PRIMAL–DUAL MULTIGRID METHODS FOR NONSMOOTH OPTIMIZATION

Felipe Guerra* Tuomo Valkonen†

Abstract In optimization, one often encounters problems of the form $\min_x F(x) + E(x) + G(Kx)$. In this work, we combine primal-dual algorithms with multigrid techniques for their solution. To link the the fine-grid and coarse-grid problems problems, we introduce a nonsmooth primal-dual coherence condition, and an efficient partially linearized line search procedure. Our work is motivated by total variation regularized inverse imaging problems, on which we demonstrate the efficacy of the method, being able to solve problems not previously possible with forward-backward multigrid methods.

1 INTRODUCTION

Several first-order methods have been developed to solve nonsmooth optimization problems to which basic forward-backward splitting is not easily applicable. These include the Primal-Dual Splitting Method (PDPS) [23] and the Alternating Direction Method of Multipliers (ADMM) [13, 25]. However, the computational cost of these methods can be high on large scale problems. A promising way to mitigate the cost is to use multigrid (or multilevel) techniques.

The main idea behind multigrid techniques is to reduce the dimension of the original problem – generally considered to be set in a finite-dimensional space – by passing to a different but related problem in a lower-dimensional space. This is commonly referred to as the *coarse problem*. By solving the coarse problem, a direction of improvement is obtained for the original *fine problem*.

Multigrid methods were popularized in [3] for the efficient solution of large-scale linear systems arising from the discretization of elliptic PDEs; see [4] for an introduction. Subsequently, it was extended to smooth optimization in [19]. Recently, various extensions of the approach have been proposed for nonsmooth optimization problems. These recent works either combine the forward-backward method with multigrid [1, 14], or smoothen the optimization problem [22], which implies that the optimization problem must have a prox-simple structure.

To bypass restrictions faced by forward-backward techniques, we propose a nonsmooth Primal-Dual Multigrid with Coarse Correction method (PDMCC). It links the PDPS applied to both the coarse and fine problems via a nonsmooth primal-dual coherence condition. This is Algorithm 1.1. We recall that the basic PDPS is sometimes called the Chambolle–Pock method [6], and the variant with an additional forward step the Condat–Vũ method [10, 28]. We review such algorithms in Section 2 through the preconditioned proximal point approach of [16], subsequently employed in [26, 9].

Algorithm 1.1 applies to problems of the form

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

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

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

Algorithm 1.1 Primal-Dual Multigrid with Coarse Correction (PDMCC)

Require: Functions F, E, G and step length parameters $\tau, \sigma > 0$ satisfying Assumption 2.1. Line search model parameters $\epsilon_k, \rho_k > 0$, $(k \in \mathbb{N})$. Line search parameter $\kappa \in (0,1)$. A trigger condition.

Choose an initial iterate $u^0 := (x^0, y^0) \in X \times Y$.
for all $k \in \mathbb{N}$ until a chosen stopping criterion is fulfilled **do**
  Perform the fine-grid PDPS update $u_+^k := (x_+^k, y_+^k)$ as
  $x_+^k := \operatorname{prox}_{\tau F}(x^k - \tau(\nabla E(x^k) + K^* y^k))$ and ▷ *Primal fine update*
  $y_+^k := \operatorname{prox}_{\sigma G^*}(y^k + \sigma K(2x^{k+1} - x^k))$. ▷ *Dual fine update*
  if the trigger condition is satisfied **then**
    Choose coarse functions $F_H^k, E_H, (G_H^k)^*$ subject to Assumptions 3.1 and 3.2.
    Form the descent direction d^k with Algorithm 4.1. ▷ *Coarse correction*
    Find a line search step length $\theta \geq 0$ satisfying (5.2) or (5.5)
    Perform the fine-grid update $u^{k+1} := u_+^k + \theta_k d^k$.
  else
    $u^{k+1} := u_+^k$.
  end if
end for

where F and G are convex, possibly nonsmooth functions, E is convex and smooth with L-Lipschitz gradient, and $K \in \mathbb{L}(X;Y)$ on Hilbert spaces X and Y. We treat in Section 3 the nonsmooth primal-dual coherence condition that lays out rules for designing F_H, E_H, G_H and K_H for a corresponding coarse problem. When a trigger condition – freely chosen by the user of the method – is satisfied, m coarse PDPS iterations are performed on these functions, starting from the restriction $v^{k,0} := I_h^H u_+^k$ of the fine primal-dual pre-iterate $u_+^k = (x_+^k, y_+^k)$, to yield a a direction $d^k = I_H^h(v^{k,m} - v^{k,0})$ from the prolonged initial coarse iterate to the prolonged final coarse iterate.

Integrating this direction into a primal-dual method presents significant challenges, because there is no obvious objective function for which to seek decrease: the Fenchel–Rockafellar duality gap can rarely be used in convergence proofs, while the more commonly employed Lagrangian duality gap depends on a base point, which is not the same on the two grids. We show in Section 4 that d^k is a descent direction for

$$\Phi^k(x,y) := F(x) + E(x) + G^*(y) + \langle y_+^k, Kx\rangle - \langle Kx_+^k, y\rangle, \tag{1.2}$$

when an ergodic or non-ergodic construction is used for the coarse algorithm. When the line search is amended with quadratic penalization, we obtain a bound on the Lagrangian gap of the fine problem, reminiscent of standard results for the PDPS. As a result, the PDMCC preserves the convergence structure of the PDPS, modified only by a controllable error term.

Our method uses line search to incorporate coarse information into the fine algorithm. In contrast to [18], where line search was first used with the PDPS – without multigrid – to adaptively adjust step length parameters without needing to know the norm of $K \in \mathbb{L}(X;Y)$, the line search we propose is more classical. In particular, while in [18] the primal and dual variables are updated using proximal steps, our method performs updates along a descent direction of Φ^k.

The convergence of primal-dual algorithms can be studied through two key concepts: Fejér monotonicity of the generated sequence with respect to the set of optimal solutions, and ergodic convergence of the Lagrangian gap. Consequently, any fine-grid update performed through line search must be carefully designed to ensure at least one of these properties—ideally, both. In Section 5 we show that we can incorporate terms relevant for Fejér monotonicity into a line search procedure on Φ^k. This then allows proving for the PDMCC the ergodic convergence of the Lagrangian gap at the rate $O(1/N)$, as

well as Fejér quasi-monotonicity. The latter establishes weak convergence via Opial's lemma. We also illustrate, in Remark 6.9, how our results can be extended to nonconvex E.

In Section 7 we discuss ways to construct the coarse functions E_H, G_H, and F_H. Finally, in Section 8, we provide a series of *numerical experiments* on total variation–regularized inverse imaging problems to validate the proposed method and highlight its computational advantages.

2 PRIMAL-DUAL PROXIMAL SPLITTING

We recall here the preconditioned proximal point approach [16, 26, 9] to the PDPS for (1.1). We follow [9, Chapter 11]. To start, we recall basic notation. We write $\overline{\mathbb{R}} = [-\infty, +\infty]$. For a convex function $F : X \to \overline{\mathbb{R}}$, $\partial F(x)$ denotes its subdifferential at x; $F^* : X \to \overline{\mathbb{R}}$ the Fenchel conjugate; and prox_F the proximal operator. If $K : X \to Y$ is linear and bounded, we write $K \in \mathbb{L}(X;Y)$.

By the Fenchel–Rockafellar theorem, primal-dual solution pairs $\hat{u} = (\hat{x}, \hat{y}) \in U := X \times Y$ on the Hilbert spaces X and Y, are characterized by

$$-K^*\hat{y} - \nabla E(\hat{x}) \in \partial F(\hat{x}) \quad \text{and} \quad K\hat{x} \in \partial G^*(\hat{y}). \tag{2.1}$$

Based on this, the PDPS method with a forward step with respect to E reads

$$\begin{cases} x^{k+1} := \operatorname{prox}_{\tau F}(x^k - \tau[\nabla E(x^k) + K^* y^k]), \\ y^{k+1} := \operatorname{prox}_{\sigma G^*}(y^k + \sigma K(2x^{k+1} - x^k)). \end{cases} \tag{2.2}$$

In implicit form,

$$\begin{aligned} &0 \in \partial F(x^{k+1}) + K^* y^{k+1} + \nabla E(x^k) - K^*(y^{k+1} - y^k) + \tau^{-1}(x^{k+1} - x^k) \quad \text{and} \\ &0 \in \partial G^*(y^{k+1}) - Kx^{k+1} - K(x^{k+1} - x^k) + \sigma^{-1}(y^{k+1} - y^k). \end{aligned}$$

These inclusions take a convenient form in the product space $U = X \times Y$. Let

$$\bar{G} : U \to \overline{\mathbb{R}}, \quad \bar{G}(u) := F(x) + G^*(y) \quad \text{and} \quad \bar{E}(u) : U \to \overline{\mathbb{R}}, \quad \bar{E}(u) := E(x). \tag{2.3}$$

We also define $M, \Xi, \Lambda \in \mathbb{L}(U;U)$ by

$$M := \begin{pmatrix} \tau^{-1}\mathrm{Id} & -K^* \\ -K & \sigma^{-1}\mathrm{Id} \end{pmatrix}, \quad \Xi := \begin{pmatrix} 0 & K^* \\ -K & 0 \end{pmatrix}, \quad \text{and} \quad \Lambda := \begin{pmatrix} L\,\mathrm{Id} & 0 \\ 0 & 0 \end{pmatrix}, \tag{2.4}$$

where $L \geq 0$ is the Lipschitz constant of the gradient of E, and Id denotes the identity operator. Then, the PDPS is characterized by the joint implicit inclusion

$$0 \in \partial \bar{G}(u^{k+1}) + \Xi u^{k+1} + \nabla \bar{E}(u^k) + M(u^{k+1} - u^k). \tag{2.5}$$

In other words, it is a preconditioned forward-backward method with the skew-adjoint perturbation Ξ, which does not arise as a differential.

To state basic results for the PDPS, we require:

Assumption 2.1. $F, E : X \to \overline{\mathbb{R}}$ and $G : Y \to \overline{\mathbb{R}}$ are proper, convex and lower semicontinous with L-Lipschitz ∇E, and $K \in \mathbb{L}(X;Y)$ on Hilbert spaces X and Y. The steps length parameters $\tau, \sigma > 0$ are chosen such that $\tau L + \tau\sigma\|K\|^2_{\mathbb{L}(X;Y)} < 1$.

Assumption 2.1 implies [9, Chapter 7] that $\bar{E}$, defined in (2.3) satisfies for any $u, \hat{u}, \tilde{u} \in U$ and $\Lambda \in \mathbb{L}(U;U)$ from (2.4) the *three-point smoothness inequality*

$$\langle \nabla \bar{E}(\hat{u}), u - \tilde{u} \rangle \geq \bar{E}(u) - \bar{E}(\tilde{u}) - \frac{1}{2}\|u - \hat{u}\|^2_\Lambda. \tag{2.6}$$

Moreover, we have:

Lemma 2.2 ([9, Lemma 9.12]). *Let Assumption 2.1 hold. Then $M-\Lambda$ and M are bounded and self-adjoint; $M-\Lambda$ is positive definite; and $\|u\|_M^2 \geq C\|u\|^2$ for all $u \in X\times Y$, where $C := (1-\sqrt{\tau\sigma}\|K\|_{\mathbb{L}(X;Y)})\min\{\tau^{-1},\sigma^{-1}\} > 0$.*

As a positive (semi-)definite and self-adjoint operator, M defines the (semi)norm $\|u\|_M := \sqrt{\langle Mu, u\rangle}$ that satisfies for all $u, \hat{u}, \tilde{u} \in U$ the *three-point identity*

$$\langle M(u-\hat{u}), u-\tilde{u}\rangle = \frac{1}{2}\|u-\hat{u}\|_M^2 - \frac{1}{2}\|\hat{u}-\tilde{u}\|_M^2 + \frac{1}{2}\|u-\tilde{u}\|_M^2. \tag{2.7}$$

To analyze convergence, we introduce the Lagrangian gap

$$\mathscr{G}_L(u;\hat{u}) := \Phi(u;\hat{u}) - \Phi(\hat{u};\hat{u}) \quad \text{for} \quad u, \hat{u} \in U, \tag{2.8}$$

where[1]

$$\Phi : U \times U \to \overline{\mathbb{R}}, \qquad \Phi(u;\hat{u}) := \bar{G}(u) + \bar{E}(u) + \langle \Xi\hat{u}, u\rangle. \tag{2.9}$$

If F, G, and E are convex, and $\hat{u}$ solves the primal-dual optimality conditions (2.1), then the Lagrangian gap is non-negative. It is zero if $u = \hat{u}$. We have $\mathscr{G}_L(u;u_+^k) = \Phi^k(u) - \Phi(u_+^k;u_+^k)$ for Φ^k of (1.2), which suggests why d^k being a descent direction of Φ^k can be useful. The basic PDPS admits the following fundamental bound:

Theorem 2.3 ([9, Theorem 11.7]). *Let Assumption 2.1 hold. Then $\{u^k = (x^k, y^k)\}_{k\in\mathbb{N}}$ generated by the PDPS (2.2) satisfies for all $\hat{u} \in U$ and $k \geq 0$ the bound*

$$\mathscr{G}_L(u^{k+1};\hat{u}) + \frac{1}{2}\|u^{k+1}-\hat{u}\|_M^2 + \frac{1}{2}\|u^{k+1}-u^k\|_{M-\Lambda}^2 \leq \frac{1}{2}\|u^k-\hat{u}\|_M^2.$$

As $M \geq \Lambda$ by Assumption 2.1 and Lemma 2.2, summing this bound over $k = 0, \dots, N-1$, and letting $N \nearrow \infty$, we get ergodic $O(1/N)$ convergence of the gap [9].

3 COARSE PROBLEM

We now construct our overall approach to coarse approximations of (1.1). The fundamental idea of our approach traces back to the smooth *coherence condition* [19], explicitly expressed as such in [22]. A nonsmooth variant was introduced for forward-backward type methods in [14]. We recall the concept in 3.2, and then extend it to the primal-dual setting in Section 3.3 after first defining basic notation and multigrid transfer operators in Section 3.1.

