Preconditioned ADMM with Nonlinear Operator Constraint

Martin Benning[1](✉), Florian Knoll[2], Carola-Bibiane Schönlieb[1],
and Tuomo Valkonen[3]

[1] Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Wilberforce Road, Cambridge CB3 0WA, UK
{mb941,cbs31}@cam.ac.uk
[2] Center for Advanced Imaging, Innovation and Research, New York University,
4th Floor 660 First Avenue, New York, NY 10016, USA
florian.knoll@nyumc.org
[3] Department of Mathematical Sciences, University of Liverpool,
Mathematical Sciences Building, Liverpool L69 7ZL, UK
tuomo.valkonen@liverpool.ac.uk

Abstract. We are presenting a modification of the well-known Alternating Direction Method of Multipliers (ADMM) algorithm with additional preconditioning that aims at solving convex optimisation problems with nonlinear operator constraints. Connections to the recently developed Nonlinear Primal-Dual Hybrid Gradient Method (NL-PDHGM) are presented, and the algorithm is demonstrated to handle the nonlinear inverse problem of parallel Magnetic Resonance Imaging (MRI).

Keywords: ADMM · Primal-dual · Nonlinear inverse problems · Parallel MRI · Proximal point method · Operator splitting · Iterative Bregman method

1 Introduction

Non-smooth regularisation methods are popular tools in the imaging sciences. They allow to promote sparsity of inverse problem solutions with respect to specific representations; they can implicitly restrict the null-space of the forward operator while guaranteeing noise suppression at the same time. The most prominent representatives of this class are total variation regularisation [19] and ℓ^1-norm regularisation as in the broader context of compressed sensing [8,10].

In order to solve convex, non-smooth regularisation methods with linear operator constraints computationally, first-order operator splitting methods have gained increasing interest over the last decade, see [3,9,11,12] to name just a few. Despite some recent extensions to certain types of non-convex problems [7,14–16] there has to our knowledge only been made little progress for nonlinear operators constraints [2,22].

In this paper we are particularly interested in minimising non-smooth, convex functionals with nonlinear operator constraints. This model covers many interesting applications; one particular application that we are going to address is the joint reconstruction of the spin-proton-density and coil sensitivity maps in parallel MRI [13,21].

The paper is structured as follows: we will introduce the generic problem formulation, then address its numerical minimisation via a generalised ADMM method with linearised operator constraints. Subsequently we will show connections to the recently proposed NL-PDHGM method (indicating a local convergence result of the proposed algorithm) and conclude with the joint spin-proton-density and coil sensitivity map estimation as a numerical example.

2   Problem Formulation

We consider the following generic constrained minimisation problem:

$$(\hat{u}, \hat{v}) = \arg\min_{u,v} \left\{ H(u) + J(v) \text{ subject to } F(u, v) = c \right\}. \tag{1}$$

Here H and J denote proper, convex and lower semi-continuous functionals, F is a nonlinear operator and c a given function. Note that for nonlinear operators of the form $F(u, v) = G(u) - v$ and $c = 0$ problem (1) can be written as

$$\hat{u} = \arg\min_{u} \left\{ H(u) + J(G(u)) \right\}. \tag{2}$$

In the following we want to propose a strategy for solving (1) that is based on simultaneous linearisation of the nonlinear operator constraint and the solution of an inexact ADMM problem.

3   Alternating Direction Method of Multipliers

We solve (1) by alternating optimisation of the augmented Lagrange function

$$\mathcal{L}_\delta(u, v; \mu) = H(u) + J(v) + \langle \mu, F(u, v) - c \rangle + \frac{\delta}{2} \|F(u, v) - c\|_2^2. \tag{3}$$

