Optimization and metamodeling of functions defined over clouds of points

based on the PhD thesis of Babacar Sow

in collaboration with:

- LE RICHE Rodolphe (CNRS)
- PELAMATTI Julien (EDF R&D)
- KELLER Merlin (EDF R&D)
- ZANNANE Sanaa (EDF UK)

This work was funded in part by the French Agence Nationale de la Recherche through the SAMOURAI project, ANR-20-CE46-0013

December 10, 2024

Table of contents

Context

Context of the thesis

Research project

- The thesis is part of the project ANR SAMOURAI (Simulation Analytics and Meta-model-based solutions for Optimization, Uncertainty and Reliability Analysis).
  - I Design MM adapted to large scale problems with limited simulations budget.
  - II Large scale sequential enrichment strategies for reliability-based design-optimization / inversion
  - III Design efficient black-box optimization methods with mixed input variables
  - IV Learn hidden constraints within adaptive design procedures.

Axis III: optimization and metamodeling

- It is an academic thesis between Mines Saint-Etienne and EDF R&D (PRISME)
- This project explores both theoretical aspects and industrial applications related to the **optimization** over clouds of points. The latter also includes the **metamodeling** of costly simulators.

Context of the work

Industrial optimization problems

- This work is inspired by industrial optimization problems such as well placement Guyaguler and Horne [11], positioning of turbines Emami and Noghreh [7] or sensors Krause et al. [13].

Illustration of wind turbines positioned in a landscape.

Statistical optimization problems

- Similar challenges can be encountered when optimizing the underlying criterion in design of experiments, see [17].

Problem formulation

An optimization problem over sets of vectors

$\boldsymbol{D}$ is a compact domain of $\mathbb{R}^d$, $n_{\min}$ and $n_{\max} \in \mathbb{N}$:

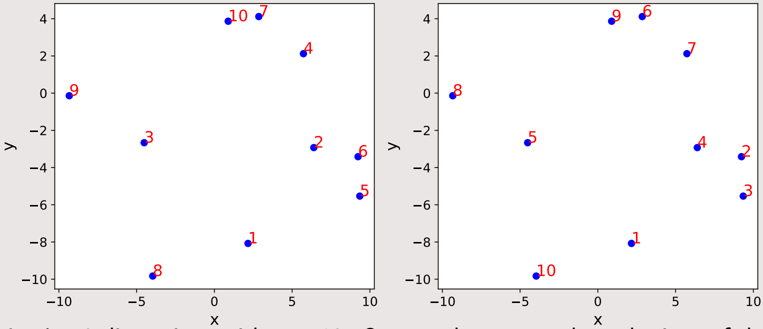
$$\max_{X=\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}} \quad F(X) \,,$$
$$n \in \{n_{\min}, \dots, n_{\max}\} \,,$$
$$\forall i, \boldsymbol{x}_i \in \boldsymbol{D} \subset \mathbb{R}^d \,.$$

- The function F is assumed to be **black box** (see Nakayama, Arakawa, and Washino [14], Xiao et al. [21]) meaning that no information related to its convexity, continuity, derivatives and smoothness is known.
- The inputs are in the form of sets (bags, clouds) of vectors (points): $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{\boldsymbol{n}}\}$, with $\boldsymbol{x}_{\boldsymbol{i}} \in \mathbb{R}^d$ and $n \in \{n_{\min}, \dots, n_{\max}\}$ (n, $n_{\min}$, $n_{\max} \in \mathbb{N}$).

Mixed aspect: no order and varying size

Clouds of points

The functions of interest are **permutation invariant** with respect to their inputs.



A cloud of points in $d = 2$ dimensions with $n = 10$. One can have several numberings of the points for the same cloud.

Operations research approaches

A general approach to mixed continuous optimization with varying dimensions can be found in the work of Hallé-Hannan, Audet, Le Digabel, Diouane, et al. ([2, 12]): meta-variables $\equiv n$. But we would like to account more closely for the specificities of the set of vectors problem.

A typical operations research formulation (for minimization) would be (thanks Renaud Chicoisne):
Assume $0 \leq F(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) \leq u_n$ and $n_{\min} = 1$

$$\min_{\eta, z_i, \boldsymbol{x}_i, i=1,\ldots,n_{\max}} \eta \qquad \text{such that}$$

$$z_n \in \{0, 1\} \ , \ \forall n \in \{1, \ldots, n_{\max}\} \quad \text{and} \quad \sum_{n=1}^{n_{\max}} z_n = 1$$

$$\eta \geq F(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n) - u_n(1 - z_n) \ , \ \forall n \in \{1, \ldots, n_{\max}\}$$

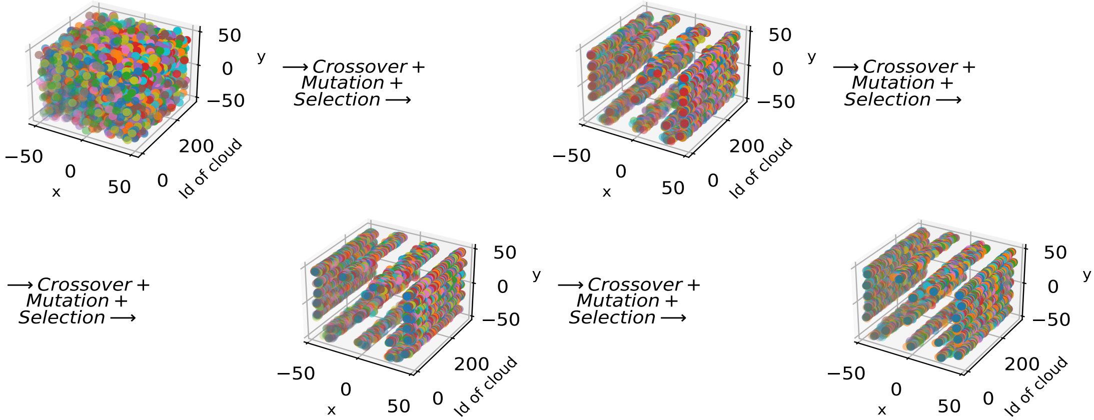
but additional assumptions must be made on F to make it MILP approximable.

