# MULTIGRID METHODS FOR TOTAL VARIATION

Felipe Guerra* Tuomo Valkonen†

Abstract Based on a *nonsmooth coherence condition*, we construct and prove the convergence of a forward-backward splitting method that alternates between steps on a fine and a coarse grid. Our focus is on total variation regularised inverse imaging problems, specifically, their dual problems, for which we develop in detail the relevant coarse-grid problems. We demonstrate the performance of our method on total variation denoising and magnetic resonance imaging.

## 1 INTRODUCTION

In this work, we consider composite optimisation problems of the form

$$\min_{x\in X} F(x) + G(x), \tag{1.1}$$

where $F$ is convex and smooth, and $G$ is convex but possibly nonsmooth on a Hilbert space $X$. We want to apply forward-backward splitting [11] to this problem, while reducing computational effort by occasionally passing to a lower-dimensional problem. Such multigrid methods make possible the computationally efficient high-precision solution of partial differential equations [3]. The authors of [12, 13, 9, 1, 10, 15, 14] have looked into applying the same principle to large-scale optimisation problems and variational inequalities, many of these to *smoothed* total variation regularised imaging problems. No method, so far, treats nonsmoothness completely satisfactorily: while [1] allow $G$ to be nonsmooth, it requires $F$ to be strongly convex in addition to smooth. Such an assumption is rarely satisfied in problems of practical interest. In [9] only constrained quadratic problems are considered. In [13, 10], less assumptions are imposed on the fine-grid problem, however, the coarse-grid problems are required to be smooth. We want to exploit the inherent nonsmooth properties of the problem on the coarse grid as well.

We focus on imaging with isotropic total variation regularisation, i.e.,

$$\min_{y\in Y} E(y) + \alpha\|\nabla_h y\|_{2,1},$$

where $E$ is a data fitting term; $\nabla_h \in \mathbb{L}(Y;X)$ a (discrete) gradient operator; and $\|\cdot\|_{2,1}$ the sum over a pixelwise 2-norms. Since, in the present work, we are limited to forward-backward splitting, and the proximal map of total variation is not prox-simple, i.e., not easily calculated, we have to work with the dual problem

$$\min_{x\in X} E^*(-\nabla_h^* x) + \delta_{\alpha B_{2,1}}(x). \tag{1.2}$$

This research has been supported by the EPN internal project PIS-23-04.

*Research Center in Mathematical Modeling and Optimization (MODEMAT), Quito, Ecuador *and* Department of Mathematics, Escuela Politécnica Nacional (EPN), Quito, Ecuador. edison.guerra@epn.edu.ec, ORCID: 0009-0003-8329-9166

†MODEMAT *and* EPN *and* Department of Mathematics and Statistics, University of Helsinki, Finland. tuomo.valkonen@iki.fi, ORCID: 0000-0001-6683-3572

To be able to numerically calculate $\nabla E^*$ efficiently, indeed, also for $E^*$ to be smooth, we will, unfortunately, need $E$ to be strongly convex. We do not, however, require $E$ be smooth, unlike in [1], and outside imaging applications that practically require dualisation, will also be able to treat non-strongly-convex problems (1.1). Typically $E(y) = \frac{1}{2}\|Ty - b\|^2$ for a forward operator $T$ mapping an image to its measurements, so we will need $T$ to be invertible. This holds with fully or over-sampled data.

The problem (1.2) is of the form (1.1). A significant step in forming a multigrid optimisation method is deciding on a coarse-grid version of the problem. While MGProx [1] allows any smooth coarse-grid problem, we will follow the approach of [13]—where gradient descent for smooth problems was considered—in proposing in Section 2 a *nonsmooth coherence condition* that locally determines the coarse-grid problem. For (1.2), this will form coarse-grid constraints more difficult than $\alpha B_{2,1}$. We analyse the projection to those constraints in Section 4 after proposing and proving the convergence of the general method in Section 3. We finish with denoising and magnetic resonance imaging (MRI) experiments in Section 5.

This work builds upon the Master's thesis of the first author [7] with simplified proofs and coarse problem, and expanded numerics with a faster implementation [6].

Notation Let $X$ and $Y$ be Hilberts spaces, $A \subset X$. We write $\mathbb{L}(X;Y)$ for the bounded linear operators between $X$ and $Y$, $\delta_A$ for the $\{0,\infty\}$-valued indicator function of $A$, and $B(x,\alpha)$ for the closed ball of radius $\alpha$ and centre $x$ in $X$. We write $A^\circ := \{z \mid \langle z, x\rangle \le 0 \ \forall x \in A\}$ for the polar, and $A^{\circ\circ} := (A^\circ)^\circ$ for the bipolar. These satisfy $A^{\circ\circ} \supset A$ and $(A^{\circ\circ})^\circ = A^\circ$. The smallest closed convex cone containing $A$ is $\operatorname{ccone} A := A^{\circ\circ}$. The normal cone to $A$ at $x$ is $N_A(x) := \{z \mid \langle z, \tilde{x} - x\rangle \le 0 \ \forall \tilde{x} \in A\}$. For a convex $F : X \to \overline{\mathbb{R}}$, the subdifferential at $x$ is $\partial F(x)$, the proximal operator $\operatorname{prox}_F$, and the Fenchel conjugate $F^*$. We have $\partial\delta_A(x) = N_A(x)$. We refer to [5] for more details on these concepts. Finally $x_{\cdot\, j} \in \mathbb{R}^D$ is the $j$:th row of $x \in \mathbb{R}^{D\times n}$.

## 2 THE COARSE PROBLEM

The first question we have to answer is how to build the coarse problem? In this section, we construct a general guideline, the *nonsmooth coherence condition*, and prove that a forward-backward method applied to a coarse-grid problem satisfying this condition, will construct a descent direction for the fine grid.

### 2.1 THE NONSMOOTH COHERENCE CONDITION

Write $I_h^H \in \mathbb{L}(X; X_H)$ for the *restriction* operator from the fine grid modelled by the Hilbert space $X$ to the coarse grid modelled by the Hilbert space $X_H$. Typically $X$ and $X_H$ are finite-dimensional with $\dim X_H \ll \dim X$. We call $I_H^h \in \mathbb{L}(X_H; X)$ satisfying $\mu I_H^h = (I_h^H)^*$ for some $\mu > 0$ the *prolongation* operator.

