



# On the rate of convergence of alternating minimization for non-smooth non-strongly convex optimization in Banach spaces

Jakub Wiktor Both<sup>1</sup> 

Received: 5 November 2020 / Accepted: 15 May 2021  
© The Author(s) 2021

## Abstract

In this paper, the convergence of the fundamental alternating minimization is established for non-smooth non-strongly convex optimization problems in Banach spaces, and novel rates of convergence are provided. As objective function a composition of a smooth, and a block-separable, non-smooth part is considered, covering a large range of applications. For the former, three different relaxations of strong convexity are considered: (i) quasi-strong convexity; (ii) quadratic functional growth; and (iii) plain convexity. With new and improved rates benefiting from both separate steps of the scheme, linear convergence is proved for (i) and (ii), whereas sublinear convergence is showed for (iii).

**Keywords** Convex optimization · Alternating minimization · Rate of convergence · Linear convergence · Sublinear convergence · Banach spaces

## 1 Introduction

The (cyclic) *block coordinate descent* (BCD), in the literature also referred to as non-linear block Gauss-Seidel or successive subspace correction method, is a fundamental optimization algorithm [4, 12]. Given a block structured minimization problem, it consists of the successive minimization with respect to the single blocks. Since numerous applications naturally inherit a block structure, the BCD and its variations have been of great interest for decades—especially whenever it is more convenient or feasible to solve the corresponding subproblems instead of the globally coupled problem. For an overview, we refer to the review paper [15].

The convergence of the BCD has been extensively studied in the literature—typically in Euclidean spaces. For instance, if already partial minimization is

---

✉ Jakub Wiktor Both  
jakub.both@uib.no

<sup>1</sup> Department of Mathematics, University of Bergen, Bergen, Norway

well-defined, any generated accumulation point is a stationary point [4]. Furthermore, convergence has been established under various convexity assumptions as, e.g., strong convexity [1], and quasi-convexity with respect to each block [7,8]. Even more strongly, linear convergence has been proved in the context of (multiplicative Schwarz) domain decomposition methods for smooth and strongly convex problems, here, in Banach spaces [14], and the context of feasible descent methods under stricter convexity assumptions (e.g., strong convexity w.r.t. single blocks) [9]; for the latter, lately the overarching class of smooth convex functions with quadratic functional growth has been identified to lead to linear convergence [11]. Commonly, smoothness assumptions of global kind are made, as e.g., global Lipschitz continuity of the Jacobian.

The BCD for two blocks is entitled *alternating minimization*. It is worth noting that two-block structured problems, appearing in various applications, constitute an important class. In view of this work, we mention an emerging interest for iterative decoupling strategies of two-way coupled partial differential equations, cf., e.g., [5] and references within.

The presence of just two blocks allows for an improved convergence analysis of the BCD in contrast to the general case. For unconstrained smooth convex problems in finite dimensional Euclidean spaces equipped with the  $l_2$  norm, linear convergence has been established under additional strong convexity [3], and moreover sublinear convergence has been showed for problems with non-smooth, block-separable contributions [2]. Both results have in common that the theoretical multiplicative constant merely depends on the minimum of the Lipschitz constants of the partial derivatives, instead of a global one. The proofs essentially utilize knowledge on first-order gradient descent methods as the (proximal) BCD. To our best knowledge, those theoretical convergence results are the finest in the literature.

The motivation for this work has been to generalize and improve the previous convergence results for the alternating minimization. For this purpose, we consider a model problem in (infinite dimensional) Banach spaces incorporating block-separable non-smooth contributions (Sect. 2). The model problem covers a large class of problems, allowing, e.g., for block-separable convex constraints or non-smooth regularization; for more examples, we refer to Beck [2]. Finally, by exploiting tailored norms in the analysis this setting can enable (A) tighter convergence results in (B) a fairly general setup. Furthermore, driven by that fact that strong convexity may be a lot to ask for, for the first time, linear convergence of the alternating minimization is investigated under two relaxations of strong convexity: quasi-strong convexity (Sect. 3), and mere quadratic functional growth without an explicitly required feasible descent property (Sect. 4). For a more complete picture, we additionally study the case of plain convex optimization but in Banach spaces (Sect. 5). An illustrative numerical PDE-based example inspired from multiphysics solved by the alternating minimization is provided in Sect. 6. The results are summarized and discussed in the concluding Sect. 7.

## 2 Alternating minimization for two-block structured model problem

We consider the two-block structured model problem

$$\min \{H(x_1, x_2) \equiv f(x_1, x_2) + g_1(x_1) + g_2(x_2) \mid (x_1, x_2) \in \mathcal{B}_1 \times \mathcal{B}_2\}, \quad (1)$$

where  $\mathcal{B}_1, \mathcal{B}_2, f, g_1, g_2$  satisfy the following properties:

- (P1)  $(\mathcal{B}_i, \|\cdot\|_i)$  is a Banach space with its dual  $(\mathcal{B}_i^*, \|\cdot\|_{i,*})$  and the duality pairing  $\langle \cdot, \cdot \rangle_i, i = 1, 2$ . The index will be omitted for duality pairings.
- (P2) The function  $g_i : \mathcal{B}_i \rightarrow \mathbb{R} \cup \{\infty\}$  is proper convex, (Fréchet) subdifferentiable with subdifferential  $\partial g_i$  on  $\text{dom } g_i, i = 1, 2$ . Let  $\mathcal{D} := \text{dom } g_1 \times \text{dom } g_2$ .
- (P3) The function  $f : \mathcal{B}_1 \times \mathcal{B}_2 \rightarrow \mathbb{R}$  is convex and (Fréchet) differentiable over  $\mathcal{D}$ . Let  $\nabla f$  denote the (Fréchet) derivative of  $f$ .
- (P4) The optimal set of problem (1), denoted by  $X \subset \mathcal{B}_1 \times \mathcal{B}_2$ , is non-empty, and the corresponding optimal value is denoted by  $H^*$ .
- (P5) For any  $(\tilde{x}_1, \tilde{x}_2) \in \mathcal{D}$ , the following problems have minimizers

$$\min_{x_1 \in \mathcal{B}_1} H(x_1, \tilde{x}_2), \quad \text{and} \quad \min_{x_2 \in \mathcal{B}_2} H(\tilde{x}_1, x_2).$$

Exploiting the particular two-block structure, we consider the iterative solution of (1) via the classical *alternating minimization*, cf. Algorithm 1.

