# The Worst Case Complexity of Direct Search and the Unexpected Mathematics in It

Zaikun Zhang

IRIT-ENSEEEIHT, Toulouse, France

Joint works with M. Dodangeh, S. Gratton, C. W. Royer, and L. N. Vicente

August 2015, Beijing

# Problem setting

## Unconstrained derivative-free optimization (DFO)

$$\min_{x \in \mathbb{R}^n} f(x)$$

$$f : \mathbb{R}^n \to \mathbb{R}$$

$f$ is bounded from below and differentiable

$\nabla f$ is Lipschitz continuous but unavailable

# Problem setting

**Unconstrained derivative-free optimization (DFO)**

$$\min_{x \in \mathbb{R}^n} f(x)$$

$$f : \mathbb{R}^n \to \mathbb{R}$$

$f$ is bounded from below and differentiable

$\nabla f$ is Lipschitz continuous but unavailable

- Many real-world problems: derivatives are expensive or unreliable.

# Problem setting

**Unconstrained derivative-free optimization (DFO)**

$$\min_{x \in \mathbb{R}^n} f(x)$$

$$f : \mathbb{R}^n \to \mathbb{R}$$

$f$ is bounded from below and differentiable

$\nabla f$ is Lipschitz continuous but unavailable

- Many real-world problems: derivatives are expensive or unreliable.
- S. Gratton, P. Laloyaux, and A. Sartenaer, "Derivative-free Optimization for Large-scale Nonlinear Data Assimilation Problems", *Quarterly Journal of the Royal Meteorological Society*, 140: 943-957, 2014

# Derivative-free optimization: What is desirable?

Algorithms that

- do not use derivatives, and

# Derivative-free optimization: What is desirable?

Algorithms that

- do not use derivatives, and
- use function evaluations as few as possible.

# Derivative-free optimization: What is desirable?

Algorithms that

- do not use derivatives, and
- use function evaluations as few as possible.

And convergence theory of the algorithms.

# Existing methods

- Two main classes of rigorous methods in DFO
  - Directional methods, like direct search (`GPS`, `GSS`, `MADS` ...)
  - Model-based methods, like trust region methods (`DFO`, `NEWUOA`, `CONDER`, `BOOSTER`, `ORBIT` ...)

# Direct search — an example

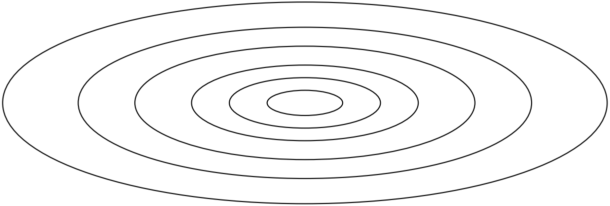
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.



Contour of an objective function

# Direct search — an example

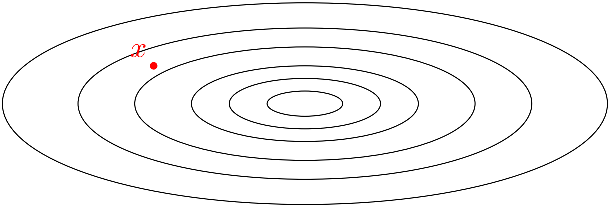
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.




Contour of an objective function

# Direct search — an example

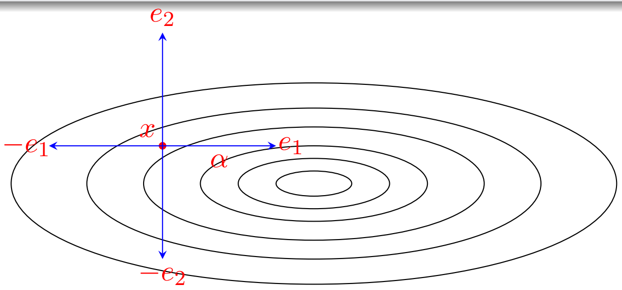
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \dots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.




Contour of an objective function

# Direct search — an example

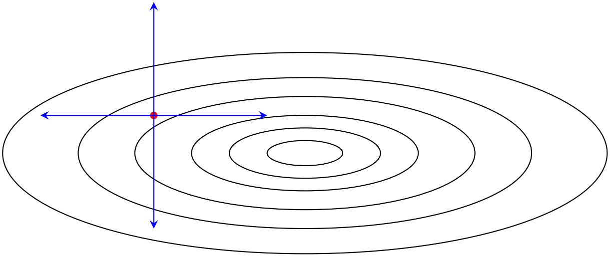
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.



Contour of an objective function

# Direct search — an example

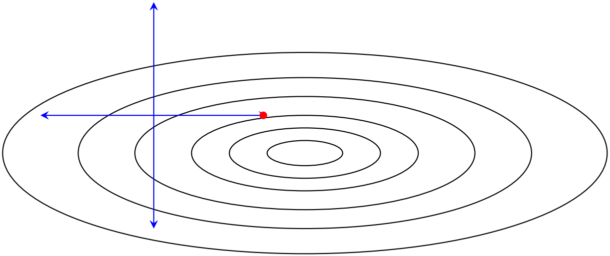
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.



Contour of an objective function

# Direct search — an example

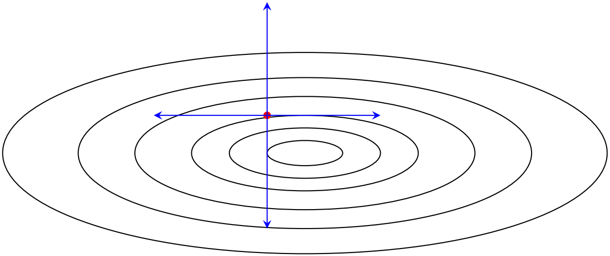
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.



Contour of an objective function

# Direct search — an example

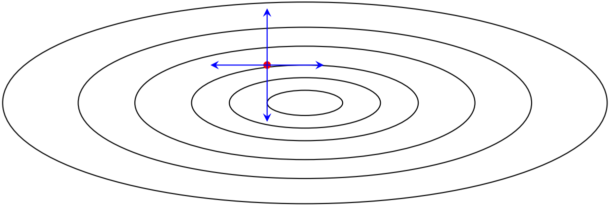
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.



Contour of an objective function

# Direct search — an example

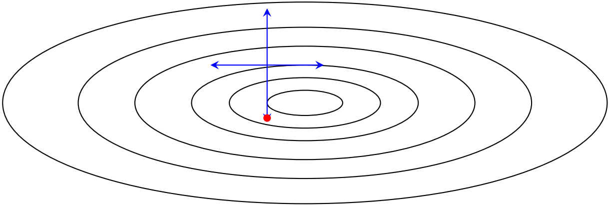
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.



