Comput Optim Appl manuscript No.
(will be inserted by the editor)

Analysis of multi-objective Kriging-based methods for constrained global optimization

Cédric Durantin · Julien Marzat · Mathieu Balesdent

Received: date / Accepted: date

Abstract Metamodeling, i.e. building surrogate models to expensive black-box functions, is an interesting way to reduce the computational burden for optimization purpose. Kriging is a popular metamodel based on Gaussian Process theory, whose statistical properties have been exploited to build efficient global optimization algorithms. Single and multi-objective extensions have been proposed to deal with constrained optimization when the constraints are also evaluated numerically. This paper first compares these methods on a representative analytical benchmark. A new multi-objective approach is then proposed to also take into account the prediction accuracy of the constraints. A numerical evaluation is provided on the same analytical benchmark and a realistic aerospace case study.

Keywords Black-box functions · Constrained global optimization · Kriging · Multi-objective optimization

1 Introduction

In the aerospace industry, as well as in other fields, it is usual to design products using computer models. This facilitates the development of new concepts and reduces the need for expensive prototypes. However, these computer models are usually very expensive-to-evaluate. For example, a single crash analysis or a finite-element code may require more than twenty-four hours of computation. A classical strategy in this case is to replace the computer simulation

C. Durantin
Univ. Grenoble Alpes, CEA, LETI, MINATEC Campus, F-38054 Grenoble, France
E-mail: cedric.durantin@cea.fr

J. Marzat and M. Balesdent
ONERA - The French Aerospace Lab, F-91123 Palaiseau, France
E-mail: julien.marzat@onera.fr, mathieu.balesdent@onera.fr

by a metamodel to emulate the statistical input-output relationship at a very reduced computational cost. These surrogate models are then very useful for optimization purpose to replace the costly process and limit the number of actual function calls.

The Kriging predictor [1,2], which is based on Gaussian Process theory, is a possible choice in this context. It has become one of the most popular methods for approximating deterministic computer models, since it also provides information on the uncertainty of the approximation. This has been exploited to develop the "Efficient Global Optimization" (EGO) algorithm [3] for unconstrained problems, which suggests iteratively new points where to evaluate the costly black-box function so as to improve the evaluation of the global optimum until convergence conditions are met or the simulation budget exhausted. Different infill criteria have been described such as the Probability of Improvement (PI) [4] or the Expected Improvement (EI) [3], the latter having been successfully applied on many realistic problems [5].

Most real world design optimization problems involve constraint functions that are also expensive-to-evaluate. They can then be approximated by Kriging models and taken into account in the surrogate-based optimization algorithms. A basic approach is to add a penalty to the objective function if the constraints are violated but this strategy does not provide accurate results when the constraints are highly nonlinear. A probability-based approach proposed in [2] uses the product of the EI criterion with the probability of feasibility of the constraints. This method smooths the landscape produced by a simple penalty function although it introduces a high conservativeness with respect to constraint boundaries, which might fail to find the actual global optimum. Improvements have been suggested by using sequential uncertainty reduction [6] or approximating the constraint boundaries by Support Vector Machines (SVM) [7].

A bi-objective approach has been described in [8] to treat objective improvement and constraint satisfaction as separate goals for improving the choice of infill points, with promising results. Related methods [9,10] proposed to aggregate the constraints using a filter, which boils down to a bi-objective problem minimizing the objective while reducing as much as possible the aggregated constraint violations. This work evolved into a progressive barrier approach to handle constraints [11], which is now integrated in the NOMAD software [12].

The present work compares single and multi-objective methods, which are based on Expected Improvement and Probability of Feasibility criteria, on a representative benchmark of test functions with complementary features in terms of smoothness, active constraints and disjoint feasible regions. A new multi-objective criterion with three objectives is also proposed to also take into account the prediction accuracy of the constraints, so as to increase confidence in the feasibility of the estimated global optimum.

A brief review of Kriging is presented in Section 2 and the principles of Kriging-based optimization in Section 3. Existing methods for constrained optimization are then reviewed and compared on an analytical benchmark in

Section 4. The efficiency of bi-objective approaches is evaluated in Section 5 and compared to the new three-objective criterion on the analytical benchmark in Section 6 and on a realistic aerospace case study in Section 7.

2 Kriging basics

The function to be approximated is defined as

$$Y : \left| \begin{array}{l} \mathbb{X} \subset \mathbb{R}^d \ \mapsto \mathbb{Y} \subset \mathbb{R} \\ \mathbf{x} \rightarrow z = y(\mathbf{x}) \end{array} \right. . \tag{1}$$

A surrogate model is built from a set of n input vectors $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and a vector of corresponding scalar evaluations $\mathbf{z} = \{z_1, z_2, \ldots, z_n\}$. Once the training of the surrogate model is achieved, the output $\widehat{y}$ of the process can be predicted at new sampled points, with a highly reduced computational cost.

2.1 Gaussian process

Consider a probability space $(\Omega, \mathcal{F}, \mathcal{P})$, where Ω is the sample space equipped with its σ-algebra $\mathcal{F}$ and a probability measure $\mathcal{P}$. A stochastic (or random) process is an application that maps the Cartesian product of the probability space and an index space $\mathbb{X}$ in an output space $\mathbb{Y}$:

$$Y(\mathbf{x}, \omega) : \left| \begin{array}{l} \mathbb{X} \times (\Omega, \mathcal{F}, \mathcal{P}) \ \mapsto \mathbb{Y} \\ (\mathbf{x}, \omega) \rightarrow y(\mathbf{x}, \omega) \end{array} \right. , \tag{2}$$

where ω is an event of Ω. A path (or realization) of this process is defined by $y(\mathbf{x}, \omega) \equiv Y(\mathbf{x}, \omega)$. $Y(\mathbf{x}, \omega_0)$ is a function of $\mathbf{x} \in \mathbb{X}$ for some given $\omega_0 \in \Omega$, while $Y(\mathbf{x}^{(0)}, \omega)$ is a random variable for some given $\mathbf{x}^{(0)} \in \mathbb{X}$. The event ω is usually implied and will now be omitted for the sake of readability.

A random process is Gaussian if any finite number of its random variables has a multivariate Gaussian distribution [13]. A Gaussian Process is thus a second-order stochastic process defined by its mean function μ and its auto-covariance function $C(\cdot)$:

$$\mu(\mathbf{x}) \equiv \mathbb{E}\left[Y(\mathbf{x})\right], \forall \mathbf{x} \in \mathbb{X}, \tag{3}$$

$$C(\mathbf{x}, \mathbf{x}') \equiv \mathbb{E}\left[\left(Y(\mathbf{x}) - \mu(\mathbf{x})\right)\left(Y(\mathbf{x}') - \mu(\mathbf{x}')\right)\right], \forall \left(\mathbf{x}, \mathbf{x}'\right) \in \mathbb{X} \times \mathbb{X}. \tag{4}$$

A stochastic process is stationary if it is invariant by translation. As a consequence, its mean function reduces to a constant,

$$\mu(\mathbf{x}) = \mu_0, \forall \mathbf{x} \in \mathbb{X}, \tag{5}$$

and its autocovariance function becomes

$$C(\mathbf{x}, \mathbf{x}') = \sigma^2 R(\mathbf{x} - \mathbf{x}'), \forall (\mathbf{x}, \mathbf{x}') \in \mathbb{X} \times \mathbb{X}, \tag{6}$$

where σ^2 is the constant process variance and $R(\cdot)$ is the autocorrelation function which only depends on the shift $\mathbf{x} - \mathbf{x}'$ in the input space.

