

# Medical University of Graz

## ${\bf Spezial For schungs Bereich}\ {\bf F\,32}$

Karl–Franzens Universität Graz Technische Universität Graz Medizinische Universität Graz





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

Stephen L. Keeling

Christian Clason

Michael Hintermüller Florian Knoll Antoine Laurain Gregory Von Winckel

SFB-Report No. 2008–022 $\,$ 

December 2008

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Supported by the Austrian Science Fund (FWF)



SFB sponsors:

- Austrian Science Fund (FWF)
- University of Graz
- Graz University of Technology
- Medical University of Graz
- Government of Styria
- City of Graz



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

Stephen L. Keeling<sup>1</sup>, Christian Clason<sup>1</sup>, Michael Hintermüller<sup>2</sup>, Florian Knoll<sup>3</sup>, Antoine Laurain<sup>1</sup>, Gregory Von Winckel<sup>1</sup>

Abstract. The Cartesian parallel magnetic imaging problem is formulated variationally using a high order penalty for modulations and a total variation like penalty for the image. Then the optimality system is derived and numerically discretized. The full problem is first considered in terms of the subproblems of modulation correction and aliasing correction. The cost functional used is non-convex, but the derivative of the cost has a bilinear residual term, and the functional is convex in each single argument. Thus, convex analysis is used to formulate the optimality condition for the image in terms of a primal-dual system. Also, a nonlinear Gauss-Seidel iteration is used to minimize with respect to one variable after the other using Newton's method. Favorable computational results are shown for artifical phantoms as well as for realistic magnetic resonance images.

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

## 1 Introduction

Magnetic Resonance Imaging (MRI) is a medical imaging method in which radio frequency coils, such as those shown in Fig. 1, are used for both nuclear excitation and for signal detection in order to measure the hydrogen atom density distribution in the human body; thus, displaying this distribution permits visualization of tissues with varying density. The density distribution is not measured directly by a coil; rather, it is encoded during the measurement process by applying external gradient magnetic fields so that the phase and the frequency of the timedependent radio pulse echo correspond to source location while the amplitude of the pulse corresponds to the density. Specifically, such raw data are shown in Fig. 2 with respect to frequency and phase axes, and the magnitude of a Fourier Transform of these raw data gives the density distribution as seen in Fig. 3. (For a full discussion of the principles of MRI, see, e.g., [13], [23].)

On the one hand, a large homogeneous coil such as the body coil shown in Fig. 1a, may be used for the uniform illumination of a volume as shown in Fig 3a. On the other hand, a smaller surface coil such as those shown in Fig. 1b can be used to resolve local details with greater sensitivity near the coil center but with an illumination falling off with the distance from the coil center as seen in Fig. 3b. Aside from spatial resolution, achieving the temporal resolution necessary for dynamic examinations, as in [17], is particularly challenging since each line in Fig. 2 must be acquired separately in the current implementation of MRI. A standard approach for accelerating measurement is to acquire only a subset of these lines. However, such Cartesian or line-wise subsampling leads to the aliasing shown in Fig. 4. (See [13], [23] for a full discussion of aliasing.) To compensate for the aliasing effect the approach of Parallel Magnetic Resonance Imaging (PMRI) is used in which multiple independent surface coils measure the radio echo simultaneously in a complementary fashion allowing reconstruction. Specifically, each image

<sup>&</sup>lt;sup>1</sup>Institut für Mathematik und Wissenschaftliches Rechnen, Karl-Franzens-Universität Graz, Heinrichstraße 36, 8010 Graz, Austria; email: firstname.lastname@uni-graz.at.

<sup>&</sup>lt;sup>2</sup>Department of Mathematics, University of Sussex, Falmer, Brighton, BN1 9RF, United Kingdom; email: M.Hintermueller@sussex.ac.uk.

<sup>&</sup>lt;sup>3</sup>Institut für Medizintechnik, Technische Universität Graz, Kronesgasse 5/II, 8010 Graz, Austria; email: florian.knoll@tugraz.at.

All authors are supported by the Austrian Science Fund Fond zur Förderung der Wissenschaftlichen Forschung (FWF) under grant SFB F032 and are linked on the SFB webpage http://math.uni-graz.at/mobis/.



**Figure 1**: (a) Shown on the left is a magnetic resonance body coil in which a reclining patient may be situated. (Used with permission of GE Medical Systems.) (b) Shown on the right are smaller surface coils mounted on a head rack. (See [15].)



**Figure 2**: Shown on the left and right respectively are (a) the real and (b) the imaginary part of the raw measurement, for which the magnitude of a Fourier Transform gives the image shown in Fig. 3a. In both (a) and (b), the horizontal axis corresponds to the measured pulse frequency while the vertical axis corresponds to the measured pulse frequency while the vertical axis corresponds to the measured pulse phase.



**Figure 3:** (a) Shown on the left is an image measured with a body coil as in Fig. 1a. This is the magnitude of a Fourier Transform of the data shown in Fig. 2. (b) Shown in the middle is an image measured with a surface coil as in Fig. 1b, where the brightest point of the image is nearest to the coil center. (c) Shown on the right is the coil sensitivity of the middle image, displayed only on the support of the first image.



**Figure 4**: Shown on the left in (a) is the subset (in white) of lines taken from Fig. 2 to compute the aliased image (b) shown on the right, as opposed to the fully sampled image in Fig. 3a. Specifically, the subsampling consists here of every fourth horizontal line, and the aliasing consists of vertically copying the fully sampled image (intensities divided by four) so that slicing the result at the arrow-marked heights gives four identical horizontal strips.



**Figure 5**: Modulated and aliased images measured by four surface coils such as shown in Fig. 1 using the subsampling shown in Fig. 4. Note that the intensities are reduced by modulation (nonuniform) as well as by division by four (uniform) as in Fig. 4.

measured by a surface coil is corrupted both by the modulation effect seen in Fig. 3 as well as by the aliasing effect seen in Fig. 4; however, the coils are typically placed in a circle as shown in Fig. 1b, so the modulations (the so-called sensitivities) are independent. The goal of PMRI is thus to reconstruct the image shown in Fig. 3a from images such as those shown in Fig. 5.

Standard reconstruction strategies currently in use include SENSE [19] and GRAPPA [8]. The SENSE approach is based in image space and it involves the use of initial reference images, such as in Fig. 3a and 3b, to estimate their quotient shown in Fig. 3c. Knowledge of these sensitivities and the subsampling strategy allows an algebraic reconstruction of images underlying subsequent coil measurements. 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. The GRAPPA approach is based



**Figure 6:** Shown at the upper left in (a) is the subset (in white) of lines taken from Fig. 2 to compute the aliased image (b) shown at the upper right. Note in relation to Fig. 4 that center-lines are included, and the resulting image (b) is nearer to the fully sampled image in Fig. 3a than is the image in Fig. 4b. Shown at the lower left in (c) is the subset (in white) of lines and points taken from Fig. 2 to compute the aliased image (d) shown at the lower right. Note in relation to the result in the first row with whole horizontal reference lines, that this result with only a small square of reference points is comparable.

in frequency space and it involves the interpolation of missing lines based upon additionally acquired so-called center-lines near the origin as shown in Fig. 6a. Specifically, the data on the lines of Fig. 4a nearest to the center-lines are used to construct an interpolation kernel which best fits the additionally measured data on the center-lines of Fig. 6a, and this interpolation kernel is used to interpolate missing data on the remaining unsampled lines. The image shown in Fig. 6b shows the improvement over that in Fig. 4b when the additional data from the centerlines are used. Note also that when measurements are performed in three dimensions, lines of data are acquired orthogonal to the image plane, and these may include just a few additional reference points near the origin as seen in the center square of Fig. 6c. The resulting image in Fig. 6d is comparable to that obtained with whole horizontal lines as shown in Fig. 6b; thus, in certain computational contexts, it may be advantageous to use only the few low frequencies in Fig. 6c, even if the whole lines of Fig. 6a are measured. The additional low frequency information used in Fig. 6 strengthens the previous image of Fig. 4b by adding a low resolution version of the full image in Fig. 3a. Nevertheless, using these additional data to interpolate the remaining unsampled data according to the GRAPPA approach is prone to errors since the points in frequency space are in general uncorrelated.

Nonlinear least squares methods for PMRI have been proposed recently [1] [21]. These methods are based in frequency space and so use a spectral formulation to impose high-order Sobolev regularity on sensitivities and  $L_2$  regularity on the reconstructed image. Note that such spectral formulations implicitly impose an unnatural periodicity on sensitivities. Also, these approaches rely on the measurement of center-lines. The present authors were motivated to consider approaches to PMRI which are free of the use of center-lines and of pre-measured reference images, partly because of the success of modulation (not aliasing) corrections without any outside information besides a given corrupted image. (See, e.g., [22] as well as the results in Subsection 2.4.) The authors' corresponding investigations are reported separately in detail in [9] and [4], where special cost functionals as well as tailored optimization techniques are studied. However, it will be seen in results below that reconstructions are significantly better with the additional data such as shown in Fig. 6. Thus, the present challenge is to reduce the cost of such additional measurements as much as possible while retaining a suitable reconstruction quality.

In particular, the approach set forth in this paper performs well by using only a very small number of reference points as seen in Fig. 6c. The proposed approach is variationally formulated with second order derivative penalties for sensitivities and a total variation like penalty for the image. Specifically, the formulation for sensitivities is based upon [15] and [16], and the total variation formulation is based upon the primal-dual formulation of [10], as opposed to others such as [3] or [11]. The use of total variation regularization is found below to yield results superior to those obtained by  $L_2$  regularization.

With respect to the required minimization, note that the least-squares residual in the cost functional for parallel imaging has a derivative which is bilinear in the unknown image and sensitivities. Using only first derivative information, as in a gradient descent scheme, can be inefficient because of poor scaling observed in the gradient components corresponding to the sensitivities as opposed to the image. On the other hand, the non-convexity of the cost functional makes its Hessian in general indefinite and thus a pure Newton's method is not suitable, as demonstrated in [4]. In this work a nonlinear Gauss-Seidel scheme is used to solve the optimality system by applying a Newton scheme to solve for one variable after the other. To regularize the computation of sensitivities, a segmentation of the image is also used, which is based on the use of topological derivatives as in the work of the authors [12].

The paper is organized as follows. Before considering the full parallel imaging problem, the subproblems of modulation correction and aliasing correction are first considered in Section 2. These subproblems are formulated variationally and their optimality systems, their numerical discretizations and their solutions are treated in subsections. For the methods formulated it is found that pure modulation corrections can be achieved reliably, but that pure aliasing corrections cannot be achieved for Cartesian subsampling. With this framework the stage is set for the analogous constructions for the complete parallel imaging problem in Section 3. As with the subproblems, the full problem is formulated variationally and its optimality system, its numerical discretization and its solution are treated in subsections. In spite of the failure of pure aliasing correction for Cartesian subsampling, it is found for the approach proposed for parallel imaging that modulation correction and aliasing correction can indeed be performed simultaneously when several measurements are available. On the other hand, it is found that the number of coils should be larger than the aliasing folding factor, and that at least a minimal number of additional measurement reference points in frequency space, as shown in Fig. 6, are necessary for accurate reconstructions.

# 2 Corrections of Single Surface Coil Images

A natural approach to reconstructing uncorrupted images, similar to that shown in Fig. 3a, from modulated and aliased images, such as those shown in Section 1, is first to correct the corrupted images individually before combining them for the reconstruction. In this section, the potentials and limitations of this approach are elucidated. In the following subsections, pure modulation corrections are first investigated, and then pure aliasing corrections are considered.

#### 2.1 Modulation Correction

For instance, consider first the estimation of the modulation  $\sigma$  of Fig. 3c, as well as the unmodulated image u in Fig. 3a, simply from the image  $\tilde{u}$  of Fig. 3b. The desired unknowns  $\sigma$  and u satisfy  $\sigma u \approx \tilde{u}$  and are determined here by minimizing the sum of a residual term plus

regularization terms:

$$J(u,\sigma) = \frac{1}{2} \int_{\Omega} |\sigma u - \tilde{u}|^2 d\boldsymbol{x} + \frac{\nu}{2} \int_{\Omega} |\nabla^2 \sigma|^2 d\boldsymbol{x} + \frac{\kappa}{2} \int_{\Omega} u^2 d\boldsymbol{x} + \mu \int_{\Omega} \phi_{\epsilon}(|\nabla u|^2) d\boldsymbol{x}$$
(2.1)

Here  $\Omega = (0, 1)^d$  is the image domain, with d = 2 in the examples of this work, but there is no fundamental restriction on the dimension. While the raw data such as in Fig. 2 are complexvalued, it is assumed here for simplicity that their inverse Fourier Transform is real-valued<sup>5</sup> and therefore agrees with the magnitude images such as shown in Fig. 3. Thus, the arguments of J in (2.1) are real-valued; furthermore, they are considered to be mappings  $u, \sigma : \Omega \to [0, 1]$ , although the restriction of range is not explicitly enforced in (2.1), as would be the case, e.g., if barrier functions were added to (2.1). For (2.1), the  $\ell_2$  norm of the *n*th order derivative is given by:

$$|\nabla^n \sigma|^2 = \nabla^n \sigma \cdot \nabla^n \sigma, \qquad \nabla^n \sigma_1 \cdot \nabla^n \sigma_2 = \sum_{|\alpha|=n} \begin{pmatrix} n \\ \alpha! \end{pmatrix} \partial^\alpha \sigma_1 \partial^\alpha \sigma_2 \tag{2.2}$$

The penalty on the second derivative of the modulation  $\sigma$  seen in (2.1) is based upon work in [15] and [16]. Among the key points is first the fact that the modulation  $\sigma$  is much smoother than the image u. Also, to avoid that  $\sigma$  have values outside [0, 1], barrier functions have been considered as seen in [9] and [12], but it is found here in practice that the modulation is nonnegative on the image support, where it can also be scaled to be less than one. The values of  $\sigma$  outside the support of u are not important except in the way that values inside the support are influenced through smoothness of the modulation. In particular, the high order natural boundary conditions on  $\sigma$  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  $\int_{\Omega} |\Delta\sigma|^2 d\mathbf{x}$ is used, then harmonic functions are in the kernel of the penalty and boundary disturbances appear as seen in [15]. Also, when a spectral penalty is used such as  $\sum_{\mathbf{k}} (1+|\mathbf{k}|^2)^2 |\omega(\mathbf{k})|^2$ , where  $\{\omega(\mathbf{k})\}$  are Fourier or trigonometric series coefficients of  $\sigma$ , then  $\sigma$  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 [10] and [14],

$$\phi_{\epsilon}(s) = \begin{cases} s/(2\epsilon), & 0 \le s \le \epsilon^2\\ \sqrt{s} - \epsilon/2, & s \ge \epsilon^2 \end{cases}$$
(2.3)

which emerges naturally from the duality formulation as shown in [10] and as seen below in (A.28). The  $L^2$  regularization in (2.1) is included so that  $\kappa I + \sigma^2$  is invertible in (2.15) below even when  $\sigma$  becomes very small.