Contour of an objective function

# Direct search — an example

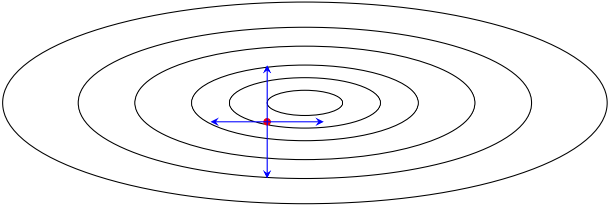
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.



Contour of an objective function

# Direct search — an example

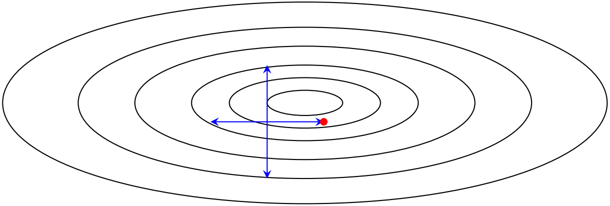
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \dots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.



Contour of an objective function

# Direct search — an example

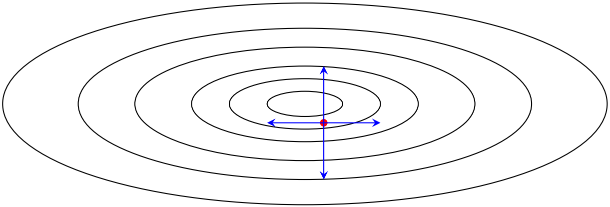
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \dots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.



Contour of an objective function

# Direct search — an example

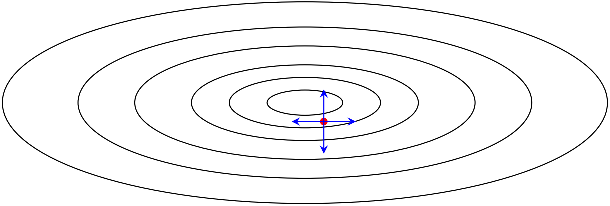
## A classical direct search algorithm (Coordinate Search)

**Input** Starting point $x$ and initial step size $\alpha$.

**Repeat** Check whether there exists $d \in \{\pm e_1, \pm e_2, \ldots, \pm e_n\}$ such that

$$f(x + \alpha d) < f(x) - \alpha^2/2.$$

If yes, $x := x + \alpha d$ and possibly expand $\alpha$; if no, contract $\alpha$.



Contour of an objective function

# Direct search — a general framework

**Choose:** $x_0$, $\alpha_0$, $\gamma \in [1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

# Direct search — a general framework

**Choose:** $x_0$, $\alpha_0$, $\gamma \in [1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

**For** $k = 0, 1, 2, \ldots$

- **Polling:** Select a polling set $D_k$ of directions, and seek $d_k \in D_k$:

$$f(x_k + \alpha_k d_k) \ < \ f(x_k) - \rho(\alpha_k).$$

# Direct search — a general framework

**Choose:** $x_0$, $\alpha_0$, $\gamma \in [1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

**For** $k = 0, 1, 2, \dots$

- **Polling:** Select a polling set $D_k$ of directions, and seek $d_k \in D_k$:

$$f(x_k + \alpha_k d_k) \ < \ f(x_k) - \rho(\alpha_k).$$

  If $d_k$ is found, the iteration is successful. Otherwise, it is unsuccessful.

# Direct search — a general framework

**Choose:** $x_0$, $\alpha_0$, $\gamma \in [1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

**For** $k = 0, 1, 2, \ldots$

- **Polling:** Select a polling set $D_k$ of directions, and seek $d_k \in D_k$:

$$f(x_k + \alpha_k d_k) \; < \; f(x_k) - \rho(\alpha_k).$$

  If $d_k$ is found, the iteration is successful. Otherwise, it is unsuccessful.

- **Update:**

$$x_{k+1} = \begin{cases} x_k + \alpha_k d_k & \text{if successful} \\ x_k & \text{if unsuccessful,} \end{cases}$$

# Direct search — a general framework

**Choose:** $x_0$, $\alpha_0$, $\gamma \in [1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

**For** $k = 0, 1, 2, \ldots$

- **Polling:** Select a polling set $D_k$ of directions, and seek $d_k \in D_k$:

$$f(x_k + \alpha_k d_k) \; < \; f(x_k) - \rho(\alpha_k).$$

  If $d_k$ is found, the iteration is successful. Otherwise, it is unsuccessful.

- **Update:**

$$x_{k+1} = \begin{cases} x_k + \alpha_k d_k & \text{if successful} \\ x_k & \text{if unsuccessful,} \end{cases}$$

$$\alpha_{k+1} = \begin{cases} \gamma \alpha_k & \text{if successful} \\ \theta \alpha_k & \text{if unsuccessful.} \end{cases}$$

# Direct search — a general framework

**Choose:** $x_0$, $\alpha_0$, $\gamma \in [1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

**For** $k = 0, 1, 2, \ldots$

- **Polling:** Select a polling set $D_k$ of directions, and seek $d_k \in D_k$:

$$f(x_k + \alpha_k d_k) \ < \ f(x_k) - \rho(\alpha_k).$$

  If $d_k$ is found, the iteration is successful. Otherwise, it is unsuccessful.

- **Update:**

$$x_{k+1} = \begin{cases} x_k + \alpha_k d_k & \text{if successful} \\ x_k & \text{if unsuccessful,} \end{cases}$$

$$\alpha_{k+1} = \begin{cases} \gamma \alpha_k & \text{if successful} \\ \theta \alpha_k & \text{if unsuccessful.} \end{cases}$$

# More ...

A forcing function $\rho$ is a positive and monotonically nondecreasing function such that

$$\lim_{\alpha \downarrow 0} \frac{\rho(\alpha)}{\alpha} \ = \ 0.$$

# More ...

A forcing function $\rho$ is a positive and monotonically nondecreasing function such that

$$\lim_{\alpha \downarrow 0} \frac{\rho(\alpha)}{\alpha} = 0.$$

For simplicity, in this talk:

$$\begin{aligned} \rho(\alpha) &= \frac{\alpha^2}{2} \\ \alpha_0 &= 1 && \text{(initial stepsize)} \\ \gamma &= 2 && \text{(increasing factor)} \\ \theta &= \frac{1}{2} && \text{(decreasing factor)} \end{aligned}$$

# Traditional polling set: PSS

- Positive spanning set (PSS):

  $D = \{d_1, \ldots, d_m\}$ is a PSS if it spans $\mathbb{R}^n$ positively:

$$\mathbb{R}^n = \left\{ \sum_{i=1}^{m} \mu_i d_i : \mu_i \geq 0 \ (1 \leq i \leq m) \right\}.$$

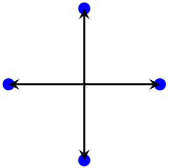
# Traditional polling set: PSS

- Positive spanning set (PSS):

  $D = \{d_1, \ldots, d_m\}$ is a PSS if it spans $\mathbb{R}^n$ positively:

$$\mathbb{R}^n = \left\{ \sum_{i=1}^{m} \mu_i d_i : \mu_i \geq 0 \ (1 \leq i \leq m) \right\}.$$

Example:



$$D_\oplus = \{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$$

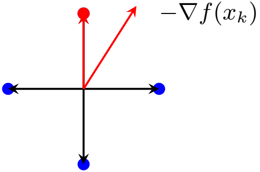
# Traditional polling set: PSS

- Positive spanning set (PSS):

  $D = \{d_1, \ldots, d_m\}$ is a PSS if it spans $\mathbb{R}^n$ positively:

$$\mathbb{R}^n = \left\{ \sum_{i=1}^{m} \mu_i d_i : \mu_i \geq 0 \ (1 \leq i \leq m) \right\}.$$

  Example:




$$D_\oplus = \{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$$

- $\exists\, d \in D$ that 'approximates' $-\nabla f(x_k)$, meaning $d^\top[-\nabla f(x_k)] > 0$.

