Non-convergence Analysis of Randomized Direct Search

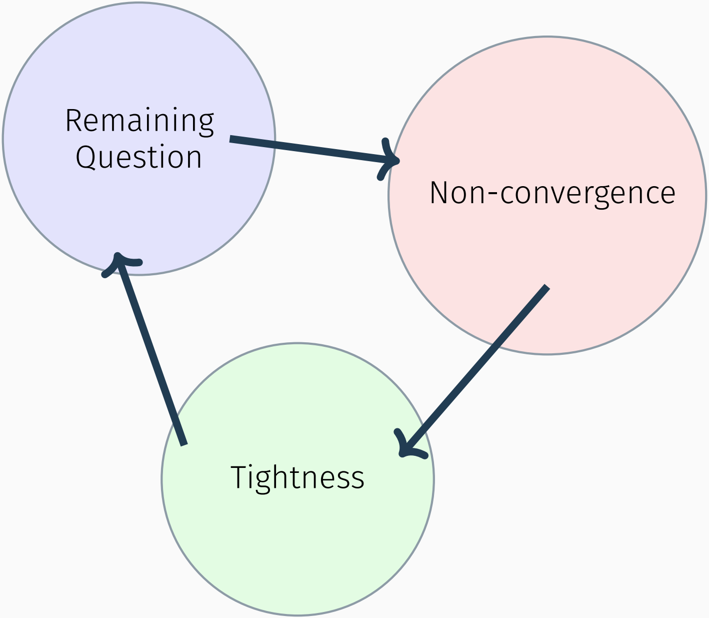
AMA613

Cunxin Huang
Co-supervised by Dr. Zaikun Zhang and Prof. Xiaojun Chen

November 6, 2023

Department of Applied Mathematics
The Hong Kong Polytechnic University

Big Picture



Remaining question

Brief review of last talk

In last talk

1. Derivative-free optimization (DFO)
   - A branch of optimization
   - Do not use derivatives (only use function evaluations)
2. Randomized direct search (RDS)
   - Make decisions based on simple comparisons of function values
   - Choose direction sets randomly

