# An image space approach to Cartesian based parallel MR imaging with total variation regularization

Stephen L. Keeling[a,*], Christian Clason[a], Michael Hintermüller[b], Florian Knoll[c], Antoine Laurain[b], Gregory von Winckel[a]

[a] *Institute for Mathematics and Scientific Computing, Karl-Franzens-Universität Graz, Heinrichstraße 36, 8010 Graz, Austria*
[b] *Institute for Mathematics, Humboldt-Universität zu Berlin, Unter den Linden 6, 10099 Berlin, Germany*
[c] *Institute for Medical Engineering, Technische Universität Graz, Kronesgasse 5/II, 8010 Graz, Austria*

---

## Abstract

The Cartesian parallel magnetic imaging problem is formulated variationally using a high order penalty for coil sensitivities and a total variation like penalty for the reconstructed image. Then the optimality system is derived and numerically discretized. The objective function used is non-convex, but it possesses a bilinear structure that allows the ambiguity among solutions to be resolved technically by regularization and practically by normalizing a pre-estimated norm of the reconstructed image. Since the objective function is convex in each single argument, convex analysis is used to formulate the optimality condition for the image in terms of a primal-dual system. To solve the optimality system, a nonlinear Gauss-Seidel outer iteration is used in which the objective function is minimized with respect to one variable after the other using an inner generalized Newton iteration. Computational results for *in vivo* MR imaging data show that a significant improvement in reconstruction quality can be obtained by using the proposed regularization methods in relation to alternative approaches.

*Keywords:* Cartesian parallel magnetic resonance imaging, aliasing correction, bilinear residual, primal-dual, total variation regularization

---

## 1. Introduction

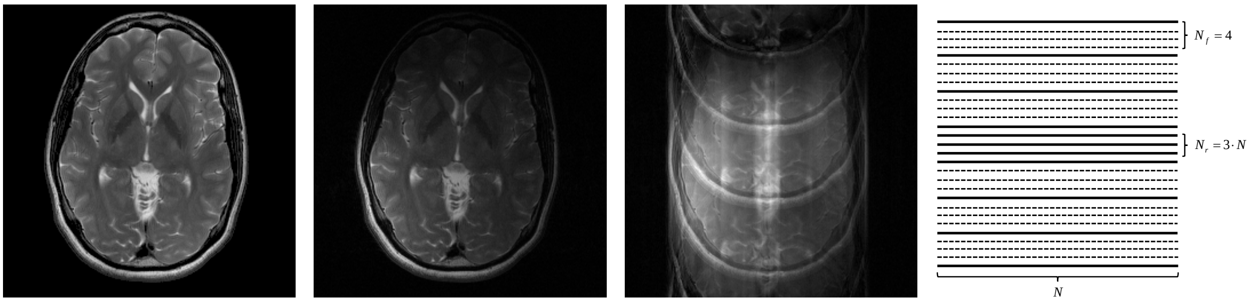
This work is concerned with the image reconstruction problem associated with parallel magnetic resonance imaging (pMRI). The strategy of pMRI is to accelerate data acquisition by sampling only a subset of the spatial Fourier frequencies necessary to represent the desired image. This subsampling leads to aliasing artifacts (see Figure 1c), but measurements are obtained simultaneously from several coils with complementary sensitivities (see Figure 1b), permitting image reconstruction of the underlying image (see Figure 1a) under conditions on the subsampling strategy. Although, for instance, radial and spiral subsampling strategies are in use for pMRI, Cartesian subsampling strategies are still predominantly used in clinical practice, and these involve to undersample whole horizontal or vertical lines in frequency space. Such subsampling strategies are the focus of this paper. For a full discussion of the principles of MRI, see, e.g., Bernstein et al. (2004). For further details of pMRI in particular, see the reviews in Bammer and Schoenberg (2004); Heidemann et al. (2003); Blaimer et al. (2004); Larkman and Nunes (2007) as well as the references cited in these works.

Standard reconstruction strategies currently in use in clinical practice include SENSE (Pruessmann et al., 2001, 1999) and GRAPPA (Griswold et al., 2002). The SENSE approach is based in image space where coil sensitivities are pre-estimated using reference images such as in Figures 1a and 1b obtained with a pre-scan. Knowledge of these sensitivities and the subsampling strategy allows an algebraic reconstruction of images from subsequent measurements

---

[*] Corresponding author

*Email addresses:* `stephen.keeling@uni-graz.at` (Stephen L. Keeling), `christian.clason@uni-graz.at` (Christian Clason), `hint@math.hu-berlin.de` (Michael Hintermüller), `florian.knoll@tugraz.at` (Florian Knoll), `laurain@math.hu-berlin.de` (Antoine Laurain), `gregory.von-winckel@uni-graz.at` (Gregory von Winckel)



(a) original image (b) sensitivity modulated (c) aliased (d) sampling pattern

Figure 1: Shown on the left is **(a)** an image which is neither modulated by a coil sensitivity nor aliased by undersampling. The image in the middle results from **(b)** the product of the first image and a smooth coil sensitivity which reaches a maximum magnitude at the lower left from the image center. The image on the right results from **(c)** aliasing obtained by projecting the first image onto the set of images formed by only a subset of frequencies. Here, these frequencies include every fourth horizontal line in frequency space as well as three horizontal center-lines near the origin of frequency space. This is schematically illustrated in **(d)**, where acquired lines of frequencies are solid and omitted lines are dashed.

using the same coils. Particularly in the course of a dynamic examination in which patient motion may occur, the sensitivities originally estimated according to the SENSE approach may not be accurate for subsequent measurements. Autocalibrated approaches for SENSE have been developed (Wang et al., 2001), which estimate coil sensitivity maps from a fully sampled block of data at the center of frequency space. These are compared here with the proposed approach.

The GRAPPA approach is based in frequency space, where the goal is to recover the frequency coefficients that were not sampled during data acquisition, rather than to unfold aliased images. The key idea of the GRAPPA method is that (due to the use of phased array coils) individual coil images consist of a modulation of the underlying image by the coil sensitivity. Thus, the acquired frequency data are a convolution of the Fourier transform of the underlying image with the Fourier transform of the corresponding coil sensitivity. This convolution operation distributes the information that is stored in a single frequency to neighboring frequencies. Due to the resulting correlation among frequencies, it is possible to recover missing points by a weighted linear combination of adjacent samples. The weights that are used during this interpolation procedure represent the influence of the coil sensitivities, and have to be estimated prior to the reconstruction. This is done with the additional acquisition of so-called autocalibration-signal lines typically near the center of frequency space. To visualize related effects, note that the image of Figure 1c is obtained from the original image in Figure 1a by using only every fourth horizontal line in frequency space as well as additional three horizontal lines near the center of frequency space. Because of the reference center-lines, the image of Figure 1c has more low frequency structure than the vertically periodic image that would otherwise result. Thus, increasing the number of reference lines for autocalibration increases the robustness of the fitting procedure that is used to estimate the GRAPPA reconstruction coefficients, at the cost of a decrease in the total acceleration factor.

Recently, nonlinear least squares approaches have been applied to the pMRI estimation problem; see, e.g., the review of iterative algorithms for model-based MR image reconstruction in Fessler (2010). In particular, comparisons are made in the present paper with the iteratively regularized Gauss-Newton (IRGN) method found in Uecker et al. (2008), where IRGN is compared to SENSE and to GRAPPA. As seen in Uecker et al. (2008), a joint estimation of the image and the sensitivities can lead to improved reconstruction quality of the image. The IRGN method uses a spectral formulation to impose high-order Sobolev regularity on coil sensitivities and $L_2$ regularity on the reconstructed image. Note that such spectral formulations implicitly impose an unnatural periodicity on the globally supported sensitivities. Also, the IRGN approach utilizes quadratic penalties for simplicity, and it relies on the measurement of a sufficient amount of reference data for accuracy. The present authors investigate here the use of non-quadratic penalties and seek thereby to reduce as much as possible the cost of reference measurements while still retaining a suitable reconstruction quality.

Specifically, in the present work the authors utilize a total variation (TV) based regularizer for the reconstructed image. TV has also been applied for SENSE (Liang et al., 2009; Liu et al., 2007), as well as for iterative approaches for radial subsampling (Block et al., 2007; Knoll et al., 2011). Indeed, the recently developed total generalized variation (TGV) (Bredies et al., 2010) has been applied to reconstructions for radial subsampling (Knoll et al., 2011). TV is known not only in the filtering context for its suitability to penalize oscillations without smoothing edges, but also in the compressed sensing context for its role in the reconstruction of images with a sparse representation (Lustig et al., 2007; Liang et al., 2009). A difficulty that arises with the use of TV type penalties is their non-differentiability. As a result, primal-dual formulations have been developed, which have been shown to be superior to purely primal formulations (Chan et al., 1999; Hintermüller and Kunisch, 2004). For further theoretical aspects of TV regularization, see Aubert and Kornprobst (2006).

The proposed approach, here named TVSENSE, has been formulated purely in image space, as with SENSE, as opposed to frequency space, as with GRAPPA. The proposed approach is formulated variationally using a high order penalty for coil sensitivities and a TV based penalty for the reconstructed image. Specifically, the formulation for sensitivities is based upon Keeling and Bammer (2004) and Keeling and Haase (2007), and the total variation formulation is based upon the primal-dual formulation of Hintermüller and Stadler (2006). On the basis of imaging rawdata obtained with *in vivo* MR-measurements, it is seen here that TVSENSE performs well by using only a very small number of reference frequencies.

The paper is organized as follows. In Section 2 the image space oriented variational framework for the parallel imaging problem is formulated. Here, the subsampling projection is introduced along with the regularization used for the modulations and for the reconstructed image. In Section 3 the optimality conditions for the objective function are derived. In particular, the high-order equations for the modulations and the primal-dual system for the reconstructed image are given. In Section 4 the numerical methods used to solve the optimality system are detailed. In particular, the coefficient matrices are partitioned to permit convenient inversion by means of the Sherman–Morrison–Woodbury formula. Also a Newton scheme is developed to solve the primal-dual optimality system for the reconstructed image. Then, the methods used are summarized algorithmically. In Section 5, experiments for measured *in vivo* data are used to compare TVSENSE with SENSE (Pruessmann et al., 2001), GRAPPA (Lustig and Pauly, 2010) and of the iteratively regularized Gauss Newton (IRGN) approach of Uecker et al. (2008). The computational experiments demonstrate that a significant improvement in reconstruction quality is possible by using TVSENSE as opposed to the cited alternatives. Finally, the paper ends with concluding remarks in Section 6.

## 2. Variational Framework

Let $\Omega = (0,1)^d$ denote the image domain with $d = 2$ in this work, although there is no fundamental restriction on the dimension. For $\boldsymbol{x} \in \Omega$, let $u(\boldsymbol{x})$ represent the image to be reconstructed. We assume that there are $N_c$ coil channels, from which the measurements $\{u_i(\boldsymbol{x})\}_{i=1}^{N_c}$ are obtained. The respective coil sensitivities are denoted by $\{\sigma_i(\boldsymbol{x})\}_{i=1}^{N_c}$, which are to be estimated in our approach. All these images are complex-valued (Bernstein et al., 2004). Furthermore, the measured images are aliased by a factor of $N_{\mathbf{f}}$, i.e., in standard Cartesian sampling, one out of every $N_{\mathbf{f}}$ phase encoding lines are acquired, corresponding to $N/N_{\mathbf{f}}$ Fourier coefficients, if $N$ is the fully sampled number of frequencies according to the Nyquist theorem. The number of additional coefficients around the center of frequency space required for auto-calibration will be denoted by $N_{\mathbf{r}}$. Note that we assume that the central point itself will always be part of the sampling pattern, so, e.g., $N_{\mathbf{r}} = 8$ corresponds to a fully acquired square of 9 coefficients in the center of frequency space. For illustration, the magnitude $|u(\boldsymbol{x})|$ of the image to be estimated is displayed in Figure 1a. Similarly, a sample modulated image $|\sigma(\boldsymbol{x})u(\boldsymbol{x})|$ is shown in Figure 1b, and a sample aliased image $|Pu(\boldsymbol{x})|$ is given in Figure 1c. Here, $P$ represents the subsampling projection operator, whose detailed formulation is given below. Each measurement $u_i$ is not only potentially noisy but also manifests simultaneously the modulation effect as well as the aliasing effect: $P\sigma_i u \approx u_i$, $i = 1, \dots, N_c$. As this residual equation hints, each measurement $u_i$ is in the range of $P$, since all the data are sampled on the same subset of frequencies. The desired unknowns are determined here by minimizing the sum of a residual term plus regularization terms:

$$J(u,\{\sigma_i\}) = \frac{1}{2}\sum_{i=1}^{N_c}\int_\Omega |P\sigma_i u - \tilde{u}_i|^2 d\boldsymbol{x} + \frac{\nu}{2}\sum_{i=1}^{N_c}\int_\Omega |\nabla^2\sigma_i|^2 d\boldsymbol{x} + \frac{\kappa}{2}\int_\Omega |u|^2 d\boldsymbol{x} + \mu\int_\Omega \phi_\epsilon(|Du|)d\boldsymbol{x}, \tag{2.1}$$

where $\kappa > 0$ and $\mu \geq 0$ are regularization parameters. The terms in (2.1) will now be described in the order of their appearance.

