Non-convergence Analysis of Probabilistic Direct Search

Huang Cunxin, Zhang Zaikun
(corresponding: cun-xin.huang@connext.polyu.hk)
Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong SAR, China

Contribution Highlights

In this work, we conduct the non-convergence analysis of the probabilistic direct search (PDS). With the help of the non-convergence theory, we

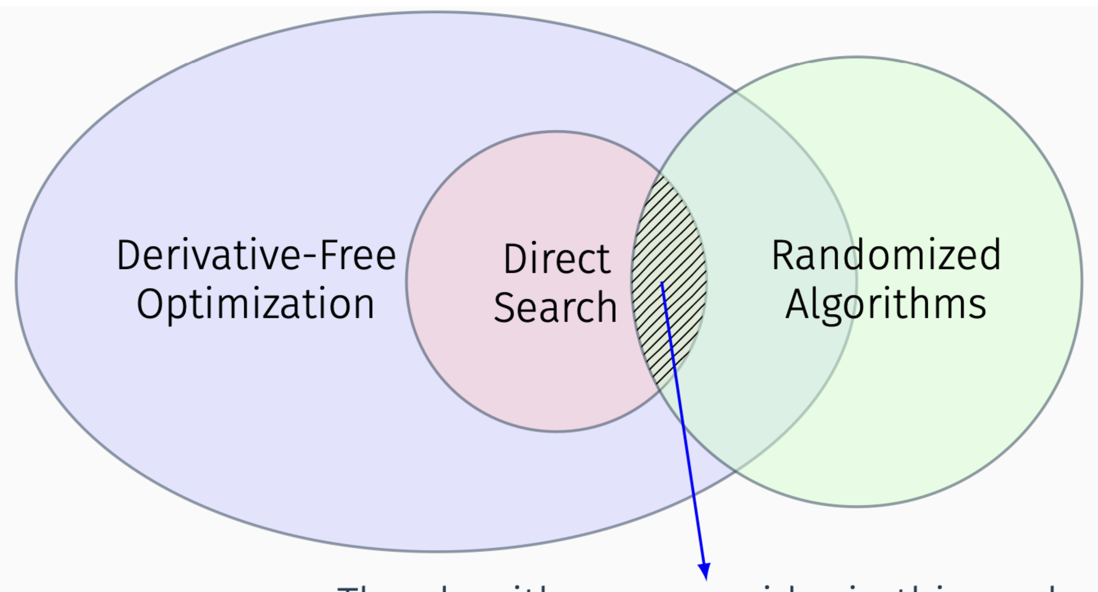
- theoretically explain the non-convergence phenomenon of PDS,
- and find out the behavior of PDS is closely related to the random series

$$S(\kappa) \; = \; \sum_{k=1}^{\infty}\prod_{\ell=0}^{k-1}\gamma^{1-Y_\ell(\kappa)}\theta^{Y_\ell(\kappa)}.$$

Non-convergence analysis can

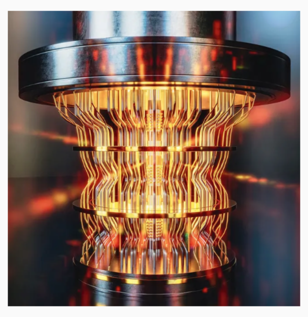
- verify whether your assumption for convergence is essential,
- deepen our understanding of mathematical tools we use,
- provide new perspectives on convergence analysis,
- guide the choice of algorithmic parameters.

Introduction



Derivative-free optimization (DFO) is a major class of optimization methods that

- do not use derivatives (first-order info.), only use function values,
- and is closely related to zeroth-order/black-box optimization,
- and have various applications such as



Quantum Computing

Machine Learning

Circuit Design

Direct-search methods are a popular class of DFO methods that decide the iterates based on "simple" comparisons of function values.

Probabilistic direct search is an efficient offspring of direct search and randomization techniques. The algorithm is shown as follows.

