Two-phase derivative-free constrained optimization

Francisco N. C. Sobral
(State University of Maringá – Brazil)

Computational Optimization in Action
The University of Edinburgh
08.05.2018

Outline

1. Introduction
2. Two-phase optimization algorithms
   - Skinny – Direct searches for "thin" domains
   - IRDF – Inexact Restoration Derivative-free
   - FIRD – Filter Inexact-Restoration Derivative-free
3. Conclusions

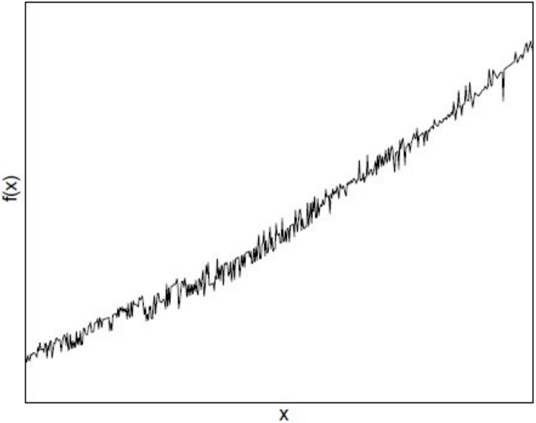
No derivatives?

- Analytic expression of f is unavailable or complex
- User do not know/want to calculate derivatives
- f is a black box function

- f is not differentiable

No derivatives?

- Derivatives do not provide useful information



Kolda, Lewis e Torczon (2003)

The problem

We are interested in solving the following optimization problem

$$\begin{array}{ll} \min & f(x) \\ \text{s. t.} & g(x) \leq 0 \\ & h(x) = 0 \\ & x \in \mathcal{X}, \end{array}$$

where $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \to \mathbb{R}^m$, $h : \mathbb{R}^n \to \mathbb{R}^p$

The problem

- f can be differentiable, but its derivatives are unavailable
- The feasible set

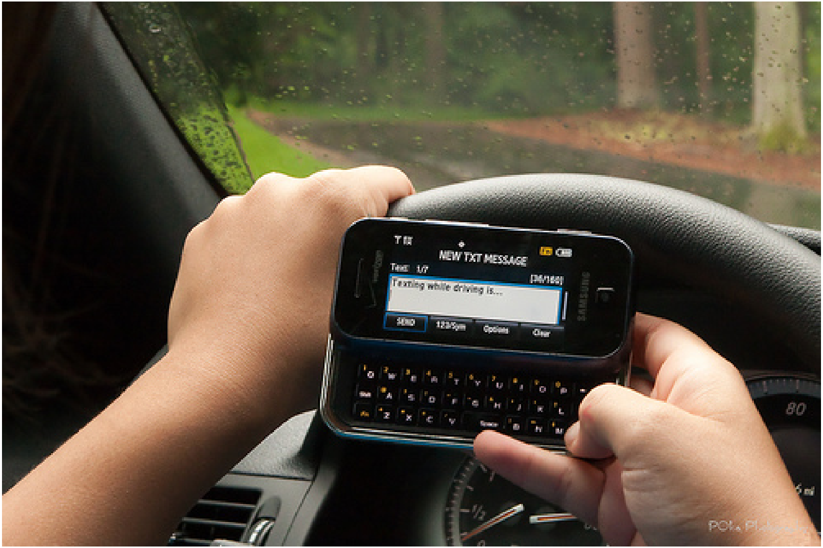
$$\Omega = \{x \in \mathbb{R}^n \mid g(x) \leq 0 \text{ and } h(x) = 0\}$$

  has constraints that may or may not have derivatives available
- $\mathcal{X} \subseteq \mathbb{R}^n$ contains unrelaxable or hard constraints (bounds, black boxes, etc.)

