My Work and Life in Coimbra

— *An Annual Report to My Professor and Friends*

Zaikun Zhang
(from here)

September 3, At home

`http//www.zhangzk.net`

Outline

1. My work in Coimbra — "Direct Search Based on Probabilistic Descent" (30 min + 5 min)
2. My life in Coimbra (10 min)
3. Something else (12 min)

Direct Search Based on Probabilistic Descent

Zaikun Zhang
University of Coimbra
(Joint work with S. Gratton, C. W. Royer, and L. N. Vicente)

September 3, ICMSEC

http//www.zhangzk.net

Direct search

Algorithm

Step 1. Initiate.Set $k = 0$

Direct search

Algorithm

Step 1. Initiate.Set $k = 0$

Step 2. Determine a set of nonzero vectors $\mathcal{D}_k$.

Direct search

Algorithm

Step 1. Initiate.Set $k = 0$

Step 2. Determine a set of nonzero vectors $\mathcal{D}_k$.

Step 3. If there exists $d_k \in \mathcal{D}_k$ such that

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \rho(\alpha_k),$$

then set

$$x_{k+1} = x_k + \alpha_k d_k, \quad \alpha_{k+1} = \min\{\mu\alpha_k, \alpha_{\max}\};$$

else set

$$x_{k+1} = x_k, \quad \alpha_{k+1} = \theta\alpha_k.$$

Direct search

Algorithm

Step 1. Initiate.Set $k = 0$

Step 2. Determine a set of nonzero vectors $\mathcal{D}_k$.

Step 3. If there exists $d_k \in \mathcal{D}_k$ such that

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \rho(\alpha_k),$$

then set

$$x_{k+1} = x_k + \alpha_k d_k, \quad \alpha_{k+1} = \min\{\mu\alpha_k, \alpha_{\max}\};$$

else set

$$x_{k+1} = x_k, \quad \alpha_{k+1} = \theta\alpha_k.$$

Step 4. Increment k by one, and go to **Step 2**.

Direct search

Algorithm

Step 1. Determine constants $\theta \in (0,1)$, $\mu \in [1,\infty)$, $\alpha_{\max} > 0$, and a function $\rho : (0,\infty) \to (0,\infty)$. Set $k = 0$.

Step 2. Determine a set of nonzero vectors $\mathcal{D}_k$.

Step 3. If there exists $d_k \in \mathcal{D}_k$ such that

$$f(x_k) - f(x_k + \alpha_k d_k) \geq \rho(\alpha_k),$$

then set

$$x_{k+1} = x_k + \alpha_k d_k, \quad \alpha_{k+1} = \min\{\mu\alpha_k, \alpha_{\max}\};$$

else set

$$x_{k+1} = x_k, \quad \alpha_{k+1} = \theta\alpha_k.$$

Step 4. Increment k by one, and go to **Step 2**.

Classical theory

Assumption

$\rho(\alpha) = c\alpha^q$, $c > 0$, $q > 1$ *(an illustration, more general* ρ *is acceptable).*

Classical theory

Assumption

$\rho(\alpha) = c\alpha^q$, $c > 0$, $q > 1$ *(an illustration, more general ρ is acceptable).*

Assumption

f is twice continuously differentiable, $\nabla^2 f$ is bounded, and f is bounded from below.

Classical theory

Assumption

$\rho(\alpha) = c\alpha^q$, $c > 0$, $q > 1$ (an illustration, more general ρ is acceptable).

Assumption

f is twice continuously differentiable, $\nabla^2 f$ is bounded, and f is bounded from below.

Theorem

Suppose that there exists a positive constant κ such that

$$\mathrm{cm}(\mathcal{D}_k) \geq \kappa,$$

then

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

Classical theory

Assumption

$\rho(\alpha) = c\alpha^q$, $c > 0$, $q > 1$ *(an illustration, more general ρ is acceptable).*

Assumption

f is twice continuously differentiable, $\nabla^2 f$ is bounded, and f is bounded from below.

Theorem

Suppose that there exists a positive constant κ such that

$$\mathrm{cm}(\mathcal{D}_k) \geq \kappa,$$

then

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

Positive spanning set

Let $\mathcal{D} = \{d_1, \ldots, d_m\}$ be a set of nonzero vectors in $\mathbb{R}^n$. Define $0/0 = 1$.

Definition (Cosine measure)

$$\operatorname{cm}(\mathcal{D}) = \min_{v \in \mathbb{R}^n} \max_{d \in \mathcal{D}} \frac{v^\top d}{\|v\| \|d\|}.$$

Positive spanning set

Let $\mathcal{D} = \{d_1, \ldots, d_m\}$ be a set of nonzero vectors in $\mathbb{R}^n$. Define $0/0 = 1$.

