PRIMAL-DUAL BLOCK-PROXIMAL SPLITTING FOR A CLASS OF NON-CONVEX PROBLEMS

Stanislav Mazurenko[*] Jyrki Jauhiainen[†] Tuomo Valkonen[‡]

Abstract We develop block structure adapted primal-dual algorithms for non-convex non-smooth optimisation problems whose objectives can be written as compositions $G(x) + F(K(x))$ of non-smooth block-separable convex functions G and F with a non-linear Lipschitz-differentiable operator K. Our methods are refinements of the non-linear primal-dual proximal splitting method for such problems without the block structure, which itself is based on the primal-dual proximal splitting method of Chambolle and Pock for convex problems. We propose individual step length parameters and acceleration rules for each of the primal and dual blocks of the problem. This allows them to convergence faster by adapting to the structure of the problem. For the squared distance of the iterates to a critical point, we show local $O(1/N)$, $O(1/N^2)$ and linear rates under varying conditions and choices of the step lengths parameters. Finally, we demonstrate the performance of the methods on practical inverse problems: diffusion tensor imaging and electrical impedance tomography.

1 INTRODUCTION

We want to solve in Hilbert spaces X and Y the problem

$$\min_{x \in X} G(x) + F(K(x)), \tag{P$_0$}$$

where $G : X \to \overline{\mathbb{R}}$ and $F : Y \to \overline{\mathbb{R}}$ are convex, proper, and lower semicontinuous, but $K \in C^1(X; Y)$ is possibly non-linear. The linear case has been considered frequently in the literature, while in our earlier work [33, 11, 9] we have developed first-order primal-dual methods for the generally non-convex problem with a non-linear K. We refer to [36] for a simplified overview of such methods. In the present work, still with a non-linear K, we consider problems of the more specific form

$$\min_{x \in X} \sum_{j=1}^{m} G_j(P_j x) + \sum_{\ell=1}^{n} F_\ell(Q_\ell K(x)), \tag{P}$$

where for all $j = 1, \ldots, m$ and $\ell = 1, \ldots, n$, the functions $G_j : X \to \overline{\mathbb{R}}$ and $F_\ell : Y \to \overline{\mathbb{R}}$ are convex, proper, and lower semicontinuous, and $P_1, \ldots, P_m \in \mathbb{L}(X; X)$ as well as $Q_1, \ldots, Q_n \in \mathbb{L}(Y; Y)$ are mutually

This work was started when T. Valkonen and S. Mazurenko were at the University of Liverpool. There their work was supported by the EPSRC First Grant EP/P021298/1. More recently, T. Valkonen has been supported by the Academy of Finland grants 314701 and 320022 as well as the Escuela Politécnica Nacional internal grant PIJ-18-03. S. Mazurenko has been supported by the Operational Programme Research, Development and Education project MSCAfellow@MUNI (CZ.02.2.69/0.0/0.0/17_050/0008496). J. Jauhiainen has been supported by the European Union Horizon 2020 research and innovation programme under grant agreement No 764810.

[*]Loschmidt Laboratories, Masaryk University, Brno, Czechia (stan.mazurenko@gmail.com, ORCID: 0000-0003-3659-4819)

[†]University of Eastern Finland, Kuopio, Finland (jyrki.jauhiainen@uef.fi, ORCID: 0000-0001-6711-6997)

[‡]ModeMat, Escuela Politécnica Nacional, Quito, Ecuador *and* Department of Mathematics and Statistics, University of Helsinki, Finland (tuomo.valkonen@iki.fi, ORCID: 0000-0001-6683-3572)

orthogonal families of linear projection operators. In other words, G and F are block-separable. More specifically, we develop spatially adaptive and block-stochastic optimisation methods for the solution of (P).

As observed in [35] for linear K, the adaptation of step lengths to individual blocks j and ℓ can speed up the convergence of optimisation methods due to blockwise Lipschitz or strong convexity factors being better than the global factor. Moreover, as now extensively studied, randomly sampling the blocks to be updated on each step can also improve convergence on very large-scale problems, in part due to the spatial adaptation, and in part due to being able to avoid communication in a cluster implementation of the algorithm. For more on stochastic block coordinate descent type methods, we refer to the review [42] and, among others, the original articles [23, 28, 16, 29, 46, 31, 13, 25, 2] on forward–backward type methods, [8, 32, 45, 12, 4, 6, 15, 35] on primal-dual methods, and [27, 26] on second-order methods, all in the convex case. For the non-convex case we point to [43, 44]. Compared to the latter, we work in the primal-dual setting and aim for spatial adaptation also in the deterministic setting. We also aim to prove convergence rates.

Several works consider, instead of a random selection of blocks, a random selection of terms of a sum of functions. In the non-convex case, recent mathematical works in this area include [14, 22], aside from more applied works in the area of neural networks. In our block-stochastic approach, for non-convex C^1 functions J_ℓ, $(\ell = 1, \ldots, n)$, we can with $K(x) := (J_1(x), \ldots, J_n(x))$ and $F(z) := \sum_{\ell=1}^n z_\ell$ write

$$\min_x G(x) + \sum_{\ell=1}^n J_\ell(x) = \min_x G(x) + F(K(x)). \tag{1.1}$$

To start describing our approach, using the conjugates F_ℓ^* of the convex, proper, lower semicontinuous functions F_ℓ, we reformulate (P) as the minmax problem

$$\min_{x\in X} \max_{y\in Y} \sum_{j=1}^m G_j(P_j x) + \langle K(x), y\rangle - \sum_{\ell=1}^n F_\ell^*(Q_\ell y). \tag{S}$$

If K is linear, and the number of blocks $n = m = 1$, a popular algorithm for solving this formulation is the primal-dual proximal splitting (PDPS) of Chambolle and Pock [7]. It consists of alternating proximal steps with respect to the dual and primal variables, with the other variable fixed, and an over-relaxation step that ensures convergence. Its extension to non-linear K (but still without blockwise structure) iterates [33, 9]

$$\begin{cases} x^{i+1} := \operatorname{prox}_{\tau_i G}(x^i - \tau_i \nabla K(x^i)^* y^i), \\ \bar{x}^{i+1} := x^{i+1} + \omega_i (x^{i+1} - x^i), \\ y^{i+1} := \operatorname{prox}_{\sigma_{i+1} F^*}(y^i + \sigma_{i+1} K(\bar{x}^{i+1})) \end{cases}$$

for some step length and over-relaxation parameters τ_i, σ_{i+1}, ω_i and $\operatorname{prox}_{\tau_i G}(x) := (I + \tau_i \partial G)^{-1}(x)$. Our purpose in this work is to randomise and adapt the method to the multi-block structure of (S): *firstly*, on each step we will only update random subsets of either or both primal and dual blocks, and, *secondly*, even when we deterministically update every block on each step, we adapt the step lengths to the local structure of the problem in each block.

We organise our work as follows: first, in Section 2, we introduce general notations, concepts, and the rough structure of the algorithm. In Section 3 we start the convergence proof by deriving several technical estimates. In Section 4 we then use these estimates to derive convergence rates of more specific algorithms when only the primal updates are randomised. Likewise, in Section 5 we study the case when only the dual updates are randomised. We finish our work in Section 6 with numerical experience in diffusion tensor imaging (DTI) and electrical impedance tomography (EIT).

2 NOTATIONS, ROUGH ALGORITHM, AND ITS TESTING

Throughout this paper, we write $\mathbb{L}(X;Y)$ for the space of bounded linear operators between Hilbert spaces X and Y; I is the identity operator; and $\langle x, x'\rangle$ is the inner product in the corresponding space. We write with $\mathcal{P}A$ for the power set of a set A and $\chi_A(a)$ for the indicator function that equals 1 if $a \in A$ and 0 otherwise. We set $\langle x, x'\rangle_T := \langle Tx, x'\rangle$, and $\|x\|_T := \sqrt{\langle x, x\rangle_T}$, where in the latter we require $T \geq 0$. For $T, S \in \mathbb{L}(X;Y)$, the inequality $T \geq S$ means that $T - S$ is positive semidefinite. If H is a set-valued operator $X \rightrightarrows X$, inequalities such as $\langle H(x), x'\rangle \geq 0$ mean that $\langle w, x'\rangle \geq 0$ for every $w \in H(x)$.

We write $(\Omega, \mathcal{O}, \mathbb{P})$ for the probability space consisting of a sample set Ω, a σ-algebra $\mathcal{O}$ on Ω, and a probability measure $\mathbb{P}$. We write $\mathcal{R}(\mathcal{O}; V)$ for the space of V-valued $\mathcal{O}$-measurable random variables. $\mathcal{R}(\mathcal{O}; U \rightrightarrows U)$ is therefore the space of $\mathcal{O}$-measurable random variables whose values are set-valued operators $U \rightrightarrows U$. Due to the iterative nature of optimisation algorithms, we introduce a sequence of σ-algebras $\{\mathcal{O}_i\}_{i\in\mathbb{N}}$ such that $\mathcal{O}_i \subseteq \mathcal{O}_{i+1}$ and $\mathcal{O}_i \subseteq \mathcal{O}$ for any $i \in \mathbb{N}$. We use $\mathcal{O}_i$ to collect all the information available before the $(i+1)$:th iteration. We write $\mathbb{E}_i[\,\cdot\,] := \mathbb{E}[\,\cdot \mid \mathcal{O}_i]$ for the corresponding conditional expectation.

Many conditions that we impose in the following sections only apply to the subspace on which the operator K from the introduction acts non-linearly. Correspondingly, we introduce

$$Y_{\mathrm{L}} := \{y \in Y \mid \text{the map } x \mapsto \langle y, K(x)\rangle \text{ is linear}\} \quad \text{and} \quad Y_{\mathrm{NL}} := Y_{\mathrm{L}}^{\perp},$$

as well as the orthogonal projection P_{NL} to Y_{NL}. See Section 6 for how such subspaces practically come about in applications. We also use the short-hand notations

$$x_j := P_j x \quad \text{and} \quad y_\ell := Q_\ell y.$$

2.1 ABSTRACT STRUCTURE OF THE ALGORITHM

We generally use the symbol x for primal variables (elements of X), and symbol y for dual variables (elements of Y). We group these variables together into $u = (x, y) \in X \times Y$. This applies to indexed variables, $u^i := (x^i, y^i)$, critical points $\widehat{u} = (\widehat{x}, \widehat{y})$, etc., without explicit introduction of the primal and dual components in each case. We define the set-valued operator $H : X \times Y \rightrightarrows X \times Y$ for $u = (x, y)$ as

$$H(u) := \begin{pmatrix} \partial G(x) + \nabla K(x)^* y \\ \partial F^*(y) - K(x) \end{pmatrix} \quad \text{with} \quad G(x) := \sum_{j=1}^m G_j(P_j x) \quad \text{and} \quad F^*(y) := \sum_{\ell=1}^n F_\ell^*(Q_\ell y). \tag{2.1}$$

Then $0 \in H(\widehat{u})$ encodes the critical point conditions for (S). These will also become the first-order necessary optimality conditions under a constraint qualification, e.g., when G is C^1 and either the null space of $\nabla K(x)^*$ is trivial or $\operatorname{dom} F = X$ [30, Example 10.8].

Following the “testing” approach to convergence analysis from [34], we introduce the primal-dual *step length*, *testing*, and *preconditioning* operators

$$W_{i+1} := \begin{pmatrix} T_i & 0 \\ 0 & \Sigma_{i+1} \end{pmatrix}, \quad Z_{i+1} := \begin{pmatrix} \Phi_i & 0 \\ 0 & \Psi_{i+1} \end{pmatrix}, \quad \text{and} \quad M_{i+1} := \begin{pmatrix} I & -\Phi_i^{-1}\Lambda_i^* \\ -\Psi_{i+1}^{-1}\Lambda_i & I \end{pmatrix}. \tag{2.2}$$

Here T_i, Φ_i and Σ_{i+1}, Ψ_{i+1} are the respective primal and dual step length and testing operators, and Λ_i is a term that we will develop to suitably decouple the updates of the primal and dual variables. In the deterministic case, $T_i, \Phi_i \in \mathbb{L}(X;X)$ and $\Sigma_{i+1}, \Psi_{i+1} \in \mathbb{L}(Y;Y)$ as well as $\Lambda_i \in \mathbb{L}(X;Y)$. Clearly, $Z_{i+1}M_{i+1}$ is self-adjoint. For the stochastic setting we will impose our formal assumptions later in (3.17). We will in particular require the tests Φ_i and Ψ_{i+1} to already be known before the start of the i:th iteration (calculating u^i), whereas the step lengths themselves will have to be known before the $(i+1)$:th iteration (calculating u^{i+1}).

Finally, we write our proposed algorithm in the implicit form

$$0 \in W_{i+1}\widetilde{H}_{i+1}(u^{i+1}) + M_{i+1}(u^{i+1} - u^i) \tag{PP}$$

for

$$\widetilde{H}_{i+1}(u^{i+1}) := H(u^{i+1}) + \begin{pmatrix} [\nabla K(x^i) - \nabla K(x^{i+1})]^* y^{i+1} \\ K(x^{i+1}) - K(x^{i+1} + \Omega^i(x^{i+1} - x^i)) + \nabla K(x^i)\Omega^i(x^{i+1} - x^i) \end{pmatrix} \tag{2.3}$$

and some over-relaxation operator Ω_i, which in the deterministic setting is in $\mathbb{L}(X;X)$. Here $\widetilde{H}_{i+1}(u)$ is a partial linearization of $H(u)$ similar to [33]. It simplifies to $H(u)$ for a linear K. In the following, by specifying the testing, step length, preconditioning, and over-relaxation operator, we develop more explicit methods from this implicit formulation, which itself is more amenable to convergence analysis.

2.2 TESTING FOR CONVERGENCE

The proximal point method iteratively solves u^{i+1} from

$$0 \in H(u^{i+1}) + \tau^{-1}(u^{i+1} - u^i) \tag{2.4}$$

given a step length parameter $\tau > 0$. If H is a γ-strongly monotone operator and $\widehat{u} \in H^{-1}(0)$. Then $\langle H(u^{i+1}), u^{i+1} - \widehat{u}\rangle \geq \gamma\|u^{i+1} - \widehat{u}\|^2$. This suggest “testing” (2.4) by the application of $\langle\,\cdot\,, u^{i+1} - \widehat{u}\rangle$. Subsequently to this testing, the strong monotonicity and Pythagoras’ identity

$$\langle u^{i+1} - u^i, u^{i+1} - \widehat{u}\rangle = \frac{1}{2}\|u^{i+1} - u^i\|^2 - \frac{1}{2}\|u^i - \widehat{u}\|^2 + \frac{1}{2}\|u^{i+1} - \widehat{u}\|^2,$$

applied to $0 \in \langle H(u^{i+1}) + \tau^{-1}(u^{i+1} - u^i), u^{i+1} - \widehat{u}\rangle$ yield

$$\frac{1 + 2\gamma\tau}{2}\|u^{i+1} - \widehat{u}\|^2 + \frac{1}{2}\|u^{i+1} - u^i\|^2 \leq \frac{1}{2}\|u^i - \widehat{u}\|^2.$$

Telescoping this inequality, it is clear that $u^N \to \widehat{u}$ at the linear rate $O(1/(1+2\gamma\tau)^N)$. The next theorem from [34] generalises these simple arguments to the more general algorithm (PP) in the stochastic setting.

Theorem 2.1 ([34, Corollary 3.1]). *On a Hilbert space U and a probability space (Ω, O), let $\widetilde{H}_{i+1} : \mathcal{R}(O; U \rightrightarrows U)$, and $M_{i+1}, Z_{i+1} \in \mathcal{R}(O; \mathbb{L}(U;U))$ for $i \in \mathbb{N}$. Suppose (PP) is solvable for $\{u^{i+1}\}_{i\in\mathbb{N}} \subset \mathcal{R}(O;U)$. If for all $i \in \mathbb{N}$ and almost all random events $\omega \in \Omega$, $(Z_{i+1}M_{i+1})(\omega)$ is self-adjoint, and the* expected fundamental condition

$$\mathbb{E}[\langle W_{i+1}\widetilde{H}_{i+1}(u^{i+1}), u^{i+1} - \widehat{u}\rangle_{Z_{i+1}}] \geq \mathbb{E}\left[\frac{1}{2}\|u^{i+1} - \widehat{u}\|^2_{Z_{i+2}M_{i+2} - Z_{i+1}M_{i+1}} - \frac{1}{2}\|u^{i+1} - u^i\|^2_{Z_{i+1}M_{i+1}}\right] \tag{2.5}$$

holds, then so does the expected descent inequality

$$\mathbb{E}\left[\frac{1}{2}\|u^N - \widehat{u}\|^2_{Z_{N+1}M_{N+1}}\right] \leq \mathbb{E}\left[\frac{1}{2}\|u^0 - \widehat{u}\|^2_{Z_1 M_1}\right] \quad (N \geq 1). \tag{2.6}$$

The condition (2.5) is simply a relaxation of the strong monotonicity we assumed above. It also includes the term $\frac{1}{2}\|u^{i+1} - u^i\|_{Z_{i+1}M_{i+1}}$ intended to be used with forward steps. In application to (2.4), we have $M_{i+1} = I$, and we can take as the testing operator $Z_{i+1} = \phi_i I$ with $\phi_{i+1} = (1 + 2\gamma\tau)\phi_i$ and $\phi_0 = 1$. Thus $Z_{N+1}M_{N+1}$ in (2.6) forms a local metric that measures rates of convergence. If we can ensure $Z_{i+1}M_{i+1} \geq \mu_i I$ for some deterministic $\mu_i \nearrow \infty$, then (2.6) shows $\mathbb{E}[\|u^N - \widehat{u}\|^2]$ to converge to zero at the rate $O(1/\mu_N)$. We will in Section 3 develop lower bounds of this kind.

2.3 Blockwise algorithm structure

We now develop a more blockwise-refined structure of our proposed algorithm. Inserting (2.2), we can expand (PP) as the pair of implicit updates (compare [35, §2.3])

$$\begin{cases} x^{i+1} = (I + T_i \partial G)^{-1}(x^i + \Phi_i^{-1}[\Lambda_i^* - \Phi_i T_i \nabla K(x^i)^*](y^{i+1} - y^i) - T_i \nabla K(x^i)^* y^i), \\ y^{i+1} = (I + \Sigma_{i+1}\partial F^*)^{-1}(y^i + \Psi_{i+1}^{-1}[\Lambda_i - \Psi_{i+1}\Sigma_{i+1}\nabla K(x^i)\Omega^i](x^{i+1} - x^i) \\ \qquad\qquad + \Sigma_{i+1} K(x^{i+1} + \Omega^i(x^{i+1} - x^i))). \end{cases} \tag{2.7}$$

Due to the block-separable structure of G and F^* in (2.1), we take for all $i \in \mathbb{N}$,

$$T_i := \sum_{j \in S(i)} \tau_j^i P_j, \qquad \Sigma_{i+1} := \sum_{\ell \in V(i+1)} \sigma_\ell^{i+1} Q_\ell, \qquad \Omega_i := \sum_{j \in S(i)} \omega_j^i P_j, \tag{2.8a}$$

$$\Phi_i := \sum_{j=1}^m \phi_j^i P_j, \qquad \Psi_{i+1} := \sum_{\ell=1}^n \psi_\ell^{i+1} Q_\ell, \quad \text{and} \qquad \Lambda_i := \sum_{j=1}^m \sum_{\ell=1}^n \lambda_{\ell,j}^i Q_\ell \nabla K(x^i) P_j, \tag{2.8b}$$

for some (random) subsets of indices $S(i) \subseteq \{1, \ldots, m\}$ and $V(i+1) \subseteq \{1, \ldots, n\}$ and (random) parameters $\tau_j^i, \phi_j^i, \sigma_\ell^{i+1}, \psi_\ell^{i+1} > 0$, and $\omega_j^i, \lambda_{j,\ell}^i \in \mathbb{R}$. We wait until (3.17) to specify the exact probabilistic setup, which we do not need before that. Due to the block-separable structures of G and F^*, the operators $(I + T_i \partial G)^{-1}$ and $(I + \Sigma_{i+1}\partial F^*)^{-1}$ are also block-separable.

We also pick further subsets of indices $\mathring{S}(i) \subseteq S(i)$ and $\mathring{V}(i+1) \subset V(i+1)$; the rough idea is that x_j^{i+1} for $j \in \mathring{S}(i)$ is updated within each step of the algorithm independently of y^{i+1}. In the linear-K case of [35] also y_ℓ^{i+1} for $\ell \in \mathring{V}(i+1)$ would be updated independently of x^{i+1}, but presently we are not able to ensure that. However, we show at the end of this subsection that the primal blocks x_j^{i+1} for $j \in S(i) \setminus \mathring{S}(i)$ still depend on y_ℓ^{i+1} only for $\ell \in \mathring{V}(i+1)$, as is the case for a linear K in [35]. Moreover we require the "nesting conditions"

$$\chi_{\mathring{S}(i)}(j)(1 - \chi_{V(i+1)}(\ell)) = 0, \qquad (1 - \chi_{S(i)}(j))\chi_{\mathring{V}(i+1)}(\ell) = 0, \tag{2.9a}$$

$$\chi_{\mathring{S}(i)}(j)\chi_{\mathring{V}(i+1)}(\ell) = 0, \quad \text{and} \qquad \chi_{S(i)\setminus\mathring{S}(i)}(j)\chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell) = 0 \tag{2.9b}$$

when

$$\ell \in \mathcal{V}_j^i := \{\ell \in \{1, \ldots, n\} \mid Q_\ell \nabla K(x^i) P_j \neq 0\}. \tag{2.9c}$$

These conditions force those dual blocks that are "connected" by K to the "independently updated" primal blocks $\mathring{S}(i)$ to also be ("dependently") updated, and vice versa. They also disallow connections between independently updated blocks and dependently updated blocks. Note that the last three equations in (2.9) are tantamount to the single equality $\chi_{V(i+1)}(\ell)\chi_{S(i)/\mathring{S}(i)}(j) = \chi_{\mathring{V}(i+1)}(\ell)$: they follow by multiplying the latter by $1 - \chi_{S(i)}$, $\chi_{\mathring{S}(i)}(j)$, and $\chi_{S(i)/\mathring{S}(i)}(j)$, respectively; and vice versa $\chi_{\mathring{V}(i+1)}(\ell) = \chi_{\mathring{V}(i+1)}(\ell)\chi_{S(i)}(j) = \chi_{\mathring{V}(i+1)}(\ell)\chi_{S(i)/\mathring{S}(i)}(j) = \chi_{V(i+1)}(\ell)\chi_{S(i)/\mathring{S}(i)}(j)$.

Example 2.2. We can trivially satisfy (2.9) by taking either $V(i+1) = \{1, \ldots, n\}$, $\mathring{V}(i+1) = \emptyset$, and $\mathring{S}(i) = S(i)$ or $S(i) = \{1, \ldots, m\}$, $\mathring{S}(i) = \emptyset$, and $\mathring{V}(i+1) = V(i+1)$. We will consider these two cases in the respective Section 4 (full dual update methods) and Section 5 (full primal update methods). We may also alternate iterations between these two choices.

Following the notations for the subsets and their complements, we also write

$$\mathring{P}_i := \sum_{j \in \mathring{S}(i)} P_j, \quad \breve{P}_i := \sum_{j \in S(i)\setminus\mathring{S}(i)} P_j, \quad \mathring{Q}_{i+1} := \sum_{\ell \in \mathring{V}(i+1)} Q_\ell, \quad \text{and} \quad \breve{Q}_{i+1} := \sum_{\ell \in V(i+1)\setminus\mathring{V}(i+1)} Q_\ell.$$

In (2.7), for the subsets $S(i)$ and $V(i+1)$ to have the intended meaning that only the corresponding blocks are updated, we need to ensure that $P_j x^{i+1} = P_j x^i$ for $j \notin S(i)$ and $Q_\ell y^{i+1} = Q_\ell y^i$ for $\ell \notin V(i+1)$. This holds if $P_j \Lambda_i^* Q_\ell = 0$ whenever $j \notin S(i), \ell \in V(i+1)$ or $j \in S(i), \ell \notin V(i+1)$ or $j \notin S(i), \ell \notin V(i+1)$. Similarly, for $\mathring{S}(i)$ to have the intended meaning that x_j^{i+1} for $j \in \mathring{S}(i)$ does not depend on y^{i+1}, studying (2.7), we are also led to require

$$\mathring{P}_i[\Lambda_i^* - \Phi_i T_i \nabla K(x^i)^*]Q_\ell = 0 \quad \text{for any } \ell \in V(i+1).$$

Finally, since $\breve{P}_i x^{i+1}$ may in (2.7) depend on y^{i+1}, we require y^{i+1} to not depend on $\breve{P}_i x^{i+1}$:

$$[\Lambda_i - \Psi_{i+1}\Sigma_{i+1}\nabla K(x^i)\Omega^i]\breve{P}_i = 0 \quad \text{and} \quad [I + \Omega^i]\breve{P}_i = 0.$$

Combining the above conditions on Λ_i and Ω_i, we arrive at

$$\begin{cases} P_j \Lambda_i^* Q_\ell = 0 & \text{whenever either } j \notin S(i) \text{ or } \ell \notin V(i+1) \text{ or both,} \\ \mathring{P}_i[\Lambda_i^* - \Phi_i T_i \nabla K(x^i)^*]Q_\ell = 0 & \text{for } \ell \in V(i+1), \\ [\Lambda_i + \Psi_{i+1}\Sigma_{i+1}\nabla K(x^i)]\breve{P}_i = 0, & \text{and} \quad (\Omega^i + I)\breve{P}_i = 0. \end{cases} \tag{2.10}$$

Substituting (2.10) into the identity

$$\Lambda_i = \sum_{\ell \in V(i+1)} Q_\ell \Lambda_i \mathring{P}_i + \sum_{\ell \notin V(i+1)} Q_\ell \Lambda_i \mathring{P}_i + \Lambda_i \breve{P}_i + \sum_{j \notin S(i)} \sum_{\ell=1}^n Q_\ell \Lambda_i P_j,$$

we are led to take

$$\Lambda_i := \sum_{\ell \in V(i+1)} Q_\ell \nabla K(x^i) T_i^* \Phi_i^* \mathring{P}_i - \Psi_{i+1}\Sigma_{i+1}\nabla K(x^i)\breve{P}_i, \tag{2.11}$$

which in terms of the components $\lambda_{\ell,j}^i$ reads

$$\lambda_{\ell,j}^i := \begin{cases} \tau_j^i \phi_j^i & \ell \in V(i+1), j \in \mathring{S}(i), \\ -\sigma_\ell^{i+1}\psi_\ell^{i+1} & \ell \in V(i+1), j \in S(i) \setminus \mathring{S}(i), \\ 0 & \text{otherwise.} \end{cases} \tag{2.12}$$

Using the coupling conditions (2.9) between $\mathring{S}(i)$ and $\mathring{V}(i+1)$ in (2.11), we deduce

$$\Lambda_i = \nabla K(x^i) T_i^* \Phi_i^* \mathring{P}_i - \mathring{Q}_{i+1}\Psi_{i+1}\Sigma_{i+1}\nabla K(x^i).$$

Plugging Λ_i into (2.7), we get two cases for the primal variable. If $j \in \mathring{S}(i)$, we have

$$\mathring{P}_i x^{i+1} = (I + \mathring{T}_i \partial G)^{-1}(\mathring{P}_i x^i - \mathring{T}_i \nabla K(x^i)^* y^i), \quad \text{where} \quad \mathring{T}_i := \mathring{P}_i T_i.$$

If $j \in S(i) \setminus \mathring{S}(i)$, given that $\Omega^i \breve{P}_i = -\breve{P}_i$ due to the last equality of (2.10), taking $\breve{T}_i := \breve{P}_i T_i$, we have

$$\breve{P}_i x^{i+1} = (I + \breve{T}_i \partial G)^{-1}(\breve{P}_i x^i - \breve{T}_i \nabla K(x^i)^* \mathring{Q}_{i+1} y^{i+1} - \breve{P}_i \Phi_i^{-1} \nabla K(x^i)^* \Sigma_{i+1}^* \Psi_{i+1}^* \mathring{Q}_{i+1}(y^{i+1} - y^i)).$$

Also $x^{i+1} = \mathring{P}_i x^{i+1} + \breve{P}_i x^{i+1} + (I - \mathring{P}_i - \breve{P}_i)x^{i+1}$, therefore, for $\bar{x}^{i+1} = x^{i+1} + \Omega^i(x^{i+1} - x^i)$ we can expand $x^{i+1} = \mathring{P}_i x^{i+1} - \Omega^i \breve{P}_i x^{i+1} + (I - \mathring{P}_i - \breve{P}_i)x^i = \mathring{P}_i x^{i+1} + (I - \mathring{P}_i)x^i - \Omega^i \breve{P}_i(x^{i+1} - x^i)$. Consequently, the implicitly defined algorithm in (2.7) expands into the explicit successive updates for each of the involved projections :

$$
(2.13)\qquad \begin{cases}
\mathring{P}_i x^{i+1} := (I + \mathring{T}_i \partial G)^{-1}(\mathring{P}_i x^i - \mathring{T}_i \nabla K(x^i)^* y^i), \\
\quad \bar{x}^{i+1} := (I - \mathring{P}_i) x^i + \mathring{P}_i x^{i+1} + \mathring{P}_i \Omega_i \mathring{P}_i (x^{i+1} - x^i), \\
\quad y^{i+1} := (I + \Sigma_{i+1}\partial F^*)^{-1}\Big(y^i + \Sigma_{i+1} K(\bar{x}^{i+1}) \\
\qquad\qquad + \breve{Q}_{i+1}\Psi_{i+1}^{-1}[\nabla K(x^i) T_i^* \Phi_i^* - \Psi_{i+1}\Sigma_{i+1}\nabla K(x^i)\Omega_i]\mathring{P}_i(x^{i+1} - x^i)\Big), \\
\breve{P}_i x^{i+1} := (I + \breve{T}_i \partial G)^{-1}\Big(\breve{P}_i x^i - \breve{T}_i \nabla K(x^i)^* \breve{Q}_{i+1} y^{i+1} \\
\qquad\qquad - \breve{P}_i \Phi_i^{-1} \nabla K(x^i)^* \Sigma_{i+1}^* \Psi_{i+1}^* \mathring{Q}_{i+1}(y^{i+1} - y^i)\Big), \\
P_j x^{i+1} := P_j x^i \quad \text{for } j \notin S(i).
\end{cases}
$$

In the following sections we will further develop and simplify this algorithm by imposing additional conditions on the step length and testing parameters through convergence analysis.