---

### Algorithm 1 Alternating minimization for model problem (1)

---

**Initialization:**  $x^0 = (x_1^0, x_2^0) \in \mathcal{D}$ , such that

$$x_2^0 \in \arg \min \{H(x_1^0, x_2) \mid x_2 \in \mathcal{B}_2\}. \quad (2)$$

**General step:** For  $k = 0, 1, \dots$ , given  $x^k \in \mathcal{D}$ , find  $x^{k+1} \in \mathcal{B}_1 \times \mathcal{B}_2$  such that

$$x_1^{k+1} \in \arg \min \{H(x_1, x_2^k) \mid x_1 \in \mathcal{B}_1\}, \quad (3)$$

$$x_2^{k+1} \in \arg \min \{H(x_1^{k+1}, x_2) \mid x_2 \in \mathcal{B}_2\}. \quad (4)$$

**Abbreviation:** For  $k \geq 0$ , define  $x^{k+1/2} := (x_1^{k+1}, x_2^k), H^k := H(x^k), H^{k+1/2} := H(x^{k+1/2})$ .

---

As in [2], the partial optimality condition (2) on the initial guess has been chosen for the sake of simpler notation in the subsequent analysis; we will analyze the convergence behavior of Algorithm 1 under the following additional assumptions on the product structure and smoothness:

- (A1)  $\mathcal{B}_1 \times \mathcal{B}_2$  is equipped with a separate norm  $\|\cdot\|$  and  $\beta_1, \beta_2 \geq 0$ , satisfying

$$\|(x_1, x_2)\|^2 \geq \beta_i \|x_i\|_i^2 \quad \text{for all } (x_1, x_2) \in \mathcal{B}_1 \times \mathcal{B}_2, \quad i = 1, 2. \quad (5)$$

Furthermore,  $\mathcal{B}_1 \times \mathcal{B}_2$  is equipped with a canonical duality pairing  $\langle \cdot, \cdot \rangle$ .

(A2) The partial (Fréchet) derivative of  $f$  with respect to the  $i$ -th component, denoted by  $\nabla_i f \in \mathcal{B}_i^*$ , is Lipschitz continuous with Lipschitz constant  $L_i \in (0, \infty]$ ,  $i = 1, 2$ , with  $\min\{L_1, L_2\} < \infty$ ; exemplarily, for  $i = 1$  (analogously for  $i = 2$ ) it holds that  $\|\nabla_1 f(x_1 + h_1, x_2) - \nabla_1 f(x_1, x_2)\|_{1,*} \leq L_1 \|h_1\|_1$  for all  $(x_1, x_2) \in \mathcal{D}$ ,  $h_1 \in \mathcal{B}_1$ , such that  $x_1 + h_1 \in \text{dom } g_1$ , equivalently by a block version of the so-called descent lemma [2,4]

$$f(x_1 + h_1, x_2) \leq f(x_1, x_2) + \langle \nabla_1 f(x_1, x_2), h_1 \rangle + \frac{L_1}{2} \|h_1\|_1^2. \tag{6}$$

**Remark 1 (Semi-normed spaces)** The following analysis does in fact not require  $\|\cdot\|$  or  $\|\cdot\|_i, i = 1, 2$ , to be positive definite. Consequently, it is sufficient to formulate (5) and (6) as well as convexity properties (specified in each section), with respect to semi-norms. Without introducing additional notation, we also subsequently allow  $\|\cdot\|$  and  $\|\cdot\|_i, i = 1, 2$ , to be merely semi-norms.

### 3 Linear convergence in the quasi-strongly convex case

In this section, linear convergence is established for the alternating minimization applied to model problem (1) under additional quasi-strong convexity for  $f$ :

(A3a) The function  $f : \mathcal{B}_1 \times \mathcal{B}_2 \rightarrow \mathbb{R}$  is quasi-strongly convex w.r.t.  $X$ , with modulus  $\sigma > 0$ , i.e., for all  $x \in \mathcal{D}$  and  $\bar{x} := \arg \min \{\|x - y\| \mid y \in X\}$ , the projection of  $x$  onto  $X$ , it holds

$$f(\bar{x}) \geq f(x) + \langle \nabla f(x), \bar{x} - x \rangle + \frac{\sigma}{2} \|x - \bar{x}\|^2.$$

Any strongly convex function is quasi-strongly convex. Moreover, by convexity of  $g_1$  and  $g_2$ ,  $H$  inherits quasi-strong convexity from  $f$  [with (A3a) stated for subdifferentiable functions].

**Theorem 1 (Q-linear convergence under quasi-strong convexity)** Assume that (P1)–(P5) and (A1), (A2), (A3a) hold. Let  $\{x^k\}_{k \geq 0}$  be the sequence generated by the alternating minimization, cf. Algorithm 1. For all  $k \geq 0$  it holds

$$H^{k+1} - H^* \leq \left(1 - \frac{\sigma\beta_1}{L_1}\right) \left(1 - \frac{\sigma\beta_2}{L_2}\right) (H^k - H^*).$$

**Proof** We consider the first half-step of the alternating minimization and show

$$H^{k+1/2} - H^* \leq \left(1 - \frac{\sigma\beta_1}{L_1}\right) (H^k - H^*) \quad \text{for all } k \in \mathbb{N}_0. \tag{7}$$

