

SCIENTIFIC ARTICLE

© R. A. Khachatryan, 2026

<https://doi.org/10.20310/2686-9667-2026-31-154-168-185>

Stochastic finite-difference schemes for nondifferentiable quasiconvex optimization: convergence rate

Rafik A. KHACHATRYAN

Yerevan State University

1 Alex Manukyan St., Yerevan 0025, Armenia

Abstract. This paper investigates stochastic approximation methods based on finite differences for the minimization of quasiconvex functions. Traditional approaches to convex optimization problems primarily rely on exact or stochastic subgradients. However, in many practical situations, only noisy information about function values is available. In such cases, stochastic quasi-gradient schemes constructed using randomized finite-difference estimators are considered. In particular, we study batch two-point schemes that provide unbiased approximations of the gradient of a smoothed objective function. Variance bounds are established for these estimators, which enable a rigorous convergence analysis of projected stochastic descent methods.

The paper considers the minimization problem of functions of the form $f(x) = \max_{i \in I} f_i(x)$, where each function $f_i(x)$, $i \in I$, is quasiconvex and has a Lipschitz-continuous gradient on a convex compact set. The main result of the paper is the derivation of convergence rate estimates for stochastic finite-difference methods in the quasiconvex setting. It is shown that the expected suboptimality decreases at the rate $O(1/\sqrt{k})$. The obtained results significantly broaden the applicability of stochastic finite-difference methods to nonsmooth quasiconvex optimization problems and provide a rigorous theoretical justification of the algorithm in black-box settings where only noisy function evaluations are available.

Keywords: optimization, quasi convex function, subgradient, finite-difference scheme, convergence rate

Mathematics Subject Classification: 65K10, 49N30, 90C15.

For citation: Khachatryan R.A. Stochastic finite-difference schemes for nondifferentiable quasiconvex optimization: convergence rate. *Vestnik Rossiyskikh Universitetov. Matematika = Russian Universities Reports. Mathematics*, **31**:154 (2026), 168–185. <https://doi.org/10.20310/2686-9667-2026-31-154-168-185>

НАУЧНАЯ СТАТЬЯ

© Хачатрян Р.А., 2026

<https://doi.org/10.20310/2686-9667-2026-31-154-168-185>

УДК 517.97, 517.988.38, 519.6



Стохастические конечно-разностные схемы для недифференцируемой квазивыпуклой оптимизации: скорость сходимости

Рафик Агасиевич ХАЧАТРЯН

Ереванский государственный университет

0025, Армения, г. Ереван, ул. Алек Манукян, 1

Аннотация. В данной работе исследуются методы стохастического приближения на основе конечных разностей для задачи минимизации квазивыпуклых функций. Традиционные подходы к задачам выпуклой оптимизации в основном опираются на использование точных или стохастических субградиентов. Однако во многих практических ситуациях доступна только зашумленная информация о значениях функции. В таких случаях рассматриваются стохастические квазиградиентные схемы, построенные на основе рандомизированных конечно-разностных оценок. В частности, изучаются пакетные двухточечные схемы, которые дают несмещенные приближения градиента сглаженной целевой функции. Для данных схем устанавливаются оценки дисперсии, что позволяет провести анализ сходимости проективных стохастических методов спуска.

В работе рассматривается задача минимизации функций вида $f(x) = \max_{i \in I} f_i(x)$, где каждая функция $f_i(x)$, $i \in I$, является квазивыпуклой и имеет липшицев градиент на выпуклом компактном множестве. Основным результатом работы является получение оценок скорости сходимости стохастических конечно-разностных методов в случае квазивыпуклых функций. Показано, что ожидаемая субоптимальность убывает со скоростью $O\left(\frac{1}{\sqrt{k}}\right)$. Полученные результаты существенно расширяют область применения стохастических конечно-разностных методов к негладким задачам квазивыпуклой оптимизации и обеспечивают строгое теоретическое обоснование алгоритма в условиях «черного ящика», когда доступны только зашумленные оценки функции.

Ключевые слова: оптимизация, квазивыпуклая функция, субградиент, конечно-разностная схема, скорость сходимости

Для цитирования: Хачатрян Р. А. Стохастические конечно-разностные схемы для недифференцируемой квазивыпуклой оптимизации: скорость сходимости // Вестник российских университетов. Математика. 2026. Т. 31. № 154. С. 168–185.

<https://doi.org/10.20310/2686-9667-2026-31-154-168-185>

Introduction. Finite-difference stochastic approximation

The study of quasiconvex functions has been an active area of research in mathematical optimization for several decades [1–3]. A substantial body of work is devoted to the analysis and solution of quasiconvex optimization problems. In particular, [4] was the first to propose an iterative method, the *Normalized Gradient Descent (NGD) method*, and to establish that it attains an ϵ -approximate solution within a finite number of iterations for differentiable quasiconvex functions. Building on this, a modification of the NGD algorithm was introduced in [5], referred to as the *Approximation Adaptive Normalized Gradient Descent (AANGD) method*. In this approach, the normalized gradient is approximated, and descent is performed along the resulting direction. Notably, the same convergence rate estimate is achieved, showing that effective minimization can be carried out even without exact gradient information.

In large-scale optimization problems, exact gradients or subgradients are often unavailable or expensive to compute. In such settings, *stochastic gradients* are frequently employed and have been widely used for solving both convex and nonconvex optimization problems [6–8]. However, constructing reliable stochastic gradient estimators is itself a nontrivial task. As a practical alternative, gradients are often approximated using *finite-difference schemes*, which enable efficient derivative-free optimization while retaining favorable convergence properties. Finite-difference methods for the minimization of nonsmooth Lipschitz functions were initially proposed by A. M. Gupal in [9] and subsequently developed in [10]. Subsequently, finite-difference stochastic schemes with batches are employed in [11, 12] for the minimization of convex functions.

The following notation is used throughout the paper:

- (x, y) denotes the scalar product of vectors $x, y \in \mathbb{R}^n$;
- $\mathbb{E}[\xi]$ denotes the mathematical expectation (expected value) of a random variable ξ ;
- $\mathbb{E}[\xi | x]$ denotes the conditional expectation of ξ given x ;
- $\Pi_M(a)$ denotes the projection of a vector a onto a closed convex set M ;
- $f'(x)$ denotes the gradient of f at the point x ;
- $B_\delta(x)$ denotes the closed ball of radius δ centered at the point x .

Consider the general optimization problem:

$$f(x) \rightarrow \min, \quad x \in M. \quad (0.1)$$

When the function f is smooth and the set M is convex, the well-known projected gradient method is commonly used to solve this problem. In the non-smooth case, namely for locally Lipschitz functions, an efficient mechanism for solving this problem was developed in [10]. The justification of numerical minimization methods for locally Lipschitz functions is carried out as follows.