The problem

- The computational cost of evaluating f is high
- Find $x \in \Omega \cap \mathcal{X}$ is computationally cheap, but demands a large amount of operations
- f cannot be evaluated outside $\mathcal{X}$

Two-phase optimization



Two-phase optimization

The optimization process is split into two phases (Martínez, 1998):

- Feasibility (or Restoration): feasibility is improved without objective function evaluations and bounded deterioration of optimality

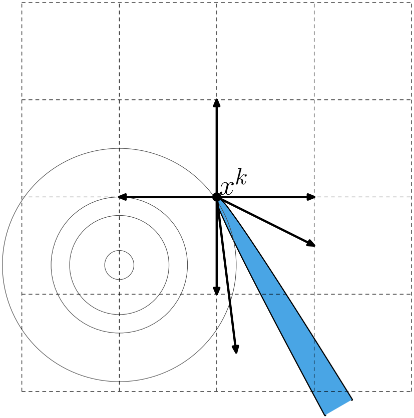
  No computational impact since f is not evaluated

- Optimality (or Minimization): objective function value is reduced on a "relaxed" subproblem

  Derivative-free subproblems can be solved by specific algorithms

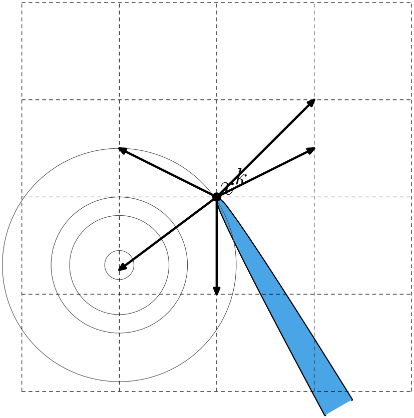
Direct searches and "thin" domains

- Direct search for general nonlinear constraints



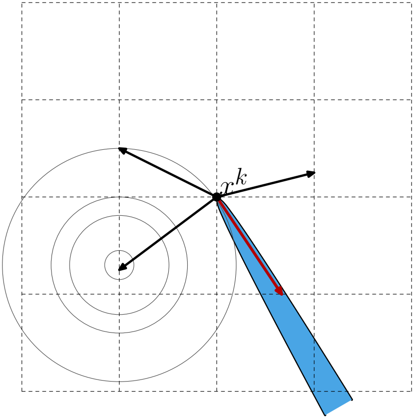
Direct searches and "thin" domains

- Direct search for general nonlinear constraints



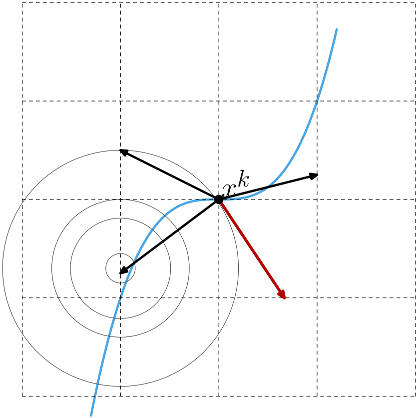
Direct searches and “thin” domains

- Direct search for general nonlinear constraints



Direct searches and "thin" domains

- Direct search for general nonlinear constraints



Skinny – Direct searches for "thin" domains

Let us consider, for simplicity,

$$\begin{array}{ll} \min & f(x) \\ \text{s. t.} & g(x) \leq 0 \\ & x \in \mathcal{X} \end{array}$$