3.1 BASIC DEFINITIONS; MULTIGRID TRANSFER OPERATORS

We write X_H and Y_H for the coarse spaces corresponding to the primal space X and the dual space Y. All the spaces are assumed to be Hilbert spaces. Usually they are finite-dimensional, and satisfy $\dim X_H < \dim X$ and $\dim Y_H < \dim Y$. We write

$$\begin{cases} u = (x, y) \in U := X \times Y \text{ for the fine (primal, dual) variables; and} \\ v = (\zeta, \xi) \in U_H := X_H \times Y_H \text{ for the coarse (primal, dual) variables.} \end{cases}$$

Transfer operators are the fundamental tool of multigrid methods: restriction from a fine grid to a coarse grid, and interpolation (or prolongation) from the coarse grid to the fine grid [4]. Typically one is the adjoint of the other. Since the PDMCC involves primal and dual variables, it is necessary to

[1] Φ is not the Lagrangian $\mathscr{L}(x, y) = [F + E](x) + \langle Kx, y\rangle - G^*(y)$ that has $\mathscr{G}_L(u;\hat{u}) = \mathscr{L}(x, \hat{y}) - \mathscr{L}(\hat{x}, y)$.

distinguish transfer operators associated with each of these spaces: $P_h^H \in \mathbb{L}(X; X_H)$ and $D_h^H \in \mathbb{L}(Y; Y_H)$ are the primal and dual restriction operators, respectively. We define the combined restriction operator $I_h^H \in \mathbb{L}(U; U_H)$ by $I_h^H(x, y) := (P_h^H x, D_h^H y)$. Then the primal and dual prolongation operators are $P_H^h := (P_h^H)^* \in \mathbb{L}(X_H; X)$ and $D_H^h := (D_h^H)^* \in \mathbb{L}(Y_H; H)$. The combined prolongation operator $I_H^h \in \mathbb{L}(U_H; H)$ is $I_H^h := (I_h^H)^*$, i.e., $I_H^h(\zeta, \xi) = (P_H^h \zeta, D_H^h \xi)$.

3.2 MOTIVATION FOR THE COHERENCE CONDITION

Consider the simple smooth problem $\min_x E(x)$. Suppose that on iteration k, at the point x^k, we want to pass to the coarse grid. To do so, we have to construct a coarse objective E_H^k. In [19, 22], this construction has to satisfy the *coherence condition*

$$P_h^H \nabla E(x^k) = \nabla E_H^k(\zeta^{k,0}) \quad \text{where} \quad \zeta^{k,0} := P_h^H x^k. \tag{3.1}$$

Then, any descent direction d_H for E_H^k at $\zeta^{k,0}$, i.e., $\langle E_h^k(\zeta^{k,0}), d_H\rangle < 0$, can easily be translated into a descent direction for E. Indeed,

$$\langle \nabla E(x^k), P_H^h d_H\rangle = \langle P_h^H \nabla E(x^k), d_H\rangle = \langle E_h^k(\zeta^{k,0}), d_H\rangle < 0.$$

One way to construct E_H^k satisfying the smooth coherence condition (3.1), is to take *any* differentiable $E_H : X_H \to \mathbb{R}$, and set

$$E_H^k(\zeta) := E_H(\zeta) + \langle r_H^k, \zeta\rangle \quad \text{for} \quad r_H^k := P_h^H \nabla E(x^k) - \nabla E_H(P_h^H x^k). \tag{3.2}$$

Intuitively, E_H should be a "coarse version" of E, whereas E_H^k corrects it for the restriction error.

Consider then the problem $\min_x J(x) := F(x) + E(x)$, where E is differentiable but F is not. We can then extend (3.1) into the *nonsmooth coherence condition*

$$P_h^H \partial J(x^k) \subset \partial J_H^k(\zeta^{k,0}), \tag{3.3}$$

This can be achieved as $J_H^k = E_H^k + F_H^k$, where E_H^k is as in (3.2), and F_H^k satisfies the nonsmooth coherence condition with respect to F. The function F_H^k can be constructed as an indicator function of a cone [14]. Again, a descent direction for J_H^k, will be a descent direction for J [14].

3.3 NONSMOOTH PRIMAL–DUAL COHERENCE CONDITION

A challenge with extending the coherence condition to primal-dual methods is that they are not necessarily monotone with respect to the Lagrangian, or the Fenchel–Rockafellar gap: they may not yield descent directions. It is, also, difficult to transfers of coarse gap to the fine gap, due to the bilinear terms in the gap, in other words, the skew-symmetric term Ξ in the generalized forward-backward formulation (2.5).

Nevertheless, motivated by the descent estimate of Theorem 2.3, we construct the coarse problem such that we decrease the function (recall (1.2) and (2.9))

$$\Phi^k(u) := \Phi(u; u_+^k) = \bar{G}(u) + \bar{E}(u) + \langle \Xi u_+^k, u\rangle. \tag{3.4}$$

In the first "ergodic" variant of our coarse method, detailed in Section 4, we pick a primal-dual *tilt vector* $w_H^k = (r_H^k, o_H^k) \in U_H$, associated with the smooth part $\bar{E} + \langle \Xi u_+^k, \cdot\rangle$ of Φ^k. We then consider the coarse problem

$$\min_{\zeta \in X_H} \max_{\xi \in Y_H} F_H^k(\zeta) + E_H(\zeta) + \langle \xi, K_H \zeta\rangle_{Y_H} - (G_H^k)^*(\xi) + \langle r_H^k, \zeta\rangle_{X_H} - \langle o_H^k, \xi\rangle_{Y_H}. \tag{3.5}$$

In the second "non-ergodic" variant of the method, the tilt vectors depend on the coarse iteration index j, i.e., $w_H^{k,j} = (r_H^{k,j}, o_H^{k,j})$. In this case, the interpretation of minimizing (3.5) cannot be directly given, although the algorithm will be analogous.

The coarse functions F_H^k and $(G_H^k)^*$ will need to satisfy the following assumptions.

Assumption 3.1 (Basic coarse structure). $E_H, F_H^k : X_H \to \mathbb{R}$ and $G_H^k : Y_H \to \overline{\mathbb{R}}$ are proper, convex, and lower semicontinuous with L_H-Lipschitz ∇E_H, and $K_H \in \mathbb{L}(X_H; Y_H)$ on Hilbert spaces X_H and Y_H. The steps length parameters $\tau_H, \sigma_H > 0$ are chosen such that $\tau_H L_H + \tau_H \sigma_H \|K_H\|^2_{\mathbb{L}(X_H;Y_H)} < 1$.

As in Section 2, we introduce the extended coarse functions $\bar{G}_H^k : U_H \to \overline{\mathbb{R}}$ and $\bar{E}_H(v) : U_H \to \overline{\mathbb{R}}$ on the product space U_H, defined by

$$\bar{G}_H^k(v) := F_H^k(\zeta) + (G_H^k)^*(\xi) \quad \text{and} \quad \bar{E}_H(v) := E_H(\zeta). \tag{3.6}$$

The operators $M_H, \Xi_H, \Lambda_H \in \mathbb{L}(U_H; U_H)$ are defined by

$$M_H := \begin{pmatrix} \tau_H^{-1}\,\mathrm{Id} & -K_H^* \\ -K_H & \sigma_H^{-1}\,\mathrm{Id} \end{pmatrix}, \quad \Xi_H := \begin{pmatrix} 0 & K_H^* \\ -K_H & 0 \end{pmatrix} \quad \text{and} \quad \Lambda_H := \begin{pmatrix} L_H\,\mathrm{Id} & 0 \\ 0 & 0 \end{pmatrix}, \tag{3.7}$$

where $L_H \geq 0$ is the Lipschitz constant of the gradient of E_H.

We can now state our primal-dual coherence condition as a variant of (3.3):

Assumption 3.2 (Nonsmooth Primal-Dual Coherence Condition). For a given $k \in \mathbb{N}$, the fine-grid pre-iterate $u_+^k \in U$, and the corresponding initial coarse iterate $v^{k,0} \in U_H$ (typically $I_h^H u_+^k$) satisfy

$$I_h^H \partial \bar{G}(u_+^k) \subset \partial \bar{G}_H^k(v^{k,0}).$$

This condition can be decomposed into its *primal* and *dual* components

$$P_h^H \partial F(x_+^k) \subset \partial F_H^k(\zeta^{k,0}) \quad \text{and} \quad D_h^H \partial G^*(y_+^k) \subset \partial (G_H^k)^*(\xi^{k,0}).$$

4 COARSE ALGORITHM

In Sections 4.1 and 4.2, we present two coarse-grid primal-dual algorithms. The first variant produces an ergodic descent direction, i.e., one that takes the average over the iterations. The second avoids this averaging, and potentially expensive operator evaluations. In Section 4.3, we derive coarse-grid descent estimates for both variants. Then, in Section 4.4, we transfer these estimates to the fine grid.

4.1 ERGODIC VARIANT

In the ergodic variant of the coarse-grid method, we take the $r_H^k \in X_H$ and $o_H^k \in Y_H$ of the coarse problem (3.5) as

$$r_H^k := P_h^H\big(\nabla E(x_+^k) + K^* y_+^k\big) - \big(\nabla E_H(\zeta^{k,0}) + K_H^* \xi^{k,0}\big) \quad \text{and} \quad o_H^k := -D_h^H K x_+^k + K_H \zeta^{k,0},$$

that is,

$$w_H^k := (r_H^k, o_H^k) = I_h^H(\nabla \bar{E}(u_+^k) + \Xi u_+^k) - (\nabla \bar{E}_H(v^{k,0}) + \Xi_H v^{k,0}).$$

On each outer iteration $k \in \mathbb{N}$, denoting the coarse iteration number by j, and the coarse iterates by $v^{k,j} = (\zeta^{k,j}, \xi^{k,j})$, the PDPS (2.2) for the general coarse problem (3.5) expands as Lines 4 and 5 of Algorithm 4.1 with $\vartheta = 1$. This is valid for any w_H^k. Using the definitions (3.6) and (3.7), we write the method in the implicit form

$$0 \in \partial \bar{G}_H^k(v^{k,j+1}) + \nabla \bar{E}_H(v^{k,j}) + \Xi_H v^{k,j+1} + w_H^k + M_H(v^{k,j+1} - v^{k,j}). \tag{4.1}$$

Algorithm 4.1 Coarse primal-dual method

Require: Coarse functions F_H^k, G_H^k, and E_H satisfying Assumptions 3.1 and 3.2. Step length parameters $\tau_H, \sigma_H > 0$. Iteration count $m \in \mathbb{N}^+$. Transfer operators P_h^H and D_h^H. Choice of $\vartheta \in \{0, 1\}$ (non-ergodic vs. ergodic variant).

1: $v^{k,0} := (P_h^H x_+^k, D_h^H y_+^k) \in U_H$.
2: $a_H^k := P_h^H\big(\nabla E(x_+^k) + K^* y_+^k\big) - \nabla E_H(\zeta^{k,0})$ and $b_H^k := -D_h^H K x_+^k$.
3: **for all** $j = 0, 1, \dots, m-1$ **do**
4: $\quad \zeta^{k,j+1} := \operatorname{prox}_{\tau_H F_H^k}(\zeta^{k,j} - \tau_H\big[\nabla E_H(\zeta^{k,j}) + \vartheta K_H^*(\xi^{k,j} - \xi^{k,0}) + a_H^k\big])$ ▷ *Primal update*
5: $\quad \xi^{k,j+1} := \operatorname{prox}_{\sigma_H (G_H^k)^*}(\xi^{k,j} + \sigma_H\big[2K_H(\zeta^{k,j+1} - \zeta^{k,j}) + \vartheta K_H(\zeta^{k,j} - \zeta^{k,0}) - b_H^k\big])$ ▷ *Dual update*
6: **end for**
7: **return** $d = \tilde{v}^{k,j+1} - v^{k,0}$, where $\tilde{v}^{k,m}$ is defined by (4.6) and $v^{k,j} = (\zeta^{k,j}, \xi^{k,j})$. ▷ *Descent direction*

4.2 NON-ERGODIC VARIANT

The solution of (4.1), i.e., Algorithm 4.1 with $\vartheta = 1$, requires the evaluation of $K_H \in \mathbb{L}(X_H; Y_H)$ as well as its adjoint on each iteration j. This can be computationally expensive and even lead to numerical instabilities. To ameliorate these issues, we now make the tilt vectors dependent on the coarse iteration j, using in place of r_H^k and o_H^k the vectors

$$r_H^{k,j} := P_h^H\big(\nabla E(x_+^k) + K^* y_+^k\big) - \big(\nabla E_H(\zeta^{k,0}) + K_H^* \xi^{k,j}\big) \quad \text{and} \quad o_H^{k,j} := -D_h^H K x_+^k + K_H \zeta^{k,j},$$

that is, in place of w_H^k, the vector

$$w_H^{k,j} := I_h^H(\nabla \bar{E}(u_+^k) + \Xi u_+^k) - (\nabla \bar{E}_H(v^{k,0}) + \Xi_H v^{k,j}). \tag{4.2}$$

We correspondingly modify the implicit algorithm (4.1) into

$$0 \in \partial \bar{G}_H^k(v^{k,j+1}) + \nabla \bar{E}_H(v^{k,j}) + \underbrace{\Xi_H v^{k,j+1} + w_H^{k,j} + M_H(v^{k,j+1} - v^{k,j})}_{(\Xi_H + M_H)(v^{k,j+1} - v^{k,j}) + I_h^H(\nabla \bar{E}(u_+^k) + \Xi u_+^k) - \nabla \bar{E}(v^{k,0})}. \tag{4.3}$$

Using the rearrangement under the brace, which follows (4.2) and the definition of the operators M_H and Ξ_H in (3.7), in explicit form, we see that (4.3) reads as Lines 4 and 5 of Algorithm 4.1 with $\vartheta = 0$.

