An error bound-based convergence analysis framework for a class of randomized algorithms

Zhichun Yang[*] Li Jiang[†] Tianxiang Liu[‡] Man-Chung Yue[§]

Abstract

Existing error-bound-based analyses for stochastic algorithms that exhibit certain descent properties, such as randomized coordinate descent and randomized projection methods, are often limited in scope and typically lead to overly conservative convergence guarantees. To address this gap, we develop an abstract framework for analyzing such stochastic algorithms based on new unified error bound (UEB) conditions. The proposed UEB conditions subsume many common error bound- and Kurdyka–Łojasiewicz-type conditions used in existing studies of algorithms for optimization, convex feasibility, and common fixed point problems. Under the global UEB condition, we establish non-asymptotic in-expectation and asymptotic almost-sure convergence rates for the stochastic algorithms in our framework. Under the local UEB condition, we also show asymptotic almost sure convergence rates. We demonstrate the strength and versatility of our framework through two applications. For the common fixed point problem, we provide comprehensive convergence guarantees for the randomized alternating Krasnoselskii-Mann method under Hölderian error bound conditions. Furthermore, for unconstrained minimization of smooth definable functions, we establish novel convergence guarantees for the randomized subspace descent method, an algorithm subsuming both randomized coordinate and block coordinate descent.

1. Introduction

In modern machine learning, randomized algorithms have emerged as the methods of choice for many real-world applications with high-dimensional decision variables, massive datasets, and/or a large number of constraints. Gaining a rigorous understanding of the convergence behavior of such algorithms is therefore a fundamental question with significant practical implications. The central goal of this paper is to systematically analyze a class of randomized algorithms satisfying certain structural properties under general notions of error bound conditions.

[*]The Hong Kong Polytechnic University (zhichun.yang@polyu.edu.hk).
[†]The Chinese University of Hong Kong, Shenzhen (lijiang@link.cuhk.edu.cn).
[‡]University of Tsukuba (liutx@sk.tsukuba.ac.jp)
[§]The Hong Kong Polytechnic University (manchung.yue@polyu.edu.hk).

To motivate our study and highlight the challenges involved, let us begin by focusing on the randomized coordinate descent (RCD), a classical and widely used randomized algorithm for solving unconstrained smooth minimization problems. Owing to its simplicity and effectiveness in tackling high-dimensional problems, RCD and its variants have been extensively studied in the literature; see, for example, [11, 33, 37, 39, 41, 50, 53, 54]. The majority of studies on RCD and its variants assume the convexity of the objective function. These works typically demonstrate that the expected objective value converges to the optimal value at a sublinear rate, which improves to a linear rate when the objective is strongly convex [5, 33, 37, 41]. In optimization, it is a folklore that the strong convexity assumption in linear convergence rate results can often be relaxed to certain error bound (EB) or Kurdyka–Łojasiewicz (KL) conditions. RCD is no exception: under either an EB or KL condition, the linear convergence rates of RCD and its variants continue to hold without assuming strong convexity [22, 39].

For non-convex minimization problems, EB- or KL-based analyses of RCD-type algorithms have also attracted increasing attention in recent years. For instance, [14, 24, 25, 54] establish convergence results based on local KL conditions under an essential cyclicity assumption, which requires that all coordinates be selected at least once within any fixed number of consecutive iterations. This assumption fails almost surely when coordinates are sampled independently according to the uniform distribution, a setting commonly adopted in practice as well as existing theoretical analyses. More recently, local KL-based analyses of RCD-type algorithms have been developed in [13,26] without relying on essential cyclicity. Nevertheless, due to the use of Egorov's Theorem [15] in their proofs, these results yield asymptotic convergence rates only up to an additive error of the order of $\sqrt{\delta}$ that hold after k_δ iterations, where $k_\delta \to \infty$ as $\delta \to 0$. Furthermore, [39] establishes asymptotic linear convergence of the expected objective value under a local Lipschitzian EB condition, but the proof requires the iterates' entry time to the EB neighborhood to be deterministic, *i.e.*, depends only on the initial point but not the random outcome, which is valid, for example, when the EB condition holds globally.

RCD is a specific instance of stochastic gradient descent (SGD). Therefore, in principle, RCD can be studied under the KL-based SGD analysis frameworks recently developed in [17,32,40]. However, since these frameworks are designed for analyzing general SGDs and do not exploit the special structures of RCD, applying them to RCD leads to conservative conclusions. Specifically, the analyses in [32] and [40] are premised on the bounded variance assumption, which necessitates polynomially decaying stepsizes. In contrast, RCD satisfies a stronger variance assumption called the strong growth condition [42]. Although this condition has been incorporated into the KL-based SGD analysis in [17], the authors established only the convergence of objective values and iterates without providing explicit convergence rates.

In view of the above discussions, one of the primary goals of this study is to develop an EB- or KL-based analysis of RCD for non-convex problems that can yield error-free convergence rates without additionally imposing strong assumptions. At a deeper level, the convergence behaviour of RCD is tied to a small set of structural properties, namely, a sufficient decrease with respect to an optimality measure and certain inequalities relating the distance between consecutive iterates to a residual function. These features are not unique to RCD but are shared by various randomized algorithms. Notable examples include the random projection method (RPM) [36] for convex feasibility problems and the random Krasnoselskii–Mann method (RKMM) for common fixed point problems. With this observation, we endeavor to analyze a family of randomized algorithms in a unified manner under general EB or KL conditions.

To develop such an analysis, a major challenge lies in formulating suitable unified EB or KL conditions. On one hand, these conditions should be sufficiently general to accommodate different classes of computational problems (*e.g.*, optimization and fixed point problems) and flexible enough to delineate the convergence rates of different instances within the same class of problems. On the other hand, they should remain concrete and amenable to theoretical analysis. To this end, we introduce two new notions: a unified global error bound and a unified local error bound. These conditions are inspired by the work of Liu and Lourenço [28, 29], who proposed consistent error bounds for analyzing deterministic Fejér monotone algorithms for convex feasibility problems and deterministic quasi-cyclic algorithms for common fixed point problems. Therefore, our work can also be viewed as an extension of this line of research to a family of randomized algorithms under a much broader error bound framework. We should emphasize that this extension is highly non-trivial and requires substantial new ideas.

We now summarize our main contributions.

- We formulate, in Assumption 1, a class of randomized algorithms that capture the key structural properties of several important methods, including RCD and its variants in optimization, RPM for convex feasibility problems, and RKMM for common fixed point problems. We introduce the unified global and local EB conditions for analyzing these algorithms. The proposed error bounds subsume a wide range of EB- or KL-type conditions studied in the literature, including classical EB and KL conditions in unconstrained optimization and consistent error bounds for fixed point problems [28, 29].

- We conduct a comprehensive analysis of the convergence behavior of algorithms prescribed in Assumption 1. Specifically, in both the global and local EB settings, we establish a spectrum of non-asymptotic in-expectation and asymptotic almost-sure convergence rates for the iterates and a certain optimality measure specified in Assumption 1. Instrumental to our analysis are several general probability-theoretic results, proved in Propositions 1-4, which could be of independent interest.

- To demonstrate the versatility and strength of our framework, we apply our general results to RKMM for common fixed point problems and the random subspace descent

(RSD) for unconstrained smooth minimization problems. Although these results are obtained as special cases of our abstract analysis, the resulting convergence guarantees for RKMM and RSD are novel and have not been reported in the existing literature.

- As a side result, Theorem 6 establishes that for any smooth definable function f with a locally Lipschitz gradient, the desingularization function φ at a non-maximizer stationary point must satisfy a non-integrability condition on $(\varphi')^2$ near the origin. An immediate consequence is that f cannot satisfy the KL property with an exponent $\kappa \in [0, 1/2)$ at any such point. This substantially generalizes existing results (*e.g.*, [6, Theorem 7]), which are restricted to local minimizers and power-type desingularization functions.

1.1. Notation

The sets of real and non-negative real numbers are denoted by $\mathbb{R}$ and $\mathbb{R}_+$, respectively. The Euclidean norm is denoted by $\|\cdot\|$. For any integer $n > 0$, we denote $[n] = \{1, \ldots, n\}$. For any $\bar{x} \in \mathbb{R}^d$ and $\delta > 0$, $\mathbb{B}(\bar{x}, \delta)$ denotes the open ball centered at $\bar{x}$ with radius δ. For any set B, we denote by $\operatorname{dist}(x, B)$ the Euclidean distance from x to the set B and by $\mathbf{1}_B$ its indicator function, *i.e.*, $\mathbf{1}_B(x) = 1$ if $x \in B$; and $\mathbf{1}_B(x) = 0$ otherwise. We also abbreviate "almost surely" as "a.s.".

2. Problem Setup and Preliminaries

2.1. A Class of Structured Monotone Randomized Algorithms

Let $\{x^k\}_{k\geq 0} \subseteq \mathbb{R}^d$ be a sequence of random vectors defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with a deterministic initial value x^0. The random sequence models the iterates generated by a randomized algorithm. Associated with the sequence is a filtration $\mathcal{F}_k = \sigma(\{x^i\}_{0\leq i\leq k})$ with the boundary cases given by $\mathcal{F}_0 = \{\emptyset, \Omega\}$ and $\mathcal{F}_\infty = \sigma\left(\bigcup_{k\geq 0} \mathcal{F}_k\right)$. For any random variable X defined on the same probability space, we denote by $\mathbb{P}(X \mid \mathcal{F}_k)$ the regular version of the conditional probability distribution, *i.e.*, $\mathbb{P}(X \mid \mathcal{F}_k)$ satisfies the definition of a conditional distribution, and for each $\omega \in \Omega$, the evaluation $\mathbb{P}(X \mid \mathcal{F}_k)(\omega)$ defines a probability distribution. The existence of a regular conditional distribution is guaranteed, *e.g.*, by [44, Page 269, Theorem 4].

This paper focuses on the following class of randomized algorithms.

Assumption 1 (Monotone randomized algorithms)**.** There exist a closed set $S \subseteq \mathbb{R}^d$, continuous functions $g, h : S \to \mathbb{R}$ with $g \geq 0$ and $h^* := \inf_{x\in S} h(x) > -\infty$, constants $c_1, c_2, c_3 > 0$ with $c_3 > c_2$, and a positive sequence $\{\alpha_k\}_{k\geq 0}$ such that the following hold.

(i) For all $k \geq 0$,

$$h(x^{k+1}) \leq h(x^k) - \frac{c_1}{\alpha_k}\|x^{k+1} - x^k\|^2, \tag{1}$$

$$\mathbb{E}[\|x^{k+1} - x^k\|^2 | \mathcal{F}_k] \geq c_2\, \alpha_k^2\, g(x^k), \quad \text{a.s.}, \tag{2}$$

$$\|x^{k+1} - x^k\|^2 \leq c_3\, \alpha_k^2\, g(x^k). \tag{3}$$

(ii) $\{x^k\}_{k\geq 0} \subseteq S$.

In Assumption 1, the sequence $\{\alpha_k\}_{k\geq 0}$ plays the role of stepsizes of the randomized algorithm, while the functions g and h represent two metrics for quantifying the progress of the algorithm. Our theoretical development will also rely on certain error bound conditions relating the metrics g and h. The presence of the set S is to allow the additional flexibility for accommodating the situation where the error bound condition holds only on a subset in $\mathbb{R}^d$. Below we illustrate our algorithmic framework with two representative examples.

Example 1. Consider the common fixed point problem

$$\text{find } x \in F = \bigcap_{i=1}^{m} \operatorname{Fix} T_i, \tag{4}$$

where each T_i is a firmly quasi-nonexpansive operator on $\mathbb{R}^d$, and $\operatorname{Fix} T_i = \{x \in \mathbb{R}^d : T_i(x) = x\}$ is the set of its fixed points. A natural algorithm for solving problem (4) is the random Krasnoselskii–Mann method (RKMM):

$$x^{k+1} = x^k + \beta_k (T_{i_k}(x^k) - x^k) \quad \forall k \geq 0, \tag{5}$$

where $\{\beta_k\}_{k\geq 0} \subseteq (0, 1]$ are the stepsizes and $\{i_k\}_{k\geq 0}$ are independent and identically distributed (i.i.d.) random indices sampled according to the uniform distribution on $[m]$. Then, as we will prove in Section 5.1, RKMM satisfies Assumption 1 with $h(x) = \operatorname{dist}^2(x, F)$ and $g(x) = \max_{i\in[m]} \|x - T_i(x)\|^2$.

Example 2. Consider the unconstrained minimization problem

$$\min_{x\in\mathbb{R}^d} \; f(x), \tag{6}$$

where the objective function f attains its minimum $f^\star$ and has a Lipschitz continuous gradient. A popular randomized algorithm for solving problem (6) is the randomized coordinate descent (RCD):

$$x^{k+1} = x^k - \alpha \frac{\partial f}{\partial x_{i_k}}(x^k)\, e_{i_k} \quad \forall k \geq 0, \tag{7}$$

where $\alpha > 0$ is the stepsize, $\{i_k\}_{k\geq 0}$ are independent and identically distributed (i.i.d.) random indices sampled according to the uniform distribution on $[d]$, and e_i denotes the i-th standard basis in $\mathbb{R}^d$. Then, as we will see in Section 5.2, RCD satisfies Assumption 1 with $h(x) = f(x)$ and $g(x) = \|\nabla f(x)\|^2$.

We list below a set of conditions on $\{\alpha_k\}_{k\geq 0}$ defined in Assumption 1.

Assumption 2. The positive sequence $\{\alpha_k\}_{k\geq 0}$ satisfies:

(i) There exists a constant $\bar{\alpha} > 0$ such that $\alpha_k \leq \bar{\alpha}$ for all $k \geq 0$, and $\sum_{k=0}^{\infty} \alpha_k = +\infty$.

(ii) It holds that
$$\sum_{k=0}^{\infty} \left(\frac{\alpha_k}{\sum_{j=0}^{k} \alpha_j} \right)^2 < \infty.$$

Assumption 2(i) serves as a standing hypothesis throughout the paper. By contrast, Assumption 2(ii) is only required for establishing almost sure convergence results.

2.2. Auxiliaries

We next collect some preparatory results for our theoretical development.

Definition 1 (Inverse smoothing and desingularization functions)**.** Given any $\tau > 0$ and function $\phi : [0, \tau) \to \mathbb{R}_+$ that is non-decreasing and continuous on $[0, \tau)$, positive on $(0, \tau)$, and satisfies $\phi(0) = 0$, we define the following functions.

(i) Fix a $t_0 \in (0, \tau)$,[1] define $\Psi : [0, \tau) \to [0, +\infty]$ such that $\Psi(t) = \int_t^{t_0} \frac{1}{\phi(s)} ds$.

(ii) If $1/\sqrt{\phi}$ is integrable near 0,[2] define $\Phi : [0, \tau) \to \mathbb{R}_+$ such that $\Phi(t) = \int_0^t \frac{1}{\sqrt{\phi(s)}} ds$.

The function Ψ is inspired by the inverse smoothing function [29, Equation (4.3)], while the function Φ serves a similar purpose to the desingularization function in the Kurdyka–Łojasiewicz (KL) condition.

To reflect the relevant context, the notation (ϕ, Ψ, Φ) will be augmented with subscripts in the following sections. Specifically, it appears as (ϕ_S, Ψ_S, Φ_S) in Section 3, and as $(\phi_{\mathcal{A}(\omega)}, \Psi_{\mathcal{A}(\omega)}, \Phi_{\mathcal{A}(\omega)})$ and $(\phi_{x^*(\omega)}, \Psi_{x^*(\omega)}, \Phi_{x^*(\omega)})$ in Section 4.

The lemma below summarizes the properties of Ψ and Φ, which is partly established in [29, Proposition 4.3]. We provide a proof for self-containedness.

