Handling of constraints

MTH8418

S. Le Digabel, Polytechnique Montréal

Winter 2020

(v2)

Plan

Taxonomy of constraints

Approaches

The Progressive Barrier (PB)

References

Taxonomy of constraints

Approaches

The Progressive Barrier (PB)

References

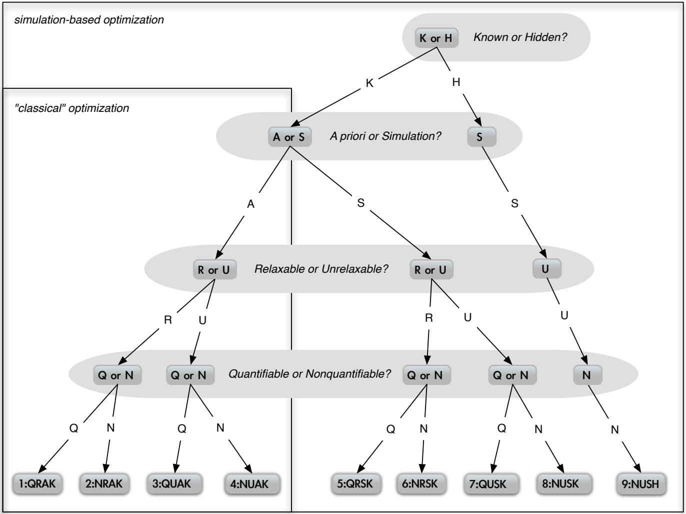
The feasible set Ω

- Optimization problem: $\min_{x \in \Omega} f(x)$
- $\Omega = \left\{ x \in \mathcal{X} \subseteq \mathbb{R}^n : c_j(x) \leq 0, j \in J \text{ and } g_j(x) = 0, j \in J_e \right\}$
- $J = \{1, 2, \dots, m\}$, $J_e = \{1, 2, \dots, m_e\}$
- n variables, m general inequalities, m_e general equalities
- Typically but not always:
  - Functions f, c_j, and g_i, for $j \in J$ and $i \in J_e$, are given by a blackbox
  - $\mathcal{X}$ contains a priori constraints (term explained later on)

Equalities

- No yet proven method can deal with general equalities, except maybe the augmented Lagrangian approach (see Slide 18)
- In practice, such an equality $g(x) = 0$ may be replaced with two inequalities $-\varepsilon \leq g(x) \leq \varepsilon$. The more ε is small, the more difficult it is to satisfy the constraint
- Treatments for specific kinds of equalities exist. An example for the linear case is described in this lesson
- From now on, we suppose $J_e = \emptyset$

Taxonomy of constraints (QRAK)



[Le Digabel and Wild, 2015]

Taxonomy of constraints: Definitions

- The taxonomy allows the classification of each constraint for a given instance, or formulation or the problem
- Other classifications are possible (linear vs nonlinear, equality vs inequality, etc.)
- Consolidates many previous terms: *Soft, virtual, hard, hidden, difficult, easy, constraints with available derivatives, derivative-free, open, closed, implicit, etc.*
- 9 main classes of constraints with general difficulty growing from left to right:

1:QRAK 2:NRAK 3:QUAK 4:NUAK | 5:QRSK 6:NRSK 7:QUSK 8:NUSK 9:NUSH = Hidden

Quantifiable (Q) versus Nonquantifiable (N)

A quantifiable constraint is a constraint for which the degree of feasibility and/or violation can be quantified. A nonquantifiable constraint is one for which the degrees of satisfying or violating the constraint are both unavailable.