Definition (Cosine measure)

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

Definition

We say $\mathcal{D}$ is a positive spanning set, if for each $x \in \mathbb{R}^n$, there exist nonnegative scalars a_1, a_2, ..., a_m such that

$$x = \sum_{i=1}^{m} a_i d_i.$$

Positive spanning set

Let $\mathcal{D} = \{d_1, \ldots, d_m\}$ be a set of nonzero vectors in $\mathbb{R}^n$. Define $0/0 = 1$.

Definition (Cosine measure)

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

Definition

We say $\mathcal{D}$ is a positive spanning set, if for each $x \in \mathbb{R}^n$, there exist nonnegative scalars a_1, a_2, ..., a_m such that

$$x = \sum_{i=1}^{m} a_i d_i.$$

Proposition

$\mathcal{D}$ is a positive spanning set iff $\mathrm{cm}(\mathcal{D}) > 0$.

Example: $\mathcal{D} = \{e_1, -e_1, \ldots, e_n, -e_n\}$, $\mathrm{cm}(D) = n^{-\frac{1}{2}}$.

Numerical results

Table: Direct search: PSS v.s. random directions

	$[Q, -Q]$	$n/2$	n	$2n$
arwhead	112	306	445	897
bdqrtic	690	197	338	599
biggs6	1847	5530	NaN	NaN
broydn3d	1094	387	562	1060
integreq	1075	310	514	926
penalty1	521	144	219	376
penalty2	309	105	161	317
powellsg	1548	1094	1475	1611
rosenbrock	5106	2702	4702	6618
srosenbr	NaN	15802	NaN	NaN
vardim	46	11	11	15
woods	3067	583	852	1866

Positive spanning set revisited

- The classical convergence theory requires $\mathrm{cm}(\mathcal{D}_k) > \kappa$, which necessitate $\#\mathcal{D}_k \geq n+1$.

Positive spanning set revisited

- The classical convergence theory requires $\mathrm{cm}(\mathcal{D}_k) > \kappa$, which necessitate $\#\mathcal{D}_k \geq n+1$.
- Why require $\mathrm{cm}(\mathcal{D}_k) > \kappa$?

Positive spanning set revisited

- The classical convergence theory requires $\mathrm{cm}(\mathcal{D}_k) > \kappa$, which necessitate $\#\mathcal{D}_k \geq n+1$.

- Why require $\mathrm{cm}(\mathcal{D}_k) > \kappa$?

  Recall that
  $$\mathrm{cm}(\mathcal{D}_k) = \min_{v\in\mathbb{R}^n} \max_{d\in\mathcal{D}_k} \frac{v^\top d}{\|v\|\|d\|}.$$

Positive spanning set revisited

- The classical convergence theory requires $\mathrm{cm}(\mathcal{D}_k) > \kappa$, which necessitate $\#\mathcal{D}_k \ge n+1$.
- Why require $\mathrm{cm}(\mathcal{D}_k) > \kappa$?

  Recall that

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

Positive spanning set revisited

- The classical convergence theory requires $\mathrm{cm}(\mathcal{D}_k) > \kappa$, which necessitate $\#\mathcal{D}_k \geq n+1$.
- Why require $\mathrm{cm}(\mathcal{D}_k) > \kappa$?

  Recall that

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

- What about

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

Positive spanning set revisited

- The classical convergence theory requires $\mathrm{cm}(\mathcal{D}_k) > \kappa$, which necessitate $\#\mathcal{D}_k \geq n+1$.
- Why require $\mathrm{cm}(\mathcal{D}_k) > \kappa$?

  Recall that

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

- What about

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

  It is nice, but have we forgotten about "derivative-free"?

Local cosine measure

Definition

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

Local cosine measure

Definition

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

What about assuming that $\mathcal{D}_k$ satisfies $\mathrm{cm}(\mathcal{D}_k, -\nabla f(x_k)) \geq \kappa$ in some probabilistic sense?

Change notations due to randomness

If the direction sets are random, then so are the iterates, the stepsizes ...

Change notations due to randomness

If the direction sets are random, then so are the iterates, the stepsizes ...

To make the discussions mathematically sound, we have to clarify the notations.