4.3 DESCENT

We can write both (4.1) and (4.3) as

$$0 \in \partial \bar{G}_H^k(v^{k,j+1}) + \Xi_H(v^{k,j+1} - v_\vartheta^j) + \nabla \bar{E}(v^{k,j}) + s_H^k + M_H(v^{k,j+1} - v^{k,j}), \tag{4.4a}$$

where $v_\vartheta^j := (1 - \vartheta) v^{k,j} + \vartheta v^{k,0}$, $\vartheta \in \{0, 1\}$ indicates the variant, and

$$s_H^k = (a_H^k, b_H^k) := I_h^H(\nabla \bar{E}(u_+^k) + \Xi u_+^k) - \nabla \bar{E}_H(v^{k,0}) \in U_H. \tag{4.4b}$$

With this, we obtain a tilted descent estimate in the coarse grid:

Theorem 4.1. *Let Assumption 3.1 hold. Then, for any initial coarse point $v^{k,0} \in U_H$, and $v^{k,1}, \dots, v^{k,m}$ generated through (4.4), we have*

$$[\bar{G}_H^k + \bar{E}_H](\tilde{v}^{k,m}) + \langle s_H^k, \tilde{v}^{k,m} - v^{k,0} \rangle + Q(v^{k,0}, \dots, v^{k,m}, m) \leq [\bar{G}_H^k + \bar{E}_H](v^{k,0}). \tag{4.5}$$

where

$$Q(v^{k,0},\dots,v^{k,m},m) := \frac{1}{2m}\left[\|v^{k,m}-v^{k,0}\|_{M_H}^2 + \sum_{j=0}^{m-1}\|v^{k,j+1}-v^{k,j}\|_{M_H-\Lambda_H}^2\right] \geq 0$$

and

$$\tilde{v}^{k,m} := \begin{cases} \frac{1}{m}\sum_{j=0}^{m-1} v^{k,j+1}, & \text{for the ergodic variant (4.1),} \\ v^{k,m}, & \text{for the non-ergodic variant (4.3).} \end{cases} \tag{4.6}$$

Proof. We write $v^j := v^{k,j}$ and $\bar{J} := \bar{G}_H^k + \bar{E}_H$ for brevity. By the convexity of $\bar{G}_H^k$ and the three-point smoothness inequality on $\bar{E}$, we know that

$$\langle \partial\bar{G}_H^k(v^{j+1}) + \nabla\bar{E}_H(v^j), v^{j+1}-v_\vartheta^j\rangle \geq \bar{J}(v^{j+1}) - \bar{J}(v_\vartheta^j) - \frac{1}{2}\|v^{j+1}-v^j\|_{\Lambda_H}^2.$$

Applying $\langle\,\cdot\,, v^{j+1}-v_\vartheta^j\rangle$ on both sides of (4.4a) and using this inequality, yields

$$\langle \Xi_H(v^{j+1}-v_\vartheta^j), v^{j+1}-v_\vartheta^j\rangle + \langle s_H^k, v^{j+1}-v_\vartheta^j\rangle + \langle M_H(v^{j+1}-v^j), v^{j+1}-v_\vartheta^j\rangle \\ \leq \bar{J}(v_\vartheta^j) - \bar{J}(v^{j+1}) + \frac{1}{2}\|v^{j+1}-v^j\|_{\Lambda_H}^2.$$

Using the skew-adjointness of Ξ_H and the *three-point identity* (2.7), this becomes

$$\bar{J}(v^{j+1}) + \frac{1}{2}\|v^{j+1}-v_\vartheta^j\|_{M_H}^2 + \langle s_H^k, v^{j+1}-v_\vartheta^j\rangle + \frac{1}{2}\|v^{j+1}-v^j\|_{M_H-\Lambda_H}^2 \leq \bar{J}(v_\vartheta^j) + \frac{1}{2}\|v^j-v_\vartheta^j\|_{M_H}^2. \tag{4.7}$$

Ergodic case: When $\vartheta = 1$, we have $v_\vartheta^j = v^0$. Summing (4.7) over $j = 0,\dots,m-1$, thus, yields

$$\sum_{j=0}^{m-1}\left[\bar{J}(v^{j+1}) + \langle s_H^k, v^{j+1}-v^0\rangle + \frac{1}{2}\|v^{j+1}-v^j\|_{M_H-\Lambda_H}^2\right] + \frac{1}{2}\|v^m-v^0\|_{M_H}^2 \leq m\bar{J}(v^0).$$

According to Assumption 3.1, M_H is positive definite, so $\|\cdot\|_{M_H}^2$ is a convex function. Thus, applying Jensen's inequality to $\bar{J}$, we obtain (4.5) for $\tilde{v}^{k,m} = \frac{1}{m}\sum_{j=0}^{m-1} v^{k,j+1}$.

Non-ergodic case: When $\vartheta = 0$, we have $v_\vartheta^j = v^j$. Summing (4.7) over $j = 0,\dots,m-1$ now yields (4.5) in the form

$$\bar{J}(v^m) + \langle s_H^k, v^m - v^0\rangle + \frac{1}{2}\sum_{j=0}^{m-1}\|v^{j+1}-v^j\|_{M_H}^2 + \frac{1}{2}\sum_{j=0}^{m-1}\|v^{j+1}-v^j\|_{M_H-\Lambda_H}^2 \leq \bar{J}(v^0).$$

□

To prove descent instead of mere non-increase, we use the next lemma.

Lemma 4.2. *Let Assumption 3.1 hold, $m \geq 1$, and define*

$$c_Q := \frac{2(1-\sqrt{\tau_H\sigma_H(1-\tau_H L_H)^{-1}}\|K_H\|)\min\{(1-\tau_H L_H)\tau_H^{-1}, \sigma_H^{-1}\}}{m(m+1)^2},$$

If $v^{k,0} \in U_H$ does not solve (3.5), then, for $\tilde{v}^{k,m}$ defined by (4.6), we have

$$Q(v^{k,0},\dots,v^{k,m},m) \geq c_Q\|\tilde{v}^{k,m}-v^{k,0}\|^2.$$

Proof. **Non-ergodic case:** By Assumption 3.1 and Lemma 2.2, $M_H - \Lambda_H$ is self-adjoint and positive definite; consequently, $\|\cdot\|^2_{M_H-\Lambda_H}$ is convex and non-negative. We have $\tilde{v}^{k,m} = v^{k,m}$ and $\frac{1}{2m} \geq \frac{2}{m(m+1)^2}$, hence $(m+1)^2 \geq 4$ for all $m \geq 1$. Therefore,

$$\begin{aligned} Q(v^{k,0},\dots,v^{k,m},m) &= \frac{1}{2m}\left[\|v^{k,m}-v^{k,0}\|^2_{M_H} + \sum_{j=0}^{m-1}\|v^{k,j+1}-v^{k,j}\|^2_{M_H-\Lambda_H}\right] \\ &\geq \frac{1}{2m}\|v^{k,m}-v^{k,0}\|^2_{M_H} \\ &\geq \frac{(1-\sqrt{\tau_H\sigma_H}\|K_H\|)\min\{\tau_H^{-1},\sigma_H^{-1}\}}{2m}\|\tilde{v}^{k,m}-v^{k,0}\|^2 \\ &\geq c_Q\|\tilde{v}^{k,m}-v^{k,0}\|^2. \end{aligned}$$

Ergodic case: Recall that, now, $\tilde{v}^{k,m} = \frac{1}{m}\sum_{j=0}^{m-1} v^{k,j+1}$. We first prove that

$$\tilde{v}^{k,m} - v^{k,0} = \sum_{j=0}^{m-1}\frac{m-j}{m}[v^{k,j+1}-v^{k,j}] \quad \text{for all} \quad m \geq 1.$$

We use induction. The base $m = 1$ is clear. Assuming the expression for $m = n$, to prove it for $m = n+1$, we use the telescoping sum $v^{k,n+1} - v^{k,0} = \sum_{j=0}^{n}[v^{k,j+1}-v^{k,j}]$ to rearrange

$$\tilde{v}^{k,n+1} - v^{k,0} = \frac{1}{n+1}\sum_{j=0}^{n-1}[v^{k,j+1}-v^{k,0}] + \frac{1}{n+1}[v^{k,n+1}-v^{k,0}] = \sum_{j=0}^{n}\frac{n+1-j}{n+1}[v^{k,j+1}-v^{k,j}].$$

This completes the induction.

Defining $\bar{\lambda} = \sum_{j=0}^{m-1}(m-j)$ and $\lambda_j = (m-j)/\bar{\lambda}$, we now obtain

$$\|\tilde{v}^{k,m}-v^{k,0}\|^2_{M_H-\Lambda_H} = \frac{\bar{\lambda}^2}{m^2}\left\|\sum_{j=0}^{m-1}\lambda_j[v^{k,j+1}-v^{k,j}]\right\|^2_{M_H-\Lambda_H}.$$

Now, since $\sum_{j=0}^{m-1}\lambda_j = 1$ with $\lambda_j \in (0,1]$ for all j, Jensen's inequality gives

$$\|\tilde{v}^{k,m}-v^{k,0}\|^2_{M_H-\Lambda_H} \leq \frac{\bar{\lambda}^2}{m^2}\sum_{j=0}^{m-1}\lambda_j\|v^{k,j+1}-v^{k,j}\|^2_{M_H-\Lambda_H} \leq \frac{\bar{\lambda}^2}{m^2}\sum_{j=0}^{m-1}\|v^{k,j+1}-v^{k,j}\|^2_{M_H-\Lambda_H}.$$

Finally, by the last inequality, the definition of $Q(v^{k,0},\dots,v^{k,m},m)$, and Assumption 3.1, which ensures that $M \geq 0$, we obtain

$$\begin{aligned} Q(v^{k,0},\dots,v^{k,m},m) &= \frac{1}{2m}\left[\|v^{k,m}-v^{k,0}\|^2_{M_H} + \sum_{j=0}^{m-1}\|v^{k,j+1}-v^{k,j}\|^2_{M_H-\Lambda_H}\right] \\ &\geq c_Q\|\tilde{v}^{k,m}-v^{k,0}\|^2. \end{aligned}$$

□

4.4 DESCENT IN THE FINE GRID

We now transfer the descent estimate of Theorem 4.1 to the fine grid.

Corollary 4.3. *Let Assumptions 3.1 and 3.2 hold. Then for any initial coarse iterate $v^{k,0} \in U_H$, which does not solve the coarse problem (3.5), the descent direction $d^k := I_H^h(\tilde{v}^{k,m} - v^{k,0}) \in U$ generated through (4.4), satisfies*

$$\|d^k\| \leq \tilde{c}_Q \inf_{g^k \in \partial\bar{G}(u_+^k)}\|g^k + \nabla\bar{E}(u_+^k) + \Xi u_+^k\| \quad \text{and} \quad [\Phi^k]'(u_+^k;d^k) < 0,$$

where $\tilde{v}^{k,m}$ is defined by (4.6) and $\tilde{c}_Q := \|I_H^h\|/c_Q$.

Proof. Let $s_H^k \in U_H$ be given by (4.4b). By its definition and the convexity of $\bar{E}_H$,

$$\begin{aligned}\langle s_H^k, \tilde{v}^{k,m} - v^{k,0}\rangle &= \langle I_h^H(\nabla \bar{E}(u_+^k) + \Xi u_+^k), \tilde{v}^{k,m} - v^{k,0}\rangle + \langle \nabla \bar{E}_H(v^{k,0}), v^{k,0} - \tilde{v}^{k,m}\rangle \\ &\geq \langle I_h^H(\nabla \bar{E}(u_+^k) + \Xi u_+^k), \tilde{v}^{k,m} - v^{k,0}\rangle + \bar{E}_H(v^{k,0}) - \bar{E}_H(\tilde{v}^{k,m}).\end{aligned} \tag{4.8}$$

We now recall the coarse descent estimate (4.5) obtained in Theorem 4.1,

$$\bar{G}_H^k(\tilde{v}^{k,m}) + \bar{E}_H(\tilde{v}^{k,m}) + \langle s_H^k, \tilde{v}^{k,m} - v^{k,0}\rangle + Q(v^{k,0}, \dots, v^{k,m}, m) \leq \bar{G}_H^k(v^{k,0}) + \bar{E}_H(v^{k,0}).$$

Combining (4.8) with the last inequality, yields

$$\langle I_h^H(\nabla \bar{E}(u_+^k) + \Xi u_+^k), \tilde{v}^{k,m} - v^{k,0}\rangle + \bar{G}_H^k(\tilde{v}^{k,m}) - \bar{G}_H^k(v^{k,0}) + Q(v^{k,0}, \dots, v^{k,m}, m) \leq 0. \tag{4.9}$$

Using the nonsmooth primal-dual coherence condition of Assumption 3.2, we obtain

$$\langle I_h^H(g^k + \nabla \bar{E}(u_+^k) + \Xi u_+^k), \tilde{v}^{k,m} - v^{k,0}\rangle + Q(v^{k,0}, \dots, v^{k,m}, m) \leq 0 \quad \text{for all} \quad g^k \in \partial \bar{G}(u_+^k).$$

Since $v^{k,0}$ does not solve the coarse problem, we have $v^{k,1} \neq v^{k,0}$. It follows that $Q(v^{k,0}, \dots, v^{k,m}, m) > 0$. Hence, using the definition of d^k, and taking the supremum over $g^k \in \partial \bar{G}(u_+^k)$, yields $[\Phi^k]'(u_+^k; d^k) < 0$. Moreover, by Lemma 4.2, it follows

$$\langle I_h^H(g^k + \nabla \bar{E}(u_+^k) + \Xi u_+^k), \tilde{v}^{k,m} - v^{k,0}\rangle + c_Q \|\tilde{v}^{k,m} - v^{k,0}\|^2 \leq 0 \quad \text{for all} \quad g^k \in \partial \bar{G}(u_+^k).$$

Since $I_H^h \in \mathbb{L}(U_H; U)$, also using the Cauchy–Schwarz inequality we get

$$\frac{c_Q}{\|I_H^h\|^2}\|d^k\|^2 \leq \|d^k\|\|g^k + \nabla \bar{E}(u_+^k) + \Xi u_+^k\| \quad \text{for all} \quad g^k \in \partial \bar{G}(u_+^k).$$

Taking the infimum over g^k, we get the claim. □

Remark 4.4 (Necessity and rejection of coarse corrections). Corollary 4.3 requires the initial coarse iterate to not solve (3.5). If this condition is not satisfied, we have can take $d^k = 0$ and $\theta = 0$ in Algorithm 1.1, rejecting the coarse correction. In fact, the nonsmooth primal-dual coherence condition ensures that if $u_+^k \in U$ is an optimal solution to the fine problem (1.1), then $v^{k,0} \in U_H$ is an optimal solution to the coarse problem; consequently, the coarse correction is unnecessary.

5 LINE SEARCH FOR THE COARSE CORRECTION

In this section, we describe the second stage of coarse correction: an inexpensive line search that provides sufficient descent in a form that we can use in our main convergence proof in Section 6. We already know from these results that d^k is a descent direction for Φ^k defined in (3.4). In Section 5.1, we use this result to construct a line search procedure for the Lagrangian gap, compatible with the PDPS convergence estimate, Theorem 2.3. Then, in Section 5.2, we explain how to reduce the computational cost of the line search by reusing the information used in the construction of the coarse problem.