The sequence $\{x^k\}$ is generated by the following recurrence relation:

$$x^{k+1} = \Pi_M(x^k - \rho_k g^k), \quad (0.2)$$

where the vector g^k is chosen according to one of the following finite-difference stochastic approximations:

$$g^k = H_1(x^k, \alpha_k) = \frac{1}{2\alpha_k} \sum_{i=1}^n \left(f(\tilde{x}_1^k, \dots, x_i^k + \alpha_k, \dots, \tilde{x}_n^k) - f(\tilde{x}_1^k, \dots, x_i^k - \alpha_k, \dots, \tilde{x}_n^k) \right) e_i,$$

$$g^k = H_2(x^k, \alpha_k, \Delta_k) = \sum_{i=1}^n \frac{f(\tilde{x}^k + \Delta_k e_i) - f(\tilde{x}^k)}{\Delta_k} e_i.$$

Here, $\tilde{x}^k = (\tilde{x}_1^k, \dots, \tilde{x}_n^k)$, where the components $\tilde{x}_i^k$ are independent random variables uniformly distributed on the segment $[x_i^k - \alpha_k, x_i^k + \alpha_k]$, $\alpha_k > 0$. The vector g^k defined by these formulas is called a stochastic quasi-gradient of the function f .

Using these schemes, it is shown that the limit points of the generated sequence are stationary points of the function with probability one (see [10, Ch. 2, §4, Theorem 4.1]).

In recent years, the following finite-difference stochastic scheme for gradient or subgradient approximation has been widely used for convex functions:

$$H_3(x, \alpha, u) = n \frac{f(x + \alpha u) - f(x - \alpha u)}{2\alpha} u, \quad (0.3)$$

where u is a random vector uniformly distributed on the n -dimensional sphere or drawn from a standard Gaussian. This is referred to as the two-point randomized finite-difference scheme (see [11, 12]). By employing the aforementioned two-point finite-difference scheme, the work [11] investigate the sequence $\{x^k\}$ generated by formula (0.3) for a convex function f and provide quantitative estimates of its convergence rate.

Now consider the problem (0.1) in the case where the objective function f is quasiconvex and the set is convex and compact. For solving this problem, stochastic finite-difference batch schemes are applied, and it is proved that the generated sequence converges to a global minimizer of the problem with probability one. The rate of convergence is also estimated. Let us introduce the necessary definitions.

Definition 0.1 (see [1]). A function $f(x)$ is called quasiconvex on a convex set M if for all $x, y \in M$ and $\lambda \in [0, 1]$ we have $f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\}$.

Note that this definition is equivalent to requiring that the level sets of the function f be convex; that is, for every $\lambda \in \mathbb{R}$, the set $A_\lambda = \{x \in M : f(x) \leq \lambda\}$ is convex.

Definition 0.2 (see [1]). A function f is said to be strongly quasiconvex on the convex set M if there exists $\mu > 0$ such that for all $x, y \in M$ and all $\lambda \in [0, 1]$, we have

$$f(\lambda x + (1 - \lambda)y) \leq \max\{f(x), f(y)\} - \mu \lambda(1 - \lambda) \|x - y\|^2.$$

In the sequel, we will make the following assumptions:

- i) f is quasiconvex and L_f -Lipschitz continuous on the set

$$M + B_1(0) = \{x \in \mathbb{R}^n : \text{dist}(x, M) \leq 1\},$$

where $M \subset \mathbb{R}^n$ is a convex compact set;

- ii) We have access only to noisy evaluations of the function $\tilde{f}(x) = f(x) + \xi$, where $\mathbb{E}[\xi | x] = 0$ and $\mathbb{E}[\xi^2 | x] \leq \sigma^2$. The noise is independent across oracle calls;
- iii) $x^* \in \arg \min_{x \in M} f(x)$.

We now describe *batch finite-difference schemes* for minimizing the function f over the convex compact set M [11]. At iteration k :

- a) Let $u_{k,1}, \dots, u_{k,m_k}$ be independent random vectors uniformly drawn from the unit sphere in $\mathbb{R}^n$. Here, m_k denotes the batch size at iteration k .
- b) For each $i = 1, 2, \dots, m_k$, we compute an estimate of the subgradient of f using $u_{k,i}$:

$$g_{k,i} = \frac{n}{2\alpha_k} \left(\tilde{f}(x^k + \alpha_k u_{k,i}) - \tilde{f}(x^k - \alpha_k u_{k,i}) \right) u_{k,i}, \quad g^k = \frac{1}{m_k} \sum_{i=1}^{m_k} g_{k,i}. \quad (0.4)$$

- c) We generate a sequence $\{x^k\}$ using the following recursive relation:

$$x^{k+1} = \Pi_M(x^k - \alpha_k g^k).$$

The function $f(x, \alpha)$ is defined as follows:

$$f(x, \alpha) = \mathbb{E}_{z \sim B_1(0)} [f(x + \alpha z)] = \frac{1}{\text{Vol}(B_1(0))} \int_{\|z\| \leq 1} f(x + \alpha z) dz.$$

The following results hold (see [7, Lemma 2.8, p. 31]):

1)

$$|f(x, \alpha) - f(x)| \leq \alpha L_f \quad \forall x \in M. \quad (0.5)$$

2) The function $f(x, \alpha)$ is differentiable and

$$f'(x, \alpha) = \frac{n}{2\alpha} \mathbb{E}_{u \sim S^{n-1}} [(f(x + \alpha u) - f(x - \alpha u))u], \quad \|f'(x, \alpha)\| \leq n L_f \quad \forall x \in M.$$

That is, $H_3(x, \alpha)$ represents the stochastic gradient of $f(x, \alpha)$.

3) If the gradient of the function f satisfies the Lipschitz condition with constant L' , then

$$\|f'(x) - f'(x, \alpha)\| \leq \alpha L'. \quad (0.6)$$

In the sequel, it will be necessary to establish an upper bound for the following quantity: $\mathbb{E}[\|g^k\|^2]$.

Since $\mathbb{E}[\xi] = 0$, we have $\mathbb{E}[g_{k,i} | x^k] = f'(x^k, \alpha_k)$, $\mathbb{E}[g^k | x^k] = f'(x^k, \alpha_k)$.

We estimate the second moment:

$$g_{k,i} = s_{k,i} + n_{k,i}, \quad s_{k,i} = \frac{n}{2\alpha_k} \left(f(x^k + \alpha_k u_{k,i}) - f(x^k - \alpha_k u_{k,i}) \right) u_{k,i}, \quad n_{k,i} = \frac{n}{2\alpha_k} (\xi_{k,i}^+ - \xi_{k,i}^-) u_{k,i},$$

where $\mathbb{E}[\xi_{k,i}^\pm] = 0$, $\mathbb{E}[(\xi_{k,i}^\pm)^2] \leq \sigma^2$.