To examine the landscape of the functional in (2.1), as well as those with similar structure used later, consider the minimization of the following model function:

$$f(x,y) = \frac{1}{2}(yx-z)^2 + \frac{1}{2}\nu y^2 + \mu|x|$$
(2.4)

A contour plot of f is shown in Fig. 7 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, it is not surprising that computational experiments using Newton's method to minimize functionals

<sup>&</sup>lt;sup>5</sup>Physically this means that there are negligible differences among the chemical environments of the hydrogen atoms. When this is not the case, real and imaginary parts of the residual  $\sigma u - \tilde{u}$  may be processed separately as seen in [15].



Figure 7: Contour plots of f in (2.4) (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.

such as (2.1) have had limited success. Furthermore, using rapidly converging schemes to minimize a convex functional with respect to one variable and then the other has performed better than carrying out line searches along gradient directions for the joint functional.

This iterative method arrives quickly at the flat elongated region of the landscape. However, any of the above iterations can stall in such a region, leading numerically to an effective nonuniqueness in minimizers. To distinguish among such numerical minimizers, iterations may be guided by additional information. For instance, since the arguments of J are expected to have range in [0, 1], the modulation is normalized here in each iteration to achieve a maximum value of 1 on the support of the image. Note that a normalization of the modulation is more stable than that of the image since the modulation is smoother. Also, the modulation is regularized in early iterations by using a segmentation of the image in the modulation computation instead of the image itself. Such projection and regularization techniques have been found here to perform better than treating the nonconvexity of J by starting iterations with larger and ending with smaller regularization parameters in (2.1).

#### 2.2 Optimality Conditions

In this work  $H^k(\Omega)$  denotes the Sobolev space of functions with distributional derivatives up to order k in  $L^p(\Omega)$ ; see [5] for further information about these function spaces. As explained in [15], the optimality condition for (2.1) with respect to  $\sigma$  for fixed u is:

$$B(u)\sigma = u\tilde{u}, \qquad \sigma \in H^2(\Omega) \tag{2.5}$$

which is given in weak form as:

$$\int_{\Omega} \left[ \nu \nabla^2 \sigma \cdot \nabla^2 \bar{\sigma} + u^2 \sigma \bar{\sigma} \right] d\boldsymbol{x} = \int_{\Omega} \bar{\sigma} u \tilde{u} d\boldsymbol{x}, \quad \forall \bar{\sigma} \in H^2(\Omega).$$
(2.6)

According to [15] there is a unique weak solution  $\sigma \in H^2(\Omega)$  when  $\tilde{u} \in L^2(\Omega)$  holds and  $u \in L^{\infty}(\Omega)$  has a support with positive measure. The additional regularity  $\sigma \in H^4(\Omega)$  is shown in [16]. In strong form,  $B(u) = \nu \Delta^2 + u^2$ , and a smooth solution  $\sigma$  satisfies the natural boundary conditions  $\partial_n^3 \sigma = \partial_n^2 \sigma = \partial_n \partial_\tau \sigma = 0$ ,  $\partial\Omega$ , where  $\partial_n$  and  $\partial_\tau$  are the normal and tangential derivatives respectively [15].

In initial iterations of the nonlinear Gauss-Seidel method (2.5) is solved by replacing the image u with a segmentation S(u) based upon [12]. Specifically, the image is approximated by:

$$S(u) = \sum_{i=1}^{M} c_i \chi_i \tag{2.7}$$

where  $\chi_i$  is the characteristic function for a subdomain  $\Omega_i \subseteq \Omega$  in which the segmentation possesses the grey level  $c_i$ . These subdomains are disjoint and are determined so that the cost

 $J(\Omega_1, \ldots, \Omega_M) = \int_{\Omega} [S(u) - u]^2 d\mathbf{x}$  is minimized and so cannot be reduced by transfering part of one subdomain to another. The topological derivative  $T_{ij}(\mathbf{x})$  measures the rate of change of  $J(\Omega_1, \ldots, \Omega_M)$  when a ball  $B(\mathbf{x}, r)$  of vanishingly small radius r is transferred from  $\Omega_i$  to  $\Omega_j$ :

$$T_{ij}(\boldsymbol{x}) = \lim_{|B(\boldsymbol{x},r)| \to 0} \frac{J(\Omega_1, \dots, \Omega_i \setminus B(\boldsymbol{x}, r), \dots, \Omega_j \cup B(\boldsymbol{x}, r), \dots, \Omega_M) - J(\Omega_1, \dots, \Omega_M)}{|B(\boldsymbol{x}, r)|}$$
(2.8)

Here and below, |B| denotes the measure of the set |B|. The segmentation (2.7) is determined so that all derivatives  $T_{ij}$  are non-negative. As shown in [12] they are given explicitly as:

$$T_{i,j}(\boldsymbol{x}) = \begin{cases} [c_j - u(\boldsymbol{x})]^2 - [c_i - u(\boldsymbol{x})]^2, & |\Omega_j| \neq 0\\ -[c_i - u(\boldsymbol{x})]^2, & |\Omega_j| = 0 \end{cases}$$
(2.9)

for  $\boldsymbol{x} \in \Omega_i$ , and  $T_{i,j}(\boldsymbol{x}) = 0$  for  $\boldsymbol{x} \notin \Omega_i$ . In order that the topological derivative approach operate in a more global fashion, a point  $\boldsymbol{x}$  is transferred from  $\Omega_i$  to  $\Omega_j$  only when

$$T_{i,j}(\boldsymbol{x}) < \gamma \min_{\boldsymbol{y} \in \Omega_j} T_{i,j}(\boldsymbol{y})$$
(2.10)

where  $\gamma \in (0, 1)$ . Given the subdomains, the grey levels  $\{c_i\}$  are given by the following:

$$c_i = \frac{1}{|\Omega_i|} \int_{\Omega_i} u d\boldsymbol{x}; \quad |\Omega_i| \neq 0, \qquad c_i = 0, \quad |\Omega_i| = 0.$$
(2.11)

See Algorithm 2 below for the details of determining S(u).

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

$$\mathcal{F}(u) = \frac{1}{2} \int_{\Omega} |\sigma u - \tilde{u}|^2 d\boldsymbol{x} + \frac{\kappa}{2} \int_{\Omega} u^2 d\boldsymbol{x}$$
(2.12)

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

$$\mathcal{G}(u) = \frac{\mu}{2} \int_{\Omega} \phi_{\epsilon}(|\nabla u|^2) d\boldsymbol{x}$$
(2.13)

so that the dependence on u in J is given by  $\mathcal{F}(u) + \mathcal{G}(u)$ .<sup>6</sup> Note that these operators satisfy the conditions of the Fenchel Duality Theorem, and the desired optimality condition is thus given by [20]:

$$\begin{cases} \mathcal{F}(u) + \mathcal{F}^*(v) = \int_{\Omega} uv d\boldsymbol{x} \\ \mathcal{G}(u) + \mathcal{G}^*(-v) = -\int_{\Omega} uv d\boldsymbol{x} \end{cases}$$
(2.14)

where the convex conjugates  $\mathcal{F}^*$  and  $\mathcal{G}^*$  in (2.14) are given as follows; see Appendix A for details. First,  $\mathcal{F}^* : L^2(\Omega) \to \mathbb{R} \cup \infty$  is given by:

$$\mathcal{F}^*(v) = \frac{1}{2} \int_{\Omega} \left[ (\kappa + \sigma^2)^{-1} (v + \sigma \tilde{u})^2 - \tilde{u}^2 \right] d\boldsymbol{x}$$
(2.15)

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

$$\mathcal{G}^*(v) = I_{S_{\mu}}(v) + \frac{\epsilon}{2\mu} \int_{\Omega} |\nabla(\Delta_{\mathrm{N}}^{-1}v)|^2 d\boldsymbol{x}$$
(2.16)

Here  $I_S$  denotes the indicator function of the set S, and in particular,

$$I_{S_{\mu}}(v) = \begin{cases} 0, & v \in S_{\mu} \\ \infty, & \text{otherwise} \end{cases} \quad S_{\mu} = \left\{ v \in L^{2}(\Omega) : |\nabla(\Delta_{N}^{-1}v)| \le \mu, \int_{\Omega} v d\boldsymbol{x} = 0 \right\}$$
(2.17)

<sup>&</sup>lt;sup>6</sup>The authors wish to thank Otmar Scherzer for his suggestion that these functionals be defined on  $L^2$ ; see also the recent book [7].

Also,  $\Delta_{\rm N}^{-1}: L^2(\Omega) \to H^2(\Omega)$  is the solution operator for the boundary value problem,

$$\begin{cases} \Delta w = v, \ \Omega \\ \partial w/\partial n = 0, \ \partial \Omega \end{cases} \qquad \int_{\Omega} w d\boldsymbol{x} = \int_{\Omega} v d\boldsymbol{x} = 0 \qquad (2.18)$$

where  $\partial w/\partial n = \nabla w \cdot n$  and n is the outwardly directed unit normal vector at  $\partial \Omega$ . As seen in Appendix A the functionals (2.12), (2.13), (2.15) and (2.16) lead to the following formulation of the optimality system (2.14):

$$\begin{cases} (\kappa + \sigma^2)u - \Delta_{\mathcal{N}}w &= \sigma \tilde{u} \\ -\mu \nabla u + \lfloor \nabla u \rfloor_{\epsilon} \nabla w &= 0 \end{cases} \qquad u \in L^2(\Omega), \quad \Delta_{\mathcal{N}}w \in S_{\mu}$$
(2.19)

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

#### 2.3 Numerical Methods

The discretization of the optimality conditions in Subsection 2.2 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  $\mathbf{j} = (j_1, j_2, \ldots)$ ,  $\mathbf{0} = (0, 0, \ldots)$ , and  $\mathbf{1} = (1, 1, \ldots)$ , the cell centroids are  $\mathbf{x}_{\mathbf{j}} = (\mathbf{j} - \frac{1}{2})h$ ,  $\mathbf{1} \leq \mathbf{j} \leq N \cdot \mathbf{1}$ . Then,  $U_{\mathbf{j}} \approx u(\mathbf{x}_{\mathbf{j}})$  and  $\mathbf{U}$  denotes the vector of values  $\{U_{\mathbf{j}}\}$  according to the lexicographic ordering in which  $j_1$  increments first from 1 to N, then  $j_2$ , and so on. Also, let  $D(\mathbf{U})$  denote the diagonal matrix with the values  $\{U_{\mathbf{j}}\}$  situated along the diagonal according to the lexicographic ordering.

Following [15], (2.5) is discretized by  $B(u) \approx \nu B_h + D(U)^2$ , 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  $\{x_j : 1 \leq j \leq 3 \cdot 1\}$ , where stencil weights are obtained by dividing the following by  $2800h^4$ :

0 0 0 0 0	0 0 0 0	$-152 \\ -592 \\ 4368 \\ -592 \\ -152$	$424 \\ -2176 \\ -2256 \\ -2176 \\ 424$	208 848 768 848 208	0 0 0 0 0	$424 \\ -2176 \\ -2256 \\ -2176 \\ 424$	$920 \\ -3920 \\ 20400 \\ -3920 \\ 920$	$848 \\ -4352 \\ -4512 \\ -4352 \\ 848$	208 848 768 848 208	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
0 0 0 0 0	0 0 0 0 0	$-152 \\ -592 \\ 3440 \\ -296 \\ 0$	$424 \\ -2176 \\ -1960 \\ -1088 \\ 0$	$208 \\ 848 \\ 920 \\ 424 \\ 0$	0 0 0 0 0	$424 \\ -2176 \\ -1960 \\ -1088 \\ 0$	$920 \\ -3920 \\ 16960 \\ -1960 \\ 0$	$848 \\ -4352 \\ -3920 \\ -2176 \\ 0$	$208 \\ 848 \\ 920 \\ 424 \\ 0$	$208 \\ 848 \\ 920 \\ 424 \\ 0$	$848 \\ -4352 \\ -3920 \\ -2176 \\ 0$	$768 \\ -4512 \\ 20400 \\ -2256 \\ 0$	$848 \\ -4352 \\ -3920 \\ -2176 \\ 0$	$208 \\ 848 \\ 920 \\ 424 \\ 0$
0 0 0 0 0	0 0 0 0		$424 \\ -1088 \\ -296 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ $	$208 \\ 424 \\ -152 \\ 0 \\ 0$	0 0 0 0 0	$424 \\ -1088 \\ -296 \\ 0 \\ 0 \\ 0$	$920 \\ -1960 \\ 3440 \\ 0 \\ 0 \\ 0$	$848 \\ -2176 \\ -592 \\ 0 \\ 0$	$208 \\ 424 \\ -152 \\ 0 \\ 0$	$208 \\ 424 \\ -152 \\ 0 \\ 0$	$848 \\ -2176 \\ -592 \\ 0 \\ 0$	$768 \\ -2256 \\ 4368 \\ 0 \\ 0 \\ 0$	$848 \\ -2176 \\ -592 \\ 0 \\ 0$	$208 \\ 424 \\ -152 \\ 0 \\ 0$

For a given u and  $\tilde{u}$  set  $U = \{u(\boldsymbol{x}_{j})\}$  and  $\tilde{U} = \{\tilde{u}(\boldsymbol{x}_{j})\}$ . Then the numerical solution to (2.5) is given as  $\boldsymbol{S} = \{S_{j}\}, S_{j} \approx \sigma(\boldsymbol{x}_{j})$  which solves:

$$[\nu B_h + D(\boldsymbol{U})^2]\boldsymbol{S} = D(\boldsymbol{U})\tilde{\boldsymbol{U}}$$
(2.20)

The matrix on the left side can be stored in sparse format, and the equation is solved in MATLAB<sup>7</sup> in the present work using backslash.