3 GENERAL ESTIMATES

With the estimate (2.6) in mind, our main task in this section is to prove (2.5). After introducing the assumptions we need for this work in Section 3.1, and bounding $Z_{i+1}M_{i+1}$ from below in Section 3.2, we do the first stage of this estimation in Section 3.3 still deterministically. Then in Section 3.4 we refine these estimates by taking the expectation. Finally in Section 3.5 we combine the various estimates and state a self-contained result on the validity of (2.6).

3.1 ASSUMPTIONS

We will need K to be sufficiently smooth and to satisfy a somewhat technical "three-point" version of standard second-order growth conditions:

Assumption 3.1 (Lipschitz $\nabla K(x)$). For some $L \geq 0$ and a neighbourhood $\mathcal{X}_K \ni \widehat{x}$,

$$
(3.1)\qquad \|\nabla K(x) - \nabla K(x')\| \leq L\|x - x'\| \quad (x, x' \in \mathcal{X}_K).
$$

Using the equality

$$
K(x') = K(x) + \nabla K(x)(x' - x) + \int_0^1 (\nabla K(x + s(x' - x)) - \nabla K(x))(x' - x)ds,
$$

we obtain for any $x, x' \in \mathcal{X}_K$ and $y \in \operatorname{dom} F^*$ as a direct consequence of Assumption 3.1 that

$$
(3.2)\qquad \langle K(x') - K(x) - \nabla K(x)(x' - x), y\rangle \leq \frac{L}{2}\|x - x'\|^2 \|y\|_{P_{\mathrm{NL}}}.
$$

The norm of y only needs to be evaluated within Y_{NL} because $x \mapsto (I - P_{\mathrm{NL}})K(x)$ is linear so the corresponding inner product with the integral term is zero.

Assumption 3.2 (three-point condition on K). For a neighbourhood $\mathcal{X}_K$ of $\widehat{x}$, some $\Gamma_K = \sum_{j=1}^m \gamma_{K,j} P_j \in \mathbb{L}(X; X)$ with $\gamma_{K,j} \in \mathbb{R}$, $L_3 \geq 0$, and $p \in [1, 2]$, for any $A = \sum_{j=1}^m a_j P_j \geq 0$ and some $\theta_A \geq 0$ the following holds

$$
(3.3)\qquad \langle [\nabla K(x) - \nabla K(\widehat{x})]^* \widehat{y}, x' - \widehat{x}\rangle_A \\
\geq \|x' - \widehat{x}\|^2_{A\Gamma_K} + \theta_A \|K(\widehat{x}) - K(x) - \nabla K(x)(\widehat{x} - x)\|^p - \frac{L_3}{2}\|x' - x\|^2_A, \quad (x, x' \in \mathcal{X}_K).
$$

This assumption is trivially satisfied for $\gamma_{K,j} = L_3 = 0$ and any $\theta_A > 0$ whenever $x \mapsto \langle K(x), \widehat{y}\rangle$ is linear. In Appendix A we also provide the constants ensuring this assumption, e.g., whenever the latter is block-separable and strongly-convex. For a less straight-forward example in the single-block case, we refer to [9]. There we verified the assumption for the reconstruction of the phase and amplitude of a complex number from a noisy measurements. That example evidently applies to the present setting in the single-block case or as a separable block of $x \mapsto \langle K(x), \widehat{y}\rangle$.

We also need pointwise monotonicity of ∂G and ∂F^* at a root $\widehat{u} \in H^{-1}(0)$:

Definition 3.3. Let U be a Hilbert space, and $\Gamma \in \mathbb{L}(U; U)$, $\Gamma \geq 0$. We say that the set-valued map $H : U \rightrightarrows U$ is *Γ-strongly monotone at $\widehat{u}$ for $\widehat{w} \in H(\widehat{u})$* if there exists a neighbourhood $\mathcal{U} \ni \widehat{u}$ such that for any $u \in \mathcal{U}$ and $w \in H(u)$,

$$\langle w - \widehat{w}, u - \widehat{u}\rangle \geq \|u - \widehat{u}\|_\Gamma^2. \tag{3.4}$$

If $\Gamma = 0$, we say that H is monotone at $\widehat{u}$ for $\widehat{w}$.

Assumption 3.4. For any $\widehat{w} = (\widehat{v}, \widehat{\xi}) \in H(\widehat{u})$, the set-valued map ∂G is $\sum_{j=1}^m \gamma_{G,j} P_j$-strongly monotone at $\widehat{x}$ for $\widehat{v} - \nabla K(\widehat{x})^* \widehat{y}$ in the neighbourhood $\mathcal{X}_G$, and the set-valued map ∂F^* is $\sum_{\ell=1}^n \gamma_{F^*,\ell} Q_\ell$-strongly monotone at $\widehat{y}$ for $\widehat{\xi} + K(\widehat{x})$ in the neighbourhood $\mathcal{Y}_{F^*}$, where the constants $\gamma_{G,j}, \gamma_{F^*,\ell} \geq 0$ for all $j = 1, \dots, m$ and $\ell = 1, \dots, n$.

3.2 A LOWER BOUND ON THE LOCAL METRIC

To estimate $Z_{i+1}M_{i+1}$ from below, we formulate a block-adapted version of the basic step length condition $\tau\sigma\|K\|^2 < 1$ from [7]. The assumptions of the following lemma replace the more abstract constructions of [35, Definition 2.2 and Examples 2.3 and 2.4]. We recall from (2.9c) the "set of connections" $\mathcal{V}_j^i$ and also introduce the set of "simultaneous connections", filtered by $\lambda_{k,j}^i$, as

$$\bar{\mathcal{V}}_j^i(\ell) := \{k \in \{1, \dots, n\} \mid Q_\ell \nabla K(x^i) P_j \nabla K(x^i)^* Q_k \neq 0,\ \lambda_{k,j}^i \neq 0\}. \tag{3.5}$$

Lemma 3.5. *Let $i \in \mathbb{N}$ and $0 \leq \delta \leq \kappa < 1$. For some factors $w_{j,\ell,k}^i = 1/w_{j,k,\ell}^i > 0$, ($\ell, k = 1, \dots, n$; $j = 1, \dots, m$), define*

$$w_{j,\ell}^i := \chi_{\mathcal{V}_j^i}(\ell) \sum_{k \in \bar{\mathcal{V}}_j^i(\ell)} w_{j,\ell,k}^i \tag{3.6}$$

and suppose

$$(1-\kappa)\psi_\ell^{i+1} \geq \left\| \sum_{j=1}^m |\lambda_{\ell,j}^i| \sqrt{w_{j,\ell}^i/\phi_j^i}\, Q_\ell \nabla K(x^i) P_j \right\|^2 \quad (\ell = 1, \dots, n). \tag{3.7}$$

Then

$$Z_{i+1}M_{i+1} \geq \begin{pmatrix} \delta\Phi_i & 0 \\ 0 & \frac{\kappa-\delta}{1-\delta}\Psi_{i+1} \end{pmatrix}. \tag{3.8}$$

Proof. Setting $\zeta_{\ell,j} := (\phi_j^i)^{-1}(\lambda_{\ell,j}^i)^2/(1-\kappa)$, we use (3.7) and the orthogonality of the projections $\{P_j\}_{j=1}^m$ to obtain for any $y \in Y$ that

$$\begin{aligned}
\sum_{\ell=1}^n \psi_\ell^{i+1}\|Q_\ell y\|^2 &\geq \sum_{\ell=1}^n \left\| \sum_{j=1}^m \sqrt{\zeta_{\ell,j} w_{j,\ell}^i}\, Q_\ell \nabla K(x^i) P_j \right\|^2 \|Q_\ell y\|^2 \geq \sum_{\ell=1}^n \left\| \sum_{j=1}^m \sqrt{\zeta_{\ell,j} w_{j,\ell}^i}\, P_j \nabla K(x^i)^* Q_\ell y \right\|^2 \\
&= \sum_{\ell=1}^n \sum_{j=1}^m \zeta_{\ell,j} w_{j,\ell}^i \|P_j \nabla K(x^i)^* Q_\ell y\|^2 \geq \sum_{j=1}^m \sum_{\ell \in \mathcal{V}_j^i} \left(\sum_{k \in \bar{\mathcal{V}}_j^i(\ell)} w_{j,\ell,k}^i \right) \zeta_{\ell,j} \|P_j \nabla K(x^i)^* Q_\ell y\|^2.
\end{aligned}$$

Since $w^i_{j,k,\ell} = 1/w^i_{j,\ell,k}$, we continue to estimate by Young's inequality

$$\sum_{\ell=1}^{n} \psi_\ell^{i+1} \|Q_\ell y\|^2 \geq \sum_{j=1}^{m} \sum_{k,\ell=1}^{n} \zeta_{\ell,j}^{1/2} \zeta_{k,j}^{1/2} \langle P_j \nabla K(x^i)^* Q_\ell y, \nabla K(x^i)^* Q_k y \rangle.$$

Here we also used (3.5) to convert the second sum to run over all $k, \ell = 1, \ldots, n$. As $y \in Y$ was arbitrary, inserting $\zeta_{k,j}$ and the structure (2.8) of Ψ_{i+1}, Φ_i, and Λ_i, we deduce $(1-\kappa)\Psi_{i+1} \geq \Lambda_i \Phi_i^{-1} \Lambda_i^*$.

On the other hand, applying Young's inequality with the factor $(1-\delta)$ we deduce that

$$Z_{i+1}M_{i+1} = \begin{pmatrix} \Phi_i & -\Lambda_i^* \\ -\Lambda_i & \Psi_{i+1} \end{pmatrix} \geq \begin{pmatrix} \delta\Phi_i & 0 \\ 0 & \Psi_{i+1} - \frac{1}{1-\delta}\Lambda_i \Phi_i^{-1} \Lambda_i^* \end{pmatrix} \tag{3.9}$$

Thus (3.8) holds.

□

The next example demonstrates a simple choice of the weights $w_{j,k,\ell}$ that is likely to work if all the dual blocks ℓ have similar roles in the problem. In Section 6 we will also consider other options when some dual blocks have different roles.

Example 3.6 (Equal weighting). Suppose $\mathcal{V}_j^i \subset \mathcal{V}_j$ and $\bar{\mathcal{V}}_j^i(\ell) \subset \bar{\mathcal{V}}_j(\ell)$ where $\mathcal{V}_j$ and $\bar{\mathcal{V}}_j(\ell)$ do not depend on the iteration. If we take $w^i_{j,\ell,k} \equiv 1$, then $w_{j,\ell} = \chi_{\mathcal{V}_j}(\ell)\#\bar{\mathcal{V}}_j(\ell)$ counts the dual blocks "simultaneously connected" with ℓ via the primal block j as defined by (3.5).

To provide further intuition into the result, let $w_{j,\ell}$ be as in Example 3.6. With only one primal block ($j, m = 1$), and assuming full connectedness ($w_{1,\ell} = n$ for all $\ell = 1, \ldots, n$), Lemma 3.5 requires $\psi_\ell \geq \zeta_{1,\ell} n \|Q_\ell \nabla K(x^i)\|^2$. Let $a := \frac{1}{n}\sum_{\ell=1}^n \|Q_\ell \nabla K(x^i)\|^2 = \frac{1}{n}\|\nabla K(x^i)\|^2$. After plugging $\lambda^i_{\ell,j}$ from (2.12) into (3.7), the lemma then says that the step length parameters can be proportionally larger compared to the single dual block case ($n = 1$) when $\|Q_\ell \nabla K(x^i)\|^2 < a$, and have to be proportionally smaller when $\|Q_\ell \nabla K(x^i)\|^2 > a$. In Section 4 and Section 5, we further transform (3.7) to obtain explicit step-length conditions. But now, for the remainder of Section 3, we assume that (3.8) holds and derive sufficient conditions to be able to apply Theorem 2.1.

3.3 Initial non-stochastic estimates

The next lemma starts the verification of (2.5).

Lemma 3.7. *Suppose Assumptions 3.1 and 3.4 hold together with (3.8) for some $L \geq 0$, $\gamma_{G,j}, \gamma_{F^*,\ell} \geq 0$ ($j = 1, \ldots, m$, $\ell = 1, \ldots, n$), and $0 \leq \delta \leq \kappa < 1$. Then with $\widetilde{H}_{i+1}$ given by (2.3) and M_{i+1} given by (2.2), we have*

$$\begin{aligned} \frac{1}{2}\|u^{i+1} - u^i\|^2_{Z_{i+1}M_{i+1}} &+ \frac{1}{2}\|u^{i+1} - \widehat{u}\|^2_{Z_{i+1}M_{i+1} - Z_{i+2}M_{i+2}} + \langle \widetilde{H}_{i+1}(u^{i+1}), u^{i+1} - \widehat{u} \rangle_{W_{i+1}Z_{i+1}} \\ &\geq \frac{1}{2}\|x^{i+1} - x^i\|^2_{R_x} + \frac{1}{2}\frac{\kappa - \delta}{1-\delta}\|y^{i+1} - y^i\|^2_{\Psi_{i+1}} + \frac{1}{2}\|u^{i+1} - \widehat{u}\|^2_{R'} + D_i^K + D_i^\Lambda, \end{aligned} \tag{3.10}$$

where for an arbitrary $\Gamma_K := \sum_{j=1}^m \gamma_{K,j} P_j \in \mathbb{L}(X;X)$ for $\gamma_{K,j} \in \mathbb{R}$ we set

$$R_x := \delta\Phi_i - L\|\Omega^i + I\|^2 \|\Psi_{i+1}^* \Sigma_{i+1}^* (y^{i+1} - \widehat{y})\|_{P_{\mathrm{NL}}} I, \tag{3.11a}$$

$$R' := \begin{pmatrix} \Phi_i - \Phi_{i+1} + 2\sum_{j \in S(i)} \phi_j^i \tau_j^i (\gamma_{G,j} + \gamma_{K,j}) P_j & 0 \\ 0 & \Psi_{i+1} - \Psi_{i+2} + 2\sum_{\ell \in V(i+1)} \psi_\ell^{i+1} \sigma_\ell^{i+1} \gamma_{F^*,\ell} Q_\ell \end{pmatrix}, \tag{3.11b}$$

$$\begin{aligned} D_i^\Lambda := &\langle [\Lambda_{i+1} - \Lambda_i](x^{i+1} - \widehat{x}), y^{i+1} - \widehat{y} \rangle \\ &+ \langle \nabla K(x^i)^* (y^{i+1} - \widehat{y}), x^{i+1} - \widehat{x} \rangle_{\Phi_i T_i - \Sigma_{i+1}^* \Psi_{i+1}^*}, \quad \textit{and} \end{aligned} \tag{3.11c}$$

$$\begin{aligned} D_i^K := &\langle [\nabla K(x^i) - \nabla K(\widehat{x})]^* \widehat{y}, x^{i+1} - \widehat{x} \rangle_{\Phi_i T_i} - \|x^{i+1} - \widehat{x}\|^2_{\Phi_i T_i \Gamma_K} \\ &+ \langle K(\widehat{x}) - K(x^i) - \nabla K(x^i)(\widehat{x} - x^i), y^{i+1} - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}}. \end{aligned} \tag{3.11d}$$

Proof. We bound from below all the terms on the left-hand side of (3.10). For the first term, we have from (3.8) that

$$Z_{i+1}M_{i+1} \geq \begin{pmatrix} \delta\Phi_i & 0 \\ 0 & \frac{\kappa-\delta}{1-\delta}\Psi_{i+1} \end{pmatrix}. \tag{3.12}$$

For the second term we use the expansion

$$Z_{i+1}M_{i+1} - Z_{i+2}M_{i+2} = \begin{pmatrix} \Phi_i - \Phi_{i+1} & \Lambda_{i+1}^* - \Lambda_i^* \\ \Lambda_{i+1} - \Lambda_i & \Psi_{i+1} - \Psi_{i+2} \end{pmatrix}. \tag{3.13}$$

We need to work more to estimate the third term on the left-hand side of (3.10). Since $0 \in H(\widehat{u})$, we have $\partial G(\widehat{x}) \ni z_G := -\nabla K(\widehat{x})^* \widehat{y}$, and $\partial F^*(\widehat{y}) \ni z_{F^*} := K(\widehat{x})$. We can therefore recall the definition of $H(u)$ from (2.1) and rewrite

$$\begin{aligned} \langle H(u), u - \widehat{u} \rangle_{W_{i+1}Z_{i+1}} &= \langle \partial G(x) - z_G, x - \widehat{x} \rangle_{\Phi_i T_i} + \langle \partial F^*(y) - z_{F^*}, y - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}} \\ &\quad + \langle \nabla K(x)^* y - \nabla K(\widehat{x})^* \widehat{y}, x - \widehat{x} \rangle_{\Phi_i T_i} + \langle K(\widehat{x}) - K(x), y - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}}. \end{aligned}$$

Recalling the definition of $\widetilde{H}_{i+1}(u^{i+1})$ in (2.3), we therefore expand the third term of (3.10) as

$$\begin{aligned} &\langle \widetilde{H}_{i+1}(u^{i+1}), u^{i+1} - \widehat{u} \rangle_{W_{i+1}Z_{i+1}} \\ &\quad = \langle \partial G(x^{i+1}) - z_G, x^{i+1} - \widehat{x} \rangle_{\Phi_i T_i} + \langle \partial F^*(y^{i+1}) - z_{F^*}, y^{i+1} - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}} \\ &\quad\quad + \langle \nabla K(x^{i+1})^* y^{i+1} - \nabla K(\widehat{x})^* \widehat{y}, x^{i+1} - \widehat{x} \rangle_{\Phi_i T_i} + \langle K(\widehat{x}) - K(x^{i+1}), y^{i+1} - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}} \\ &\quad\quad + \langle [\nabla K(x^i) - \nabla K(x^{i+1})]^* y^{i+1}, x^{i+1} - \widehat{x} \rangle_{\Phi_i T_i} \\ &\quad\quad + \langle K(x^{i+1}) - K(x^{i+1} + \Omega^i(x^{i+1} - x^i)) + \nabla K(x^i)\Omega^i(x^{i+1} - x^i), y^{i+1} - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}}. \end{aligned}$$

Due to Assumption 3.4 and (3.2), we have

$$\begin{aligned} D_i^\Gamma &:= \langle \partial G(x^{i+1}) - z_G, x^{i+1} - \widehat{x} \rangle_{\Phi_i T_i} + \|x^{i+1} - \widehat{x}\|^2_{\Phi_i T_i \Gamma_K} + \langle \partial F^*(y^{i+1}) - z_{F^*}, y^{i+1} - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}} \\ &\geq \sum_{j \in S(i)} \phi_j^i \tau_j^i \|x^{i+1} - \widehat{x}\|^2_{P_j \Gamma_G P_j} + \|x^{i+1} - \widehat{x}\|^2_{\Phi_i T_i \Gamma_K} + \sum_{\ell \in V(i+1)} \psi_\ell^{i+1} \sigma_\ell^{i+1} \|y^{i+1} - \widehat{y}\|^2_{Q_\ell \Gamma_{F^*} Q_\ell}, \end{aligned} \tag{3.14}$$

and

$$\begin{aligned} D_i^\Omega &:= \langle K(x^i) - K(x^{i+1} + \Omega^i(x^{i+1} - x^i)) + \nabla K(x^i)(\Omega^i + I)(x^{i+1} - x^i), y^{i+1} - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}} \\ &\geq -\frac{L}{2} \|\Omega^i + I\|^2 \|\Psi_{i+1}^* \Sigma_{i+1}^* (y^{i+1} - \widehat{y})\|_{P_{\mathrm{NL}}} \|x^{i+1} - x^i\|^2. \end{aligned} \tag{3.15}$$

Hence, recalling D_i^K from (3.11d), we deduce

$$\begin{aligned} &\langle \widetilde{H}_{i+1}(u^{i+1}), u^{i+1} - \widehat{u} \rangle_{W_{i+1}Z_{i+1}} \\ &= \langle [\nabla K(x^i) - \nabla K(\widehat{x})]^* \widehat{y}, x^{i+1} - \widehat{x} \rangle_{\Phi_i T_i} - \|x^{i+1} - \widehat{x}\|^2_{\Phi_i T_i \Gamma_K} \\ &\quad + \langle K(\widehat{x}) - K(x^i) - \nabla K(x^i)(\widehat{x} - x^i), y^{i+1} - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}} \\ &\quad + \langle \partial G(x^{i+1}) - z_G, x^{i+1} - \widehat{x} \rangle_{\Phi_i T_i} + \|x^{i+1} - \widehat{x}\|^2_{\Phi_i T_i \Gamma_K} + \langle \partial F^*(y^{i+1}) - z_{F^*}, y^{i+1} - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}} \\ &\quad + \langle K(x^i) - K(x^{i+1} + \Omega^i(x^{i+1} - x^i)) + \nabla K(x^i)(\Omega^i + I)(x^{i+1} - x^i), y^{i+1} - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}} \\ &\quad + \langle \nabla K(x^i)^* (y^{i+1} - \widehat{y}), x^{i+1} - \widehat{x} \rangle_{\Phi_i T_i} - \langle \nabla K(x^i)(x^{i+1} - \widehat{x}), y^{i+1} - \widehat{y} \rangle_{\Psi_{i+1}\Sigma_{i+1}} \\ &= D_i^K + D_i^\Gamma + D_i^\Omega + \langle \nabla K(x^i)^* (y^{i+1} - \widehat{y}), x^{i+1} - \widehat{x} \rangle_{\Phi_i T_i - \Sigma_{i+1}^* \Psi_{i+1}^*}. \end{aligned} \tag{3.16}$$

Inserting the lower bounds from (3.12), (3.14), and (3.15) into (3.16), and using (3.11d) and (3.13), we obtain

$$\begin{aligned}\frac{1}{2}\|u^{i+1}-u^i\|^2_{Z_{i+1}M_{i+1}} &+ \frac{1}{2}\|u^{i+1}-\widehat{u}\|^2_{Z_{i+1}M_{i+1}-Z_{i+2}M_{i+2}} + \langle \widetilde{H}_{i+1}(u^{i+1}), u^{i+1}-\widehat{u}\rangle_{W_{i+1}Z_{i+1}} \\ &\geq \frac{1}{2}\|x^{i+1}-x^i\|^2_{\delta\Phi_i} + \frac{1}{2}\frac{\kappa-\delta}{1-\delta}\|y^{i+1}-y^i\|^2_{\Psi_{i+1}} + \frac{1}{2}\|u^{i+1}-\widehat{u}\|^2_{R'} + D_i^\Lambda + D_i^K \\ &\qquad - \frac{L}{2}\|\Omega^i+I\|^2\|\Psi^*_{i+1}\Sigma^*_{i+1}(y^{i+1}-\widehat{y})\|_{P_{\mathrm{NL}}}\|x^{i+1}-x^i\|^2\end{aligned}$$

for D_i^Λ as in (3.11c). Finally, using the definitions of R_x in (3.11), we observe

$$\frac{1}{2}\|x^{i+1}-x^i\|^2_{\delta\Phi_i} - L\|\Omega^i+I\|^2\|\Psi^*_{i+1}\Sigma^*_{i+1}(y^{i+1}-\widehat{y})\|_{P_{\mathrm{NL}}}\|x^{i+1}-x^i\|^2 = \|x^{i+1}-x^i\|^2_{R_x}.$$

This yields the claim. □

3.4 EXPECTATION ESTIMATES

To further estimate D_i^K and D_i^Λ, we have to take the expectation with respect to O_{i-1}. We will use a split definition of the step lengths, writing

$$\tau_j^i = \begin{cases} \mathring{\tau}_j^i, & j \in \mathring{S}(i), \\ \breve{\tau}_j^i, & j \in S(i)\setminus\mathring{S}(i),\end{cases} \qquad \text{and} \qquad \sigma_\ell^{i+1} = \begin{cases} \mathring{\sigma}_\ell^{i+1}, & \ell \in \mathring{V}(i+1), \\ \breve{\sigma}_\ell^{i+1}, & \ell \in V(i+1)\setminus\mathring{V}(i+1),\end{cases}$$

where we make for all $i \in \mathbb{N}$ the conditionality assumptions

$$\phi_j^i, \psi_\ell^{i+1} \in \mathcal{R}(O_{i-1};(0,\infty)), \qquad \mathring{\tau}_j^i, \breve{\tau}_j^i, \mathring{\sigma}_\ell^{i+1}, \breve{\sigma}_\ell^{i+1} \in \mathcal{R}(O_{i-1};(0,\infty)), \tag{3.17a}$$

$$S(i), \mathring{S}(i) \in \mathcal{R}(O_i;\mathcal{P}\{1,\ldots,m\}), \quad \text{and} \quad V(i+1), \mathring{V}(i+1) \in \mathcal{R}(O_i;\mathcal{P}\{1,\ldots,n\}). \tag{3.17b}$$

Thus $\mathring{\tau}_j^i$ always refers to what τ_j^i would be if $j \in \mathring{S}(i)$, and similarly for the other variables. Moreover, these step lengths are already known on iteration $i-1$, prior to their use. The only part that is not known about T_i and Σ_{i+1} before commencing iteration i are the subsets of blocks to be updated. Observe that (3.17) and (2.13) imply

$$x^{i+1} \in \mathcal{R}(O_i;X) \quad \text{and} \quad y^{i+1} \in \mathcal{R}(O_i;Y) \quad (i \in \mathbb{N}). \tag{3.18}$$

Also, for brevity, we write

$$\begin{aligned}\pi_j^i &:= \mathbb{P}[j \in S(i) \mid O_{i-1}], & \mathring{\pi}_j^i &:= \mathbb{P}[j \in \mathring{S}(i) \mid O_{i-1}], \\ \nu_\ell^{i+1} &:= \mathbb{P}[\ell \in V(i+1) \mid O_{i-1}], \quad \text{and} & \mathring{\nu}_\ell^{i+1} &:= \mathbb{P}[\ell \in \mathring{V}(i+1) \mid O_{i-1}].\end{aligned}$$

Lemma 3.8. *Suppose Assumption 3.2 and (3.17) hold for some $L_3 \geq 0$, $p \in [1,2]$, and $\theta_A \geq 0$. For some $\rho_\ell > 0$ assume*

$$1 = \mathbb{P}[\|y_\ell^{i+1} - \widehat{y}_\ell\|_{P_{\mathrm{NL}}} \leq \rho_\ell \mid O_{i-1}] \quad (\ell = 1,\ldots,m). \tag{3.19}$$

Then D_i^K defined in (3.11c) satisfies for any $\zeta_\ell > 0$ with $\sum_{\ell=1}^n \nu_\ell^{i+1}\psi_\ell^{i+1}\sigma_\ell^{i+1}\zeta_\ell^{1-p}\rho_\ell^{2-p} \leq p^p\mathbb{E}_{i-1}[\theta_{\Phi_i T_i}]$ the lower bound

$$\begin{aligned}\mathbb{E}_{i-1}[D_i^K] \geq &-\frac{L_3}{2}\mathbb{E}_{i-1}[\|x^{i+1}-x^i\|^2_{\Phi_i T_i}] \\ &- \sum_{\ell=1}^n \mathbb{E}_{i-1}\left[\psi_\ell^{i+1}\sigma_\ell^{i+1}(p-1)\zeta_\ell\|y_\ell^{i+1}-\widehat{y}_\ell\|^2_{P_{\mathrm{NL}}}\right].\end{aligned} \tag{3.20}$$

Proof. Setting $A = \Phi_i T_i$ in Assumption 3.2, we obtain

$$\begin{aligned}\langle[\nabla K(x^i) - \nabla K(\widehat{x})]^*\widehat{y}, x^{i+1} - \widehat{x}\rangle_{\Phi_i T_i} \\ \geq \|x^{i+1} - \widehat{x}\|^2_{\Phi_i T_i \Gamma_K} + \theta_{\Phi_i T_i}\|K(\widehat{x}) - K(x^i) - \nabla K(x^i)(\widehat{x} - x^i)\|^p - \frac{L_3}{2}\|x^{i+1} - x^i\|^2_{\Phi_i T_i}.\end{aligned}$$