Illustration of how RDS works

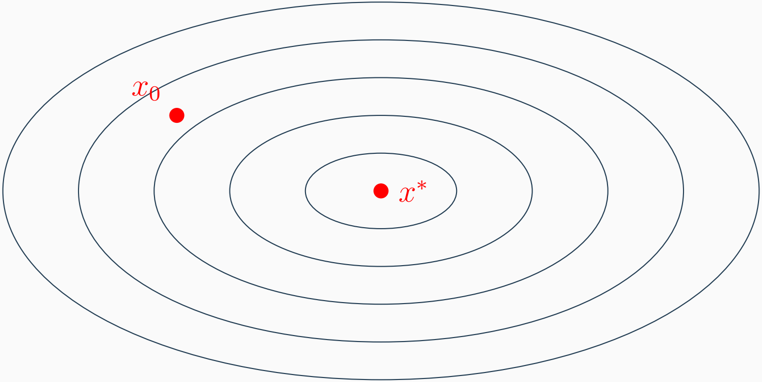


Illustration of how RDS works

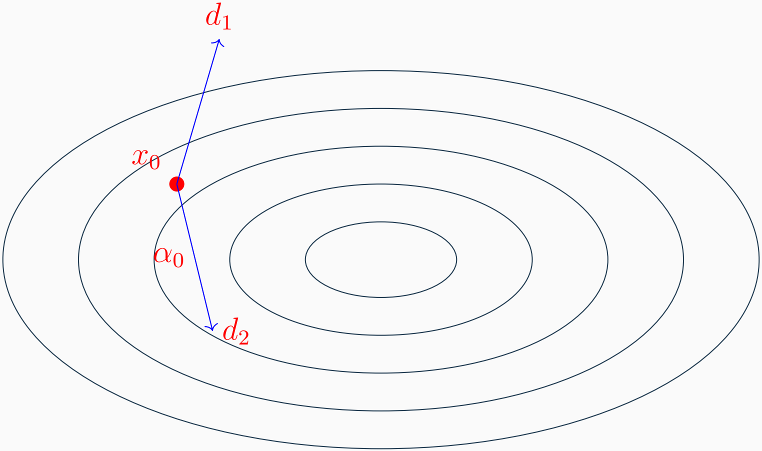


Illustration of how RDS works

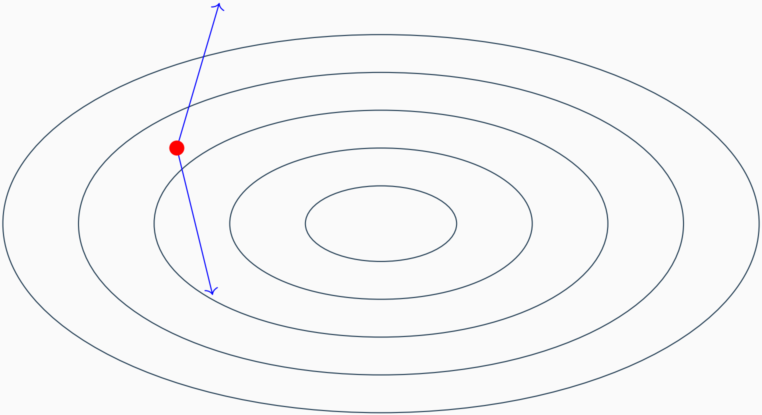


Illustration of how RDS works

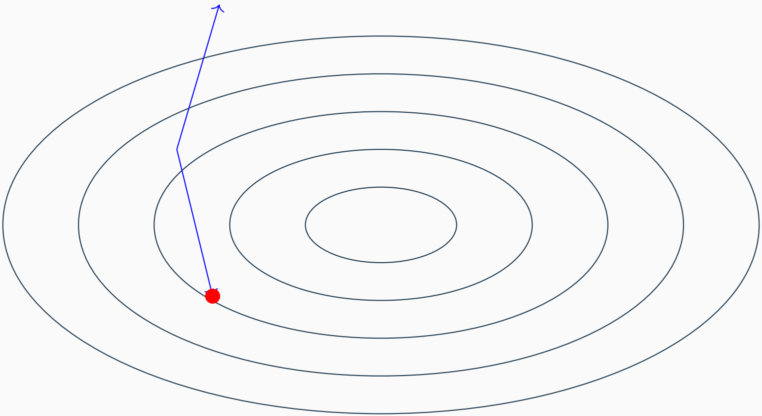


Illustration of how RDS works

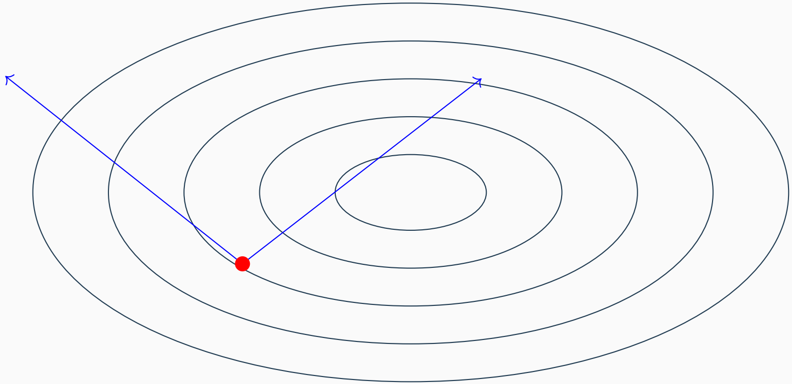


Illustration of how RDS works

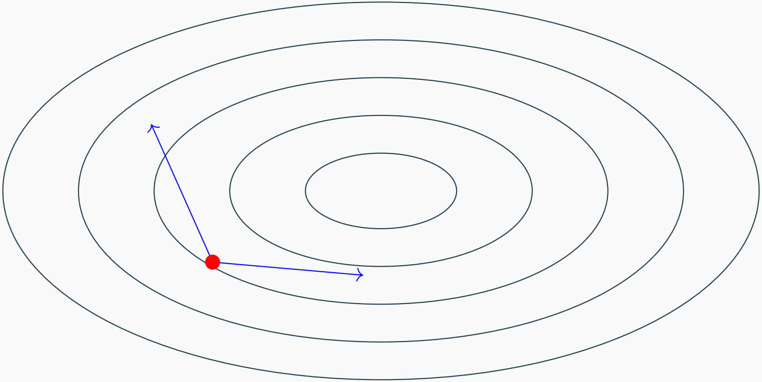


Illustration of how RDS works

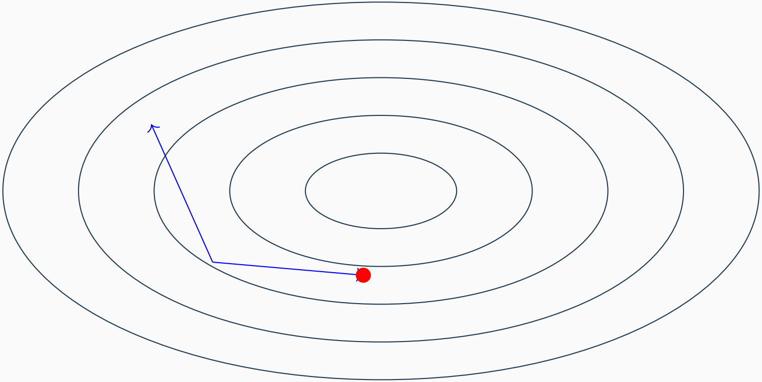


Illustration of how RDS works

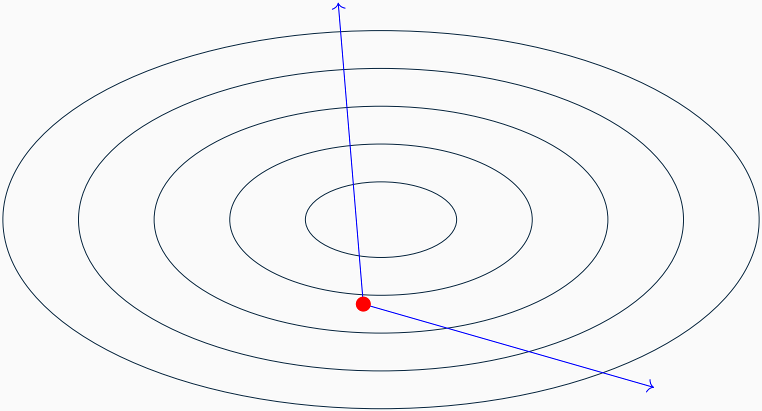


Illustration of how RDS works

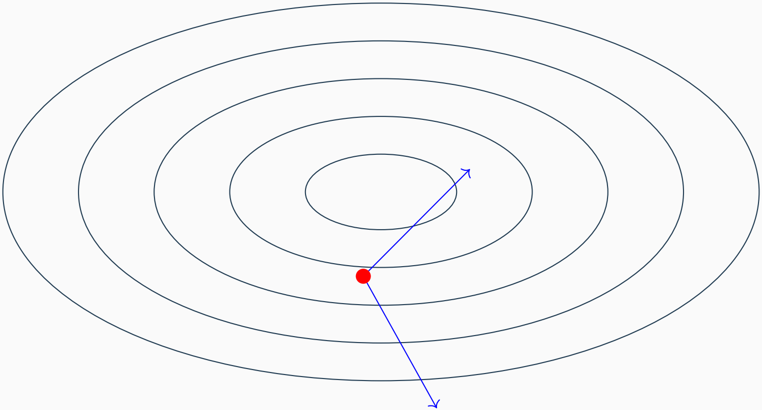


Illustration of how RDS works

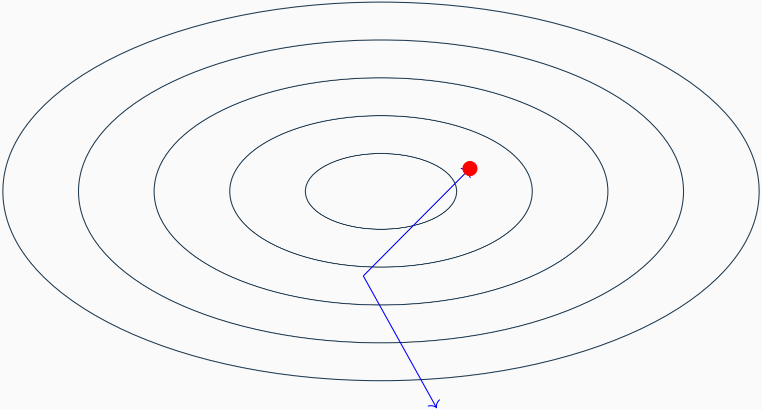


Illustration of how RDS works

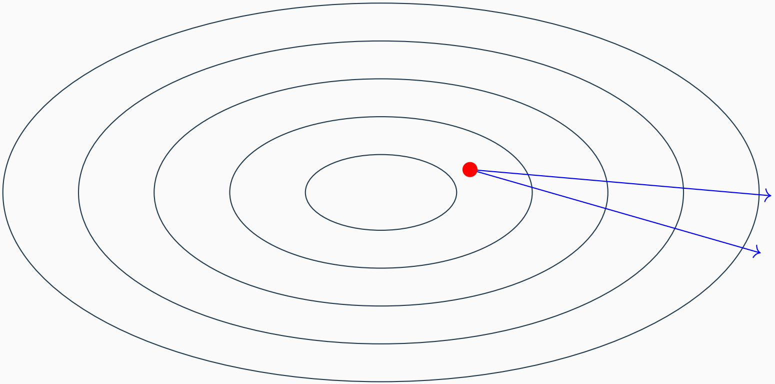


Illustration of how RDS works