By definition, it holds  $\frac{\beta_1}{L_1} \geq 0$ , whereas equality holds if  $\beta_1 = 0$  or  $L_1 = \infty$ . W.l.o.g. we assume that  $\frac{\beta_1}{L_1} > 0$  (since  $H^{k+1/2} \leq H^k$  by construction, the statement (7) follows

immediately for  $\frac{\beta_1}{L_1} = 0$ ). We first utilize: (i) (A3a) and the definition of  $\beta_1$ , cf. (5); (ii) a simple rescaling; and (iii) the fact that  $\frac{\sigma\beta_1}{L_1} \in (0, 1]$  [by (5), (6), (A3a)] and Lipschitz continuity of  $\nabla_1 f$ , cf. (6). For this, let  $\bar{x}^k = (\bar{x}_1^k, \bar{x}_2^k) := \arg \min \{ \|x - x^k\| \mid x \in X \} \in \mathcal{D}$ , with  $H^* = H(\bar{x}^k)$ . Ultimately, it holds

$$\begin{aligned} f(x^k) - f(\bar{x}^k) &\stackrel{(i)}{\leq} \left\langle \nabla f(x^k), x^k - \bar{x}^k \right\rangle - \frac{\sigma\beta_1}{2} \|x^k - \bar{x}_1^k\|_1^2 \\ &\stackrel{(ii)}{=} \frac{L_1}{\sigma\beta_1} \left[ \left\langle \nabla_1 f(x^k), \frac{\sigma\beta_1}{L_1} (x_1^k - \bar{x}_1^k) \right\rangle - \frac{L_1}{2} \left\| \frac{\sigma\beta_1}{L_1} (x_1^k - \bar{x}_1^k) \right\|_1^2 \right] \\ &\quad + \left\langle \nabla_2 f(x^k), x_2^k - \bar{x}_2^k \right\rangle \\ &\stackrel{(iii)}{\leq} \frac{L_1}{\sigma\beta_1} \left[ f(x^k) - f \left( x_1^k + \frac{\sigma\beta_1}{L_1} (\bar{x}_1^k - x_1^k), x_2^k \right) \right] \\ &\quad + \left\langle \nabla_2 f(x^k), x_2^k - \bar{x}_2^k \right\rangle. \end{aligned} \tag{8}$$

Furthermore, by convexity of  $g_1$ , it holds with  $\frac{\sigma\beta_1}{L_1} \in (0, 1]$  that

$$g_1 \left( \frac{\sigma\beta_1}{L_1} \bar{x}_1^k + \left( 1 - \frac{\sigma\beta_1}{L_1} \right) x_1^k \right) \leq \frac{\sigma\beta_1}{L_1} g_1(\bar{x}_1^k) + \left( 1 - \frac{\sigma\beta_1}{L_1} \right) g_1(x_1^k),$$

or equivalently after reordering terms

$$g_1(x_1^k) - g_1(\bar{x}_1^k) \leq \frac{L_1}{\sigma\beta_1} \left[ g_1(x_1^k) - g_1 \left( x_1^k + \frac{\sigma\beta_1}{L_1} (\bar{x}_1^k - x_1^k) \right) \right]. \tag{9}$$

Furthermore, the optimality condition corresponding to the second step of Algorithm 1 reads:  $x_2^k \in \text{dom } g_2$  and  $0 \in \nabla_2 f(x^k) + \partial g_2(x_2^k)$  for all  $k \geq 0$ , which by definition of a subdifferential together with  $\bar{x}_2^k \in \text{dom } g_2$  implies

$$g_2(x_2^k) - g_2(\bar{x}_2^k) \leq - \left\langle \nabla_2 f(x^k), x_2^k - \bar{x}_2^k \right\rangle. \tag{10}$$

Combining (i) Eqs. (8)–(10), and (ii) the optimality of  $x_1^{k+1}$ , cf. (3), yields

$$H^k - H^* \stackrel{(i)}{\leq} \frac{L_1}{\sigma\beta_1} \left[ H^k - H \left( x_1^k + \frac{\sigma\beta_1}{L_1} (\bar{x}_1^k - x_1^k), x_2^k \right) \right] \stackrel{(ii)}{\leq} \frac{L_1}{\sigma\beta_1} (H^k - H^{k+1/2}).$$

Reordering terms finally yields Eq. (7). By symmetry (incl. discussion of  $\frac{\beta_2}{L_2} \geq 0$ ), it holds

$$H^{k+1} - H^* \leq \left( 1 - \frac{\sigma\beta_2}{L_2} \right) (H^{k+1/2} - H^*). \tag{11}$$

Ultimately, combining Eqs. (7) and (11), proves the assertion.  $\square$

### 3.1 Numerical test for quasi-strongly convex minimization in a Euclidean space

To assess the sharpness of Theorem 1 under the use of suitable problem-dependent norms, we consider a two-block structured, unconstrained, quadratic, convex optimization problem in a Euclidean space (here  $\mathbb{R}^{n+m}$ ,  $n, m \in \mathbb{N}$ )

$$\min \left\{ H(\mathbf{x}_1, \mathbf{x}_2) \equiv \frac{1}{2} \left\| \underbrace{\begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 \end{bmatrix}}_{=: \mathbf{A}} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} - \mathbf{b} \right\|_{l_2}^2 \mid \begin{matrix} \mathbf{x}_1 \in \mathbb{R}^n, \\ \mathbf{x}_2 \in \mathbb{R}^m \end{matrix} \right\} \tag{12}$$

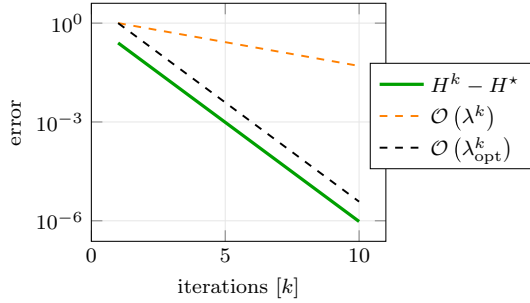
with  $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}, \mathbf{b}$  properly dimensioned. We assume that  $\mathbf{A}$  is non-zero. Then by Theorem 8 in [11], the problem (12) is quasi-strongly convex w.r.t. the Euclidean  $l_2$  norm, with  $\sigma = \sigma_{\min}(\mathbf{A})^2$ , where  $\sigma_{\min}(\cdot)$  denotes the minimal singular value. Furthermore, it satisfies the smoothness and convexity assumptions of Theorem 1 with  $\beta_1 = \beta_2 = 1$ ,  $L_1 = \sigma_{\max}(\mathbf{A}_1)^2$ ,  $L_2 = \sigma_{\max}(\mathbf{A}_2)^2$ , where  $\sigma_{\max}(\cdot)$  denotes the maximal singular value. Ultimately, by Theorem 1, q-linear convergence is guaranteed for all  $k \geq 0$

$$H^{k+1} - H^* \leq \underbrace{\prod_{i=1}^2 \left( 1 - \frac{\sigma_{\min}(\mathbf{A})^2}{\sigma_{\max}(\mathbf{A}_i)^2} \right)}_{=: \lambda} (H^k - H^*). \tag{13}$$