Alternating minimisation of (3) in u, v and subsequent maximisation of μ via a step of gradient ascent yields this nonlinear version of ADMM [11]:

$$u^{k+1} \in \arg\min_{u} \left\{ \frac{\delta}{2} \|F(u, v^k) - c\|_2^2 + \langle \mu^k, F(u, v^k) \rangle + H(u) \right\}, \tag{4}$$

$$v^{k+1} \in \arg\min_{v} \left\{ \frac{\delta}{2} \|F(u^{k+1}, v) - c\|_2^2 + \langle \mu^k, F(u^{k+1}, v) \rangle + J(v) \right\}, \tag{5}$$

$$\mu^{k+1} = \mu^k + \delta \left(F(u^{k+1}, v^{k+1}) - c \right). \tag{6}$$

Not having to deal with nonlinear subproblems, we replace $F(u, v^k)$ and $F(u^{k+1}, v)$ by their Taylor linearisations around u^k and v^k, which yields $F(u, v^k) \approx F(u^k, v^k) + \partial_u F(u^k, v^k) \left(u - u^k \right)$ and $F(u^{k+1}, v) \approx F(u^{k+1}, v^k) + \partial_v F(u^{k+1}, v^k) \left(v - v^k \right)$, respectively. The updates (4) and (5) modify to

$$u^{k+1} \in \arg\min_u \left\{ \frac{\delta}{2} \left\| A^k u - c_1^k \right\|_2^2 + \langle \mu^k, A^k u \rangle + H(u) \right\}, \tag{7}$$

$$v^{k+1} \in \arg\min_v \left\{ \frac{\delta}{2} \left\| B^k v - c_2^k \right\|_2^2 + \langle \mu^k, B^k v \rangle + J(v) \right\}, \tag{8}$$

with $A^k := \partial_u F(u^k, v^k)$, $B^k := \partial_v F(u^{k+1}, v^k)$, $c_1^k := c + A^k u^k - F(u^k, v^k)$ and $c_2^k := c + B^k v^k - F(u^{k+1}, v^k)$. Note that the updates (7) and (8) are still implicit, regardless of H and J. In the following, we want to modify the updates such that they become simple proximity operations.

4   Preconditioned ADMM

Based on [23], we modify (7) and (8) by adding the surrogate terms $\|u^{k+1} - u^k\|_{Q_1^k}^2/2$ and $\|v^{k+1} - v^k\|_{Q_2^k}^2/2$, with $\|w\|_Q := \sqrt{\langle Qw, w \rangle}$ (note that if Q is chosen to be positive definite, $\|\cdot\|_Q$ becomes a norm). We then obtain

$$u^{k+1} \in \arg\min_u \left\{ \frac{\delta}{2} \left\| A^k u - c_1^k \right\|_2^2 + \langle \mu^k, A^k u \rangle + H(u) + \frac{1}{2}\|u - u^k\|_{Q_1^k}^2 \right\},$$

$$v^{k+1} \in \arg\min_v \left\{ \frac{\delta}{2} \left\| B^k v - c_2^k \right\|_2^2 + \langle \mu^k, B^k v \rangle + J(v) + \frac{1}{2}\|v - v^k\|_{Q_2^k}^2 \right\}.$$

If we choose $Q_1^k := \tau_1^k I - \delta A^{k*} A^k$ with $\tau_1^k \delta < 1/\|A^k\|^2$ and $Q_2^k := \tau_2^k I - \delta B^{k*} B^k$ with $\tau_2^k \delta < 1/\|B^k\|^2$ and if we define $\overline{\mu}^k := 2\mu^k - \mu^{k-1}$ we obtain

$$u^{k+1} = \left(I + \tau_1^k \partial H \right)^{-1} \left(u^k - \tau_1^k A^{k*} \overline{\mu}^k \right), \tag{9}$$

$$v^{k+1} = \left(I + \tau_2^k \partial J \right)^{-1} \left(v^k - \tau_2^k B^{k*} \left(\mu^k + \delta \left(F(u^{k+1}, v^k) - c \right) \right) \right), \tag{10}$$

with $(I + \alpha \partial E)^{-1}(w)$ denoting the proximity or resolvent operator

$$(I + \alpha \partial E)^{-1}(w) := \arg\min_u \left\{ \frac{1}{2}\|u - w\|_2^2 + \alpha E(u) \right\}.$$

The entire proposed algorithm with updates (9), (10) and (6) reads as

---

Algorithm 1. Preconditioned ADMM with nonlinear operator constraint

---

Parameters: H, J, F, c
Initialization: u^0, v^0, μ^0, δ
$\overline{\mu}^0 = \mu^0$
while convergence criterion is not met **do**