Lemma 1. For the function ϕ defined in Definition 1, the following properties hold:

(i) The function Ψ is convex and decreasing on $[0, \tau)$, and real-valued and differentiable with $\Psi'(t) = -1/\phi(t)$ on $(0, \tau)$. If $1/\phi$ is integrable near 0, then $\Psi(t) \uparrow \Psi(0) < +\infty$ as $t \downarrow 0$; otherwise, $\Psi(t) \uparrow \Psi(0) = +\infty$ as $t \downarrow 0$.

(ii) If $1/\sqrt{\phi}$ is integrable near 0, then Φ is continuous, concave, and increasing on $[0, \tau)$, with $\Phi(0) = 0$, and differentiable with $\Phi'(t) = 1/\sqrt{\phi(t)}$ on $(0, \tau)$.

[1] The results in this paper are independent of the choice of t_0.

[2] Given a function $f : [0, \tau) \to [0, +\infty]$ with $\tau > 0$, we say f is integrable near 0 if there exists a number $\tau' \in (0, \tau)$ such that f is integrable on $[0, \tau']$.

Proof. We prove only assertion (i) as assertion (ii) can be proved similarly. By definition, Ψ is real-valued and satisfies $\Psi' = -1/\phi$ on the open interval $(0, \tau)$. Since the function ϕ is positive and non-decreasing, it follows that Ψ' is negative and non-decreasing. Consequently, Ψ is decreasing and convex. By the monotone convergence theorem, we have $\Psi(t) \uparrow \Psi(0)$ as $t \downarrow 0$. The proof is completed by noting that $\Psi(0) < +\infty$ if and only if $1/\phi$ is integrable near zero. □

Lemma 1 implies that when $1/\phi$ is not integrable near zero, the range of Ψ, and thus the domain of Ψ^{-1}, is exactly $(\lim_{t\uparrow\tau} \Psi(t), +\infty)$. Therefore, for any $t \in (0, \tau)$, the interval $[\Psi(t), +\infty)$ lies entirely within the domain of Ψ^{-1}.

We will also use the following elementary result.

Lemma 2. For the function ϕ defined in Definition 1, the following properties hold:

(i) For any $a, \tilde{a} \in (0, \tau)$ and $b \geq 0$, if $\tilde{a} \leq a - b\phi(a)$, then $\Psi(\tilde{a}) \geq \Psi(a) + b$. If $1/\phi$ is integrable near 0, then this conclusion also holds for $\tilde{a} = 0$, $a \in (0, \tau)$.

(ii) Suppose further that $1/\sqrt{\phi}$ is integrable near 0. For any $a \in (0, \tau)$, $\tilde{a} \geq 0$ and $\Delta \geq 0$, if $\tilde{a} \leq a - \Delta\sqrt{\phi(a)}$, then $\Delta \leq \Phi(a) - \Phi(\tilde{a})$.

Proof. The recursion in (i) can be reformulated as

$$-\frac{1}{\phi(a)}(\tilde{a} - a) \geq b.$$

By Lemma 1, Ψ is convex and $\Psi' = -1/\phi$. We thus have $\Psi(\tilde{a}) - \Psi(a) \geq [-1/\phi(a)](\tilde{a} - a) \geq b$. For the case where $1/\phi$ is integrable near 0 and $\tilde{a} = 0$, we use the continuity of Ψ at 0 established in Lemma 1. Taking the limit as $\tilde{a} \downarrow 0$ in the preceding inequality yields the result. As for the recursion in (ii), again by Lemma 1, Φ is concave and $\Phi' = 1/\sqrt{\phi}$. Hence,

$$\Delta \leq \frac{a - \tilde{a}}{\sqrt{\phi(a)}} = \Phi'(a)(a - \tilde{a}) \leq \Phi(a) - \Phi(\tilde{a}),$$

which completes the proof. □

The following lemma quantifies the random descent of h.

Lemma 3. Suppose that Assumption 1 holds. Let $p = \frac{c_2}{2c_3 - c_2}$. Then, for any $k \geq 0$, there exists a $\mathcal{F}_{k+1}$-measurable binary random variable z_k such that

$$h(x^{k+1}) \leq h(x^k) - \frac{c_1 c_2}{2} \alpha_k \, z_k \, g(x^k), \tag{8}$$

where $\mathbb{P}(z_k = 1 | \mathcal{F}_k) \geq p$ almost surely.

Proof. To begin, we define the random variable

$$Z_k = \begin{cases} c_3, & \text{if } g(x^k) = 0, \\ \frac{\|x^{k+1}-x^k\|^2}{\alpha_k^2 g(x^k)}, & \text{otherwise.} \end{cases}$$

By the definition of Z_k, the identity $\|x^{k+1} - x^k\|^2 = Z_k \, \alpha_k^2 \, g(x^k)$ always holds. Also, it follows from (3) that $Z_k \le c_3$. These two observations together yield

$$\begin{aligned} \mathbb{E}[\|x^{k+1} - x^k\|^2 | \mathcal{F}_k] &= \mathbb{E}[Z_k|\mathcal{F}_k] \, \alpha_k^2 \, g(x^k) \\ &\le \left[\mathbb{P}(Z_k \ge \tfrac{c_2}{2}|\mathcal{F}_k) \, c_3 + \mathbb{P}(Z_k < \tfrac{c_2}{2}|\mathcal{F}_k) \, \frac{c_2}{2}\right] \alpha_k^2 \, g(x^k), \quad \text{a.s..} \end{aligned} \tag{9}$$

Next, since

$$\int_{g(x^k)=0} \mathbb{P}(Z_k \ge \tfrac{c_2}{2}|\mathcal{F}_k) = \int_{g(x^k)=0} \mathbb{E}[\mathbf{1}_{\{Z_k \ge \frac{c_2}{2}\}}|\mathcal{F}_k] = \mathbb{P}(g(x^k) = 0, Z_k \ge \tfrac{c_2}{2}) = \mathbb{P}(g(x^k) = 0),$$

we have that $\mathbb{P}(Z_k \ge \frac{c_2}{2}|\mathcal{F}_k)(\omega) = 1$ almost surely on the set $\{g(x^k) = 0\}$. On the complement $\{g(x^k) \ne 0\}$, by (2) and (9), it holds that

$$\mathbb{P}(Z_k \ge \tfrac{c_2}{2}|\mathcal{F}_k) \, c_3 + \mathbb{P}(Z_k < \tfrac{c_2}{2}|\mathcal{F}_k) \, \frac{c_2}{2} \ge c_2, \quad \text{a.s.,}$$

which implies that $\mathbb{P}(Z_k \ge \frac{c_2}{2}|\mathcal{F}_k) \ge \frac{c_2}{2c_3 - c_2} = p$ almost surely. Combining the two cases, we have

$$\mathbb{P}(Z_k \ge \tfrac{c_2}{2}|\mathcal{F}_k) \ge p = \frac{c_2}{2c_3 - c_2}, \quad \text{a.s..}$$

Define the random variable

$$z_k = \begin{cases} 1, & \text{if } Z_k \ge \frac{c_2}{2}, \\ 0, & \text{otherwise.} \end{cases}$$

Then, $\mathbb{P}(z_k = 1|\mathcal{F}_k) = \mathbb{P}(Z_k \ge \frac{c_2}{2}|\mathcal{F}_k) \ge p$. Moreover, since $Z_k \ge \frac{c_2}{2} z_k$, we obtain

$$\|x^{k+1} - x^k\|^2 \ge \frac{c_2}{2} z_k \, \alpha_k^2 \, g(x^k),$$

which yields the desired inequality when combined with Assumption 1(i). Finally, the $\mathcal{F}_{k+1}$-measurability of z_k follows from that of the Z_k. □

3. Convergence Analysis under Global Error Bounds

The main purpose of this paper is to study the convergence behaviour of the algorithms defined in Assumption 1 under general error bound conditions relating g and h. We first focus on the situation where the error bound condition holds globally over the domain S.

Assumption 3 (Global error bound)**.** There exists a function $\phi_S : [0, \tau) \to \mathbb{R}_+$, where $\tau > 0$ is some constant, such that the following hold.

(i) For any $x \in S$,

$$\phi_S(h(x) - h^*) \le g(x). \tag{10}$$

(ii) The function ϕ_S is continuous and non-decreasing on $[0, \tau)$, strictly positive on $(0, \tau)$, and satisfies $\phi_S(0) = 0$.

(iii) $0 \le h(x) - h^* < \tau$ for all $x \in S$.

Assumption 3 encompasses many existing EB- or KL-type conditions across various contexts. For instance, consider the common fixed point problem (4) in Example 1 and recall our choice of $h(x) = \mathrm{dist}^2(x, F)$ and $g(x) = \max_{i\in[m]} \|x - T_i(x)\|^2$. Under some mild growth conditions on ϕ_S, Assumption 3 reduces to the joint Karamata regularity [28] of the operators $\{T_i\}_{i=1}^m$. In the special case of convex feasibility problems, where each T_i is the projection operator of a convex set and the aim is to find a point in the intersection of all the convex sets, with ϕ_S being a linear map, Assumption 3 further reduces to the bounded linear regularity [3, definition 5.6]. Revisit the smooth unconstrained minimization problem (6) in Example 2 and recall the choice of $h(x) = f(x)$ and $g(x) = \|\nabla f(x)\|^2$. In this case, Assumption 3 is equivalent to a global KL condition [16,18] with the desingularization function given by $\Phi(t) = \int_0^t \frac{1}{\sqrt{\phi_S(s)}} ds$. When ϕ_S is linear, then it further reduces to the Polyak-Łojasiewicz condition [30, 31, 38].

3.1. The case when ϕ_S is convex

A standard technique for analyzing stochastic algorithms under a global error bound condition is as follows. First, conditional expectations are taken on both sides of inequality (1). Then, together with Assumption 1(i) and Assumption 3(i), one obtains

$$\mathbb{E}[h(x^{k+1}) - h^* | \mathcal{F}_k] \le h(x^k) - h^* - c_1 c_2 \alpha_k \phi_S(h(x^k) - h^*). \tag{11}$$

In most existing studies adopting this technique [5, 33, 35–37, 39, 41, 45], the function ϕ_S is either linear, which corresponds to bounded linear regularity, strong convexity, or the Polyak-Łojasiewicz (PL) condition; or quadratic, which corresponds to the case of (non-strongly) convex objectives. Below we extend the technique to convex ϕ_S.

Theorem 1. Suppose that Assumptions 1, 2(i), and 3 hold and that ϕ_S is convex.

(i) If $1/\phi_S$ is integrable near 0, then there exists an index K_0 such that $x^k = x^{K_0} \in \arg\min_{x\in S} h(x)$ for all $k \ge K_0$ almost surely.

(ii) If $1/\phi_S$ is not integrable near 0, then for any $k \ge 1$,

$$\mathbb{E}[h(x^k) - h^*] \le \Psi_S^{-1}\left(\Psi_S(h(x^0) - h^*) + c_1 c_2 \sum_{i=0}^{k-1} \alpha_i\right). \tag{12}$$

Proof. Taking the expectation on both sides of (11), we obtain

$$\mathbb{E}[h(x^{k+1}) - h^*] \leq \mathbb{E}[h(x^k) - h^*] - c_1 c_2 \alpha_k \mathbb{E}[\phi_S(h(x^k) - h^*)]. \tag{13}$$

Since ϕ_S is convex, Jensen's inequality yields $\mathbb{E}[\phi_S(h(x^k) - h^*)] \geq \phi_S(\mathbb{E}[h(x^k) - h^*])$. It follows that

$$\mathbb{E}[h(x^{k+1}) - h^*] \leq \mathbb{E}[h(x^k) - h^*] - c_1 c_2 \alpha_k \phi_S(\mathbb{E}[h(x^k) - h^*]).$$

For any $k \geq 0$ where $\mathbb{E}[h(x^{k+1}) - h^*] > 0$, Assumption 1(i) guarantees that $\mathbb{E}[h(x^i) - h^*] > 0$ for all $i \leq k$. Consequently, Lemma 2 gives us that for all $i \leq k$,

$$\Psi_S(\mathbb{E}[h(x^{i+1}) - h^*]) \geq \Psi_S(\mathbb{E}[h(x^i) - h^*]) + c_1 c_2 \alpha_i.$$

Summing this inequality from $i = 0$ to $k - 1$ yields

$$\Psi_S(\mathbb{E}[h(x^k) - h^*]) \geq \Psi_S(\mathbb{E}[h(x^0) - h^*]) + c_1 c_2 \sum_{i=0}^{k-1} \alpha_i. \tag{14}$$

If $1/\phi_S$ is integrable near 0, Lemma 1(i) ensures that $\Psi_S(t) \leq \Psi_S(0) < +\infty$ for all $t \in [0, \tau)$. Suppose, for the sake of contradiction, that $\mathbb{E}[h(x^k) - h^*] > 0$ for all $k \geq 0$. Consequently, inequality (14) holds for all $k \geq 0$. Taking the limit as $k \to \infty$, the condition $\sum_{i=0}^{\infty} \alpha_i = +\infty$ from Assumption 2(i) forces the right-hand side of (14) to diverge to $+\infty$. This contradicts the fact that the left-hand side, $\Psi_S(\mathbb{E}[h(x^k) - h^*])$, is bounded above by $\Psi_S(0) < +\infty$. Thus, there must exist a finite index K_0 such that $\mathbb{E}[h(x^{K_0}) - h^*] = 0$. By Assumption 1(i), it then follows that $h(x^k) = h^*$ and $x^k = x^{K_0}$ almost surely for all $k \geq K_0$.

Next, consider the case where $1/\phi_S$ is not integrable near 0. For any $k \geq 0$ such that $\mathbb{E}[h(x^k) - h^*] > 0$, applying the strictly decreasing inverse function Ψ_S^{-1} to (14) yields (12). Furthermore, if $\mathbb{E}[h(x^k) - h^*] = 0$, (12) holds trivially. This completes the proof. □

3.2. The case with a general ϕ_S

Theorem 1 is unsatisfactory for the following reasons. First, Theorem 1 does not address the convergence rate of the iterates x^k, but only that of the optimality metric $h(x^k)$. Second, only the in-expectation convergence rate of the optimality metric $h(x^k)$ is established. It is unclear whether almost-sure convergence rates are possible. Third, the analysis in Theorem 1 requires ϕ_S to be convex. However, the example below shows that allowing non-convex ϕ_S could potentially be useful in some applications.

Example 3. Consider the univariate function $f(x) = \exp(-|x|^{-1}) \mathbf{1}_{\{x \neq 0\}}(x)$ defined on the set $S = [-\frac{1}{2}, \frac{1}{2}]$. It is well known that f is C^∞, and

$$|f'(x)| = \exp(-|x|^{-1})\, |x|^{-2} \cdot \mathbf{1}_{\{x \neq 0\}}(x) = f(x)(\log(f(x)))^2.$$

Taking $g(x) = |f'(x)|^2$ and $h(x) = f(x)$, then

$$g(x) = h(x)^2(\log(h(x)))^4.$$

In respect of Assumption 3, we take $\phi_S(t) = t^2(\log t)^4$ with domain $[0, e^{-2}]$, which is non-decreasing on $[0, e^{-2}]$; however, it is convex only on the subinterval $[0, e^{-3-\sqrt{3}}]$ and concave on $[e^{-3-\sqrt{3}}, e^{-2}]$.

In view of the limitations of Theorem 1, we develop a new convergence rate analysis for the randomized algorithms in Assumption 1 under the global EB condition in Assumption 3 that can yield in-expectation and almost-sure convergence rates for both the iterates x^k and the optimality metric $h(x^k)$ without requiring the convexity of ϕ_B. We first present the case where $1/\phi_S$ is integrable near 0, which is analogous to the part (i) in Theorem 1.

Theorem 2. Suppose that Assumptions 1, 2(i) and 3 hold and that $1/\phi_S$ is integrable near 0. Then, there exist an almost surely finite random index K_1 and a random variable x^* such that almost surely, $x^* \in \arg\min_{x\in S} h(x)$ and $x^k = x^*$ for all $k \geq K_1$.