Two parts in this presentation depending on the computational cost of F

- Inexpensive: stochastic (evolutionary) algorithms [6]
- Expensive: replacing the true function F by a Gaussian process [20, 9, 10]

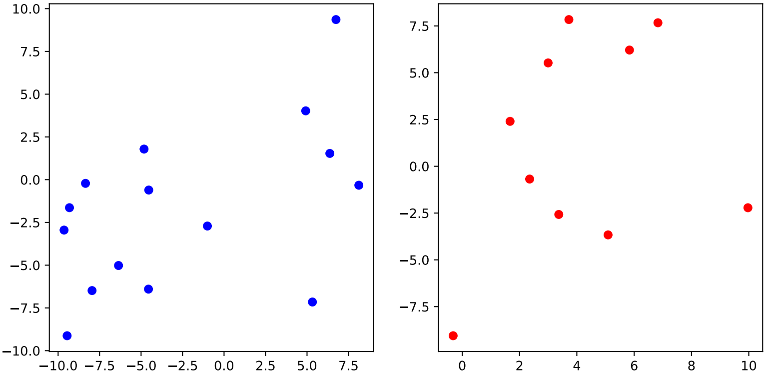
Evolutionary optimization over sets of vectors: inexpensive case

Principle: perturb and update an initial set (population) of sets (clouds) of vectors through evolutionary operators.



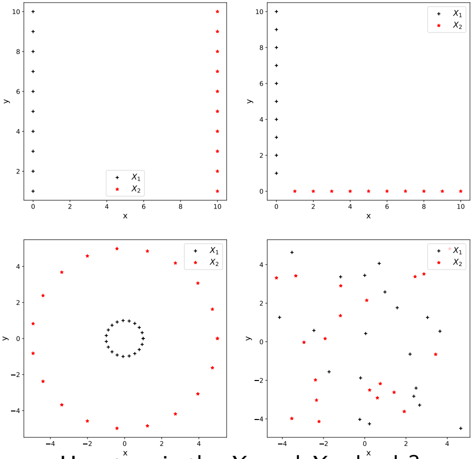
Populations corresponding to iterations 1, 50, 100, 200 (left to right, top to bottom)

Question 1: which metric for the input space?



Two sets of points with different sizes: which distance between the two sets?

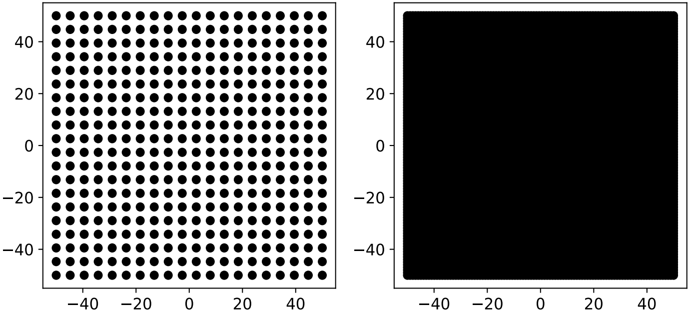
Question 2: which evolutionary operators for sets of points?



How to mix the X_1 and X_2 clouds?

State of the art for wind farm layout

- [4], [16], and [15] describe algorithms for optimizing sets of vectors based on, respectively, simulated annealing, genetic algorithm and particle swarm.
- Generally authors suppose predefined fixed points and use binary encoding. Our work differs by letting points vary continuously.



Grid of points on the left, continuous domain on the right

Contributions to questions 1 and 2

- A set of continuous vectors can be modeled as a uniform discrete measure with finite support.
- This model helps defining a topology in the space of sets of points using the Wasserstein distance between measures.
- We propose evolutionary crossover and mutation operators relying on the concept of the Wasserstein barycenter.

(for more details, see Sow et al. [19])

Input model and Wasserstein metric

Sets of vectors model: discrete uniform measures

- To two cloud of points $X^{(j)} = \{\boldsymbol{x}_1^{(j)}, \ldots, \boldsymbol{x}_n^{(j)}\}$, $j = 1, 2$ we associate $P_{X^{(j)}} = \frac{1}{n} \sum_{i=1}^{n} \delta_{\boldsymbol{x}_i^{(j)}}$
- We can compute a new cloud of points by finding an intermediary uniform measure.

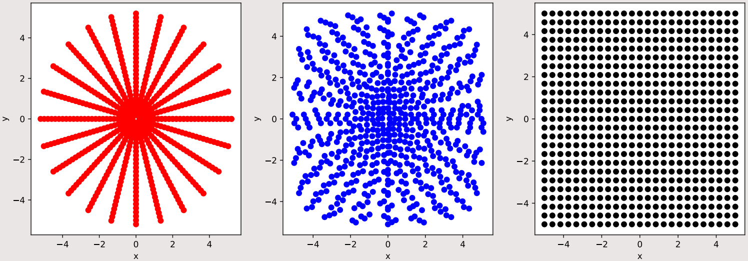
Wasserstein distance

- To two measures μ and ν defined over $\mathbb{R}^d$, the Wasserstein distance of order p is defined as follows : $W_p^p(\mu, \nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \rho(\boldsymbol{x}, \boldsymbol{x}')^p \mathrm{d}\pi(\boldsymbol{x}, \boldsymbol{x}')$
  - $\rho(\boldsymbol{x}, \boldsymbol{x}')$ corresponds to the Euclidean distance between $\boldsymbol{x}$ and $\boldsymbol{x}'$
  - $\Pi(\mu, \nu)$ is the set of all probability measures defined over $\mathbb{R}^d \times \mathbb{R}^d$ with marginals μ and ν.

Wasserstein barycenters

- *The barycenter of N measures* $\nu_1, ..., \nu_N$ *is the measure* ν^* *that minimizes* $f(\nu) = \sum_{i=1}^{N} \epsilon_i W_p^p(\nu, \nu_i)$, with $\epsilon_i \geq 0, \sum_{i=1}^{N} \epsilon_i = 1$ see Agueh and Carlier [1].
- Relatively fast calculation with the method of Cuturi and Doucet [5] implemented in [8].