- N constraint: Binary indicator for constraint being satisfied
- Q constraint: No guarantee that measures of both feasibility and violation are available
- A constraint for which both the degrees of feasibility and violation are available can be referred to as fully quantifiable

$\mathtt{Q}$ versus $\mathtt{N}$: Examples

- Quantifiable feasibility: *"Max time for simulation to complete is 10s."* Code automatically terminated if it takes more than 10 seconds
- Quantifiable violation: *"Time-stepping simulation should run to completion time T."* If the simulation stops at time $t < T$, then $T - t$ measures how close one was to satisfying the constraint
- Models: $\mathtt{Q}$ implies interpolation while $\mathtt{N}$ implies classification

Relaxable ($\mathtt{R}$) versus Unrelaxable ($\mathtt{U}$)

A relaxable constraint is a constraint that does not need to be satisfied in order to obtain meaningful outputs from the simulation(s) in order to compute the objective and the constraints. An unrelaxable constraint is one which must be satisfied for meaningful outputs to be obtained

- Meaningful means that the values are coherent with the logic of the model, can be trusted as valid by an optimization algorithm, and rightly interpreted when observed in a solution
- $\mathtt{R}$ constraint: Generally not part of a physical model (budget)
- Optimization:
  - All iterates must satisfy $\mathtt{U}$ constraints
  - $\mathtt{R}$ constraints need to be satisfied only at the proposed solution

A Priori (A) versus Simulation-Based (S)

An a priori constraint is a constraint for which feasibility can be confirmed without running a simulation. A simulation-based constraint requires running a simulation to verify feasibility

- A constraints:
  - Bounds, linear equations, analytic expressions, etc.
  - Solver should evaluate *UA* (unrelaxable, a priori) constraints first to avoid a simulation execution for infeasible points
  - Less clear for *RA* (relaxable, a priori) constraints
- S constraints:
  - Constraints expressed as a function of some simulator output
  - Typically implies a costly run, but simulation may include a constraint that is cheap to evaluate and can be used as a flag to avoid any further computation (*USK)

Known (K) versus Hidden (H)

A known constraint is a constraint that is explicitly given in the problem formulation. A hidden constraint is not known to the solver

- Do not find out about until violated
- Must be satisfied to obtain full set of meaningful outputs from the simulation
- Not necessarily a bug in the simulator
- Implicit assumption inside simulator, unstated to optimizer
- Flag indicating if the simulation failed: If not given in the considered instance, or raised by the solver, it is a H. Otherwise it is a NUSK

Taxonomy of constraints

Approaches

The Progressive Barrier (PB)

References

Handling of constraints specific to the problem

- For a specific problem, one can come up with some heuristic or reparation procedure that always generates feasible solutions from infeasible ones
- Example: Spiral walk for finding feasible coordinates on a map

Linear equalities

- [Audet et al., 2015]
- J_e is not empty anymore, but just for the present slide
- g_j functions are linear for all $j \in J_e$ and the problem can be re-written as

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
$$\text{subject to} \quad \begin{cases} c(x) & \leq & 0 \\ Ax & = & 0 \\ \ell \leq & x & \leq u \end{cases}$$

- Bounds and linear constraints are `Q*AK`. They should be treated by the solver
- Idea: Work in a different subspace by finding solutions implicitly satisfying $Ax = 0$. The equalities are removed and the dimension is reduced

Extreme Barrier (EB) approach

- Minimize f_Ω instead of f, with $f_\Omega(x) = \begin{cases} f(x) \text{ if } x \in \Omega \\ \infty \text{ otherwise} \end{cases}$

- Works with any unconstrained algorithm

- Not efficient because it does not exploit the knowledge of relaxable constraints

Penalty methods

For $x \in \mathcal{X}$ with $\mathcal{X}$ containing `**AK` constraints, and for all constraints $c_j \leq 0$ that are `QRSK`, $j \in J$:

- Solve $\min_{x \in \mathcal{X}} f(x) + \rho \sum_{j \in J} \max\{0, c_j(x)\}^2$
- $\rho > 0$: Fixed penalty parameter
- We can consider a sequence of such problems with growing penalty parameters ρ^k
- How to choose/update the penalty values? Very unstable numerically

