



HAL
open science

Combining Cross-Entropy and MADS Methods for Inequality Constrained Global Optimization

Charles Audet, Jean Bigeon, Romain Couderc

► **To cite this version:**

Charles Audet, Jean Bigeon, Romain Couderc. Combining Cross-Entropy and MADS Methods for Inequality Constrained Global Optimization. SN Operations Research Forum, 2021, 2 (3), 10.1007/s43069-021-00075-y . hal-03447123

HAL Id: hal-03447123

<https://hal.science/hal-03447123v1>

Submitted on 24 Nov 2021

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.

Combining Cross Entropy and MADS methods for inequality constrained global optimization

Charles Audet[†] Jean Bigeon[‡] Romain Couderc^{†§}

July 27, 2021

Abstract:

This paper proposes a way to combine the Mesh Adaptive Direct Search (**Mads**) algorithm with the Cross-Entropy (CE) method for nonsmooth constrained optimization. The CE method is used as an exploration step by the **Mads** algorithm. The result of this combination retains the convergence properties of **Mads** and allows an efficient exploration in order to move away from local minima. The CE method samples trial points according to a multivariate normal distribution whose mean and standard deviation are calculated from the best points found so far. Numerical experiments show the efficiency of this method compared to other global optimization heuristics. Moreover, applied on complex engineering test problems, this method allows an important improvement to reach the feasible region and to escape local minima.

Keywords: Cross Entropy; MADS; Global optimization; Derivative-free optimization; Black-box optimization; Constrained optimization.

1 Introduction

This work studies inequality constrained blackbox optimization problems of the form:

$$\min_{\mathbf{x} \in \Omega \subseteq \mathbb{R}^n} f(\mathbf{x}), \quad (1)$$

with

$$\Omega = \{\mathbf{x} \in \mathcal{X} : c_j(\mathbf{x}) \leq 0, j = 1, 2, \dots, m\},$$

[†]GERAD and Département de mathématiques et génie industriel, École Polytechnique de Montréal, C.P. 6079, Succ. Centre-ville, Montréal, Québec, Canada H3C 3A7.

Mail: charles.audet@gerad.ca

[‡]Jean Bigeon, CNRS, LS2N, Nantes, France.

Mail: jean.bigeon@ls2n.fr

[§]Univ. Grenoble Alpes, CNRS, Grenoble INP*, G-SCOP, 38000 Grenoble, France.

Mail: romain.couderc@grenoble-inp.fr

*Institute of Engineering Univ. Grenoble Alpes

where $f : \mathbb{R}^n \rightarrow \bar{\mathbb{R}} = \mathbb{R} \cup \pm\infty$, $c : \mathbb{R}^n \rightarrow \bar{\mathbb{R}}^m$ and the set \mathcal{X} represents bound constraints of type $\ell \leq \mathbf{x} \leq \mathbf{u}$ with $\ell, \mathbf{u} \in \bar{\mathbb{R}}^n$. The specificity of this work is due to the form of the objective function f and of the constraints c_j . They can be the result of a simulation of complex physical phenomena. These simulations can take an important amount of time or present some discontinuities and therefore classical optimization methods are difficult to apply. Especially, when the gradient of the objective function and/or of the constraints are not explicitly known, hard to compute or its estimation is time consuming. This field is called derivative-free optimization (DFO). In the worst case, the gradient does not even exist, which is called blackbox optimization (BBO).

Specialized BBO and DFO algorithms have been developed in order to solve this kind of problem. There are two main categories: model based algorithms [15] and direct search algorithms [5]. This work deals with direct search algorithms which benefit from theoretical convergence results and adding some modifications may improve their performance. In particular, the Mesh Adaptive Direct Search (Mads) algorithm [2] ensures convergence to a point satisfying necessary conditions based on the Clarke calculus [13]. This theoretical guarantee is a solid basis for blackbox optimization. However, blackbox optimization algorithms must take into account two other types of difficulties. First, algorithms must be efficient in terms of simulation evaluations (constraints and objective function). Indeed, the simulation in an engineering context is often time consuming. Second, blackbox simulations may involve multi-extrema functions, so algorithms must be able to escape from local minimum. Mads may be trapped in a local minimum.

To address the second difficulty, the Mads algorithm may be combined with Variable Neighborhood Search (VNS) [1] and Latin Hypercube Sampling (LHS) [43] techniques to escape local minima. Other heuristics of global optimization with no convergence guarantees exist including Evolution Strategy with Covariance Matrix Adaptation (CMA-ES) [20], Genetic Algorithm (GA) [18], Differential Evolution (DE) [39] or Particule Swarm Optimization (PSO) [25]. However, these heuristics often require a large number of function evaluations which is incompatible with the first difficulty. Moreover, no specific mechanism has been developed to enable these methods to deal with inequality constraints. In contrast, some methods have been developed to address the first problem, by reducing the overall number of simulation evaluation such as: the use of ensembles of surrogate [6] or of quadratic models [14] and the integration of the Nelder-Mead (NM) algorithm [9]. These different methods improve the efficiency of the Mads algorithm but do not address the difficulty of local optima.

The objective of the present research is to propose an alternative strategy: the Cross Entropy (CE) [42] method in hopes of escaping local minima. This method is a trade-off between a more global search and a limited number of blackbox evaluations. It was introduced in 1997, first in a context of rare event in discrete optimization [41] and then adapted to continuous optimization [27]. The main benefit of using CE is that it often converges rapidly to a promising region in the space of variables. However, this method does not benefit from theoretical guarantees, and once that it has found a promising region, it requires a large number of

simulation evaluations to improve the local accuracy. The two aims of this work are: global exploration with limited number of iterations while preserving the theoretical convergence guarantees. In this purpose, CE is used as a step in the **Mads** algorithm. The present work proposes a way to include a CE global exploration strategy within the **Mads** algorithm.

This paper is divided as follow, Section 2 proposes an overview of the **Mads** and Cross Entropy methods. Section 3 presents an algorithm combining CE and **Mads**. Finally, Section 4 shows the main numerical results comparing the proposed method with other **Mads** type algorithm and state-of-the-art heuristics. Section 5 concludes on future work and on the contributions of this paper.

2 Description of Mads and Cross Entropy algorithms

This Section describes the **Mads** and CE algorithms.

2.1 The Mads constrained optimization algorithm

The present work considers the **Mads** algorithm with the progressive barrier (PB) [3] to handle inequality constraints and with dynamic scaling [7] to handle the varying magnitudes of the variables. The PB uses the constraint violation function [16]

$$h(\mathbf{x}) := \begin{cases} \sum_{j=1}^m (\max\{c_j(\mathbf{x}), 0\})^2 & \text{if } \mathbf{x} \in \mathcal{X} \subset \mathbb{R}^n \\ \infty & \text{otherwise.} \end{cases}$$

The constraints violation function value $h(\mathbf{x})$ is equal to 0 if and only if the point \mathbf{x} belongs to Ω and is nonnegative otherwise.

Mads is a direct search algorithm (see algorithm 1 below). It proceeds iteratively where the blackbox functions are evaluated at some trial points. These points are either accepted as new iterates or rejected, depending on the value of the objective function and of the constraints violations. A key principle of the **Mads** algorithm is that the candidate points may only be chosen on a discretization of the space of variables called the mesh. This discretization is adaptative and its fineness is controlled by the mesh size vector $\boldsymbol{\delta}^k \in \mathbb{R}_+^n$. In its simplest form, the mesh [7] is defined as follows:

$$M^k = V^k + \{\text{diag}(\boldsymbol{\delta}^k)\mathbf{z} : \mathbf{z} \in \mathbb{Z}^n\}$$

where V^k is the cache containing all points visited by the start of iteration k . A point \mathbf{x} belongs to V^k if and only if both the objective function and the constraints were evaluated by the start of iteration k . The first set V^0 may be initialized by the user or by a collection of points generated by LHS for instance.

Each iteration includes three steps. The first is an optional step called the `SEARCH` where various strategies may be used to explore the space of variables. In practice, the `SEARCH` accelerates the convergence to an optimum and it may attempt to escape from local minima. The only rules to follow are that the trial points must remain on the current mesh M^k and that the search terminates in finite time. It is in this step of the algorithm that CE was integrated.

The second step is mandatory and called the `POLL`. It is the where the space of variable is locally explored by following strict rules guaranteeing convergence. The `POLL` is confined to a region delimited by the so called poll size vector $\Delta^k \in \mathbb{R}_+^n$ which is taken such that $\Delta^k \leq \delta^k$. This region is defined as follows:

$$F^k = \{\mathbf{x} \in M^k : |\mathbf{x}_j - \mathbf{x}_j^k| \leq \Delta_j^k, \forall j \in [1, n]\}$$

with \mathbf{x}^k the current incumbent solution . Then, this step only consist to select a positive spanning set $\mathbb{D}_{\Delta^k}^k$ such that

$$P^k = \{\mathbf{x}^k + \delta^k \mathbf{d} : \mathbf{d} \in \mathbb{D}_{\Delta^k}^k\}$$

is a subset of F^k of extent $\mathbb{D}_{\Delta^k}^k$ and to evaluate the objective functions and the different constraints at these points.

Finally, the last step updates the mesh and the poll size vectors at the end of each iteration. The values of both vectors are reduced when an iteration fails to improve the current solution and they are increased or remain at the same value otherwise. Algorithm 1 provides a description of `Mads`, the reader may consult [7] for complete details.

The fundamental convergence result [3] of the `Mads` algorithm states that if the entire sequence of trial points belongs to a bounded set, then there exists an accumulation point \mathbf{x}^* such that the generalized directional derivative $f^\circ(\mathbf{x}^*; \mathbf{d})$ of Clarke [13] is nonnegative in every hypertangent [23] direction \mathbf{d} to the domain Ω at \mathbf{x}^* provided that \mathbf{x}^* is feasible. A similar result holds for the constraint violation function h over the set \mathcal{X} in situations where the iterates never approach the feasible region.