The subsampling operator is a projection $P = F^{-1}\mathcal{X}F$, where $F$ is the Fourier Transform,

$$\begin{aligned} Fu = \omega, \quad \omega(\boldsymbol{k}) &= \int_\Omega u(\boldsymbol{x}) e^{-2\pi\imath \boldsymbol{k}\cdot\boldsymbol{x}} d\boldsymbol{x}, \\ F^{-1}\omega = u, \quad u(\boldsymbol{x}) &= \sum_{\boldsymbol{k}=-\infty}^{\infty} \omega(\boldsymbol{k}) e^{2\pi\imath \boldsymbol{k}\cdot\boldsymbol{x}}, \end{aligned}$$

and $\mathcal{X}$ denotes a pointwise multiplication in frequency space with a characteristic function $\chi = \chi(\boldsymbol{k})$ assuming the value 1 for sampled frequencies and the value 0 otherwise, i.e., $Pu$ amounts to taking the Fourier transform of $u$, setting the value of all not coefficients corresponding to not acquired frequencies to zero, and then taking the inverse Fourier transform. The projection also has the direct image space representation $P = Q + R$ where $Q$ is given by

$$(Qu)(x_1, x_2) = \frac{1}{N_\mathrm{f}} \sum_{n=1}^{N_\mathrm{f}} u(x_1, \lceil x_2 + (n-1)/N_\mathrm{f} \rceil_1), \quad \lceil x \rceil_1 = x - \lfloor x \rfloor, \tag{2.2}$$

and $R$ is given by

$$(Ru)(\boldsymbol{x}) = 2 \sum_{l=1}^{N_\mathrm{r}/2} \int_\Omega u(\boldsymbol{y}) \cos(2\pi \boldsymbol{k}_l \cdot (\boldsymbol{x} - \boldsymbol{y})) d\boldsymbol{y}. \tag{2.3}$$

Note that $Q$ represents the folding effect that is responsible for the periodicity as seen partially in Figure 1c. Specifically, without center-lines, the image of Figure 1c could be sliced horizontally three times to create four identical horizontal strips. Thus, the folding factor $N_\mathrm{f}$ of (2.2) is in this case $N_\mathrm{f} = 4$. The effect of the center-lines, discussed above in relation to Figure 1c, is represented by $R$. More generally, $R$ represents the effect of $N_\mathrm{r}$ additionally measured reference points which are assumed to be symmetrically situated in frequency space. Thus, according to (2.3), $(Ru)^* = Ru^*$ holds, where $u^*$ denotes the complex conjugate of the function $u$ and $R^*$ denotes the adjoint of the operator $R$. From (2.2) and (2.3) it follows that $P$ satisfies

$$(Pu(\boldsymbol{x}))^* = Pu(\boldsymbol{x})^*, \quad P^* = P, \tag{2.4}$$

i.e., the projection $P$ is self-adjoint.

The image space representation of $Q$ in (2.2) is obtained by noting (Hintermüller et al., 2008) that

$$\begin{aligned} \frac{1}{N_\mathrm{f}} \sum_{j=1}^{N_\mathrm{f}} u(x_1, \lceil x_2 + (j-1)/N_\mathrm{f} \rceil_1) &= \sum_{\boldsymbol{k}=-\infty}^{\infty} \omega(\boldsymbol{k}) e^{2\pi\imath k_1 x_1} \left( \sum_{j=1}^{N_\mathrm{f}} e^{2\pi\imath k_2 (x_2 + (j-1)/N_\mathrm{f})} \right) \\ &= \sum_{\boldsymbol{k}=-\infty}^{\infty} \omega(k_1, N_\mathrm{f} k_2) e^{2\pi\imath (k_1 x_1 + N_\mathrm{f} k_2 x_2)}, \end{aligned} \tag{2.5}$$

where the right side in (2.5) includes only every $N_\mathrm{f}$ line of frequencies, and the left side in (2.5) agrees with the right side of (2.2). Similarly, when Cartesian subsampling is performed in both the vertical and horizontal directions, the operator $Q$ is given by the following instead of (2.2),

$$(Qu)(x_1, x_2) = \frac{1}{N_\mathrm{f}^2} \sum_{n=1}^{N_\mathrm{f}} u(\lceil x_1 + (n-1)/N_\mathrm{f} \rceil_1, \lceil x_2 + (n-1)/N_\mathrm{f} \rceil_1), \tag{2.6}$$

where $N_\mathrm{f}$ is here the same in the vertical and horizontal directions. Again, (2.4) holds. The representation of $R$ in (2.3) is obtained by writing the symmetrically situated reference points as $\boldsymbol{k}_l = -\boldsymbol{k}_{N_\mathrm{r}/2+l}$, $l = 1, \ldots, N_\mathrm{r}/2$ so that (Hintermüller et al., 2008)

$$(Ru)(\boldsymbol{x}) = \sum_{l=1}^{N_\mathrm{r}} \int_\Omega u(\boldsymbol{y}) e^{2\pi\imath \boldsymbol{k}_l \cdot (\boldsymbol{x}-\boldsymbol{y})} d\boldsymbol{y} = \sum_{l=1}^{N_\mathrm{r}/2} \int_\Omega u(\boldsymbol{y}) \left[ e^{2\pi\imath \boldsymbol{k}_l \cdot (\boldsymbol{x}-\boldsymbol{y})} + e^{-2\pi\imath \boldsymbol{k}_l \cdot (\boldsymbol{x}-\boldsymbol{y})} \right] d\boldsymbol{y}$$

holds, which agrees with (2.3). The degree of subsampling is quantified by the acceleration factor $N_\mathrm{a}$, which is the quotient of the number of fully sampled frequencies $N$ in an unaliased image divided by the number of subsampled frequencies in an aliased image.

Note that the magnitudes of all images are considered to be mappings $|u|, |\sigma_i|, |u_i| : \Omega \to [0, 1]$ (where the plotted images are black for magnitudes near 0 and white for magnitudes near 1). However, the restriction of range is not explicitly enforced here, as would be the case, e.g., if barrier functions were added to (2.1) as seen in Hintermüller et al. (2008) and Hintermüller and Laurain (2009).

For the regularization of $\sigma_i$ seen in (2.1), the $\ell_2$ norm of the $n$th order derivative is given by

$$|\nabla^n \sigma|^2 = \nabla^n \sigma^* \cdot \nabla^n \sigma, \qquad \nabla^n \sigma_1^* \cdot \nabla^n \sigma_2 = \sum_{|\alpha|=n} \binom{n}{\alpha!} \partial^\alpha \sigma_1^* \partial^\alpha \sigma_2.$$

The penalty on the second derivative of the modulation $\sigma_i$ seen in (2.1) is based upon work in Keeling and Bammer (2004) and Keeling and Haase (2007). Among the key points is first the fact that the modulation $\sigma_i$ is much smoother than the image $u$. Although it is desirable to avoid that $|\sigma_i|$ have values outside its assumed range $[0, 1]$, it is found here in practice that the modulation remains in the desired range on the image support $\{\boldsymbol{x} \in \Omega : |u(\boldsymbol{x}| > 0\} = (|u| > 0)$ without the use of barrier functions as discussed above. The values of $\sigma_i$ outside the support of $u$ are not important except in the way that they influence values inside the support indirectly through the smoothness of the modulation. In particular, the high order natural boundary conditions on $\sigma_i$ reduce disturbances of the modulation at the domain boundary and thus also at the boundary of the image support. For example, when a penalty such as $\int_\Omega |\Delta \sigma_i|^2 d\boldsymbol{x}$ is used, then harmonic functions are in the kernel of the penalty and boundary disturbances appear as seen in Keeling and Bammer (2004). Also, when a spectral penalty is used such as $\sum_{\boldsymbol{k}} (1 + |\boldsymbol{k}|^2)^2 |\omega(\boldsymbol{k})|^2$, where $\{\omega(\boldsymbol{k})\}$ are Fourier or trigonometric series coefficients of $\sigma_i$ (cf. e.g., Uecker et al. (2008)), then $\sigma_i$ is implicitly continued by periodicity outside of $\Omega$, and finite dimensional approximations lead to boundary disturbances.

To regularize the image $u$, the function $\phi_\epsilon$ in (2.1) is the Gauss-TV penalty used by the authors in Hintermüller and Stadler (2006) and Keeling (2003),

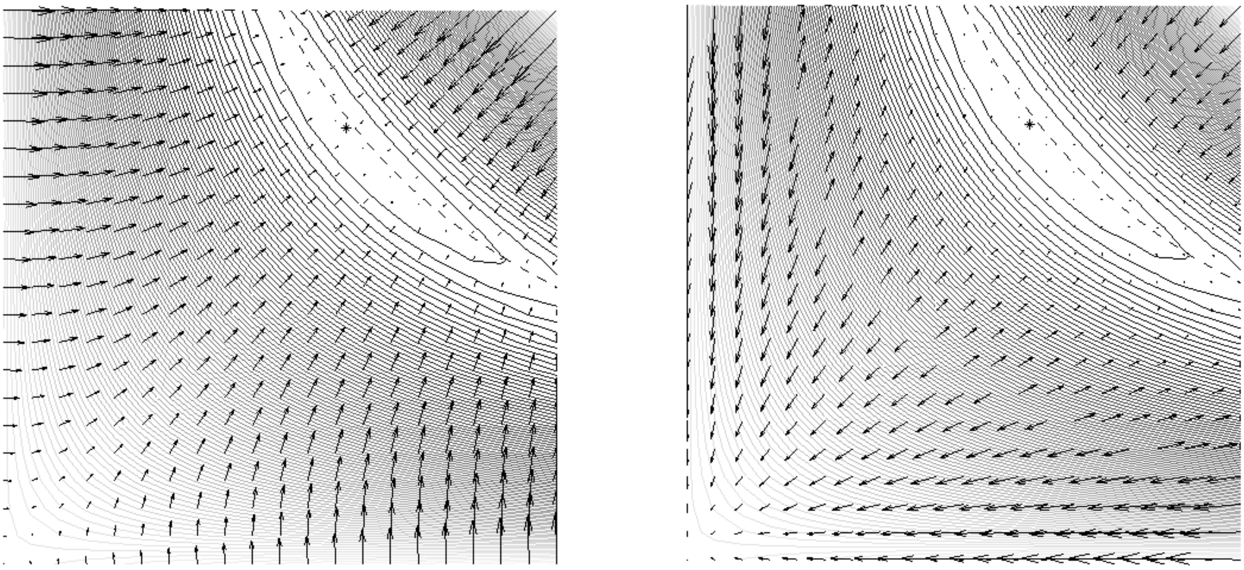
$$\phi_\epsilon(s) = \begin{cases} s^2/(2\epsilon), & |s| \le \epsilon \\ |s| - \epsilon/2, & |s| \ge \epsilon, \end{cases}$$

which emerges naturally from the duality formulation as shown in Keeling et al. (2011). For $u \in \mathrm{BV}(\Omega, \mathbb{C})$, $Du$ in $\phi_\epsilon(|Du|)$ is understood as a measure (Aubert and Kornprobst, 2006), and otherwise $J$ is unbounded. The $L^2$ regularization in (2.1) is included so that the optimality system is well defined even when modulations become very small (see (3.6) and (3.7) below).

