > 0$, which renders the overall objective function strictly convex. Thus, there exists a unique global minimizer. □

We denote the changes in the value of the augmented Lagrangian for the three consecutive ADMM steps by

$$\begin{aligned}
\Delta_{\mathbf{b}} \mathcal{L}_\rho &= \mathcal{L}_\rho(\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k) - \mathcal{L}_\rho(\mathbf{b}^{k+1}, \mathbf{z}^k, \mathbf{y}^k), \\
\Delta_{\mathbf{z}} \mathcal{L}_\rho &= \mathcal{L}_\rho(\mathbf{b}^{k+1}, \mathbf{z}^k, \mathbf{y}^k) - \mathcal{L}_\rho(\mathbf{b}^{k+1}, \mathbf{z}^{k+1}, \mathbf{y}^k), \\
\Delta_{\mathbf{y}} \mathcal{L}_\rho &= \mathcal{L}_\rho(\mathbf{b}^{k+1}, \mathbf{z}^{k+1}, \mathbf{y}^k) - \mathcal{L}_\rho(\mathbf{b}^{k+1}, \mathbf{z}^{k+1}, \mathbf{y}^{k+1}),
\end{aligned}$$

respectively. Next we show that the augmented Lagrangian decreases sufficiently after the first ADMM step (20).

Lemma 4. *If $\rho \geq \frac{2}{h^3} L_f$, then the update $\mathbf{b}^k \to \mathbf{b}^{k+1}$ does not increase the augmented Lagrangian. More precisely,*

$$\Delta_{\mathbf{b}} \mathcal{L}_\rho \geq \left(\frac{\rho h^3}{2} - L_f \right) \left\| \mathbf{b}^k - \mathbf{b}^{k+1} \right\|^2.$$

Proof. Denote $\mathbf{d}^{k+1} = \mathbf{y}^{k+1} + \rho h^3 (\mathbf{z}^{k+1} - \mathbf{z}^k)$ and for brevity $\nabla f^{k+1} = \nabla f(\mathbf{b}^{k+1})$. From the optimality condition for (20) we conclude that $\mathbf{b}^{k+1}$ satisfies

$$\begin{aligned}
& \mathbf{0} \in \nabla f^{k+1} + \partial \iota_C(\mathbf{b}^{k+1}) + \mathbf{y}^k + \rho h^3 (\mathbf{b}^{k+1} - \mathbf{z}^k) \\
\Leftrightarrow \quad & \mathbf{0} \in \nabla f^{k+1} + \partial \iota_C(\mathbf{b}^{k+1}) + \mathbf{y}^{k+1} + \rho h^3 (\mathbf{z}^{k+1} - \mathbf{z}^k) \\
\Leftrightarrow \quad & -\left(\nabla f^{k+1} + \mathbf{d}^{k+1} \right) \in \partial \iota_C(\mathbf{b}^{k+1}) \\
\Leftrightarrow \quad & \left(\nabla f^{k+1} + \mathbf{d}^{k+1} \right)^\top \left(\mathbf{w} - \mathbf{b}^{k+1} \right) \geq 0 \quad \forall \mathbf{w} \in C.
\end{aligned} \tag{31}$$

Here, we used that the subgradient of an indicator function of a convex set C is

$$\partial \iota_C(\mathbf{b}^{k+1}) = \left\{ \mathbf{d} \in \mathbb{R}^n : \mathbf{d}^\top (\mathbf{w} - \mathbf{b}^{k+1}) \leq 0, \, \forall \mathbf{w} \in C \right\},$$

see also [23, Ch. 8]. Then we get

$$\begin{aligned}
\Delta_{\mathbf{b}} \mathcal{L}_\rho &= f(\mathbf{b}^k) - f(\mathbf{b}^{k+1}) + (\mathbf{y}^k)^\top (\mathbf{b}^k - \mathbf{b}^{k+1}) \\
&\quad + \frac{\rho h^3}{2} \left(\left\| \mathbf{b}^k - \mathbf{z}^k \right\|^2 - \left\| \mathbf{b}^{k+1} - \mathbf{z}^k \right\|^2 \right) \\
&= f(\mathbf{b}^k) - f(\mathbf{b}^{k+1}) + (\mathbf{d}^{k+1})^\top (\mathbf{b}^k - \mathbf{b}^{k+1}) \\
&\quad + \frac{\rho h^3}{2} \left\| \mathbf{b}^k - \mathbf{b}^{k+1} \right\|^2 \\
&= f(\mathbf{b}^k) - f(\mathbf{b}^{k+1}) - (\nabla f^{k+1})^\top (\mathbf{b}^k - \mathbf{b}^{k+1}) \\
&\quad + \left(\nabla f^{k+1} + \mathbf{d}^{k+1} \right)^\top (\mathbf{b}^k - \mathbf{b}^{k+1}) \\
&\quad + \frac{\rho h^3}{2} \left\| \mathbf{b}^k - \mathbf{b}^{k+1} \right\|^2 \\
&\geq \left(\frac{\rho h^3}{2} - L_f \right) \left\| \mathbf{b}^k - \mathbf{b}^{k+1} \right\|^2,
\end{aligned}$$

where we used (31) and the Lipschitz continuity of ∇f in the last step. □

A similar result holds for the second ADMM step (21).