The segmentation computations in (2.7) - (2.11) are carried out by evaluating each function at the cell centroids  $\boldsymbol{x} = \boldsymbol{x}_{j}$ . See Algorithm 2 below for the details of determining  $S(\boldsymbol{U})$ .

The optimality system (2.19) is solved using a generalized Newton method. The system,

$$\begin{bmatrix} (\kappa + \sigma^2) & -\Delta_{\rm N} \\ [-\mu I + \frac{(|\nabla u| > \epsilon)}{|\nabla u|_{\epsilon}} \nabla u \nabla w^{\rm T}] \nabla & [\nabla u]_{\epsilon} \nabla \end{bmatrix} \begin{bmatrix} \delta u \\ \delta w \end{bmatrix} = -\begin{bmatrix} (\kappa + \sigma^2)u - \Delta_{\rm N}w - \sigma \tilde{u} \\ -\mu \nabla u + [\nabla u]_{\epsilon} \nabla w \end{bmatrix}$$
(2.21)

 $<sup>^{7}\</sup>mathrm{See} \ \mathtt{http://www.mathworks.com}$ 

is simplified by first eliminating the second equation to obtain:

$$\left\{ (\kappa + \sigma^2) - \nabla \cdot \left[ \frac{1}{\lfloor \nabla u \rfloor_{\epsilon}} \left( \mu I - \frac{(|\nabla u| > \epsilon)}{2 \lfloor \nabla u \rfloor_{\epsilon}} [\nabla u \, \boldsymbol{p}^{\mathrm{T}} + \boldsymbol{p} \, \nabla u^{\mathrm{T}}] \right) \right] \nabla \right\} \delta u = -(\kappa + \sigma^2) u + \sigma \tilde{u} + \mu \nabla \cdot \left( \frac{\nabla u}{\lfloor \nabla u \rfloor_{\epsilon}} \right)$$

$$(2.22)$$

and the eliminated equation becomes:

$$\boldsymbol{\delta p} = \frac{1}{\left\lfloor \nabla u \right\rfloor_{\epsilon}} \left( \mu I - \frac{\left( \left| \nabla u \right| > \epsilon \right)}{2 \left\lfloor \nabla u \right\rfloor_{\epsilon}} \left[ \nabla u \boldsymbol{p}^{\mathrm{T}} + \boldsymbol{p} \,\nabla u^{\mathrm{T}} \right] \right) \nabla \delta u + \mu \frac{\nabla u}{\left\lfloor \nabla u \right\rfloor_{\epsilon}} - \boldsymbol{p}$$
(2.23)

where  $\boldsymbol{p} = \nabla w \in \boldsymbol{H}_0(\text{div}) = \{ \boldsymbol{p} \in \boldsymbol{L}^2(\Omega) : \nabla \cdot \boldsymbol{p} \in L^2(\Omega), \ \boldsymbol{n} \cdot \boldsymbol{p} = 0, \ \partial \Omega \}.$  Note that the term  $\nabla u \boldsymbol{p}^{\mathrm{T}}$  has been symmetrized with  $[\nabla u \boldsymbol{p}^{\mathrm{T}} + \boldsymbol{p} \nabla u^{\mathrm{T}}]/2.$ 

To discretize (2.22) and (2.23) the discrete derivative matrices  $\nabla_h^{(x_i)}$  have stencils given explicitly as follows in  $\mathbb{R}^2$  for the cells with centroids  $\{x_j : 1 \leq j \leq 2 \cdot 1\}$ , where stencil weights are obtained by dividing the following by  $2h^2$ :

0 0 0	$\begin{smallmatrix}&0\\-1\\0\end{smallmatrix}$	$\begin{array}{c} 0 \\ 1 \\ 0 \end{array}$	$\begin{smallmatrix}&0\\-1\\0\end{smallmatrix}$	0 0 0	$\begin{array}{c} 0 \\ 1 \\ 0 \end{array}$	0 0 0	1 0 -1	000000000000000000000000000000000000000	0 0 0	$\begin{array}{c}1\\0\\-1\end{array}$	0 0 0
$\begin{array}{c} 0 \\ 0 \\ 0 \end{array}$	$\begin{smallmatrix}&0\\-1\\0\end{smallmatrix}$	$\begin{array}{c} 0 \\ 1 \\ 0 \end{array}$	$\begin{smallmatrix}&0\\-1\\0\end{smallmatrix}$	0 0 0	$\begin{array}{c} 0 \\ 1 \\ 0 \end{array}$	0 0 0	$-1 \\ -1 \\ 0$	0 0 0	0 0 0	$\begin{smallmatrix}&1\\-1\\0\end{smallmatrix}$	0 0 0

so that Neumann boundary conditions are implicitly implemented. Thus, the gradient and divergence are approximated by:

$$\nabla \approx \nabla_h = \begin{pmatrix} \nabla_h^{(x_1)} \\ \nabla_h^{(x_2)} \end{pmatrix}, \quad \nabla \cdot \approx -\nabla_h^{\mathrm{T}}$$
(2.24)

Also, functions of  $\nabla_h U$  are defined according to:

$$\begin{aligned} |\nabla_{h}\boldsymbol{U}| &= \left\{ [(\nabla_{h}^{(1)}\boldsymbol{U})_{\boldsymbol{j}}^{2} + (\nabla_{h}^{(2)}\boldsymbol{U})_{\boldsymbol{j}}^{2}]^{\frac{1}{2}} \right\} \qquad (|\nabla_{h}\boldsymbol{U}| > \epsilon) &= \left\{ (|\nabla_{h}\boldsymbol{U}|)_{\boldsymbol{j}} > \epsilon \right\} \\ \frac{1}{[\nabla_{h}\boldsymbol{U}]_{\epsilon}} &= \left\{ \frac{1}{\max\{(|\nabla_{h}\boldsymbol{U}|)_{\boldsymbol{j}},\epsilon\}} \right\} \qquad \frac{(|\nabla_{h}\boldsymbol{U}| > \epsilon)}{[\nabla_{h}\boldsymbol{U}]_{\epsilon}} &= D(|\nabla_{h}\boldsymbol{U}| > \epsilon)\frac{1}{[\nabla_{h}\boldsymbol{U}]_{\epsilon}} \end{aligned}$$
(2.25)

and  $P = (P_1; P_2)$  is understood below as a column vector of column vectors  $P_1$  and  $P_2$ . Thus, (2.22) is discretized as:

$$\left\{ \kappa I + D(\mathbf{S})^{2} + \nabla_{h}^{\mathrm{T}} \left[ D\left(\frac{1}{\left\lfloor \nabla_{h} \mathbf{U} \right\rfloor_{\epsilon}}\right) \left( \mu I - D\left(\frac{\left(\left|\nabla_{h} \mathbf{U}\right| > \epsilon\right)}{2\left\lfloor \nabla_{h} \mathbf{U} \right\rfloor_{\epsilon}}\right) \left[\nabla_{h} \mathbf{U} \mathbf{P}_{\mu}^{\mathrm{T}} + \mathbf{P}_{\mu} \nabla_{h} \mathbf{U}^{\mathrm{T}} \right] \right) \right] \nabla_{h} \right\} \boldsymbol{\delta} \mathbf{U} = - \left[ \kappa I + D(\mathbf{S})^{2} \right] \mathbf{U} + D(\mathbf{S}) \tilde{\mathbf{U}} - \mu \nabla_{h}^{\mathrm{T}} \left[ D\left(\frac{1}{\left\lfloor \nabla_{h} \mathbf{U} \right\rfloor_{\epsilon}}\right) \nabla_{h} \mathbf{U} \right]$$
(2.26)

where, following [10],  $P_{\mu}$  is a version of P truncated to have a cellwise magnitude not more than  $\mu$ :

$$\begin{aligned}
\mathbf{P}_{\mu} &= (D(|\mathbf{P}| < \mu)\mathbf{P}_{1}; D(|\mathbf{P}| < \mu)\mathbf{P}_{2}) \\
&+ \mu(D(|\mathbf{P}| \ge \mu)D(|\mathbf{P}|)^{-1}\mathbf{P}_{1}; D(|\mathbf{P}| \ge \mu)D(|\mathbf{P}|)^{-1}\mathbf{P}_{2}), \\
|\mathbf{P}| &= \{\sqrt{(\mathbf{P}_{1})_{j}^{2} + (\mathbf{P}_{2})_{j}^{2}}\}
\end{aligned}$$
(2.27)

As shown in [10], using the truncation  $P_{\mu}$  in (2.26) guarantees that  $\delta U$  provides a descent direction for the cost functional with fixed modulation. The matrix on the left in (2.26) can be stored in sparse format and the system is solved in MATLAB in the present work using backslash. The update for P itself is given by a discretization of (2.23):

$$\delta \boldsymbol{P} = D\left(\frac{1}{\lfloor \nabla \boldsymbol{U} \rfloor_{\epsilon}}\right) \left(\mu I - D\left(\frac{(|\nabla \boldsymbol{U}| > \epsilon)}{2\lfloor \nabla \boldsymbol{U} \rfloor_{\epsilon}}\right) \left[\nabla_{h} \boldsymbol{U} \boldsymbol{P}_{\mu}^{\mathrm{T}} + \boldsymbol{P}_{\mu} \nabla_{h} \boldsymbol{U}^{\mathrm{T}}\right]\right) \nabla_{h} \delta \boldsymbol{U} + \mu D\left(\frac{1}{\lfloor \nabla \boldsymbol{U} \rfloor_{\epsilon}}\right) \nabla_{h} \boldsymbol{U} - \boldsymbol{P}$$

$$(2.28)$$

The above numerical formulations are applied algorithmically as follows.

Algorithm 1: Correction of a single modulated image without aliasing

Input:  $\tilde{U}$ ,  $\epsilon$ ,  $\kappa$ ,  $\mu$ ,  $\nu$ , M,  $\gamma$ ,  $\delta$ Output: U, S

Initialization:  $U = (\tilde{U} - \min\{\tilde{U}_{j}\})/(\max\{\tilde{U}_{j}\} - \min\{\tilde{U}_{j}\}), S = 1$ **Outer Iteration**: start with  $t_0 = ||\boldsymbol{U}||, t = 2\delta \cdot t_0$ while  $(t > \delta \cdot t_0)$ save  $\hat{U} = U$ compute S(U) with Algorithm 2 below, in the first iteration with 2 grey levels, increasing later to Msolve (2.20) where in initial iterations U is replaced by S(U)determine the support of U from S(U) by setting  $0 = c_k = \min\{c_i\}$  and  $1 = c_{i \neq k}$ **normalize**  $S = S/\sigma$ ,  $\sigma = \max\{S_j : \text{over } j \text{ for which } S(U)_j \neq 0\}$ **Inner Iteration**: start with  $s_0 = ||\boldsymbol{U}||, s = 2\delta \cdot s_0, \boldsymbol{P} = 0$ in initial outer iterations with  $\mu = 0$ , later with the input value of  $\mu$ while  $(s > \delta \cdot s_0)$ solve (2.26) for  $\delta U$  and set  $U = U + \delta U$ set  $\delta P$  with (2.28), set  $P = P + \delta P$  and set  $P_{\mu}$  according to (2.27) update  $s = \|\delta U\|$ end update  $t = \|\boldsymbol{U} - \hat{\boldsymbol{U}}\|$ end Algorithm 2: Image Segmentation Input:  $U, M, \gamma, \delta$ Output:  $\Omega_i, c_i, i = 1, \ldots, M$ **Initialization:**  $\chi_1 = 1, \chi_j = 0$  for  $j \neq 1, c_j$  determined from (2.11) for j = 2, ..., Mfor  $i = 1, \ldots, (j - 1)$ **compute**  $T_{i,j}$  by (2.9) with  $|\Omega_j| = 0$ **transfer**  $\{x \in \Omega_i : (2.10) \text{ holds}\}$  from  $\Omega_i$  to  $\Omega_j$ 

**update**  $c_i$  and  $c_j$  with (2.11) end

end

**Iteration**: start with  $t_0 = 1 + \sum_{i,j}^M ||T_{i,j}||, t = 2\delta \cdot t_0$ while  $(t > \delta \cdot t_0)$  for i = 1, ..., Mcompute  $T_{i,j}$  for  $i \neq j = 1, ..., M$  with (2.9) if  $|\Omega_i| \neq 0$  and otherwise  $T_{i,j} = 0$ set  $T_{i,j}(\boldsymbol{x}) = \min\{T_{i,j}(\boldsymbol{x}), 0\}$ define:  $T_i(\boldsymbol{x}) = \min_{1 \leq j \leq M} T_{i,j}(\boldsymbol{x})$ transfer:  $\{\boldsymbol{x} \in \Omega_i : T_{i,j}(\boldsymbol{x}) = T_i(\boldsymbol{x}) \text{ and } T_i(\boldsymbol{x}) < \gamma \cdot \min_{\boldsymbol{y} \in \Omega_i} T_i(\boldsymbol{y})\}$ from  $\Omega_i$  to  $\Omega_j, i \neq j = 1, ..., M$ update  $c_j, j = 1, ..., M$ , with (2.11) end update  $t = \sum_{i,j}^M ||T_{i,j}||$ 

 $\mathbf{end}$ 

Computational results using these algorithms are shown in the next subsection.

