Globalization strategies for Mesh Adaptive Direct Search

Charles Audet, École Polytechnique de Montréal
John Dennis, Rice University
Sébastien Le Digabel, École Polytechnique de Montréal

July 2009

Thanks to: AFOSR, ExxonMobil, Boeing,
LANL, FQRNT, NSERC, IREQ.

Presentation outline

1. Handling constraints in real problems
   - Three types of constraints
   - Strategies to deal with constraints
   - Three instantiations of mesh adaptive direct searches
   - Hierarchical convergence analysis

2. Numerical results on engineering problems
   - Three real test problems
   - A feasible starting point
   - An infeasible starting point
   - Multiple runs

3. Discussion

Blackbox optimization problems

My main research interest is nonsmooth optimization:

$$\begin{array}{ll} \text{minimize} & f(x) \\ \text{subject to} & x \in \Omega = \{x \in X : c_j(x) \leq 0, j \in J\} \subset \mathbb{R}^n, \end{array}$$

where

- $f, c_j : X \to \mathbb{R} \cup \{\infty\}$ for all $j \in J = \{1, 2, \ldots, m\}$,
- X is a subset of $\mathbb{R}^n$,
- evaluation of the functions are usually the result of a computer code (a black box) – costly to evaluate.

Presentation outline

1. Handling constraints in real problems
   - Three types of constraints
   - Strategies to deal with constraints
   - Three instantiations of mesh adaptive direct searches
   - Hierarchical convergence analysis

2. Numerical results on engineering problems
   - Three real test problems
   - A feasible starting point
   - An infeasible starting point
   - Multiple runs

3. Discussion

Three types of constraints

The domain: $\Omega = \{x \in X : c_j(x) \leq 0, j \in J\} \subset \mathbb{R}^n$

- Unrelaxable constraints define X

  Cannot be violated by any trial point.
  For example, logical conditions on the variables indicating if the simulation may be launched.

Three types of constraints

The domain: $\Omega = \{x \in X : c_j(x) \leq 0, j \in J\} \subset \mathbb{R}^n$

- Unrelaxable constraints define X
- Relaxable constraints $c_j(x) \leq 0$

  Can be violated, and $c_j(x)$ provides a measure of how much the constraint is violated. A budget for example.

Three types of constraints

The domain: $\Omega = \{x \in X : c_j(x) \leq 0, j \in J\} \subset \mathbb{R}^n$

- Unrelaxable constraints define X
- Relaxable constraints $c_j(x) \leq 0$
- Hidden constraints

  Is a convenient term to exclude the set of points in the feasible region for the relaxable or unrelaxable constraints at which the black box fails to return a value for one of the problem functions. A typical example is when the simulation crashes unexpectedly.

Three strategies to deal with constraints

- Extreme barrier (EB)

Treats the problem as being unconstrained,
by replacing the objective function $f(x)$ by

$$f_\Omega(x) \;:=\; \begin{cases} f(x) & \text{if } x \in \Omega, \\ \infty & \text{otherwise.} \end{cases}$$

The problem

$$\min_{x \in \mathbb{R}^n} f_\Omega(x)$$

is then solved.
Remark : If $x \notin X$ (the non-relaxable constraints), then the costly evaluation of $f(x)$ is not performed.

Three strategies to deal with constraints

- Extreme barrier (EB)
- Progressive barrier (PB)

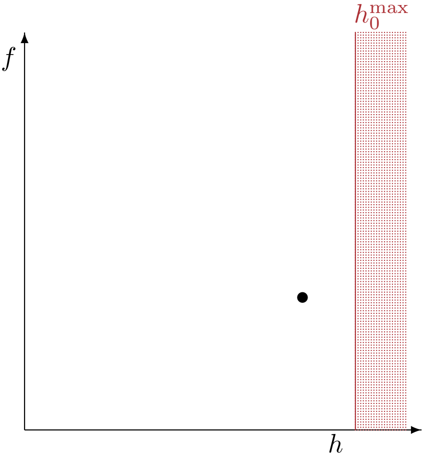
  Defined for the relaxable constraints.
  As in the filter methods of Fletcher and Leyffer, it uses the non-negative constraint violation function $h : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$

$$h(x) \;:=\; \begin{cases} \sum_{j \in J} (\max(c_j(x), 0))^2 & \text{if } x \in X, \\ \infty, & \text{otherwise.} \end{cases}$$

  At iteration k, points with $h(x) > h_k^{\max}$ are rejected by the algorithm, and $h_k^{\max} \to 0$ as $k \to \infty$.