Lemma 5. *For any $\rho > 0$, the update $\mathbf{z}^k \to \mathbf{z}^{k+1}$ does not increase the value of the augmented Lagrangian. More precisely,*

$$\Delta_{\mathbf{z}} \mathcal{L}_\rho \geq \frac{\rho h^3}{2} \left\| \mathbf{z}^k - \mathbf{z}^{k+1} \right\|^2.$$

Proof. From the optimality condition for (21) we conclude that $\mathbf{z}^{k+1}$ satisfies

$$\begin{aligned}
0 &= \nabla g(\mathbf{z}^{k+1}) - \mathbf{y}^k - \rho h^3 (\mathbf{b}^{k+1} - \mathbf{z}^{k+1}) \\
&= \nabla g(\mathbf{z}^{k+1}) - \rho h^3 \mathbf{y}^{k+1}
\end{aligned}$$

and therefore $\mathbf{y}^{k+1} = \nabla g(\mathbf{z}^{k+1})$. Hence, we get

$$\begin{aligned}
\Delta_{\mathbf{z}} \mathcal{L}_\rho &= g(\mathbf{z}^k) - g(\mathbf{z}^{k+1}) - (\mathbf{y}^k)^\top (\mathbf{z}^k - \mathbf{z}^{k+1}) \\
&\quad + \frac{\rho h^3}{2} \left(\left\| \mathbf{b}^{k+1} - \mathbf{z}^k \right\|^2 - \left\| \mathbf{b}^{k+1} - \mathbf{z}^{k+1} \right\|^2 \right) \\
&= g(\mathbf{z}^k) - g(\mathbf{z}^{k+1}) - (\mathbf{y}^{k+1})^\top (\mathbf{z}^k - \mathbf{z}^{k+1}) \\
&\quad + \frac{\rho h^3}{2} \left\| \mathbf{z}^k - \mathbf{z}^{k+1} \right\|^2 \\
&= g(\mathbf{z}^k) - g(\mathbf{z}^{k+1}) - \nabla g(\mathbf{z}^{k+1})^\top (\mathbf{z}^k - \mathbf{z}^{k+1}) \\
&\quad + \frac{\rho h^3}{2} \left\| \mathbf{z}^k - \mathbf{z}^{k+1} \right\|^2 \\
&\geq \frac{\rho h^3}{2} \left\| \mathbf{z}^k - \mathbf{z}^{k+1} \right\|^2
\end{aligned}$$

where we used the convexity of g in the last step. □

We now show an analogous result for the third ADMM step (22).

Lemma 6. *For any $\rho > 0$, the update $\mathbf{y}^k \to \mathbf{y}^{k+1}$ does not increase the value of the augmented Lagrangian, i.e. we have $\Delta_{\mathbf{y}} \mathcal{L}_\rho \geq 0$.*

Proof. This directly follows from

$$\begin{aligned}\Delta_{\mathbf{y}}\mathcal{L}_\rho &= (\mathbf{y}^k - \mathbf{y}^{k+1})^\top(\mathbf{b}^{k+1} - \mathbf{z}^{k+1}) \\ &= \rho h^3 \left\|\mathbf{b}^{k+1} - \mathbf{z}^{k+1}\right\|^2 \geq 0. \end{aligned}$$ □

Having established the above results, we can now verify that our problem satisfies the three properties in Theorem 2.

Proof of Theorem 2. We note that (P2), the sufficient decrease, follows immediately from Lemmas 4 to 6 with the constant $C_1 = \frac{\rho h^3}{2} - L_f > 0$. For brevity we denote $\mathcal{L}_\rho^k = \mathcal{L}_\rho(\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k)$ and $\mathcal{L}_\rho^{k+1} = \mathcal{L}_\rho(\mathbf{b}^{k+1}, \mathbf{z}^{k+1}, \mathbf{y}^{k+1})$.

(P1) The well-definedness of the updates $\mathbf{b}^k \to \mathbf{b}^{k+1}$ and $\mathbf{z}^k \to \mathbf{z}^{k+1}$ and thus also the update $\mathbf{y}^k \to \mathbf{y}^{k+1}$ follows from Lemma 3. We next show that the sequence $\{\mathcal{L}(\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k)\}_k$ is bounded below. First, note that for each iteration k we have

$$f(\mathbf{b}^k) + g(\mathbf{b}^k) \geq 0.$$

Using that $\mathbf{y}^k = \nabla g(\mathbf{z}^k)$ for all k and the Lipschitz continuity of ∇g we note that

$$\left\|\mathbf{y}^{k+1} - \mathbf{y}^k\right\| \leq L_g \left\|\mathbf{z}^{k+1} - \mathbf{z}^k\right\|. \tag{32}$$

Combining this with $\rho h^3 \geq 2L_g$ we obtain

$$\begin{aligned}\mathcal{L}_\rho^k &= f(\mathbf{b}^k) + g(\mathbf{z}^k) + (\mathbf{y}^k)^\top(\mathbf{b}^k - \mathbf{z}^k) \\ &\quad + \frac{\rho h^3}{2}\left\|\mathbf{b}^k - \mathbf{z}^k\right\|^2 \\ &= f(\mathbf{b}^k) + g(\mathbf{b}^k) + g(\mathbf{z}^k) - g(\mathbf{b}^k) \\ &\quad - \nabla g(\mathbf{z}^k)^\top(\mathbf{z}^k - \mathbf{b}^k) + \frac{\rho h^3}{2}\left\|\mathbf{b}^k - \mathbf{z}^k\right\|^2 \\ &\geq f(\mathbf{b}^k) + g(\mathbf{b}^k) + \left(\frac{\rho h^3}{2} - L_g\right)\left\|\mathbf{b}^k - \mathbf{z}^k\right\|^2 \\ &\geq 0, \end{aligned}\tag{33}$$

which shows that $\left\{\mathcal{L}_\rho(\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k)\right\}_k$ is lower bounded by zero. Finally, we need to show the boundedness of $\{\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k\}_k$. For $\{\mathbf{b}^k\}_k$ this directly follows from the boundedness of C. From property (P2) we know that $\mathcal{L}_\rho(\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k)$ is monotonically decreasing and therefore bounded above by $\mathcal{L}_\rho(\mathbf{b}^0, \mathbf{z}^0, \mathbf{y}^0)$. Using this, the boundedness of $\{\mathbf{b}^k\}_k$, (33), and the lower boundedness of $\{f(\mathbf{b}^k) + g(\mathbf{b}^k)\}_k$, we conclude the boundedness of $\{\mathbf{z}^k\}_k$. Finally, again using $\mathbf{y}^k = \nabla g(\mathbf{z}^k)$ and the Lipschitz continuity of ∇g, we have

$$\begin{aligned}\left\|\mathbf{y}^k\right\| &= \left\|\nabla g(\mathbf{z}^k)\right\| \\ &\leq \left\|\nabla g(\mathbf{z}^k) - \nabla g(\mathbf{0})\right\| + \left\|\nabla g(\mathbf{0})\right\| \\ &\leq L_g \left\|\mathbf{z}^k\right\| + \left\|\nabla g(\mathbf{0})\right\|, \end{aligned}$$

which shows the boundedness of $\{\mathbf{y}^k\}_k$.