Proof. By Lemma 3 and Assumption 3, we have

$$h(x^{k+1}) \leq h(x^k) - \frac{c_1 c_2}{2}\alpha_k\, z_k\, \phi_S(h(x^k) - h^*). \tag{15}$$

Applying Lemma 2(i) to (15), we obtain

$$\Psi_S(h(x^{k+1}) - h^*) \geq \Psi_S(h(x^k) - h^*) + \frac{c_1 c_2}{2}\alpha_k\, z_k\, \mathbf{1}_{\{h(x^k) > h^*\}}. \tag{16}$$

Taking conditional expectations on both sides of (16) yields

$$\mathbb{E}[\Psi_S(h(x^{k+1}) - h^*) \mid \mathcal{F}_k] \geq \Psi_S(h(x^k) - h^*) + \frac{c_1 c_2}{2}\alpha_k\, \mathbb{E}[z_k \mid \mathcal{F}_k]\mathbf{1}_{\{h(x^k) > h^*\}}.$$

Since $\mathbb{E}[z_k \mid \mathcal{F}_k] \geq p$ almost surely, it follows that

$$\mathbb{E}[\Psi_S(h(x^{k+1}) - h^*)] \geq \mathbb{E}[\Psi_S(h(x^k) - h^*)] + \frac{c_1 c_2 p}{2}\alpha_k\, \mathbb{E}[\mathbf{1}_{\{h(x^k) > h^*\}}].$$

Since $\mathbb{E}[\Psi_S(h(x^k) - h^*)]$ is non-decreasing with respect to k and $\Psi_S(t) \leq \Psi_S(0) < +\infty$ for all $t \in [0, \tau)$, it follows that

$$\mathbb{E}\left[\sum_{k=0}^{\infty} \alpha_k \mathbf{1}_{\{h(x^k) > h^*\}}\right] = \sum_{k=0}^{\infty} \alpha_k\, \mathbb{E}\left[\mathbf{1}_{\{h(x^k) > h^*\}}\right] \leq \frac{2}{c_1 c_2 p}\Psi_S(0) < +\infty,$$

which implies that $\sum_{k=0}^{\infty} \alpha_k \mathbf{1}_{\{h(x^k) > h^*\}} < \infty$ almost surely. Combined with $\sum_k \alpha_k = +\infty$ and the non-increasing property of $h(x^k)$, we conclude that $h(x^k)$ reaches h^* in a finite number of iterations almost surely. According to Assumption 1(i), once $h(x^k)$ attains h^*, the algorithm terminates in finitely many iterations[3] starting from iteration k. This completes the proof. □

[3] Given any $\omega \in \Omega$, we say that the sequence $\{x^k(\omega)\}_k$ terminates in finitely many iterations if there exists some $K(\omega) \in \mathbb{N}$ such that $x^k(\omega) = x^K(\omega)$ for all $k \geq K(\omega)$.

We now consider the case where $1/\phi_S$ is not integrable near 0. This case is analogous to part (ii) in Theorem 1 and considerably more challenging.

Theorem 3. Suppose that Assumptions 1 and 3 hold and $1/\phi_S$ is not integrable near 0.

(i) If Assumption 2(i) holds, we have that for all $k \geq 1$,

$$\mathbb{E}[h(x^k)-h^*] \leq \exp\left(-\frac{p^2}{2\bar{\alpha}}\sum_{i=0}^{k-1}\alpha_i\right)(h(x^0)-h^*)+\Psi_S^{-1}\left(\Psi_S(h(x^0)-h^*)+\frac{c_1c_2p}{4}\sum_{i=0}^{k-1}\alpha_i\right). \tag{17}$$

Suppose additionally that $1/\sqrt{\phi_S}$ is integrable near 0. Then, there exists a random variable x^* such that x^k converges to x^* both almost surely and in L^1, that $x^* \in \arg\min_{x\in S} h(x)$ almost surely, and that

$$\mathbb{E}[\|x^k - x^*\|] \leq \frac{\sqrt{c_3}}{c_1c_2}\Phi_S(\mathbb{E}[h(x^k)-h^*]) \quad \forall k \geq 0. \tag{18}$$

(ii) If Assumption 2 holds, then there exists an almost surely finite random index K_2 such that

$$h(x^k) - h^* \leq \Psi_S^{-1}\left(\frac{c_1c_2p}{3}\sum_{i=0}^{k-1}\alpha_i\right) \quad \forall k \geq K_2 \text{ a.s..} \tag{19}$$

Suppose additionally that $1/\sqrt{\phi_S}$ is integrable near 0. Then, for the limit x^* established in (i), there exists an almost surely finite random index K_3 such that

$$\|x^k - x^*\| \leq \frac{4}{c_1\sqrt{c_2p}}\Phi_S\left(\Psi_S^{-1}\left(\frac{c_1c_2p}{4}\sum_{i=0}^{k-1}\alpha_i\right)\right) \quad \forall k \geq K_3 \text{ a.s..} \tag{20}$$

The proof of Theorem 3 is based on four novel analytical tools, Propositions 1-4. The first two facilitates the establishment of in-expectation and almost-sure convergence rates of $h(x^k)$; while the third and fourth transfer the convergence rates of $h(x^k)$ to x^k. As a preparation, we present the following lemma.

Lemma 4. Suppose that Assumptions 1 and 3 hold and $1/\phi_S$ is not integrable near 0. Then,

$$h(x^k) - h^* \leq \Psi_S^{-1}\left(\Psi_S(h(x^0)-h^*) + \frac{c_1c_2}{2}\sum_{i=0}^{k-1}\alpha_i z_i\right) \quad \forall k \geq 1. \tag{21}$$

Proof. Given any $\omega \in \Omega$ and integer $k > 0$, if $h(x^{k-1}(\omega)) = h^*$, then (21) holds trivially. Otherwise, if $h(x^{k-1}(\omega)) > h^*$, it follows that $h(x^i(\omega)) > h^*$ for all $i \leq k-1$. By inequality (16), we have that for all $0 \leq i \leq k-1$,

$$\Psi_S(h(x^{i+1}(\omega)) - h^*) \geq \Psi_S(h(x^i(\omega)) - h^*) + \frac{c_1c_2}{2}\alpha_i z_i(\omega).$$

Summing both sides of this inequality from $i = 0$ to $k - 1$ yields

$$\Psi_S(h(x^k(\omega)) - h^*) \geq \Psi_S(h(x^0(\omega)) - h^*) + \frac{c_1 c_2}{2} \sum_{i=0}^{k-1} \alpha_i z_i(\omega),$$

which implies (21). □

3.2.1. An Apparatus for In-Expectation Convergence Rate Analysis

To estimate $\mathbb{E}[h(x^k) - h^*]$, by Lemma 4, it suffices to estimate

$$\mathbb{E}\left[\Psi_S^{-1}\left(\Psi_S(h(x^0) - h^*) + \frac{c_1 c_2}{2} \sum_{i=0}^{k-1} \alpha_i z_i\right)\right].$$

This task is abstracted as follows.

Question 1. Consider a $\{0, 1\}$-valued random sequence $\{X_k\}_{k \geq 0}$ such that X_k is $\mathcal{F}_{k+1}$-measurable and $\mathbb{P}(X_k = 1|\mathcal{F}_k) \geq p$ for all $k \geq 0$, where $p \in (0, 1]$. Let $\{\alpha_k\}_{k \geq 0}$ be a sequence of positive coefficients. Define the partial sum $S_k = \sum_{i=0}^{k-1} \alpha_i X_i$, and let $\chi : \mathbb{R}_+ \to \mathbb{R}$ be a given non-increasing function. Can we upper bound $\mathbb{E}[\chi(S_k)]$ for a given k?

Intuitively, since the conditional success probability of X_k is uniformly bounded from below by p, we should be able to construct an independent sequence $\{Y_k\}_{k \geq 0}$ that serves as a lower bound for $\{X_k\}_{k \geq 0}$. Specifically, we seek a sequence $\{Y_k\}_{k \geq 0}$ such that

$$Y_k \sim \text{Bernoulli}(p) \quad \text{i.i.d.} \quad \text{and} \quad X_k \geq Y_k \quad \text{a.s. for all } k \geq 0.$$

Define $T_k = \sum_{i=0}^{k-1} \alpha_i Y_i$. Since $\alpha_i > 0$ and $X_k \geq Y_k$ almost surely, it follows that $S_k \geq T_k$ almost surely. Furthermore, because χ is non-increasing, we have $\chi(S_k) \leq \chi(T_k)$ almost surely, which implies $\mathbb{E}[\chi(S_k)] \leq \mathbb{E}[\chi(T_k)]$. Since $\{Y_k\}$ is an i.i.d. Bernoulli sequence, estimating $\mathbb{E}[\chi(T_k)]$ is much more tractable than estimating $\mathbb{E}[\chi(S_k)]$. The existence of such a coupling is guaranteed by stochastic dominance results established in [27].

Definition 2. Given two Borel probability measures μ and ν on a partially ordered space Ω, we say that μ stochastically dominates ν—denoted as $\mu \succeq \nu$—if for every continuous, non-decreasing function $f : \Omega \to \mathbb{R}$,

$$\int f d\mu \geq \int f d\nu.$$

Lemma 5. Consider the sequence $\{X_k\}_{k \geq 0}$ defined in Question 1. There exists a sequence of independent Bernoulli random variables $\{Y_k\}_{k \geq 0}$ with success probability p such that $X_k \geq Y_k$ almost surely for all $k \geq 0$.

Proof. Let μ denote the joint distribution of the sequence $\{X_k\}_{k\geq 0}$ and let π_p denote the product measure corresponding to an independent sequence of Bernoulli(p) trials. The condition $\mathbb{P}(X_k = 1 \mid \mathcal{F}_k) \geq p$ almost surely implies, by [27, Lemma 1.1], that μ stochastically dominates π_p (*i.e.*, $\mu \succeq \pi_p$). Consequently, [27, Lemma 1.0] guarantees the existence of a sequence $\{Y_k\}_{k\geq 0}$ defined on the same probability space as $\{X_k\}_{k\geq 0}$ such that $\{Y_k\}_{k\geq 0}$ has law π_p and $X_k \geq Y_k$ almost surely. □

Proposition 1. Consider Question 1. Suppose that the sequence $\{\alpha_k\}_{k\geq 0}$ is upper bounded by some constant $\bar{\alpha} > 0$. Then,

$$\mathbb{E}[\chi(S_k)] \leq \chi(0) \exp\Big(-\frac{p^2}{2\bar{\alpha}} \sum_{i=0}^{k-1} \alpha_i\Big) + \chi\Big(\frac{p}{2} \sum_{i=0}^{k-1} \alpha_i\Big) \quad \forall k \geq 1. \tag{22}$$

Proof. By Lemma 5, we have $\mathbb{E}[\chi(S_k)] \leq \mathbb{E}[\chi(T_k)]$. Thus, it suffices to estimate the upper bound of $\mathbb{E}[\chi(T_k)]$. Let $t_k = \mathbb{E}[T_k] = p\sum_{i=0}^{k-1} \alpha_i$ be the mean. Partitioning the sample space, we obtain

$$\mathbb{E}[\chi(T_k)] \leq \chi(0)\, \mathbb{P}(T_k \leq \frac{t_k}{2}) + \chi\Big(\frac{t_k}{2}\Big)\, \mathbb{P}(T_k > \frac{t_k}{2}).$$

Since $\mathbb{P}(T_k > \frac{t_k}{2}) \leq 1$, we can simplify this to

$$\mathbb{E}[\chi(T_k)] \leq \chi(0)\, \mathbb{P}(T_k \leq \frac{t_k}{2}) + \chi\Big(\frac{t_k}{2}\Big).$$

Next, we bound the remaining probability term. Note that $\mathbb{P}(T_k \leq \frac{t_k}{2}) = \mathbb{P}(T_k - \mathbb{E}[T_k] \leq -\frac{t_k}{2})$. Since $\alpha_i Y_i \in [0, \alpha_i]$, applying Hoeffding's inequality yields

$$\mathbb{P}(T_k - \mathbb{E}[T_k] \leq -\frac{t_k}{2}) \leq \exp\Big(-\frac{2\cdot(t_k/2)^2}{\sum_{i=0}^{k-1}\alpha_i^2}\Big) = \exp\Big(-\frac{p^2(\sum_{i=0}^{k-1}\alpha_i)^2}{2\sum_{i=0}^{k-1}\alpha_i^2}\Big).$$

which, together with the fact that $\sum_{i=0}^{k-1}\alpha_i^2 \leq \bar{\alpha}\sum_{i=0}^{k-1}\alpha_i$, implies

$$\mathbb{P}(T_k - \mathbb{E}[T_k] \leq -\frac{t_k}{2}) \leq \exp\Big(-\frac{p^2}{2\bar{\alpha}}\sum_{i=0}^{k-1}\alpha_i\Big).$$

Combining these bounds, we arrive at the desired bound

$$\mathbb{E}[\chi(T_k)] \leq \chi(0)\, \exp\Big(-\frac{p^2}{2\bar{\alpha}}\sum_{i=0}^{k-1}\alpha_i\Big) + \chi\Big(\frac{p}{2}\sum_{i=0}^{k-1}\alpha_i\Big).$$

This completes the proof. □

3.2.2. An Apparatus for Almost-Sure Convergence Rate Analysis

Based on (21), establishing an almost-sure upper bound for $h(x^k)$ reduces to deriving an almost sure lower bound for $\sum_{i=0}^{k-1} \alpha_i z_i$. This task is abstracted as follows.

Question 2. Consider a $\{0,1\}$-valued random sequence $\{X_k\}_{k\geq 0}$ such that X_k is $\mathcal{F}_{k+1}$-measurable and $\mathbb{P}(X_k = 1|\mathcal{F}_k) \geq p$ for all $k \geq 0$, where $p \in (0,1]$. Given a positive sequence $\{\alpha_k\}_{k\geq 0}$, can we obtain an almost-sure asymptotic lower bound on $\sum_{i=0}^{k-1} \alpha_i X_i$?

We answer this question in the proposition below, whose proof closely follows that of [51, Theorem 1].

Proposition 2. Consider Question 2. Suppose that sequence $\{\alpha_k\}_{k\geq 0}$ satisfies Assumption 2. Then,

$$\liminf_{k\to\infty} \frac{\sum_{j=0}^{k-1} \alpha_j X_j}{\sum_{j=0}^{k-1} \alpha_j} \geq p \text{ a.s.}.$$

Proof. Denote $b_j = \sum_{i=0}^{j} \alpha_i$ and $Z_j = X_j - \mathbb{E}[X_j|\mathcal{F}_j]$ for $j \geq 0$. Then, $\frac{\alpha_j}{b_j} Z_j$ is a martingale difference and $|Z_j| \leq 1$ almost surely. Define $U_k = \sum_{j=0}^{k-1} \frac{\alpha_j}{b_j} Z_j$ for $k \geq 1$. Then, U_k is a martingale. Observing that

$$\mathbb{E}\left[\left(\frac{\alpha_j}{b_j} Z_j\right)^2 \middle| \mathcal{F}_j\right] \leq \left(\frac{\alpha_j}{b_j}\right)^2 \quad \text{a.s.},$$

Assumption 2(ii) then implies that $\sum_{j=0}^{\infty} \mathbb{E}[(\frac{\alpha_j}{b_j} Z_j)^2|\mathcal{F}_j] < \infty$ almost surely. By [19, Theorem 2.17], U_k converges almost surely. Furthermore, since $0 < b_0 \leq b_1 \leq \cdots$ and $b_j \to +\infty$, an application of Kronecker's lemma [43, Lemma IV.3.2] yields

$$\frac{1}{b_{k-1}} \sum_{j=0}^{k-1} b_j \cdot \frac{\alpha_j}{b_j} Z_j = \sum_{j=0}^{k-1} \frac{\alpha_j Z_j}{b_{k-1}} \to 0 \quad \text{a.s.},$$

or equivalently,

$$\sum_{j=0}^{k-1} \frac{\alpha_j X_j}{b_{k-1}} - \sum_{j=0}^{k-1} \frac{\alpha_j \mathbb{E}[X_j|\mathcal{F}_j]}{b_{k-1}} \to 0 \quad \text{a.s.}.$$

The proof is then completed by noting that $\mathbb{E}[X_j|\mathcal{F}_j] = \mathbb{P}(X_j = 1|\mathcal{F}_j) \geq p$ almost surely. $\square$

3.2.3. Apparatuses for Transferring Convergence Rates of $h(x^k)$ to x^k

The aim of this part is devise a mechanism for establishing the convergence rate of $h(x^k)$ using that of x^k. We begin with the simpler case of transferring in-expectation rates.