Since f is Lipschitz, we have $|f(x^k + \alpha_k u_{k,i}) - f(x^k - \alpha_k u_{k,i})| \leq 2\alpha_k L_f$, thus $\|s_{k,i}\| \leq n L_f$, $\mathbb{E}[\|n_{k,i}\|^2] = \frac{n^2}{4\alpha_k^2} \mathbb{E}[(\xi_{k,i}^+ - \xi_{k,i}^-)^2] \leq \frac{n^2 \sigma^2}{2\alpha_k^2}$. Hence,

$$\mathbb{E}[\|g_{k,i}\|^2] \leq 2\mathbb{E}[\|s_{k,i}\|^2] + 2\mathbb{E}[\|n_{k,i}\|^2] \leq 2n^2 L_f^2 + \frac{n^2 \sigma^2}{\alpha_k^2}.$$

As a result, we obtain

$$\begin{aligned} \mathbb{E}[\|g^k\|^2] &= \frac{1}{m_k^2} \left(\sum_{i=1}^{m_k} \mathbb{E}[\|g_{k,i}\|^2] + \sum_{i \neq j} \mathbb{E}[(g_{k,i}, g_{k,j})] \right) \\ &= \frac{1}{m_k^2} \left(m_k \mathbb{E}[\|g_{k,1}\|^2] + m_k(m_k - 1) \|f'(x^k, \alpha_k)\|^2 \right) \\ &\leq \|f'(x^k, \alpha_k)\|^2 + \frac{1}{m_k} \left(2n^2 L_f^2 + \frac{n^2 \sigma^2}{\alpha_k^2} \right) \leq n^2 L_f^2 + \frac{1}{m_k} \left(2n^2 L_f^2 + \frac{n^2 \sigma^2}{\alpha_k^2} \right). \end{aligned}$$

Thus, we have obtained the following estimate, which will play a key role in deriving the convergence rate:

$$\mathbb{E}[\|g^k\|^2] \leq n^2 L_f^2 + \frac{1}{m_k} \left(2n^2 L_f^2 + \frac{n^2 \sigma^2}{\alpha_k^2} \right). \quad (0.7)$$

The batch schemes are important: they reduce stochastic noise, ensure stable and balanced convergence, and allow efficient use of computational resources.

In the present paper, we apply this scheme to the class of quasiconvex functions of the form $f(x) = \max_{1 \leq i \leq r} f_i(x)$, where each function f_i is quasiconvex and has a Lipschitz-continuous gradient. For this class of problems, we derive convergence rate estimates for the proposed stochastic finite-difference method. Specifically, using this scheme for quasiconvex functions, we study the convergence of the sequence $\{x^k\}$ and obtain convergence rate estimates when the feasible set M is convex and compact. In this case, the convergence rate is of order $O\left(\frac{1}{\sqrt{k}}\right)$. Here, we choose the batch size on the order of k .

The paper also discusses the convergence of finite-difference stochastic schemes without noise, where the step sizes are chosen according to Polyak's method. In this case, the sequence $\{x_k\}$ converges to a solution of problem (0.1) at a rate of order $O\left(\frac{1}{\sqrt{k}}\right)$. Moreover, in the convex case, the sequence constructed using Numinskii's method converges with probability one to the set of minimizers.

1 The necessary information about quasiconvex functions

Let $M \subset \mathbb{R}^n$ be a convex compact subset. We denote by f^* the optimal value of the objective function in problem (0.1), and by M^* the set of all $x \in M$ at which this value is attained:

$$M^* = \{x \in M : f(x) = f^* = \min_{u \in M} f(u)\}.$$

Let us note that a quasiconvex continuous function defined on convex sets may have local minima that are not global. For example:

$$f(x) = \begin{cases} 1, & \text{if } x \in [0, 1], \\ 2 - x, & \text{if } x \in [1, 2]. \end{cases}$$

The function is quasiconvex, and $x = 1$ is a local minimum point that is not global, since the global minimum is attained at $x = 2$.

However, for strongly quasiconvex functions, the following statement is known.

Proposition 1.1 (see [1]). *Any strongly quasiconvex continuous function defined on a closed convex set possesses a unique minimum point, which is necessarily global.*

We emphasize that in this paper, problem (0.1) concerns the task of identifying a global minimum of a quasiconvex, Lipschitz function f over a convex compact set M .

For arbitrary $x, y \in M$ and $g(x) \in \partial f(x)$ with $g(x) \neq 0$, we introduce the auxiliary quantity

$$v_f(x, y) := \left(\frac{g(x)}{\|g(x)\|}, x - y \right).$$

Here $\partial f(x)$ denotes the Clarke subdifferential of the quasiconvex function f at x , defined as

$$\partial f(x) = \{v \in \mathbb{R}^n : f^o(x, h) \geq (v, h) \quad \forall h \in \mathbb{R}^n\},$$

where

$$f^o(x, h) = \lim_{\lambda \rightarrow 0, y \rightarrow x} \sup \frac{f(y + \lambda h) - f(y)}{\lambda}.$$

It is known (see [13]) that if f is locally Lipschitz, then $x \mapsto \partial f(x)$ is upper semicontinuous with convex and compact values.

Proposition 1.2. *Let f be quasiconvex, locally Lipschitz on a convex set M , and regular, i. e., it admits an ordinary directional derivative in every direction v :*

$$f'(x, v) = f^o(x, v).$$

If $f(y) \leq f(x)$, then for every $g(x) \in \partial f(x)$,

$$(g(x), y - x) \leq 0.$$

P r o o f. By definition of quasiconvexity,

$$f(\alpha y + (1 - \alpha)x) \leq \max\{f(y), f(x)\}.$$

Since $f(y) \leq f(x)$,

$$\frac{f(x + \alpha(y - x)) - f(x)}{\alpha} \leq 0.$$

Due to local Lipschitz continuity, we have

$$f^o(x, y - x) = \lim_{\alpha \rightarrow 0} \frac{f(x + \alpha(y - x)) - f(x)}{\alpha} \leq 0.$$

Hence

$$f^o(x, y - x) = \max_{g \in \partial f(x)} (g, y - x) \leq 0,$$

i. e.,

$$(g(x), y - x) \leq 0, \quad \forall g(x) \in \partial f(x).$$

Examples of regular functions include

$$f(x) = \max_{1 \leq i \leq r} f_i(x),$$

where the gradient f'_i of each function f_i is locally Lipschitz continuous on M . □

Corollary 1.1. *Let f be a locally Lipschitz, regular, and quasiconvex function on the set*

$$D = M + B_r(0), \quad r = \max\{1, \text{diam}(M)\},$$

Then for any $x \in M$ and any $g(x) \in \partial f(x)$ with $g(x) \neq 0$, the following holds: inequality holds:

$$f(x) - f(x^*) \leq L_d \frac{\langle g(x), x - x^* \rangle}{\|g(x)\|}, \quad (1.1)$$

where L_d is the Lipschitz constant of f on the set D .