- Equality constraints are given by two inequalities
- f and g are continuous

- Code: `http://fsobral.github.io/skinny`

Algorithm

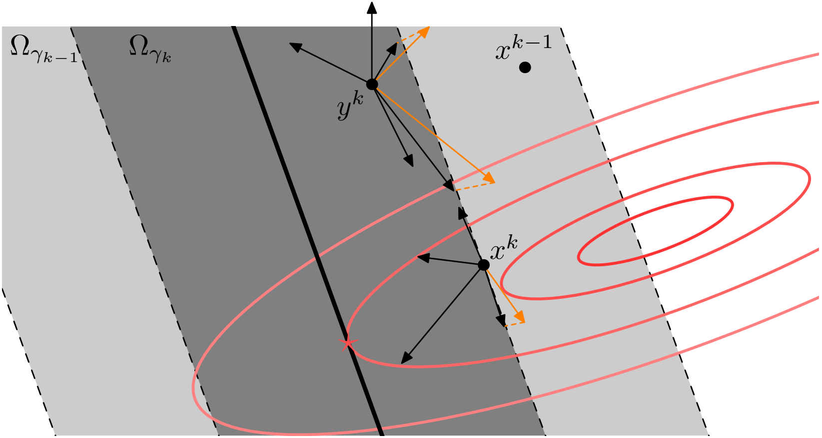
Feasibility. We try to find y such that:

$$y \in \mathcal{X} \quad \text{and} \quad y \in \Omega_\gamma = \{x \in \mathbb{R}^n \mid g(x) \leq \gamma\}, \ \gamma \geq 0$$

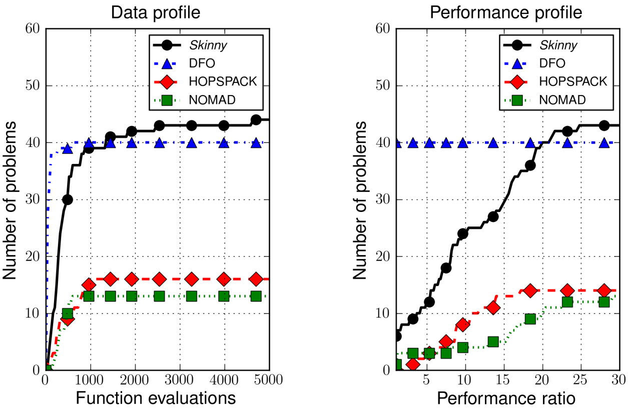
Optimality. We use well-established algorithms to solve the following derivative-free subproblem:

$$\begin{array}{ll} \min & f(x) \\ \text{s. t.} & x \in \Omega_\gamma \cap \mathcal{X} \end{array}$$

Algorithm



Results



Martínez, JM and Sobral, FNC, (2013). "Constrained derivative-free optimization on thin domains". Journal of Global Optimization.

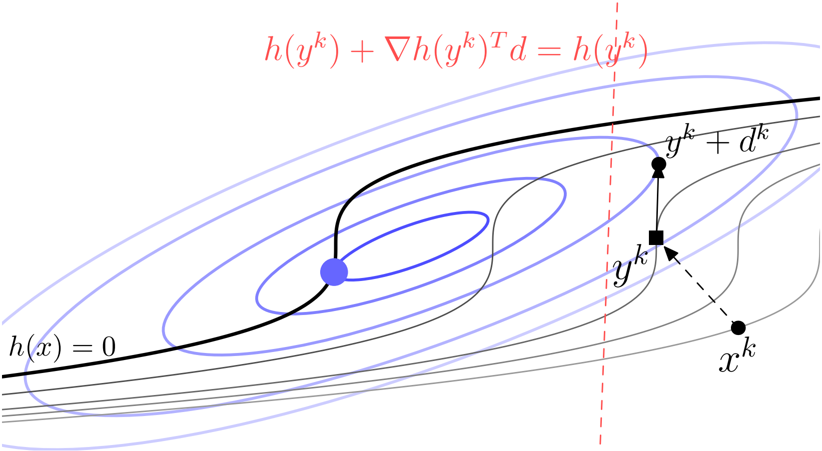
IRDF – Inexact-Restoration Derivative-free

Let us consider the following problem

$$\begin{array}{ll} \min & f(x) \\ \text{s. t.} & h(x) = 0 \\ & x \in \mathcal{X} \end{array}$$

- f and h are continuously differentiable
- The derivatives of h are available
- $\mathcal{X}$ contains linear and bound constraints
- The aim is to extend the classical Inexact Restoration results (Fischer and Friedlander, 2010) to the derivative-free case

IRDF – Inexact-Restoration Derivative-free



Algorithm

Parameter initialization. $k \leftarrow 1$.

