Direct Search based on Probabilistic Feasible Descent for Bound and Linearly Constrained Problems

Clément Royer - University of Wisconsin-Madison

Joint work with S. Gratton, L. N. Vicente and Z. Zhang

SIAM Conference on Optimization, Vancouver - May 24, 2017

Outline

1. Direct search for unconstrained optimization
2. Deterministic direct search on linearly constrained problems
3. Probabilistic feasible descent
4. Numerical results

Introductory assumptions and definitions

We consider an unconstrained smooth problem:

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

Assumptions on f

- f bounded from below.
- f continuously differentiable, ∇f Lipschitz continuous.

Introductory assumptions and definitions

We consider an unconstrained smooth problem:

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

Assumptions on f

- f bounded from below.
- f continuously differentiable, ∇f Lipschitz continuous.

Derivative-free optimization (DFO) context

- *Simulation code:* gradient too expensive to be computed.
- *Black-box objective function:* no derivative code available.
- *Automatic differentiation:* inapplicable.

The gradient exists but cannot be used for algorithmic purposes.

Our algorithmic framework: Direct Search

1. **Initialization:** Set $x_0 \in \mathbb{R}^n, \alpha_0 > 0, 0 < \theta < 1 \leq \gamma$.
2. **For** $k = 0, 1, 2, \ldots$
   - Choose a (polling) set D_k of r vectors.
   - If it exists $d_k \in D_k$ so that

     $$f(x_k + \alpha_k d_k) < f(x_k) - \alpha_k^2,$$

     then declare k *successful*, set $x_{k+1} := x_k + \alpha_k d_k$ and update $\alpha_{k+1} := \gamma \alpha_k$.
   - Otherwise declare k *unsuccessful*, set $x_{k+1} := x_k$ and update $\alpha_{k+1} := \theta \alpha_k$.

Our algorithmic framework: Direct Search

1. **Initialization:** Set $x_0 \in \mathbb{R}^n, \alpha_0 > 0, 0 < \theta < 1 \leq \gamma$.
2. **For** $k = 0, 1, 2, \ldots$
   - Choose a (polling) set D_k of r vectors.
   - If it exists $d_k \in D_k$ so that

     $$f(x_k + \alpha_k \, d_k) < f(x_k) - \alpha_k^2,$$

     then declare k *successful*, set $x_{k+1} := x_k + \alpha_k \, d_k$ and update $\alpha_{k+1} := \gamma \, \alpha_k$.
   - Otherwise declare k *unsuccessful*, set $x_{k+1} := x_k$ and update $\alpha_{k+1} := \theta \, \alpha_k$.

Performance indicator: number of evaluations of f.

Direction choice in deterministic direct search

A measure of set quality

For a set of vectors D and $v \in \mathbb{R}^n \setminus \{0\}$, the cosine measure of D at v is

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

Direction choice in deterministic direct search

A measure of set quality

For a set of vectors D and $v \in \mathbb{R}^n \setminus \{0\}$, the cosine measure of D at v is

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

- When $\mathrm{cm}(D, v) > 0$, v makes an acute angle with some $d \in D$.
- If $v = -\nabla f(x) \neq 0$, D contains a descent direction for f at x.

Direction choice in deterministic direct search

A measure of set quality

For a set of vectors D and $v \in \mathbb{R}^n \setminus \{0\}$, the cosine measure of D at v is

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

- When $\mathrm{cm}(D, v) > 0$, v makes an acute angle with some $d \in D$.
- If $v = -\nabla f(x) \neq 0$, D contains a descent direction for f at x.
- Ensuring $\mathrm{cm}(D, v) > 0 \;\; \forall v \neq 0$ requires at least $n + 1$ vectors.

Example

Coordinate set: $D_\oplus = \{e_1, \ldots, e_n, \text{-}e_1, \ldots, \text{-}e_n\}$.

- $|D_\oplus| = 2n$.
- $\forall v, \quad \mathrm{cm}\left(D_\oplus, v\right) \geq \frac{1}{\sqrt{n}}$.