Augmented Lagrangian (1/2)

For $x \in \mathcal{X}$ with $\mathcal{X}$ containing **AK constraints, and for all constraints $c_j \leq 0$ that are QRSK, $j \in J$:

- Lagrangian: $\mathcal{L}(x;\lambda) = f(x) + \lambda^\top c(x)$
- $\lambda \in \mathbb{R}^m_+$: Lagrange multipliers
- Nonlinear constrained optimization $\simeq \nabla \mathcal{L} = 0$
- Augmented Lagrangian (`auglag`):

$$\mathcal{L}_A(x;\lambda,\rho) = f(x) + \lambda^\top c(x) + \frac{1}{2\rho} \sum_{j \in J} \max\left\{0, c_j(x)\right\}^2$$

- $\rho > 0$: Penalty parameter

Note: Equalities can be considered

Augmented Lagrangian (2/2)

- Auglag-based methods: The original problem is transformed into a sequence of problems where only the constraints in $\mathcal{X}$ are imposed
- Given the current values for ρ^{k-1} and the approximate Lagrange multipliers λ^{k-1}, approximately solve the subproblem: $\min_{x \in \mathcal{X}} \left\{ \mathcal{L}_A(x; \lambda^{k-1}, \rho^{k-1}) \right\}$
- Set $\lambda_j^k = \max \left\{ 0, \lambda_j^{k-1} + \frac{1}{\rho^{k-1}} c_j(x^k) \right\}$ for all $j \in J$
- If $c_j(x^k) \leq 0$ for all $j \in J$, set $\rho^k = \rho^{k-1}$;
  Otherwise, set $\rho^k = \frac{1}{2} \rho^{k-1}$

Note: Auglag with MADS in [Gramacy et al., 2016, Picheny et al., 2016]

Taxonomy of constraints

Approaches

The Progressive Barrier (PB)

References

Progressive Barrier (PB) approach

- [Audet and Dennis, Jr., 2009]
- Suppose that the constraints c_j, $j \in J$ are `QR*K` (quantifiable and relaxable)
- Constraint violation function $h : \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$:

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

- Another norm than ℓ_2 can be used
- $x \in \Omega \Leftrightarrow h(x) = 0$ and $x \notin \Omega \Leftrightarrow h(x) > 0$
- At iteration k, points with $h(x) > h_k^{\max}$ are rejected by the algorithm, and $h_k^{\max} \to 0$ as $k \to \infty$

Note: Application of the PB to DFTR in [Audet et al., 2018]

PB with MADS: Filter

- Filter $=$ Set of non-dominated solutions
- Principle: x dominates y, or $x \prec y$, if $f(x) \leq f(y)$ and $h(x) \leq h(y)$ with at least one strict inequality
- Feasible incumbent: Feasible point with the best f value
- Infeasible incumbent: Non-dominated infeasible point with the best f value and with its h value below $h_k^{\max}$