#### 2.4 Computational Results

In this section given images are artificially modulated, and Algorithm 1 is used to correct the modulation; thereby, a reconstruction can be compared to a known desired result. Based upon the authors' considerations in [4] and [9] of appropriate sensitivity parameterizations founded on the Biot-Savart Law, functions of the following form are used for artificial modulations:

$$\sigma(\boldsymbol{x}) = \frac{1}{[1 + \alpha \|\boldsymbol{x} - \boldsymbol{x}_0\|^2]^{\frac{3}{2}}}$$
(2.29)

For the example of Fig. 8, a modulation (2.29) was used with parameters,

$$x_0 = (\frac{1}{2}, \frac{1}{2}) + r(\cos(\theta), \sin(\theta)), \quad r = \frac{3}{4}\frac{\sqrt{2}}{2}, \quad \theta = \frac{\pi}{2}, \quad \alpha = 5$$
 (2.30)

This modulation is shown in Fig. 8b on the support of the given exact image  $U^*$  appearing in Fig. 3a. The product of the two images, denoted by  $\tilde{U}$ , is shown in Fig. 8a. This  $\tilde{U}$  was used as input for Algorithm 1 along with  $\epsilon = 10^{-3}$ ,  $\kappa = 10^{-5}$ ,  $\mu = 10^{-4}$ ,  $\nu = 10^2$ , M = 5,  $\gamma = 0.5$  and  $\delta = 10^{-3}$ . Also, the dimension of the images is N = 128, but the domain is normalized to  $(1, N)^2$  so that h = 1. The reconstruction U is shown in Fig. 8c with 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,

$$d_{2}(U, U^{*}) = \frac{1}{N} \min_{s \in \mathbf{R}} \|sU - U^{*}\|_{\ell_{2}} = \frac{1}{N} \|s^{*}U - U^{*}\|_{\ell_{2}}, \quad s^{*} = U \cdot U^{*} / \|U\|_{\ell_{2}}^{2}$$

$$d_{\infty}(U, U^{*}) = \|\Delta U\|_{\ell_{\infty}}, \quad \Delta U = s^{*}U - U^{*}$$
(2.31)

Then the error image  $\Delta U$  is shown in Fig. 8d with the intensity scale [-1, 1]. The quantitative errors are  $d_2(U, U^*) = 0.053$  and  $d_{\infty}(U, U^*) = 0.40$ . The simulation was then repeated with 5% noise added as follows with n = 0.05:

$$\operatorname{FFT}(\tilde{U}) \to \hat{U}, \quad \hat{U} + \frac{n}{N} \|\hat{U}\|_2 X \to \hat{U}, \quad \operatorname{FFT}^{-1}(\hat{U}) \to \tilde{U}$$
 (2.32)

where  $\boldsymbol{X}$  is  $N \times N$  with elements normally distributed around mean 0 with variance 1. With noise added,  $\mu = 5 \cdot 10^{-4}$  was used on input. The reconstruction resulting from Algorithm 1 is shown in Fig. 8e and the reconstruction error image is shown in Fig. 8f. The quantitative errors in this case are  $d_2(\boldsymbol{U}, \boldsymbol{U}^*) = 0.052$  and  $d_{\infty}(\boldsymbol{U}, \boldsymbol{U}^*) = 0.42$ .

On the basis of the positive results shown in Fig. 8, simulations with phantoms are next considered to demonstrate how the method performs on images with a very different spectrum than that of Fig. 8. In the following example, the exact image  $U^*$  is that shown in Fig. 9b. The product of this image with the same modulation of Fig. 8 gives the data image  $\tilde{U}$  shown in Fig. 9a. This  $\tilde{U}$  was used as input for Algorithm 1, and other input parameters were the same as



**Figure 8:** The upper left image (a) is the product of the exact image of Fig. 3a and the modulation in the lower left image (b). The upper middle image (c) is the reconstruction obtained by Algorithm 1, and the corresponding reconstruction error image is shown in the lower middle image (d). The upper right image (e) is the reconstruction when 5% noise is added to (a), and the lower right image (f) is the corresponding reconstruction error image.



**Figure 9:** The upper left image (a) is the product of the exact image in the lower left (b) and the modulation of Fig. 8. The upper middle image (c) is the reconstruction obtained by Algorithm 1, and the corresponding reconstruction error image is shown in the lower middle image (d). The upper right image (e) is the reconstruction when 5% noise is added to (a), and the lower right image (f) is the corresponding reconstruction error image.

with Fig. 8. The reconstruction U is shown in Fig. 9c, and the reconstruction error image  $\Delta U$  is shown in Fig. 9d. The quantitative errors are  $d_2(U, U^*) = 0.0045$  and  $d_{\infty}(U, U^*) = 0.060$ . The simulation was then repeated with 5% noise added to the image in Fig. 9a. The reconstruction resulting from Algorithm 1 is shown in Fig. 9e and the reconstruction error image is shown in Fig. 9f. The quantitative errors in this case are  $d_2(U, U^*) = 0.015$  and  $d_{\infty}(U, U^*) = 0.24$ .

On the basis of these positive results, the approach of Section 3 could well be simplified by first correcting the modulations of data images which are not corrupted by aliasing. The potential for aliasing correction is considered in the next subsection.

#### 2.5 Aliasing Correction

Now consider the estimation of an image u such as in Fig. 3a from an aliased image  $\tilde{u}$  obtained from Cartesian subsampling as shown in Figs. 4 or 6. As in Subsection 2.1 u is considered to be a mapping  $u: \Omega \to [0, 1]$ .

The subsampling operator is a projection  $P = F^{-1}\chi F$ , where

$$Fu = \omega, \quad \omega(\mathbf{k}) = \int_{\Omega} u(\mathbf{x}) e^{-2\pi i \mathbf{k} \cdot \mathbf{x}} d\mathbf{x}$$
(2.33)

$$F^{-1}\omega = u, \quad u(\boldsymbol{x}) = \sum_{\boldsymbol{k}=-\infty}^{\infty} \omega(\boldsymbol{k}) e^{2\pi i \boldsymbol{k} \cdot \boldsymbol{x}}$$
(2.34)

and  $\chi = \chi(\mathbf{k})$  is a characteristic function in frequency space. The projection also has the direct image space representation P = Q + R where Q represents the folding effect as seen in Fig. 4 without reference points,

$$(Qu)(x_1, x_2) = \frac{1}{N_{\rm f}} \sum_{n=1}^{N_{\rm f}} u(x_1, \lceil x_2 + (n-1)/N_{\rm f} \rceil_1), \quad \lceil x \rceil_1 = x - [x]$$
(2.35)

and R represents the effect of the  $N_{\rm r}$  reference points as seen additionally in Fig. 6,

$$(Ru)(\boldsymbol{x}) = 2\sum_{l=1}^{N_{r}/2} \int_{\Omega} u(\boldsymbol{y}) \cos(2\pi \boldsymbol{k}_{l} \cdot (\boldsymbol{x} - \boldsymbol{y})) d\boldsymbol{y}$$
(2.36)

In (2.35)  $N_{\rm f}$  is the folding factor, i.e.,  $N_{\rm f} = 4$  in Fig. 4, and the simplification of the projection in (2.35) is obtained by noting [9]:

$$\frac{1}{N_{\rm f}} \sum_{j=1}^{N_{\rm f}} u(x_1, \lceil x_2 + (j-1)/N_{\rm f} \rceil_1) = \sum_{k=-\infty}^{\infty} \omega(k) e^{2\pi i k_1 x_1} \left( \sum_{j=1}^{N_{\rm f}} e^{2\pi i k_2 (x_2 + (j-1)/N_{\rm f})} \right) \\
= \sum_{k=-\infty}^{\infty} \omega(k_1, N_{\rm f} k_2) e^{2\pi i (k_1 x_1 + N_{\rm f} k_2 x_2)}$$
(2.37)

where the right side in (2.37) includes only every  $N_{\rm f}$  line of frequencies as illustrated in Fig. 4, and the left side in (2.37) agrees with the right side of (2.35). The representation of R in (2.36) is obtained under the assumption that the reference points are symmetrically situated, as in Fig. 6, so that  $\mathbf{k}_l = -\mathbf{k}_{N_{\rm r}/2+l}, \ l = 1, \ldots, N_{\rm r}/2$ . Thus, [9]

$$(Ru)(\boldsymbol{x}) = \sum_{l=1}^{N_{\mathrm{r}}} \int_{\Omega} u(\boldsymbol{y}) e^{2\pi i \boldsymbol{k}_l \cdot (\boldsymbol{x} - \boldsymbol{y})} d\boldsymbol{y} = \sum_{l=1}^{N_{\mathrm{r}}/2} \int_{\Omega} u(\boldsymbol{y}) \left[ e^{2\pi i \boldsymbol{k}_l \cdot (\boldsymbol{x} - \boldsymbol{y})} + e^{-2\pi i \boldsymbol{k}_l \cdot (\boldsymbol{x} - \boldsymbol{y})} \right] d\boldsymbol{y}$$
(2.38)

which agrees with (2.36).

Since the data are measured only on the support of  $\chi$ ,  $\tilde{u}$  is in the range of P. Since the data may be noisy on the support of  $\chi$ ,  $\tilde{u} \approx Pu$  holds, and it is considered to estimate u from  $\tilde{u}$  by minimizing:

$$J(u) = \frac{1}{2} \int_{\Omega} |Pu - \tilde{u}|^2 d\boldsymbol{x} + \frac{\kappa}{2} \int_{\Omega} u^2 d\boldsymbol{x} + \mu \int_{\Omega} \phi_{\epsilon}(|\nabla u|^2) d\boldsymbol{x}$$
(2.39)

As with (2.40) seen below for the more general case with more than one coil, the primal-dual optimality system for (2.39) is:

$$\begin{cases} \kappa u + Pu - \Delta_{\mathcal{N}} w = \tilde{u} \\ -\mu \nabla u + \lfloor \nabla u \rfloor_{\epsilon} \nabla w = 0 \end{cases} \quad u \in L^{2}(\Omega), \quad \Delta_{\mathcal{N}} w \in S_{\mu}$$
(2.40)

The  $L^2$  regularization in (2.39) is included so that  $\kappa I + P$  is invertible in (3.7) below no matter which subsampling P is used.

The images reconstructed from the aliased images shown in Figs. 4b and 6d by solving (2.40) are shown respectively in Figs. 10a and 10b. Unfortunately, the aliasing artifacts resulting from



Figure 10: The images shown (a) on the left and (b) in the middle are obtained by solving (2.40) with  $\tilde{u}$  as shown in Figs. 4b and 6d, respectively. For comparison, the exact image in Fig. 3a has also been aliased by subsampling along a limited set of radial lines, and the resulting image is shown (c) on the right.

Cartesian subsampling contain so much structure that a TV like penalty is not suitable for artifact removal. As shown for instance in [2], reconstruction methods for undersampled data require at least that the aliasing artifacts be relatively incoherent or noise like. For example, the image shown in Fig. 10b is obtained by subsampling the exact image in Fig. 3a along a limited set of radial lines. Note that the aliasing artifacts from radial subsampling have much less structure than those from Cartesian subsampling. Indeed, the authors have successfully carried out TV based image reconstruction from radially subsampled data as reported separately in [18].

On the basis of the unsuccessful result in Fig. 10 of attempting a pure aliasing correction when the data are Cartesian subsampled, it is not surprising that the method is even less successful to correct data which are aliased *and* modulated as in Fig. 5. Nevertheless, when several such modulated and aliased images are available, they can be used together to successfully reconstruct an uncorrupted image as seen in the next section.

# 3 Image Reconstruction from Parallel Coil Measurements

In this section it is considered to reconstruct an image u such as in Fig. 3a from several aliased and modulated images  $\{\tilde{u}_i\}$  such as in Fig. 5 measured simultaneously from coils using the same Cartesian subsampling represented by the projection P described in Subsection 2.5. The number of coils is denoted by  $N_c$  and satisfies  $N_c \geq N_f$ . While the subsampling projection P

is assumed to be known and identical for each coil, the sensitivity  $\sigma_i$  of the *i*th coil is not known and must be estimated along with u. The desired unknowns  $\{\sigma_i\}$  and u satisfy  $P\sigma_i u \approx u_i$ ,  $i = 1, \ldots, N_c$  and are determined by minimizing:

$$J(u, \{\sigma_i\}) = \frac{1}{2} \sum_{i=1}^{N_c} \int_{\Omega} |P\sigma_i u - \tilde{u}_i|^2 dx + \frac{\nu}{2} \sum_{i=1}^{N_c} \int_{\Omega} |\nabla^2 \sigma_i|^2 dx + \frac{\kappa}{2} \int_{\Omega} u^2 dx + \mu \int_{\Omega} \phi_\epsilon(|\nabla u|^2) dx \quad (3.1)$$

where the penalty terms of (3.1) are explained in Subsections 2.1 and 2.5. The landscape of the cost functional (3.1) is also conceptualized as in Fig. 7, and a nonlinear Gauss-Seidel scheme is used to solve the optimality system by applying a Newton scheme to solve for one variable after the other. As with modulation correction, additional information is used to orient iterations. Specifically, all sensitivities are scaled by the same constant in each iteration so that the largest value among them achieves a maximum value of 1 on the support of the image. Also, the sensitivities are regularized in early iterations by using a segmentation of the image in the computation of sensitivities instead of the image itself.

#### 3.1 Optimality Conditions

The optimality condition for (3.1) with respect to  $\sigma_i$  for fixed u is:

$$B(u)\sigma_i = u\tilde{u}_i, \qquad \sigma_i \in H^2(\Omega) \tag{3.2}$$

which is given in weak form as:

$$\int_{\Omega} \left[ \nu \nabla^2 \sigma_i \cdot \nabla^2 \bar{\sigma} + \sigma_i u P u \bar{\sigma} \right] d\boldsymbol{x} = \int_{\Omega} \bar{\sigma} u \tilde{u}_i d\boldsymbol{x}, \quad \forall \bar{\sigma} \in H^2(\Omega).$$
(3.3)

That (3.3) is solvable is established as follows.

**Theorem 1** Suppose  $\tilde{u}_i \in L^2(\Omega)$  and  $u \in L^{\infty}(\Omega)$ . Suppose further that if a linear function  $\sigma$  satisfies  $\int_{\Omega} u\sigma d\mathbf{x} = 0$  then  $\sigma = 0$ . Then there exists a unique solution  $\sigma_i \in H^2(\Omega)$  to (3.3).

Proof: Define the bilinear form on the left in (3.3) as  $F(\sigma, \bar{\sigma})$ . If  $\sigma \in H^2(\Omega)$  satisfies  $F(\sigma, \sigma) = 0$ , then  $\int_{\Omega} |\nabla^2 \sigma|^2 d\boldsymbol{x} = 0$  implies that  $\sigma$  is a linear function. From Parsevals's identity,  $0 = \int_{\Omega} |Pu\sigma|^2 d\boldsymbol{x} \geq \int_{\Omega} |Qu\sigma|^2 d\boldsymbol{x}$  with Q from (2.35). Then integrating  $Qu\sigma = 0$  pointwise over  $(0, 1) \times (0, 1/N_{\rm f})$  gives:

$$0 = \int_0^1 \left[ \int_0^{1/N_{\rm f}} \sum_{n=1}^{N_{\rm f}} u(x_1, x_2 + (n-1)/N_{\rm f}) \sigma(x_1, x_2 + (n-1)/N_{\rm f}) dx_2 \right] dx_1 = \int_\Omega u \sigma d\boldsymbol{x} \quad (3.4)$$

Therefore,  $\sigma = 0$ . Thus, as in the proof of Poincarè's Inequality [5],  $[F(\sigma, \sigma)]^{\frac{1}{2}}$  is equivalent to the norm  $\|\sigma\|_{H^2(\Omega)}$ , and hence the bilinear form  $F(\sigma, \bar{\sigma})$  is coercive and bounded on  $H^2(\Omega) \times H^2(\Omega)$ . The linear form  $G(\bar{\sigma})$  on the right in (3.3) is bounded in  $L^2(\Omega)$  and hence in  $H^2(\Omega)$ . Thus, the claim follows with the Lax Milgram Lemma [5].