Convergence for deterministic direct search

Lemma

Independently of $\{D_k\}$,

$$\lim_{k \to \infty} \alpha_k = 0.$$

Lemma

If the k-th iteration is unsuccessful and $\mathrm{cm}(D_k, -\nabla f(x_k)) \geq \kappa > 0$, *then*

$$\kappa \, \|\nabla f(x_k)\| \; \leq \; \mathcal{O}\left(\alpha_k\right).$$

Convergence for deterministic direct search

Lemma

Independently of $\{D_k\}$,

$$\lim_{k \to \infty} \alpha_k = 0.$$

Lemma

If the k-th iteration is unsuccessful and $\mathrm{cm}(D_k, -\nabla f(x_k)) \geq \kappa > 0$, *then*

$$\kappa \, \|\nabla f(x_k)\| \;\leq\; \mathcal{O}\left(\alpha_k\right).$$

Convergence Theorem

If $\forall k,\ \mathrm{cm}(D_k, -\nabla f(x_k)) \geq \kappa$, we have

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

Worst-case complexity in deterministic direct search

Evaluation complexity

Let $\epsilon \in (0,1)$ and N_ϵ be the number of function evaluations needed to reach an point such that $\inf_{0 \le l \le k} \|\nabla f(x_l)\| < \epsilon$. Then,

$$N_\epsilon \;\le\; \mathcal{O}\left(r\,(\kappa\,\epsilon)^{-2}\right).$$

Choosing $D_k = D_\oplus$, one has $\kappa = 1/\sqrt{n}$, $r = 2n$, and the bound becomes

$$N_\epsilon \;\le\; \mathcal{O}\left(n^2\,\epsilon^{-2}\right).$$

Direct search based on probabilistic properties

Probabilistic properties

- Main idea: relax essential properties for convergence...
- ...and satisfy them only with a certain probability.

Direct search based on probabilistic properties

Probabilistic properties

- Main idea: relax essential properties for convergence...
- ...and satisfy them only with a certain probability.

In direct search

- Generate directions sets **independently at random**;
- Use less vectors than deterministic case;
- Those sets may be “good" with some probability.

(p, κ)-descent sets

From deterministic to probabilistic notations

- Polling sets/directions: $D_k = \mathfrak{D}_k(\omega)$, $d_k = \mathfrak{d}_k(\omega)$;
- Iterates: $x_k = X_k(\omega)$;
- Step sizes: $\alpha_k = A_k(\omega)$.

- We want to look at

$$\mathbb{P}\left(\mathrm{cm}\left(\mathfrak{D}_k, -\nabla f(X_k)\right) \geq \kappa\right).$$

  where X_k depends on $\mathfrak{D}_0, \ldots, \mathfrak{D}_{k-1}$ but not on $\mathfrak{D}_k$.
- We exploit conditional probabilities/conditioning to the past.

(p, κ)-descent sets

From deterministic to probabilistic notations

- Polling sets/directions: $D_k = \mathfrak{D}_k(\omega)$, $d_k = \mathfrak{d}_k(\omega)$;
- Iterates: $x_k = X_k(\omega)$;
- Step sizes: $\alpha_k = A_k(\omega)$.

- We want to look at

$$\mathbb{P}\left(\mathrm{cm}\left(\mathfrak{D}_k, -\nabla f(X_k)\right) \geq \kappa\right).$$

where X_k depends on $\mathfrak{D}_0, \ldots, \mathfrak{D}_{k-1}$ but not on $\mathfrak{D}_k$.
- We exploit conditional probabilities/conditioning to the past.

Probabilistic descent property

A random set sequence $\{\mathfrak{D}_k\}$ is said to be (p, κ)-descent if:

$$\begin{aligned} \mathbb{P}\left(\mathrm{cm}\left(\mathfrak{D}_0, -\nabla f(x_0)\right) \geq \kappa\right) &\geq p \\ \forall k \geq 1, \quad \mathbb{P}\left(\mathrm{cm}\left(\mathfrak{D}_k, -\nabla f(X_k)\right) \geq \kappa \mid \mathfrak{D}_0, \ldots, \mathfrak{D}_{k-1}\right) &\geq p, \end{aligned}$$

Convergence results

Lemma

For all realizations $\{\alpha_k\}$ *of* $\{A_k\}$, *independently of* $\{\mathfrak{D}_k\}$, $\lim_{k\to\infty} \alpha_k = 0$.