2.2 The Cross Entropy method for continuous optimization

The Cross Entropy method was introduced by Rubinstein in 1997 in the context of a minimization algorithm for estimating probabilities of rare events [41]. Later, it was modified to solve combinatorial optimization problems [42] and then in 2006 to solve continuous problems [27]. The main idea of this method is as follows. First, each optimization problem is transformed into a rare event estimation problem called associated stochastic problem (ASP). For instance, the deterministic problem (1) is transformed as the minimization of the expectation:

$$\min_{\mathbf{x} \in \Omega, \gamma \in \mathbb{R}} P(f(\mathbf{X}) \leq \gamma) = \min_{\mathbf{x} \in \Omega, \gamma \in \mathbb{R}} E(I_{\{f(\mathbf{x}) \leq \gamma\}}) \quad (2)$$

where \mathbf{X} is a random vector and $I_{\{f(\mathbf{x}) \leq \gamma\}}$ is the indicator function. Then, this ASP is tackled efficiently by an adaptive algorithm. This algorithm constructs a sequence of solutions

Algorithm 1: The Mesh Adaptive Direct Search algorithm (Mads)

0. Initialization:
 - A set of starting point: $V^0 \subset \mathbb{R}^n$,
 - An initial poll size vector: Δ^0 ,
 - The iteration counter: $k \leftarrow 0$,
 - The mesh size adjustment parameter: $\tau \in \mathbb{Q} \cap (0, 1)$,
 - Define $\delta_j^0 = \min\{\Delta_j^0, (\Delta_j^0)^2\}$, $\forall j \in [1, n]$.
 1. Search step (optional):
 - Launch the simulation on a finite set S^k of mesh points.
 - If successful, go to 3.
 2. Poll step:
 - Launch the simulation on the set P^k of poll points.
 3. Updates:
 - Update the cache V^{k+1} .
 - If the SEARCH or the POLL were successful:
 - update \mathbf{x}^{k+1} and $\Delta^{k+1} = \tau^{-1} \Delta^k$.
 - Otherwise:
 - $\Delta^{k+1} = \tau \Delta^k$.
 - Set $\delta_j^{k+1} = \min\{\Delta_j^{k+1}, (\Delta_j^{k+1})^2\}$, $\forall j \in [1, n]$.
 - Increase the iteration counter $k \leftarrow k + 1$ and go to 1.
-

which converging to the optimal solution of the ASP. The CE method is composed of two iterative steps:

- generation a sample of random data according to a density of probability;
- density parameters update thanks to the data sampled to create a new sample in the next iteration.

It results that this method often escapes from local minima.

2.2.1 An introductory example

For clarity, consider the example from [27] of minimizing the function:

$$f(x) = -e^{-(x-2)^2} - 0.8e^{-(x+2)^2}, x \in \mathbb{R}. \quad (3)$$

The function f has two local minima and a single global minimum at $x = 2$.

Using a normal distribution the CE procedure is the following:

- Initialization : at the first iteration $k = 0$, a mean $\mu^0 \in \mathbb{R}^n$ and a standard deviation $\sigma^0 \in \mathbb{R}^n$ (with n the dimension of the problem) are arbitrarily chosen. A large value of σ^0 is taken in order to escape from local solutions.

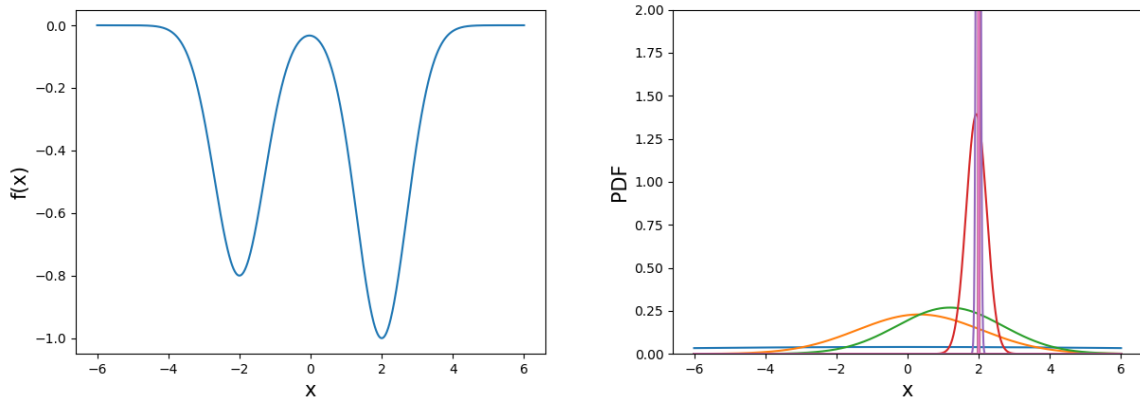


Figure 1: (Figure inspired by [27]) Graph of objective function f (left) and evolution of the normal distribution during the seven first iterations with $\mu_0 = 0$, $\sigma_0 = 10$, $N_e = 10$ and $N_s = 50$ (right).

- Iterative part: at each iteration $k \geq 1$:
 - First, a sample $\mathbf{X}_1, \dots, \mathbf{X}_{N_s}$ of points in \mathbb{R}^n is generated from a normal law $\mathcal{V}(\boldsymbol{\mu}^{k-1}, \boldsymbol{\sigma}^{k-1})$ where N_s is the number of samples.
 - Then, f is evaluated at each sampled points and a number of *elite* points N_e , with the lowest value of f . $\boldsymbol{\mu}^k$ and $\boldsymbol{\sigma}^k$ are the mean and standard deviation of these N_e points.
 - Termination: once the standard deviation becomes sufficiently small, the procedure is stopped.

The sequence of normal distribution is illustrated in the right part of Figure 1. This example shows how the CE procedure escapes from the local minimum at $x = -2$ and converges in seven iterations to the neighborhood of $x = 2$.

2.2.2 The general CE method

Before presenting the algorithm, the ASP is considered and the two iterative steps of the algorithm are precised. Problem 1 is transformed into an ASP. Using a family of probability distribution functions (pdf) $\{g(\cdot; \mathbf{v}) : \mathbf{v} \in \mathcal{V}\}$ where g is the law chosen to sample the different points at each iteration. \mathcal{V} is the set of vector parameters of the pdf g which are calculated at each iteration. In the previous example g is taken as the normal law and the $\mathbf{v}^k \in \mathcal{V}$ is composed of the mean and standard deviation $\mathbf{v}^k = (\boldsymbol{\mu}^k, \boldsymbol{\sigma}^k)$. Having explained the law and its parameters, the ASP related to problem 1 can be defined as follows:

$$\min_{\mathbf{x} \in \Omega, \gamma \in \mathbb{R}} P_{\mathbf{v}}(f(\mathbf{X}) \leq \gamma) = \min_{\mathbf{x} \in \Omega, \gamma \in \mathbb{R}} E_{\mathbf{v}}(I_{\{f(\mathbf{x}) \leq \gamma\}}) \quad (4)$$

where $\mathbf{v} \in \mathcal{V}$ is a vector of parameter, \mathbf{X} is a random vector with a pdf $g(\cdot; \mathbf{v})$ and γ is a variable. At this stage, for a given value of γ , the parameter \mathbf{v} may be estimated. Conversely,

given a vector of parameters \mathbf{v} , the value γ may be also estimated. The CE method is based on these two estimations, at each iterations, the algorithm estimates one then the other. In the iterative part of Example 3, the first item corresponds to the estimation of γ and the second one to the estimation of \mathbf{v} . More precisely, we denote $\gamma^* \in \mathbb{R}$ as the infimum of the objective function, \mathbf{v}^* the parameters and $g(\cdot; \mathbf{v}^*)$ the pdf associated to this infimum. The goal is to generate a sequence (γ^k, \mathbf{v}^k) converging to (γ^*, \mathbf{v}^*) . To achieve this goal, a sequence of pdf $g(\cdot; \mathbf{v}^0), g(\cdot; \mathbf{v}^1), \dots$ converging to $g(\cdot; \mathbf{v}^*)$ is created. To assure the convergence, one must have a “measure” of the difference between the iterate pdf $g(\cdot; \mathbf{v}^k)$ and the objective one $g(\cdot; \mathbf{v}^*)$. The Kullback-Leibler (KL) divergence [28] is used:

$$D(g(\cdot; \mathbf{v}^*) || g(\cdot; \mathbf{v}^k)) = \int_{-\infty}^{\infty} g(\mathbf{x}; \mathbf{v}^*) \ln \left(\frac{g(\mathbf{x}; \mathbf{v}^*)}{g(\mathbf{x}; \mathbf{v}^k)} \right) d\mathbf{x}. \quad (5)$$

The iterative steps may now be described. ρ is defined as a very small quantity, corresponding to the proportion of elite points which are kept from an iteration to another. The procedure is:

- **Adaptive update of γ^k .** With a fixed parameter of pdf \mathbf{v}^{k-1} , γ^k is defined such that it is the $(1 - \rho)$ -quantile of $f(\mathbf{X})$ under v_{k-1} . Then, γ^k satisfies:

$$P_{\mathbf{v}^{k-1}}(f(\mathbf{X}) \leq \gamma^k) \geq \rho, \quad (6)$$

$$P_{\mathbf{v}^{k-1}}(f(\mathbf{X}) \geq \gamma^k) \geq 1 - \rho \quad (7)$$

where $X \sim g(\cdot; \mathbf{v}^{k-1})$. The γ^k is denoted $\hat{\gamma}^k$. To obtain this estimator, a sample $\mathbf{X}_1, \dots, \mathbf{X}_{N_s}$ is drawn from $g(\cdot; \mathbf{v}^{k-1})$ and evaluated. Then, the $(1 - \rho)$ quantile is:

$$\hat{\gamma}^k = f_{\lceil (1-\rho)N_s \rceil}. \quad (8)$$

- **Adaptive update of \mathbf{v}^k .** With a fixed γ^k and knowing \mathbf{v}^{k-1} , \mathbf{v}^k is a solution of:

$$\begin{aligned} \max_{\mathbf{v}} D(\mathbf{v}) &= \max_{\mathbf{v}} E_{\mathbf{v}^{k-1}} \left(I_{\{f(\mathbf{X}) \leq \gamma^k\}} \ln(g(\mathbf{X}; \mathbf{v})) \right) \\ &= \min_{\mathbf{v}} E_{\mathbf{v}^{k-1}} \left(I_{\{f(\mathbf{X}) \leq \gamma^k\}} \ln \left(\frac{I_{\{f(\mathbf{X}) \leq \gamma^k\}}}{g(\mathbf{X}; \mathbf{v})} \right) \right) \end{aligned} \quad (9)$$

which is the minimization of the KL divergence at iteration k (with the convention $0 \ln(0) = 0$). Nevertheless, in practice, the real expectation and the real γ^k are not known, estimators must be used and the following equation is solved:

$$\tilde{\mathbf{v}}^k \in \operatorname{argmax}_{\mathbf{v}} \hat{D}(\mathbf{v}) = \frac{1}{N_s} \sum_{i=1}^{N_s} I_{\{f(\mathbf{X}_i) \leq \hat{\gamma}^k\}} \ln(g(\mathbf{X}_i; \mathbf{v})). \quad (10)$$

Last but not least, $\hat{\mathbf{v}}^k$ is not set to $\tilde{\mathbf{v}}^k$. There are two reasons for that: first the value of $\tilde{\mathbf{v}}^k$ are smoothed. Second, some components of $\tilde{\mathbf{v}}^k$ could be set to 0 or 1 at the first few iterations and the algorithm could converge to non optimal solution. To avoid these problems, the authors of [27] propose to use the following convex combination:

$$\hat{\mathbf{v}}^k = \alpha \tilde{\mathbf{v}}^k + (1 - \alpha) \hat{\mathbf{v}}^{k-1} \quad (11)$$

with $0 < \alpha \leq 1$. Theoretically, any distribution converging in the neighborhood where the global maximum is attained can be used including normal, exponential or beta distribution. Nevertheless, in practice, the updating step is quite simple with the normal distribution that is often chosen. Therefore, the detailed algorithm is presented next:

Algorithm 2: The Cross Entropy (CE) algorithm with normal law

Choose $\hat{\boldsymbol{\mu}}^0 \in \mathbb{R}^n$ and $\hat{\boldsymbol{\sigma}}^0 \in \mathbb{R}^n$

Set the iteration counter: $k \leftarrow 0$.

N_s number of sampled data at each iteration

N_e number of elite population

α the parameter of convex combination

1. Estimation of γ^k :

Generate a random sample $\mathbf{X}_1, \dots, \mathbf{X}_{N_s}$ from $N(\hat{\boldsymbol{\mu}}^{k-1}, \hat{\boldsymbol{\sigma}}^{k-1})$ distribution.

Evaluation of the N_s points by the simulation and then go to 2.

2. Estimation of mean and standard deviation

Let E^k be the indices of the N_e best performing samples.

$$\text{Set } \tilde{\boldsymbol{\mu}}^k = \frac{1}{N_e} \sum_{j \in E^k} \mathbf{X}_j$$

$$\text{and } [\tilde{\boldsymbol{\sigma}}^k]_i = \sqrt{\frac{1}{N_e - 1} \sum_{j \in E^k} [\mathbf{X}_j - \tilde{\boldsymbol{\mu}}^k]_i^2}, \quad \forall i \in [1, n]$$

3. Updates:

Apply the convex combinations:

$$\hat{\boldsymbol{\mu}}^k = \alpha \tilde{\boldsymbol{\mu}}^k + (1 - \alpha) \hat{\boldsymbol{\mu}}^{k-1}$$

$$\hat{\boldsymbol{\sigma}}^k = \alpha \tilde{\boldsymbol{\sigma}}^k + (1 - \alpha) \hat{\boldsymbol{\sigma}}^{k-1}$$

Increase the iteration counter $k \leftarrow k + 1$ and go to 1.

3 The CE-MADS constrained optimization algorithm

This section presents the CE-inspired SEARCH step of Mads. Section 3.1 describes how to handle constraints, the update of the mean $\boldsymbol{\mu}$ and the standard deviation $\boldsymbol{\sigma}$ and the condition to enter the CE search step. Section 3.2 presents the algorithm of the CE SEARCH step.

3.1 The CE-search step

3.1.1 Handling the constraints

Section 2 presented the CE method for unconstrained optimization. In [27], the bound constrained case is treated using a truncated normal law and a penalty approach is used for inequality constraints. In our work, the truncated normal law is also used to treat the bound constraints. For general inequality constraints, the algorithm does not use the penalty approach. In fact, when the algorithm chooses the elite sample, it uses the following function **Best** (defined in [9] and recalled here). Thus, any points in the cache may be selected even if its value of constraint violation is over the threshold of progressive barrier [3]. The definition relies on both the objective and the constraint violation functions f and h .

Definition 3.1. *The function $\text{Best} : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{R}^n$*

$$\text{Best}(\mathbf{x}, \mathbf{y}) = \begin{cases} \mathbf{x} & \text{if } \mathbf{x} \text{ dominates } \mathbf{y} \text{ or if } h(\mathbf{x}) < h(\mathbf{y}), \\ \mathbf{y} & \text{if } \mathbf{y} \text{ dominates } \mathbf{x} \text{ or if } h(\mathbf{y}) < h(\mathbf{x}), \\ \text{Older}(\mathbf{x}, \mathbf{y}) & \text{Otherwise} \end{cases}$$

returns the best of two points.

The function **Older** gives the point which was generated before the former one. Thanks to this definition, CE may treat the general inequality constraints with the terminology used in **Mads**.

3.1.2 Update of the mean and standard deviation

Three elements differ compared to classical CE method concerning the mean and the standard deviation. First, the elite points taken to compute the mean and the standard deviation are not only the N_s points generated by the normal law. The elite points are chosen from the cache at the iteration k , denotes $V^k \subset \mathbb{R}^n$, so any points generated by the **Mads** algorithm may be selected. This set is ordered with the function **Best**, in order to select the N_e elite points, it is sufficient to take the N_e first points of V^k .

Second, the mean and the standard deviation initialization procedure differs from the CE method proceeds. Indeed, **Mads** always begins with a starting point, thus there is at least one point in the cache (the set of evaluated points). Moreover, to avoid generating trial points far from the current solution, bounds are added to the problem as follows (using $\bar{\mathbf{x}}^k$ the poll center at iteration k):

$$(\ell_i^k, u_i^k) = \begin{cases} (\ell_i, u_i) & \text{if } \ell_i \neq -\infty \text{ and } u_i \neq \infty, \\ (\bar{x}_i^k - 10 \times \Delta_i^k, u_i) & \text{if } \ell_i = -\infty \text{ and } u_i \neq \infty, \\ (\ell_i, \bar{x}_i^k + 10 \times \Delta_i^k) & \text{if } \ell_i \neq -\infty \text{ and } u_i = \infty, \end{cases} \quad \forall i \in [1, n]. \quad (12)$$

Once the problem has finite bound constraints, there are two cases to calculate the mean $\boldsymbol{\mu}^k \in \mathbb{R}^n$ and the standard deviation $\boldsymbol{\sigma}^k \in \mathbb{R}^n$:

- In case where the number of points in the cache is too small to be relevant, i.e. fewer points than the number N_e required, then the mean and the standard deviation are determined such that:

$$\boldsymbol{\mu}^k = \bar{\mathbf{x}}^k \quad (13)$$

$$\boldsymbol{\sigma}^k = 2(\mathbf{u}^k - \boldsymbol{\ell}^k) \quad (14)$$

- In the others cases, the same calculations are made that in the original CE process:

$$\boldsymbol{\mu}^k = \frac{1}{N_e} \sum_{j \in E^k} \mathbf{X}_j$$

$$[\boldsymbol{\sigma}^k]_i = \sqrt{\frac{1}{N_e - 1} \sum_{j \in E^k} [\mathbf{X}_j - \boldsymbol{\mu}^k]_i^2}, \quad \forall i \in [1, n].$$

Third, to generate the point during the CE search, the truncated normal law was always used with the bounds created in (12). Moreover, the elite points come not only from the previous normal sampling but also of the other kind of search step. That gives a vector of standard deviation which tends to zero very quickly, the other methods doing generally a local search. That is why, the standard deviation is calculated as in 11 with a coefficient $\alpha = 0.7$.

3.1.3 The condition to pass in the CE-SEARCH step

The goal of the CE method is to explore in few evaluations the space to determine the promising region. The number of evaluations used by the CE SEARCH step must be quite small. For this purpose, **Mads** does not perform the CE SEARCH step at every iteration. The standard deviation can be seen as a measure of the incertitude on the data and is used to determine whether to launch the SEARCH step or not. First, a new variable called $\boldsymbol{\sigma}^p$ is introduced, it represents the incertitude measured the last time the algorithm passed through the CE SEARCH step and generated trial points. This variable is initialized to ∞ . Then, the condition to launch the CE SEARCH is the following:

$$\|\boldsymbol{\sigma}^k\| < \|\boldsymbol{\sigma}^p\| \quad (15)$$

This conditions means that the current incertitude is smaller than the previous one. Each time this conditions is respected, $\boldsymbol{\sigma}^p$ is updated with the standard deviation obtained after the CE step. Last but not least, there is a special case. The CE method being associate with **Mads** which is a local search, it is possible that the points become rapidly close to each others, reducing the standard deviation. In some cases, that avoids to escape from unfeasible region. Thus, in case where several iterations of **Mads** algorithm are passed and the feasible region is still not reached, then the CE- SEARCH is launched with a mean equal to the current best point and a standard deviation equal to 2 times the initial standard deviation until a feasible point is found.