However, the generality of Theorem 1 also allows for utilizing problem-dependent norms, allowing for improving the straight forward result (13). Having Remark 1 in mind, set  $\|\cdot\|_i := \|\cdot\|_{\mathbf{A}_i^\top \mathbf{A}_i}$ ,  $i = 1, 2$ , where  $\|\mathbf{x}\|_{\mathbf{S}}^2 := \mathbf{x}^\top \mathbf{S} \mathbf{x}$  for any symmetric, suitably dimensioned matrix  $\mathbf{S}$ . Consequently, it is  $L_1 = L_2 = 1$ . In addition, let  $\eta > 0$ ,  $\mathbf{I}$  be the identity matrix (in any dimension), and define the norm on the product space by  $\|\cdot\| := \|\cdot\|_{\mathbf{A}_\eta^2}$ , with  $\mathbf{A}_\eta := (\eta \mathbf{I} + \mathbf{A}^\top \mathbf{A})^{1/2}$ . Similarly, set  $\mathbf{A}_{i\eta} := (\eta \mathbf{I} + \mathbf{A}_i^\top \mathbf{A}_i)^{1/2}$ , and the Schur complement  $\mathbf{S}_{\mathbf{A}_{i\eta}^2} := \mathbf{A}_{i\eta}^2 - \mathbf{A}_i^\top \mathbf{A}_j \mathbf{A}_{j\eta}^{-2} \mathbf{A}_j^\top \mathbf{A}_i$ , where  $j \in \{1, 2\}$ ,  $j \neq i$ ,  $i = 1, 2$ . In order to determine  $\sigma$  and  $\beta_i$ , it follows from standard linear algebra that

$$\begin{aligned} \left\| \mathbf{A} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \right\|_{l_2}^2 &\geq \sigma_{\min} \left( \mathbf{A} \mathbf{A}_\eta^{-1} \right)^2 \left\| \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} \right\|_{\mathbf{A}_\eta^2}^2 = \sigma_{\min} \left( \mathbf{A} \mathbf{A}_\eta^{-1} \right)^2 \|\mathbf{x}_1, \mathbf{x}_2\|^2, \quad \text{and} \\ \|\mathbf{x}_1, \mathbf{x}_2\|_{\mathbf{A}_\eta^2}^2 &\geq \|\mathbf{x}_i\|_{\mathbf{S}_{\mathbf{A}_{i\eta}^2}}^2 \geq \sigma_{\min} \left( \mathbf{S}_{\mathbf{A}_{i\eta}^2}^{1/2} \mathbf{A}_{i\eta}^{-1} \right)^2 \|\mathbf{x}_1\|_{\mathbf{A}_{i\eta}^2}^2 \geq \sigma_{\min} \left( \mathbf{S}_{\mathbf{A}_{i\eta}^2}^{1/2} \mathbf{A}_{i\eta}^{-1} \right)^2 \|\mathbf{x}_1\|_i^2. \end{aligned}$$

**Fig. 1** Practical and theoretical convergence for the alternating minimization solving (12)



Finally,  $\sigma$  and  $\beta_i$  are obtained by maximizing the singular values w.r.t.  $\eta$ , equivalent with the limit  $\eta \rightarrow 0$ . Thus, Theorem 1 predicts that for all  $k \geq 0$  it holds

$$H^{k+1} - H^* \leq \underbrace{\prod_{i=1}^2 \left( 1 - \lim_{\eta \rightarrow 0} \sigma_{\min} \left( \mathbf{A} \mathbf{A}_{i\eta}^{-1} \right)^2 \sigma_{\min} \left( \mathbf{S}_{\mathbf{A}_{i\eta}^2}^{1/2} \mathbf{A}_{i\eta}^{-1} \right)^2 \right)}_{=:\lambda_{opt}} (H^k - H^*). \tag{14}$$

Using a small example, we demonstrate the sharpness of (14) opposing to (13). Let

$$\mathbf{A}_1 := \begin{bmatrix} 0 & 0 \\ 1 & -2 \\ 1 & 1 \end{bmatrix}, \quad \mathbf{A}_2 := \begin{bmatrix} 1 & -1 \\ 0 & 0 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{b} := \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

For this choice, the two bounds in (13) and (14) are given by  $\lambda \approx 0.717$  and  $\lambda_{opt} \approx 0.245$ , respectively. In Fig. 1, the theoretical and actual performances of the alternating minimization applied to (12) are visualized for the initial guess  $\mathbf{x}_1^0 := \mathbf{0}$ . We observe a good agreement between the practical convergence rate and the theoretical bound  $\lambda_{opt}$ , stemming from the analysis using problem-dependent norms.

### 4 Linear convergence in the quadratic functional growth case

In this section, linear convergence is established for the alternating minimization applied to model problem (1) under additional quadratic growth for  $H$ :

- (A3b) The objective function  $H : \mathcal{B}_1 \times \mathcal{B}_2 \rightarrow \mathbb{R}$  has *quadratic functional growth* w.r.t.  $X$  with modulus  $\kappa > 0$ ; i.e., for all  $x \in \mathcal{D}$  and  $\bar{x}$  (as in (A3a)), it holds

$$H(x) - H(\bar{x}) \geq \frac{\kappa}{2} \|x - \bar{x}\|^2.$$

Quasi-strong convexity implies quadratic functional growth [11], but not vice versa; functions satisfying (A3b) do not require to be convex [16]. We refer to [11] to examples.

Following a similar strategy as in the proof of Theorem 1, we show q-linear convergence. We stress that opposing to the analysis of general feasible descent methods for problems with quadratic functional growth, cf., e.g., [11], a feasible descent property—ensured e.g. for block coordinatewise strongly convex functions—is not explicitly required for a mere two-block structure.