(P3) We need to bound the derivatives of $\mathcal{L}_\rho$. First, we note that

$$\begin{aligned}\partial_{\mathbf{b}}\mathcal{L}_\rho^{k+1} &= \nabla f(\mathbf{b}^{k+1}) + \partial\iota_C(\mathbf{b}^{k+1}) \\ &\quad + \mathbf{y}^{k+1} + \rho h^3(\mathbf{b}^{k+1} - \mathbf{z}^{k+1}) \\ &= \nabla f(\mathbf{b}^{k+1}) + \partial\iota_C(\mathbf{b}^{k+1}) \\ &\quad + \mathbf{y}^k + \rho h^3(\mathbf{b}^{k+1} - \mathbf{z}^k) \\ &\quad + \mathbf{y}^{k+1} - \mathbf{y}^k + \rho h^3(\mathbf{z}^k - \mathbf{z}^{k+1}). \end{aligned}$$

The optimality condition of the first ADMM step (20) implies $\mathbf{0} \in f(\mathbf{b}^{k+1}) + \partial\iota_C(\mathbf{b}^{k+1}) + \mathbf{y}^k + \rho h^3(\mathbf{b}^{k+1} - \mathbf{z}^k)$ and thus

$$\mathbf{y}^{k+1} - \mathbf{y}^k + \rho h^3(\mathbf{z}^k - \mathbf{z}^{k+1}) \in \partial_{\mathbf{b}}\mathcal{L}_\rho(\mathbf{b}^{k+1}, \mathbf{z}^{k+1}, \mathbf{y}^{k+1}),$$

which has a norm bounded by $(L_g + \rho h^3)\left\|\mathbf{z}^k - \mathbf{z}^{k+1}\right\|$ due to (32). Second, we note that

$$\begin{aligned}\left\|\partial_{\mathbf{z}}\mathcal{L}_\rho^{k+1}\right\| &= \left\|\nabla g(\mathbf{z}^{k+1}) - \mathbf{y}^{k+1} - \rho h^3(\mathbf{b}^{k+1} - \mathbf{z}^{k+1})\right\| \\ &= \left\|\mathbf{y}^k - \mathbf{y}^{k+1}\right\| \\ &\leq L_g \left\|\mathbf{z}^k - \mathbf{z}^{k+1}\right\|. \end{aligned}$$

Finally, we note that

$$\begin{aligned}\left\|\partial_{\mathbf{y}}\mathcal{L}_\rho^{k+1}\right\| &= \left\|\mathbf{b}^{k+1} - \mathbf{z}^{k+1}\right\| \\ &= \left\|\frac{1}{\rho h^3}(\mathbf{y}^{k+1} - \mathbf{y}^k)\right\| \\ &\leq \frac{L_g}{\rho h^3}\left\|\mathbf{z}^k - \mathbf{z}^{k+1}\right\|. \end{aligned}$$

Therefore, with constant $C_2 = \max\{\frac{1}{2}, 3L_g\}$ there exists a subgradient $\mathbf{d}^{k+1} \in \partial\mathcal{L}_\rho(\mathbf{b}^{k+1}, \mathbf{z}^{k+1}, \mathbf{y}^{k+1})$ as claimed. □

Finally, we conclude by proving the main result, which is done exactly as in [33] by using the properties (P1)–(P3).

Proof of Theorem 1. In Theorem 2 we have established that the properties (P1)–(P3) hold for the iterates generated by (20) to (22), provided that the augmentation parameter satisfies $\rho > \frac{2}{h^3}\max\{L_f, L_g\}$. From (P1) we know that the set of iterates $\{\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k\}_k$ is bounded, so it has a convergent subsequence. We denote a limit point by $\{\mathbf{b}^*, \mathbf{z}^*, \mathbf{y}^*\}$. Also from (P1) we know that the sequence $\left\{\mathcal{L}_\rho(\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k)\right\}_k$ is bounded below. By (P2) it is also monotonically and sufficiently decreasing and this implies $\left\|\mathbf{b}^k - \mathbf{b}^{k+1}\right\| \to 0$ and $\left\|\mathbf{z}^k - \mathbf{z}^{k+1}\right\| \to 0$. Finally, by (P3), we get that there exists a subgradient $\mathbf{d}^k \in \partial\mathcal{L}_\rho(\mathbf{b}^k, \mathbf{z}^k, \mathbf{y}^k)$ with $\left\|\mathbf{d}^k\right\| \to 0$, which shows that $\mathbf{0} \in \partial\mathcal{L}_\rho(\mathbf{b}^*, \mathbf{z}^*, \mathbf{y}^*)$, and thus $\{\mathbf{b}^*, \mathbf{z}^*, \mathbf{y}^*\}$ is a stationary point. □

The lower bound for ρ depends on the Lipschitz constants L_f and L_g of the distance and regularization function, respectively, as well as the voxel size h. The Lipschitz constants are commonly not available in practice. In our numerical experiments in Section 5 we use a modified version of the adaptive augmentation parameter choice described in [2], which ensures that the augmentation parameter ρ remains larger than an experimentally defined lower bound $\rho_{\min}$, which we chose equal for all steps in the multilevel optimization. We compare this parameter choice method with a constant choice of the augmentation parameter and the unmodified adaptive scheme proposed for convex problems in [2].