Note that an additive decomposition, $\tilde{u} \to u + v$, has been accomplished previously through different regularizations on $u$ and $v$; see, e.g., Aujol and Chambolle (2005). In the present work, as well as in Clason and von Winckel (2010), Hintermüller and Laurain (2009), Keeling et al. (2011) and Uecker et al. (2008), the multiplicative decomposition, $\tilde{u} \to \sigma u$, is also accomplished through different regularizations on $\sigma$ and $u$. This structure makes the derivative of the residual in (2.1) bilinear in the unknown image and sensitivities. To examine the landscape of the functional in (2.1), consider the minimization of the model function

$$f(x, y) = \tfrac{1}{2}(yx - z)^2 + \tfrac{1}{2}\nu y^2 + \mu|x|.$$

A contour plot of $f$ is shown in Figure 2 along with the vector fields $-\nabla f$ and $-[\nabla^2 f]^{-1}\nabla f$. Without regularization from $\mu$ and $\nu$, the whole curve $yx = z$ would minimize $f$; however, with positive regularization a unique minimizer exists, although it lies in a flat and elongated region of the landscape. The location of the unique minimizer depends of course entirely on the regularization. Note that the gradient direction field $-\nabla f$ points strongly toward $yx = z$, but the field is rather weak in a near neighborhood of the curve. On the other hand, the Newton direction field $-[\nabla^2 f]^{-1}\nabla f$ actually points away from $yx = z$ unless sufficiently near to the curve, where the Newton direction field is actually weaker than the gradient direction field. Thus, the application of Newton's method to minimize functionals such as (2.1) is not feasible. (However, quasi-Newton methods can still be applied (Clason and von Winckel, 2010).) Since the objective function is convex in each single argument, it can be minimized rapidly with respect to one argument while the other is held fixed; furthermore, employing such a procedure alternately has performed better than carrying out line searches along gradient directions for the joint functional. Thus, a nonlinear Gauss-Seidel outer iteration is



(a) gradient field

(b) Newton field

Figure 2: Contour plots of $f$ in (2) (with $z = 0.5$, $\mu = \nu = 0.01$) are shown with the vector fields (a) $-\nabla f$ on the left and (b) $-[\nabla^2 f]^{-1}\nabla f$ on the right. The curve $yx = z$ is shown dashed in both plots, and the minimizer for $f$ is shown as the asterisk near the dashed curve.

used in which the formulated objective function is minimized with respect to one variable after the other using an inner generalized Newton iteration.

This iterative method arrives quickly at the flat elongated region of the landscape. However, iterations can stall in such a region, leading numerically to an effective non-uniqueness in minimizers. To distinguish among such numerical minimizers, iterations may be guided by additional information. In particular, the $L^2$-norm of the image can be estimated from the data, and the reconstructed image is normalized here in each iteration to possess the norm estimated below in (3.11). This normalization has been found in practice to have a very positive influence on the iterative scheme. The full approach has been found to perform better, for instance, than treating the nonconvexity of $J$ by starting iterations with larger and ending with smaller regularization parameters in (2.1).

## 3. Optimality Conditions

In this work $H^k(\Omega, \mathbb{C})$ denotes the Sobolev space of complex-valued functions with distributional derivatives up to order $k$ in $L^2(\Omega, \mathbb{C})$; see Evans (2010) for further information about these function spaces. The optimality condition for (2.1) with respect to $\sigma_i$ for fixed $u$ is

$$B(u)\sigma_i = u^* \tilde{u}_i, \qquad \sigma_i \in H^2(\Omega, \mathbb{C}). \tag{3.1}$$

where $B(u)$ is formally given in the strong form by $B(u) = \nu\Delta^2 + u^* P u$ with natural boundary conditions $\partial_n^3 \sigma_i = \partial_n^2 \sigma_i = \partial_n \partial_\tau \sigma_i = 0$, $\partial\Omega$, where $\partial_n$ and $\partial_\tau$ are the normal and tangential derivatives respectively (Keeling, 2003). The optimality condition (3.1) is given in the weak form according to

$$\int_\Omega \left[\nu \nabla^2 \bar{\sigma}^* \cdot \nabla^2 \sigma_i + \bar{\sigma}^* u^* P u \sigma_i\right] d\boldsymbol{x} = \int_\Omega \bar{\sigma}^* u^* \tilde{u}_i d\boldsymbol{x}, \quad \forall \bar{\sigma} \in H^2(\Omega, \mathbb{C}). \tag{3.2}$$

See Keeling and Bammer (2004) for a discussion of existence and regularity of solutions to (3.2) under the conditions that $N_\mathrm{c} = 1$ and $P = I$ hold and that $u$ has a support with positive measure. For the case $N_\mathrm{c} > 1$, the solvability of (3.2) is established as follows.

**Theorem 1.** *Suppose* $\tilde{u}_i \in L^2(\Omega, \mathbb{C})$ *and* $u \in L^\infty(\Omega, \mathbb{C})$. *Suppose further that the support of* $u$ *is such that if a linear function* $\sigma$ *satisfies* $\int_\Omega u\sigma d\boldsymbol{x} = 0$ *then* $\sigma = 0$. *Then there exists a unique solution* $\sigma_i \in H^2(\Omega, \mathbb{C})$ *to* (3.2).

Proof. Define the bilinear form on the left in (3.2) as $F(\sigma, \bar{\sigma})$. If $\sigma \in H^2(\Omega, \mathbb{C})$ satisfies $F(\sigma, \sigma) = 0$, then $\int_\Omega |\nabla^2\sigma|^2 d\boldsymbol{x} = 0$ implies that $\sigma$ is a linear function. From Parseval's identity it follows that that $Qu\sigma = 0$ since $0 = \int_\Omega |Pu\sigma|^2 d\boldsymbol{x} \geq \int_\Omega |Qu\sigma|^2 d\boldsymbol{x}$. Assume that $Q$ is given by (2.2). Then integrating $Qu\sigma = 0$ pointwise over $(0, 1) \times (0, 1/N_\mathrm{f})$ gives

$$0 = \int_0^1 \left[ \int_0^{1/N_\mathrm{f}} \sum_{n=1}^{N_\mathrm{f}} u(x_1, x_2 + (n-1)/N_\mathrm{f})\sigma(x_1, x_2 + (n-1)/N_\mathrm{f}) dx_2 \right] dx_1 = \int_\Omega u\sigma d\boldsymbol{x}. \tag{3.3}$$

The calculation is similar when $Q$ is given by (2.6). Therefore, by the assumption on $u$ and by (3.3) it follows that $\sigma = 0$. Thus, as in the proof of Poincarè's Inequality (Evans, 2010), $[F(\sigma, \sigma)]^{\frac{1}{2}}$ is equivalent to the norm $\|\sigma\|_{H^2(\Omega,\mathbb{C})}$, and hence the bilinear form $F(\sigma, \bar{\sigma})$ is coercive and bounded on $H^2(\Omega, \mathbb{C}) \times H^2(\Omega, \mathbb{C})$. The linear form $G(\bar{\sigma})$ on the right in (3.2) is bounded in $L^2(\Omega, \mathbb{C})$ and hence in $H^2(\Omega, \mathbb{C})$. Thus, the claim follows with the Lax–Milgram lemma (Evans, 2010).

To establish an optimality condition for (2.1) with respect to $u$ for fixed sensitivities $\{\sigma_i\}$, define the functionals $\mathcal{F} : L^2(\Omega, \mathbb{C}) \to \mathbb{R} \cup \{\infty\}$,

$$\mathcal{F}(u) = \frac{1}{2} \sum_{i=1}^{N_\mathrm{c}} \int_\Omega |P\sigma_i u - \tilde{u}_i|^2 d\boldsymbol{x} + \frac{\kappa}{2} \int_\Omega |u|^2 d\boldsymbol{x}, \tag{3.4}$$

and $\mathcal{G} : L^2(\Omega, \mathbb{C}) \to \mathbb{R} \cup \{\infty\}$,

$$\mathcal{G}(u) = \frac{\mu}{2} \int_\Omega \phi_\epsilon(|Du|) d\boldsymbol{x}, \tag{3.5}$$

so that the dependence on $u$ in $J$ is given by $\mathcal{F}(u) + \mathcal{G}(u)$.[1] Note that these operators satisfy the conditions of the Fenchel Duality Theorem, and the desired optimality condition is thus given by (Ekeland and Témam, 1999)

$$\begin{cases} \mathcal{F}(u) + \mathcal{F}^*(v) = \int_\Omega u^* v d\boldsymbol{x}, \\ \mathcal{G}(u) + \mathcal{G}^*(-v) = -\int_\Omega u^* v d\boldsymbol{x}, \end{cases} \tag{3.6}$$

where the convex conjugates $\mathcal{F}^*$ and $\mathcal{G}^*$ in (3.6) are given as follows; see Keeling et al. (2011) for details. First, $\mathcal{F}^* : L^2(\Omega, \mathbb{C}) \to \mathbb{R} \cup \{\infty\}$ is given by

$$\mathcal{F}^*(v) = \frac{1}{2} \int_\Omega \left\{ \left( v^* + \sum_{i=1}^{N_\mathrm{c}} \sigma_i \tilde{u}_i^* \right) \left[ \kappa I + \sum_{i=1}^{N_\mathrm{c}} \sigma_i^* P \sigma_i \right]^{-1} \left( v + \sum_{i=1}^{N_\mathrm{c}} \sigma_i^* \tilde{u}_i \right) - \sum_{i=1}^{N_\mathrm{c}} |\tilde{u}_i|^2 \right\} d\boldsymbol{x}. \tag{3.7}$$

Note that (3.7) is obtained under the assumption that (2.4) holds; otherwise, $\sigma_i^* P \sigma_i$ must be replaced by $\sigma_i^* P^* P \sigma_i$. The invertibility of the operator $\kappa I + \sum_{i=1}^{N_\mathrm{c}} \sigma_i^* P \sigma_i$ is established as follows.

**Theorem 2.** *Suppose that* $\{\sigma_i\} \subset L^\infty(\Omega, \mathbb{C})$. *Then for every* $\kappa > 0$ *the operator* $\kappa I + \sum_{i=1}^{N_\mathrm{c}} \sigma_i^* P \sigma_i$ *is invertible on* $L^2(\Omega, \mathbb{C})$.

Proof. Define the bilinear form $F$ on $L^2(\Omega, \mathbb{C}) \times L^2(\Omega, \mathbb{C})$ and the linear form $G$ on $L^2(\Omega, \mathbb{C})$ by

$$F(u, \bar{u}) = \int_\Omega \bar{u}^* \left[ \kappa I + \sum_{i=1}^{N_\mathrm{c}} \sigma_i^* P \sigma_i \right] u d\boldsymbol{x}, \quad G(\bar{u}) = \int_\Omega \bar{u}^* v d\boldsymbol{x}, \quad u, \bar{u}, v \in L^2(\Omega, \mathbb{C}).$$

---

[1] The authors wish to thank Otmar Scherzer for his suggestion that these functionals be defined on $L^2$; see also the recent book Scherzer et al. (2009).

According to (2.4) $P^*P = P^2 = P$ holds and it follows that

$$\langle \sigma_i u, P\sigma_i u\rangle_{L^2(\Omega,\mathbb{C})} = \langle \sigma_i u, P^* P\sigma_i u\rangle_{L^2(\Omega,\mathbb{C})} = \langle P\sigma_i u, P\sigma_i u\rangle_{L^2(\Omega,\mathbb{C})} \geq 0.$$

Thus, $F$ is coercive. Since $\{\sigma_i\} \subset L^\infty(\Omega, \mathbb{C})$, $F$ is also bounded. Given any $v \in L^2(\Omega, \mathbb{C})$, $G(\bar{u})$ is bounded. Thus, the claim follows with the Lax–Milgram lemma (Evans, 2010).

Secondly, $\mathcal{G}^* : L^2(\Omega, \mathbb{C}) \to \mathbb{R} \cup \{\infty\}$ is given by

$$\mathcal{G}^*(v) = \begin{cases} \frac{\epsilon}{2\mu} \int_\Omega |\boldsymbol{p}|^2 d\boldsymbol{x}, & \text{if } v = \nabla \cdot \boldsymbol{p}, \quad \boldsymbol{p} \in S_\mu, \\ +\infty, & \text{else.} \end{cases} \tag{3.8}$$

where

$$S_\mu = \{\boldsymbol{p} \in H_0(\mathrm{div}, \mathbb{C}) : |\boldsymbol{p}| \leq \mu, \text{ a.e. in } \Omega\},$$

$$H_0(\mathrm{div}, \mathbb{C}) = \{\boldsymbol{p} \in L^2(\Omega, \mathbb{C}^d) : \nabla \cdot \boldsymbol{p} \in L^2(\Omega, \mathbb{C}),\ \boldsymbol{n} \cdot \boldsymbol{p} = 0,\ \partial\Omega\},$$

and $\boldsymbol{n}$ is the outwardly directed unit normal vector at $\partial\Omega$. As seen in Keeling et al. (2011), the functionals (3.4), (3.5), (3.7) and (3.8) lead to the following formulation of the optimality system (3.6):

$$\begin{cases} \left[\kappa I + \sum_{i=1}^{N_c} \sigma_i^* P \sigma_i\right] u - \nabla \cdot \boldsymbol{p} = \sum_{i=1}^{N_c} \sigma_i^* \tilde{u}, & u \in \mathrm{BV}(\Omega, \mathbb{C}), \quad \boldsymbol{p} \in S_\mu, \\ -\mu D u + \lfloor D u \rfloor_\epsilon \boldsymbol{p} = 0, \end{cases} \tag{3.9}$$

where $\lfloor Du \rfloor_\epsilon = \max\{\epsilon, |Du|\}$.