**Theorem 2** (Q-linear convergence under quadratic functional growth) *Assume (P1)–(P5) and (A1), (A2), (A3b) hold. Let  $\{x^k\}_{k \geq 0}$  be the sequence generated by the alternating minimization, cf. Algorithm 1. For all  $k \geq 0$  it holds*

$$H^{k+1} - H^* \leq \left(1 - \frac{\kappa\beta_1}{8L_1}\right) \left(1 - \frac{\kappa\beta_2}{8L_2}\right) (H^k - H^*).$$

**Proof** We consider the first half-step of the alternating minimization and show

$$H^{k+1/2} - H^* \leq \left(1 - \frac{\kappa\beta_1}{8L_1}\right) (H^k - H^*). \tag{15}$$

W.l.o.g. we assume that  $\frac{\beta_1}{L_1} > 0$ . Let  $\bar{x}^k := \arg \min \{\|x - x^k\| \mid x \in X\} \in \mathcal{D}$ , with  $H^* = H(\bar{x}^k)$ . Utilizing the convexity and smoothness of  $f$ , we then obtain

$$f(x^k) - f(\bar{x}^k) \leq \left\langle \nabla_1 f(x^k), x_1^k - \bar{x}_1^k \right\rangle + \left\langle \nabla_2 f(x^k), x_2^k - \bar{x}_2^k \right\rangle. \tag{16}$$

By (i) introducing  $\gamma \in (0, 1]$  to be specified later, (ii) using the Lipschitz continuity of  $\nabla_1 f$ , cf. (A2), and the convexity of  $f$ , and (iii) the definition of  $\beta_1$ , cf. Eq. (5), we moreover obtain

$$\begin{aligned} & \left\langle \nabla_1 f(x^k), x_1^k - \bar{x}_1^k \right\rangle \\ & \stackrel{(i)}{=} \left\langle \nabla_1 f(x_1^k, x_2^k) - \nabla_1 f\left(x_1^k + \gamma\left(\bar{x}_1^k - x_1^k\right), x_2^k\right), x_1^k - \bar{x}_1^k \right\rangle \\ & \quad + \frac{1}{\gamma} \left\langle \nabla_1 f\left(x_1^k + \gamma\left(\bar{x}_1^k - x_1^k\right), x_2^k\right), \gamma\left(x_1^k - \bar{x}_1^k\right) \right\rangle \\ & \stackrel{(ii)}{\leq} L_1 \gamma \|x_1^k - \bar{x}_1^k\|_1^2 + \frac{1}{\gamma} \left[ f\left(x_1^k, x_2^k\right) - f\left(x_1^k + \gamma\left(\bar{x}_1^k - x_1^k\right), x_2^k\right) \right] \\ & \stackrel{(iii)}{\leq} \frac{L_1}{\beta_1} \gamma \|x^k - \bar{x}^k\|^2 + \frac{1}{\gamma} \left[ f\left(x_1^k, x_2^k\right) - f\left(x_1^k + \gamma\left(\bar{x}_1^k - x_1^k\right), x_2^k\right) \right]. \end{aligned} \tag{17}$$

Based on same grounds as utilized for deriving (9) and (10), it holds

$$g_1(x_1^k) - g_1(\bar{x}_1^k) \leq \frac{1}{\gamma} \left[ g_1(x_1^k) - g_1\left(x_1^k + \gamma\left(\bar{x}_1^k - x_1^k\right)\right) \right], \tag{18}$$

$$g_2(x_2^k) - g_2(\bar{x}_2^k) \leq -\left\langle \nabla_2 f(x^k), x_2^k - \bar{x}_2^k \right\rangle. \tag{19}$$



By definition of  $H$  and (16)–(19), we obtain

$$H^k - H^* \leq \frac{L_1}{\beta_1} \gamma \|x^k - \bar{x}^k\|^2 + \frac{1}{\gamma} \left[ H(x^k) - H \left( x_1^k + \gamma (\bar{x}_1^k - x_1^k), x_2^k \right) \right]. \quad (20)$$

Thus, by utilizing (A3b), the optimality property of  $x_1^{k+1}$  based on the first step of the alternating minimization, cf. (3), and choosing  $\gamma = \frac{\kappa\beta_1}{4L_1}$ , it follows

$$H^k - H^* \leq \frac{1}{2} (H^k - H^*) + \frac{4L_1}{\kappa\beta_1} (H^k - H^{k+1/2}),$$

which yields (15), after reordering. By symmetry, it analogously follows that

$$H^{k+1} - H^* \leq \left( 1 - \frac{\kappa\beta_2}{8L_2} \right) (H^{k+1/2} - H^*). \quad (21)$$

Finally, combining Eqs. (15) and (21) proves the assertion. □

### 5 Sublinear convergence in the plain convex case

In this section, sublinear convergence is established for the alternating minimization applied to model problem (1) under the mild assumption of a compact level set of  $H$  w.r.t. the initial value, inspired by Beck [2]:

(A3c) The functions  $g_i : \mathcal{B}_i \rightarrow \mathbb{R} \cup \{\infty\}$ ,  $i = 1, 2$ , are closed convex (and thereby  $H$ ). Furthermore, the level set of  $H$  with respect to  $H(x^0)$ ,  $\mathcal{L} := \{x \in \mathcal{D} \mid H(x) \leq H(x^0)\}$ , be compact; let  $R := \text{diam}(\mathcal{L}, X)$ .

The following result predicts a two-stage behavior: first, the error decreases q-linearly until sufficiently small; after that, sublinear convergence is initiated. The shift depends on the smoothness properties of the problem.

**Theorem 3** (Sublinear convergence for the non-smooth convex case) *Assume that (P1)–(P5) and (A1), (A2), (A3c) are satisfied. Let  $\{x^k\}_{k \geq 0}$  be the sequence generated by the alternating minimization, cf. Algorithm 1. Define*

$$m^* := \left[ -1 + \left[ \log_2 \left( \frac{H^0 - H^*}{\min \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\} R^2} \right) \right] \right]_+, \quad p^* := \frac{2 \left( \frac{\beta_1}{L_1} + \frac{\beta_2}{L_2} \right)^{-1}}{\min \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\}} \in [1, 2],$$

where  $\lceil \cdot \rceil$  and  $[\cdot]_+$  respectively denote the ceiling function and the restriction to the positive part. It holds for all  $k \geq 0$

$$H^k - H^* \leq \max \left\{ \left( \frac{1}{2} \right)^k (H^0 - H^*), \frac{4R^2 \left( \frac{\beta_1}{L_1} + \frac{\beta_2}{L_2} \right)^{-1}}{[k - m^*]_+ + p^*} \right\}.$$

In particular, for  $k \geq m^*$  at the earliest, sublinear convergence kicks in.

