Multiobjective Optimization

MTH8418

S. Le Digabel, Polytechnique Montréal

Winter 2020

(v2)

Plan

Multiobjective optimization problem

- The multiobjective optimization problem (MOP) can be formally stated as

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

- where

$$F : \Omega \to \{\mathbb{R} \cup \{+\infty\}\}^p$$

  and

$$F(x) = \left(f^{(1)}(x), f^{(2)}(x), \ldots, f^{(p)}(x)\right)$$

- p is the number of objective functions
- Case $p = 2$: Biobjective optimization problem (BOP)
- The feasible set Ω remains unchanged
- Typically, the different objectives are contradictory: A decrease in one objective causes an increase in the other objectives

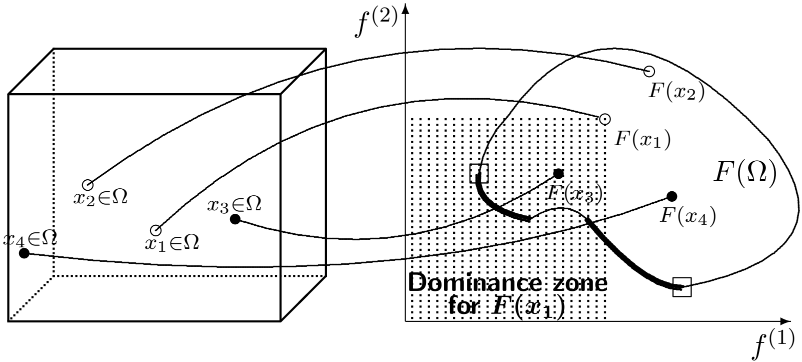
Difficulty

- Single-objective optimization: Particular case where $p = 1$. An optimal solution typically consists of one single vector $x \in \Omega$
- Multiobjective optimization: There is usually no such vector that simultaneously minimizes all of the $p \geq 2$ objective functions
- The solution consists of a set of trade-off solutions in Ω, the *Pareto solutions*
- The methods presented in this lesson construct an approximation to this set

Pareto notion

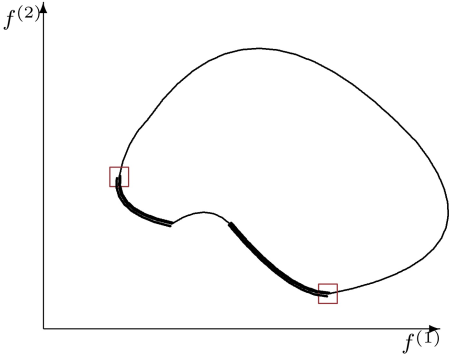
- Single-objective: $u, v \in \Omega$ can be trivially ranked by comparing $f(u)$ and $f(v)$
- Generalization with $p > 1$:
  - u dominates v, denoted $u \prec v$, if and only if $F(u) \leq F(v)$ and $f^{(q)}(u) < f^{(q)}(v)$ for at least one index q in $\{1, 2, \ldots, p\}$
  - u is indifferent to v, denoted $u \sim v$, if and only if u does not dominate v and v does not dominate u
- A point $u \in \Omega$ is *Pareto optimal* if and only if there is no $w \in \Omega$ such that $w \prec u$
- The set of Pareto (optimal) solutions is the *Pareto set* $\Omega_{\mathcal{P}}$
- The image of $\Omega_{\mathcal{P}}$ under the mapping F defines the solution to the problem and is called the *Pareto front* $F_{\mathcal{P}} \subseteq \mathbb{R}^p$

Pareto front example



Feasible region : $\Omega \subset \mathbb{R}^3$ Image of Ω in objective space $\mathbb{R}^2$

Individual minima



The individual minima of F are the solutions to the single-objective optimization problems

$$\min_{x \in \Omega} \quad f^{(q)}(x) \text{ , for } q \in \{1, 2, \ldots p\}$$

How to choose one solution?