Therefore, recalling the definition of D^K_k in (3.11d) and using (3.18),

$$\begin{aligned}\mathbb{E}_{i-1}[D^K_i] \geq \mathbb{E}_{i-1}[\theta_{\Phi_i T_i}]\|K(\widehat{x}) - K(x^i) - \nabla K(x^i)(\widehat{x} - x^i)\|^p - \frac{L_3}{2}\mathbb{E}_{i-1}\big[\|x^{i+1} - x^i\|^2_{\Phi_i T_i}\big] \\ + \langle K(\widehat{x}) - K(x^i) - \nabla K(x^i)(\widehat{x} - x^i), \mathbb{E}_{i-1}[\Sigma^*_{i+1}\Psi^*_{i+1}(y^{i+1} - \widehat{y})]\rangle.\end{aligned} \tag{3.21}$$

By Young's inequality and (3.19) as in [9, (3.16) and (3.17)], for any $\zeta_\ell > 0$,

$$\begin{aligned}&\langle K(\widehat{x}) - K(x^i) - \nabla K(x^i)(\widehat{x} - x^i), \Sigma^*_{i+1}\Psi^*_{i+1}(y^{i+1} - \widehat{y})\rangle \\ &\geq -\sum_{\ell\in V(i+1)} \psi^{i+1}_\ell \sigma^{i+1}_\ell \|y^{i+1}_\ell - \widehat{y}_\ell\|_{P_{\mathrm{NL}}} \cdot \|K(\widehat{x}) - K(x^i) - \nabla K(x^i)(\widehat{x} - x^i)\| \\ &\geq -\sum_{\ell\in V(i+1)} \psi^{i+1}_\ell \sigma^{i+1}_\ell (p-1)\zeta_\ell \|y^{i+1}_\ell - \widehat{y}_\ell\|^2_{P_{\mathrm{NL}}} \\ &\quad - \sum_{i=1}^n \frac{\chi_{V(i+1)}(\ell)\psi^{i+1}_\ell \sigma^{i+1}_\ell \|y^{i+1}_\ell - \widehat{y}_\ell\|^{2-p}_{P_{\mathrm{NL}}}}{p^p \zeta_\ell^{p-1}} \cdot \|K(\widehat{x}) - K(x^i) - \nabla K(x^i)(\widehat{x} - x^i)\|^p.\end{aligned}$$

Taking the expectation $\mathbb{E}_{i-1}$, applying the assumption $\sum_{\ell=1}^n \nu^{i+1}_\ell \psi^{i+1}_\ell \sigma^{i+1}_\ell \zeta_\ell^{1-p} \rho_\ell^{2-p} \leq p^p \mathbb{E}_{i-1}[\theta_{\Phi_i T_i}]$, and inserting the result in (3.21), we obtain the claim (3.20). ◻

Lemma 3.9. *Suppose Assumption 3.1 and (3.17) are satisfied for some $L \geq 0$, and the nesting conditions (2.9) hold for any j and ℓ on both iterations i and $i+1$. For some $\eta^{i+1} > 0$ assume*

$$\mathring{\pi}^{i+1}_j \phi^{i+1}_j \mathring{\tau}^{i+1}_j = \eta^{i+1} - \chi_{S(i)\setminus\mathring{S}(i)}(j)\phi^i_j \breve{\tau}^i_j, \tag{3.22a}$$

$$\mathring{\nu}^{i+2}_\ell \psi^{i+2}_\ell \mathring{\sigma}^{i+2}_\ell = \eta^{i+1} - \chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell)\psi^{i+1}_\ell \breve{\sigma}^{i+1}_\ell. \tag{3.22b}$$

Then D^Λ_i defined in (3.11c) satisfies for any given $\alpha_x, \alpha_y > 0$ the lower bound

$$\begin{aligned}\mathbb{E}_i[D^\Lambda_i] + \frac{d^{i+1}}{2}\|x^{i+1} - x^i\|^2 \geq -\alpha_x \sum_{j=1}^m \chi_{S(i)\setminus\mathring{S}(i)}(j)\phi^i_j \breve{\tau}^i_j \|x^{i+1}_j - \widehat{x}_j\|^2 \\ - \alpha_y \sum_{\ell=1}^n \chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell)\psi^{i+1}_\ell \breve{\sigma}^{i+1}_\ell \|y^{i+1}_\ell - \widehat{y}_\ell\|^2_{P_{\mathrm{NL}}},\end{aligned} \tag{3.23}$$

where

$$d^{i+1} := \frac{L^2}{2\alpha_x}\left(\sum_{j\in S(i)\setminus\mathring{S}(i)} \phi^i_j \breve{\tau}^i_j\right)\|y^{i+1} - \widehat{y}\|^2_{P_{\mathrm{NL}}} + \frac{L^2}{2\alpha_y}\left(\sum_{\ell\in V(i+1)\setminus\mathring{V}(i+1)} \psi^{i+1}_\ell \breve{\sigma}^{i+1}_\ell\right)\|x^{i+1} - \widehat{x}\|^2.$$

Moreover, if

$$\mathbb{P}[\|x^{i+1} - \widehat{x}\| \leq \rho_x,\ \|Q_\ell(y^{i+1} - \widehat{y})\|_{P_{\mathrm{NL}}} \leq \rho_\ell, (\ell = 1, \ldots, n) \mid O_{i-1}] = 1, \tag{3.24}$$

then

$$\mathbb{E}_{i-1}[d^{i+1}\|x^{i+1} - x^i\|^2] \leq \mathbb{E}_{i-1}[c^i_*\|x^{i+1} - x^i\|^2] \tag{3.25}$$

for

$$c_*^i := \frac{L^2}{2\alpha_x\alpha_y}\Big(\alpha_y \sum_{\ell=1}^{n} \rho_\ell^2 \#(S(i) \setminus \mathring{S}(i)) \max_{j=1,\dots,m} \phi_j^i \check{\tau}_j^i + \alpha_x \rho_x^2 \#(V(i+1) \setminus \mathring{V}(i+1)) \max_{\ell=1,\dots,n} \psi_\ell^{i+1} \check{\sigma}_\ell^{i+1}\Big). \tag{3.26}$$

Proof. We recall from (3.11c) that

$$\begin{aligned}
D_i^\Lambda &:= \langle \nabla K(x^i)^*(y^{i+1} - \widehat{y}), x^{i+1} - \widehat{x}\rangle_{\Phi_i T_i - \Sigma_{i+1}^* \Psi_{i+1}^*} \\
&\quad + \Bigg\langle \Bigg[\sum_{\ell \in V(i+2)} Q_\ell \nabla K(x^{i+1}) T_{i+1}^* \Phi_{i+1}^* \mathring{P}_{i+1} - \Psi_{i+2}\Sigma_{i+2} \nabla K(x^{i+1}) \check{P}_{i+1} \Bigg] (x^{i+1} - \widehat{x}), y^{i+1} - \widehat{y} \Bigg\rangle \\
&\quad - \Bigg\langle \Bigg[\sum_{\ell \in V(i+1)} Q_\ell \nabla K(x^{i}) T_{i}^* \Phi_{i}^* \mathring{P}_{i} - \Psi_{i+1}\Sigma_{i+1} \nabla K(x^{i}) \check{P}_{i} \Bigg] (x^{i+1} - \widehat{x}), y^{i+1} - \widehat{y} \Bigg\rangle.
\end{aligned}$$

Defining for brevity

$$k_{\ell,j} := \langle \nabla K(x^i)^*(y_\ell^{i+1} - \widehat{y}_\ell), x_j^{i+1} - \widehat{x}_j\rangle \quad \text{and} \quad k_{\ell,j}^+ := \langle \nabla K(x^{i+1})^*(y_\ell^{i+1} - \widehat{y}_\ell), x_j^{i+1} - \widehat{x}_j\rangle,$$

and using (3.17), which implies $\phi_j^i \tau_j^i, \psi_\ell^{i+1}\sigma_\ell^{i+1} \in \mathcal{R}(O_i; (0,\infty))$, we expand

$$\begin{aligned}
\mathbb{E}_i[D_i^\Lambda] = \sum_{\ell=1}^{n}\sum_{j=1}^{m} \Big[&(\chi_{S(i)}(j)\phi_j^i \tau_j^i - \chi_{V(i+1)}(\ell)\psi_\ell^{i+1}\sigma_\ell^{i+1}) k_{\ell,j} \\
&+ \mathbb{E}_i[\chi_{V(i+2)}(\ell)(\chi_{\mathring{S}(i+1)}(j)\phi_j^{i+1}\mathring{\tau}_j^{i+1} - \chi_{S(i+1)\setminus\mathring{S}(i+1)}(j)\psi_\ell^{i+2}\sigma_\ell^{i+2})k_{\ell,j}^+] \\
&- \chi_{V(i+1)}(\ell)(\chi_{\mathring{S}(i)}(j)\phi_j^i\mathring{\tau}_j^i - \chi_{S(i)\setminus\mathring{S}(i)}(j)\psi_\ell^{i+1}\sigma_\ell^{i+1})k_{\ell,j}\Big].
\end{aligned}$$

Writing in the first term $\chi_{S(i)}(j)\phi_j^i\tau_j^i = \chi_{\mathring{S}(i)}(j)\phi_j^i\mathring{\tau}_j^i + \chi_{S(i)\setminus\mathring{S}(i)}(j)\phi_j^i\check{\tau}_j^i$, this rearranges as

$$\begin{aligned}
\mathbb{E}_i[D_i^\Lambda] = \sum_{\ell=1}^{n}\sum_{j=1}^{m} \Bigg(&\Big[\chi_{S(i)\setminus\mathring{S}(i)}(j)\phi_j^i\check{\tau}_j^i + (1 - \chi_{V(i+1)}(\ell))\chi_{\mathring{S}(i)}(j)\phi_j^i\mathring{\tau}_j^i \\
&\quad + \chi_{V(i+1)}(\ell)(\chi_{S(i)\setminus\mathring{S}(i)}(j) - 1)\psi_\ell^{i+1}\sigma_\ell^{i+1}\Big] k_{\ell,j} \\
&+ \mathbb{E}_i\Big[\chi_{V(i+2)}(\ell)\chi_{\mathring{S}(i+1)}(j)\phi_j^{i+1}\mathring{\tau}_j^{i+1} \\
&\quad - \chi_{V(i+2)}(\ell)\chi_{S(i+1)\setminus\mathring{S}(i+1)}(j)\psi_\ell^{i+2}\sigma_\ell^{i+2}\Big] k_{\ell,j}^+ \Bigg).
\end{aligned}$$

Using (2.9), we continue

$$\begin{aligned}
\mathbb{E}_i[D_i^\Lambda] = \sum_{\ell=1}^{n}\sum_{j=1}^{m} \Bigg(&[\chi_{S(i)\setminus\mathring{S}(i)}(j)\phi_j^i\check{\tau}_j^i - \chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell)\psi_\ell^{i+1}\check{\sigma}_\ell^{i+1}] k_{\ell,j} \\
&+ \mathbb{E}_i[\chi_{\mathring{S}(i+1)}(j)\phi_j^{i+1}\mathring{\tau}_j^{i+1} - \chi_{\mathring{V}(i+2)}(\ell)\psi_\ell^{i+2}\mathring{\sigma}_\ell^{i+2}] k_{\ell,j}^+ \Bigg),
\end{aligned}$$

after which a use of (3.22) rearranges this as

$$\begin{aligned}
\mathbb{E}_i[D_i^\Lambda] &= \sum_{\ell=1}^{n}\sum_{j=1}^{m} (\mathring{\pi}_j^{i+1}\phi_j^{i+1}\mathring{\tau}_j^{i+1} - \mathring{\nu}_\ell^{i+2}\psi_\ell^{i+2}\mathring{\sigma}_\ell^{i+2})(k_{\ell,j}^+ - k_{\ell,j}) \\
&= \sum_{\ell=1}^{n}\sum_{j=1}^{m} (\chi_{S(i)\setminus\mathring{S}(i)}(j)\phi_j^i\check{\tau}_j^i - \chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell)\psi_\ell^{i+1}\check{\sigma}_\ell^{i+1})(k_{\ell,j} - k_{\ell,j}^+).
\end{aligned}$$

Expanding $k_{\ell,j} - k^+_{\ell,j}$, using Assumption 3.1, and continuing with Young's inequality, for any $\alpha_x, \alpha_y > 0$,

$$
\begin{aligned}
\mathbb{E}_i[D_i^\Lambda] &= \sum_{\ell=1}^{n}\sum_{j=1}^{m}\Big[(\chi_{S(i)\setminus\mathring{S}(i)}(j)\phi_j^i\check{\tau}_j^i - \chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell)\psi_\ell^{i+1}\check{\sigma}_\ell^{i+1})\\
&\qquad\qquad \cdot \langle y_\ell^{i+1} - \widehat{y}_\ell, [\nabla K(x^i) - \nabla K(x^{i+1})](x_j^{i+1} - \widehat{x}_j)\rangle\Big]\\
&\geq -\sum_{j=1}^{m}\chi_{S(i)\setminus\mathring{S}(i)}(j)\phi_j^i\check{\tau}_j^i \cdot \|y^{i+1} - \widehat{y}\|_{P_{\mathrm{NL}}} L\|x^{i+1} - x^i\|\|x_j^{i+1} - \widehat{x}_j\|\\
&\quad -\sum_{\ell=1}^{n}\chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell)\psi_\ell^{i+1}\check{\sigma}_\ell^{i+1} \cdot \|y_\ell^{i+1} - \widehat{y}_\ell\|_{P_{\mathrm{NL}}} L\|x^{i+1} - x^i\|\|x^{i+1} - \widehat{x}\|\\
&\geq -\sum_{j=1}^{m}\chi_{S(i)\setminus\mathring{S}(i)}(j)\phi_j^i\check{\tau}_j^i\left(\alpha_x\|x_j^{i+1} - \widehat{x}_j\|^2 + \frac{L^2}{4\alpha_x}\|y^{i+1} - \widehat{y}\|^2_{P_{\mathrm{NL}}}\|x^{i+1} - x^i\|^2\right)\\
&\quad -\sum_{\ell=1}^{n}\chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell)\psi_\ell^{i+1}\check{\sigma}_\ell^{i+1}\left(\alpha_y\|y_\ell^{i+1} - \widehat{y}_\ell\|^2_{P_{\mathrm{NL}}} + \frac{L^2}{4\alpha_y}\|x^{i+1} - x^i\|^2\|x^{i+1} - \widehat{x}\|^2\right).
\end{aligned}
$$

This rearranges as (3.23). By (3.24), $\mathbb{P}[d^{i+1} \leq c_*^i \mid O_{i-1}] = 1$. Hence (3.25) follows. □

Remark 3.10. For slightly stronger results, it would in (3.24) and throughout the rest of the manuscript, be possible to take $\rho_x = \rho_x^{i+1}$ and $\rho_\ell = \rho_\ell^{i+1}$ dependent on the iteration.

3.5 PUTTING IT ALL TOGETHER

We are now ready to state our main generic result providing the tool to estimate convergence rates based on growth rates of ϕ_j^i and ψ_ℓ^{i+1}.

Theorem 3.11. *Suppose Assumptions 3.1, 3.2 and 3.4 hold for some $0 < \delta \leq \kappa < 1$, $\gamma_{G,j}, \gamma_{F^*,\ell} \geq 0$, $\gamma_{K,j} \in \mathbb{R}$ ($j = 1, \dots, m$, $\ell = 1, \dots, n$), $L, L_3 \geq 0$, $p \in [1, 2]$, $\theta_A \geq 0$ together with the nesting conditions (2.9), the lower bound (3.8) on the local metric, and the conditionality assumptions (3.17) for all $i \leq N - 1$. For some sequence of $\eta^{i+1} > 0$ assume the* coupling conditions

$$\mathring{\pi}_j^{i+1}\phi_j^{i+1}\mathring{\tau}_j^{i+1} + \chi_{S(i)\setminus\mathring{S}(i)}(j)\phi_j^i\check{\tau}_j^i = \eta^{i+1} \quad (j = 1, \dots, m) \quad \textit{and} \tag{3.27a}$$

$$\mathring{\nu}_\ell^{i+2}\psi_\ell^{i+2}\mathring{\sigma}_\ell^{i+2} + \chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell)\psi_\ell^{i+1}\check{\sigma}_\ell^{i+1} = \eta^{i+1} \quad (\ell = 1, \dots, n). \tag{3.27b}$$

Also assume for some $\rho_x, \rho_\ell \geq 0$ and $\zeta_\ell \geq 0$,

$$1 = \mathbb{P}[\|x^{i+1} - \widehat{x}\| \leq \rho_x, \|Q_\ell(y^{i+1} - \widehat{y})\|_{P_{\mathrm{NL}}} \leq \rho_\ell, (\ell = 1, \dots, n) \mid O_{i-1}] \quad \textit{and} \tag{3.28a}$$

$$\mathbb{E}_{i-1}[\theta_{\Phi_i T_i}] \geq p^{-p}\textstyle\sum_{\ell=1}^{n}\nu_\ell^{i+1}\psi_\ell^{i+1}\sigma_\ell^{i+1}\zeta_\ell^{1-p}\rho_\ell^{2-p} \quad (\ell = 1, \dots, n). \tag{3.28b}$$

Finally, for c_*^i defined in (3.26) for some $\alpha_x, \alpha_y > 0$ let

$$L_j^i := L_3 + (L\|\Omega^i + I\|^2\textstyle\sum_{\ell=1}^{m}\psi_\ell^{i+1}\sigma_\ell^{i+1}\rho_\ell + c_*^i)/\phi_j^i\tau_j^i, \tag{3.29}$$

$$\bar{\gamma}_{GK,j}^i := \gamma_{G,j} + \gamma_{K,j} - \chi_{S(i)\setminus\mathring{S}(i)}(j)\alpha_x, \tag{3.30}$$

$$\bar{\gamma}_{F^*,\ell}^{i+1} := \begin{cases}\gamma_{F^*,\ell}, & Q_\ell P_{\mathrm{NL}} = 0,\\ \gamma_{F^*,\ell} - (p-1)\zeta_\ell - \chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell)\alpha_y, & Q_\ell P_{\mathrm{NL}} \neq 0.\end{cases} \tag{3.31}$$

Then

$$
\begin{aligned}
\delta\sum_{j=1}^{m}\mathbb{E}\left[\phi_j^i\|P_j(x^N - \widehat{x})\|^2\right] + \frac{\kappa - \delta}{1 - \delta}\sum_{\ell=1}^{n}\mathbb{E}\left[\psi_\ell^{i+1}\|Q_\ell(y^N - \widehat{y})\|^2\right]&\\
\leq \mathbb{E}[\|u^N - \widehat{u}\|^2_{Z_{N+1}M_{N+1}}] \leq \mathbb{E}[\|u^0 - \widehat{u}\|^2_{Z_1 M_1}]&
\end{aligned}
\tag{3.32}
$$

holds provided for every $i \le N-1$ both (i) and (ii) are true:

(i) *Either of the* primal test update conditions *holds for every $j = 1, \dots, m$:*

(a) *both $\phi_j^{i+1} \le (1 + 2\chi_{S(i)}(j)\tau_j^i\bar\gamma_{GK,j}^i)\phi_j^i$ and $\delta \ge \chi_{S(i)}(j)L_j^i\tau_j^i$; or*

(b) *for some $\widetilde\gamma_{G,j}^i \in \mathcal{R}(O_{i-1}, \mathbb{R})$, $\widetilde\tau_j^i := (\mathring\pi_j^i\mathring\tau_j^i + (\pi_j^i - \mathring\pi_j^i)\breve\tau_j^i)/\pi_j^i$,*

$$\phi_j^{i+1} = (1 + 2\widetilde\tau_j^i\widetilde\gamma_{G,j}^i)\phi_j^i, \quad \widetilde\tau_j^i\widetilde\gamma_{G,j}^i < \mathbb{E}_{i-1}[\chi_{S(i)}(j)\tau_j^i\bar\gamma_{GK,j}^i], \quad \textit{and} \tag{3.33a}$$

$$\delta \ge \chi_{S(i)}(j)\left(L_j^i\tau_j^i + \frac{2(\tau_j^i\bar\gamma_{GK,j}^i - \mathbb{E}_{i-1}[\chi_{S(i)}(j)\tau_j^i\bar\gamma_{GK,j}^i])(\tau_j^i\bar\gamma_{GK,j}^i - \widetilde\tau_j^i\widetilde\gamma_{G,j}^i)}{\mathbb{E}_{i-1}[\chi_{S(i)}(j)\tau_j^i\bar\gamma_{GK,j}^i] - \widetilde\tau_j^i\widetilde\gamma_{G,j}^i}\right). \tag{3.33b}$$

(ii) *Either of the* dual test update conditions *holds for every $\ell = 1, \dots, n$:*

(a) *$\psi_\ell^{i+1} \le (1 + 2\chi_{V(i+1)}(\ell)\sigma_\ell^{i+1}\bar\gamma_{F^*,\ell}^{i+1})\psi_\ell^{i+1}$; or*

(b) *for some $\widetilde\gamma_{F^*,\ell}^{i+1} \in \mathcal{R}(O_{i-1}, \mathbb{R})$, $\widetilde\sigma_\ell^{i+1} := (\mathring\nu_\ell^{i+1}\mathring\sigma_\ell^{i+1} + (\nu_\ell^{i+1} - \mathring\nu_\ell^{i+1})\breve\sigma_\ell^{i+1})/\nu_\ell^{i+1}$:*

$$\psi_\ell^{i+2} = (1 + 2\widetilde\sigma_\ell^{i+1}\widetilde\gamma_{F^*,\ell}^{i+1})\psi_\ell^{i+1}, \quad \widetilde\sigma_\ell^{i+1}\widetilde\gamma_{F^*,\ell}^{i+1} < \mathbb{E}_{i-1}[\chi_{V(i+1)}(\ell)\sigma_\ell^{i+1}\bar\gamma_{F^*,\ell}^{i+1}], \tag{3.34a}$$

$$\begin{aligned}\frac{\kappa - \delta}{1 - \delta} \ge 2(\sigma_\ell^{i+1}\bar\gamma_{F^*,\ell}^{i+1} - \mathbb{E}_{i-1}[\chi_{V(i+1)}(\ell)\sigma_\ell^{i+1}\bar\gamma_{F^*,\ell}^{i+1}])&\\ \cdot \frac{\chi_{V(i+1)}(\ell)(\sigma_\ell^{i+1}\bar\gamma_{F^*,\ell}^{i+1} - \widetilde\sigma_\ell^{i+1}\widetilde\gamma_{F^*,\ell}^{i+1})}{\mathbb{E}_{i-1}[\chi_{V(i+1)}(\ell)\sigma_\ell^{i+1}\bar\gamma_{F^*,\ell}^{i+1}] - \widetilde\sigma_\ell^{i+1}\widetilde\gamma_{F^*,\ell}^{i+1}}&.\end{aligned} \tag{3.34b}$$

Proof. We first apply Lemma 3.7. Recalling R' from (3.11b), let us set

$$\begin{aligned} R'' &:= R' - 2\begin{pmatrix} \sum_{j=1}^m \tau_j^i\phi_j^i\chi_{S(i)\setminus\mathring S(i)}(j)\alpha_x P_j & 0 \\ 0 & \sum_{\ell=1}^n \sigma_\ell^{i+1}\psi_\ell^{i+1}(\chi_{V(i+1)}(\ell)(p-1)\zeta_\ell + \chi_{V(i+1)\setminus\mathring V(i+1)}(\ell)\alpha_y)Q_\ell P_{\mathrm{NL}} \end{pmatrix} \\ &= \begin{pmatrix} \Phi_i - \Phi_{i+1} + 2\sum_{j\in S(i)}\phi_j^i\tau_j^i\bar\gamma_{GK,j}P_j & 0 \\ 0 & \Psi_{i+1} - \Psi_{i+2} + 2\sum_{\ell\in V(i+1)}\psi_\ell^{i+1}\sigma_\ell^{i+1}\bar\gamma_{F^*,\ell}Q_\ell \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^m q_j^i P_j & 0 \\ 0 & \sum_{\ell=1}^n h_\ell^{i+1}Q_\ell \end{pmatrix} \end{aligned} \tag{3.35}$$

for

$$q_j^i := (1 + 2\chi_{S(i)}(j)\tau_j^i\bar\gamma_{GK,j}^i)\phi_j^i - \phi_j^{i+1} \quad \text{and} \quad h_\ell^{i+1} := (1 + 2\chi_{V(i+1)}(\ell)\sigma_\ell^{i+1}\bar\gamma_{F^*,\ell}^{i+1})\psi_\ell^{i+1} - \psi_\ell^{i+2}.$$

Thus

$$\mathbb{E}_{i-1}[\|u^{i+1} - \widehat u\|_{R''}^2] = \sum_{j=1}^m \mathbb{E}_{i-1}[q_j^i\|P_j(x^{i+1} - \widehat x)\|^2] + \sum_{\ell=1}^n \mathbb{E}_{i-1}[h_\ell^{i+1}\|Q_\ell(y^{i+1} - \widehat y)\|^2]. \tag{3.36}$$

Estimation of q_j^i Suppose $j \in \{1, \dots, m\}$ satisfies (i)(a). Then $q_j^i \ge 0$ and $\delta \ge \chi_{S(i)}(j)L_j^i\tau_j^i$, so we immediately estimate

$$\mathbb{E}_{i-1}[q_j^i\|P_j(x^{i+1} - \widehat x)\|^2] \ge -\mathbb{E}_{i-1}[\chi_{S(i)}(j)(\delta\phi_j^i - L_j^i\phi_j^i\tau_j^i)\|P_j(x^{i+1} - x^i)\|^2]. \tag{3.37}$$

Otherwise, if $j \in \{1, \dots, m\}$ satisfies (i)(b), using (3.18) and that $q_j^i = \mathbb{E}_i[q_j^i]$ due to (3.17) and (3.30), we decompose

$$\begin{aligned}\mathbb{E}_{i-1}[q_j^i\|P_j(x^{i+1} - \widehat x)\|^2] = \mathbb{E}_{i-1}\big[q_j^i\|P_j(x^{i+1} - x^i)\|^2 + \mathbb{E}_{i-1}[q_j^i]\|P_j(x^i - \widehat x)\|^2&\\ + 2q_j^i\langle P_j(x^{i+1} - x^i), x^i - \widehat x\rangle\big]&.\end{aligned}$$

Using $(1-\chi_{S(i)}(j))P_j(x^{i+1}-x^i)=0$ and Young's inequality with the factor $\alpha>0$, we obtain

$$\begin{aligned}\mathbb{E}_{i-1}\big[q_j^i\|P_j(x^{i+1}-\widehat{x})\|^2\big] \ge \mathbb{E}_{i-1}\big[&\chi_{S(i)}(j)(q_j^i-\alpha|q_j^i|)\|P_j(x^{i+1}-x^i)\|^2 \\ &+ (\mathbb{E}_{i-1}[q_j^i]-\chi_{S(i)}(j)\alpha^{-1}|q_j^i|)\|P_j(x^i-\widehat{x})\|^2\big].\end{aligned} \tag{3.38}$$

Since $\phi_j^{i+1}=(1+2\widetilde{\tau}_j^i\widetilde{\gamma}_{G,j}^i)\phi_j^i$ with $\widetilde{\gamma}_{G,j}^i\in\mathcal{R}(O_{i-1};\mathbb{R})$, we have from (3.33a)

$$\mathbb{E}_{i-1}[q_j^i]=(1+2\mathbb{E}_{i-1}[\chi_{S(i)}(j)\tau_j^i\overline{\gamma}_{GK,j}^i])\phi_j^i-\mathbb{E}_{i-1}[\phi_j^{i+1}]=2\phi_j^i(\mathbb{E}[\chi_{S(i)}(j)\tau_j^i\overline{\gamma}_{GK,j}^i]-\widetilde{\tau}_j^i\widetilde{\gamma}_{G,j}^i)>0,$$

and rearranging (3.33b) for $j\in S(i)$:

$$q_j^i=2\phi_j^i(\chi_{S(i)}(j)\tau_j^i\overline{\gamma}_{GK,j}^i-\widetilde{\tau}_j^i\widetilde{\gamma}_{G,j}^i)\ge(\mathbb{E}_{i-1}[q_j^i])^{-1}|q_j^i|^2-\delta\phi_j^i+L_j^i\phi_j^i\tau_j^i.$$

Therefore, taking $\alpha:=(\mathbb{E}_{i-1}[q_j^i])^{-1}|q_j^i|$ for $j\in S(i)$ in (3.38), we verify (3.37) for the case (i)(b) as well.