2.2 Kriging predictor

The best linear unbiased predictor (in the least-square sense) of the unobserved quantity of interest $Y_0 \equiv Y(\mathbf{x}^{(0)})$ is also known as the Kriging predictor [1, 2]. The black-box function z is modeled as a sample path from a Gaussian process Y to be characterized for any $\mathbf{x} \in \mathbb{X}$:

$$Y(\mathbf{x}) = \beta(\mathbf{x}) + \varepsilon(\mathbf{x}), \tag{7}$$

where the mean function $\beta(\cdot)$ is an optional regression model that can be estimated from available data (here, a constant mean model is chosen) and the noise $\varepsilon(\cdot)$ is a zero-mean Gaussian process with stationary autocovariance. The Kriging predictor has many interesting properties: it interpolates the observations in the dataset, it is asymptotically consistent if the autocovariance function is regular and its predictor is a Gaussian variable [14]. For the dataset $\{\mathcal{X}, \mathbf{z}\}$, the mean of the prediction $\widehat{y}$ is

$$\widehat{y}(\mathbf{x}) = \widehat{\beta}(\mathbf{x}) + \boldsymbol{\psi}(\mathbf{x}, \mathcal{X})^T \mathbf{R}^{-1}(\mathcal{X}) \left(\mathbf{z} - \mathbf{1}_n \widehat{\beta}(\mathbf{x}) \right), \tag{8}$$

where

$$\begin{cases} \mathbf{R}_{|ij}(\mathcal{X}) = R(\mathbf{x}_i - \mathbf{x}_j), \text{ for } i, j = 1, \ldots, n \\ \boldsymbol{\psi}(\mathbf{x}, \mathcal{X}) = [R(\mathbf{x} - \mathbf{x}_1), \ldots, R(\mathbf{x} - \mathbf{x}_n)]^T \end{cases}. \tag{9}$$

Among many possible choices for the autocorrelation function $R(\cdot)$ [15], the widely-used *p-exponential* function (10) has been chosen here. It involves two types of shape parameters: the exponents $0 < p_i \leq 2$ reflect the smoothness of the interpolation (2 is the smoothest) while the correlation lengths θ_i are scale factors which translate the spatial influence area of a sample point.

$$R(\mathbf{x} - \mathbf{x}') = \exp\left(- \sum_{i=1}^{d} \theta_i |\mathbf{x}[i] - \mathbf{x}'[i]|^{\, p_i} \right), \tag{10}$$

where $\mathbf{x}[i]$ is the $i-$th component of $\mathbf{x}$. Since the prediction is a Gaussian variable, a mean squared error formula can be derived to estimate the variance of the prediction error. It is written as

$$\widehat{s}^2(\mathbf{x}) = \widehat{\sigma}^2 \left(1 - \boldsymbol{\psi}(\mathbf{x}, \mathcal{X})^T \mathbf{R}^{-1}(\mathcal{X}) \boldsymbol{\psi}(\mathbf{x}, \mathcal{X}) + \frac{\left(1 - \mathbf{1}_n^T \mathbf{R}^{-1}(\mathcal{X}) \boldsymbol{\psi}(\mathbf{x}, \mathcal{X})\right)^2}{\mathbf{1}_n^T \mathbf{R}^{-1}(\mathcal{X}) \mathbf{1}_n} \right). \tag{11}$$

This measure is one of the advantages of Gaussian process models. It has the intuitive property that its value is null at any sampled point. If the variance is high at some point, it means that there are not enough points in the training dataset to build an accurate prediction at this location.

Maximum likelihood estimation (MLE) is generally used to learn the parameters $\boldsymbol{\Theta} = \{\theta_i, p_i | i = 1, \ldots, n\}$ of the autocorrelation function (10), the mean regression function $\beta(\cdot)$ and the process variance σ^2. It is well suited to computer experiments because it does not depend on the dimension of the

input space. The likelihood of the observations $\mathbf{z}$ is defined from the normal distribution

$$L(\beta, \sigma^2, \boldsymbol{\Theta}|\mathbf{z}) = \frac{1}{(2\pi\sigma^2)^{n/2}\sqrt{\det(\mathbf{R}(\boldsymbol{\Theta}))}} \exp\left[-\frac{(\mathbf{z}-\beta)^T\mathbf{R}^{-1}(\boldsymbol{\Theta})(\mathbf{z}-\beta)}{2\sigma^2}\right]. \tag{12}$$

The maximum-likelihood estimates are then obtained from the first-order optimality conditions of this unconstrained optimization problem,

$$\widehat{\beta} = \frac{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{z}}{\mathbf{1}^T\mathbf{R}^{-1}\mathbf{1}}, \qquad \widehat{\sigma}^2 = \frac{(\mathbf{z}-\widehat{\beta})^T\mathbf{R}^{-1}(\mathbf{z}-\widehat{\beta})}{n}. \tag{13}$$

By substituting these two solutions in (12), it leads to a new expression called *reduced likelihood function* which is used to estimate correlation parameters:

$$\widehat{\boldsymbol{\Theta}} \in \arg\min_{\theta,p} \frac{n}{2}\ln(\widehat{\sigma}^2) + \frac{1}{2}\ln[\det(\mathbf{R}(\boldsymbol{\Theta}))]. \tag{14}$$

The shape of this function is usually multimodal with several local minima, especially when both $\boldsymbol{\theta}$ and $\mathbf{p}$ are searched for. In our experiments, this optimization problem was solved using the Covariance Matrix Adaptation Evolution Strategy (CMA-ES) [16]. CMA-ES is a second-order evolutionary search strategy based on the propagation of a covariance matrix, which has obtained excellent results on multimodal and high-dimensional test problems [16].

3 Unconstrained optimization

This section considers the unconstrained global optimization problem

$$\widehat{\mathbf{x}} \in \arg\min_{\mathbf{x}\in\mathbb{X}} z(\mathbf{x}). \tag{15}$$

Several strategies have been studied to use the Kriging model for estimating $\widehat{\mathbf{x}}$ with a limited budget of evaluations. The simplest strategy that minimizes iteratively the Kriging mean prediction is not efficient, since it does not take into account the variance information and thus discards areas where uncertainty is high [17]. Two main alternative infill criteria have been proposed and are now widespread: PI (*Probability of Improvement*) and EI (*Expected Improvement*). These criteria are used to suggest a point where the actual function should be evaluated to improve the estimation of the global minimum.

The function value z is considered as the realization of a Gaussian variable $Y(\mathbf{x})$ with mean $\widehat{y}(\mathbf{x})$, and standard deviation $\widehat{s}(\mathbf{x})$. The *improvement* from the current sampled minimum $z_{\min} = \min\{z_1, \ldots, z_n\}$ is defined as

$$I(\mathbf{x}) = \begin{cases} z_{\min} - \widehat{y}(\mathbf{x}), \text{ if } \widehat{y}(\mathbf{x}) < z_{\min} \\ 0, \text{ otherwise} \end{cases}. \tag{16}$$

The PI for the optimum estimate can then be computed as the probability that $Y(\mathbf{x}) \leq z_{\min}$,

$$\begin{aligned} \mathrm{PI}(\mathbf{x}) &= \mathcal{P}[Y(\mathbf{x}) \leq z_{\min}] = \frac{1}{\widehat{s}(\mathbf{x})\sqrt{2\pi}} \int_0^{\infty} \exp\left(\frac{-\left(z_{\min} - \widehat{y}(\mathbf{x})\right)^2}{2\widehat{s}^2(\mathbf{x})} \right) \mathrm{d}\mathbf{x} \\ \mathrm{PI}(\mathbf{x}) &= \Phi\left(\frac{z_{\min} - \widehat{y}(\mathbf{x})}{\widehat{s}(\mathbf{x})} \right), \end{aligned} \tag{17}$$

where $\Phi(\cdot)$ is the cumulative density function of the normal distribution [4]. Maximizing this criterion provides sample points that are mostly located near the current best point, where the probability is already high. However, it does not allow the exploration of areas where there are not many training points, which makes it more suitable for local refinement.