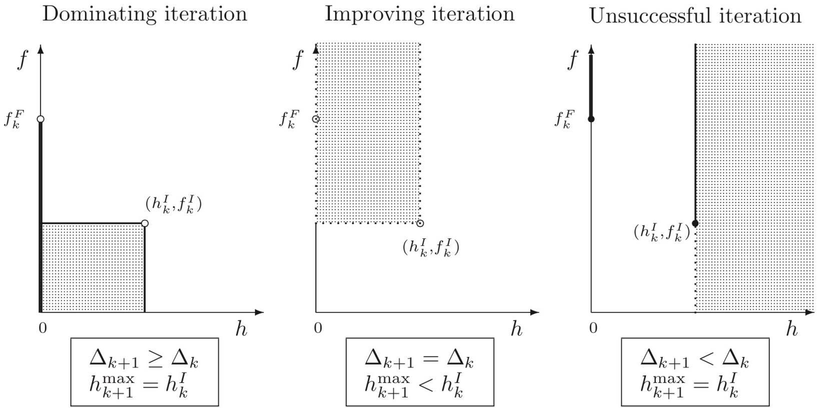
PB with MADS: Reduced poll

- At the beginning of the iteration, the PB defines primary and secondary poll centers
- The primary poll center is taken as the feasible incumbent if $f(\text{feas. inc.}) - \rho \leq f(\text{inf. inc.})$. Otherwise it is the infeasible incumbent
- $\rho > 0$ is the frame center trigger
- A regular poll is performed around the primary poll center
- A reduced poll is performed around the secondary poll center, with a reduced number of directions
- Typically: Regular poll with $n + 1$ directions and reduced poll with two directions

PB with MADS: Success types

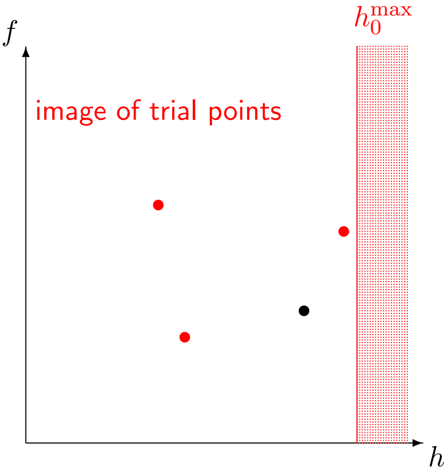
- At the end of each iteration, the PB provides a mechanism to define the type of success of the iteration:
  - Full success, or dominating iteration: New feasible incumbent or new infeasible incumbent with a better h value.
  - Partial success or improving iteration: New non-dominated infeasible point with a better h value.
  - Failure or unsuccessful iteration: Other situations.

PB with MADS: Illustration of the success types

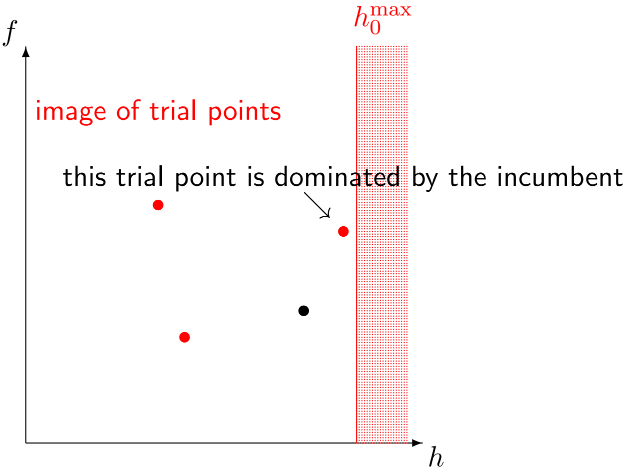


Taken from [Audet and Dennis, Jr., 2009]. Δ_k corresponds to the mesh size parameter Δ_k^m