In [12, 13], treating the smooth problem $\min_x F(x)$, the coarse-grid problems $\min_x F_H^k$ were built by introducing an arbitrary coarse objective $F_H$ that satisfies the smooth *coherence condition* $I_h^H \nabla F(x^k) = \nabla F_H(\zeta^{k,0})$, for an initial coarse point $\zeta^{k,0}$, typically $\zeta^{k,0} = I_h^H x^k \in X_H$, and setting

$$F_H^k(\zeta) := F_H(\zeta) + \langle w_H^k, \zeta - \zeta^{k,0}\rangle \quad \text{for} \quad w_H^k := I_h^H \nabla F(x^k) - \nabla F_H(\zeta^{k,0}) \in X_H.$$

Then $\nabla F_H^k(\zeta^{k,0}) = \nabla F(x^k)$, so if $x^k$ solves the fine-grid problem, the coarse grid problem triggers no change: it is solved by $\zeta^{k,0}$. A descent direction for the coarse-grid problem also allows constructing a fine-grid descent direction [12].

We extend this approach to the nonsmooth problems (1.1). Specifically, for $G_H^k$ satisfying the following two assumptions, we take as our *coarse-grid problem*

$$\min_{\zeta \in X_H} F_H^k(\zeta) + G_H^k(\zeta). \tag{2.1}$$

Assumption 2.1 (Basic coarse structure). $F_H : X_H \to \mathbb{R}$ is convex with $L_H$-Lipschitz gradient. For all $k \in \mathbb{N}$, $G_H^k : X \to \overline{\mathbb{R}}$ is convex, proper, lower semicontinuous. The step length parameter $\tau_H > 0$ satisfies $\epsilon := 2 - \tau_H L_H > 0$.

Assumption 2.2 (Nonsmooth coherence condition). The fine-grid iterate $x^k$ and the initial coarse iterate $\zeta^{k,0}$ (typically $I_h^H x^k$) satisfy $I_h^H \partial G(x^k) \subseteq \partial G_H^k(\zeta^{k,0})$.

Example 2.3. Take $G_H^k = \delta_{\Omega^k}$ for $\Omega^k := \zeta^{k,0} + (I_h^H \partial G(x^k))^\circ$. Obviously, $\Omega^k \subset X_H$ is nonempty and convex, and $\partial G_H^k(\zeta^{k,0}) = N_{\Omega_k}(\zeta^{k,0}) = (I_h^H \partial G(x^k))^{\circ\circ} \supset I_h^H \partial G(x^k)$.

### 2.2 COARSE-GRID ALGORITHM AND DESCENT DIRECTIONS

We apply forward-backward splitting to the coarse-grid problems (2.1). For an initial coarse point $\zeta^{k,0}$ and an iteration count $m \in \mathbb{N}$, we thus iterate

$$\zeta^{k,j+1} := \operatorname{prox}_{\tau_H G_H^k}(\zeta^{k,j} - \tau_H \nabla F_H^k(\zeta^{k,j})) \quad (j = 0, \dots, m-1). \tag{2.2}$$

In the following we show that $d = I_H^h(\zeta^{k,m} - \zeta^{k,0})$ is a fine-grid descent direction.

Lemma 2.4. *If Assumption 2.1 holds, and we apply (2.2) for any $\zeta^{k,0} \in X_H$, then*

$$\langle I_h^H \nabla F(x^k), \zeta^{k,m} - \zeta^{k,0} \rangle + \frac{\epsilon}{2\tau_H} \sum_{j=0}^{m-1} \|\zeta^{k,j+1} - \zeta^{k,j}\|_{X_H}^2 \le G_H^k(\zeta^{k,0}) - G_H^k(\zeta^{k,m}).$$

*Proof.* We abbreviate $J_H^k := G_H^k + F_H$ and $\zeta^j := \zeta^{k,j}$. In implicit form, (2.2) reads

$$0 \in \partial G_H^k(\zeta^{j+1}) + \nabla F_H(\zeta^j) + w_H^k + \tau_H^{-1}(\zeta^{j+1} - \zeta^j). \tag{2.3}$$

By the subdifferentiability of $G_H^k$ and the descent inequality $F_H(\zeta + h) \le F_H(\zeta) + \langle \nabla F_H(\zeta), h \rangle + \frac{L_H}{2}\|h\|_X^2$, valid for any $\zeta, h$ (see, e.g., [5, Theorem 7.1]),

$$\langle \partial G_H^k(\zeta^{j+1}) + \nabla F_H(\zeta^j), \zeta^{j+1} - \zeta^j \rangle \ge J_H^k(\zeta^{j+1}) - J_H^k(\zeta^j) - \frac{L_H}{2}\|\zeta^{j+1} - \zeta^j\|_{X_H}^2,$$

where the inequality and inner product are applied to all elements of $\partial G_H^k(\zeta^{j+1})$. Applying $\langle \cdot, \zeta^{j+1} - \zeta^j \rangle$ on both sides of (2.3) and using $\epsilon := 2 - \tau_H L_H$ thus yields

$$\langle w_H^k, \zeta^{j+1} - \zeta^j \rangle + \frac{\epsilon}{2\tau_H}\|\zeta^{j+1} - \zeta^j\|_{X_H}^2 \le J_H^k(\zeta^j) - J_H^k(\zeta^{j+1}).$$

Summing over $j = 0, \dots, m-1$, it follows

$$\langle w_H^k, \zeta^m - \zeta^0 \rangle + \frac{\epsilon}{2\tau_H} \sum_{j=0}^{m-1} \|\zeta^{j+1} - \zeta^j\|_{X_H}^2 \le J_H^k(\zeta^0) - J_H^k(\zeta^m).$$

Since $\epsilon > 0$, by the construction of $w_H^k$, we get using the convexity of $F_H$,

$$\langle w_H^k, \zeta^m - \zeta^0 \rangle \ge \langle I_h^H \nabla F(x^k), \zeta^m - \zeta^0 \rangle + F_H(\zeta^0) - F_H(\zeta^m).$$

Combining these two estimates and simplifying, we obtain the claim. □

Corollary 2.5. *Suppose Assumptions 2.1 and 2.2 hold. Let* $d := I_H^h(\zeta^{k,m} - \zeta^{k,0})$. *Then*

$$[G+F]'(x^k;d) = \sup_{g\in\partial G(x^k)} \langle g + \nabla F(x^k), d\rangle \leq -\frac{\epsilon}{2\mu\tau_H}\sum_{j=0}^{m-1} \|\zeta^{k,j+1} - \zeta^{k,j}\|^2 \leq 0.$$

*Proof.* By Assumption 2.2, $\langle I_h^H g, \zeta - \zeta^{k,0}\rangle \leq G_H^k(\zeta) - G_H^k(\zeta^{k,0})$ for all $\zeta$ and $g \in \partial G(x^k)$. Taking $\zeta = \zeta^{k,m}$, combining with Lemma 2.4, we obtain

$$\langle I_h^H g + I_h^H \nabla F(x^k), \zeta - \zeta^{k,0}\rangle + \frac{\epsilon}{2\tau_H}\sum_{j=0}^{m-1} \|\zeta^{k,j+1} - \zeta^{k,j}\|_{X_H}^2 \leq 0 \quad \forall\, g \in \partial G(x^k).$$

Since $(I_h^H)^* = \mu I_H^h$, taking the supremum over $g$, we obtain the middle inequality of the claim. The equality is standard, e.g., [5, Lemma 4.4]. □

Now, as $d$ is aligned with a negative subdifferential of the fine-grid objective, we readily show that it is a fine-grid descent direction.