Expected Improvement (EI) has been proposed in [3] to achieve a better balance between exploration and exploitation of the surrogate model. Instead of simply finding the probability that there will be some improvement, the amount of improvement expected is calculated. The improvement (16) can be rewritten as

$$I(\mathbf{x}) = \begin{cases} \widehat{s}(z'_{\min}(\mathbf{x}) - u(\mathbf{x})), \text{ if } u(\mathbf{x}) < z'_{\min}(\mathbf{x}) \text{ and } \widehat{s} > 0, \\ 0, \text{ otherwise} \end{cases}, \tag{18}$$

where $u(\mathbf{x}) = \frac{z - \widehat{y}(\mathbf{x})}{\widehat{s}}$ and $z'_{\min}(\mathbf{x}) = \frac{z_{\min} - \widehat{y}(\mathbf{x})}{\widehat{s}}$. Using the assumed normal distribution for the Kriging predictor, the mathematical expectation of the improvement is derived in (19) where $\phi(\cdot)$ is the normal probability density function.

$$\begin{aligned} \mathrm{EI}(\mathbf{x}) &= \widehat{s} \int_{-\infty}^{z'_{\min}} (z'_{\min}(\mathbf{x}) - u(\mathbf{x}))\, \phi(u(\mathbf{x})) d\mathbf{x}, \\ \mathrm{EI}(\mathbf{x}) &= \widehat{s} \left[z'_{\min} \Phi(z'_{\min}) + \left[\frac{e^{-u^2/2}}{\sqrt{2\pi}} \right]_{-\infty}^{z'_{\min}} \right], \\ \mathrm{EI}(\mathbf{x}) &= \widehat{s} \left[z'_{\min} \Phi(z'_{\min}) + \phi(z'_{\min}) \right], \\ \mathrm{EI}(\mathbf{x}) &= (z_{\min} - \widehat{y})\, \Phi\left(\frac{z_{\min} - \widehat{y}}{\widehat{s}} \right) + \widehat{s} \phi\left(\frac{z_{\min} - \widehat{y}}{\widehat{s}} \right). \end{aligned} \tag{19}$$

Inspection of the formula reveals two trends. The first part is nearly the same as the PI criterion with an additional factor that scales the EI value on the improvement value. This part is large when $\widehat{y}$ is smaller than $z_{\min}$. The second part tends to be large when the uncertainty on the prediction is high. The fact that EI is large for areas of improvement and regions of high uncertainty shows its interesting global refinement properties.

An algorithm called *Efficient Global Optimization*(EGO) was also defined in [3] to address (15). It proceeds as follows:

1. Compute an initial space-filling design $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and evaluate $\mathbf{z} = \{z_1, z_2, \ldots, z_n\}$. Learn the Kriging parameters.

2. Find the argument $\mathbf{x}_{\text{new}}$ of the maximum of the infill criterion, either PI (17) or EI (19), and compute the corresponding function value z_{new}.

3. Append $\mathbf{x}_{\text{new}}$ to $\mathcal{X}$ and z_{new} to $\mathbf{z}$. Learn the Kriging parameters again.
4. Repeat the procedure from 2 until stopping criterion is met (either on the number of function evaluations or if the value of the infill criterion remains below some threshold at successive iterations).

As already suggested, and further confirmed by many publications, the EI criterion is better suited for global optimization with EGO [2,18,19] and will be the basis of the work presented here. Figure 1 illustrates the behavior of the two criteria on a simple example and highlights the fact that EI takes the uncertainty on the prediction more into account than PI, which leads its maximum to be closer to the minimum of the real function. The shape of the EI criterion is multimodal with several local extrema, thus its optimization at Step 2 of EGO should be performed by an appropriate global search algorithm. For example, the optimum might be found using CMA-ES [16] or DIRECT [20]. Note that, at this step, there is no constraint on the number of evaluations of EI since the Kriging predictor is very cheap to compute. It has been shown in [21] that EGO generates a dense sequence of samples which converges to the global optimum, under the assumption of a known and fixed autocorrelation function.

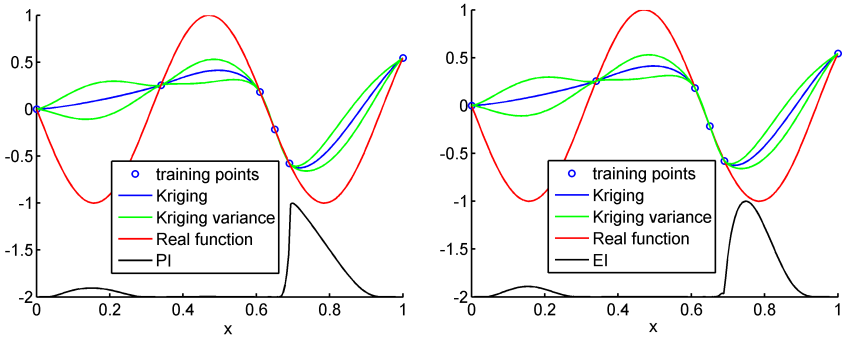


Fig. 1: Simple example of PI (left) and EI (right) criteria

4 Single-objective constrained optimization

Constrained global optimization of the black-box function z is now considered.

$$\begin{aligned} \widehat{\mathbf{x}} \in \arg\min_{\mathbf{x}\in\mathbb{X}} \quad & z\left(\mathbf{x}\right), \\ \text{s.t.} \quad & \mathbf{g}\left(\mathbf{x}\right) \leq \mathbf{0}. \end{aligned} \tag{20}$$

The constraint vector $\mathbf{g}(\cdot)$ is also assumed to be known only via costly evaluations of the same simulation, and each constraint will thus be approximated by a Kriging surrogate model. The sub-optimization problem of the infill criteria (Step 2 of EGO) is usually where the constraints are taken into account. There are two main strategies to incorporate constraint information: the first one is to evaluate the probability of constraint feasibility [8], the second one is to directly use the Kriging mean value of the constraints in the optimization of the infill criterion [19].

4.1 Probability of feasibility (EI×PF)

This measure is analogous to PI, with $\widehat{g}_i(\cdot)$ the Kriging predicted mean of the i-th constraint and $\widehat{s}_{g_i}(\cdot)$ its variance,

$$\mathrm{PF}_i(\mathbf{x}) = \Phi\left(-\frac{\widehat{g}_i(\mathbf{x})}{\widehat{s}_{g_i}(\mathbf{x})}\right). \tag{21}$$

This method transforms the constrained optimization problem of EI at Step 2 of EGO into an unconstrained one by multiplying the value of EI by the probability that the point is feasible [2]. The infill criteria is then

$$\mathrm{EI} \times \mathrm{PF}(\mathbf{x}) = \mathrm{EI}(\mathbf{x}) \prod_{i=1}^{m} \mathrm{PF}_i(\mathbf{x}), \tag{22}$$

with m the number of constraints. The magnitude of EI will be driven to zero where there is a very low probability of feasibility for any of the constraints. One concern often noticed is that it impacts the value of EI too strongly and keeps the algorithm from exploring points close to the constraint boundaries.

4.2 Constrained EI (CEI)

This method for handling constraints, originally proposed in [22], is simply to solve the EI optimization as a constrained optimization problem. It is highly dependent on the accuracy of the constraint surrogate models because the Kriging mean value is directly used as constraints. If constraints are active, it will add points close to the value $\widehat{\mathbf{g}}(\mathbf{x}) = \mathbf{0}$:

$$\begin{array}{ll} \max\limits_{\mathbf{x}\in\mathbb{X}} & \mathrm{EI}(\mathbf{x}), \\ \text{s.t.} & \widehat{\mathbf{g}}(\mathbf{x}) \leq 0. \end{array} \tag{23}$$

An auxiliary global optimization algorithm with constraint handling is needed here, possible choices being DIRECT [20], MADS [23] or CMA-ES with additional penalization [24]. Note that when disconnected feasible regions occur and no feasible point has been found, [22] suggested to use the probability of feasibility in order to get a starting point for this criterion.

4.3 Expected violation (EV)

An alternative approach uses the Expected Violation (EV) as a constraint instead of the Kriged constraint mean values [25],

$$\mathrm{EV}_i(\mathbf{x}) = -\widehat{g}_i(\mathbf{x})\Phi\left(-\frac{\widehat{g}_i(\mathbf{x})}{\widehat{s}_{g_i}(\mathbf{x})}\right) + \widehat{s}_{g_i}\phi\left(-\frac{\widehat{g}_i(\mathbf{x})}{\widehat{s}_{g_i}(\mathbf{x})}\right). \tag{24}$$

EV is analogous to the EI criterion, using the Kriging model of the constraints. It is high in regions where the constraints are likely to be non-violated or where there is a high uncertainty on the constraint models. If EV is higher than a given threshold t_{EV}, the point can be considered as feasible. The new EI optimization problem is defined as

$$\begin{array}{ll} \max\limits_{\mathbf{x}\in\mathbb{X}} & \mathrm{EI}(\mathbf{x}), \\ \text{s.t.} & \mathrm{EV}_i(\mathbf{x}) \geq t_{\mathrm{EV}} \text{ for } i = 1, \ldots, m. \end{array} \tag{25}$$

4.4 Analytical benchmark

A benchmark of four analytical optimization test problems with various characteristics has been defined to evaluate the methods. The first optimization problem is the Branin function with a smooth constraint. This function is usually evaluated on the square $x_1 \in [-5\ ,\ 10]\,, x_2 \in [0\ ,\ 15]$ (rescaled in $[0,1]^2$ for display). In Figure 2, the constraint limits are in red, the non-feasible zone is white and the black point is the true global optimum.