# Direct search with PSS: Global convergence

**Global convergence (Torczon 1997, Kolda, Lewis, and Torczon 2003)**

*If $\{D_k\}$ is a sequence of PSSs with 'uniformly good quality', then*

$$\liminf_{k\to\infty} \|\nabla f(x_k)\| = 0.$$

# The quality of a PSS: Cosine measure

- Cosine measure: the ability of $D$ to 'approximate' directions in $\mathbb{R}^n$.

$$\mathrm{cm}(D) \;=\; \min_{0 \neq v \in \mathbb{R}^n} \max_{d \in D} \frac{d^\top v}{\|d\| \|v\|}.$$

- Cosine of the largest angle 'between $D$ and the vectors in $\mathbb{R}^n$'.

# The quality of a PSS: Cosine measure

- Cosine measure: the ability of $D$ to 'approximate' directions in $\mathbb{R}^n$.

$$\mathrm{cm}(D) \; = \; \min_{0 \neq v \in \mathbb{R}^n} \max_{d \in D} \frac{d^\top v}{\|d\| \|v\|}.$$

- Cosine of the largest angle 'between $D$ and the vectors in $\mathbb{R}^n$'.

# The quality of a PSS: Cosine measure

- Cosine measure: the ability of $D$ to 'approximate' directions in $\mathbb{R}^n$.

$$\mathrm{cm}(D) \;=\; \min_{0 \neq v \in \mathbb{R}^n} \max_{d \in D} \frac{d^\top v}{\|d\|\|v\|}.$$

- Cosine of the largest angle 'between $D$ and the vectors in $\mathbb{R}^n$'.

- For each $v \in \mathbb{R}^n$, there exists $d \in D$ such that

$$d^\top v \;\geq\; \mathrm{cm}(D)\|d\|\|v\|.$$

We are particularly interested in the case with $v = -\nabla f(x_k)$.

# The quality of a PSS: Cosine measure

- Cosine measure: the ability of $D$ to 'approximate' directions in $\mathbb{R}^n$.

$$\mathrm{cm}(D) \;=\; \min_{0\neq v\in\mathbb{R}^n} \max_{d\in D} \frac{d^\top v}{\|d\|\|v\|}.$$

- Cosine of the largest angle 'between $D$ and the vectors in $\mathbb{R}^n$'.

- For each $v \in \mathbb{R}^n$, there exists $d \in D$ such that

$$d^\top v \;\geq\; \mathrm{cm}(D)\|d\|\|v\|.$$

  We are particularly interested in the case with $v = -\nabla f(x_k)$.

- Example:

$$\mathrm{cm}(D_\oplus) \;=\; \frac{1}{\sqrt{n}}.$$

# Direct search with PSS: Worst case complexity (WCC)

## Worst case complexity for iterations (Vicente 2013)

*If*

$$\mathrm{cm}(D_k) \;\geq \kappa \; > 0,$$

*then*

- $\min_{0 \leq \ell \leq k} \|\nabla f(x_\ell)\| \leq \mathcal{O}(\kappa^{-1} k^{-1/2})$,
- $\|\nabla f(x_k)\|$ *is driven under* $\epsilon$ *within* $\mathcal{O}(\kappa^{-2}\epsilon^{-2})$ *iterations.*

# Direct search with PSS: Worst case complexity (WCC)

## Worst case complexity for iterations (Vicente 2013)

*If*

$$\mathrm{cm}(D_k) \;\geq \kappa \; > 0,$$

*then*

- $\min_{0\leq \ell \leq k} \|\nabla f(x_\ell)\| \leq \mathcal{O}(\kappa^{-1} k^{-1/2})$,
- $\|\nabla f(x_k)\|$ *is driven under* $\epsilon$ *within* $\mathcal{O}(\kappa^{-2}\epsilon^{-2})$ *iterations.*

## Worst case complexity for function evalautions (Vicente 2013)

*If*

$$\mathrm{cm}(D_k) \;\geq \kappa \; > 0 \quad \textit{and} \quad |D_k| \leq m,$$

*then* $\|\nabla f(x_k)\|$ *is driven under* $\epsilon$ *within* $\mathcal{O}(m\kappa^{-2}\epsilon^{-2})$ *function evalations.*

# Direct search with PSS: Worst case complexity (WCC)

## Worst case complexity for iterations (Vicente 2013)

*If*

$$\mathrm{cm}(D_k) \;\geq \kappa \;> 0,$$

*then*

- $\min_{0 \leq \ell \leq k} \|\nabla f(x_\ell)\| \leq \mathcal{O}(\kappa^{-1} k^{-1/2})$,
- $\|\nabla f(x_k)\|$ *is driven under* $\epsilon$ *within* $\mathcal{O}(\kappa^{-2}\epsilon^{-2})$ *iterations.*

## Worst case complexity for function evalautions (Vicente 2013)

*If*

$$\mathrm{cm}(D_k) \;\geq \kappa \;> 0 \quad \textit{and} \quad |D_k| \leq m,$$

*then* $\|\nabla f(x_k)\|$ *is driven under* $\epsilon$ *within* $\mathcal{O}(m\kappa^{-2}\epsilon^{-2})$ *function evalations.*

*In particular, if* $D_k \equiv D_\oplus$*, then the complexity bound is* $\mathcal{O}(n^2\epsilon^{-2})$.

# Direct search with PSS: Worst case complexity (WCC)

## Worst case complexity for iterations (Vicente 2013)

*If*

$$\mathrm{cm}(D_k) \;\geq \kappa \; > 0,$$

*then*

- $\min_{0 \leq \ell \leq k} \|\nabla f(x_\ell)\| \leq \mathcal{O}(\kappa^{-1} k^{-1/2})$,
- $\|\nabla f(x_k)\|$ *is driven under* $\epsilon$ *within* $\mathcal{O}(\kappa^{-2}\epsilon^{-2})$ *iterations.*

## Worst case complexity for function evalautions (Vicente 2013)

*If*

$$\mathrm{cm}(D_k) \;\geq \kappa \; > 0 \quad \textit{and} \quad |D_k| \leq m,$$

*then* $\|\nabla f(x_k)\|$ *is driven under* $\epsilon$ *within* $\mathcal{O}(m\kappa^{-2}\epsilon^{-2})$ *function evalations.*

*In particular, if* $D_k \equiv D_\oplus$*, then the complexity bound is* $\mathcal{O}(n^2\epsilon^{-2})$.