- Can be done visually with $p = 2$ and some knowledge of the problem. Large and small slopes should be identified
- For $p \geq 2$, engineers use carpet plots
- More generally, this is the subject of multicriteria optimization

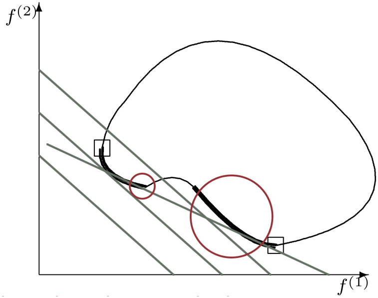
The ε-constraint method

- The most commonly used method
- It transforms objectives into constraints: The original problem with p objectives becomes a problem with one objective and $p-1$ constraints
- Then, change the bounds on the constraints in order to grasp the Pareto front

Weighted sums of objectives for BOP ($p = 2$)

Natural single-objective reformulation: Solve

$$\min_{x \in \Omega} \quad \alpha f^{(1)}(x) + (1-\alpha) f^{(2)}(x) \tag{1}$$

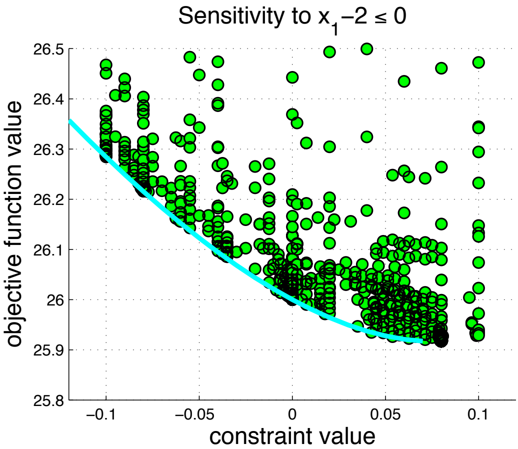


Inconvenient: Some regions of the Pareto front are never optimal for (1), regardless of α

Application: Constraint sensitivity analysis

- Biobjective optimization can be used in order to conduct sensitivity analyses relative to constraints
- The constraint of interest is transformed as an objective function
- The analysis of the approximated Pareto front allows to interpret the impact of this constraint on the original objective
- Two different tools are available within `NOMAD`:
  - A post-optimization analysis. Cheap and rough approximation of the sensitivities
  - An additional biobjective execution. More expensive, but gives a good approximation of the sensitivities

Sensitivity analysis: Example



Metrics

- [Audet et al., 2018]
- How to compare the approximations to the Pareto front obtained by different solvers?
- $\mathcal{S}$: Set of solvers; $\mathcal{P}$: Set of problems
- To draw performance and data profiles, we need a performance measure $t_{p,s} > 0$ for each $p \in \mathcal{P}$ and $s \in \mathcal{S}$
- $F_{p,s}$: Approximated Pareto front determined by the solver $s \in \mathcal{S}$ for problem $p \in \mathcal{P}$
- F_p: Approximated Pareto front for problem p. Obtained by $\cup_{s \in \mathcal{S}} F_{p,s}$ and by removing the dominated points

Purity metric

- Purity metric:

$$purity_{p,s} = \frac{|F_{p,s} \cap F_p|}{|F_{p,s}|} \in [0;1]$$

- The higher the better
- Take $t_{p,s} = 1/purity_{p,s}$ if $purity_{p,s} \neq 0$, $+\infty$ otherwise
- Problem: The purity is equal to one (i.e. perfect) for a solver that gives only one non-dominated solution

Largest hole

- These measures compute the spread of an approximated Pareto front with the maximum size of the “holes” in the front. We need $|F_{p,s}| > 1$
- $t_{p,s} = \Gamma_{p,s} = \max_{q \in \{1,2,\ldots,p\}} \left(\max_{i \in \{1,2,\ldots,|F_{p,s}|\}} \left\{ \delta_i^{(q)} \right\} \right)$ where $\delta_i^{(q)}$ represents the distance between the ith point of $F_{p,s}$ and its closest neighbor, in terms of $f^{(q)}$
- HRS (*Hole Relative Size*): $t_{p,s} = \max_{i \in \{1,2,\ldots,|F_{p,s}|\}} \left\{ d_i / \overline{d} \right\}$ where d_i represents the distance between the ith point of $F_{p,s}$ and its closest neighbor, and $\overline{d} = \sum_{i=1}^{|F_{p,s}|} d_i / |F_{p,s}|$
- Standard deviation: $t_{p,s} = \sqrt{\dfrac{\sum\limits_{i=1}^{|F_{p,s}|} (d_i - \overline{d})^2}{|F_{p,s}| - 1}}$