$$f_{Branin}(x_1, x_2) = \left(x_2 - \frac{5.1x_1^2}{4\pi^2} + \frac{5x_1}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(x_1) + \frac{1}{8\pi}. \tag{26}$$

Problem 1:

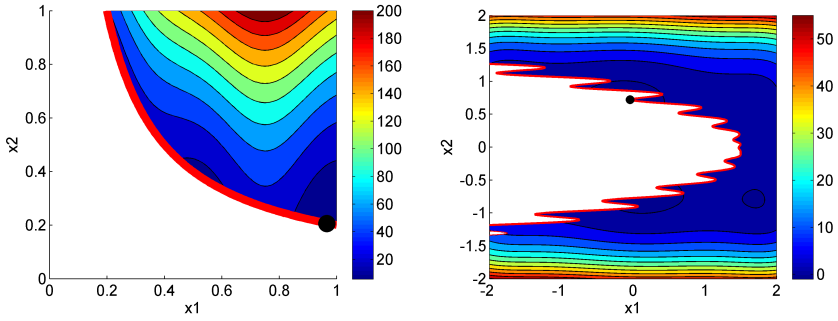
$$\begin{array}{ll} \min\limits_{\mathbf{x}\in\mathbb{X}} & f_{Branin}(x_1, x_2), \\ \text{s.t.} & g_{Branin}(x_1, x_2) = 0.2 - x_1 x_2 \leq 0. \end{array} \tag{27}$$

The second optimization problem is the six-hump Camel function with a multimodal constraint defined on the square $[-2, 2]^2$. This problem has also a local minimum with a function value close to the global minimum (see Figure 2b).

Problem 2:

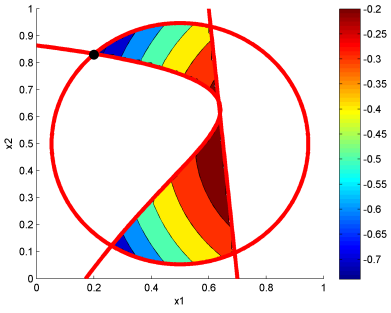
$$\begin{array}{ll} \min\limits_{\mathbf{x}\in\mathbb{X}} & f_{Camel}(x_1, x_2) = \left(4 - 2.1x_1^2 + \frac{x_1^4}{3}\right)x_1^2 + x_1x_2 + \left(-4 + 4x_2^2\right), \\ \text{s.t.} & g_{Camel}(x_1, x_2) = 1.5 - \left(1.5x_2 - \frac{\cos(31x_2)}{6}\right)^2 - x_1 \leq 0. \end{array} \tag{28}$$

The third optimization problem is a function used in [19] with three constraints (two active and one inactive, see Figure 2c). It is defined on the square $[0, 1]^2$. This test problem challenges methods on their ability to handle multiple constraints with some of them inactive.

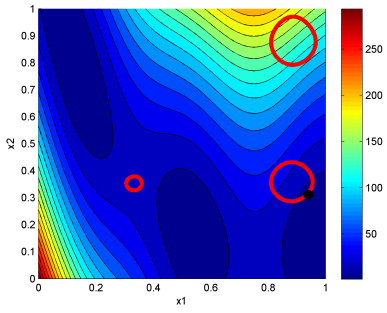


(a) Problem 1

(b) Problem 2



(c) Problem 3



(d) Problem 4

Fig. 2: Benchmark of analytical functions

Problem 3:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{X}} \quad & f_{Sasena}(x_1,x_2) = -(x_1-1)^2-(x_2-0.5)^2, \\ \text{s.t.} \quad & g_1(x_1,x_2) = (x_1-3)^2+(x_2+2)^2 e^{-x_2^7} - 12 \leq 0, \\ & g_2(x_1,x_2) = 10x_1+x_2-7 \leq 0, \\ & g_3(x_1,x_2) = (x_1-0.5)^2+(x_2+0.5)^2 - 0.2 \leq 0. \end{aligned} \tag{29}$$

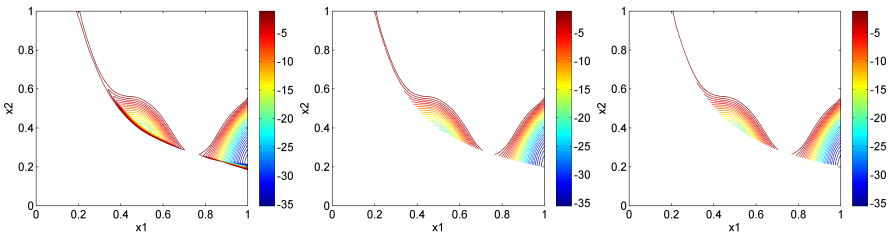
The fourth optimization problem is the Branin function with a Gomez function as constraint and an additional sine function to increase modality of the constraint. It is defined on the square $[0,1]^2$ where the constraint has disjoint domains. In Figure 2d, the full Branin function is displayed for readability sake but areas allowed by the constraint are only inside circles.

$$\begin{aligned} g_{Gomez}(x_1,x_2) = 6 - \left(4-2.1x_1^2+\tfrac{1}{3}x_1^4\right)x_1^2 - x_1x_2 \\ + \left(4-4x_2^2\right)x_2^2 - 3\sin\left[6\left(1-x_1\right)\right] - 3\sin\left[6\left(1-x_2\right)\right]. \end{aligned} \tag{30}$$

Problem 4:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{X}} \quad & f_{Sasena}(x_1,x_2), \\ \text{s.t.} \quad & g_{Gomez}(x_1,x_2) \leq 0. \end{aligned} \tag{31}$$

A first illustration of the behavior of the three single-objective criteria is available in Figure 3 for Problem 1 at final iteration, their maximal value being located in the blue area. For the EI×PF criterion, the constraint is well delimited by the Probability of feasibility but the contour level shows that the maximum of the EI does not lie on the boundary, unlike the actual function minimum. For CEI, if constraints are badly modeled the criterion may add points toward a local minimum. There is nothing to drive the criterion to add points in high uncertainty areas of the constraint Kriging prediction. For EV (with $t_{\mathrm{EV}} = 10^{-3}$), the boundary is nearly the same as for CEI. It is slightly more conservative because the Kriging prediction variance is also taken into account. The optimum of the EI criterion is very close to the actual constraint.



(a) EIxPF (b) Constrained EI (c) EV

Fig. 3: Behavior of the single-objective criteria on Problem 1

Numerical results on the four test problems, averaged on ten initial random space-filling designs, are presented in Table 1 and Figure 4. For each problem, the number of function calls needed and the distance to the real optimum are evaluated. To compare the accuracy of the approximation of the constraints for each method, the root mean squared error around the region of the minimum is calculated between actual constraint functions and the mean value of their Kriging approximations. Another root mean squared error is also calculated in the same area but only on constraints boundaries ($g_i = 0$). These two evaluation criteria indicate whether the method under study has provided an accurate evaluation of the constraint around the optimum and therefore if the estimated optimum can be trusted.

The EV method has the worst results on the four problems. Taking into account the prediction variance to reduce the allowed regions leads this method to be too conservative. It found local minima several times for the three last problems. CEI provides good results with less function calls than the Probability of feasibility method. But this method failed to find the global minimum for Problems 3 and 4, due to disjoint regions. When it does not fail, both RMSE on constraint Kriging approximations are lower than those obtained with the EIxPF method. This might be explained by the fact that points added with CEI are closer to the constraint actual boundaries than points

added using EIxPF. The only method that does not fail on the four problems is EIxPF. It also obtains fair results in terms of constraint prediction accuracy, although less accurate than CEI. These benchmark results are consistent with and complementary to those presented in [8, 19].