Proposition 3. Suppose that Assumptions 1 and 3 hold and that $1/\sqrt{\phi_S}$ is integrable near 0. Then, there exists a random variable x^* such that x^k converges to x^*, both almost surely and in L^1, and that

$$\mathbb{E}[\|x^k - x^*\|] \leq \frac{\sqrt{c_3}}{c_1 c_2} \Phi_S(\mathbb{E}[h(x^k) - h^*]) \quad \forall k \geq 0. \tag{23}$$

Proof. Taking the conditional expectation on both sides of inequality (1) and applying Assumption 1(i), we obtain

$$\mathbb{E}[h(x^{k+1}) - h^* \mid \mathcal{F}_k] \le h(x^k) - h^* - c_1 c_2 \alpha_k g(x^k).$$

By Assumption 3 and inequality (3), we see that

$$\sqrt{g(x^k)} \ge \frac{1}{\sqrt{c_3}\alpha_k}\|x^{k+1} - x^k\| \quad \text{and} \quad \sqrt{g(x^k)} \ge \sqrt{\phi_S(h(x^k) - h^*)}.$$

Combining these inequalities, we have

$$\mathbb{E}[h(x^{k+1}) - h^* \mid \mathcal{F}_k] \le h(x^k) - h^* - \frac{c_1 c_2}{\sqrt{c_3}}\|x^{k+1} - x^k\|\sqrt{\phi_S(h(x^k) - h^*)}.$$

By inequality (1), $\|x^{k+1} - x^k\| = 0$ whenever $h(x^k) - h^\star = 0$. Applying Lemma 2(ii) with $a = h(x^k) - h^*$, $\tilde{a} = \mathbb{E}[h(x^{k+1}) - h^* \mid \mathcal{F}_k]$, and $\Delta = \frac{c_1 c_2}{\sqrt{c_3}}\|x^{k+1} - x^k\|$ yields

$$\|x^{k+1} - x^k\| \le \frac{\sqrt{c_3}}{c_1 c_2}(\Phi_S(h(x^k) - h^*) - \Phi_S(\mathbb{E}[h(x^{k+1}) - h^* \mid \mathcal{F}_k]))$$

Using the concavity of Φ_S, we obtain

$$\Phi_S(\mathbb{E}[h(x^{k+1}) - h^* \mid \mathcal{F}_k]) \ge \mathbb{E}[\Phi_S(h(x^{k+1}) - h^*) \mid \mathcal{F}_k] \quad \text{a.s.}.$$

Thus,

$$\mathbb{E}[\|x^{k+1} - x^k\|] \le \frac{\sqrt{c_3}}{c_1 c_2}(\mathbb{E}[\Phi_S(h(x^k) - h^*)] - \mathbb{E}[\Phi_S(h(x^{k+1}) - h^*)]), \tag{24}$$

which, together with Assumption 3(iii) and the continuity of Φ_S, yields $\sum_{k=0}^{\infty} \mathbb{E}[\|x^{k+1} - x^k\|] < +\infty$. Therefore, there exists a random variable x^* such that $x^k \to x^*$ almost surely and in L^1. Furthermore,

$$\mathbb{E}[\|x^k - x^*\|] \le \frac{\sqrt{c_3}}{c_1 c_2}\mathbb{E}[\Phi_S(h(x^k) - h^*)] \le \frac{\sqrt{c_3}}{c_1 c_2}\Phi_S(\mathbb{E}[h(x^k) - h^*]),$$

where the first inequality follows from a telescoping trick on (24) the second inequality follows directly from the concavity of Φ_S. □

A more challenging task is to transfer almost-sure convergence rates. This cannot be achieved by merely taking expectations; instead, a more refined, trajectory-based analysis is required. Our starting point is Assumption 1(i), which yields

$$\|x^{k+1} - x^k\| \le \sqrt{\frac{\alpha_k}{c_1}(h(x^k) - h(x^{k+1}))}.$$

If $h(x^k) - h^*$ converges linearly, the goal can be achieved by a simple relaxation:

$$\|x^{k+1} - x^k\| \le \sqrt{\frac{\bar{\alpha}}{c_1}(h(x^k) - h^*)}.$$

This approach is employed in [34, 46, 47]. However, the linear convergence of $h(x^k) - h^*$ typically holds only when ϕ_S is linear. To address this limitation, we derive a trajectory-based rate transference result applicable to nonlinear ϕ_S.

Proposition 4. Suppose that Assumptions 1, 2(i) and 3 hold and that $1/\sqrt{\phi_S}$ is integrable near 0. Given a fixed sample path ω, if there exists $\epsilon > 0$ and an integer $K \geq 1$ such that $\epsilon \sum_{i=0}^{k-1} \alpha_i$ lies in the domain of Ψ^{-1} for all $k \geq K$ and

$$h(x^k(\omega)) - h^* \leq \Psi_S^{-1}\left(\epsilon \sum_{i=0}^{k} \alpha_i\right) \quad \forall k \geq K, \tag{25}$$

then $\{x^k(\omega)\}_{k\geq 0}$ converges to some $x^*(\omega) \in \arg\min_{x\in S} h(x)$ and

$$\|x^k(\omega) - x^*(\omega)\| \leq \frac{2}{\sqrt{c_1\epsilon}} \Phi_S\left(\Psi_S^{-1}\left(\epsilon \sum_{i=0}^{k-1} \alpha_i\right)\right) \quad \forall k \geq K.$$

Proof. For any positive, non-increasing sequence $\{\beta_k\}_{k\geq 0}$, Assumption 1(i) implies that

$$\|x^{k+1}(\omega) - x^k(\omega)\| \leq \sqrt{\frac{\alpha_k}{c_1}[h(x^k(\omega)) - h(x^{k+1}(\omega))]} \leq \frac{\alpha_k\beta_k}{c_1} + \frac{h(x^k(\omega)) - h(x^{k+1}(\omega))}{\beta_k}.$$

Define $L(t) = \Phi_S(\Psi_S^{-1}(t))$. Then,

$$L'(t) = \frac{1}{\sqrt{\phi_S(\Psi_S^{-1}(t))}} \cdot (-\phi_S(\Psi_S^{-1}(t))) = -\sqrt{\phi_S(\Psi_S^{-1}(t))}.$$

Since L' is non-decreasing, it follows that L is convex. Let $b_k = \sum_{i=0}^{k} \alpha_i$ and choose $\beta_k = \sqrt{c_1\epsilon\phi_S(h(x^k(\omega)) - h^*)}$. Then, for all $k \geq K$, we have from (25) that

$$\alpha_k\beta_k \leq (b_k - b_{k-1})\sqrt{c_1\epsilon\phi_S(\Psi_S^{-1}(\epsilon b_k))} = \sqrt{c_1\epsilon}(b_{k-1} - b_k)L'(\epsilon b_k) \leq \sqrt{\frac{c_1}{\epsilon}}(L(\epsilon b_{k-1}) - L(\epsilon b_k)).$$

Meanwhile, by the concavity of Φ_S, we have

$$\begin{aligned}
&\beta_k^{-1}(h(x^k(\omega)) - h(x^{k+1}(\omega))) \\
= \;& \frac{1}{\sqrt{c_1\epsilon}}\left[\Phi_S'(h(x^k(\omega)) - h^*)\left((h(x^k(\omega)) - h^*) - (h(x^{k+1}(\omega)) - h^*)\right)\right] \\
\leq \;& \frac{1}{\sqrt{c_1\epsilon}}\left[\Phi_S(h(x^k(\omega)) - h^*) - \Phi_S(h(x^{k+1}(\omega)) - h^*)\right].
\end{aligned}$$

Finally, by inequality (25) and the monotonicity of Φ_S and Ψ_S, we get

$$\begin{aligned}
\sum_{k=K}^{\infty} \|x^{k+1}(\omega) - x^k(\omega)\| &\leq \sum_{k=K}^{\infty}\left[\frac{\alpha_k\beta_k}{c_1} + \beta_k^{-1}[h(x^k(\omega)) - h(x^{k+1}(\omega))]\right] \\
&\leq \frac{1}{c_1}\sqrt{\frac{c_1}{\epsilon}}\Phi_S(\Psi_S^{-1}(\epsilon b_{K-1})) + \frac{1}{\sqrt{c_1\epsilon}}\Phi_S(h(x^K(\omega)) - h^*) \\
&\leq \frac{2}{\sqrt{c_1\epsilon}}\Phi_S(\Psi_S^{-1}(\epsilon b_{K-1})),
\end{aligned}$$

which shows that $\{x^k(\omega)\}_{k\geq 0}$ converges to some $x^*(\omega) \in \arg\min_{x\in S} h(x)$ and then yields the desired inequality. This completes the proof. □

3.2.4. Proof of Theorem 3

We begin with assertion (i). By inequality (21), applying Proposition 1 with $\chi(t) = \Psi_S^{-1}(\Psi_S(h(x^0) - h^*) + \frac{c_1 c_2}{2} t)$ and $X_k = z_k$, we obtain inequality (17). The existence of the limit x^* and inequality (18) follow directly from Proposition 3. It remains to prove that $x^* \in \arg\min_{x \in S} h(x)$ almost surely. From (17) and Lemma 1(i), we have $\mathbb{E}[h(x^k) - h^*] \to 0$, which implies $h(x^k) \to h^*$ in L^1. Therefore, by [21, Theorem 4.3.3], there exists a subsequence $\{k_n\}_n$ such that $h(x^{k_n}) \to h^*$ almost surely, which in turn implies that $x^* \in \arg\min_{x \in S} h(x)$.

Next, we prove assertion (ii). Define the event

$$E = \left\{ \omega : \liminf_{k \to \infty} \frac{\sum_{i=0}^{k-1} \alpha_i z_i}{\sum_{i=0}^{k-1} \alpha_i} \geq p \right\}.$$

Under Assumption 2, applying Proposition 2 with $X_k = z_k$ yields $\mathbb{P}(E) = 1$. Define

$$K_2 = \min \left\{ N : \Psi_S(h(x^0) - h^*) + \frac{c_1 c_2}{2} \sum_{i=0}^{k-1} \alpha_i z_i \geq \frac{c_1 c_2 p}{3} \sum_{i=0}^{k-1} \alpha_i, \quad \forall\, k \geq N \right\}.$$

Then, K_2 is finite on E. By the definition of K_2, Lemma 4 and the monotonicity of Ψ_S^{-1}, we obtain (19). Moreover, by Assumption 2(i), there exists a deterministic N such that

$$\frac{1}{3} \sum_{i=0}^{k-1} \alpha_i \geq \frac{1}{4} \sum_{i=0}^{k} \alpha_i, \quad \forall\, k \geq N.$$

Define $K_3 = \max\{K_2, N\}$, which is also finite on E. By inequality (19), we have on E that

$$h(x^k) - h^* \leq \Psi_S^{-1} \left(\frac{c_1 c_2 p}{4} \sum_{i=0}^{k} \alpha_i \right) \quad \forall k \geq K_3.$$

Applying Proposition 4 with $\epsilon = \frac{c_1 c_2 p}{4}$, we obtain (20).

4. Convergence Analysis under Local Error Bounds

Our next goal is to analyze the algorithms in Assumption 1 under a local error bound condition defined below. From now on, we denote by $F = g^{-1}(\{0\}) \cap S$ the target set.

Assumption 4 (Local error bound)**.** For any $\bar{x} \in F$, there exists a function $\phi_{\bar{x}} : [0, \eta_{\bar{x}}) \to \mathbb{R}_+$, where $\eta_{\bar{x}} > 0$ is some constant, such that the following hold.

(i) There exist $\delta_{\bar{x}} > 0$ such that

$$\phi_{\bar{x}}(h(x) - h(\bar{x})) \leq g(x) \quad \forall x \in \{x \in \mathbb{R}^d : h(\bar{x}) < h(x) < h(\bar{x}) + \eta_{\bar{x}}\} \cap \mathbb{B}(\bar{x}, \delta_{\bar{x}}) \cap S.$$

(ii) The function $\phi_{\bar{x}}$ is continuous and non-decreasing on $[0, \eta_{\bar{x}})$, strictly positive on $(0, \eta_{\bar{x}})$, and satisfies $\phi_{\bar{x}}(0) = 0$.

A key distinction in this setting is that although $h(x^k)$ remains non-increasing, the limit $h^\infty = \lim_{k\to\infty} h(x^k)$ is a random variable. We first show that the local error bound condition in Assumption 4 can be uniformized in the sense of [8, Lemma 6].

Lemma 6 (Uniformized error bound)**.** Suppose that Assumption 4 holds. Define $V = h(F)$. Then, for any $v \in h(F)$ and compact set $C \subseteq h^{-1}(v) \cap F$, there exist $\delta_C, \eta_C > 0$ and a finite set $\{\bar{x}_j\}_{j=1,\ldots,l} \subseteq C$ such that

$$\phi_C(h(x) - v) \leq g(x) \quad \forall x \in \{x \in \mathbb{R}^d : v < h(x) < v + \eta_C,\ \operatorname{dist}(x, C) < \delta_C\} \cap S,$$

where $\phi_C : [0, \eta_C) \to \mathbb{R}_+$ is defined by $\phi_C = \min_{j=1,\ldots,l} \phi_{\bar{x}_j}$ with $\phi_{\bar{x}}$ defined in Assumption 4.

Proof. By compactness, C can be covered by a finite number of open balls with centers $\bar{x}_1, \ldots, \bar{x}_l \in C$ and the corresponding radii $\delta_{\bar{x}_1}, \ldots, \delta_{\bar{x}_l}$ given in Assumption 4(i). Hence, there exists a constant $\delta_C > 0$ such that the δ_C-extension $\{x \in \mathbb{R}^d : \operatorname{dist}(x, C) < \delta_C\}$ can be covered by the same collection of open balls. The proof is then completed by taking $\eta_C = \min_{j=1,\ldots,l} \eta_{\bar{x}_j}$, and $\phi_C = \min_{j=1,\ldots,l} \phi_{\bar{x}_j}$. □

The following two events will be crucial to our analysis in this section:

$$E_0 = \left\{\omega : g(x^k) \to 0,\ \liminf_{k\to\infty} \frac{\sum_{i=1}^k \alpha_i z_i}{\sum_{i=1}^k \alpha_i} \geq p\right\} \quad \text{and} \quad E_1 = \{\omega : x^k \text{ is bounded}\}. \tag{26}$$

We now present our main convergence rate result in the local error bound setting.

Theorem 4. Suppose that Assumptions 1, 2(i) and 4 hold and that $1/\sqrt{\phi_{\bar{x}}}$ is integrable near 0 for all $\bar{x} \in F$. Then, there exists a random variable x^* such that $x^* \in F$ and $x^k \to x^*$ on $E_0 \cap E_1$. Moreover, the following hold on $E_0 \cap E_1$.

(i) If $1/\phi_{x^*(\omega)}$ is integrable near 0, then the sequence $\{x^k(\omega)\}_{k\geq 0}$ terminates in finitely many iterations.