Estimation of h_ℓ^{i+1} Similarly, if $\ell\in\{1,\ldots,n\}$ satisfies (ii)(a), we have $h_\ell^{i+1}\ge 0$, hence

$$\mathbb{E}_{i-1}[h_\ell^{i+1}\|Q_\ell(y^{i+1}-\widehat{y})\|^2]\ge-\mathbb{E}_{i-1}\Big[\chi_{V(i+1)}(\ell)\frac{\kappa-\delta}{1-\delta}\psi_\ell^{i+1}\|Q_\ell(y^{i+1}-y^i)\|^2\Big]. \tag{3.39}$$

Otherwise, when $\ell\in\{1,\ldots,n\}$ satisfies (ii)(b), using (3.18) and that $h_\ell^{i+1}=\mathbb{E}_i[h_\ell^{i+1}]$ due to (3.17) and (3.31), we estimate for arbitrary $\alpha>0$ that

$$\begin{aligned}\mathbb{E}_{i-1}[h_\ell^{i+1}\|Q_\ell(y^{i+1}-\widehat{y})\|^2]\ge\mathbb{E}_{i-1}[&\chi_{V(i+1)}(\ell)(h_\ell^{i+1}-\alpha|h_\ell^{i+1}|)\|Q_\ell(y^{i+1}-y^i)\|^2 \\ &+(\mathbb{E}_{i-1}[h_\ell^{i+1}]-\chi_{V(i+1)}(\ell)\alpha^{-1}|h_\ell^{i+1}|)\|Q_\ell(y^i-\widehat{y})\|^2].\end{aligned} \tag{3.40}$$

Since $\psi_\ell^{i+2}=(1+2\widetilde{\sigma}_\ell^{i+1}\widetilde{\gamma}_{F^*,\ell}^{i+1})\psi_\ell^{i+1}$ with $\widetilde{\gamma}_{F^*,\ell}^{i+1}\in\mathcal{R}(O_{i-1};\mathbb{R})$, from (3.34a) we have

$$\mathbb{E}_{i-1}[h_\ell^{i+1}]=(1+2\mathbb{E}_{i-1}[\chi_{V(i+1)}(\ell)\sigma_\ell^{i+1}\overline{\gamma}_{F^*,\ell}^{i+1}])\psi_\ell^{i+1}-\mathbb{E}_{i-1}[\psi_\ell^{i+2}]>0$$

and rearranging (3.34b) for $\ell\in V(i+1)$:

$$h_\ell^{i+1}\ge(\mathbb{E}_{i-1}[h_\ell^{i+1}])^{-1}|h_\ell^{i+1}|^2-\frac{\kappa-\delta}{1-\delta}\psi_\ell^{i+1}.$$

Consequently, taking $\alpha:=(\mathbb{E}_{i-1}[h_\ell^{i+1}])^{-1}|h_\ell^{i+1}|$ for $\ell\in V(i+1)$ in (3.40), we obtain (3.39) for the case (ii)(b) as well.

Combining the estimates Since (3.37) and (3.39) hold for all $j=1,\ldots,m$ and $\ell=1,\ldots,n$, respectively, continuing from (3.36), we get

$$\begin{aligned}\mathbb{E}_{i-1}[\|u^{i+1}-\widehat{u}\|_{R''}^2]\ge-\mathbb{E}_{i-1}\Big[&\sum_{j=1}^m(\chi_{S(i)}(j)(\delta\phi_j^i-L_j^i\phi_j^i\tau_j^i)\|P_j(x^{i+1}-x^i)\|^2 \\ &+\sum_{\ell=1}^n\Big(\chi_{V(i+1)}(\ell)\frac{\kappa-\delta}{1-\delta}\psi_\ell^{i+1}\|Q_\ell(y^{i+1}-y^i)\|^2\Big)\Big].\end{aligned}$$

Plugging L_j^i from (3.29), thus

$$\begin{aligned}\mathbb{E}_{i-1}[\|u^{i+1}-\widehat{u}\|_{R''}^2]\ge-\mathbb{E}_{i-1}\Bigg[&\sum_{j=1}^m\chi_{S(i)}(j)\left(\delta\phi_j^i-L\|\Omega^i+I\|^2\sum_{\ell=1}^m\psi_\ell^{i+1}\sigma_\ell^{i+1}\rho_\ell\right)\|P_j(x^{i+1}-x^i)\|^2 \\ &+\sum_{\ell=1}^n\Big(\chi_{V(i+1)}(\ell)\frac{\kappa-\delta}{1-\delta}\psi_\ell^{i+1}\|Q_\ell(y^{i+1}-y^i)\|^2\Big)-\sum_{j=1}^m(\chi_{S(i)}(j)(L_3\phi_j^i\tau_j^i+c_*^i)\|P_j(x^{i+1}-x^i)\|^2\Bigg].\end{aligned}$$

By the definitions of R_x in (3.11) and ρ_ℓ in (3.28a), we continue

$$\mathbb{E}_{i-1}[\|u^{i+1}-\widehat{u}\|^2_{R''}] \ge -\mathbb{E}_{i-1}\Big[\|x^{i+1}-x^i\|^2_{R_x} + \frac{\kappa-\delta}{1-\delta}\|y^{i+1}-y^i\|^2_{\Psi_{i+1}} - \sum_{j=1}^m \chi_{S(i)}(j)(L_3\phi^i_j\tau^i_j + c^i_*)\|P_j(x^{i+1}-x^i)\|^2\Big]. \tag{3.41}$$

On the other hand, by the definition of R'' in (3.35),

$$\begin{aligned}\mathbb{E}_{i-1}[\|u^{i+1}-\widehat{u}\|^2_{R''}] = \mathbb{E}_{i-1}\Big[&\|u^{i+1}-\widehat{u}\|^2_{R'} - 2\alpha_x\sum_{j=1}^m \tau^i_j\phi^i_j\chi_{S(i)\setminus\mathring{S}(i)}(j)\|P_j(x^{i+1}-\widehat{x})\|^2 \\ &- 2\sum_{\ell=1}^n(\chi_{V(i+1)}(\ell)(p-1)\zeta_\ell + \chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell)\alpha_y)\sigma^{i+1}_\ell\psi^{i+1}_\ell\|Q_\ell(y^{i+1}-\widehat{y})\|^2_{P_{\mathrm{NL}}}\Big].\end{aligned}$$

Combining with (3.41) and rearranging the terms, we therefore have

$$\mathbb{E}_{i-1}[\|u^{i+1}-\widehat{u}\|^2_{R'} + \|x^{i+1}-x^i\|^2_{R_x} + \frac{\kappa-\delta}{1-\delta}\|y^{i+1}-y^i\|^2_{\Psi_{i+1}}] \ge \mathbb{E}_{i-1}[b_1+b_2] \tag{3.42}$$

for

$$b_1 := \sum_{j=1}^n \chi_{S(i)}(j)L_3\phi^i_j\tau^i_j\|P_j(x^{i+1}-x^i)\|^2 + 2\sum_{\ell=1}^n \sigma^{i+1}_\ell\psi^{i+1}_\ell\chi_{V(i+1)}(\ell)(p-1)\zeta_\ell\|Q_\ell(y^{i+1}-\widehat{y})\|^2_{P_{\mathrm{NL}}},$$

and

$$\begin{aligned}b_2 := {}& 2\alpha_x\sum_{j=1}^m \tau^i_j\phi^i_j\chi_{S(i)\setminus\mathring{S}(i)}(j)\|P_j(x^{i+1}-\widehat{x})\|^2 \\ &+ 2\alpha_y\sum_{\ell=1}^n \sigma^{i+1}_\ell\psi^{i+1}_\ell\chi_{V(i+1)\setminus\mathring{V}(i+1)}(\ell)\|Q_\ell(y^{i+1}-\widehat{y})\|^2_{P_{\mathrm{NL}}} + \sum_{j=1}^n \chi_{S(i)}(j)c^i_*\|P_j(x^{i+1}-x^i)\|^2.\end{aligned}$$

Our conditions (3.28) and $\delta \ge \chi_{S(i)}(j)L^i_j\tau^i_j$ ensure the conditions of Lemmas 3.8 and 3.9. By Lemma 3.8 thus $\mathbb{E}_{i-1}[b_1+2D^K_i] \ge 0$ while using both (3.23) and (3.25) of Lemma 3.9 establishes $\mathbb{E}_{i-1}[b_2+2D^\Lambda_i] = \mathbb{E}_{i-1}[b_2+2\mathbb{E}_i[D^\Lambda_i]] \ge 0$. Consequently (3.42) yields

$$\mathbb{E}_{i-1}[\|u^{i+1}-\widehat{u}\|^2_{R'} + \|x^{i+1}-x^i\|^2_{R_x} + \frac{\kappa-\delta}{1-\delta}\|y^{i+1}-y^i\|^2_{\Psi_{i+1}} + 2D^\Lambda_i + 2D^K_i] \ge 0.$$

We now use Lemma 3.7 to verify (2.5). Minding that each $Z_{i+1}M_{i+1}$ is self-adjoint by Lemma 3.5, a referral to Theorem 2.1 establishes (2.6). Using (3.8) as well as $\phi^N_j, \psi^{N+1}_\ell \in \mathcal{R}(O_{N-1};(0,\infty))$ and $u^N \in \mathcal{R}(O_{N-1}; X\times Y)$ that follow from (3.17), we estimate

$$\begin{aligned}\mathbb{E}[\|u^N-\widehat{u}\|^2_{Z_{N+1}M_{N+1}} \mid O_{N-1}] &= \|u^N-\widehat{u}\|^2_{\mathbb{E}[Z_{N+1}M_{N+1}|O_{N-1}]} \\ &\ge \delta\sum_{j=1}^m \phi^N_j\|P_j(x^N-\widehat{x})\|^2 + \frac{\kappa-\delta}{1-\delta}\sum_{\ell=1}^n \psi^{N+1}_\ell\|Q_\ell(y^N-\widehat{y})\|^2.\end{aligned}$$

Taking the full expectation and using (2.6) establishes the claim. □

Remark 3.12. The conditions (i)(a) and (ii)(a) differ from (i)(b) and (ii)(b) by larger $\overline{\gamma}^i_{GK,j}$ and $\overline{\gamma}^i_{F^*,\ell}$, and updating ϕ^{i+1}_j and $\psi^{i+2}_\ell \in \mathcal{R}(O_i;\mathbb{R})$ potentially non-deterministically.

In Section 4 we have $\mathring{\pi}_j^i = \pi_j^i$, $\tau_j^i = \mathring{\tau}_j^i$, $\mathring{\nu}_\ell^{i+1} = 0$, and $\sigma_\ell^{i+1} = \breve{\sigma}_\ell^{i+1}$. In Section 5 we take $\mathring{\pi}_j^i = 0$, $\tau_j^i = \breve{\tau}_j^i$, $\mathring{\nu}_\ell^{i+1} = \nu_\ell^{i+1}$, and $\sigma_\ell^{i+1} = \mathring{\sigma}_\ell^{i+1}$. Also (i)(b) and (ii)(b) then simplify for $\widetilde{\gamma}_{G,j}^i < \pi_j^i \overline{\gamma}_{GK,j}^i$ to

$$\text{(3.43a)} \qquad \phi_j^{i+1} = (1 + 2\tau_j^i \widetilde{\gamma}_{G,j}^i)\phi_j^i, \qquad \delta \geq \chi_{S(i)}(j)\tau_j^i \left(L_j^i + 2(1-\pi_j^i)\overline{\gamma}_{GK,j}^i \frac{\overline{\gamma}_{GK,j}^i - \widetilde{\gamma}_{G,j}^i}{\pi_j^i \overline{\gamma}_{GK,j}^i - \widetilde{\gamma}_{G,j}^i} \right)$$

and, respectively, for $\widetilde{\gamma}_{F^*,\ell}^{i+1} < \nu_\ell^{i+1} \overline{\gamma}_{F^*,\ell}^{i+1}$ to

$$\text{(3.43b)} \qquad \psi_\ell^{i+2} = (1 + 2\sigma_\ell^{i+1}\widetilde{\gamma}_{F^*,\ell}^{i+1})\psi_\ell^{i+1}, \qquad \frac{\kappa - \delta}{1-\delta} \geq 2\chi_{V(i+1)}(\ell)(1-\nu_\ell^{i+1})\sigma_\ell^{i+1}\overline{\gamma}_{F^*,\ell}^{i+1} \frac{\overline{\gamma}_{F^*,\ell}^{i+1} - \widetilde{\gamma}_{F^*,\ell}^{i+1}}{\nu_\ell^{i+1}\overline{\gamma}_{F^*,\ell}^{i+1} - \widetilde{\gamma}_{F^*,\ell}^{i+1}}.$$

Remark 3.13. Another quite restrictive requirement that we will need in the next sections is the almost sure boundedness of the iterates in (3.28a). We already had this requirement in the deterministic single-block algorithm in [9, Section 4.3] and [10, Section 5]. We verified in [9, Proposition 4.8.] that this requirement can be restated in terms of the sufficiently close initialisation of iterations to the critical point, which is often required in non-convex optimisation.

In this work, the rates for convergence are in expectation, hence, the required boundedness is in the almost sure terms. Moreover, in order to be able to update only some primal blocks on each iteration, (3.28a) now also requires the primal variable to be bounded. Through the simplified algorithms of Sections 4 and 5, treating respective non-randomised dual updates and non-randomised primal updates, we will somewhat relax these restrictions:

- Algorithm 4.2 of Section 4 will not require the dual variable to be bounded if Assumption 3.2 holds with $p = 2$; see Corollaries 4.6 and 4.8.
- In Section 5, we will not require any bound on the primal variable.

In some cases, boundedness can, moreover, be checked analytically based on the explicit formula for F. For example, for $F(x) = |\langle a, x\rangle|$ or $F(x) = \|ax\|$ the support of $F^*(y)$ is bounded by $\|a\|$. Hence the range of the corresponding proximal operator is also bounded. In particular, if the F is of such a form, the boundedness assumptions of Section 5 are automatically satisfied.

4 METHODS WITH FULL DUAL UPDATES

We now develop more specific methods based on (2.13) and study their convergence based on Theorem 3.11. In this section we take $\mathring{V}(i+1) = \emptyset$, $V(i+1) = \{1, \ldots, n\}$, and $\mathring{S}(i) = S(i)$ for all iterations i. The nesting conditions (2.9) of Theorem 3.11 then hold, and the coupling conditions (3.27) become

$$\text{(4.1)} \qquad \mathring{\pi}_j^{i+1}\phi_j^{i+1}\mathring{\tau}_j^{i+1} = \eta^{i+1} = \psi_\ell^{i+1}\breve{\sigma}_\ell^{i+1}.$$

The dual update of (2.13) involves $\Psi_{i+1}^{-1}[\nabla K(x^i)T_i^*\Phi_i^* - \Psi_{i+1}\Sigma_{i+1}\nabla K(x^i)\Omega_i]$, in scalar form

$$\text{(4.2)} \qquad \frac{\phi_j^i\mathring{\tau}_j^i - \omega_j^i\breve{\sigma}_\ell^{i+1}\psi_\ell^{i+1}}{\psi_\ell^{i+1}} = \breve{\sigma}_\ell^{i+1}\left(\frac{\eta^i}{\mathring{\pi}_j^i\eta^{i+1}} - \omega_j^i\right) = \breve{\sigma}_\ell^{i+1}\left(\frac{\bar{\omega}^i}{\mathring{\pi}_j^i} - \omega_j^i\right) \quad \text{for} \quad \bar{\omega}^i := \frac{\eta^i}{\eta^{i+1}}.$$

Therefore, with $\omega_j^i = \frac{\bar{\omega}^i}{\mathring{\pi}_j^i}$, the updates (2.13) simplify to those of Algorithm 4.1. Moreover, (2.12) reduces to $\lambda_{j,\ell}^i = \phi_j^i\tau_j^i\chi_{\mathring{S}(i)}(j)$. We thus verify (3.8) via:

Lemma 4.1. *Suppose $\mathring{V}(i+1) = \emptyset$, $V(i+1) = \{1, \ldots, n\}$, $\mathring{S}(i) = S(i)$ for $i \in \mathbb{N}$; the coupling condition (4.1) holds; $\bar{\omega}^i \leq 1$; as well as, for all $\ell = 1, \ldots, n$ and $j = 1, \ldots, m$,*

$$\text{(4.3)} \qquad \bar{\omega}^i\breve{\sigma}_\ell^{i+1}\mathring{\tau}_j^i \leq \breve{\sigma}_\ell^0\mathring{\tau}_j^0 \quad \textit{and} \quad 1 - \kappa \geq \left\| \sum_{j \in \mathring{S}(i)} \sqrt{\frac{w_{j,\ell}^i\breve{\sigma}_\ell^0\mathring{\tau}_j^0}{\mathring{\pi}_j^i}} Q_\ell \nabla K(x^i) P_j \right\|^2$$

Algorithm 4.1 Full dual updates #1

Assume the problem structure (P), equivalently (S). For each iteration $i \in \mathbb{N}$, choose a sampling pattern for generating the random set of updated primal blocks $S(i) \in \mathcal{R}(O_i; \mathcal{P}\{1, \dots, m\})$ with corresponding blockwise probabilities $\mathring{\pi}_j^i := \mathbb{P}[j \in S(i) \mid O_{i-1}] > 0$. Also choose a rule for the iteration and block-dependent step length parameters $\mathring{\tau}_j^i, \breve{\sigma}_\ell^i, \bar{\omega}^i > 0$ from one of Theorem 4.5, 4.4, or 4.7. Pick an initial iterate (x^0, y^0) and on each iteration $i \in \mathbb{N}$ update all blocks $x_j^{i+1} = P_j x^{i+1}$, $(j = 1, \dots, m)$, and $y_\ell^{i+1} = Q_\ell y^{i+1}$, $(\ell = 1, \dots, n)$, of x^{i+1} and y^{i+1} as:

$$
\begin{aligned}
x_j^{i+1} &:= \begin{cases} (I + \mathring{\tau}_j^i P_j \partial G_j P_j)^{-1}(x_j^i - \mathring{\tau}_j^i P_j \nabla K(x^i)^* y^i), & j \in S(i), \\ x_j^i, & j \notin S(i), \end{cases} \\
\bar{x}_j^{i+1} &:= \begin{cases} x_j^{i+1} + \bar{\omega}^i (x_j^{i+1} - x_j^i)/\mathring{\pi}_j^i, & j \in S(i), \\ x_j^i, & j \notin S(i), \end{cases} \\
y_\ell^{i+1} &:= (I + \breve{\sigma}_\ell^{i+1} Q_\ell \partial F_\ell^* Q_\ell)^{-1}(y_\ell^i + \breve{\sigma}_\ell^{i+1} Q_\ell K(\bar{x}^{i+1})).
\end{aligned}
$$

for some $0 \le \kappa \le 1$ *and* $w_{j,\ell,k} = 1/w_{j,k,\ell} > 0$ *such that*

$$
w_{j,\ell}^i := \chi_{\mathcal{V}_j^i}(\ell) \sum_{k \in \mathcal{V}_j^i(\ell)} w_{j,\ell,k} \tag{4.4a}
$$

with

$$
\mathcal{V}_j^i(\ell) = \{k \in \{1, \dots, n\} \mid Q_\ell \nabla K(x^i) P_j \nabla K(x^i)^* Q_k \neq 0,\ j \in \mathring{S}(i)\}. \tag{4.4b}
$$

Then the lower bound (3.8) *holds.*

Proof. By the first part of (4.3), (4.1), and $\lambda_{j,\ell}^i = \phi_j^i \tau_j^i \chi_{\mathring{S}(i)}(j)$, we have

$$
\breve{\sigma}_\ell^0 \mathring{\tau}_j^0 \ge \frac{\eta^i \breve{\sigma}_\ell^{i+1} \mathring{\tau}_j^i}{\eta^{i+1}} = \frac{\mathring{\pi}_j^i \phi_j^i (\mathring{\tau}_j^i)^2}{\psi_\ell^{i+1}} = \frac{\mathring{\pi}_j^i (\lambda_{j,\ell}^i)^2}{\psi_\ell^{i+1} \phi_j^i} \quad (j \in \mathring{S}(i)).
$$

By the orthogonality of the projections P_j, we may insert this estimation into the second part of (4.3), obtaining (3.7); compare the proof of Lemma 3.5. The definition of $\mathcal{V}_j^i(\ell)$ in (3.5) also reduces to that in (4.4b), while the definition of $w_{j,\ell}^i$ in (4.4a) is exactly that in (3.6). We finish by applying Lemma 3.5 to verify (3.8). ◻

Remark 4.2. The first part of (4.3) relaxes the property $\tau^i \sigma^i = \tau^0 \sigma^0$ of the basic PDPS [7].

Remark 4.3. With deterministic updates ($\mathring{\pi}_j^i \equiv 1$), (4.1) couples $\mathring{\tau}_j^i \phi_j^i = \breve{\sigma}_\ell^i \psi_\ell^i$. With $\psi_\ell^i \equiv \psi_\ell^0$, (4.3) therefore becomes a block-coupled variant of the condition $\tau_i \sigma_i \|K\|^2 < 1$ from [7].

Finally, we also remind that (3.30) and (3.31) for this section simplify to

$$
\bar{\gamma}_{GK,j}^i \equiv \bar{\gamma}_{GK,j} := \gamma_{G,j} + \gamma_{K,j}, \quad \text{and} \quad \bar{\gamma}_{F^*,\ell}^{i+1} \equiv \bar{\gamma}_{F^*,\ell} := \begin{cases} \gamma_{F^*,\ell}, & Q_\ell P_{\mathrm{NL}} = 0, \\ \gamma_{F^*,\ell} - (p-1)\zeta_\ell - \alpha_y, & Q_\ell P_{\mathrm{NL}} \neq 0. \end{cases} \tag{4.5}
$$

4.1 ACCELERATED RATES

We start with simple step length rules for $O(1/N)$ rates on the blocks admitting second-order growth ($\gamma_{G,j} + \gamma_{K,j} > 0$ for primal blocks j or $\gamma_{F^*,\ell} > 0$ for dual blocks ℓ). Throughout, for simplicity, we assume iteration-independent probabilities, $\mathring{\pi}_j^i = \pi_j^i \equiv \mathring{\pi}_j$ for all $i \in \mathbb{N}$.

Theorem 4.4. *Suppose Assumptions 3.1, 3.2 and 3.4 hold with $L, L_3 \geq 0$; $p \in [1,2]$; $\gamma_{G,j} + \gamma_{K,j} \geq 0$, $(j = 1, \ldots, m)$, and $\overline{\gamma}_{F^*,\ell} \geq 0$, $(\ell = 1, \ldots, n)$, for some $\alpha_y, \zeta_\ell \geq 0$ as defined in (4.5). Let the iterates $\{u^i = (x^i, y^i)\}_{i \in \mathbb{N}}$ be generated by Algorithm 4.1 with iteration-independent probabilities $\mathring{\pi}_j^i \equiv \mathring{\pi}_j$ and step length parameters*

$$\breve{\sigma}_\ell^{i+1} := \frac{\breve{\sigma}_\ell^i}{1 + 2\breve{\sigma}_\ell^i \overline{\gamma}_{F^*,\ell}}, \quad \overline{\omega}^i \equiv 1, \quad \text{and} \quad \mathring{\tau}_j^{i+1} := \frac{\mathring{\tau}_j^i}{1 + 2\mathring{\tau}_j^i \widetilde{\gamma}_{G,j}}, \tag{4.6a}$$

with either $0 \leq \widetilde{\gamma}_{G,j} < \mathring{\pi}_j(\gamma_{G,j} + \gamma_{K,j})$ or $\widetilde{\gamma}_{G,j} = \gamma_{G,j} + \gamma_{K,j} = 0$ for each $j = 1, \ldots, m$; and initial $\mathring{\tau}_j^0, \breve{\sigma}_\ell^0 > 0$ satisfying for some $0 < \delta < \kappa < 1$, $\rho_x, \rho_\ell \geq 0$, $(\ell = 1, \ldots, n)$, and $w_{j,\ell}^i$ as in (4.4) the bounds $(i \in \mathbb{N};\ j = 1, \ldots, m)$

$$1 - \kappa \geq \left\| \sum_{j \in \mathring{S}(i)} \sqrt{\frac{w_{j,\ell}^i \breve{\sigma}_\ell^0 \mathring{\tau}_j^0}{\mathring{\pi}_j}} Q_\ell \nabla K(x^i) P_j \right\|^2 \tag{4.7a}$$

and

$$\delta \geq \mathring{\tau}_j^0 \overline{L} + \mathring{\tau}_j^0 \cdot \begin{cases} 2(1 - \mathring{\pi}_j)(\gamma_{G,j} + \gamma_{K,j}) \frac{\gamma_{G,j} + \gamma_{K,j} - \widetilde{\gamma}_{G,j}}{\mathring{\pi}_j(\gamma_{G,j} + \gamma_{K,j}) - \widetilde{\gamma}_{G,j}} & \gamma_{G,j} + \gamma_{K,j} > 0 \\ 0 & \gamma_{G,j} + \gamma_{K,j} = 0 \end{cases} \tag{4.7b}$$

with

$$\overline{L} := L_3 + L\Big(\max_{j=1\ldots m} \Big(\frac{1}{\mathring{\pi}_j} + 1\Big)^2 \textstyle\sum_{\ell=1}^n \rho_\ell + \frac{nL}{2\alpha_y} \rho_x^2 \Big). \tag{4.7c}$$

Assume for $A := \sum_{j \in S(i)} (\mathring{\pi}_j)^{-1} P_j$ that

$$\mathbb{E}_{i-1}[\theta_A] \geq p^{-p} \textstyle\sum_{\ell=1}^n \zeta_\ell^{1-p} \rho_\ell^{2-p} \quad \text{and} \tag{4.8a}$$

$$1 = \mathbb{P}[\|x^{i+1} - \widehat{x}\| \leq \rho_x, \|Q_\ell(y^{i+1} - \widehat{y})\|_{P_{\mathrm{NL}}} \leq \rho_\ell, (\ell = 1, \ldots, n) \mid O_{i-1}]. \tag{4.8b}$$

Then $\mathbb{E}[\|P_j(x^N - \widehat{x})\|^2] \to 0$ at the rate $O(1/N)$ for all j such that $\widetilde{\gamma}_{G,j} > 0$ and $\mathbb{E}[\|Q_\ell(y^N - \widehat{y})\|^2] \to 0$ at the rate $O(1/N)$ for all ℓ such that $\overline{\gamma}_{F^*,\ell} > 0$.

Proof. We use Theorem 3.11 whose conditions we need to verify. We have already verified the nesting condition (2.9) for $\mathring{V}(i+1) = \emptyset$, $V(i+1) = \{1, \ldots, n\}$, and $\mathring{S}(i) = S(i)$ in Algorithm 4.1. The coupling condition (3.27) we have reduced to (4.1), which we now verify. For some $\eta^0 > 0$ we set $\eta^i \equiv \eta^0$, $\phi_j^0 := \eta^0(\mathring{\pi}_j \mathring{\tau}_j^0)^{-1}$, and $\psi_\ell^0 := \eta^0 / \breve{\sigma}_\ell^0$. Then we update

$$\phi_j^{i+1} = (1 + 2\mathring{\tau}_j^i \widetilde{\gamma}_{G,j}) \phi_j^i, \qquad \psi_\ell^{i+2} = (1 + 2\breve{\sigma}_\ell^{i+1} \overline{\gamma}_{F^*,\ell}) \psi_\ell^{i+1}. \tag{4.9}$$

By (4.6), consequently, $\breve{\sigma}_\ell^{i+1} \psi_\ell^{i+1} = \eta^{i+1} = \mathring{\pi}_j \phi_j^{i+1} \mathring{\tau}_j^{i+1}$ for all ℓ and j. Consequently (4.1) holds. Clearly so does (3.17) due the deterministic step length and testing parameter updates. The conditions (3.28) follow from (4.8) given that $\theta_{\Phi_i T_i} = \eta^i \theta_A = \eta^{i+1} \theta_A = \breve{\sigma}_\ell^{i+1} \psi_\ell^{i+1} \theta_A$.

The step length parameters $\mathring{\tau}_j^i$ and $\breve{\sigma}_\ell^{i+1}$ are non-increasing in i by the defining (4.6). Also using (4.7a), we thus verify (4.3). Now Lemma 4.1 verifies (3.8).

We still need to verify Theorem 3.11 (i) and (ii). Regarding the latter, $\psi_\ell^{i+2} \leq (1 + 2\breve{\sigma}_\ell^{i+1} \overline{\gamma}_{F^*,\ell}^{i+1}) \psi_\ell^{i+1}$ trivially as long as $\overline{\gamma}_{F^*,\ell}^{i+1} \geq 0$, which follows from the assumptions on $\gamma_{F^*,\ell}$. Therefore Theorem 3.11 (ii) option (a) holds. Regarding Theorem 3.11 (i), we first of all observe that (3.26) reduces to $c_*^i = nL^2 \eta^{i+1} \rho_x^2 / (2\alpha_y)$. Moreover, in Algorithm 4.1 we took $\omega_j^i := \overline{\omega}^i / \mathring{\pi}_j = 1/\mathring{\pi}_j$ by (4.6). Consequently (3.29) becomes

$$\begin{aligned} L_j^i &:= L_3 + \Big(L \max_{j \in S(i)} (\omega_j^i + 1)^2 \textstyle\sum_{\ell=1}^m \psi_\ell^{i+1} \breve{\sigma}_\ell^{i+1} \rho_\ell + \frac{nL^2 \eta^{i+1} \rho_x^2}{2\alpha_y} \Big) \frac{1}{\phi_j^i \mathring{\tau}_j^i} \\ &= L_3 + L\mathring{\pi}_j \Big(\max_{j \in S(i)} (1/\mathring{\pi}_j + 1)^2 \textstyle\sum_{\ell=1}^n \rho_\ell + \frac{nL}{2\alpha_y} \rho_x^2 \Big) \frac{\eta^{i+1}}{\eta^i} \leq \overline{L}. \end{aligned} \tag{4.10}$$

We now consider two cases for the satisfaction of Theorem 3.11 (i) option (a) or (b):

(A) If $\gamma_{G,j} + \gamma_{K,j} = 0$, then $\widetilde{\gamma}_{G,j} = 0$ and $\phi_j^{i+1} = \phi_j^i$ by (4.9), so option (a) holds.

(B) If $\gamma_{G,j} + \gamma_{K,j} > 0$, then (4.7b), (4.10), and $\mathring{\tau}_j^i \le \mathring{\tau}_j^0$ show (3.43a), hence (b).