3.2 The complete algorithm

The CE SEARCH step of Mads algorithm is presented here:

Algorithm 3: The CE SEARCH step

1. Calculation of $\boldsymbol{\mu}^k$ and $\boldsymbol{\sigma}^k$:

if $\text{card}(V^k) < N_e$:

$$\begin{aligned} \boldsymbol{\mu}^k &= \bar{\mathbf{x}}_0 \\ \boldsymbol{\sigma}^k &= 2(\mathbf{u}^k - \boldsymbol{\ell}^k) \end{aligned}$$

else:

$$\boldsymbol{\mu}^k = \frac{1}{N_e} \sum_{j \in E^k} \mathbf{X}_j$$

$$[\boldsymbol{\sigma}^k]_i = \sqrt{\frac{1}{N_e - 1} \sum_{j \in E^k} [\mathbf{X}_j - \boldsymbol{\mu}^k]_i^2}, \quad \forall i \in [1, n]$$

2. CE SEARCH

If $\|\boldsymbol{\sigma}^k\| < \|\boldsymbol{\sigma}^p\|$:

Generate a random sample $\mathbf{X}_1, \dots, \mathbf{X}_{N_s}$ from $\mathcal{N}(\boldsymbol{\mu}^k, 2\boldsymbol{\sigma}^k)$ distribution and project them on the mesh.

Evaluation of the N_s points by the simulation.

Update:

$$\boldsymbol{\mu}^{k+1} = \frac{1}{N_e} \sum_{j \in E^k} \mathbf{X}_j$$

$$[\boldsymbol{\sigma}^{k+1}]_i = \sqrt{\frac{1}{N_e - 1} \sum_{j \in E^k} [\mathbf{X}_j - \boldsymbol{\mu}^{k+1}]_i^2}$$

$$(\boldsymbol{\sigma}^p)^2 = (\boldsymbol{\sigma}^{k+1})^2$$

4 Computational experiments

The present work uses data profiles to compare the different algorithm. Data profiles [36] allow to assess if algorithms are successful in generating solution values close to the best objective function values. To identify a successful run, a convergence test is required. Let denote \mathbf{x}_e the best iterates obtained by one algorithm on one problem after e evaluations, f_{fea} a common reference for a given problem obtained by taking the max feasible objective function values on all run instances of that problem for all algorithms and f^* the best solution obtained by all tested algorithms on all run instances of that problem. Then, the problem is said to be solved within the convergence tolerance τ when:

$$f_{fea} - f(\mathbf{x}_e) \geq (1 - \tau)(f_{fea} - f^*).$$

Different initial points constitute different problems. Moreover, an instance of a problem corresponds to a particular pseudo-random generator seeds. The horizontal axis of a data profile

represents the number of evaluations for problems of fixed dimension, and represents group of $n + 1$ evaluations when problems of different dimension are involved. The vertical axis corresponds to the proportion of problems solved within a given tolerance τ . Each algorithm has its curve to allow comparison of algorithms capability to converge to the best objective function value.

This section presents the numerical experiments. It is divided in two subsections. The numerical experiments of Section 4.1 are performed on analytical test problems to calibrate the CE-SEARCH parameters. Section 4.2 compares CE-Mads with others state-of-the-art global optimization method. Finally, section 4.3 compares Mads, LH-Mads, VNS-Mads and CE-Mads without the use of models on three real engineering problems.

4.1 Preliminary experiments to calibrate parameters

Computational experiments are conducted using the version 3.9.1 of NOMAD [29] software package. All tests use the Mads strategy with the use of the NM search [9] and without the use of models [6] [14]. When the CE-SEARCH is used, it is the first SEARCH step to be applied.

Numerical experiments on analytical test problems are conducted to set default values for the two algorithmic parameters: the number of sampled data at each iteration N_s and the number of elite population N_e . CE-Mads is tested on 100 analytical problems from the optimization literature. The characteristics and sources of these problems are summarized in Table 1 in appendix A. The number of variables ranges from 2 to 60; 28 problems have constraints other than bound constraints. In order to have a more precise idea of the effect between the hyper-parameters (n_e and n_s), three series of tests are conducted:

- A series of tests on the 69 unconstrained test problems having a dimension from 2 to 20.
- A series of tests on the 25 constrained test problems having a dimension from 2 to 20.
- A series of tests on the 6 larger problems in term of dimension (from 50 to 60), three are constrained and three are not.

For each test, the maximal number of function evaluations is set to $1000(n+1)$, where n is the number of variables and each problem is run with 3 different random seeds. For each series of tests, the five following CE-MADS setup of hyper-parameters are compared: $(N_e, N_s) \in \{(2, n), (4, 2n), (6, 3n), (8, 4n), (10, 5n)\}$ with n the dimension of the test problem. A run called NOMAD default is added in each series of test to compare our results with the current NOMAD software. Data profiles are presented on Figure 2, 3 and 4 with different values of the tolerance τ .

These results are analysed by series of problems:

- On the unconstrained problems (see Figure 2), no algorithm really stands out regardless of the value of τ , it is difficult to choose one hyper-parameter rather than another one even if the couple $N_e = 4$ and $N_s = 2n$ appears to be more efficient.

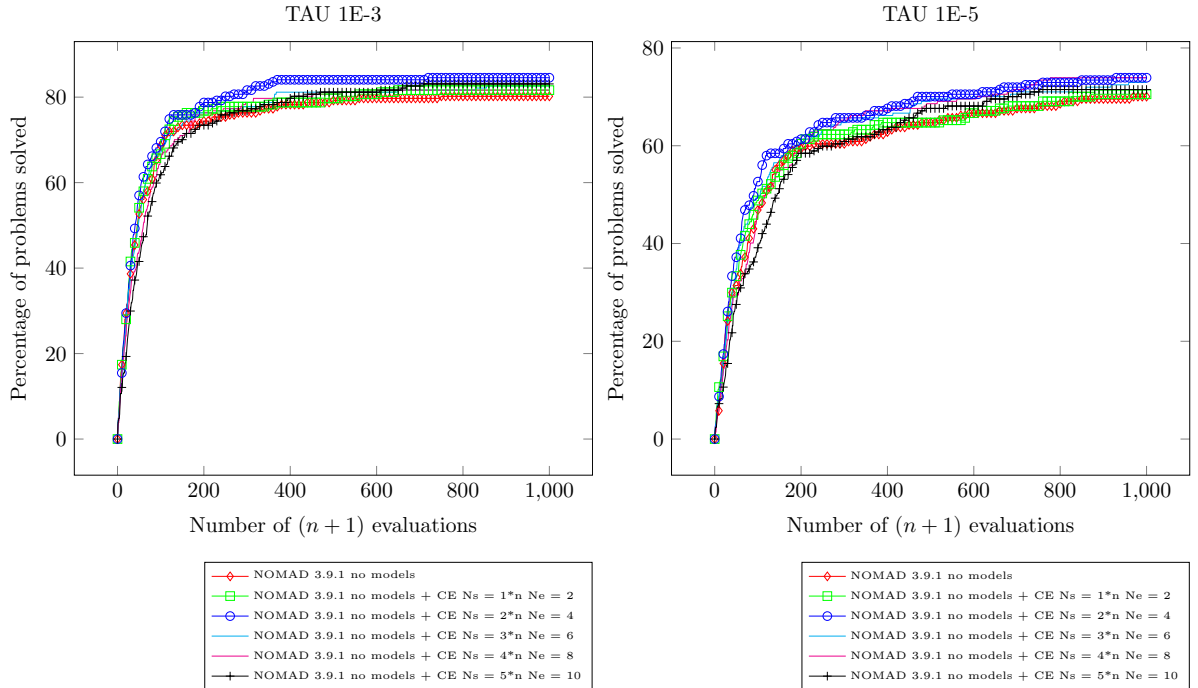


Figure 2: Result of calibration of the hyper-parameters of CE-MADS on the 69 unconstrained test problems

- On the constrained problems (see Figure 3), there are different behaviors according to the value of τ . For $\tau = 10^{-3}$, no algorithm appears to be dominant. However, for $\tau = 10^{-5}$, it happens that greater are the values of N_s and N_e , higher is the percentage of problems solved finally. That can be explained because great N_s and N_e allow a better exploration of the space, and so a more precise result at the end.
- Finally (see Figure 4), on the large test problems, and for small values of the tolerance τ the CE-Mads is outperformed by the Mads algorithm with default values. It seems that the CE method is not useful for problems with a large number of variables.

Inspection of the logs of the hyper-parameter calibration reveals the two following observations:

- The CE-Mads performance is not very sensitive to the hyper-parameter values. This allows to avoid some calibration experiments before applying the algorithm on a new test problem.
- For problems with a large number of variables, our tests suggest to avoid using of the CE-SEARCH. Nevertheless, this point has not been confirmed on real engineering problems given that we do not have access to engineering test problems with large dimension.

In the remainder of the paper, the CE-SEARCH values are set to $N_e = 4$ and $N_s = 2n$ as they often perform well.