(ii) If $1/\phi_{x^*(\omega)}$ is not integrable near 0, then there exists a positive integer $K_4(\omega)$ such that for all $k \geq K_4(\omega)$,

$$h(x^k(\omega)) - h^\infty(\omega) \leq \Psi^{-1}_{x^*(\omega)}\left(\frac{c_1 c_2 p}{3} \sum_{i=0}^{k-1} \alpha_i\right)$$

and

$$\|x^k - x^*\| \leq \frac{4}{c_1\sqrt{c_2 p}} \Phi_{x^*(\omega)}\left(\Psi^{-1}_{x^*(\omega)}\left(\frac{c_1 c_2 p}{4} \sum_{i=0}^{k-1} \alpha_i\right)\right).$$

Proof. For any $\omega \in E_0 \cap E_1$, denote by $\mathcal{A}(\omega)$ the set of limit points of $\{x^k(\omega)\}_{k\geq 0}$. Then, $\mathcal{A}(\omega)$ is compact, $\mathcal{A}(\omega) \subseteq F$ by the continuity of g, and $h(x) = h^\infty(\omega) \in h(F)$ for all $x \in \mathcal{A}(\omega)$. By Lemma 6, there exist positive constants $\delta_{\mathcal{A}(\omega)}$, $\eta_{\mathcal{A}(\omega)}$ and a finite set $\{\bar{x}_j\}_{j=1,\ldots,l} \subseteq \mathcal{A}(\omega)$ such that

$$\phi_{\mathcal{A}(\omega)}(h(x) - h^\infty(\omega)) \leq g(x)$$

for any $x \in S$ satisfying $h^\infty(\omega) < h(x) < h^\infty(\omega) + \eta_{\mathcal{A}(\omega)}$ and $\operatorname{dist}(x, \mathcal{A}(\omega)) < \delta_{\mathcal{A}(\omega)}$, where $\phi_{\mathcal{A}(\omega)} = \min_{j=1,\ldots,l} \phi_{\bar{x}_j}$ is defined on $[0, \eta_{\mathcal{A}(\omega)})$. Since $1/\sqrt{\phi_{\bar{x}_j}}$ is integrable near 0 for all $j \in [l]$, it follows that $1/\sqrt{\phi_{\mathcal{A}(\omega)}}$ is also integrable near 0. Define

$$N_1(\omega) = \min\{N : h^\infty(\omega) \leq h(x^k(\omega)) < h^\infty(\omega) + \eta_{\mathcal{A}(\omega)},\ \operatorname{dist}(x^k(\omega), \mathcal{A}(\omega)) < \delta_{\mathcal{A}(\omega)}\ \forall k \geq N\}.$$

Then, $N_1(\omega) < \infty$. For any $k \geq N_1(\omega)$, by Lemma 2(i) and Lemma 3, we have

$$\Psi_{\mathcal{A}(\omega)}(h(x^{k+1}(\omega)) - h^\infty(\omega)) \geq \Psi_{\mathcal{A}(\omega)}(h(x^k(\omega)) - h^\infty(\omega)) + \frac{c_1 c_2}{2}\alpha_k z_k(\omega),$$

whenever $h(x^k(\omega)) > h^\infty(\omega)$. Since $\alpha_k z_k$ is not summable on E_0, if $1/\phi_{\mathcal{A}(\omega)}$ is integrable near 0, then the algorithm terminates in finitely many iterations. If $1/\phi_{\mathcal{A}(\omega)}$ is not integrable near 0, following similar arguments as in the proof of Lemma 4 but replacing the global error bound in Assumption 3 by the uniformized error bound in Lemma 6, we obtain

$$h(x^k(\omega)) - h^\infty(\omega) \leq \Psi^{-1}_{\mathcal{A}(\omega)}\left(\Psi_{\mathcal{A}(\omega)}(h(x^{N_1(\omega)}(\omega)) - h^\infty(\omega)) + \frac{c_1 c_2}{2}\sum_{i=N_1(\omega)}^{k-1} \alpha_i z_i(\omega)\right) \quad \forall k > N_1(\omega).$$

Since Ψ is non-negative and $\omega \in E_0$, there exists a integer $N_2(\omega) > N_1(\omega)$ such that

$$\Psi_{\mathcal{A}(\omega)}(h(x^{N_1(\omega)}(\omega)) - h^\infty(\omega)) + \frac{c_1 c_2}{2}\sum_{i=N_1(\omega)}^{k-1} \alpha_i z_i \geq \frac{c_1 c_2 p}{3}\sum_{i=0}^{k-1}\alpha_i \geq \frac{c_1 c_2 p}{4}\sum_{i=0}^{k}\alpha_i \quad \forall k \geq N_2(\omega).$$

Thus, for all $k \geq N_2(\omega)$, we have

$$h(x^k(\omega)) - h^\infty(\omega) \leq \Psi^{-1}_{\mathcal{A}(\omega)}\left(\frac{c_1 c_2 p}{3}\sum_{i=0}^{k-1}\alpha_i\right) \leq \Psi^{-1}_{\mathcal{A}(\omega)}\left(\frac{c_1 c_2 p}{4}\sum_{i=0}^{k}\alpha_i\right).$$

Following similar arguments as in the proof of Proposition 4 but replacing the global error bound in Assumption 3 by the uniformized error bound in Lemma 6, $\{x^k(\omega)\}_{k\geq 0}$ converges to some $x^*(\omega) \in F$ such that

$$\|x^k(\omega) - x^*(\omega)\| \leq \frac{4}{c_1\sqrt{c_2 p}}\Phi_{\mathcal{A}(\omega)}\left(\Psi^{-1}_{\mathcal{A}(\omega)}\left(\frac{c_1 c_2 p}{4}\sum_{i=0}^{k-1}\alpha_i\right)\right) \quad \forall k \geq N_2(\omega).$$

In particular, $\mathcal{A}(\omega) = \{x^*(\omega)\}$. Substituting $\mathcal{A}(\omega)$ with $x^*(\omega)$ in the previous argument and setting $K_4(\omega) = N_2(\omega)$ yields the desired inequality. Finally, extending the definition of x^* by setting $x^*(\omega) = 0$ for $\omega \notin E_0 \cap E_1$ completes the proof. $\square$

The next lemma shows that the progress metric $g(x^k)$ converges to zero almost surely.

Lemma 7. Suppose that Assumptions 1(i) and 2(i) hold and that $\Gamma \circ g$ is uniformly continuous for some strictly increasing function $\Gamma : \mathbb{R}_+ \to \mathbb{R}$. Then, $g(x^k) \to 0$ almost surely.

Proof. By inequalities (1) and (2), we have

$$\sum_{k=0}^{\infty} \frac{1}{\alpha_k} \|x^{k+1} - x^k\|^2 < +\infty \quad \text{and} \quad \sum_{k=0}^{\infty} \alpha_k \, \mathbb{E}[g(x^k)] < +\infty,$$

which implies $\sum_{k=0}^{\infty} \alpha_k g(x^k) < +\infty$ almost surely. Define the event

$$E = \left\{ \omega : \sum_{k=0}^{\infty} \alpha_k g(x^k(\omega)) < +\infty \right\}.$$

Then $\mathbb{P}(E) = 1$. Fix any $\omega \in E$. Since $\sum_{k=1}^{\infty} \alpha_k = \infty$, we have $\liminf_{k\to\infty} g(x^k(\omega)) = 0$. Suppose that $\limsup_{k\to\infty} g(x^k(\omega)) > 0$. Then, these exists an $\epsilon > 0$ and two index sequences $\{l_j\}_{j\geq 0}$ and $\{u_j\}_{j\geq 0}$ such that

$$l_j < u_j < l_{j+1} \quad \forall j \geq 0,$$

and that the corresponding subsequences $\{x^{l_j}\}_{j\geq 0}$ and $\{x^{u_j}\}_{j\geq 0}$ satisfy

$$g(x^{l_j}(\omega)) \geq \epsilon, \quad g(x^{u_j}(\omega)) \leq \frac{\epsilon}{2} \quad \text{and} \quad g(x^k(\omega)) \geq \frac{\epsilon}{2} \quad \forall l_j < k < u_j.$$

Define $\gamma_j = \sum_{k=l_j}^{u_j - 1} \alpha_k$. Since $\sum_{k=0}^{\infty} \alpha_k g(x^k) < +\infty$, γ_j is summable. Then, as $j \to \infty$,

$$\|x^{u_j}(\omega) - x^{l_j}(\omega)\| \leq \sum_{k=l_j}^{u_j-1} \|x^{k+1}(\omega) - x^k(\omega)\| \leq \sqrt{\gamma_j} \left(\sum_{k=l_j}^{u_j-1} \frac{1}{\alpha_k} \|x^{k+1}(\omega) - x^k(\omega)\|^2 \right)^{\frac{1}{2}} \to 0.$$

On the other hand, by the uniform continuity of $\Gamma \circ g$, we have

$$0 < \Gamma(\epsilon) - \Gamma(\frac{\epsilon}{2}) \leq \Gamma(g(x^{l_j}(\omega))) - \Gamma(g(x^{u_j}(\omega))) \to 0, \quad \text{as } j \to \infty,$$

which yields a contradiction. Thus, $g(x^k(\omega)) \to 0$ for all $\omega \in E$. □

The following result is immediate from Theorem 4 and Lemma 7.

Corollary 1. Suppose that Assumptions 1, 2 and 4 hold with a compact S and that $1/\sqrt{\phi_{\bar{x}}}$ is integrable near 0 for all $\bar{x} \in F$. Then, there exists a random variable x^* such that $x^* \in F$ and $x^k \to x^*$ almost surely. Moreover, $\mathbb{P}(E_0) = 1$ and the following hold on E_0.

(i) If $1/\phi_{x^*(\omega)}$ is integrable near 0, then the sequence $\{x^k(\omega)\}_{k\geq 0}$ terminates in finitely many iterations.

(ii) If $1/\phi_{x^*(\omega)}$ is not integrable near 0, then there exists a positive integer $K_4(\omega)$ such that for all $k \geq K_4(\omega)$,

$$h(x^k(\omega)) - h^\infty(\omega) \leq \Psi_{x^*(\omega)}^{-1}\left(\frac{c_1 c_2 p}{3}\sum_{i=0}^{k-1}\alpha_i\right)$$

and

$$\|x^k(\omega) - x^*(\omega)\| \leq \frac{4}{c_1\sqrt{c_2 p}}\Phi_{x^*(\omega)}\left(\Psi_{x^*(\omega)}^{-1}\left(\frac{c_1 c_2 p}{4}\sum_{i=0}^{k-1}\alpha_i\right)\right).$$

Proof. Note that $E_1 = \Omega$ since S is bounded and that $\mathbb{P}(E_0) = 1$ by Lemma 7. The conclusions then follow directly from Theorem 4. □

5. Applications

5.1. Application to RKMM

Next, we analyze RKMM for solving the common fixed point problem (4) using our theoretical framework. To begin, we recall the notion of firm quasi-nonexpansiveness [4, Definition 4.1].

Definition 3 (Firm quasi-nonexpansiveness)**.** Let $B \subseteq \mathbb{R}^d$ be a nonempty subset. Then, $T : B \to \mathbb{R}^d$ is said to be firmly quasi-nonexpansive if

$$\|Tx - y\|^2 + \|Tx - x\|^2 \leq \|x - y\|^2 \quad \forall y \in \operatorname{Fix} T,\ x \in B.$$

Algorithms for common fixed point problems or convex feasibility problems are often analyzed under Hölderian error bound conditions; see, *e.g.*, [9, 10, 28, 29].

Definition 4 (Hölderian error bounds)**.** A finite family of operators $T_1, \ldots, T_m : \mathbb{R}^d \to \mathbb{R}^d$ with a non-empty common fixed point set $F = \cap_{i\in[m]} \operatorname{Fix} T_i$ is said to satisfy the Hölderian error bound on a set $B \subseteq \mathbb{R}^d$ with exponent $\theta \in (0, 1]$ if there exists $r > 0$ such that

$$\operatorname{dist}(x, F) \leq r \max_{i\in[m]} \|T_i(x) - x\|^\theta \quad \forall x \in B.$$

Note that we have restricted the exponent θ to $(0, 1]$. This is because the property of firm quasi-nonexpansiveness contradicts the case of $\theta > 1$.

Lemma 8. Let $T_1, \ldots, T_m : \mathbb{R}^d \to \mathbb{R}^d$ be a family of firmly quasi-nonexpansive operators with a non-empty common fixed point set $F = \bigcap_{i\in[m]} \operatorname{Fix} T_i$ such that Assumption 3 is satisfied with $h(x) = \operatorname{dist}^2(x, F)$, $g(x) = \max_{i\in[m]} \|T_i(x) - x\|^2$, $S = \{x \in \mathbb{R}^d : \operatorname{dist}(x, F) \leq \operatorname{dist}(x^0, F)\}$ for some $x^0 \notin F$, $\tau > \operatorname{dist}^2(x^0, F)$, and some $\phi_S : [0, \tau) \to \mathbb{R}_+$. Then, $1/\phi_S$ must be non-integrable near 0.

Proof. For any $t \in (0, \operatorname{dist}^2(x^0, F))$, there exists a point $x_t \in S$ such that $h(x_t) = t$. By the definition of ϕ_S, we have

$$\phi_S(t) \leq g(x_t). \tag{27}$$

Next, the firm quasi-nonexpansiveness of T_i guarantees that for all $i \in [m]$,

$$\operatorname{dist}^2(T_i(x_t), F) + \|T_i(x_t) - x_t\|^2 \leq \operatorname{dist}^2(x_t, F).$$

This implies that $h(x_t) = \operatorname{dist}^2(x_t, F) \geq \max_{i\in[m]} \|T_i(x_t) - x_t\|^2 = g(x_t)$. Combining this bound with (27), we obtain

$$\phi_S(t) \leq t \quad \text{for all } t \in (0, \operatorname{dist}^2(x^0, F)).$$

Taking the reciprocal yields $1/\phi_S(t) \geq 1/t$. Since the function $1/t$ is non-integrable near 0, it follows that $1/\phi_S$ is also non-integrable near 0. □

For Hölderian error bounds defined in Definition 4, $\phi_S(t) = (r^{-2}t)^{\frac{1}{\theta}}$. An immediate consequence of Lemma 8 is the impossibility of $\theta > 1$ under the firm quasi-nonexpansiveness of the operators.

The implications of Hölderian error bounds on RKMM are summarized below.

Proposition 5. Consider RKMM (see (5)) for solving the common fixed point problem (4) with an arbitrary initial point $x^0 \in \mathbb{R}^d$ and $\beta_k \in (0, 1]$ for all $k \geq 0$. Take $h(x) = \operatorname{dist}^2(x, F)$, $g(x) = \max_{i\in[m]} \|T_i(x) - x\|^2$, and $S = \{x \in \mathbb{R}^d : \operatorname{dist}(x, F) \leq \operatorname{dist}(x^0, F)\}$, where F is the set of common fixed points.

(i) If T_i is firmly quasi-nonexpansive for all $i \in [m]$ and there exists $\rho > 0$ such that $\mathbb{P}(i_k = i \mid \mathcal{F}_k) \geq \rho$ for all $i \in [m]$ and $k \geq 0$, then Assumption 1 holds with $c_1 = 1$, $c_2 = \rho$, $c_3 = 1$ and $\alpha_k = \beta_k$.

(ii) If a finite family of operators $\{T_i\}_{i\in[m]}$ satisfy the Hölderian error bound on S with exponent $\theta \in (0, 1]$, then Assumption 3 holds with $\phi_S(t) = (r^{-2}t)^{\frac{1}{\theta}}$.