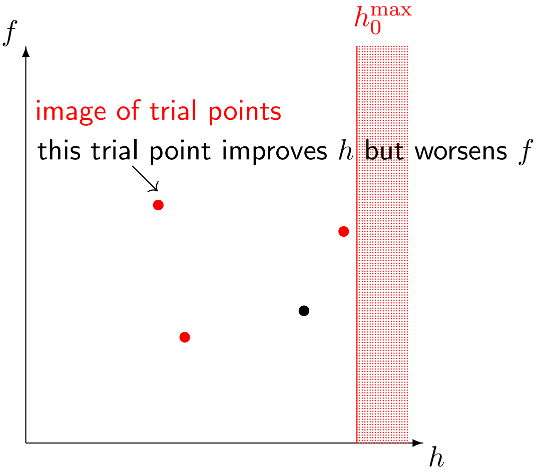
PB illustration



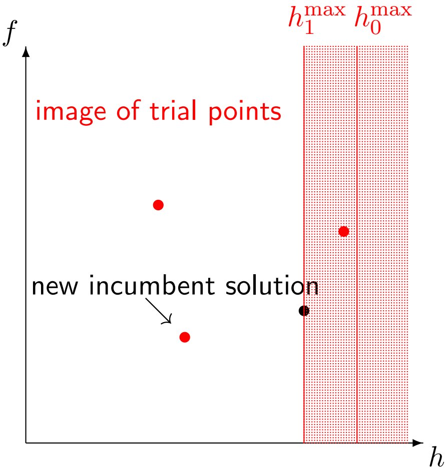
PB illustration



PB illustration



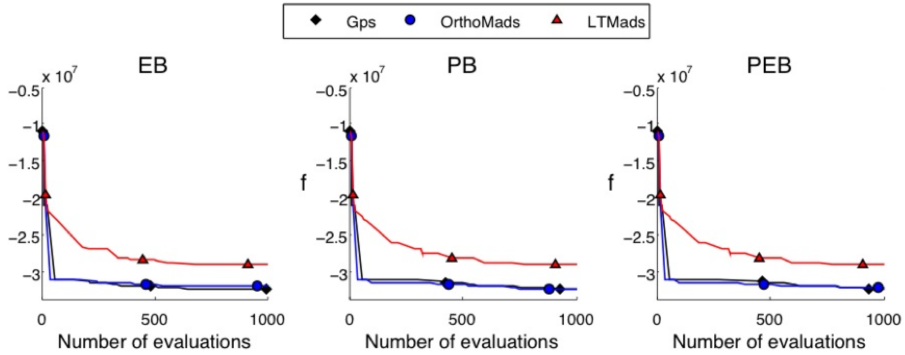
PB illustration



PEB approach

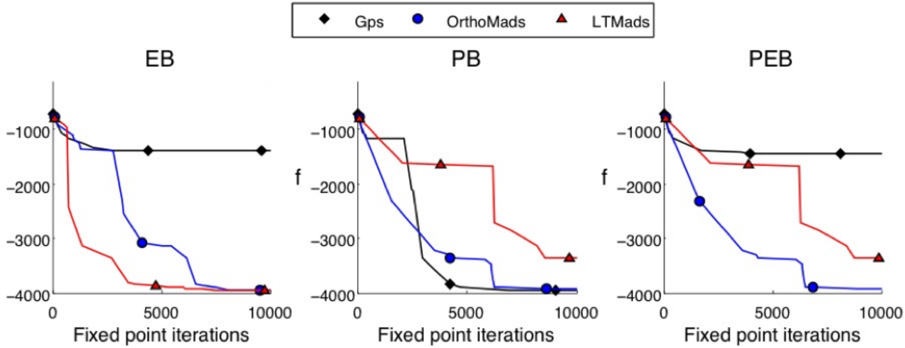
- [Audet et al., 2012]
- Progressive to Extreme Barrier
- Hybrid method between the EB and the PB
- Initially treats a `QR*K` constraint with the progressive barrier
- Then, if polling around the infeasible poll center generates a new infeasible incumbent that satisfies a constraint violated by the poll center, then that constraint moves from being treated by the progressive barrier to the extreme barrier

STYRENE problem from a feasible starting point



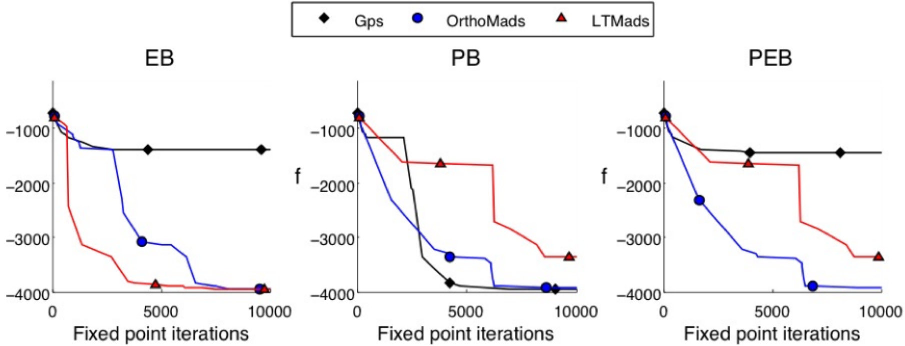
- GPS and OrthoMADS perform better than LT-MADS
- Treatment of constraints has no significant effect because initial point is feasible