As with (2.5), (3.2) is solved in initial iterations of the nonlinear Gauss-Seidel method by replacing the image u with the segmentation S(u) of (2.7). It was also considered to compute a segmentation according to the minimization of  $J(\Omega_1, \ldots, \Omega_M) = \sum_{i=1}^{N_c} \int_{\Omega} [P\sigma_i S(u) - \tilde{u}_i]^2 dx$ , whose topological derivatives are given by:

$$T_{i,j}(\boldsymbol{x}) = \frac{1}{N_{\rm f}} \sum_{k=1}^{N_{\rm c}} \begin{cases} [(c_i - c_j)\sigma_k(\boldsymbol{x}) - N_{\rm f}(P\sigma_k S(u) - \tilde{u}_k)(\boldsymbol{x})]^2 - [N_{\rm f}(P\sigma_k S(u) - \tilde{u}_k)(\boldsymbol{x})]^2, & |\Omega_j| \neq 0\\ -[N_{\rm f}(P\sigma_k S(u) - \tilde{u}_k)(\boldsymbol{x})]^2, & |\Omega_j| = 0 \end{cases}$$

$$(3.5)$$

for  $\boldsymbol{x} \in \Omega_i$ , and  $T_{i,j}(\boldsymbol{x}) = 0$  for  $\boldsymbol{x} \notin \Omega_i$ . However, this approach was not found to perform better than that given following (2.7). See Algorithm 2 above for the details of determining S(u).

To establish an optimality condition for (3.1) with respect to u for fixed sensitivities  $\{\sigma_i\}$ , define the functionals  $\mathcal{F}: L^2(\Omega) \to \mathbb{R} \cup \{\infty\}$ ,

$$\mathcal{F}(u) = \frac{1}{2} \sum_{i=1}^{N_c} \int_{\Omega} |P\sigma_i u - \tilde{u}_i|^2 d\boldsymbol{x} + \frac{\kappa}{2} \int_{\Omega} u^2 d\boldsymbol{x}$$
(3.6)

and  $\mathcal{G}: L^2(\Omega) \to \mathbb{R} \cup \{\infty\}$  as in (2.13) so that the dependence on u in J is given by  $\mathcal{F}(u) + \mathcal{G}(u)$ . Note that these operators satisfy the conditions of the Fenchel Duality Theorem, and the desired optimality condition is thus given by (2.14) [20], where the convex conjugates  $\mathcal{F}^*$  and  $\mathcal{G}^*$  in (2.14) are given as follows; see Appendix A for details. First,  $\mathcal{G}^*$  is given in (2.16). Then,  $\mathcal{F}^*: L^2(\Omega) \to \mathbb{R} \cup \infty$  is given by:

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

That the operator  $\kappa I + \sum_{i=1}^{N_c} \sigma_i P \sigma_i$  is invertible is established as follows.

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

*Proof*: Define the bilinear form F on  $L^2(\Omega) \times L^2(\Omega)$  and the linear form G on  $L^2(\Omega)$ :

$$F(u,\bar{u}) = \int_{\Omega} \left[ \kappa u\bar{u} + \sum_{i=1}^{N_c} u\sigma_i P\sigma_i \bar{u} \right] d\boldsymbol{x}, \quad G(\bar{u}) = \int_{\Omega} v\bar{u}d\boldsymbol{x}, \quad u,\bar{u},v \in L^2(\Omega)$$
(3.8)

Since  $\sum_{i=1}^{N_c} \sigma_i P \sigma_i$  is non-negative, F is coercive. Since  $\{\sigma_i\} \subset L^{\infty}(\Omega)$ , F is also bounded. Given any  $v \in L^2(\Omega)$ ,  $G(\bar{u})$  is bounded. Thus, the claim follows with the Lax Milgram Lemma [5].

As seen in Appendix A the functionals (3.6), (2.13), (3.7) and (2.16) lead to the following formulation of the optimality system (2.14):

$$\begin{cases} \left[\kappa I + \sum_{i=1}^{N_{c}} \sigma_{i} P \sigma_{i}\right] u - \Delta_{N} w = \sum_{i=1}^{N_{c}} \sigma_{i} \tilde{u} \\ -\mu \nabla u + \lfloor \nabla u \rfloor_{\epsilon} \nabla w = 0 \end{cases} \qquad u \in L^{2}(\Omega), \quad \Delta_{N} w \in S_{\mu}$$
(3.9)

#### 3.2 Numerical Methods

The discretization of the optimality conditions in Subsection 3.1 involves the same constructions detailed in Subsection 2.3. As with (2.20), the numerical solution to (3.2) is given as  $S_i = \{S_{i,j}\}, S_{i,j} \approx \sigma_i(x_j)$ , which solves:

$$[\nu B_h + D(\boldsymbol{U})P_h D(\boldsymbol{U})]\boldsymbol{S}_i = D(\boldsymbol{U})\tilde{\boldsymbol{U}}_i$$
(3.10)

where, in parallel to (2.35) and (2.36),  $P_h$  is defined as follows:

$$P_h = Q_h + V_h V_h^{\rm T} \tag{3.11}$$

Here, the discrete counterparts to (2.35) and (2.36) are

$$Q_h = \frac{1}{N_f} K_h^T K_h, \quad K_h = (\underbrace{I_k, I_k, \dots, I_k}_{N_f}), \quad k = N^2 / N_f$$
(3.12)

and [9]

$$V_h = [\boldsymbol{V}_1, \dots, \boldsymbol{V}_{N_{\rm r}}], \qquad \boldsymbol{V}_i \cdot \boldsymbol{V}_j = \delta_{ij}$$
(3.13)

$$\boldsymbol{V}_{l} = \sqrt{\frac{N}{2}} \frac{1}{N^{d}} \begin{cases} \left\{ \cos(2\pi \boldsymbol{k}_{l} \cdot \boldsymbol{x}_{\boldsymbol{j}-\frac{1}{2}}) : \mathbf{1} \leq \boldsymbol{j} \leq N \cdot \mathbf{1} \right\}, \quad l = 1, \dots, N_{\mathrm{r}}/2 \\ \left\{ \sin(2\pi \boldsymbol{k}_{l} \cdot \boldsymbol{x}_{\boldsymbol{j}-\frac{1}{2}}) : \mathbf{1} \leq \boldsymbol{j} \leq N \cdot \mathbf{1} \right\}, \quad l = N_{\mathrm{r}}/2 + 1, \dots, N_{\mathrm{r}} \end{cases}$$
(3.14)

The matrices on the left side in (3.10) 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  $Q_h$ :

$$C = [\nu B_h + D(\boldsymbol{U})Q_h D(\boldsymbol{U})]$$
(3.15)

Thus, the matrix on the left in (3.10) is partitioned as:

$$C + D(\boldsymbol{U})V_h V_h^{\mathrm{T}} D(\boldsymbol{U}) = C + X X^{\mathrm{T}}, \qquad X = D(\boldsymbol{U})V_h$$
(3.16)

Then  $C + XX^{T}$  can be inverted using the Sherman-Morrison-Woodbury formula [6]:

$$(C + XX^{\mathrm{T}})^{-1} = C^{-1} + C^{-1}X(I + X^{\mathrm{T}}CX)^{-1}X^{\mathrm{T}}C^{-1}$$
(3.17)

This formula is implemented in MATLAB in the present work using backslash. In order that (3.17) be computationally useful, it is crucial that the dimension of the full matrix  $(I + X^{T}CX)$  be small, i.e., that only a few reference points be used as in Fig. 6c. If whole center-lines are used, as in Fig. 6a, the dimension of  $(I + X^{T}CX)$  becomes too large for direct inversion methods to be used. Thus, considering the discussion in Section 1 of Fig. 6, it is computationally advantageous to use only the few low frequencies in Fig. 6c, even if the whole lines of Fig. 6a are measured. If conditions for the convenient use of (3.17) are not met, (3.10) must be solved iteratively.

The segmentation computations in (2.9) - (2.11) are carried out by evaluating each function at the cell centroids  $x = x_j$ . See Algorithm 2 above for the details of determining S(U).

The optimality system (2.40) is solved using a generalized Newton method. The system,

$$\begin{bmatrix} \kappa I + \sum_{i=1}^{N_{c}} \sigma_{i} P \sigma_{i} & -\Delta_{N} \\ \left[ -\mu I + \frac{(|\nabla u| > \epsilon)}{[\nabla u]_{\epsilon}} \nabla u \nabla w^{T} \right] \nabla \quad [\nabla u]_{\epsilon} \nabla \end{bmatrix} \begin{bmatrix} \delta u \\ \delta w \end{bmatrix} =$$

$$-\begin{bmatrix} \left[ \kappa I + \sum_{i=1}^{N_{c}} \sigma_{i} P \sigma_{i} \right] u - \Delta_{N} w - \sum_{i=1}^{N_{c}} \sigma_{i} \tilde{u}_{i} \\ -\mu \nabla u + [\nabla u]_{\epsilon} \nabla w \end{bmatrix}$$

$$(3.18)$$

is simplified by first eliminating the second equation to obtain:

$$\left\{\kappa I + \sum_{i=1}^{N_{c}} \sigma_{i} P \sigma_{i} - \nabla \cdot \left[\frac{1}{\lfloor \nabla u \rfloor_{\epsilon}} \left(\mu I - \frac{(|\nabla u| > \epsilon)}{2\lfloor \nabla u \rfloor_{\epsilon}} \left[\nabla u \, \boldsymbol{p}^{\mathrm{T}} + \boldsymbol{p} \, \nabla u^{\mathrm{T}}\right]\right)\right] \nabla\right\} \delta u = -\left[\kappa I + \sum_{i=1}^{N_{c}} \sigma_{i} P \sigma_{i}\right] u + \sum_{i=1}^{N_{c}} \sigma_{i} \tilde{u}_{i} + \mu \nabla \cdot \left(\frac{\nabla u}{\lfloor \nabla u \rfloor_{\epsilon}}\right)$$

$$(3.19)$$

and the eliminated equation becomes:

$$\boldsymbol{\delta p} = \frac{1}{\left\lfloor \nabla u \right\rfloor_{\epsilon}} \left( \mu I - \frac{\left( \left| \nabla u \right| > \epsilon \right)}{2 \left\lfloor \nabla u \right\rfloor_{\epsilon}} \left[ \nabla u \, \boldsymbol{p}^{\mathrm{T}} + \boldsymbol{p} \, \nabla u^{\mathrm{T}} \right] \right) \nabla \delta u + \mu \frac{\nabla u}{\left\lfloor \nabla u \right\rfloor_{\epsilon}} - \boldsymbol{p}$$
(3.20)

where  $\boldsymbol{p} = \nabla w \in \boldsymbol{H}_0(\text{div})$  as with (2.23). Also the term  $\nabla u \boldsymbol{p}^{\text{T}}$  has been symmetrized with  $[\nabla u \boldsymbol{p}^{\text{T}} + \boldsymbol{p} \nabla u^{\text{T}}]/2$ . As with (2.22) and (2.23), (3.19) and (3.20) are discretized as follows:

$$\begin{cases} \kappa I + \sum_{i=1}^{N_{c}} D(\boldsymbol{S}_{i}) P_{h} D(\boldsymbol{S}_{i}) + \\ \nabla_{h}^{T} \left[ D\left(\frac{1}{\lfloor \nabla_{h} \boldsymbol{U} \rfloor_{\epsilon}}\right) \left( \mu I - D\left(\frac{(|\nabla \boldsymbol{U}| > \epsilon)}{2\lfloor \nabla \boldsymbol{U} \rfloor_{\epsilon}}\right) \left[ \nabla_{h} \boldsymbol{U} \boldsymbol{P}_{\mu}^{T} + \boldsymbol{P}_{\mu} \nabla_{h} \boldsymbol{U}^{T} \right] \right) \right] \nabla_{h} \right\} \boldsymbol{\delta} \boldsymbol{U} = \\ - \left[ \kappa I + \sum_{i=1}^{N_{c}} D(\boldsymbol{S}_{i}) P_{h} D(\boldsymbol{S}_{i}) \right] \boldsymbol{U} + \sum_{i=1}^{N_{c}} D(\boldsymbol{S}_{i}) \tilde{\boldsymbol{U}}_{i} - \mu \nabla_{h}^{T} \left[ D\left(\frac{1}{\lfloor \nabla \boldsymbol{U} \rfloor_{\epsilon}}\right) \nabla_{h} \boldsymbol{U} \right] \end{cases}$$
(3.21)

and:

$$\delta \boldsymbol{P} = D\left(\frac{1}{\lfloor \nabla \boldsymbol{U} \rfloor_{\epsilon}}\right) \left(\mu I - D\left(\frac{(|\nabla \boldsymbol{U}| > \epsilon)}{2\lfloor \nabla \boldsymbol{U} \rfloor_{\epsilon}}\right) \left[\nabla_{h} \boldsymbol{U} \boldsymbol{P}_{\mu}^{\mathrm{T}} + \boldsymbol{P}_{\mu} \nabla_{h} \boldsymbol{U}^{\mathrm{T}}\right]\right) \nabla_{h} \delta \boldsymbol{U} + \mu D\left(\frac{1}{\lfloor \nabla \boldsymbol{U} \rfloor_{\epsilon}}\right) \nabla_{h} \boldsymbol{U} - \boldsymbol{P}$$
(3.22)

As shown in [10], using the truncation  $P_{\mu}$  of (2.27) in (3.21) guarantees that  $\delta U$  provides a descent direction for the cost functional with fixed modulations. The matrices on the left in (3.21) 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  $Q_h$ :

$$A = \left\{ \kappa I + \sum_{i=1}^{N_{c}} D(\boldsymbol{S}_{i}) Q_{h} D(\boldsymbol{S}_{i}) + \nabla_{h}^{\mathrm{T}} \left[ D\left(\frac{1}{\left\lfloor \nabla_{h} \boldsymbol{U} \right\rfloor_{\epsilon}}\right) \left( \mu I - D\left(\frac{\left(\left|\nabla \boldsymbol{U}\right| > \epsilon\right)}{2\left\lfloor \nabla \boldsymbol{U} \right\rfloor_{\epsilon}}\right) \left[\nabla_{h} \boldsymbol{U} \boldsymbol{P}_{\mu}^{\mathrm{T}} + \boldsymbol{P}_{\mu} \nabla_{h} \boldsymbol{U}^{\mathrm{T}} \right] \right) \right] \nabla_{h} \right\}$$
(3.23)

so the matrix on the left in (3.21) can be partitioned as:

$$A + \sum_{i=1}^{N_{\rm c}} D(\mathbf{S}_i) V_h V_h^{\rm T} D(\mathbf{S}_i) = A + Y Y^{\rm T}, \qquad Y = (D(\mathbf{S}_1) V_h, \dots, D(\mathbf{S}_{N_{\rm c}}) V_h)$$
(3.24)

Thus,  $A + YY^{T}$  can be inverted using the Sherman-Morrison-Woodbury formula [6]:

$$(A + YY^{\mathrm{T}})^{-1} = A^{-1} + A^{-1}Y(I + Y^{\mathrm{T}}AY)^{-1}Y^{\mathrm{T}}A^{-1}$$
(3.25)

This formula is implemented in MATLAB in the present work using backslash. As in the discussion of (3.17), for (3.25) to be computationally useful, it is crucial that the dimension of the full matrix  $(I + Y^{T}AY)$  be small, i.e., that only a few reference points be used as in Fig. 6c. Otherwise, (3.21) must be solved iteratively.