We can now apply Theorem 3.11 to obtain (3.32). From (4.9) we have

$$\begin{aligned}\phi_j^{i+1} &= \phi_j^i + 2\widetilde{\gamma}_{G,j}\eta^i/\mathring{\pi}_j = \phi_j^i + 2\widetilde{\gamma}_{G,j}\eta^1/\mathring{\pi}_j = \ldots = \phi_j^1 + 2i\widetilde{\gamma}_{G,j}\eta^1/\mathring{\pi}_j \quad \text{and}\\ \psi_\ell^{i+2} &= \psi_\ell^{i+1} + 2\bar{\gamma}_{F^*,\ell}\eta^{i+1} = \psi_\ell^{i+1} + 2\bar{\gamma}_{F^*,\ell}\eta^1 = \ldots = \psi_\ell^1 + 2(i+1)\bar{\gamma}_{F^*,\ell}\eta^1.\end{aligned}$$

Therefore, for any j such that $\widetilde{\gamma}_{G,j} > 0$ and ℓ such that $\bar{\gamma}_{F^*,\ell} > 0$, ϕ_j^N and ψ_ℓ^{N+1} grow as $\Omega(N)$. This together with (3.32) gives the claim. ◻

We can improve the convergence to $O(1/N^2)$ in the primal variable if all the primal blocks exhibit second-order growth. This is achieved by making the dual step lengths grow as in the basic single-block convex case of [7].

Theorem 4.5. *Suppose Assumptions 3.1, 3.2 and 3.4 hold with $L, L_3 \ge 0$; $p \in [1,2]$; $\gamma_{G,j} + \gamma_{K,j} > 0$, $(j = 1,\ldots,m)$, and $\bar{\gamma}_{F^*,\ell} \ge 0$, $(\ell = 1,\ldots,n)$, for some $\alpha_y, \zeta_\ell \ge 0$ as defined in (4.5). Let the iterates $\{u^i = (x^i, y^i)\}_{i\in\mathbb{N}}$ be generated by Algorithm 4.1 with iteration-independent probabilities $\mathring{\pi}_j^i \equiv \mathring{\pi}_j$ and step length parameters*

$$\breve{\sigma}_\ell^{i+1} := \frac{\breve{\sigma}_\ell^i}{\bar{\omega}^i}, \quad \mathring{\tau}_j^{i+1} := \frac{1}{1+2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j}}\frac{\mathring{\tau}_j^i}{\bar{\omega}^i}, \quad \text{and} \quad \bar{\omega}^i := \max_{j=1,\ldots,m}\frac{1}{\sqrt{1+2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j}}} \tag{4.11}$$

with $0 < \widetilde{\gamma}_{G,j} < \mathring{\pi}_j(\gamma_{G,j} + \gamma_{K,j})$; and initial $\mathring{\tau}_j^0, \breve{\sigma}_\ell^0 > 0$ satisfying for some $0 < \delta \le \kappa < 1$, $\rho_x, \rho_\ell \ge 0$, $(\ell = 1,\ldots,n)$, and $w_{j,\ell}^i$ as in (4.4) the bounds

$$1-\kappa \ge \left\| \sum_{j\in\mathring{S}(i)} \sqrt{\frac{w_{j,\ell}^i\breve{\sigma}_\ell^0\mathring{\tau}_j^0}{\mathring{\pi}_j}} Q_\ell \nabla K(x^i) P_j \right\|^2 \quad (i \in \mathbb{N}) \quad \text{and} \tag{4.12a}$$

$$\delta \ge \mathring{\tau}_j^0 \left(\bar{L} + 2(1-\mathring{\pi}_j)(\gamma_{G,j}+\gamma_{K,j}) \frac{\gamma_{G,j}+\gamma_{K,j}-\widetilde{\gamma}_{G,j}}{\mathring{\pi}_j(\gamma_{G,j}+\gamma_{K,j}) - \widetilde{\gamma}_{G,j}} \right) \quad \text{with} \tag{4.12b}$$

$$\bar{L} := L_3 + \frac{L}{\bar{\omega}^0}\left(\max_{j=1\ldots m}\left(\frac{1}{\mathring{\pi}_j}+1\right)^2 \textstyle\sum_{\ell=1}^n \rho_\ell + \frac{nL}{2\alpha_y}\rho_x^2 \right). \tag{4.12c}$$

Assume for $A := \sum_{j\in S(i)}(\mathring{\pi}_j)^{-1}P_j$ that

$$\mathbb{E}_{i-1}[\theta_A] \ge p^{-p}\textstyle\sum_{\ell=1}^n \zeta_\ell^{1-p}\rho_\ell^{2-p}/\bar{\omega}^0 \quad \text{and} \tag{4.13a}$$

$$1 = \mathbb{P}[\|x^{i+1}-\widehat{x}\| \le \rho_x, \|Q_\ell(y^{i+1}-\widehat{y})\|_{P_{\mathrm{NL}}} \le \rho_\ell, (\ell = 1,\ldots,n) \mid O_{i-1}]. \tag{4.13b}$$

Then $\mathbb{E}[\|P_j(x^N - \widehat{x})\|^2] \to 0$ at the rate $O(1/N^2)$ for all j.

Proof. We use Theorem 3.11 whose conditions we need to verify. We have already verified the nesting conditions (2.9) for the choices $\mathring{V}(i+1) = \emptyset$, $V(i+1) = \{1,\ldots,n\}$, and $\mathring{S}(i) = S(i)$ in Algorithm 4.1. The coupling condition (3.27) we have reduced to (4.1). To verify (4.1), we initialise $\phi_j^0 := \eta^0(\mathring{\pi}_j^0\mathring{\tau}_j^0)^{-1}$ and $\psi_\ell^0 := \eta^0/\breve{\sigma}_\ell^0$ for some $\eta^0 > 0$, and update

$$\phi_j^{i+1} := (1+2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j})\phi_j^i, \quad \psi_\ell^{i+1} := \psi_\ell^i, \quad \text{and} \quad \eta^{i+1} := \eta^i/\bar{\omega}^i. \tag{4.14}$$

Then from (4.11), $\psi_\ell^{i+1}\breve{\sigma}_\ell^{i+1} = \psi_\ell^i\breve{\sigma}_\ell^i/\bar{\omega}^i$ and $\phi_j^{i+1}\mathring{\tau}_j^{i+1} = \phi_j^i\mathring{\tau}_j^i/\bar{\omega}^i$. Therefore, (4.1) holds by induction. Clearly also (3.17) holds due to the step length and testing parameters being updated deterministically. The conditions (3.28) follow from (4.13) and (4.1) given that $\mathring{\tau}_j^i$ decreases so $\bar{\omega}^i \geq \bar{\omega}_0$ and $\theta_{\Phi_i T_i} = \eta^i \theta_A = \eta^{i+1}\bar{\omega}^i\theta_A$.

We now verify (3.8). By (4.11) and (4.14), we get $\phi_j^{i+1}(\mathring{\tau}_j^{i+1})^2 \leq \phi_j^i(\mathring{\tau}_j^i)^2$. This and (4.1) yield

$$\bar{\omega}^i \breve{\sigma}_\ell^{i+1}\mathring{\tau}_j^i = \frac{\eta^i\mathring{\tau}_j^i}{\psi_\ell^{i+1}} = \frac{\phi_j^i(\mathring{\tau}_j^i)^2}{\psi_\ell^{i+1}\mathring{\pi}_j} \leq \frac{\phi_j^0(\mathring{\tau}_j^0)^2}{\psi_\ell^{i+1}\mathring{\pi}_j} = \frac{\eta^0\mathring{\tau}_j^0}{\psi_\ell^0} = \breve{\sigma}_\ell^0\mathring{\tau}_j^0.$$

Combining this estimate with (4.12a) we verify (4.3). Thus Lemma 4.1 establishes (3.8).

We still need to verify Theorem 3.11 (i) and (ii). Regarding the dual test, $\psi_\ell^{i+2} = \psi_\ell^{i+1} \leq (1 + 2\breve{\sigma}_\ell^{i+1}\bar{\gamma}_{F^*,\ell}^{i+1})\psi_\ell^{i+1}$ trivially as long as $\bar{\gamma}_{F^*,\ell}^{i+1} \geq 0$, which follows from the assumptions on $\gamma_{F^*,\ell}$. Therefore Theorem 3.11 (ii) option (a) holds. As far as Theorem 3.11 (i) is concerned, we observe that (3.26) reduces to $c_*^i = nL^2\eta^{i+1}\rho_x^2/(2\alpha_y)$. Consequently (3.29) becomes

$$L_j^i := L_3 + L\mathring{\pi}_j\big(\max_{j\in S(i)}(\omega_j^i+1)^2\textstyle\sum_{\ell=1}^n\rho_\ell + \frac{nL}{2\alpha_y}\rho_x^2\big)\eta^{i+1}/\eta^i \leq \bar{L} \tag{4.15}$$

thanks to $\omega_j^i := \bar{\omega}^i/\mathring{\pi}_j \leq 1/\mathring{\pi}_j$ and $\bar{\omega}^i \geq \bar{\omega}^0$. Also, with $\widetilde{\gamma}_{G,j}^i < (\mathring{\pi}_j\gamma_{G,j} + \gamma_{K,j})$, (4.12b), (4.15), and $\mathring{\tau}_j^i \leq \mathring{\tau}_j^0$ show (3.43a), hence, (3.33). Therefore, Theorem 3.11(i) option (b) holds for every $j = 1,\ldots,m$.

We can thus apply Theorem 3.11 to obtain (3.32). Multiplying the τ update of (4.11) by $2\widetilde{\gamma}_{G,j}$, plugging in $\bar{\omega}^i$, and taking the inverse, we have

$$(2\mathring{\tau}_j^{i+1}\widetilde{\gamma}_{G,j})^{-1} = \frac{1+2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j}}{2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j}\sqrt{1+\min_{j=1\ldots m}(2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j})}} = \frac{1+(2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j})^{-1}}{\sqrt{1+(\max_{j=1\ldots m}(2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j})^{-1})^{-1}}}$$

We now apply Lemma B.1 with $z_j^i = (2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j})^{-1}$ to get $\max_{j=1\ldots m}(2\mathring{\tau}_j^N\widetilde{\gamma}_{G,j})^{-1} \leq \bar{z}_0 + N/2$ with $\bar{z}_0 > 0$. Then from (4.14), we have

$$\begin{aligned}\phi_j^{N+1} &\geq (1+\min_{j=1\ldots m}(2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j}))\phi_j^N \geq \left(1+\frac{1}{\bar{z}_0+N/2}\right)\phi_j^N = \frac{2\bar{z}_0+N+2}{2\bar{z}_0+N}\phi_j^N \\ &= \frac{2\bar{z}_0+N+2}{2\bar{z}_0+N}\frac{2\bar{z}_0+N+1}{2\bar{z}_0+N-1}\phi_j^{N-1} = \ldots = \frac{(2\bar{z}_0+N+2)(2\bar{z}_0+N+1)}{2\bar{z}_0(2\bar{z}_0+1)}\phi_j^0.\end{aligned}$$

Therefore, ϕ_j^N grows as $\Omega(N^2)$, and we obtain the claimed convergence rates from (3.32). □

In Algorithm 4.1, we chose ω_j^i to eliminate the $\nabla K(x^i)$ term from the dual step. Selecting $\omega_j^i = -1$ keeps this term, but eliminates the necessity to have a finite ρ_ℓ as long as $p = 2$ as (3.29) and (3.28b) will no longer depend on it. This yields Algorithm 4.2 and the following:

Corollary 4.6. *Theorems 4.4 and 4.5 apply to Algorithm 4.2 if Assumption 3.2 holds with $p = 2$, and instead of (4.7c), (4.12c), (4.8b), and (4.13b), we assume*

$$\bar{L} := L_3 + nL^2\rho_x^2/(2\alpha_y) \quad \textit{and} \quad \mathbb{P}[\|x^{i+1} - \widehat{x}\| \leq \rho_x \mid O_{i-1}] = 1.$$

Proof. The proof remains exactly the same those of Theorems 4.4 and 4.5. Inserting $\omega_j^i = -1$, (4.10) and (4.15) as well as (4.8a) and (4.13a) lose their dependency on ρ_ℓ. Hence ρ_ℓ can be taken infinitely large. □

4.2 LINEAR CONVERGENCE

If all the primal and dual blocks exhibit second-order growth, i.e., $\bar{\gamma}_{F^*,\ell} > 0$ and $\gamma_{G,j} + \gamma_{K,j} > 0$, we obtain linear convergence:

Algorithm 4.2 Full dual updates #2

Assume the problem structure (P), equivalently (S). . For each iteration $i \in \mathbb{N}$, choose a sampling pattern for generating the random set of updated dual blocks $V(i+1) \in \mathcal{R}(O_i; \mathcal{P}\{1,\ldots,n\})$ with corresponding blockwise probabilities $\mathring{\pi}_j^i := \mathbb{P}[j \in S(i) \mid O_{i-1}] > 0$. Choose a rule for the iteration and block-dependent step length parameters $\mathring{\tau}_j^i, \breve{\sigma}_\ell^{i+1}, \bar{\omega}^i > 0$ based on one of Theorem 4.5, 4.4, or 4.7. Pick an initial iterate (x^0, y^0) and on each iteration $i \in \mathbb{N}$ update all blocks $x_j^{i+1} = P_j x^{i+1}$, $(j = 1, \ldots, m)$, and $y_\ell^{i+1} = Q_\ell y^{i+1}$, $(\ell = 1, \ldots, n)$, of x^{i+1} and y^{i+1} as:

$$
x_j^{i+1} := \begin{cases} (I + \mathring{\tau}_j^i P_j \partial G_j P_j)^{-1}(x_j^i - \mathring{\tau}_j^i P_j \nabla K(x^i)^* y^i), & j \in S(i), \\ x_j^i, & j \notin S(i), \end{cases}
$$

$$
y_\ell^{i+1} := (I + \breve{\sigma}_\ell^{i+1} Q_\ell \partial F_\ell^* Q_\ell)^{-1}\left(y_\ell^i + \breve{\sigma}_\ell^{i+1} Q_\ell K(x^i) + \breve{\sigma}_\ell^{i+1} \sum_{j \in S(i)} \left(\frac{\bar{\omega}^i}{\mathring{\pi}_j^i} + 1 \right) Q_\ell \nabla K(x^i)(x_j^{i+1} - x_j^i) \right).
$$

Theorem 4.7. *Suppose Assumptions 3.1, 3.2 and 3.4 hold with $L, L_3 \geq 0$; $p \in [1,2]$; $\gamma_{G,j} + \gamma_{K,j} > 0$, $(j = 1, \ldots, m)$, and $\bar{\gamma}_{F^*,\ell} > 0$, $(\ell = 1, \ldots, n)$, for some $\alpha_y, \zeta_\ell \geq 0$ as defined in (4.5). Let the iterates $\{u^i = (x^i, y^i)\}_{i \in \mathbb{N}}$ be generated by Algorithm 4.1 with iteration-independent probabilities $\mathring{\pi}_j^i \equiv \mathring{\pi}_j$ and step length parameters*

$$
\mathring{\tau}_j^{i+1} := \frac{\mathring{\tau}_j^i}{(1 + 2\mathring{\tau}_j^i \widetilde{\gamma}_{G,j})\bar{\omega}}, \quad \breve{\sigma}_\ell^{i+1} := \frac{\breve{\sigma}_\ell^i}{(1 + 2\breve{\sigma}_\ell^i \bar{\gamma}_{F^*,\ell})\bar{\omega}}, \quad \text{and} \tag{4.16a}
$$

$$
\bar{\omega}^i \equiv \bar{\omega} := \max\left\{ \max_{j=1\ldots m} \frac{1}{1 + 2\mathring{\tau}_j^0 \widetilde{\gamma}_{G,j}}, \max_{\ell=1\ldots n} \frac{1}{1 + 2\breve{\sigma}_\ell^0 \bar{\gamma}_{F^*,\ell}} \right\} \tag{4.16b}
$$

with $0 < \widetilde{\gamma}_{G,j} < \mathring{\pi}_j(\gamma_{G,j} + \gamma_{K,j})$; and initial $\mathring{\tau}_j^0, \breve{\sigma}_\ell^0 > 0$ satisfying for some $0 < \delta < \kappa < 1$, $\rho_x, \rho_\ell \geq 0$, $(\ell = 1, \ldots, n)$, and $w_{j,\ell}^i$ as in (4.4) the bounds

$$
1 - \kappa \geq \left\| \sum_{j \in \mathring{S}(i)} \sqrt{\frac{w_{j,\ell}^i \breve{\sigma}_\ell^0 \mathring{\tau}_j^0}{\mathring{\pi}_j}} Q_\ell \nabla K(x^i) P_j \right\|^2 \quad (i \in \mathbb{N}) \quad \text{and} \tag{4.17a}
$$

$$
\delta \geq \mathring{\tau}_j^0 \left(\bar{L} + 2(1 - \mathring{\pi}_j)(\gamma_{G,j} + \gamma_{K,j}) \frac{\gamma_{G,j} + \gamma_{K,j} - \widetilde{\gamma}_{G,j}}{\mathring{\pi}_j(\gamma_{G,j} + \gamma_{K,j}) - \widetilde{\gamma}_{G,j}} \right) \quad (j \in S(i)), \quad \text{with} \tag{4.17b}
$$

$$
\bar{L} := L_3 + \frac{L}{\bar{\omega}} \left(\max_{j=1\ldots m} \left(\frac{\bar{\omega}}{\mathring{\pi}_j} + 1 \right)^2 \textstyle\sum_{\ell=1}^n \rho_\ell + \frac{nL}{2\alpha_y} \rho_x^2 \right). \tag{4.17c}
$$

Further assume for $A := \sum_{j \in S(i)} (\mathring{\pi}_j)^{-1} P_j$ that

$$
\mathbb{E}_{i-1}[\theta_A] \geq p^{-p} \textstyle\sum_{\ell=1}^n \zeta_\ell^{1-p} \rho_\ell^{2-p} / \bar{\omega} \quad \text{and} \tag{4.18a}
$$

$$
1 = \mathbb{P}[\|x^{i+1} - \widehat{x}\| \leq \rho_x, \|Q_\ell(y^{i+1} - \widehat{y})\|_{P_{\mathrm{NL}}} \leq \rho_\ell, (\ell = 1, \ldots, n) \mid O_{i-1}]. \tag{4.18b}
$$

Then $\mathbb{E}[\|P_j(x^N - \widehat{x})\|^2]$ and $\mathbb{E}[\|Q_\ell(y^N - \widehat{y})\|^2]$ converge to zero at the linear rate $O((1/\bar{\omega})^N)$ for all $j \in \{1, \ldots, m\}$ and $\ell \in \{1, \ldots, n\}$.

Proof. We use Theorem 3.11 whose conditions we need to verify. We have already verified the nesting condition (2.9) for the choices $\mathring{V}(i+1) = \emptyset$, $V(i+1) = \{1, \ldots, n\}$, and $\mathring{S}(i) = S(i)$ in Algorithm 4.1. The coupling condition (3.27) we have reduced to (4.1). To verify (4.1), we initialise $\phi_j^0 := \eta^0(\mathring{\pi}_j^0\mathring{\tau}_j^0)^{-1}$ and $\psi_\ell^0 := \eta^0/\breve{\sigma}_\ell^0$ for some $\eta^0 > 0$, and update

$$\phi_j^{i+1} := (1+2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j})\phi_j^i, \quad \psi_\ell^{i+1} := (1+2\breve{\sigma}_\ell^i\overline{\gamma}_{F^*,\ell})\psi_\ell^i, \quad \text{and} \quad \eta^{i+1} := \eta^i/\bar{\omega}. \tag{4.19}$$

Then from (4.16), $\psi_\ell^{i+1}\breve{\sigma}_\ell^{i+1} = \psi_\ell^i\breve{\sigma}_\ell^i/\bar{\omega}$ and $\phi_j^{i+1}\mathring{\tau}_j^{i+1} = \phi_j^i\mathring{\tau}_j^i/\bar{\omega}$. Therefore, (4.1) holds by induction. Clearly also (3.17) holds as the step length and testing parameters are updated deterministically. The conditions (3.28) follow from (4.18) given that $\theta_{\Phi_i T_i} = \eta^i\theta_A = \bar{\omega}\eta^{i+1}\theta_A$.

We now prove (3.8). We start by proving by induction that

$$\bar{\omega} = \max\left\{\max_{j=1\ldots m}\frac{1}{1+2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j}}, \max_{\ell=1\ldots n}\frac{1}{1+2\breve{\sigma}_\ell^i\overline{\gamma}_{F^*,\ell}}\right\}, \tag{4.20}$$

in other words

$$\bar{\omega}^{-1} = 1 + \min\left\{\min_{j=1\ldots m} 2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j}, \min_{\ell=1\ldots n} 2\breve{\sigma}_\ell^i\overline{\gamma}_{F^*,\ell}\right\}.$$

The inductive base for $i = 0$ is clear from (4.16b). Using (4.16a), we obtain

$$\begin{aligned}\min\left\{\min_{j=1\ldots m} 2\mathring{\tau}_j^{i+1}\widetilde{\gamma}_{G,j}, \min_{\ell=1\ldots n} 2\breve{\sigma}_\ell^{i+1}\overline{\gamma}_{F^*,\ell}\right\} &= \frac{1}{\bar{\omega}}\min\left\{\min_{j=1\ldots m}\frac{1}{1+(2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j})^{-1}}, \min_{\ell=1\ldots n}\frac{1}{1+(2\breve{\sigma}_\ell^i\overline{\gamma}_{F^*,\ell})^{-1}}\right\}\\ &= \frac{1}{\bar{\omega}}\frac{1}{1+\min^{-1}\left\{\min_{j=1\ldots m} 2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j}, \min_{\ell=1\ldots n} 2\breve{\sigma}_\ell^i\overline{\gamma}_{F^*,\ell}\right\}} = \min\left\{\min_{j=1\ldots m} 2\mathring{\tau}_j^i\widetilde{\gamma}_{G,j}, \min_{\ell=1\ldots n} 2\breve{\sigma}_\ell^i\overline{\gamma}_{F^*,\ell}\right\},\end{aligned}$$

This establishes the inductive step, hence (4.20). By (4.20) and (4.16a), $\mathring{\tau}_j^{i+1}$ and $\breve{\sigma}_\ell^{i+1}$ are non-increasing in i. Also using (4.17a), this verifies (4.3). Thus Lemma 4.1 verifies (3.8).

We need to verify Theorem 3.11 (i) and (ii). Option (a) of the latter is trivially satisfied for every $\ell = 1, \ldots, n$ based on (4.19). Regarding Theorem 3.11 (i), we first of all observe that (3.26) reduces to $c_*^i = nL^2\eta^{i+1}\rho_x^2/(2\alpha_y)$. Consequently (3.29) becomes

$$L_j^i := L_3 + L\mathring{\pi}_j\big(\max_{j\in S(i)}(\omega_j^i+1)^2\textstyle\sum_{\ell=1}^n\rho_\ell + \frac{nL}{2\alpha_y}\rho_x^2\big)\eta^{i+1}/\eta^i \le \bar{L} \tag{4.21}$$

for $\omega_j^i := \bar{\omega}^i/\mathring{\pi}_j$ as in Algorithm 4.1. And with $\widetilde{\gamma}_{G,j} < \mathring{\pi}_j(\gamma_{G,j}+\gamma_{K,j})$, (4.17b), (4.21), and $\mathring{\tau}_j^{i+1} \le \mathring{\tau}_j^0$ show (3.43a). Therefore, Theorem 3.11(i) option (b) holds for every $j = 1, \ldots, m$.

We can now apply Theorem 3.11 to obtain (3.32). By (4.19) and (4.20) we have

$$\begin{aligned}&\phi_j^{N+1} = (1+2\mathring{\tau}_j^N\widetilde{\gamma}_{G,j})\phi_j^N \ge \phi_j^N/\bar{\omega} \ge \ldots \ge \phi_j^0/\bar{\omega}^{N+1} \quad \text{and}\\ &\psi_\ell^{N+1} = (1+2\breve{\sigma}_\ell^N\overline{\gamma}_{F^*,\ell})\psi_\ell^N \ge \psi_\ell^N/\bar{\omega} \ge \ldots \ge \psi_\ell^0/\bar{\omega}^{N+1}.\end{aligned}$$

Applying these estimates in (3.32) establishes the claimed linear convergence rates. □

Similarly to Algorithm 4.2, we could in the derivation of Algorithm 4.1 set $\omega_j^i = -1$ to remove any dependencies on ρ_ℓ from (4.17c) and (4.18a). This yields Algorithm 4.2 and:

Corollary 4.8. *Theorem 4.7 applies to Algorithm 4.2 if Assumption 3.2 holds with $p = 2$, and* (4.17c) *and* (4.18b) *are replaced with*

$$\bar{L} \ge L_3 + nL^2\rho_x^2/(2\alpha_y\bar{\omega}) \quad \text{and} \quad \mathbb{P}[\|x^{i+1}-\widehat{x}\| \le \rho_x \mid O_{i-1}] = 1.$$

Proof. The proof remains exactly the same as Theorem 4.7 given all $\omega_j^i = -1$ in (4.21) and (4.18a) no longer depend on ρ_ℓ, hence ρ_ℓ can be taken infinitely large. □

Remark 4.9 (Stochastic block-coordinate forward–backward splitting). Let $F(z) := z$ for $z \in \mathbb{R}$ and $K \in C^1(X)$. Then $F^*(y) = \delta_{\{1\}}(y)$. Taking $n = 1$ and $Q_1 = I$ results in $(I + \breve{\sigma}_1^{i+1} Q_1 \partial F^* Q_1)^{-1} \equiv 1$. Consequently $y^i \equiv 1$ on all iterations, so that the updates of Algorithms 4.1 and 4.2 reduce to

$$
x_j^{i+1} := \begin{cases} (I + \mathring{\tau}_j^i P_j \partial G_j P_j)^{-1}(x_j^i - \mathring{\tau}_j^i P_j \nabla K(x)), & j \in S(i), \\ x_j^i, & j \notin S(i), \end{cases} \tag{4.22}
$$

In the step length conditions of Theorems 4.4, 4.5 and 4.7, we can moreover take $\rho_1 = 0$ and let $\gamma_{F^*,1} \nearrow \infty$, consequently $a_y \nearrow \infty$. In particular, in all the theorems, $\bar{L} = L_3$, so that when $\mathring{\pi}_j = 1$, the upper bounds on the primal step lengths reduce to $\delta \geq \mathring{\tau}_j^0 L_3$ for some $\delta \in (0, 1)$ similarly to the standard condition in forward–backward splitting type methods. Moreover, by (A.1), $\gamma_{K,1}$ is simply a (reduced) factor of strong monotonicity of K at $\widehat{x}$ as defined in Assumption 3.4. Finally, since we can take $\breve{\sigma}_1^0 > 0$ arbitrarily small without affecting the updates (4.22), the conditions in the theorems corresponding to (3.7) become irrelevant.

5 METHODS WITH FULL PRIMAL UPDATES

We continue with developing more specific methods and their convergence results based on the updates of (2.13) and the conditions of Theorem 3.11. We now take $\mathring{S}(i) = \emptyset$, $S(i) = \{1, \ldots, m\}$, and $\mathring{V}(i+1) = V(i+1)$ for all iterations i. Then the nesting condition (2.9) of Theorem 3.11 holds and the coupling condition (3.27) becomes

$$
\phi_j^i \breve{\tau}_j^i = \eta^{i+1} = \mathring{\nu}_\ell^{i+2} \psi_\ell^{i+2} \mathring{\sigma}_\ell^{i+2}. \tag{5.1}
$$

Taking $\Omega_i = -I$, the updates of (2.13) simplify to those of Algorithm 5.1 since for the last two terms in the primal update

$$
\breve{\tau}_j^i y_\ell^{i+1} + \frac{\psi_\ell^{i+1} \sigma_\ell^{i+1}}{\phi_j^i}(y_\ell^{i+1} - y_\ell^i) = \breve{\tau}_j^i \left(y_\ell^{i+1} + \frac{\bar{\omega}^i}{\mathring{\nu}_\ell^{i+1}}(y_\ell^{i+1} - y_\ell^i) \right) \quad \text{for} \quad \bar{\omega}^i := \frac{\eta^i}{\eta^{i+1}}.
$$

Moreover, (2.12) reduces to $\lambda_{j,\ell}^i = -\sigma_\ell^{i+1} \psi_\ell^{i+1}$. We thus verify (3.8) via:

Lemma 5.1. *Suppose $\mathring{S}(i) = \emptyset$, $S(i) = \{1, \ldots, m\}$, and $\mathring{V}(i+1) = V(i+1)$ for $i \in \mathbb{N}$; the coupling condition (5.1) holds; $\bar{\omega}^i \leq 1$; as well as, for all $\ell = 1, \ldots, n$; $j = 1, \ldots, m$,*

$$
\mathring{\sigma}_\ell^{i+1} \breve{\tau}_j^i \leq \mathring{\sigma}_\ell^1 \breve{\tau}_j^0, \quad \text{and} \quad 1 - \kappa \geq \left\| \sum_{j=1}^m \sqrt{\frac{w_{j,\ell}^i \mathring{\sigma}_\ell^1 \breve{\tau}_j^0}{\mathring{\nu}_\ell^{i+1}}} Q_\ell \nabla K(x^i) P_j \right\|^2 \tag{5.2}
$$

for some $0 \leq \kappa \leq 1$ and $w_{j,\ell,k} = 1/w_{j,k,\ell} > 0$ such that

$$
w_{j,\ell}^i := \chi_{\mathcal{V}_j^i}(\ell) \sum_{k \in \mathcal{V}_j^i(\ell)} w_{j,\ell,k} \tag{5.3a}
$$

with

$$
\mathcal{V}_j^i(\ell) = \{k \in \{1, \ldots, n\} \mid Q_\ell \nabla K(x^i) P_j \nabla K(x^i)^* Q_k \neq 0,\ \ell \in \mathring{V}(i+1)\}. \tag{5.3b}
$$

Then the lower bound (3.8) holds.