$\quad A^k = \partial_u F(u^k, v^k)$
$\quad$ Set τ_1^k such that $\tau_1^k \delta < 1/\|A^k\|^2$
$\quad u^{k+1} = \left(I + \tau_1^k \partial H \right)^{-1} \left(u^k - \tau_1^k A^{k*} \overline{\mu}^k \right)$
$\quad B^k = \partial_v F(u^{k+1}, v^k)$
$\quad$ Set τ_2^k such that $\tau_2^k \delta < 1/\|B^k\|^2$
$\quad v^{k+1} = \left(I + \tau_2^k \partial J \right)^{-1} \left(v^k - \tau_2^k B^{k*} \left(\mu^k + \delta \left(F(u^{k+1}, v^k) - c \right) \right) \right)$
$\quad \mu^{k+1} = \mu^k + \delta \left(F(u^{k+1}, v^{k+1}) - c \right)$
$\quad \overline{\mu}^{k+1} = 2\mu^{k+1} - \mu^k$

end while
return u^k, v^k, μ^k, $\overline{\mu}^k$

---

5   Connection to NL-PDHGM

In the following we want to show how the algorithm simplifies in case the non-linear operator constraint is only nonlinear in one variable, which is sufficient for problems of the form (2). Without loss of generality we consider constraints of the form $F(u,v) = G(u) - v$, where G represents a nonlinear operator in u. Then we have $A^k = \mathcal{J}G(u^k)$ (with $\mathcal{J}G(u^k)$ denoting the Jacobi matrix of G at u^k), $B^k = -I$ and if we further choose $\tau_2^k = 1/\delta$ for all k, update (10) reads

$$v^{k+1} = \left(I + \frac{1}{\delta}\partial J\right)^{-1}\left(G(u^{k+1}) + \frac{1}{\delta}\mu^k\right).$$

Applying Moreau's identity [18] $b = \left(I + \frac{1}{\delta}\partial J\right)^{-1}(b) + \frac{1}{\delta}(I + \delta\partial J^*)^{-1}(\delta b)$ yields

$$\mu^{k+1} = (I + \delta\partial J^*)^{-1}\left(\mu^k + \delta G(u^{k+1})\right).$$

If we further change the order of the updates, starting with the update for μ, the whole algorithm reads

$$\mu^{k+1} = (I + \delta\partial J^*)^{-1}\left(\mu^k + \delta G(u^k)\right),$$
$$\overline{\mu}^{k+1} = 2\mu^{k+1} - \mu^k,$$
$$u^{k+1} = \left(I + \tau_1^k\partial H\right)^{-1}\left(u^k - \tau_1^k\,\mathcal{J}G(u^k)^*\overline{\mu}^{k+1}\right).$$

Note that this algorithm is almost the same as NL-PDHGM proposed in [22] for $\theta = 1$, except that the extrapolation step is carried out on the dual variable μ instead of the primal variable u. In the following we want to briefly sketch how to prove convergence for this algorithm in analogy to [22]. We define

$$N(\mu^{k+1}, u^{k+1}) := \begin{pmatrix} \partial J^*(\mu^{k+1}) - \nabla G(u^k)u^{k+1} - c^k \\ \partial H(u^{k+1}) + \mathcal{J}G(u^k)^*\mu^{k+1} \end{pmatrix},$$

$$L^k := \begin{pmatrix} \frac{1}{\delta}I & \mathcal{J}G(u^k) \\ \mathcal{J}G(u^k)^* & \frac{1}{\tau_1^k}I \end{pmatrix},$$

with $c^k := G(u^k) - \mathcal{J}G(u^k)u^k$. Now the algorithm is: find (μ^{k+1}, u^{k+1}) such that