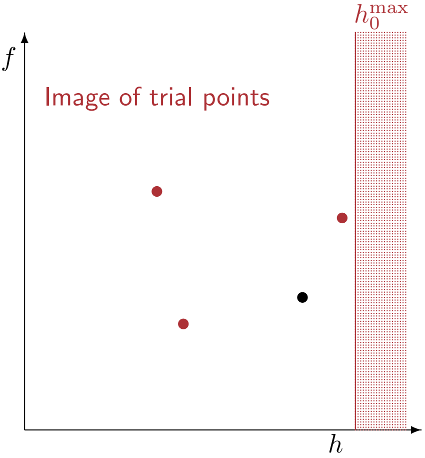
Three strategies to deal with constraints

- Extreme barrier (EB)
- Progressive barrier (PB)



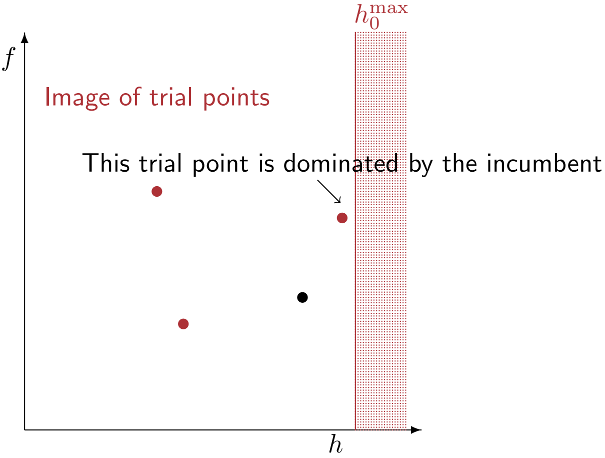
Three strategies to deal with constraints

- Extreme barrier (EB)
- Progressive barrier (PB)

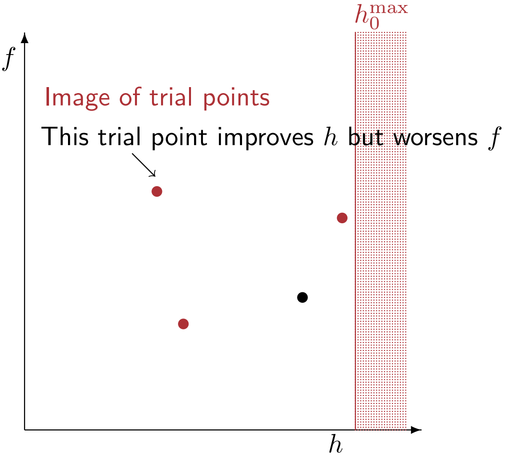


Three strategies to deal with constraints

- Extreme barrier (EB)
- Progressive barrier (PB)

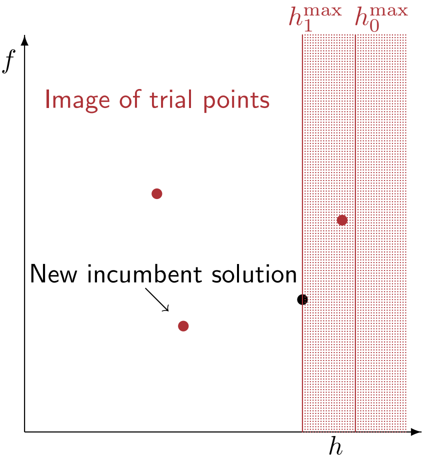


- Extreme barrier (EB)
- Progressive barrier (PB)



Three strategies to deal with constraints

- Extreme barrier (EB)
- Progressive barrier (PB)



Three strategies to deal with constraints

- Extreme barrier (EB)
- Progressive barrier (PB)
- Progressive-to-Extreme Barrier (PEB)

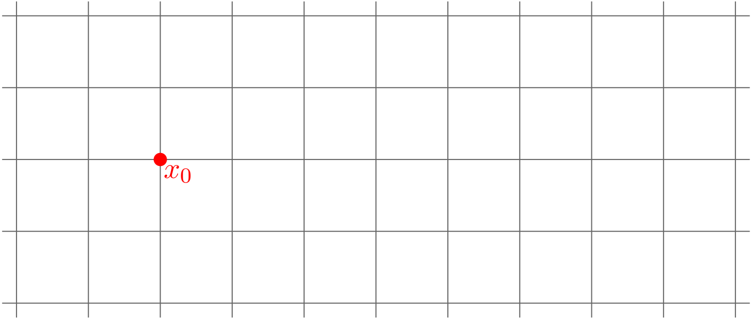
  Initially treats a relaxable constraint by the progressive barrier. Then, if polling around the infeasible poll center generates a new infeasible incumbent that satisfies a constraint violated by the poll center, then that constraint moves from being treated by the progressive barrier to the extreme barrier.

Infeasible starting point

- The progressive and progressive-to-extreme barrier approaches allow initial points that violate the relaxable constraints $c_j(x) \leq 0$.
- A two-phase method can be ran on the relaxable constraints that we want to treat by the extreme barrier approach.
  - The first phase minimizes the constraint violation function subject to $x \in X$, the unrelaxable constraints.
  - Avoids expensive computations of f.
  - The first phase terminates as soon as a $h = 0$, providing an initial point for the second phase.

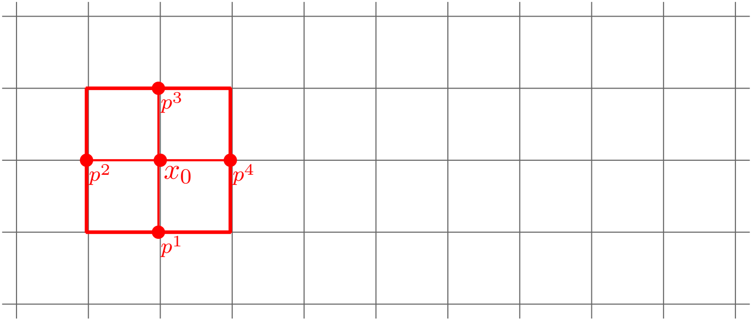
Three instantiations of mesh adaptive direct searches

- GPS with coordinate search.