The above numerical formulations are applied algorithmically as follows.

Algorithm 3: Parallel image reconstruction

Input:  $\{\tilde{U}_i\}, \epsilon, \gamma, M, \delta, \kappa, \mu, \nu$ Output:  $U, \{S_i\}$ 

**Initialization:** set  $\tilde{\boldsymbol{U}} = \sum_{i=1}^{N_c} \tilde{\boldsymbol{U}}_i / N_c$  and  $\boldsymbol{U} = (\tilde{\boldsymbol{U}} - \min\{\tilde{U}_{\boldsymbol{j}}\}) / (\max\{\tilde{U}_{\boldsymbol{j}}\} - \min\{\tilde{U}_{\boldsymbol{j}}\})$  as well as  $\boldsymbol{S}_i = 1, i = 1, \dots, N_c$ 

**Outer Iteration**: start with  $t_0 = ||\boldsymbol{U}||, t = 2\delta \cdot t_0$ while  $(t > \delta \cdot t_0)$ save  $\hat{U} = U$ compute S(U) with Algorithm 2 solve (3.10) using (3.17) where in initial iterations U is replaced by S(U)**determine** the support of U from S(U) by setting  $0 = c_k = \min\{c_i\}$  and  $1 = c_{i \neq k}$ normalize  $S_i = S_i / \sigma, i = 1, ..., N_c, \sigma = \max\{(S_i)_j : S(U)_j \neq 0, i = 1, ..., N_c\}$ Inner Iteration: start with  $s_0 = ||\boldsymbol{U}||, s = 2\delta \cdot s_0, \boldsymbol{P} = 0$ in initial outer iterations with  $\mu = 0$ , later with the input value of  $\mu$ while  $(s > \delta \cdot s_0)$ solve (3.21) for  $\delta U$  using (3.25) and set  $U = U + \delta U$ set  $\delta P$  with (3.22) set  $P = P + \delta P$  and set  $P_{\mu}$  according to (2.27) update  $s = \|\delta U\|$ end update  $t = \|\boldsymbol{U} - \hat{\boldsymbol{U}}\|$ end

Computational results using this algorithm are shown in the next subsection.

### 3.3 Computational Results

Here as in Subsection 2.4, given images are artificially modulated and aliased, for varying numbers of virtual coils, and Algorithm 3 is used to reconstruct the original image; thereby a reconstruction can be compared to a known desired result. The parameterization (2.29) is used for  $N_c$  coils according to:

$$\sigma_{i}(\boldsymbol{x}) = \frac{1}{[1+\alpha\|\boldsymbol{x}-\boldsymbol{x}_{i}\|^{2}]^{\frac{3}{2}}}, \quad i = 1, \dots, N_{c}$$

$$\boldsymbol{x}_{i} = (\frac{1}{2}, \frac{1}{2}) + r(\cos(\theta_{i}), \sin(\theta_{i})), \quad \theta_{i} = \theta_{0} + \frac{2\pi(i-1)}{N_{c}}$$
(3.26)

Note that the distribution of the virtual coil centers  $\{x_i\}$ , equally spaced around a circle of radius r, corresponds to the geometry seen in Fig. 1b. The criterion for the selection of  $\theta_0$  is that the conditioning of the matrix  $\sum_{i=1}^{N_c} D(\mathbf{S}_i) P_h D(\mathbf{S}_i)$  be as favorable as possible. Direct calculations with a variety of values for  $N_c$ ,  $N_f$  and  $N_r$  leads to the choice  $\theta_0 = \frac{\pi}{2} + \frac{\pi}{2N_c}$  when aliasing is vertical as seen in Figs. 4 - 6 [9].

The purpose of Fig. 11 is to demonstrate with a simple phantom example the difficulty of parallel image reconstruction when the number of coils is equal to the folding factor  $N_{\rm c} = N_{\rm f}$ and when no reference points are used  $N_r = 0$ . Specifically, the parameters used for the three cases considered in Fig. 11 were, respectively,  $N_{\rm c} = N_{\rm f} = 2$ ,  $N_{\rm c} = N_{\rm f} = 4$  and  $N_{\rm c} = 8 > 4 = N_{\rm f}$ , where  $N_{\rm r} = 0$  holds in every case. Also, for each example, modulations (3.26) were used with parameters r and  $\alpha$  from (2.30). After multiplying these modulations by the exact phantom image in Fig. 9b and applying a subsampling projection such as seen in Fig. 4, i.e., without reference points, data images  $\{U_i\}$  were obtained, and no noise was added to these artificial data. Such data images were used as input for Algorithm 3 along with  $\epsilon = 10^{-3}$ ,  $\kappa = 10^{-4}$ ,  $\nu = 10^2$ , M = 5,  $\gamma = 0.5$  and  $\delta = 10^{-3}$ . Also, the dimension of the images is N = 128, but the domain is normalized to  $(1, N)^2$  so that h = 1. For simplicity, the input  $\mu = 0$  was used for these noise-free examples, and the results were not improved with  $\mu > 0$ ; nevertheless, the role of the segmentation was found to be crucial for convergence in these cases. The reconstructed images U are shown in the top row of Fig. 11 with the intensity scale [0, 1], while the reconstruction error images  $\Delta U$  of (2.31) are shown in the bottom row of Fig. 11 with the intensity scale [-1, 1]. The case of  $N_{\rm c} = N_{\rm f} = 2$  is shown in the first column of Fig. 11. Although this result is positive,



Figure 11: For the three reconstructions demonstrated in the three columns, the exact image is shown in Fig. 9b, and virtual data for Algorithm 3 were simulated using the modulations in (3.26). Reconstructions are shown in the top row and the corresponding reconstruction error images are shown in the bottom row. In each case,  $N_{\rm r} = 0$ . The top left image (a) and bottom left image (b) correspond to the case  $N_{\rm c} = N_{\rm f} = 2$ . The top middle image (c) and bottom middle image (d) correspond to the case  $N_{\rm c} = N_{\rm f} = 4$ . The top right image (e) and bottom right image (f) correspond to the case  $N_{\rm c} = 8 > 4 = N_{\rm f}$ .

it is desirable for practical applications that the subsampling factor be higher. So the case of  $N_{\rm c} = N_{\rm f} = 4$  is shown in the second column of Fig. 11. Apparently, the reconstruction of the image of Fig. 3a with the data of Fig. 5 is not practical. This negative result already suggests the importance of using enough coils so that  $N_{\rm c} > N_{\rm f}$  holds. Thus, the case of  $N_{\rm c} = 8 > 4 = N_{\rm f}$  is shown in the third column of Fig. 11. While this result is more positive than that in the second column, it is clearly not satisfactory. On the basis of these examples it is concluded that it is advantageous for  $N_{\rm c} > N_{\rm f}$  and  $N_{\rm r} > 0$  to hold.

For all the examples shown in Fig. 12, Algorithm 3 was used to reconstruct the exact image  $U^*$  shown in Fig. 3a using simulated data with  $N_{\rm c} = 8$ ,  $N_{\rm f} = 4$  and  $N_{\rm r} = 54$ , where these reference points are contained in a  $9 \times 9$  square centered at the origin in frequency space. The modulations are given by (3.26) with parameters r and  $\alpha$  from (2.30). The data images  $\{U_i\}$ were obtained by multiplying the exact image by the indicated modulations and applying the subsampling projection as illustrated in Fig. 6c. For the first column of Fig. 12, no noise was added to these data, but for the other two columns, 5% noise was added with n = 0.05 in (2.32). For the third column of Fig. 12, only  $L^2$  regularization was used without assistance from TV regularization or from segmentation. On the other hand, along with segmentation,  $\mu = 10^{-4}$  and  $\mu = 3 \cdot 10^{-4}$  were used for the first and second columns of Fig. 12 respectively. All other parameters were the same as those used for the examples of Fig. 11. The reconstructed images U and the reconstruction error images  $\Delta U$  of (2.31) are shown respectively in the top and bottom rows of Fig. 12 with the same format as in Fig. 11. For the first, second and third columns of Fig. 12, the quantitative errors are  $d_2(\boldsymbol{U}, \boldsymbol{U}^*) = 0.033, 0.041$  and 0.055 and  $d_{\infty}(\boldsymbol{U},\boldsymbol{U}^*) = 0.18, 0.24$  and 0.37 respectively. Note that errors are larger and the effect of noise is stronger in the middle horizontal strip of the images, and this effect diminishes as  $\alpha$  in (3.26) is reduced, i.e., as the slope of the sensitivities becomes smaller.

To demonstrate the effect of using even less information, Fig. 13 shows the results of using



Figure 12: In all three examples shown here, Algorithm 3 was used to reconstruct the exact image shown in Fig. 3a using simulated data with  $N_c = 8$ ,  $N_f = 4$  and  $N_r = 54$ . The upper left image (a) is the reconstruction obtained with noise-free data, and the corresponding reconstruction error image is shown in the lower left image (b). The upper middle image (c) is the reconstruction obtained with 5% noise in the data, and the corresponding reconstruction error image is shown in the lower middle image (d). The upper right image (e) is the reconstruction obtained with neither TV regularization nor segmentation and with 5% noise in the data, and the corresponding reconstruction error image is shown in the lower right image (f).

 $N_{\rm r} = 20$ . Specifically, these reference points are contained in a 5×5 square centered at the origin in frequency space. All other parameters are the same as those used for the examples of Fig. 12, and the formats of the two figures are also the same. For the first, second and third columns of Fig. 13, the quantitative errors are  $d_2(\boldsymbol{U}, \boldsymbol{U}^*) = 0.041$ , 0.047 and 0.057 and  $d_{\infty}(\boldsymbol{U}, \boldsymbol{U}^*) = 0.26$ , 0.28 and 0.34 respectively. Note that the errors in Figs. 12 and 13 are comparable, even though the number of reference points is reduced in Fig. 13.

Nevertheless, as  $N_r$  or  $N_c$  are decreased further, as seen in Fig. 14, significant errors emerge, particularly in the purely  $L^2$ -regularized reconstructions appearing in the top row, but also in the TV-regularized reconstructions appearing in the bottom row. Specifically, the results shown in the first, second and third columns of Fig. 14 were obtained using  $N_r = 54$ , 20 and 6, respectively, where these reference points are contained in  $9 \times 9$ ,  $5 \times 5$  and  $3 \times 3$  squares centered at the origin in frequency space. In all cases,  $N_c = 6$  and  $N_f = 4$  were used, and all other parameters are the same as those used for the examples of Figs. 12 and 13. The quantitative errors for the  $L^2$  regularized reconstructions in Figs. 14a, b and c are  $d_2(\boldsymbol{U}, \boldsymbol{U}^*) = 0.074$ , 0.078 and 0.087 and  $d_{\infty}(\boldsymbol{U}, \boldsymbol{U}^*) = 0.56$ , 0.51 and 0.52 respectively. The quantitative errors for the TV regularized reconstructions in Figs. 14d, e and f are  $d_2(\boldsymbol{U}, \boldsymbol{U}^*) = 0.052$ , 0.056 and 0.076 and  $d_{\infty}(\boldsymbol{U}, \boldsymbol{U}^*) = 0.35$ , 0.32 and 0.56 respectively. Note that the TV-regularized reconstructions are superior, but they also begin to manifest artifacts as one sees in an exaggerated fashion in Fig. 11.

On the other hand, when the Cartesian subsampling is performed in both the vertical and horizontal directions, Fig. 15 shows that successful reconstructions can be performed with more noise and with considerably less information than that necessary for Figs. 12 - 14. Specifically, the results shown in Fig. 15 were obtained using  $N_c = 4$  coils and  $N_r = 8$  reference points contained in a  $3 \times 3$  square centered at the origin in frequency space. Also the folding factor



Figure 13: In all three examples shown here, Algorithm 3 was used to reconstruct the exact image shown in Fig. 3a using simulated data with  $N_c = 8$ ,  $N_f = 4$  and  $N_r = 20$ . The upper left image (a) is the reconstruction obtained with noise-free data, and the corresponding reconstruction error image is shown in the lower left image (b). The upper middle image (c) is the reconstruction obtained with 5% noise in the data, and the corresponding reconstruction error image is shown in the lower middle image (d). The upper right image (e) is the reconstruction obtained with neither TV regularization nor segmentation and with 5% noise in the data, and the corresponding reconstruction error image is shown in the lower right image (f).

was  $N_{\rm f} = 2$  in both the horizontal and vertical directions. In this case, the operator Q is given by the following instead of (2.35),

$$(Qu)(x_1, x_2) = \frac{1}{N_{\rm f}} \sum_{n=1}^{N_{\rm f}} u(\lceil x_1 + (n-1)/N_{\rm f} \rceil_1, \lceil x_2 + (n-1)/N_{\rm f} \rceil_1), \quad \lceil x \rceil_1 = x - [x] \quad (3.27)$$

For the first column of Fig. 15, no noise was added to the simulated data, but for the other two columns, 10% noise was added with n = 0.10 in (2.32). For the third column of Fig. 12, only  $L^2$  regularization without the assistance of TV regularization or segmentation. On the other hand, along with segmentation,  $\mu = 10^{-4}$  and  $\mu = 3 \cdot 10^{-4}$  were used for the first and second columns, respectively. All other parameters are the same as those used for the examples of Figs. 11 - 14. The reconstructed images U and the reconstruction error images  $\Delta U$  are shown respectively in the top and bottom rows of Fig. 15. For the first, second and third columns of Fig. 15, the quantitative errors are  $d_2(U, U^*) = 0.030$ , 0.040 and 0.048 and  $d_{\infty}(U, U^*) = 0.19$ , 0.24 and 0.28 respectively.