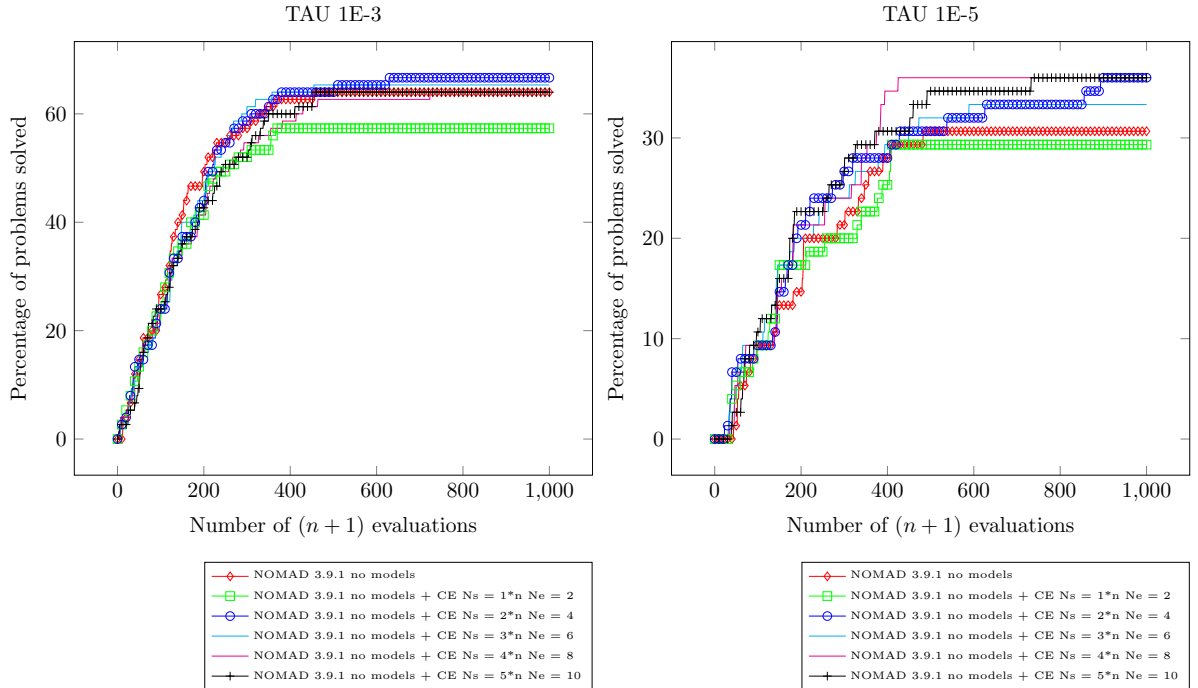


Figure 3: Result of calibration of the hyper-parameter of CE-MADS on the 25 constrained test problems

4.2 Comparisons between CE-Mads and some of state-of-the art global optimization heuristics

In this section, we compared CE-Mads with other well-known global optimization heuristics designed to escape local minima on a collection of unconstrained global optimization benchmark problems. The framework used to compare the CE-Mads algorithm with others global optimization method is pymoo [10]. This framework proposes a variety of global optimization algorithm. The CE-Mads method is compared to four of them :

- Genetic Algorithm [18], a method based on biological inspired operators such as mutation, cross-over and selection. No special advice on the hyper-parameter are given in the pymoo framework, however after few tests, it seems that a population size of 40 performs well. We run the different tests with this value.
- Differential Evolution [38], a method which combines evolutionary strategies with geometrical search techniques. In the pymoo framework, the authors advise to test with the following settings: the crossover constant $CR = 0.9$, the select weighting factor $F = 0.8$ and the method is “DE/rand/1/bin”. The size of the population is set to 20 which seems to perform well.
- Covariance Matrix Adaptation-Evolutionary Strategy (CMA-ES) [20], a method based also on biological inspired operators. Its name comes from the adaptation of the co-

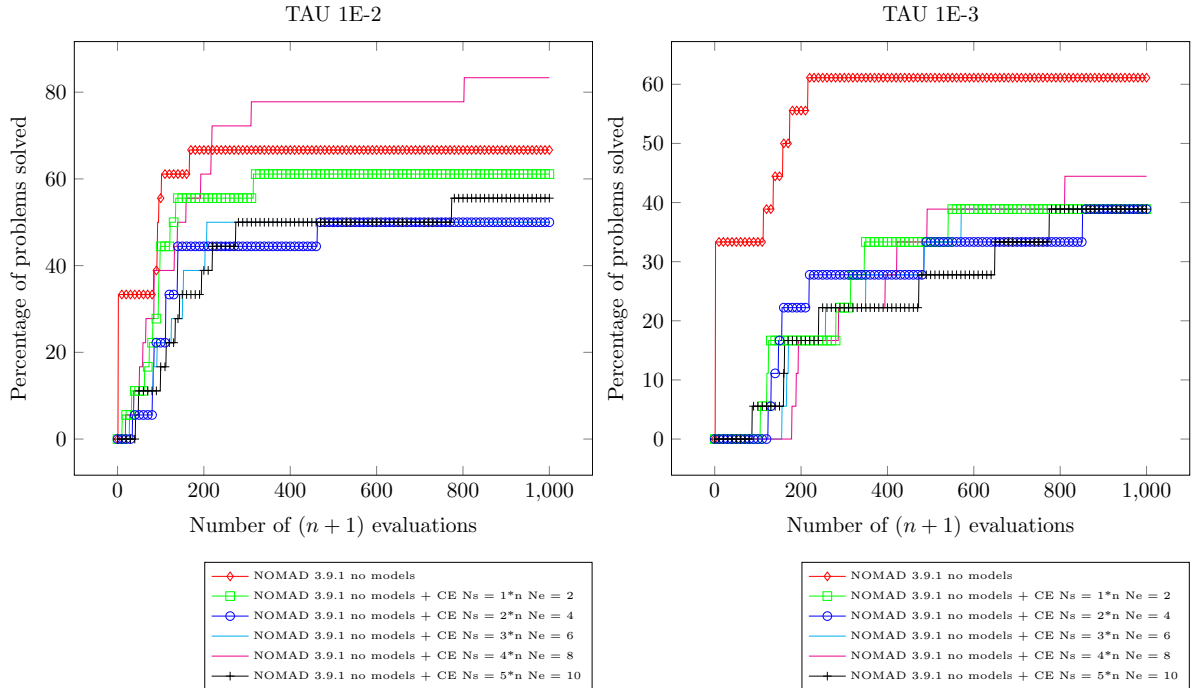


Figure 4: Result of calibration of the hyper-parameter of CE-MADS on the 6 large test problems

variance matrix of the multivariate normal distribution used during the mutation. The setting used for the tests are the default setting.

- Particle Swarm Optimization [25], a method inspired by the birds movement and more generally on the collaboration between the individuals. No indication are given in the pymoo framework but it seems that a population size of 15 performs well.

To compare the different algorithms, we use the test problem common between the global optimization benchmark problems of pymoo and the problems provided in the appendix A. That gives 19 unconstrained test problems (marked in with an asterisk in the appendix) and we run the algorithms with five different seeds in order to reduce the effect of randomness. The maximum number of function evaluations is fixed to 3000. Results are provided on Figure 5. The MADS-type algorithms are more efficient than the heuristic of global optimization on this test set. The heuristic’s performances are comparable when the required accuracy is 10^{-3} . If a higher accuracy is desired, it would seem that CMA-ES is the more appropriate method. The gap between heuristics and MADS-type algorithms seems to widen when greater precision is required. This is normal considering that MADS is a local search algorithm originally. The interest of CE-MADS stands out since it allows to combine both a global and a local search, which explains its better performance.

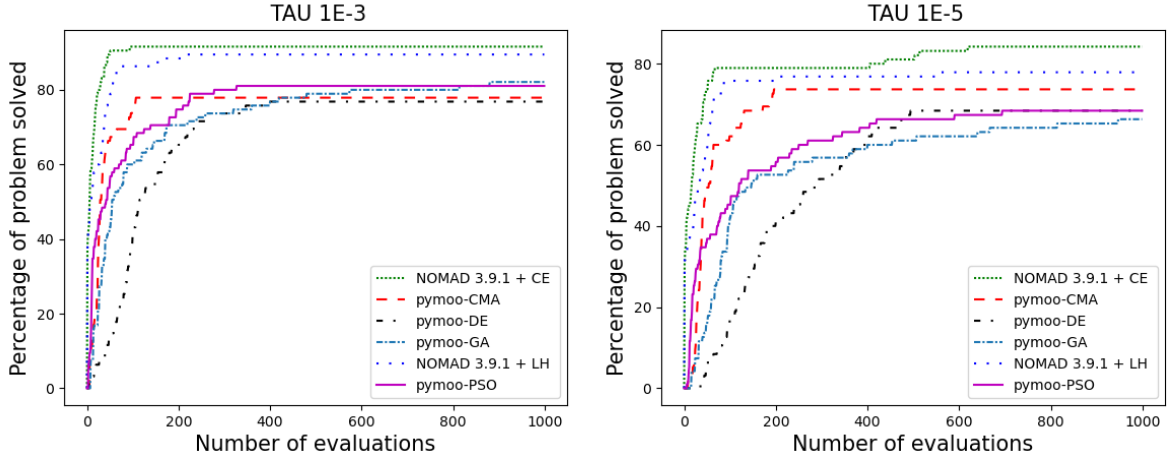


Figure 5: Result on the 19 global optimization test problems between CE-Mads, LH-Mads, GA, DE, CMA-ES and PSO for $\tau = 10^{-3}$ (left) and $\tau = 10^{-5}$ (right).

4.3 Test on engineering problems

In this section, the CE-Mads algorithm is tested on three different engineering problems and compared to three algorithms: the Mads-default (without models), the VNS-Mads where a VNS-SEARCH is used and the LH-Mads which is a default Mads with in addition a LHS search. The comparison with the two last algorithms is crucial because they are methods aiming to explore the space of design variables. The Latin Hypercube SEARCH strategy is used with two parameters $n_{init} = 100$ and $n_{iter} = 10$: n_{init} is the number of LH trial points generated at the first iteration of Mads and n_{iter} the number of LH trial points generated at each subsequent iteration. The Variable Neighbour Search is used with the default parameters [1]. It is an metaheuristic allowing to explore distant neighborhoods of the current incumbent solution.

4.3.1 The MDO problem

The Mads-default (no models), CE-Mads, VNS-Mads and LH-Mads are tested to solve a simple multidisciplinary wing design optimization problem [17]. Each initial point defines a MDO problem. Solving the problem consists in maximizing the range of an aircraft subject to 10 general constraints. The problem has 10 scaled design variables bounded in $[0; 100]$. Figure 6 shows the result on a data profile when solving 20 MDO problems on different initial points using 3000 function evaluations or less. The initial points are real randomly selected within the bounds. Each run is done with three different seeds in order to minimize the impact of the seed.