Table 1: Results of single-objective methods on the four benchmark problems, averaged on 10 different space-filling designs (mean ± standard deviation)

Benchmark	Method	Distance to minimizer	Function calls	RMSE $\widehat{g}$ area	RMSE $\widehat{g}=0$
Problem 1	EIxPF	0.002 ± 0.001	31.8 ± 1.6	0.042 ± 0.011	0.0015 ± 0.001
	CEI	0.018 ± 0.014	28.7 ± 4.7	0.036 ± 0.011	0.0014 ± 0.001
	EV	0.002 ± 0.000	33.8 ± 3.3	0.038 ± 0.009	0.0016 ± 0.001
Problem 2	EIxPF	0.001 ± 0.000	53.1 ± 9.5	8.236 ± 3.882	0.3444 ± 0.443
	CEI	0.187 ± 0.195	49.6 ± 6.6	12.240 ± 6.554	0.2959 ± 0.309
	EV	0.155 ± 0.189	37.9 ± 3.3	13.442 ± 8.702	0.8872 ± 1.026
Problem 3	EIxPF	0.082 ± 0.222	25.3 ± 0.9	1.368 ± 0.540	0.0567 ± 0.040
	CEI	0.003 ± 0.001	24.6 ± 1.5	1.130 ± 0.403	0.0458 ± 0.033
	EV	0.360 ± 0.374	25 ± 0.9	1.369 ± 0.829	0.0484 ± 0.021
Problem 4	EIxPF	0.005 ± 0.006	33.7 ± 4.1	6.339 ± 1.840	0.2295 ± 0.149
	CEI	0.081 ± 0.124	30.8 ± 2.4	6.858 ± 2.182	0.6401 ± 0.028
	EV	0.074 ± 0.203	30.3 ± 2.7	7.706 ± 2.478	0.5404 ± 0.013

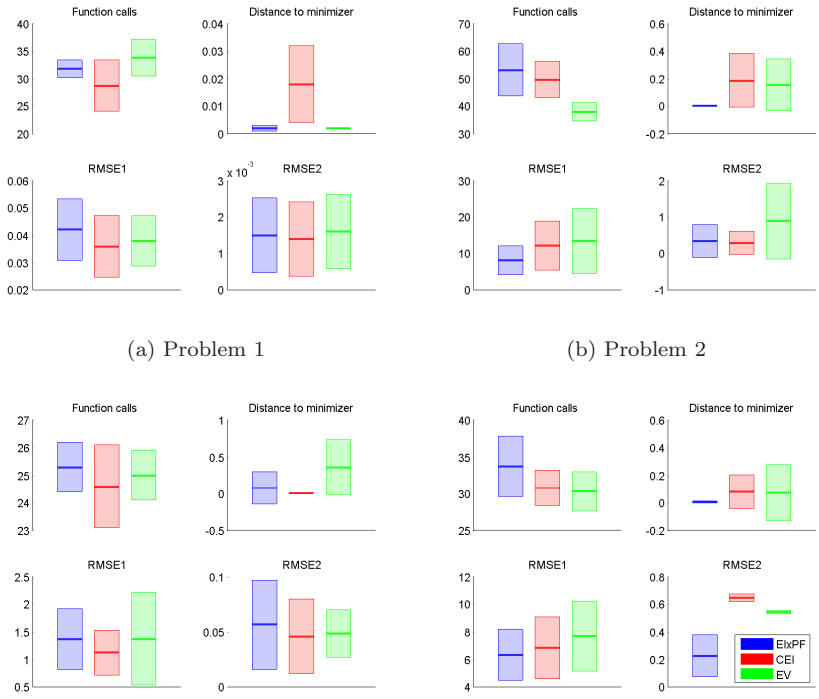
5 Bi-objective EI for constrained optimization

It has been seen in the previous section that single-objective methods have difficulties to achieve a good trade-off between an accurate evaluation of the optimum, the feasibility of the solution and an accurate approximation of the constraints. Therefore, multi-objective criteria are now examined, starting in this section with the bi-objective criterion proposed in [8] and then introducing the new three-objective criterion developed in this work in the next section.

Rather than transforming the constrained problem into an unconstrained one, an alternative is to treat EI and PF as individual goals to optimize,

$$\max_{\mathbf{x}\in\mathbb{X}} \left\{ \mathrm{EI}(\mathbf{x}), \prod_{i=1}^{m} \mathrm{PF}_i(\mathbf{x}) \right\}. \tag{32}$$

The trade-off between obtaining the minimum of the objective function and fulfilling the constraints is then explicitly considered within a bi-objective formulation, forming a set of potential update points as a Pareto set [8]. An example of the two goals presented as a Pareto front for Problem 1 is available in Figure 5. The bottom-left area of the plot is where the interesting values for both objectives lie. This clearly shows that any increase in EI is detrimental



(a) Problem 1

(b) Problem 2

(c) Problem 3

(d) Problem 4

Fig. 4: Summary of numerical results (normalized) for single-objective methods (mean ± standard deviations)

to PF and conversely.

Constructing this Pareto set gives the user a large choice of possible infill points. In the EGO algorithm, only one point that fulfills an acceptable trade-off between both objectives should be selected. Following [26], this point is chosen by maximizing the product of EI and PF. It is important to highlight that a point determined using the single objective approach EIxPF may not correspond to the point found using the Pareto front. The single objective function is severely multimodal, so treating the problem as multi-objective may obtain better solutions [27]. The multi-objective problem might be solved using NSGA-II [28], which builds the Pareto front with a genetic strategy (other multi-objective solvers based on the MADS algorithm could also be considered [29]). In our experiments, a population size of 250 with 150 generations was used. NSGA-II finds a set of solutions clustered in different areas of the design space, then the point with the largest product of EI and PF is selected as the global solution.

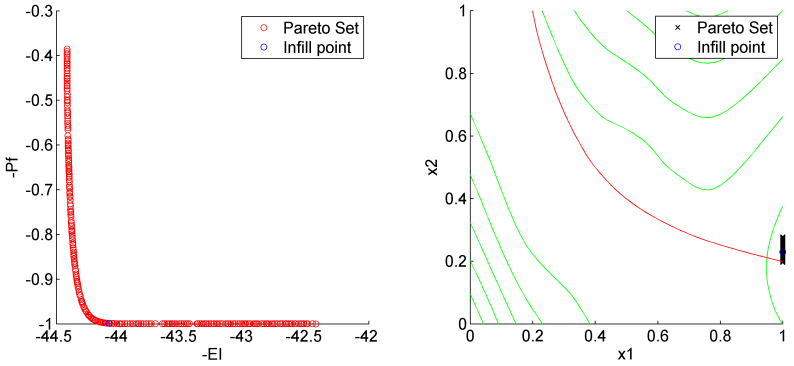
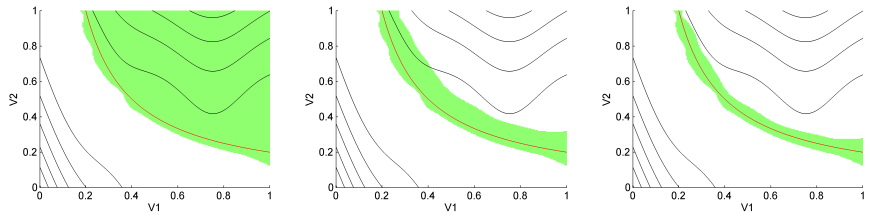


Fig. 5: Example of Pareto front on Problem 1 (left) and location of these points in the input space (right, black line)

In [26], a variant of (32) has also been derived to restrict the area allowed by PF, for problems where inactive constraints can be ignored during optimization. Note that this restricts its applicability because active constraints are not known before the optimization process. The standard formulation of Probability of feasibility identifies regions where the predicted values of constraints will be lower than constraint limits. If all inactive constraints are ignored, the remaining constraints might be treated as equality constraints and the accuracy of the model away from constraint boundaries is considered as less important. To encourage exploitation of regions close to the constraints and on the feasible side, the PDF is integrated over the interval $[0, \varepsilon]$, as

$$Pf_{\varepsilon}(\mathbf{x}) = \Phi\left(-\frac{\widehat{g}(\mathbf{x})}{\widehat{s}_g(\mathbf{x})}\right) - \Phi\left(-\frac{\varepsilon + \widehat{g}(\mathbf{x})}{\widehat{s}_g(\mathbf{x})}\right). \tag{33}$$

The value assigned to ε influences the size of the allowed area. Some examples of regions where $\mathrm{PF}_{\varepsilon}$ is non-zero are drawn on Figure 6 for Problem 1. At high values of ε, the enhanced probability of feasibility is close to the PF function. The authors chose an ε equal to 5% of the known output range of the Kriging approximation of the constraints (difference between $\widehat{g}_{i_{\min}}$ and $\widehat{g}_{i_{\max}}$).