Finally, Fig. 16 shows a comparison between the presently proposed methods and the authors' implementation of the methods used in [21] combined with an iteratively regularized Gauss-Newton method as proposed in [4]. Results of the former methods are shown in the left column and results of the latter methods are shown in the right column. As in Fig. 15 these results were obtained using  $N_c = 4$  coils and  $N_r = 8$  reference points contained in a  $3 \times 3$ square centered at the origin in frequency space. Also the folding factor was  $N_f = 2$  in both the horizontal and vertical directions. No additional noise was added to the simulated data. For the first column in Fig. 16 only  $L^2$  regularization was used without assistance from TV regularization or from segmentation, and all other parameters were the same as those used for



Figure 14: In all three examples shown here, Algorithm 3 was used to reconstruct the exact image shown in Fig. 3a using simulated data containing 5% noise and with  $N_c = 6$  and  $N_f = 4$ . Also, the number of reference points is  $N_r = 54$ , 20 and 6 for the results obtained in the first, second and third columns, respectively. The images in the top row, (a), (b) and (c), are the reconstructions obtained with neither TV regularization nor segmentation, and the images in the bottom row, (d), (e) and (f), were obtained with TV regularization and segmentation.

the examples of Fig. 15. The regularization used for the result in the right column of Fig. 16 was a spectral implementation of a high-order Sobolev penalty on sensitivities and an  $L_2$  penalty on the reconstructed image. The reconstructed images are shown in the top row of Fig. 16 and the reconstruction error images are shown in the bottom row. The quantitative errors for the first column in Fig. 16 are  $d_2(\boldsymbol{U}, \boldsymbol{U}^*) = 0.034$  and  $d_{\infty}(\boldsymbol{U}, \boldsymbol{U}^*) = 0.20$ . The quantitative errors for the second column in Fig. 16 are  $d_2(\boldsymbol{U}, \boldsymbol{U}^*) = 0.061$  and  $d_{\infty}(\boldsymbol{U}, \boldsymbol{U}^*) = 0.36$ . As mentioned in Section 1, the spectral regularization of sensitivities implicitly imposes an unnatural periodicity which can lead to the artifacts seen in the second column of Fig. 16. On the other hand, the results shown in Fig. 16 are a test of an extreme case in which very little data are given, and both methods provide better results when  $N_c$  and  $N_r$  are higher. Furthermore, the authors' MATLAB codes used to obtain the left and right columns of Fig. 16 differ by an order of magnitude in speed with the spectral approach being faster. Nevertheless, optimized implementations on specialized hardware can narrow this gap in speed.

Thus, the developed approach for PMRI is accurate, and reconstructions from noisy data are advantageous with the TV regularization used along with the segmentation approach. However, since there is no trace of modulation or aliasing errors in the purely  $L^2$  regularized reconstructions, e.g., in Figs. 12 - 15, one finds, in contrast to Fig. 11, that additional regularization methods are not crucial for anti-aliasing when  $N_r$  is sufficiently large and when  $N_c$  is sufficiently larger than  $N_f$ , depending upon the subsampling strategy. Nevertheless, these regularization methods improve image quality based upon direct processing of raw data as opposed to postprocessing reconstructions obtained by other means possibly with less quality.

## References

[1] F. BAUER and S. KANNENGIESSER, An alternative approach to the image reconstruction



Figure 15: In all three examples shown here, Algorithm 3 was used to reconstruct the exact image shown in Fig. 3a using simulated data with  $N_c = 4$  and  $N_r = 8$  and with  $N_f = 2$  in both the vertical and horizontal directions. The upper left image (a) is the reconstruction obtained with noise-free data, and the corresponding reconstruction error image is shown in the lower left image (b). The upper middle image (c) is the reconstruction obtained with 10% noise in the data, and the corresponding reconstruction error image is shown in the lower middle image (d). The upper right image (e) is the reconstruction obtained with neither TV regularization nor segmentation and with 10% noise in the data, and the corresponding reconstruction error image is shown in the lower right image (f).

for parallel data acquisition in MRI, Math. Methods Appl. Sci., Vol. 30, pp. 1437–1451, 2007.

- [2] E.J. CANDÈS, J. ROMBERG and T. TAO, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Transactions on Information Theory, Vol. 52, No. 2, pp. 489 – 509, 2006.
- [3] T. CHAN, G. GOLUB and P. MULET, A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration, SIAM J. Sci. Comput., Vol. 20, No. 6, pp. 1964 – 1977, 1999.
- [4] C. CLASON and G. VON WINCKEL, On a bilinear optimization problem in parallel magnetic resonance imaging, SFB-Report 2008-009, Graz, Austria, 2008, submitted.
- [5] L.C. EVANS, *Partial Differential Equations*, American Mathematical Society, Providence, 1998.
- [6] G.H. GOLUB and C.F. VAN LOAN, *Matrix computations*, (3rd ed.), page 50, Johns Hopkins University Press, 1996.
- [7] M. GRASMAIR, H. GROSSAUER, and M. HALTMEIER, F. LENZEN and O. SCHERZER, Variational Methods in Imaging, Applied Mathematical Sciences, Vol. 167, Springer Verlag, New York, 2008.
- [8] M.A. GRISWOLD, P.M. JAKOB, and R.M. HEIDEMANN, Generalized autocalibrating partially parallel acquisitions (GRAPPA), Magnetic Resonance in Medicine, Vol. 47, pp. 1202– 1210, 2002.



Figure 16: In the examples shown here, two separate algorithms were used to reconstruct the exact image shown in Fig. 3a using noise free simulated data with  $N_c = 4$  and  $N_r = 8$  and with  $N_f = 2$  in both the vertical and horizontal directions. The upper left image (a) is the reconstruction obtained from Algorithm 3 with neither TV regularization nor segmentation, and the corresponding reconstruction error image is shown in the lower left image (b). The upper right image (c) is the reconstruction obtained with the authors' implementation of the methods of [21], and the corresponding reconstruction error image is shown in the lower right image (d).

- [9] M. HINTERMÜLLER, S.L. KEELING and A. LAURAIN, Modulation recovery and image reconstruction in MRI: a structural study by parameterization, SFB-Report No. 2008-023, Graz, Austria, 2008, submitted.
- [10] M. HINTERMÜLLER and G. STADLER, A primal-dual algorithm for TV-based inf-convolution-type image restoration, SIAM J. Sci. Comput., Vol. 28, No. 1, pp. 1–23, 2006.
- [11] M. HINTERMÜLLER and K. KUNISCH, Total bounded variation regularization as a bilaterally constrained optimization problem, SIAM J. Appl. Math., Vol. 64, pp. 1311–1333, 2004.
- [12] M. HINTERMÜLLER and A. LAURAIN, Multiphase image segmentation and modulation recovery based on shape and topological sensitivity, SFB Report No. 2008-015, Graz, Austria, 2008, submitted.
- [13] J. HORNAK, The Basics of MRI, Interactive Learning Software, Henrietta, NY, 1996-2008. online: http://www.cis.rit.edu/htbooks/mri/.
- [14] S.L. KEELING, Total Variation Based Convex Filters for Medical Imaging, Applied Mathematics and Computation, Vol. 139, pp. 101-119, 2003.
- [15] S.L. KEELING and R. BAMMER, A Variational Approach to Magnetic Resonance Coil Sensitivity Estimation, Applied Mathematics and Computation, Vol. 158, No. 2, pp. 53– 82, 2004.
- [16] S.L. KEELING and G. HAASE, Geometric Multigrid for High-Order Regularizations of Early Vision Problems, Applied Mathematics and Computation, Vol. 184, pp. 536-556, 2007.

- [17] S.L. KEELING, T. KOGLER and R. STOLLBERGER, Nonparametric Deconvolution Approaches to Dynamic Contrast Enhanced Magnetic Resonance Imaging, to appear in Medical Image Analysis.
- [18] F. KNOLL, M. UNGER, F. EBNER, R. STOLLBERGER, Real Time Elimination of Undersampling Artifacts in CE MRA using Variational Denoising on Graphics Hardware, 20th Annual Int. Conference on Magnetic Resonance Angiography, 2008.
- [19] K.P. PRUESSMANN, M. WEIGER, M.B. SCHEIDEGGER, and P. BOESIGER, SENSE: Sensitivity encoding for fast MRI, Magnetic Resonance in Medicine, Vol. 42, pp. 952–962, 1999.
- [20] I. EKELAND and R. TÉMAM, Convex Analysis and Variational Problems, Vol. 28 of Classics in Applied Mathematics, SIAM, Philadelphia, PA, English ed., 1999. (Translated from French.)
- [21] M. UECKER, T. HOHAGE, K.T. BLOCK, and J. FRAHM, Image reconstruction by regularized nonlinear inversion - joint estimation of coil sensitivities and image content, Magnetic Resonance in Medicine, Vol. 60, pp. 674 – 682, 2008.
- [22] L.L. WALD, L. CARVAJAL, S.E. MOYHER, S.J. NELSON, P.E. GRANT, A.J. BARKOVICH and D.B. VIGNERON, *Phased array detectors and an automated intensity-correction algorithm for high-resolution MR imaging of the human brain*, Magnetic Resonance in Medicine, Vol. 34, No. 3, pp. 433 – 439, 1995.
- [23] D. WEISHAUPT, V.D. KOECHLI and B. Marincek, How does MRI work? An Introduction to the Physics and Function of Magnetic Resonance Imaging, Spinger, Berlin, 2nd ed., 2006.

# A Derivation of Primal-Dual Optimality System

For completeness, the Fenchel Duality Theorem is stated as follows; see [20] for the details. Let  $J(u) = \mathcal{F}(\Lambda u) + \mathcal{G}(u)$  be a decomposition of the cost J, where the operators  $\mathcal{F} : Y \to \mathbb{R} \cup \{\infty\}$ ,  $\mathcal{G} : V \to \mathbb{R} \cup \{\infty\}$  and  $\Lambda : V \to Y$  satisfy the following. First  $\Lambda$  is a bounded linear operator from V into Y. Then  $\mathcal{F}$  and  $\mathcal{G}$  are convex, lower semicontinuous functionals on Y and V, and there exists a  $u_0 \in V$  such that  $J(u_0) < \infty$ , and  $\mathcal{F}$  is continuous in Y at  $\Lambda u_0$ . Then the minimization of J is characterized by  $\inf_{u \in V} [\mathcal{F}(\Lambda u) + \mathcal{G}(u)] = \sup_{p \in Y^*} [-\mathcal{F}^*(-p) - \mathcal{G}^*(\Lambda^*p)]$  where  $\mathcal{F}^*$  and  $\mathcal{G}^*$  are the convex conjugates defined for instance by  $\mathcal{F}^*(p) = \sup_{q \in Y} [\langle q, p \rangle_{Y,Y^*} - \mathcal{F}(q)]$ . Also,  $\Lambda^*$  is the adjoint of  $\Lambda$ . Furthermore, the optimality system for this problem is given by  $\mathcal{F}(p) + \mathcal{F}^*(-\Lambda^*p) = -\langle p, \lambda u \rangle_{Y,Y^*}$  and  $\mathcal{G}(u) + \mathcal{F}^*(\Lambda u) = \langle u, \Lambda^*p \rangle_{V,V^*}$ .

In the present work,  $V = Y = L^2(\Omega)$  and  $\Lambda = I$ , so the pairings  $\langle \cdot, \cdot \rangle_{Y,Y^*}$  and  $\langle \cdot, \cdot \rangle_{V,V^*}$  are both the inner product on  $L^2(\Omega)$ . Then  $\mathcal{G}$  and  $\mathcal{F}$  are convex, lower semicontinuous functionals defined by (2.13) and (2.12) or (3.6). For  $u_0 = 0$  the cost satisfies  $J(u_0) < \infty$ , and the residual  $\mathcal{F}$  is continuous in  $L^2(\Omega)$  at  $u_0$ .