Figure 6 shows that the CE-Mads outperforms the other algorithms for all values of τ . Given that the computational time of engineering test problem is relatively low, the comparison between the heuristics and the MADS-type algorithm can be done. The inequality constraints are handled by the default setting in pymoo which is a penalization method. Given

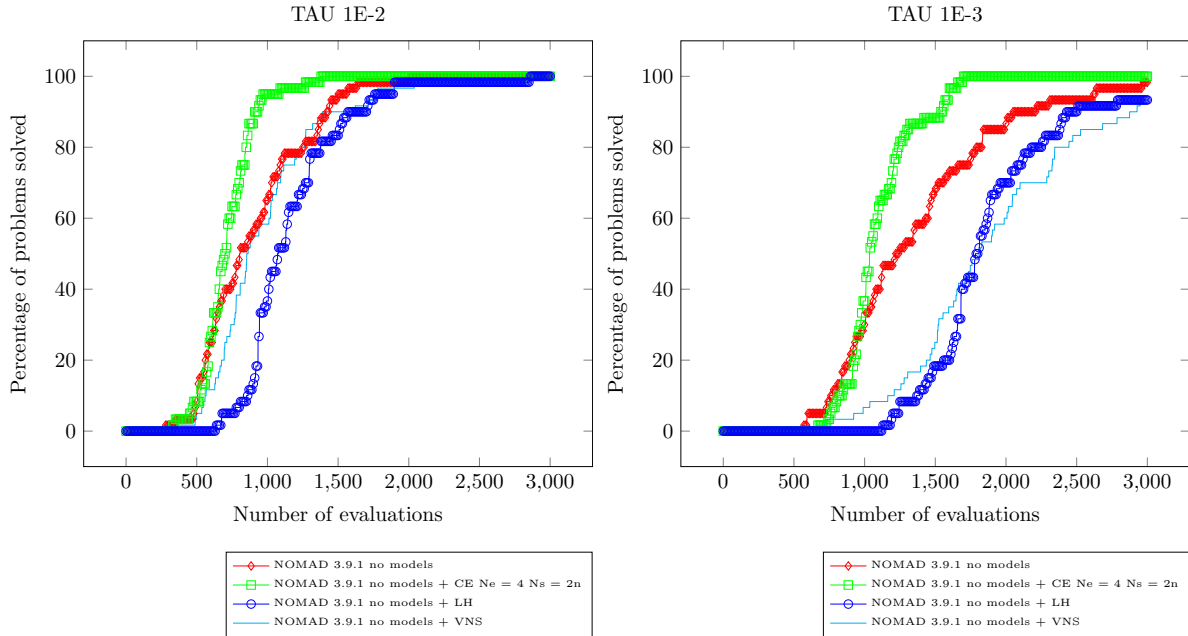


Figure 6: Result on the 60 MDO instances between Mads (no models), CE-Mads, VNS-Mads and LH-Mads for $\tau = 10^{-2}$ (left) and $\tau = 10^{-3}$ (right).

that Mads and VNS-Mads require a starting point to be run which can impact their performance, they are not used in this comparison. Therefore, we do not use any starting point but the different algorithms are run with 20 different seeds. The result are given on the figure 7. The heuristics perform poorly compared to the Mads-type algorithm. That is not surprising given that Mads benefits of a specialized method to handle the constraints and is specialized in blackbox optimization. The tests on the other engineering test problems are not presented given the great computational time required and the even harder optimization of the blackbox.

4.3.2 The STYRENE problem

The Mads-default (no models), CE-Mads, VNS-Mads and LH-Mads algorithms are tested to optimize a styrene production process [1]. This problem is a simulation of a chemical process. This process relies on a series of interdependent calculation of blocks using common numerical tools as Runge-Kutta, Newton, fixed point and also chemical related solver. The particularity of this problem is the presence of “hidden” constraints, i.e. sometimes the process does not finish and just return an error. In the case where the chemical process ends, the constraints (not hidden) and the objective functions may be evaluated during a post-processing. The objective is to maximize the net value of the styrene production process with 9 industrial and environmental regulations constraints. In this work, a STYRENE problem possesses eight independent variables influencing the styrene production process. The variables considered during the optimization process are all scaled and bounded in $X = [0, 100]^8$.

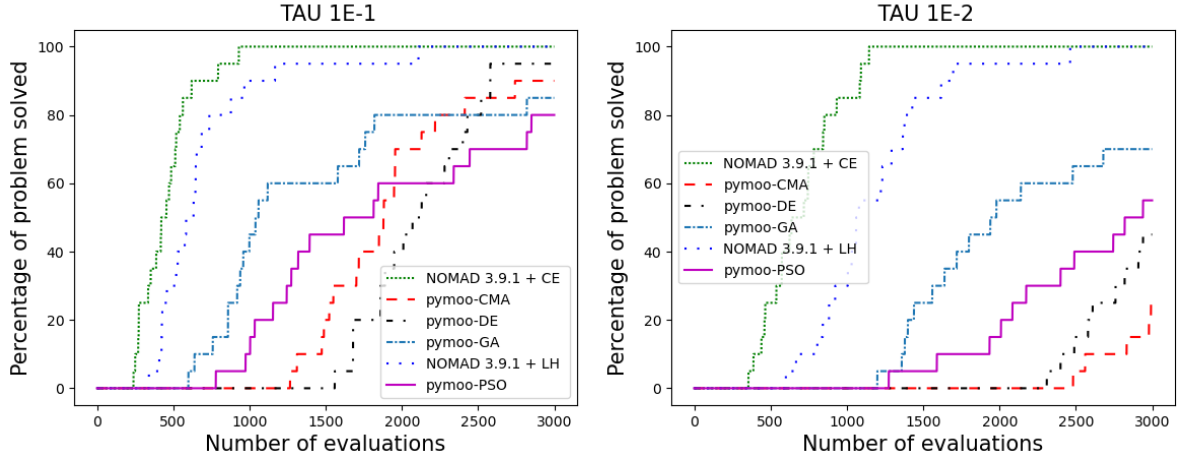


Figure 7: Result on the 20 MDO instances between CE-Mads, LH-Mads, GA, DE, CMA-ES and PSO for $\tau = 10^{-1}$ (left) and $\tau = 10^{-2}$ (right).

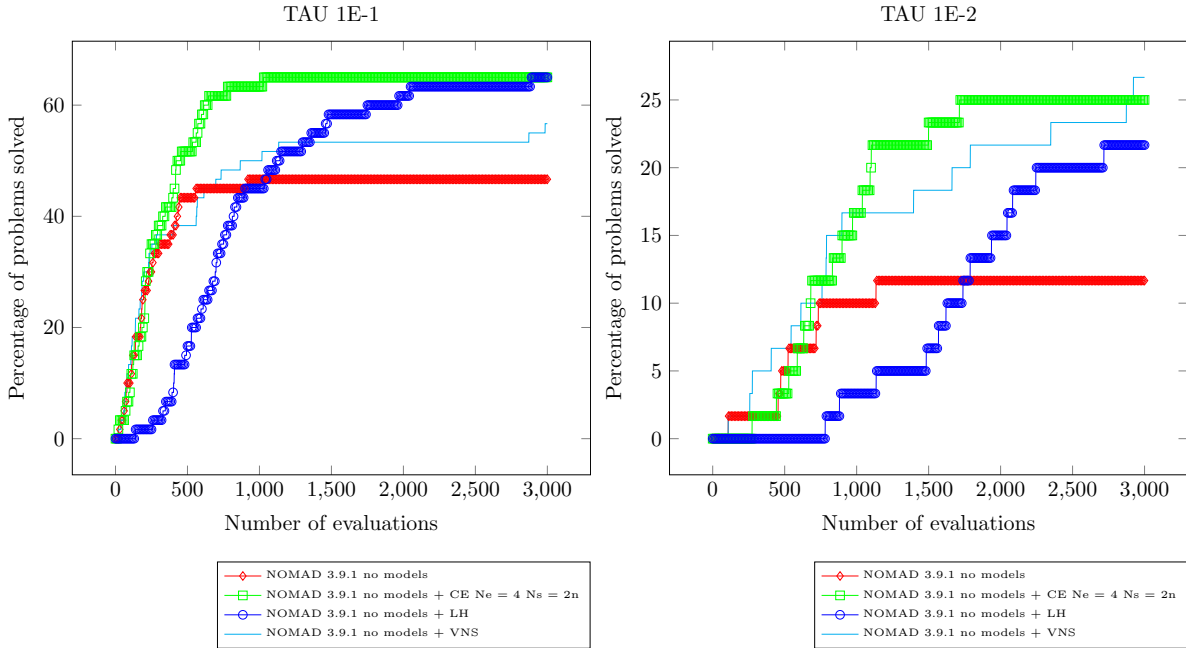


Figure 8: Result on the 60 STYRENE instances between Mads (no models), CE-Mads and LH-Mads for $\tau = 10^{-1}$ (left) and $\tau = 10^{-2}$ (right).

As it was done for the MDO test problems, the four algorithms are tested with 20 different starting points taken in \mathcal{X} . A maximal number of evaluations of 3000 is used and each problem is run with three different seeds. The STYRENE problems is particularly interesting in this study, because there are two minima as it is shown in [9]. The results with $\tau = 10^{-1}$ allow to know the percentage of problems having found the global minimum. The results are provided on Figure 8. On the left plot, it is interesting to notice that the CE-Mads algorithm

find the global minimum the same number of times that the LH-Mads algorithm but is more efficient. On the right plot, the CE-Mads algorithm seems to have the same accuracy that the VNS-Mads algorithm and is slightly more efficient.