Proof. By the first part of (5.2), (5.1), and $\lambda_{j,\ell}^i = -\sigma_\ell^{i+1} \psi_\ell^{i+1} = -\mathring{\sigma}_\ell^{i+1} \psi_\ell^{i+1}$, we have

$$
\mathring{\sigma}_\ell^1 \breve{\tau}_j^0 \geq \mathring{\sigma}_\ell^{i+1} \breve{\tau}_j^i = \frac{(\mathring{\sigma}_\ell^{i+1} \psi_\ell^{i+1})^2 \breve{\tau}_j^i}{\mathring{\sigma}_\ell^{i+1} (\psi_\ell^{i+1})^2} = \frac{(\lambda_{j,\ell}^i)^2 \mathring{\nu}_\ell^{i+1}}{\psi_\ell^{i+1} \phi_j^i} \quad (j = 1, \ldots, m).
$$

Algorithm 5.1 Full primal updates

Assume the problem structure (P), equivalently (S). For each iteration $i \in \mathbb{N}$, choose a sampling pattern for generating the random set of updated dual blocks $V(i+1) \in \mathcal{R}(O_i; \mathcal{P}\{1,\ldots,n\})$ with corresponding blockwise probabilities $\mathring{v}_\ell^{i+1} := \mathbb{P}[\ell \in V(i+1) \mid O_{i-1}] > 0$. Also choose a rule for the iteration and block-dependent step length parameters $\mathring{\sigma}_\ell^{i+1}, \breve{\tau}_j^i, \bar{\omega}^i > 0$ from one of Theorem 5.3, 5.4 or 5.5. Pick an initial iterate (x^0, y^0) and on each iteration $i \in \mathbb{N}$ update all blocks $x_j^{i+1} = P_j x^{i+1}$, $(j = 1,\ldots,m)$, and $y_\ell^{i+1} = Q_\ell y^{i+1}$, $(\ell = 1,\ldots,n)$, of x^{i+1} and y^{i+1} as:

$$
\begin{aligned}
y_\ell^{i+1} &:= \begin{cases} (I + \mathring{\sigma}_\ell^{i+1} Q_\ell \partial F_\ell^* Q_\ell)^{-1}(y_\ell^i + \mathring{\sigma}_\ell^{i+1} Q_\ell K(x^i)), & \ell \in V(i+1), \\ y_\ell^i, & \ell \notin V(i+1), \end{cases} \\
x_j^{i+1} &:= (I + \breve{\tau}_j^i P_j \partial G_j P_j)^{-1} P_j \left(x_j^i - \breve{\tau}_j^i \nabla K(x^i)^* \sum_{\ell \in V(i+1)} \left(y_\ell^{i+1} + \frac{\bar{\omega}^i}{\mathring{v}_\ell^{i+1}} (y_\ell^{i+1} - y_\ell^i) \right) \right).
\end{aligned}
$$

By the orthogonality of the projections P_j, we may insert this estimation into the second part of (5.2), obtaining (3.7); compare the proof of Lemma 3.5. The definition of $\bar{\mathcal{V}}_j^i(\ell)$ in (3.5) also reduces to that in (5.3b), while the definition of $w_{j,\ell}^i$ in (5.3a) is exactly that in (3.6). We finish by applying Lemma 3.5 to verify (3.8). ◻

Remark 5.2. The first part of (5.2) is a relaxation of the property $\tau^i \sigma^{i+1} = \tau^0 \sigma^1$ that would be satisfied by a dual-first variant of the basic PDPS; compare Remark 4.2.

Finally, we also remind that (3.30) and (3.31) for this section simplify to

$$
\tag{5.4} \bar{\gamma}_{GK,j}^i := \gamma_{G,j} + \gamma_{K,j} - \alpha_x, \quad \text{and} \quad \bar{\gamma}_{F^*,\ell}^{i+1} \equiv \bar{\gamma}_{F^*,\ell} := \begin{cases} \gamma_{F^*,\ell}, & Q_\ell P_{\mathrm{NL}} = 0, \\ \gamma_{F^*,\ell} - (p-1)\zeta_\ell, & Q_\ell P_{\mathrm{NL}} \neq 0. \end{cases}
$$

5.1 ACCELERATED RATES

As in Section 4, we start with simple step length rules that yield $O(1/N)$ convergence rates for those blocks that exhibit second-order growth.

Theorem 5.3. *Suppose Assumptions 3.1, 3.2 and 3.4 hold with $L, L_3 \geq 0$; $p \in [1,2]$; $\gamma_{G,j} + \gamma_{K,j} > 0$ $(j = 1,\ldots,m)$; and $\gamma_{F^*,\ell} \geq (p-1)\zeta_\ell$ for some $\zeta_\ell \geq 0$ when $Q_\ell P_{\mathrm{NL}} \neq 0$, $(\ell = 1,\ldots,n)$. Let the iterates $\{u^i = (x^i, y^i)\}_{i \in \mathbb{N}}$ be generated by Algorithm 5.1 with iteration-independent probabilities $\mathring{v}_\ell^i \equiv \mathring{v}_\ell$ and step lengths*

$$
\tag{5.5} \mathring{\sigma}_\ell^{i+1} := \frac{\mathring{\sigma}_\ell^i}{1 + 2\mathring{\sigma}_\ell^i \widetilde{\gamma}_{F^*,\ell}}, \quad \bar{\omega}^i \equiv 1, \quad \text{and} \quad \breve{\tau}_j^{i+1} := \frac{\breve{\tau}_j^i}{1 + 2\breve{\tau}_j^i \widetilde{\gamma}_{G,j}}
$$

with $0 \leq \widetilde{\gamma}_{G,j} < \gamma_{G,j} + \gamma_{K,j}$, $(j = 1,\ldots,m)$, and either $0 \leq \widetilde{\gamma}_{F^*,\ell} < \mathring{v}_\ell \bar{\gamma}_{F^*,\ell}$ or $\widetilde{\gamma}_{F^*,\ell} = \bar{\gamma}_{F^*,\ell} = 0$ for each $\ell = 1,\ldots,n$, $\bar{\gamma}_{F^*,\ell}$ defined in (5.4); and initial $\breve{\tau}_j^0, \mathring{\sigma}_\ell^1 > 0$ satisfying for some $\rho_\ell \geq 0$, $(\ell = 1,\ldots,n)$,

$0 < \delta < \kappa < 1$, and $w^i_{j,\ell}$ as in (5.3) the bounds

$$1 - \kappa \geq \left\| \sum_{j=1}^{m} \sqrt{\frac{w^i_{j,\ell}\mathring{\sigma}^1_\ell \check{\tau}^0_j}{\mathring{\nu}_\ell}} Q_\ell \nabla K(x^i) P_j \right\|^2 \tag{5.6a}$$

$$\delta \geq \check{\tau}^0_j \left(L_3 + \frac{mL^2}{2\min_{j=1\ldots m}(\gamma_{G,j} + \gamma_{K,j} - \widetilde{\gamma}_{G,j})} \sum_{\ell=1}^{n} \rho_\ell^2 \right), \quad \textit{and} \tag{5.6b}$$

$$\frac{\kappa - \delta}{1 - \delta} \geq 2\chi_{V(i+1)}(\ell)(1 - \mathring{\nu}_\ell)\overline{\gamma}_{F^*,\ell}\mathring{\sigma}^1_\ell \frac{\overline{\gamma}_{F^*,\ell} - \widetilde{\gamma}_{F^*,\ell}}{\mathring{\nu}_\ell \overline{\gamma}_{F^*,\ell} - \widetilde{\gamma}_{F^*,\ell}} \quad (i \in \mathbb{N};\ j = 1, \ldots, m). \tag{5.6c}$$

Assume that

$$\theta_I \geq p^{-p} \textstyle\sum_{\ell=1}^n (\mathring{\nu}_\ell)^2 \zeta_\ell^{1-p} \rho_\ell^{2-p} \quad \textit{and} \tag{5.7a}$$

$$1 = \mathbb{P}[\|Q_\ell(y^{i+1} - \widehat{y})\|_{P_{\mathrm{NL}}} \leq \rho_\ell, (\ell = 1, \ldots, n) \mid O_{i-1}]. \tag{5.7b}$$

Then $\mathbb{E}[\|P_j(x^N - \widehat{x})\|^2] \to 0$ at the rate $O(1/N)$ for all j such that $\widetilde{\gamma}_{G,j} > 0$ and $\mathbb{E}[\|Q_\ell(y^N - \widehat{y})\|^2] \to 0$ at the rate $O(1/N)$ for all ℓ such that $\widetilde{\gamma}_{F^*,\ell} > 0$.

Proof. We will use Theorem 3.11, whose conditions we need to verify. With the choice of $\mathring{S}(i) = \emptyset$, $S(i) = \{1, \ldots, m\}$, and $\mathring{V}(i+1) = V(i+1)$ in Algorithm 5.1, we have already verified the nesting conditions (2.9) and reduced the coupling conditions (3.27) to (5.1). To verify (5.1), we set $\phi^0_j = \eta^1/\check{\tau}^0_j$, $\psi^2_\ell = \eta^1/(\mathring{\sigma}^2_\ell \mathring{\nu}_\ell)$ for some $\eta^1 > 0$, and update

$$\phi^{i+1}_j = (1 + 2\check{\tau}^i_j \widetilde{\gamma}_{G,j})\phi^i_j, \quad \psi^{i+2}_\ell = (1 + 2\mathring{\sigma}^{i+1}_\ell \widetilde{\gamma}_{F^*,\ell})\psi^{i+1}_\ell, \quad \text{and} \quad \eta^{i+1} := \eta^i. \tag{5.8}$$

Then $\mathring{\nu}_\ell \mathring{\sigma}^{i+2}_\ell \psi^{i+2}_\ell = \eta^{i+1} = \phi^i_j \check{\tau}^i_j$ due to (5.5) for all ℓ and j, and (5.1) follows. Clearly also (3.17) holds because the step length and testing parameters are updated deterministically. The conditions (3.28) follow from (5.7) given that in Assumption 3.2 we can take $\theta_{\Phi_i T_i} = \eta^{i+1}\theta_I = \eta^i \theta_I = \psi^{i+1}_\ell \check{\sigma}^{i+1}_\ell \theta_I / \mathring{\nu}_\ell$, and ρ_x can be taken infinitely large.

The step length parameters $\mathring{\sigma}^{i+1}$ and $\check{\tau}^i_j$ are non-increasing in i by the defining (5.5). Also using (5.6a), we thus verify (5.2). Hence Lemma 5.1 establishes (3.8).

We still need to verify Theorem 3.11 (i) and (ii). As far as the former is concerned, $\phi^{i+1}_j \leq (1+2\check{\tau}^i_j \widetilde{\gamma}_{G,j})\phi^i_j$ from (5.8). Moreover, after applying (5.1), (3.26) and (3.29) reduce to

$$c^i_* = \frac{mL^2\eta^{i+1}}{2\alpha_x} \sum_{\ell=1}^{n} \rho_\ell^2 \quad \text{and} \quad L^i_j = L_3 + \frac{mL^2}{2\alpha_x} \sum_{\ell=1}^{n} \rho_\ell^2,$$

which Thus, setting $\alpha_x = \min_{j=1\ldots m}(\gamma_{G,j} + \gamma_{K,j} - \widetilde{\gamma}_{G,j}) > 0$, Theorem 3.11 (i) option (a) follows for every j from (5.6b) and $\check{\tau}^{i+1}_j$ being non-increasing. Regarding the dual test, we have $\psi^{i+2}_\ell \leq (1+2\mathring{\sigma}^{i+1}_\ell \widetilde{\gamma}^{i+1}_{F^*,\ell})\psi^{i+1}_\ell$ which together with (5.6c) leads to (3.43b). Therefore, Theorem 3.11 (ii) option (b) holds for every ℓ.

We can now apply Theorem 3.11 to obtain (3.32). From (5.8) we have

$$\begin{aligned}
&\phi^{i+1}_j = \phi^i_j + 2\widetilde{\gamma}_{G,j}\eta^{i+1} = \phi^i_j + 2\widetilde{\gamma}_{G,j}\eta^1 = \ldots = \phi^0_j + 2i\widetilde{\gamma}_{G,j}\eta^1 \quad \text{and} \\
&\psi^{i+2}_\ell = \psi^{i+1}_\ell + 2\widetilde{\gamma}_{F^*,\ell}\eta^i/\mathring{\nu}_\ell = \psi^{i+1}_\ell + 2\widetilde{\gamma}_{F^*,\ell}\eta^1/\mathring{\nu}_\ell = \ldots = \psi^1_\ell + 2(i+1)\widetilde{\gamma}_{F^*,\ell}\eta^1/\mathring{\nu}_\ell.
\end{aligned}$$

Therefore, for any primal block j with $\widetilde{\gamma}_{G,j} > 0$ and dual block ℓ with $\widetilde{\gamma}_{F^*,\ell} > 0$, ϕ^N_j and ψ^{N+1}_ℓ grow as $\Omega(N)$, respectively. This together with (3.32) gives the claim. □

We get improved $O(1/N^2)$ rates if all primal blocks exhibit second-order growth:

Theorem 5.4. *Suppose Assumptions 3.1, 3.2 and 3.4 hold with $L, L_3 \geq 0$; $p \in [1,2]$; $\gamma_{G,j} + \gamma_{K,j} > 0$, $(j = 1, \dots, m)$; and $\gamma_{F^*,\ell} \geq (p-1)\zeta_\ell$ for some ζ_ℓ when $Q_\ell P_{\mathrm{NL}} \neq 0$, $(\ell = 1, \dots, n)$. Let the iterates $\{u^i = (x^i, y^i)\}_{i \in \mathbb{N}}$ be generated by Algorithm 5.1 with iteration-independent probabilities $\mathring{\nu}_\ell^i \equiv \mathring{\nu}_\ell$ and step length parameters*

$$\mathring{\sigma}_\ell^{i+2} = \frac{\mathring{\sigma}_\ell^{i+1}}{\bar{\omega}^i}, \quad \check{\tau}_j^{i+1} = \frac{1}{1 + 2\check{\tau}_j^i \widetilde{\gamma}_{G,j}} \frac{\check{\tau}_j^i}{\bar{\omega}^{i+1}}, \quad \text{and} \quad \bar{\omega}^{i+1} := \max_{j=1\dots m} \frac{1}{\sqrt{1 + 2\check{\tau}_j^i \widetilde{\gamma}_{G,j}}}; \tag{5.9}$$

with $0 < \widetilde{\gamma}_{G,j} < \gamma_{G,j} + \gamma_{K,j}$, $(j = 1, \dots, m)$, and the initial $\bar{\omega}^0 = 1$, $\check{\tau}_j^0$ and $\mathring{\sigma}_\ell^1$ satisfying for some $\rho_\ell \geq 0$, $(\ell = 1, \dots, n)$, $0 < \delta \leq \kappa < 1$, and $w_{j,\ell}^i$ as in (5.3) the bounds $(i \in \mathbb{N}; j = 1, \dots, m)$

$$1 - \kappa \geq \left\| \sum_{j=1}^m \sqrt{\frac{w_{j,\ell}^i \mathring{\sigma}_\ell^1 \check{\tau}_j^0}{\mathring{\nu}_\ell}} Q_\ell \nabla K(x^i) P_j \right\|^2 \qquad \delta \geq \check{\tau}_j^0 \left(L_3 + \frac{mL^2}{2 \min_{j=1\dots m}(\gamma_{G,j} + \gamma_{K,j} - \widetilde{\gamma}_{G,j})} \sum_{\ell=1}^n \rho_\ell^2 \right). \tag{5.10}$$

Also assume

$$\theta_I \geq p^{-p} \textstyle\sum_{\ell=1}^n (\mathring{\nu}_\ell)^2 \zeta_\ell^{1-p} \rho_\ell^{2-p} \quad \text{and} \tag{5.11a}$$

$$1 = \mathbb{P}[\|Q_\ell(y^{i+1} - \widehat{y})\|_{P_{\mathrm{NL}}} \leq \rho_\ell, (\ell = 1, \dots, n) \mid O_{i-1}]. \tag{5.11b}$$

Then $\mathbb{E}[\|P_j(x^N - \widehat{x})\|^2] \to 0$ at the rate $O(1/N^2)$ for all j.

Proof. We will use Theorem 3.11 whose conditions we need to verify. With the choice of $\mathring{S}(i) = \emptyset$, $S(i) = \{1, \dots, m\}$, and $\mathring{V}(i+1) = V(i+1)$ in Algorithm 5.1, we have already verified the nesting conditions (2.9) and reduced the coupling conditions (3.27) to (5.1). To verify (5.1), we set $\phi_j^0 = \eta^1 / \check{\tau}_j^0$ and $\psi_\ell^2 := \eta^1 / (\mathring{\nu}_\ell \mathring{\sigma}_\ell^2)$ for some $\eta^1 > 0$, and update

$$\phi_j^{i+1} := (1 + 2\check{\tau}_j^i \widetilde{\gamma}_{G,j}) \phi_j^i, \quad \psi_\ell^{i+1} := \psi_\ell^i, \quad \text{and} \quad \eta^{i+1} = \eta^i / \bar{\omega}^i. \tag{5.12}$$

Then from (5.9), we inductively get $\mathring{\nu}_\ell \psi_\ell^{i+2} \mathring{\sigma}_\ell^{i+2} = \mathring{\nu}_\ell \psi_\ell^{i+1} \mathring{\sigma}_\ell^{i+1} / \bar{\omega}^i = \eta^{i+1}$ for all ℓ. From (5.9), we also have inductively for all j, $\phi_j^{i+1} \check{\tau}_j^{i+1} = \phi_j^i \check{\tau}_j^i / \bar{\omega}^{i+1} = \eta^{i+2}$. Therefore (5.1) holds. Then, the conditions (3.28) follow from (5.11) given that $\bar{\omega}^i \leq 1$ and in Assumption 3.2 we can take $\theta_{\Phi_i T_i} = \eta^{i+1} \theta_I = \eta^i \theta_I / \bar{\omega}^i = \psi_\ell^{i+1} \sigma_\ell^{i+1} \theta_I / (\mathring{\nu}_\ell \bar{\omega}^i)$, and ρ_x can be taken infinitely large. Clearly also (3.17) holds because the step length and testing parameters are updated deterministically.

We now verify (3.8). From (5.9) we obtain

$$\begin{aligned}
\mathring{\sigma}_\ell^{i+2} \check{\tau}_j^{i+1} &= \frac{\mathring{\sigma}_\ell^{i+1} \check{\tau}_j^i}{\bar{\omega}^i \bar{\omega}^{i+1} (1 + 2\widetilde{\gamma}_{G,j} \check{\tau}_j^i)} \leq \mathring{\sigma}_\ell^{i+1} \check{\tau}_j^i \sqrt{\frac{1 + 2\widetilde{\gamma}_{G,j} \check{\tau}_j^{i-1}}{1 + 2\widetilde{\gamma}_{G,j} \check{\tau}_j^i}} \leq \ldots \leq \mathring{\sigma}_\ell^2 \check{\tau}_j^1 \sqrt{\frac{1 + 2\widetilde{\gamma}_{G,j} \check{\tau}_j^0}{1 + 2\widetilde{\gamma}_{G,j} \check{\tau}_j^i}} \\
&= \mathring{\sigma}_\ell^1 \check{\tau}_j^0 \frac{1}{\bar{\omega}^1 \sqrt{1 + 2\widetilde{\gamma}_{G,j} \check{\tau}_j^0}} \frac{1}{1 + 2\widetilde{\gamma}_{G,j} \check{\tau}_j^i} \leq \mathring{\sigma}_\ell^1 \check{\tau}_j^0.
\end{aligned}$$

This and (5.10) verify (5.2). Thus Lemma 5.1 establishes (3.8).

We still need to verify Theorem 3.11 (i) and (ii). Regarding the former, $\phi_j^{i+1} \leq (1 + 2\check{\tau}_j^i \widetilde{\gamma}_{G,j}) \phi_j^i$ from (5.12). Moreover, after applying (5.1), equalities (3.26) and (3.29) reduce to

$$c_*^i = \frac{mL^2 \eta^{i+1}}{2\alpha_x} \sum_{\ell=1}^n \rho_\ell^2 \quad \text{and} \quad L_j^i = L_3 + \frac{mL^2}{2\alpha_x} \sum_{\ell=1}^n \rho_\ell^2.$$

Thus, setting $\alpha_x = \min_{j=1\dots m}(\gamma_{G,j} + \gamma_{K,j} - \widetilde{\gamma}_{G,j}) > 0$, Theorem 3.11 (i) option (a) follows for every j from the second inequality in (5.10) and $\check{\tau}_j^{i+1}$ being decreasing. As for Theorem 3.11 (ii), $\psi_\ell^{i+1} = \psi_\ell^{i+2} \leq$

$(1 + 2\chi_{V(i+1)}(\ell)\sigma_\ell^{i+1}\overline{\gamma}_{F^*,\ell}^{i+1})\psi_\ell^{i+1}$ trivially as we have assumed $\overline{\gamma}_{F^*,\ell}^{i+1} \geq 0$. Thus Theorem 3.11 (ii) option (a) holds for every ℓ.

We can now use Theorem 3.11 to verify (3.32). Multiplying the τ update of (5.9) by $2\widetilde{\gamma}_{G,j}$, plugging in $\overline{\omega}^{i+1}$, and taking the inverse, we get

$$(2\check{\tau}_j^{i+1}\widetilde{\gamma}_{G,j})^{-1} = \frac{1 + 2\check{\tau}_j^i\widetilde{\gamma}_{G,j}}{2\check{\tau}_j^i\widetilde{\gamma}_{G,j}\sqrt{1 + \min_{j=1\ldots m}(2\check{\tau}_j^i\widetilde{\gamma}_{G,j})}} = \frac{1 + (2\check{\tau}_j^i\widetilde{\gamma}_{G,j})^{-1}}{\sqrt{1 + (\max_{j=1\ldots m}(2\check{\tau}_j^i\widetilde{\gamma}_{G,j})^{-1})^{-1}}}.$$

We then apply Lemma B.1 with $z_j^i = (2\check{\tau}_j^i\widetilde{\gamma}_{G,j})^{-1}$ to obtain $\max_{j=1\ldots m}(2\check{\tau}_j^N\widetilde{\gamma}_{G,j})^{-1} \leq \overline{z}_0 + N/2$ with $\overline{z}_0 > 0$. Then from (5.12), we have

$$\begin{aligned}\phi_j^{N+1} &\geq (1 + \min_{j=1\ldots m}(2\check{\tau}_j^i\widetilde{\gamma}_{G,j}))\phi_j^N \geq \left(1 + \frac{1}{\overline{z}_0 + N/2}\right)\phi_j^N = \frac{2\overline{z}_0 + N + 2}{2\overline{z}_0 + N}\phi_j^N \\ &= \frac{2\overline{z}_0 + N + 2}{2\overline{z}_0 + N}\frac{2\overline{z}_0 + N + 1}{2\overline{z}_0 + N - 1}\phi_j^{N-1} = \ldots = \frac{(2\overline{z}_0 + N + 2)(2\overline{z}_0 + N + 1)}{2\overline{z}_0(2\overline{z}_0 + 1)}\phi_j^0.\end{aligned}$$

Therefore, ϕ_j^N grows as $\Omega(N^2)$, and we obtain the claimed convergence rates from (3.32). □

5.2 LINEAR CONVERGENCE

If all the primal and dual blocks exhibit second-order growth, i.e., $\overline{\gamma}_{F^*,\ell} > 0$ and $\gamma_{G,j} + \gamma_{K,j} > 0$, we obtain linear convergence:

Theorem 5.5. *Suppose Assumptions 3.1, 3.2 and 3.4 hold with $L, L_3 \geq 0$; $p \in [1, 2]$; $\gamma_{G,j} + \gamma_{K,j} > 0$, $(j = 1, \ldots, m)$. Let the iterates $\{u^i = (x^i, y^i)\}_{i\in\mathbb{N}}$ be generated by Algorithm 5.1 with iteration-independent $\mathring{\nu}_\ell^i \equiv \mathring{\nu}_\ell$ and step lengths*

$$\check{\tau}_j^{i+1} := \frac{\check{\tau}_j^i}{(1 + 2\check{\tau}_j^i\widetilde{\gamma}_{G,j})\overline{\omega}}, \quad \mathring{\sigma}_\ell^{i+2} := \frac{\mathring{\sigma}_\ell^{i+1}}{(1 + 2\mathring{\sigma}_\ell^{i+1}\widetilde{\gamma}_{F^*,\ell})\overline{\omega}}, \quad \text{and} \tag{5.13a}$$

$$\overline{\omega}^i \equiv \overline{\omega} := \max\left\{\max_{j=1\ldots m}\frac{1}{1 + 2\check{\tau}_j^0\widetilde{\gamma}_{G,j}}, \max_{\ell=1\ldots n}\frac{1}{1 + 2\mathring{\sigma}_\ell^1\widetilde{\gamma}_{F^*,\ell}}\right\} \tag{5.13b}$$

with $0 < \widetilde{\gamma}_{G,j} < \gamma_{G,j} + \gamma_{K,j}$, $(j = 1, \ldots, m)$, and $0 < \widetilde{\gamma}_{F^*,\ell} < \mathring{\nu}_\ell\overline{\gamma}_{F^*,\ell}$, $(\ell = 1, \ldots, n)$, $\overline{\gamma}_{F^*,\ell}$ defined in (5.4); and initial $\check{\tau}_j^0, \mathring{\sigma}_\ell^1 > 0$ satisfying for some $0 < \delta < \kappa < 1$, $\rho_\ell \geq 0$ $(\ell = 1, \ldots, n)$, with $w_{j,\ell}^i$ as in (5.3) the bounds

$$1 - \kappa \geq \left\|\sum_{j=1}^m \sqrt{\frac{w_{j,\ell}^i\mathring{\sigma}_\ell^1\check{\tau}_j^0}{\mathring{\nu}_\ell}}Q_\ell\nabla K(x^i)P_j\right\|^2, \tag{5.14a}$$

$$\delta \geq \check{\tau}_j^0\left(L_3 + \frac{mL^2}{2\min_{j=1\ldots m}(\gamma_{G,j} + \gamma_{K,j} - \widetilde{\gamma}_{G,j})}\sum_{\ell=1}^n\rho_\ell^2\right), \quad \text{and} \tag{5.14b}$$

$$\frac{\kappa - \delta}{1 - \delta} \geq 2(1 - \mathring{\nu}_\ell)\overline{\gamma}_{F^*,\ell}\mathring{\sigma}_\ell^1\frac{\overline{\gamma}_{F^*,\ell} - \widetilde{\gamma}_{F^*,\ell}}{\mathring{\nu}_\ell\overline{\gamma}_{F^*,\ell} - \widetilde{\gamma}_{F^*,\ell}} \qquad (\ell \in V(i+1);\ j = 1, \ldots, m;\ i \in \mathbb{N}). \tag{5.14c}$$

Further assume that

$$\theta_I \geq p^{-p}\overline{\omega}\textstyle\sum_{\ell=1}^n(\mathring{\nu}_\ell)^2\zeta_\ell^{1-p}\rho_\ell^{2-p} \quad \text{and} \tag{5.15a}$$

$$1 = \mathbb{P}[\|Q_\ell(y^{i+1} - \widehat{y})\|_{P_{\mathrm{NL}}} \leq \rho_\ell, (\ell = 1, \ldots, n) \mid O_{i-1}]. \tag{5.15b}$$

Then $\mathbb{E}[\|P_j(x^N - \widehat{x})\|^2] \to 0$ and $\mathbb{E}[\|Q_\ell(y^N - \widehat{y})\|^2] \to 0$ at the linear rate $O((1/\overline{\omega})^N)$ for all $j \in \{1, \ldots, m\}$ and $\ell \in \{1, \ldots, n\}$.

Proof. We will use Theorem 3.11, whose conditions we need to verify. With the choice of $\mathring{S}(i) = \emptyset$, $S(i) = \{1, \ldots, m\}$, and $\mathring{V}(i+1) = V(i+1)$ in Algorithm 5.1, we have already verified the nesting conditions (2.9) and reduced the coupling conditions (3.27) to (5.1). To verify (5.1), we set $\phi_j^0 = \eta^1/\check{\tau}_j^0$ and $\psi_\ell^2 := \eta^1/(\mathring{\nu}_\ell \mathring{\sigma}_\ell^2)$ for some $\eta^1 > 0$, and update

$$\phi_j^{i+1} := (1 + 2\check{\tau}_j^i \widetilde{\gamma}_{G,j})\phi_j^i, \quad \psi_\ell^{i+1} := (1 + 2\mathring{\sigma}_\ell^i \widetilde{\gamma}_{F^*,\ell})\psi_\ell^i, \quad \text{and} \quad \eta^{i+1} = \eta^i/\bar{\omega}. \tag{5.16}$$

Then from (5.13), we inductively get $\mathring{\nu}_\ell \psi_\ell^{i+2} \mathring{\sigma}_\ell^{i+2} = \mathring{\nu}_\ell \psi_\ell^{i+1} \mathring{\sigma}_\ell^{i+1}/\bar{\omega} = \eta^{i+1}$ for all ℓ and $\phi_j^{i+1}\check{\tau}_j^{i+1} = \phi_j^i \check{\tau}_j^i/\bar{\omega} = \eta^{i+2}$ for all j, therefore, (5.1) holds. Then, the conditions (3.28) follow from (5.15) given that in Assumption 3.2 we can take $\theta_{\Phi_i T_i} = \eta^{i+1}\theta_I = \eta^i \theta_I/\bar{\omega} = \psi_\ell^{i+1}\sigma_\ell^{i+1}\theta_I/(\mathring{\nu}_\ell \bar{\omega})$, and ρ_x can be taken infinitely large. Clearly also (3.17) holds because the step length and testing parameters are updated deterministically.