MDO problem from a feasible starting point



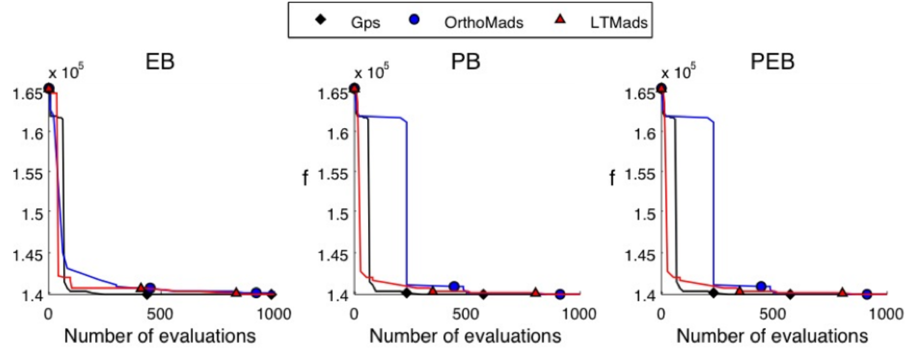
Remark: The horizontal axis is the number of fixed points iterations of the truth and surrogate. 10,000 corresponds to about 650 evaluations of f

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 $f \simeq -1500$

WELL problem from a feasible starting point

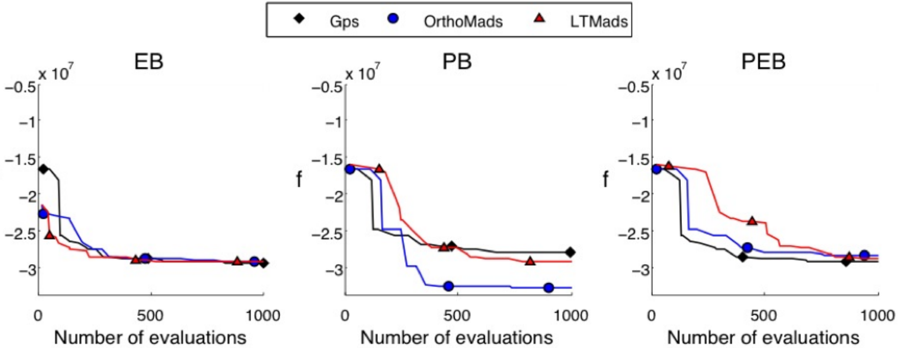


- OrthoMADS PB and PEB explore infeasible domain

An infeasible starting point

This is where things get interesting...

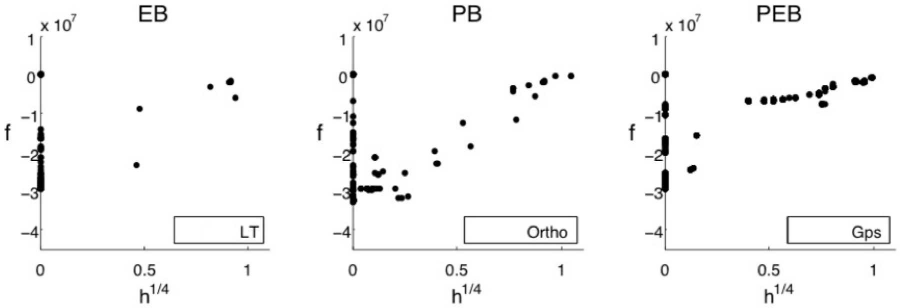
STYRENE problem from an infeasible starting point