4.3.3 The LOCKWOOD problem

Finally, the Mads default (no models), LH-Mads and CE-Mads algorithms without quadratic models are tested to solve the basic version of a pump-and-treat groundwater remediation problem from Montana Lockwood Solvent Groundwater Plume Site [31]. The problem has 6 design variables bounded in $X = [0, 20000]^6$ and 4 constraints. A particularity of this problem is that each simulation run take several seconds, so the maximum number of blackbox evaluations is set to 1500. The algorithms are started from 20 different randomly selected initial points in X and three different seeds are used as previously. The results are provided on Figure 9. In this problem, reach the feasible region is not easy. Here again, the results

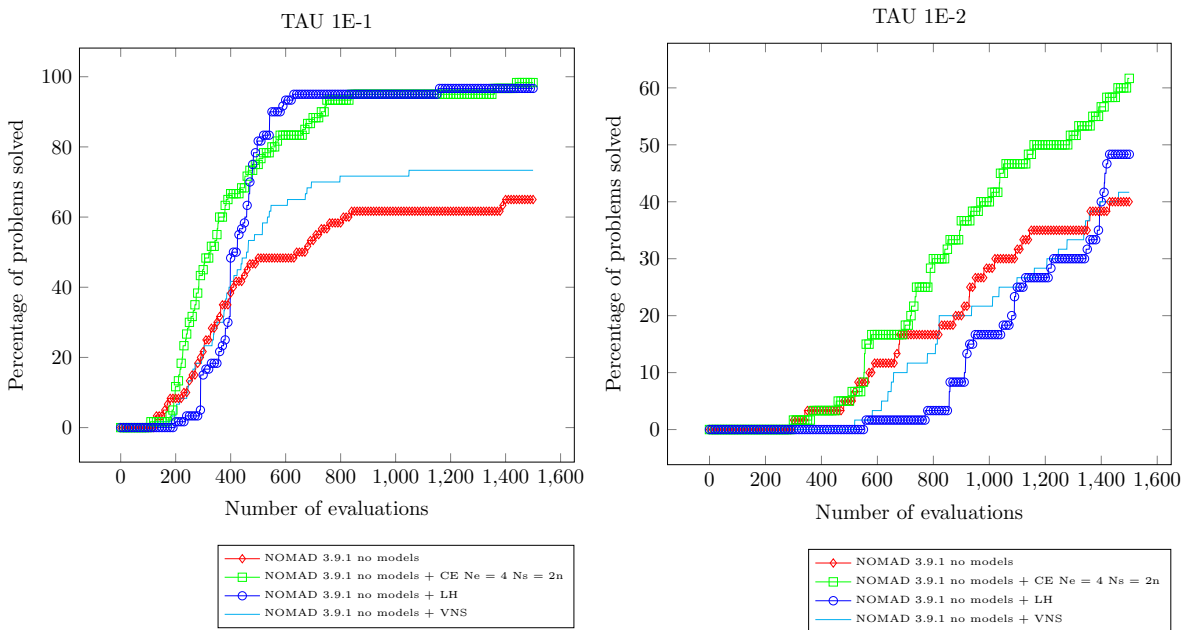


Figure 9: Result on the 20 LOCKWOOD instances between Mads-default (no models), CE-Mads, VNS-Mads and LH-Mads for $\tau = 10^{-1}$ (left) and $\tau = 10^{-2}$ (right).

at $\tau = 10^{-1}$ allows to give an idea of the number of times the algorithm reach the feasible region. For instance, CE-Mads and LH-Mads always reach the feasible region while Mads default reaches the feasible only 41 times on 60 instances and VNS-Mads only 46 times. The efficiency of CE-Mads and LH-Mads is comparable. However, on the right plot, a better accuracy is reached with a greater efficiency by the CE-Mads algorithm.

5 Discussion

This paper introduces a way to combine the CE algorithm and the **Mads** algorithm so that it allows a better space exploration. This is achieved by defining a CE-SEARCH step within the **Mads** algorithm. The CE-SEARCH generates some points according to a normal distribution whose mean and standard deviation is calculated from the best points stored in the cache. This approach allows to handle the constraints in a different ways. Moreover, the particularity of this SEARCH is that it is not performed at each iteration of the **Mads** algorithm, but according to a criterion based on the value of the norm of the standard deviation of the best points.

Numerical experiments show that in case where the problem has different minima or a feasible region hard to reach, the CE-**Mads** algorithm performs well. Indeed, it attains as often as the LH-**Mads** the feasible region or the global minimum but it is far more efficient, especially when a tight accuracy is considered. Moreover, even on problem, as MDO, where the classical exploration SEARCH, LH and VNS, does not work well, the CE-**Mads** algorithm gives interesting results. Finally, comparison with other global algorithms has been made, two conclusions are drawn. First, CE-**Mads** works better than the heuristics on the unconstrained global optimization test problems. That shows its real ability to escape local optima. Second, CE-**Mads** outperforms the heuristic on the engineering problems, which is not particularly relevant because it benefits to the **Mads** ability of performing well on this kind of problems.

Further works will be devoted to improve the link between the **Mads** algorithm and the CE algorithm by adjusting the size of the mesh with the standard deviation calculated in CE.

A Appendix

#	Name	Source	n	m	Bnds	#	Name	Source	n	m	Bnds
1	ARWHEAD10	[19]	10	0	no	51	MXHILB	[30]	50	0	no
2	ARWHEAD20	[19]	20	0	no	52	OPTENG_RBF	[26]	3	4	yes
3	BARD	[35]	3	0	no	53	OSBORNE1	[35]	5	0	no
4	BDQRTIC10	[19]	10	0	no	54	OSBORNE2	[30]	11	0	no
5	BDQRTIC20	[19]	20	0	no	55	PBC1	[30]	5	0	no
6	BEALE*	[35]	2	0	no	56	PENALTY1_4*	[19]	4	0	no
7	BIGGS	[19]	6	0	no	57	PENALTY1_10*	[19]	10	0	no
8	BOX	[35]	3	0	no	58	PENALTY1_20*	[19]	20	0	no
9	BRANIN*	[21]	2	0	yes	59	PENALTY2_4*	[19]	4	0	no
10	BROWNAL5	[19]	5	0	no	60	PENALTY2_10*	[19]	10	0	no
11	BROWNAL7	[19]	7	0	no	61	PENALTY2_20*	[19]	20	0	no
12	BROWNAL10	[19]	10	0	no	62	PENTAGON	[30]	6	15	no
13	BROWNAL20	[19]	20	0	no	63	PIGACHE_X00	[37]	4	11	yes
14	BROWNDENNIS	[35]	4	0	no	64	PIGACHE_X01	[37]	4	11	yes
15	BROWN_BS	[35]	2	0	no	65	POLAK2	[30]	10	0	no
16	B250	[11]	60	1	yes	66	POWELL_BS	[35]	2	0	no
17	B500	[11]	60	1	yes	67	POWELLSG4*	[19]	4	0	no
18	CHENWANG_F2_X0	[12]	8	6	yes	68	POWELLSG8	[19]	8	0	no
19	CHENWANG_F2_X1	[12]	8	6	yes	69	POWELLSG12	[19]	12	0	no
20	CHENWANG_F3_X0	[12]	10	8	yes	70	POWELLSG20	[19]	20	0	no
21	CHENWANG_F3_X1	[12]	10	8	yes	71	RADAR	[34]	7	0	yes
22	CRESCENT	[3]	10	2	no	72	RANA*	[24]	2	0	yes
23	DISK	[3]	10	1	no	73	RASTRIGIN*	[21]	2	0	yes
24	DIFFICULT2	[3]	10	0	no	74	RHEOLOGY	[5]	3	0	no
25	ELATTAR	[30]	6	0	no	75	ROSENBROCK*	[35]	2	0	yes
26	EVD61	[30]	6	0	no	76	SHOR	[30]	5	0	no
27	FILTER	[30]	9	0	no	77	SNAKE	[3]	2	2	no
28	FREUDENSTEINROTH*	[35]	2	0	no	78	SPRING_X00	[40]	3	4	yes
29	GAUSSIAN	[35]	3	0	no	79	SPRING_X01	[40]	3	4	yes
30	G2_10	[4]	10	2	yes	80	SROSENBR6	[19]	6	0	no
31	G2_20	[4]	20	2	yes	81	SROSENBR8	[19]	8	0	no
32	G2_50	[4]	50	2	yes	82	SROSENBR10	[19]	10	0	no
33	GOFFIN	[30]	50	0	no	83	SROSENBR20	[19]	20	0	no
34	GRIEWANK*	[21]	10	0	yes	84	TAOWANG_F2_X00	[44]	7	4	yes
35	GULFRD*	[8]	3	0	no	85	TAOWANG_F2_X01	[44]	7	4	yes
36	HELICALVALLEY*	[35]	3	0	no	86	TREFETHEN*	[24]	2	0	yes
37	HS19	[22]	2	2	yes	87	TRIDIA10	[19]	10	0	no
38	HS78	[30]	5	0	no	88	TRIDIA20	[19]	20	0	no
39	HS83_X0	[22]	5	6	yes	89	TRIGONOMETRIC	[35]	10	0	no
40	HS83_X1	[22]	5	6	yes	90	VARDIM8	[19]	8	0	no
41	HS114_X0	[30]	9	6	yes	91	VARDIM10	[19]	10	0	no
42	HS114_X1	[30]	9	6	yes	92	VARDIM20	[19]	20	0	no
43	JENNRICHSAMPSON	[35]	2	0	no	93	WANGWANG_F3	[45]	2	0	yes
44	KOWALIKOSBORNE*	[35]	4	0	no	94	WATSON9	[35]	9	0	no
45	L1HILB	[30]	50	0	no	95	WATSON12	[35]	12	0	yes
46	MAD6_X0	[30]	5	7	no	96	WONG1	[30]	7	0	no
47	MAD6_X1	[30]	5	7	no	97	WONG2	[30]	10	0	no
48	MCKINNON	[32]	2	0	no	98	WOODS4	[19]	4	0	no
49	MEYER*	[35]	3	0	no	99	WOODS12	[19]	12	0	no
50	MEZMONTES	[33]	2	2	yes	100	WOODS20	[19]	20	0	no

Table 1: Description of the set of 100 analytical problems.

Nomenclature

The following list describes symbols used within the body of the document. In what follows, if the symbol is bold then it is a vector otherwise it is a scalar.