Three instantiations of mesh adaptive direct searches

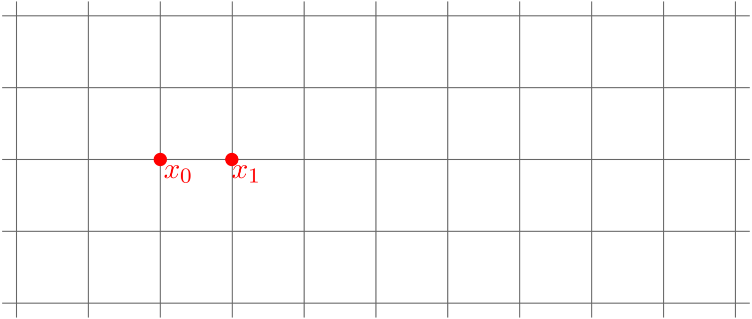
- GPS with coordinate search.



$$f(p^4) < f(x_0)$$

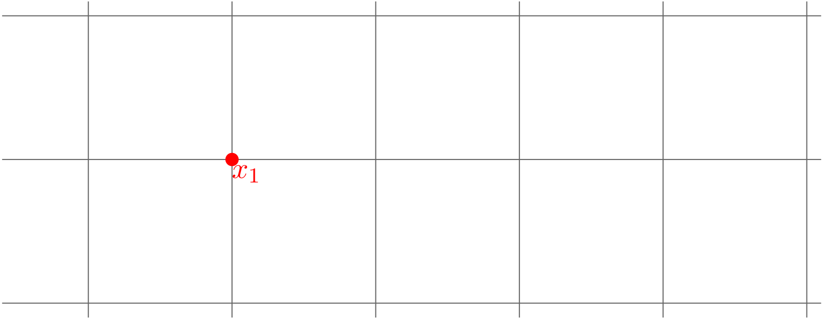
Three instantiations of mesh adaptive direct searches

- GPS with coordinate search.

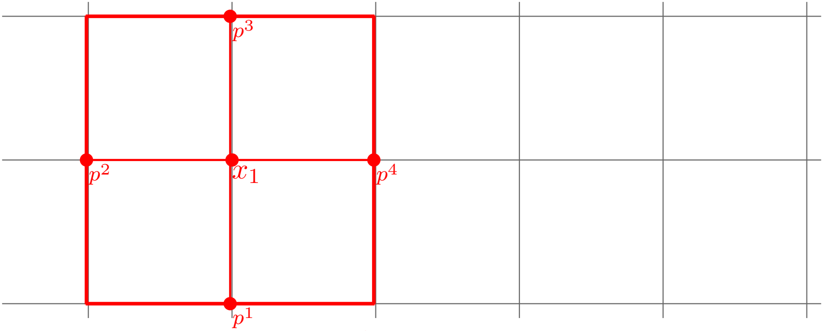


Three instantiations of mesh adaptive direct searches

- GPS with coordinate search.

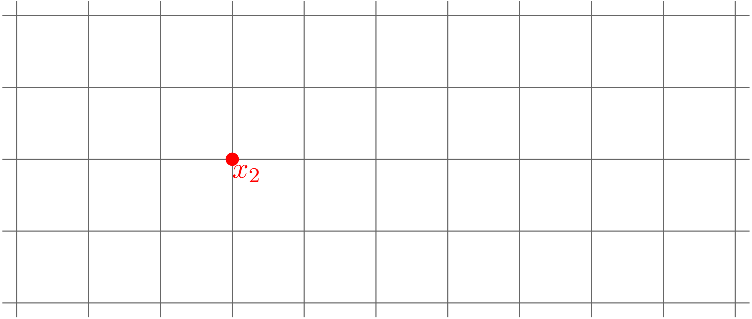


- GPS with coordinate search.



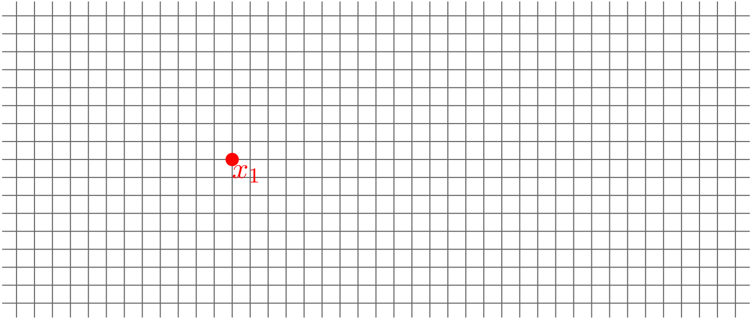
$$f(p^i) \geq f(x_1)$$

- GPS with coordinate search.



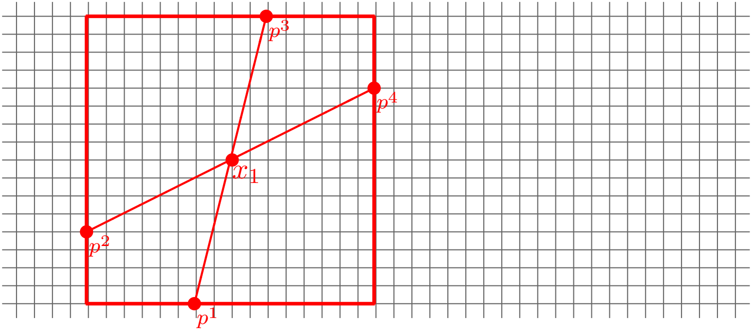
Three instantiations of mesh adaptive direct searches

- GPS with coordinate search.
- LTMADS a non-deterministic implementation of MADS. Union of normalized polling directions grows dense in the unit sphere with probability one.



