



**HAL**  
open science

# Clouds of points optimization in convex polygons with evolutionary algorithms based on Wasserstein barycenters

Babacar Sow, Rodolphe Le Riche, Julien Pelamatti, Merlin Keller, Sanaa Zannane

► **To cite this version:**

Babacar Sow, Rodolphe Le Riche, Julien Pelamatti, Merlin Keller, Sanaa Zannane. Clouds of points optimization in convex polygons with evolutionary algorithms based on Wasserstein barycenters. MASCOT-NUM2024, Apr 2024, Giens, France. hal-04645929

**HAL Id: hal-04645929**

**<https://hal.science/hal-04645929>**

Submitted on 12 Jul 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



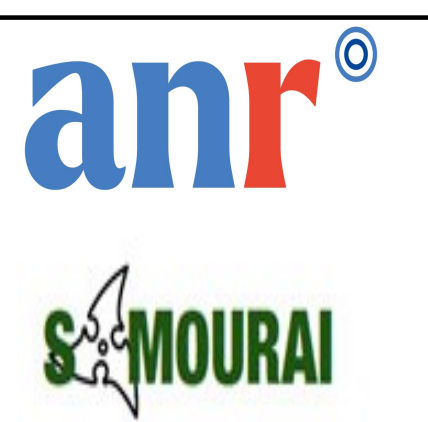
# Clouds of points optimization in convex polygons with evolutionary algorithms based on Wasserstein barycenters

Babacar SOW (EMSE, LIMOS), Rodolphe LE RICHE (CNRS LIMOS)

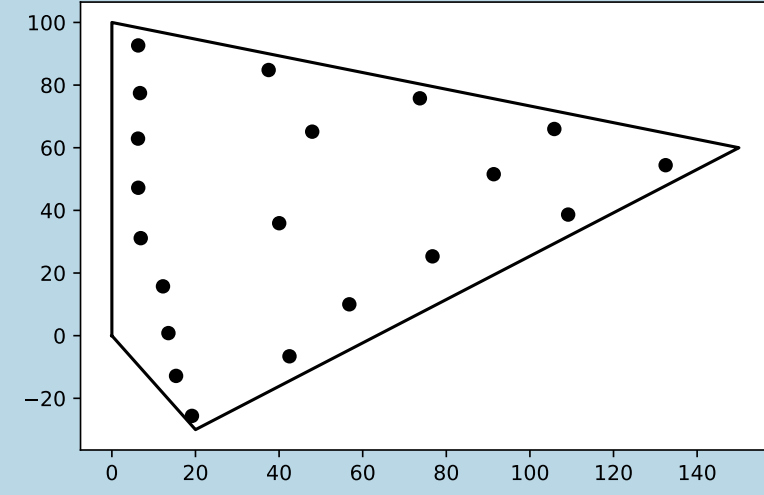
Julien PELAMATTI(EDF R&D), Merlin KELLER (EDF R&D), Sanaa ZANNANE (EDF R&D)

Universités: Ecole Nationale Supérieure des Mines de Saint-Etienne(EMSE), Université Clermont Auvergne(UCA)

Laboratoire: LIMOS Email of contact: babacar.sow@emse.fr



## Context and problem



- A cloud of points, denoted by  $X$ , is a set of  $n$  unordered points:  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  with  $\mathbf{x}_i \in \mathbb{R}^d$  ( $d = 2$  for the rest). The points of the clouds belong to a convex polygon.
- First aim: Use evolutionary algorithms with operators based on the Wasserstein barycenter when functions evaluations are cheap.
- Second aim: Perform Bayesian Optimization in case of expensive functions evaluations.

## An evolutionary algorithm

Clouds can be viewed as measures through the bijective mapping:  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\} \mapsto P_X = \frac{1}{n} \sum_{i=1}^n \delta_{\mathbf{x}_i}$ .

Given two clouds of points  $X_1$  and  $X_2$ , we denote  $W_2^2(P_{X_1}, P_{X_2})$  the 2-Wasserstein distance between  $P_{X_1}$  and  $P_{X_2}$  (see [2] for details).

