On a Bilinear Optimization Problem in Parallel Magnetic Resonance Imaging

Christian Clason, Gregory von Winckel

Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, A-8010 Graz, Austria

Abstract

This work is concerned with the structure of bilinear minimization problems arising in recovering sub-sampled and modulated images in parallel magnetic resonance imaging. By considering a physically reasonable simplified model exhibiting the same fundamental mathematical difficulties, it is shown that such problems suffer from poor gradient scaling and non-convexity, which causes standard optimization methods to perform inadequately. A globalized quasi-Newton method is proposed which is able to reconstruct both image and the unknown modulations without additional a priori information. Thus the present paper serves as a first contribution toward understanding and solving such bilinear optimization problems.

Keywords: bilinear optimization, quasi-Newton method, medical imaging

1. Introduction

Magnetic resonance imaging (MRI) is a medical imaging method that employs radio pulse echoes to measure the hydrogen atom density, which allows the discrimination of different types of tissue. The spatial information is encoded, using a combination of gradient magnetic fields, in the phase and frequency of the time-dependent echo, which is then measured by coils surrounding the patient. A Fourier transform of the recorded signal will therefore yield an image of the investigated area (for a full discussion of the principles of MRI, see, e.g., [1, 2]). Mathematically, MRI can thus be thought of as direct measurement of the Fourier coefficients of the image. One of the major drawbacks of MRI in current practice is the speed of the image acquisition, since in principle each (discrete) Fourier coefficient $k(i, j)$ has to be acquired separately: Each coordinate pair (i, j) needs to be encoded by the gradient fields and measured by a separate radio pulse. By employing a phase instead of frequency encoding for one coordinate, a "line" of Fourier coefficients $k(i, \cdot)$ can be read out in parallel for every radio excitation. The standard approach for further speeding up the process acquires only a subset of these lines (e.g., every second, $k(2i, \cdot)$, or every fourth, $k(4i, \cdot)$). This, however, leads to aliasing, as the signal is now sampled below the Nyquist frequency (cf. Figure 1 and the above references; note that the field of view is effectively halved by the sub-sampling). As a remedy, parallel magnetic resonance imaging (PMRI) measures the radio echo using multiple independent coils, which are usually placed in a circle around the patient, in this way hoping to compensate for the lost information. Since these coils have only limited aperture compared to a single coil, the resulting measurements are non-uniformly modulated. It is, therefore, necessary to recover both the missing Fourier coefficients and the unknown modulations (the so-called sensitivities) from a set of modulated and aliased coil images.

Email addresses: christian.clason@uni-graz.at (Christian Clason), gregory.von-winckel@uni-graz.at (Gregory von Winckel)

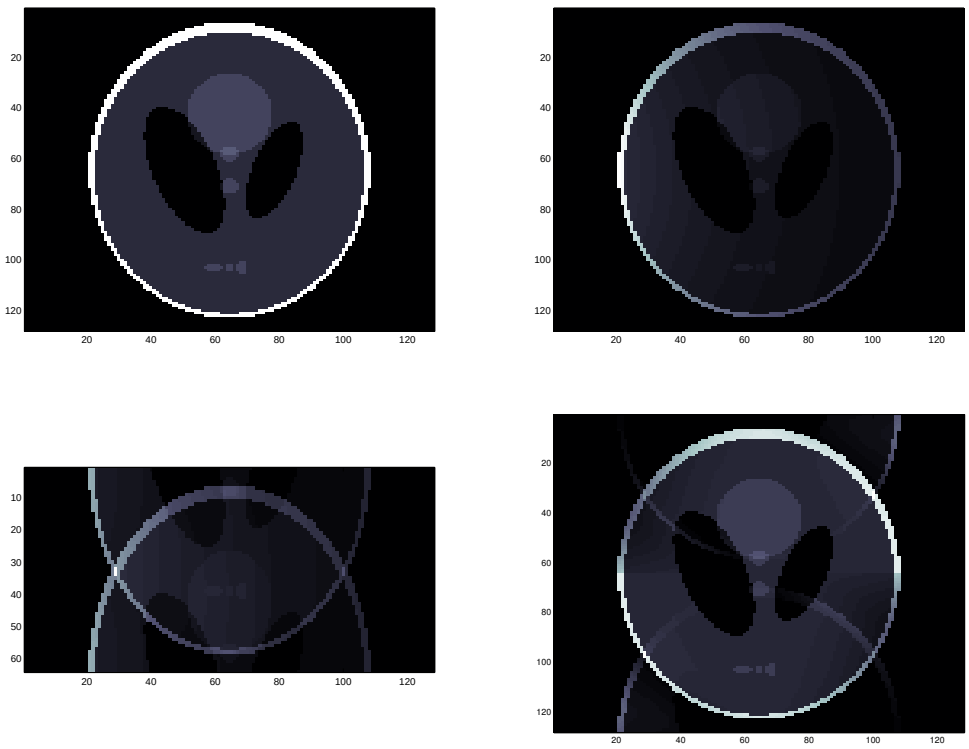

Figure 1: Illustration of modulation and aliasing applied to a phantom. From top left: True image; modulated coil image; modulated and sub-sampled coil image (factor 2, hence field of view is halved); naive reconstruction suffering from ghost artifacts.

Reconstruction strategies currently in use include SENSE [3], which is an algebraic linear least-squares recovery of the unaliased image using sensitivities taken from a fully sampled reference scan, and GRAPPA [4], where the missing Fourier coefficients are interpolated using an interpolation kernel fitted to additionally acquired lines around the zero frequency (in effect, a fully sampled low resolution image). The quality of reconstruction therefore depends on the number of these so-called center lines, which reduces the speed-up gained from sub-sampling. Recently, nonlinear least squares methods for PMRI have been proposed [5, 6], but these also rely on the presence of center lines. Common to all algebraic reconstruction techniques is the danger of ghost artifacts, which arise from incomplete separation of the superimposed parts in the aliased image (illustrated in Figure 1 for a naive Tikhonov-regularized least squares solution using known sensitivities).