5.1 BASIC PROCEDURE

We first prove the existence of an interval of line search parameters θ such that

$$\mathcal{G}_L(u_+^k + \theta d^k; \hat{u}) + \frac{1}{2}\|u_+^k + \theta d^k - \hat{u}\|_M^2 \leq \frac{1}{2}\|u^k - \hat{u}\|_M^2 + \frac{\varepsilon_k}{2}\|u_+^k - \hat{u}\|_M^2 + \rho_k$$

for chosen penalty parameters $\varepsilon_k, \rho_k > 0$, Choosing these parameters small, we can thus bound the Lagrangian gap arbitrarily well by the the squared distance of u^k to a minimiser $\hat{u}$. The additive penalty $\rho_k > 0$ prevents the line search step length θ from becoming arbitrarily small.

To start, we recall the basic Armijo line search result for Φ^k [21]:

Lemma 5.1. *On the given fine iteration $k \in \mathbb{N}$, suppose Assumptions 3.1 and 3.2 hold for the initial coarse iterate $v^{k,0} = I_h^H u_+^k$, which does not solve (3.5). Let d^k be defined by Corollary 4.3. Then, for any $\kappa \in (0,1)$, there exist $\theta_0 > 0$ such that*

$$\Phi^k(u_+^k + \theta d^k) \le \Phi^k(u_+^k) + \kappa\theta[\Phi^k]'(u_+^k; d^k) \quad \text{for all} \quad 0 < \theta \le \theta_0. \tag{5.1}$$

Proof. By Corollary 4.3, we know that d^k is a descent direction for Φ^k at u_+^k, that is, $[\Phi^k]'(u_+^k; d^k) < 0$. The rest follows from the definition of the directional derivative. □

The next lemma provides the basis for our basic line search procedure. There, using the constant $C > 0$ be the from Lemma 2.2, we define $W, Z \in \mathbb{L}(U;U)$ as

$$Z := \begin{pmatrix} \tau^{-1}\mathrm{Id} & -2K^* \\ 0 & \sigma^{-1}\mathrm{Id} \end{pmatrix} \quad \text{and} \quad W = \frac{1}{C} ZZ^*.$$

Lemma 5.2. *On the given fine iteration $k \in \mathbb{N}$, suppose Assumptions 3.1 and 3.2 hold for the initial coarse iterate $v^{k,0} = I_h^H u_+^k$, which does not solve (3.5). Let $\varepsilon_k, \rho_k > 0$ and $\kappa \in (0,1)$. Then there exists $\theta_0 > 0$ such that*

$$\Phi^k(u_+^k + \theta d^k) + \frac{1}{2}\|\theta d^k\|_M^2 + \frac{1}{2\varepsilon_k}\|\theta d^k\|_W^2 \le \Phi^k(u_+^k) + \kappa\theta[\Phi^k]'(u_+^k; d^k) + \frac{\rho_k}{2} \tag{5.2}$$

for all $0 < \theta \le \theta_0$ and d^k defined in Corollary 4.3.

Proof. We have $[\tilde{\Phi}^k]'(u_+^k) = [\Phi^k]'(u_+^k) < 0$ for

$$\tilde{\Phi}^k(u) := \Phi^k(u) + \frac{1}{2}\|u - u_+^k\|_M^2 + \frac{1}{2\varepsilon_k}\|u - u_+^k\|_W^2.$$

Corollary 4.3 shows that $d^k = I_H^h(\tilde{v}^{k,m} - v^{k,0})$ is a descent direction for Φ^k at u_+^k, hence also for $\tilde{\Phi}^k$. Lemma 5.1 applied to $\tilde{\Phi}^k$, thus, shows the existence of $\theta_0 > 0$ such that $\tilde{\Phi}^k(u_+^k + \theta d^k) \le \tilde{\Phi}^k(u_+^k) + \kappa\theta[\tilde{\Phi}^k]'(u_+^k, d^k)$ for all $0 < \theta \le \theta_0$. Now, applying the definition of $\tilde{\Phi}$ and $\rho_k/2 > 0$ then yields (5.2). □

The next result transforms some terms in the line search criterion.

Lemma 5.3. *Let $\varepsilon_k > 0$. Suppose that Assumption 2.1 holds. Then, for any $u, \hat{u}, \tilde{u} \in U$, we have*

$$\langle \Xi(u - \hat{u}), \tilde{u}\rangle + \frac{1}{2}\|u - \hat{u}\|_M^2 \ge \frac{1}{2}\|u + \tilde{u} - \hat{u}\|_M^2 - \frac{\varepsilon_k}{2}\|u - \hat{u}\|_M^2 - \frac{1}{2\varepsilon_k}\|\tilde{u}\|_W^2 - \frac{1}{2}\|\tilde{u}\|_M^2.$$

Proof. The operator M of (2.4) can be split as $M = Z + \Xi$. With this, we obtain

$$\langle \Xi(u - \hat{u}), \tilde{u}\rangle + \frac{1}{2}\|u - \hat{u}\|_M^2 = \langle M(u - \hat{u}), \tilde{u}\rangle + \frac{1}{2}\|u - \hat{u}\|_M^2 - \langle Z(u - \hat{u}), \tilde{u}\rangle. \tag{5.3}$$

By the Pythagorean identity, we have

$$\langle M(u - \hat{u}), \tilde{u}\rangle + \frac{1}{2}\|u - \hat{u}\|_M^2 = \frac{1}{2}\big[\|u + \tilde{u} - \hat{u}\|_M^2 - \|\tilde{u}\|_M^2\big],$$

and since $\|u\|_W^2 = \|C^{-1/2}Z^*u\|^2$, we have

$$\begin{aligned} -\langle Z(u - \hat{u}), \tilde{u}\rangle &= -\langle u - \hat{u}, \theta Z^* d^k\rangle \\ &= \frac{1}{2}\big[\|(C\varepsilon_k)^{1/2}(u - \hat{u}) - (C\varepsilon_k)^{-1/2}\theta Z^* d^k\|^2 - C\varepsilon_k\|u - \hat{u}\|^2 - \frac{1}{\varepsilon_k}\|\tilde{u}\|_W^2\big]. \end{aligned}$$

Replacing the last two identities in (5.3) yields

$$\langle \Xi(u-\hat{u}), \tilde{u}\rangle + \frac{1}{2}\|u-\hat{u}\|_M^2 = \frac{1}{2}\Big[\|u+\tilde{u}-\hat{u}\|_M^2 - C\varepsilon_k\|u-\hat{u}\|^2 - \frac{1}{\varepsilon_k}\|\tilde{u}\|_W^2 - \|\tilde{u}\|_M^2 + \|(C\varepsilon_k)^{1/2}(u-\hat{u}) - (C\varepsilon_k)^{-1/2}\theta Z^* d^k\|^2\Big].$$

Finally, we discard undesired non-negative terms and use $C\|u\|^2 \le \|u\|_M^2$. □

Finally, we have our desired result.

Theorem 5.4. *On the given fine iteration $k \in \mathbb{N}$, suppose Assumptions 2.1 to 3.2 hold for the initial coarse iterate $v^{k,0} = I_h^H u_+^k$, which does not solve (3.5). Moreover, let $\varepsilon_k, \rho_k > 0$, and let $\theta > 0$ (which exists, by Lemma 5.2) satisfy (5.2). Then, for all $\hat{u} \in U$ and d^k defined in Corollary 4.3, we have*

$$\mathcal{G}_L(u_+^k + \theta d^k; \hat{u}) + \frac{1}{2}\|u_+^k + \theta d^k - \hat{u}\|_M^2 + \frac{1}{2}\|u_+^k - u^k\|_{M-\Lambda}^2 \le \frac{1}{2}\|u^k - \hat{u}\|_M^2 + \frac{\varepsilon_k}{2}\|u_+^k - \hat{u}\|_M^2 + \frac{\rho_k}{2}.$$

Proof. By Theorem 2.3 and Line 3 of Algorithm 1.1, the fine-grid pre-iterate u_+^k satisfies

$$\mathcal{G}_L(u_+^k; \hat{u}) + \frac{1}{2}\|u_+^k - \hat{u}\|_M^2 + \frac{1}{2}\|u_+^k - u^k\|_{M-\Lambda}^2 \le \frac{1}{2}\|u^k - \hat{u}\|_M^2.$$

Using (2.8), (5.2) rewrites as

$$\mathcal{G}_L(u_+^k + \theta d^k; \hat{u}) + \langle \Xi(u_+^k - \hat{u}), \theta d^k\rangle + \frac{1}{2}\Big[\|\theta d^k\|_M^2 + \frac{1}{\varepsilon_k}\|\theta d^k\|_W^2\Big] \le \mathcal{G}_L(u_+^k; \hat{u}) + \frac{\rho_k}{2}.$$

Combining the last last two inequalities yields

$$\mathcal{G}_L(u_+^k + \theta d^k; \hat{u}) + \langle \Xi(u_+^k - \hat{u}), \theta d^k\rangle + \frac{1}{2}\|u_+^k - \hat{u}\|_M^2 + \frac{1}{2}\|u_+^k - u^k\|_{M-\Lambda}^2 + \frac{1}{2}\|\theta d^k\|_M^2 + \frac{1}{2\varepsilon_k}\|\theta d^k\|_W^2 \le \frac{1}{2}\|u^k - \hat{u}\|_M^2 + \frac{\rho_k}{2}.$$

Applying Lemma 5.3 for $u = u_+^k$ and $\tilde{u} = \theta d^k$, establishes the claim. □

5.2 EFFICIENT LINEARISED LINE SEARCH

In many applications, including the examples of Section 8, the evaluation of the smooth function E in the line search criterion (5.2) can be computationally highly expensive. To avoid the full evaluation of E, we will now perform line search on the E-linearisation of (a quadratically penalized version of) Φ^k of (3.4), i.e.,

$$\Psi^k(u) := \bar{G}(u) + \bar{E}(u_+^k) + \langle \nabla \bar{E}(u_+^k), u - u_+^k\rangle + \langle \Xi u_+^k, u\rangle.$$

To prove that this works, we exploit the descent inequality

$$E(x+h) \le E(x) + \langle \nabla E(x), h\rangle + \frac{L}{2}\|h\|^2, \tag{5.4}$$

which holds when E has L-Lipschitz gradient, even without convexity [9, Chapter 7].

Lemma 5.5. *Assume that the conditions of Lemma 5.1 hold. Let $\varepsilon_k, \rho_k > 0$, and $\kappa \in (0,1)$. With d^k defined by Corollary 4.3, there then exists $\theta_0 > 0$ such that for all $\theta \in [0, \theta_0]$, we have*

$$\Psi^k(u_+^k + \theta d^k) + \frac{1}{2}\|\theta d^k\|_{M+\Lambda}^2 + \frac{1}{2\varepsilon_k}\|\theta d^k\|_W^2 \le \Psi^k(u_+^k) + \kappa\theta[\Psi^k]'(u_+^k; d^k) + \frac{\rho_k}{2}. \tag{5.5}$$

Proof. Corollary 4.3 shows that $d^k = I_H^h(\tilde{v}^{k,m} - v^{k,0})$ is a descent direction for Φ^k. Let

$$\tilde{\Psi}^k(u) := \Psi^k(u) + \frac{1}{2}\|u - u_+^k\|_{M+\Lambda}^2 + \frac{1}{2\varepsilon_k}\|u - u_+^k\|_W^2.$$

Then $[\tilde{\Psi}^k]'(u_+^k; d^k) = [\Psi^k]'(u_+^k; d^k) = [\Phi^k]'(u_+^k; d^k)$. Therefore, (5.5) rewrites as

$$\tilde{\Psi}^k(u_+^k + \theta d^k) \leq \tilde{\Psi}^k(u_+^k) + \kappa\theta[\tilde{\Psi}^k]'(u_+^k; d^k) + \frac{\rho_k}{2}.$$

Now we simply use Lemma 5.1 on $\tilde{\Psi}^k$. □

This result allows us to formulate an efficient line search procedure.

Theorem 5.6. *Assume that ∇E is L-Lipschitz. Then (5.5) implies the sufficient descent condition (5.2) of Lemma 5.2.*

Proof. Since $\Psi^k(u_+^k) = \Phi^k(u_+^k)$ as well as $[\Psi^k]'(u_+^k; d^k) = [\Phi^k]'(u_+^k; d^k)$, it suffices to show that $\Phi^k(u_+^k + \theta d^k) \leq \Psi^k(u_+^k + \theta d^k) + \frac{1}{2}\|\theta d^k\|_\Lambda^2$. But this is immediate from to the descent inequality (5.4), that is $\bar{E}(u_+^k + \theta d^k) \leq \bar{E}(u_+^k) + \langle \nabla\bar{E}(u_+^k), \theta d^k\rangle + \frac{1}{2}\|\theta d^k\|_\Lambda^2$. □

Corollary 5.7. *We can replace (5.2) by (5.5) in Theorem 5.4.*

6 CONVERGENCE ANALYSIS

We finally prove the convergence of Algorithm 1.1. We start in Section 6.1 by presenting a bound on the ergodic gap of the fine problem. This bound still depends on a uniform bound on $\{\|u_+^k\|\}_{k\in\mathbb{N}}$, which we derive in Section 6.2. With that in hand, we can can then, in Section 6.3, establish the ergodic convergence of the Lagrangian gap, and, via Féjer quasi-monotonicity and Opial's lemma, the weak convergence of the iterates. We also illustrate in Remark 6.9, how our results can be extended to nonconvex E.