Lemma

If k *is an unsuccessful iteration; then*

$$\{\mathrm{cm}\left(\mathfrak{D}_k, -\nabla f(X_k)\right) \geq \kappa\} \subset \{\kappa \|\nabla f(X_k)\| \leq \mathcal{O}\left(A_k\right)\}.$$

We can identify conditions under which $\{\mathrm{cm}\left(\mathfrak{D}_k, -\nabla f(X_k)\right) \geq \kappa\}$ happens sufficiently often.

Convergence results

Lemma

For all realizations $\{\alpha_k\}$ *of* $\{A_k\}$*, independently of* $\{\mathfrak{D}_k\}$, $\lim_{k\to\infty} \alpha_k = 0$.

Lemma

If k *is an unsuccessful iteration; then*

$$\{\mathrm{cm}\left(\mathfrak{D}_k, -\nabla f(X_k)\right) \geq \kappa\} \subset \{\kappa \|\nabla f(X_k)\| \leq \mathcal{O}\left(A_k\right)\}.$$

We can identify conditions under which $\{\mathrm{cm}\left(\mathfrak{D}_k, -\nabla f(X_k)\right) \geq \kappa\}$ happens sufficiently often.

Almost-sure Convergence Theorem

If $\{\mathfrak{D}_k\}$ is (p, κ)-descent with $p > p_0 = \ln(\theta)/\ln(\theta\gamma^{-1})$, then

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

Probabilistic complexity bound

Probabilistic worst-case complexity

Let $\{\mathfrak{D}_k\}$ be (p, κ)-descent, $\epsilon \in (0,1)$ and N_ϵ the number of function evaluations needed to have $\inf_{0 \leq l \leq k} \|\nabla f(X_l)\| \leq \epsilon$. Then

$$\mathbb{P}\left(N_\epsilon \leq \mathcal{O}\left(\frac{r\,(\kappa\epsilon)^{-2}}{p-p_0}\right)\right) \geq 1 - \exp\left(-\mathcal{O}\left(\frac{p-p_0}{p}(\kappa\,\epsilon)^{-2}\right)\right).$$

Probabilistic complexity bound

Probabilistic worst-case complexity

Let $\{\mathfrak{D}_k\}$ be (p,κ)-descent, $\epsilon \in (0,1)$ and N_ϵ the number of function evaluations needed to have $\inf_{0 \leq l \leq k} \|\nabla f(X_l)\| \leq \epsilon$. Then

$$\mathbb{P}\left(N_\epsilon \leq \mathcal{O}\left(\frac{r\,(\kappa\epsilon)^{-2}}{p-p_0}\right)\right) \geq 1 - \exp\left(-\mathcal{O}\left(\frac{p-p_0}{p}(\kappa\,\epsilon)^{-2}\right)\right).$$

Practical technique

Use

$$r > \log_2\left(1 - \frac{\ln\theta}{\ln\gamma}\right)$$

directions uniformly distributed over the unit sphere.

- Defines a $(p, \tau/\sqrt{n})$-descent sequence, $p > p_0$.
- Complexity: $\mathcal{O}(n\,\epsilon^{-2})$.
- Ex) $\gamma = \theta^{-1} = 2$, only needs $r = 2$!

Outline

1 Direct search for unconstrained optimization

2 Deterministic direct search on linearly constrained problems

3 Probabilistic feasible descent

4 Numerical results

General linearly constrained problems

Chosen formulation:

$$\begin{cases} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t.} & Ax = b \\ & l \leq x \leq u. \end{cases}$$

Assumptions

- $A \in \mathbb{R}^{m \times n}$, $\text{rank}(A) = m$,
- $(l, u) \in (\mathbb{R} \cup \{-\infty, \infty\})^n$, $l < u$.

Associated direct-search algorithm

- Require **feasible iterates and directions**.
- Same framework otherwise.

Two linearly constrained problems

Linear equality constraints - Done ?

$$\begin{cases} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t.} & Ax = b. \end{cases}$$

- Equivalent to the unconstrained problem $\min_{\tilde{x} \in \mathbb{R}^{n-m}} f(x_0 + W\tilde{x})$ with $W \in \mathbb{R}^{n \times (n-m)}$ orthonormal basis for null(A) and $Ax_0 = b$.
- Deterministic and probabilistic analyses apply !