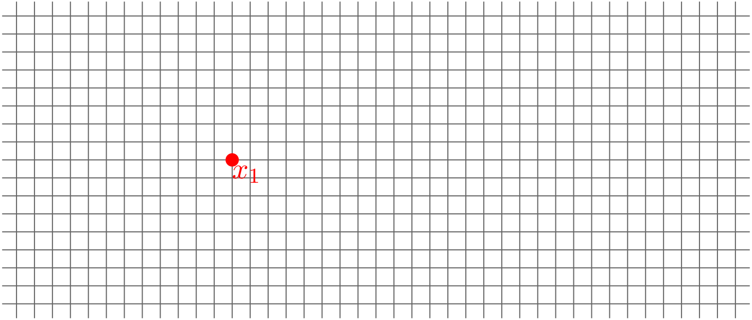
Three instantiations of mesh adaptive direct searches

- GPS with coordinate search.
- LTMADS a non-deterministic implementation of MADS. Union of normalized polling directions grows dense in the unit sphere with probability one.



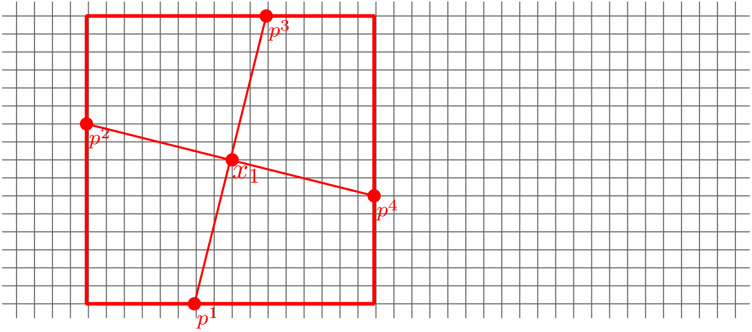
Three instantiations of mesh adaptive direct searches

- GPS with coordinate search.
- LTMADS a non-deterministic implementation of MADS. Union of normalized polling directions grows dense in the unit sphere with probability one.
- ORTHOMADS a deterministic implementation of MADS with orthogonal polling directions. Union of normalized polling directions grows dense in the unit sphere.



Three instantiations of mesh adaptive direct searches

- GPS with coordinate search.
- LTMADS a non-deterministic implementation of MADS. Union of normalized polling directions grows dense in the unit sphere with probability one.
- ORTHOMADS a deterministic implementation of MADS with orthogonal polling directions. Union of normalized polling directions grows dense in the unit sphere.



Convergence analysis of MADS

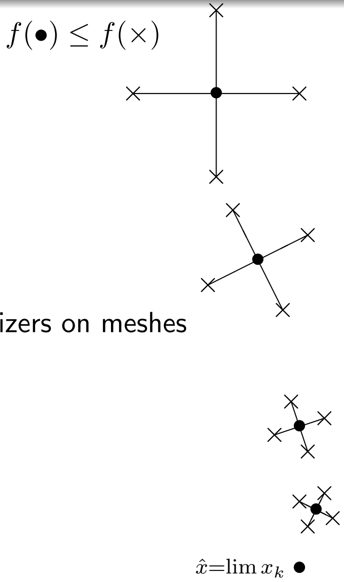
Assumptions

- At least one initial point in X is provided
  – but not required to be in Ω.
- All iterates belong to some compact set
  – it is sufficient to assume that level sets of f in X are bounded.

Key to the analysis

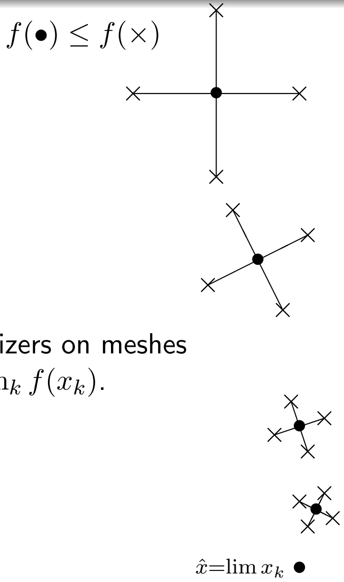
- These assumptions ensure that there is a convergent subsequence of poll centers on meshes that get infinitely fine.
- The analysis is divided in two: the limit of feasible poll centers, and the limit of infeasible poll centers.

- If nothing is known about f.