4.3 Multilevel Strategy

As common in image registration and also suggested in [24], we employ a multilevel approach for solving (6); see [20] for details. In a nutshell, we start by solving a discrete version of (6) on a relatively coarse grid. Then, we prolongate the estimated field inhomogeneity to a finer grid to serve as a starting guess for the next discrete optimization problem. On each level, the resolution of the image data is increased as well and the procedure is repeated until the desired resolution is achieved. Apart from reducing computational costs on a coarse grid and obtaining excellent starting guesses, multilevel approaches have been observed to be more robust against local minima, which are less likely to occur in the coarse grid discretization; see Figure 3 for an example.

For ADMM there are several options how to initialize the optimization on a finer grid discretization, given the results on the coarser grid. We have tested and compared three strategies:

1. Prolongate all three coarse mesh vectors $\mathbf{b}_c^{k+1}$, $\mathbf{z}_c^{k+1}$, and $\mathbf{u}_c^{k+1}$ and use the resulting fine mesh vectors $\mathbf{b}_f$, $\mathbf{z}_f$, and $\mathbf{u}_f$ as initial guesses for the next level.
2. Restart ADMM using the fine mesh variable $\mathbf{b}_f$ for both $\mathbf{z}^0$ and $\mathbf{b}^0$ and set the dual variable to zero, i.e., $\mathbf{u}^0 = \mathbf{0}$.
3. Restart ADMM on the fine level using an average $\mathbf{b}^0 = \mathbf{z}^0 = \frac{1}{2}(\mathbf{b}_f + \mathbf{z}_f)$ and $\mathbf{u}^0 = \mathbf{0}$ as initial guesses.

In our examples, we obtained comparable results for all three strategies, however, the third strategy performed best and is used in the subsequent experiments.

5 Numerical Experiments

In this section, we perform numerical experiments to demonstrate the effectiveness and the performance of the proposed ADMM method using high-resolution 3D data acquired using a state-of-the-art MRI scanner. We conclude the section by comparing the novel method to two existing state-of-the-art methods for susceptibility artifact correction. A MacBook Pro laptop with 2.8 GHz Intel Core i7 processor and 16 GB 1600 MHz DD3 memory running MATLAB 2016B is used for all numerical experiments.

Test Data. The 3D data sets are provided by the Human Connectome Project [30]. We use several pairs of unprocessed $b = 0$ diffusion-weighted images of female and male subjects aged 26-35 (cf. Table 1 for subject IDs) acquired using reversed phase-encoding directions on a 7T scanner. The image data is obtained using a commonly used single-shot spin-echo EPI sequence. In all cases the rectangular field of view is 210 mm $\times$ 210 mm $\times$ 138.5 mm and the voxel size is 1.05 mm in all spatial directions. We use a three-level multilevel strategy using grid sizes of $50 \times 50 \times 33$, $100 \times 100 \times 66$, and $200 \times 200 \times 132$.

Performance of ADMM. We study the performance of the proposed ADMM scheme and its dependence on the choice of the augmentation parameter ρ using one of the 3D data sets described above (subject ID 111312). The convergence of the primal residual, dual residual, and ρ for the adaptive method, the fixed parameter method, and the adaptive method with lower bound, are visualized in Figure 4. In the fixed case we use $\rho = 10^2$, which is also the lower bound in the bounded adaptive method. For both adaptive schemes the initial parameter is $\rho^0 = 10^6$. As discussed in Section 4.3, we average $\mathbf{b}$ and $\mathbf{z}$ after prolongation to start the next level. Clearly, larger values for ρ result in a smaller primal residual, whereas smaller values for ρ result in a smaller dual residual. The benefit of the adaptive method with lower bound is, that it forces the primal residual to converge fast during the first iterations and afterwards keeps the primal residual small while the dual residual converges as well. This way the equality constraint in (18) is almost satisfied during most iterations, meaning that we indeed solve the problem (18).

Note, that even though ρ changes dramatically on the coarsest level, it remains constant on the finer levels, where computations are more costly. The separability of the first – and in our experience most expensive – ADMM subproblem (20) provides a way for substantial speed up by using parallel computing. In principle, the data of each image column can be processed completely in parallel. To strike a balance between communication overhead and computations, in our current MATLAB implementation we correct all image slices in parallel.

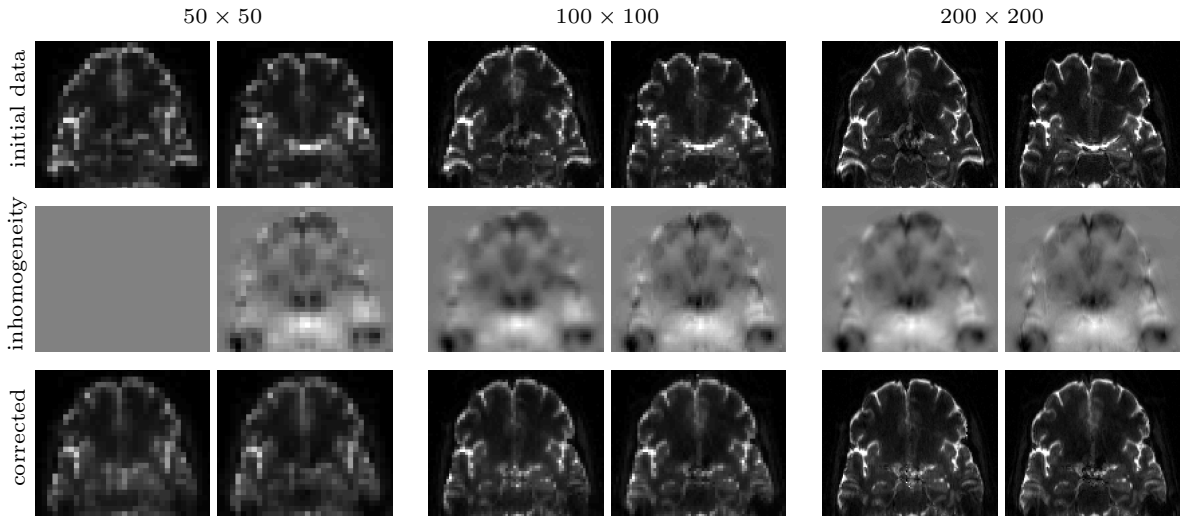

Fig. 3 Multilevel susceptibility artifact correction for EPI-MRI brain data provided by the Human Connectome Project [30] (subject ID 111312); see also Section 5. Deformed measurements $\mathcal{I}_{+v}$ and $\mathcal{I}_{-v}$ (top row), estimated field inhomogeneity (middle row) and corrected images (bottom row) are visualized for three different discretization levels (coarse to fine, left to right). Image data is shown in pairs corresponding to the different phase-encoding directions. Left plots for the field inhomogeneity visualize the starting guesses and right plots the refined solutions on each level.