Algorithm 1: Probabilistic Direct Search based on sufficient decrease

Input: $x_0 \in \mathbb{R}^n$, $\alpha_0 \in (0,\infty)$, $c \in (0,\infty)$, $0 < \theta < 1 < \gamma$.

for $k = 0, 1, \ldots$ **do**

  Select a finite set of directions $\mathcal{D}_k \subset \mathbb{R}^n$ randomly.
  (In this work, assume $\mathcal{D}_k$ is a set of unit vectors for simplicity)
  Set $d_k = \arg\min\{f(x_k + \alpha_k d) : d \in \mathcal{D}_k\}$.
  (complete polling for simplicity)
  if $f(x_k + \alpha_k d_k) < f(x_k) - c\alpha_k^2$ **then**
    Set $x_{k+1} = x_k + \alpha_k d_k$ and $\alpha_{k+1} = \gamma\alpha_k$.
    (Move and expand step size)
  else
    Set $x_{k+1} = x_k$ and $\alpha_{k+1} = \theta\alpha_k$.
    (Stay and shrink step size)

Typical choice of $\{\mathcal{D}_k\}$ (GRVZ 2015): $\mathcal{D}_k = \{d_1, \ldots, d_m\}$ with $d_\ell \overset{\text{i.i.d.}}{\sim} \mathrm{U}(\mathcal{S}^{n-1})$

Tools for analysis

- **Cosine measure w.r.t. a vector**
  Cosine measure of a finite set $\mathcal{D} \subseteq \mathbb{R}^n \backslash \{0\}$ and a vector $v \in \mathbb{R}^n \backslash \{0\}$ is defined as
$$\mathrm{cm}(\mathcal{D}, v) \; = \; \max_{d \in \mathcal{D}} \frac{d^\top v}{\|d\|\|v\|}.$$
- ***p*-probabilistically κ-descent** [GRVZ 2015]
  $\{\mathcal{D}_k\}$ is p-probabilistically κ-descent if for all $k \ge 0$,
$$\mathbb{P}\left(\mathrm{cm}(\mathcal{D}_k, -g_k) \ge \kappa \mid \mathcal{F}_{k-1}\right) \; \ge \; p,$$
  where $g_k = \nabla f(x_k)$ and $\mathcal{F}_{k-1} = \sigma(\mathcal{D}_0, \ldots, \mathcal{D}_{k-1})$.

Existing convergence results

Global convergence:
If $\{\mathcal{D}_k\}$ is p_0-probabilistically κ-descent with $\kappa > 0$ and

$$p_0 \; = \; \frac{\log\theta}{\log(\gamma^{-1}\theta)},$$

then PDS converges a.s. when f is L-smooth and lower-bounded.

Corollary under the typical case:
If $\mathcal{D}_k = \{d_1, \ldots, d_m\}$, where $d_\ell \overset{\text{i.i.d.}}{\sim} \mathrm{U}(\mathcal{S}^{n-1})$, then PDS converges a.s. if

$$m \; > \; \log_2\left(1 - \frac{\log\theta}{\log\gamma}\right).$$

Naive question & simple test

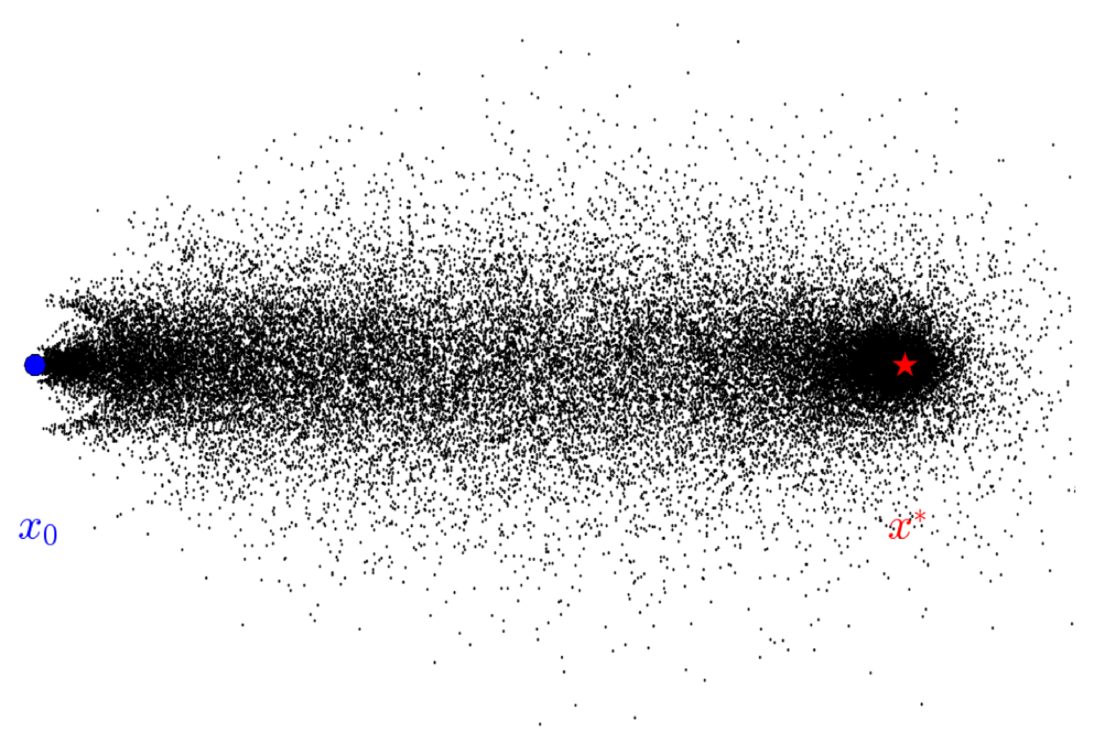
Questions:

- Is p_0-probabilistically κ-descent an essential assumption or a technical one?
- What will happen if
$$m \le \log_2\left(1 - \frac{\log\theta}{\log\gamma}\right)?$$

Simple test:

- Objective function: $f(x) = x^\top x/2$,
- Initial point: $x_0 = (-10, 0)^\top$,
- Stopping criterion: $\alpha_k \le$ machine epsilon,
- Number of experiments: 100,000,
- Parameters of PDS: $\alpha_0 = 1$, $\theta = 0.25$, $\gamma = 1.5$, $m = 2$, which render
$$m \; = \; 2 < 2.143 \approx \log_2\left(1 - \frac{\log\theta}{\log\gamma}\right).$$

Test results:



Note: each **black dot** represents the **output point** of one run of PDS.

Key ingredients

A important series:

$$S(\kappa) \; = \; \sum_{k=1}^{\infty}\prod_{\ell=0}^{k-1}\gamma^{1-Y_\ell(\kappa)}\theta^{Y_\ell(\kappa)}.$$

***p*-probabilistically ascent:** for each $k \ge 0$,

$$\mathbb{P}\left(\mathrm{cm}\left(\mathcal{D}_k, -g_k\right) \le 0 \mid \mathcal{F}_{k-1}\right) \; \ge \; p.$$

Main results w.r.t. $S(\kappa)$

New convergence result w.r.t. $S(\kappa)$:
If there exists a $\kappa > 0$ such that $S(\kappa) = \infty$ a.s., then PDS converges a.s.

Convergence of $S(0)$:
If $\{\mathcal{D}_k\}$ is p-probabilistically ascent with $p > p_*$, where

$$p_* \; = \; 1 - p_0 \; = \; \frac{\log\gamma}{\log(\theta^{-1}\gamma)},$$

then we have $S(0) < \infty$ a.s. and

$$\mathbb{P}\left(S(0) < \zeta\right) \; > \; 0 \quad \iff \quad \zeta > \frac{\theta}{1-\theta}.$$

Main theorems

Relation between convergence results:

p_0-probabilistically κ-descent $\quad\Rightarrow\quad S(\kappa) = \infty$ a.s.

Non-convergence theorem:
Assume that f is smooth, convex, and has a solution set $\mathcal{S}^*$. If $\{\mathcal{D}_k\}$ is p-probabilistically ascent with $p > p_*$, then

$$\mathbb{P}\left(\liminf_{k\to\infty} \mathrm{dist}(x_k, \mathcal{S}^*) > 0\right) \; > \; 0,$$

provided that $\mathrm{dist}(x_0, \mathcal{S}^*) > \alpha_0/(1-\theta)$.

Non-convergence under the typical case:
Let $\mathcal{D}_k = \{d_1, \ldots, d_m\}$, where $d_\ell \overset{\text{i.i.d.}}{\sim} \mathrm{U}(\mathcal{S}^{n-1})$. Then PDS is non-convergent if

$$m \; < \; \log_2\left(1 - \frac{\log\theta}{\log\gamma}\right).$$

Tightness of the analysis

Question: Can p-probabilistically ascent with $p \ge p_*$ instead of $p > p_*$ lead to non-convergence?

Answer: No. Following is an implementation of PDS that is p_*-probabilistically ascent but converges a.s.

- $\theta = 1/2$ and $\gamma = 2$, which implies $q_* = 1/2$;
- $\mathcal{D}_k = \{g_k/\|g_k\|\}$ or $\{-g_k/\|g_k\|\}$ with probability $1/2$, respectively.

Resources

Website Website Website

References

[1] S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang (GRVZ 2015). Direct search based on probabilistic descent. *SIAM J. Optim.*, 25:1515–1541, 2015.

[2] R. Durrett. *Probability: Theory and Examples.* Camb. Ser. Stat. Probab. Math. Cambridge University Press, Cambridge, fourth edition, 2010.