Progress measures (1/2)

- These measures are focused on the convergence of the methods. Useful for plotting simplified data profiles
- Progress for objective $q \in \{1, 2, \ldots, p\}$ at evaluation k: $prog_k^{(q)} = \log \sqrt{\frac{f_1^{(q)}}{f_k^{(q)}}}$ where $f_k^{(q)}$ represents the best value obtained after the kth evaluation, in terms of $f^{(q)}$
- We need feasible starting solutions, and all objective values need to be > 0
- We could consider $t_{p,s} = \max\limits_{q \in \{1,2,\ldots,p\}} \{prog_k^{(q)}$ for s and $p\}$

Progress measures (2/2)

- Number of non-dominated points at each evaluation: For k the number of evaluation or a group of $n+1$ evaluations, consider $|F_{p,s}|$
- Or consider the number of *new* non-dominated points between two values of k
- Number of waves: Consider all the solutions produced by solver s on problem p. Recursively remove the non-dominated points, and W is the number of times that this operation is necessary to consider all the points. The more W is close to 1, the better is s

Generational Distance (GD)

- Measures a distance between $F_{p,s}$ and F_p
- $GD_{p,s} = \frac{\sqrt{\sum_{i=1}^{|F_{p,s}|} d_{i,p}^2}}{|F_{p,s}|}$
- $d_{i,p}$ represents the distance between the ith point in $F_{p,s}$ and the closest point of F_p
- The standard deviation of the GD measures the deformation of the front obtained by $s \in \mathcal{S}$ compared to the global approximation: $STDGD_{p,s} = \frac{\sum_{i=1}^{|F_{p,s}|}(d_{i,p} - GD_{p,s})^2}{|F_{p,s}|-1}$
- Maximum Pareto Front Error: $ME_{p,s} = \max_{i \in \{1,2,\ldots,|F_{p,s}|\}} d_{i,p}$

Hypersurface for $p = 2$

- Consider $t_{p,s} = HS_{p,s} = \frac{S_{p,s}}{S_p}$
- $S_{p,s}$ represents the surface under the plot of $F_{p,s}$ and S_p the surface under the plot of F_p
- Not easy to generalize/compute for $p > 2$ (hypervolume)

BiMADS: Series of single-optimization executions

- [Audet et al., 2008]
- Based on a single-objective optimization algorithm: **MADS**
- MADS is launched on a series of **subproblems**
- Constraints are handled by MADS with EB/PB/PEB techniques
- Each subproblem is obtained by a single-objective reformulation that is **not based on weights**
- The solutions of each of these subproblems produces a local approximation of the Pareto set
- The set of undominated solutions produces an approximation of the **entire** Pareto set
- BiMADS is implemented in `NOMAD`

- Reference point in the objective space: $r \in \mathbb{R}^2$
- Reformulated objective:

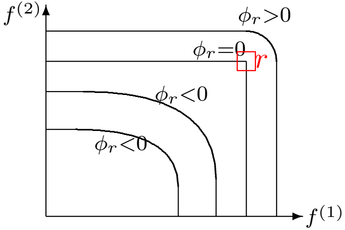
$$\phi_r(F(x)) := \begin{cases} -\prod_{q=1}^{p}(r_q - f^{(q)}(x))^2 & \text{if } F(x) \leq r \\ \sum_{q=1}^{p}\left((f^{(q)}(x) - r_q)_+\right)^2 & \text{otherwise} \end{cases}$$

- When minimized on $x \in \Omega$, starting from x_r (with $F(x_r) = r$), it potentially gives a solution that dominates r