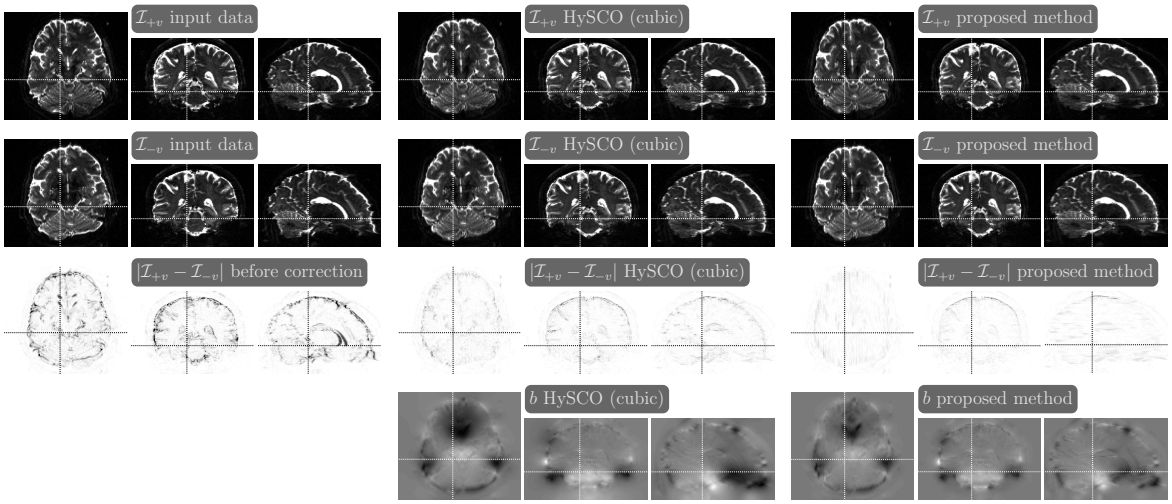

Fig. 5 3D correction results for high-resolution EPI-MRI brain data provided by the Human Connectome Project [30] (subject ID 111312); see also Section 5. Initial data (left) and results of the two for this data set most accurate reversed-gradient based correction methods are visualized using orthogonal slice views. The color axis are chosen identically in each row. First and second rows show initial data and corrected images for both phase-encoding directions. Third row shows the absolute difference between both images using an inverted color scale for improved visualization. A considerable improvement of image similarity can be observed for both correction methods, however the proposed method leads to a slightly smaller residual. The bottom row shows the estimated field inhomogeneity that is comparable among the correction techniques.

Comparison. We compare the quality of the proposed correction method to two established state-of-the-art methods for susceptibility artifact correction for thirteen 3D data sets provided by the Human Connectome Project [30]. Exemplarily, we consider HySCO [26] since it is based on the same variational formulation and uses the same three-level approach as our proposed method, and Topup [1] as implemented in the FMRIB Software Library (FSL) [28] since it is part of the standard processing pipeline for diffusion-weighted images used by the Human Connectome Project. In contrast to the proposed method, HySCO uses a nodal discretization, which leads to coupling across slices and image columns (see second row in Figure 2). HySCO uses a penalty term $P(\mathbf{b})$ scaled by a parameter β to enforce the invertibility constraints and a Gauss-Newton-PCG itera-

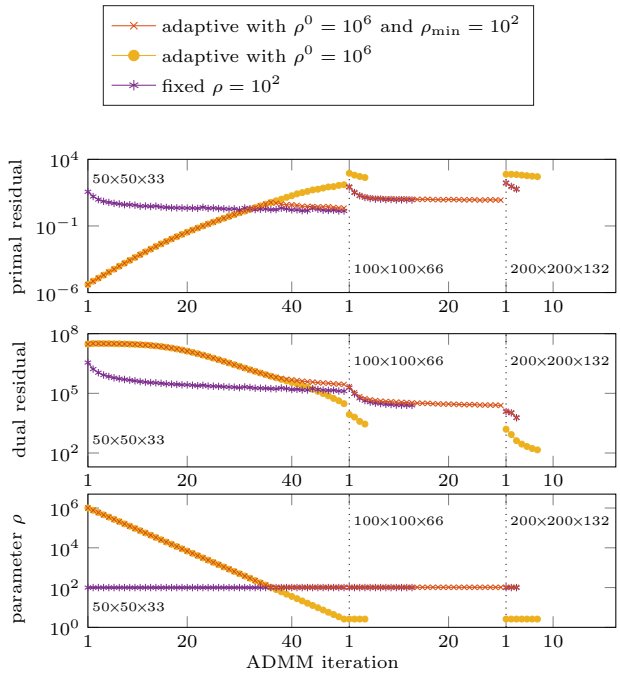

Fig. 4 Convergence of ADMM for a multilevel 3D EPI susceptibility artifact correction. Two adaptive and one fixed strategies for choosing the augmentation parameter ρ in (19) are compared. First row shows the norm of the primal residual (cf. (29)) for each iteration of ADMM and each level in the coarse to fine hierarchy. Second row shows the norm of the dual residual (cf. (30)) and the bottom row shows the augmentation parameter for each iteration. The regularization parameter for the smoothing term is $\alpha = 50$.

