Pareto Front Approximation through a Multi-objective Augmented Lagrangian Method

EUROPT 2021
18th Workshop on Advances in Continuous Optimization
Toulouse, July 9, 2021

G. Cocchi, M. Lapucci, **P. Mansueto**

Department of Information Engineering
University of Florence

Single-point descent methods

- Extensions of classical scalar optimization techniques.
- Production of sequence of points asymptotically driven to optimality.
- `MultiObjectiveSteepestDescent` (`MOSD`) [Fliege and Svaiter, 2000]: one of the most famous.
- Drawback:
  - Single solution, no Pareto front approximation.
  - Pareto front useful to choose the most appropriate trade-off.

Our goals

- Focus on Augmented Lagrangian Algorithm for Multi-objective Optimization (`ALAMO`) [Cocchi and Lapucci, 2020]:
  - Single-point extension of scalar Augmented Lagrangian Method (`ALM`).
- Definition of an extended version of `ALAMO`:
  - Management of set of mutually non-dominated points;
  - Properties of convergence to Pareto-stationarity.

Problem

Definition

$$\min_{x \in \mathbb{R}^n} F(x) = (f_1(x), \ldots, f_m(x))^T$$
$$\text{s.t. } g(x) \leq 0 \qquad \text{(OPT)}$$

- $F : \mathbb{R}^n \to \mathbb{R}^m$
  - Continuously differentiable.
- $g : \mathbb{R}^n \to \mathbb{R}^p$
  - Continuously differentiable.
  - Component-wise convex.

Multi-objective Augmented Lagrangian

Definition

$$\mathcal{L}_\tau(x,\mu) = F(x) + \frac{\tau}{2}\left(\sum_{i=1}^{p}\left(\max\left\{0, g_i(x) + \frac{\mu_i}{\tau}\right\}\right)^2\right) e$$

- $\tau > 0 \in \mathbb{R}$: penalty parameter.
- $\mu \geq 0 \in \mathbb{R}^p$: Lagrange multipliers.

- Problem with `MO` Augmented Lagrangian.

$$\min_{x\in\mathbb{R}^n} \mathcal{L}_\tau(x,\mu) = (\ell_{\tau 1}(x,\mu), \ldots, \ell_{\tau m}(x,\mu))^T \qquad \text{(LAG)}$$

Pareto-stationarity

Definition

$\bar{x} \in \mathbb{R}^n$ *Pareto-stationary* for (LAG) if

$$\max_{j=1,\dots,m} \nabla \ell_{\tau j}(\bar{x}, \mu)^T d \geq 0 \ \forall d \in \mathbb{R}^n.$$

- Equivalent condition.

$$\min_{d \in \mathbb{R}^n} \max_{j=1,\dots,m} \nabla \ell_{\tau j}(\bar{x}, \mu)^T d = 0$$

- $\bar{x}$ not Pareto-stationary $\to \exists d \in \mathbb{R}^n$, descent for all objectives.

Steepest common descent direction

Definition

$$v(\bar{x}) = \arg\min_{\substack{d \in \mathbb{R}^n \\ \|d\| \leq 1}} \max_{j=1,\dots,m} \nabla \ell_{\tau j}(\bar{x}, \mu)^T d \qquad \text{(SC)}$$

- $\theta : \mathbb{R}^n \to \mathbb{R}$
  - $\theta(\bar{x})$: optimal value of (SC) at $\bar{x}$.

Steepest partial descent direction

Definition

Let $I \subseteq \{1, \ldots, m\}$ subset of objectives.

$$v^I(\bar{x}) = \underset{\substack{d \in \mathbb{R}^n \\ \|d\| \leq 1}}{\arg\min} \max_{j \in I} \nabla \ell_{\tau j}(\bar{x}, \mu)^T d \qquad \text{(SP)}$$

- $\theta^I(\bar{x})$: optimal value of (SP) at $\bar{x}$.

MOSD – Scheme

Input: $x^0 \in \mathbb{R}^n$.

While x^k not Pareto-stationary

- Compute
$$d^k \in \operatorname*{arg\,min}_{\substack{d \in \mathbb{R}^n \\ \|d\| \leq 1}} \max_{j=1,\dots,m} \nabla \ell_{\tau j}(x^k, \mu)^T d$$
- Find step size α_k (`ArmijoTypeLineSearch`)
- $x^{k+1} = x^k + \alpha_k d^k$

Output: x^k.

ArmijoTypeLineSearch – Scheme and properties

Inputs:

- $x, d \in \mathbb{R}^n$;
- $\alpha_0 > 0$;
- $\delta, \beta \in (0, 1)$.

While $\mathcal{L}_\tau(x + \alpha d, \mu) \not\leq \mathcal{L}_\tau(x, \mu) + \beta\alpha J_{\mathcal{L}_\tau}(x, \mu)d$

- $\alpha = \delta\alpha$

Output: α.