6.1 A PRELIMINARY ESTIMATE

The next lemma almost shows ergodic gap convergence, however, the right hand side will still need to be bounded. There, we set

$$\mathcal{J} := \{k \in \mathbb{N} \mid \text{the trigger condition of Algorithm 1.1 holds on iteration } k\}.$$

Lemma 6.1. *Let $\{\varepsilon_k\}_{k\in\mathcal{J}}$ and $\{\rho_k\}_{k\in\mathcal{J}}$ be two sequences such that $\varepsilon_k, \rho_k > 0$ for all $k \in \mathcal{J}$. Suppose that Assumptions 2.1 to 3.2 hold. Then, for any initial iterate $u^0 = (x^0, y^0) \in U$, the sequence $\{u^k\}_{k\in\mathbb{N}}$ generated by Algorithm 1.1 satisfies*

$$\sum_{k=0}^{N-1} \mathcal{G}_L(u^{k+1}; \hat{u}) + \frac{1}{2}\|u^N - \hat{u}\|_M^2 + \frac{1}{2}\sum_{k=0}^{N-1}\|u_+^k - u^k\|_{M-\Lambda}^2 \leq \frac{1}{2}\|u^0 - \hat{u}\|_M^2 + \frac{1}{2}\sum_{k=0}^{N-1}\gamma_{k+1} \tag{6.1}$$

for all $\hat{u} \in U$ and

$$\gamma_{k+1} := \begin{cases} 0, & k \notin \mathcal{J}, \\ \varepsilon_k\|u_+^k - \hat{u}\|_M^2 + \rho_k, & k \in \mathcal{J}. \end{cases} \tag{6.2}$$

Proof. Algorithm 1.1 alternates between fine PDPS iterations and coarse corrections. By Theorems 2.3 and 5.4 and Corollary 5.7

$$\mathcal{G}_L(u^{k+1};\hat{u}) + \frac{1}{2}\|u^{k+1}-\hat{u}\|_M^2 + \frac{\mu_{k+1}}{2} \le \frac{1}{2}\|u^k-\hat{u}\|_M^2 + \frac{\gamma_{k+1}}{2} \quad \text{for all} \quad k\in\mathbb{N} \tag{6.3}$$

and

$$\mu_{k+1} := \begin{cases} \|u^{k+1}-u^k\|_{M-\Lambda}^2, & k\notin\mathcal{J}, \\ \|u_+^k-u^k\|_{M-\Lambda}^2, & k\in\mathcal{J}. \end{cases}$$

However, by Line 12, we have $\mu_{k+1} = \|u_+^k-u^k\|_{M-\Lambda}^2$ for any $k\notin\mathcal{J}$, and hence for all $k\ge 0$. Summing (6.3) over $k=0,\dots,N-1$ establishes the claim. □

6.2 UNIFORM BOUNDEDNESS

For (6.1) to establish convergence of the Lagrangian gaps, we now need to bound $\sum_{k\in\mathcal{J}}\varepsilon_k\|u_+^k-\hat{u}\|_M^2+\rho_k$. This follows when $\{u_+^k\}_{k\in\mathcal{J}}$ is uniformly bounded, which is what we now prove. We start with technical results on real sequences.

Lemma 6.2. *Let $\{\varepsilon_k\}_{k\in\mathbb{N}}\subset(0,\infty)$ satisfy $\sum_{k\in\mathbb{N}}\varepsilon_k<\infty$. Then $\prod_{k\in\mathbb{N}}(1+\varepsilon_k)<\infty$.*

Proof. Since the exponential function is convex, we have $1+x\le\exp(x)$ for all $x\in\mathbb{R}$. In addition, given that $1+\varepsilon_k>0$ for all $k\ge0$, we obtain $\prod_{k=0}^n(1+\varepsilon_k)\le\exp\left(\sum_{k=0}^n\varepsilon_k\right)\le\exp\left(\sum_{k\in\mathbb{N}}\varepsilon_k\right)$. Therefore, the sequence of partial products is uniformly bounded, and consequently, the desired result holds. □

Lemma 6.3. *Let $\{\varepsilon_i\}_{i\in\mathbb{N}},\{\rho_i\}_{i\in\mathbb{N}}\subset(0,\infty)$ satisfy $\sum_{i\in\mathbb{N}}\varepsilon_i<\infty$ and $\sum_{i\in\mathbb{N}}\rho_i<\infty$. Let $\{\omega_k\}_{k\in\mathbb{N}}\subset(0,\infty)$ be such that*

$$\omega_{k+1}\le\varpi_k := \omega_0+\sum_{i=0}^k[\varepsilon_i\omega_i+\rho_i] \tag{6.4}$$

for all $k\ge0$ and a $\omega_0\ge0$. Then, $\omega_k\le(\omega_0+\sum_{i\in\mathbb{N}}\rho_i)[\prod_{i\in\mathbb{N}}(1+\varepsilon_i)]<\infty$ for all $k\ge0$.

Proof. Since $\sum_{i\in\mathbb{N}}\varepsilon_i<\infty$ and $\sum_{i\in\mathbb{N}}\rho_i<\infty$, Lemma 6.2 guarantees that $\prod_{i=0}^k(1+\varepsilon_i)\le\prod_{i\in\mathbb{N}}(1+\varepsilon_i)<\infty$ and also $\sum_{i=0}^k\rho_i\le\sum_{i\in\mathbb{N}}\rho_i<\infty$. Thus, our claim follows if we prove that $\varpi_k\le\beta_k$, where we define and estimate

$$\beta_k := \omega_0\prod_{i=0}^k(1+\varepsilon_i)+\sum_{i=0}^k\rho_i\prod_{p=i+1}^k(1+\varepsilon_p)\le(\omega_0+\sum_{i\in\mathbb{N}}\rho_i)\left[\prod_{i\in\mathbb{N}}(1+\varepsilon_i)\right].$$

The proof is by induction. For $k=0$, the definition of ϖ_0 yields

$$\varpi_0=\omega_0+\varepsilon_0\omega_0+\rho_0\le\omega_0+\varepsilon_0\omega_0+\rho_0=\omega_0(1+\varepsilon_0)+\rho_0=\beta_0.$$

Now suppose it holds for $k-1$. Firstly, based on the definition of β_k, we have

$$\beta_k=\left[\omega_0\prod_{i=0}^{k-1}(1+\varepsilon_i)+\sum_{i=0}^{k-1}\rho_i\prod_{p=i+1}^{k-1}(1+\varepsilon_p)\right](1+\varepsilon_k)+\rho_k=\beta_{k-1}(1+\varepsilon_k)+\rho_k,$$

for all $k\ge1$. Now, by (6.4), we have $\omega_k\le\varpi_{k-1}\le\beta_{k-1}$. Then,

$$\varpi_k=\omega_0+\sum_{i=0}^k[\varepsilon_i\omega_i+\rho_i]=\varpi_{k-1}+\varepsilon_k\omega_k+\rho_k\le\beta_{k-1}(1+\varepsilon_k)+\rho_k=\beta_k.$$ □

The next result establishes the required uniform bound, subject to the following control on the line search model parameters.

Assumption 6.4. $\{\varepsilon_k\}_{k\in\mathcal{J}} \subset (0,\infty)$ and $\{\rho_k\}_{k\in\mathcal{J}} \subset (0,\infty)$ satisfy

$$\sum_{k\in\mathcal{J}} \varepsilon_k := \varepsilon < \infty, \quad \sum_{k\in\mathcal{J}} \rho_k := \rho < \infty, \quad \text{and} \quad \prod_{k\in\mathcal{J}} (1+\varepsilon) := \kappa_\varepsilon < \infty.$$

Corollary 6.5. *Let Assumptions 2.1 to 3.2 and 6.4 hold. Then for any initial $u^0 \in \Omega$ and any primal-dual solution $\hat{u} \in \hat{U} := [\partial\bar{G} + \nabla\bar{E} + \Xi]^{-1}(0)$ of the fine problem (1.1), we have*

$$\sup_{k\in\mathcal{J}} \|u_+^k - \hat{u}\|_M^2 \le \kappa_\varepsilon(\|u^0 - \hat{u}\|_M^2 + \rho). \tag{6.5}$$

Furthermore,

$$\sum_{k=0}^{N-1} \gamma_{k+1} \le \sum_{k\in\mathcal{J}} \varepsilon_k \|u_+^k - \hat{u}\|_M^2 + \rho_k \le \kappa_\varepsilon(\|u^0 - \hat{u}\|_M^2 + \rho)\varepsilon + \rho < \infty. \tag{6.6}$$

Proof. We order the trigger iteration indices $k_n \in \mathcal{J}$ as $k_0 < k_1 < k_2 < \dots$. Since Line 3 of Algorithm 1.1 performs a standard PDPS pre-step performed each coarse correction, Theorem 2.3 yields

$$\|u_+^{k_n} - \hat{u}\|_M^2 \le \|u^{k_n} - \hat{u}\|_M^2 \quad \text{for all} \quad n \ge 0. \tag{6.7}$$

Moreover, after updating the fine variable with the coarse correction Line 10, PDPS steps are performed until reaching the next coarse correction. Thus, (6.3) from Lemma 6.1 holds with $\gamma_{k+1} = 0$ for all the non-trigger iterations $k = k_{n-1}+1, \dots, k_n - 1$ for every $n \ge 0$, where we set $k_{-1} := -1$. By Assumption 2.1 and Lemma 2.2, we have $\mathcal{G}_L(u^{k+1}; \hat{u}) \ge 0$ and $M \ge \Lambda$, which implies that $\mu_{k+1} \ge 0$. Consequently,

$$\|u^{k_0} - \hat{u}\|_M^2 \le \|u^0 - \hat{u}\|_M^2 \quad \text{and} \quad \|u^{k_n} - \hat{u}\|_M^2 \le \|u^{k_{n-1}+1} - \hat{u}\|_M^2, \quad \text{for all} \quad n \ge 1. \tag{6.8}$$

On each trigger iteration $k_{n-1} \in \mathcal{J}$, $(n \ge 1)$, it follows from the definition of $\gamma_{k_{n-1}+1}$ in (6.3), from Lemma 6.1, that

$$\|u^{k_{n-1}+1} - \hat{u}\|_M^2 \le \|u^{k_{n-1}} - \hat{u}\|_M^2 + \varepsilon_{k_{n-1}} \|u_+^{k_{n-1}} - \hat{u}\|_M^2 + \rho_{k_{n-1}}, \quad \text{for all} \quad n \ge 1. \tag{6.9}$$

Combining the second inequality of (6.8) with (6.9) yields

$$\|u^{k_n} - \hat{u}\|_M^2 \le \|u^{k_{n-1}} - \hat{u}\|_M^2 + \varepsilon_{k_{n-1}} \|u_+^{k_{n-1}} - \hat{u}\|_M^2 + \rho_{k_{n-1}} \quad \text{for all} \quad n \ge 1.$$

Summing this over $n = 1, \dots, N$ and using (6.7) and the first inequality of (6.8), we obtain

$$\|u_+^{k_N} - \hat{u}\|_M^2 \le \|u^{k_N} - \hat{u}\|_M^2 \le \|u^0 - \hat{u}\|_M^2 + \sum_{i=0}^{N-1} [\varepsilon_{k_i} \|u_+^{k_i} - \hat{u}\|_M^2 + \rho_{k_i}].$$

This reads as $\omega_N \le \omega_0 + \sum_{i=0}^{N-1} [\varepsilon_{k_i}\omega_i + \rho_{k_i}]$ for $\omega_i := \|u_+^{k_i} - \hat{u}\|_M^2$ and $\omega_0 := \|u^0 - \hat{u}\|_M^2$. Assumption 6.4 and Lemma 6.3 show the uniform boundedness of the sequence $\{u_+^k - \hat{u}\}_{k\in\mathcal{J}}$, i.e., (6.5). Since $\{u_+^k\}_{k\in\mathcal{J}}$ is uniformly bounded, it follows that

$$\sum_{k=0}^{N-1} \gamma_{k+1} = \sum_{k\in\mathcal{J}_N} \varepsilon_k \|u_+^k - \hat{u}\|_M^2 + \rho_k \le \sup_{k\in\mathcal{J}} \|u_+^k - \hat{u}\|_M^2 \sum_{k\in\mathcal{J}_N} [\varepsilon_k + \rho_k],$$

where $\mathcal{J}_N := \{0, \dots, N-1\} \cap \mathcal{J}$. Consequently, (6.5) implies (6.6). □

6.3 CONVERGENCE

We can now prove the Fejér quasi-monotonicity of the sequence generated by Algorithm 1.1, and establish the ergodic convergence of the Lagrangian gap, as well as the weak convergence of the iterates.

Corollary 6.6. *Assume that Assumptions 2.1 to 3.2 and 6.4 hold. Then for any initial $u^0 \in U$ the iterates generated by Algorithm 1.1 satisfy for any $\hat{u} \in := [\partial\bar{G} + \nabla\bar{E} + \Xi]^{-1}(0)$ the ergodic gap estimate*

$$\mathcal{G}_L(\tilde{u}^N; \hat{u}) \leq \frac{\varepsilon\kappa_\varepsilon + 1}{2N}\left[\|u^0 - \hat{u}\|_M^2 + \rho\right] \quad \text{where} \quad \tilde{u}^N := \frac{1}{N}\sum_{k=0}^{N-1} u^{k+1}. \tag{6.10}$$

In particular $\mathcal{G}_L(\tilde{u}^N; \hat{U}) \to 0$ at the rate $O(1/N)$.

Proof. Combining Corollary 6.5 and Lemma 6.1 and using $M \geq \Lambda$ from Lemma 2.2, we get

$$\sum_{k=0}^{N-1} \mathcal{G}_L(u^{k+1}; \hat{u}) \leq \frac{1}{2}\left[\|u^0 - \hat{u}\|_M^2 + \kappa_\varepsilon(\|u^0 - \hat{u}\|_M^2 + \rho)\varepsilon + \rho\right] = \frac{\varepsilon\kappa_\varepsilon + 1}{2}\left[\|u^0 - \hat{u}\|_M^2 + \rho\right] < \infty.$$

An application of Jensen's inequality now establishes (6.10). □

Theorem 6.7. *Assume that Assumptions 2.1 to 3.2 and 6.4 hold. Suppose $[\partial\bar{G} + \nabla\bar{E} + \Xi]^{-1}(0) \neq \emptyset$, i.e., the fine problem (1.1) has a minimizer. Then for any initial $u^0 \in U$ the sequence generated by Algorithm 1.1 is quasi-Fejér monotone, and converges weakly to a root $\hat{u} \in [\partial\bar{G} + \nabla\bar{E} + \Xi]^{-1}(0)$.*

Proof. We note $\hat{U} := [\partial\bar{G} + \nabla\bar{E} + \Xi]^{-1}(0)$. Let $\hat{u} \in \hat{U}$ be an optimal solution to the problem (1.1). Let γ_{k+1} be given by (6.2). Since $\mathcal{G}_L(u; \hat{u}) \geq 0$ for all $u \in U$ and $M \geq \Lambda$, the inequality (6.3) reduces to

$$\|u^{k+1} - \hat{u}\|_M^2 \leq \|u^k - \hat{u}\|_M^2 + \gamma_{k+1} \quad \text{for all} \quad k \geq 0.$$

We already know from Corollary 6.5 that $\sup_{k\in\mathcal{J}} \|u_+^k - \hat{u}\|_M^2 < \infty$, hence

$$\gamma_{k+1} \leq \tilde{\gamma}_{k+1} = \begin{cases} 0 & k \in \mathcal{J}, \\ \kappa_\varepsilon(\|u^0 - \hat{u}\|_M^2 + \rho)\varepsilon_k + \rho_k & k \notin \mathcal{J}. \end{cases}$$

Since Assumption 6.4 holds, it follows that $\sum_{k\in\mathbb{N}} \tilde{\gamma}_{k+1} < \infty$. Thus the sequence $\{u^k\}_{k\in\mathbb{N}}$ exhibits quasi-Fejér monotonicity respect to $\hat{U}$.

