Informatics and Control Problems 45 Issue 2 (2025)
journal homepage: www.icp.az

Derivative-free methods in nonconvex optimization

Boris S. Mordukhovich

Department of Mathematics Wayne State University, Detroit, Michigan, USA

ARTICLE INFO

Article history:
Received 20.05.2025
Received in revised form 16.06.2025
Accepted 28.11.2025
Available online 26.12.2026

Keywords:
Derivative-free optimization
Nonconvex smooth objective functions
Finite differences
Convergent algorithms
Convergence rates

ABSTRACT

This paper discusses some new directions and results, obtained in joint work with Pham Duy Khanh (Ho Chi Minh City University of Education, Vietnam) and Dat Ba Tran (Rowan University, USA), for models of derivative-free optimization (DFO) with nonconvex data. We overview several approaches to DFO problems and focus on finite-difference approximation schemes. Our algorithms address the two major classes: objective functions with globally Lipschitzian and locally Lipschitzian gradients, respectively. Global convergence results with constructive convergence rates are established for both cases in noiseless and noisy environments. The developed algorithms in the noiseless case are based on the backtracking line search and achieves fundamental convergence properties. The noisy version is essentially more involved being based on the novel dynamic step line search. Numerical experiments demonstrate higher robustness of the proposed algorithms compared with other finite-difference-based schemes.

1. Introduction

We consider here standard unconstrained optimization problem (P) defined by

$$\text{minimize} \quad f(x) \quad \text{subject to} \;\; x \in \mathbb{R}^n,$$

where $f: \mathbb{R}^n \to \mathbb{R}$ is a continuously differentiable (i.e., $\mathcal{C}^1$-smooth) function that may not be convex. Derivative-free optimization means that there is no available information about ∇f but only about either f (*noiseless case*), or its approximation $\phi = f + \xi$ (*noisy case*). Over the years, problems of this type have drawn strong attention in optimization theory and applications; see, e.g., [1-5] with the references therein. Among various approaches to DFO, we concentrate below on optimization algorithms involving *finite-difference approximations*, which are less investigated while showing to be very promising from both viewpoints of optimization theory and applications; see [4], [5] among other recent publications.

The noiseless case is more direct to handle by finite-difference approximations since a fixed mesh interval can be used in discretization. However, known results obtained in this way for DFO have not yet achieved desired convergence properties as for standard gradient descent methods. One of the reasons for this is that derivative-free methods based on finite-difference approximations may not produce descent directions without careful adaptive modifications. Dealing with DFO models in the presence of noise is much more complicated, and the usage of finite-difference approximations may significantly increase computational costs.

To address these issues in both noiseless and noisy cases, we first consider objective functions f of *class* $\mathcal{C}_L^{1,1}$, i.e., such that ∇f is globally Lipschitz continuous on $\mathbb{R}^n$ with constant L. The proposed

E-mail address: aa1086@wayne.edu (Boris S. Mordukhovich)

www.icp.az/2025/2-01.pdf https://doi.org/10.54381/icp.2025.2.01

method is labeled as a *derivative-free method with constant stepsize* (DFC). In the noiseless case, the proposed DFC exhibits global convergence with constructive convergence rates under natural conditions including *Polyak-Łojasiewicz-Kurdyka* (PLK) properties; see [6] for more details. In the noisy case, DFC provides finite convergence to an explicitly defined *near-stationary point* by estimating the number of iterates needed to reach such a point. In contrast to known methods, a *crucial feature* of DFC is a simultaneous adjustment, at each iteration, of the Lipschitz constant of the objective function with the stepsize of the finite-difference approximation.

