Direct Search Using Probabilistic Descent

S. Gratton[1,2], C. W. Royer[1,2], L. N. Vicente[3] and Z. Zhang[3]
[1]INP-ENSEEIHT [2]IRIT-CERFACS, Toulouse, France [3] CMUC, Coimbra, Portugal

Introduction

We aim to solve

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

where f is smooth but **its gradient is not available**.
Direct-search algorithms are derivative-free methods that are based on the exploration of the space through the use of *direction* or *polling sets*. Classic direct-search schemes require these sets to be **Positive Spanning Sets** (PSS), i.e. sets that generate $\mathbb{R}^n$ by positive linear combination. We study here how this assumption can be replaced while using random directions, thus leading to cheaper computations.

Algorithm

1. **Initialization:** Choose $x_0 \in \mathbb{R}^n$, $\alpha_0 > 0$, $\theta < 1 \leq \gamma$ and $\rho(t) = o(t)$ when $t \searrow 0$.
2. **For** $k = 0, 1, 2, \ldots$
   - Choose a set D_k of m *unitary* vectors.
   - If it exists $d_k \in D_k$ so that

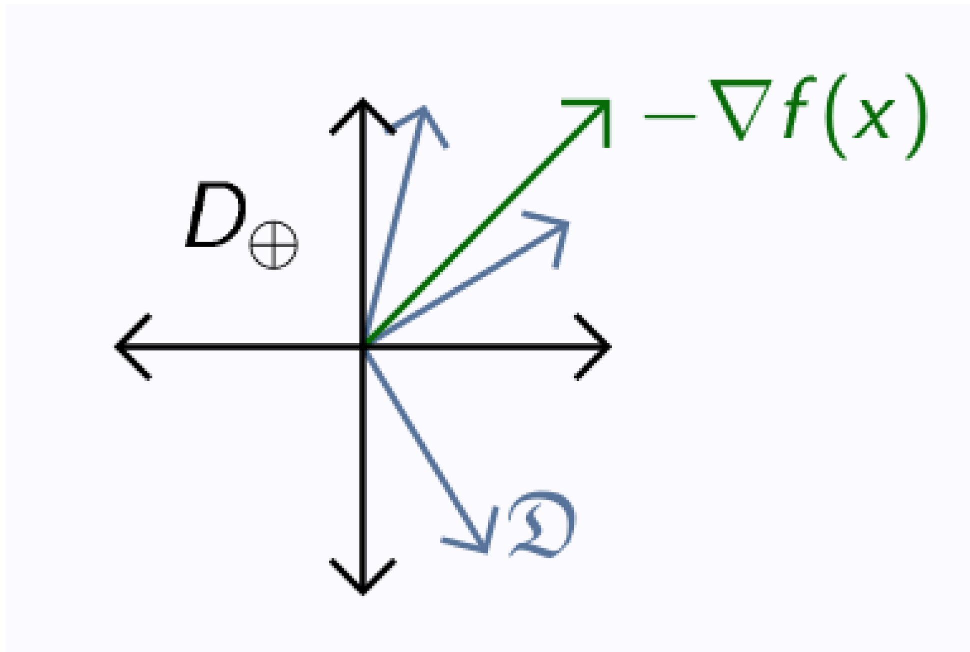
     $$f(x_k + \alpha_k\, d_k) < f(x_k) - \rho(\alpha_k),$$

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

Remarks:

* When D_k is a random set $\mathfrak{D}_k$, it implies that the iterates x_k become random vectors X_k.
* In the rest, $\rho(t) = c\, t^2$ with $c > 0$.

Theoretical results



The interest of PSS is to provide at least one vector that makes an acute angle with the negative gradient, hence is a descent direction. However, being a PSS is not necessary to obtain this property.

The figure on the left shows that the PSS $D_\oplus$ does not approximate the negative gradient as well as $\mathfrak{D}$, although this set is clearly not a PSS.

Probabilistic descent
We say that a sequence $\{\mathfrak{D}_k\}_k$ is p-probabilistically κ-descent (or (p, κ)-descent) if

$$\mathbb{P}\left(E_0\right) \geq p \quad \text{and} \quad \forall k \geq 1,\ \mathbb{P}\left(E_k | S_{k-1}\right) \geq p,$$

where

$$E_k \;=\; \left\{ \max_{d \in \mathfrak{D}_k} \frac{-d^T\, \nabla f(X_k)}{\|\nabla f(X_k)\|} > \kappa \right\}$$

and S_{k-1} is the σ-algebra generated by the previous random sets $\mathfrak{D}_0, \ldots, \mathfrak{D}_{k-1}$.

Convergence theorem
Using a p-probabilistic κ-descent set sequence with $p > \ln\theta / \ln(\gamma^{-1}\theta)$, one can show that

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

Worst-case complexity results
Let N_ϵ the number of function evaluations needed to obtain $\|\nabla f(X_k)\| \leq \epsilon \in (0,1)$, then:

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

- For PSS, $p = 1, m \geq n+1, \kappa \geq 1/\sqrt{n}$, and we recover the deterministic rate in $\mathcal{O}\left(n^2\,\epsilon^{-2}\right)$;
- For uniformly random sets of size m, this reduces to $\mathcal{O}\left(m\,n\,\epsilon^{-2}\right)$ with overwhelming probability.

High probability iteration complexity
For ϵ sufficiently small and $P \in (0,1)$, it holds

$$\mathbb{P}\left(\min_{0 \leq l \leq k} \|\nabla f(X_l)\| \leq \epsilon \right) \geq P \quad \text{whenever} \quad k \;\geq\; \mathcal{O}\left(\kappa^{-2}\epsilon^{-2}\right) - \mathcal{O}(\ln(1-P)).$$

Numerical experiments

Problem	Dim	Set choice 1	Set choice 2
arglina	40	6.35	1.00
arglina	100	18.00	1.00
dqrtic	40	28.32	1.00
dqrtic	100	failed	1.00
vardim	40	2.97	1.00
vardim	100	8.04	1.00

Table 1: Ratio of function evaluations

The tolerance is 10^{-3}, $\rho(t) = 10^{-3}\,t^2$. The problems are from the CUTER package. The ratio corresponds to a mean on 100 runs. At iteration k, there are two different choices for the polling set:

1. A PSS of the form $[Q_k\ {-Q_k}]$ with Q_k orthogonal, $\gamma = 1$;
2. Two random directions i.i.d. uniformly distributed on the unit sphere, $\gamma = 2$.

Further improvements

- Using $[d\ -d]$ with d random is the best choice in terms of sets with two directions.
- Other probability distributions may lead to better results.

Conclusions

- Direct search can be proven convergent without the use of PSS.
- Using (p, κ)-descent sets reduces both theoretical complexity and numerical cost.
- The reasoning for global rates can be used for other algorithms (already applied to a trust-region framework).
- Future work: extending these results to the nonsmooth/convex cases.

References

[1] S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang. Direct search using probabilistic descent. Technical Report 14-11, Dept. Mathematics, Univ. Coimbra, 2014.

[2] A. S. Bandeira, K. Scheinberg, and L. N. Vicente. Convergence of trust-region methods based on probabilistic models. Technical Report 13-11, Dept. Mathematics, Univ. Coimbra, 2013.

[3] L. N. Vicente. Worst Case Complexity of Direct Search. *EURO Journal on Computational Optimization*, 1:143–153, 2013.

Acknowledgments

Support for this research was provided by FCT under grants PTDC/MAT/116736/2010 and PEst-C/MAT/UI0324/2011 and by RTRA under the grant ADTAO.