Proof. We first prove assertion (i). By the definition (5) of RKMM and [4, Proposition 4.3],

$$\operatorname{dist}^2(x^{k+1}, F) \leq \operatorname{dist}^2(x^k, F) - \frac{2 - \beta_k}{\beta_k}\|x^{k+1} - x^k\|^2,$$

which together with $\beta_k \in (0, 1]$ yields (1) with $\alpha_k = \beta_k$ and $c_1 = 1$. Also,

$$\mathbb{E}[\|x^{k+1} - x^k\|^2|\mathcal{F}_k] = \sum_{i\in[m]} \mathbb{P}(i_k = i \mid \mathcal{F}_k) \cdot \beta_k^2 \|T_i(x^k) - x^k\|^2 \geq \rho\, \beta_k^2 \max_{i\in[m]} \|T_i(x^k) - x^k\|^2,$$

which yields (2) with $\alpha_k = \beta_k$ and $c_2 = \rho$. Furthermore, we obviously have

$$\|x^{k+1} - x^k\|^2 = \beta_k^2 \|T_{i_k}(x^k) - x^k\|^2 \leq \beta_k^2 \max_{i\in[m]} \|T_i(x^k) - x^k\|^2,$$

which yields (3) with $\alpha_k = \beta_k$ and $c_3 = 1$. This proves assertion (i).

Next, it is easy to see that Assumption 3 holds with $\phi_S : \mathbb{R}_+ \to \mathbb{R}_+$ defined by $\phi_S(t) = (r^{-2}t)^{\frac{1}{\theta}}$, which proves assertion (ii). □

We can now present the convergence rate result for RKMM.

Theorem 5. Consider RKMM (5) for solving the common fixed point problem (4) with an arbitrary initial point $x^0 \in \mathbb{R}^d$ and stepsizes β_k satisfying Assumption 2. Suppose that the operators $\{T_i\}_{i\in[m]}$ are firmly quasi-nonexpansive and satisfy the Hölderian error bound on S with exponent $\theta \in (0,1]$. Furthermore, assume there exists $\rho > 0$ such that $\mathbb{P}(i_k = i \mid \mathcal{F}_k) \geq \rho$ for all $i \in [m]$ and $k \geq 0$. Let $F = \bigcap_{i\in[m]} \operatorname{Fix} T_i$, $p = \frac{\rho}{2-\rho}$, and $A_k = \sum_{i=0}^{k-1} \alpha_i$. Then the following hold.

(i) If $\theta \in (0,1)$, then there exist constants $C_1, C_2 > 0$ and an almost surely finite random index J_1 such that

$$\mathbb{E}[\operatorname{dist}^2(x^k, F)] \leq C_1 A_k^{-\frac{\theta}{1-\theta}} \quad \forall k \geq 1,$$

and

$$\operatorname{dist}^2(x^k, F) \leq C_2 A_k^{-\frac{\theta}{1-\theta}} \quad \forall k \geq J_1 \text{ a.s..}$$

(ii) If $\theta \in (\frac{1}{2}, 1)$, there exists a random variable x^* such that $x^* \in F$ almost surely and that that x^k converges to x^* both in L^1 and almost surely. Furthermore, there exist constants $C_3, C_4 > 0$ and an almost surely finite random index J_2 such that

$$\mathbb{E}[\|x^k - x^*\|] \leq C_3 A_k^{-\frac{2\theta-1}{2-2\theta}} \quad \forall k \geq 0,$$

and

$$\|x^k - x^*\| \leq C_4 A_k^{-\frac{2\theta-1}{2-2\theta}} \quad \forall k \geq J_2 \text{ a.s..}$$

(iii) If $\theta = 1$, there exists a random variable x^* such that $x^* \in F$ almost surely and that x^k converges to x^* both in L^1 and almost surely. Furthermore, there exist constants $C_5, \dots, C_{12} > 0$ and almost surely finite random indices J_3, J_4 such that

$$\begin{aligned}
\mathbb{E}[\operatorname{dist}^2(x^k, F)] &\leq C_5 \exp(-C_6 A_k) \quad \forall k \geq 1, \\
\operatorname{dist}^2(x^k, F) &\leq C_7 \exp(-C_8 A_k) \quad \forall k \geq J_3 \text{ a.s.,} \\
\mathbb{E}[\|x^k - x^*\|] &\leq C_9 \exp(-C_{10} A_k) \quad \forall k \geq 0, \\
\|x^k - x^*\| &\leq C_{11} \exp(-C_{12} A_k) \quad \forall k \geq J_4 \text{ a.s..}
\end{aligned}$$

Proof. By Proposition 5, Assumption 1 holds with $c_1 = 1$, $c_2 = \rho$, $c_3 = 1$, and $\alpha_k = \beta_k$. Additionally, Assumption 3 holds with $\phi_S = (r^{-2} t)^{\frac{1}{\theta}}$. Recall from Lemma 3 the definition of p; in this case, we have $p = \frac{c_2}{2c_3 - c_2} = \frac{\rho}{2-\rho}$. Because $\theta \in (0,1]$, the function $1/\phi_S(t) = r^{\frac{2}{\theta}} t^{-\frac{1}{\theta}}$ is not integrable near 0.

For $\theta \in (0,1)$, the inverse smoothing function (with $t_0 = 1$) and its inverse are given, respectively, by

$$\Psi_S(t) = \int_t^1 r^{\frac{2}{\theta}} u^{-\frac{1}{\theta}} \mathrm{d}u = \frac{\theta\, r^{\frac{2}{\theta}}}{1-\theta}(t^{1-\frac{1}{\theta}} - 1) \quad \text{and} \quad \Psi_S^{-1}(s) = \left(1 + \frac{1-\theta}{\theta\, r^{\frac{2}{\theta}}} s\right)^{-\frac{\theta}{1-\theta}}.$$

Then, it follows from inequality (17) that

$$\mathbb{E}[\mathrm{dist}^2(x^k,F)] \le \exp\Big(-\frac{p^2}{2\bar{\alpha}}A_k\Big)\,[\mathrm{dist}^2(x^0,F)] + \Psi_S^{-1}\Big(\Psi_S(\mathrm{dist}^2(x^0,F)) + \frac{c_1c_2p}{4}A_k\Big)$$
$$= \exp\Big(-\frac{p^2}{2\bar{\alpha}}A_k\Big)\,[\mathrm{dist}^2(x^0,F)] + \left(\mathrm{dist}^{2-\frac{2}{\theta}}(x^0,F) + \frac{p\rho(1-\theta)}{4\theta\, r^{\frac{2}{\theta}}}A_k\right)^{-\frac{\theta}{1-\theta}} \le C_1 A_k^{-\frac{\theta}{1-\theta}},$$

where $C_1 > 0$ is a constant. Moreover, Theorem 3(ii) guarantees the existence of an almost surely finite random index J_1 such that for any $k \ge J_1$, we have almost surely

$$\mathrm{dist}^2(x^k,F) \le \Psi_S^{-1}\Big(\frac{c_1c_2p}{3}A_k\Big) = \left(1 + \frac{p\rho(1-\theta)}{3\theta\, r^{\frac{2}{\theta}}}A_k\right)^{-\frac{\theta}{1-\theta}} \le C_2 A_k^{-\frac{\theta}{1-\theta}},$$

for some $C_2 > 0$.

When $\theta \in (\frac{1}{2},1)$, the function $1/\sqrt{\phi_S(t)} = r^{\frac{1}{\theta}}t^{-\frac{1}{2\theta}}$ is integrable near 0. In this case, the desingularization function is

$$\Phi_S(t) = \int_0^t r^{\frac{1}{\theta}}u^{-\frac{1}{2\theta}}\mathrm{d}u = \frac{2\theta r^{\frac{1}{\theta}}}{2\theta-1}t^{1-\frac{1}{2\theta}}.$$

By Theorem 3, there exists a random variable x^* such that $x^* \in F$ almost surely, and x^k converges to x^* both in L^1 and almost surely. Furthermore, it follows from inequalities (18) and (20) that

$$\mathbb{E}[\|x^k - x^*\|] \le \frac{\sqrt{c_3}}{c_1c_2}\Phi_S(\mathbb{E}[\mathrm{dist}^2(x^k,S^*)]) \le \frac{\sqrt{c_3}}{c_1c_2}\frac{2\theta r^{\frac{1}{\theta}}}{2\theta-1}\left(C_1A_k^{-\frac{\theta}{1-\theta}}\right)^{1-\frac{1}{2\theta}} \le C_3A_k^{-\frac{2\theta-1}{2-2\theta}} \quad \forall k \ge 0,$$

and that there exists an almost surely finite random index J_2 such that

$$\|x^k - x^*\| \le \frac{4}{c_1\sqrt{c_2p}}\Phi_S\left(\Psi_S^{-1}\left(\frac{c_1c_2p}{4}\sum_{i=0}^{k-1}\beta_i\right)\right) \le C_4A_k^{-\frac{2\theta-1}{2-2\theta}} \quad \forall k \ge J_2 \text{ a.s.},$$

where $C_3, C_4 > 0$ are constants.

Finally, if $\theta = 1$, the inverse smoothing function (with $t_0 = 1$) and its inverse are, respectively,

$$\Psi_S(t) = \int_t^1 r^2u^{-1}\mathrm{d}u = -r^2\ln t \quad \text{and} \quad \Psi_S^{-1}(s) = \exp\left(-\frac{s}{r^2}\right),$$

and the desingularization function is

$$\Phi_S(t) = \int_0^t ru^{-\frac{1}{2}}\mathrm{d}u = 2rt^{\frac{1}{2}}.$$

Then, it follows from inequality (17) that

$$\mathbb{E}[\mathrm{dist}^2(x^k,F)] \le \exp\big(-p^2A_k\big)\,[\mathrm{dist}^2(x^0,F)] + \Psi_S^{-1}\Big(\Psi_S(\mathrm{dist}^2(x^0,F)) + \frac{c_1c_2p}{4}A_k\Big)$$
$$= \exp\big(-p^2A_k\big)\,[\mathrm{dist}^2(x^0,F)] + \exp\left(\ln(\mathrm{dist}^2(x^0,F)) - \frac{p\rho}{4r^2}A_k\right) \le C_5\exp\left(-C_6A_k\right),$$

for some $C_5, C_6 > 0$. Moreover, Theorem 3(i) guarantees the existence of a random variable x^* such that $x^* \in F$ almost surely, and x^k converges to x^* both in L^1 and almost surely. It then follows from inequality (18) that

$$\mathbb{E}[\|x^k - x^*\|] \leq \frac{\sqrt{c_3}}{c_1 c_2}\Phi_S(\mathbb{E}[\text{dist}^2(x^k, S^*)]) \leq C_9 \exp\left(-C_{10}A_k\right) \quad \forall k \geq 0,$$

for some constants $C_9, C_{10} > 0$. Furthermore, by Theorem 3(ii), there exist almost surely finite random indices J_3 and J_4 such that

$$\text{dist}^2(x^k, F) \leq \Psi_S^{-1}\left(\frac{c_1 c_2 p}{3}A_k\right) \leq C_7 \exp\left(-C_8 A_k\right) \quad \forall k \geq J_3 \text{ a.s.},$$

and

$$\|x^k - x^*\| \leq \frac{4}{c_1\sqrt{c_2 p}}\Phi_S\left(\Psi_S^{-1}\left(\frac{c_1 c_2 p}{4}A_k\right)\right) \leq C_{11} \exp\left(-C_{12}A_k\right) \quad \forall k \geq J_4 \text{ a.s.},$$

where $C_7, C_8, C_{11}, C_{12} > 0$ are constants. □

Note that the constants $C_1, \ldots, C_{12}$ in Theorem 5 are deterministic and depend only on ρ, θ, r, and $\text{dist}(x^0, F)$.

5.2. Application to Randomized Subspace Descent Method

In this section, we apply our framework to the randomized subspace descent (RSD) [12, 20, 52], an algorithm for solving the smooth unconstrained optimization problem (6). To begin, let V denote the Euclidean space $\mathbb{R}^d$ equipped with the inner product $\langle \cdot, \cdot \rangle_V$ with the decomposition

$$V = V_1 + V_2 + \cdots + V_m.$$

In particular, we do not assume the subspaces are mutually orthogonal; in fact, we do not even require the sum to be a direct sum. We define the embedding operator $I_i : V_i \to V$ by $I_i v = v$ for all $v \in V_i$, and the restriction operator $R_i : V^* \to V_i^*$ by

$$R_i(v^*)(v_i) = v^*(I_i v_i) \quad \forall v^* \in V^*,\ v_i \in V_i.$$

The restriction operators allow us to define restricted gradients. Specifically, let $\langle \cdot, \cdot \rangle_{V_i}$ be an inner product defined on V_i, and let $df(x)$ denote the differential of the function f at a point $x \in \mathbb{R}^d$. The restricted gradient in the subspace V_i with respect to the inner product $\langle \cdot, \cdot \rangle_{V_i}$, denoted by $\nabla_i f(x)$, is defined through

$$R_i(df(x))(h) = \langle \nabla_i f(x), h \rangle_{V_i} \quad \forall h \in V_i. \tag{28}$$

Each iteration of RSD then takes the form

$$x^{k+1} = x^k - \beta_k I_{i_k} \nabla_{i_k} f(x),$$

where i_k is randomly sampled from $[m]$ and $\beta_k > 0$. Notably, by taking each subspace V_i as the span of a subset of standard coordinate bases and equipping both V and V_i with the standard Euclidean inner product, we see that RSD subsumes randomized block coordinate descent (RBCD), which in turn subsumes RCD. More interestingly, by choosing different inner products, we can obtain a generalized RBCD featuring a fixed preconditioner in each subspace. The following proposition is a standard result on RSD (see, *e.g.*, [12, 20, 52]). We provide a proof here for completeness.

Proposition 6. There exist positive constants γ_i, Γ_i and Λ such that

$$\gamma_i \|I_i h\|_V^2 \leq \|h\|_{V_i}^2 \leq \Gamma_i \|I_i h\|_V^2 \quad \text{for all } h \in V_i, \tag{29}$$

and

$$\sum_{i=1}^{m} \|I_i \nabla_i f(x)\|_V^2 \geq \Lambda \|\nabla f(x)\|_V^2. \tag{30}$$

Proof. Inequality (29) follows directly from the equivalence of norms in a finite-dimensional space.

To prove (30), consider the product space $\tilde{V} = V_1 \times V_2 \times \cdots \times V_m$ equipped with the norm $\|(v_1, \ldots, v_m)\|_{\tilde{V}}^2 = \sum_{i=1}^m \|v_i\|_{V_i}^2$. Define the linear mapping $T : \tilde{V} \to V$ by $T(v_1, v_2, \ldots, v_m) = \sum_{i=1}^m I_i v_i$. Because $V = V_1 + V_2 + \cdots + V_m$, the mapping T is surjective. By the open mapping theorem, there exists a constant $\sigma > 0$ such that for any $v \in V$, there exists a tuple $(v_1, v_2, \ldots, v_m) \in \tilde{V}$ satisfying $v = \sum_{i=1}^m I_i v_i$ and

$$\sum_{i=1}^{m} \|v_i\|_{V_i}^2 \leq \sigma^2 \|v\|_V^2.$$

Using this decomposition, the definition of the restricted gradient and Cauchy-Schwarz inequality, we can bound the directional derivative:

$$\begin{aligned}
|df(x)(v)| &= \left| \sum_{i=1}^{m} R_i(df(x))(v_i) \right| = \left| \sum_{i=1}^{m} \langle \nabla_i f(x), v_i \rangle_{V_i} \right| \\
&\leq \sum_{i=1}^{m} \|\nabla_i f(x)\|_{V_i} \|v_i\|_{V_i} \\
&\leq \left(\sum_{i=1}^{m} \|\nabla_i f(x)\|_{V_i}^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{m} \|v_i\|_{V_i}^2 \right)^{\frac{1}{2}} \\
&\leq \sigma \left(\sum_{i=1}^{m} \|\nabla_i f(x)\|_{V_i}^2 \right)^{\frac{1}{2}} \|v\|_V.
\end{aligned}$$

Taking the supremum over all $v \neq 0$ implies that

$$\|\nabla f(x)\|_V \leq \sigma \left(\sum_{i=1}^{m} \|\nabla_i f(x)\|_{V_i}^2 \right)^{\frac{1}{2}}.$$

Squaring both sides and applying the upper bound Γ_i from (29), we obtain

$$\|\nabla f(x)\|_V^2 \le \sigma^2 \sum_{i=1}^m \|\nabla_i f(x)\|_{V_i}^2 \le \sigma^2 \left(\max_i \Gamma_i\right) \sum_{i=1}^m \|I_i \nabla_i f(x)\|_V^2.$$

Thus, (30) follows by setting $\Lambda = \frac{1}{\sigma^2 \max_i \Gamma_i}$. This completes the proof. □

To facilitate the convergence analysis of RSD, we make the following smoothness assumption regarding f.