As stated in the previous section, the ambiguity in the product $\sigma_i^* u$ is settled not only by regularization but also by using additional information to guide iterations. One possibility is to normalize the sensitivities to satisfy the following standard assumption (Pruessmann et al., 1999):

$$\sum_{i=1}^{N_c} |\sigma_i(\boldsymbol{x})|^2 = 1, \quad \forall \boldsymbol{x} \in \Omega. \tag{3.10}$$

However, this normalization involves to change the sensitivities in a distributed fashion while modifying an energy norm of the reconstructed image involves only to change the image by a multiplicative constant. The latter normalization has been found to yield results far superior to those obtained with the former normalization. To estimate the $L^2$-norm of $u$ from the given data, note from Parseval's identity and from the residual equations $P\sigma_i u \approx u_i$, $i = 1, \ldots, N_c$ that

$$\begin{aligned} \sum_{i=1}^{N_c} \int_\Omega |\tilde{u}_i(\boldsymbol{x})|^2 d\boldsymbol{x} &= \sum_{i=1}^{N_c} \sum_{\boldsymbol{k}=-\infty}^{\infty} |F\tilde{u}_i|^2(\boldsymbol{k}) \approx \sum_{i=1}^{N_c} \sum_{\boldsymbol{k}=-\infty}^{\infty} |FP\sigma_i u|^2(\boldsymbol{k}) = \sum_{\boldsymbol{k}=-\infty}^{\infty} |F[F^{-1}\mathcal{X}F]\sigma_i u|^2(\boldsymbol{k}) \\ &\approx \sum_{i=1}^{N_c} \frac{1}{N_a} \sum_{\boldsymbol{k}=-\infty}^{\infty} |F\sigma_i u|^2(\boldsymbol{k}) = \frac{1}{N_a} \int_\Omega \left[\sum_{i=1}^{N_c} |\sigma_i(\boldsymbol{x})|^2\right] |u(\boldsymbol{x})|^2 d\boldsymbol{x} = \frac{1}{N_a} \int_\Omega |u(\boldsymbol{x})|^2 d\boldsymbol{x}. \end{aligned} \tag{3.11}$$

Here, it is assumed that the ratio of the energy in the undersampled frequencies $\chi(F\sigma_i u)$ to the energy in the fully sampled frequencies $F(\chi_i u)$ is roughly $1/N_a$, since the acceleration factor $N_a$ appears in the estimate. For the final equality, Parseval's identity is used together with (3.10).

## 4. Numerical Methods

The discretization of the optimality conditions in Subsection 3 begins with a division of $\Omega$ into $N^d = 2^{pd}$ (dimension $d = 2$) cells, each with unit aspect ratio and width $h = 2^{-p}$. Specifically, with the integer component multi-indices

$\boldsymbol{\jmath} = (\jmath_1, \jmath_2, \dots)$, $\mathbf{0} = (0, 0, \dots)$, and $\mathbf{1} = (1, 1, \dots)$, the cell centroids are $\boldsymbol{x}_{\boldsymbol{\jmath}} = (\boldsymbol{\jmath} - \frac{1}{2})h$, $\mathbf{1} \leq \boldsymbol{\jmath} \leq N \cdot \mathbf{1}$. Then, $U_{\boldsymbol{\jmath}} \approx u(\boldsymbol{x}_{\boldsymbol{\jmath}})$ and $\boldsymbol{U}$ denotes the vector of values $\{U_{\boldsymbol{\jmath}}\}$ according to the lexicographic ordering in which $\jmath_1$ increments first from 1 to $N$, then $\jmath_2$, and so on. Also, let $D(\boldsymbol{U})$ denote the diagonal matrix with the values $\{U_{\boldsymbol{\jmath}}\}$ situated along the diagonal according to the lexicographic ordering. The conjugate transpose is denoted by $\boldsymbol{U}^*$ and $D(\boldsymbol{U}^*) = D(\boldsymbol{U})^*$.

Following Keeling and Bammer (2004), (3.1) is discretized using $\Delta^2 \approx B_h$, where $B_h$ is a finite difference approximation to the biharmonic operator with natural boundary conditions. Specifically, the stencil values (weights for neighboring cells) for $B_h$ are given explicitly as follows for the cells with centroids $\{\boldsymbol{x}_{\boldsymbol{\jmath}} : \mathbf{1} \leq \boldsymbol{\jmath} \leq 3 \cdot \mathbf{1}\}$, where stencil weights are obtained by dividing the following by $2800h^4$:

$$
\begin{array}{rrrrr}
0 & 0 & -152 & 424 & 208 \\
0 & 0 & -592 & -2176 & 848 \\
0 & 0 & 4368 & -2256 & 768 \\
0 & 0 & -592 & -2176 & 848 \\
0 & 0 & -152 & 424 & 208
\end{array}
\qquad
\begin{array}{rrrrr}
0 & 424 & 920 & 848 & 208 \\
0 & -2176 & -3920 & -4352 & 848 \\
0 & -2256 & 20400 & -4512 & 768 \\
0 & -2176 & -3920 & -4352 & 848 \\
0 & 424 & 920 & 848 & 208
\end{array}
\qquad
\begin{array}{rrrrr}
208 & 848 & 768 & 848 & 208 \\
848 & -4352 & -4512 & -4352 & 848 \\
768 & -4512 & 24768 & -4512 & 768 \\
848 & -4352 & -4512 & -4352 & 848 \\
208 & 848 & 768 & 848 & 208
\end{array}
$$

$$
\begin{array}{rrrrr}
0 & 0 & -152 & 424 & 208 \\
0 & 0 & -592 & -2176 & 848 \\
0 & 0 & 3440 & -1960 & 920 \\
0 & 0 & -296 & -1088 & 424 \\
0 & 0 & 0 & 0 & 0
\end{array}
\qquad
\begin{array}{rrrrr}
0 & 424 & 920 & 848 & 208 \\
0 & -2176 & -3920 & -4352 & 848 \\
0 & -1960 & 16960 & -3920 & 920 \\
0 & -1088 & -1960 & -2176 & 424 \\
0 & 0 & 0 & 0 & 0
\end{array}
\qquad
\begin{array}{rrrrr}
208 & 848 & 768 & 848 & 208 \\
848 & -4352 & -4512 & -4352 & 848 \\
920 & -3920 & 20400 & -3920 & 920 \\
424 & -2176 & -2256 & -2176 & 424 \\
0 & 0 & 0 & 0 & 0
\end{array}
$$

$$
\begin{array}{rrrrr}
0 & 0 & -152 & 424 & 208 \\
0 & 0 & -296 & -1088 & 424 \\
0 & 0 & 928 & -296 & -152 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{array}
\qquad
\begin{array}{rrrrr}
0 & 424 & 920 & 848 & 208 \\
0 & -1088 & -1960 & -2176 & 424 \\
0 & -296 & 3440 & -592 & -152 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{array}
\qquad
\begin{array}{rrrrr}
208 & 848 & 768 & 848 & 208 \\
424 & -2176 & -2256 & -2176 & 424 \\
-152 & -592 & 4368 & -592 & -152 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{array}
$$

For a given $u$ and $\tilde{u}$ set $\boldsymbol{U} = \{u(\boldsymbol{x}_{\boldsymbol{\jmath}})\}$ and $\tilde{\boldsymbol{U}} = \{\tilde{u}(\boldsymbol{x}_{\boldsymbol{\jmath}})\}$. Then the numerical solution to (3.1) is given as $\boldsymbol{S}_i = \{S_{i,\boldsymbol{\jmath}}\}$, $S_{i,\boldsymbol{\jmath}} \approx \sigma_i(\boldsymbol{x}_{\boldsymbol{\jmath}})$, which solves

$$[\nu B_h + D(\boldsymbol{U}^*) P_h D(\boldsymbol{U})]\boldsymbol{S}_i = D(\boldsymbol{U}^*)\tilde{\boldsymbol{U}}_i. \tag{4.1}$$

where $P_h = Q_h + V_h V_h^{\mathrm{T}}$ for discretizations $Q \approx Q_h$ and $R \approx V_h V_h^{\mathrm{T}}$. The discrete counterparts to (2.2) and (2.6) are, respectively,

$$Q_h = \frac{1}{N_{\mathrm{f}}} K_2^{\mathrm{T}} K_2, \quad K_2 = \Pi_1^{\mathrm{T}} K_h \Pi_1, \quad K_h = \mathbf{1}_{N_{\mathrm{f}}}^{\mathrm{T}} \otimes I_{N*N/N_{\mathrm{f}}}$$

and

$$Q_h = \frac{1}{N_{\mathrm{f}}^2} [K_1^{\mathrm{T}} K_1][K_2^{\mathrm{T}} K_2], \quad K_1 = \Pi_2^{\mathrm{T}} K_h \Pi_2.$$

where $\Pi_1$ permutes so that cells are numbered horizontally and then vertically while $\Pi_2$ permutes so that cells are numbered vertically and then horizontally. Also, $I_k \in \mathbb{R}^{k \times k}$ is the identity matrix and $\mathbf{1}_k \in \mathbb{R}^k$ has all unit entries. The discrete counterpart to (2.3) is (Hintermüller et al., 2008)

$$V_h = [\boldsymbol{V}_1, \dots, \boldsymbol{V}_{N_{\mathrm{r}}}], \qquad \boldsymbol{V}_i \cdot \boldsymbol{V}_j = \delta_{ij},$$

$$\boldsymbol{V}_l = \sqrt{\frac{N}{2} \frac{1}{N^d}} \begin{cases} \left\{ \cos(2\pi \boldsymbol{k}_l \cdot \boldsymbol{x}_{\boldsymbol{\jmath}-\frac{1}{2}}) : \mathbf{1} \leq \boldsymbol{\jmath} \leq N \cdot \mathbf{1} \right\}, & l = 1, \dots, N_{\mathrm{r}}/2, \\ \left\{ \sin(2\pi \boldsymbol{k}_l \cdot \boldsymbol{x}_{\boldsymbol{\jmath}-\frac{1}{2}}) : \mathbf{1} \leq \boldsymbol{\jmath} \leq N \cdot \mathbf{1} \right\}, & l = N_{\mathrm{r}}/2 + 1, \dots, N_{\mathrm{r}}. \end{cases}$$

The matrices on the left side in (4.1) can be stored in sparse format except for the term with $P_h$, and this term can be stored in sparse format if $P_h$ is replaced by $Q_h$:

$$C = [\nu B_h + D(\boldsymbol{U}^*) Q_h D(\boldsymbol{U})].$$

Thus, the matrix on the left in (4.1) is partitioned as

$$C + D(\boldsymbol{U}^*) V_h V_h^{\mathrm{T}} D(\boldsymbol{U}) = C + XX^*, \qquad X = D(\boldsymbol{U}^*) V_h.$$

Then $C + XX^*$ can be inverted using the Sherman–Morrison–Woodbury formula (Golub and Van Loan, 1996):

$$(C + XX^*)^{-1} = C^{-1} - C^{-1}X(I + X^*C^{-1}X)^{-1}X^*C^{-1}. \tag{4.2}$$

Although the iterative solution to systems involving $(C + XX^*)$ has been considered, it has been found in practice that the total computational time is significantly less when implementing the formula of (4.2) in MATLAB using backslash. Note that the matrix $(I + X^*C^{-1}X)$ is full and its size depends upon the number of reference points, which in practice tends to be small in relation to the size of images.