Step 1. (**Feasibility**) Find $y^k \in \mathcal{X}$ which reduces $\|h(x^k)\|$.

Step 2. Update the penalty parameter of a certain merit function.

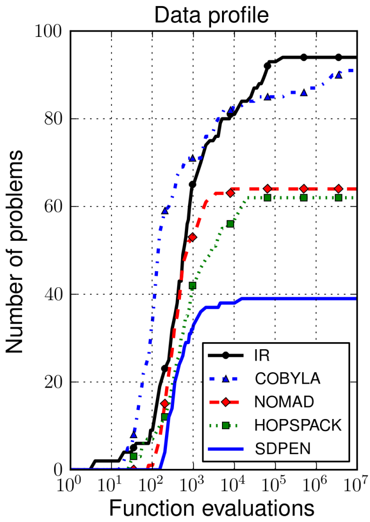
Step 3. (**Optimality and regularization**) Compute d^k, the approximate solution of

$$\begin{array}{ll} \min & f(y^k + d) + \mu \|d\|_2^2 \\ \text{s. t.} & \nabla h(y^k)^T d = 0 \\ & y^k + d \in \mathcal{X} \end{array}$$

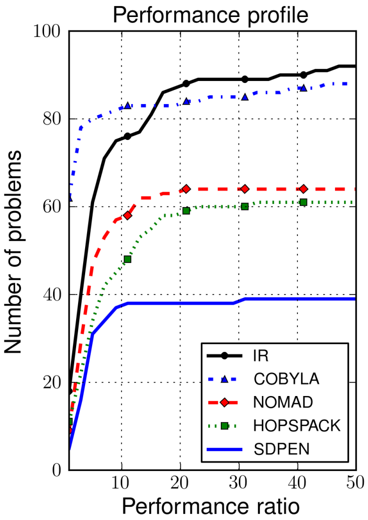
Step 4. If $y^k + d^k$ decreases the merit function and the objective function, then $x^{k+1} = x^k + d^k$.
Otherwise, increase μ and go back to **3**.

Step 5. $k \leftarrow k + 1$ and go back to **1**.

Results



Data profile
Number of problems
Function evaluations
100
80
60
40
20
0
10^0 10^1 10^2 10^3 10^4 10^5 10^6 10^7
IR
COBYLA
NOMAD
HOPSPACK
SDPEN



Performance profile
Number of problems
Performance ratio
100
80
60
40
20
0
10 20 30 40 50
IR
COBYLA
NOMAD
HOPSPACK
SDPEN

Results

- The algorithm is well defined
  - Feasibility step is always possible
  - Penalty parameter is bounded away from zero
  - Direction d^k is found in a finite number of iterations
- $\lim_{k\to\infty} \|h(x^k)\|_2 = \lim_{k\to\infty} \|h(y^k)\|_2 = \lim_{k\to\infty} \|d^k\|_2 = 0$
- Under weak constraint qualifications, every limit point is stationary

Bueno, LF, Friedlander, A, Martínez, JM and Sobral, FNC (2013). "Inexact Restoration Method for Derivative-Free Optimization with Smooth Constraints". SIAM Journal on Optimization.

FIRD – Filter Inexact-Restoration Derivative-free

Let us now consider the following problem

$$\begin{array}{ll} \min & f(x) \\ \text{s. t.} & c_{\mathcal{E}}(x) = 0 \\ & c_{\mathcal{I}}(x) \leq 0 \end{array}$$

- $f, \; c_i : \mathbb{R}^n \longrightarrow \mathbb{R}$, for $i \in \mathcal{E} \cup \mathcal{I}$ are continuously differentiable
- The derivatives of the constraints are available
- The aim is to replace
  - merit function by filters
  - direct search by a derivative-free trust region algorithm
  - extend the results of (Gonzaga, Karas, Vanti, 2003) to the derivative-free case