By (6.3) from Lemma 6.1, we have, for all $k \in \mathbb{N}$ and $\hat{u} \in \hat{U}$,

$$\mathcal{G}_L(u^{k+1}; \hat{u}) + \frac{1}{2}\|u^{k+1} - \hat{u}\|_M^2 + \frac{1}{2}\|u_+^k - u^k\|_{M-\Lambda}^2 \leq \frac{1}{2}\|u^k - \hat{u}\|_M^2 + \frac{\gamma_{k+1}}{2}.$$

Since $\mathcal{G}_L(u^k; \hat{u}) \geq 0$ and summing over $k = 0, \ldots, N-1$, we obtain

$$\sum_{k=0}^{N-1} \|u_+^k - u^k\|_{M-\Lambda}^2 + \|u^N - \hat{u}\|_M^2 \leq \|u^0 - \hat{u}\|_M^2 + \sum_{k=0}^{N-1} \gamma_{k+1}.$$

Using (6.6) from Corollary 6.5, we have

$$\sum_{k=1}^{N-1} \|u_+^k - u^k\|_{M-\Lambda}^2 \leq \frac{\varepsilon\kappa_\varepsilon + 1}{2}[\|u^0 - \hat{u}\|_M^2 + \rho] < \infty.$$

This proves that $\|u_+^k - u^k\|_{M-\Lambda} \to 0$. By Assumption 2.1 and Lemma 2.2, we conclude that $u_+^k - u^k \to 0$. Let $\hat{u}$ be a limit point of $\{u_+^k\}_{k\in\mathbb{N}}$, i.e., there exists a subsequence $\{k_i\}_{i\in\mathbb{N}}$ such that $u_+^{k_i} \to \hat{u}$. To prove that $\hat{u} \in \hat{U}$, we consider the implicit equation, related to Line 3 of Algorithm 1.1,

$$0 \in \partial\bar{G}(u_+^{k_i}) + \Xi u_+^{k_i} + \nabla\bar{E}(u^{k_i}) + M(u_+^{k_i} - u^{k_i}).$$

Since $\partial\bar{G} + \nabla\bar{E} + \Xi$ is weak-to-strong outer semicontinous, M is self-adjoint, bounded and positive definite, and $u_+^k - u^k \to 0$, we conclude that $\hat{u} \in \hat{U}$, (cf. [9, Chapter 9]). However, since $u_+^k - u^k \to 0$ it follows that $u^{k_i} = (u^{k_i} - u_+^{k_i}) + u_+^{k_i} \to \hat{u}$, i.e., both sequences, $\{u^k\}_{k\in\mathbb{N}}$ and $\{u_+^k\}_{k\in\mathbb{N}}$, share the same set of limit points, and consequently all limit points of $\{u^k\}_{k\in\mathbb{N}}$ are solutions of the problem. Finally, applying Opial's lemma for quasi-Fejér sequences [12, Lemma A.2.], the result follows. □

The next result shows that also the coarse corrections converge, to zero.

Corollary 6.8. *Let Assumptions 2.1 to 3.2 and 6.4 hold. Suppose $\hat{U} := [\partial\bar{G} + \nabla\bar{E} + \Xi]^{-1}(0) \neq \emptyset$. Then the coarse-grid correction vanishes asymptotically, i.e., $d^k \to 0$.*

Proof. By Corollary 4.3, we have

$$\|d^k\| \leq \tilde{c}_Q \inf_{g\in\partial\bar{G}(u_+^k)} \|g + \nabla\bar{E}(u_+^k) + \Xi u_+^k\| \quad \text{for all} \quad k \in \mathcal{J}. \tag{6.11}$$

On the other hand, by Line 3 of Algorithm 1.1, there exists $g^k \in \partial\bar{G}(u_+^k)$ such that

$$g^k + \nabla\bar{E}(u_+^k) + \Xi u_+^k = \nabla\bar{E}(u_+^k) - \nabla\bar{E}(u^k) - \tau^{-1}M(u_+^k - u^k).$$

Since ∇E is L-Lipschitz continuous and, by Assumption 2.1, M is bounded, combining this identity with (6.11), we obtain

$$\|d^k\| \leq \tilde{c}_Q\|\nabla\bar{E}(u_+^k) - \nabla\bar{E}(u^k) - \tau^{-1}M(u_+^k - u^k)\| \leq \tilde{c}_Q(L + \tau^{-1}\|M\|)\|u_+^k - u^k\|.$$

We finish by observing from the proof of Theorem 6.7 that $u_+^k - u^k \to 0$ □

Remark 6.9 (Nonconvex fine problems). The convexity of E has been required only in Theorem 2.3, and in Corollary 6.5 to have $\mathcal{G}_L(u^{k+1}; \hat{u}) \geq 0$. In the first case, this arises through a three-point descent inequality [9, Chapters 7 and 11]

$$\langle\nabla E(x^k), x^{k+1} - \hat{x}\rangle \geq E(x^{k+1}) - E(\hat{x}) - \frac{L}{2}\|x^{k+1} - x^k\|^2,$$

where L is the Lipschitz factor of ∇E. The two-point descent inequality (5.4), i.e., $\hat{x} = x^k$ here, holds even without convexity. The three-point version holds for nonconvex functions provided x^k is in a local neighborhood of $\hat{x}$ with some second-order growth [26, 12]. There is no requirement for x^{k+1} to be in that neighborhood on iteration k.

However, ensuring that $\mathcal{G}_L(u^{k+1}; \hat{u}) \geq 0$ is somewhat more involved. We need to repeat the arguments of the proof *a priori* without involving the gap, using monotonicity and three-point co-coercivity. Though the summability of $\tilde{\gamma}_k$, we can then obtain an a priori bound on $\|x^{k+1} - \hat{x}\|$, which will guarantee that we stay in a given local neighborhood, if we start close enough to a solution. We can then improve this bound by repeating the above gap-based arguments *a posteriori*. See [8, 12, 11] for details.

With this, we can, locally, extend our gap convergence results to nonconvex E. In [12, §7.3], it is also shown how to obtain convergence of the convex envelope from gap convergence. Weak convergence of iterates iterates requires, e.g., explicitly assuming weak-to-strong continuity of ∇E; compare [8] and [12, §5.6].

7 CONSTRUCTION OF COARSE FUNCTIONS

We now provide several examples on the construction of the smooth coarse function E_H (Section 7.1), and recall the approach of [14] for the nonsmooth component functions $(G_H^*)^k$ and F_H^k (Section 7.2).

7.1 DATA TERM

In inverse problems applications, the smooth function E is typically a data term, which involves expensive-to-evaluate operators mapping a desired reconstruction to measurable data. An important objective in the construction of the coarse variant E_H is to reduce the operator evaluation cost. Our first "ideal" example often fails this, as it requires evaluating the original fine function.

Example 7.1 (Reparametrisation of the fine function). Given the initial primal coarse iterate $\zeta^{k,0} = P_h^H x_+^k$, the conceptually ideal choice for $E_H : X_H \to \overline{\mathbb{R}}$ is

$$E_H(\zeta) := E(x_+^k + P_H^h(\zeta - \zeta^{k,0})).$$

That is, we use the original E, adding the restriction error $x_+^k - P_H^h \zeta^{k,0}$ to the prolongation of ζ. Then, at the initial iterate, $\nabla E_H(\zeta^{k,0}) = P_h^H \nabla E(x_+^k)$. Thus, in Algorithm 4.1, $a_H^k = P_h^H K^* y_+^k$. For the primal coarse step on Line 4 of Algorithm 4.1, we use the definition (3.2) of E_H^k to construct

$$\begin{cases} r_H^k = P_h^H(\nabla E(x_+^k) + K^* y_+^k) - (\nabla E_H(\zeta^{k,0}) + K_H^*(\xi^{k,0})) = P_h^H K^* y_+^k - K_H^* \xi^{k,0} & \text{ergodic,} \\ r_H^{k,j} = P_h^H(\nabla E(x_+^k) + K^* y_+^k) - (\nabla E_H(\zeta^{k,0}) + K_H^*(\xi^{k,j})) = P_h^H K^* y_+^k - K_H^* \xi^{k,j} & \text{non-ergodic.} \end{cases}$$

Hence, the gradient of E_H^k reads

$$\nabla E_H^k(\zeta) = \begin{cases} P_h^H \nabla E(x_+^k - P_H^h(\zeta - \zeta^{k,0})) + P_h^H K^* y_+^k - K_H^* \xi^{k,0} & \text{ergodic,} \\ P_h^H \nabla E(x_+^k - P_H^h(\zeta - \zeta^{k,0})) + P_h^H K^* y_+^k - K_H^* \xi^{k,j} & \text{non-ergodic.} \end{cases} \tag{7.1}$$

Moreover, ∇E_H is $L_H = \|P_H^h\|\|P_h^H\| L$-Lipshitz, if ∇E is L-Lipschitz, so the relevant parts of Assumption 3.1 hold.

Example 7.2 (Linearisation). In the previous example, computing $\nabla E_H(\zeta)$ for $\zeta \neq \zeta^{k,0}$ can be expensive. For this reason, we introduce its linearisation

$$E_H(\zeta) := E(x_+^k) + \langle \nabla E(x_+^k), P_H^h(\zeta - \zeta^{k,0}) \rangle$$

Again, $\nabla E_H(\zeta^{k,0}) = P_h^H \nabla E(x_+^k)$ holds, and r_H^k and $r_H^{k,j}$ coincide with Example 7.1. Now, ∇E_H^k is Lipshitz with factor $L_H = 0$, and

$$\nabla E_H^k(\zeta) = \begin{cases} P_h^H(\nabla E(x_+^k) + K^* y_+^k) - K_H^* \xi^{k,0} & \text{ergodic,} \\ P_h^H(\nabla E(x_+^k) + K^* y_+^k) - K_H^* \xi^{k,j} & \text{non-ergodic.} \end{cases}$$

Example 7.3 (Quadratic data terms). For $A \in \mathbb{L}(X; V)$ and data $e \in V$, consider

$$E(x) := \frac{1}{2}\|Ax - e\|_V^2$$

in an Euclidean space V. It seems then reasonable to take E_H of the same form,

$$E_H(\zeta) := \frac{1}{2}\|A_H\zeta - e_H\|^2_{V_H}$$

for some linear operator $A_H \in \mathbb{L}(X_H; V_H)$ and data $e_H \in V_H$ in an Euclidean space V_H. Recall from Example 7.1 that an ideal choice of E_H would generally be

$$E_H(\zeta) := E(x^k_+ + I^h_H(\zeta - \zeta^{k,0})) = \frac{1}{2}\|A(x^k_+ + P^h_H(\zeta - \zeta^{k,0})) - e\|^2_V$$

i.e. $e_H = e + P^h_H\zeta^{k,0} - Ax^k_+$ and $A_H = AP^h_H$. Computing $\nabla E_H = A^*_H(A_H\zeta - e_H)$ only requires one application of $A^*_H A_H$ per iteration and a single pre-computation of the coarse data $A^*_H e_H$. However, we do not wish to compute the possibly expensive A, so require an efficient presentation for AP^h_H, or an A_H that approximates AP^h_H.

7.2 NONSMOOTH FUNCTIONS

The construction of the coarse functions F^k_H and $(G^k_H)^*$ that satisfy Assumption 3.2 can be carried out using polar cone indicators. For a set $A \subset X$, we recall that the polar cone $A^\circ := \{z \in X \mid \langle z, x\rangle \leq 0 \text{ for all } x \in X\}$ and the bipolar cone $A^{\circ\circ} := (A^\circ)^\circ$. We have $A^{\circ\circ} \supset A$ with equality if A is non-empty, convex, and closed [9, Theorem 1.8].

Lemma 7.4. *Take $\zeta^{k,0} = P^h_H x^k_+$, where $x^k_+ \in \operatorname{dom} F$, and F is convex, proper, and lower semicontinuous. Then $P^H_h \partial F(x^k_+) \subset \partial F^k_H(\zeta^{k,0})$ for $F^k_H = \delta_{\Omega^k}$ with $\Omega^k := \zeta^{k,0} + (P^H_h \partial F(x^k_+))^\circ$.*

Proof. Indeed, Ω^k is clearly nonempty and convex. Furthermore, since the subdifferential of the indicator function is the normal cone, $\partial F^k_H(\zeta^{k,0}) = N_{\Omega^k}(\zeta^{k,0}) = (P^H_h \partial F(x^k_+))^{\circ\circ} \supset (P^H_h \partial F(x^k_+))^\circ$. □

More details on the construction can be found in [14] for $G^* = \delta_{B(0,\alpha)^n}$, where $B(0,\alpha)$ is the Euclidean ball. This arises when $G(K\,\cdot)$ models total variation.

8 NUMERICAL EXPERIMENTS

We now compare the two PDMCC variants against the PDPS on three inverse imaging problems: magnetic resonance imaging (MRI), positron emission tomography (PET), and electrical impedance tomography (EIT), all with total variation (TV) regularization. All involve expensive-to-evaluate operators. The EIT problem is nonconvex. These problems share the structure

$$\min_{x \in X} J(x) := F(x) + E(x) + \alpha\|\nabla_h x\|_{2,1} \tag{8.1}$$

for a regularization parameter $\alpha > 0$. Before treating the specifics of each problem in Sections 8.3 to 8.5, first, in Section 8.1, we formulate the trigger condition that we use to pass to the coarse grid in Algorithm 1.1. Then, in Section 8.2, we discuss general numerical setup, shared by our example problem. We finish the paper with our conclusions from the experiments in Section 8.6.