The optimality system (3.9) is solved using a generalized Newton method, which is given by

$$\begin{bmatrix} \kappa I + \sum_{i=1}^{N_c} \sigma_i^* P \sigma_i & -\nabla\cdot \\ \left[-\mu I + \frac{(|\nabla u|>\epsilon)}{\lfloor \nabla u \rfloor_\epsilon} \boldsymbol{p}(\nabla u)^*\right]\nabla & \lfloor \nabla u \rfloor_\epsilon \end{bmatrix} \begin{bmatrix} \delta u \\ \delta \boldsymbol{p} \end{bmatrix} = -\begin{bmatrix} \left[\kappa I + \sum_{i=1}^{N_c} \sigma_i^* P \sigma_i\right] u - \nabla \cdot \boldsymbol{p} - \sum_{i=1}^{N_c} \sigma_i^* \tilde{u}_i \\ -\mu \nabla u + \lfloor \nabla u \rfloor_\epsilon \boldsymbol{p} \end{bmatrix}. \tag{4.3}$$

where $(|\nabla u| > \epsilon)$ is the indicator function of the set $\{x \in \Omega : |\nabla u(x)| > \epsilon\}$. Note that the singular part of the measure $Du$ in (3.9) is neglected here in anticipation of the natural numerical approximation seen below in (4.6). Also, $(\nabla u)^*$ denotes the conjugate transpose of $\nabla u$. The system above in (4.3) is simplified by first eliminating the second equation to obtain

$$\left\{\kappa I + \sum_{i=1}^{N_c} \sigma_i^* P \sigma_i - \nabla \cdot \left[\frac{1}{\lfloor \nabla u \rfloor_\epsilon}\left(\mu I - \frac{(|\nabla u| > \epsilon)}{2\lfloor \nabla u \rfloor_\epsilon}\left[(\nabla u)\,\boldsymbol{p}^* + \boldsymbol{p}\,(\nabla u)^*\right]\right)\right]\nabla\right\}\delta u = \\ -\left[\kappa I + \sum_{i=1}^{N_c} \sigma_i^* P \sigma_i\right] u + \sum_{i=1}^{N_c} \sigma_i^* \tilde{u}_i + \mu \nabla \cdot \left(\frac{\nabla u}{\lfloor \nabla u \rfloor_\epsilon}\right). \tag{4.4}$$

and the eliminated equation becomes

$$\delta \boldsymbol{p} = \frac{1}{\lfloor \nabla u \rfloor_\epsilon}\left(\mu I - \frac{(|\nabla u| > \epsilon)}{2\lfloor \nabla u \rfloor_\epsilon}\left[(\nabla u)\boldsymbol{p}^* + \boldsymbol{p}(\nabla u)^*\right]\right)\nabla \delta u + \mu \frac{\nabla u}{\lfloor \nabla u \rfloor_\epsilon} - \boldsymbol{p}. \tag{4.5}$$

Note that the term $\boldsymbol{p}(\nabla u)^*$ in (4.3) has been symmetrized in (4.4) and (4.5) with $[(\nabla u)\boldsymbol{p}^* + \boldsymbol{p}(\nabla u)^*]/2$. To discretize (4.4) and (4.5) the discrete derivative matrices $\nabla_h^{(x_i)}$ are formed using forward differences, where Neumann boundary conditions are implemented implicitly by setting the stencil weights of $\nabla_h^{(x_i)}$ to zero at cells adjacent to the boundary where $x_i = 1$. Thus, in $\mathbb{R}^2$ the gradient and divergence are approximated by

$$\nabla \approx \nabla_h = \begin{pmatrix} \nabla_h^{(x_1)} \\ \nabla_h^{(x_2)} \end{pmatrix}, \quad \nabla\cdot \approx -\nabla_h^{\mathrm{T}}. \tag{4.6}$$

Also, functions of $\nabla_h \boldsymbol{U}$ are defined according to

$$|\nabla_h \boldsymbol{U}| = \left\{\sqrt{|\nabla_h^{(1)} \boldsymbol{U}|_J^2 + |\nabla_h^{(2)} \boldsymbol{U}|_J^2}\right\}, \qquad (|\nabla_h \boldsymbol{U}| > \epsilon) = \left\{|\nabla_h \boldsymbol{U}|_J > \epsilon\right\},$$

$$\frac{1}{\lfloor \nabla_h \boldsymbol{U} \rfloor_\epsilon} = \left\{\frac{1}{\max\{|\nabla_h \boldsymbol{U}|_J, \epsilon\}}\right\}, \qquad \frac{(|\nabla_h \boldsymbol{U}| > \epsilon)}{\lfloor \nabla_h \boldsymbol{U} \rfloor_\epsilon} = D(|\nabla_h \boldsymbol{U}| > \epsilon)\frac{1}{\lfloor \nabla_h \boldsymbol{U} \rfloor_\epsilon},$$

and $\boldsymbol{P} = (\boldsymbol{P}_1; \boldsymbol{P}_2)$ is understood below as a column vector of column vectors $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$. Thus, (4.4) and (4.5) are discretized as

$$\left\{\kappa I + \sum_{i=1}^{N_c} D(\boldsymbol{S}_i^*) P_h D(\boldsymbol{S}_i) + \nabla_h^{\mathrm{T}}\left[D\left(\frac{1}{\lfloor \nabla_h \boldsymbol{U} \rfloor_\epsilon}\right)\left(\mu I - D\left(\frac{(|\nabla \boldsymbol{U}| > \epsilon)}{2\lfloor \nabla \boldsymbol{U} \rfloor_\epsilon}\right)\left[(\nabla_h \boldsymbol{U})\boldsymbol{P}_\mu^* + \boldsymbol{P}_\mu(\nabla_h \boldsymbol{U})^*\right]\right)\right]\nabla_h\right\}\boldsymbol{\delta U} = \\ -\left[\kappa I + \sum_{i=1}^{N_c} D(\boldsymbol{S}_i^*) P_h D(\boldsymbol{S}_i)\right]\boldsymbol{U} + \sum_{i=1}^{N_c} D(\boldsymbol{S}_i^*)\tilde{\boldsymbol{U}}_i - \mu \nabla_h^{\mathrm{T}}\left[D\left(\frac{1}{\lfloor \nabla \boldsymbol{U} \rfloor_\epsilon}\right)\nabla_h \boldsymbol{U}\right] \tag{4.7}$$

and

$$\boldsymbol{\delta P} = D\left(\frac{1}{\lfloor \nabla \boldsymbol{U} \rfloor_\epsilon}\right)\left(\mu I - D\left(\frac{(|\nabla \boldsymbol{U}| > \epsilon)}{2\lfloor \nabla \boldsymbol{U} \rfloor_\epsilon}\right)\left[(\nabla_h \boldsymbol{U})\boldsymbol{P}_\mu^* + \boldsymbol{P}_\mu(\nabla_h \boldsymbol{U})^*\right]\right)\nabla_h \boldsymbol{\delta U} + \mu D\left(\frac{1}{\lfloor \nabla U \rfloor_\epsilon}\right)\nabla_h \boldsymbol{U} - \boldsymbol{P}, \tag{4.8}$$

where, following Hintermüller and Stadler (2006), $\boldsymbol{P}_\mu$ is a projection of $\boldsymbol{P}$ truncated to have a cellwise magnitude not more than $\mu$:

$$\begin{aligned} \boldsymbol{P}_\mu &= \begin{pmatrix} D(|\boldsymbol{P}| < \mu)\boldsymbol{P}_1 \\ D(|\boldsymbol{P}| < \mu)\boldsymbol{P}_2 \end{pmatrix} + \mu \begin{pmatrix} D(|\boldsymbol{P}| \geq \mu)D(|\boldsymbol{P}|)^{-1}\boldsymbol{P}_1 \\ D(|\boldsymbol{P}| \geq \mu)D(|\boldsymbol{P}|)^{-1}\boldsymbol{P}_2 \end{pmatrix}, \\ |\boldsymbol{P}| &= \left\{ \sqrt{|\boldsymbol{P}_1|_J^2 + |\boldsymbol{P}_2|_J^2} \right\}. \end{aligned} \tag{4.9}$$

As shown in Hintermüller and Stadler (2006), using the truncation $\boldsymbol{P}_\mu$ of (4.9) in (4.7) guarantees that $\boldsymbol{\delta U}$ provides a descent direction for the objective function with fixed modulations. The matrices on the left in (4.7) can be stored in sparse format except for the term with $P_h$. Nevertheless, this term can be stored in sparse format if $P_h$ is replaced by $Q_h$:

$$A = \left\{ \kappa I + \sum_{i=1}^{N_c} D(\boldsymbol{S}_i^*) Q_h D(\boldsymbol{S}_i) + \nabla_h^{\mathrm{T}} \left[ D\left( \frac{1}{\lfloor \nabla_h \boldsymbol{U} \rfloor_\epsilon} \right) \left( \mu I - D\left( \frac{(|\nabla \boldsymbol{U}| > \epsilon)}{2\lfloor \nabla \boldsymbol{U} \rfloor_\epsilon} \right) \left[ (\nabla_h \boldsymbol{U}) \boldsymbol{P}_\mu^* + \boldsymbol{P}_\mu (\nabla_h \boldsymbol{U})^* \right] \right) \right] \nabla_h \right\}, \tag{4.10}$$

so the matrix on the left in (4.7) can be partitioned as

$$A + \sum_{i=1}^{N_c} D(\boldsymbol{S}_i^*) V_h V_h^{\mathrm{T}} D(\boldsymbol{S}_i) = A + YY^*, \qquad Y = (D(\boldsymbol{S}_1^*)V_h, \dots, D(\boldsymbol{S}_{N_c}^*)V_h).$$

Thus, $A + YY^*$ can be inverted using the Sherman–Morrison–Woodbury formula (Golub and Van Loan, 1996):

$$(A + YY^*)^{-1} = A^{-1} - rbA^{-1}Y(I + Y^*A^{-1}Y)^{-1}Y^*A^{-1}. \tag{4.11}$$

As with (4.2), the iterative solution to systems involving $(A + YY^*)$ has been considered, and it has been found in practice that the total computational time is significantly less when implementing the formula of (4.11) in MATLAB using backslash. Note that the matrix $(I + Y^*A^{-1}Y)$ is full and its size depends upon the number of reference points, which in practice tends to be small in relation to the size of images.

Note that if only $L^2$ regularization is used ($\kappa > 0$) and no TV regularization is used ($\mu = 0$), then (4.10) is simplified according to

$$A = \kappa I + \sum_{i=1}^{N_c} D(\boldsymbol{S}_i^*) Q_h D(\boldsymbol{S}_i), \tag{4.12}$$

and $\boldsymbol{U}$ is determined by simplifying (4.7) according to

$$(A + YY^*)\boldsymbol{U} = \sum_{i=1}^{N_c} D(\boldsymbol{S}_i^*) \tilde{U}_i, \tag{4.13}$$

where the formula (4.11) is used as described above to solve this system. To reduce computational expense, the algorithm begins with purely $L^2$ regularization and then continues with TV regularization. The complete algorithm is summarized as Algorithm 1. Computational results using this algorithm are shown in the next section.

## 5. Computational Results

In this section, the approach developed in the previous sections is applied to *in vivo* raw data obtained from MR scanner. A fully sampled T2 weighted Turbo Spin Echo Scan of the brain of a healthy volunteer was acquired on a 3T system. Written informed consent was obtained prior to the examination. Sequence parameters were repetition time TR=5000ms, echo time TE=99ms, turbo factor 10, slice thickness 4mm, matrix size $(x, y) = 256 \times 256$, covering a field of view of 220mm × 220mm. A conventional 4 channel head coil was used, so the number of coils is $N_c = 4$. Rawdata were exported from the scanner to obtain fully sampled images $\{\boldsymbol{W}_i^\dagger\}_{i=1}^4$. These were subsampled retrospectively to produce the data set $\{\tilde{\boldsymbol{U}}_i\}_{i=1}^4$. For instance, the image $|\tilde{\boldsymbol{U}}_1|$ is shown in Figure 3c. This approach was used in order to have a fully sampled gold standard available to compare the reconstruction and to compute the quantitative errors according to (5.1) below. The gold standard image was determined according to $\boldsymbol{U}^\dagger = [\sum_{i=1}^4 |\boldsymbol{W}_i^\dagger|^2]^{\frac{1}{2}}$ and $|\boldsymbol{U}^\dagger|$ is shown