The evolutionary optimizer over clouds of points is based on a mutation alternating between the following two operators (randomly chosen with a Bernoulli law of parameter 0.5):

- **Full domain mutation:**  $X_m = \text{W-barycenter}(X_1, X_r, \epsilon)$  where  $P_{X_m}$  is the discrete measure minimizing  $(\epsilon W_2^2(P_X, P_{X_1}) + (1 - \epsilon) W_2^2(P_X, P_{X_r}))$ .  $X_r$  contains points randomly sampled in the convex domain and  $\epsilon$  is uniformly chosen in  $[0, 1]$ . If  $X_r$  and  $X_1$  have the same sizes then  $X_m$  has the same one. Otherwise, the mutation produces two clouds with the different two sizes.
- **Boundary Mutation:**  $X_b = \text{W-barycenter}(X_1, (X_1 \cup \text{Bound}), \epsilon)$ .  $\text{Bound}$  contains points randomly sampled in the domain boundary. The size of  $X_b$  is chosen to be equal to the one of  $X_1$ .

The first one has a contracting property while the second increases systematically the spread of clouds.

## Test functions over clouds of points

$F_0$  is an analytical function mimicking the power production of wind-farms in a  $0^\circ$ -directed wind zone.  $F_{0\_pen}$  is its variant, penalized by the size of the input. Additional test functions are:  $F_{inert}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\}) = \sum_{i=1}^n \|\mathbf{x}_i - \bar{X}\|^2$  and  $F_{minDist}(\{\mathbf{x}_1, \dots, \mathbf{x}_n\}) = \min_{i \neq j} \|\mathbf{x}_i - \mathbf{x}_j\|$ .

## References

- [1] Muandet, Krikamol and Fukumizu, Kenji and Sriperumbudur, Bharath and Schölkopf, Bernhard and others: *Kernel mean embedding of distributions: A review and beyond*, Foundations and Trends® in Machine Learning
- [2] Cuturi, Marco and Doucet, Arnaud Fast computation of Wasserstein barycenters. *International conference on machine learning*

## Evolutionary algorithm: Wasserstein barycenter vs Gaussian distribution

- The evolutionary algorithm is solely based on mutation because the presence of crossover tends to worsen the results. The size of the population is denoted by  $\lambda$ , and at each iteration  $2\lambda$  clouds of points are generated. The algorithm selects the best  $\lambda$  clouds for the next generation.
- We compare the Wasserstein Barycenter Generator based Evolutionary Algorithm denoted by WBGEA with an algorithm having the same structure but based on a Gaussian mutation (denoted by Gauss). For the latter, each point  $\mathbf{x}_i$  is disturbed with  $\mathcal{N}(\mathbf{x}_i, \sigma^2)$  with  $\sigma^2$  equal to  $(0.01) \mathbb{E}\|X - X'\|^2$ , where  $X$  and  $X'$  are points sampled uniformly in the domain.
- We fix  $\lambda$  to 300. The sizes of the clouds vary between 10 and 20. The results indicate that the algorithm based on Wasserstein operators yield better results except on  $F_{minDist}$ .