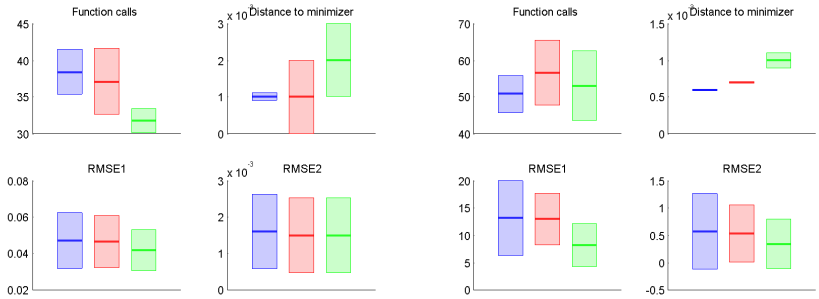
(a) $\varepsilon = 100$ (b) $\varepsilon = 1$ (c) $\varepsilon = 0.1$

Fig. 6: Bi-objective EI vs PFε with various thresholds

The single-objective results from the EIxPF method are compared to the bi-objective methods EIvsPF and EIvsPF$_\varepsilon$ described above, on the same analytical benchmark. Numerical results averaged on ten initial designs are reported in Table 2 and Figure 7. The limitations of the EIvsPF$_\varepsilon$ criterion are highlighted on Problem 3: it is not suitable for optimization with inactive constraints. Since constraints are not considered to be known before the optimization, the practical applicability of this method is questionable. The EIvsPF criterion provides more accurate results on all problems, although EIxPF converges a little faster on Problems 1 and 4.

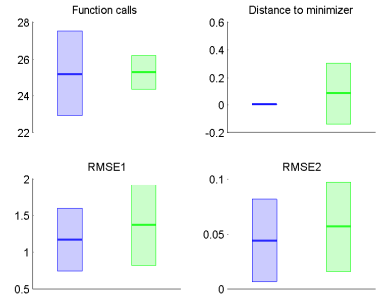
Table 2: Results of bi-objective methods on four benchmark problems, averaged on 10 different space-filling design (mean $\pm$ standard deviation)

Benchmark	Method	Distance to minimizer	Function calls	RMSE g area	RMSE g=0
Problem 1	EIvsPF	0.001 ± 0.000	38.4 ± 3.1	0.0471 ± 0.015	0.0016 ± 0.001
	EIvsPF$_\varepsilon$	0.001 ± 0.001	37.1 ± 4.5	0.0465 ± 0.014	0.0015 ± 0.001
	EIxPF	0.002 ± 0.001	31.8 ± 1.6	0.042 ± 0.011	0.0015 ± 0.001
Problem 2	EIvsPF	0.0006 ± 0.000	50.8 ± 5.1	13.142 ± 6.806	0.5711 ± 0.679
	EIvsPF$_\varepsilon$	0.0007 ± 0.000	56.6 ± 8.9	12.934 ± 4.622	0.5341 ± 0.510
	EIxPF	0.001 ± 0.000	53.1 ± 9.5	8.236 ± 3.882	0.3444 ± 0.443
Problem 3	EIvsPF	0.002 ± 0.002	25.2 ± 2.3	1.172 ± 0.420	0.0437 ± 0.037
	EIvsPF$_\varepsilon$	convergence not reached			
	EIxPF	0.082 ± 0.222	25.3 ± 0.9	1.368 ± 0.540	0.0567 ± 0.040
Problem 4	EIvsPF	0.002 ± 0.002	40.4 ± 3.1	5.130 ± 1.400	0.1052 ± 0.114
	EIvsPF$_\varepsilon$	0.003 ± 0.001	34.8 ± 2.1	6.176 ± 2.409	0.1707 ± 0.257
	EIxPF	0.005 ± 0.006	33.7 ± 4.1	6.339 ± 1.840	0.2295 ± 0.149

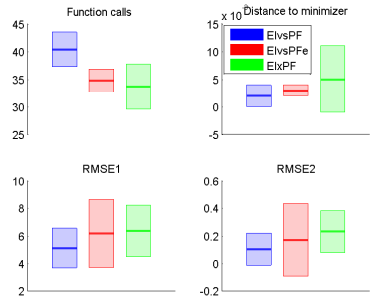


(a) Problem 1

(b) Problem 2



(c) Problem 3



(d) Problem 4

Fig. 7: Summary of numerical results (normalized) for bi-objective methods (mean ± standard deviations)

6 New three-objective criterion: EI, PF and Prediction variance of constraints

The prediction variance of the Kriging approximation of the constraints is not explicitly taken into account in the single and bi-objective approaches. If the variance is high in some areas and the global minimum lies there, it might prevent the optimization process to find this optimum with enough confidence. Reducing the prediction variance leads also to a global minimum that is closer to the actual constraint boundaries in the case of active constraints. A third objective translating the need to minimize the prediction variance of the Kriging approximation of the constraints is thus advocated in addition to the two previous objectives. The prediction variance of constraints $\widehat{s}_{g_i}$ might be directly used as the third objective. An alternative solution is to use the second

part of the EV criterion which governs exploration, $\widehat{s}_{g_i}\phi\left(-\frac{\widehat{g}_i(\mathbf{x})}{\widehat{s}_{g_i}}\right)$,

$$\max_{\mathbf{x}\in\mathbb{X}}\left\{\mathrm{EI}(\mathbf{x}),\prod_{i=1}^{m}\mathrm{PF}_i(\mathbf{x}),-\sum_{i=1}^{m}\widehat{s}_{g\,i}\right\}, \tag{34}$$

or

$$\max_{\mathbf{x}\in\mathbb{X}}\left\{\mathrm{EI}(\mathbf{x}),\prod_{i=1}^{m}\mathrm{PF}_i(\mathbf{x}),-\sum_{i=1}^{m}\widehat{s}_{g_i}\phi\left(-\frac{\widehat{g}_i(\mathbf{x})}{\widehat{s}_{g_i}}\right)\right\}. \tag{35}$$

The point selected from the three-objective Pareto set is still the one that maximizes the product of EI and PF. An example of new Pareto fronts is displayed in Figure 8. The influence of the third objective removes some points from the corresponding bi-objective front. The Pareto set seems to have dominated points only because the front is projected on the plane of EI and PF. In the input space, points from the Pareto set are more spread where the accuracy of the constraint Kriging prediction is high. This may help to reduce uncertainty on constraint boundaries.

The two new criteria are compared (see results averaged on ten initial designs in Table 3 and Figure 9) with the bi-objective method EIvsPF which gave the better results so far. The three-objective criterion (34) provided equivalent optimum estimation as the bi-objective methods but with a much better accuracy for the constraint prediction, except for Problem 1 where the constraint is very smooth and thus easily approximated by all the methods even without explicitly taking into account this objective. The second three-objective criterion (35) obtained the best constraint accuracy on the entire space, however it is less accurate for estimating the optimum value because it gives too much importance to constraint exploration. Nevertheless, these estimates remain more accurate than those obtained with the single-objective methods.

7 Solid propulsion design test problem

A four-dimensional aerospace case study is considered as a validation problem for the multi-objective methods (Figure 10). Its goal is to maximize the propulsive speed increment provided by a rocket motor under minimum thrust-to-weight ratio and geometrical and physical feasibility constraints.

The design is based on a simplified analysis of a cylindrical solid propellant rocket motor which involves the three following disciplines:

- Propulsion: for a given set of propellant characteristics (density, combustion speed, flame temperature, gas composition), and taking nozzle shape and combustion pressure as inputs, thrust, mass flow rate and subsequently specific impulse, combustion time are computed under the assumption of constant thrust.
- Structure, mass and sizing: the tank walls should comply with the design pressure, and for a given mass of propellant the other dry masses are computed, together with geometrical feasibility constraints (combustion area, packaging ratio, size of central channel).