Two linearly constrained problems

Linear equality constraints - Done ?

$$\begin{cases} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t.} & Ax = b. \end{cases}$$

- Equivalent to the unconstrained problem $\min_{\tilde{x} \in \mathbb{R}^{n-m}} f(x_0 + W\tilde{x})$ with $W \in \mathbb{R}^{n \times (n-m)}$ orthonormal basis for $\text{null}(A)$ and $Ax_0 = b$.
- Deterministic and probabilistic analyses apply !

Bound constrained case - Unclear

$$\begin{cases} \min_{x \in \mathbb{R}^n} & f(x) \\ \text{s.t.} & l \leq x \leq u. \end{cases}$$

Deterministic practice relies on $D_\oplus = \{e_1, \ldots, e_n, -e_1, \ldots, -e_n\}$.

- What makes $D_\oplus$ a good choice ?
- Probabilistic case ?

The bound constrained case

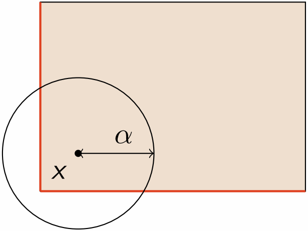
- Feasible set: $\mathcal{F} = \{l \leq x \leq u\}$.

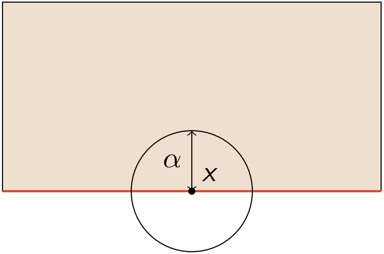
Nearby constraints

The indexes

$$\begin{aligned} I_u(x,\alpha) &= \{i : |u_i - [x]_i| \leq \alpha\} \\ I_l(x,\alpha) &= \{i : |l_i - [x]_i| \leq \alpha\} \end{aligned}$$

define the nearby constraints at $x \in \mathcal{F}$ given $\alpha > 0$.



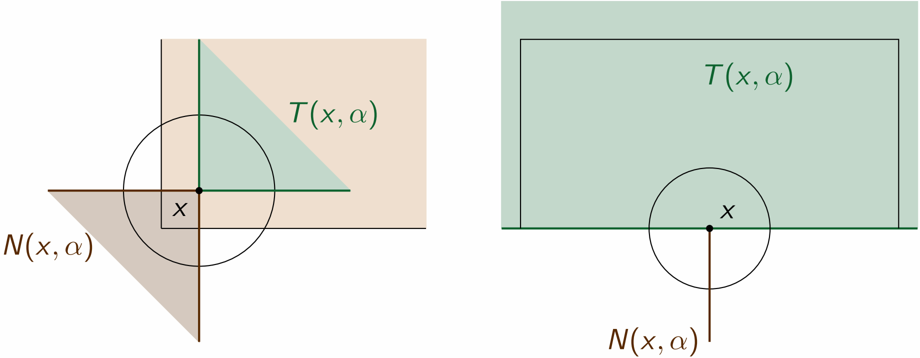


The bound constrained case (2)

- **Approximate normal cone $N(x,\alpha)$:** Positive span of

$$\{e_i\}_{i \in I_u(x,\alpha)} \cup \{-e_i\}_{i \in I_l(x,\alpha)} .$$

- **Approximate tangent cone $T(x,\alpha)$:** polar of $N(x,\alpha)$.



Feasible descent property

Feasible descent property

D is a κ-feasible descent set for $T(x,\alpha)$ if $D \subset T(x,\alpha)$ and

$$\mathrm{cm}_{T(x,\alpha)}(D, -\nabla f(x)) \quad = \quad \max_{d \in D} \frac{d^\top[-\nabla f(x)]}{\|d\| \|P_{T(x,\alpha)}[-\nabla f(x)]\|} \;\geq\; \kappa.$$

Feasible descent property

Feasible descent property

D is a κ-feasible descent set for $T(x,\alpha)$ if $D \subset T(x,\alpha)$ and

$$\mathrm{cm}_{T(x,\alpha)}(D,-\nabla f(x)) \;=\; \max_{d \in D} \frac{d^\top[-\nabla f(x)]}{\|d\|\|P_{T(x,\alpha)}[-\nabla f(x)]\|} \;\geq\; \kappa.$$

- Using κ-feasible descent sets guarantees both convergence and complexity in $\mathcal{O}\left(r(\kappa\epsilon)^{-2}\right)$ (similar analysis than unconstrained case).