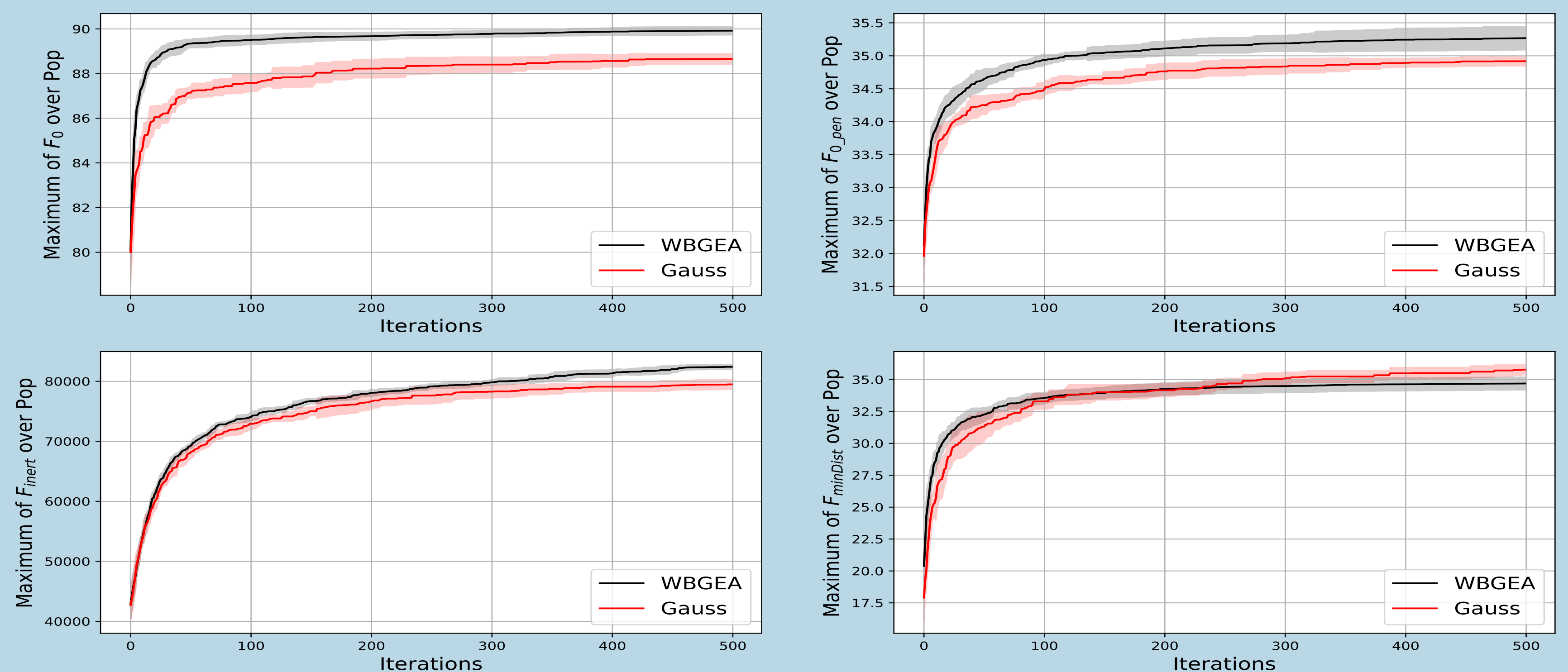


Figure 1: Average over 10 (+/- std. deviation) of the evolutions of the maximum of the functions in each population over the evolutionary algorithms iterations.

## Bayesian optimization: MMD as kernel and WBGEA for the optimization of the EI

- Let  $\mathcal{H}$  be a Reproducing Kernel Hilbert Space with a characteristic kernel  $k_{\mathcal{H}}$  such as Matérn  $5/2$ . The characteristic nature guarantees the injectivity of the embedding map [1]:  $P_X \mapsto \mu_X = \int P_X(x) k_{\mathcal{H}}(x, \cdot) dx$ .  $\theta_1$  and  $\theta_2$  are 2 hyper-parameters scaling the dimensions between two points  $\mathbf{x}$  and  $\mathbf{x}'$  through  $|x_1 - x'_1|/\theta_1$  and  $|x_2 - x'_2|/\theta_2$ .
- The correlation kernel  $K(X, X') = \sigma \exp(-\frac{\|\mu_X - \mu_{X'}\|_{\mathcal{H}}^2}{2\theta^2})$  is symmetric and semi-definite positive. We denote this kernel by MMD for Maximum Mean Discrepancy and use it to define a Gaussian process over clouds of points.
- The acquisition function is the expected improvement (EI) with the following formula:  $EI(X) = (y^{min} - \mu(X)) \Phi(\frac{(y^{min} - \mu(X))}{\sqrt{\Sigma(X, X)}}) + \sqrt{\Sigma(X, X)} \phi(\frac{(y^{min} - \mu(X))}{\sqrt{\Sigma(X, X)}})$  where  $\mu(X)$  and  $\Sigma(X, X)$  denote respectively the mean and the variance of the prediction.  $\Phi$  and  $\phi$  are respectively the cumulative distribution and density function of the standard normal law. The EI is defined over the same search space as the functions to optimize. It is optimized with WBGEA.
- We fix the budget of evaluations to 100. A random initial set of 50 clouds is chosen. The hyper-parameters of the kernel are updated every 5 iterations by maximization of the likelihood. The acquisition criterion is optimized at each iteration. We present below the percentage of the maximum value attained by Bayesian optimization (BO) with respect to that attained by WBGEA (denoted by Percentage\_BO\_WBGEA):