ℓ	The lower bound of a decision variable
Δ^k	The frame size parameter at iteration k
δ^k	The mesh size parameter at iteration k
ϵ	The stopping criterion
γ	A parameter to estimate in an associated stochastic problem
u	The upper bound of a decision variable
\mathcal{E}	The expectation
\mathcal{N}	The normal distribution
\mathcal{V}	Set of parameters of a probability density function
\mathcal{X}	The bounded constraints set of type $\ell \leq x \leq u$
μ	The mean
Ω	The feasible set
ρ	A percentage of quantile
σ	The standard deviation
τ	The mesh size adjustment parameter
\mathbf{X}	A random vector
c_j	The j^{th} constraint
D	A positive spanning set
E^k	The set of indices of elite points
F^k	The frame at iteration k
$g(\cdot; \cdot)$	A probability density function
h	The measure of constraints violation
I_x	Indicator function of x

k	The iteration counter
M^k	The mesh at iteration k
n	The dimension of a problem
N_s	Number of sampled data at each iteration
N_e	Number of elite population
V	The cache
v	A parameter of a probability density function

References

- [1] C. Audet, V. Béchar, and S. Le Digabel. Nonsmooth optimization through Mesh Adaptive Direct Search and Variable Neighborhood Search. *Journal of Global Optimization*, 41(2):299–318, 2008.
- [2] C. Audet and J.E. Dennis, Jr. Mesh Adaptive Direct Search Algorithms for Constrained Optimization. *SIAM Journal on Optimization*, 17(1):188–217, 2006.
- [3] C. Audet and J.E. Dennis, Jr. A Progressive Barrier for Derivative-Free Nonlinear Programming. *SIAM Journal on Optimization*, 20(1):445–472, 2009.
- [4] C. Audet, J.E. Dennis, Jr., and S. Le Digabel. Parallel Space Decomposition of the Mesh Adaptive Direct Search Algorithm. *SIAM Journal on Optimization*, 19(3):1150–1170, 2008.
- [5] C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization*. Springer Series in Operations Research and Financial Engineering. Springer International Publishing, Berlin, 2017.
- [6] C. Audet, M. Kokkolaras, S. Le Digabel, and B. Talgorn. Order-based error for managing ensembles of surrogates in mesh adaptive direct search. *Journal of Global Optimization*, 70(3):645–675, 2018.
- [7] C. Audet, S. Le Digabel, and C. Tribes. Dynamic scaling in the mesh adaptive direct search algorithm for blackbox optimization. *Optimization and Engineering*, 17(2):333–358, 2016.
- [8] C. Audet, S. Le Digabel, and C. Tribes. The Mesh Adaptive Direct Search Algorithm for Granular and Discrete Variables. *SIAM Journal on Optimization*, 29(2):1164–1189, 2019.
- [9] C. Audet and C. Tribes. Mesh-based Nelder-Mead algorithm for inequality constrained optimization. *Computational Optimization and Applications*, 71(2):331–352, 2018.
- [10] J. Blank and K. Deb. Pymoo: Multi-objective optimization in python. *IEEE Access*, 8:89497–89509, 2020.
- [11] A.J. Booker, E.J. Cramer, P.D. Frank, J.M. Gablonsky, and J.E. Dennis, Jr. Movars: Multidisciplinary optimization via adaptive response surfaces. AIAA Paper 2007–1927, 2007.
- [12] X. Chen and N. Wang. Optimization of short-time gasoline blending scheduling problem with a DNA based hybrid genetic algorithm. *Chemical Engineering and Processing: Process Intensification*, 49(10):1076–1083, 2010.

- [13] F.H. Clarke. *Optimization and Nonsmooth Analysis*. John Wiley & Sons, New York, 1983. Reissued in 1990 by SIAM Publications, Philadelphia, as Vol. 5 in the series Classics in Applied Mathematics.
- [14] A.R. Conn and S. Le Digabel. Use of quadratic models with mesh-adaptive direct search for constrained black box optimization. *Optimization Methods and Software*, 28(1):139–158, 2013.
- [15] A.R. Conn, K. Scheinberg, and L.N. Vicente. *Introduction to Derivative-Free Optimization*. MOS-SIAM Series on Optimization. SIAM, Philadelphia, 2009.
- [16] R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical Programming, Series A*, 91:239–269, 2002.
- [17] A.A. Giunta. *Aircraft Multidisciplinary Optimization using Design of Experiments Theory and Response Surface Modeling Methods*. PhD thesis, Virginia Tech, Houston, Texas, 1997; available as Tech. Rep. MAD 97-05-01, May 1997, Department of Aerospace and Ocean Engineering, Virginia Tech, 215 Randolph Hall, Blacksburg, Virginia 24061.
- [18] D.E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 1st edition, 1989.
- [19] N.I.M. Gould, D. Orban, and Ph.L. Toint. CUTer (and SifDec): A constrained and unconstrained testing environment, revisited. *ACM Transactions on Mathematical Software*, 29(4):373–394, 2003.
- [20] N. Hansen. The CMA Evolution Strategy: A Comparing Review. In J. Lozano, P. Larrañaga, I. Inza, and E. Bengoetxea, editors, *Towards a New Evolutionary Computation*, volume 192 of *Studies in Fuzziness and Soft Computing*, pages 75–102. Springer Berlin Heidelberg, 2006.
- [21] A.-R. Hedar. Global Optimization Test Problems. http://www-optima.amp.i.kyoto-u.ac.jp/member/student/hedar/Hedar_files/TestGO.htm. (last accessed on 2017-10-20).
- [22] W. Hock and K. Schittkowski. *Test Examples for Nonlinear Programming Codes*, volume 187 of *Lecture Notes in Economics and Mathematical Systems*. Springer, Berlin, Germany, 1981.
- [23] J. Jahn. *Introduction to the Theory of Nonlinear Optimization*. Springer, Berlin, 1994.
- [24] M. Jamil and X.-S. Yang. A literature survey of benchmark functions for global optimisation problems. *International Journal of Mathematical Modelling and Numerical Optimisation*, 4(2):150–194, 2013.

- [25] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of the 1995 IEEE International Conference on Neural Networks*, pages 1942–1948, Perth, Australia, 1995. IEEE Service Center, Piscataway.
- [26] S. Kitayama, M. Arakawa, and K. Yamazaki. Sequential approximate optimization using radial basis function network for engineering optimization. *Optimization and Engineering*, 12(4):535–557, 2011.
- [27] D. P. Kroese, S. Porotsky, and R. Y. Rubinstein. The Cross-Entropy method for continuous and multi-extremal optimization. *Metodol Comput Appl Probab*, 8:383–407, 2006.
- [28] S. Kullback and R. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, 22:79–86, 1951.
- [29] S. Le Digabel. Algorithm 909: NOMAD: Nonlinear Optimization with the MADS algorithm. *ACM Transactions on Mathematical Software*, 37(4):44:1–44:15, 2011.
- [30] L. Lukšan and J. Vlček. Test problems for nonsmooth unconstrained and linearly constrained optimization. Technical Report V-798, ICS AS CR, 2000.
- [31] L.S. Matott, A.J. Rabideau, and J.R. Craig. Pump-and-treat optimization using analytic element method flow models. *Advances in Water Resources*, 29(5):760–775, 2006.
- [32] K.I.M. McKinnon. Convergence of the Nelder-Mead simplex method to a nonstationary point. *SIAM Journal on Optimization*, 9(1):148–158, 1998.
- [33] E. Mezura-Montes and C.A. Coello. Useful Infeasible Solutions in Engineering Optimization with Evolutionary Algorithms. In *Proceedings of the 4th Mexican International Conference on Advances in Artificial Intelligence, MICAI’05*, pages 652–662, Berlin, Heidelberg, 2005. Springer-Verlag.
- [34] N. Mladenović, J. Petrović, V. Kovačević-Vujčić, and M. Čangalović. Solving spread spectrum radar polyphase code design problem by tabu search and variable neighbourhood search. *European Journal of Operational Research*, 151(2):389–399, 2003.
- [35] J.J. Moré, B.S. Garbow, and Kenneth E. Hillstom. Testing unconstrained optimization software. *ACM Transactions on Mathematical Software*, 7(1):17–41, 1981.
- [36] J.J. Moré and S.M. Wild. Benchmarking derivative-free optimization algorithms. *SIAM Journal on Optimization*, 20(1):172–191, 2009.
- [37] F. Pigache, F. Messine, and B. Nogarede. Optimal Design of Piezoelectric Transformers: A Rational Approach Based on an Analytical Model and a Deterministic Global Optimization. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 54(7):1293–1302, 2007.

- [38] Kenneth Price, Rainer M Storn, and Jouni A Lampinen. *Differential evolution: a practical approach to global optimization*. Springer Science & Business Media, 2006.
- [39] K. Price R. Storne. Differential Evolution - A simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global optimization*, 11:341–359, 1997.
- [40] J.F. Rodríguez, J.E. Renaud, and L.T. Watson. Trust Region Augmented Lagrangian Methods for Sequential Response Surface Approximation and Optimization. *Journal of Mechanical Design*, 120(1):58–66, 1998.
- [41] R. Y. Rubinstein. Optimization of computer simulation models with rare events . *European Journal of Operational Research*, 99:89–112, 1997.
- [42] R. Y. Rubinstein and D. P. Kroese. *The Cross-Entropy Method: A unified Approach to Combinatorial Optimization, Monte-Carlo Simulation and Machine Learning*. Springer: Berlin Heidelberg, New York, USA, 2004.
- [43] M. Stein. Large sample properties of simulations using latin hypercube sampling. *Technometrics*, 29(2):143–151, 1987.
- [44] J. Tao and N. Wang. DNA Double Helix Based Hybrid GA for the Gasoline Blending Recipe Optimization Problem. *Chemical Engineering and Technology*, 31(3):440–451, 2008.
- [45] K. Wang and N. Wang. A novel RNA genetic algorithm for parameter estimation of dynamic systems. *Chemical Engineering Research and Design*, 88(11):1485–1493, 2010.