Theorem 2.6. *Suppose Assumptions 2.1 and 2.2 hold. Then for* $d = I_H^h(\zeta^{k,m} - \zeta^{k,0})$ *and any* $\kappa \in (0,1)$, *there exists* $\theta_0 > 0$ *such that, for all* $0 < \theta \leq \theta_0$, *we have*

$$[G+F](x^k + \theta d) < [G+F](x^k) + \kappa\theta [G+F]'(x^k;d).$$

*Proof.* In Corollary 2.5, we use the definition of the directional derivative. □

## 3 FORWARD-BACKWARD MULTIGRID

We now develop the overall multigrid algorithm for (1.1). Our starting point, again, is the classical forward-backward splitting

$$x^{k+1} = \operatorname{prox}_{\tau G}(x^k - \tau\nabla F(x^k)).$$

As this is a monotone descent method, Theorem 2.6 suggests that performing coarse-grid iterations between its iterations would not ruin the convergence. However, details remain to attend to. To prove convergence, we adapt the technique of [1].

### 3.1 ALGORITHM

Our proposed Algorithm 3.1 is, for simplicity, limited to two grid levels, but, subject to designing deeper coarse versions of $G_H$, etc., it can, in principle, easily be extended to multiple levels. These functions need to again satisfy the nonsmooth coherence condition, which may require some effort. However, for the TV application, as the $G_H$ designed in Section 4.1 will be a polyhedral constraint, also deeper coarsenings remain polyhedral.

The algorithm depends on *line search* to guarantee (2.6), as well as a *trigger condition* to perform coarse corrections. Options for the latter include:

1. heuristic, after enough progress since the previous correction [13];
2. a fixed number of fine iterations between coarse corrections; or
3. a bounded number of coarse corrections during the algorithm runtime.

Although a standard line search procedure can be used, in practise, for efficiency, we either take at each coarse correction a fixed $\theta_k = \bar{\theta}_k$, or, if this does not satisfy the sufficient decrease condition (2.6), reject the coarse grid result with $\theta_k = 0$.

**Algorithm 3.1** Forward-backward multigrid (FBMG)

**Require:** $F, G, F_H$ and $\tau, \tau_H > 0$ satisfying Assumptions 2.1 and 3.1. Sufficient descent parameter $\kappa \in (0,1)$ as well as a line search procedure and trigger condition.

1: Choose an initial iterate $x^0 \in X$. Set $J := G + F$.
2: **for all** $k = 0, 1, 2, \ldots$ until a chosen stopping criterion is fulfilled **do**
3: **if** a trigger condition is satisfied **then**
4: Choose an initial coarse point $\zeta^{k,0}$ (e.g., $I_h^H x^k$).
5: Design $G_H^k$ satisfying the nonsmooth coherence condition (Assumption 2.2)
6: Set $w_H^k := I_h^H \nabla F(x^k) - \nabla F_H(\zeta^{k,0})$,
7: **for all** $j = 1, \ldots, m-1$ **do**
8: $\zeta^{k,j+1} := \operatorname{prox}_{\tau_H G_H^k}(\zeta^{k,j} - \tau_H[\nabla F_H(\zeta^{k,j}) + w_H^k])$ ▷ *Coarse-grid FB*
9: **end for**
10: Set $d := I_H^h(\zeta^{k,m} - \zeta^{k,0})$.
11: Find $\theta_k \geq 0$ such that $J(x^k + \theta_k d) \leq J(x^k) + \kappa\theta_k J'(x^k; d)$. ▷ *Line search*
12: $x^{k+1} := \operatorname{prox}_{\tau G}(z^k + \nabla F(z^k))$ for $z^k := x^k + \theta_k d$ ▷ *Fine-grid FB*
13: **else**
14: $x^{k+1} := \operatorname{prox}_{\tau G}(x^k + \nabla F(x^k))$ ▷ *Fine-grid FB*
15: **end if**
16: **end for**

### 3.2 CONVERGENCE

Assumption 3.1. $F, G : X \to \overline{\mathbb{R}}$ are proper, convex and lower semicontinuous, $F$ Fréchet differentiable with $L$-Lipschitz gradient. The step length $\tau \in (0, 1/L)$.

Theorem 3.2 (Sublinear convergence). *Suppose Assumption 3.1 holds, and $x^* \in [\partial J]^{-1}(0)$ where $J := G+F$. Let $\{x^k\}_{k\geq 1}$ be generated by Algorithm 3.1 for an initial $x^0 \in X$ such that the corresponding sublevel set is bounded, i.e.,*

$$\varepsilon := \operatorname{diam}(\operatorname{sub}_{J(x^0)} J) := \sup\{\|x - y\|_2 \mid J(x), J(y) \leq J(x^0)\} < \infty.$$

*Then, for any minimiser $x^*$ of $J$,*

$$J(x^{k+1}) - J(x^*) \leq \frac{1}{k}\max\{4C, J(x^0) - J(x^*)\} \quad \text{for} \quad C = \frac{2\varepsilon^2}{\tau(2-\tau L)} > 0. \tag{3.1}$$

*Proof.* Theorem 2.6 guarantees the line search on Line 11 of Algorithm 3.1 to be satisfiable (strictly for a $\theta_k > 0$, although we also allow $\theta_k = 0$ and mere non-increase). By standard arguments based on convexity, the descent lemma, and the Pythagoras identity (see the proof of [5, Theorem 11.4]), Line 12 satisfies

$$J(x^{k+1}) - J(\tilde{x}) + \frac{1}{2\tau}\|x^{k+1} - \tilde{x}\|_X^2 + \frac{1-\tau L}{2\tau}\|x^{k+1} - z^k\|_X^2 \leq \frac{1}{2\tau}\|z^k - \tilde{x}\|_X^2 \tag{3.2}$$

for any $\tilde{x}$. Since $\tau L < 1$, taking $\tilde{x} = x^*$, we get

$$\begin{aligned} J(x^{k+1}) - J(x^*) &\leq \frac{1}{2\tau}(\|z^k - x^*\|_X^2 - \|x^{k+1} - x^*\|_X^2) \\ &= \frac{1}{2\tau}(\|z^k - x^*\|_X - \|x^{k+1} - x^*\|_X)(\|z^k - x^*\|_X + \|x^{k+1} - x^*\|_X) \\ &\leq \frac{\varepsilon}{\tau}(\|z^k - x^*\|_X - \|x^{k+1} - x^*\|_X) \\ &\leq \frac{\varepsilon}{\tau}\|x^{k+1} - z^k\|_X. \end{aligned}$$