To obtain the desired optimality system, the convex conjugates are calculated first. The convex conjugate of  $\mathcal{F}$  in (3.6) is given as follows:

$$\mathcal{F}^*(v) = \sup_{u \in L^2(\Omega)} F_v(u) := \int_{\Omega} \left[ uv - \frac{1}{2} \sum_{i=1}^{N_c} |P\sigma_i u - \tilde{u}_i|^2 - \frac{\kappa}{2} u^2 \right] d\boldsymbol{x}$$

The directional derivative of  $F_v(u)$  with respect to a perturbation  $\bar{u}$  is given by:

$$\frac{\delta F_v}{\delta u}(u;\bar{u}) = \int_{\Omega} \left[ \bar{u}v - \sum_{i=1}^{N_c} \left( P\sigma_i u - \tilde{u}_i \right) P\sigma_i \bar{u} - \kappa u \bar{u} \right] d\boldsymbol{x} = \int_{\Omega} \left[ v - \sum_{i=1}^{N_c} \left( \sigma_i P\sigma_i u - \sigma_i \tilde{u}_i \right) - u \right] \bar{u} d\boldsymbol{x}$$

where  $P\tilde{u}_i = \tilde{u}_i$ . Under the conditions of Theorem 2 the operator  $\kappa I + \sum_{k=1} \sigma_k P \sigma_k$  is invertible, so setting the critical

$$u^* = \left[\kappa I + \sum_{k=1}^{N_c} \sigma_k P \sigma_k\right]^{-1} \left(v + \sum_{l=1}^{N_c} \sigma_l \tilde{u}_l\right)$$

in  $F_v(u)$  gives the maximum of this convex functional:

$$F_{v}(u^{*}) = \frac{1}{2} \int_{\Omega} \left\{ \left( v + \sum_{l=1}^{N_{c}} \sigma_{l} \tilde{u}_{l} \right) \left[ \kappa I + \sum_{i=1}^{N_{c}} \sigma_{i} P \sigma_{i} \right]^{-1} \left( v + \sum_{l=1}^{N_{c}} \sigma_{l} \tilde{u}_{l} \right) - \tilde{u}_{i}^{2} \right\} d\boldsymbol{x}$$

as seen in (3.7). Also, for the case that  $N_c = 1$  and P = I hold, this formula reduces to (2.15). Next, the convex conjugate of  $\mathcal{G}$  in (2.13) is given as follows:

$$\mathcal{G}^*(v) = \sup_{v \in L^2(\Omega)} G_v(u) := \int_{\Omega} \left[ uv - \mu \phi_{\epsilon}(|\nabla u|^2) \right] d\boldsymbol{x}$$

For a fixed  $v \in L^2(\Omega)$ , let the constant  $\bar{v}$  be chosen so that  $v_0 = v - \bar{v}$  has zero average value over  $\Omega$ . Then using (2.18) let  $w = \Delta_N^{-1} v_0$ , so that w also has zero average value over  $\Omega$ . A lower bound for  $\mathcal{G}_v^*(u)$  is first obtained by maximizing  $G_v(u)$  over the finite dimensional set  $\{u = -sw + r : r, s \in \mathbf{R}\}$ . For elements in this set,  $G_v(u)$  satisfies:

$$G_{v}(u) = \int_{\Omega} \left[ u\bar{v} - \nabla u \cdot \nabla w - \mu \phi_{\epsilon}(|\nabla u|^{2}) \right] d\mathbf{x}$$

If  $|\nabla w| > \mu$  holds, let u = -sw + r be chosen with  $s > \epsilon/\mu$  and  $r\bar{v} > 0$ , so that  $|\nabla u| > \epsilon$  holds, and according to (2.3):

$$\int_{\Omega} \left[ u\bar{v} - \nabla u \cdot \nabla w - \mu \phi_{\epsilon}(|\nabla u|^2) \right] d\boldsymbol{x} = \left( r\bar{v} + \frac{\epsilon\mu}{2} \right) |\Omega| + s \int_{\Omega} |\nabla w| [|\nabla w| - \mu] d\boldsymbol{x} - s\bar{v} \int_{\Omega} w d\boldsymbol{x}$$

where the last integral is zero. Note that the right side becomes infinitely large as s and  $r\bar{v}$ become infinitely large. Thus, the indicator function  $I_{S_{\mu}}$  appears in (2.16). Henceforth it is assumed that  $v \in S_{\mu}$ , where  $S_{\mu}$  is defined in (2.17), so  $\bar{v} = 0$  and  $|\nabla w| \leq \mu$  hold. Now an upper bound for  $G_v(u)$  is determined over the dense subset of  $L^2(\Omega)$  with  $u \in C^{\infty}(\Omega)$ . For elements in this set,  $G_v(u)$  can be written as:

$$\begin{array}{lll} G_{v}(u) &=& \int_{|\nabla u| \leq \epsilon} \left[ u \Delta w - \frac{\mu}{2\epsilon} |\nabla u|^{2} \right] d\boldsymbol{x} + \int_{|\nabla u| > \epsilon} \left[ u \Delta w - \mu |\nabla u| - \frac{\mu\epsilon}{2} \right] d\boldsymbol{x} \\ &=& \int_{|\nabla u| \leq \epsilon} \left[ -\nabla u \cdot \nabla w - \frac{\mu}{2\epsilon} |\nabla u|^{2} \right] d\boldsymbol{x} + \int_{|\nabla u| > \epsilon} \left[ -\nabla u \cdot \nabla w - \mu |\nabla u| + \frac{\mu\epsilon}{2} \right] d\boldsymbol{x} \\ &\leq& \int_{|\nabla u| \leq \epsilon} \left[ |\nabla u| |\nabla w| - \frac{\mu}{2\epsilon} |\nabla u|^{2} \right] d\boldsymbol{x} + \int_{|\nabla u| > \epsilon} \left[ |\nabla u| |\nabla w| - \mu |\nabla u| + \frac{\mu\epsilon}{2} \right] d\boldsymbol{x} \end{array}$$

It will now be shown that both integrands can be bounded by  $\epsilon |\nabla w|^2/(2\mu)$ . The integrand of the first integral is a quadratic function of  $|\nabla u|$  defined on  $|\nabla u| \leq \epsilon$  which achieves its maximum value of  $\epsilon |\nabla w|^2/(2\mu)$  at  $|\nabla u| = \epsilon |\nabla w|/\mu$ . Because of the non-negative slope  $(|\nabla w| - \mu)$ , the integrand of the second integral is a linear function of  $|\nabla u|$  defined on  $|\nabla u| > \epsilon$  which approaches a maximum value as  $|\nabla u| \to \epsilon$ . Thus,

$$|\nabla u||\nabla w| - \mu |\nabla u| + \frac{\mu \epsilon}{2} \le \epsilon |\nabla w| - \frac{\epsilon \mu}{2} \le \frac{\epsilon}{2\mu} |\nabla w|^2$$

where the difference between the rightmost term and the term next to it is  $\epsilon(|\nabla w| - \mu)^2/(2\mu) \ge 0$ . These calculations show that

$$G_v(u) \le \frac{\epsilon}{2\mu} \int_{\Omega} |\nabla w|^2 d\boldsymbol{x}$$

Also, setting  $u^* = -\epsilon w/\mu$  in  $G_v(u)$  gives:

$$G_{v}(u^{*}) = \int_{\Omega} \left[ \frac{\epsilon}{\mu} |\nabla w|^{2} - \frac{\mu}{2\epsilon} (\frac{\epsilon}{\mu})^{2} |\nabla w|^{2} \right] d\boldsymbol{x} = \frac{\epsilon}{2\mu} \int_{\Omega} |\nabla w|^{2} d\boldsymbol{x}$$

as seen in (2.16).

With these convex conjugates the primal problem is related to the dual problem as follows:

$$\inf_{u\in L^{2}(\Omega)} \left\{ \frac{1}{2} \sum_{i=1}^{N_{c}} \int_{\Omega} |P\sigma_{i}u - \tilde{u}_{i}|^{2} d\boldsymbol{x} + \frac{\kappa}{2} \int_{\Omega} u^{2} d\boldsymbol{x} \frac{\mu}{2} \int_{\Omega} \phi_{\epsilon}(|\nabla u|^{2}) d\boldsymbol{x} \right\}$$

$$= \inf_{u\in L^{2}(\Omega)} [\mathcal{F}(u) + \mathcal{G}(u)] = \sup_{v\in L^{2}(\Omega)} [-\mathcal{F}^{*}(-v) - \mathcal{G}^{*}(v)] =$$

$$\sup_{v\in S_{\mu}} \left\{ \frac{1}{2} \int_{\Omega} \left\{ \left(v + \sum_{l=1}^{N_{c}} \sigma_{l} \tilde{u}_{l}\right) \left[\kappa I + \sum_{i=1}^{N_{c}} \sigma_{i} P\sigma_{i}\right]^{-1} \left(v + \sum_{l=1}^{N_{c}} \sigma_{l} \tilde{u}_{l}\right) - \tilde{u}_{i}^{2} \right\} d\boldsymbol{x} + \frac{\epsilon}{2\mu} \int_{\Omega} |\nabla \Delta_{N} v|^{2} d\boldsymbol{x} \right\}$$
(A.28)

where  $S_{\mu}$  is defined in (2.17). Here it can be seen from the last term that  $\epsilon > 0$  in (2.3) provides a regularization in the dual problem, which motivates its use in [10].

Now the optimality system (2.14) is obtained with the constructions above. For  $\mathcal{F}$  in (3.6) and  $\mathcal{F}^*$  in (3.7) the equation  $\mathcal{F}(u) + \mathcal{F}^*(v) = \int_{\Omega} uv d\boldsymbol{x}$  can be written as follows:

$$0 = \frac{1}{2} \int_{\Omega} \left\{ \sum_{i=1}^{N_{c}} [P\sigma_{i}u - \tilde{u}_{i}]^{2} + \frac{\kappa}{2}u^{2} + \left(v + \sum_{l=1}^{N_{c}} \sigma_{l}\tilde{u}_{l}\right) \left[\kappa I + \sum_{i=1}^{N_{c}} \sigma_{i}P\sigma_{i}\right]^{-1} \left(v + \sum_{l=1}^{N_{c}} \sigma_{l}\tilde{u}_{l}\right) - \tilde{u}_{i}^{2} - 2uv \right\} d\boldsymbol{x}$$
$$= \frac{1}{2} \int_{\Omega} \left(\kappa u \sum_{i=1}^{N_{c}} \sigma_{i}P\sigma_{i}u + v - \sum_{l=1}^{N_{c}} \sigma_{l}\tilde{u}_{l}\right) \left[\kappa I + \sum_{i=1}^{N_{c}} \sigma_{i}P\sigma_{i}\right]^{-1} \times \left(\kappa u + \sum_{i=1}^{N_{c}} \sigma_{i}P\sigma_{i}u + v - \sum_{l=1}^{N_{c}} \sigma_{l}\tilde{u}_{l}\right) d\boldsymbol{x}$$

which is the first equation in (3.9) with  $v = \Delta_{\rm N} w$ . Also, for the case that  $N_{\rm c} = 1$  and P = I hold, this formula reduces to the first equation in (2.19) with  $v = \Delta_{\rm N} w$ .

For  $\mathcal{G}$  in (2.13) and  $\mathcal{G}^*$  in (2.16) the equation  $\mathcal{G}(u) + \mathcal{G}^*(-v) = -\int_{\Omega} uv dx$  can be written as follows in terms of  $\Delta_N w = v \in S_{\mu}$ :

$$0 = \int_{\Omega} \left[ \mu \phi_{\epsilon} (|\nabla u|^{2}) + \frac{\epsilon}{2\mu} |\nabla (\Delta_{N}^{-1} v)|^{2} + uv \right] dx = \int_{\Omega} \left[ \mu \phi_{\epsilon} (|\nabla u|^{2}) + \frac{\epsilon}{2\mu} |\nabla w|^{2} - \nabla u \cdot \nabla w \right] dx$$

$$= \int_{|\nabla u| < \epsilon} \left[ \frac{\mu}{2\epsilon} |\nabla u|^{2} + \frac{\epsilon}{2\mu} |\nabla w|^{2} - \nabla u \cdot \nabla w \right] dx$$

$$+ \int_{|\nabla u| \ge \epsilon} \left[ \mu |\nabla u| - \frac{\mu\epsilon}{2} + \frac{\epsilon}{2\mu} |\nabla w|^{2} - \nabla u \cdot \nabla w \right] dx$$

$$= \int_{|\nabla u| < \epsilon} \frac{\epsilon}{2\mu} \left| \nabla w - \frac{\mu}{\epsilon} \nabla u \right|^{2} dx \quad (\ge 0)$$
(A.29)
$$+ \int_{|\nabla u| < \epsilon} \left[ \frac{\epsilon}{2\mu} |\nabla w - \frac{\mu}{\epsilon} \nabla u \right]^{2} - \frac{\mu}{\epsilon} (|\nabla u| - v)^{2} dx$$

$$+ \int_{|\nabla u| \ge \epsilon} \left[ \frac{\epsilon}{2\mu} \left| \nabla w - \frac{\mu}{\epsilon} \nabla u \right|^2 - \frac{\mu}{2\epsilon} \left( |\nabla u| - \epsilon \right)^2 \right] d\boldsymbol{x} \quad (\ge 0)$$
(A.30)

The non-negativity of the integrand in (A.29) is evident, and the non-negativity of the integrand in (A.30) can be seen from the following estimate:

$$\int_{|\nabla u| \ge \epsilon} \left[ \frac{\epsilon}{2\mu} \left| \nabla w - \frac{\mu}{\epsilon} \nabla u \right|^2 - \frac{\mu}{2\epsilon} \left( |\nabla u| - \epsilon \right)^2 \right] d\boldsymbol{x}$$
(A.31)

$$\geq \int_{|\nabla u| \ge \epsilon} \frac{\mu}{2\epsilon} \left[ \left( \frac{\epsilon}{\mu} |\nabla w| - |\nabla u| \right)^2 - \left( |\nabla u| - \epsilon \right)^2 \right] d\boldsymbol{x}$$
(A.32)

$$= \int_{|\nabla u| \ge \epsilon} (\mu - |\nabla w|) \left[ \frac{\epsilon}{2\mu} (\mu - |\nabla w|) + (|\nabla u| - \epsilon) \right] d\mathbf{x} \ge 0$$
 (A.33)

Since each integrand in (A.29) and (A.30) is non-negative and the sum is zero, each integrand must be pointwise zero. That the integrand in (A.33) must vanish means that  $|\nabla w| = \mu$  must hold when  $|\nabla u| \ge \epsilon$  holds. That the two integrals in (A.31) and (A.32) are zero implies, after removing identical terms in each, that

$$0 = \int_{|\nabla u| \ge \epsilon} [|\nabla u| |\nabla w| - \nabla u \cdot \nabla w] d\boldsymbol{x}$$

where the integrand is non-negative. Thus,  $\nabla u \cdot \nabla w = |\nabla u| |\nabla w|$  holds, i.e.,  $\nabla u/|\nabla u|$  and  $\nabla w/|\nabla w|$  are parallel unit vectors, so  $|\nabla u| \nabla w = |\nabla w| \nabla u = \mu \nabla u$  holds when  $|\nabla u| \ge \epsilon$  holds. That the integrand in (A.29) is non-negative means that  $\epsilon \nabla w = \mu \nabla u$  holds when  $|\nabla u| < \epsilon$  holds, and it follows that  $\epsilon |\nabla w| = \mu |\nabla u| < \epsilon \mu$  or  $|\nabla w| < \mu$  holds. These conditions can be summarized as follows:

$$\begin{array}{ll} (|\nabla u| \geq \epsilon \Rightarrow) & |\nabla w| = \mu, & |\nabla u| \nabla w = \mu \nabla u \\ (|\nabla u| < \epsilon \Rightarrow) & |\nabla w| < \mu, & \epsilon \nabla w = \mu \nabla u \end{array}$$

or that  $\max\{\epsilon, |\nabla u|\} \nabla w = \mu \nabla u$ , as seen in the second equation of (2.19) and (3.9).