Finite termination

If $\mathcal{L}_\tau(\cdot, \mu)$ continuously differentiable and $J_{\mathcal{L}_\tau}(x, \mu)d < 0$ $(\theta(x) < 0)$, then $\exists\epsilon > 0$ such that

$$\mathcal{L}_\tau(x + td, \mu) < \mathcal{L}_\tau(x, \mu) + \beta t J_{\mathcal{L}_\tau}(x, \mu)d \quad \forall t \in (0, \epsilon].$$

ε-Pareto-stationarity

Definition

Let $\varepsilon \geq 0$. $\bar{x} \in \mathbb{R}^n$ ε-*Pareto-stationary* for (LAG) if

$$\min_{\substack{d \in \mathbb{R}^n \\ \|d\| \leq 1}} \max_{j=1,\ldots,m} \nabla \ell_{\tau j}(\bar{x}, \mu)^T d \geq -\varepsilon.$$

- ε-Pareto-stationarity $\rightarrow$ finite termination of `MOSD`.

FRONT-ALAMO – Scheme

Inputs:

- X^0, feasible non-dominated points for (OPT);
- $\{\varepsilon_k\} \subset \mathbb{R}$, decreasing sequence.

For $k = 0, 1, \ldots$

- Inner loop
  - Exploration and refinement phase for X^k
  - `MOSD` and `ArmijoTypeLineSearch`
  - Output $\rightarrow X^{k+1}$
- Update for multipliers and penalty parameter

FRONT-ALAMO – Inner loop

- $X^{k+1} = X^k$
- $\forall x_c \in X^k$, $\forall I \subseteq \{1, \dots, m\}$, if $\theta^I_k(x_c) < 0$
  - $d \in v^I_k(x_c)$
  - Find α (ArmijoTypeLineSearch) considering $\mathcal{L}^I_{\tau_k}(\cdot, \mu^k)$
  - Find z through MOSD with $\mathcal{L}_{\tau_k}(\cdot, \mu^k), x_c + \alpha d, \varepsilon_k$
  - If $\nexists y \in X^{k+1} : \mathcal{L}_{\tau_k}(y, \mu^k) \lneqq \mathcal{L}_{\tau_k}(z, \mu^k)$
    - $X^{k+1} = X^{k+1} \setminus \{x \in X^{k+1} \mid \mathcal{L}_{\tau_k}(z, \mu^k) \lneqq \mathcal{L}_{\tau_k}(x, \mu^k)\}$
    - $X^{k+1} = X^{k+1} \cup \{z\}$

FRONT-ALAMO – Update phase

- μ update
  - $\mu_i^{k+1} = \Pi_{[0,\bar{\mu}]}\left[\mu_i^k + \tau_k \max_{x\in X^{k+1}}\{g_i(x)\}\right] \quad \forall i = 1, \dots, p$
- τ update
  - $V_i^{k+1} = \min\left\{\min_{x\in X^{k+1}}\{-g_i(x)\}, \frac{\mu_i^k}{\tau_k}\right\} \qquad \forall i = 1, \dots, p$
  - $\left\|V^{k+1}\right\| > \sigma\left\|V^k\right\| \to \tau_{k+1} = \rho\tau_k$

FRONT-ALAMO – Properties – 1

- MOSD and line search finite termination $\rightarrow$ FRONT-ALAMO well-defined.

X^{k+1} features

$\forall k$, $\forall x^{k+1} \in X^{k+1}$:

- $\nexists y \in X^{k+1} : \mathcal{L}_{\tau_k}(y, \mu^k) \lneqq \mathcal{L}_{\tau_k}(x^{k+1}, \mu^k)$;
- x^{k+1} is ε_k-Pareto-stationary w.r.t. $\mathcal{L}_{\tau_k}(\cdot, \mu^k)$.

FRONT-ALAMO – Properties – 2

Let

- $\varepsilon_k \to 0$,
- $\{x^{k+1}\}$ such that $x^{k+1} \in X^{k+1}$.
- $K \subseteq \{0, 1, \dots\}$ and $\lim_{\substack{k\to\infty \\ k\in K}} x^{k+1} = \bar{x}$.

Limit points feasibility

$g(\bar{x}) \leq 0$.

Limit points Pareto-stationarity

$\bar{x}$ is Pareto-stationary for (OPT).

State-of-the-art algorithms

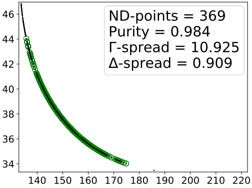
- MOSQP [Fliege and Vaz, 2016]: gradient-based method.
- DMS [Custódio et al., 2011]: derivative-free methodology.
- NSGA-II [Deb et al., 2002]: non-dominated sorting-based evolutionary algorithm.

Settings

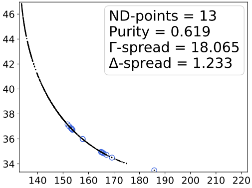
- Hyper-parameters values chosen by preliminary results or according to the algorithms authors.
- Code written in `Python3`.
- For each algorithm and problem, time limit of 2 minutes.
- For each problem, `NSGA-II` settings:
  - 10 runs;
  - 10 fronts compared based on *purity*;
  - Best front is chosen.

M-BNH2

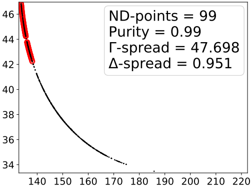
- $m = 2$; $n = 2$; 2 non-linear constraints.