Feasible descent property

Feasible descent property

D is a κ-feasible descent set for $T(x,\alpha)$ if $D \subset T(x,\alpha)$ and

$$\mathrm{cm}_{T(x,\alpha)}(D, -\nabla f(x)) \quad = \quad \max_{d \in D} \frac{d^\top[-\nabla f(x)]}{\|d\| \|P_{T(x,\alpha)}[-\nabla f(x)]\|} \;\geq\; \kappa.$$

- Using κ-feasible descent sets guarantees both convergence and complexity in $\mathcal{O}\left(r(\kappa\epsilon)^{-2}\right)$ (similar analysis than unconstrained case).
- **General case:** Use generators of $T(x,\alpha)$.
- **Bounds:** $D_\oplus \cap T(x,\alpha)$ generates $T(x,\alpha)$ and is $\frac{1}{\sqrt{n}}$-feasible descent.

Outline

1. Direct search for unconstrained optimization
2. Deterministic direct search on linearly constrained problems
3. **Probabilistic feasible descent**
4. Numerical results

Probabilistic feasible descent

Definition

A random set sequence $\{\mathfrak{D}_k\}$ is said to be (p,κ)-feasible descent if:

$$\begin{aligned} \mathbb{P}\left(\mathrm{cm}_{T_0}\left(\mathfrak{D}_0, -\nabla f(x_0)\right) \geq \kappa\right) &\geq p \\ \forall k \geq 1, \quad \mathbb{P}\left(\mathrm{cm}_{T_k}\left(\mathfrak{D}_k, -\nabla f(X_k)\right) \geq \kappa \mid \mathfrak{D}_0, \dots, \mathfrak{D}_{k-1}\right) &\geq p. \end{aligned}$$

where $T_k = T(X_k, A_k)$.

Theoretical guarantees

If $\{\mathfrak{D}_k\}$ is (p,κ)-feasible descent with $p > p_0$,

- **Almost-sure convergence towards stationary point**;
- **Complexity bound for ϵ-stationarity:**

$$\mathbb{P}\left(N_\epsilon \leq \mathcal{O}\left(\frac{r(\kappa\,\epsilon)^{-2}}{p-p_0}\right)\right) \geq 1 - \exp\left(-\mathcal{O}\left(\frac{p-p_0}{p}(\kappa\epsilon)^{-2}\right)\right).$$

From theory to pratice

Main concerns

- How to define probabilistic feasible descent sets ?
- What are the orders of r and κ ?
- Can we use **less directions** than in the deterministic case ?

From theory to pratice

Main concerns

- How to define probabilistic feasible descent sets ?
- What are the orders of r and κ ?
- Can we use **less directions** than in the deterministic case ?

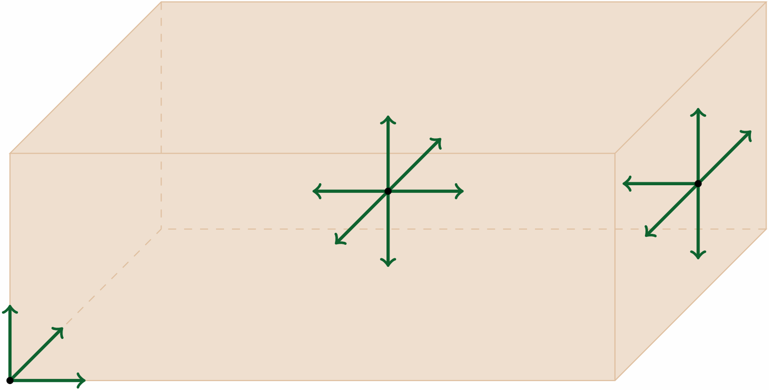
Our approach

- Based on tangent cone generators;
- Choice of a random feasible polling set;
- **At worst as expensive as the deterministic case**.

