JOPT 2012

Reducing the number of function evaluations in MADS algorithms

Sébastien Le Digabel, École Polytechnique de Montréal

2012–05–08

C. Audet, C. Tribes, École Polytechnique de Montréal

A. Ianni, Università di Roma La Sapienza

Presentation outline

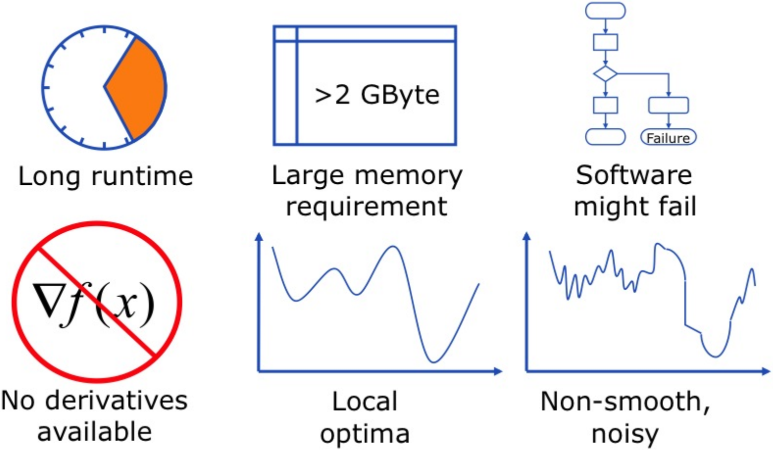
Blackbox optimization problems

We consider the optimization problem:

$$\min_{x \in \Omega} \quad f(x)$$

where evaluations of f and the functions defining Ω are usually the result of a computer code (a blackbox).

Blackboxes as illustrated by J. Simonis [ISMP 2009]



Blackbox optimization problems

The MADS algorithm

Reducing the size of the poll set

Numerical results

Discussion

Mesh Adaptive Direct Search (MADS)

- Audet and Dennis [SIOPT, 2006]
- Iterative algorithm that evaluates the blackbox at some trial points on a spatial discretization called the mesh.
- One iteration = search and poll.
- The search allows trial points generated anywhere on the mesh.
- The poll consists in generating a list of trial points constructed from poll directions. These directions grow dense.
- At the end of the iteration, the mesh size is reduced if no new iterate is found.
- Algorithm is backed by a convergence analysis based on the Clarke Calculus for nonsmooth functions.

The poll (1/2)

- Mesh at iteration k:
  - $M_k = \bigcup_{x \in V_k} \{x + \Delta_k D z : z \in \mathbb{N}^{n_D}\}$.
  - V_k is the "cache".
  - $\Delta_k > 0$ is the mesh size parameter.
  - D is a fixed set of directions typically set to $[I \ \ -I]$.

- Poll directions: A positive spanning set $D_k \subset \mathbb{R}^n$ where each direction $d \in D_k$ can be written as a nonnegative integer combination of directions of D.

- The directions correspond typically to a minimal positive basis ($n+1$ directions) or a maximal positive basis ($2n$ directions).

The poll (2/2)

- Poll set: $P_k = \{x_k + \Delta_k d : d \in D_k\}$ where x_k is the current incumbent, or the poll center.
- The trial points in P_k are evaluated following the opportunistic strategy: evaluations are interrupted as soon as a new better solution is found.
- Trial points ordering is then crucial in practice. It can be based on:
  - Model or surrogate values.
  - Angle with the gradient of a model.
  - Angle with the last direction of success.

[0] Initializations (x_0, Δ_0)

[1] Iteration k

- **[1.1] Search**
  - select a finite number of mesh points
  - evaluate candidates opportunistically
- **[1.2] Poll** (if the Search failed)
  - construct poll set $P_k = \{x_k + \Delta_k d : d \in D_k\}$
  - sort(P_k)
  - evaluate candidates opportunistically

[2] Updates

- if success
  - $x_{k+1} \leftarrow$ success point
  - increase Δ_k
- else
  - $x_{k+1} \leftarrow x_k$
  - decrease Δ_k
- $k \leftarrow k + 1$, stop or go to **[1]**

Motivation

- MADS is a generic algorithmic framework with two practical implementations: LT-MADS and Ortho-MADS.
- For various reasons, Ortho-MADS is preferred to LT-MADS. LT-MADS defines $n+1$ and $2n$ types of directions, and Ortho-MADS has only the $2n$ variant:

	LT-MADS(2006)	Ortho-MADS (2009)
$n+1$	✓	X
$2n$	✓	✓

- Some tests suggest that the LT-MADS implementation is more efficient with $n+1$ directions. We want to define an Ortho-MADS variant using $n+1$ directions.