Question: How to choose $D_k$ to minimize the WCC for function evalations?

# The optimal PSS?

To find the PSS that minimizes the bound $\mathcal{O}(m\kappa^{-2}\epsilon^{-2})$, we have to solve

$$\begin{aligned} \min_{D\in\mathcal{D}} \quad & m\kappa^{-2} \\ \text{s.t.} \quad & \operatorname{cm}(D) \geq \kappa, \\ & |D| \leq m, \end{aligned}$$

where $\mathcal{D}$ is the set consisting of all the PSSs in $\mathbb{R}^n$.

# The optimal PSS?

To find the PSS that minimizes the bound $\mathcal{O}(m\kappa^{-2}\epsilon^{-2})$, we have to solve

$$\begin{aligned} \min_{D\in\mathcal{D}} \quad & m\kappa^{-2} \\ \text{s.t.} \quad & \mathrm{cm}(D) \geq \kappa, \\ & |D| \leq m, \end{aligned}$$

where $\mathcal{D}$ is the set consisting of all the PSSs in $\mathbb{R}^n$.

It is equivalent to solve

$$\min_{D\in\mathcal{D}} \quad \frac{|D|}{\mathrm{cm}^2(D)}.$$

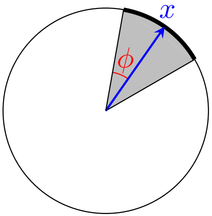
# Cosine measure and sphere covering

Suppose that $D$ is a PSS consisting of unit vectors. Recall that

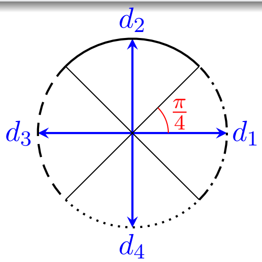
$$\mathrm{cm}(D) \;=\; \min_{\|v\|=1}\max_{d\in D} d^\top v.$$

## Lemma

*Let $\mathbb{C}(d,\phi)$ be the spherical cap centered at $d$ with geodesic radius $\phi$, and $\mathrm{cm}(D)=\kappa$, then $\mathbb{S}^{n-1} \subseteq \bigcup_{d\in D}\mathbb{C}(d,\arccos\kappa)$.*




Spherical cap $\mathbb{C}(d,\phi)$




$\mathbb{S}^1 \subseteq \bigcup_{d\in D_\oplus}\mathbb{C}(d,\pi/4)$

# A sphere covering problem

Solving

$$\min_{D \in \mathcal{D}} \ \frac{|D|}{\mathrm{cm}^2(D)}$$

is equivalent to solve

$$\begin{aligned} \min \ & \frac{|D|}{\cos^2 \phi} \\ \text{s.t.} \ & \mathbb{S}^{n-1} \subseteq \bigcup_{d \in D} \mathbb{C}(d, \phi). \end{aligned}$$

# A sphere covering problem

Solving

$$\min_{D \in \mathcal{D}} \ \frac{|D|}{\mathrm{cm}^2(D)}$$

is equivalent to solve

$$\begin{aligned} \min \ & \frac{|D|}{\cos^2 \phi} \\ \text{s.t.} \ & \mathbb{S}^{n-1} \subseteq \bigcup_{d \in D} \mathbb{C}(d, \phi). \end{aligned}$$

One possible approch is to study

$$\begin{aligned} \min_{|D|=m} \ & \phi \\ \text{s.t.} \ & \mathbb{S}^{n-1} \subseteq \bigcup_{d \in D} \mathbb{C}(d, \phi). \end{aligned}$$

$\Longrightarrow$ What is the most 'economical' covering of $\mathbb{S}^{n-1}$ by $m$ identical caps?

# A sphere covering problem

- Good news: Sphere covering is a classical topic of Discrete Geometry.

# A sphere covering problem

- Good news: Sphere covering is a classical topic of Discrete Geometry.
- Bad news: The most 'economical' covering is still unknown.

# A sphere covering problem

- Good news: Sphere covering is a classical topic of Discrete Geometry.
- Bad news: The most 'economical' covering is still unknown.
- Worse news: It is unknown even when $m = 2n$ $(n \geq 5)$.

# A sphere covering problem

- Good news: Sphere covering is a classical topic of Discrete Geometry.

- Bad news: The most 'economical' covering is still unknown.

- Worse news: It is unknown even when $m = 2n$ $(n \geq 5)$.

  When $m = 2n$, it is reasonable to conjecture that $D_{\oplus}$ gives the most economical covering, but the proof is non-trivial even for $n = 3$.

  
  - K. Böröczky, Jr. *Finite Packing and Covering*, Cambridge University Press, New York, 2004
  - K. Böröczky, Jr. and G. Wintsche, "Covering the sphere by equal spherical balls", In *Discrete and Computational Geometry*, volume 25 of Algorithms and Combinatorics, pages 235–251. Springer Berlin, 2003
  

# A sphere covering problem

- Good news: Sphere covering is a classical topic of Discrete Geometry.
- Bad news: The most 'economical' covering is still unknown.
- Worse news: It is unknown even when $m = 2n$ $(n \geq 5)$.

  When $m = 2n$, it is reasonable to conjecture that $D_{\oplus}$ gives the most economical covering, but the proof is non-trivial even for $n = 3$.

  - K. Böröczky, Jr. *Finite Packing and Covering*, Cambridge University Press, New York, 2004
  - K. Böröczky, Jr. and G. Wintsche, "Covering the sphere by equal spherical balls", In *Discrete and Computational Geometry*, volume 25 of Algorithms and Combinatorics, pages 235–251. Springer Berlin, 2003

Hopeless?

# A recent bound

**Lemma** (Tikhomirov 2014)

*Any covering of* $\mathbb{S}^{n-1}$ *by* $m \geq n+1$ *spherical caps of geodesic radius* $\phi$ *satisfies*

$$\cos\phi \;\leq\; \zeta\sqrt{n^{-1}\log(n^{-1}m)}$$

*for some universal constant* $\zeta > 0$.

# A recent bound

**Lemma** (Tikhomirov 2014)

*Any covering of* $\mathbb{S}^{n-1}$ *by* $m \geq n+1$ *spherical caps of geodesic radius* $\phi$ *satisfies*

$$\cos\phi \;\leq\; \zeta\sqrt{n^{-1}\log(n^{-1}m)}$$

*for some universal constant* $\zeta > 0$.