tion to solve the discrete optimization problem. In our numerical experiments we use the default penalty parameter of $\beta = 10$. We provide results for two variants of HySCO. First, we report results using the default interpolation settings in HySCO. By default, HySCO uses a smoothed cubic spline based approximation of the data described in [20], which aims at adding robustness to image noise. Since this comes at additional computational costs, results for a second variant using the same linear interpolation model also used in the proposed ADMM method are provided. For Topup we choose the same settings as used in the processing pipeline of the Human Connectome Project. Quantitative results are summarized in Table 1. All three methods effectively correct for susceptibility artifacts improving the similarity of the corrected image pairs. However, in all cases the proposed method outperforms the other methods in terms of improvement in the measure of image similarity and runtime. The parallelized implementation of ADMM yields an average speedup factor of around 30x as compared to the default settings of HySCO, and an average speedup factor of around 60x as compared to

Table 1 Comparison of improvements in image similarity and runtime for different reversed-gradient methods (HySCO [26] with cubic interpolation (c) and linear interpolation (l), Topup [1], and the novel ADMM method) applied to several 3D-MRI data provided by the Human Connectome Project [30] with $\alpha = 50$ and $\beta = 10$ (HySCO only). Image improvement is assessed using the reduction in the sum of squared differences (SSD) distance [20] between the initial and transformed blip-up and blip-down data. The speedup factors of the novel ADMM method over both HySCO variants and Topup are shown in brackets next to the corresponding runtimes.

subject	SSD reduction & runtime (ADMM speedup factor)			
ID	HySCO (l)	HySCO (c)	Topup [1]	ADMM
102816	71.3% 114s (2.1x)	75.5% 926s (16.8x)	82.4% 3617s (65.8x)	95.4% 55s (1.0x)
111312	78.9% 198s (4.6x)	87.1% 2059s (47.9x)	83.0% 3810s (88.6x)	93.1% 43s (1.0x)
156334	90.8% 118s (2.2x)	92.7% 1648s (31.1x)	85.3% 3573s (67.4x)	96.8% 53s (1.0x)
176542	89.9% 142s (2.7x)	91.7% 3023s (57.0x)	85.6% 3705s (69.9x)	97.0% 53s (1.0x)
199655	79.0% 137s (2.4x)	82.9% 900s (15.8x)	87.1% 3703s (65.0x)	97.8% 57s (1.0x)
221319	88.6% 148s (2.9x)	90.3% 1859s (36.5x)	83.8% 3703s (72.6x)	95.1% 51s (1.0x)
246133	88.5% 118s (2.2x)	90.8% 2504s (46.4x)	83.9% 3708s (68.7x)	96.1% 54s (1.0x)
251833	92.2% 115s (2.0x)	93.7% 1979s (34.7x)	85.5% 3700s (64.9x)	98.0% 57s (1.0x)
467351	86.5% 187s (2.7x)	89.7% 1804s (26.1x)	83.4% 3706s (53.7x)	95.9% 69s (1.0x)
541943	86.1% 117s (1.6x)	88.1% 4911s (67.3x)	82.4% 3693s (50.6x)	93.4% 73s (1.0x)
680957	81.2% 100s (1.2x)	84.5% 3532s (43.1x)	84.9% 3649s (44.5x)	91.1% 82s (1.0x)
899885	90.7% 104s (1.8x)	92.5% 1697s (29.8x)	87.4% 3804s (66.7x)	96.3% 57s (1.0x)
951457	71.0% 143s (3.0x)	76.5% 765s (16.3x)	82.8% 3688s (78.5x)	96.2% 47s (1.0x)
average	84, 2% 134s (2.3x)	87, 4% 2124s (31.6x)	84, 4% 3697s (63.7x)	95, 6% 58s (1.0x)

FSL Topup. For one data set (subject ID 111312) the 3D correction results are also visualized in Figure 5.

A notable difference is observed comparing the performance of HySCO with cubic B-spline and linear interpolation of the image data. While the former achieves a higher quality correction, the average time-to-solution is around 35 minutes. With an average runtime of less than 3 minutes, the linear interpolation model is considerably faster, but it achieves considerably inferior correction results. Changing from the nodal discretization employed in HySCO to the face-staggered discretization proposed here, leads to a highly accurate correction (see improvement in image similarity), even with a linear interpolation model.

Anatomical Reference. Finally, for a 3D data set (subject ID 102816) we compare the correction results of

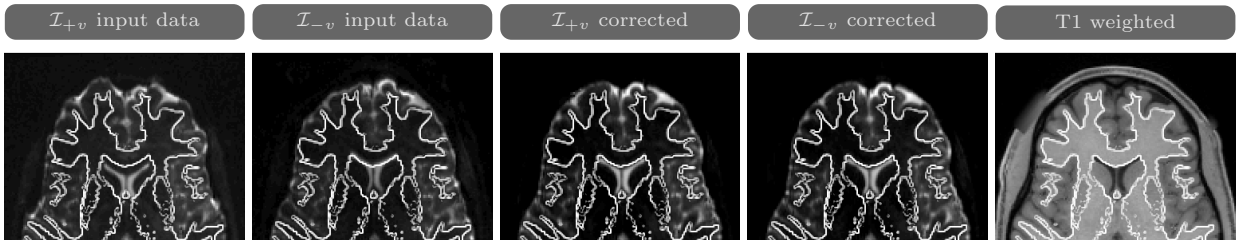

Fig. 6 Visualization of the image improvement obtained with the novel ADMM method. The initial and corrected blip-up and blip-down data (first four, left to right) for one 3D-MRI data set provided by the Human Connectome Project [30] (subject ID 102816) are independently registered to a corresponding T1 weighted image (right) using only rigid-body transforms. A zoomed in frontal brain region in one axial slice with an overlay of the white matter contours obtained from a segmentation of the T1 weighted image is shown for improved visibility.

the proposed method with an anatomical T1 weighted reference images of the same subject. To this end we register the uncorrected and corrected reversed gradient image pairs to the T1 weighted image by rigid-body transforms using the Statistical Parametric Mapping Toolbox (SPM) [6]. The results can be seen in Figure 6. For improved visualization we add an overlay of the white matter contours, which are obtained from a SPM segmentation of the T1 weighted image, and show only the frontal brain region of a single axial slice.