Possible direction generation techniques

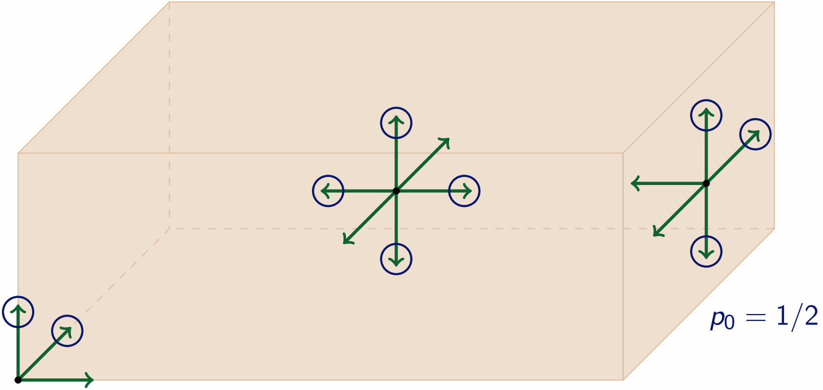
A random generator sampling approach

1. Compute a deterministic generating set V_k for T_k;



A random generator sampling approach

1. Compute a deterministic generating set V_k for T_k;
2. Take a random sample $\mathfrak{D}_k$ of V_k of size $> |V_k| p_0$;
3. $\{\mathfrak{D}_k\}$ is (p, κ)-descent with $p > p_0$.



Using fewer directions by exploiting subspaces

Idea

- Unconstrained case: probabilistic descent can use less directions;
- Also for linear equalities: unconstrained problem in the null space of A;
- Benefit in exploiting unconstrained subspaces ?

Decomposition

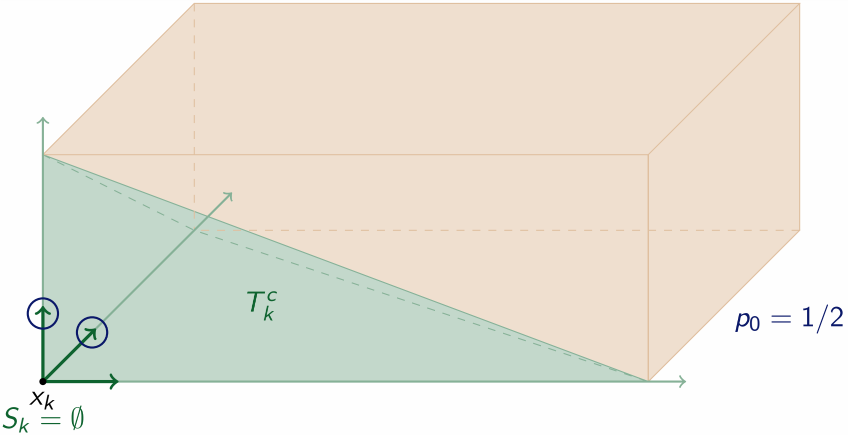
Let S_k be a linear subspace within the cone T_k. Then $T_k = S_k + T_k^c$, where T_k^c is a cone lying in $S_k^\perp$.

Separate strategies for S_k and T_k^c lead to (p, κ)-descent!

Second technique: Illustration

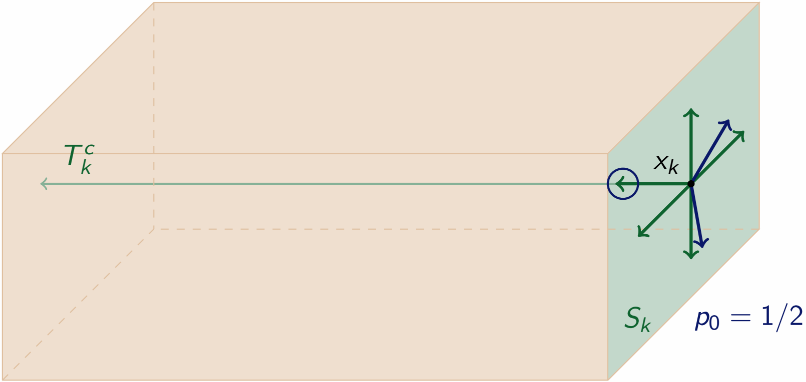
Two types of directions

- Subspace S_k: Directions uniformly at random;
- Orthogonal part T_k^c: Random subset of generators.