P r o o f. Define

$$\Delta = f(x) - f(x^*),$$

and

$$z = x^* + \frac{\Delta}{L_d} \frac{g(x)}{\|g(x)\|}.$$

Since f is L_d -Lipschitz continuous on D , we have

$$f(z) \leq f(x^*) + L_d \|z - x^*\|.$$

By the definition of z ,

$$\|z - x^*\| = \frac{\Delta}{L_d},$$

hence

$$f(z) \leq f(x^*) + \Delta = f(x).$$

Because f is quasiconvex and regular, by Proposition 1.2 it follows from

$$f(z) \leq f(x)$$

that

$$\langle g(x), z - x \rangle \leq 0.$$

Substituting the expression for z , we obtain

$$\langle g(x), x^* - x \rangle + \frac{\Delta}{L_d} \left\langle g(x), \frac{g(x)}{\|g(x)\|} \right\rangle \leq 0.$$

Since

$$\left\langle g(x), \frac{g(x)}{\|g(x)\|} \right\rangle = \|g(x)\|,$$

we get

$$\langle g(x), x^* - x \rangle + \frac{\Delta}{L_d} \|g(x)\| \leq 0.$$

Equivalently,

$$-\langle g(x), x - x^* \rangle + \frac{\Delta}{L_d} \|g(x)\| \leq 0,$$

which implies

$$\langle g(x), x - x^* \rangle \geq \frac{\Delta}{L_d} \|g(x)\|.$$

Therefore,

$$\Delta = f(x) - f(x^*) \leq L_d \frac{\langle g(x), x - x^* \rangle}{\|g(x)\|}.$$

This completes the proof. □

R e m a r k 1.1. From (1.1), if for some $x \in M$,

$$(g(x), x - x^*) \leq 0, \quad g(x) \in \partial f(x), \quad g(x) \neq 0,$$

then x is a minimizer.

Even if $0 \in \partial f(x)$, x may fail to be a minimizer, e. g., $f(x) = x^3$ on $[-1, 1]$ has $f'(0) = 0$, but the minimizer is $x^* = -1$. Also, quasiconvex continuous functions may fail to have directional derivatives at some points, e. g., $f(x) = \sqrt{|x|}$ at $x = 0$.

Now consider the optimization problem (0.1), where

$$f(x) = \max_{1 \leq i \leq r} f_i(x).$$

If each f_i is continuously differentiable and quasiconvex, then f is quasiconvex and regular.

The finite-difference scheme is

$$x^{k+1} = \Pi_M(x^k - \rho_k g^k),$$

with

$$g^k = \frac{1}{m_k} \sum_{i=1}^{m_k} g_{k,i}, \quad g_{k,i} = \frac{n}{2\alpha_k} \left(\tilde{f}_{j_k}(x^k + \alpha_k u_{k,i}) - \tilde{f}_{j_k}(x^k - \alpha_k u_{k,i}) \right) u_{k,i},$$

where $j_k \in I(x^k) = \{i : f_i(x^k) = f(x^k)\}$.

2 Convergences rate for quasiconvex functions

Theorem 2.1. *Suppose that M is a convex and compact set, and the quasiconvex functions f_i , $i = 1, 2, \dots, r$ defined on $M + B_r(0)$ satisfy the following conditions:*

- i) The functions f_i , $i = 1, 2, \dots, r$ and their gradients satisfy the Lipschitz condition on the convex compact set $M + B_r(0)$, i. e., there exist constants $L_1 > 0$ and $L_2 > 0$ such that for any $i = 1, 2, \dots, r$*

$$|f_i(x) - f_i(y)| \leq L_1 \|x - y\|, \quad \|f'_i(x) - f'_i(y)\| \leq L_2 \|x - y\| \quad \forall x, y \in M + B_r(0).$$

- ii) Suppose that for any $i = 1, 2, \dots, r$*

$$f'_i(x) \neq 0 \quad \forall x \in M.$$

- iii) We minimize the function f over the compact convex set M using a finite-difference stochastic scheme with batching based on formula (0.4), and choose the parameters as follows:*

$$\rho_k = k^{-1/2}, \quad \alpha_k = k^{-1/2}, \quad m_k = k.$$

Then

$$\mathbb{E} \left[\min_{1 \leq k \leq T} f(x^k) \right] - f(x^*) = O\left(\frac{1}{\sqrt{T}}\right).$$

P r o o f. Using the update rule for x^{k+1} , we get

$$\|x^{k+1} - x^*\|^2 \leq \|x^k - \rho_k g^k - x^*\|^2 \leq \|x^k - x^*\|^2 - 2\rho_k (g^k, x^k - x^*) + \rho_k^2 \|g^k\|^2.$$

Hence

$$(g^k, x^k - x^*) \leq \frac{1}{2} \left(\frac{\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2}{\rho_k} \right) + \frac{1}{2} \rho_k \|g^k\|^2.$$

Taking the mathematical expectations of both sides of this inequality and taking into account

$$\mathbb{E} [g^k | x^k] = f'_{j_k}(x^k, \alpha_k),$$

we get

$$(f'_{j_k}(x^k, \alpha_k), x^k - x^*) \leq \frac{1}{2} \mathbb{E} \left[\frac{\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2}{\rho_k} \mid x^k \right] + \frac{1}{2} \rho_k \mathbb{E} [\|g^k\|^2 \mid x^k]. \quad (2.1)$$

From (0.6) and (1.1) we deduce the estimate:

$$(f'_{j_k}(x^k, \alpha_k), x^k - x^*) \geq (f'_{j_k}(x^k), x^k - x^*) - L_2 \text{diam}(M) \alpha_k \geq \frac{\gamma}{L_1} (f(x^k) - f(x^*)) - L_2 \text{diam}(M) \alpha_k,$$

where

$$\gamma = \min_{x \in M} \min_{1 \leq i \leq r} f'_i(x) > 0.$$

Consequently, using the inequality (2.1), we derive:

$$\mathbb{E} [f(x^k) - f(x^*)] \leq \frac{L_1}{2\gamma} \left(\mathbb{E} \left[\frac{\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2}{\rho_k} \right] + \frac{1}{2} \rho_k \mathbb{E} [\|g^k\|^2] \right) + C \alpha_k,$$

where $C = \frac{L_1 L_2 \text{diam}(M)}{\gamma}$.