Using the fact that $x^*$ is a minimiser, so that the left-hand-side is non-negative, squaring gives

$$(J(x^{k+1}) - J(x^*))^2 \leq (\varepsilon\tau^{-1})^2 \|x^{k+1} - z^k\|_X^2. \tag{3.3}$$

On the other hand, taking $\tilde{x} = z^k$ in (3.2), and continuing with $J(z^k) \leq J(x^k)$ established by the line search procedure on Line 11 of Algorithm 3.1, we obtain

$$\frac{2 - \tau L}{2\tau} \|x^{k+1} - z^k\|_X^2 \leq J(z^k) - J(x^{k+1}) \leq J(x^k) - J(x^{k+1}).$$

Combining with (3.3) yields $(J(x^{k+1}) - J(x^*))^2 \leq C[J(x^k) - J(x^{k+1})]$. Repeating the analysis with $z^k = x^k$ establishes the same result for Line 14. Now, according to [8, Lemma 4], the monotonically decreasing sequence $\{J(x^k)\}_{k\in\mathbb{N}}$ satisfies (3.1). □

## 4 TOTAL VARIATION REGULARISED IMAGING PROBLEMS

We will in Section 5 apply Algorithm 3.1 to image processing problems of the form

$$\min_{y\in\mathbb{R}^n} \phi(y) + \alpha\|\nabla_h y\|_{2,1} \quad \text{for} \quad \phi(y) := \frac{1}{2}\sum_{s=1}^{t} \|T_s y - b_s\|_2^2, \tag{4.1}$$

where $\nabla_h \in \mathbb{L}(\mathbb{R}^n; \mathbb{R}^{D\times n})$ for some dimension $D$. Since the nonsmooth total variation regulariser is not prox-simple, to derive an efficient method, we will need to work with the dual problem. Since MRI involves complex numbers, we allow $T_s \in \mathbb{L}(\mathbb{R}^n; \mathbb{C}^n)$ and $b_s \in \mathbb{C}^n$. Thus the dual formulation is

$$\min_{x\in\mathbb{R}^{D\times n}} \phi^*(-\nabla_h^* x) + G(x) \quad \text{for} \quad G(x) := (\alpha\|\cdot\|_{2,1})^*(x) = \sum_{i=1}^{n} \delta_{B(0,\alpha)}(x_{\cdot i}). \tag{4.2}$$

We construct $\phi^* : \mathbb{R}^n \to \mathbb{R}$ in Section 4.3, after we have first constructed the coarse nonsmooth function $G_H^k$ and its proximal operator in Sections 4.1 and 4.2.

### 4.1 THE COARSE PROBLEM

We use the standard restriction operator[1] $I_h^H = R \otimes R$, where, in stencil notation, $R = [\ \frac{1}{2}\ 1\ \frac{1}{2}\ ]$; see [3]. The prolongation operator is then $I_h^H = \frac{1}{4}(I_H^h)^*$. We take

$$G_H^k(\zeta) := \sum_{l=1}^{N} \delta_{\Omega_l}(\zeta_{\cdot l}) \quad \text{for} \quad \Omega_l := \zeta_{\cdot l}^{k,0} + \Gamma_l^\circ, \tag{4.3}$$

where for all coarse pixel indices $l = 1, \ldots, N$, we define

$$\Gamma_l := [I_h^H \partial G(x^k)]_l = \left\{ \sum\nolimits_{p\in A_l} q_{\cdot p}^k \,\middle|\, q_{\cdot p}^k \in [\partial G(x^k)]_p \ \forall p \in A_l \right\}, \tag{4.4}$$

with $A_l \subset \{1, \ldots, n\}$ the subset of fine pixel indices $i$ that contribute to the coarse pixel $l$ via $I_h^H$, that is, $[I_h^H]_{li} \neq 0$. Note that $\Omega_l$ is nonempty, closed and convex for all $l = 1, \ldots, N$. Apart from the Lipschitz gradient, there are no theoretical restrictions on $F_H$, but we take it as a coarse version of $\phi^* \circ -\nabla_h^*$, as we describe later in Example 4.7, after forming $\phi^*$.

[1] For MRI we could replace $I_H^h$ by $\mathcal{F}^* I_H^h \mathcal{F}$, where $\mathcal{F}$ is the Fourier transform in the relevant grid, but do not currently use this form.

Lemma 4.1. *Assumption 2.2 holds for $G$ and $G_H^k$ as in (4.2) and (4.3).*

*Proof.* We recall for all fine pixel indices $i = 1, .., n$ that

$$\left[\partial G(x^k)\right]_i = \partial\delta_{B(0,\alpha)}(x^k_{\cdot i}) = \begin{cases} \{0\}, & x^k_{\cdot i} \in \operatorname{int} B(0,\alpha), \\ \{\beta x^k_{\cdot i} \mid \beta \geq 0\}, & x^k_{\cdot i} \in \operatorname{bd} B(0,\alpha), \\ \emptyset & \text{otherwise.} \end{cases} \tag{4.5}$$

It follows that $\Gamma_l$ is a (possibly non-convex) cone in $\mathbb{R}^D$. We then deduce

$$\partial\delta_{\Omega_l}(\zeta^{k,0}_{\cdot l}) = N_{\Omega_l}(\zeta^{k,0}_{\cdot l}) = N_{\Gamma_l^\circ}(0) = (\Gamma_l^\circ)^\circ \supset \Gamma_l,$$

and further

$$I_h^H \partial G(x^k) = \Gamma_1 \times \cdots \times \Gamma_N \subset \partial\delta_{\Omega_l}(\zeta^{k,0}_{\cdot 1}) \times \cdots \times \partial\delta_{\Omega_l}(\zeta^{k,0}_{\cdot N}) = \partial G_H^k(\zeta^{k,0}). \qquad \square$$

### 4.2 THE COARSE PROXIMAL OPERATOR

By (4.3), $\Omega_l - \zeta_l^{k,0} = \Gamma_l^\circ = (\Gamma_l^{\circ\circ})^\circ$ is a closed convex cone for each $l$. When $D = 2$, it is therefore defined by at most two "director" vectors. If we can identify them, it will be possible to write the proximal operator of $G_H^k$ in a simple form.

Lemma 4.2. *Suppose $D = 2$. For any coarse pixel index $l \in \{1, \ldots, N\}$, $\Gamma_l^{\circ\circ}$ can only take for some fine pixel indices $j, s \in A_l$ one of the forms*

$$\{0\}, \quad \operatorname{ccone}\{x_{\cdot j}\}, \quad \operatorname{ccone}\{x_{\cdot j}, x_{\cdot s}\}, \quad \operatorname{ccone}\{x_{\cdot j}, x_{\cdot s}, -x_{\cdot j}\}, \quad \mathbb{R}^2.$$

*Proof.* We construct the director vectors algorithmically, starting with empty sets $V_l$ and $O_l$. The former will eventually contain the directors, while $O_l$ tracks subspaces generated by opposing vectors. For all $p \in A_l$, we repeat steps 1 and 2:

1. We omit $x^k_{\cdot p} \in \operatorname{int} B(0,\alpha)$ since, by (4.4), the subgradient $q^k_{\cdot p} = 0$, does not contribute to $\Gamma_l$. By contrast, if $x^k_{\cdot p} \in \operatorname{bd} B(0,\alpha)$, we add this vector to $V_l$.