Two types of directions

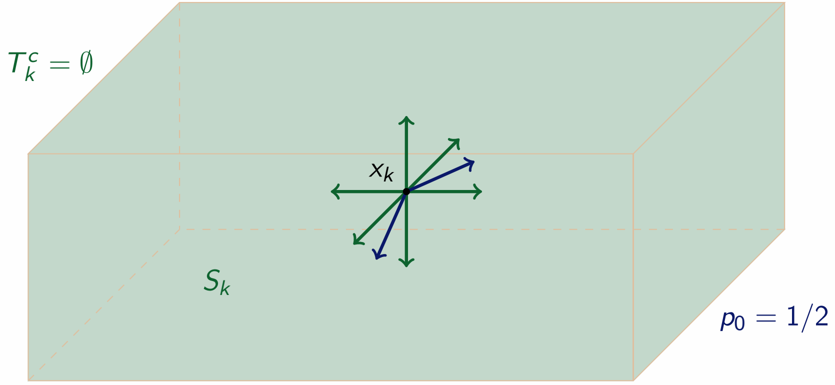
- Subspace S_k: Directions uniformly at random;
- Orthogonal part T_k^c: Random subset of generators.



Second technique: Illustration

Two types of directions

- Subspace S_k: Directions uniformly at random;
- Orthogonal part T_k^c: Random subset of generators.



Complexity aspects

- The general bound is $\mathcal{O}\left(r\kappa^{-2}\epsilon^{-2}\right)$.

Comparison of results - Linear constraints only

Method	r	κ	Bound
Determ.	$2(n-m)$	$\frac{1}{\sqrt{n-m}}$	$\mathcal{O}\left((n-m)^2\epsilon^{-2}\right)$
Proba. 1	$\mathcal{O}(2(n-m)p_0)$	$\frac{1}{\sqrt{n-m}}$	$\mathcal{O}\left((n-m)^2\epsilon^{-2}\right)$
Proba. 2 (subspace)	$\mathcal{O}(1)$	$\frac{\tau}{\sqrt{n-m}}$	$\mathcal{O}\left((n-m)\epsilon^{-2}\right)$

Comparison of results - Bounds on $n_b < n$ variables only

Method	r	κ	Bound
Determ.	$2n$	$\frac{1}{\sqrt{n}}$	$\mathcal{O}\left(n^2\epsilon^{-2}\right)$
Proba. 1	$\mathcal{O}(2np_0)$	$\frac{1}{\sqrt{n}}$	$\mathcal{O}\left(n^2\epsilon^{-2}\right)$
Proba. 2 (subspace)	$\mathcal{O}(1)+\mathcal{O}\left(n_b\,p_0\right)$	$\frac{1}{\sqrt{n}}$	$\mathcal{O}\left(n\,n_b\epsilon^{-2}\right)$

Outline

1. Direct search for unconstrained optimization
2. Deterministic direct search on linearly constrained problems
3. Probabilistic feasible descent
4. Numerical results

Numerical experiments-Bound constraints

- Comparison with MATLAB built-in `patternsearch` function.

Four solvers

Name	Polling in $T(x_k, \alpha_k) = T_k = S_k + T_k^c$	Guarantees
`dspfd-0`	Shuffled $D_\oplus \cap T_k$	Deterministic
`dspfd-1`	Random subset of $D_\oplus \cap T_k$	Probabilistic
`dspfd-2`	Random vectors in S_k/subset of $D_\oplus \cap T_k^c$	Probabilistic
`matlab`	$D_\oplus \cap T(x_k, t\alpha_k), t \in (0, 1]$	Deterministic

Performance profiles

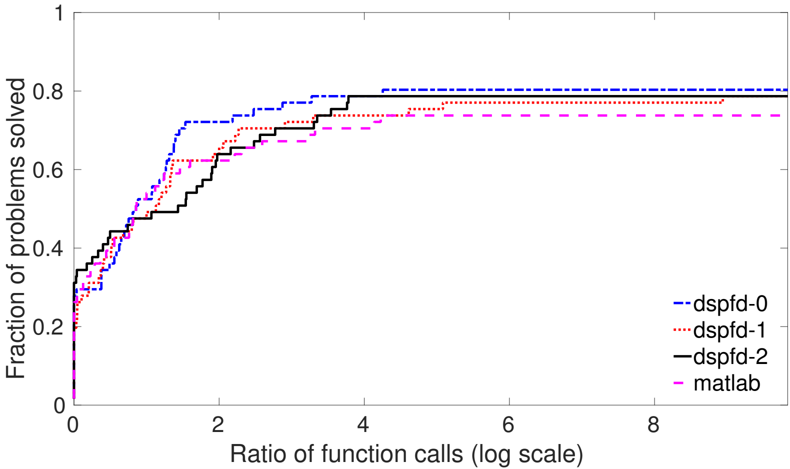
- Criterion: # of function evaluations (budget of $2000n$) to satisfy

$$f(x_k) - f_{best} < 10^{-3}(f(x_0) - f_{best}).$$

- Benchmark: Problems from the CUTEst collection.