**Algorithm 1** TVSENSE

```
Input: $\{\tilde{\boldsymbol{U}}_i\}$, $\epsilon$, $\delta$, $\kappa$, $\mu$, $\nu$, $k_1$, $k_2$
Output: $\boldsymbol{U}$, $\{\boldsymbol{S}_i\}$

Initialization:
    Starting Image: According to (3.11),
      $\boldsymbol{U} \leftarrow \sum_{i=1}^{N_c} \tilde{\boldsymbol{U}}_i / N_c$,
      $\boldsymbol{U} \leftarrow U_n \boldsymbol{U}/\|\boldsymbol{U}\|$ where $U_n^2 = N_a \sum_{i=1}^{N_c} \|\tilde{U}_i\|^2$
    Starting Sensitivities: According to (3.10),
      $\boldsymbol{S}_i = 1/\sqrt{N_c}$, $i = 1, \dots, N_c$
for $m = 1, 2$ do                                  ▷ ($m = 1 \Rightarrow$ purely $L_2$, $m = 2 \Rightarrow$ with TV)
    Outer Iteration: start with $t_0 = \|\boldsymbol{U}\|$, $t = 2\delta \cdot t_0$, $k = 0$
    while ($t > \delta \cdot t_0$) and ($k < k_m$) do
        save $\hat{\boldsymbol{U}} = \boldsymbol{U}$
        solve (4.1) using (4.2) to obtain $\boldsymbol{S}_i$, $i = 1, \dots, N_c$
        if $m = 1$ then
            solve (4.13) using (4.11) and (4.12) to obtain $\boldsymbol{U}$
        else if $m = 2$ then
            Inner Iteration: start with $s_0 = \|\boldsymbol{U}\|$, $s = 2\delta \cdot s_0$, $\boldsymbol{P} = 0$
            while ($s > \delta \cdot s_0$) do
                solve (4.7) using (4.11) to obtain $\boldsymbol{\delta U}$ and set $\boldsymbol{U} = \boldsymbol{U} + \boldsymbol{\delta U}$
                set $\boldsymbol{\delta P}$ with (4.8) set $\boldsymbol{P} = \boldsymbol{P} + \boldsymbol{\delta P}$ and set $\boldsymbol{P}_\mu$ according to (4.9)
                update $s = \|\boldsymbol{\delta U}\|$
            end while
        end if
        normalize $\boldsymbol{U} \leftarrow U_n \boldsymbol{U}/\|\boldsymbol{U}\|$
        update $t = \|\boldsymbol{U} - \hat{\boldsymbol{U}}\|$
        update $k = k + 1$
    end while
end for
```

in Figure 3a. The gold standard sensitivities were determined according to $\boldsymbol{S}_i^\dagger = \boldsymbol{W}_i/U^\dagger$, and $|\boldsymbol{S}_1^\dagger|$ for instance is shown in Figure 3d. The data images $\{\tilde{\boldsymbol{U}}_i\}$ were simulated as follows. Cartesian subsampling was performed in both the vertical and horizontal directions, so the operator $Q$ is given by (2.6), where $N_f = 2$ in the vertical and horizontal directions. Also, only $N_r = 8$ additional reference points were used, which were contained in a $3 \times 3$ square centered at the origin in frequency space. Note that the point in the center is automatically sampled due to the regular sampling pattern (2.6). Neglecting the acceleration effect of this small number of reference points gives an acceleration factor $N_a \approx N_f^2 = 4$. While this acceleration is routinely used in practice, $N_r$ is chosen significantly smaller to illustrate the effectiveness of the proposed approach.

The data images $\{\tilde{\boldsymbol{U}}_i\}$ were used as input for the TVSENSE algorithm along with the parameters $\epsilon = 10^{-3}$, $\kappa = 10^{-4}$, $\nu = 10^2$, $\mu = 10^{-4}$ and $\delta = 10^{-3}$ which were chosen based on visual inspection. (Also, $k_1$ and $k_2$ are chosen sufficiently large so that the $\delta$-tolerance must be met for termination.) The domain was normalized to $(1, N)^2$ so that $h = 1$. Results of the TVSENSE algorithm for this example are shown in the second row of Figure 3. To demonstrate the effect of TV regularization, $k_2 = 0$ was also used to perform only $L_2$ regularization, and the results appear in the third row of Figure 3. All images are shown with the intensity scale $[0, 1]$, where lower intensities are shown darker while higher intensities are shown brighter. To avoid the effect of an error in intensity scale, the reconstruction error is measured with the metrics,

$$\begin{cases} d_2(\boldsymbol{U}, \boldsymbol{U}^*) = \dfrac{1}{N} \min_{s \in \boldsymbol{R}} \|s\boldsymbol{U} - \boldsymbol{U}^*\|_{\ell_2} = \dfrac{1}{N}\|s^*\boldsymbol{U} - \boldsymbol{U}^*\|_{\ell_2}, \quad s^* = \boldsymbol{U} \cdot \boldsymbol{U}^*/\|\boldsymbol{U}\|_{\ell_2}^2, \\ d_\infty(\boldsymbol{U}, \boldsymbol{U}^*) = \|\Delta \boldsymbol{U}\|_{\ell_\infty}, \quad \Delta \boldsymbol{U} = s^*\boldsymbol{U} - \boldsymbol{U}^*. \end{cases} \tag{5.1}$$

For the two cases of Figure 3, the reconstruction $|\boldsymbol{U}|$ is shown in the first two columns, the scaled image reconstruction error $5|\Delta\boldsymbol{U}|$ is shown in the third column and the scaled sensitivity reconstruction error $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$ is shown in the last column of Figure 3. Note that the multiplicative factor of 5 is used to emphasize differences. The quantitative errors for these cases are listed in Table 1 along with the computational costs. The result with $\mu > 0$ is evidently more accurate. In particular, the noise suppression properties of TV regularization can be seen clearly.

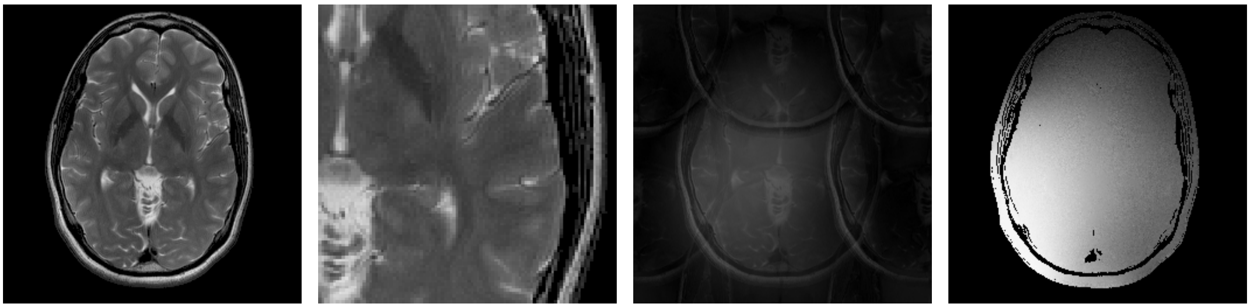
The TVSENSE result is also compared with IRGN (Uecker et al., 2008), SENSE (Pruessmann et al., 1999) and GRAPPA (Griswold et al., 2002). The authors' IRGN code was implemented in cooperation with the authors of Uecker et al. (2008). The main parameter influencing the reconstruction quality for IRGN is the number of iterations, which was chosen (= 3) to minimize $d_2(\boldsymbol{U}, \boldsymbol{U}^\dagger)$. The authors' SENSE code was an in-house implementation of iterative SENSE according to Pruessmann et al. (2001). In the SENSE code, sensitivities were estimated from the low resolution fully sampled $3 \times 3$ square at the center of frequency space according to an autocalibrated approach like mSENSE (Wang et al., 2001). This fully sampled block was transformed to image space and a pointwise sum-of-squares combination of the images was performed. The individual coil images were then normalized by the sum-of-squares combination, which yielded the estimations of the sensitivities that were then used for the iterative SENSE reconstruction. The main parameter influencing the reconstruction quality for the SENSE code is the Tikhonov regularization parameter and the maximum number of CG iterations, which were chosen (= $10^{-3}$ and 500, respectively) to minimize $d_2(\boldsymbol{U}, \boldsymbol{U}^\dagger)$. The authors' GRAPPA code was an implementation provided by Michael Lustig and described in Lustig and Pauly (2010). In the GRAPPA code, a $3 \times 3$ reconstruction kernel was used, and Tikhonov regularization was used in the calibration step. The unaliased images $\{\boldsymbol{W}_i \approx \boldsymbol{W}_i^\dagger\}_{i=1}^4$ are estimated by GRAPPA, and the reconstructed GRAPPA image is given by $\boldsymbol{U} = [\sum_{i=1}^4 |\boldsymbol{W}_i|^2]^{\frac{1}{2}}$ while the reconstructed sensitivities are given by $\{\boldsymbol{S}_i = \boldsymbol{W}_i/\boldsymbol{U}\}_{i=1}^4$. The main parameter influencing the reconstruction quality for the GRAPPA code is the interpolation kernel size, which was chosen (= $3 \times 3$) to minimize $d_2(\boldsymbol{U}, \boldsymbol{U}^\dagger)$. All computations of this work were performed on an Intel Core 2 Q9400 (four cores) with 4GB RAM.

Reconstructed images $|\boldsymbol{U}|$ obtained by the four methods are shown in the first two columns of Figure 4. The image reconstruction errors $5|\Delta\boldsymbol{U}|$ and the sensitivity reconstruction errors $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$ are shown respectively in the third and fourth columns of Figure 4, where the multiplicative factor of 5 is used to emphasize differences. All resulting images are shown on the scale $[0, 1]$, where lower intensities are shown darker while higher intensities are shown brighter. To avoid the effect of an error in intensity scale, the reconstruction error is measured with the metrics (5.1). The quantitative errors for these methods are listed in Table 2 along with the computational costs. From the quantitative measures in Table 2 one notices that in all metrics the errors associated with the TVSENSE image reconstruction are smaller than those associated with the other methods. However, the images of Figure 4 indicate the accuracy much more clearly than do the values of Table 2. According to a visual inspection, the results shown in Figure 4 appear evidently in the order of decreasing accuracy.

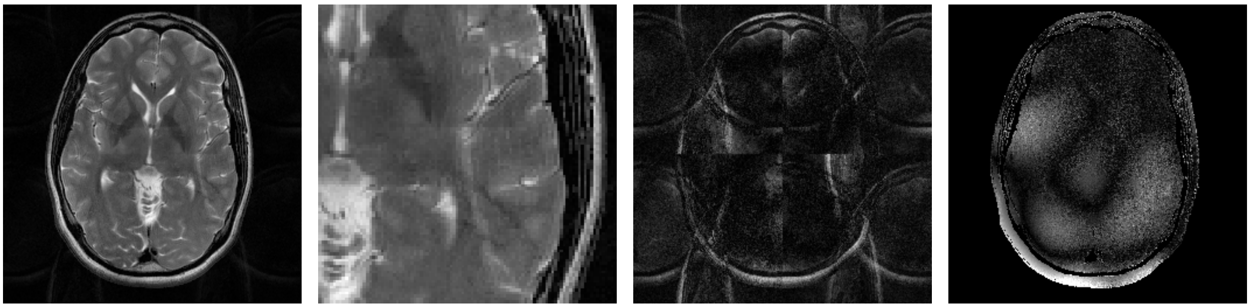
Figure 5 demonstrates the effect of additional reference data as would be used in routine applications. For this, each method was applied using $N_\mathrm{r} = 56$ additional reference points in an $11 \times 11$ square centered at the origin in frequency space. Again, the number of iterations (= 4) for IRGN, the Tikhonov regularization parameter (= $10^{-4}$) for SENSE and the interpolation kernel size (= $5 \times 5$) for GRAPPA were chosen to minimize $d_2(\boldsymbol{U}, \boldsymbol{U}^\dagger)$. Parameters for TVSENSE were the same as in the previous example. The same format is used for Figure 5 as is used for Figure 4. The quantitative errors associated with Figure 5 are listed in Table 3 along with the computational costs. From the quantitative measures in Table 2 one notices that the errors of the four methods are comparable. On the other hand, the results in Figure 5 are evidently shown in the order of decreasing accuracy.

Table 1: Shown for each case of Figure 3 are the quantitative errors, measured in the metrics of (5.1), as well as the computational time in seconds and the number of iterations where applicable. Here, $\hat{\boldsymbol{S}}_1 = \boldsymbol{S}_1 \cdot (|\boldsymbol{U}^\dagger| > 0)$ and $\hat{\boldsymbol{S}}_1^\dagger = \boldsymbol{S}_1^\dagger \cdot (|\boldsymbol{U}^\dagger| > 0)$.