- Reference point in the objective space: $r \in \mathbb{R}^2$
- Reformulated objective:

$$\phi_r(F(x)) := \begin{cases} -\prod_{q=1}^{p} (r_q - f^{(q)}(x))^2 & \text{if } F(x) \le r \\ \sum_{q=1}^{p} \left((f^{(q)}(x) - r_q)_+\right)^2 & \text{otherwise} \end{cases}$$

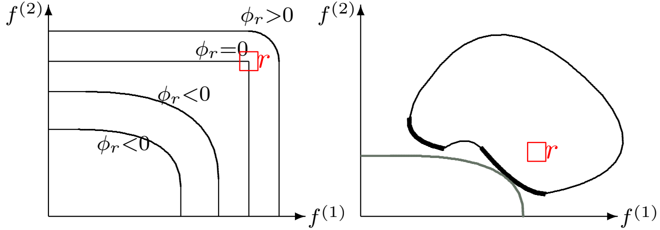
- When minimized on $x \in \Omega$, starting from x_r (with $F(x_r) = r$), it potentially gives a solution that dominates r



- Reference point in the objective space: $r \in \mathbb{R}^2$
- Reformulated objective:

$$\phi_r(F(x)) := \begin{cases} -\prod_{q=1}^{p} (r_q - f^{(q)}(x))^2 & \text{if } F(x) \leq r \\ \sum_{q=1}^{p} \left((f^{(q)}(x) - r_q)_+ \right)^2 & \text{otherwise} \end{cases}$$

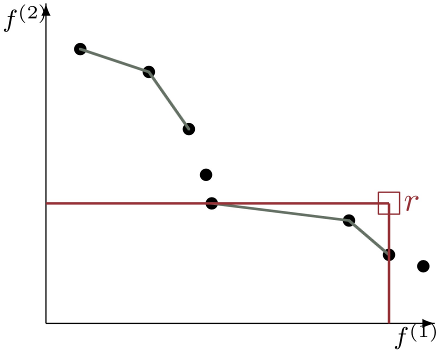
- When minimized on $x \in \Omega$, starting from x_r (with $F(x_r) = r$), it potentially gives a solution that dominates r



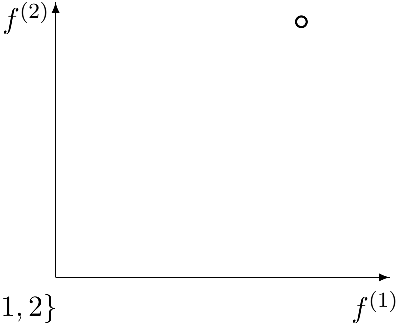
- Every Pareto solution can be obtained with some r

Reference point selection

- Use the ordering property inherent to $p = 2$ to compute gaps between 3 successive undominated solutions in the objective space
- Choose r with the largest gap
- Associate a weight to r to decrease the probability of choosing it again and prevent stalling when the Pareto front is discontinuous

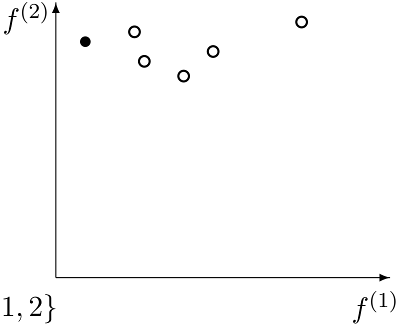


BiMADS: successive MADS runs



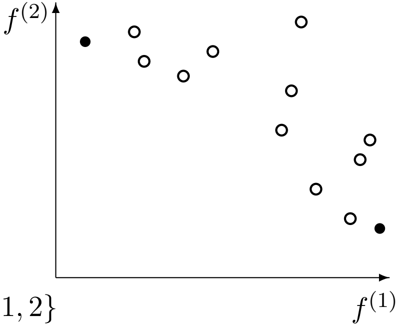
- INITIALIZATION:
  Solve $\min\limits_{x \in \Omega} f^{(q)}(x)$ for $q \in \{1, 2\}$