We now verify (3.8). We start by proving by induction that

$$\bar{\omega} = \max\left\{ \max_{j=1\ldots m} \frac{1}{1 + 2\check{\tau}_j^i \widetilde{\gamma}_{G,j}}, \max_{\ell=1\ldots n} \frac{1}{1 + 2\mathring{\sigma}_\ell^{i+1}\widetilde{\gamma}_{F^*,\ell}} \right\}, \tag{5.17}$$

in other words

$$\bar{\omega}^{-1} = 1 + \min\left\{ \min_{j=1\ldots m} 2\check{\tau}_j^i \widetilde{\gamma}_{G,j}, \min_{\ell=1\ldots n} 2\mathring{\sigma}_\ell^{i+1}\widetilde{\gamma}_{F^*,\ell} \right\}.$$

The inductive base for $i = 0$ holds by (5.13b). Using (5.13a),

$$\begin{aligned}
\min\left\{ \min_{j=1\ldots m} 2\check{\tau}_j^{i+1}\widetilde{\gamma}_{G,j}, \min_{\ell=1\ldots n} 2\mathring{\sigma}_\ell^{i+2}\widetilde{\gamma}_{F^*,\ell} \right\} &= \frac{1}{\bar{\omega}} \min\left\{ \min_{j=1\ldots m} \frac{1}{1 + (2\check{\tau}_j^i \widetilde{\gamma}_{G,j})^{-1}}, \min_{\ell=1\ldots n} \frac{1}{1 + (2\mathring{\sigma}_\ell^{i+1}\widetilde{\gamma}_{F^*,\ell})^{-1}} \right\} \\
&= \frac{1}{\bar{\omega}} \frac{1}{1 + \min^{-1}\left\{ \min_{j=1\ldots m} 2\check{\tau}_j^i \widetilde{\gamma}_{G,j}, \min_{\ell=1\ldots n} 2\mathring{\sigma}_\ell^{i+1}\widetilde{\gamma}_{F^*,\ell} \right\}} = \min\left\{ \min_{j=1\ldots m} 2\check{\tau}_j^i \widetilde{\gamma}_{G,j}, \min_{\ell=1\ldots n} 2\mathring{\sigma}_\ell^{i+1}\widetilde{\gamma}_{F^*,\ell} \right\}.
\end{aligned}$$

This establishes the inductive step, hence (5.17), which in turn shows that $\check{\tau}_j^i$ and $\mathring{\sigma}_\ell^{i+1}$ as updated according to (5.13a) are non-increasing in i. Also using (5.14), this proves (5.2). Thus Lemma 5.1 verifies (3.8).

We need to verify Theorem 3.11 (i) and (ii). As for the former, (3.26) and (3.29) reduce to

$$c_*^i = \frac{mL^2}{2\alpha_x} \sum_{\ell=1}^n \rho_\ell^2 \eta^{i+1} \quad \text{and} \quad L_j^i = L_3 + \frac{mL^2}{2\alpha_x} \sum_{\ell=1}^n \rho_\ell^2,$$

so (5.14), together with non-increasing $\check{\tau}_j^i$ and the update rule for ϕ_j^{i+1} in (5.16), verify Theorem 3.11 (i) option (a) for every j and $\alpha_x = \min_{j=1\ldots m}(\gamma_{G,j} + \gamma_{K,j} - \widetilde{\gamma}_{G,j})$. Regarding the latter, since we take $\widetilde{\gamma}_{F^*,\ell} < \mathring{\nu}_\ell \bar{\gamma}_{F^*,\ell}$, we obtain (3.43b) using the last inequality of (5.14) and that $\mathring{\sigma}_\ell^{i+1}$ is non-increasing by definition in (5.13). Hence Theorem 3.11 (ii) option (b) holds for every ℓ.

Therefore, we can apply Theorem 3.11 to obtain (3.32). By (5.16) and (5.17),

$$\begin{aligned}
\phi_j^{N+1} &= (1 + 2\check{\tau}_j^N \widetilde{\gamma}_{G,j})\phi_j^N \geq \phi_j^N/\bar{\omega} \geq \ldots \geq \phi_j^0/\bar{\omega}^{N+1} \quad \text{and} \\
\psi_\ell^{N+1} &= (1 + 2\mathring{\sigma}_\ell^N \widetilde{\gamma}_{F^*,\ell})\psi_\ell^N \geq \psi_\ell^N/\bar{\omega} \geq \ldots \geq \psi_\ell^1/\bar{\omega}^N.
\end{aligned}$$

Applying these estimates in (3.32) establishes the claimed linear convergence rates. □

Remark 5.6 (Stochastic sum-sampling forward–backward splitting). Consider the problem (1.1) with $F^*(y) = \delta_{\{\mathbb{1}\}}$ for $\mathbb{1} := (1, \ldots, 1) \in \mathbb{R}^n$ and $\nabla K(x)^* y = \sum_{\ell=1}^n \nabla J_\ell(x) y_{(\ell)}$ with $y = (y_{(1)}, \ldots, y_{(n)})$. Taking $Q_\ell y := (0, \ldots, 0, y_{(\ell)}, 0, \ldots, 0)$, it follows that $(I + \check{\sigma}_\ell^{i+1} Q_\ell \partial F_\ell^* Q_\ell)^{-1} \equiv (0, \ldots, 0, 1, 0, \ldots, 0)$. Consequently

$y^i \equiv \mathbb{1}$ on all iterations, so that with just a single primal block with corresponding step length $\breve{\tau}^i = \breve{\tau}_1^i$, Algorithm 5.1 reduces to

$$x^{i+1} := (I + \breve{\tau}^i \partial G)^{-1}\left(x^i - \breve{\tau}^i \sum_{\ell \in V(i+1)} \nabla J_\ell(x^i)\right).$$

With random $V(i+1)$, this is a forward–backward splitting method that stochastically samples $\sum_\ell J_\ell$ in (1.1). We can take any $\gamma_{F^*,\ell} \in (0, \infty)$, which in Theorems 5.3 to 5.5 also allows us to take ζ_ℓ arbitrarily large and $\mathring{\sigma}_\ell^i > 0$ arbitrarily small. Consequently, the systems of step length bounds (5.6) and (5.14) reduce to their second part (with first and third part unnecessary), and (5.10) reduces to its second part. In other words, we only need to choose $\breve{\tau}^0$ sufficiently small.

6 NUMERICAL EXPERIENCE

We will now study the performance of our proposed methods on two application problems: diffusion tensor imaging (DTI), which is a form of magnetic resonance imaging (MRI), and electrical impedance tomography (EIT).

6.1 DIFFUSION TENSOR IMAGING

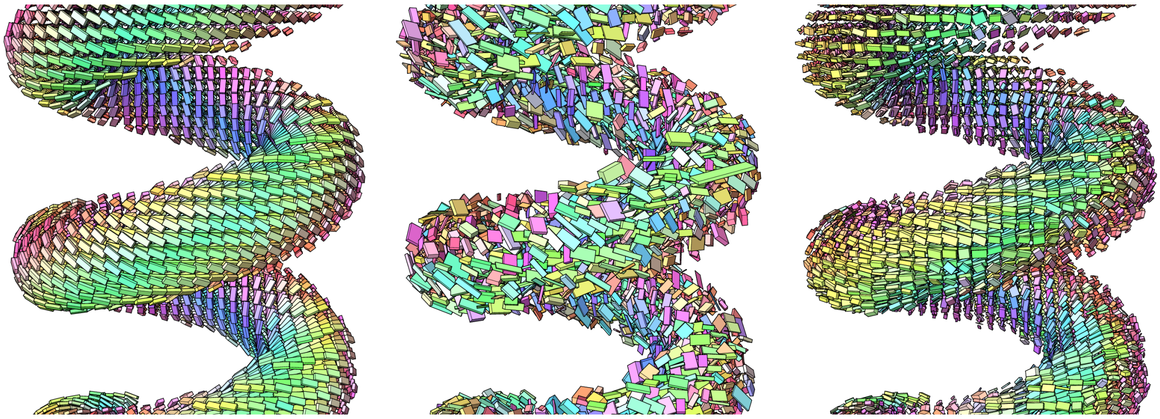
Diffusion tensor imaging is covered by the Stejskal–Tanner equation: given a tensor field $x : \Omega \to \mathrm{Sym}^2(\mathbb{R}^3)$, associating each point on the domain $\Omega \subset \mathbb{R}^3$ with a of symmetric 2-tensor (presentable as a symmetric 3×3 matrix), and a non-diffusion-weighted image $s_0 : \Omega \to \mathbb{R}$, the diffusion-weighted image $s_k : \Omega \to \mathbb{R}$ corresponding to a diffusion-sensitising gradient $b_k \in \mathbb{R}^3$ is given by

$$s_k(\xi) = s_0(\xi) e^{-\langle x(\xi) b_k, b_k \rangle} \quad (\xi \in \Omega). \tag{6.1}$$

At each spatial point ξ, the tensor $x(\xi)$ models the covariance of a Gaussian probability distribution for the spatial directions of the diffusion of water at that point. Models more advanced than DTI, such as HARDI, consider composite probability distributions at each ξ. For our purposes a simplified DTI model will be sufficient. One can measure s_k and s_0 by suitable MRI pulse sequences, inversion of a Fourier transform, and taking the absolute value of a complex number; for details we refer to [1, 19], among others. We recommend [24] as an introduction to MRI.

We want to determine x from noisy measurements of s_0 and s_k, $(k = 1, \ldots, N)$. Clearly, (6.1) can be converted into an invertible system of linear equations with respect to x if $N \geq 6$ and the tensors $b_k \otimes b_k$ are linearly independent. With noise involved, to get a good-quality image, we want to obtain a regularised solution. We therefore consider a problem of the form (P_0) where G is a data term modelling (6.1) along with any noise, and $F \circ K$ is the regulariser. Ideally, our data term would model the Rician noise distribution, which is the distribution of the absolute value of a complex number when the latter has Gaussian noise distribution. However, the numerical treatment of the Rician distribution is quite involved – we refer to [20, 17] for some variational approaches – and instead of modelling it directly, a more fruitful approach may be to work with complex data directly, even incorporating the Fourier transform into our model. For the purposes of the present work, since we only use synthetic data, we will therefore assume that the noise in s_k is Gaussian. We note that (6.2) in infinite dimensions requires the use of the Banach space of functions of bounded deformation, so, since our algorithms require Hilbert spaces, only discretised versions of the model can be considered. Consequently, taking the discretised domain $\Omega_d := \{1, \ldots, n_1\} \times \{1, \ldots, n_2\} \times \{1, \ldots n_3\}$ and incorporating total deformation regularisation with parameter $\alpha > 0$, we seek to solve

$$\min_{x:\Omega_d \to \mathrm{Sym}^2(\mathbb{R}^3)} \frac{1}{2}\|T(x)\|^2 + \alpha \|\mathcal{E}_d x\|_{F,1}, \quad [T(x)]_k := s_k(\xi) - s_0(\xi) e^{-\langle x(\xi) b_k, b_k \rangle} \quad (k = 1, \ldots, N). \tag{6.2}$$



(a) Original helix (b) Least squares reconstruction (c) Regularised reconstruction

Figure 1: Visualisation of original helix data (a) and the reconstruction from noisy diffusion-weighted measurements. The reference least squares reconstruction in (b) is based on linearising (6.1) with respect to x by taking the logarithm. The regularised reconstruction (c) is the numerical solution of (6.2) for $\alpha = 0.005$ with the variant (d2) of our method after 10000 iterations. The visualisation, generated with Teem [40], displays the tensor at each voxel of the 3D volume as a cuboid oriented along the eigenvectors of the tensor, size of each side proportional to the corresponding eigenvalue. The cuboids are also colour-coded based on the principal eigenvector. Tensors with too small eigenvalues are suppressed; in essence this suppresses the background outside the helix, letting the latter to be inspected unobstructedly.

Here $[\mathcal{E}_d x](\xi) \in \mathrm{Sym}^3(\mathbb{R}^3)$ is forward–differences discretisation of the symmetrised gradient, a symmetric third-order tensor. The $F, 1$-norm is based on taking pointwise the Frobenius norm of $[\mathcal{E}_d x](\xi)$ and integration of the space (1-norm). This model is sightly simplified from our previous work in [37, 39, 38], where second-order total generalised variation regularisation was considered and we included a positivity semi-definiteness constraint on $x(\xi)$.

To write (6.2) in the form (S), we take with $y = (\mu, \lambda)$ the functions

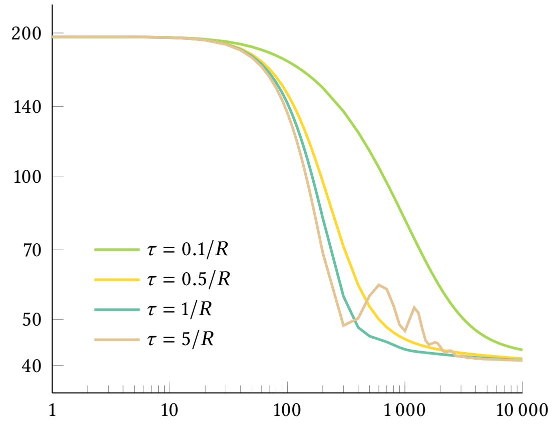
$$G(x) := 0, \;\; K(x) := (\mathcal{E}_d x, T(x)), \;\; F^*(y) := F^*_\mu(\mu) + F^*_\lambda(\lambda), \;\; F^*_\mu(\mu) := \delta_{\alpha\mathbb{B}}(\mu), \;\; F^*_\lambda(\lambda) := \frac{1}{2}\|\lambda\|^2.$$

Here $\mathbb{B}$ is the product of the voxelwise unit balls of $\mathrm{Sym}^3(\mathbb{R}^3)$ over Ω_d. To better satisfy the conditions of our convergence theorems, we replace F^*_μ by $F^*_{\mu,\gamma}(\mu) := \delta_{\alpha\mathbb{B}}(\mu) + \gamma\alpha^{-1}\|\mu\|^2$ with $\gamma = 10^{-9}$. This is the same as applying Moreau–Yosida regularisation to $\|\cdot\|_{F,1}$ in (6.2).

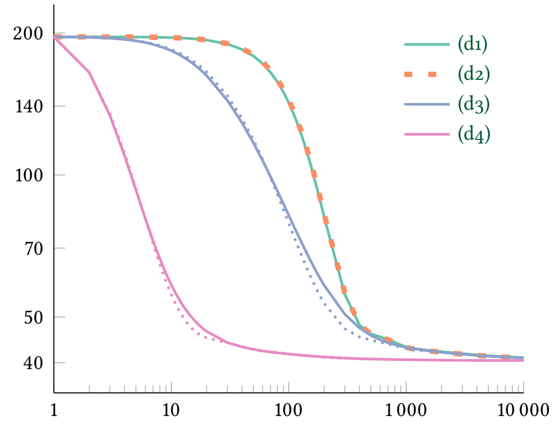
We generated our test data, a simple helix depicted in Figure 1, with the Teem toolkit [40]. The dimensions are $n_1 \times n_2 \times n_3 = 38 \times 39 \times 40$. In the background, outside the helix, the tensors are fully isotropic with the eigenvalues of 10% of the maximal eigenvalue of the tensors within the helix. The exact generation details can be deciphered from our codes [21] written in Julia [3]. After generating the helix data, we took $s_0(\xi) = \|x(\xi)\|_F$. Then we generated s_k, $(k = 1, \ldots, 6)$, from the Stejskal–Tanner equation (6.1) with the diffusion-sensitising gradients $b_1 = (1, 0, 0)$, $b_2 = (0, 1, 0)$, $b_3 = (0, 0, 1)$, $b_4 = (\sqrt{2}, \sqrt{2}, 0)$, $b_5 = (\sqrt{2}, 0, \sqrt{2})$, and $b_6 = (0, \sqrt{2}, \sqrt{2})$. To these diffusion-weighted images we added synthetic Gaussian noise of standard deviation 30% of the mean magnitude of s_0. As the regularisation parameter in the model (6.2) we took $\alpha = 0.005$.

We only consider deterministic updates. We develop step length rules for Algorithm 4.1 based on Theorem 4.4, however, although F^*_λ is strongly convex, and the Moreau–Yosida regularisation makes also $F^*_{\mu,\gamma}$ strongly convex, we generally do *not* employ acceleration and instead keep the step length parameters fixed throughout the iterations. Therefore the theorem does not generally provide any convergence claims.

For convenience, we will identify the linear primal indices j and dual indices ℓ (used for arbitrary blocks) with symbolic indices corresponding to the different variables x, μ, λ and their sub-blocks (used



(a) Multiple step length parametrisations of the non-block-adapted reference algorithm (d1) to justify the choice $\tau = 1/R$.



(b) Comparison of the algorithm variants (d1)–(d4). The dotted lines show the effect of accelerating the dual blocks in (d3) and (d4) following Theorem 4.4.

Figure 2: Reference algorithm step length justification (a) and algorithm performance (b) on the DTI problem. Function values are on the vertical axis, and iteration counts are on the horizontal axis. Based on (a), we take $\tau = 1/R$ in (b): $\tau = 5/R$ appears to have convergence issues and $\tau = 0.5/R$ yields slower convergence.

for specific blocks). The primal variable will be just a single block "x", or be divided into voxelwise blocks "x_ξ" for $\xi \in \Omega_d$. The dual variable will consist of just a single block "y", the two blocks corresponding to the variables "μ" and "λ", or "μ" and the sub-blocks "$\lambda_{k,\xi}$" over $k = 1, \ldots, N$ and $\xi \in \Omega_d$.

Of the conditions of Theorem 4.4, we will not seek to satisfy the boundedness (4.8); following Remark 3.13 this seems likely to hold if we initialise close enough to a solution and take the primal step length parameters $\mathring{\tau}_j^0$ small enough. However, we do not know, how small and how close would be theoretically required. Likewise, (4.7b), which with deterministic updates simplifies to $\delta \geq \mathring{\tau}_j^0 \bar{L}$, is satisfied by taking $\mathring{\tau}_j^0$ small enough. To do this exactly, we would need to calculate the constant L that satisfies the Lipschitz requirement of Assumption 3.1. Assumption 3.4 readily holds (with Moreau–Yosida regularisation, as discussed above) with $\gamma_{G,x} = 0$ and any $0 \leq \gamma_{F^*,\mu} \leq \gamma\alpha^{-1}$ and $0 \leq \gamma_{F^*,\lambda} \leq 1$. We take the latter as well as α_y and ζ_ℓ such that (4.5) yields $\bar{\gamma}_{F^*,\ell} \equiv 0$ for all ℓ. Assumption 3.2 we do not hope to verify in the confines of the present manuscript. With (4.7b) out of the way, for the calculation of the step lengths, it would only be needed for the constants $\gamma_{K,j}$. We simply make the reasonable assumption that we start close enough to a local minimiser satisfying the "second-order necessary condition" $\gamma_{G,j} + \gamma_{K,j} \geq 0$, i.e., $\gamma_{K,j} \geq 0$. Then we may simply assume $\gamma_{K,j} = 0$ and are justified in taking $\widetilde{\gamma}_{G,j} = 0$.

It remains to satisfy the relationship (4.7a) between the primal and dual step lengths. Taking the weights $w_{j,\ell,k} = w^i_{j,\ell,k}$ and the set of connections $\bar{\mathcal{V}}^i_j(\ell) = \bar{\mathcal{V}}_j(\ell)$ given in (4.4b) independent of the iteration and inserting $w_{j,k}$ from (4.4a) into (4.7a), the latter holds if

$$1 - \kappa \geq \left\| \sum_{j=1}^{m} \sqrt{\breve{\sigma}_\ell^0 \mathring{\tau}_j^0 \chi_{\mathcal{V}_j^i}(\ell) \sum_{\ell' \in \bar{\mathcal{V}}_j^i(\ell)} w_{j,\ell,\ell'}} Q_\ell \nabla K(x^i) P_j \right\|^2. \tag{6.3}$$

In particular, with just a single primal block x, we then satisfy (6.3) by taking

$$\breve{\sigma}_\ell^0 = \frac{1-\kappa}{\mathring{\tau}_x^0 \sum_{\ell' \in \bar{\mathcal{V}}_j(\ell)} w_{x,\ell,\ell'} R_\ell^2} \quad \text{where we need the estimate} \quad R_\ell \geq \|Q_\ell \nabla K(x^i)\|. \tag{6.4}$$

Similarly to [5] we estimate $\|\mathcal{E}_d\| \leq R_{\mathcal{E}} := \sqrt{12}$. Assuming that each $x^i(\xi)$ for $\xi \in \Omega_d$ is positive semi-definite, we also estimate with $r_{k,\xi} := |s_0(\xi)|\|b_k\|_2^2$ that

$$\|\nabla T(x^i)\| \le R_T := \sqrt{\sum_{k=1}^{N} \sum_{\xi \in \Omega_d} r_{k,\xi}^2} \quad \text{and} \quad \|\nabla K(x^i)\| \le R := \sqrt{R_{\mathcal{E}}^2 + R_T^2}.$$

We obtain R_ℓ for (6.4) from the same constituents $r_{k,\xi}$ and $R_{\mathcal{E}}$, depending on the exact block structure.

It then remains to choose the primal step lengths and the weights $w_{j,k,\ell}$. We consider the following four block structures and choices of weights:

(d1) As our reference case, corresponding to earlier non-block-adapted works [33, 9], a single primal block x ($m = 1$) and a single dual block y ($n = 1$). Based on the rough optimisation of the step length parameters illustrated in Figure 2a, for a range of $\tau = \mathring{\tau}_x^0$ with $\breve{\sigma}_y^0 = \sigma := (1-\kappa)/(\tau R^2)$ with $\kappa = 0.05$, we take $\tau := 1/R$.

(d2) A single primal block x ($m = 1$) and the two dual blocks μ and λ ($n = 2$). We take $\tau = \mathring{\tau}_1^0$ as in (d1) and with $w_{x,\lambda,\mu} := R_{\mathcal{E}}/(R - R_{\mathcal{E}})$ calculate from (6.4) the dual step length parameters as $\breve{\sigma}_\mu^0 = (1-\kappa)/(\tau(1 + w_{x,\lambda,\mu}^{-1})R_{\mathcal{E}}^2)$ and $\breve{\sigma}_\lambda^0 = (1-\kappa)/(\tau(1 + w_{x,\lambda,\mu})R_T^2)$. Thus $\breve{\sigma}_\mu^0 R_{\mathcal{E}}$ equals σR of (d1).

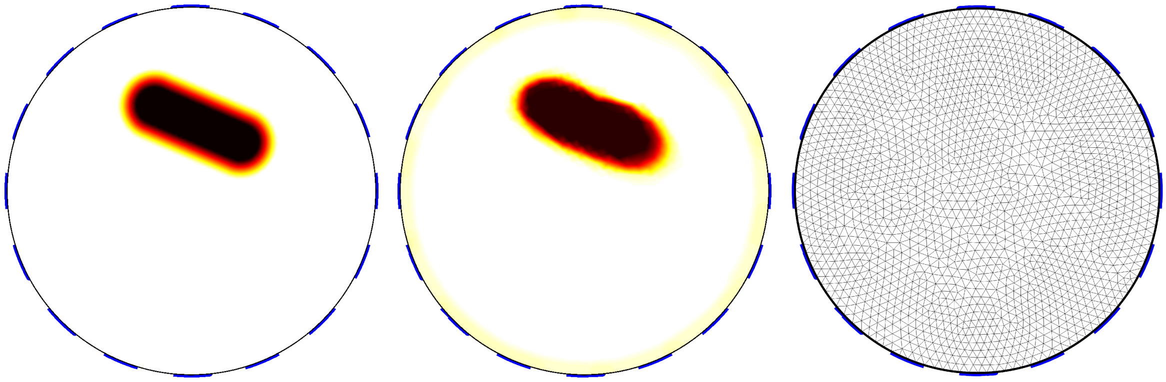
(d3) A single primal block x ($m = 1$) and in addition to the dual block μ, we split λ into voxelwise and b_k-wise blocks $\lambda_{k,\xi}$ ($n = 1 + N n_1 n_2 n_3$) indexed by $k = 1, \ldots, N$ and $\xi \in \Omega_d$. We still take $\tau = \mathring{\tau}_1^0$ as in (d1) and with $w_{x,\lambda_{(k,\xi)},\mu} := \sum_{k',\xi'} r_{k',\xi'} R_{\mathcal{E}}/((R - R_{\mathcal{E}}) r_{k,\xi})$ and $w_{x,\lambda_{(k,\xi)},\lambda_{(k',\xi')}} \equiv 1$ calculate from (6.4) the dual step length parameters as $\breve{\sigma}_\mu^0 := (1-\kappa)/(\tau(1 + \sum_{k,\xi} w_{x,\lambda_{(k,\xi)},\mu}^{-1})R_{\mathcal{E}}^2)$ and $\breve{\sigma}_{\lambda_{k,\xi}}^0 := (1-\kappa)/(\tau(N + w_{x,\lambda_{(k,\xi)},\mu}) r_{k,\xi}^2)$. This also keeps $\breve{\sigma}_\mu^0 R_{\mathcal{E}}$ equal to σR of (d1).

(d4) Voxelwise primal blocks x_ξ for $\xi \in \Omega$ ($n = n_1 n_2 n_3$) in addition to dual blocks as in (d3). We take the blockwise primal step length parameters $\mathring{\tau}_\xi^0 = \tau_\xi := R\tau/(1+N \max_{k=1,\ldots,N} r_{k,\xi})$ for $\xi \in \Omega_d$, where τ is as in (d1). Then we take $w_{x_\xi,\lambda_{(k,\xi)},\mu} := r_{k,\xi}$ and $w_{x_\xi,\lambda_{(k,\xi)},\lambda_{(k',\xi')}} = 1$. Observe that according to the definition of the connection set $\mathcal{V}_j(\ell)$ in (4.4b) that the dual block (k, ξ) is not connected by K to (k', ξ') for $\xi' \neq \xi$. Therefore, we satisfy (6.3) by taking $\breve{\sigma}_\mu^0 = (1-\kappa)/(\max_{\xi\in\Omega_d} \tau_\xi(1+\sum_{k=1}^N r_{k,\xi})R_{\mathcal{E}}^2)$ and $\breve{\sigma}_{\lambda_{k,\xi}}^0 = (1-\kappa)/(\tau_\xi(N + r_{k,\xi}^{-1})r_{k,\xi}^2)$. The maximum comes from estimating the norm in (6.3).

We report in Figure 2b for the first 10000 iterations the function value achieved by each algorithm variant. For (d3) and (d4) we also display the effect of the $O(1/N)$ acceleration of Theorem 4.4; on (d1) and (d2) this has no notable effect.

On a mid-2014 MacBook Pro with a 2.8GHz Intel Core i5 processor and 16GB RAM running Julia 1.1.0, each iteration of (d1)–(d3) takes roughly 0.048 seconds. For (d4) this is roughly 0.062 seconds due to a more complicated primal update.[1] However, in terms of computational times, (d4) is clearly much faster than the other variants: 0.77s against 14.7–19.2s for (d1) and 13.6–18.1s for (d2) and (d3) to reach function value 50. The time ranges account for us sampling the function values only every 100 iterations after the first 100. The visual character of the approximate solution provided by (d4) is on closer inspection slightly smoothed out compared to the other variants. This may be due to non-optimal α in the model (6.2) or due to a different local solution.

[1] In the Julia code [21], we update $x^{i+1}(\xi) := x^i(\xi) - \tau_\xi \Delta x^i(\xi)$ and $\lambda^{i+1}(k,\xi) := (\lambda^i(k,\xi) + \sigma_{k,\xi}\Delta\lambda^i(k,\xi))/(1+\sigma_{k,\xi})$ for some temporary Δx^i and $\Delta\lambda^i$ and all $\xi \in \Omega_d$ and $k = 1, \ldots, N$. The latter does not appear to cause a notable performance penalty compared to a spatially constant σ while the former does. However, each $x^{i+1}(\xi)$ is a tensor consisting of multiple floating point numbers while $\lambda^{i+1}(k,\xi)$ is a single floating point number. Our guess is that, due to uneven memory indexing when τ is spatially varying, the tensor update cannot make as good use of processor SIMD instructions.



(a) Synthetic conductivity (b) Reconstructed conductivity (c) Finite element mesh

Figure 3: Synthetic true conductivity and reconstructed conductivity for the EIT example. The reconstruction is the one obtained with the block structure and dual step length setup of (e3) with $\tau = 500/R$ after 15000 iterations. The blue patches on the boundary of the domain indicate the electrodes. We display in (c) the finite element mesh used to represent the conductivity.