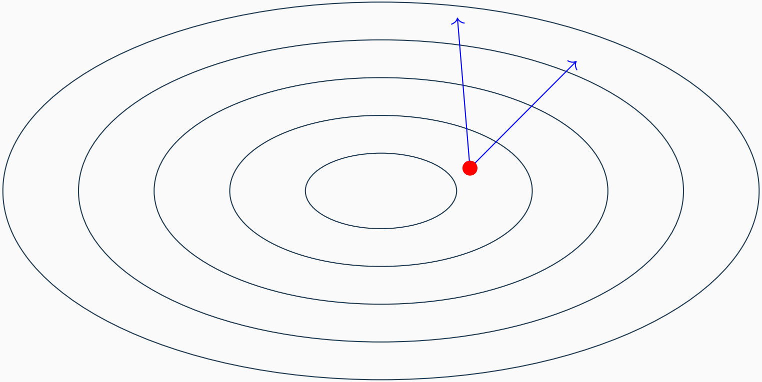
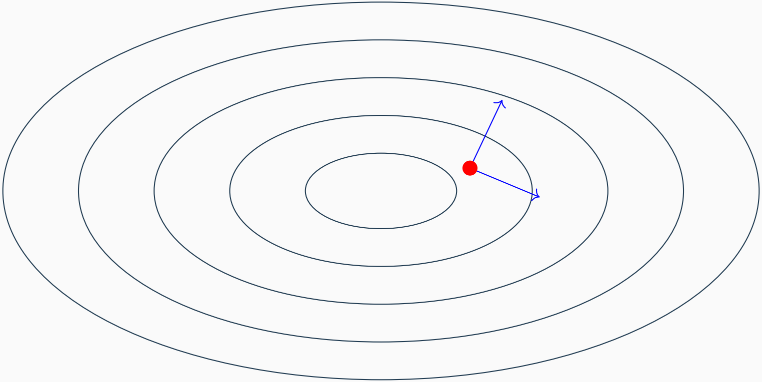


Illustration of how RDS works



Remaining question

Recall that

Theorem (Gratton et al. 2015)

If $D_k = \{d_1, \ldots, d_m\}$, *where* $d_i \overset{\text{i.i.d.}}{\sim}$ *Unif*$(\mathcal{S}^{n-1})$, *then RDS is convergent if*

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

Reminder: θ *is shrinking factor,* γ *is expanding factor.*

Remaining question

Recall that

Theorem (Gratton et al. 2015)

If $D_k = \{d_1, \dots, d_m\}$, where $d_i \overset{i.i.d.}{\sim}$ Unif$(\mathcal{S}^{n-1})$, then RDS is convergent if

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

Reminder: θ is shrinking factor, γ is expanding factor.

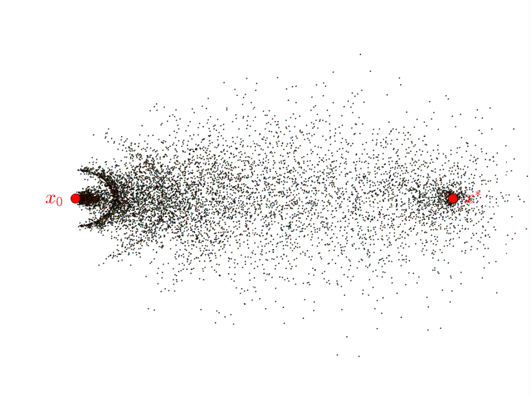
A natural question: what if

$$m \leq \log_2\left(1 - \frac{\log\theta}{\log\gamma}\right)?$$

Remaining question

A natural question: what if

$$m \leq \log_2\left(1 - \frac{\log\theta}{\log\gamma}\right)?$$



Non-convergence

Motivation: non-convergence analysis matters

Many well-known algorithms have non-convergence analysis.

- S. Reddi, S. Kale, and S. Kumar. On the convergence of Adam and beyond. In Y. Bengio, Y. LeCun, T. Sainath, I. Murray, M. Ranzato, and O. Vinyals, editors, *International Conference on Learning Representations (ICLR 2018)*. Curran Associates, Inc., 2018.
- C. Chen, B. He, Y. Ye, and X. Yuan. The direct extension of ADMM for multi-block convex minimization problems is not necessarily convergent. *Math. Program.*, 155:57-79, 2016.
- W. Mascarenhas. The divergence of the BFGS and Gauss Newton methods. *Math. Program.*, 147:253-276, 2014.
- ...