Illustration



Two initial sets at left and right, and their equal weight Wasserstein barycenter in the middle

Contracting effect

Theorem

Consider $\mathcal{P}'$ to be the set of discrete measures over $\mathbb{R}^d$ with finite support and $\epsilon \in [0,1]$. Let P_{X_1}, P_{X_2} and P_{X^*} be defined respectively as

- $P_{X_1} = \sum_{i=1}^n \alpha_i \delta_{\mathbf{x}_i^1}, \sum_{i=1}^n \alpha_i = 1, \alpha_i > 0,$
- $P_{X_2} = \sum_{j=1}^m \beta_j \delta_{\mathbf{x}_j^2}, \sum_{j=1}^m \beta_j = 1, \beta_j > 0$,
- $P_{X^*} = \sum_{l=1}^k \lambda_l \delta_{\mathbf{x}_l^*}, \sum_{l=1}^k \lambda_l = 1, \lambda_l > 0$,

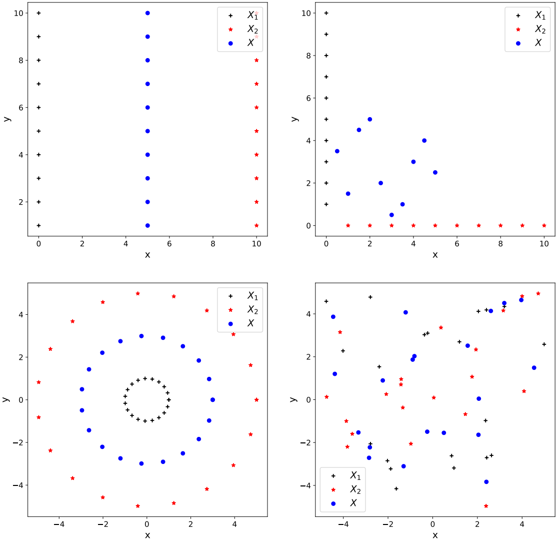
with P_{X^*} the unique minimizer of $\underset{P_X \in \mathcal{P}'}{\arg} \min \epsilon W_2^2(P_X, P_{X_1}) + (1-\epsilon) W_2^2(P_X, P_{X_2})$.

We have:

$$\forall l \in \{1, ..., k\}, \mathbf{x}_l^* \in \overline{Conv(\mathbf{x}_1^1, ..., \mathbf{x}_n^1, \mathbf{x}_1^2, ..., \mathbf{x}_m^2)}$$

where $\overline{Conv(\mathbf{x}_1^1, ..., \mathbf{x}_n^1, \mathbf{x}_1^2, ..., \mathbf{x}_m^2)}$ is the closed convex hull of the set $\{\mathbf{x}_1^1, ..., \mathbf{x}_n^1, \mathbf{x}_1^2, ..., \mathbf{x}_m^2\}$

Wasserstein barycenters of some sets of points



X_1 and X_2 are two sets and X represents their equal weight Wasserstein barycenter

Evolutionary operators: crossovers

- Given $\epsilon \sim \mathcal{U}[0,1]$
- **Equal weights crossover:** For two measures (P_{X_1} and P_{X_2}), X_c is defined as

$$P_{X_c} = \arg\min_{P_X} W_2^2(P_X, P_{X_1}) + W_2^2(P_X, P_{X_2})$$

- **Random weights crossover:** For two measures (P_{X_1} and P_{X_2}), X_c is defined as

$$P_{X_c} = \arg\min_{P_X} \epsilon W_2^2(P_X, P_{X_1}) + (1-\epsilon) W_2^2(P_X, P_{X_2})$$

- Number of vectors: card(X_c) equal both card(X_1) and card(X_2)
- Which crossover ?

Evolutionary operators: mutations

- Given $\epsilon \sim \mathcal{U}[0,1]$
- **Full Domain mutation:** given X_c and X_{rand} a cloud of points randomly sampled in the domain, X_m is defined as

$$P_{X_m} = \arg\min_{P_X} \epsilon W_2^2(P_X, P_{X_c}) + (1-\epsilon) W_2^2(P_X, P_{X_{rand}})$$

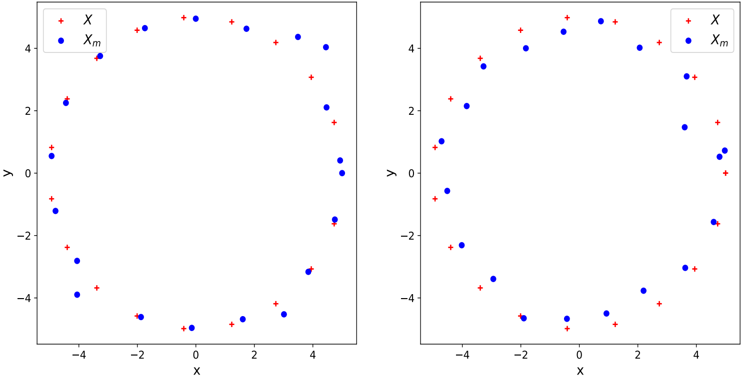
$$\mathrm{card}(X) = \mathrm{card}(X_{rand}) = \mathrm{card}(X_c) \pm (-1, 0, 1) \text{ uniform, if feasible}$$

- **Escape from contraction:** To define operators allowing to counteract the contracting property, we introduce the following mutation over clouds of points:
  - **Boundary mutation:** given X_c and X_{bound} a cloud of points randomly sampled at the domain boundary (a point on each side), X_m is defined as

$$P_{X_m} = \arg\min_{P_X \ ,\ \mathrm{card}(X)=\mathrm{card}(X_c)} \epsilon W_2^2(P_X, P_{X_c}) + (1-\epsilon) W_2^2(P_X, P_{X_c \cup X_{bound}}).$$

- How to arrange the two mutations ?

Escaping the contraction



X and X_m in red and blue, respectively, $\epsilon = 0.5$. Boundary mutation on left and Full Domain mutation on right.

Alternating mutation

Do a Boundary Mutation with probability *prob*, Full Domain Mutation otherwise.