2. If $V_l = \{z_j, z_s, z_p\}$ for three distinct vectors, one of them must be superfluous for forming $\Gamma_l^\circ$. To discard it, we form the linear system $z_j\beta_1 + z_s\beta_2 = z_p$ for the unknowns $\beta_1$ and $\beta_2$, and consider several cases:

   (a) For a non-unique solution, $z_s = cz_j$ for some $c \neq 0$. When $c > 0$, $z_s$ is superfluous. When $c < 0$, $\Gamma_l$ must be contained in the subspace orthogonal to $z_s$. In both cases, we take $V_l = \{z_j, z_p\}$, and in the latter add $z_s$ to $O_l$.

   Otherwise the system has a unique solution, and we continue with the cases:

   (b) If $\beta_1 \geq 0$ and $\beta_2 \geq 0$, then $z_p \in \operatorname{ccone}\{z_j, z_s\}$, so we remove $z_p$ from $V_l$.
   (c) If $\beta_1 > 0$ and $\beta_2 < 0$, then $z_j \in \operatorname{ccone}\{z_s, z_p\}$, so we remove $z_j$ from $V_l$.
   (d) If $\beta_1 < 0$ and $\beta_2 > 0$, then $z_s \in \operatorname{ccone}\{z_j, z_p\}$, so we remove $z_s$ from $V_l$.
   (e) If $\beta_1 = 0$ and $\beta_2 < 0$ or $\beta_1 < 0$ and $\beta_2 = 0$, then $z_s = -z_p$ or $z_j = -z_p$, so we eliminate $z_p$ from $V_l$ but add it to $O_l$.
   (f) If $\beta_1 < 0$ and $\beta_2 < 0$, then $\Gamma_l^\circ = \{0\}$, so we terminate with $\Gamma_l^{\circ\circ} = \mathbb{R}^2$.

If the construction did not terminate in the above loop, we consider:

(i) If $|V_l| + |O_l| \leq 2$, then by construction $O_l = \emptyset$. If $V_l = \emptyset$, also $\Gamma_l^{\circ\circ} = \{0\}$. Otherwise $\Gamma_l^{\circ\circ} = \operatorname{ccone} V_l$.

(ii) If $|V_l| + |O_l| = 3$, then $O_l = \{z_p\}$ and $V_l = \{z_j, z_s\}$ with $z_p = -z_j$ or $z_p = -z_s$. In other words, the three vectors define the half-space $\Gamma_l^{\circ\circ} = \operatorname{ccone}\{z_j, z_s, -z_j\}$.

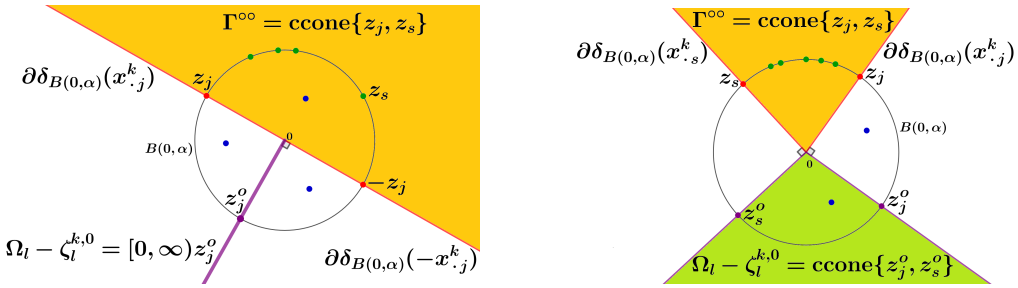



Figure 1: Illustration of the middle case of (4.6a) (left) and (4.6b) (right).

(iii) $|V_l| + |O_l| \geq 4$ then $O_l$ defines two distinct subspaces that contain $\Gamma_l^\circ$, which must then by $\{0\}$. Hence $\Gamma_l^{\circ\circ} = \mathbb{R}^2$. □

In the next proposition, $z_j^o$ denotes a vector orthogonal to $z_j$ with $\sup_{z\in\Gamma}\langle z, z_j^o\rangle \leq 0$. We also write $p(\zeta, z) := \max\{0, \langle\zeta, z\rangle\}z/\|z\|_2^2$ for the projection of $\zeta$ to $z[0,\infty)$.

Proposition 4.3. *The proximal operator of $G_H^k$ defined in (4.3) is given by*

$$[\operatorname{prox}_{\gamma G_H^k}(\zeta)]_l = \begin{cases} \zeta_l, & \Gamma_l^{\circ\circ} = \{0\}, \\ \zeta_l - p(\zeta_l - \zeta_l^{k,0}, z_j), & \Gamma_l^{\circ\circ} = \operatorname{ccone}\{z_j\}, \\ \zeta_l^{k,0} + p(\zeta_l - \zeta_l^{k,0}, z_j^o), & \Gamma_l^{\circ\circ} = \operatorname{ccone}\{z_j, z_s, -z_j\}, \\ \zeta_l^{k,0}, & \Gamma_l^{\circ\circ} = \mathbb{R}^2, \\ \textit{see below} & \Gamma_l^{\circ\circ} = \operatorname{ccone}\{z_j, z_s\}, \end{cases} \tag{4.6a}$$

*for each component $l = 1, \ldots, m$, where in the final case,*

$$[\operatorname{prox}_{\gamma G_H^k}(\zeta)]_l = \begin{cases} \zeta_l^{k,0}, & \zeta_l \in \operatorname{ccone}\{z_j, z_s\}, \\ \zeta_l, & \zeta_l \in \operatorname{ccone}\{z_j^o, z_s^o\}, \\ \zeta_l^{k,0} + p(\zeta_l - \zeta_l^{k,0}, z_j^o), & \zeta_l \in \operatorname{ccone}\{z_j, z_j^o\}, \\ \zeta_l^{k,0} + p(\zeta_l - \zeta_l^{k,0}, z_s^o), & \zeta_l \in \operatorname{ccone}\{z_s, z_s^o\}. \end{cases} \tag{4.6b}$$

*Proof.* The proximal map of $G_H^k$ separates into individual Euclidean projections onto each $\Omega_l$. These are determined by $\Gamma_l^{\circ\circ}$, so we consider the cases of Lemma 4.2:

(a) If $\Gamma_l^{\circ\circ} = \{0\}$, we have $\Omega_l = \mathbb{R}^2$. Hence the proximal map is the identity.