Functions	$F_0$	$F_{0\_pen}$	$F_{inert}$	$F_{minDist}$
Percentage_BO_WBGEA	96.05%	90.12%	69.35%	65.80%

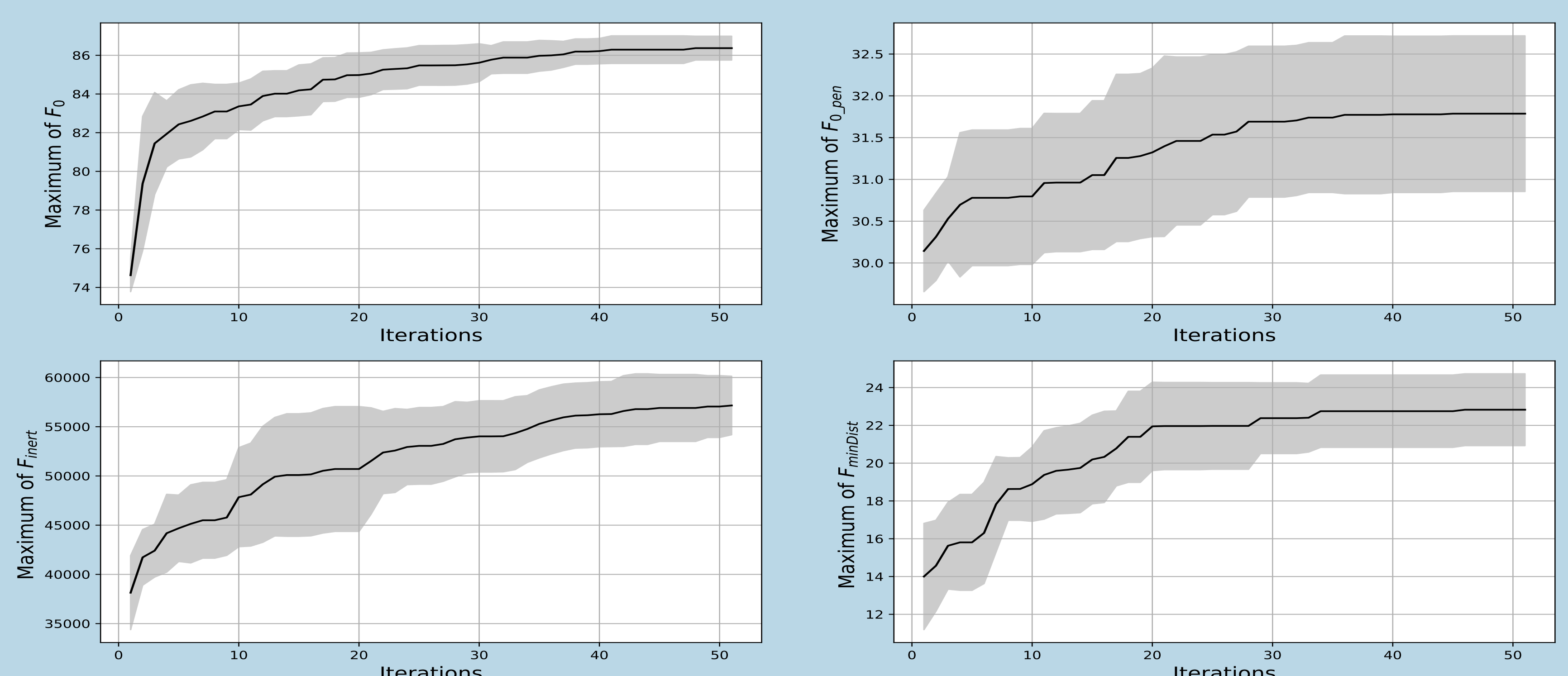


Figure 2: Average over 10 (+/- std. deviation) of the evolutions of the maximum of the functions over the BO iterations.