$$N(\mu^{k+1}, u^{k+1}) + L^k(\mu^{k+1} - \mu^k, u^{k+1} - u^k) \ni 0.$$

If we exchange the order of μ and u here, i.e., reorder the rows of N, and the rows and columns of L^k, we obtain *almost* the "linearised" NL-PDHGM of [22]. The difference is that the sign of $\mathcal{J}G$ in L^k is inverted. The only points in [22] where the exact structure of L^k (M_{x^k} therein) is used, are Lemma 3.1, Lemma 3.6 and Lemma 3.10. The first two go through exactly as before with the negated structure. Reproducing Lemma 3.10 demands bounding actual step lengths $\|u^k - u^{k+1}\|$ and $\|\mu^k - \mu^{k+1}\|$ from below, near a solution for arbitrary $\epsilon > 0$. A proof would go beyond the page limit of this proceeding. Let us just point out that this can be done, implying that the convergence results of [22] apply for this algorithm as well. This means that under somewhat technical regularity conditions, which for TV type problems amount to Huber regularisation, local convergence in a neighbourhood of the true solution can be guaranteed.

6 Joint Estimation of the Spin-Proton Density and Coil Sensitivities in Parallel MRI

We want to demonstrate the numerical capabilities of Algorithm 1 by applying it to the nonlinear problem of joint estimation of the spin-proton density and the coil sensitivities in parallel MRI. The discrete problem of joint reconstruction from sub-sampled k-space data on a rectangular grid reads

$$
\begin{pmatrix} \hat{u} \\ \hat{c}_1 \\ \vdots \\ \hat{c}_2 \end{pmatrix} \in \operatorname*{arg\,min}_{\mathbf{v}=(u,c_1,\ldots,c_n)} \left\{ \frac{1}{2} \sum_{j=1}^{n} \| S\mathcal{F}(G(\mathbf{v}))_j - f_j \|_2^2 + \alpha_0 R_0(u) + \sum_{j=1}^{n} \alpha_j R_j(c_j) \right\},
$$

where $\mathcal{F}$ is the 2D discrete Fourier transform, f_j are the k-space measurements for each of the n coils, S is the sub-sampling operator and R_j denote appropriate regularisation functionals. The nonlinear operator G maps the unknown spin-proton density u and the different coil sensitivities c_j as follows [21]:

$$
G(u, c_1, \ldots, c_n) = (uc_1, uc_2, \ldots, uc_n)^T. \tag{11}
$$

In order to compensate for sub-sampling artefacts in sub-sampled MRI it is common practice to use total variation as a regulariser [6,17]. Coil sensitivities are assumed to be smooth, cf. Fig. 1, motivating a reconstruction model similar to the one proposed in [13]. We therefore choose the discrete isotropic total variation, $R_0(u) = \|\nabla u\|_{2,1}$, and the smooth 2-norm of the discretised gradient, i.e. $R_j(c_j) := \|\nabla c_j\|_{2,2}$, for all $j > 0$, following the notation in [4]. We further introduce regularisation parameters λ_j in front of the data fidelities and rescale all regularisation parameters such that $\alpha_0 + \frac{1}{n}\left(\sum_{j=1}^{n}\lambda_j + \sum_{j=1}^{n}\alpha_j\right) = 1$. In order to realise this model via Algorithm 1 we consider the following operator splitting strategy. We define $F(u_0, \ldots, u_n, v_0, \ldots, v_{2n})$ as

$$
F(u_0,\ldots,u_n,v_1,\ldots,v_n) := \begin{pmatrix} G(u_0,\ldots,u_n) \\ \nabla u_0 \quad 0 \quad \cdots \quad 0 \\ 0 \quad \nabla u_1 \ddots \quad \vdots \\ \vdots \quad \ddots \quad \ddots \quad 0 \\ 0 \quad \cdots \quad 0 \quad \nabla u_n \end{pmatrix} - \begin{pmatrix} v_0 \\ \vdots \\ v_n \\ \vdots \\ v_{2n} \end{pmatrix},
$$

set $H(u_0, \ldots, u_n) \equiv 0$, and $J(v_0, \ldots, v_{2n}) = \sum_{j=0}^{2n} J_j(v_j)$ with $J_j(v_j) := \frac{\lambda_j}{2}\|S\mathcal{F}v_j - f_j\|_2^2$ for $j \in \{0, \ldots, n-1\}$, $J_n(v_n) = \alpha_0\|v_n\|_{2,1}$ and $J_j(v_j) = \alpha_{j-n}\|v_j\|_{2,2}$ for $j \in \{n+1, \ldots, 2n\}$. Note that with these choices of functions, all the resolvent operations can be carried out easily. In particular, we obtain