Mathematically, the reconstruction problem can be formulated as a minimization problem:

$$\min_{u,\sigma_i} J(u, \sigma_1, \ldots, \sigma_n) = \sum_{j=1}^{n} \int_\Omega (K(\sigma_j u) - v_j)^2 \, dx. \tag{1}$$

Here, u denotes the image, $\sigma_1, \ldots, \sigma_n$ are the coil sensitivities, $v_1, \ldots, v_n$ stand for the measured coil images, and K is a subsampling operator (which will be made more precise in Section 2). The fundamental mathematical difficulty in parallel imaging is due to the bilinearity of the functional J: It holds that $J(u, \sigma_i) = J(u/p, \sigma_i p)$ for all functions $p \neq 0$, hence the problem is highly non-convex. The aim of this

work is to investigate the characteristics of this problem and present a strategy for its solution. Therefore, we consider a simplified model, which nonetheless exhibits the full difficulties inherent in the bilinear structure of parallel imaging, allowing us to concentrate on this central mathematical aspect of the problem.

It will be shown that a gradient descent scheme for this problem will be highly inefficient, due to bad scaling of the gradient components. However, since the functional J is highly non-convex, the Hessian will in general be severely rank deficient and even indefinite, making Newton's method also inapplicable. Therefore we employ a quasi-Newton method with limited storage, which—together with a suitable line search—guarantees a positive definite approximation of the Hessian in every step. To the authors' best knowledge, this is the first treatment of an optimization problem for such a bilinear functional, which is different from the optimization of bilinear systems as arising, e.g., in quantum control problems. Hence, our paper is a first contribution toward understanding and solving this challenging problem.

This paper is organized as follows. We give a precise statement of a sufficiently realistic model problem in Section 2. Section 3 deals with the solution of this minimization problem, where the optimality system is given (§3.1), the issues with poor scaling and non-convexity are discussed (§3.2), and a quasi-Newton method is proposed (§3.3). The details of the implementation are given in Section 4. Section 5 contains numerical results showing the efficacy of the proposed approach.

2. Problem Formulation

We wish to reconstruct the true image $u \in L^2([-1,1]^2)$, given n modulated and aliased coil images $v_1, \ldots, v_n \in L^2(\Omega)$. As the MRI method produces a sampled image with pixel-wise constant data, we will interpret the image to be reconstructed as a sum of characteristic functions on the square tensor product grid.

$$u(x,y) = \sum_{i,j} u_{ij}\chi_i(x)\chi_j(y), \quad \chi_i(x) = \begin{cases} 1 & \text{if } x_i < x < x_{i+1} \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $x_1, x_2, \ldots$ are the locations of the pixel boundaries in the x direction. To make the optimality system more readable, we shall leave the image as simply an L^2 function, but it should be understood for the numerical calculations as being constant on each pixel.

The exact sensitivity functions, which modulate the image, are determined by the coil geometry, orientation, and placement and are governed by the Biot-Savart law which relates the current through the coil to the magnetic field. Using the magneto-static approximation, a differential magnetic field $d\mathbf{B}$ will induce a constant current I in a differential element of wire with length $d\mathbf{l}$ according to the relationship

$$d\mathbf{B} = \frac{\mu_0 I}{4\pi} \frac{d\mathbf{l} \times \mathbf{r}}{r^3} \tag{3}$$

where μ_0 is the permeability of free space, $\mathbf{r}$ is the displacement vector between the point at which the magnetic field is given and the wire element, and $r = |\mathbf{r}|$. Since the details of the coil geometry and location are not provided with raw image data, it is necessary to develop a reasonable and tractable model for the sensitivity. Instead of incorporating the equations of electromagnetics and concerning ourselves with exact shapes of the coils, we utilize a few characteristics of the sensitivity to construct a model. We know that sensitivity is a very smooth function which is uniformly bounded and exhibits asymptotically algebraic decay. We also expect that if the coil has a relatively small aspect ratio, such as approximately square or

circular, then away from the coil, the sensitivity should be approximately radially isotropic. By this we mean that if the coil were at the center of a polar coordinate system, for sufficiently large radial distance, there should be little fluctuation in the sensitivity in the angular direction. We use these observations to model coil sensitivities σ_i by two-dimensional Lorentzian functions which exhibit the above behavior:

$$\sigma_i(x,y) = \frac{1}{c_{i,1}^2 + c_{i,2}^2((x-c_{i,3})^2 + (y-c_{i,4})^2)}, \tag{4}$$

where $\vec{c} = (c_{1,1}, \ldots, c_{1,4}, c_{2,1}, \ldots, c_{n,4})$ are parameters to be identified. While such a parameterization is of course inadequate to describe sensitivities occurring in real-world data, it is sufficient for the purpose of investigating the bilinear structure of the parallel imaging problem.

The effect of sub-sampling with an acceleration factor l is expressed by a linear convolution operator K:

$$Ku = \int_{-1}^{1}\int_{-1}^{1}\sum_{k=-\infty}^{\infty} \delta(\eta - lkT)\hat{u}(\xi,\eta)e^{-2\pi i(x\xi+y\eta)}\,d\xi d\eta, \tag{5}$$

where T is the ideal sampling interval according to Shannon's theorem and $\hat{u}$ is the two-dimensional inverse Fourier transform of u.

We therefore define for $\mu > 0$ the functional

$$J(u,\vec{c}) = \sum_{j=1}^{n}\int_\Omega (K(\sigma_j u) - v_j)^2\,dx + \mu\int_\Omega |\nabla u|\,dx, \tag{6}$$

n being again the number of coils, and pose the following minimization problem:

$$\min_{u\in L^2(\Omega),\vec{c}\in\mathbb{R}^{4n}} J(u,\vec{c}). \tag{P}$$

The regularization term penalizes the total variation of the image, which is known to preserve edges better than standard L^2- or H^1-regularization (cf. [7]). Due to the parameterization of the sensitivities, no additional regularization term (e.g., enforcing smoothness) for the σ_i is necessary.