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

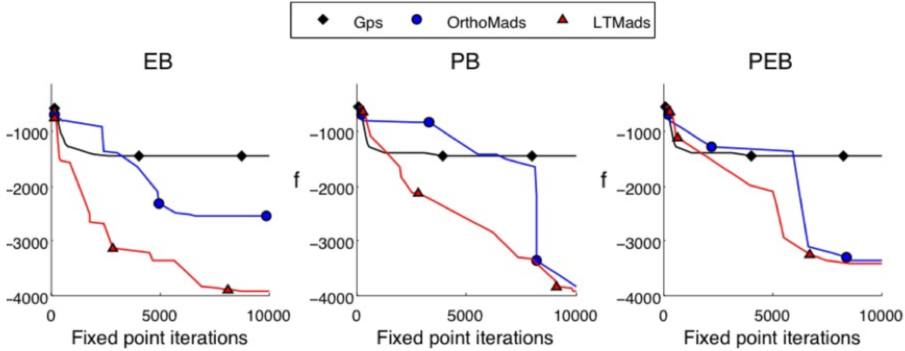
STYRENE problem from an infeasible starting point

Plots of the objective function value versus the constraint violation.



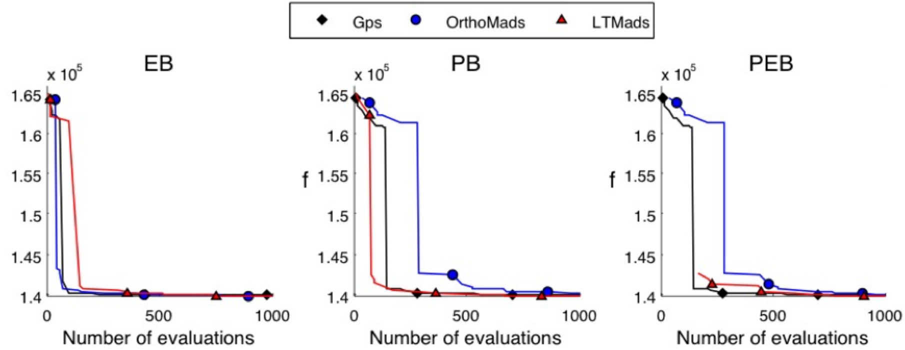
- Feasible solutions are where $h = 0$
- PB finds a way to move across the infeasible region to a better solution
- PEB moves across the infeasible region, but switches to EB

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

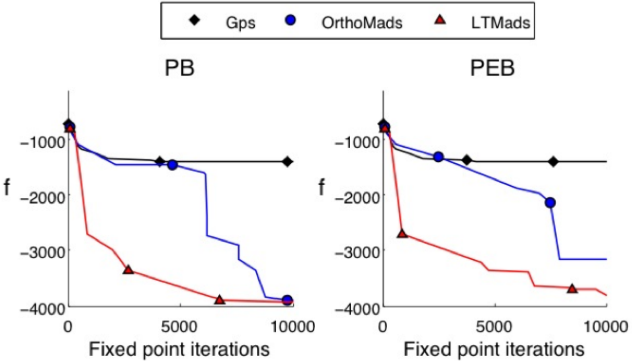
WELL problem from an infeasible starting point



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

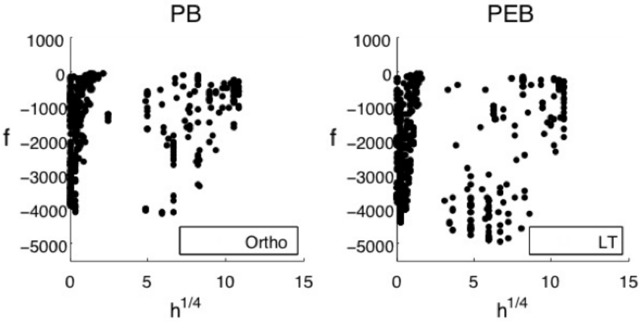
MDO from a feasible and an infeasible starting point

EB is redundant with the run from only a feasible point.