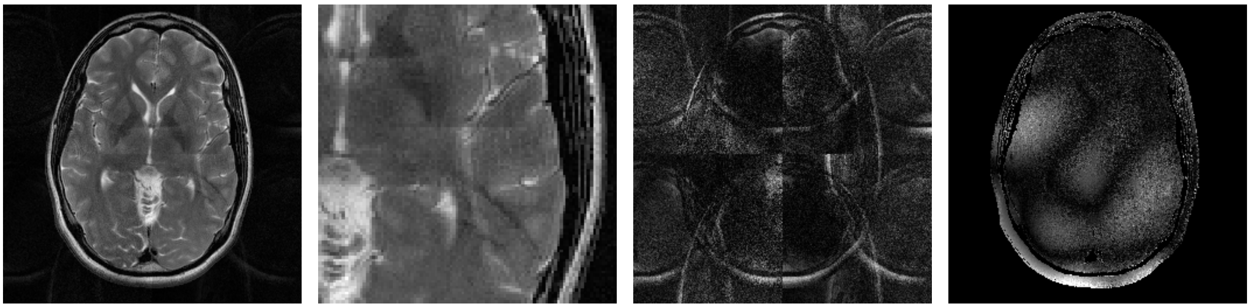
| TVSENSE | $d_2(\boldsymbol{U}, \boldsymbol{U}^\dagger)$ | $d_\infty(\boldsymbol{U}, \boldsymbol{U}^\dagger)$ | $d_2(\hat{\boldsymbol{S}}_1, \hat{\boldsymbol{S}}_1^\dagger)$ | $d_\infty(\hat{\boldsymbol{S}}_1, \hat{\boldsymbol{S}}_1^\dagger)$ | time | iterations |
|---|---|---|---|---|---|---|
| $\mu = 10^{-4}$ | 0.0283 | 0.172 | 0.0384 | 0.379 | 11544 | $244_{L^2} + 38_{\mathrm{TV}}$ |
| $\mu = 0$ | 0.0350 | 0.179 | 0.0354 | 0.355 | 9918 | $244_{L^2}$ |



(a) $|\boldsymbol{U}^{\dagger}|$ (b) $|\boldsymbol{U}^{\dagger}|$ (zoom) (c) $|\tilde{\boldsymbol{U}}_1|$ (d) $|\boldsymbol{S}_1^{\dagger}|$

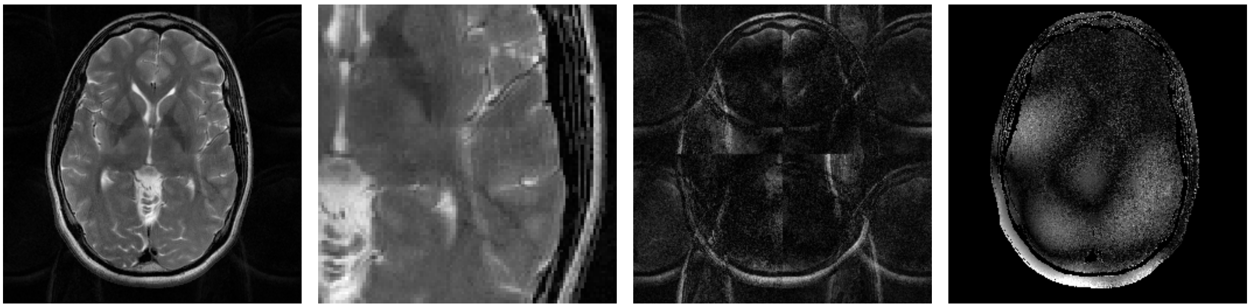


(e) $|\boldsymbol{U}|$ ($\mu > 0$) (f) $|\boldsymbol{U}|$ ($\mu > 0$, zoom) (g) $5|\Delta \boldsymbol{U}|$ (h) $5|\Delta \boldsymbol{S}_1| \cdot (|\boldsymbol{U}^{\dagger}| > 0)$

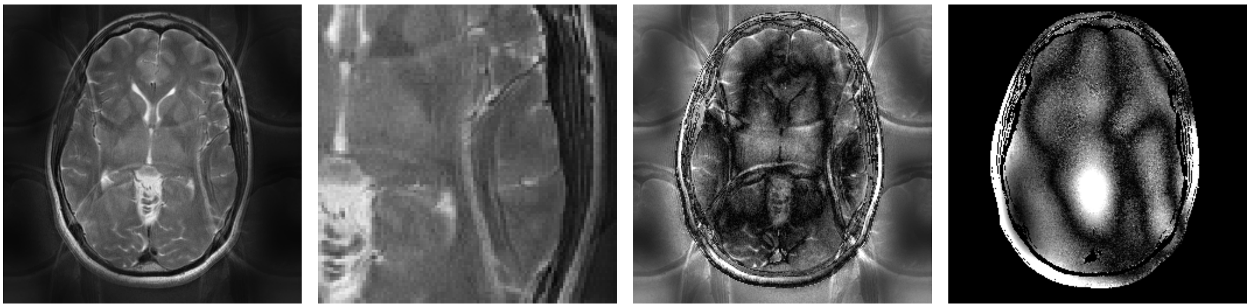


(i) $|\boldsymbol{U}|$ ($\mu = 0$) (j) $|\boldsymbol{U}|$ ($\mu = 0$, zoom) (k) $5|\Delta \boldsymbol{U}|$ (l) $5|\Delta \boldsymbol{S}_1| \cdot (|\boldsymbol{U}^{\dagger}| > 0)$

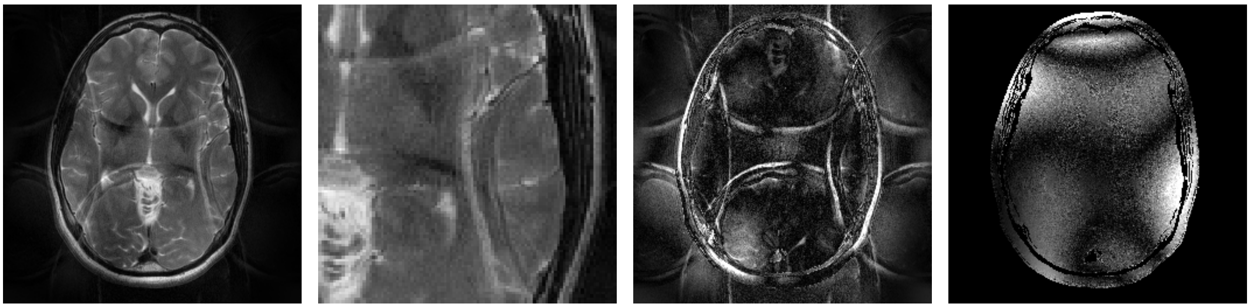
Figure 3: The top row shows the fully sampled gold standard image $|\boldsymbol{U}^{\dagger}|$, the data $|\tilde{\boldsymbol{U}}_1|$ and the gold standard sensitivity $|\boldsymbol{S}_1^{\dagger}|$. In the subsequent rows, the TVSENSE approach is compared for the cases that TV regularization is used or not. The reconstruction $|\boldsymbol{U}|$ is shown in the first two columns for the indicated method, and the corresponding error images $5|\Delta \boldsymbol{U}|$ and $5|\Delta \boldsymbol{S}_1| \cdot (|\boldsymbol{U}^{\dagger}| > 0)$ are shown respectively in the third and fourth columns, where the multiplicative factor of 5 is used to emphasize differences. All resulting images are shown on the scale $[0, 1]$.



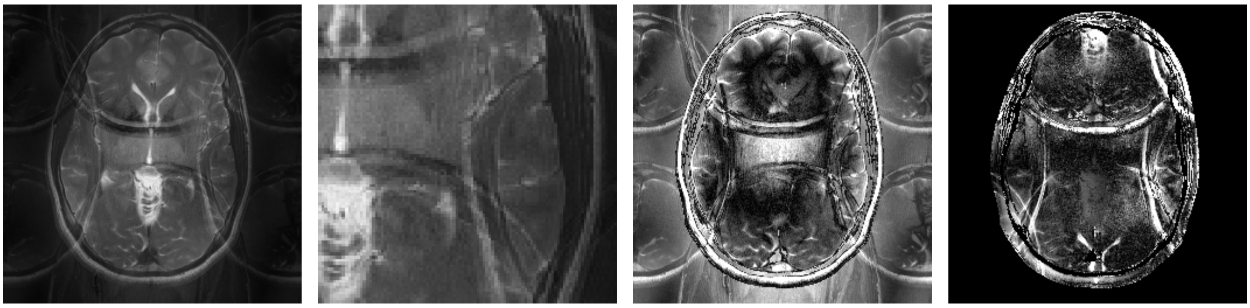
(a) $|\boldsymbol{U}|$ (TVSENSE) (b) $|\boldsymbol{U}|$ (zoom) (c) $5|\Delta\boldsymbol{U}|$ (d) $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$



(e) $|\boldsymbol{U}|$ (IRGN) (f) $|\boldsymbol{U}|$ (zoom) (g) $5|\Delta\boldsymbol{U}|$ (h) $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$

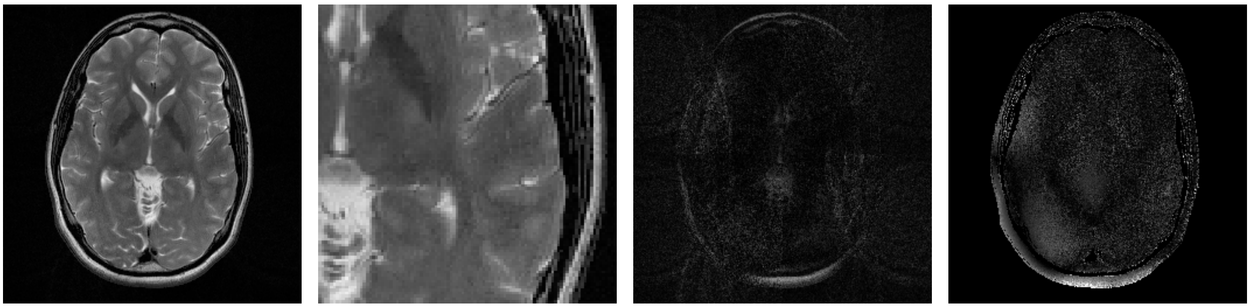


(i) $|\boldsymbol{U}|$ (SENSE) (j) $|\boldsymbol{U}|$ (zoom) (k) $5|\Delta\boldsymbol{U}|$ (l) $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$

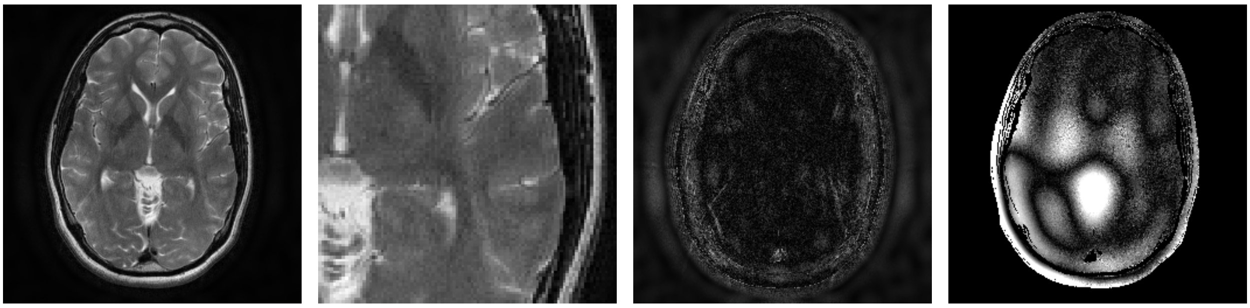


(m) $|\boldsymbol{U}|$ (GRAPPA) (n) $|\boldsymbol{U}|$ (zoom) (o) $5|\Delta\boldsymbol{U}|$ (p) $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$

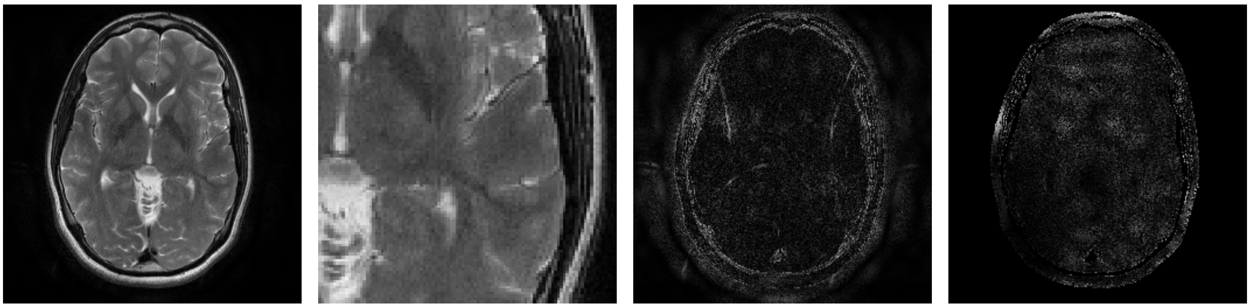
Figure 4: For $N_\mathrm{r} = 8$ the methods TVSENSE, IRGN, SENSE and GRAPPA are compared, where results are shown in the respective rows. The fully sampled gold standard image $|\boldsymbol{U}^\dagger|$ is shown in Figure 3a. The reconstruction $|\boldsymbol{U}|$ is shown in the first two columns for the indicated method, and the corresponding error images $5|\Delta\boldsymbol{U}|$ and $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$ are shown respectively in the third and fourth columns, where the multiplicative factor of 5 is used to emphasize differences. All resulting images are shown on the scale $[0, 1]$.