⇒ Then $\hat{x}$ is the limit of mesh local optimizers on meshes that get infinitely fine,

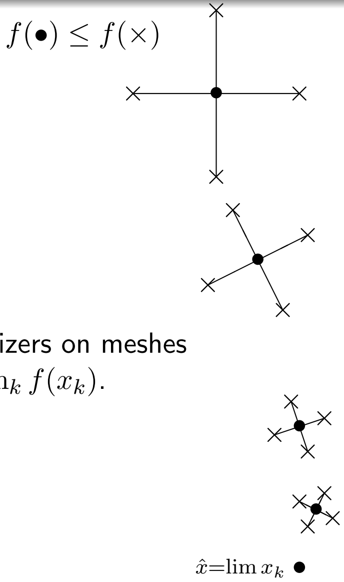
- If nothing is known about f.
- If f is lower semi-continuous near $\hat{x}$,



$\Rightarrow$ Then $\hat{x}$ is the limit of mesh local optimizers on meshes that get infinitely fine, and $f(\hat{x}) \leq \lim_k f(x_k)$.

Hierarchical convergence analysis - feasible iterates

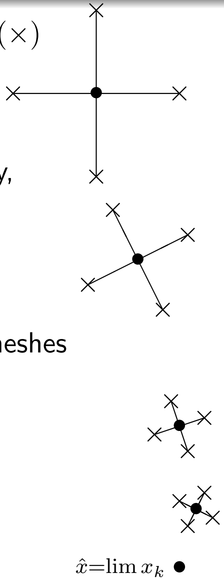
- If nothing is known about f.
- If f is lower semi-continuous near $\hat{x}$,
- and if f is Lipschitz near $\hat{x}$,



$\Rightarrow$ Then $\hat{x}$ is the limit of mesh local optimizers on meshes that get infinitely fine, and $f(\hat{x}) \leq \lim_k f(x_k)$.

$\Rightarrow$ Then $f^\circ(\hat{x}; v) \geq 0$ for all $v \in T_\Omega^H(\hat{x})$.

- If nothing is known about f.
- If f is lower semi-continuous near $\hat{x}$,
- and if f is Lipschitz near $\hat{x}$,
- and if the hypertangent cone $T_\Omega^H(\hat{x})$ is non-empty,

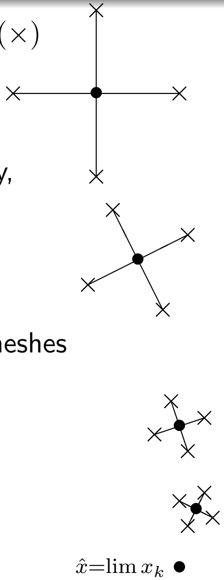


$\Rightarrow$ Then $\hat{x}$ is the limit of mesh local optimizers on meshes that get infinitely fine, and $f(\hat{x}) \le \lim_k f(x_k)$.

$\Rightarrow$ Then $f^\circ(\hat{x}; v) \ge 0$ for all $v \in T_\Omega^H(\hat{x})$.

$\Rightarrow$ Then $f^\circ(\hat{x}; v) \ge 0$ for all $v \in T_\Omega^{Cl}(\hat{x})$.

- If nothing is known about f.
- If f is lower semi-continuous near $\hat{x}$,
- and if f is Lipschitz near $\hat{x}$,
- and if the hypertangent cone $T^H_\Omega(\hat{x})$ is non-empty,
- and if f is regular near $\hat{x}$,



$\Rightarrow$ Then $\hat{x}$ is the limit of mesh local optimizers on meshes that get infinitely fine, and $f(\hat{x}) \le \lim_k f(x_k)$.

$\Rightarrow$ Then $f^\circ(\hat{x}; v) \ge 0$ for all $v \in T^H_\Omega(\hat{x})$.

$\Rightarrow$ Then $f^\circ(\hat{x}; v) \ge 0$ for all $v \in T^{Cl}_\Omega(\hat{x})$.

$\Rightarrow$ Then $f'(\hat{x}; v) \ge 0$ for all $v \in T^{Cl}_\Omega(\hat{x})$.

$f(\bullet) \leq f(\times)$

- If nothing is known about f.
- If f is lower semi-continuous near $\hat{x}$,
- and if f is Lipschitz near $\hat{x}$,
- and if the hypertangent cone $T_\Omega^H(\hat{x})$ is non-empty,
- and if f is regular near $\hat{x}$,
- and if f is strictly differentiable near $\hat{x}$,

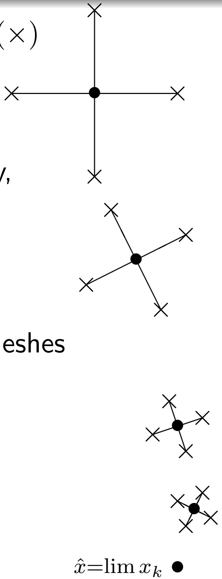
$\Rightarrow$ Then $\hat{x}$ is the limit of mesh local optimizers on meshes that get infinitely fine, and $f(\hat{x}) \leq \lim_k f(x_k)$.

$\Rightarrow$ Then $f^\circ(\hat{x}; v) \geq 0$ for all $v \in T_\Omega^H(\hat{x})$.

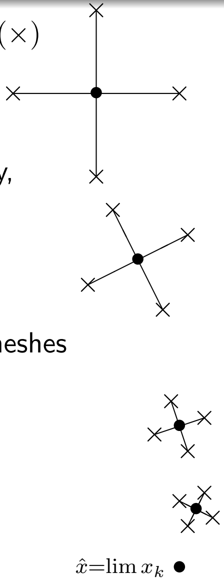
$\Rightarrow$ Then $f^\circ(\hat{x}; v) \geq 0$ for all $v \in T_\Omega^{Cl}(\hat{x})$.