Remark 2.1. *While the parameterized functional (6) is of course no longer bilinear in the strict sense, it still exhibits the multiplicative structure which makes the problem non-convex. Similarly, adding a regularization term is not sufficient to make the functional convex for reasonable values of μ. Showing this is the purpose of Section 3.2.*

3. Justification and description of the quasi-Newton method

3.1. Optimality system

A minimizer $(u^*, \vec{c}^*)$ of Problem (P) will satisfy the first order necessary optimality conditions

$$\nabla J(u^*,\vec{c}^*) = 0. \tag{7}$$

Straightforward formal differentiation yields the following explicit derivatives:

$$J_u(u,\vec{c}) = \sum_{i=1}^{n} \sigma_i K^*(K(\sigma_i u) - v_i) - \mu\nabla\cdot\left(\frac{\nabla u}{|\nabla u|}\right), \tag{8}$$

$$J_{c_{i,1}}(u,\vec{c}) = -2c_{i,1}\iint u K^*(K(\sigma_i u) - v_i)\sigma_i^2\, dxdy, \tag{9}$$

$$J_{c_{i,2}}(u,\vec{c}) = -2c_{i,2}\iint u K^*(K(\sigma_i u) - v_i)\sigma_i^2((x-c_{i,3})^2 + (y-c_{i,4})^2)\, dxdy, \tag{10}$$

$$J_{c_{i,3}}(u,\vec{c}) = 2c_{i,2}^2\iint u K^*(K(\sigma_i u) - v_i)\sigma_i^2(x-c_{i,3})\, dxdy, \tag{11}$$

$$J_{c_{i,4}}(u,\vec{c}) = 2c_{i,2}^2\iint u K^*(K(\sigma_i u) - v_i)\sigma_i^2(y-c_{i,4})\, dxdy. \tag{12}$$

Here, K^* denotes the adjoint of the linear operator K with respect to the inner product in L^2.

3.2. Problem characterization

A basic approach for the minimization of (6) would be a gradient descent scheme for $x_n = (u_n, \vec{c}_n)$:

$$x_{n+1} = x_n - a\nabla J(x_n), \tag{13}$$

where a is determined from a line search. However, such schemes can perform very poorly if the optimal step length is too small. This is true in the considered problem, due to bad scaling of the different components of the gradient. Because of the inherently bilinear structure of the objective functional, this can be demonstrated by partitioning the gradient into an image component J_u given by (8) and a sensitivity component $J_c = (J_{c_{1,1}}, J_{c_{1,2}}, \ldots, J_{c_{n,4}})^T$ given by (9)–(12). If we selectively perform a descent step in each of these two directions independently, the local minimum along the search directions occurs at widely disparate step lengths.

To make this more precise, we introduce a bivariate functional which is equal to the full functional on a two-dimensional manifold. Let the two components of the steepest descent direction be denoted as $p_u = -J_u(u,\vec{c})$ and $p_c = -J_c(u,\vec{c})$. The cost along these two directions is

$$j(a,b) = J(u_n + ap_u, \vec{c}_n + bp_c) \tag{14}$$

The gradient of $j(a,b)$ and gives us a simpler view of the full gradients' scaling. It consists of the following two components:

$$j_a(a,b) = p_u^\top J_u(u_n + ap_u, \vec{c}_n + bp_c), \quad j_b(a,b) = p_c^\top J_c(u_n + ap_u, \vec{c}_n + bp_c) \tag{15}$$

Now a large aspect ratio of the minimizer (a^*, b^*) of (14) would imply that the gradient of the full problem is poorly scaled. Hence, a line search in the steepest descent direction of the full functional (6)—which is equivalent to minimizing the reduced functional $j(a,b)$ with respect to a and fixing $b = a$—will lead to a bad choice of step length for either the image or the sensitivity direction (or possibly both).

Usually one remedies such problems by using second order methods, such as Newton's method:

$$x_{n+1} = x_n - (\nabla^2 J(x_n))^{-1}\nabla J(x_n), \tag{16}$$

This, however, requires the Hessian $\nabla^2 J$ to be positive definite at every point, which cannot hold for bilinear functionals of the class considered. Indeed, the Hessian is necessarily rank deficient: Any constant factor (in

the kernel of the penalty term) can arbitrarily be shifted between image and sensitivities without changing the functional value, and hence even local uniqueness is impossible. However, simply fixing the norm of u (or setting its maximal value to one) is not enough to remove the singularity, since the Hessian can be indefinite. Furthermore, constant functions are not necessarily the only functions which leave the functional value unchanged: Since the TV semi-norm is rotationally invariant, there can be smooth functions for which different rotations give the same functional value.

To illustrate these issues quantitatively, we examine the cost functional and its derivatives at a sequence of points x_n in the image-sensitivity parameter space. To give a sense of how the scaling depends on proximity to a minimum, we apply the minimization Algorithm 1 described in section 4 to the functional and compute the functional $J(x_n)$ and gradient values $\nabla J(x_n)$ for several iterates. Table 1 shows the successive decrease of the cost and the norm of the gradient.