$$(I + \tau_1^k \partial H)^{-1}(w) = w,$$

$$(I + \tau_2^k \partial J_j)^{-1}(w) = \mathcal{F}^{-1} \left(\frac{\mathcal{F}w_j + \tau_2^k \lambda_j S^T f_j}{1 + \tau_2^k \lambda_j \mathrm{diag}(S^T S)} \right) \text{ for } j \in \{0, \ldots, n-1\},$$

$$(I + \tau_2^k \partial J_n)^{-1}(w) = \frac{w_n}{\|w_n\|_2} \max\left(\|w_n\|_2 - \alpha_0 \tau_2^k, 0\right),$$

$$(I + \tau_2^k \partial J_j)^{-1}(w) = \frac{w_j}{\|w_j\|_{2,2}} \max\left(\|w_j\|_{2,2} - \alpha_{j-n}\tau_2^k, 0\right) \text{ for } j \in \{n+1, \ldots, 2n\}.$$

Moreover, as $B_k = -I$ (and thus, $\|B^k\| = 1$) for all k, we can simply eliminate τ_2^k by replacing it with $1/\delta$, similar to Sect. 5.

(a) Brain phantom   (b) 25% sub-sampling

(c) 1st coil   (d) 2nd coil   (e) 3rd coil   (f) 4th coil

(g) 5th coil   (h) 6th coil   (i) 7th coil   (j) 8th coil
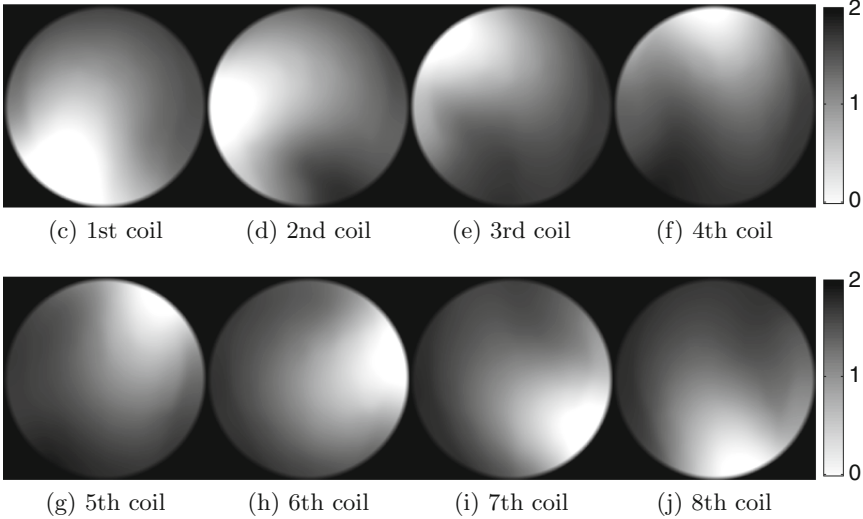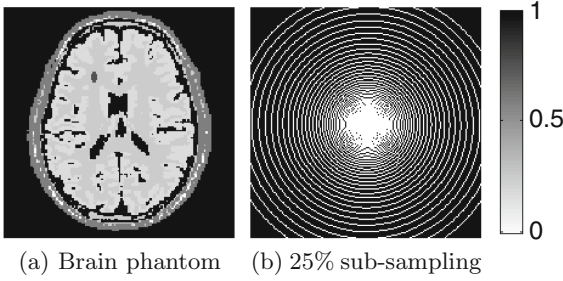
Fig. 1. (a) shows the brain phantom as described in Sect. 6.1. (c)–(j) show visualisations of the measured coil sensitivities of a water bottle. (b) shows the simulated, spiral-shaped sub-sampling scheme used to sub-sample the k-space data.
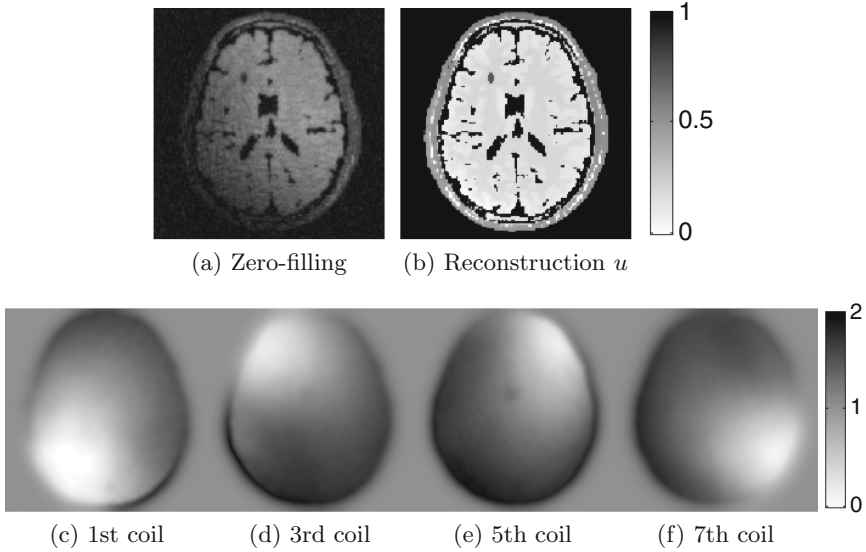
(a) Zero-filling   (b) Reconstruction u