6.2 ELECTRICAL IMPEDANCE TOMOGRAPHY

In this problem, we want to solve

$$\min_{x\in V} \sum_{k=1}^{N} \frac{1}{2}\|A_k(x)\|^2 + \alpha\|\nabla x\|_{2,1} \tag{6.5}$$

on a finite-dimensional subspace $V \subset L^2(\Omega)$ with $\Omega \subset \mathbb{R}^2$ and each $A_k : V \to \mathbb{R}^N$ a non-linear operator corresponding to the fit of the solution of a partial differential equation controlled by x to measured data. We specifically use the complete electrode model of EIT [41]. Our implementation of the model will be described in detail in [18]. The rough idea is that N *electrodes* are placed on the boundary of the domain Ω inside which we want to reconstruct an unknown conductivity x; see Figure 3, which presents a synthetic 2D slice model of an object in a cylindrical water tank. As our data, we only have N boundary measurements corresponding to exciting in turn each of the electrodes $k = 1, \dots, N$ with a positive electric potential. In each of these excitations, the remaining electrodes are grounded, and the electric current generated by these excitations is measured at each electrode, yielding N measurements. The operators A_k correspond to each such excitation setup. In the example of Figure 3, the number of electrodes $N = 16$.

We can again write this problem in the form (S) with

$$G(x) := 0, \quad K(x) := (\nabla x, A_1(x), \dots, A_N(x)), \quad \text{and} \quad F^*(y) = \delta_{\alpha\mathbb{B}}(\mu) + \sum_{k=1}^{N} \|\lambda_k\|_2^2,$$

where $y = (\mu, \lambda_1, \dots, \lambda_N)$ and $\mathbb{B}$ is the product of the pointwise Euclidean unit balls of $\mathbb{R}^2$ over Ω.

As a first case of the dual blocks, we take y_0 corresponding to the total variation term, and the full measurement vectors y_k corresponding to each excitation $k = 1, \dots, N$. We estimate $\|\nabla\| \le R_\nabla$ for R_∇ being the largest singular value of ∇ on V. We do not have exact estimates on the norm of $\nabla A_k(x^i)$. Therefore, we take a dynamic norm estimate $r_k = r_k(i)$ over the last 100 iterations,

$$\|\nabla A_k(x^i)\| \le r_k := 1.05 \max_{\max\{i-99,0\}\le\iota\le i} \|\nabla A_k(x^\iota)\| \quad (k = 1, \dots, N).$$

We may then estimate $\|\nabla K(x^i)\| \le R := \sqrt{R_\nabla^2 + r_1^2 + \cdots + r_N^2}$. As a second case, we further split each y_k into sub-blocks $y_{k,j} \in \mathbb{R}$ corresponding to each individual electrode $j = 1, \dots, N$ being measured.

We then take norm estimates $r_{k,j} = r_{k,j}(i)$ over the last 100 iterations,

$$|[\nabla A_k(x^i)]_j| \le r_{k,j} := 1.05 \max_{\max\{i-99,0\}\le\iota\le i} |[\nabla A_k(x^\iota)]_j| \quad (k, j = 1, \dots, N).$$

We work in the setting of Section 5. Note that unlike Algorithm 4.1 in the DTI experiments of Section 6.1, Algorithm 5.1 allows partial calculation of K in both the primal and dual updates, which should in principle be beneficial in stochastic methods. We develop step length rules for Algorithm 5.1 based on Theorem 5.3. Similarly to (6.4), with $w_{j,\ell,k} = w^i_{j,\ell,k}$ and $\bar{\mathcal{V}}^i_j(\ell) = \bar{\mathcal{V}}_j(\ell)$ independent of the iteration, for non-stochastic methods with a single primal block x, (5.6a) in particular holds by taking

$$\mathring{\sigma}^1_\ell = \frac{1-\kappa}{\tilde{\tau}^0_x \sum_{\ell'\in\bar{\mathcal{V}}_j(\ell)} w_{x,\ell,\ell'} R^2_\ell} \quad \text{where we estimate} \quad R_\ell \ge \|Q_\ell \nabla K(x^i)\|. \tag{6.6}$$

Again, for convenience, we identify the linear primal indices j and dual indices ℓ and ℓ' with symbolic indices x, μ, and λ_k. It then remains to choose $\tilde{\tau}^0_x$ and the weights $w_{x,\ell,\ell'}$. For this we consider four different block and weight setups:

(e1) Again, as our reference case, corresponding to earlier non-block-adapted works [33, 9], a single primal block x ($m = 1$) and a single dual block y ($n = 1$). Based on rough optimisation of the step length parameters, illustrated in Figure 4a for a range of $\tau = \tilde{\tau}^0_x$ with $\mathring{\sigma}^1_y = (1-\kappa)/(\tau R^2)$ with $\kappa = 0.05$, we take $\tau := 5/R$ for R computed using just the initial iterate x^0 as explained above.

(e2) A single primal block x ($m = 1$) and the dual blocks $\mu, \lambda_1, \dots, \lambda_N$. We take $\tau = \tilde{\tau}^0_x$ as in (e1) and with $w_{x,\lambda_p,\mu} := \sum_k r_k R_\nabla/((R - R_\nabla) r_p)$ and $w_{x,\lambda_p,\lambda_k} := 1$ for $p, k = 1, \dots, N$, solve from (6.6) that $\mathring{\sigma}^1_\mu := (1-\kappa)/(\tau(1 + \sum_k w^{-1}_{x,\lambda_k,\mu}) R^2_\nabla)$ and $\mathring{\sigma}^1_{\lambda_p} := (1-\kappa)/(\tau(N + w_{x,\lambda_p,\mu}) r^2_p)$ for $p = 1, \dots, N$. This case and the step length rules are analogous to (d3) for DTI.

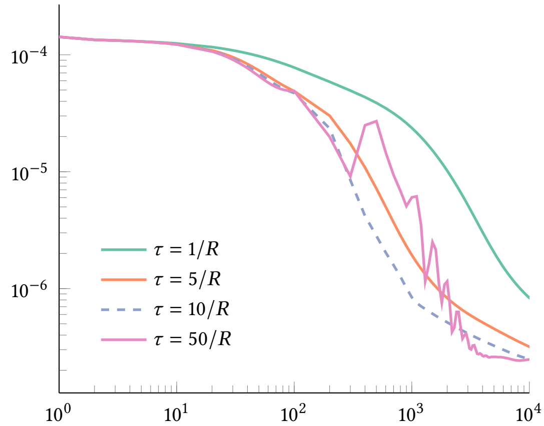
(e3) As (e2) but split each λ_p into further measurement-wise dual blocks $y_{p,j}$ ($p, j = 1, \dots, N$), replacing in the expressions of (e2) the indices p and k by (p, j) and (k, j') with $j, j' \in \{1, \dots, N\}$. Thus r_k becomes $r_{k,j'}$, etc.

(e4) Measurement-wise dual blocks as in (e3) but $w_{x,\lambda_{(p,j)},\mu} := r^{-1}_{p,j}$.

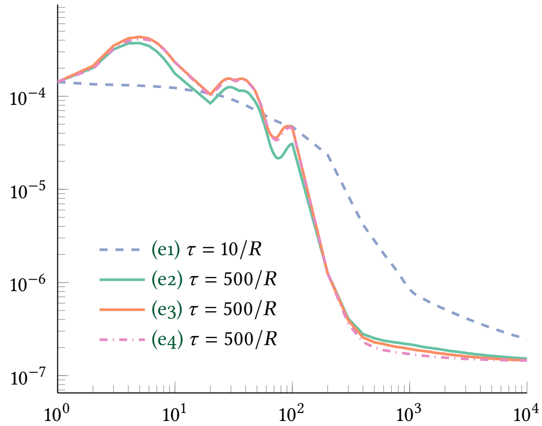
The performance of the algorithm variants (e1)–(e4) is depicted in Figure 4, and a sample reconstruction in Figure 3b. Observe how the block-adapted algorithms allow in practise larger τ than the reference algorithm without block-adaptation. This has significant performance benefits: To reach and stay below objective function value in the order 10^{-7}, (e4) with $\tau = 500/R$ requires 208 iterations while (e1) with $\tau = 10/R$ requires 906 iterations. (With $\tau = 500/R$ the latter requires 3544 iterations, no longer converging well with high τ.) We also tested stochastic variants of the algorithms for the EIT problem, updating on each iteration only a random subset of the dual blocks. This did not, however, offer any performance benefits over the block-adapted variants, neither in terms of epoch count (iteration count scaled by the fraction of updated blocks) nor actual computational time.

7 CONCLUSION

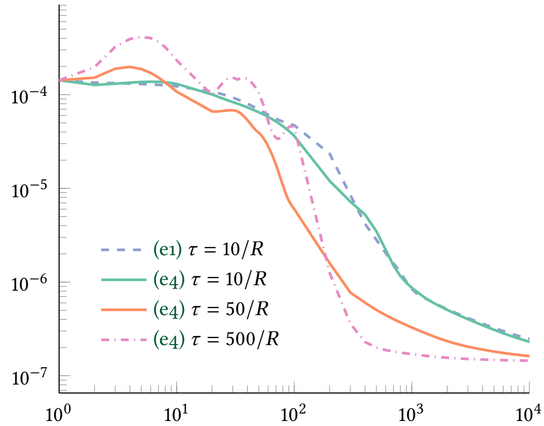
In this paper, we studied block-proximal primal-dual splitting methods for non-convex non-smooth optimisation. From an abstract starting point—also able to model doubly-stochastic methods—we derived explicit algorithms and step-length bounds for two particular cases: methods with full dual updates and methods with full primal updates. For both of the cases, we derived rules ensuring local $O(1/N)$, $O(1/N^2)$ and linear rates under varying conditions and choices of the step lengths parameters.



(a) Reference algorithm (e1), multiple step lengths



(b) Comparison of algorithm variants (e1)–(e4)



(c) Blocked algorithm (e4), multiple step lengths

Figure 4: EIT reconstruction performance: iteration counts are on the x axis and primal objective function values (6.5) are on the y axis. We start with step length justification for the non-blocked reference algorithm (e1) in (a). Based on this we use step length $\tau = 10/R$ for the reference algorithm as higher step lengths become unstable. Comparison of the different blocked algorithm variants is given in (b) for $\tau = 500/R$: with lower parameters the differences are less noticeable, and with higher parameters insignificant improvement is obtained. Based on this, in (c) we represent the performance of (e4) for multiple step lengths.

We demonstrated the performance of the methods on practical inverse problems. Based on our experience with both the DTI and EIT examples, the block-adaptation provides significant performance benefits. Random updates, by contrast, did not offer benefits in our sample problems. We suspect they might be more beneficial on very large scale problems that do not share work between the blocks, yet the blocks have overlapping information, or where communication delays within a computing cluster become significant. This may be one of the possible directions for further research on the presented methods and their application.

A DATA STATEMENT FOR THE EPSRC

The codes and data for the DTI experiments are available at [21]. The codes for EIT, based on historical work of several people, cannot be made available at this point.

APPENDIX A SATISFACTION OF THE THREE-POINT CONDITION

The following lemma provides simplified conditions under which Assumption 3.2 holds, e.g., whenever $x \mapsto \langle K(x), \widehat{y}\rangle$ is block-separable and strongly-convex.

Lemma A.1. *Suppose Assumption 3.1 holds and the following is true for the given neighbourhood $\mathcal{X}_K$ of $\widehat{x}$,*

$\Gamma_K = \sum_{j=1}^m \gamma_{K,j} P_j \in \mathbb{L}(X;X)$, $\gamma_{K,j} \in \mathbb{R}$, some $\gamma_x > 0$:

$$\langle [\nabla K(x') - \nabla K(\widehat{x})]^* \widehat{y}, x' - \widehat{x} \rangle \geq \|x' - \widehat{x}\|^2_{\Gamma_K} + \gamma_x \|x' - \widehat{x}\|^2, \tag{A.1a}$$

$$\langle [P_j \nabla K(x') - P_j \nabla K(\widehat{x})]^* \widehat{y}, x'_j - \widehat{x}_j \rangle \geq \gamma_{K,j} \|x'_j - \widehat{x}_j\|^2 \quad (j = 1, \ldots, m). \tag{A.1b}$$

Let $\beta_1, \beta_2 > 0$, $A = \sum_{j=1}^m a_j P_j$, and $\underline{a} := \min_j a_j$. Then Assumption 3.2 holds for $p = 1$ when

$$L\theta_A \leq \underline{a}(\gamma_x - \beta_1) - \beta_2 \max_j (a_j - \underline{a}) \quad \text{and}$$
$$L_3 \geq L^2 \|P_{\text{NL}} \widehat{y}\| (\beta_1^{-1} + (\beta_2 \underline{a})^{-1} \textstyle\sum_{j=1}^m (a_j - \underline{a}))/2 + 2L\theta_A.$$

Proof. We need to study (3.3). We have

$$\begin{aligned}
R^K &:= \langle [\nabla K(x) - \nabla K(\widehat{x})]^* \widehat{y}, x' - \widehat{x} \rangle_A - \|x' - \widehat{x}\|^2_{A\Gamma_K} \\
&= \underline{a}(\langle [\nabla K(x) - \nabla K(\widehat{x})]^* \widehat{y}, x' - \widehat{x} \rangle - \|x' - \widehat{x}\|^2_{\Gamma_K}) \\
&\quad + \textstyle\sum_{j=1}^m (a_j - \underline{a})(\langle [\nabla K(x) - \nabla K(\widehat{x})]^* \widehat{y}, x'_j - \widehat{x}_j \rangle - \gamma_{K,j} \|x'_j - \widehat{x}_j\|^2).
\end{aligned}$$

We now apply (A.1a), Young's inequality with the factor $\beta_1 > 0$, and Assumption 3.1 to bound

$$\begin{aligned}
&\langle [\nabla K(x) - \nabla K(\widehat{x})]^* \widehat{y}, x' - \widehat{x} \rangle - \|x' - \widehat{x}\|^2_{\Gamma_K} \\
&\quad = \langle [\nabla K(x') - \nabla K(\widehat{x})]^* \widehat{y}, x' - \widehat{x} \rangle - \|x' - \widehat{x}\|^2_{\Gamma_K} + \langle [\nabla K(x) - \nabla K(x')]^* \widehat{y}, x' - \widehat{x} \rangle \\
&\quad \geq (\gamma_x - \beta_1) \|x' - \widehat{x}\|^2 - L^2 \|P_{\text{NL}} \widehat{y}\|^2 (4\beta_1)^{-1} \|x' - x\|^2.
\end{aligned}$$

Similarly, for any $\beta_2 > 0$, we have

$$\begin{aligned}
&\langle [\nabla K(x) - \nabla K(\widehat{x})]^* \widehat{y}, x'_j - \widehat{x}_j \rangle \\
&\quad = \langle [P_j \nabla K(x') - P_j \nabla K(\widehat{x})]^* \widehat{y}, x'_j - \widehat{x}_j \rangle + \langle [\nabla K(x) - \nabla K(x')]^* \widehat{y}, x'_j - \widehat{x}_j \rangle \\
&\quad \geq \gamma_{K,j} \|x'_j - \widehat{x}_j\|^2 - L^2 \|P_{\text{NL}} \widehat{y}\|^2 (4\beta_2)^{-1} \|x' - x\|^2 - \beta_2 \|x'_j - \widehat{x}_j\|^2.
\end{aligned}$$

Combining the two estimates, we arrive at

$$\begin{aligned}
R^K &\geq \underline{a}(\gamma_x - \beta_1) \|x' - \widehat{x}\|^2 - \underline{a} L^2 \|P_{\text{NL}} \widehat{y}\|^2 (4\beta_1)^{-1} \|x' - x\|^2 \\
&\quad - \textstyle\sum_{j=1}^m (a_j - \underline{a})(\beta_2 \|x'_j - \widehat{x}_j\| + L^2 \|P_{\text{NL}} \widehat{y}\|^2 \|x' - x\|^2) \\
&= \textstyle\sum_{j=1}^m (\underline{a}(\gamma_x - \beta_1) - (a_j - \underline{a})\beta_2) \|x'_j - \widehat{x}_j\|^2 \\
&\quad - \underline{a} L^2 \|P_{\text{NL}} \widehat{y}\| (\beta_1^{-1} + (\beta_2 \underline{a})^{-1} \textstyle\sum_{j=1}^m (a_j - \underline{a})) \|x' - x\|^2 / 4.
\end{aligned}$$

At the same time, using Assumption 3.1, we get for the right-hand side of (3.3) the bound

$$\|K(\widehat{x}) - K(x) - \nabla K(x)(\widehat{x} - x)\| \leq \frac{L}{2} \|x - \widehat{x}\|^2 \leq L \|x' - \widehat{x}\|^2 + L \|x' - x\|^2.$$

So Assumption 3.2 holds if we take $p = 1$, $L\theta_A \leq \min_j \underline{a}(\gamma_x - \beta_1) - (a_j - \underline{a})\beta_2$, and $L_3 \geq L^2 \|P_{\text{NL}} \widehat{y}\| (\beta_1^{-1} + (\beta_2 \underline{a})^{-1} \sum_{j=1}^m (a_j - \underline{a}))/2 + 2L\theta_A$. □

APPENDIX B TECHNICAL LEMMA

Lemma B.1. *We have $\bar{z}_N \leq \bar{z}_0 + N/2$ whenever $z^i_j > 0$, $(i = 1, \ldots, N; j = 1, \ldots, m)$ satisfy*

$$z^{i+1}_j = \frac{1 + z^i_j}{\sqrt{1 + \bar{z}_i^{-1}}} \quad \text{with} \quad \bar{z}_i := \max_{j=1,\ldots,m} z^i_j. \tag{B.1}$$

Proof. Taking $\max_{j=1...m}$ on both sides of the first part of (B.1), we obtain

$$\bar{z}_{i+1} = (1 + \bar{z}_i)\sqrt{\frac{\bar{z}_i}{\bar{z}_i + 1}} = \sqrt{\bar{z}_i^2 + \bar{z}_i}.$$

We thus obtain the claim by telescoping

$$\bar{z}_{i+1} - \bar{z}_i = \sqrt{\bar{z}_i^2 + \bar{z}_i} - \bar{z}_i = \frac{\bar{z}_i}{\sqrt{\bar{z}_i^2 + \bar{z}_i} + \bar{z}_i} = \frac{1}{\sqrt{1 + \bar{z}_i^{-1}} + 1} \le \frac{1}{2}. \qquad \square$$

REFERENCES

[1] P. J. Basser and D. K. Jones, Diffusion-tensor MRI: theory, experimental design and data analysis – a technical review., *NMR in Biomedicine* 15 (2002), 456–467, doi:10.1002/nbm.783.

[2] D. P. Bertsekas, Incremental aggregated proximal and augmented Lagrangian algorithms, 2015, arXiv:1509.09257.

[3] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, Julia: A fresh approach to numerical computing, *SIAM Review* 59 (2017), 65–98, https://doi.org/10.1137/141000671.

[4] P. Bianchi, W. Hachem, and I. Franck, A stochastic coordinate descent primal-dual algorithm and applications, in *2014 IEEE International Workshop on Machine Learning for Signal Processing (MLSP)*, IEEE, 2014, 1–6, doi:10.1109/mlsp.2014.6958866.

[5] A. Chambolle, An algorithm for total variation minimization and applications, *Journal of Mathematical Imaging and Vision* 20 (2004), 89–97, doi:10.1023/b:jmiv.0000011325.36760.1e.

[6] A. Chambolle, M. Ehrhardt, P. Richtárik, and C. Schönlieb, Stochastic primal dual hybrid gradient algorithm with arbitrary sampling and imaging applications, *SIAM Journal on Optimization* 28 (2018), 2783–2808, doi:10.1137/17m1134834.

[7] 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, doi:10.1007/s10851-010-0251-1.

[8] R. Chan, H. Yang, and T. Zeng, A two-stage image segmentation method for blurry images with Poisson or multiplicative gamma noise, *SIAM Journal on Imaging Sciences* 7 (2014), 98–127, doi:10.1137/130920241.

[9] C. Clason, S. Mazurenko, and T. Valkonen, Acceleration and global convergence of a first-order primal-dual method for nonconvex problems, *SIAM Journal on Optimization* 29 (2019), 933–963, doi:10.1137/18m1170194, arXiv:1802.03347, https://tuomov.iki.fi/m/nlpdhgm_redo.pdf.

[10] C. Clason, S. Mazurenko, and T. Valkonen, Primal-dual proximal splitting and generalized conjugation in nonsmooth nonconvex optimization, *Applied Mathematics and Optimization* (2020), arXiv:1901.02746, https://tuomov.iki.fi/m/nlpdhgm_general.pdf. accepted.

[11] C. Clason and T. Valkonen, Primal-dual extragradient methods for nonlinear nonsmooth PDE-constrained optimization, *SIAM Journal on Optimization* 27 (2017), 1313–1339, doi:10.1137/16m1080859, arXiv:1606.06219, https://tuomov.iki.fi/m/pdex2_nlpdhgm.pdf.

[12] P. L. Combettes and J. Pesquet, Stochastic forward-backward and primal-dual approximation algorithms with application to online image restoration, in *2016 24th European Signal Processing Conference (EUSIPCO)*, EUSIP, 2016, 1813–1817, doi:10.1109/eusipco.2016.7760561.

[13] D. Csiba, Z. Qu, and P. Richtarik, Stochastic Dual Coordinate Ascent with Adaptive Probabilities, in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach and D. Blei (eds.), volume 37 of Proceedings of Machine Learning Research, PMLR, Lille, France, 2015, 674–683, arXiv:1502.08053, http://proceedings.mlr.press/v37/csiba15.html.

[14] D. Davis and D. Drusvyatskiy, Stochastic model based minimization of weakly convex functions, *SIAM Journal on Optimization* 29 (2019), 207–239, doi:10.1137/18m1178244.

[15] O. Fercoq and P. Bianchi, A coordinate descent primal dual algorithm with large step size and possibly nonseparable functions, *SIAM Journal on Optimization* 29 (2019), 100–134, doi:10.1137/18m1168480.

[16] O. Fercoq and P. Richtárik, Accelerated, parallel, and proximal coordinate descent, *SIAM Journal on Optimization* 25 (2015), 1997–2023, doi:10.1137/130949993.

[17] P. Getreuer, M. Tong, and L. A. Vese, A variational model for the restoration of MR images corrupted by blur and Rician noise, in *Advances in Visual Computing*, volume 6938 of Lecture Notes in Computer Science, Springer, 2011, 686–698, doi:10.1007/978-3-642-24028-7_63.

[18] J. Jauhiainen, P. Kuusela, A. Seppänen, and T. Valkonen, Relaxed Gauss–Newton methods with applications to electrical impedance tomography, 2020, arXiv:2002.08044, https://tuomov.iki.fi/m/gn_overrelax.pdf. submitted.

[19] P. Kingsley, Introduction to diffusion tensor imaging mathematics: Parts I-III, *Concepts in Magnetic Resonance Part A* 28 (2006), 101–179, doi:10.1002/cmr.a.20048.

[20] A. Martín and E. Schiavi, Automatic total generalized variation based DTI Rician denoising, in *Image Analysis and Recognition*, volume 7950 of Lecture Notes in Computer Science, Springer, 2013, 581–588, doi:10.1007/978-3-642-39094-4_66.

[21] S. Mazurenko, J. Jauhiainen, and T. Valkonen, Diffusion tensor imaging codes from "Primal-dual block-proximal splitting for a class of non-convex problems", Software on Zenodo, 2019, doi:10.5281/zenodo.3528137.

[22] A. Milzarek, X. Xiao, S. Cen, Z. Wen, and M. Ulbrich, A stochastic semismooth Newton method for nonsmooth nonconvex optimization, 2018, arXiv:1803.03466.

[23] Y. Nesterov, Efficiency of coordinate descent methods on huge scale optimization problems, *SIAM Journal on Optimization* 22 (2012), 341–362, doi:10.1137/100802001.

[24] D. Nishimura, *Principles of Magnetic Resonance Imaging*, Stanford University, 1996.

[25] Z. Peng, Y. Xu, M. Yan, and W. Yin, ARock: An algorithmic framework for asynchronous parallel coordinate updates, *SIAM Journal on Scientific Computation* 38 (2016), A2851–A2879, doi:10.1137/15m1024950.

[26] M. Pilanci and M. J. Wainwright, Iterative Hessian sketch: fast and accurate solution approximation for constrained least squares, *Journal of Machine Learning Research* 17 (2016), 1–38.

[27] Z. Qu, P. Richtárik, M. Takáč, and O. Fercoq, SDNA: Stochastic dual newton ascent for empirical risk minimization, in *Proceedings of The 33rd International Conference on Machine Learning*, M. F. Balcan and K. Q. Weinberger (eds.), volume 48 of Proceedings of Machine Learning Research, PMLR, New York, New York, USA, 2016, 1823–1832, http://proceedings.mlr.press/v48/qub16.html.

[28] P. Richtárik and M. Takáč, Parallel coordinate descent methods for big data optimization, *Mathematical Programming* (2015), 1–52, doi:10.1007/s10107-015-0901-6.

[29] P. Richtárik and M. Takáč, Distributed coordinate descent method for learning with big data, *Journal of Machine Learning Research* 17 (2016), 1–25, arXiv:1310.2059, http://jmlr.org/papers/v17/15-001.html.

[30] R. T. Rockafellar and R. J. B. Wets, *Variational Analysis*, Springer, 1998, doi:10.1007/978-3-642-02431-3.

[31] S. Shalev-Shwartz and T. Zhang, Accelerated proximal stochastic dual coordinate ascent for regularized loss minimization, *Mathematical Programming* 155 (2014), 105–145, doi:10.1007/s10107-014-0839-0.

[32] T. Suzuki, Stochastic dual coordinate ascent with alternating direction method of multipliers, in *Proceedings of the 31st International Conference on Machine Learning*, E. P. Xing and T. Jebara (eds.), volume 32 of Proceedings of Machine Learning Research, PMLR, Bejing, China, 2014, 736–744, http://proceedings.mlr.press/v32/suzuki14.html.

[33] T. Valkonen, A primal-dual hybrid gradient method for non-linear operators with applications to MRI, *Inverse Problems* 30 (2014), 055012, doi:10.1088/0266-5611/30/5/055012, arXiv:1309.5032, https://tuomov.iki.fi/m/nl-pdhgm.pdf.

[34] T. Valkonen, Testing and non-linear preconditioning of the proximal point method, *Applied Mathematics and Optimization* (2018), doi:10.1007/s00245-018-9541-6, arXiv:1703.05705, https://tuomov.iki.fi/m/proxtest.pdf.

[35] T. Valkonen, Block-proximal methods with spatially adapted acceleration, *Electronic Transactions on Numerical Analysis* 51 (2019), 15–49, doi:10.1553/etna_vol51s15, arXiv:1609.07373, https://tuomov.iki.fi/m/blockcp.pdf.

[36] T. Valkonen, First-order primal-dual methods for nonsmooth nonconvex optimisation, 2019, arXiv:1910.00115, https://tuomov.iki.fi/m/firstorder.pdf. submitted.

[37] T. Valkonen, K. Bredies, and F. Knoll, Total generalised variation in diffusion tensor imaging, *SIAM Journal on Imaging Sciences* 6 (2013), 487–525, doi:10.1137/120867172, https://tuomov.iki.fi/m/dtireg.pdf.

[38] T. Valkonen, F. Knoll, and K. Bredies, TGV for diffusion tensors: A comparison of fidelity functions, *Journal of Inverse and Ill-Posed Problems* 21 (2013), 355–377, doi:10.1515/jip-2013-0005, https://tuomov.iki.fi/m/ipmsproc.pdf. Special issue for IP:M&S 2012, Antalya, Turkey.

[39] T. Valkonen and M. Liebmann, GPU-accelerated regularisation of large diffusion tensor volumes, *Computing* 95 (2013), 771–784, doi:10.1007/s00607-012-0277-x, https://tuomov.iki.fi/m/escoproc.pdf. Special issue for ESCO 2012, Pilsen, Czech Republic.

[40] Various authors, Teem toolkit, version 1.11.0, 2012, http://teem.sourceforge.net.

[41] P. J. Vauhkonen, M. Vauhkonen, T. Savolainen, and J. P. Kaipio, Three-dimensional electrical impedance tomography based on the complete electrode model, *IEEE Transactions on Biomedical Engineering* 46 (1999), 1150–1160, doi:10.1109/10.784147.

[42] S. J. Wright, Coordinate descent algorithms, *Mathematical Programming* 151 (2015), 3–34, doi:10.1007/s10107-015-0892-3.

[43] Y. Xu and W. Yin, Block stochastic gradient iteration for convex and nonconvex optimization, *SIAM Journal on Optimization* 25 (2015), 1686–1716, doi:10.1137/140983938.

[44] Y. Xu and W. Yin, A globally convergent algorithm for nonconvex optimization based on block coordinate update, *Journal of Scientific Computing* 72 (2017), 700–734, doi:10.1007/s10915-017-0376-0.

[45] Y. Zhang and X. Lin, Stochastic primal dual coordinate method for regularized empirical risk minimization, in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach and D. Blei (eds.), volume 37 of Proceedings of Machine Learning Research, PMLR, Lille, France, 2015, 353–361, http://proceedings.mlr.press/v37/zhanga15.html.

[46] P. Zhao and T. Zhang, Stochastic optimization with importance sampling for regularized loss minimization, in *Proceedings of the 32nd International Conference on Machine Learning*, F. Bach and D. Blei (eds.), volume 37 of Proceedings of Machine Learning Research, PMLR, Lille, France, 2015, 1–9, http://proceedings.mlr.press/v37/zhaoa15.html.