Substituting the values of the parameters ρ_k, α_k, m_k here and taking into account the estimate (0.7), we obtain

$$\begin{aligned} \sum_{k=1}^T \mathbb{E} [f(x^k) - f(x^*)] &\leq \frac{L_1}{2\gamma \rho_1} \mathbb{E} [\|x^1 - x^*\|^2] + \frac{L_1}{2\gamma} \sum_{k=1}^{T-1} \mathbb{E} [\|x^{k+1} - x^*\|^2] (\sqrt{k+1} - \sqrt{k}) \\ &\quad + C \sum_{k=1}^T \frac{1}{k^{1/2}} + \frac{L_1}{2\gamma} \sum_{k=1}^T \left(2n^2 L_1^2 \frac{1}{k^{3/2}} + n^2 \sigma^2 \frac{k}{k^{3/2}} + \frac{1}{\sqrt{k}} n^2 L_1^2 \right) \\ &\leq \frac{L_1}{2\gamma} (\text{diam}(M))^2 + \frac{L_1}{2\gamma} (\text{diam}(M))^2 \sqrt{T} + Q \sum_{k=1}^T \frac{1}{\sqrt{k}} \\ &\leq \left(\frac{L_1}{\gamma} (\text{diam}(M))^2 + 2Q \right) \sqrt{T}, \end{aligned}$$

where

$$Q \equiv C L_1 + \frac{3n^2 L_1^3}{2\gamma} + \frac{n^2 \sigma^2 L_1}{2\gamma}.$$

From this, it follows that

$$\mathbb{E} \left[\min_{1 \leq k \leq T} f(x^k) \right] - f(x^*) \leq \frac{P}{\sqrt{T}},$$

where

$$P = \frac{L_1}{\gamma} \text{diam}(M)^2 + 2Q,$$

i. e.,

$$\mathbb{E} \left[\min_{1 \leq k \leq T} f(x^k) \right] - f(x^*) = O\left(\frac{1}{\sqrt{T}}\right).$$

□

Theorem 2.2. Suppose that f is quasiconvex and regular on the compact convex set $M + B_1(0)$ and $0 \notin \partial f(x)$, $\forall x \in M$. Assume further that the problem (0.1) is solved using a finite-difference stochastic batch scheme, with parameters ρ_k, α_k, m_k chosen as in Theorem 2.1. Then, under these conditions, the sequence $F_T \equiv \min_{1 \leq k \leq T} f(x^k)$ converges to the minimal value f^* with probability 1 as the number of iterations T tends to infinity.

P r o o f. The proof of this theorem is analogous to the proof of Theorem 2.1. We outline the main steps below.

Analogously to inequality (2.1), we have

$$(f'(x^k, \alpha_k), x^k - x^*) \leq \frac{1}{2} \mathbb{E} \left[\frac{\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2}{\rho_k} \mid x^k \right] + \frac{1}{2} \rho_k \mathbb{E} [\|g^k\|^2 \mid x^k]. \quad (2.2)$$

Since the Clarke subdifferential is upper semicontinuous, for every $\varepsilon > 0$ there exists $\delta > 0$ such that, for all $\alpha \in (0, \delta)$, $x \in M$, and $u \in B_1(0)$, one has

$$\partial f(x + \alpha u) \subset \partial f(x) + \varepsilon B_1(0). \quad (2.3)$$

It is also known that the gradient $f'(x, \alpha)$ admits the representation

$$f'(x, \alpha) = \frac{1}{\text{Vol}(B_1(0))} \int_{\|u\| \leq 1} g(x + \alpha u) du,$$

where $g(x)$ is some measurable selection from the set-valued mapping $\partial f(x)$.

Combining (2.2) and (2.3) with the quasiconvexity property (inequality (1.1)), we obtain

$$f(x) - f(x^*) \leq \frac{L_f}{\nu} (f'(x, \alpha), x - x^*) + \frac{L_f \text{diam}(M) \varepsilon}{\nu}, \quad (2.4)$$

where

$$\nu := \min_{x \in M} \text{dist}(0, \partial f(x)) > 0.$$

Combining the results above, we deduce that there exists an index k_0 such that for all $k \geq k_0$, using the law of total expectation together with inequalities (2.2) and (2.4), we have

$$\mathbb{E}[f(x^k) - f(x^*)] \leq \frac{L_f}{2\nu} \mathbb{E} \left[\frac{\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2}{\rho_k} \right] + \frac{1}{2} \rho_k \mathbb{E} [\|g^k\|^2] + \frac{L_f \text{diam}(M) \varepsilon}{\nu}.$$

Hence, by an argument identical to the proof of Theorem 2.1, using the parameter choices $\rho_k = k^{-1/2}$, $\alpha_k = k^{-1/2}$, $m_k = k$, and standard estimates for the stochastic gradient variance, we conclude that

$$\min_{1 \leq k \leq T} f(x^k) \rightarrow f^* \quad \text{with probability 1 as } T \rightarrow \infty.$$

This completes the proof. $\square$

3 The case of a strongly quasiconvex function

P r o p o s i t i o n 3.1 (see [2]). *Let M be a closed convex set in $\mathbb{R}^n$ and $f : M \rightarrow \mathbb{R}$ be continuous and strongly quasiconvex with modulus $\mu > 0$. Then $\arg \min_{x \in M} f(x) = x^*$ and*

$$f(x) - f(x^*) \geq \frac{\mu}{8} \|x - x^*\|^2 \quad \forall x \in M. \quad (3.1)$$

Theorem 3.1. *Assume that the conditions of Theorem 2.1 are satisfied, and in addition that all functions $f_i(x)$, $i = 1, 2, \dots, r$, are strongly quasiconvex with modulus $\mu > 0$ over a compact convex set M . Consider solving problem (0.1) using finite-difference stochastic batch schemes based on formula (0.4) parameters:*

$$\rho_k = \frac{1}{\beta k}, \quad \alpha_k = \frac{1}{\sqrt{k}}, \quad m_k = \lfloor \sqrt{k} \rfloor + 1,$$

where

$$\beta = \frac{\mu\gamma}{4L_1}, \quad \gamma = \min_i \min_{x \in M} \|f'_i(x)\|.$$

Then

$$\mathbb{E}[\|x^T - x^*\|^2] = O\left(\frac{1}{\sqrt{T}}\right).$$

P r o o f. Arguing as in Theorem 2.1 and additionally taking inequality (3.1) into consideration, we obtain

$$\begin{aligned} \mathbb{E}[\|x^{k+1} - x^*\|^2] &\leq \mathbb{E}[\|x^k - x^*\|^2] - 2\rho_k \mathbb{E}[(g^k, x^k - x^*)] + \rho_k^2 \mathbb{E}[\|g^k\|^2] \\ &\leq \mathbb{E}[\|x^k - x^*\|^2] - 2\rho_k \mathbb{E}[(f'_{j_k}(x^k, \alpha_k), x^k - x^*)] + \rho_k^2 \mathbb{E}[\|g^k\|^2] \\ &\leq \mathbb{E}[\|x^k - x^*\|^2] - 2\rho_k \left(\frac{\gamma}{L_1} \mathbb{E}[f(x^k) - f(x^*)] - L_2 \text{diam}(M) \alpha_k \right) + \rho_k^2 \mathbb{E}[\|g^k\|^2] \\ &\leq \left(1 - \frac{\mu\gamma}{4L_1} \rho_k\right) \mathbb{E}[\|x^k - x^*\|^2] + 2L_2 \text{diam}(M) \rho_k \alpha_k \\ &\quad + \rho_k^2 \frac{1}{m_k} \left(2n^2 L_1^2 + \frac{n^2 \sigma^2}{\alpha_k^2}\right) + n^2 L_1^2 \rho_k^2 \\ &\leq \left(1 - \frac{1}{k}\right) \mathbb{E}[\|x^k - x^*\|^2] + \frac{G^2}{\beta^2 k \sqrt{k}}, \end{aligned}$$

where

$$G^2 = 2L_2 \text{diam}(M) \beta + 3n^2 L_1^2 + n^2 \sigma^2.$$

Define $a_k = \mathbb{E}[\|x^k - x^*\|^2]$. Then

$$a_{k+1} \leq \left(1 - \frac{1}{k}\right) a_k + \frac{b}{k^{3/2}}, \quad b = \frac{G^2}{\beta^2}.$$