(c) 1st coil   (d) 3rd coil   (e) 5th coil   (f) 7th coil

Fig. 2. Reconstructions for noise with low noise level $\sigma = 0.05$. Despite the sub-sampling, features of the brain phantom are very well preserved. In addition, the coil sensitivities seem to correspond well to the original ones, despite a slight loss of contrast. Note that coil sensitivities remain the initial value where the signal is zero.

6.1 Experimental Setup

We now want to discuss the experimental setup. We want to reconstruct the synthetic brain phantom in Fig. 1a from sub-sampled k-space measurements. The numerical phantom is based on the design in [1] with a matrix size of 190×190. It consists of several different tissue types like cerebrospinal fluid (CSF), gray matter (GM), white matter (WM) and cortical bone. Each pixel is assigned a set of MR tissue properties: Relaxation times $T_1(x,y)$ and $T_2(x,y)$ and spin density $\rho(x,y)$. These parameters were also selected according to [1]. The MR signal $s(x,y)$ in each pixel was then calculated by using the signal equation of a fluid attenuation inversion recovery (FLAIR) sequence [5]:

$$s(x,y) = \rho(x,y)(1 - 2\,e^{-TI/T_1(x,y)})(1 - e^{-TR/T_1(x,y)})\,e^{-TE/T_2(x,y)}.$$

The sequence parameters were selected: $TR = 10000$ ms, $TE = 90$ ms. TI was set to 1781 ms to achieve signal nulling of CSF ($T_1^{csf} \log(2)$ with $T_1^{csf} = 2569$ ms).

In order to generate artificial k-space measurements for each coil, we proceed as follows. First, we produce 8 images of the brain phantom multiplied by the measured coil sensitivity maps shown in Fig. 1c–j. The coil sensitivity maps were generated from the measurements of a water bottle with an 8-channel head coil array. Then we produce artificial k-space data by applying the 2D discrete Fourier-transform to each of those individual images. Subsequently, we sub-sample only approx. 25% of each of the k-space datasets via the spiral shown in Fig. 1b. Finally, we add Gaußian noise with standard deviation σ to the sub-sampled data.