Assumption 5. The objective function $f \in C^1(\mathbb{R}^d)$ has a locally Lipschitz continuous gradient. For each $i \in [m]$, the restricted gradient $\nabla_i f$ is L_i-Lipschitz continuous with respect to V_i, *i.e.*,

$$\|\nabla_i f(x + I_i h) - \nabla_i f(x)\|_{V_i} \le L_i \|h\|_{V_i} \quad \forall x \in V,\ h \in V_i.$$

To guarantee a sufficient descent, a standard requirement on the stepsize is $\beta_k \in (0, \frac{1}{L_{i_k}}]$. However, this imposes a stochastic upper bound on β_k, which is incompatible with our framework since Assumption 1 requires a deterministic stepsize schedule. To resolve this discrepancy, we normalize the restricted gradient. Each iteration of RSD then reads

$$x^{k+1} = x^k - \beta_k \frac{I_{i_k} \nabla_{i_k} f(x)}{L_{i_k}}. \tag{31}$$

Under this normalized formulation, only the deterministic condition $\beta_k \in (0, 1]$ is needed.

Proposition 7. Consider RSD (31) for solving the unconstrained optimization problem (6) with an arbitrary initial point $x^0 \in \mathbb{R}^d$ and $\beta_k \in (0, 1]$ for all $k \ge 0$. Suppose that Assumption 5 holds and that there exists $\rho > 0$ such that $\mathbb{P}(i_k = i \mid \mathcal{F}_k) \ge \rho$ for all $i \in [m]$ and $k \ge 0$. Take $h(x) = f(x)$, $g(x) = \|\nabla f(x)\|^2$, and $S = \{x \in \mathbb{R}^d : f(x) \le f(x^0)\}$. Then Assumption 1 holds with $c_1 = \frac{\min_i \gamma_i L_i}{2}$, $c_2 = \frac{\rho \Lambda}{\max_i L_i^2}$, $c_3 = \frac{1}{\min_i L_i^2 \gamma_i^2}$ and $\alpha_k = \beta_k$.

Proof. Define $d_i(x) = \frac{\nabla_i f(x)}{L_i}$. For any $\alpha \le 1$, we have

$$\begin{aligned}
& f(x - \alpha I_i d_i(x)) - f(x) - df(x)(-\alpha I_i d_i(x)) \\
= & \int_0^1 [df(x - t\alpha I_i d_i(x)) - df(x)](-\alpha I_i d_i(x)) dt \\
= & \int_0^1 [R_i df(x - t\alpha I_i d_i(x)) - R_i df(x)](-\alpha d_i(x)) dt \\
\le & \int_0^1 \langle \nabla_i f(x - t\alpha I_i d_i(x)) - \nabla_i f(x), -\alpha d_i(x) \rangle_{V_i} dt \\
\le & \frac{L_i \alpha^2}{2} \|d_i(x)\|_{V_i}^2.
\end{aligned}$$

Note that $df(x)(-\alpha I_i d_i(x)) = \langle \nabla_i f(x), -\alpha d_i(x)\rangle_{V_i} = -L_i\alpha\|d_i(x)\|_{V_i}^2$. Thus, if $\beta_k \le 1$, the update step $x^{k+1} = x^k - \beta_k I_{i_k} d_{i_k}(x^k)$ yields

$$\begin{aligned} f(x^{k+1}) \le&\ f(x^k) - (1 - \frac{\beta_k}{2})\beta_k L_{i_k}\|d_{i_k}(x)\|_{V_{i_k}}^2 \\ \le&\ f(x^k) - \frac{L_{i_k}}{2\beta_k}\|\beta_k d_{i_k}(x^k)\|_{V_{i_k}}^2 \\ \le&\ f(x^k) - \frac{\min_i \gamma_i L_i}{2\beta_k}\|x^{k+1} - x^k\|_V^2, \end{aligned}$$

which verifies inequality (1). Next, taking the conditional expectation gives

$$\begin{aligned} \mathbb{E}[\|x^{k+1} - x^k\|_V^2 \mid \mathcal{F}_k] \ge&\ \rho\beta_k^2 \sum_{i=1}^m \frac{1}{L_i^2}\|I_i \nabla_i f(x)\|_V^2 \\ \ge&\ \frac{\rho\Lambda}{\max_i L_i^2}\beta_k^2\|\nabla f(x)\|_V^2. \end{aligned}$$

which verifies inequality (2). Finally, taking $h = \nabla_i f(x)$ in (28), for any x, we have

$$\|\nabla_i f(x)\|_{V_i}^2 = df(x)(I_i\nabla_i f(x)) = \langle \nabla f(x), I_i \nabla_i f(x)\rangle_V \le \|\nabla f(x)\|_V \|I_i \nabla_i f(x)\|_V.$$

By (29), we have $\|\nabla_i f(x)\|_{V_i}^2 \ge \gamma_i \|I_i \nabla_i f(x)\|_V^2$, which implies

$$\|I_i \nabla_i f(x)\|_V \le \frac{1}{\gamma_i}\|\nabla f(x)\|_V.$$

Thus,

$$\|x^{k+1} - x^k\|_V^2 = \frac{\beta_k^2}{L_{i_k}^2}\|I_{i_k}\nabla_{i_k} f(x)\|_V^2 \le \frac{1}{\min_i L_i^2\gamma_i^2}\beta_k^2\|\nabla f(x)\|_V^2,$$

which verifies inequality (3). This completes the verification of condition (i). Condition (ii) follows directly from the descent property of RSD and the definition of S. □

We then analyze the convergence rate of RSD. However, the analysis under global error bounds closely mirrors that in Section 5.1 and is therefore omitted. Instead, we focus on its convergence analysis under local error bounds. To this end, we first recall the definition of the KL property.

Definition 5. A smooth function f is said to satisfy the Kurdyka–Łojasiewicz (KL) property at a point $\bar{x} \in \mathbb{R}^d$ if there exist constants $\delta_{\bar{x}}, \eta_{\bar{x}} \in (0, \infty]$, and a continuous, concave function $\varphi_{\bar{x}} : [0, \eta_{\bar{x}}) \to [0, \infty)$ that is continuously differentiable on $(0, \eta_{\bar{x}})$ with $\varphi_{\bar{x}}(0) = 0$ and $\varphi'_{\bar{x}} > 0$ on $(0, \eta_{\bar{x}})$, such that for all $x \in \mathbb{B}(\bar{x}, \delta_{\bar{x}})$ satisfying $0 < f(x) - f(\bar{x}) < \eta_{\bar{x}}$, the following inequality holds:

$$\varphi'_{\bar{x}}(f(x) - f(\bar{x}))\|\nabla f(x)\| \ge 1.$$

The function $\varphi_{\bar{x}}$ is called a desingularization function. Furthermore, we say that f satisfies the KL property at $\bar{x}$ with exponent κ if the desingularization function can be chosen as $\varphi_{\bar{x}}(t) = \bar{c}t^{1-\kappa}$ for some constant $\bar{c} > 0$.

A very broad class of functions, known as definable functions, satisfies the KL property. Furthermore, its subclass of subanalytic functions satisfies the KL property with an explicit exponent. Further details on definable and subanalytic functions can be found in [1, 2, 7, 8, 23]. The following proposition asserts the validity of Assumption 4 under the KL condition, whose proof is trivial and therefore omitted.

Proposition 8. Consider the unconstrained minimization problem (6). Suppose that the objective function f is definable. Let h, g, and S be defined as in Proposition 7. Then, Assumption 4 is satisfied with $\phi_{\bar{x}} = 1/(\varphi'_{\bar{x}})^2$ for any critical point $\bar{x}$ of f in S.

To proceed, we define the inverse smoothing function $\psi_{\bar{x}} : (0, \eta_{\bar{x}}) \to \mathbb{R}_+$ for $1/(\varphi'_{\bar{x}})^2$ as

$$\psi_{\bar{x}}(t) = \int_t^{t_0} (\varphi'_{\bar{x}}(s))^2 ds,$$

where $t_0 \in (0, \eta_{\bar{x}})$. The following theorem asserts that $(\varphi'_{\bar{x}}(s))^2$ is not integrable near 0.

Theorem 6. Suppose $f \in C^1(\mathbb{R}^d)$ is a definable function with a locally Lipschitz continuous gradient ∇f and that $\bar{x}$ is a critical point of f that is not a local maximizer. If f satisfies the KL property at $\bar{x}$ with a desingularization function φ, then $(\varphi')^2$ is not integrable near 0. In particular, f cannot satisfy the KL property with an exponent $\kappa \in [0, 1/2)$.

Proof. Without loss of generality, we assume $\bar{x} = 0$ and $f(0) = 0$. Assume the contrary. Then there exist $\eta > 0$, $r > 0$, and a desingularization function $\varphi : [0, \eta) \to [0, \infty)$ such that $(\varphi')^2$ is integrable on $(0, \eta)$ and

$$\varphi'(f(x))\|\nabla f(x)\| \geq 1 \quad \forall x \in \mathbb{B}(0, r) \text{ with } 0 < f(x) < \eta. \tag{32}$$

By the assumption that 0 is not a local maximizer of f, we may select $x^k \to 0$ such that $f(x^k) > 0$. Without loss of generality, we may also assume $x^k \in \mathbb{B}(0, r)$ and $f(x^k) < \eta$ for all $k \in \mathbb{N}$. For each $t \in [0, r)$, we define the set

$$H(t) = \arg\max\{f(x) : x \in \overline{\mathbb{B}}(0, t)\}, \tag{33}$$

where $\overline{\mathbb{B}}(0, t)$ closed ball of radius t centered at the origin. Let ι_C denote the indicator function in the sense of convex analysis (*i.e.*, $\iota_C(x) = 0$ if $x \in C$ and $\iota_C(x) = +\infty$ otherwise), and define

$$W(x, t) = -f(x) + \iota_{\|x\| \leq t} \quad \text{and} \quad Q(t) = \inf_{x \in \mathbb{R}^d} W(x, t).$$

Then, W is definable by [23, Remark 2], Q is definable by [23, Lemma 1], and $\mathrm{gph}(H) = (W - Q)^{-1}(0)$ is definable by [48, Chapter 1, Lemma 2.3(iii)]. Hence, we can use [49, 4.5] to show that we are able to select a definable curve $z : [0, r) \to \mathbb{R}^d$ with $z(t) \in H(t)$ for all $t \in [0, r)$. Next, we claim that for all $y \in H(t)$ it holds that $\|y\| = t$. Otherwise, we have $y \in \operatorname{int} \bar{\mathbb{B}}(0, t)$, and by the first-order necessary condition of maximization, we would have $\nabla f(y) = 0$. Then, either $f(y) > 0$, which the contradicts (32) (since $\varphi'(s)$ is finite for $s > 0$); or $f(y) = 0$, which contradicts local maximality of 0. Moreover, by the optimality condition of (33), we have that

$$\forall y \in H(t),\ \exists a > 0 \text{ s.t. } \nabla f(y) = ay. \tag{34}$$

Using the monotonicity lemma [49, 4.1] on each component of z and reducing r if necessary, we may assume that $z \in C^2((0,r))$. Notice also that z is continuous at 0 by the definition of H. By (34), we know there exists $a(t)$ such that

$$\forall t \in [0,r),\ \nabla f(z(t)) = a(t)z(t), \quad \forall t \in (0,r),\ a(t) > 0, \tag{35}$$

where $a(0)$ can be arbitrary since $z(0) = \nabla f(0) = 0$. Also, by (33) and the fact that $\|z(t)\|^2 = t^2$ for all $t \in [0,r)$, we have

$$\langle z(t), z'(t)\rangle = t \quad \text{and} \quad \langle \frac{z(t)}{\|z(t)\|}, z'(t)\rangle = 1 \quad \forall t \in (0,r).$$

Then, by using (35), we get

$$\langle \frac{\nabla f(z(t))}{\|\nabla f(z(t))\|},\ z'(t)\rangle = 1. \tag{36}$$

Define the function $g : [0,r) \to \mathbb{R}$ as

$$g(t) = \varphi(f(z(t))).$$

Since $f(z(t)) > 0$ for all $t \in (0,r)$ and $z \in C^2(0,r)$, we see that $g \in C^2(0,r)$ and for any $t \in (0,r)$,

$$\begin{aligned} g'(t) &= \varphi'(f(z(t)))\langle \nabla f(z(t)),\ z'(t)\rangle \\ &\overset{(a)}{=} \varphi'(f(z(t)))\|\nabla f(z(t))\| \\ &\overset{(b)}{\geq} 1, \end{aligned}$$

where for (a) we have used (36), and for (b) we have used (32). Since z is continuous at 0, we know $f \circ z$ and g are both continuous at 0. Integrating $g'(t) \geq 1$ from 0 to t and using $g(0) = \varphi(f(0)) = \varphi(0) = 0$, we obtain

$$\varphi(f(z(t))) \geq t \quad \forall t \in [0,r). \tag{37}$$

On the other hand, since ∇f is locally Lipschitz continuous, we may assume ∇f is Lipschitz continuous on $\mathbb{B}(0,r)$ with a modulus L. By the descent lemma and $\nabla f(0) = 0$, we have

$$f(z(t)) \leq f(0) + \frac{L}{2}\|z(t)\|^2 = \frac{L}{2}t^2,$$

which further implies

$$t \geq \sqrt{\frac{2}{L} f(z(t))}. \tag{38}$$

Combining (37) and (38), we deduce that for all $t \in (0, r)$,

$$\varphi(f(z(t))) \geq \sqrt{\frac{2}{L}} \sqrt{f(z(t))}. \tag{39}$$

However, by the Cauchy-Schwarz inequality, for any $s \in (0, \eta)$,

$$\varphi(s) = \int_0^s \varphi'(u) \mathrm{d}u \leq \left(\int_0^s 1^2 \mathrm{d}u \right)^{1/2} \left(\int_0^s (\varphi'(u))^2 \mathrm{d}u \right)^{1/2} = \sqrt{s} \left(\int_0^s (\varphi'(u))^2 \mathrm{d}u \right)^{1/2}.$$

Substituting $s = f(z(t))$ into this inequality and dividing by $\sqrt{f(z(t))}$, we get

$$\int_0^{f(z(t))} (\varphi'(u))^2 \mathrm{d}u \geq \frac{(\varphi(f(z(t))))^2}{f(z(t))} \geq \frac{2}{L}.$$

As $t \to 0$, we have $f(z(t)) \to 0$. Because $(\varphi')^2$ is assumed to be integrable on $(0, \eta)$, the integral $\int_0^s (\varphi'(u))^2 \mathrm{d}u$ must converge to 0 as $s \to 0$. However, the last display inequality implies that it is strictly below $2/L > 0$, which leads to a contradiction. Thus, $(\varphi')^2$ cannot be integrable near 0. In particular, when $\varphi_{\bar{x}}(t) = \bar{c} t^{1-\kappa}$ for some constants $\bar{c} > 0$ and $\kappa \in [0, 1)$, the non-integrability of $(\varphi')^2$ near 0 implies that $\kappa \in [1/2, 1)$. □

It is proved in [6, Theorem 7] that the sum of a globally Lipschitz smooth function and a convex function cannot have a KL exponent $\kappa \in (0, \frac{1}{2}]$ at a local minimizer in the interior of its domain. Our Theorem 6 does not require such a sum structure but considers a general locally Lipschitz smooth function. It is advantageous in two senses: it covers general desingularization functions and does not require the target point to be a local minimizer. The former advantage allows us to work with general definable functions but not just semi-algebraic or subanalytic functions; while the latter accommodates nonconvex functions, where limit points are not guaranteed to be local minimizers.