Practically meaningful: guide the choice of algorithmic parameters

Recall cosine measure

Definition (Cosine measure)

Cosine measure for a finite set of nonzero vectors $D \subseteq \mathbb{R}^n$ w.r.t. a given vector $v \in \mathbb{R}^n$:

$$\mathrm{cm}(D, v) \; = \; \max_{d \in D} \frac{d^\top v}{\|d\|\|v\|}.$$

Recall cosine measure

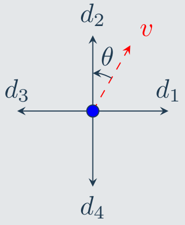
Definition (Cosine measure)

Cosine measure for a finite set of nonzero vectors $D \subseteq \mathbb{R}^n$ w.r.t. a given vector $v \in \mathbb{R}^n$:

$$\mathrm{cm}(D, v) \ = \ \max_{d \in D} \frac{d^\top v}{\|d\|\|v\|}.$$

Example

$$\mathrm{cm}(D, v) \ = \ \cos\theta$$



Recall cosine measure

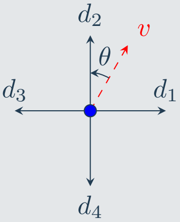
Definition (Cosine measure)

Cosine measure for a finite set of nonzero vectors $D \subseteq \mathbb{R}^n$ w.r.t. a given vector $v \in \mathbb{R}^n$:

$$\mathrm{cm}(D, v) \;=\; \max_{d \in D} \frac{d^\top v}{\|d\|\|v\|}.$$

Example

$$\mathrm{cm}(D, v) \;=\; \cos\theta$$



Measure the ability that "D approximates v"

Establishment of non-convergence

p-probabilistically κ-descent

$$\mathbb{P}\left(\mathrm{cm}\left(D_k, -g_k\right) \geq \kappa \mid D_0, \ldots, D_{k-1}\right) \geq p \quad \text{for each } k \geq 0.$$

Establishment of non-convergence

p-probabilistically κ-descent

$$\mathbb{P}\left(\mathrm{cm}\left(D_k, -g_k\right) \geq \kappa \mid D_0, \ldots, D_{k-1}\right) \geq p \quad \text{for each } k \geq 0.$$

q-probabilistically ascent

$$\mathbb{P}\left(\mathrm{cm}\left(D_k, -g_k\right) > 0 \mid D_0, \ldots, D_{k-1}\right) \leq q \quad \text{for each } k \geq 0.$$

Establishment of non-convergence

q-probabilistically ascent

$$\mathbb{P}\left(\mathrm{cm}\left(D_k,-g_k\right) > 0 \mid D_0,\ldots,D_{k-1}\right) \le q \quad \text{for each } k \geq 0.$$

$$\begin{tikzcd}
\mathrm{cm}\left(D_k,-g_k\right) \le 0 \arrow[rr] & & \text{No descent direction} \arrow[d, "f \text{ convex}"] \\
\alpha_k \text{ shrinks} & & f(x_k+\alpha_k d_k) \geq f(x_k)\ \forall d_k \in D_k \arrow[ll]
\end{tikzcd}$$

q-probabilistically ascent

$$\mathbb{P}\left(\mathrm{cm}\left(D_k, -g_k\right) > 0 \mid D_0, \ldots, D_{k-1}\right) \leq q \quad \text{for each } k \geq 0.$$

$$\begin{tikzcd}
\mathrm{cm}\left(D_k, -g_k\right) \leq 0 \arrow[r] & \text{No descent direction} \arrow[d, "f \text{ convex}"] \\
\alpha_k \text{ shrinks} & f(x_k + \alpha_k d_k) \geq f(x_k) \; \forall d_k \in D_k \arrow[l]
\end{tikzcd}$$

non-convergence for convex functions

$\Downarrow$

non-convergence in general

Establishment of non-convergence

$(D_k)_{k\in\mathbb{N}}$ is
q-probabilistically ascent

Establishment of non-convergence

$(D_k)_{k\in\mathbb{N}}$ is
q-probabilistically ascent

$\Downarrow$

α_k shrinks with lower-
bounded probability

Establishment of non-convergence

$(D_k)_{k\in\mathbb{N}}$ is
q-probabilistically ascent

$\Downarrow$

α_k shrinks with lower-
bounded probability

$\Downarrow$

$$\sum_{k=0}^{\infty} \alpha_k \;<\; \infty \quad \text{a.s. ?}$$

Establishment of non-convergence

$(D_k)_{k\in\mathbb{N}}$ is
q-probabilistically ascent

$\Downarrow$

α_k shrinks with lower-
bounded probability

$\Downarrow$

$$\sum_{k=0}^{\infty} \alpha_k \;<\; \infty \quad \text{a.s. ?}$$

$\Downarrow$

$$x_0 \;\notin\; \bar{\mathcal{B}}(x^*, \textstyle\sum_{k=0}^{\infty} \alpha_k)$$