While the full Hessian is rather complicated and expensive to compute, we can nonetheless get an estimate on the non-convexity of the functional by approximately computing the minimal and maximal eigenvalues of the Hessian. This is performed using the MATLAB function `eigs`, which serves as an interface for ARPACK [8]. As it is a Krylov method, ARPACK allows for the computation of eigenvalues without requiring the explicit construction of the Hessian matrix. Therefore, it is only necessary to provide a function which performs the action of the Hessian on a test vector, φ, so we approximate its action using the centered difference formula

$$\nabla^2 J(x_n)\varphi = \frac{\nabla J(x_n + \epsilon\varphi) - \nabla J(x_n - \epsilon\varphi)}{2\epsilon} + O(\epsilon^2). \tag{17}$$

The value of ϵ is chosen sufficiently small such that the computed eigenvalues are approximately independent of ϵ. We note that at no stage does the problem appear to be strictly convex, as indicated by the smallest eigenvalue of the Hessian $\nabla^2 J(x_n)$ never being greater than zero. Indeed, in each step the Hessian generally has a large number of zero or even negative eigenvalues. We also see that the problem is poorly scaled by evaluating the reduced functional $j(a, b)$ and noting that the optimal step length in the image and sensitivity descent directions are dissimilar by orders of magnitude, even very close to the minimum. As Table 2 shows, this still holds if we chose a fairly large value for the regularization parameter μ.

The poor scaling of the gradient coupled with lack of positive definiteness of the Hessian suggests that both gradient descent and Newton methods are an unfavorable choice for solving the optimization problem. For this reason, we employ a limited-memory BFGS method which maintains a positive definite approximation to the Hessian and guarantees a descent direction at every step.

3.3. Quasi-Newton method

We therefore wish to make use of second order information to reduce the effects of bad scaling, while still guaranteeing a descent direction in every step. Hence, we use a quasi-Newton method

$$x_{n+1} = x_n - \alpha H_n^{-1} \nabla J(x_n), \tag{18}$$

where H_n is an iteratively constructed approximation to the Hessian $\nabla^2 J(x_n)$ and $\alpha \in (0, 1]$ is chosen such that the Armijo condition

$$J(x_n + \alpha d_n) \leq J(x_n) + a_1 \alpha \nabla J(x_n) \cdot d_n, \tag{19}$$

Table 1: Different iterates in a minimization of J with $\mu = 0$. From left: Functional value, norm of gradient, smallest and largest eigenvalue of the Hessian, optimal step lengths for minimization of $j(a, b)$, optimal value.

J	$\|\nabla J\|$	$\lambda_{\min}$	$\lambda_{\max}$	a^*	b^*	$j(a^*, b^*)$
$3.37 \cdot 10^4$	$1.51 \cdot 10^4$	$-7.82 \cdot 10^3$	$7.74 \cdot 10^3$	$5.43 \cdot 10^{-5}$	$1.29 \cdot 10^{-2}$	$1.83 \cdot 10^4$
$2.95 \cdot 10^1$	$5.96 \cdot 10^1$	$-2.77 \cdot 10^1$	$2.35 \cdot 10^5$	$-8.54 \cdot 10^{-6}$	$6.68 \cdot 10^{-2}$	$2.82 \cdot 10^1$
$1.13 \cdot 10^1$	$5.96 \cdot 10^0$	0	$2.36 \cdot 10^5$	$1.54 \cdot 10^{-5}$	$3.66 \cdot 10^{-3}$	$1.13 \cdot 10^1$
$5.10 \cdot 10^0$	$3.68 \cdot 10^0$	0	$2.35 \cdot 10^5$	$8.48 \cdot 10^{-6}$	$4.84 \cdot 10^{-3}$	$5.09 \cdot 10^0$
$7.64 \cdot 10^{-2}$	$1.20 \cdot 10^0$	0	$2.23 \cdot 10^5$	$8.75 \cdot 10^{-6}$	$1.23 \cdot 10^{-2}$	$7.57 \cdot 10^{-2}$
$7.60 \cdot 10^{-2}$	$7.16 \cdot 10^{-1}$	0	$2.20 \cdot 10^5$	$1.21 \cdot 10^{-5}$	$1.58 \cdot 10^{-2}$	$7.54 \cdot 10^{-2}$
$5.81 \cdot 10^{-5}$	$3.27 \cdot 10^{-1}$	0	$2.24 \cdot 10^5$	$2.96 \cdot 10^{-6}$	$9.47 \cdot 10^{-3}$	$5.55 \cdot 10^{-5}$
$5.72 \cdot 10^{-5}$	$7.90 \cdot 10^{-2}$	0	$2.27 \cdot 10^5$	$4.19 \cdot 10^{-6}$	$9.33 \cdot 10^{-3}$	$5.51 \cdot 10^{-5}$

Table 2: Different iterates in a minimization of J with $\mu = 10^{-2}$. From left: Functional value, norm of gradient, smallest and largest eigenvalue of the Hessian, optimal step lengths for minimization of $j(a, b)$, optimal value.

J	$\|\nabla J\|$	$\lambda_{\min}$	$\lambda_{\max}$	a^*	b^*	$j(a^*, b^*)$
$3.37 \cdot 10^4$	$1.51 \cdot 10^4$	$-7.82 \cdot 10^3$	$7.74 \cdot 10^3$	$5.43 \cdot 10^{-5}$	$1.29 \cdot 10^{-2}$	$1.83 \cdot 10^4$
$3.47 \cdot 10^1$	$2.84 \cdot 10^1$	0	$2.34 \cdot 10^5$	$-3.71 \cdot 10^{-5}$	$1.06 \cdot 10^{-1}$	$3.18 \cdot 10^1$
$1.17 \cdot 10^1$	$5.75 \cdot 10^0$	$-4.20 \cdot 10^0$	$2.37 \cdot 10^5$	$1.75 \cdot 10^{-5}$	$3.01 \cdot 10^{-3}$	$1.16 \cdot 10^1$
$4.89 \cdot 10^0$	$3.48 \cdot 10^0$	0	$2.36 \cdot 10^5$	$-3.84 \cdot 10^{-6}$	$8.95 \cdot 10^{-3}$	$4.87 \cdot 10^0$
$7.34 \cdot 10^{-2}$	$1.40 \cdot 10^0$	0	$2.22 \cdot 10^5$	$4.76 \cdot 10^{-6}$	$1.08 \cdot 10^{-2}$	$7.12 \cdot 10^{-2}$
$7.31 \cdot 10^{-2}$	$8.44 \cdot 10^{-1}$	0	$2.20 \cdot 10^5$	$9.92 \cdot 10^{-6}$	$1.34 \cdot 10^{-2}$	$7.11 \cdot 10^{-2}$
$1.45 \cdot 10^{-3}$	$3.33 \cdot 10^{-1}$	0	$2.24 \cdot 10^5$	$3.99 \cdot 10^{-6}$	$9.92 \cdot 10^{-3}$	$4.65 \cdot 10^{-5}$
$1.40 \cdot 10^{-3}$	$7.54 \cdot 10^{-2}$	0	$2.27 \cdot 10^5$	$4.08 \cdot 10^{-6}$	$9.59 \cdot 10^{-3}$	$3.73 \cdot 10^{-5}$