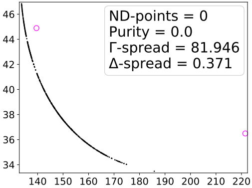
(a) `FRONT-ALAMO`



(b) `DMS`



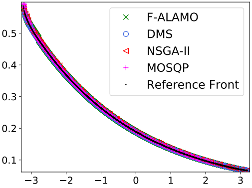
(c) `NSGA-II`



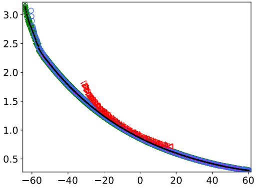
(d) `MOSQP`

LAP2

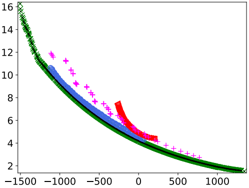
- $m = 2$; 1 non-linear constraint.



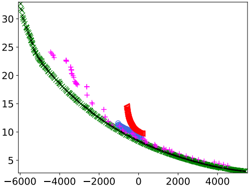
(a) LAP2, $n = 2$



(b) LAP2, $n = 10$



(c) LAP2, $n = 50$

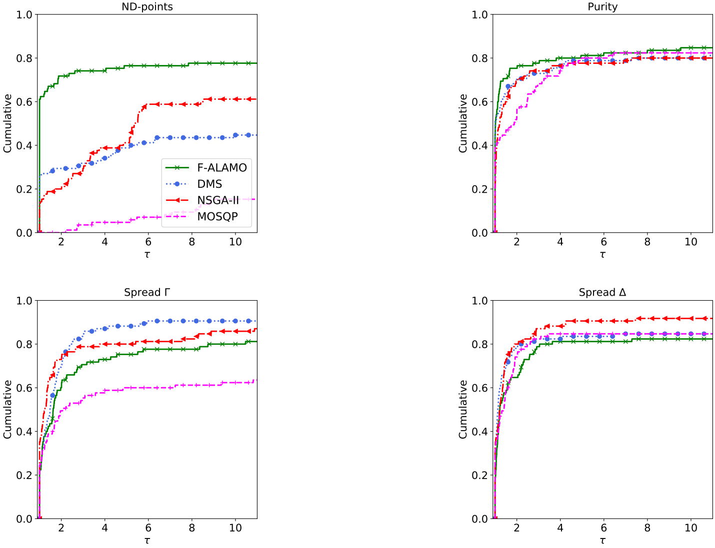


(d) LAP2, $n = 100$

Bound-constrained problems – 1

- CEC [Zhang et al., 2008].
  - $m = 2$: CEC1, CEC2, CEC3, CEC7.
  - $m = 3$: CEC8, CEC9, CEC10.
  - $n \in \{5, 10, 20, 30, 40, 50, 100, 200\}$.
- ZDT [Zitzler et al., 2000].
  - $m = 2$: ZDT1, ZDT2.
  - $n \in \{2, 5, 10, 20, 30, 40, 50, 100, 200\}$.
- MOP [Huband et al., 2006].
  - MOP1: $m = 2$, $n = 1$.
  - MOP2: $m = 2$, $n \in \{2, 5, 10, 20, 30, 40, 50, 100, 200\}$.
  - MOP3: $m = 2$, $n = 2$.

Bound-constrained problems – 2



Conclusions

- Smooth multi-objective optimization problems with convex constraints.
- Augmented Lagrangian method designed to produce Pareto front approximations.
- Properties of global convergence to Pareto-stationarity.
- Great performance w.r.t. state-of-the-art algorithms.

References I

Cocchi, G. and Lapucci, M. (2020).
An augmented Lagrangian algorithm for multi-objective optimization.
Computational Optimization and Applications, 77(1):29–56.

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

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.

Fliege, J. and Svaiter, B. F. (2000).
Steepest descent methods for multicriteria optimization.
Mathematical Methods of Operations Research, 51(3):479–494.

References II

Fliege, J. and Vaz, A. I. F. (2016).
A method for constrained multiobjective optimization based on SQP techniques.
SIAM Journal on Optimization, 26(4):2091–2119.

Huband, S., Hingston, P., Barone, L., and While, L. (2006).
A review of multiobjective test problems and a scalable test problem toolkit.
IEEE Transactions on Evolutionary Computation, 10(5):477–506.

Zhang, Q., Zhou, A., Zhao, S., Suganthan, P., Liu, W., and Tiwari, S. (2008).
Multiobjective optimization test instances for the cec 2009 special session and competition.
Mechanical Engineering.

Zitzler, E., Deb, K., and Thiele, L. (2000).
Comparison of multiobjective evolutionary algorithms: Empirical results.
Evol. Comput., 8(2):173–195.

Pareto Front Approximation through a Multi-objective Augmented Lagrangian Method

EUROPT 2021

18th Workshop on Advances in Continuous Optimization

Toulouse, July 9, 2021

G. Cocchi, M. Lapucci, **P. Mansueto**

Department of Information Engineering

University of Florence