BiMADS: successive MADS runs



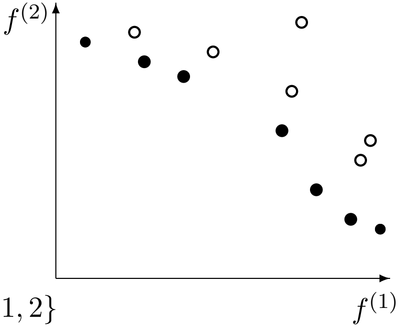
- Initialization:
  Solve $\min_{x \in \Omega} f^{(q)}(x)$ for $q \in \{1, 2\}$

BiMADS: successive MADS runs



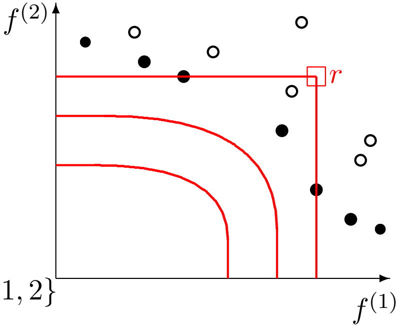
- INITIALIZATION:
  Solve $\min_{x \in \Omega} f^{(q)}(x)$ for $q \in \{1, 2\}$

BiMADS: successive MADS runs



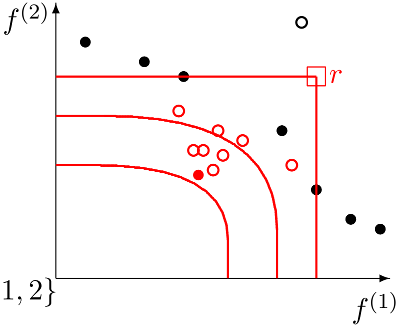
- Initialization:
  Solve $\min_{x \in \Omega} f^{(q)}(x)$ for $q \in \{1, 2\}$
- Main iterations:
  - Reference point determination:
    Use the set of feasible ordered undominated points generated so far to generate a reference point r

BiMADS: successive MADS runs



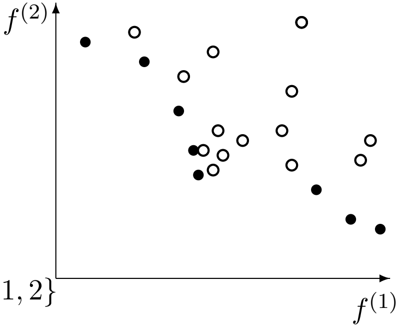
- Initialization:
  Solve $\min_{x \in \Omega} f^{(q)}(x)$ for $q \in \{1, 2\}$
- Main iterations:
  - Reference point determination:
    Use the set of feasible ordered undominated points generated so far to generate a reference point r
  - Single-objective minimization:
    Solve the subproblem $\min_{x \in \Omega} \phi_r(F(x))$

BiMADS: successive MADS runs



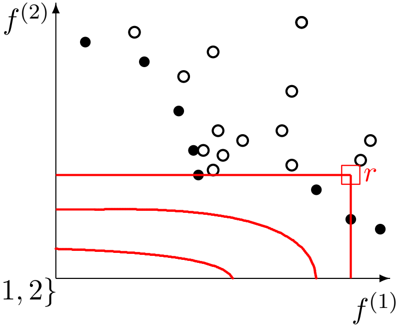
- Initialization:
  Solve $\min_{x \in \Omega} f^{(q)}(x)$ for $q \in \{1, 2\}$
- Main iterations:
  - Reference point determination:
    Use the set of feasible ordered undominated points generated so far to generate a reference point r
  - Single-objective minimization:
    Solve the subproblem $\min_{x \in \Omega} \phi_r(F(x))$

BiMADS: successive MADS runs



- INITIALIZATION:
  Solve $\min\limits_{x\in\Omega} f^{(q)}(x)$ for $q \in \{1, 2\}$
- MAIN ITERATIONS:
  - REFERENCE POINT DETERMINATION:
    Use the set of feasible ordered undominated points generated so far to generate a reference point r
  - SINGLE-OBJECTIVE MINIMIZATION:
    Solve the subproblem $\min\limits_{x\in\Omega} \phi_r(F(x))$