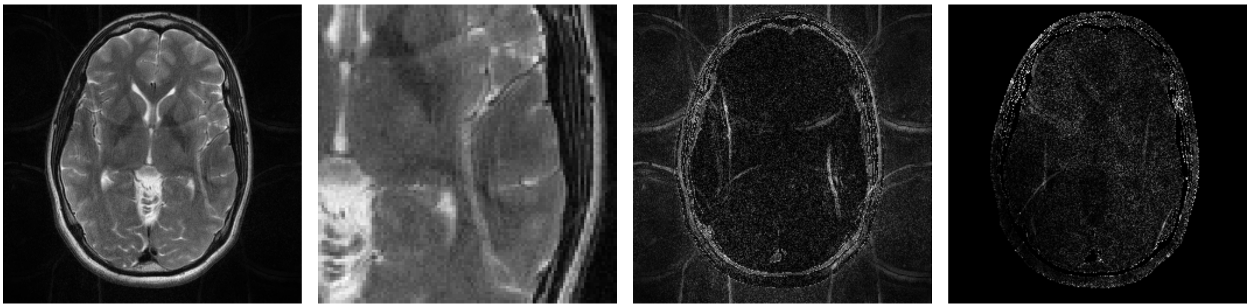
(a) $|\boldsymbol{U}|$ (TVSENSE) (b) $|\boldsymbol{U}|$ (zoom) (c) $5|\Delta\boldsymbol{U}|$ (d) $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$



(e) $|\boldsymbol{U}|$ (IRGN) (f) $|\boldsymbol{U}|$ (zoom) (g) $5|\Delta\boldsymbol{U}|$ (h) $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$



(i) $|\boldsymbol{U}|$ (SENSE) (j) $|\boldsymbol{U}|$ (zoom) (k) $5|\Delta\boldsymbol{U}|$ (l) $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$



(m) $|\boldsymbol{U}|$ (GRAPPA) (n) $|\boldsymbol{U}|$ (zoom) (o) $5|\Delta\boldsymbol{U}|$ (p) $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$

Figure 5: For $N_\mathrm{r} = 56$ the methods TVSENSE, IRGN, SENSE and GRAPPA are compared, where results are shown in the respective rows. The fully sampled gold standard image $|\boldsymbol{U}^\dagger|$ is shown in Figure 3a. The reconstruction $|\boldsymbol{U}|$ is shown in the first two columns for the indicated method, and the corresponding error images $5|\Delta\boldsymbol{U}|$ and $5|\Delta\boldsymbol{S}_1| \cdot (|\boldsymbol{U}^\dagger| > 0)$ are shown respectively in the third and fourth columns, where the multiplicative factor of 5 is used to emphasize differences. All resulting images are shown on the scale $[0, 1]$.

While the computing time for methods other than TVSENSE are comparable, the computational costs for TVSENSE are significantly higher. Nevertheless, optimized implementations on specialized hardware such as GPUs can narrow this gap in speed and thereby yield practically feasible computing times.

## 6. Conclusion

In this work the Cartesian parallel magnetic imaging problem has been formulated variationally in image space, as with SENSE (Pruessmann et al., 1999), as opposed to frequency space, as with GRAPPA (Griswold et al., 2002). The proposed approach employs a high order penalty for coil sensitivities and a total variation like penalty for the reconstructed image.

Based upon the authors' previous works, the optimality system for the formulated objective function is derived, numerically discretized and efficiently solved. The objective function used is non-convex, but it is seen to possess a bilinear structure as illustrated in Figure 2. Specifically, regularization resolves the technical ambiguity among possible solutions, but to resolve the practical ambiguity due to the flat landscape, the reconstructed image is normalized to possess a pre-estimated norm as seen in (3.11).

Computational results for *in vivo* MR imaging data show that a significant improvement in reconstruction quality is possible using TVSENSE in relation to GRAPPA, SENSE or IRGN, particularly when the amount of additional reference information is quite limited. Further work will be concerned with efficient GPU implementations of TVSENSE as well as its extension to different sampling strategies, e.g., radial or pseudorandom.

## Acknowledgments

The authors wish to express their thanks to the referees for their constructive comments which helped to improve the manuscript significantly. The authors also gratefully acknowledge support from the Austrian Science Fund *Fond zur Förderung der Wissenschaftlichen Forschung* (FWF) for financial support under grant SFB F032 ("Mathematical Optimization and Applications in Biomedical Sciences", `http://math.uni-graz.at/mobis`).

## References

Aubert, G., Kornprobst, P., 2006. Mathematical problems in image processing. volume 147 of *Applied Mathematical Sciences*. Springer, New York. second edition.

Aujol, J., Chambolle, A., 2005. Dual norms and image decomposition models. International Journal of Computer Vision 63, 85–104.

Bammer, R., Schoenberg, S.O., 2004. Current concepts and advances in clinical parallel magnetic resonance imaging. Top Magn Reson Imaging 15, 129–158.

Bernstein, M.A., King, K.F., Zhou, X.J., 2004. Handbook of MRI Pulse Sequences. Academic Press.

Blaimer, M., Breuer, F., Mueller, M., Heidemann, R.M., Griswold, M.A., Jakob, P.M., 2004. SMASH, SENSE, PILS, GRAPPA: how to choose the optimal method. Top Magn Reson Imaging 15, 223–236.

Block, K.T., Uecker, M., Frahm, J., 2007. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. Magn Reson Med 57, 1086–1098.

Bredies, K., Kunisch, K., Pock, T., 2010. Total generalized variation. SIAM Journal on Imaging Sciences 3, 492–526.

Chan, T.F., Golub, G.H., Mulet, P., 1999. A nonlinear primal-dual method for total variation-based image restoration. SIAM J. Sci. Comput. 20, 1964–1977.

Table 2: Shown for each case of Figure 4 are the quantitative errors, measured in the metrics of (5.1), as well as the computational time in seconds and the number of iterations where applicable. Here, $\hat{\boldsymbol{S}}_1 = \boldsymbol{S}_1 \cdot (|\boldsymbol{U}^\dagger| > 0)$ and $\hat{\boldsymbol{S}}_1^\dagger = \boldsymbol{S}_1^\dagger \cdot (|\boldsymbol{U}^\dagger| > 0)$. Since GRAPPA is not iterative, no entry is given in the last cell of the table.

| method | $d_2(\boldsymbol{U}, \boldsymbol{U}^\dagger)$ | $d_\infty(\boldsymbol{U}, \boldsymbol{U}^\dagger)$ | $d_2(\hat{\boldsymbol{S}}_1, \hat{\boldsymbol{S}}_1^\dagger)$ | $d_\infty(\hat{\boldsymbol{S}}_1, \hat{\boldsymbol{S}}_1^\dagger)$ | time | iterations |
|---|---|---|---|---|---|---|
| TVSENSE | 0.0283 | 0.172 | 0.0384 | 0.379 | 11544 | $244_{L^2} + 38_{\text{TV}}$ |
| IRGN | 0.0782 | 0.366 | 0.0749 | 0.485 | 5.7 | 3 |
| SENSE | 0.0526 | 0.329 | 0.0546 | 0.353 | 3.4 | 43 |
| GRAPPA | 0.0956 | 0.442 | 0.0706 | 0.680 | 6.5 | – |

Table 3: Shown for each case of Figure 5 are the quantitative errors, measured in the metrics of (5.1), as well as the computational time in seconds and the number of iterations where applicable. Here, $\hat{\boldsymbol{S}}_1 = \boldsymbol{S}_1 \cdot (|\boldsymbol{U}^\dagger| > 0)$ and $\hat{\boldsymbol{S}}_1^\dagger = \boldsymbol{S}_1^\dagger \cdot (|\boldsymbol{U}^\dagger| > 0)$. Since GRAPPA is not iterative, no entry is given in the last cell of the table.

| method | $d_2(\boldsymbol{U}, \boldsymbol{U}^\dagger)$ | $d_\infty(\boldsymbol{U}, \boldsymbol{U}^\dagger)$ | $d_2(\hat{\boldsymbol{S}}_1, \hat{\boldsymbol{S}}_1^\dagger)$ | $d_\infty(\hat{\boldsymbol{S}}_1, \hat{\boldsymbol{S}}_1^\dagger)$ | time | iterations |
|---|---|---|---|---|---|---|
| TVSENSE | 0.0283 | 0.137 | 0.0246 | 0.301 | 6307 | $65_{L^2} + 11_{\mathrm{TV}}$ |
| IRGN | 0.0269 | 0.156 | 0.0714 | 0.526 | 22.0 | 4 |
| SENSE | 0.0229 | 0.183 | 0.0204 | 0.237 | 13.4 | 179 |
| GRAPPA | 0.0321 | 0.204 | 0.0210 | 0.378 | 9.6 | – |

Clason, C., von Winckel, G., 2010. On a bilinear optimization problem in parallel magnetic resonance imaging. Appl. Math. Comput. 216, 1443–1452.

Ekeland, I., Témam, R., 1999. Convex Analysis and Variational Problems. volume 28 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. english edition.

Evans, L.C., 2010. Partial Differential Equations. volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI. second edition.

Fessler, J., 2010. Model-based image reconstruction for MRI. Signal Processing Magazine, IEEE 27, 81–89.

Golub, G.H., Van Loan, C.F., 1996. Matrix computations. Johns Hopkins Studies in the Mathematical Sciences, Johns Hopkins University Press, Baltimore, MD. third edition.

Griswold, M.A., Jakob, P.M., Heidemann, R.M., Nittka, M., Jellus, V., Wang, J., Kiefer, B., Haase, A., 2002. Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 47, 1202–1210.

Heidemann, R.M., Ozsarlak, O., Parizel, P.M., Michiels, J., Kiefer, B., Jellus, V., Mller, M., Breuer, F., Blaimer, M., Griswold, M.A., Jakob, P.M., 2003. A brief review of parallel magnetic resonance imaging. Eur Radiol 13, 2323–2337.

Hintermüller, M., Keeling, S.L., Laurain, A., 2008. Modulation recovery and image reconstruction in MRI: a structural study by parameterization. Technical Report 2008-023. MOBIS.

Hintermüller, M., Kunisch, K., 2004. Total bounded variation regularization as a bilaterally constrained optimization problem. SIAM J. Appl. Math. 64, 1311–1333 (electronic).

Hintermüller, M., Laurain, A., 2009. Multiphase image segmentation and modulation recovery based on shape and topological sensitivity. J. Math. Imaging Vision 35, 1–22.

Hintermüller, M., Stadler, G., 2006. An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration. SIAM J. Sci. Comput. 28, 1–23 (electronic).

Keeling, S.L., 2003. Total variation based convex filters for medical imaging. Appl. Math. Comput. 139, 101–119.

Keeling, S.L., Bammer, R., 2004. A variational approach to magnetic resonance coil sensitivity estimation. Appl. Math. Comput. 158, 359–388.

Keeling, S.L., Haase, G., 2007. Geometric multigrid for high-order regularizations of early vision problems. Appl. Math. Comput. 184, 536–556.

Keeling, S.L., Hintermller, M., Knoll, F., Kraft, D., Laurain, A., 2011. A total variation based approach to correcting surface coil magnetic resonance images. Applied Mathematics and Computation In Press, Corrected Proof.

Knoll, F., Bredies, K., Pock, T., Stollberger, R., 2011. Second order total generalized variation (TGV) for MRI. Magn Reson Med 65, 480–491.

Larkman, D.J., Nunes, R.G., 2007. Parallel magnetic resonance imaging. Phys Med Biol 52, R15–R55.

Liang, D., Liu, B., Wang, J., Ying, L., 2009. Accelerating SENSE using compressed sensing. Magn Reson Med 62, 1574–1584.

Liu, B., Ying, L., Steckner, M., Xie, J., Sheng, J., 2007. Regularized SENSE reconstruction using iteratively refined total variation method, in: Proc. 4th IEEE Int. Symp. Biomedical Imaging: From Nano to Macro ISBI 2007, pp. 121–124.

Lustig, M., Donoho, D., Pauly, J.M., 2007. Sparse MRI: The application of compressed sensing for rapid MR imaging. Magn Reson Med 58, 1182–1195.

Lustig, M., Pauly, J.M., 2010. SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space. Magn Reson Med 64, 457–471.

Pruessmann, K.P., Weiger, M., Börnert, P., Boesiger, P., 2001. Advances in sensitivity encoding with arbitrary k-space trajectories. Magn Reson Med 46, 638–651.

Pruessmann, K.P., Weiger, M., Scheidegger, M.B., Boesiger, P., 1999. SENSE: sensitivity encoding for fast MRI. Magn Reson Med 42, 952–962.

Scherzer, O., Grasmair, M., Grossauer, H., Haltmeier, M., Lenzen, F., 2009. Variational Methods in Imaging. volume 167 of *Applied Mathematical Sciences*. Springer, New York.

Uecker, M., Hohage, T., Block, K.T., Frahm, J., 2008. Image reconstruction by regularized nonlinear inversion – joint estimation of coil sensitivities and image content. Magn Reson Med 60, 674–682.

Wang, J., Kluge, T., Nittka, M., Jellus, V., Kuehn, B., Kiefer, B., 2001. Parallel acquisition techniques with modified SENSE reconstruction mSENSE, in: Proceedings of the First Wurzburg Workshop on Parallel Imaging, p. 92.