General framework (1)

Idea: Given a poll set of $2n$ trial points, prune it to n points and add a direction to obtain $n+1$ points.

Poll at iteration k

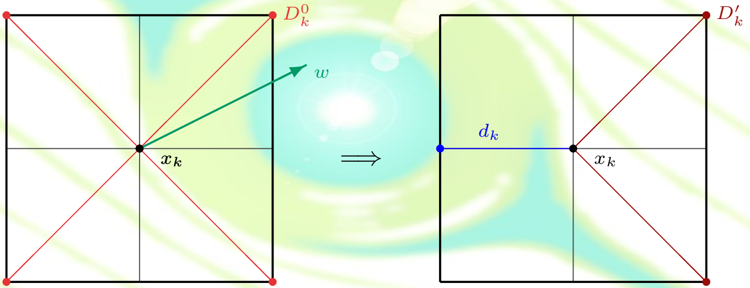
$P_k^o = \{x_k + \Delta_k d : d \in D_k^o\}$ (original poll set)
extract $D'_k \subset D_k^o$
compute new direction d_k
$D_k = D'_k \cup \{d_k\}$
construct $P_k = \{x_k + \Delta_k d : d \in D_k\}$ (reduced poll set)
sort(P_k)
evaluate(P_k) (opportunistically)

Ortho-MADS $n+1$ is a simple first implementation of this general framework.

Ortho-Mads $n+1$

- $D_k^o = [H_k \ \ -H_k]$ is the original Ortho-MADS spanning set with $2n$ directions and $H_k \in \mathbb{Z}^{n\times n}$ an orthogonal basis with integer coefficients.
- The selection of n columns of D_k^o to obtain D'_k is based on a target direction $w \in \mathbb{R}^n$.
- The target direction is taken as the last direction of success.
- The $(n+1)$th direction is $d_k = -\sum_{d \in D'_k} d$.

Ortho-MADS $n+1$: Idea



$2n$ directions D_k^0 $\qquad$ $n+1$ directions $D_k = D'_k \cup \{d_k\}$

Completion using function values

- We now present a second and more general framework.
- This version is not limited to Ortho-MADS and may be applied to any poll sets. For example hybrid versions with more that $2n$ points.
- The first framework is decomposed allowing to evaluate n trial points in a first step and possibly a last $(n+1)$th point $y_k = x_k + d_k\Delta_k$.
- y_k is constructed by exploiting the function values at the first n points.
- y_k must lie on the mesh and d_k must be inside the cone of the negative directions of D'_k so that the poll directions remain a positive spanning set.

General framework (2)

Poll at iteration k

- $P_k^o = \{x_k + \Delta_k d : d \in D_k^o\}$ (original poll set)
- extract $D'_k \subset D_k^o$ and construct P'_k
- sort (P'_k)
- evaluate (P'_k) (opportunistically)
- Success
  - interrupt iteration
- Failure
  - compute new direction d_k
  - evaluate $(x_k + \Delta_k d_k)$

Use of quadratic models

Quadratic models may be used at two different levels:

1. In Ortho-MADS $n+1$: The simplex gradient is taken as the target direction w.
2. In Framework (2): Optimize a model to determine the last trial point $y_k = x_k + \Delta_k d_k$.