Filters

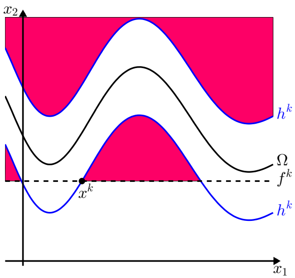
Let

$$h(x) = \left\| c^+(x) \right\|,$$

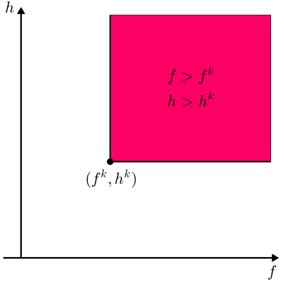
where

$$c_i^+(x) = \begin{cases} c_i(x), & \text{for } i \in \mathcal{E} \\ \max\{0, c_i(x)\}, & \text{for } i \in \mathcal{I} \end{cases}$$

A filter is a set of pairs $F = \{(f^j, h^j), \quad j = 1, \ldots, n_F\}$



Domain of f

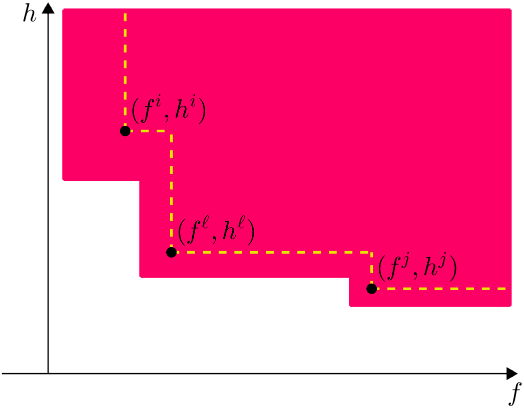


Plane $f \times h$

Filters

Flat filter (Fletcher and Leyffer, 2002)

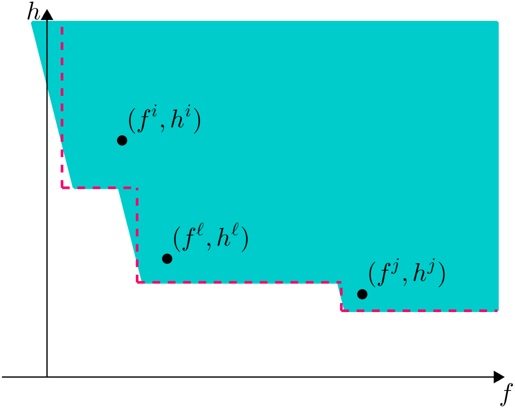
$$\mathcal{R}_j = \left\{ x \in \mathbb{R}^n \,\middle|\, f(x) \geq f^j - \alpha h^j \quad \text{and} \quad h(x) \geq (1-\alpha) h^j \right\},\ \alpha \in (0,1)$$



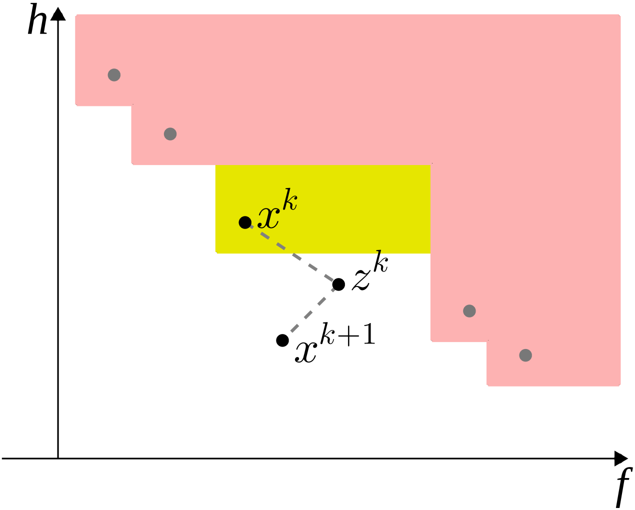
Filters

Slanting filter (Chin and Fletcher, 2003)

$$\mathcal{R}_j = \left\{x \in \mathbb{R}^n \,\middle|\, f(x) + \alpha h(x) \geq f^j \quad \text{and} \quad h(x) \geq (1-\alpha)h^j \right\},\ \alpha \in (0,1)$$



Algorithm



Algorithm

Step 1. Define $\widehat{F}_k = F_k \cup \{(f^k, h^k)\}$ and $\widehat{\mathcal{F}}_k = \mathcal{F}_k \cup \mathcal{R}_k$.