The proof utilizes two auxiliary results: general descent properties for each subiteration of the alternating minimization, and a criterion for concluding sublinear convergence. Those are summarized in the following two lemmas.

**Lemma 1** *Under the assumptions of Theorem 3, it holds for all  $k \geq 0$  that*

$$H^k - H^{k+1/2} \geq \min \left\{ \frac{1}{2}, \frac{\beta_1}{4L_1R^2}(H^k - H^*) \right\} (H^k - H^*), \tag{22}$$

$$H^{k+1/2} - H^{k+1} \geq \min \left\{ \frac{1}{2}, \frac{\beta_2}{4L_2R^2}(H^{k+1/2} - H^*) \right\} (H^{k+1/2} - H^*). \tag{23}$$

**Proof** We show Eq. (22), assuming w.l.o.g.  $\frac{\beta_1}{L_1} > 0$ . As in the proof of Theorem 2, Eq. (20) can be derived under given assumptions; i.e., for  $\gamma \in (0, 1]$  it holds

$$H^k - H^* \leq \frac{L_1}{\beta_1} \gamma \|x^k - \bar{x}^k\|^2 + \frac{1}{\gamma} \left[ H(x^k) - H \left( x_1^k + \gamma (\bar{x}_1^k - x_1^k), x_2^k \right) \right].$$

By definition of  $R$ , cf. (A3c), and the monotonicity of  $\{H(x^k)\}_{k=0, \frac{1}{2}, 1, \dots}$ , it holds  $\|x^k - \bar{x}^k\| \leq R$ . Thus, with the definition of  $x^{k+1/2}$ , cf. Eq. (3), it follows

$$H^k - H^* \leq \frac{L_1R^2}{\beta_1} \gamma + \frac{1}{\gamma} (H^k - H^{k+1/2}).$$

We distinguish two cases: If  $H^k - H^* > \frac{2L_1R^2}{\beta_1}$ , we choose  $\gamma = 1$ ; otherwise, we choose  $\gamma = \frac{\beta_1}{2L_1R^2}(H^k - H^*)$ . This finally proves the first part of the assertion (22). The second part (23) analogously follows by symmetry.  $\square$

The following auxiliary convergence criterion, inspired by a similar result in [3], will allow for effectively making use of the energy descent of both steps of the alternating minimization.

**Lemma 2** *Let  $\{A_k\}_{k=0, \frac{1}{2}, 1, \dots} \subset \mathbb{R}_{\geq 0}$  and  $\gamma_1, \gamma_2, p \geq 0$  satisfy*

$$A_k - A_{k+1/2} \geq \gamma_1 A_k^2 \quad \text{for all } k \geq 0, \tag{24a}$$

$$A_{k+1/2} - A_{k+1} \geq \gamma_2 A_{k+1/2}^2 \quad \text{for all } k \geq 0, \tag{24b}$$

$$A_0 \leq (p(\gamma_1 + \gamma_2))^{-1}. \tag{24c}$$

*Then it holds for all  $k \geq 0$  that  $A_k \leq [(k + p)(\gamma_1 + \gamma_2)]^{-1}$ .*

**Proof** By (24a) and (24b),  $\{A_k\}_{k=0, \frac{1}{2}, 1, \frac{3}{2}, \dots}$  is non-increasing, and it holds

$$\frac{1}{A_{k+1}} - \frac{1}{A_k} = \frac{A_k - A_{k+1/2}}{A_k A_{k+1/2}} + \frac{A_{k+1/2} - A_{k+1}}{A_{k+1/2} A_{k+1}} \geq \gamma_1 + \gamma_2,$$

for  $k \geq 0$ . Thus, by utilizing a telescope sum and applying Eq. (24c), we obtain

$$\frac{1}{A_{k+1}} = \left( \frac{1}{A_{k+1}} - \frac{1}{A_k} \right) + \left( \frac{1}{A_k} - \frac{1}{A_{k-1}} \right) + \dots + \frac{1}{A_0} \geq (k + 1 + p)(\gamma_1 + \gamma_2).$$

This proves the assertion for  $k \geq 1$ ; for  $k = 0$  it follows directly from (24c).  $\square$

Finally, we are able to prove Theorem 3.

**Proof of Theorem 3** As long as  $H^k - H^* > 2 \min \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\} R^2$  for some  $k \in \mathbb{N}_0$ , by Lemma 1 and the monotonicity of  $\{H^k\}_{k=0,1,\dots}$ , it holds that

$$H^k - H^* \leq \left( \frac{1}{2} \right)^k (H^0 - H^*). \tag{25}$$

Thereby, there exists a minimal  $m \geq 0$  such that  $H^k - H^* \leq 2 \min \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\} R^2$  for all  $k \geq m$ . Assuming  $m \geq 1$ , Eq. (25) holds for all  $k \leq m - 1$ , and it holds

$$2 \min \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\} R^2 < H^{m-1} - H^* \leq \frac{1}{2^{m-1}} (H^0 - H^*).$$

Thus, it holds that  $m < \log_2 \left( \frac{H^0 - H^*}{\min \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\} R^2} \right)$ , and consequently (including the case  $m = 0$ ),  $m \leq m^*$ , with  $m^*$  as defined above.

Since  $\{H^k\}_{k=0, \frac{1}{2}, 1, \dots}$  is non-increasing, it also holds for  $k \geq m$  that  $H^{k+1/2} - H^* \leq 2 \min \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\} R^2$ . Hence, by Lemma 1 it follows for all  $k \geq m$  that

$$H^k - H^{k+1/2} \geq \frac{\beta_1}{4L_1R^2} (H^k - H^*)^2, \tag{26a}$$

$$H^{k+1/2} - H^{k+1} \geq \frac{\beta_2}{4L_2R^2} (H^{k+1/2} - H^*)^2. \tag{26b}$$

Using the notation of Lemma 2, we define the sequence  $\{A_n\}_{n=0, \frac{1}{2}, 1, \dots}$  with  $A_n := H^{n+m} - H^*$ , satisfying the assumptions of Lemma 2 with  $\gamma_1 = \frac{\beta_1}{4L_1R^2}$ ,  $\gamma_2 = \frac{\beta_2}{4L_2R^2}$ ,  $p = p^*$ . Finally, the application of Lemma 2 yields

$$H^k - H^* \leq \frac{4R^2 \left( \frac{\beta_1}{L_1} + \frac{\beta_2}{L_2} \right)^{-1}}{k - m + p^*} \leq \frac{4R^2 \left( \frac{\beta_1}{L_1} + \frac{\beta_2}{L_2} \right)^{-1}}{[k - m^*]_+ + p^*} \text{ for all } k \geq m. \tag{27}$$