**Corollary** (Dodangeh, Vicente, Zhang 2014)

- *If* $\mathbb{S}^{n-1} \subseteq \bigcup_{d\in D} \mathbb{C}(d,\phi)$, *then*

$$\frac{|D|}{\cos^2\phi} \;\geq\; \zeta^{-2}n^2.$$

- *For each PSS* $D$ *in* $\mathbb{R}^n$,

$$\frac{|D|}{\mathrm{cm}^2(D)} \;\geq\; \zeta^{-2}n^2.$$

# The optimality of $D_\oplus$

## Worst case complexity for function evalautions (Recalling)

*If*

$$\mathrm{cm}(D_k) \ \geq \kappa \ > 0 \quad \textit{and} \quad |D_k| \leq m,$$

*then* $\|\nabla f(x_k)\|$ *is driven under* $\epsilon$ *within* $\mathcal{O}(m\kappa^{-2}\epsilon^{-2})$ *function evalations.*

*In particular, if* $D_k \equiv D_\oplus$*, then the complexity bound is* $\mathcal{O}(n^2\epsilon^{-2})$*.*

# The optimality of $D_\oplus$

## Worst case complexity for function evalautions (Recalling)

*If*

$$\mathrm{cm}(D_k) \ \geq \kappa \ > 0 \quad \textit{and} \quad |D_k| \leq m,$$

*then* $\|\nabla f(x_k)\|$ *is driven under* $\epsilon$ *within* $\mathcal{O}(m\kappa^{-2}\epsilon^{-2})$ *function evalations.*

*In particular, if* $D_k \equiv D_\oplus$*, then the complexity bound is* $\mathcal{O}(n^2\epsilon^{-2})$.

## Theorem (Dodangeh, Vicente, Zhang 2014)

*In the theorem above, it holds that*

$$m\kappa^{-2} \geq \zeta^{-2} n^2.$$

*Hence the complexity bound achives its optimal order when* $D_k \equiv D_\oplus$.

# The optimality of $D_\oplus$

**Worst case complexity for function evalautions (Recalling)**

*If*

$$\mathrm{cm}(D_k) \geq \kappa > 0 \quad \textit{and} \quad |D_k| \leq m,$$

*then* $\|\nabla f(x_k)\|$ *is driven under* $\epsilon$ *within* $\mathcal{O}(m\kappa^{-2}\epsilon^{-2})$ *function evalations.*

*In particular, if* $D_k \equiv D_\oplus$*, then the complexity bound is* $\mathcal{O}(n^2\epsilon^{-2})$*.*

**Theorem (Dodangeh, Vicente, Zhang 2014)**

*In the theorem above, it holds that*

$$m\kappa^{-2} \geq \zeta^{-2} n^2.$$

*Hence the complexity bound achives its optimal order when* $D_k \equiv D_\oplus$*.*

Is it possible to do even better than $D_\oplus$?

# The optimality of $D_\oplus$

**Worst case complexity for function evalautions (Recalling)**

*If*

$$\mathrm{cm}(D_k) \;\geq \kappa \;> 0 \quad \textit{and} \quad |D_k| \leq m,$$

*then* $\|\nabla f(x_k)\|$ *is driven under* $\epsilon$ *within* $\mathcal{O}(m\kappa^{-2}\epsilon^{-2})$ *function evalations.*

*In particular, if* $D_k \equiv D_\oplus$*, then the complexity bound is* $\mathcal{O}(n^2\epsilon^{-2})$*.*

**Theorem (Dodangeh, Vicente, Zhang 2014)**

*In the theorem above, it holds that*

$$m\kappa^{-2} \geq \zeta^{-2} n^2.$$

*Hence the complexity bound achives its optimal order when* $D_k \equiv D_\oplus$*.*

Is it possible to do even better than $D_\oplus$? Yes, by randomization.

# A competition

Relative performance: PSS v.s. Random polling sets ($n = 40$)

| | $D_\oplus$ | $2n$ | $n+1$ | $n/4$ | $2$ | $1$ |
|---|---|---|---|---|---|---|
| arglina | 3.42 | 10.30 | 6.01 | 1.88 | 1.00 | – |
| arglinb | 20.50 | 7.38 | 2.81 | 1.85 | 1.00 | 2.04 |
| broydn3d | 4.33 | 6.54 | 3.59 | 1.28 | 1.00 | – |
| dqrtic | 7.16 | 9.10 | 4.56 | 1.70 | 1.00 | – |
| engval1 | 10.53 | 11.90 | 6.48 | 2.08 | 1.00 | 2.08 |
| freuroth | 56.00 | 1.00 | 1.67 | 1.67 | 1.00 | 4.00 |
| integreq | 16.04 | 12.44 | 6.76 | 2.04 | 1.00 | – |
| nondquar | 6.90 | 7.56 | 4.23 | 1.87 | 1.00 | – |
| sinquad | – | 1.65 | 2.01 | 1.00 | 1.55 | – |
| vardim | 1.00 | 1.80 | 2.40 | 1.80 | 1.80 | 4.30 |

Solution accuracy was $10^{-3}$. Averages were taken over $10$ independent runs.

# Direct search — recalling

**Choose:** $x_0$, $\alpha_0$, $\gamma \in (1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

# Direct search — recalling

**Choose:** $x_0$, $\alpha_0$, $\gamma \in (1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

**For** $k = 0, 1, 2, \ldots$

- **Polling:** Select a polling set $D_k$ of directions, and seek $d_k \in D_k$:

$$f(x_k + \alpha_k d_k) \ < \ f(x_k) - \rho(\alpha_k).$$

# Direct search — recalling

**Choose:** $x_0$, $\alpha_0$, $\gamma \in (1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

**For** $k = 0, 1, 2, \ldots$

- **Polling:** Select a polling set $D_k$ of directions, and seek $d_k \in D_k$:

$$f(x_k + \alpha_k d_k) \; < \; f(x_k) - \rho(\alpha_k).$$

  If $d_k$ is found, the iteration is successful. Otherwise, it is unsuccessful.

# Direct search — recalling

**Choose:** $x_0$, $\alpha_0$, $\gamma \in (1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

**For** $k = 0, 1, 2, \ldots$

- **Polling:** Select a polling set $D_k$ of directions, and seek $d_k \in D_k$:

$$f(x_k + \alpha_k d_k) \; < \; f(x_k) - \rho(\alpha_k).$$

  If $d_k$ is found, the iteration is successful. Otherwise, it is unsuccessful.

- **Update:**

$$x_{k+1} = \begin{cases} x_k + \alpha_k d_k & \text{if successful} \\ x_k & \text{if unsuccessful,} \end{cases}$$

## Direct search — recalling

**Choose:** $x_0$, $\alpha_0$, $\gamma \in (1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

**For** $k = 0, 1, 2, \ldots$

- **Polling:** Select a polling set $D_k$ of directions, and seek $d_k \in D_k$:

$$f(x_k + \alpha_k d_k) \; < \; f(x_k) - \rho(\alpha_k).$$

  If $d_k$ is found, the iteration is successful. Otherwise, it is unsuccessful.