(b) If $\Gamma_l^{\circ\circ} = \mathbb{R}^2$, we have $\Omega_l = \{\zeta_l^{k,0}\}$, so the proximal map is the constant $\zeta_l^{k,0}$.

(c) If $\Gamma_l^{\circ\circ} = \operatorname{ccone}\{z_j\}$, then $\Omega_l = \zeta_l^{k,0} + (\Gamma_l^{\circ\circ})^\circ = \zeta_l^{k,0} + \{z_j\}^\circ$. Thus $[\operatorname{prox}_{\gamma G_H^k}(\zeta)]_l = \zeta_l - p(\zeta_l - \zeta_l^{k,0}, z_j)$.

(d) If $\Gamma_l^{\circ\circ} = \operatorname{ccone}\{z_j, z_s, -z_j\}$, then, likewise, $\Omega_l = \zeta_l^{k,0} + [0,\infty)z_j^o$, resulting in $[\operatorname{prox}_{\gamma G_H^k}(\zeta)]_l = \zeta_l^{k,0} - p(\zeta_l - \zeta_l^{k,0}, z_j^o)$; see Figure 1.

(e) When $\Gamma_l^{\circ\circ} = \operatorname{ccone}\{z_j, z_s\}$, we can divide $\mathbb{R}^2$ into four cones, each with distinct projection operator agreeing with (4.6b); see Figure 1. □

### 4.3 THE FENCHEL CONJUGATE OF THE (COMPLEX) DATA TERM

To form the Fenchel conjugate of $\phi$, we start with its derivative. We write $\|\cdot\|_{\mathbb{R}^n}$ for the Euclidean norm in $\mathbb{R}^n$ and $\|x\|_{\mathbb{C}^n} := \sqrt{\sum_{k=1}^n |x_i|^2}$ for $x = (x_1, \dots, x_n) \in \mathbb{C}^n$.

Lemma 4.4. *$\mathbb{R}^n \ni \nabla\phi(y) = Ty - e$ for $T := \sum_{s=1}^t \operatorname{Re} T_s^* T_s$ and $e := \sum_{s=1}^t \operatorname{Re} T_s^* b_s$.*

*Proof.* Due to properties of the complex inner product, for any $y, h \in \mathbb{R}^n$,

$$\begin{aligned}\|T_s(y+h) - b_s\|_{\mathbb{C}^n}^2 - \|T_s y - b_s\|_{\mathbb{C}^n}^2 &= 2\operatorname{Re}\langle T_s y - b_s, T_s h\rangle_{\mathbb{C}^n} + \|T_s h\|_{\mathbb{C}^n}^2 \\ &= 2\langle \operatorname{Re} T_s^*(T_s y - b_s), h\rangle_{\mathbb{R}^n} + \|T_s h\|_{\mathbb{C}^n}^2.\end{aligned}$$

Dividing by 2 and summing over $s = 1, \dots, t$, therefore

$$\phi(y+h) - \phi(y) = \langle Ty - e, h\rangle_{\mathbb{R}^n} + \frac{1}{2}\sum_{s=1}^t \|T_s h\|_{\mathbb{C}^n}^2.$$

By the definition of the Fréchet derivative, the claim follows. □

From the definition of the Fenchel conjugate and the Fermat principle, now

$$\phi^*(z) = \langle z, T^{-1}(z+e)\rangle_{\mathbb{R}^n} - \phi(T^{-1}(z+e)). \tag{4.7}$$

We next develop a more tractable form, which shows (4.2) to have equivalent form

$$\min_{x \in \mathbb{R}^{D\times n}} \frac{1}{2}\|T^{-1/2}(\nabla_h^* x - e)\|_{\mathbb{R}^n}^2 + \sum_i \delta_{B(0,\alpha)}(x_{\cdot i}). \tag{4.8}$$

Lemma 4.5. *If $A \in \mathbb{L}(\mathbb{C}^n; \mathbb{C}^n)$. Then $\|Ay\|_{\mathbb{C}^n}^2 = \langle \operatorname{Re}(A^*A)y, y\rangle_{\mathbb{R}^n}$ for all $y \in \mathbb{R}^n$.*

*Proof.* Writing $A = A_1 + iA_2$ for $A_1, A_2 \in \mathbb{R}^{n\times n}$, we have $A^* = A_1^T - iA_2^T$. Thus

$$\begin{aligned}\|Ay\|_{\mathbb{C}^n}^2 = \langle A^*Ay, y\rangle_{\mathbb{C}^n} &= \langle (A_1^T - iA_2^T)(A_1 + iA_2)y, y\rangle_{\mathbb{C}^n} \\ &= \langle (A_1^T A_1 + A_2^T A_2)y, y\rangle_{\mathbb{R}^n} + i\langle (A_1^T A_2 - A_2^T A_1)y, y\rangle_{\mathbb{R}^n}.\end{aligned}$$

The last term is zero by the properties of the real inner product and transpose. □

Lemma 4.6. *Let $r_s = b_s - T_s T^{-1} e$ for $T$ and $e$ from Lemma 4.4. Then*

$$\phi^*(z) = \frac{1}{2}\|T^{-1/2}(z+e)\|_{\mathbb{R}^n}^2 - \frac{1}{2}\sum_{s=1}^t \|r_s\|_{\mathbb{C}^n}^2 - \|T^{-1/2}e\|_{\mathbb{R}^n}^2 \quad \forall z \in \mathbb{R}^n.$$

*Proof.* Lemma 4.5 and the properties of the complex inner product yield

$$\begin{aligned}\|T_s(T^{-1}(z+e)) - b_s\|_{\mathbb{C}^n}^2 &= \|T_s T^{-1} z - r_s\|_{\mathbb{C}^n}^2 \\ &= \|T_s T^{-1} z\|_{\mathbb{C}^n}^2 - \langle T_s T^{-1} z, r_s\rangle_{\mathbb{C}^n} - \langle r_s, T_s T^{-1} z\rangle_{\mathbb{C}^n} + \|r_s\|_{\mathbb{C}^n}^2 \\ &= \langle \operatorname{Re}(T_s^* T_s) T^{-1} z, T^{-1} z\rangle_{\mathbb{R}^n} - 2\operatorname{Re}\langle T^{-1} z, T_s^* r_s\rangle_{\mathbb{C}^n} + \|r_s\|_{\mathbb{C}^n}^2.\end{aligned} \tag{4.9}$$

We have $\sum_{s=1}^t T_s^* r_s = \sum_{s=1}^t T_s^* b_s - \sum_{s=1}^t T_s^* T_s T^{-1} e$, hence $\operatorname{Re}\sum_{s=1}^t T_s^* r_s = 0$. Since $T^{-1}z \in \mathbb{R}^n$, it follows that $\sum_{s=1}^t \operatorname{Re}\langle T^{-1}z, T_s^* r_s\rangle_{\mathbb{C}^n} = 0$. Dividing (4.9) by 2 and summing over $s = 1, \dots, t$, therefore

$$\phi(T^{-1}(z+e)) = \frac{1}{2}\langle z, T^{-1}z\rangle_{\mathbb{R}^n} + \frac{1}{2}\sum_{s=1}^t \|r_s\|_{\mathbb{C}^n}^2.$$