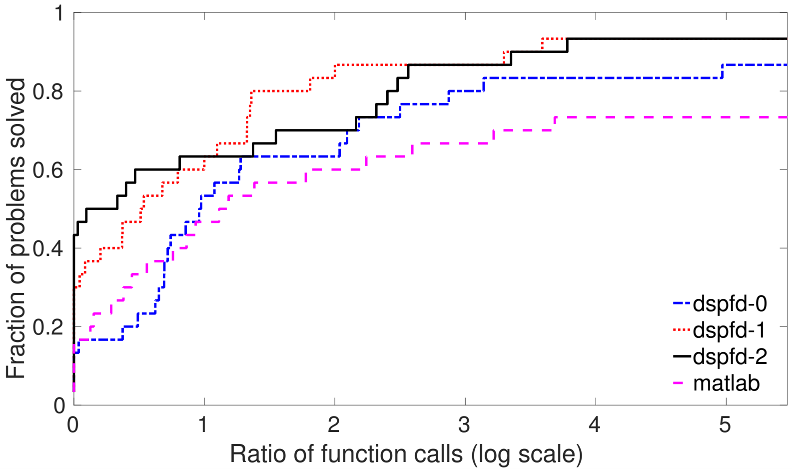
Profiles for bound-constrained problems

- Performance on 63 problems with bounds, small dimensions: $2 \leq n \leq 20$.



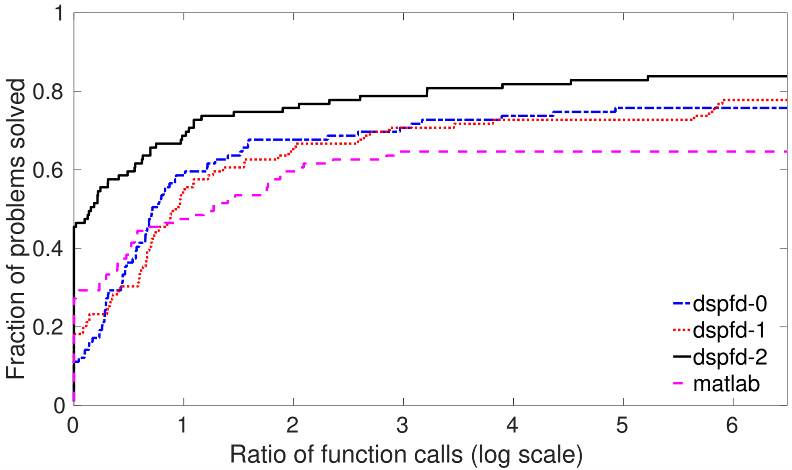
Profiles for bound-constrained problems (2)

- Performance on 31 problems with bounds constraints, **larger dimensions:** $20 \leq n \leq 52$.



Profiles for linearly constrained problems

- Performance on 106 problems with linear constraints, $2 \leq n \leq 96$.



Conclusions

Main contributions

- Use of (probabilistic) feasible descent.
- Random generation in "unconstrained" subspaces.
- Practically efficient method (code available upon request).

Direct search based on probabilistic feasible descent for bound and linearly constrained problems. Gratton, Royer, Vicente and Zhang, *Submitted*, 2017.

Direct search based on probabilistic descent. Gratton, Royer, Vicente and Zhang, *SIAM J. Optim.*, 2015.

- Follow-up: **Nonlinear constraints**.

Conclusions

Main contributions

- Use of (probabilistic) feasible descent.
- Random generation in "unconstrained" subspaces.
- Practically efficient method (code available upon request).

Direct search based on probabilistic feasible descent for bound and linearly constrained problems. Gratton, Royer, Vicente and Zhang, *Submitted*, 2017.

Direct search based on probabilistic descent. Gratton, Royer, Vicente and Zhang, *SIAM J. Optim.*, 2015.

- Follow-up: **Nonlinear constraints**.

Thank you for your attention !

`croyer2@wisc.edu`