Establishment of non-convergence

$(D_k)_{k\in\mathbb{N}}$ is
q-probabilistically ascent

⬇

α_k shrinks with lower-bounded probability

⬇

$$\sum_{k=0}^{\infty} \alpha_k \;<\; \infty \quad \text{a.s. ?}$$

⬇

$$x_0 \;\notin\; \bar{\mathcal{B}}(x^*, \textstyle\sum_{k=0}^{\infty} \alpha_k)$$

⬇

$$\mathbb{P}\,(\text{Convergence}) \;<\; 1$$

Key for analysis

- Define indicator function $Y_k = \mathbb{1}_{\{\mathrm{cm}(D_k, -g_k) > 0\}}$
  Indicator for “good” event

Key for analysis

- Define indicator function $Y_k = \mathbb{1}_{\{\mathrm{cm}(D_k, -g_k) > 0\}}$
  Indicator for “good” event
- $\alpha_{k+1} \leq \gamma^{Y_k} \theta^{1-Y_k} \alpha_k$, when f is convex

Key for analysis

- Define indicator function $Y_k = \mathbb{1}_{\{\mathrm{cm}(D_k, -g_k) > 0\}}$
  Indicator for “good” event
- $\alpha_{k+1} \leq \gamma^{Y_k}\theta^{1-Y_k}\alpha_k$, when f is convex
- $\alpha_k \leq \alpha_0 \prod_{i=0}^{k-1} \gamma^{Y_i}\theta^{1-Y_i} =: \alpha_0 U_k$

Key for analysis

- Define indicator function $Y_k = \mathbb{1}_{\{\mathrm{cm}(D_k, -g_k) > 0\}}$
  Indicator for "good" event
- $\alpha_{k+1} \leq \gamma^{Y_k} \theta^{1-Y_k} \alpha_k$, when f is convex
- $\alpha_k \leq \alpha_0 \prod_{i=0}^{k-1} \gamma^{Y_i} \theta^{1-Y_i} =: \alpha_0 U_k$
- $\sum_{k=1}^{\infty} \alpha_k \leq \alpha_0 \sum_{k=1}^{\infty} U_k$

Key for analysis

- Define indicator function $Y_k = \mathbb{1}_{\{\mathrm{cm}(D_k, -g_k) > 0\}}$
  Indicator for "good" event
- $\alpha_{k+1} \leq \gamma^{Y_k} \theta^{1-Y_k} \alpha_k$, when f is convex
- $\alpha_k \leq \alpha_0 \prod_{i=0}^{k-1} \gamma^{Y_i} \theta^{1-Y_i} =: \alpha_0 U_k$
- $\sum_{k=1}^{\infty} \alpha_k \leq \alpha_0 \sum_{k=1}^{\infty} U_k < \infty$ a.s.?

Key for analysis

- Define indicator function $Y_k = \mathbb{1}_{\{\mathrm{cm}(D_k, -g_k) > 0\}}$
  Indicator for "good" event
- $\alpha_{k+1} \le \gamma^{Y_k} \theta^{1-Y_k} \alpha_k$, when f is convex
- $\alpha_k \le \alpha_0 \prod_{i=0}^{k-1} \gamma^{Y_i} \theta^{1-Y_i} =: \alpha_0 U_k$
- $\sum_{k=1}^{\infty} \alpha_k \le \alpha_0 \sum_{k=1}^{\infty} U_k < \infty$ a.s.?

Under q-probabilistically ascent assumption, can we find a constant ζ such that

$$\mathbb{P}\left(\sum_{k=1}^{\infty} U_k < \zeta \right) \; > \; 0?$$

Main results

Assumption

$$\mathbb{P}\left(\mathrm{cm}\left(D_k, -g_k\right) \leq 0 \mid D_0, \ldots, D_{k-1}\right) \;\geq\; q \;>\; q_0 \quad \text{for each } k \geq 0,$$

where $q_0 = 1 - p_0 = \log\gamma / \log(\theta^{-1}\gamma)$.

Main results

Assumption

$$\mathbb{P}\left(Y_k = 0 \mid Y_0, \ldots, Y_{k-1}\right) \;\geq\; q \;>\; q_0 \quad \text{for each } k \geq 0,$$

where $q_0 = 1 - p_0 = \log\gamma / \log(\theta^{-1}\gamma)$.