$\Rightarrow$ Then $f'(\hat{x}; v) \geq 0$ for all $v \in T_\Omega^{Cl}(\hat{x})$.

$\Rightarrow$ Then $\nabla f(\hat{x})^T v \geq 0$ for all $v \in T_\Omega^{Cl}(\hat{x})$.



$f(\bullet) \leq f(\times)$



- If nothing is known about f.
- If f is lower semi-continuous near $\hat{x}$,
- and if f is Lipschitz near $\hat{x}$,
- and if the hypertangent cone $T_\Omega^H(\hat{x})$ is non-empty,
- and if f is regular near $\hat{x}$,
- and if f is strictly differentiable near $\hat{x}$,
- and if Ω is regular at $\hat{x}$.

$\Rightarrow$ Then $\hat{x}$ is the limit of mesh local optimizers on meshes that get infinitely fine, and $f(\hat{x}) \leq \lim_k f(x_k)$.

$\Rightarrow$ Then $f^\circ(\hat{x}; v) \geq 0$ for all $v \in T_\Omega^H(\hat{x})$.

$\Rightarrow$ Then $f^\circ(\hat{x}; v) \geq 0$ for all $v \in T_\Omega^{Cl}(\hat{x})$.

$\Rightarrow$ Then $f'(\hat{x}; v) \geq 0$ for all $v \in T_\Omega^{Cl}(\hat{x})$.

$\Rightarrow$ Then $\nabla f(\hat{x})^T v \geq 0$ for all $v \in T_\Omega^{Cl}(\hat{x})$.

$\Rightarrow$ Then $\nabla f(\hat{x})^T v \geq 0$ for all $v \in T_\Omega^{Co}(\hat{x})$: i.e., $\hat{x}$ is a KKT point.

Hierarchical convergence analysis - infeasible iterates

A similar hierarchical analysis holds for

$$\min_{x \in X} \; h(x)$$

for the infeasible iterates.

Hierarchical convergence analysis - infeasible iterates

A similar hierarchical analysis holds for

$$\min_{x \in X} \; h(x)$$

for the infeasible iterates.

For the case where $\hat{x} \in \Omega$, to analyse

$$\min_{x \in \Omega} \; f(x),$$

we need the constraint qualification:

> Suppose that for every $v \in T^H_\Omega(\hat{x}) \neq \emptyset$, there exists an $\epsilon > 0$ for which $h^\circ(x; v) < 0$ for all $x \in X \cap B_\epsilon(\hat{x})$ such that $h(x) > 0$.

Presentation outline

1. Handling constraints in real problems
   - Three types of constraints
   - Strategies to deal with constraints
   - Three instantiations of mesh adaptive direct searches
   - Hierarchical convergence analysis

2. Numerical results on engineering problems
   - Three real test problems
   - A feasible starting point
   - An infeasible starting point
   - Multiple runs

3. Discussion

Three engineering test problems

- Styrene production simulation JOGO 2008

  - Maximize the net present value while satisfying industrial and environmental regulations.
  - Written by a chemical engineer.
  - Uses some common methods such as Runge-Kutta, Newton, fixed points, secant, bisection, and many other chemical engineering related solvers.
  - 8 bound constrained variables,
    4 boolean unrelaxable constraints,
    7 relaxable constraints.
  - 14% of trial points violate a hidden constraint.
  - A surrogate is obtained by using greater tolerances and smaller maximum number of iterations in the numerical methods.

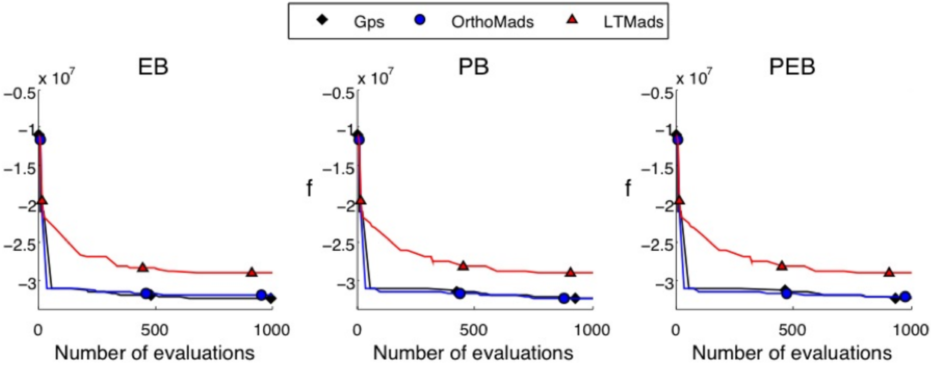
Three engineering test problems

- Styrene production simulation JOGO 2008
- Multidisciplinary design optimization AIAA/ISSMO 2004
  - Mechanical engineering literature.
  - Three coupled disciplines to maximize the aircraft range – structure – aerodynamics – propulsion.
  - Simplified aircraft model, with 10 bound constrained variables under 10 relaxable constraints.
  - Fixed point iterations through the different disciplines.
  - Surrogate consists in stopping the simulation at a larger relative error and a smaller limit on the number of fixed point iterations.