The class of $\mathcal{C}^{1,1}$ functions (i.e., those with locally Lipschitzian gradients) is rather new in the DFO literature; see [5] for detailed discussions. This framework allows us to efficiently investigate not only global but also local convergence and convergence rates under the corresponding PLK conditions. The most difficult case of *large noise* in DFO problems with $\mathcal{C}^{1,1}$ objectives is now handled by using a *dynamic step line search*, where an approximate Lipschitz constant is used to determine both the stepsize and the finite-difference interval. The conducted *numerical experiments* for various problems arising in machine learning, statistics, artificial intelligence, etc. demonstrate high efficiency of the proposed algorithms to solve practical models.

In what follows, we consider separately DFO algorithms and convergence results for classes of $\mathcal{C}_L^{1,1}$ and $\mathcal{C}^{1,1}$ objective functions in both noiseless and noisy cases.

2. Derivative-Free methods for $\mathcal{C}_L^{1,1}$ objectives

Given a $\mathcal{C}^1$-smooth function $f: \mathbb{R}^n \to \mathbb{R}$, we say that a mapping $\mathcal{G}: \mathbb{R}^n \times (0,\infty) \to \mathbb{R}^n$ is a *global approximation* of ∇f if there exist a constant $C > 0$ such that

$$\|\mathcal{G}(x,\delta) - \nabla f(x)\| \le C\delta \quad \text{for any} \quad (x,\delta) \in \mathbb{R}^n \times (0,\infty).$$

$\mathcal{G}$ is a *local approximation* of ∇f if for any bounded set $\Omega \subset \mathbb{R}^n$ and any $\Delta > 0$, there exists $C > 0$ with

$$\|\mathcal{G}(x,\delta) - \nabla f(x)\| \le C\delta \quad \text{for any} \quad (x,\delta) \in \Omega \times (0,\Delta].$$

For $f \in \mathcal{C}_L^{1,1}$, design the following *derivative-free method with constant stepsize* (DFC) in the noiseless case.

Algorithm DFC-noiseless

Step 1. *Choose a global approximation $\mathcal{G}$ of ∇f. Select an initial point $x^1 \in \mathbb{R}^n$, an initial sampling radius $\delta_1 > 0$, a constant $C_1 > 0$, a reduction factor $\theta \in (0,1)$, and scaling factors $\mu > 2, \eta > 1, \kappa > 0$. Set $k := 1$.*

Step 2. *Find g^k and the smallest nonnegative integer i_k such that $g^k = \mathcal{G}(x^k, \theta^{i_k}\delta_k)$* and $\|g^k\| > \mu C_k \theta^{i_k}\delta_k$. *Then set $\delta_{k+1} := \theta^{i_k}\delta_k$.*

Step 3. *If $f\left(x^k - \frac{\kappa}{C_k} g^k\right) \le f(x^k) - \frac{\kappa(\mu-2)}{2C_k\mu}\|g^k\|^2$, then $x^{k+1} := x^k - \frac{\kappa}{C_k} g^k$ and $C_{k+1} := C_k$. Otherwise, set $x^{k+1} := x^k$ and $C_{k+1} := \eta C_k$.*

Recall that the *PLK condition* holds for a differentiable function $f: \mathbb{R}^n \to \mathbb{R}$ at $\bar{x}$ if there are $\eta > 0$, a neighborhood U of $\bar{x}$, and a nondecreasing function $\psi: (0,\eta) \to (0,\infty)$ for which $1/\psi$ is integrable over $(0,\eta)$ and

$$\|\nabla f(x)\| \ge \psi(f(x) - f(\bar{x})) \quad \text{when} \quad x \in U \quad \text{with} \quad f(\bar{x}) < f(x) < f(\bar{x}) + \eta.$$

The *exponential PLK condition* holds if $\psi(t) = Mt^q$ with some $M > 0$ and $q \in [0,1)$.

Theorem 1. Let $\{x^k\}$ be the sequence generated by Algorithm DFC-noiseless with $\nabla f(x^k) \ne 0$ for all $k \in \mathbb{N}$. We have that either $f(x^k) \to -\infty$ as $k \to \infty$, or:

(i) The gradient sequence $\{\nabla f(x^k)\}$ converges to 0 as $k \to \infty$.