Blackbox optimization problems

The MADS algorithm

Reducing the size of the poll set

Numerical results

Discussion

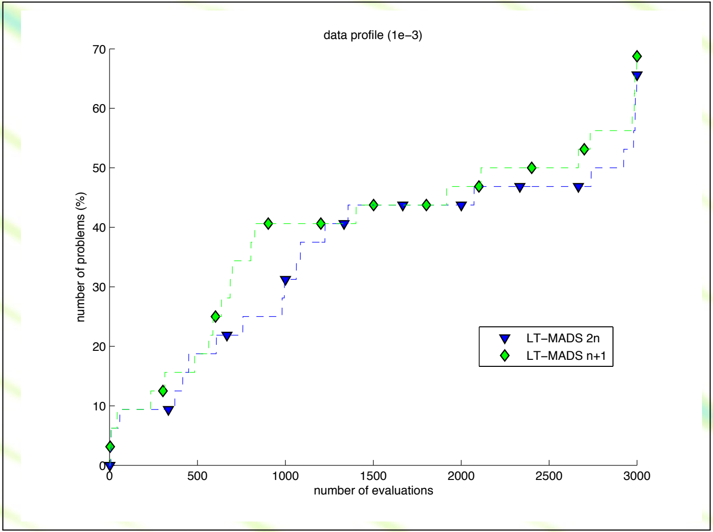
Set of problems and conditions

- These are preliminary tests.
- 32 analytic problems from the DFO literature.
- Number of variables: n ranges from 2 to 12.
- Constraints: 7 problems have from 1 to 15 constraints.
- A budget of 3000 evaluations is considered.
- Results are illustrated by data profiles with a precision of 10^{-3}.

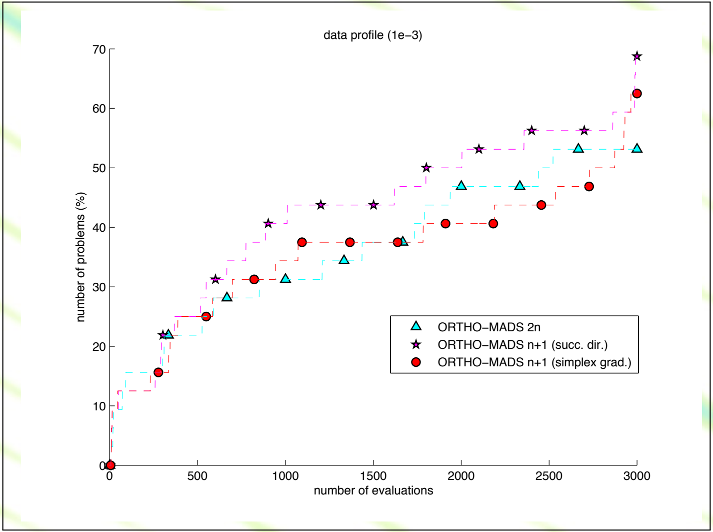
Summary of the tests

1. LT-MADS $2n$ vs $n+1$.
2. Ortho-MADS $2n$ vs Ortho-MADS $n+1$ with two variants: w success direction or simplex gradient.
3. LT-MADS vs Ortho-MADS.

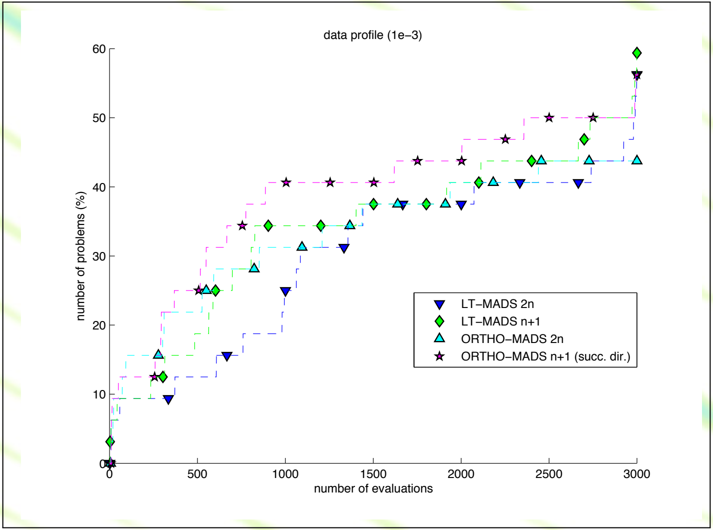
1. LT-MADS $2n$ vs $n+1$



2. Ortho-MADS $2n$ vs $n+1$



3. LT-MADS vs Ortho-MADS



Discussion

- This work is based on the remark that using minimal positive bases instead of maximal positive bases seems preferable.
- It defines several methods in order to achieve the reduction of poll directions from $2n$ to $n+1$ points.
- Preliminary tests demonstrate significative improvements.
- Future work includes more testing on true applications and development of smarter ways to use quadratic models and to treat the constraints.
- Our blackbox optimization solver: www.gerad.ca/nomad.