Three engineering test problems

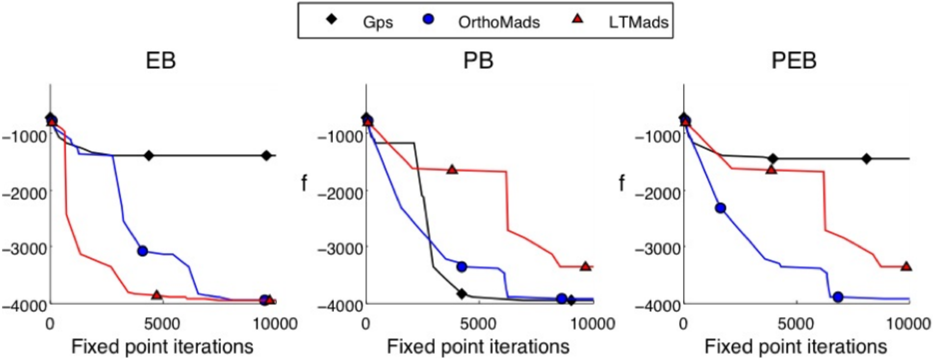
- Styrene production simulation JOGO 2008
- Multidisciplinary design optimization AIAA/ISSMO 2004
- Well positioning community problem Adv. Water Resources 2008
  - Fowler, Kelley and 13 others.
  - Minimize the cost to prevent an initial contaminant plume from spreading by using wells to control the direction and extent of advective fluid flow.
  - Requires running a Fortran solver to simulate groundwater flow.
  - Six wells and nonlinear head constraints.
  - Replace a linear constraint by an equality to eliminate the pumping rate of the sixth well as an explicit variable.
  - 17 bound constrained variables: locations and pumping rates
    12 relaxable non-linear constraints on the allowable head.

Styrene problem from a feasible starting point



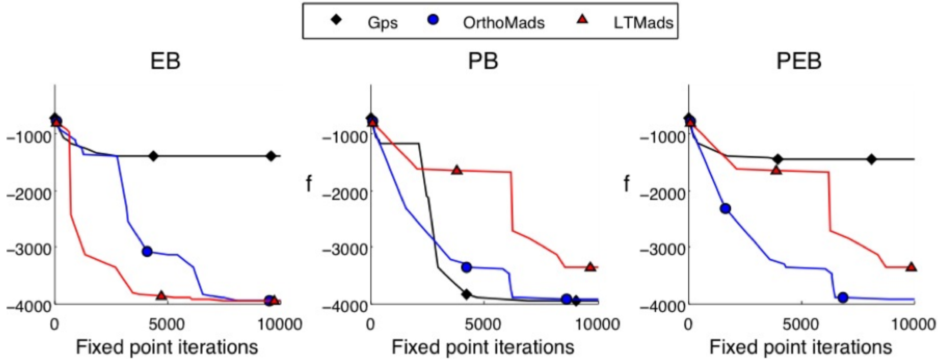
- GPS and ORTHOMADS perform better than LTMADS.
- Treatment of constraints has no significant effect because initial point is feasible.

MDO problem from a feasible starting point



Remark: The horizontal axis is the number of fixed points iterations of the truth and surrogate. 10000 corresponds to about 650 evaluations of f.

MDO problem from a feasible starting point

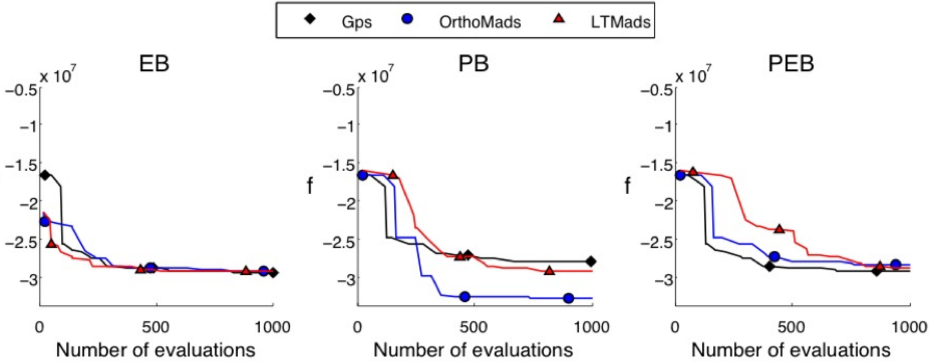


- OrthoMads performs well in all 3 cases.
- Gps gets stuck at a local solution.
- PB allows all three algorithms to escape the local solution at $f \approx -1500$.

An infeasible starting point

This is where things get interesting...

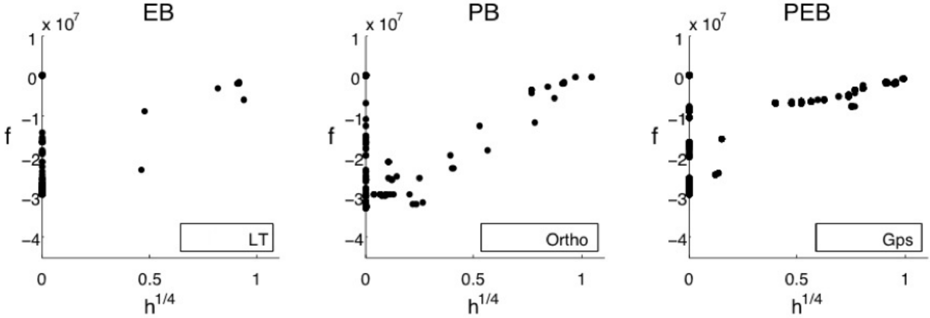
Styrene problem from an infeasible starting point



- Feasibility is reached rapidly.
- Only OrthoMads PB escapes from a local solution.