Combining Eqs. (25) and (27) proves the assertion.  $\square$

**Remark 2** (*Exponential decay during the first iterations*) In the case it holds  $\max \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\} < \infty$ , and the initial error satisfies  $H^0 - H^* > 2 \max \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\} R^2$ , the result of Theorem 3 can be in fact improved. By an analogous line of argumentation as in the above proof, one can conclude that  $H^k - H^*$  first contracts with a rate of  $\frac{1}{4}$  for the first  $k_1$  iterations, until  $H^{k_1} - H^* \leq 2 \max \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\} R^2$  for some  $k_1 \in \mathbb{N}_0$ . Afterwards, the convergence behavior can be qualitatively predicted as in Theorem 3. Ultimately,  $m^*$  is of the order

$$m^* \approx \left\lceil \log_4 \left( \frac{H^0 - H^*}{2 \max \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\} R^2} \right) \right\rceil + \left\lceil \log_2 \left( \frac{\max \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\}}{\min \left\{ \frac{L_1}{\beta_1}, \frac{L_2}{\beta_2} \right\}} \right) \right\rceil.$$

### 6 Numerical example inspired by multiphysics

Sequential solution strategies are widely used in the context of multiphysics applications. Provided a multiphysics problem enjoys a minimization structure, a sequential solution is closely related (or even equivalent) to applying alternating minimization to the underlying minimization problem.

In the following, we numerically demonstrate the efficacy of alternating minimization to a problem, inspired by poroelasticity applications, i.e., flow in deformable porous media. The following model problem corresponds to an elasticity-like vectorial p-Laplace equation coupled with a Darcy-type equation for non-Newtonian fluids, with a Biot–Darcy-type coupling, see [5,10] for more details. For instance, we consider the representative coupled problem

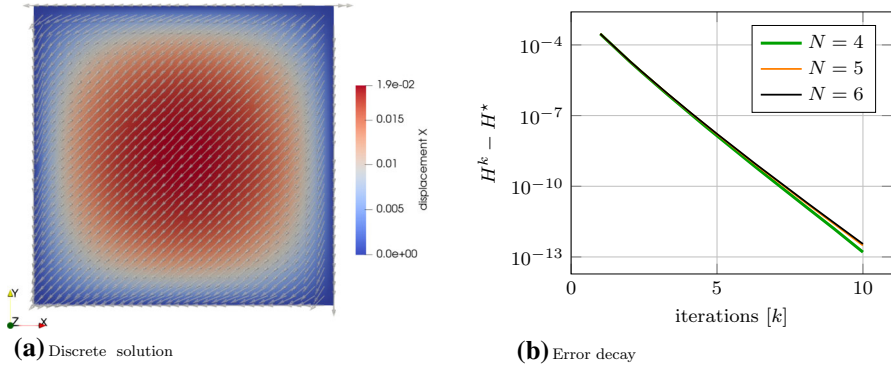
$$\min \left\{ \int_{\Omega} \left[ \frac{1}{2} |\nabla \cdot (\alpha \mathbf{u} + \beta \mathbf{q})|^2 + \frac{\mu}{p} |\nabla \mathbf{u}|^p + \frac{\kappa}{q} |\mathbf{q}|^q - \mathbf{f} \cdot \mathbf{u} \right] dx \mid (\mathbf{u}, \mathbf{q}) \in \mathcal{U} \times \mathcal{Q} \right\}, \tag{28}$$

where  $\Omega = (0, 1) \times (0, 1) \subset \mathbb{R}^2$  denotes the domain,  $\alpha, \beta \in \mathbb{R}, \mu, \kappa \in \mathbb{R}_{>0}, \mathbf{f} \in \mathbb{R}^2$  are model parameters,  $p, q \in (1, \infty)$ , and the solution spaces are defined by

$$\begin{aligned} \mathcal{U} &:= \left\{ \mathbf{v} \in L^p(\Omega; \mathbb{R}^2) \mid \nabla \mathbf{v} \in L^p(\Omega; \mathbb{R}^{2 \times 2}), \nabla \cdot \mathbf{v} \in L^2(\Omega; \mathbb{R}), \mathbf{v}|_{\partial\Omega} = \mathbf{0} \right\}, \\ \mathcal{Q} &:= \left\{ \mathbf{v} \in L^q(\Omega; \mathbb{R}^2) \mid \nabla \cdot \mathbf{v} \in L^2(\Omega; \mathbb{R}), \mathbf{v}|_{\partial\Omega} \cdot \mathbf{n}_{\partial\Omega} = 0 \right\} \end{aligned}$$

where  $L^p$  (resp.  $L^q$ ) denotes the standard Lebesgue space and  $\mathbf{n}_{\partial\Omega}$  is the outer normal vector on the boundary  $\partial\Omega$  of  $\Omega$ . We note the solution spaces  $\mathcal{U}$  and  $\mathcal{Q}$  are closely related to the standard Sobolev spaces  $W_0^{1,p}(\Omega)$  and  $H_0(\text{div}; \Omega)$ , respectively. We fix  $\alpha = 1, \beta = 10, \mu = 1, \kappa = 0.1, \mathbf{f} = (1, 1), p = q = 1.5$ . The corresponding solution is displayed in Fig. 2a.

For the numerical solution, the problem (28) is discretized using the Galerkin method and linear finite elements for  $\mathbf{u}$  and  $\mathbf{q}$  on a Cartesian grid with uniform mesh



**Fig. 2** **a** Discrete solution, displacement-like  $u_x$  ( $x$ -component of  $\mathbf{u}$ , whereas  $y$ -component is identical) and flux-like  $\mathbf{q}$  (visualized by arrows). **b** Error decay for different mesh sizes  $2^{-N}$

size  $2^{-N}$  with  $N \in \{4, 5, 6\}$ . The corresponding discrete minimization problem is then solved using Alg. 1 with an initial guess  $(\mathbf{u}^0, \mathbf{q}^0) = (\mathbf{0}, \mathbf{0})$ . For the implementation, the DUNE project [13] and in particular the dune-functions module [6] have been utilized.