$0 < a_1 < 1$, is satisfied for the search direction $d_n = -H_n^{-1} \nabla J(x_n)$. If the approximation H_n is constructed according to the BFGS method

$$\begin{cases} s_n = x_n - x_{n-1}, \\ y_n = \nabla J(x_n) - \nabla J(x_{n-1}), \\ H_n = H_{n-1} + \dfrac{y_n \otimes y_n}{y_n \cdot s} - \dfrac{(H_{n-1} s_n) \otimes (H_{n-1} s_n)}{s_n \cdot H_{n-1} s_n}, \end{cases} \tag{20}$$

(where $\otimes$ denotes the dyadic product) and H_0 is symmetric positive definite, the approximation of the Hessian will be positive definite as long as $y_n \cdot s_n > 0$ holds. This can be guaranteed by using a line search for α which additionally enforces the Wolfe condition

$$\nabla J(x_n + \alpha d_n) \cdot d_n \geq a_2 \nabla J(x_n) \cdot d_n, \tag{21}$$

$a_1 < a_2 < 1$, or by simply restarting the approximation from $H_{n+1} = H_0$ whenever $y_n \cdot s_n \leq 0$ occurs. The iteration then converges locally super-linearly (cf., e.g., [9]).

4. Implementation

Discretization. In the application to medical images we deal directly with discrete (pixel-valued) images $u_h \in \mathbb{R}^{N\times N}$, and the Fourier transform in (5) is replaced by the discrete Fourier transform. For standard (uniform) sub-sampling patterns with no center lines and acceleration factor l, the sub-sampling operator K has a simple matrix representation $K_h \in \mathbb{R}^{N/l\times N}$ given by

$$K_h = \begin{pmatrix} I_l & \dots & I_l \end{pmatrix} \tag{22}$$

with l copies of the $\frac{N}{l} \times \frac{N}{l}$-identity matrix I_l. The discretization of the adjoint operator K^* is then the same as the matrix transpose K_h^T. Accordingly, the aliased coil images $v_{k,h}$ are elements $\mathbb{R}^{N/l\times N}$. Since the coil images $v_{k,h}$ are given as pixel-valued images, and the final image will be discrete as well, we chose to implement the discrete analog of the total variation of u directly:

$$TV_h(u_h) = \frac{1}{N^2}\sum_{i,j=1}^{N} \sqrt{(u_{i+1,j} - u_{i-1,j})^2 + (u_{i,j+1} - u_{i,j-1})^2 + \varepsilon}, \tag{23}$$

where $\varepsilon > 0$ is a small parameter commonly used to have a differentiable objective functional, and is chosen small enough such that further reductions do not show significant difference. The image is assumed to have compact support. Because the values taken by the aliased and modulated images $v_{k,h}$ are given only pixel-wise, they are interpreted as an average value over the pixel. The continuous functionals are then replaced by discrete functionals which arise from approximating the area integrals with a two-dimensional composite midpoint rule. Of course it is possible to use higher order quadrature, but given that additional evaluation points do not resolve $v_{k,h}$ any further, the midpoint rule is adequate.

Let w^k be an $N \times N$ matrix representing the product of the image and the sensitivity σ_k, and u_{ij} denote the value of u_h on the ijth pixel. The midpoint of the ijth pixel is $(\hat{x}_i, \hat{y}_j) = (\frac{x_i+x_{i+1}}{2}, \frac{y_j+y_{j+1}}{2})$. This means that the elements of w^k are

$$w_{ij}^k = \frac{u_{ij}}{c_{k,1}^2 + c_{k,2}^2((\hat{x}_i - c_{k,3})^2 + (\hat{y}_j - c_{k,4})^2)}, \quad 1 \le i, j \le N \tag{24}$$

Sub-sampling is then performed by multiplication with the matrix K_h. Thus, the discretized cost functional to be minimized is

$$J(u_h, \bar{c}) = \frac{2}{N^2}\sum_{k=1}^{n} \|K_h w^k - v_k\|_F^2 + \mu TV_h(u_h) \tag{25}$$

where $\|\cdot\|_F$ denotes the Frobenius norm.