Using this expression in (4.7), the claim readily follows. □

We can now finally suggest one way to form the coarse function $F_H$:

Table 1: Time (seconds) to reach relative error $\rho_1$ and $\rho_2$ for both experiments.

| Experiment | $\rho_1$ | FB | FBMG | $\rho_2$ | FB | FBMG |
|---|---|---|---|---|---|---|
| Denoising | $10^{-2}$ | $3.05 \cdot 10^{-2}$ | $7.40 \cdot 10^{-3}$ | $10^{-3}$ | $2.32 \cdot 10^{-1}$ | $1.03 \cdot 10^{-1}$ |
| MRI | $10^{-2}$ | $6.93 \cdot 10^{-2}$ | $5.34 \cdot 10^{-3}$ | $10^{-3}$ | $1.42 \cdot 10^{-1}$ | $4.68 \cdot 10^{-2}$ |

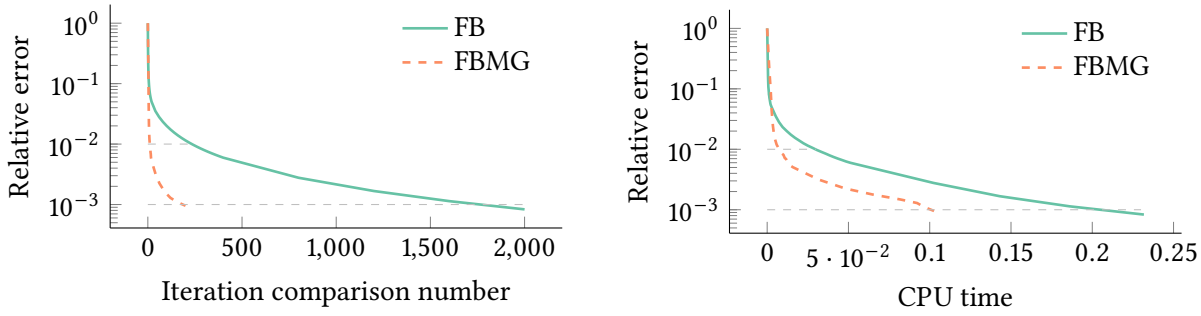



Figure 2: Relative error (5.1) versus iteration count and CPU time for denoising. The dashed grey lines indicate the levels $\rho = 10^{-2}$ and $\rho = 10^{-3}$ as in Table 1.

Example 4.7. Form a coarse discrete gradient operator $\nabla_H \in \mathbb{L}(\mathbb{R}^N; X_H)$ and set

$$F_H(\zeta) := \frac{1}{2}\|T_H^{-1/2}(\nabla_H^*\zeta - b_H)\|_{\mathbb{R}^N}^2 \quad \text{for} \quad T_H := I_h^H T \quad \text{and} \quad b_H := I_h^H b.$$

## 5 NUMERICAL EXPERIENCE

We now report our numerical experience with denoising and MRI. Both problems have the primal form (4.1). We work with the equivalent dual problem (4.2). Our Julia implementation is available on Zenodo [6].

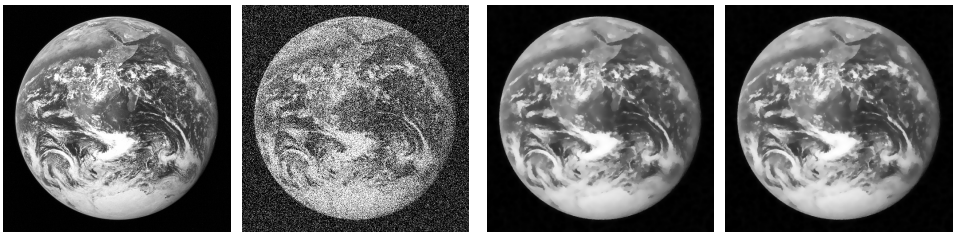
### 5.1 DENOISING

For denoising we use one full sample, i.e., solve (4.1) with $t = 1$ and $T_s = I$. The dual problem (4.8) is then $\min_{x \in \mathbb{R}^{D\times n}} \|\nabla_h^* x - e\|_{\mathbb{R}^n}^2 + \sum_i \delta_{B(0,\alpha)}(x_{\cdot,i})$. We use the *Blue Marble* public domain test image with resolution $3002 \times 3000$. We add pixelwise Gaussian noise of standard deviation $\sigma = 0.4$. The Lipschitz constant $L = L_H = 8$ [4]. We take $\alpha = 0.85$ and $\tau = 0.95/L$ and $\tau_H = 1.95/L$. In FBMG, we perform $m = 6$ coarse steps, based on trial and error, before the first 110 fine iterations only. For line search, we try $\theta_k = \bar{\theta}_k := \omega_k \langle T^{-1}(e - \nabla^* x^k), \nabla^* d\rangle / \|T^{-1/2}\nabla^* d\|_2^2 \geq 0$ for the scaling factor $\omega_k = 2/5$, and otherwise fail with $\theta_k = 0$.[2] We illustrate the data and reconstructions in Figure 3, and the performance in Figure 2 and Table 1, where, for $x^*$ computed by 100000 iterations of FBMGs, the *relative error*

$$\rho = \rho^k := (v(x^k) - v(x^*))/(v(x^0) - v(x^*)) \quad \text{with} \quad v(x) := \phi^*(-\nabla_h^* x) + G(x). \tag{5.1}$$

The *iteration comparison number* in Figure 2 scales coarse iterations by the ratio of the number of coarse to fine pixels.

[2] When $\omega_k < 2$, $\bar{\theta}_k$ is a scaled-down exact solution to (2.6) for $F = \phi^*$ and $G = 0$. Small $\omega_k$ attempts to ensure $x^k + \bar{\theta}_k d \in B(0,\alpha)^n$. By convexity, this check guarantees descent.



(a) Original (b) Noisy (c) FB (d) FBMG

Figure 3: Denoising data and results at relative error $\rho = 10^{-3}$.