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

MDO from a feasible and an infeasible starting point



- In the plots, the density of infeasible points (with $h > 0$ small) is greater with PB than PEB
- PEB spends a lot of energy at infeasible points near ($h^{1/4} = 5$) but cannot escape that local min of h

Multiple runs

Problem • Method	EB worst	EB median (out of 60 runs)	EB best	PB worst	PB median (out of 90 runs)	PB best	PEB worst	PEB median (out of 90 runs)	PEB best
STYRENE	$\times 10^7$								
• LT	-2.89	-2.93	-3.31	-2.60	-3.25	-3.36	-2.60	-3.25	-3.35
• Ortho	-2.88	-2.93	-3.31	-2.64	-3.13	-3.32	-2.64	-3.08	-3.32
MDO									
• LT	$\emptyset$	-1754.7	-3964.1	$\emptyset$	-3871.7	-3963.6	$\emptyset$	-3664.6	-3962.9
• Ortho	$\emptyset$	-1385.9	-3964.0	$\emptyset$	-3929.6	-3963.6	$\emptyset$	-3890.9	-3964.1
WELL	$\times 10^5$								
• LT	1.402	1.399	1.399	1.403	1.399	1.399	1.403	1.399	1.399
• Ortho	1.602	1.399	1.399	1.602	1.399	1.399	1.602	1.399	1.399

- Little difference in the best solutions (though there is some)
  - OrthoMADS found a better solution than LT-MADS only once
  - LT-MADS found a better solution than OrthoMADS 3 times
- OrthoMADS is preferable to LT-MADS in a worst case scenario

Taxonomy of constraints

Approaches

The Progressive Barrier (PB)

References

References I

Alarie, S., Audet, C., Garnier, V., Le Digabel, S., and Leclaire, L.-A. (2013).
Snow water equivalent estimation using blackbox optimization.
Pacific Journal of Optimization, 9(1):1–21.
(Spiral walk).

Audet, C., Conn, A., Le Digabel, S., and Peyrega, M. (2018).
A progressive barrier derivative-free trust-region algorithm for constrained optimization.
Computational Optimization and Applications, 71(2):307–329.

Audet, C. and Dennis, Jr., J. (2009).
A Progressive Barrier for Derivative-Free Nonlinear Programming.
SIAM Journal on Optimization, 20(1):445–472.

Audet, C., Dennis, Jr., J., and Le Digabel, S. (2012).
Trade-off studies in blackbox optimization.
Optimization Methods and Software, 27(4–5):613–624.
(PEB).

Audet, C., Le Digabel, S., and Peyrega, M. (2015).
Linear equalities in blackbox optimization.
Computational Optimization and Applications, 61(1):1–23.

Gramacy, R., Gray, G., Le Digabel, S., Lee, H., Ranjan, P., Wells, G., and Wild, S. (2016).
Modeling an Augmented Lagrangian for Blackbox Constrained Optimization.
Technometrics, 58(1):1–11.

References II

Le Digabel, S. and Wild, S. (2015).
A Taxonomy of Constraints in Simulation-Based Optimization.
Technical Report G-2015-57, Les cahiers du GERAD.

Nocedal, J. and Wright, S. (2006).
Numerical Optimization.
Springer Series in Operations Research and Financial Engineering. Springer, Berlin, second edition. (Augmented Lagrangian).

Picheny, V., Gramacy, R., Wild, S., and Le Digabel, S. (2016).
Bayesian optimization under mixed constraints with a slack-variable augmented Lagrangian.
In Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., and Garnett, R., editors, *Advances in Neural Information Processing Systems (NIPS) 29*, pages 1435–1443. Curran Associates, Inc.