In a similar way, the discrete gradient is obtained from (8)–(12) by applying the midpoint rule and replacing the second term in (8) by the matrix $\nabla TV_h(u_h)$ with elements

$$\begin{aligned} T_{ij} = \frac{1}{N^2}\Bigg(&\frac{u_{i,j} - u_{i,j-2}}{\sqrt{\varepsilon + (u_{i,j} - u_{i,j-2})^2 + (u_{i+1,j-1} - u_{i-1,j-1})^2}} + \frac{u_{i,j} - u_{i,j+2}}{\sqrt{\varepsilon + (u_{i,j+2} - u_{i,j})^2 + (u_{i+1,j+1} - u_{i-1,j+1})^2}} \\ &+ \frac{u_{i,j} - u_{i-2,j}}{\sqrt{\varepsilon + (u_{i-1,j+1} - u_{i+1,j+1})^2 + (u_{i,j-1} - u_{i-2,j-1})^2}} + \frac{u_{i,j} - u_{i+2,j}}{\sqrt{\varepsilon + (u_{i,j} - u_{i,j-2})^2 + (u_{i+1,j-1} - u_{i-1,j-1})^2}}\Bigg). \end{aligned} \tag{26}$$

Setting now $x = (\vec{c}^T, u_h^T)^T$,

$$\nabla J(x) = (J_{c_{1,1}}(x), \dots, J_{c_{1,4}}(x), J_{c_{2,1}}(x), \dots, J_{c_{n,4}}(x), J_u(x)^T)^T, \tag{27}$$

the implementation of the BFGS method is straightforward.

Algorithm 1 Globalized BFGS-method

Choose $x_0 = (u_0, c_0), H_0, \tau_0, m_0$, set $k = 0$
while $J(x_k) > \varepsilon_1$ and $J(x_{k-1}) - J(x_k) > \varepsilon_2$ **do**
  Set $z_0 = x_k$, $j = 0$
  while $\|\nabla J(z_j)\| > \tau_k$ **do** {BFGS}
    $d_j = -H_j^1 \nabla J(z_j)$
    Calculate α_j satisfying (19) using line search
    Set $s_j = \alpha_j d_j$, $z_{j+1} = z_j + s_j$, $y_j = \nabla J(z_{j+1}) - \nabla J(z_j)$
    if $y_j \cdot s_j > 0$ and $j < m_k$ **then**
      Update H_{j+1}^{-1}
    else
      Set $H_{j+1} = H_0$
    end if
    Set $j = j + 1$
  end while
  Set $x_{k+1} = x_{j-1}$
  Set $\tau_{k+1} = \rho\tau_k, m_{k+1} = \rho^{-1} m_k$, $k = k + 1$.
end while

Limited memory BFGS. We employ a limited memory BFGS method with restarts and reduced storage of H_n^{-1} ([9, Algorithm `bfgswopt`]). Starting with a symmetric positive definite H_0, every iteration we compute a new search direction d_n, step length α_n, iterate x_n and approximate inverse Hessian H_n^{-1}, until $\|\nabla J(x_n)\|$ drops below a specified tolerance $\tau > 0$ or the maximum number of iterations is reached. Every m iterations, or if $y_n \cdot s_n \leq 0$, we reinitialize and set $H_n = H_0$.

Globalization. Since the objective functional is non-convex, we need to globalize the method in order to avoid falling into local minima. To this end, we use an approach motivated by trust region methods, where we start with a very coarse local model (i.e., m small) and high tolerance τ. We then restart the iteration with a new $m' = \rho m$, $\tau' = \rho^{-1}\tau$. In this way, we reduce the effect of the bad scaling of the gradient away from the global minimum, while avoiding the problem of approximating the (indefinite) Hessian too well. Close to the minimum, the Hessian will be better conditioned, and we can make use of this information to get better search directions. The basic algorithm is given as Algorithm 1.

5. Results

We tested the proposed method with a 128 by 128 pixel slice of a human brain data set (cf. Figure 2, top). We chose twelve sets of parameters to simulate a standard arrangement of twelve receiver coils uniformly distributed in a circle surrounding the head, with il. Specifically, the parameters were set to $c_{i,1} = 0.1$, $c_{i,2} = 1.0$, $c_{i,3} = 1.1\cos(\theta_i)$, $c_{i,4} = -1.1\sin(\theta_i)$ for $i = 1, \ldots, 12$, where $\theta_i = \frac{2\pi}{12} i + 0.3$. The simulated measurements v_i were then generated by adding uniformly distributed Gaussian noise to $K_h(u_{ex}\sigma_i)$, where u_{ex} denotes the exact image and K_h is the sub-sampling operator with a sub-sampling factor of 2 and 4. For the 2 times sub-sampled case, we used a subset of 4 uniformly spaced coils ($i \in \{3, 6, 9, 12\}$), while the 4 times sub-sampled image was reconstructed from all twelve coil images.

Using the following parameters, we applied the globalized BFGS algorithm to this data: $\mu = 10^{-2}$ in (6); $\varepsilon = 10^{-4}$ in (23); $\tau_0 = 10$, $m_0 = 10$, $\varepsilon_1 = 10^{-6}$, $\varepsilon_2 = 10^{-9}$, and $\rho = \frac{2}{1+\sqrt{5}}$ in Algorithm 1. To ensure

that the algorithm will always terminate, we added a maximum number of 3000 iterations as an additional stopping criterion. As an initial guess, we started with a sum of the squares of the coil images for the image (cf. Figure 2, center), and the parameterization $c_i = (1, 1, 0, 0)^T$ for all coils i, which means that we initially supposed the coils to be identical and centered in the domain. The initial approximation of the Hessian was taken as $H_0 = I$, the identity. As mentioned, the result can only be unique up to reciprocal scaling of image and sensitivity, so we normalized the reconstructed image in each outer step such that the maximum value is one, and rescale the sensitivities accordingly. A reference reconstruction using the proposed method without noise and regularization is shown in Figure 2, bottom.

Figure 3 shows the reconstructions from four coil images with double sub-sampling and 3%, 5%, and 10% noise respectively. The image is well reconstructed without any artifacts. Only the last of these shows visible noise in the recovered image.

Figure 4 shows the reconstructions from twelve coil images with 4 times sub-sampling and 2%, 3%, and 5% noise respectively. The image is well reconstructed, apart from residual noise in the center of the image, where the combined sensitivity is very low. Ghost artifacts are not visible.