- **Update:**

$$x_{k+1} = \begin{cases} x_k + \alpha_k d_k & \text{if successful} \\ x_k & \text{if unsuccessful,} \end{cases}$$

$$\alpha_{k+1} = \begin{cases} \gamma \alpha_k & \text{if successful} \\ \theta \alpha_k & \text{if unsuccessful.} \end{cases}$$

# Direct search — recalling

**Choose:** $x_0$, $\alpha_0$, $\gamma \in (1, \infty)$, $\theta \in (0, 1)$, and a forcing function $\rho$.

**For** $k = 0, 1, 2, \ldots$

- **Polling:** Select a polling set $D_k$ of directions, and seek $d_k \in D_k$:

$$f(x_k + \alpha_k d_k) \;<\; f(x_k) - \rho(\alpha_k).$$

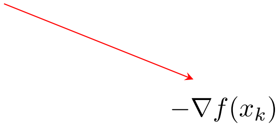
  If $d_k$ is found, the iteration is successful. Otherwise, it is unsuccessful.

- **Update:**

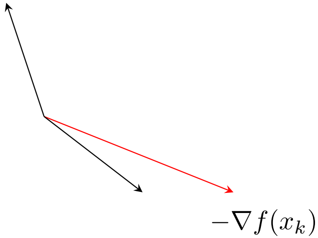
$$x_{k+1} = \begin{cases} x_k + \alpha_k d_k & \text{if successful} \\ x_k & \text{if unsuccessful,} \end{cases}$$

$$\alpha_{k+1} = \begin{cases} \gamma \alpha_k & \text{if successful} \\ \theta \alpha_k & \text{if unsuccessful.} \end{cases}$$

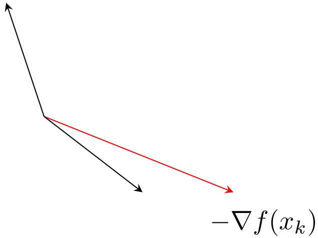
# Using random polling sets: What will happen?



# Using random polling sets: What will happen?



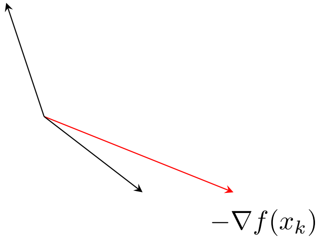
# Using random polling sets: What will happen?




$\leq n$ random polling directions

certainly not a PSS

# Using random polling sets: What will happen?




$\leq n$ random polling directions

certainly not a PSS

Yet $D_k$ is 'good' in some probabilistic sense ...

# What do we mean by 'good'?

If derivatives were available, it would have been sufficient to require

$$\max_{d \in D} \frac{-d^\top \nabla f(x_k)}{\|d\| \|\nabla f(x_k)\|} \geq \kappa.$$

# What do we mean by 'good'?

If derivatives were available, it would have been sufficient to require

$$\max_{d \in D} \frac{-d^\top \nabla f(x_k)}{\|d\| \|\nabla f(x_k)\|} \;\geq\; \kappa.$$

Define

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

Then $\mathrm{cm}(D, -\nabla f(x_k)) \geq \kappa$ would have been enough.

# What do we mean by 'good'?

If derivatives were available, it would have been sufficient to require

$$\max_{d \in D} \frac{-d^\top \nabla f(x_k)}{\|d\| \|\nabla f(x_k)\|} \;\geq\; \kappa.$$

Define

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

Then $\mathrm{cm}(D, -\nabla f(x_k)) \geq \kappa$ would have been enough.

But derivatives are not available!

# Random variables v.s. realizations

From now on, we suppose that the polling directions are not defined deterministically but taken at random from the unit sphere $\mathbb{S}^{n-1}$.

## Random variables v.s. realizations

From now on, we suppose that the polling directions are not defined deterministically but taken at random from the unit sphere $\mathbb{S}^{n-1}$.

Distinguish random variables from realizations

| | Iterate | Polling set |
|---|---|---|
| Random variables | $X_k$ | $\mathfrak{D}_k$ |
| Realizations | $x_k$ | $D_k$ |

# What to do?

- Global convergence:

$$\left\{\liminf_{k\to\infty} \|\nabla f(X_k)\| > 0\right\} \subset E$$

with $\mathbb{P}(E) = 0$.

# What to do?

- Global convergence:

$$\left\{ \liminf_{k\to\infty} \|\nabla f(X_k)\| > 0 \right\} \subset E$$

with $\mathbb{P}(E) = 0$.

- Worst case complexity:

$$\left\{ \min_{0\leq \ell \leq k} \|\nabla f(X_k)\| > \epsilon \right\} \subset E_{k,\epsilon},$$

with $\mathbb{P}(E_{k,\epsilon})$ being 'low' when $k$ is 'large'.

# What to do?

- Global convergence:

$$\left\{\liminf_{k\to\infty} \|\nabla f(X_k)\| > 0\right\} \subset E$$

with $\mathbb{P}(E) = 0$.

- Worst case complexity:

$$\left\{\min_{0\le\ell\le k} \|\nabla f(X_k)\| > \epsilon\right\} \subset E_{k,\epsilon},$$

with $\mathbb{P}(E_{k,\epsilon})$ being 'low' when $k$ is 'large'.

It remains to find $E$ and $E_{k,\epsilon}$ ...

# Global convergence: An intuitive lemma

Let $Z_k$ be the indicator function of $\left\{\mathrm{cm}\big(\mathfrak{D}_k, -\nabla f(X_k)\big) \geq \kappa\right\}$, and

$$p_0 \;=\; \frac{\ln\theta}{\ln(\gamma^{-1}\theta)} \;=\; \frac{1}{2}.$$

# Global convergence: An intuitive lemma

Let $Z_k$ be the indicator function of $\left\{\mathrm{cm}\big(\mathfrak{D}_k, -\nabla f(X_k)\big) \ge \kappa\right\}$, and

$$p_0 \;=\; \frac{\ln\theta}{\ln(\gamma^{-1}\theta)} \;=\; \frac{1}{2}.$$

Without imposing any assumption on the probabilistic behavior of $\{\mathfrak{D}_k\}$:

**Lemma**

$$\left\{\liminf_{k\to\infty} \|\nabla f(X_k)\| > 0\right\} \;\subset\; \left\{\sum_{k=0}^{\infty} (Z_k - p_0) = -\infty\right\} \;(\equiv E).$$

## Global convergence: An intuitive lemma

Let $Z_k$ be the indicator function of $\{\mathrm{cm}(\mathfrak{D}_k, -\nabla f(X_k)) \geq \kappa\}$, and

$$p_0 = \frac{\ln\theta}{\ln(\gamma^{-1}\theta)} = \frac{1}{2}.$$

Without imposing any assumption on the probabilistic behavior of $\{\mathfrak{D}_k\}$:

**Lemma**

$$\left\{\liminf_{k\to\infty} \|\nabla f(X_k)\| > 0\right\} \subset \left\{\sum_{k=0}^{\infty}(Z_k - p_0) = -\infty\right\} \ (\equiv E).$$

Meaning:

If convergence does not hold, the ‘frequency’ of $\{Z_k\}_{k\geq 0}$ is ‘less than $p_0$’.