Our software implementation is available on Zenodo [15]. The EIT component is based on [27, 17].

8.1 THE TRIGGER CONDITION

There are different ways to formulate the trigger condition of Algorithm 1.1. Some are discussed in [22]. Theoretically, there are no restrictions on the trigger condition, due to the fine-grid pre-iterate u^k_+: If there is no descent on the coarse grid, θ, such as when, $v^{k,0}$ already solves the coarse problem, Line 10 of the algorithm reduces to a standard fine-grid PDPS step.

We use a *switching strategy*, maintaining a *counter* $k_{\text{switch}} \in \mathbb{N}$ and a *switch* that can be set to *coarse-priority* or *fine-priority*, starting with coarse-priority. The trigger condition is satisfied if, for a parameter $n \in \mathbb{N}$ (100 in our experiments):

(a) The switch is on fine-priority, and $k \in k_{\text{switch}} + n\mathbb{N}$, or
(b) The switch is on coarse-priority, and $k \notin k_{\text{switch}} + n\mathbb{N}$.

After a coarse correction, we flip the switch and update $k_{\text{switch}} := k$ if :

(i) If $\|d^k\|_U \geq \eta$ for a parameter $\eta > 0$, and the switch is on fine-priority, or
(ii) if $\|d^k\|_U < \eta$, and the switch is on coarse-priority.

8.2 GENERAL SETUP

Our primal variable x or ζ generally represent a two-dimensional image on a domain Ω, and the dual variable y or ξ a corresponding vector field. For MRI and PET, they are the nodal values of a finite differences scheme, and lie in $X := \mathbb{R}^n$ (primal, fine), $X_H := \mathbb{R}^N$ (primal, coarse), $Y := \mathbb{R}^{2\times n}$ (dual, fine) and $Y_H := \mathbb{R}^{2\times N}$ (dual, coarse), respectively, where $N < n$. For EIT, the primal variables represent the nodal values of piecewise linear continuous finite elements (P_1), while the dual variables represent the elementwise values of piecewise constant functions on the same mesh.

To compare algorithm performance, we use the relative performance measure

$$\text{RP}_k := \begin{cases} \mathscr{G}_L(u^k;\hat{u})/\mathscr{G}_L(u^0;\hat{u}), & \text{for the convex MRI and PET problems,} \\ J(x^k)/J(x^0), & \text{for the nonconvex EIT problem.} \end{cases} \tag{8.2}$$

Here the Lagrangian gap is defined in (2.8), and $\hat{u} = (\hat{x}, \hat{y})$ is estimated by taking 500000 iterations of the PDMCC using the parameter settings specified for each experiment in Sections 8.3 and 8.4. For EIT, we use the relative primal objective value since the Lagrangian gap is not an appropriate performance measure in the nonconvex setting (see Remark 6.9). We use $\eta = 10^{-5}$ as the tolerance of the trigger condition of Section 8.1. Finally, with $a_v = 10^{-2}$ and $a_r = 10^{-4}$, we choose slowly decaying summable sequences satisfying Assumption 6.4:

$$\varepsilon_k := \frac{C\min\{\tau^{-2},\sigma^{-2}\}}{(1+a_v k)^{1.1}}, \quad \text{and} \quad \rho_k := \frac{\|d^k\|_M^2 + \|d^k\|_W^2}{(1+a_r k)^{1.08}}.$$

In our reports, the *iteration comparison number* scales the number of coarse-grid iterations proportionally to the ratio of the numbers of nodes in the fine and coarse grids. This provides a rough measure of the computational effort, allowing the comparison of the basic PDPS against the PDMCC. Additionally, we report the CPU time.

8.3 MRI

In MRI, the observables are Fourier transforms $\mathscr{F}$ of a two-dimensional image x. Assuming complex Gaussian noise, and sampling this transform t times with different subsampling masks presented by the operators $S_p \in \mathbb{L}(\mathbb{C}^n; C^{m_p})$, $(p = 1, \dots, t)$, we express the reconstruction problem for the data $b_p \in \mathbb{C}^{m_p}$ as (8.1) with

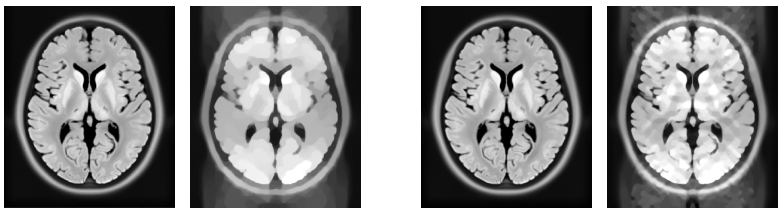
$$F \equiv 0, \quad \text{and} \quad E(x) = \frac{1}{2}\sum_{p=1}^{t} \|S_p\mathscr{F}x - b_p\|^2_{\mathbb{C}^{m_p}},$$

Let $\operatorname{Sym} S$ denote the symmetrisation of $S := \sum_{p=1}^t S_p^* S_p$ over positive and negative frequencies on both axes. To avoid complex numbers, we can then rewrite [14]

$$E(x) = \frac{1}{2}\langle Tx, x\rangle_{\mathbb{R}^n} - \langle x, e\rangle_{\mathbb{R}^n} \text{ for } T = \mathscr{F}^* \operatorname{Sym} S\, \mathscr{F} \text{ and } e = \sum_{p=1}^{t} \operatorname{Re} \mathscr{F}^* S_p^* b_p.$$

Table 8.1: CPU time (s) to reach $\mathrm{RP}_k = 10^{-4}$ and $\mathrm{RP}_k = 10^{-5}$ for the MRI problem.

MRI	$\mathrm{RP}_k = 10^{-4}$			$\mathrm{RP}_k = 10^{-5}$		
Resolution	PDPS	PDMCC-E	PDMCC-NE	PDPS	PDMCC-E	PDMCC-NE
583×493	$2.59 \cdot 10^2$	$1.88 \cdot 10^1$	$2.99 \cdot 10^1$	$7.39 \cdot 10^2$	$1.23 \cdot 10^2$	$1.42 \cdot 10^2$
2048×1732	$6.75 \cdot 10^3$	$4.02 \cdot 10^2$	$5.42 \cdot 10^2$	$2.02 \cdot 10^4$	$1.82 \cdot 10^3$	$1.81 \cdot 10^3$



(a) Original (b) Solution (c) Original (d) Solution

Figure 8.1: MRI ground truth and reconstructions at relative Lagrangian gap $\mathrm{RP}_k = 10^{-5}$ for image resolutions 583×493 (left group) and 2048×1732 (right group).

We use standard multigrid transfer operators, consistent with the forward-difference structure of ∇_h on a rectangular grid. The primal restriction operator is $P_h^H := \mathscr{R} \otimes \mathscr{R}$, with stencil $\mathscr{R} = \left(\frac{1}{2} \quad 1 \quad \frac{1}{2}\right)$ (see, e.g., [4]), and the dual restriction operator is given by $D_h^H y := [P_h^H y_1, P_h^H y_2]^\top$.

E has the structure of Example 7.3 with $A = (\operatorname{Sym} S)^{1/2} \mathscr{F}$. We take E_H following that example. The ideal coarse operator $A_H = (\operatorname{Sym} S_H)^{1/2} \mathscr{F}_H$ satisfies $A P_h^H = A_H$, but is computationally expensive. We take $A_H = \mathscr{F}_H$. Then $A_H^* A_H = \mathrm{Id}$, making the valuation of ∇E_H on each coarse iteration very cheap. It has Lipschitz factor $L_H = 1$.

As our ground truth image $\hat{x}$, we use the phantom of [2] with resolutions 583×493 and 2048×1732. The subsampling masks S_p are formed by uniform sampling of horizontal lines of the Fourier transform. We add complex Gaussian noise to $S_p \mathscr{F} \hat{x}$ to form the data b_p. The noise levels, number of lines per mask, the value of the Lipschitz constant for the fine grid, and the PDMCC parameters are as follows:

				Ergodic			Non-ergodic	
Resolution	$L = \|\operatorname{Sym} S\|_\infty$	t	ν	Lines/mask	m	(τ_0, σ_0)	m	(τ_0, σ_0)
583×493	4	8	50	100	4	$(0.05,\ 0.85)$	3	$(0.05,\ 0.75)$
2048×1732	5	15	100	200	7	$(0.10,\ 0.85)$	4	$(0.10,\ 0.75)$

We have $\|\nabla_h\|_{\mathbb{L}(X)}, \|\nabla_H\|_{\mathbb{L}(X)} \leq \sqrt{8}$ [5]. We take as the step length parameters $\tau = 0.9/L$, $\sigma = 0.09/(8\tau)$ on the fine grid, and $\tau_H = \tau_0/L$, $\sigma_H = \sigma_0(1 - \tau_0)/(8\tau_H)$ on the coarse grid, with $\tau_0, \sigma_0 \in (0, 1)$ as in the table. We set $\alpha = 0.8$.

The reconstructions are shown in Figure 8.1, while performance is reported in Figure 8.2 and Table 8.1. In the performance plots, the iteration count is scaled by the ratio between the numbers of coarse- and fine-grid pixels.

8.4 PET

In PET [20], a measurement device detects pairs of gamma photons emanating from positron-electron annihilation. Mathematically, this process is described by the Radon transform, whereas the noise follows the Poisson distribution. More precisely, denoting the partial discrete Radon transform by

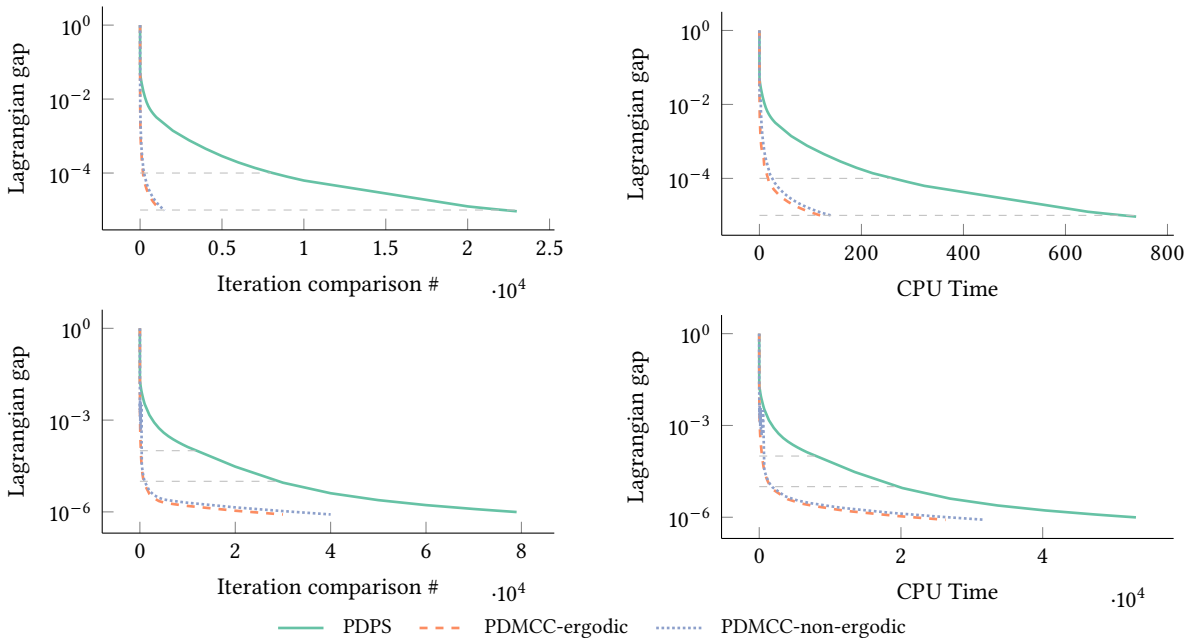


Figure 8.2: Relative performance measure (8.2) versus iterations and CPU time for the MRI problem across both resolutions (583×493 in the top row and 2048×1732 in the bottom row). Dashed gray lines correspond to the levels reported in Table 8.1.

$A \in \mathbb{R}^{t\times n}$, then for every sampling index $p = 1,\dots,t$, the measurement $b_p \sim \text{Poisson}\big((Ax+c)_p\big)$, where $c \in \mathbb{R}^t_+$ is a noise parameter. The data vector $b = \big(b_1,\dots,b_t\big)^\top \in \mathbb{R}^t$ is commonly known as the "sinogram". In the reconstruction problem (8.1) we, thus, take

$$F(x) := \delta_{[0,+\infty)^n}(x) \quad \text{and} \quad E(x) := \langle \mathbf{1}_t, Ax\rangle - \langle b, \log(Ax+c)\rangle,$$

where the logarithm is to be understood componentwise.

The transfer and discrete gradient operators are the same as in Section 8.3. On the coarse grid, we construct E_H following Example 7.2. Then $L_H = 0$. We use Lemma 7.4 to construct $F_H^k := \delta_{\prod_{i=1}^N \Omega_i^k}$. Denoting by $A_i \subset \{1,\dots,n\}$ the subset of fine-grid pixel indices j that contribute to the coarse pixel i, i.e., $[P_h^H]_{ij} \neq 0$, it gives

$$\Omega_i^k = \zeta_i^{k,0} + [P_h^H \partial F(x_+^k)]_i^\circ = \begin{cases} \mathbb{R}, & \text{if } [x_+^k]_l > 0 \text{ for all } l \in A_i, \\ [\zeta_i^{k,0}, \infty) & \text{if there exist } l \in A_i \text{ such that } [x_+^k]_l = 0. \end{cases}$$

As our ground-truth image $\hat{x}$, we take the Shepp-Logan phantom [24] at the resolutions 512×512 and 1024×1024. We add Poisson noise of parameter $\nu = 0.1$ to the sinogram $A\hat{x}$. As the regularization parameter we take $\alpha = 0.7$. The fine-grid step length parameters for both PDMCC and PDPS are taken as in the MRI experiments in Section 8.3, for the Lipschitz factor estimate $L = \|e \oslash c^2\|_\infty \sqrt{n_x^2 + n_y^2}$. For the coarse problem, we set $\tau_H = \tau_0$, and $\sigma_H = \sigma_0/(8\tau_H)$, where $\tau_0, \sigma_0 \in (0,1)$. The dimensions of the sinogram and the PDMCC parameters are as follows:

		Ergodic		Non-ergodic	
Resolution	Sinogram dimensions	m	(τ_0, σ_0)	m	(τ_0, σ_0)
512×512	256×128	4	$(0.999,\ 0.99)$	7	$(0.999,\ 0.9)$
1024×1024	512×384	2	$(0.37,\ 0.9)$	7	$(0.999,\ 0.9)$

Table 8.2: CPU time (s) to reach $\mathrm{RP}_k = 10^{-4}$ and $\mathrm{RP}_k = 10^{-6}$ for the PET problem.