Table: Random variables v.s. realizations

Random variables	X_k	$\mathscr{D}_k$
Realizations	x_k	$\mathcal{D}_k$

Global convergence

Assumption

There exists a positive constant κ *such that*

$$\Pr\left(\mathrm{cm}(\mathscr{D}_k, -\nabla f(X_k)) \geq \kappa \mid \mathscr{D}_0, \dots \mathscr{D}_{k-1}\right) \geq \frac{\log\theta}{\log(\mu^{-1}\theta)}.$$

Global convergence

Assumption

There exists a positive constant κ such that

$$\Pr\left(\mathrm{cm}(\mathscr{D}_k, -\nabla f(X_k)) \geq \kappa \mid \mathscr{D}_0, \ldots \mathscr{D}_{k-1}\right) \geq \frac{\log \theta}{\log(\mu^{-1}\theta)}.$$

Global convergence

Assumption

There exists a positive constant κ such that

$$\Pr\left(\mathrm{cm}(\mathscr{D}_k, -\nabla f(X_k)) \geq \kappa \mid \mathscr{D}_0, \ldots \mathscr{D}_{k-1}\right) \geq \frac{\log\theta}{\log(\mu^{-1}\theta)}.$$

Theorem

$$\Pr\left(\liminf_{k\to\infty} \|\nabla f(X_k)\| = 0\right) = 1.$$

Global convergence (cont.)

Sketch of proof.

Lemma

$$\left\{\liminf_{k\to\infty} \|\nabla f(X_k)\| \neq 0\right\} \subset \left\{\sum_{k=0}^{\infty} [Z_k \log \mu + (1 - Z_k) \log \theta] = -\infty\right\},$$

where Z_k is the indicator function of the event

$$\{\mathrm{cm}(\mathscr{D}_k, -\nabla f(X_k)) \geq \kappa\}.$$

Lemma

The random process $\left\{\sum_{\ell=0}^{k} [Z_\ell \log \mu + (1 - Z_\ell) \log \theta]\right\}$ is a submartigale.

Worst case complexity (global rate)

Assumption

There exist constants κ and p such that

$$\Pr\left(\mathrm{cm}(\mathscr{D}_k, -\nabla f(X_k)) \geq \kappa \mid \mathscr{D}_0, \ldots \mathscr{D}_{k-1}\right) \geq p > \frac{\log\theta}{\log(\mu^{-1}\theta)}.$$

Worst case complexity (global rate)

Assumption

There exist constants κ and p such that

$$\Pr\left(\mathrm{cm}(\mathscr{D}_k, -\nabla f(X_k)) \geq \kappa \mid \mathscr{D}_0, \ldots \mathscr{D}_{k-1}\right) \geq p > \frac{\log\theta}{\log(\mu^{-1}\theta)}.$$

Theorem

If ε is small enough, and $k \geq (1+\delta)\eta/\{(p-p_0)\rho[\varphi(\kappa\varepsilon)]\}$ $(\delta > 0)$, then

$$\Pr\left(\min_{0\leq\ell\leq k} \|\nabla f(X_\ell)\| \geq \varepsilon\right) \leq \exp\left[-\frac{(p-p_0)^2\delta^2}{2p(1+\delta)^2}k\right].$$

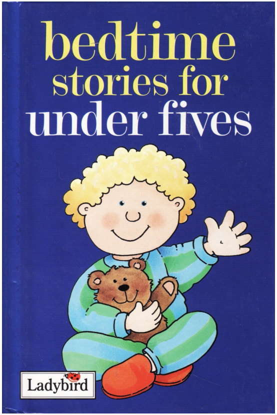
$\varphi(t)$ is the unique positive number satisfies $\frac{\rho(\alpha)}{\alpha} + \frac{1}{2}M\alpha = t$, and $p_0 = \frac{\log\theta}{\log(\mu^{-1}\theta)}$.

Worst case complexity (global rate, cont.)

Proof: See

S. Gratton, C. W. Royer, L. N. Vicent, and Z. Direct Search Based on Probabilistic Descent (in preparation), 2013

Story time



Looooong time ago, there was a princess who was kept in captivity by an evil witch in a cave. One day, a prince came to save her . . .

Story time (cont.)

WITCH:

There is a 100-dimensional nonzero vector written on this paper.

Story time (cont.)

WITCH:

There is a 100-dimensional nonzero vector written on this paper.

You give me 20 vectors of 100 dimension.

Story time (cont.)

WITCH:

There is a 100-dimensional nonzero vector written on this paper.

You give me 20 vectors of 100 dimension.

If the one of the angles between your vectors and mine is smaller than $\arccos(1/10)$, then I will set the princess free. Otherwise, the princess will stay here and you will go to the hell.

Story time (cont.)

WITCH:

There is a 100-dimensional nonzero vector written on this paper.