# Worst case complexity: Another intuitive lemma

Without imposing any assumption on the probabilistic behavior of $\{\mathfrak{D}_k\}$:

**Lemma**

$$\left\{\max_{0\leq\ell\leq k}\|\nabla f(X_k)\| > \epsilon\right\} \subset \left\{\sum_{\ell=0}^{k-1} Z_\ell \leq \left[\frac{(\nu+1)^2\beta}{2\kappa^2\epsilon^2 k} + p_0\right]k\right\} \ (\equiv E_{k,\epsilon}).$$

$\beta < \infty$ is an upper bound for $\sum_{k=0}^{\infty}\rho(\alpha_k)$ (existence guaranteed).

$\nu < \infty$ is a Lipshitz constant of $\nabla f$ in $\mathbb{R}^n$.

# Worst case complexity: Another intuitive lemma

Without imposing any assumption on the probabilistic behavior of $\{\mathfrak{D}_k\}$:

**Lemma**

$$\left\{\max_{0\le\ell\le k}\|\nabla f(X_k)\| > \epsilon\right\} \subset \left\{\sum_{\ell=0}^{k-1} Z_\ell \le \left[\frac{(\nu+1)^2\beta}{2\kappa^2\epsilon^2 k} + p_0\right] k\right\} \; (\equiv E_{k,\epsilon}).$$

$\beta < \infty$ is an upper bound for $\sum_{k=0}^{\infty} \rho(\alpha_k)$ (existence guaranteed).

$\nu < \infty$ is a Lipshitz constant of $\nabla f$ in $\mathbb{R}^n$.

Meaning:

If $\{\|\nabla f(X_0)\|\}_{0\le\ell\le k}$ are all above $\epsilon$, the 'frequency' of $\{Z_\ell\}_{0\le\ell\le k-1}$ is 'not more than $p_0 + \mathcal{O}(\epsilon^{-2}k^{-1})$'.

# What assumptions to impose?

Until now, no assumption is imposed on the probabilistic behavior of $\{\mathfrak{D}_k\}$.

## What assumptions to impose?

Until now, no assumption is imposed on the probabilistic behavior of $\{\mathfrak{D}_k\}$.

**Definition**

*The sequence* $\{\mathfrak{D}_k\}$ *is* $p$*-probabilistically* $\kappa$*-descent if, for each* $k \geq 0$,

$$\mathbb{P}\Big(\mathrm{cm}(\mathfrak{D}_k, -\nabla f(X_k)) \ \geq \ \kappa \,|\, \mathfrak{D}_0, \ldots, \mathfrak{D}_{k-1}\Big) \geq \ p.$$

# Global convergence

## Lemma

*If $\{\mathfrak{D}_k\}$ is $p_0$-probabilistically $\kappa$-descent, then $\left\{ \sum_{\ell=0}^{k-1} (Z_\ell - p_0) \right\}$ is a submartingale, and*

$$\mathbb{P}\left(\sum_{k=0}^{\infty} (Z_k - p_0) = -\infty\right) = 0.$$

# Global convergence

## Lemma

*If $\{\mathfrak{D}_k\}$ is $p_0$-probabilistically $\kappa$-descent, then $\left\{ \sum_{\ell=0}^{k-1} (Z_\ell - p_0) \right\}$ is a submartingale, and*

$$\mathbb{P}\left( \sum_{k=0}^{\infty} (Z_k - p_0) = -\infty \right) = 0.$$

## Theorem

*If $\{\mathfrak{D}_k\}$ is $p_0$-probabilistically $\kappa$-descent, then*

$$\mathbb{P}\Big( \liminf_{k\to\infty} \|\nabla f(X_k)\| = 0 \Big) \; = \; 1.$$

## Global rate

### Lemma (Chernoff bound)

*Suppose that* $\{\mathfrak{D}_k\}$ *is $p$-probabilistically $\kappa$-descent and* $\lambda \in (0, p)$. *Then*

$$\mathbb{P}\left(\sum_{\ell=0}^{k-1} Z_\ell \le \lambda k\right) \;\le\; \exp\left[-\frac{(p-\lambda)^2}{2p}k\right].$$

# Global rate

## Lemma (Chernoff bound)

*Suppose that* $\{\mathfrak{D}_k\}$ *is* *$p$-probabilistically $\kappa$-descent* *and* $\lambda \in (0, p)$. *Then*

$$\mathbb{P}\left(\sum_{\ell=0}^{k-1} Z_\ell \leq \lambda k\right) \;\leq\; \exp\left[-\frac{(p-\lambda)^2}{2p}k\right].$$

## Theorem

*Suppose that* $\{\mathfrak{D}_k\}$ *is* *$p$-probabilistically $\kappa$-descent* *with* $p > p_0$. *Then*

$$\mathbb{P}\left(\min_{0\leq\ell\leq k} \|\nabla f(X_\ell)\| \leq \left[\frac{(\nu+1)\beta^{\frac{1}{2}}}{(p-p_0)^{\frac{1}{2}}\kappa}\right]\frac{1}{\sqrt{k}}\right) \;\geq\; 1 - \exp\left[-\frac{(p-p_0)^2}{8p}k\right].$$

# Global rate

## Lemma (Chernoff bound)

*Suppose that* $\{\mathfrak{D}_k\}$ *is p-probabilistically κ-descent and* $\lambda \in (0,p)$. *Then*

$$\mathbb{P}\left(\sum_{\ell=0}^{k-1} Z_\ell \leq \lambda k\right) \leq \exp\left[-\frac{(p-\lambda)^2}{2p}k\right].$$

## Theorem

*Suppose that* $\{\mathfrak{D}_k\}$ *is p-probabilistically κ-descent with* $p > p_0$. *Then*

$$\mathbb{P}\left(\min_{0\leq\ell\leq k}\|\nabla f(X_\ell)\| \leq \left[\frac{(\nu+1)\beta^{\frac{1}{2}}}{(p-p_0)^{\frac{1}{2}}\kappa}\right]\frac{1}{\sqrt{k}}\right) \geq 1-\exp\left[-\frac{(p-p_0)^2}{8p}k\right].$$

$\implies$ $\mathcal{O}(\kappa^{-1}k^{-1/2})$ decaying rate for gradient holds with overwhelmingly high probability, matching the deterministic case (Vicente 2013).