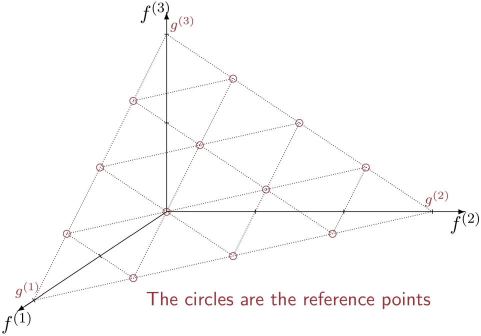
BiMADS: successive MADS runs



- INITIALIZATION:
  Solve $\min_{x\in\Omega} f^{(q)}(x)$ for $q \in \{1, 2\}$
- MAIN ITERATIONS:
  - REFERENCE POINT DETERMINATION:
    Use the set of feasible ordered undominated points generated so far to generate a reference point r
  - SINGLE-OBJECTIVE MINIMIZATION:
    Solve the subproblem $\min_{x\in\Omega} \phi_r(F(x))$

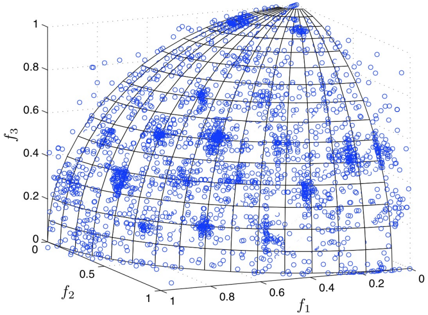
MultiMADS

- [Audet et al., 2010]
- Based on the natural boundary intersection (NBI) framework, and the convex hull of individual minima
- Consider the simplex $\{g^{(1)}, g^{(2)}, g^{(3)}\}$ obtained from the individual minima

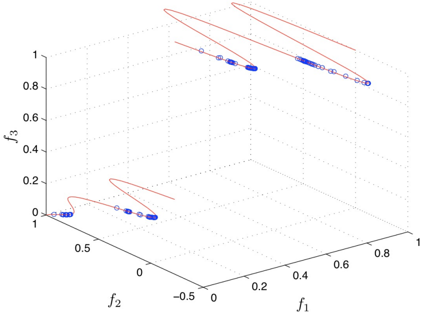


The circles are the reference points

MultiMADS: Example of solution



Spherical Pareto front



Pareto front composed of 4 disjoint segments in $\mathbb{R}^3$

Convergence analysis

- MADS solves the single objective subproblems
- These solutions $\hat{x}$ are such that if $\phi_r(F(.))$ is Lipschitz near $\hat{x}$, then $\phi_r^\circ(F(\hat{x}); d) \geq 0$ for every direction d in the hypertangent cone $T_\Omega^H(\hat{x})$ to the domain Ω at $\hat{x}$
- Every Pareto point is the optimal solution of a reformulation
- Let $\hat{x} \in \Omega$ be a refining point produced by MADS on a single-objective subproblem for some $r \in \mathbb{R}^p$. If F is Lipschitz near $\hat{x}$, then for any direction $d \in T_\Omega^H(\hat{x})$, there exists a $q \in \{1, 2, \ldots, p\}$ such that $\left(f^{(q)}\right)^\circ(\hat{x}; d) \geq 0$
- When the functions are regular, it means that moving in a feasible direction deteriorates at least one objective: This is a tradeoff solution (Pareto)

Direct Multisearch (DMS)

- [Custódio et al., 2011]
- Native adaptation of GPS to the unconstrained multiobjective case ($p \geq 2$ and use of the EB)
- Intensification with a poll step in which the acceptation criteria are based on the Pareto dominance
- Diversification with a search step
- Convergence based on the Clarke derivatives
- Differences with biMADS and multiMADS:
  - BiMADS is a framework using MADS in a subproblem while DMS is a native multiobjective method
  - At each step, DMS tries to improve the entire front, while biMADS focuses on a specific part of it