(ii) If f satisfies the PLK condition at some accumulation point $\bar{x}$ of $\{x^k\}$, then $x^k \to \bar{x}$ as $k \to \infty$.

(iii) If the exponential PLK conditions holds with some $M > 0$ and $q \in [1/2,1)$, then we have the following convergence rates for $\{x^k\}$:

• For $q = 1/2$, the sequences $\{x^k\}$, $\{\nabla f(x^k)\}$, and $\{f(x^k)\}$ converge linearly to $\bar{x}$, 0, and $f(\bar{x})$, respectively.

• For $q \in (1/2,1)$, the convergence rates are

$$\|x^k - \bar{x}\| = \mathcal{O}\left(k^{-\frac{1-q}{2q-1}}\right),$$

$$\|\nabla f(x^k)\| = \mathcal{O}\left(k^{-\frac{1-q}{2q-1}}\right), \quad f(x^k) - f(\bar{x}) = \mathcal{O}\left(k^{-\frac{2-2q}{2q-1}}\right).$$

Consider now the *noisy case* $\phi = f + \xi$ with a bounded noise function $|\xi(x)| \le \xi_f$ (which is not assumed to be known) on $\mathbb{R}^n$ and construct, by using the basic vectors $e_i$in $\mathbb{R}^n$, the gradient approximation

$$\tilde{\mathcal{G}}(x,\delta) := \frac{1}{\delta}\sum_{i=1}^{n}(\phi(x+\delta e_i) - \phi(x))\, e_i \quad \text{for} \quad (x,\delta) \in \mathbb{R}^n \times (0,\infty).$$

Algorithm DFC-noise

Step 1. *Select some* $x^1 \in \mathbb{R}^n, \delta_1 > 0,\ L_1 > 0,\ \theta \in (0,1),$ *and* $\eta > 1$.
Step 2. *Find* g^k *and the smallest nonnegative integer* i_k *satisfying* $g^k = \tilde{\mathcal{G}}(x^k, \theta^{i_k}\delta_k)$ and $\|g^k\| > 2L_k\sqrt{n}\theta^{i_k}\delta_k$ *and then set* $\delta_{k+1} := \theta^{i_k}\delta_k$.
Step 3. *If* $\phi\left(x^k - \frac{1}{L_k}g^k\right) \le \phi(x^k) - \frac{1}{24L_k}\|g^k\|^2$, *then set* $x^{k+1} := x^k - \frac{1}{L_k}g^k$ *and* $L_{k+1} := L_k$. *Otherwise, set* $x^{k+1} := x^k$ *and* $L_{k+1} := \eta L_k$.

Theorem 2. Let the sequence $\{x^k\}$ be generated by Algorithm DFC-noise with $\delta_1 \ge \sqrt{\frac{4\xi_f}{L}}$ and $L_1 < \eta L$. Then the number N of iterations that Algorithm DFC-noise takes until $\|\nabla f(x^N)\| < 16\sqrt{Ln\xi_f}$ is bounded by

$$N \le N_{\text{opt}} := 1 + \left\lceil \frac{f(x^1) - f^* + 2\xi_f}{M\xi_f} \right\rceil + \left\lceil \log_\eta\left(\frac{\eta L}{L_1}\right)\right\rceil,$$

where $M := \frac{15nL_1^2}{\eta(L+4L_1)^2}$.The total number N_{fval} of function evaluations needed to achieve this goal is bounded by

$$N_{\text{ftal}} \le (n+2)N_{\text{opt}} + n\left\lceil \log_\theta\left(\frac{2\sqrt{\xi_f}}{\delta_1\sqrt{L}}\right)\right\rceil.$$

III. Derivative-Free methods with $\mathcal{C}^{1,1}$ objectives

Consider first the *noiseless case* and design the following *derivative-free method with backtracking stepsize* (DFB).