Main results

Assumption

$$\mathbb{P}\left(Y_k = 0 \mid Y_0, \ldots, Y_{k-1}\right) \;\geq\; q \;>\; q_0 \quad \text{for each } k \geq 0,$$

where $q_0 = 1 - p_0 = \log\gamma / \log(\theta^{-1}\gamma)$.

Result

1.
$$\mathbb{P}\left(\sum_{k=1}^{\infty} U_k < \infty\right) \;=\; 1$$

2.
$$\mathbb{P}\left(\sum_{k=1}^{\infty} U_k < \zeta\right) \;>\; 0 \quad \Longleftrightarrow \quad \zeta > \frac{\theta}{1-\theta}$$

Main results

Assumption

$$\mathbb{P}\left(Y_k = 0 \mid Y_0, \ldots, Y_{k-1}\right) \;\geq\; q \;>\; q_0 \quad \text{for each } k \geq 0,$$

where $q_0 = 1 - p_0 = \log\gamma / \log(\theta^{-1}\gamma)$.

Result

1. $$\mathbb{P}\left(\sum_{k=1}^{\infty} U_k < \infty\right) \;=\; 1$$
2. $$\mathbb{P}\left(\sum_{k=1}^{\infty} U_k < \zeta\right) \;>\; 0 \quad \iff \quad \zeta > \frac{\theta}{1-\theta}$$

Note that $\sum_{k=1}^{\infty} U_k = \sum_{k=1}^{\infty} \prod_{i=0}^{k-1} \gamma^{Y_i}\theta^{1-Y_i} \geq \theta/(1-\theta)$

Main results

Assumption

$$\mathbb{P}\left(Y_k = 0 \mid Y_0, \ldots, Y_{k-1}\right) \;\geq\; q \;>\; q_0 \quad \text{for each } k \geq 0,$$

where $q_0 = 1 - p_0 = \log\gamma / \log(\theta^{-1}\gamma)$.

Result

1. $$\mathbb{P}\left(\sum_{k=1}^{\infty} U_k < \infty\right) \;=\; 1$$
2. $$\cancel{\mathbb{P}\left(\sum_{k=1}^{\infty} U_k < \zeta\right) \;>\; 0 \quad \Longleftrightarrow \quad \zeta > \frac{\theta}{1-\theta}}$$

Main results

Assumption

$$\liminf_{k\to\infty} \mathbb{P}\left(Y_k = 0 \mid Y_0, \ldots, Y_{k-1}\right) \;>\; q_0,$$

where $q_0 = 1 - p_0 = \log\gamma / \log(\theta^{-1}\gamma)$.

Result

1.
$$\mathbb{P}\left(\sum_{k=1}^{\infty} U_k < \infty\right) \;=\; 1$$

2.
$$\cancel{\mathbb{P}\left(\sum_{k=1}^{\infty} U_k < \zeta\right) \;>\; 0 \quad\Longleftrightarrow\quad \zeta > \frac{\theta}{1-\theta}}$$

Tightness

Almost zero gap

Let $D_k = \{d_1, \ldots, d_m\}$, where $d_i \overset{\text{i.i.d.}}{\sim}$ Unif$(\mathcal{S}^{n-1})$.

Recall that RDS is convergent if

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

With our non-convergence analysis, RDS is non-convergent if

Almost zero gap

Let $D_k = \{d_1, \dots, d_m\}$, where $d_i \overset{\text{i.i.d.}}{\sim} \text{Unif}(\mathcal{S}^{n-1})$.

Recall that RDS is convergent if

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

With our non-convergence analysis, RDS is non-convergent if

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

Tightness of assumption

Natural question:

$$\mathbb{P}\left(\mathrm{cm}\left(D_k, -g_k\right) \leq 0 \mid D_0, \ldots, D_{k-1}\right) \ \geq \ q \cancel{>} \geq q_0,$$

Tightness of assumption

Natural question:

$$\mathbb{P}\left(\mathrm{cm}\left(D_k, -g_k\right) \le 0 \mid D_0, \ldots, D_{k-1}\right) \;\ge\; q \cancel{>} \ge q_0,$$

Answer: NO

Tightness of assumption

Natural question:

$$\mathbb{P}\left(\mathrm{cm}\left(D_k, -g_k\right) \leq 0 \mid D_0, \ldots, D_{k-1}\right) \;\geq\; q \cancel{>} \geq q_0,$$

Answer: NO

Example

We assume

- $f : \mathbb{R}^n \to \mathbb{R}$ be L-smooth and strongly convex,
- $\theta = 1/2$ and $\gamma = 2$, $\Rightarrow q_0 = 1/2$
- $D_k = \{g_k/\|g_k\|\}$ or $D_k = \{-g_k/\|g_k\|\}$ with probability $1/2$, respectively,