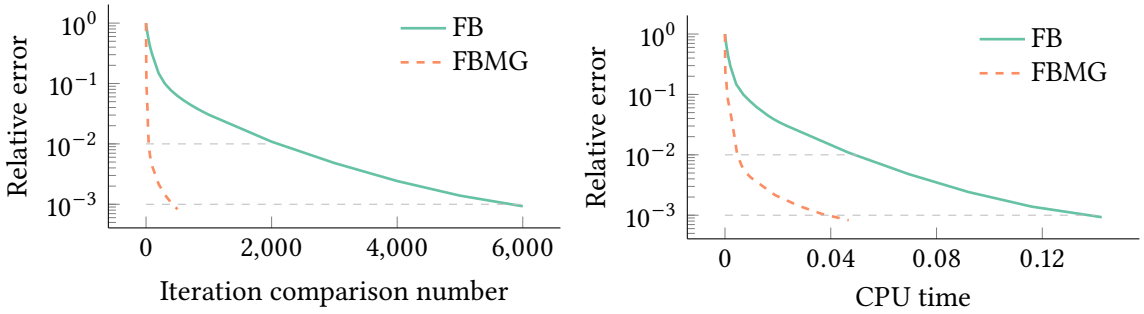


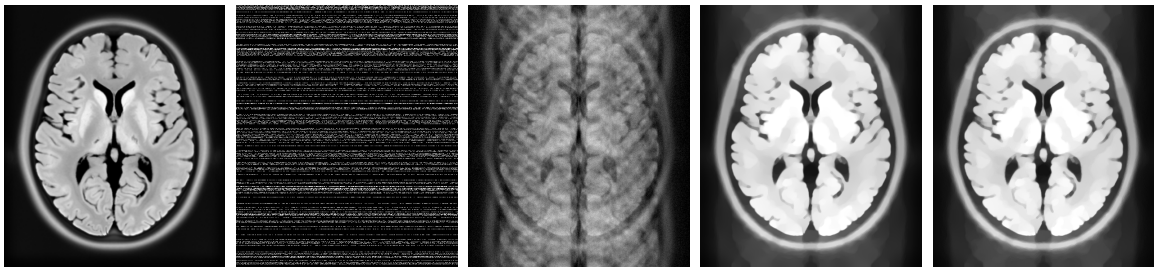

Figure 4: Relative error (5.1) versus both iteration count and CPU time for MRI. The dashed grey lines indicate the levels $\rho = 10^{-2}$ and $\rho = 10^{-3}$ as in Table 1.

### 5.2 MAGNETIC RESONANCE IMAGING

We take $T_s := S_s\mathcal{F}$, where $\mathcal{F}$ is the discrete Fourier transform and $S_s$ is a frequency subsampling operator. Then $T$ of Lemma 4.4 becomes $T = \operatorname{Re}\mathcal{F}^*S\mathcal{F} = \mathcal{F}^*\operatorname{Sym}S\mathcal{F}$ for $S := \sum_{s=1}^{t} S_s^*S_s$ and $\operatorname{Sym}S$ its symmetrisation over positive and negative frequencies in both axes. Thus $T^{-1}$, required for the dual problem (4.8), exists and is easily calculated when each frequency is sampled by some $S_s$. With $t = 21$, we form each subsampling mask $S_1, \dots, S_t$ by random sampling 150 lines in the Fourier space from a uniform distribution of such subsets of lines. We use the MRI phantom of [2] with resolution $583 \times 493$, and add complex Gaussian noise with standard deviation $\sigma = 50$. We take $\alpha = 1.15$, and the step length parameters and line search as for denoising with $L = 8\|T^{-1}\|$, $L_H = 8\|T_H^{-1}\|$. We perform $m = 6$ coarse steps before the first 500 fine iterations only. We illustrate the data, reconstructions, and performance in Figures 4 and 5 and Table 1.

### 5.3 CONCLUSIONS

Figures 2 and 4 and Table 1 indicate that while the performance improvements in denoising are noticeable, they are *very significant* for the much more expensive MRI problem. This can be expected, as the fine grid Fourier transform is an expensive operation. The situation is comparable to [13], who do deblurring directly with the primal problem. This requires proximal map of total variation to be solved numerically (as a denoising problem) on each fine-grid step, while in the coarse grid they avoid this by using a smooth problem and gradient steps. For multigrid optimisation methods to be meaningful, it therefore appears that the coarse-grid problems have to be significantly cheaper than the fine-grid problems.



(a) Original (b) Sample (c) Backproj. (d) FB (e) FBMG

Figure 5: MRI data and results at relative error $\rho = 10^{-2}$. (c) is the backprojection of the Fourier line sample (b). There are altogether $t = 21$ such samples.

## REFERENCES


[1] A. Ang, H. De Sterck, and S. Vavasis, MGProx: A nonsmooth multigrid proximal gradient method with adaptive restriction for strongly convex optimization, *SIAM J. Optim.* 34 (2024), 2788–2820.

[2] M. A. Belzunce, High-Resolution Heterogeneous Digital PET [18F]FDG Brain Phantom based on the BigBrain Atlas, 2018, doi:10.5281/zenodo.1190598.

[3] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A multigrid tutorial*, SIAM, 2000.

[4] A. Chambolle, An algorithm for total variation minimization and applications, *Journal of Mathematical imaging and vision* 20 (2004), 89–97.

[5] C. Clason and T. Valkonen, Introduction to Nonsmooth Analysis and Optimization (2020), arXiv:2001.00216.

[6] F. Guerra and T. Valkonen, Codes for "Multigrid methods for total variation", 2025, doi:10.5281/zenodo.14927401. Software.

[7] E. F. Guerra Urgiles, *Algoritmo forward-backward multimalla con aplicación al problema de supresión de ruido en una imagen*, Master's thesis, Escuela Politécnica Nacional, Quito, Ecuador, 2023, http://bibdigital.epn.edu.ec/handle/15000/23675.

[8] S. Karimi and S. Vavasis, IMRO: A proximal quasi-Newton method for solving $\ell_1$-regularized least squares problems, *SIAM J. Optim.* 27 (2017), 583–615.

[9] R. Kornhuber, Monotone multigrid methods for elliptic variational inequalities I, *Numerische Mathematik* 69 (1994), 167–184.

[10] G. Lauga, E. Riccietti, N. Pustelnik, and P. Gonçalves, IML FISTA: A multilevel framework for inexact and inertial forward-backward. application to image restoration, *SIAM J. Imaging Sci.* 17 (2024), 1347–1376.

[11] P. L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM J. Numer. Anal.* 16 (1979), 964–979.

[12] S. G. Nash, A multigrid approach to discretized optimization problems, *Optim. Methods. Software* 14 (2000), 99–116.

[13] P. Parpas, A multilevel proximal gradient algorithm for a class of composite optimization problems, *SIAM J. Optim.* 39 (2017), 681–701.

[14] C. R. Vogel, A multigrid method for total variation-based image denoising, in *Computation and Control IV: Proceedings of the Fourth Bozeman Conference, Bozeman, Montana, August 3–9, 1994*, 1995, 323–331.

[15] Z. Zhang, X. Li, Y. Duan, K. Yin, and X. C. Tai, An efficient multi-grid method for TV minimization problems, *Inverse Problems and Imaging* 15 (2021), 1199–1221.