(a) Zero-filling   (b) Reconstruction u

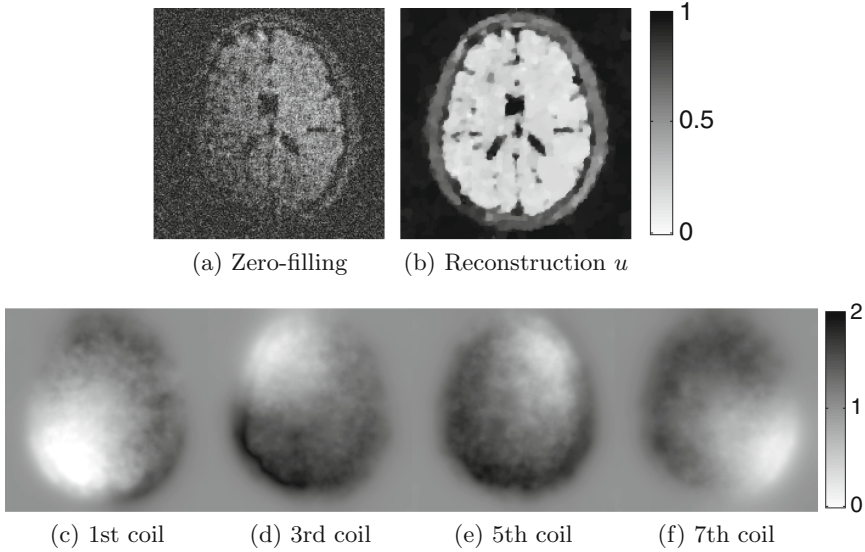(c) 1st coil   (d) 3rd coil   (e) 5th coil   (f) 7th coil

Fig. 3. Reconstructions for noise with high noise level $\sigma = 0.95$. Due to the large amount of noise, higher regularisation parameters are necessary. As a consequence, fine structures are smoothed out and in contrast to the case of little noise, compensation of sub-sampling artefacts is less successful.

6.2   Computations

For the actual computations we use two noisy versions f_j of the simulated k-space data; one with small noise ($\sigma = 0.05$) and one with a high amount of noise ($\sigma = 0.95$). As stopping criterion we simply choose a fixed number of iterations; for both the low noise level as well as the high noise level dataset we have fixed the number of iterations to 1500. The initial values used for the algorithm are $u_j^0 = \mathbf{1}$ with $\mathbf{1} \in \mathbb{R}^{l \times 1}$ being the constant one-vector, for all $j \in \{0, \ldots, n\}$. All other initial variables $(v^0, \mu^0, \overline{\mu}^0)$ are set to zero.

Low Noise Level. We have computed reconstructions from the noisy data with noise level $\sigma = 0.05$ via Algorithm 1, with regularisation parameters set to $\lambda_j = 0.0621$, $\alpha_0 = 0.0062$ and $\alpha_j = 0.9317$ for $j \in \{1, \ldots, n\}$. We have further created a naïve reconstruction by averaging the individual inverse Fourier-transformed images obtained from zero-filling the k-space data. The modulus images of the results are visualised in Fig. 2. The PSNR values for the averaged zero-filled reconstruction is 10.2185, whereas the PSNR of the reconstruction with the proposed method is 24.5572.

High Noise Level. We proceeded as in the previous section, but for noisy data with noise level $\sigma = 0.95$. The regularisation parameters were set to $\lambda_j =$

0.0149, $\alpha_0 = 0.0135$ and $\alpha_j = 0.9716$ for $j \in \{1, \ldots, n\}$. The modulus images of the results are visualised in Fig. 3. The PSNR values for the averaged zero-filled reconstruction is 9.9621, whereas the PSNR of the reconstruction with the proposed method is 16.672.

7   Conclusions and Outlook

We have presented a novel algorithm that allows to compute minimisers of a sum of convex functionals with nonlinear operator constraint. We have shown the connection to the recently proposed NL-PDHGM algorithm which implies local convergence results in analogy to those derived in [22]. Subsequently we have demonstrated the computational capabilities of the algorithm by applying it to a nonlinear joint reconstruction problem in parallel MRI.

For future work, the convergence of the algorithm in the general setting has to be verified, and possible extensions to guarantee global convergence have to be studied. Generalisation of stopping criteria such as a linearised primal-dual gap will also be of interest as well. With respect to the presented parallel MRI application, exact conditions for the convergence (like the exact norm of the bounds) have to be verified. The impact of the algorithm - as well as the regularisation-parameters on the reconstruction has to be analysed, and a rigorous study with artificial and real data would also be desirable. Moreover, future research will focus on alternative regularisation functions, e.g. based on spherical harmonics motivated by [20]. Last but not least, other applications that can be modelled via (1) should be considered in future research.