Algorithm 1 Alternating Wasserstein mutation

Input: X cloud to mutate, *prob* the probability to perform a Boundary mutation
Output: The mutated cloud(s)

1: Draw ϵ and r uniformly in $[0, 1]$
2: **if** $r \geq prob$ **then**
3: Do Full Domain mutation with weight ϵ
4: **else**
5: Do Boundary mutation with weight ϵ
6: **end if**

(the two mutations have also been tested in a deterministic successive way) default evolutionary implementation for comparison

Classic crossover

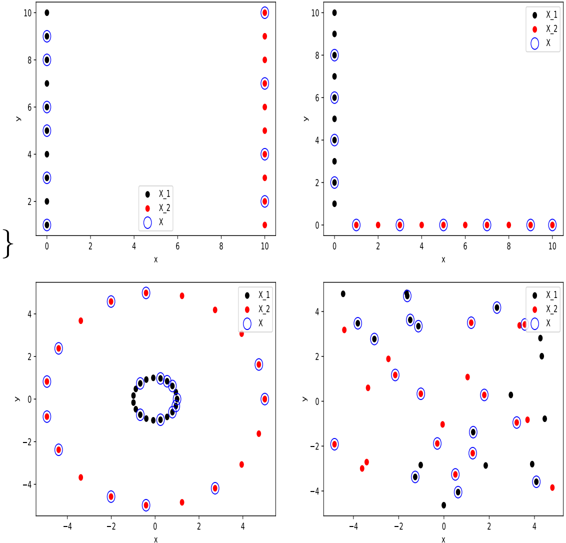
Uniform crossover, component per component of a vector encoding

$$X^1 = \{\boldsymbol{x}_1^1, \ldots, \boldsymbol{x}_{n_1}^1, \emptyset_{n_1+1}, \ldots, \emptyset_{n_{max}}\}$$

$$X^2 = \{\boldsymbol{x}_1^2, \quad \ldots \quad, \boldsymbol{x}_{n_2}^2, \ \ldots, \emptyset_{n_{max}}\}$$

$$X^c = \{\mathcal{U}(X_i^1, X_i^2)\ ,\ i = 1, \ldots, n_{max}\}$$

X^1 and X^2 in black and red and X^c (circled)



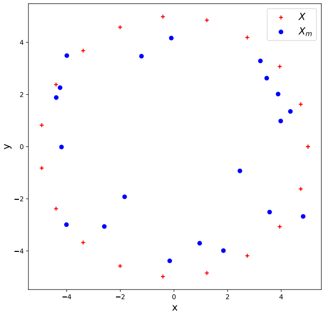
Classic mutation

$$X = \{\boldsymbol{x}_1, ..., \boldsymbol{x}_n, \emptyset_{n+1}, ..., \emptyset_{n_{max}}\}$$

$$X_m = \{\boldsymbol{x}_1 + \mathcal{N}(0,\, \sigma^2 \boldsymbol{I_d}), ..., \boldsymbol{x}_n + \mathcal{N}(0,\, \sigma^2 \boldsymbol{I_d}), \emptyset_{n+1}, ..., \emptyset_{n_{max}}\}$$

The Gaussian distributions are truncated in order to yield points inside the domain.

Add or remove a point or leave unchanged, uniformly, if feasible.



X (red crosses) and X_m (blue points), σ^2 is 3.333

Test functions

Mimicking wind-farms productions

- $F_\theta(\{\mathbf{x_1}, ..., \mathbf{x_n}\}) = \sum_{i=1}^{n} \left(\prod_{j, j \neq i} f_{x_j,\theta}(\mathbf{x_i}) \right) f_0(\mathbf{x_i})$
- $F_{\theta\_pen}(\{\mathbf{x_1}, ..., \mathbf{x_n}\}) = F_\theta(\{\mathbf{x_1}, ..., \mathbf{x_n}\}) - n\sqrt{n} + 1.5n$

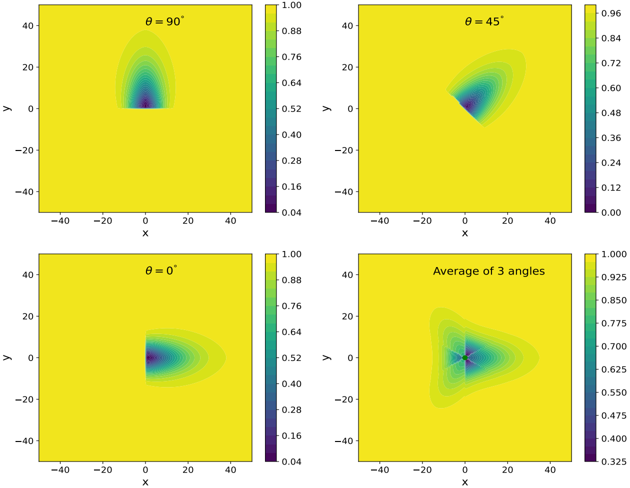
Mindist and Inertia

- $F_{minDist}(\{\mathbf{x_1}, ..., \mathbf{x_n}\}) = \min_{i \neq j} ||\mathbf{x}_i - \mathbf{x}_j||$.
- $F_{inert}(\{\mathbf{x_1}, ..., \mathbf{x_n}\}) = \sum_{i=1}^{n} ||\mathbf{x}_i - \bar{x}||^2$ with $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} \mathbf{x_i}$.

Numerical tests parameters

- The number of points varies between 10 and 20, in a fixed square domain. The number of iterations and populations sizes are respectively 500 and 300.

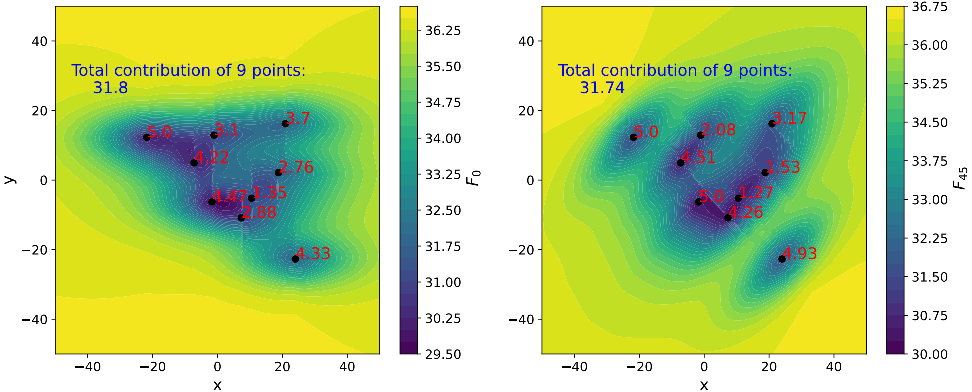
Example of $f_{x,\theta}(.)$ (I)



Representation of f_p with $\theta = 90^\circ$ at top left, $\theta = 45^\circ$ top right, $\theta = 0^\circ$ bottom left, and averaged directions at bottom right. We denote the corresponding functions respectively F_{90}, F_{45}, F_0, F_{4d}.