6. Summary and conclusions

We have considered the problem of reconstructing under-sampled and modulated images in parallel MRI as a bilinear optimization problem, which was shown to suffer from bad gradient scaling and non-convexity of the functional. We proposed an iterative quasi-Newton method which avoided these difficulties. With this method, we were able to solve the minimization problem for a reasonable restriction of admissible sensitivities.

For future work, it is planned to generalize the parameterization of the sensitivities in order to include more realistic modulations and different imaging set-ups such as channel-combined images, where the acquired data are superpositions from multiple coils, and the resulting sensitivities are therefore determined by multiple coil locations at the same time.

Acknowledgments

The authors wish to thank Stephen Keeling and Florian Knoll for helpful discussions and comments. Support by the Austrian Science Fund (FWF) under grant SFB F032 (”‘Mathematical Optimization and Applications in Biomedical Sciences”’) is gratefully acknowledged.

[1] J. Hornak, The Basics of MRI, Interactive Learning Software, Henrietta, NY, 1996–2008, online: http://www.cis.rit.edu/htbooks/mri/.

[2] D. Weishaupt, V. D. Koechli, B. Marincek, How does MRI work? An Introduction to the Physics and Function of Magnetic Resonance Imaging, 2nd Edition, Spinger, Berlin, 2006.

[3] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, P. Boesiger, SENSE: Sensitivity encoding for fast MRI, Magnetic Resonance in Medicine 42 (1999) 952–962.

[4] M. A. Griswold, P. M. Jakob, R. M. Heidemann, Generalized autocalibrating partially parallel acquisitions (GRAPPA), Magnetic Resonance in Medicine 47 (2002) 1202–1210.

[5] F. Bauer, S. Kannengiesser, An alternative approach to the image reconstruction for parallel data acquisition in MRI, Math. Methods Appl. Sci. 30 (12) (2007) 1437–1451.

[6] M. Uecker, T. Hohage, K. T. Block, J. Frahm, Image reconstruction by regularized nonlinear inversion - joint estimation of coil sensitivities and image content, Magnetic Resonance in Medicine 60 (2008) 674–682.

[7] L. I. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms., Physica D 60 (1-4) (1992) 259–268.

[8] R. B. Lehoucq, D. C. Sorensen, C. Yang, ARPACK users’ guide, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1998.

[9] C. T. Kelley, Iterative methods for optimization, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1999.

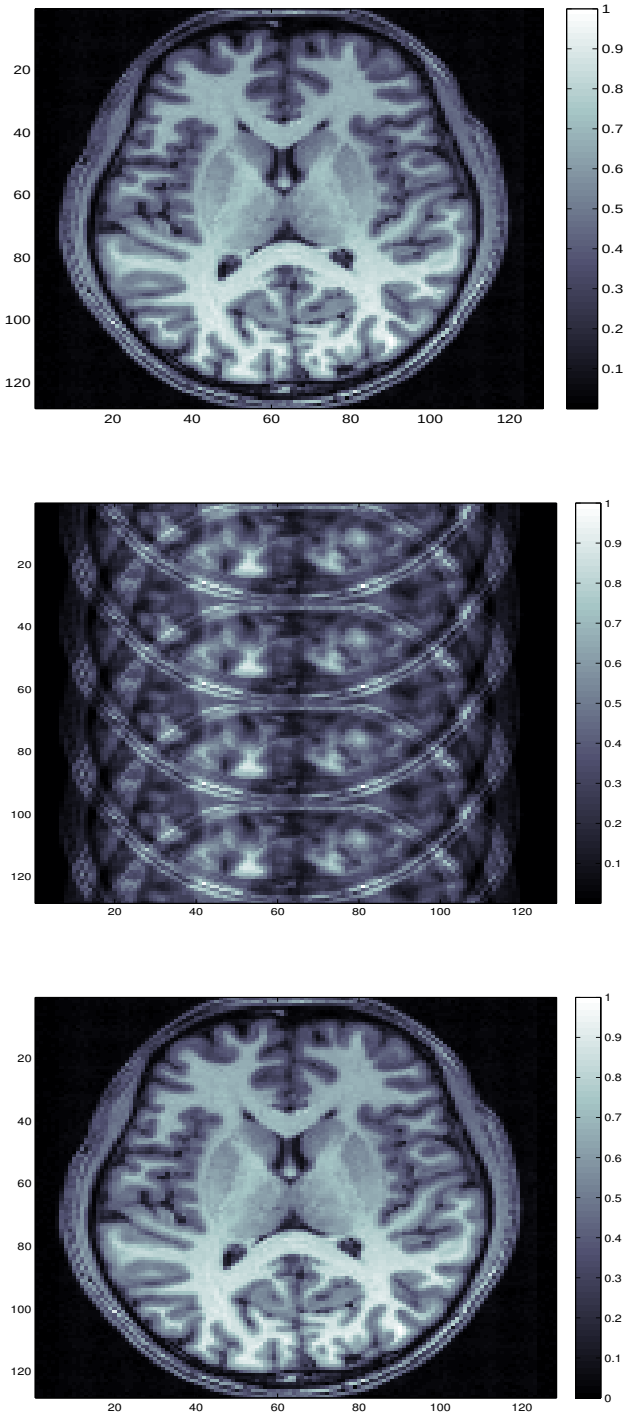

Figure 2: Top: exact image, center: initial guess (4 times sub-sampled, 12 coils), bottom: reconstruction in the absence of noise ($\mu = 0$).