NSGA-II

- NSGA-II: **N**on-dominated **S**orting **G**enetic **A**lgorithm, for BOP ($p \geq 2$) [Deb et al., 2002]
- Constraints are treated with the inclusion of the violation in the dominance relation
- Each objective parameter is treated separately
- Mutation and crossover are performed on the population
- Selection based on "non-dominated sorting" (**intensification**), and "crowded-distance sorting" (**diversification**)
- Heuristic: No guarantee on the quality of the approximated Pareto front
- From the same team: **A**rchive-based **M**icro **G**enetic **A**lgorithm (AMGA) [Tiwari et al., 2008]

Multiobjective solvers

- `NOMAD` $(p = 2)$
- NSGA-II: Several implementations can be found:
  - MATLAB version
  - C versions
- AMGA2 [Tiwari et al., 2011]
- DMS: MATLAB version (by email)
- DFL toolbox:
  - DFMO: Linesearch, constraints, FORTRAN
  - MODIR: DIRECT, constraints, FORTRAN
  - MOIF: Implicit filtering, bounds, MATLAB

References I

Audet, C., Bigeon, J., Cartier, D., Le Digabel, S., and Salomon, L. (2018).
Performance indicators in multiobjective optimization.
Technical Report G-2018-90, Les cahiers du GERAD.

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.
(Constraint Sensitivity Analysis).

Audet, C., Savard, G., and Zghal, W. (2008).
Multiobjective Optimization Through a Series of Single-Objective Formulations.
SIAM Journal on Optimization, 19(1):188–210.
(biMADS).

Audet, C., Savard, G., and Zghal, W. (2010).
A mesh adaptive direct search algorithm for multiobjective optimization.
European Journal of Operational Research, 204(3):545–556.
(multiMADS).

Collette, Y. and Siarry, P. (2002).
Optimisation multiobjectif.
Eyrolles.
(metrics).

References II

Custódio, A., Madeira, J., Vaz, A., and Vicente, L. (2011).
Direct multisearch for multiobjective optimization.
SIAM Journal on Optimization, 21(3):1109–1140.
(DMS, metrics, profiles).

Das, I. and Dennis, Jr., J. (1998).
Normal-Boundary Intersection: A New Method for Generating the Pareto Surface in Nonlinear Multicriteria Optimization Problems.
SIAM Journal on Optimization, 8(3):631–657.
(NBI).

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002).
A fast and elitist multiobjective genetic algorithm: NSGA-II.
IEEE Transactions on Evolutionary Computation, 6(2):182–197.

E.F.Campana, M.Diez, G.Liuzzi, S.Lucidi, R.Pellegrini, V.Piccialli, F.Rinaldi, and A.Serani (2018).
A Multi-objective DIRECT algorithm for ship hull optimization.
Computational Optimization and Applications, 71(1):53–72.
(MODIR).

G.Cocchi, G.Liuzzi, A.Papini, and M.Sciandrone (2018).
An implicit filtering algorithm for derivative-free multiobjective optimization with box constraints.
Computational Optimization and Applications, 69(2):267–296.
(MOIF).

References III

Laumanns, M., Zitzler, E., and Thiele, L. (2000).
A Unified Model for Multi-Objective Evolutionary Algorithms with Elitism.
In *Congress on Evolutionary Computation*, volume 1, pages 46–53, Piscataway, New Jersey, USA.
(metrics).

Liuzzi, G., Lucidi, S., and Rinaldi, F. (2016).
A Derivative-Free Approach to Constrained Multiobjective Nonsmooth Optimization.
SIAM Journal on Optimization, 26(4):2744–2774.
(DFMO).

Tiwari, S., Fadel, G., and Deb, K. (2011).
AMGA2: improving the performance of the archive-based micro-genetic algorithm for multi-objective optimization.
Engineering Optimization, 43(4):377–401.

Tiwari, S., Koch, P., Fadel, G., and Deb, K. (2008).
AMGA: An Archive-based Micro Genetic Algorithm for Multi-objective Optimization.
In *Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation*, GECCO '08, pages 729–736, New York, NY, USA. ACM.

Zitzler, E., Deb, K., and Thiele, L. (2000).
Comparison of Multiobjective Evolutionary Algorithms: Empirical Results.
Evolutionary Computation, 8(2):173–195.
(metrics).