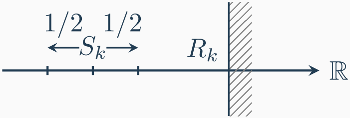
then we have

$$\mathbb{P}\left(\lim_{k\to\infty} \|g_k\| = 0\right) \;=\; 1.$$

Transformed to a special random walk problem

Two stochastic processes
$(R_k)_{k\in\mathbb{N}}$ uniformly upper bounded, $(S_k)_{k\in\mathbb{N}}$ a special random walk satisfying:

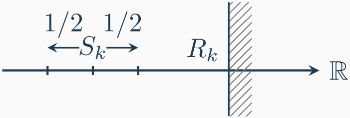
When $S_k < R_k$,



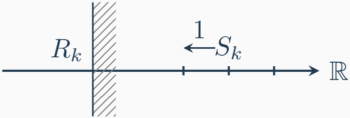
Transformed to a special random walk problem

Two stochastic processes
$(R_k)_{k\in\mathbb{N}}$ uniformly upper bounded, $(S_k)_{k\in\mathbb{N}}$ a special random walk satisfying:

When $S_k < R_k$,



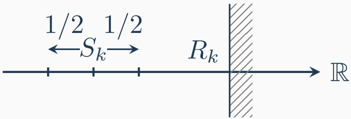
When $S_k \geq R_k$,



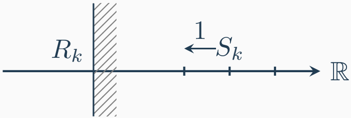
Transformed to a special random walk problem

Two stochastic processes
$(R_k)_{k\in\mathbb{N}}$ uniformly upper bounded, $(S_k)_{k\in\mathbb{N}}$ a special random walk satisfying:

When $S_k < R_k$,



When $S_k \geq R_k$,



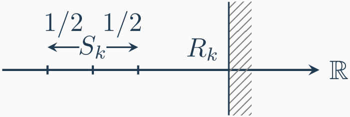
Does $\|g_k\| \to 0$?

Transformed to a special random walk problem

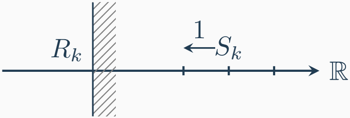
Two stochastic processes

$(R_k)_{k\in\mathbb{N}}$ uniformly upper bounded, $(S_k)_{k\in\mathbb{N}}$ a special random walk satisfying:

When $S_k < R_k$,



When $S_k \geq R_k$,

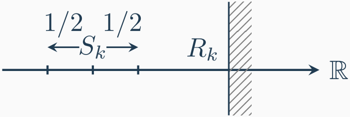


Does S_k go beyond the “wall” R_k i.o. with probability 1?

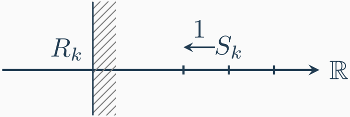
Transformed to a special random walk problem

Two stochastic processes
$(R_k)_{k\in\mathbb{N}}$ uniformly upper bounded, $(S_k)_{k\in\mathbb{N}}$ a special random walk satisfying:

When $S_k < R_k$,



When $S_k \geq R_k$,



Does S_k go beyond the “wall” R_k i.o. with probability 1? YES

Call for ideas!

Two remaining interesting questions

- RDS converges or not when $\log_2(1 - \log\theta / \log\gamma) \in \mathbb{N}_+$,

  especially when $\gamma = 1/\theta = 2$ and $m = 1$.
- Estimate the CDF of $\sum_{k=1}^{\infty} U_k$

$$F(x) = \mathbb{P}\left(\sum_{k=1}^{\infty}\prod_{i=0}^{k-1}\gamma^{Y_i}\theta^{1-Y_i} \le x\right),$$

where $Y_i \overset{\text{i.i.d.}}{\sim} \text{Bern}(p)$.

Thank you!

References I

- Conn, A. R., Scheinberg, K., and Vicente, L. N. (2009). *Introduction to Derivative-Free Optimization*. Vol. 8. MOS-SIAM Ser. Optim. Philadelphia: SIAM.
- Durrett, R. (2010). *Probability: Theory and Examples*. Fourth. Camb. Ser. Stat. Probab. Math. Cambridge: Cambridge University Press.
- Gratton, S. et al. (2015). “Direct search based on probabilistic descent”. *SIAM J. Optim.* 25, pp. 1515–1541.