PET	$\mathrm{RP}_k = 10^{-4}$			$\mathrm{RP}_k = 10^{-6}$		
Resolution	PDPS	PDMCC-E	PDMCC-NE	PDPS	PDMCC-E	PDMCC-NE
512×512	$3.13 \cdot 10^3$	$1.19 \cdot 10^2$	$5.94 \cdot 10^1$	$3.80 \cdot 10^4$	$1.27 \cdot 10^4$	$1.26 \cdot 10^4$
1024×1024	$7.63 \cdot 10^4$	$7.41 \cdot 10^2$	$5.13 \cdot 10^2$	$6.45 \cdot 10^5$	$2.89 \cdot 10^5$	$2.81 \cdot 10^5$

(a) Original (b) Solution (c) Original (d) Solution

Figure 8.3: PET ground truth and reconstructions at relative Lagrangian gap $\mathrm{RP}_k = 10^{-6}$ for image resolutions 512×512 (left group) and 1024×1024 (right group).

The observed data and reconstructions are in Figure 8.3, while we report performance in Figure 8.4 and Table 8.2. In the performance plots, the iteration count is scaled by the ratio between the number of coarse- and fine-grid pixels.

8.5 EIT

We now take in (8.1)

$$E(x) := \frac{1}{2}\sum_{i=1}^{d} \|I_i(x) - \mathcal{I}_i\|^2, \quad F(x) = \delta_{[x_{\min}, x_{\max}]^n}(x),$$

where, for a given conductivity $x \in L^2(\Omega)$, $I_i(x) \in \mathbb{R}^d$ are simulated electrical currents at $d \in \mathbb{N}$ electrodes on the boundary of the domain Ω, when the same electodes are excited with the electrical potentials $U_i \in \mathbb{R}^d$. The measured currents are $\mathcal{I}_i \in \mathbb{R}^d$. Multiple measurements $i = 1, \ldots, N$ are made. The relationship is governed by the *Complete Electrode Model* (CEM) partial differential equation (PDE), [7]. For further details on our specific approach, see [12].

We work in the circular domain $\Omega = B(0, r)$ with $r = 15$, equipped with $d = 16$ equally spaced boundary electrodes. The coarse mesh $\mathcal{T}_H$ has 4897 nodes and 9024 elements, while the fine mesh $\mathcal{T}_h$ is obtained by uniform refinement, yielding 18817 nodes and 36096 elements. We define the primal prolongation $P_H^h : X_H \hookrightarrow X_h$ as the canonical inclusion, and the dual restriction $D_h^H : Y_h \to Y_H$ by as the average over the four fine elements contained in each coarse element, i.e., in stencil notation, $D_h^H = \frac{1}{4}\begin{pmatrix} 1 & 1 & 1 & 1 \end{pmatrix}$. We set $K = \mathcal{M}_h \nabla_h$ and $K_H = \mathcal{M}_H \nabla_H$, where $\mathcal{M}_h$ and $\mathcal{M}_H$ denote the corresponding dual mass matrices. As in Section 8.4, we construct the coarse objective E_H according to Example 7.1. The construction of F_H is also similar to Section 8.4. The Lipschitz constant of ∇E cannot be computed explicitly [12], unlike in Sections 8.3 and 8.4. Based on dynamic estimation from initial experiments, we use $L \approx 18.2036$ and $\|K\| \approx 0.03041$. This yields $\tau = 0.675/L \approx 0.03708$ y $\sigma = 0.07/(\|K\|^2\tau) \approx 2041.452$. Meanwhile, both the ergodic and non-ergodic variants perform $m = 5$ coarse iterations with $\tau_H = 0.12$ and $\sigma_H = 0.9/(\|K_H\|^2\tau_H)$ for $\|K_H\| \approx 0.2416$.

The observed data and reconstructed images are shown in Figure 8.5, while performance comparisons are presented in Figure 8.6 and Table 8.3.

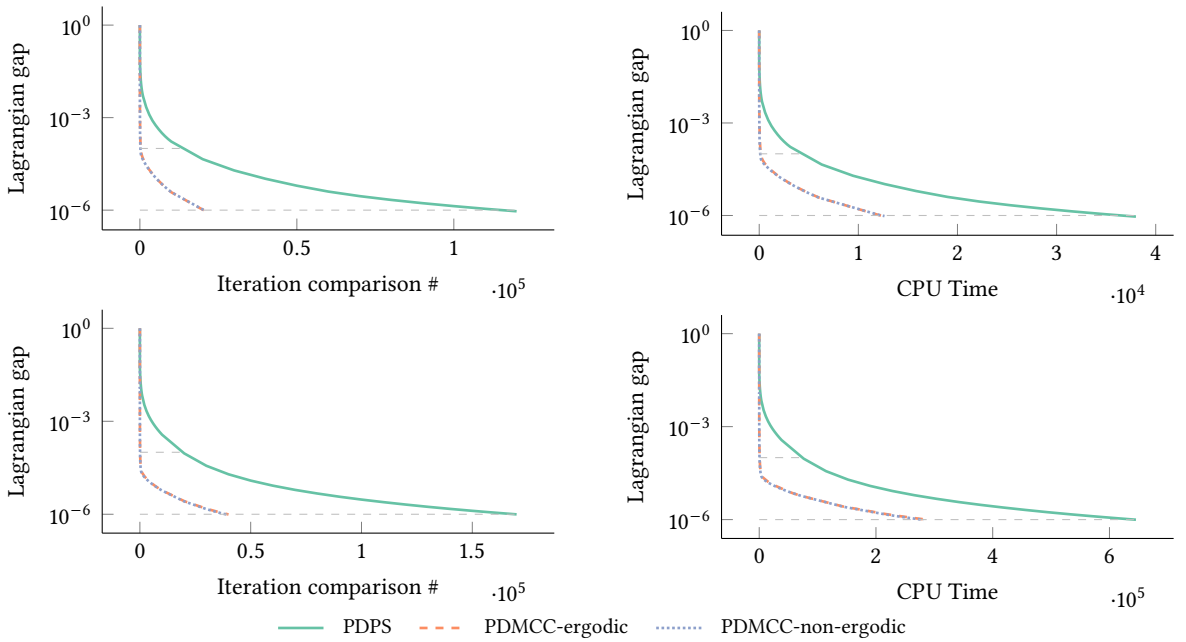


Figure 8.4: Relative performance measure (8.2) versus iterations and CPU time for the PET problem across both resolutions (512×512 in the top row and 1024×1024 in the bottom row). Dashed gray lines correspond to the levels reported in Table 8.2.

Table 8.3: CPU time (s) to reach $\mathrm{RP}_k = 3 \cdot 10^{-4}$ and $\mathrm{RP}_k = 5 \cdot 10^{-5}$ for EIT.

EIT	$\mathrm{RP}_k = 3 \cdot 10^{-4}$			$\mathrm{RP}_k = 5 \cdot 10^{-5}$		
Fine nodes	PDPS	PDMCC-E	PDMCC-NE	PDPS	PDMCC-E	PDMCC-NE
18817	$2.49 \cdot 10^4$	$5.07 \cdot 10^3$	$5.07 \cdot 10^3$	$8.41 \cdot 10^4$	$1.88 \cdot 10^4$	$1.95 \cdot 10^4$

8.6 CONCLUSIONS

Figures 8.2, 8.4 and 8.6 illustrate the convergence behavior of PDMCC, which is consistent with the $O(1/N)$ convergence rate established in Corollary 6.6. The same behavior is observed for PDPS, in agreement with Theorem 2.3. Furthermore, Tables 8.1 to 8.3 show that the proposed PDMCC variants require substantially less CPU time than PDPS to reach the same reference value of RP_k. Specifically, the reduction ranges from 81% to 93% for MRI, from 67% to 98% for PET, and from 76% to 79% for EIT. These results confirm the theoretical results presented in Section 6, indeed, show much faster convergence than that of the reference PDPS. In conclusion, our proposed method appears to provide the advantages that multigrid methods generally have. In future research, it would be desirable to produce a more theoretical analysis of the reduced computational cost.

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.

(a) Original (b) Solution

Figure 8.5: EIT ground-truth and reconstruction at relative primal objective function $\mathrm{RP}_k = 4 \cdot 10^{-5}$ for 18817 node grid.

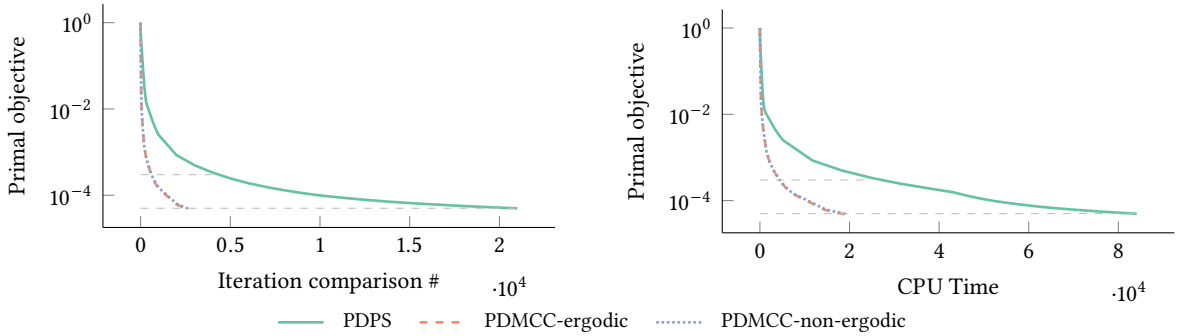


Figure 8.6: Relative performance measure (8.2) versus iterations and CPU time for the EIT problem. Dashed gray lines correspond to the levels reported in Table 8.3.

[3] A. Brandt, Multi-Level Adaptive Solutions to Boundary-Value Problems, *Mathematics of Computation* 31 (1977), 333–390, doi:10.1090/s0025-5718-1977-0431719-x.

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

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

[6] A. Chambolle and T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, *Journal of mathematical imaging and vision* 40 (2011), 120–145.

[7] K. S. Cheng, D. Isaacson, J. C. Newell, and D. G. Gisser, Electrode models for electric current computed tomography, *IEEE Transactions on Biomedical Engineering* 36 (1989), 918–924.

[8] C. Clason and T. Valkonen, Primal-dual extragradient methods for nonlinear nonsmooth PDE-constrained optimization, *SIAM J. Optim.* 27 (2017), 1313–1339, doi:10.1137/16m1080859, arXiv:1606.06219.

[9] C. Clason and T. Valkonen, *Introduction to Nonsmooth Analysis and Optimization*, MOS-SIAM Series on Optimization, SIAM, 2026, doi:10.1137/1.9781611978995.

[10] L. Condat, A Primal–Dual Splitting Method for Convex Optimization Involving Lipschitzian, Proximable and Linear Composite Terms, *J. Optim. Theory Appl.* 158 (2013), 460–479, doi:10.1007/s10957-012-0245-9.

[11] N. Dizon, J. Jauhiainen, and T. Valkonen, Online optimisation for dynamic electrical impedance tomography, *Inv. Prob.* 41 (2025), 055005, doi:10.1088/1361-6420/adcb66, arXiv:2412.12944.

[12] N. Dizon and T. Valkonen, Differential estimates for fast first-order multilevel nonconvex optimisation, 2024, arXiv:2412.01481. submitted.

[13] D. Gabay, Chapter ix applications of the method of multipliers to variational inequalities, in *Studies in mathematics and its applications*, volume 15, Elsevier, 1983, 299–331.

[14] F. Guerra and T. Valkonen, Multigrid methods for total variation, in *International Conference on Scale Space and Variational Methods in Computer Vision*, 2025, 3–16.

[15] F. Guerra and T. Valkonen, Multigrid methods for nonsmooth optimization, 2026, doi:10.5281/zenodo.21731057. Software on Zenodo.

[16] B. He, Y. You, and X. Yuan, On the Convergence of Primal-Dual Hybrid Gradient Algorithm, *SIAM J. Imaging Sci.* 7 (2014), 2526–2537, doi:10.1137/140963467.

[17] J. Jauhiainen, N. Dizon, T. Valkonen, and Y. Nabou, Online Optimisation Codes for Dynamic Electrical Impedance Tomography, 2026, doi:10.5281/zenodo.19154746. Software.

[18] Y. Malitsky and T. Pock, A first-order primal-dual algorithm with linesearch, *SIAM Journal on Optimization* 28 (2018), 411–432.

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

[20] F. Natterer, *The mathematics of computerized tomography*, SIAM, 2001.

[21] J. Nocedal and S. Wright, *Numerical optimization*, Springer Science & Business Media, 2006.

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

[23] T. Pock, D. Cremers, H. Bischof, and A. Chambolle, An algorithm for minimizing the Mumford-Shah functional, in *2009 IEEE 12th International Conference on Computer Vision*, 2009, 1133–1140.

[24] L. A. Shepp and B. F. Logan, The Fourier Reconstruction of a Head Section, *IEEE Transactions on Nuclear Science* NS-21 (1974), 21–43.

[25] T. Valkonen, Preconditioned ADMM with nonlinear operator constraint, in *System Modeling and Optimization: 27th IFIP TC 7 Conference, CSMO 2015, Sophia Antipolis, France, June 29-July 3, 2015, Revised Selected Papers*, volume 494, 2017, 117.

[26] T. Valkonen, Testing and non-linear preconditioning of the proximal point method, *Appl. Math. Optim.* 82 (2020), doi:10.1007/s00245-018-9541-6.

[27] T. Valkonen, Codes for Differential Estimates for Fast First-Order Multilevel Nonconvex Optimisation, 2026, doi:10.5281/zenodo.19154665. Software on Zenodo.

[28] B. C. Vũ, A splitting algorithm for dual monotone inclusions involving cocoercive operators, *Advances in Computational Mathematics* 38 (2013), 667–681, doi:10.1007/s10444-011-9254-8.