Multiplying both sides by k and defining $b_k = (k-1)a_k$, we get

$$b_{k+1} \leq b_k + \frac{b}{\sqrt{k}}.$$

Summing from $k = 1$ to T gives

$$b_{T+1} \leq b_1 + b \sum_{j=1}^T \frac{1}{\sqrt{j}} \leq 2b\sqrt{T} + b_1.$$

Therefore,

$$a_T = O\left(\frac{1}{\sqrt{T}}\right).$$

□

4 Finite-difference stochastic schemes for minimizing a convex non-differentiable function on $\mathbb{R}^n$: the Nurminskii method

We consider the following iterative minimization procedure for a locally Lipschitz function f , as proposed by Nurminskii in [14]:

$$x^{k+1} = \begin{cases} x^k - \rho_k g^k, & f(x^k) \leq f(x^0) + c, \\ x^0, & f(x^k) > f(x^0) + c, \end{cases}$$

where x^0 is the initial approximation, $c > 0$ is a constant, and the vector g^k is iteratively updated following the finite-difference scheme outlined in the introduction, i. e., by formula (0.3).

Now, let us consider the problem of minimizing a function

$$f(x) = \max_{1 \leq i \leq r} f_i(x)$$

of a given form over the entire space $\mathbb{R}^n$. Denote

$$M = \{x \in \mathbb{R}^n : f(x) \leq f(x^0) + c\}.$$

It is obvious that the set of minimizers of the function f over the whole space coincides with the set of its minimizers on M .

For the proof of the convergence theorem, we require the following fundamental statement.

Lemma 4.1 (see [15]). *Let $(\mathcal{F}_k)$, $k \geq 0$ be a filtration, and let*

$$\{\mathbb{X}_k\}, \{A_k\}, \{B_k\}$$

be sequences of nonnegative, $\mathcal{F}_k$ -adapted random variables satisfying the following conditions:

- 1) $\mathbb{E}[\mathbb{X}_{k+1} \mid \mathcal{F}_k] \leq \mathbb{X}_k - A_k + B_k, \quad k \geq 0.$
- 2) $\sum_{k=1}^{\infty} B_k < \infty$ (almost surely).
- 3) $\mathbb{X}_k \geq 0, \quad k \geq 0.$

Then $\mathbb{X}_k$ converges almost surely to a finite limit, and $\sum_{k=1}^{\infty} A_k < \infty$.

R e m a r k 4.1. This statement remains valid if the condition requiring $\mathbb{X}_k$ to be nonnegative is replaced by the assumption that it is bounded from below. Indeed, suppose a lower bound is known

$$\mathbb{X}_k \geq m, \quad \forall k.$$

Let us introduce

$$\mathbb{Z}_k := \mathbb{X}_k - m.$$

Then, if there was already an inequality for the original variable, substituting gives the same type of inequality for $\mathbb{Z}_k$. Consequently,

$$\mathbb{Z}_k \rightarrow \mathbb{Z}_{\infty} \quad \text{almost surely,}$$

and therefore

$$\mathbb{X}_k = \mathbb{Z}_k + m \rightarrow \mathbb{Z}_{\infty} + m \quad \text{almost surely.}$$

The following convergence theorem holds:

Theorem 4.1. *Let the following assumptions hold:*

- 1) *Assume that the functions $f_i(x)$, $i = 1, 2, \dots, r$ are convex on the whole space, and their gradients are locally Lipschitz continuous. Also assume that the set M is bounded.*

2) We minimize f over the entire space using scheme (0.2), where the random vector g^k is chosen according to formula (0.3). The parameters ρ_k, α_k are selected as follows.

$$\rho_k = \frac{1}{k}, \quad \alpha_k = \frac{1}{\sqrt{k}}.$$

Then there exists $\bar{x} : \Omega \rightarrow M^*$ such that $\{x^k\}$ converges to $\bar{x}$ almost surely, that is

$$P\left(\{\omega \in \Omega : \{x^k(\omega)\} \text{ converges to } \bar{x}(\omega)\}\right) = 1.$$

P r o o f. Since the set M is compact by assumption, and the gradients $f'_i, i = 1, 2, \dots, r$, are locally Lipschitz continuous, it follows that the function f is Lipschitz continuous on the set $M + B_1(0)$ with Lipschitz constant L_f . Moreover, the gradients $f'_i, i = 1, 2, \dots, r$, are also Lipschitz continuous with some constant L' .

First, we show that the sequence $\{x^k\}$ leaves the set

$$M = \{x \in \mathbb{R}^n : f(x) \leq f(x^0) + c\}$$

only finitely many times with probability one, and that the sequence $f(x^k)$ converges almost surely.

It is known [10, Lemma 2.4, Chapter 1, Section 2, p. 36] that the gradient of the function $f(x, \alpha)$ is Lipschitz on the set M with a Lipschitz constant $L_\alpha = D/\alpha$, where D is some constant. Therefore,

$$f(x^{k+1}, \alpha_k) \leq f(x^k, \alpha_k) - \rho_k(f'(x^k, \alpha_k), g^k) + \frac{L_{\alpha_k}}{2} \rho_k^2 \|g^k\|^2.$$

Taking the conditional expectation, we obtain

$$\mathbb{E}[f(x^{k+1}, \alpha_k) \mid x^k] \leq f(x^k, \alpha_k) - \rho_k \|f'(x^k, \alpha_k)\|^2 + \frac{n^2 D L_f^2}{\alpha_k} \rho_k^2.$$

To bring this into the Robbins–Siegmund form, note that

$$\begin{aligned} \mathbb{E}[f(x^{k+1}, \alpha_{k+1}) \mid x^k] &= \mathbb{E}[f(x^{k+1}, \alpha_k) \mid x^k] + \mathbb{E}[f(x^{k+1}, \alpha_{k+1}) - f(x^{k+1}, \alpha_k) \mid x^k] \\ &\leq f(x^{k+1}, \alpha_k) + L_f |\alpha_{k+1} - \alpha_k|. \end{aligned}$$

Thus, the Robbins–Siegmund inequality becomes

$$\mathbb{E}[\mathbb{X}_{k+1} \mid \mathcal{F}_k] \leq \mathbb{X}_k - A_k + B_k,$$

with

$$\mathbb{X}_k = f(x^k, \alpha_k), \quad A_k = \rho_k \|f'(x^k, \alpha_k)\|^2, \quad B_k = \frac{C_1}{k^{3/2}},$$

where $C_1 = n^2 D L_f^2$.