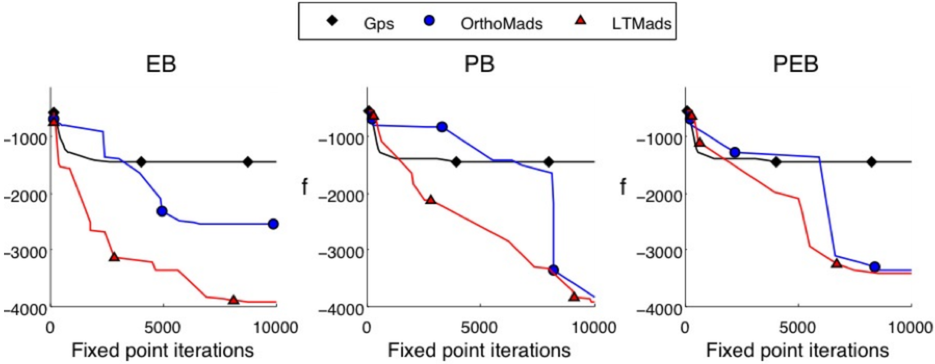
Styrene problem from an infeasible starting point

Plots of the objective function value versus the constraint violation.



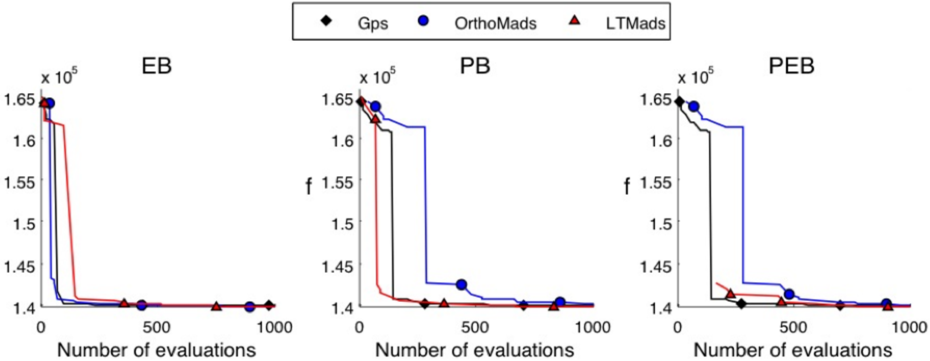
- Feasible solutions are where $h = 0$.
- PB finds a way to move across the infeasible region to a better solution.
- PEB moves across the infeasible region, but switches to EB.

MDO problem from an infeasible starting point



- GPS gets stuck at a local solution with the three approaches.
- PB allows the MADS instances to approach the best known solution.

WELL problem from an infeasible starting point



- It took a long time for LTMADS-PEB to reach feasibility, but it did at a very good solution.
- All approaches reach the same solution.

Multiple runs

Problem • Method	EB worst	EB best (out of 60 runs)	PB worst	PB best (out of 90 runs)	PEB worst	PEB best (out of 90 runs)
Styrene	$\times 10^7$					
• LT	-2.89	-3.31	-2.60	-3.36	-2.60	-3.35
• ORTHO	-2.88	-3.31	-2.64	-3.32	-2.64	-3.32
MDO						
• LT	∅	-3964.1	∅	-3963.6	∅	-3962.9
• ORTHO	∅	-3964.0	∅	-3963.6	∅	-3964.1
Well	$\times 10^5$					
• LT	1.402	1.399	1.403	1.399	1.403	1.399
• ORTHO	1.602	1.399	1.602	1.399	1.602	1.399

- ∅ indicates that no feasible solution was found.
- Little difference in the best solutions (though there is some).
  - ORTHOMADS found a better solution than LTMADS only once.
  - LTMADS found a better solution than ORTHOMADS 3 times.
- Strategies are comparable in a worst case scenario.

Discussion

- We have tested
  - three algorithms GPS, LTMADS and ORTHOMADS,
  - using three strategies to handle the constraints EB, PB and PEB,
  - on three real test problems Styrene, MDO and Well.

Discussion

- We have tested
  - three algorithms GPS, LTMADS and ORTHOMADS,
  - using three strategies to handle the constraints EB, PB and PEB,
  - on three real test problems Styrene, MDO and Well.
- The main differences show up with an infeasible initial point.
  - The progressive barrier gives the best results, as it moves across the infeasible region, while trying to retain good values of f.

Discussion

- We have tested
  - three algorithms GPS, LTMADS and ORTHOMADS,
  - using three strategies to handle the constraints EB, PB and PEB,
  - on three real test problems Styrene, MDO and Well.
- The main differences show up with an infeasible initial point.
  - The progressive barrier gives the best results, as it moves across the infeasible region, while trying to retain good values of f.
- For a single run, ORTHOMADS gave the best results.
  It is less sensitive to randomness than LTMADS.
- In a multi-start framework, this sensitivity turns into an advantage for LTMADS (however, for these types of problems, we cannot usually afford multi-starts).
- `www.gerad.ca/nomad`