Example of $f_{x,\theta}(.)$ (II)

The position of 9 points is displayed. The map shows the total contribution as a function of the position of a new point.



Representation of F_0 and F_{45} with 9 fixed points and a one varying. The maximal contribution of a point is fixed to 5.

Population diversity metric

Wasserstein-based diversity

The diversity can be calculated at each iteration in the following way:

$$\mathrm{Div}(\mathrm{pop}) \;=\; \frac{1}{\lambda} \sum_{X^i \in \mathrm{pop}} W_2^2(P_{\bar{X}}, P_{X^i}) \;,$$

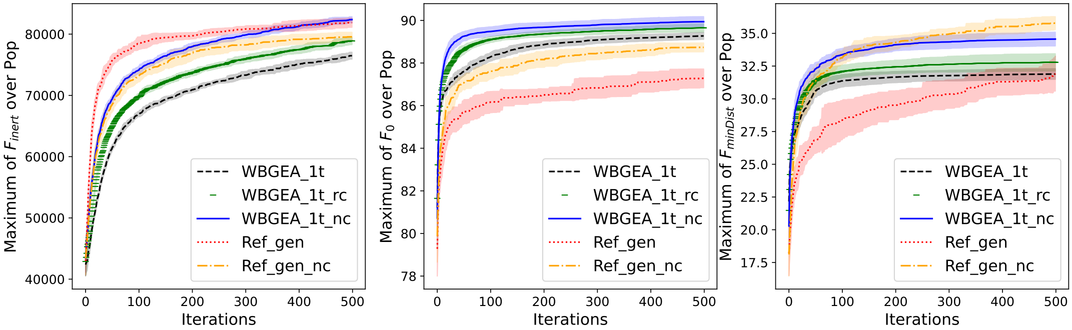
where $\mathrm{pop} = \{X^i, i = 1, ..., \lambda\}$ is a population of sets, P_{X^i} the associated discrete measures, and $P_{\bar{X}}$ is the Wasserstein barycenter of the clouds of pop.

Algorithms names

- WBGEA_1t denotes the algorithm based on Wasserstein operators with equal weights crossover, WBGEA_1t_rc (random weight crossover) and WBGEA_1t_nc (no crossover)
- Ref_gen denotes the baseline comparison algorithm and Ref_gen_nc its version without crossover.

Wasserstein vs classic operators: performances

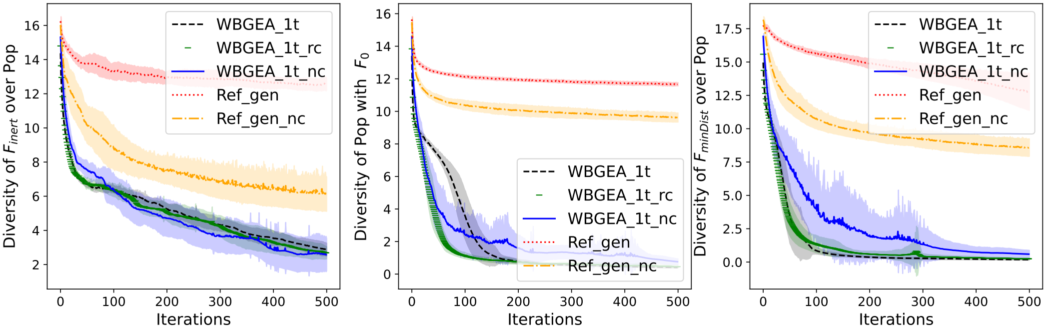
The algorithm based on Wasserstein operators denoted as WBGEA yields better results except on $F_{minDist}$.



Average over 20 (+/- std. deviation) of the evolutions of the maximum of the functions in each population over the evolutionary algorithms iterations.

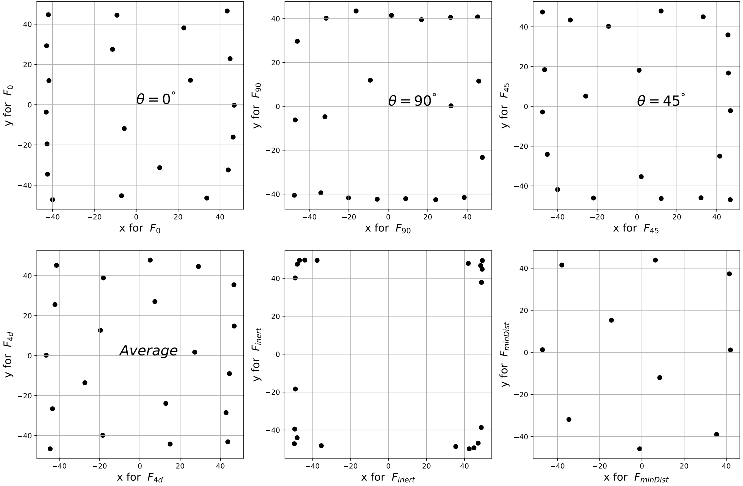
Wasserstein vs classic operators: diversities

The diversities of the populations in the algorithms based on Wasserstein operators vanish to zero more quickly.



Average over 20 (+/- std. deviation) of the evolutions of the diversities of the populations over the evolutionary algorithms iterations.