If $x^{k+1} = x^0$ (reset case), then

$$\mathbb{X}_{k+1} = f(x^0, \alpha_{k+1}) \leq f(x^0) + L_f \alpha_{k+1},$$

and

$$\mathbb{X}_k = f(x^k, \alpha_k) \geq f(x^k) - L_f \alpha_k > f(x^0) + c - L_f \alpha_k,$$

so that

$$\mathbb{X}_{k+1} - \mathbb{X}_k \leq -c + L_f(\alpha_k + \alpha_{k+1}).$$

Hence, the Robbins–Siegmund terms are

$$A_k = \max\{0, c - L_f(\alpha_k + \alpha_{k+1})\}, \quad B_k = \max\{0, L_f(\alpha_k + \alpha_{k+1}) - c\}.$$

Therefore,

$$\mathbb{E}[\mathbb{X}_{k+1} \mid \mathcal{F}_k] \leq \mathbb{X}_k - A_k + B_k,$$

with $A_k \geq 0$, $B_k \geq 0$, and $\sum_{k=1}^{\infty} B_k < \infty$.

If the number of resets were infinite, there would exist a subsequence $\{k_j\}$ with

$$\mathbb{X}_{k_j} > f(x^0) + c - L_f\alpha_{k_j}, \quad \mathbb{X}_{k_j+1} \leq f(x^0) + L_f\alpha_{k_j+1}.$$

Since $\mathbb{X}_k \rightarrow \mathbb{X}^\infty$ almost surely, this would imply

$$f(x^0) + c \leq \mathbb{X}^\infty \leq f(x^0),$$

which is a contradiction. Therefore, the sequence leaves M only finitely many times.

By the Robbins–Siegmund lemma, $f(x^k, \alpha_k)$ converges almost surely. Since

$$|f(x^k, \alpha_k) - f(x^k)| \leq L_f\alpha_k \rightarrow 0,$$

it follows that $f(x^k)$ converges almost surely. Since $f(x)$ is convex, the function $f(x, \alpha)$ is also convex with respect to x . Therefore, using this fact and inequality (0.5), we also obtain the following inequality

$$\mathbb{E}[\|x^{k+1} - x^*\|^2 \mid x^k] \leq \|x^k - x^*\|^2 - 2\rho_k(f(x^k) - f(x^*)) + \frac{\tilde{G}^2}{k\sqrt{k}},$$

where $\tilde{G}^2 = 2L' \text{diam}(M) + n^2 L_\dagger^2$. Applying the Robbins–Siegmund lemma again, we obtain that $\|x^k - x^*\|^2$ converges almost surely, and

$$\sum_{k=1}^{\infty} \frac{f(x^k) - f(x^*)}{k} < \infty \quad \text{almost surely.}$$

Hence, there exists a subsequence $f(x^{k_j}) \rightarrow f(x^*)$ almost surely, and since $f(x^k)$ converges, it follows that

$$f(x^k) \rightarrow f(x^*) \quad \text{almost surely.}$$

Let Z be a countable dense subset of M^* , and define

$$\Theta_z = \{\omega \in \Omega : \{\|x^k - z\|\} \text{ converges}\}, \quad z \in Z.$$

Then

$$P\left(\Theta = \bigcap_{z \in Z} \Theta_z\right) = 1.$$

Define $\bar{x} : \Omega \rightarrow \mathbb{R}^n$ such that $\bar{x}(\omega)$ is a cluster point of $\{x^k(\omega)\}$ for $\omega \in \Theta$. Since f is continuous and

$$\lim_{k \rightarrow \infty} f(x^k(\omega)) = f(x^*),$$

we have $\bar{x}(\omega) \in M^*$, $\omega \in \Theta$.

Fix $\varepsilon > 0$ and $\omega \in \Theta$. There exists $z(\omega) \in Z$ such that

$$\|\bar{x}(\omega) - z(\omega)\| \leq \frac{\varepsilon}{3}.$$

Let $x^{k_j}(\omega) \rightarrow \bar{x}(\omega)$. Then there exists N such that

$$\|x^k(\omega) - z(\omega)\| \leq \frac{2\varepsilon}{3}, \quad k \geq N.$$

Hence,

$$\|x^k(\omega) - \bar{x}(\omega)\| \leq \|x^k(\omega) - z(\omega)\| + \|z(\omega) - \bar{x}(\omega)\| \leq \varepsilon, \quad \forall k \geq N.$$

This proves the almost sure convergence of $\{x^k\}$ to $\bar{x} \in M^*$. $\square$

R e m a r k 4.2. If the functions f_i , $i = 1, 2, \dots, r$, are strongly convex, then the function f is also strongly convex. Moreover, the level sets of a strongly convex function are bounded. Therefore, in this case, the set M is bounded. Furthermore, such functions possess a unique minimizer. Let us present an example of a convex function defined on the entire space whose sublevel sets are bounded, which possesses a unique minimizer, but which is not strongly convex. The following is an example of such a function:

$$f(x_1, x_2) = \max\{x_1^2, x_2^2\}, \quad (x_1, x_2) \in \mathbb{R}^2.$$

5 Polyak's step selection method in finite-difference stochastic schemes

Let us reconsider problem (0.1). It is assumed that the minimal value f^* is known. To solve this problem, we apply a *finite-difference stochastic scheme* without noise, that is, the function values are assumed to be available at each iteration. The iterative process is given by

$$x^{k+1} = \Pi_M(x^k - \rho_k H_3(x^k, \alpha_k, u_k)). \quad (5.1)$$

We now show that the parameters can be chosen *dynamically*, in a manner similar to the *Polyak step size*, and that the sequence $\{x^k\}$ *converges with probability one* to a solution of problem (0.1). Let (Ω, F, P) be a probability space.