6 Summary And Conclusion

In this paper, we present an efficient method for susceptibility artifact correction of EPI-MRI. We consider a variational formulation of a reversed gradient based correction scheme similar to [24,26,13]. Our method requires one additional EPI-MRI acquisition with opposite phase-encoding direction and, thus, opposite distortion.

We follow a discretize-then-optimize paradigm and propose a face-staggered discretization of the field inhomogeneity. This choice leads to a separable discrete distance function and constraints. While the overall optimization problem is, due to the smoothness regularizer, not separable we consider an optimization scheme that exploit the partially separable structure. We split the separable and non-separable parts of the objective function by adding an artificial variable and apply ADMM to compute a saddle-point of the associated augmented Lagrangian. The resulting subproblems can be solved efficiently. The first subproblem, albeit being non-convex, is separable and can be broken down into several smaller problems with only a few hundreds of unknowns that can be solved in parallel using sequential quadratic programming. The second subproblem consists of minimizing a convex quadratic with a block-diagonal Hessian with BTTB structure and, thus, can be solved efficiently and in parallel using DCTs.

We provide a detailed convergence result that is similar to [33] but exploits the smoothness of our problem. We also derive a theoretical lower bound for the augmentation parameter in ADMM. Using a numerical experiment, we compare different adaption strategies for the augmentation parameter in Figure 4. In our experience, adaptive choice of the parameter is possible, however, in view of the presented convergence result, we recommend using at least an empirically tuned lower bound to ensure convergence. We found that choosing a sufficiently large lower bound is critical to ensure smoothness of the solution.

Both the correction quality and the time-to-solution of the novel method are superior to state-of-the-art methods as shown for high-resolution EPI-MRI examples in Figure 5 and Table 1. Average speedup factors of around 30x and 60x over HySCO [26] and Topup [1], respectively, are achieved in our experiments using thirteen high-resolution EPI-MRI datasets. Even further improvements of the efficiency of our method are possible using more advanced parallelization schemes. Here, the separable structure of the computationally first, and most challenging, subproblem in ADMM is used by parallelizing over all image slices. Given our promising results, the ADMM method is an ideal candidate for implementation on massively parallel hardware such as Graphics Processing Units (GPU) and can be attractive for real-time applications such as [4].

7 Acknowledgements

The 3D data is provided by the Human Connectome Project, WU-Minn Consortium (PIs: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. We also like to thank Siawoosh Mohammadi (University Hospital Hamburg Eppendorf, Ger-

many) and Harald Kugel (Department of Clinical Radiology, University Hospital Münster, Germany) for fruitful discussions and helpful suggestions on a first draft of this manuscript. This work is partially supported by the National Science Foundation (NSF) award DMS 1522599.