Algorithm DFB-noiseless

Step 7. *Choose a local approximation* $\mathcal{G}$ *of* ∇f *and select an initial point* $x^1 \in \mathbb{R}^n$ *and initial radius* $\delta_1 > 0$, *a constant* $C_1 > 0$, *factors* $\theta \in (0,1)$, $\mu > 2,\ \eta > 1$, *line search constants* $\beta \in (0,1/2), \gamma \in (0,1)$, $\bar{\tau} > 0$, *and an initial bound* $t_1^{\min} \in (0,\bar{\tau})$. *Choose a sequence of errors* $\{\nu_k\} \subset [0,\infty)$ *with* $\nu_k \downarrow 0$ *as* $k \to \infty$ *and set* $k := 1$.
Step 1. *Select* g^k *and the smallest nonnegative integer* i_k *satisfying* $g^k = \mathcal{G}\big(x^k, \min\{\theta^{i_k}\delta_k, \nu_k\}\big)$ and $\|g^k\| > \mu C_k\theta^{i_k}\delta_k$ *and then set* $\delta_{k+1} := \theta^{i_k}\theta_k$.

Step 2. *Set stepsize* $t_k := \bar{\tau}$ *and then take* $t_k := \gamma t_k$ *while* $f(x^k - t_k g^k) > f(x^k) - \beta t_k \|g^k\|^2$ and $t_k \geq t_k^{\min}$,

Step 3. *If* $t_k \geq t_k^{\min}$, *set* $\tau_k := t_k$, $C_{k+1} := C_k$ *and* $t_{k+1}^{\min} := t_k^{\min}$. *Otherwise, set* $\tau_k := 0$, $C_{k+1} := \eta C_k$ *and* $t_{k+1}^{\min} := \gamma t_k^{\min}$.

Step 4. *Set* $x^{k+1} := x^k - \tau_k g^k$. *Increase* k *by* 1 *and go back to Step 1.*

Theorem 3. Let $\{x^k\}$ be generated by Algorithm DFB-noiseless with $\nabla f(x^k) \neq 0$ for all $k \in \mathbb{N}$. Then either $f(x^k) \to -\infty$ as $k \to \infty$, or we have:

(i) Every accumulation point of $\{x^k\}$ is a stationary point of f.

(ii) If $\{x^k\}$ is bounded, then the collection of accumulation points of $\{x^k\}$ is nonempty, compact, and connected in $\mathbb{R}^n$.

(iii) If $\{x^k\}$ has an isolated accumulation point $\bar{x}$, then this sequence converges to $\bar{x}$ as $k \to \infty$.

Imposing the PLK conditions on f as in Theorem 1 allows us to establish global convergence and convergence rates for Algorithm DFB-noiseless that are similar to Algorithm DFC-noiseless in Theorem 1.

Next we continue the study of problem (P) with the objective function $f: \mathbb{R}^n \to \mathbb{R}$ of class $\mathcal{C}^{1,1}$, where only a *noisy approximation* $\phi(x) = f(x) + \xi(x)$ of f is available. Using the forward finite difference $\tilde{\mathcal{G}}(x, \delta)$ as in Section II, we design the following *derivative-free method with dynamic step line search* (DFD), whose crucial property is a dynamic step line search to construct the stepsize together with the finite-difference interval.

Algorithm DFD-noise

Step 1 (initialization). *Select an initial point* $x^1 \in \mathbb{R}^n$, $\eta > 1$, *and* $L_1 > 0$. *Set* $k := 1$.

Step 2. *Find an integer number* i_k *with the smallest absolute value such that for* $g^k := \tilde{\mathcal{G}}\left(x^k, \sqrt{\frac{4\xi_f}{\eta^{i_k} L_k}}\right)$ *and* $\tau_k = \frac{1}{\eta^{i_k} L_k}$, *we have the estimate* $\phi(x^k - \tau_k g^k) \leq \phi(x^k) - \frac{\tau_k}{9}\|g^k\|^2$.