Best designs

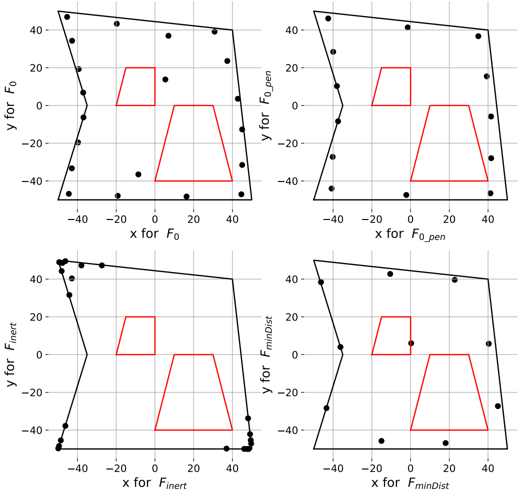


Best observed designs corresponding, respectively, to the test cases F_0, F_{90}, F_{45}, F_{4d}, F_{inert} and $F_{minDist}$ (left to right, up to bottom).

Non-convex domains: best designs with projection

The introduced operators can be effectively applied to non-convex domains, by repairing the generated sets by projection.

The PhD manuscript contains a proof that projecting an infeasible set onto the feasible domain yields the closest feasible set in terms of Wasserstein distance.



Feasible designs generated by the WBGEA evolutionary optimizer with projection as a repair method.

Partial conclusions

- Numerical tests suggest a mutation independence principle = risk mitigation as certain operators are more adapted to certain functions: the boundary and the full domain mutations should be alternated randomly.
- For WBGEA, crossover with random weights yields better results but the absence of crossover is more competitive on the test functions.
- The Wasserstein operators seem to be adapted to optimize functions where the optimal design present a global geometrical pattern (such as alignments, cf. kernels later).
- (The Wasserstein crossover reduces diversity and the classic crossover keeps diversity.)

Metamodeling functions defined over clouds of points

Computationally intensive case

Metamodeling

- When the objectife function is expensive, one may need to approximate it with a *surrogate model* to perform **Bayesian optimization**
- Some examples of such functions are simulations such as computational fluid dynamics (CFD), finite element structural analysis, ...

Requirements

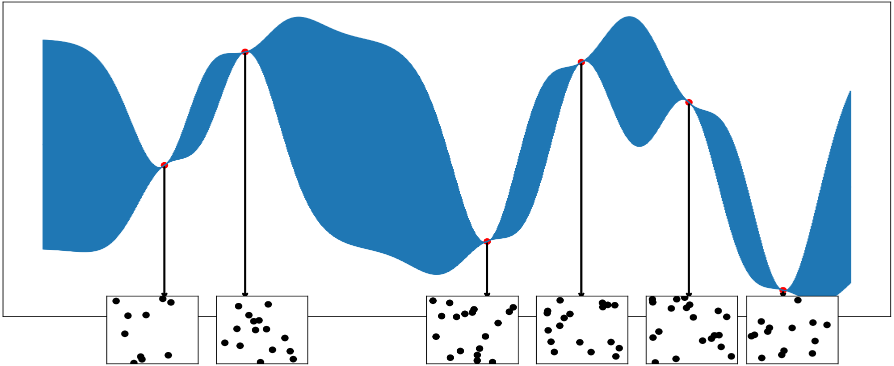
The surrogate models are required to be

- extendable to inputs such as sets of vectors,
- predictive for unexplored input values (not in the training data set)

Gaussian processes (see Williams and Rasmussen [20])

A Gaussian process prior

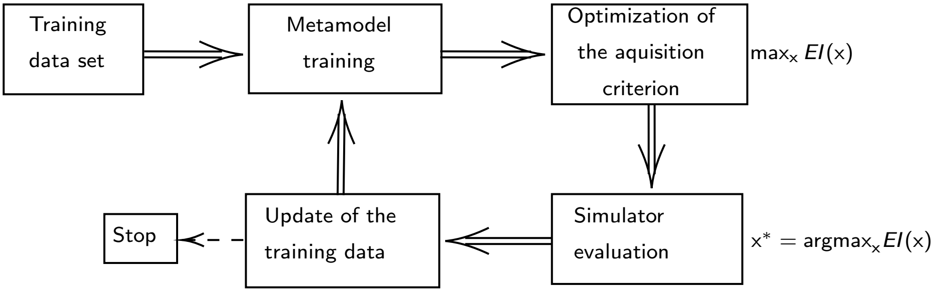
Gaussian processes are defined by a mean function m and a covariance kernel k over the input spaces $\mathcal{X}$. The kernel must be semi-definite positive.



Bayesian optimization

Generic concept

Iterative enrichment of the training set in order to maximize the information regarding the global optimum location



Generic concept

The *Expected Improvement* can be optimized with WBGEA

Research questions and contributions

Research questions

- **Research question 4: Which semi-definite positive kernel between sets of vectors to use for the construction of Gaussian processes?**
- **Research question 5: Can some kernels perform better than others depending on the geometry of the clouds in the training set?**

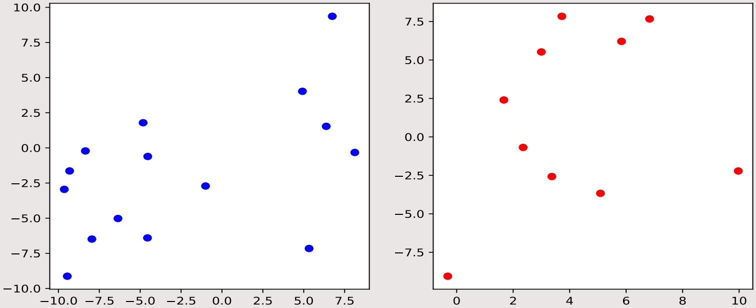
Contributions: for more details, see Sow et al. [18]

- We discuss the modeling of sets of vectors as distributions or feature vectors and present the associated relevant methods for defining kernels.
- We test alternative kernels in a set of numerical experiments. In particular, the extrapolation properties of the kernels are investigated by considering geometrical transformations of possible training data sets.

Comparing the outputs based on the inputs!

Similarity

Define a function of similarity between X_{blue} and X_{red}: $k(X_{blue}, X_{red})$



Two clouds of points in $d = 2$ dimensions with $n = 15$ points for the blue cloud and $n = 10$ points for the red one.