Theorem 5.1. *Let $f_i(x)$, $i = 1, 2, \dots, r$, satisfy the assumptions of Theorem 2.1. Assume that the sequence $\{x^k\}$ is generated by iteration formula (5.1). and that the parameters α_k and ρ_k are chosen as*

$$\alpha_k = \rho_k = \frac{f(x^k) - f(x^*)}{\eta},$$

where

$$\eta = \frac{n^2 L_1^3 + 2\gamma L_2 \text{diam}(M)}{\gamma}. \quad (5.2)$$

Then there exists a random variable $\bar{x} : \Omega \rightarrow M^*$ such that

$$P(\{\omega \in \Omega : x^k(\omega) \rightarrow \bar{x}(\omega)\}) = 1,$$

and the following convergence rate estimate holds:

$$\mathbb{E}\left[\min_{k \leq T} f(x^k)\right] - f(x^*) = O\left(\frac{1}{\sqrt{T}}\right).$$

P r o o f. Similarly to the proof of Theorem 2.1, for any $x^* \in M^*$, we have

$$\begin{aligned} \mathbb{E}[\|x^{k+1} - x^*\|^2 \mid x^k] &\leq \|x^k - x^*\|^2 - 2\rho_k(f'_{j_k}(x^k, \alpha_k), x^k - x^*) + \rho_k^2 \mathbb{E}[\|g^k\|^2 \mid x^k] \\ &\leq \|x^k - x^*\|^2 - 2\rho_k \left(\frac{\gamma}{L_1} (f(x^k) - f(x^*)) - L_2 \text{diam}(M) \alpha_k \right) + \rho_k^2 n^2 L_1^2 \\ &= \|x^k - x^*\|^2 - \left(\frac{2\gamma}{L_1 \eta} - \frac{2\gamma L_2 \text{diam}(M)}{L_1 \eta^2} - \frac{n^2 L_1^2}{\eta^2} \right) (f(x^k) - f(x^*))^2. \end{aligned}$$

Let

$$\tilde{C} = \left(\frac{2\gamma}{L_1 \eta} - \frac{2\gamma L_2 \text{diam}(M)}{L_1 \eta^2} - \frac{n^2 L_1^2}{\eta^2} \right).$$

It is easy to verify that if the parameter η is chosen according to formula (5.2), then $\tilde{C} > 0$. By assumption, the coefficient of $(f(x^k) - f(x^*))^2$ is positive. Therefore, by the Robbins–Siegmund Lemma, the sequence $\{\|x^k - x^*\|\}$ converges almost surely, and

$$\sum_{k=1}^{\infty} (f(x^k) - f(x^*))^2 < \infty \quad \text{almost surely.}$$

Hence $\lim_{k \rightarrow \infty} f(x^k) = f(x^*)$ almost surely.

From the main inequality above, summing over $k = 1$ to T gives

$$\sum_{k=1}^T \mathbb{E}[(f(x^k) - f(x^*))^2] \leq \frac{1}{\tilde{C}} \sum_{k=1}^T \mathbb{E}[\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2] \leq \frac{(\text{diam}(M))^2}{\tilde{C}}.$$

Applying the standard inequality between the minimum and the sum, we obtain

$$\mathbb{E} \left[\min_{k \leq T} (f(x^k) - f(x^*)) \right] = O \left(\frac{1}{\sqrt{T}} \right).$$

Moreover, the fact that the sequence $\{x^k\}$ converges with probability one to the set of minimizers M^* is proved in the same manner as in Theorem 4.1. $\square$

R e m a r k 5.1. For the practical implementation of Theorem 5.1, it is, of course, assumed that the parameters L_1 , L_2 , γ , and $\text{diam}(M)$ are known.

References

- [1] F. Lara, “On strongly quasiconvex functions: existence results and proximal point algorithm”, *Journal of Optimization Theory and Applications*, **192**:3 (2022), 891–911.
- [2] S. M. Grad, F. Lara, R. T. Marcavillaca, “Strongly quasiconvex functions: what we know so far”, 2024, arXiv: [2410.23055v1](#).
- [3] M. Jovanović, “A note on strongly convex and quasiconvex functions”, *Math. Notes*, **60**:5 (1996), 584–585.
- [4] Yu. E. Nesterov, “Minimization methods for non-smooth convex and quasiconvex functions”, *Matekon*, **29** (1984), 519–531.
- [5] A. Gasnikov, M. Alkousa, A. Lobanov, Y. Dorn, F. Stonyakin, I. Kuruzov, S. Singh, “On quasiconvex smooth optimization problems by a comparison oracle”, 2024, arXiv: [2411.16745v1](#).
- [6] O. Shamir, T. Zhang, “Stochastic gradient descent for non-smooth optimization: Convergence results and Averaging Schemes”, 2012, arXiv: [1212.1824](#).

- [7] E. Hazan, “Introduction to Online Convex Optimization: Essential Knowledge”, *Foundations and Trends in Optimization*, **2**:3–4 (2015), 157–325.
- [8] G. Garrigos, R.M. Gower, “Handbook of convergence theorems for (stochastic) gradient methods”, 2024, arXiv: [2301.11235v3](https://arxiv.org/abs/2301.11235v3).
- [9] A. M. Gupal, *Algorithms for finding the extremum of nondifferentiable functions with constraint*, Diss. ... Cand. of Physical and Mathematical Sciences, Institute of Cybernetics, Academy of Science of the USSR, Kyiv, 1976, 18 pp.
- [10] V. S. Mikhalevich, A. M. Gupal, V. I. Norkin, *Methods of Nonconvex Optimization*, Nauka Publ., Moscow, 1987 (In Russian), 280 pp.
- [11] Yu. E. Nesterov, V. Spokoiny, “Random gradient-free minimization of convex functions”, *Foundations of Computational Mathematics*, **17**:2 (2017), 527–566.
- [12] J. C. Duchi, M. I. Jordan, “Optimal rate for zero-order convex optimization: The power of two function evaluation”, 2014, arXiv: [1312.2139](https://arxiv.org/abs/1312.2139).
- [13] F. H. Clarke, “Generalized gradients and applications”, *Trans. Amer. Math. Soc.*, **205** (1975), 247–262.
- [14] E. A. Nurminskii, “Convergence of stochastic analogs of deterministic procedures of nonlinear programming”, *Theory of Optimal Processes*, 1973, 60–73 (In Russian).
- [15] H. Robbins, D. Siegmund, “A convergence theorem for non negative almost supermartingales and some applications”, *Optimizing Methods in Statistics*, Proceedings of a Symposium Held at the Center for Tomorrow, the Ohio State University (Ohio State University, June 14–16, 1971), Academic Press, 1971, 233–257.

Информация об авторе

Хачатрян Рафик Агасиевич, доктор физико-математических наук, профессор кафедры численного анализа и математического моделирования, Ереванский государственный университет, г. Ереван, Армения. E-mail: khrafik@ysu.am
ORCID: <https://orcid.org/0000-0002-7908-0562>

Поступила в редакцию 09.11.2025 г.
 Поступила после рецензирования 03.05.2026 г.
 Принята к публикации 05.06.2026 г.

Information about the author

Rafik A. Khachatryan, Doctor of Physical and Mathematical Sciences, Professor of the Numerical Analysis and Mathematical Modeling Department, Yerevan State University, Yerevan, Armenia. E-mail: khrafik@ysu.am
ORCID: <https://orcid.org/0000-0002-7908-0562>

Received 09.11.2025
 Reviewed 03.05.2026
 Accepted for press 05.06.2026