Step 2. (**Feasibility**) Obtain $z^k \notin \widehat{\mathcal{F}}_k$ close to x^k which improves the infeasibility measure.

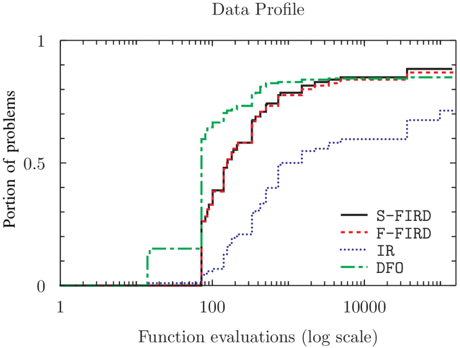
Step 3. (**Optimality**) Compute d^k by approximately solving

$$\begin{array}{ll} \min & f(z^k) + d^T g + \frac{1}{2} d^T B d \\ \text{s. t.} & J c_{\mathcal{E}}(z^k) d = 0 \\ & c_{\mathcal{I}}(z^k) + J c_{\mathcal{I}}(z^k) d \leq c_{\mathcal{I}}^{+}(z^k) \\ & \|d\| \leq \Delta \end{array}$$

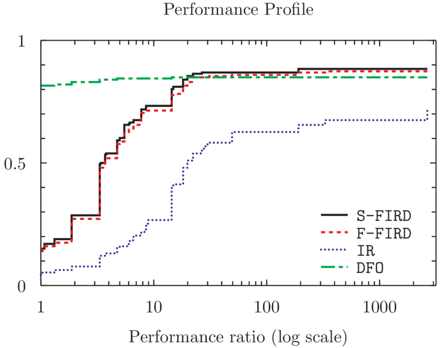
Step 4. If $z^k + d^k$ belongs to $\widehat{\mathcal{F}}_k$ or increases f, then reduce Δ, recompute g and H and go to **Step 3**.

Step 5. Compute $x^{k+1} = z^k + d^k$, update the filter and the forbidden region, F_{k+1} and $\mathcal{F}_{k+1}$ and go to **Step 1**.

Results



Data Profile
Portion of problems
Function evaluations (log scale)
1
0.5
0
1
100
10000
S-FIRD
F-FIRD
IR
DFO



Performance Profile
Performance ratio (log scale)
1
0.5
0
1
10
100
1000
S-FIRD
F-FIRD
IR
DFO

Results

- Under reasonable assumptions about the quadratic model, **Step 3** is executed a finite number of times at each iteration
- The algorithm satisfies an Efficiency Condition (Gonzaga, Karas, Vanti, 2003)
  - The sequence has a stationary accumulation point
  - If the slanting filter is used, any accumulation point of the sequence is stationary
- Code: `https://github.com/fsobral/fird`

Ferreira, PS, Karas, EW, Sachine, M, Sobral, FNC (2017). "Global convergence of a derivative-free inexact restoration filter algorithm for nonlinear programming". Optimization.

Conclusions

- Three two-phase derivative-free algorithms for constrained problems were presented
- Under usual hypotheses, global convergence to stationary points has been achieved
- Two-phase algorithms can explore the structure of derivative-free problems when the computation of the objective function is much more expensive than the constraints
- Different methods for the feasibility and optimality phases are possible

Thank you!

fncsobral@uem.br

fsobral.github.io

Credits

- Slide 5: `https://www.flickr.com/photos/vcucns/8662668483`