Substitution kernels

Substitution with Exponential

- Firstly, we consider correlation kernels of the form: $k(X, Y) = exp(-\frac{\Psi(X,Y)}{2\theta^2})$.
- We know that k defined above is a valid kernel (symmetric and positive semi-definite) if and only if Ψ is **Hermitian** (symmetric in the real case) and **conditionally negative semi-definite** Berg, Christensen, and Ressel [3].
- In other words, for any M distinct points and $c \in R^M$ with $\sum_{i=1}^{M} c_i = 0$, the following inequality must hold: $\sum_{i=1}^{M} \sum_{j=1}^{M} c_i c_j \Psi(X_i, X_j) \leq 0$

Metric Cases

- If $\Psi(X, Y) = d(\phi(X), \phi(Y))^2$, with d the distance between, $\phi(X)$ and $\phi(Y)$ the respective images of X and Y in an metric space, the above conditions are equivalent to the fact that the metric be Hilbertian.

Kernels over sets of vectors

Substitution kernels

We focus here on two possible definition of $\Psi(\cdot,\cdot)$

- The MMD
  - $\mu_X(.) = \int P_X(\boldsymbol{x}) k_{\mathcal{H}}(\boldsymbol{x}, .) \mathrm{d}\boldsymbol{x}$.
  - $MMD(P_X, P_{X'}) = ||\mu_X - \mu'_X||$
- The Sliced-Wasserstein distance
  - $SW_2(P_X, P_{X'}) = \int_{\mathcal{S}} W_2^2(\alpha * P_X, \alpha * P_{X'}) \mathrm{d}\alpha$

Kernel through measures and features

- In Sow et al. [18], we discussed kernels based on the features of the set (denoted RFK), sliced-Wasserstein distance (denoted Slice-Wass) and embedding models for regression problems.

Predictive performances for various kernels (see Sow et al. [18])

Function \ Kernels	MMD	RFK	Slice-Wass
F_0	**0.906**	0.897	0.828
F_{45}	0.868	**0.893**	0.821
F_{90}	**0.899**	0.871	0.843
F_{40d}	**0.906**	0.799	0.824

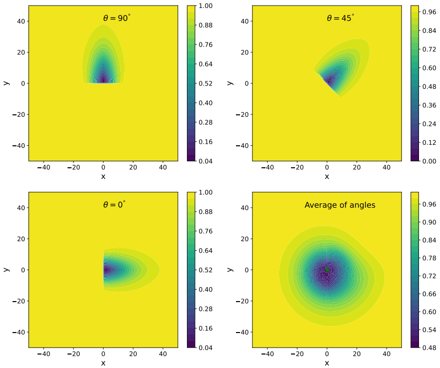
Table: Q^2 of 3 kernels on all the wind farm proxy functions, the testing clouds come from a random design.

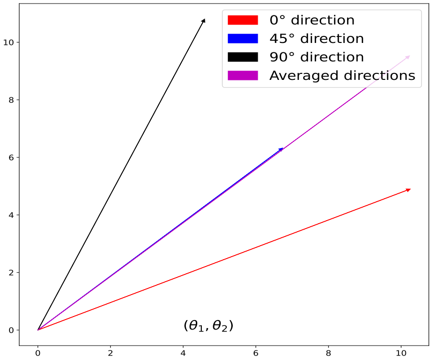
Function \ Kernels	MMD	RFK	Slice-Wass
F_{inert}	0.734	0.988	0.905
$F_{minDist}$	-0.051	0.997	0.587

Table: Q^2 of 3 kernels on F_{inert} and $F_{minDist}$, the testing clouds come from a random design.

Fitting the kernels to the test functions

θ_1 and θ_2 are 2 hyper-parameters in $k_{\mathcal{H}}$ that scale the distance between two points $\boldsymbol{x}$ and $\boldsymbol{x}'$ through $|x_1 - x_1'|/\theta_1$ and $|x_2 - x_2'|/\theta_2$. MMD-based kernels adapt to the geometrical properties of wind-farms functions through the θ_1 and θ_2.





Left: reminder of the turbine contributions for winds at 90°,45°,0° and 40 directions (left to right, top to bottom). Right: $(\theta_1, \theta_2)^\top$ vectors of length scales of the embedding kernel.

Exploiting a metamodel for Bayesian optimization

Performances of Bayesian optimization (BO) with respect to Evolutionary Algorithm (EA)

- We fix the budget to $T = 100$. A random initial set of 50 clouds is chosen. The hyper-parameters of the kernel are updated every 5 iterations.
- 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	95.95%	90.07%	71.81%	65.28%

Main result

- With fewer number of evaluations (inferior to 10^{-1}%), BO can attain more than 95% of the estimated optimum by EA on wind-farm function.

Conclusions

Modeling a set of vectors as a discrete measure

- Modeling a set of vectors as a discrete measure helps having more possibilities of defining kernels and can yield interpretable results.

Kernels comparison

- The various kernel performances are highly dependent on the relation between the set geometry and the output value
- MMD based kernels seem to be more adapted to functions with different directions of variations (anisotropy) than other kernels.

Bayesian optimization

- The combination of the MMD kernel with the evolutionary algorithm produces promising results. However, the performance of Bayesian optimization appears to be closely influenced by the kernel choice.

Limitations and perspectives

Contributions

- **Defining a general optimization problem over sets of vectors.**
- **Evolutionary algorithms over sets of vectors in convex domains.** We have proposed evolutionary operators based on the Wasserstein barycenter.
- **Extension to non-convex domains.** We have extended the operators to non-convex domains relying on optimal projections onto the feasible domain
- **Gaussian processes over sets of points.** We have introduced and/or studied several semi-definite positive kernels over sets of points and benchmarked them over functions. They are combined with evolutionary operators in the context of **Bayesian optimization**.