Let  $H^*$  denote the energy corresponding to the (converged discrete) solution of (28), and  $H^k$  the energy of the approximation  $(\mathbf{u}^k, \mathbf{q}^k)$  of the  $k$ -th step of Algorithm 1. The decay  $H^k - H^*$  is displayed in Fig. 2b for the three mesh sizes. We observe linear, essentially mesh-independent convergence. In addition, we mention a decreasing trend for the energy values  $H^*$  for consecutively refined grid, as expected due to the consecutively more accurate discretization. In particular, it is  $H^* \approx -7.077e - 3$  for  $N = 4$ ,  $H^* \approx -7.137e - 3$  for  $N = 5$ ,  $H^* \approx -7.153e - 03$  for  $N = 6$ .

We note the choices for  $p$  and  $q$  lead to a non-quadratic problem, whose coupling however is governed by a quadratic, merely semi-definite contribution. Hence, the considered problem is closely related with the small algebraic problem in Sect. 3.1, and after all leads to consistent observations. The in principle mesh-independent convergence demonstrates that convergence is most adequately described in problem-dependent, i.e., not standard Euclidean norms, which would in contrast suggest mesh-dependent convergence.

## 7 Discussion and concluding remarks

In this paper, we have established convergence of the alternating minimization applied to a two-block structured model problem within the class of non-smooth non-strongly convex optimization in general Banach spaces – a fairly broad setting. We have considered three cases of relaxed strong convexity: (i) quasi-strong convexity, (ii) quadratic functional growth, and (iii) plain convexity and a compact initial level set. Convergence rates have been provided, of linear type for the first two cases, and of sublinear type for the third case. To the best of the author’s knowledge, all results are novel.

Our results are direct extensions of previous results in the literature [2,3,11], agreeing with or partially refining them if put in the same context, and being valid also in more general scenarios. The key for arriving at our results has been the exploitation of describing smoothness properties (of the two single blocks) and convexity properties (of the full objective function) wrt. different (semi-)norms; these enter the novel rates predicting in particular that both steps of the alternating minimization separately lead to an error decrease. For the subclass of quasi-strongly convex problems, we demonstrate the sharpness of our convergence result, based on a simple numerical example. In addition, an illustrative numerical example inspired by multiphysics demonstrates the efficacy of alternating minimization for PDE-based problems. Finally, we highlight that for the first time, it is proved that quadratic functional growth is sufficient for linear convergence – without any feasible descent property as commonly required in the analysis of the general block coordinate descent [9,11].

Ultimately, it is noteworthy that the provided results allow for a systematic development and analysis of iterative block-partitioned solvers based on the alternating minimization for problems in applied variational calculus – in particular two-way coupled PDEs arising from a convex minimization problem, see, e.g., [5].

**Acknowledgements** The author thanks the anonymous reviewers for providing constructive comments resulting in significant improvements of the paper.

**Funding** Open access funding provided by University of Bergen (incl Haukeland University Hospital).

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

1. Auslender, A.: *Optimisation*. Masson, Paris (1999)
2. Beck, A.: On the convergence of alternating minimization for convex programming with applications to iteratively reweighted least squares and decomposition schemes. *SIAM J. Optim.* **25**, 185–209 (2015). <https://doi.org/10.1137/13094829X>
3. Beck, A., Tetrushvili, L.: On the convergence of block coordinate descent type methods. *SIAM J. Optim.* **23**, 2037–2060 (2013). <https://doi.org/10.1137/120887679>
4. Bertsekas, D.: *Nonlinear Programming*, Athena Scientific Optimization and Computation Series. Athena Scientific (1999)
5. Both, J.W., Kumar, K., Nordbotten, J.M., Radu, F.A.: The gradient flow structures of thermo-poro-visco-elastic processes in porous media (2019). ArXiv e-prints [arXiv:1907.03134](https://arxiv.org/abs/1907.03134)
6. Engwer, C., Gräser, C., Müthing, S., Sander, O.: Function space bases in the dune-functions module (2018). arXiv e-prints [arXiv:1806.09545](https://arxiv.org/abs/1806.09545)
7. Grippo, L., Sciandrone, M.: On the convergence of the block nonlinear Gauss–Seidel method under convex constraints. *Oper. Res. Lett.* **26**, 127–136 (2000)
8. Grippo, L., Sciandrone, M.: Globally convergent block-coordinate techniques for unconstrained optimization. *Optim. Method Softw.* **10**, 587–637 (1999). <https://doi.org/10.1080/10556789908805730>

9. Luo, Z.-Q., Tseng, P.: Error bounds and convergence analysis of feasible descent methods: a general approach. *Ann. Oper. Res.* **46**, 157–178 (1993). <https://doi.org/10.1007/BF02096261>
10. Mieke, C., Mauthe, S., Teichtmeister, S.: Minimization principles for the coupled problem of Darcy–Biot-type fluid transport in porous media linked to phase field modeling of fracture. *J. Mech. Phys. Solids* **82**, 186–217 (2015). <https://doi.org/10.1016/j.jmps.2015.04.006>
11. Necoara, I., Nesterov, Y., Glineur, F.: Linear convergence of first order methods for non-strongly convex optimization. *Math. Progr.* **175**, 69–107 (2019). <https://doi.org/10.1007/s10107-018-1232-1>
12. Ortega, J.M., Rheinboldt, W.C.: *Iterative Solution of Nonlinear Equations in Several Variables*, vol. 30. SIAM (1970)
13. Sander, O.: *DUNE-The Distributed and Unified Numerics Environment*, vol. 140. Springer (2020)
14. Tai, X.-C., Espedal, M.: Rate of convergence of some space decomposition methods For linear and nonlinear problems. *SIAM J. Numer. Anal.* **35**, 1558–1570 (1998)
15. Wright, S.J.: Coordinate descent algorithms. *Math. Progr.* **151**, 3–34 (2015). <https://doi.org/10.1007/s10107-015-0892-3>
16. Zhang, H., Cheng, L.: Restricted strong convexity and its applications to convergence analysis of gradient-type methods in convex optimization. *Optim. Lett.* **9**, 961–979 (2015). <https://doi.org/10.1007/s11590-014-0795-x>

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.