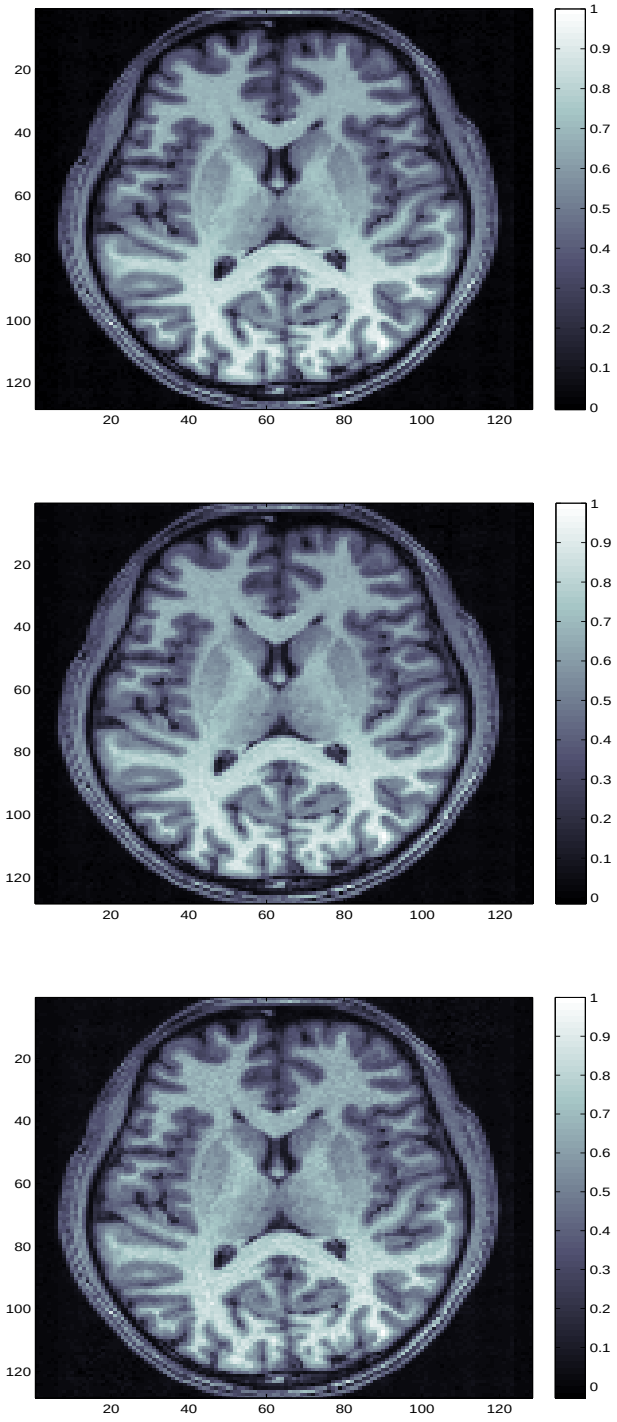

Figure 3: Reconstructed image using 4 coils, acceleration factor 2. Top: 3% noise ($\mu = 10^{-3}$), center: 5% noise ($\mu = 10^{-2}$), bottom: 10% noise ($\mu = 10^{-2}$).

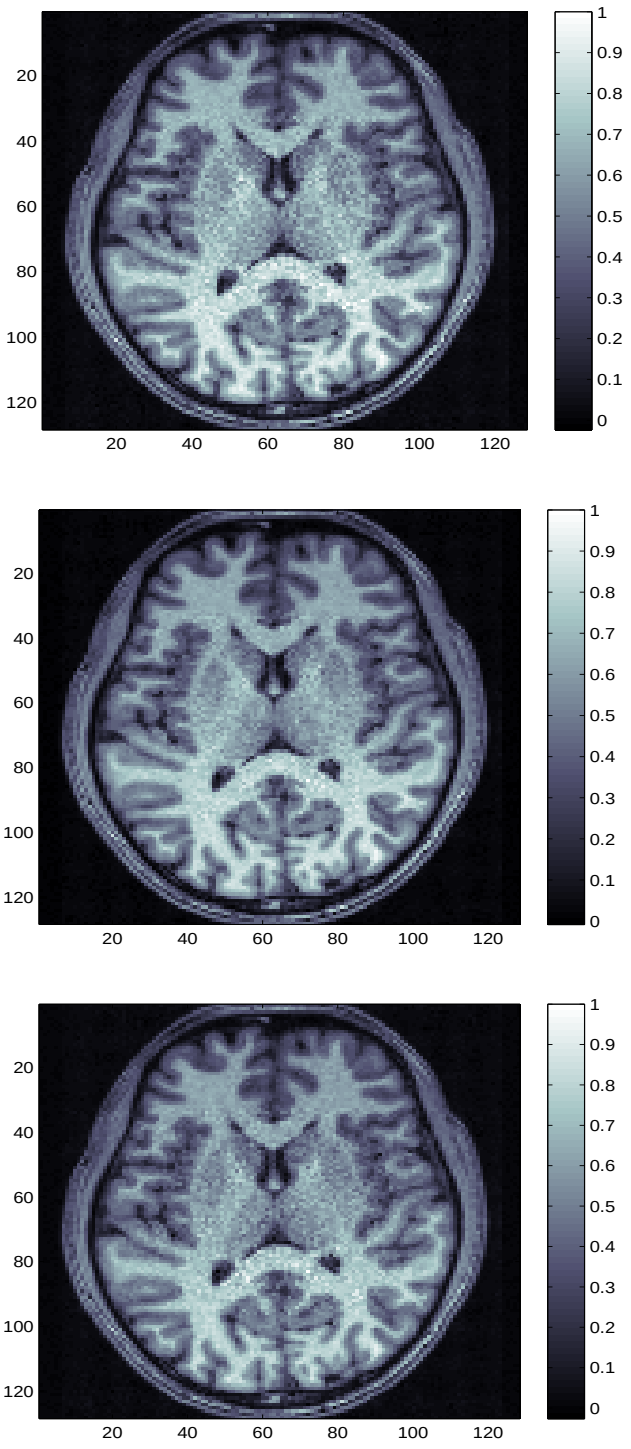

Figure 4: Reconstructed image using 12 coils, acceleration factor 4. Top: 2% noise, center: 3% noise, bottom: 5% noise.