Perspectives

- **Sampling set of points:** what is a good design of experiments over sets of points ? And more generally, how to efficiently sample sets of points ?
  - Surrogate model initial training sets, evolutionary algorithms starting populations, etc...
- Analyzing the algorithms scaling capabilities for sets of vectors of dimension higher than 2.
- Compare the Bayesian algorithm based on kernels on probability measures with extensions of MADS that have meta-variables.
- **Complexity of the algorithms:** reduce the complexity of the algorithms with an emphasis on those of the Wasserstein operators.
- **Other types of constraints:** Geometrical constraints such as a minimal distance between the points or alignments can be included in the problem.
- Considering other approaches based on a definition of a gradient in the set of vectors space

Scientific communications: publications

- Babacar Sow, Rodolphe Le Riche, Julien Pelamatti, Merlin Keller and Sanaa Zannane. Learning functions defined over sets of vectors with kernel methods. *In 5 th ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2023), 2023.*
- Babacar Sow, Rodolphe Le Riche, Julien Pelamatti, Merlin Keller and Sanaa Zannane. Active learning for the optimization of functions defined over clouds of points. *55es Journées de Statistique de la SFdS (JdS 2024), May 2024, Bordeaaux, France*
- Babacar Sow, Rodolphe Le Riche, Julien Pelamatti, Merlin Keller and Sanaa Zannane. Wasserstein-Based Evolutionary Operators for Optimizing Sets of Points: Application to Wind-Farm Layout Design. *Journal: Applied Sciences, 14, 2024, 17*

Scientific communications: conferences communications

- Babacar Sow, Rodolphe Le Riche, Julien Pelamatti, Sanaa Zannane and Merlin Keller. Gaussian Processes Indexed by Clouds of Points: a study. *MASCOT-NUM, Jun 2022, Clermont Ferrand, France*
- Babacar Sow, Rodolphe Le Riche, Sanaa Zannane, Merlin Keller and Julien Pelamatti. Wasserstein Barycenter-based Evolutionary Algorithm for the optimization of sets of points. *PGMO DAYS 2023, Nov 2023, Paris-Saclay, France*
- Babacar Sow, Rodolphe Le Riche, Sanaa Zannane, Merlin Keller and Julien Pelamatti. Cloud of points as discrete measures for Gaussian models and stochastic optimization. *MASCOT-NUM2023, Apr 2023, LE CROISIC, France*
- Babacar Sow, Rodolphe Le Riche, Julien Pelamatti, Sanaa Zannane and Merlin Keller. Clouds of points optimization in convex polygons with evolutionary algorithms based on Wasserstein barycenters. *MASCOT-NUM2024, Apr 2024, Giens, France*
- Babacar Sow, Rodolphe Le Riche, Julien Pelamatti, Sanaa Zannane and Merlin Keller. Optimization of functions defined over sets of points in polygons with evolutionary algorithms based on Wasserstein barycenters. *FGS Conference on Optimization 2024 : French German Spanish conference on optimization, Jun 2024, Gijon, Spain*.

Martial Agueh and Guillaume Carlier. "Barycenters in the Wasserstein space". In: *SIAM Journal on Mathematical Analysis* 43.2 (2011), pp. 904–924.

Charles Audet, Edward Hallé-Hannan, and Sébastien Le Digabel. "A general mathematical framework for constrained mixed-variable blackbox optimization problems with meta and categorical variables". In: *Operations Research Forum*. Vol. 4. 1. Springer. 2023, p. 12.

Christian Berg, Jens Peter Reus Christensen, and Paul Ressel. *Harmonic analysis on semigroups: theory of positive definite and related functions*. Vol. 100. Springer, 1984.

Martin Bilbao and Enrique Alba. "Simulated annealing for optimization of wind farm annual profit". In: *2009 2nd International symposium on logistics and industrial informatics*. IEEE. 2009, pp. 1–5.

Marco Cuturi and Arnaud Doucet. "Fast computation of Wasserstein barycenters". In: *International conference on machine learning*. PMLR. 2014, pp. 685–693.

Agoston E Eiben and James E Smith. *Introduction to evolutionary computing*. Springer, 2015.

- Alireza Emami and Pirooz Noghreh. "New approach on optimization in placement of wind turbines within wind farm by genetic algorithms". In: *Renewable Energy* 35.7 (2010), pp. 1559–1564.
- Rémi Flamary et al. "Pot: Python optimal transport". In: *Journal of Machine Learning Research* 22.78 (2021), pp. 1–8.
- Peter I Frazier. "A tutorial on Bayesian optimization". In: *arXiv preprint arXiv:1807.02811* (2018).
- Roman Garnett. *Bayesian Optimization*. Cambridge University Press, 2023.
- Baris Guyaguler and Roland Horne. "Optimization of well placement". In: *J. Energy Resour. Technol.* 122.2 (2000), pp. 64–70.
- Edward Hallé-Hannan et al. "A graph-structured distance for heterogeneous datasets with meta variables". In: *arXiv preprint arXiv:2405.13073* (2024).
- Andreas Krause et al. "Efficient sensor placement optimization for securing large water distribution networks". In: *Journal of Water Resources Planning and Management* 134.6 (2008), pp. 516–526.

Hirotaka Nakayama, Masao Arakawa, and Koji Washino. "Optimization for black-box objective functions". In: *Optimization and optimal control*. World Scientific, 2003, pp. 185–210.

Ajit C Pillai et al. "Comparison of offshore wind farm layout optimization using a genetic algorithm and a particle swarm optimizer". In: *International Conference on Offshore Mechanics and Arctic Engineering*. Vol. 49972. American Society of Mechanical Engineers. 2016, V006T09A033.

Ajit C Pillai et al. "Optimisation of offshore wind farms using a genetic algorithm". In: *International Journal of Offshore and Polar Engineering* 26.03 (2016), pp. 225–234.

Luc Pronzato. "Minimax and maximin space-filling designs: some properties and methods for construction". In: *Journal de la Société Française de Statistique* 158.1 (2017), pp. 7–36.

Babacar Sow et al. "Learning functions defined over sets of vectors with kernel methods". In: *5 th ECCOMAS Thematic Conference on Uncertainty Quantification in Computational Sciences and Engineering (UNCECOMP 2023)*. 2023.

Babacar Sow et al. "Wasserstein-Based Evolutionary Operators for Optimizing Sets of Points: Application to Wind-Farm Layout Design". In: *Applied Sciences* 14.17 (2024). ISSN: 2076-3417. DOI: 10.3390/app14177916. URL: https://www.mdpi.com/2076-3417/14/17/7916.

Christopher KI Williams and Carl Edward Rasmussen. *Gaussian processes for machine learning*. Vol. 2. 3. MIT press Cambridge, MA, 2006.

Jin-ke Xiao et al. "Optimization on black box function optimization problem". In: *Mathematical Problems in Engineering* 2015.1 (2015), p. 647234.