# Practical probabilistic descent sets

For each $k \geq 0$,

- $\mathfrak{D}_k$ is independent of the previous iterations,

# Practical probabilistic descent sets

For each $k \geq 0$,

- $\mathfrak{D}_k$ is independent of the previous iterations,
- $\mathfrak{D}_k$ is a set $\{\mathfrak{d}_1, \ldots, \mathfrak{d}_m\}$ of independent random vectors uniformly distributed on the unit sphere.

# Practical probabilistic descent sets

$\{\mathfrak{D}_k\}$ generated in this way is probabilistically descent.

## Proposition

*Given* $\tau \in [0, \sqrt{n}]$, $\{\mathfrak{D}_k\}$ *is* $p$*-probabilistically* $(\tau/\sqrt{n})$*-descent with*

$$p = 1 - \left(\frac{1}{2} + \frac{\tau}{\sqrt{2\pi}}\right)^m.$$

# Practical probabilistic descent sets

$\{\mathfrak{D}_k\}$ generated in this way is probabilistically descent.

## Proposition

*Given* $\tau \in [0, \sqrt{n}]$, $\{\mathfrak{D}_k\}$ *is* $p$*-probabilistically* $(\tau/\sqrt{n})$*-descent with*

$$p = 1 - \left(\frac{1}{2} + \frac{\tau}{\sqrt{2\pi}}\right)^m.$$

For instance,

$$\left.\begin{array}{rcl} m & = & 2 \\ \tau & = & \frac{1}{2} \end{array}\right\} \implies p > \frac{1}{2} = p_0 \quad (\kappa = 1/(2\sqrt{n})).$$

# Practical probabilistic descent sets: WCC bounds

Plugging $m = 2$ and $\kappa = 1/(2\sqrt{n})$ into the global rate, one obtains

## WCC for function evaluations

Let $K_\epsilon^f$ be the least number of function evaluations that is sufficient to drive $\|\nabla f(X_k)\|$ under $\epsilon$. Then

$$\mathbb{P}\left(K_\epsilon^f \le 2\left\lceil \frac{4(\nu+1)^2\beta}{p-p_0}(n\epsilon^{-2})\right\rceil\right) \ge 1 - \exp\left[-\frac{\beta(p-p_0)(\nu+1)^2}{2p}(n\epsilon^{-2})\right].$$

# Practical probabilistic descent sets: WCC bounds

Plugging $m = 2$ and $\kappa = 1/(2\sqrt{n})$ into the global rate, one obtains

**WCC for function evaluations**

Let $K_\epsilon^f$ be the least number of function evaluations that is sufficient to drive $\|\nabla f(X_k)\|$ under $\epsilon$. Then

$$\mathbb{P}\left(K_\epsilon^f \leq 2\left\lceil \frac{4(\nu+1)^2\beta}{p-p_0}(n\epsilon^{-2})\right\rceil\right) \geq 1 - \exp\left[-\frac{\beta(p-p_0)(\nu+1)^2}{2p}(n\epsilon^{-2})\right].$$

$\implies$ $\mathcal{O}(n\epsilon^{-2})$ with overwhelmingly high probability.

# Practical probabilistic descent sets: WCC bounds

Plugging $m = 2$ and $\kappa = 1/(2\sqrt{n})$ into the global rate, one obtains

**WCC for function evaluations**

Let $K_\epsilon^f$ be the least number of function evaluations that is sufficient to drive $\|\nabla f(X_k)\|$ under $\epsilon$. Then

$$\mathbb{P}\left(K_\epsilon^f \leq 2\left\lceil \frac{4(\nu+1)^2\beta}{p-p_0}(n\epsilon^{-2})\right\rceil\right) \geq 1 - \exp\left[-\frac{\beta(p-p_0)(\nu+1)^2}{2p}(n\epsilon^{-2})\right].$$

$\Longrightarrow$ $\mathcal{O}(n\epsilon^{-2})$ with overwhelmingly high probability.

- It is better than the optimal order of the deterministic case $\mathcal{O}(n^2\epsilon^{-2})$ (Dodangeh, Vicente, Zhang 2014).
- No matter how big $n$ is, using 2 random directions is sufficient to guarantee the convergence of direct search.

## Concluding remarks

- Deterministic direct search
  - The optimal order of the worst case complexity for function evaluations is $\mathcal{O}(n^2\epsilon^{-2})$.
  - The optimal order is achived when using $D_\oplus$ as the polling set.
  - M. Dodangeh, L. N. Vicente, Z. Zhang, On the optimal order of worst case complexity of direct search, to appear in *Optim. Lett.*

# Concluding remarks

- Deterministic direct search
  - The optimal order of the worst case complexity for function evaluations is $\mathcal{O}(n^2\epsilon^{-2})$.
  - The optimal order is achived when using $D_\oplus$ as the polling set.
  - M. Dodangeh, L. N. Vicente, Z. Zhang, On the optimal order of worst case complexity of direct search, to appear in *Optim. Lett.*
- Randomized direct search
  - No matter how big $n$ is, using 2 random directions is sufficient to guaranttee the convergence of direct search.
  - The worst case complexity for function evaluations is $\mathcal{O}(n\epsilon^{-2})$, better than the optimal order in the deterministic case.
  - S. Gratton, C. W. Royer, L. N. Vicente, Z. Zhang, "Direct search based on probabilistic descent", *SIAM J. Optim.*, 25(3): 1515-1541, 2015

# Concluding remarks

- Deterministic direct search
  - The optimal order of the worst case complexity for function evaluations is $\mathcal{O}(n^2\epsilon^{-2})$.
  - The optimal order is achived when using $D_\oplus$ as the polling set.
  - M. Dodangeh, L. N. Vicente, Z. Zhang, On the optimal order of worst case complexity of direct search, to appear in *Optim. Lett.*
- Randomized direct search
  - No matter how big $n$ is, using 2 random directions is sufficient to guaranttee the convergence of direct search.
  - The worst case complexity for function evaluations is $\mathcal{O}(n\epsilon^{-2})$, better than the optimal order in the deterministic case.
  - S. Gratton, C. W. Royer, L. N. Vicente, Z. Zhang, "Direct search based on probabilistic descent", *SIAM J. Optim.*, 25(3): 1515-1541, 2015
- What was 'unexpected'?

# For taking back home: A seemingly easy open problem

## Open problem

*Prove that $D_\oplus$ gives the most 'economical' covering of the unit sphere, or equivalently, for any $2n$ unit vectors $\{d_1, d_2, \ldots, d_{2n}\} \subset \mathbb{R}^n$, there exists a unit vector $v \in \mathbb{R}^n$ such that*

$$\max_{1 \le i \le 2n} d^\top v \le \frac{1}{\sqrt{n}}.$$

References:


- K. Böröczky, Jr. *Finite Packing and Covering*, Cambridge University Press, New York, 2004
- K. Böröczky, Jr. and G. Wintsche, "Covering the sphere by equal spherical balls", In *Discrete and Computational Geometry*, volume 25 of Algorithms and Combinatorics, pages 235–251. Springer Berlin, 2003