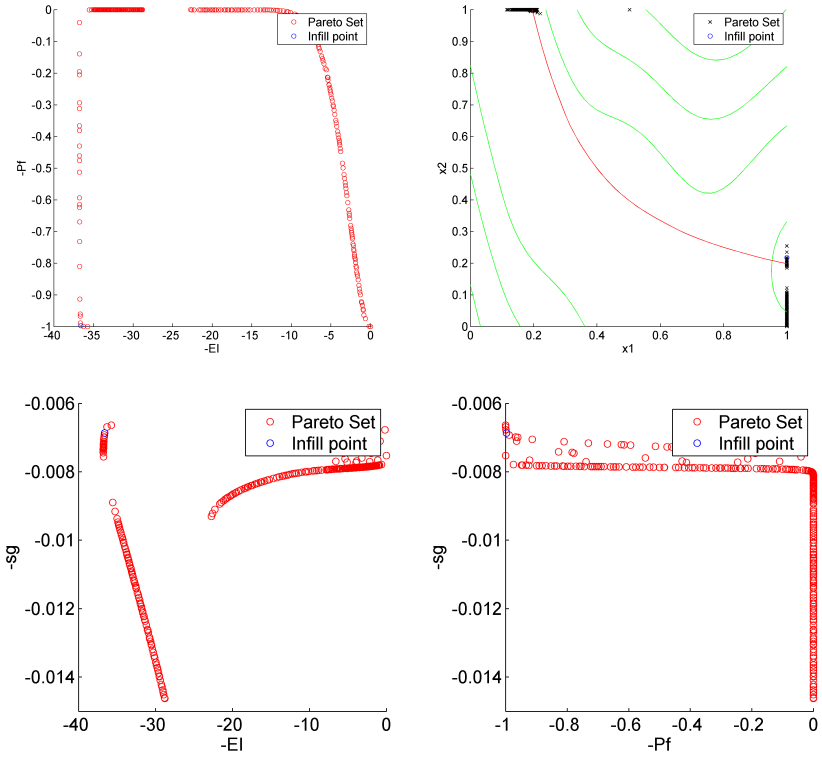
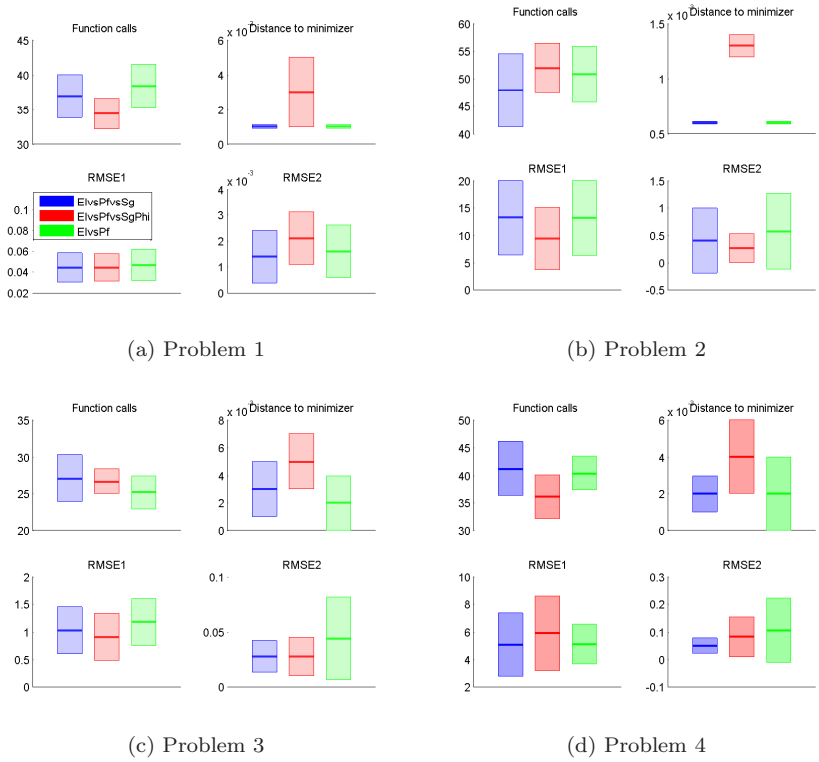


Fig. 8: Projections of Pareto front for three-objective method on Problem 1

Table 3: Results of three-objective methods on four benchmark problems, averaged on 10 different space-filling design (mean ± standard deviation)

Benchmark	Method	Distance to minimizer	Function calls	RMSE g area	RMSE g=0
Problem 1	EI, PF, $\widehat{s}_g$	0.001 ± 0.000	36.9 ± 3.1	0.0447 ± 0.014	0.0014 ± 0.001
	EI, PF, $\widehat{s}_g\phi$	0.003 ± 0.002	34.4 ± 2.2	0.0447 ± 0.013	0.0021 ± 0.001
	EIvsPF	0.001 ± 0.000	38.4 ± 3.1	0.0471 ± 0.015	0.0016 ± 0.001
Problem 2	EI, PF, $\widehat{s}_g$	0.0006 ± 0.000	47.9 ± 6.6	13.203 ± 6.709	0.4063 ± 0.581
	EI, PF, $\widehat{s}_g\phi$	0.0013 ± 0.000	51.9 ± 4.5	9.422 ± 5.544	0.2700 ± 0.258
	EIvsPF	0.0006 ± 0.000	50.8 ± 5.1	13.142 ± 6.806	0.5711 ± 0.679
Problem 3	EI, PF, $\widehat{s}_g$	0.003 ± 0.002	27.1 ± 3.2	1.023 ± 0.416	0.0281 ± 0.014
	EI, PF, $\widehat{s}_g\phi$	0.005 ± 0.002	26.7 ± 1.7	0.909 ± 0.415	0.0281 ± 0.017
	EIvsPF	0.002 ± 0.002	25.2 ± 2.3	1.172 ± 0.420	0.0437 ± 0.037
Problem 4	EI, PF, $\widehat{s}_g$	0.002 ± 0.001	41.2 ± 4.9	5.082 ± 2.244	0.0506 ± 0.027
	EI, PF, $\widehat{s}_g\phi$	0.004 ± 0.002	36.1 ± 4.0	5.894 ± 2.648	0.0820 ± 0.070
	EIvsPF	0.002 ± 0.001	40.4 ± 3.1	5.130 ± 1.400	0.1052 ± 0.114



(a) Problem 1

(b) Problem 2

(c) Problem 3

(d) Problem 4

Fig. 9: Summary of numerical results (normalized) for three-objective methods (mean ± standard deviations)

– Performance estimation: the performance estimation discipline computes the propulsive speed increment of the rocket motor, from the outputs given by the propulsion and mass budget disciplines.

The structure of the design process is illustrated in Figure 11. In this test case, only simplified models are used in the different disciplines. The computation time required to compute the objective function and constraints with respect to the design variables is less than one second. Considering fixed overall dimensions (length $L = 11$ m and diameter $D = 1.07$ m, similar to Ariane 4 additional solid-propellant motor for validation purposes), the design problem is to maximize

$$\Delta V = g.ISP.\ln\left(\frac{m_i}{m_f}\right), \tag{36}$$

where $g = 9.81\text{m.s}^{-2}$ (gravity acceleration), $ISP = \frac{\text{Thrust}}{\dot{m}.g}$ (specific impulse in s), m_i is the mass at lift-off and m_f the mass at propellant burn-out, $\dot{m}$ the mass flow rate. The three constraints of this optimization problem are:

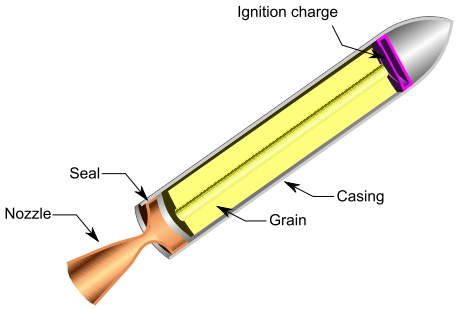


Fig. 10: Solid propulsion rocket motor

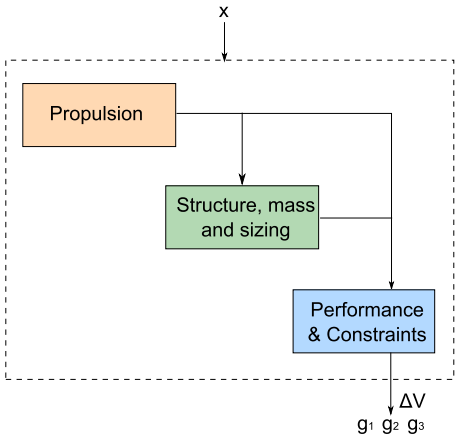


Fig. 11: Disciplines involved in the design problem

- g_1: ensure packaging ratio (Propellant volume/Available volume) $\leq$87%,
- g_2: ensure that central diameter is 30% greater than nozzle throat diameter,
- g_3: ensure that combustion area is greater than the minimum feasible area of central channel walls.