Step 3. *Set* $x^{k+1} := x^k - \tau_k g^k$ *and* $L_{k+1} := \eta^{i_k} L_k$.

We say that Step 1 of Algorithm DFD-noise is *successful* if the integer number i_k can be found. The next theorem provides a constructive condition for the successful step and allows us to estimate a number of iterations needed to reach an almost-stationary point.

Theorem 4. Let ℓ_k be a constant of the Lipschitz continuity of ∇f on $\mathbb{B}\left(x^k, \max\left\{\frac{3}{2\ell_k}\|\nabla f(x^k)\|, \sqrt{\frac{4\xi_f}{\ell_k}}\right\}\right)$ at the k th iteration of Algorithm DFD-noise. Then the condition

$$\|\nabla f(x^k)\| \geq 8\sqrt{\ell_k \eta n \xi_f}$$

ensures that Step 1 of Algorithm DFD-noise is successful. Suppose in addition that there is $L > 0$ such that ∇f is Lipschitz continuous with constant L on $\bigcup_{k=1}^{\infty} \mathbb{B}\left(x^k, \max\left\{\frac{3}{2L}\|\nabla f(x^k)\|, \sqrt{\frac{4\xi_f}{L}}\right\}\right)$. If in this case $\inf_{k \in \mathbb{N}} f(x^k) > -\infty$ and for some $K \in \mathbb{N}$ we have $L_K \in [L, \eta L)$, then the following hold:

(i) There exists $N \in \mathbb{N}$ for which

$$\|\nabla f(x^N)\| < 8\sqrt{L \eta n \xi_f}.$$

(ii) If furthermore f admits a global minimizer with the minimum value $\overline{f} < f(x^K)$ and f satisfies the exponential PLK condition with some $\mu > 0$ and $q = 1/2$, then the number N ensuring (i) is upper estimated by

$$N \leq \max\left\{1+K, 1+K+\log_{1-\frac{3\mu}{16\eta L}}\left(\frac{32\eta n\xi_f}{f(x^K)-\bar{f}}\right)\right\}.$$

A large number of conducted numerical experiments confirm the efficiency of the proposed algorithms and their advantages in applications to practical models.

3. Conclusion

This presentation describes several efficient algorithms to solve problems of smooth nonconvex derivative-free optimization with global and local Lipschitzian gradients of the objective functions. The proposed algorithms are separately designed in both noiseless and noisy cases for the minimizing functions.

Acknowledgment

This research was partly supported by the US National Science Foundation under grant DMS-22045519. The author is grateful to Pham Duy Khanh and Dat Ba Tran for the fruitful collaboration on this project, and particularly to Professor Tran for his technical help to prepare this presentation.

References

[1] M.J.D. Powell, An efficient method for finding the minimum of a function of several variables without calculating derivatives, Comput. J. 7 (1964) pp.155-162.

[2] A.R. Conn, K. Scheinberg, L.N. Vicente, Introduction to Derivative-Free Optimization, SIAM, (2009).

[3] Yu. Nesterov, V. Spokoiny, Random gradient-free minimization of convex functions, Found. Comput. Math. 17 (2017) pp.527-566.

[4] A.S. Berahas, L. Cao, K. Choromanski, K. Scheinberg, A theoretical and empirical comparison of gradient approximations in derivative-free optimization, Found. Comput. Math. 22 (2022) pp.507-560.

[5] D.H. Cuong, P.D. Khanh, B.S. Mordukhovich, D.B. Tran, Local convergence analysis for nonisolated solutions to derivative-free methods of optimization. https://optimization-online.org/?p=27216 (to appear in Optimization, 2025).

[6] G. Bento, B. S. Mordukhovich, T. Mota, Yu. Nesterov, Convergence of descent optimization algorithms under Polyak-Łojasiewicz-Kurdyka conditions. https://optimization-online.org/2025/02 (to appear in J. Optim. Theory Appl., 2025).