References

1. Andersson, J.L.R., Skare, S., Ashburner, J.: How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. NeuroImage **20**(2), 870–888 (2003). DOI 10.1016/S1053-8119(03)00336-7
2. Boyd, S., Parikh, N., Chu, E., Peleato, B., Eckstein, J.: Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers. Foundations and Trends® in Machine Learning **3**(1), 1–122 (2011). DOI 10.1561/2200000016
3. Chang, H., Fitzpatrick, J.M.: A Technique for Accurate Magnetic-Resonance-Imaging in the Presence of Field Inhomogeneities. Medical Imaging, IEEE Transactions on **11**(3), 319–329 (1992). DOI 10.1109/42.158935
4. Daga, P., Pendse, T., Modat, M., White, M., Mancini, L., Winston, G.P., McEvoy, A.W., Thornton, J., Yousry, T., Drobnjak, I., Duncan, J.S., Ourselin, S.: Susceptibility artefact correction using dynamic graph cuts: Application to neurosurgery. Medical Image Analysis **18**(7), 1132–1142 (2014). DOI 10.1016/j.media.2014.06.008
5. Eckstein, J.: Augmented Lagrangian and alternating direction methods for convex optimization: A tutorial and some illustrative computational results. RUTCOR Research Reports, Piscataway, NJ (2012)
6. Friston, K.J.: Chapter 31 - Experimental Design and Statistical Parametric Mapping. In: R.S. Frackowiak, K.J. Friston, C.D. Frith, R.J. Dolan, C.J. Price, S. Zeki, J.T. Ashburner, W.D. Penny (eds.) Human Brain Function, 2nd edn., pp. 599 – 632. Academic Press, Burlington, MA (2004). DOI 10.1016/B978-012264841-0/50033-0
7. Gabay, D., Mercier, B.: A dual algorithm for the solution of nonlinear variational problems via finite element approximation. Computers & Mathematics with Applications **2**(1), 17–40 (1976). DOI 10.1016/0898-1221(76)90003-1
8. Glowinski, R.: On Alternating Direction Methods of Multipliers: A Historical Perspective. In: Modeling, Simulation and Optimization for Science and Technology, pp. 59–82. Springer Netherlands, Dordrecht (2014). DOI 10.1007/978-94-017-9054-3_4
9. Golub, G.H., Heath, M., Wahba, G.: Generalized cross-validation as a method for choosing a good ridge parameter. Technometrics **21**(2), 215–223 (1979). DOI 10.1080/00401706.1979.10489751
10. Haber, E., Oldenburg, D.: A GCV based method for nonlinear ill-posed problems. Computational Geosciences **4**, 41–63 (2000). DOI 10.1023/A:1011599530422
11. Hansen, P.C.: Rank-deficient and discrete ill-posed problems. SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (1998). DOI 10.1137/1.9780898719697
12. Hansen, P.C., Nagy, J.G., O'Leary, D.P.: Deblurring Images: Matrices, Spectra and Filtering. Matrices, Spectra, and Filtering. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2006). DOI 10.1137/1.9780898718874
13. Holland, D., Kuperman, J.M., Dale, A.M.: Efficient correction of inhomogeneous static magnetic field-induced distortion in Echo Planar Imaging. NeuroImage **50**(1), 175–183 (2010). DOI 10.1016/j.neuroimage.2009.11.044
14. Hong, X., To, X.V., Teh, I., Soh, J.R., Chuang, K.H.: Evaluation of EPI distortion correction methods for quantitative MRI of the brain at high magnetic field. Magnetic Resonance Imaging **33**(9), 1098–1105 (2015). DOI 10.1016/j.mri.2015.06.010
15. Irfanoglu, M.O., Modi, P., Nayak, A., Hutchinson, E.B., Sarlls, J., Pierpaoli, C.: DR-BUDDI (Diffeomorphic Registration for Blip-Up blip-Down Diffusion Imaging) method for correcting echo planar imaging distortions. NeuroImage **106**(C), 284–299 (2015). DOI 10.1016/j.neuroimage.2014.11.042
16. Jezzard, P.: Correction of geometric distortion in fMRI data. NeuroImage **62**(2), 648–651 (2012). DOI 10.1016/j.neuroimage.2011.09.010
17. Jezzard, P., Balaban, R.S.: Correction for geometric distortion in echo planar images from B0 field variations. Magnetic Resonance in Medicine **34**(1), 65–73 (1995). DOI 10.1002/mrm.1910340111
18. Le Bihan, D., Mangin, J.F., Poupon, C., Clark, C.A., Pappata, S., Molko, N., Chabriat, H.: Diffusion tensor imaging: Concepts and applications. Journal of Magnetic Resonance Imaging **13**(4), 534–546 (2001). DOI 10.1002/jmri.1076
19. Madai, V.I., Martin, S.Z., von Samson-Himmelstjerna, F.C., Herzig, C.X., Mutke, M.A., Wood, C.N., Thamm, T., Zweynert, S., Bauer, M., Hetzer, S., Günther, M., Sobesky, J.: Correction for Susceptibility Distortions Increases the Performance of Arterial Spin Labeling in Patients with Cerebrovascular Disease. Journal of Neuroimaging (2016). DOI 10.1111/jon.12331
20. Modersitzki, J.: FAIR: flexible algorithms for image registration, *Fundamentals of Algorithms*, vol. 6. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2009). DOI 10.1137/1.9780898718843
21. de Munck, J.C., Bhagwandien, R., Muller, S.H., Verster, F.C., Van Herk, M.B.: The computation of MR image distortions caused by tissue susceptibility using the boundary element method. Medical Imaging, IEEE Transactions on **15**(5), 620–627 (1996). DOI 10.1109/42.538939
22. Nocedal, J., Wright, S.: Numerical Optimization. Springer Series in Operations Research and Financial Engineering. Springer Science & Business Media, New York, NY (2006). DOI 10.1007/978-0-387-40065-5
23. Rockafeller, R.T., Roger, J.B.W.: Variational Analysis, *Grundlehren der mathematischen Wissenschaft*, vol. 317. Springer-Verlag Berlin Heidelberg, Berlin (1998, (3rd printing 2009)). DOI 10.1007/978-0-387-40065-5
24. Ruthotto, L., Kugel, H., Olesch, J., Fischer, B., Modersitzki, J., Burger, M., Wolters, C.H.: Diffeomorphic susceptibility artifact correction of diffusion-weighted magnetic resonance images. Physics in Medicine and Biology **57**(18), 5715–5731 (2012). DOI 10.1088/0031-9155/57/18/5715
25. Ruthotto, L., Modersitzki, J.: Non-linear Image Registration. In: Handbook of Mathematical Methods in Imaging, pp. 2005–2051. Springer New York, New York, NY (2015). DOI 10.1007/978-1-4939-0790-8_39
26. Ruthotto, L., Mohammadi, S., Heck, C., Modersitzki J, W.N.: Hyperelastic Susceptibility Artifact Correction of DTI in SPM. Bildverarbeitung für die Medizin (2013). DOI 10.1007/978-3-642-36480-8_60

27. Skare, S., Andersson, J.L.R.: Correction of MR image distortions induced by metallic objects using a 3D cubic B-spline basis set: Application to stereotactic surgical planning. Magnetic Resonance in Medicine **54**(1), 169–181 (2005). DOI 10.1002/mrm.20528
28. Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R.K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M.: Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage **23**, S208–S219 (2004). DOI 10.1016/j.neuroimage.2004.07.051
29. Stehling, M.K., Turner, R., Mansfield, P.: Echo-Planar Imaging - Magnetic-Resonance-Imaging in a Fraction of a Second. Science **254**(5028), 43–50 (1991). DOI 10.1126/science.1925560
30. Van Essen, D.C., Ugurbil, K., Auerbach, E., Barch, D., Behrens, T.E.J., Bucholz, R., Chang, A., Chen, L., Corbetta, M., Curtiss, S.W., Della Penna, S., Feinberg, D., Glasser, M.F., Harel, N., Heath, A.C., Larson-Prior, L., Marcus, D., Michalareas, G., Moeller, S., Oostenveld, R., Petersen, S.E., Prior, F., Schlaggar, B.L., Smith, S.M., Snyder, A.Z., Xu, J., Yacoub, E., Consortium, W.M.H.: The Human Connectome Project: A data acquisition perspective. NeuroImage **62**(4), 2222–2231 (2012). DOI 10.1016/j.neuroimage.2012.02.018
31. Vardal, J., Salo, R.A., Larsson, C., Dale, A.M., Holland, D., Groote, I.R., Bjørnerud, A.: Correction of B0-Distortions in Echo-Planar-Imaging-Based Perfusion-Weighted MRI. Journal of Magnetic Resonance Imaging **39**(3), 722–728 (2013). DOI 10.1002/jmri.24213
32. Vogel, C.R.: Computational Methods for Inverse Problems. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, Philadelphia, PA (2002). DOI 10.1137/1.9780898717570
33. Wang, Y., Yin, W., Zeng, J.: Global Convergence of ADMM in Nonconvex Nonsmooth Optimization. arXiv.org (2015)