Theorem 7. Consider RSD (31) for solving the unconstrained minimization problem (6) with an arbitrary initial point $x^0 \in \mathbb{R}^d$ and stepsizes $\beta_k \in (0, 1]$ satisfying Assumption 2. Suppose that the objective function f is coercive, definable, and satisfies Assumption 5. Then there exists a random variable x^* such that $x^* \in F$ and $x^k \to x^*$ almost surely. Furthermore, there exists an almost surely finite random index J_5 such that for all $k \geq J_5$,

$$f(x^k) - f(x^*) \leq \psi_{x^*}^{-1}\left(\frac{c_1 c_2 p}{3} A_k \right)$$

and

$$\|x^k - x^*\| \leq \frac{4}{c_1 \sqrt{c_2 p}} \varphi_{x^*}\left(\psi_{x^*}^{-1}\left(\frac{c_1 c_2 p}{4} A_k \right) \right),$$

where $A_k = \sum_{i=0}^{k-1} \alpha_i$, $c_1 = \frac{\min_i \gamma_i L_i}{2}$, $c_2 = \frac{\rho \Lambda}{\max_i L_i^2}$ and $p = \frac{\rho}{2-\rho}$.

Proof. By Proposition 7, Assumption 1 is satisfied with $c_1 = \frac{\min_i \gamma_i L_i}{2}$, $c_2 = \frac{\rho \Lambda}{\max_i L_i^2}$, and $p = \frac{\rho}{2-\rho}$. Furthermore, the coercivity of f and the descent property of RSD guarantee that the set S is compact. The desired result then follows from Corollary 1. □

Corollary 2. Suppose that the conditions of Theorem 7 hold and that f is a subanalytic function. Let $A_k = \sum_{i=0}^{k-1} \alpha_i$. Then, for the event E_0 defined in (26), we have $P(E_0) = 1$. Furthermore, for any $\omega \in E_0$, the function f satisfies the KL property at $x^*(\omega)$ with an exponent $\kappa \in [1/2, 1)$, and there exist constants $C_1, \ldots, C_6 > 0$ and a finite index $J_5(\omega)$ such that the following hold.

(i) If $\kappa = 1/2$, we have that for all $k \geq J_5(\omega)$,

$$f(x^k(\omega)) - f(x^*(\omega)) \leq C_1 \exp\left(-C_2 A_k\right) \quad \text{and} \quad \|x^k(\omega) - x^*(\omega)\| = C_3 \exp\left(-C_4 A_k\right).$$

(ii) If $\kappa \in (1/2, 1)$, we have that for all $k \geq J_5(\omega)$,

$$f(x^k(\omega)) - f(x^*(\omega)) \leq C_5\, A_k^{-\frac{1}{2\kappa-1}} \quad \text{and} \quad \|x^k(\omega) - x^*(\omega)\| \leq C_6\, A_k^{-\frac{1-\kappa}{2\kappa-1}}.$$

Proof. Because f is subanalytic, it is guaranteed to satisfy the KL property at the limit point $x^*(\omega) \in F$ with some exponent $\kappa \in [0, 1)$. It follows from Theorem 6 that $\kappa \geq 1/2$. Thus, we have $\kappa \in [1/2, 1)$, and the desingularization function takes the form $\varphi_{x^*(\omega)}(t) = \bar{c}t^{1-\kappa}$ for some constant $\bar{c} > 0$. The derivative is given by $\varphi'_{x^*(\omega)}(t) = \bar{c}(1-\kappa)t^{-\kappa}$. Let $r = \bar{c}(1-\kappa)$. The inverse smoothing function $\psi_{x^*(\omega)}(t)$ is defined as

$$\psi_{x^*(\omega)}(t) = \int_t^{t_0} (\varphi'_{x^*(\omega)}(s))^2 ds = \int_t^{t_0} r^2 s^{-2\kappa} ds. \tag{40}$$

To prove assertion (i), evaluating the integral (40), we obtain $\psi_{x^*(\omega)}(t) = r^2(\ln t_0 - \ln t)$. Inverting this function yields

$$\psi^{-1}_{x^*(\omega)}(y) = t_0 \exp\left(-\frac{y}{r^2}\right).$$

Substituting $y = \frac{c_1 c_2 p}{3} A_k$ into the function value bound of Theorem 7, we obtain

$$f(x^k(\omega)) - f(x^*(\omega)) \leq t_0 \exp\left(-\frac{c_1 c_2 p}{3r^2} A_k\right) = C_1 \exp\left(-C_2 A_k\right),$$

for some constants $C_1, C_2 > 0$. Next, substituting $y = \frac{c_1 c_2 p}{4} A_k$ into the sequence bound and using $\varphi_{x^*(\omega)}(t) = \bar{c}t^{1/2}$, we have

$$\|x^k(\omega) - x^*(\omega)\| \leq \frac{4\bar{c}}{c_1\sqrt{c_2 p}} \left(t_0 \exp\left(-\frac{c_1 c_2 p}{4r^2} A_k\right)\right)^{1/2} = C_3 \exp\left(-C_4 A_k\right),$$

for some constants $C_3, C_4 > 0$.

We next prove assertion (ii). Evaluating the integral (40) yields

$$\psi_{x^*(\omega)}(t) = \frac{r^2}{2\kappa - 1} \left(t^{-(2\kappa-1)} - t_0^{-(2\kappa-1)}\right).$$

Inverting this function gives

$$\psi_{x^*(\omega)}^{-1}(y) = \left(\frac{2\kappa-1}{r^2}y + t_0^{-(2\kappa-1)}\right)^{-\frac{1}{2\kappa-1}}.$$

Substituting $y = \frac{c_1 c_2 p}{3} A_k$ into Theorem 7, we immediately deduce that

$$f(x^k(\omega)) - f(x^*(\omega)) \le C_5\, A_k^{-\frac{1}{2\kappa-1}},$$

for some $C_5 > 0$. Similarly, for the distance bound, we evaluate $\varphi_{x^*(\omega)}(\psi_{x^*(\omega)}^{-1}(y))$:

$$\varphi_{x^*(\omega)}(\psi_{x^*(\omega)}^{-1}(y)) = \bar{c}\left(\frac{2\kappa-1}{r^2}y + t_0^{-(2\kappa-1)}\right)^{-\frac{1-\kappa}{2\kappa-1}}.$$

Substituting $y = \frac{c_1 c_2 p}{4} A_k$, we deduce that

$$\|x^k(\omega) - x^*(\omega)\| \le C_6\, A_k^{-\frac{1-\kappa}{2\kappa-1}},$$

for some $C_6 > 0$. □

Note that the constants $C_1, \ldots, C_6$ in Corollary 2 depend only on c_1, c_2, p and $x^*(\omega)$.

Acknowledgment

This work is supported in part by the Hong Kong Research Grants Council under the GRF project 17310124.

References

1. P. A. Absil, R. Mahony, and B. Andrews. Convergence of the iterates of descent methods for analytic cost functions. *SIAM Journal on Optimization*, 16(2):531–547, 2005.

2. H. Attouch, J. Bolte, P. Redont, and A. Soubeyran. Proximal alternating minimization and projection methods for nonconvex problems: An approach based on the Kurdyka-Łojasiewicz inequality. *Mathematics of Operations Research*, 35(2):438–457, 2010.

3. H. H. Bauschke and J. M. Borwein. On projection algorithms for solving convex feasibility problems. *SIAM Review*, 38(3):367–426, 1996.

4. H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2017.

5. A. Beck and L. Tetruashvili. On the convergence of block coordinate descent type methods. *SIAM Journal on Optimization*, 23(4):2037–2060, 2013.

6. G. Bento, B. Mordukhovich, T. Mota, and Yu. Nesterov. Convergence of descent methods under Kurdyka-Łojasiewicz properties. *arXiv preprint arXiv:2407.00812*, 2025.

7. J. Bolte, A. Daniilidis, A. Lewis, and M. Shiota. Clarke subgradients of stratifiable functions. *SIAM Journal on Optimization*, 18(2):556–572, 2007.

8. J. Bolte, S. Sabach, and M. Teboulle. Proximal alternating linearized minimization for nonconvex and nonsmooth problems. *Mathematical Programming*, 146:459 – 494, 2013.

9. J. M. Borwein, G. Li, and M. K. Tam. Convergence rate analysis for averaged fixed point iterations in common fixed point problems. *SIAM Journal on Optimization*, 27(1):1–33, 2017.

10. J. M. Borwein, G. Li, and L. Yao. Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets. *SIAM Journal on Optimization*, 24(1):498–527, 2014.

11. X. Cai, C. Song, S. Wright, and J. Diakonikolas. Cyclic block coordinate descent with variance reduction for composite nonconvex optimization. In *Proceedings of the 40th International Conference on Machine Learning*, pages 3469–3494. PMLR, 2023.

12. L. Chen, X. Hu, and H. Wu. Randomized fast subspace descent methods, 2020.

13. F. Chorobura and I. Necoara. Random coordinate descent methods for nonseparable composite optimization. *SIAM Journal on Optimization*, 33(3):2160–2190, 2023.

14. E. Chouzenoux, J.-C. Pesquet, and A. Repetti. A block coordinate variable metric forward-backward algorithm. *Journal of Global Optimization*, 66(3):457–485, 2016.

15. D. Egoroff. Sur les suites de fonctions mesurables. *Comptes Rendus de l'Académie des Sciences*, 152:244–246, 1911.

16. I. Fatkhullin, J. Etesami, N. He, and N. Kiyavash. Sharp analysis of stochastic optimization under global Kurdyka-Łojasiewicz inequality. In *Advances in Neural Information Processing Systems*, volume 35, pages 15836–15848, 2022.

17. J.-B. Fest, A. Repetti, and E. Chouzenoux. A stochastic use of the Kurdyka-Łojasiewicz property: Investigation of optimization algorithms behaviours in a non-convex differentiable framework. *Foundations of Data Science*, 9(0):164–191, 2026.

18. X. Fontaine, V. De Bortoli, and A. Durmus. Convergence rates and approximation results for SGD and its continuous-time counterpart. In *Conference on Learning Theory*, pages 1965–2058. PMLR, 2021.

19. P. Hall and C. C. Heyde. *Martingale limit theory and its application*. Academic press, 2014.

20. B. Jiang, J. Park, and J. Xu. Randomized subspace correction methods for convex optimization. *arXiv preprint arXiv:2507.01415*, 2025.

21. H. D. Junghenn. *Principles of Analysis: Measure, Integration, Functional Analysis, and Applications*. Chapman and Hall/CRC, 2018.

22. H. Karimi, J. Nutini, and M. Schmidt. Linear convergence of gradient and proximal-gradient methods under the Polyak-Łojasiewicz condition. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, pages 795–811. Springer, 2016.

23. K. Kurdyka. On gradients of functions definable in o-minimal structures. *Annales de l'Institut Fourier*, 48:769–783, 1998.

24. P. Latafat, A. Themelis, M. Ahookhosh, and P. Patrinos. Bregman Finito/MISO for nonconvex regularized finite sum minimization without Lipschitz gradient continuity. *SIAM Journal on Optimization*, 32(3):2230–2262, 2022.

25. P. Latafat, A. Themelis, and P. Patrinos. Block-coordinate and incremental aggregated proximal gradient methods for nonsmooth nonconvex problems. *Mathematical Programming*, 193(1):195–224, 2022.

26. X. Li and W. Bian. Smoothing randomized block-coordinate proximal gradient algorithms for nonsmooth nonconvex composite optimization. *Numerical Algorithms*, 100(1):395–424, 2025.

27. T. M. Liggett, R. H. Schonmann, and A. M. Stacey. Domination by product measures. *The Annals of Probability*, 25(1):71–95, 1997.

28. T. Liu and B. F. Lourenço. Concrete convergence rates for common fixed point problems under Karamata regularity. *arXiv preprint arXiv:2407.13234*, 2024.

29. T. Liu and B. F. Lourenço. Convergence analysis under consistent error bounds. *Foundations of Computational Mathematics*, 24(2):429–479, 2024.

30. S. Łojasiewicz. Sur le problème de la division. *Studia Mathematica*, 18:87–136, 1959.

31. S. Łojasiewicz. Une propriété topologique des sous-ensembles analytiques réels. *Les équations aux dérivées partielles*, 117:87–89, 1963.

32. A. Louzi. Stochastic gradient descent revisited. *arXiv preprint arXiv:2412.06070*, 2024.

33. Z. Lu and L. Xiao. On the complexity analysis of randomized block-coordinate descent methods. *Mathematical Programming*, 152(1-2):615–642, 2015.

34. Z.-Q. Luo and P. Tseng. Error bounds and convergence analysis of feasible descent methods: A general approach. *Annals of Operations Research*, 46–47(1):157–178, 1993.

35. I. Necoara, P. Richtárik, and A. Patrascu. Randomized projection methods for convex feasibility: Conditioning and convergence rates. *SIAM Journal on Optimization*, 29(4):2814–2852, 2019.

36. A. Nedić. Random projection algorithms for convex set intersection problems. In *49th IEEE Conference on Decision and Control (CDC)*, pages 7655–7660, 2010.

37. Yu. Nesterov. Efficiency of coordinate descent methods on huge-scale optimization problems. *SIAM Journal on Optimization*, 22(2):341–362, 2012.

38. B. T. Polyak. Gradient methods for the minimisation of functionals. *USSR Computational Mathematics and Mathematical Physics*, 3(4):864–878, 1963.

39. A. Pătraşcu and I. Necoara. Efficient random coordinate descent algorithms for large-scale structured nonconvex optimization. *Journal of Global Optimization*, 61:19 – 46, 2013.

40. J. Qiu, B. Ma, X. Li, and A. Milzarek. A KL-based analysis framework with applications to non-descent optimization methods. *arXiv preprint arXiv:2406.02273*, 2024.

41. P. Richtárik and M. Takáč. Iteration complexity of randomized block-coordinate descent methods for minimizing a composite function. *Mathematical Programming*, 144(1-2):1–38, 2014.

42. M. Schmidt and N. L. Roux. Fast convergence of stochastic gradient descent under a strong growth condition. *arXiv preprint arXiv:1308.6370*, 2013.

43. A. N. Shiryaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer, second edition, 1996.

44. A. N. Shiryaev. *Probability-1*, volume 95. Springer, 2016.

45. T. Strohmer and R. Vershynin. A randomized Kaczmarz algorithm with exponential convergence. *Journal of Fourier Analysis and Applications*, 15(2):262–278, 2009.

46. P. Tseng and S. Yun. Block-coordinate gradient descent method for linearly constrained nonsmooth separable optimization. *Journal of Optimization Theory and Applications*, 140(3):513–535, 2009.

47. P. Tseng and S. Yun. A coordinate gradient descent method for linearly constrained smooth optimization and support vector machines training. *Computational Optimization and Applications*, 47(2):179–206, 2010.

48. L. van den Dries. *Tame topology and o-minimal structures*, volume 248. Cambridge University Press, 1998.

49. L. van den Dries and C. Miller. Geometric categories and o-minimal structures. *Duke Mathematical Journal*, 84:497–540, 1996.

50. S. J. Wright. Coordinate descent algorithms. *Mathematical Programming*, 151:3 – 34, 2015.

51. L. Wu, Y. Qi, and J. Yang. Asymptotics for dependent Bernoulli random variables. *Statistics & Probability Letters*, 82(3):455–463, 2012.

52. J. Xu. Iterative methods by space decomposition and subspace correction. *SIAM Review*, 34(4):581–613, 1992.

53. Y. Xu and W. Yin. A block coordinate descent method for regularized multiconvex optimization with applications to nonnegative tensor factorization and completion. *SIAM Journal on Imaging Sciences*, 6:1758–1789, 2013.

54. Y. Xu and W. Yin. A globally convergent algorithm for nonconvex optimization based on block coordinate update. *Journal of Scientific Computing*, 72(2):700–734, 2017.