You give me 20 vectors of 100 dimension.

If the one of the angles between your vectors and mine is smaller than $\arccos(1/10)$, then I will set the princess free. Otherwise, the princess will stay here and you will go to the hell.

PRINCE: ...

Mission: Impossible?

- Your mission, should you choose to accept it.
- No, I do NOT accept. I SAID it is impossible! (if the prince was a pure mathematician)

Impossible but doable

- Your mission, should you choose to accept it.
- It is impossible. But doable. (if the prince was an applied mathematician)

A practical definition of $\mathscr{D}_k$

Theorem

Suppose that V is a nonzero random vector in $\mathbb{R}^n$, and $\mathscr{D}$ is a set of m random vectors independent of V and i.i.d. uniformly distributed on the unit sphere in $\mathbb{R}^n$, then

$$\Pr(\mathrm{cm}(\mathscr{D}, V) \le \kappa) = \left\{ \frac{1}{2} + \frac{\operatorname{sign}\kappa}{2} \left[1 - \mathrm{I}\left(1 - \kappa^2, \frac{n-1}{2}, \frac{1}{2}\right) \right] \right\}^m,$$

where $\mathrm{I}(\cdot,\cdot,\cdot)$ is the regularized incomplete beta function.

A practical definition of $\mathscr{D}_k$

Theorem

Suppose that V is a nonzero random vector in $\mathbb{R}^n$, and $\mathscr{D}$ is a set of m random vectors independent of V and i.i.d. uniformly distributed on the unit sphere in $\mathbb{R}^n$, then

$$\Pr(\mathrm{cm}(\mathscr{D}, V) \le \kappa) = \left\{ \frac{1}{2} + \frac{\operatorname{sign}\kappa}{2} \left[1 - \mathrm{I}\left(1 - \kappa^2, \frac{n-1}{2}, \frac{1}{2}\right)\right]\right\}^m,$$

where $\mathrm{I}(\cdot,\cdot,\cdot)$ is the regularized incomplete beta function.

If $n = 100$, $m = 20$, and $\kappa = 1/10$, then

$$\Pr(\mathrm{cm}(\mathscr{D}, V) > \kappa) \approx 0.97 \ !$$

A practical definition of $\mathscr{D}_k$

Theorem

Suppose that V *is a nonzero random vector in* $\mathbb{R}^n$*, and* $\mathscr{D}$ *is a set of* m *random vectors independent of* V *and i.i.d. uniformly distributed on the unit sphere in* $\mathbb{R}^n$*, then*

$$\Pr(\mathrm{cm}(\mathscr{D}, V) \le \kappa) = \left\{\frac{1}{2} + \frac{\operatorname{sign}\kappa}{2}\left[1 - \mathrm{I}\left(1-\kappa^2, \frac{n-1}{2}, \frac{1}{2}\right)\right]\right\}^m,$$

where $\mathrm{I}(\cdot,\cdot,\cdot)$ *is the regularized incomplete beta function.*

If $n = 100$, $m = 20$, and $\kappa = 1/10$, then

$$\Pr(\mathrm{cm}(\mathscr{D}, V) > \kappa) \approx 0.97 \ !$$

Should the prince take the risk?

A practical definition of $\mathscr{D}_k$

Theorem

Suppose that V is a nonzero random vector in $\mathbb{R}^n$, and $\mathscr{D}$ is a set of m random vectors independent of V and i.i.d. uniformly distributed on the unit sphere in $\mathbb{R}^n$, then

$$\Pr(\mathrm{cm}(\mathscr{D}, V) \le \kappa) = \left\{ \frac{1}{2} + \frac{\operatorname{sign}\kappa}{2} \left[1 - \mathrm{I}\left(1 - \kappa^2, \frac{n-1}{2}, \frac{1}{2} \right) \right] \right\}^m,$$

where $\mathrm{I}(\cdot,\cdot,\cdot)$ is the regularized incomplete beta function.

If $n = 100$, $m = 20$, and $\kappa = 1/10$, then

$$\Pr(\mathrm{cm}(\mathscr{D}, V) > \kappa) \approx 0.97\ !$$

Should the prince take the risk?

(Should we call that a risk?)

Finally, the prince and the princess . . .

My life in Coimbra

- Learn Portuguese

My life in Coimbra

- Learn Portuguese
- Coimbra — city of student parties

My life in Coimbra

- Learn Portuguese
- Coimbra — city of student parties
- My trip to Poland

Quiz time :(

- ICCOPT Student Social, the quiz

Thanks!

zhang@mat.uc.pt
www.zhangzk.net