EPSRC Data Statement: the corresponding code and data are available for download at https://www.repository.cam.ac.uk/handle/1810/256221.

References

1. Aubert-Broche, B., Evans, A.C., Collins, L.: A new improved version of the realistic digital brain phantom. NeuroImage **32**(1), 138–145 (2006)
2. Bachmayr, M., Burger, M.: Iterative total variation schemes for nonlinear inverse problems. Inverse Prob. **25**(10), 105004 (2009)
3. Beck, A., Teboulle, M.: Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems. IEEE Trans. Image Process. **18**(11), 2419–2434 (2009)
4. Benning, M., Gladden, L., Holland, D., Schönlieb, C.-B., Valkonen, T.: Phase reconstruction from velocity-encoded MRI measurements-a survey of sparsity-promoting variational approaches. J. Magn. Reson. **238**, 26–43 (2014)
5. Bernstein, M.A., King, K.F., Zhou, X.J.: Handbook of MRI Pulse Sequences. Elsevier, Amsterdam (2004)

6. Block, K.T., Uecker, M., Frahm, J.: Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. Magn. Reson. Med. **57**(6), 1086–1098 (2007)
7. Bonettini, S., Loris, I., Porta, F., Prato, M.: Variable metric inexact line-search based methods for nonsmooth optimization. Siam J. Optim. **26**, 891–921 (2015)
8. Candes, E.J., et al.: Compressive sampling. In: Proceedings of the International Congress of Mathematicians, vol. 3, Madrid, Spain, pp. 1433–1452 (2006)
9. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**(1), 120–145 (2011)
10. Donoho, D.L.: Compressed sensing. IEEE Trans. Inf. Theory **52**(4), 1289–1306 (2006)
11. Gabay, D.: Applications of the method of multipliers to variational inequalities. Stud. Math. Appl. **15**, 299–331 (1983)
12. Goldstein, T., Osher, S.: The split Bregman method for L1-regularized problems. SIAM J. Imaging Sci. **2**(2), 323–343 (2009)
13. Knoll, F., Clason, C., Bredies, K., Uecker, M., Stollberger, R.: Parallel imaging with nonlinear reconstruction using variational penalties. Magn. Reson. Med. **67**(1), 34–41 (2012)
14. Möllenhoff, T., Strekalovskiy, E., Möller, M., Cremers, D.: The primal-dual hybrid gradient method for semiconvex splittings. SIAM J. Imaging Sci. **8**(2), 827–857 (2015)
15. Möller, M., Benning, M., Schönlieb, C., Cremers, D.: Variational depth from focus reconstruction. IEEE Trans. Image Process. **24**(12), 5369–5378 (2015)
16. Ochs, P., Chen, Y., Brox, T., Pock, T.: iPiano: inertial proximal algorithm for nonconvex optimization. SIAM J. Imaging Sci. **7**(2), 1388–1419 (2014)
17. Ramani, S., Fessler, J., et al.: Parallel MR image reconstruction using augmented Lagrangian methods. IEEE Trans. Med. Imaging **30**(3), 694–706 (2011)
18. Rockafellar, R.T.: Convex Analysis. Princeton Mathematical Series, 46:49. Princeton University Press, Princeton (1970)
19. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Phys. D: Nonlinear Phenom. **60**(1), 259–268 (1992)
20. Sbrizzi, A., Hoogduin, H., Lagendijk, J.J., Luijten, P., den Berg, C.A.T.: Robust reconstruction of B1+ maps by projection into a spherical functions space. Magn. Reson. Med. **71**(1), 394–401 (2014)
21. Uecker, M., Hohage, T., Block, K.T., Frahm, J.: Image reconstruction by regularized nonlinear inversion joint estimation of coil sensitivities and image content. Magn. Reson. Med. **60**(3), 674–682 (2008)
22. Valkonen, T.: A primal-dual hybrid gradient method for nonlinear operators with applications to MRI. Inverse Prob. **30**(5), 055012 (2014)
23. Zhang, X., Burger, M., Osher, S.: A unified primal-dual algorithm framework based on Bregman iteration. J. Sci. Comput. **46**(1), 20–46 (2011)