The four design variables are:

- x_1: nozzle throat diameter (m),
- x_2: nozzle exit diameter (m),
- x_3: combustion pressure (bar),
- x_4: propellant mass (kg).

The domain definition and the baseline of the different design variables are given in Table 4.

The three-objective criteria presented in the previous section and the bi-objective criterion EIvsPF have been evaluated on this test problem for five random initial designs (see Table 5). A reference optimum value was previously determined via expensive runs of CMA-ES. The three multi-objective

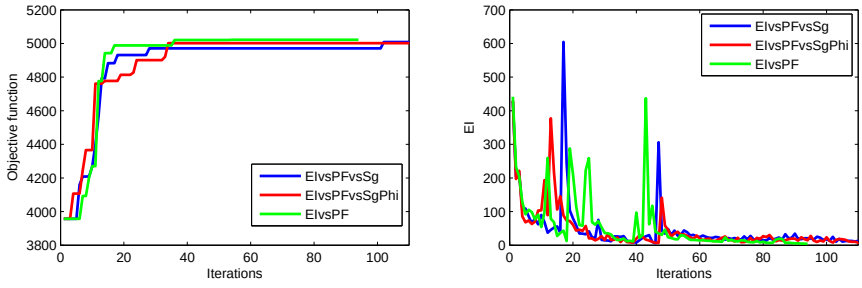
Variable	Domain definition	Baseline
x_1	[0.05,1] (m)	0.75m
x_2	[0.5,1.4] (m)	1m
x_3	[1,500] (bar)	100bar
x_4	[2000,15000] (kg)	8000kg

Table 4: Design variable description

methods obtain similar results on this test problem, and an acceptable solution is always found. The three-objective criterion $\{\mathrm{EI}, \mathrm{PF}, \widehat{s}_g\}$ provides less scattered results than the bi-objective method for a comparable number of runs, while the $\{\mathrm{EI}, \mathrm{PF}, \widehat{s}_g\phi\}$ criterion is a little less accurate but also requires less evaluations. A convergence illustration of the three methods for one initialization is provided in Figure 12. In our test problems, the bi-objective and three-objective optimization required approximately 15s per iteration (standard 3.0GHz Windows PC with MATLAB).

Table 5: Numerical results on solid propulsion design problem, averaged on five initial designs (mean ± standard deviation)

Method	Distance to reference	Function calls	Maximum value
EI, PF, $\widehat{s}_g$	1.208 ± 0.053	117.8 ± 29.2	5011.4 ± 36.8
EI, PF, $\widehat{s}_g\phi$	1.220 ± 0.060	107.4 ± 22.2	4998.8 ± 32.8
EIvsPF	1.260 ± 0.066	115.6 ± 20.1	5012.4 ± 59.6



(a) Objective function (b) EI

Fig. 12: Evolution of objective function and EI for one initialization

8 Conclusion

The main methods for Kriging-based constrained global optimization of black-box functions have been recalled and evaluated numerically in this paper, on an analytical benchmark with various features (smoothness, active constraints, disjoint feasible areas) and an aerospace design problem. These algorithms are all evolutions of the well-known EGO procedure, which is based on the Expected Improvement criterion, and on the probability of feasibility computed from the Kriging approximation of the constraints.

It has been highlighted that multi-objective methods are more appropriate than single-objective ones to achieve a good trade-off between the accuracy of the optimum evaluation and constraint feasibility. In particular, a new three-objective criterion has been proposed to also take into account the accuracy of the constraint prediction, so that the estimated optimum could be confidently identified as feasible.

The multi-objective criteria offer significant improvements for constraint handling in Kriging-based black-box optimization, although more function calls might be needed than with single objective methods on smooth and low-dimensional problems. Further experiments need to be conducted on higher-dimensional problems to confirm the benefits of the bi- and three-objective approaches.

Acknowledgements The authors are indebted to Sébastien Defoort (ONERA) for the definition of the booster benchmark model.

References

1. D. G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, 52(6):119–139, 1951.
2. M. Schonlau and W. J. Welch. Global optimization with nonparametric function fitting. *Proceedings of the ASA, Section on Physical and Engineering Sciences*, pages 183–186, 1996.
3. D. R. Jones, M. J. Schonlau, and W. J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):455–492, 1998.
4. H. J. Kushner. A new method of locating the maximum point of an arbitrary multipeak curve in the presence of noise. *Journal of Fluids Engineering*, 86(1):97–106, 1964.
5. J. P. C. Kleijnen. Kriging metamodeling in simulation: A review. *European Journal of Operational Research*, 192(3):707–716, 2009.
6. V. Picheny. A stepwise uncertainty reduction approach to constrained global optimization. In *Proceedings of the Seventeenth International Conference on Artificial Intelligence and Statistics, Reykjavik, Iceland*, pages 787–795, 2014.
7. A. Basudhar, C. Dribusch, S. Lacaze, and S. Missoum. Constrained efficient global optimization with support vector machines. *Structural and Multidisciplinary Optimization*, 46(2):201–221, 2012.
8. J. M. Parr, A. J. Keane, A. I. J. Forrester, and C. M. E. Holden. Infill sampling criteria for surrogate-based optimization with constraint handling. *Engineering Optimization*, 44(10):1147–1166, 2012.
9. R. Fletcher and S. Leyffer. Nonlinear programming without a penalty function. *Mathematical programming*, 91(2):239–269, 2002.

10. C. Audet and J. E. Dennis Jr. A pattern search filter method for nonlinear programming without derivatives. *SIAM Journal on Optimization*, 14(4):980–1010, 2004.
11. C. Audet and J. E. Dennis Jr. A progressive barrier for derivative-free nonlinear programming. *SIAM Journal on Optimization*, 20(1):445–472, 2009.
12. S. Le Digabel. Algorithm 909: NOMAD: Nonlinear optimization with the MADS algorithm. *ACM Transactions on Mathematical Software*, 37(4):44, 2011.
13. C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Springer-Verlag, New York, 2006.
14. E. Vazquez. *Modélisation comportementale de systèmes non-linéaires multivariables par méthodes à noyaux et applications*. PhD thesis, Université Paris Sud-Paris XI, 2005.
15. T. J. Santner, B. J. Williams, and W. Notz. *The Design and Analysis of Computer Experiments*. Springer-Verlag, Berlin-Heidelberg, 2003.
16. N. Hansen and S. Kern. Evaluating the CMA evolution strategy on multimodal test functions. In *Parallel Problem Solving from Nature-PPSN VIII*, pages 282–291. Springer, 2004.
17. D. R. Jones. A taxonomy of global optimization methods based on response surfaces. *Journal of Global Optimization*, 21(4):345–383, 2001.
18. A. I. J. Forrester, A. Sobester, and A. J. Keane. *Engineering Design via Surrogate Modelling: a Practical Guide*. Wiley, Chichester, 2008.
19. M. J. Sasena. *Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations*. PhD thesis, University of Michigan, USA, 2002.
20. D. R. Jones, C. D. Perttunen, and B. E. Stuckman. Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1):157–181, 1993.
21. E. Vazquez and J. Bect. Convergence properties of the expected improvement algorithm with fixed mean and covariance functions. *Journal of Statistical Planning and Inference*, 140(11):3088–3095, 2010.
22. M. J. Sasena, P. Y. Papalambros, and P. Goovaerts. The use of surrogate modeling algorithms to exploit disparities in function computation time within simulation-based optimization. *Constraints*, 2:5, 2001.
23. C. Audet and J. E. Dennis Jr. Mesh adaptive direct search algorithms for constrained optimization. *SIAM Journal on optimization*, 17(1):188–217, 2006.
24. N. Hansen. The CMA evolution strategy: a comparing review. In *Towards a New Evolutionary Computation*, pages 75–102. Springer, 2006.
25. C. Audet, A. J. Booker, J. E. Dennis Jr, P. D. Frank, and D. W. Moore. A surrogate-model-based method for constrained optimization. In *Proceedings of the 8th AIAA/NASA/USAF/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Long Beach, CA, USA*. Paper No. AIAA-2000-4891, 2000.
26. J. M. Parr, A. I. J. Forrester, A. J. Keane, and C. M. E. Holden. Enhancing infill sampling criteria for surrogate-based constrained optimization. *Journal of Computational Methods in Science and Engineering*, 12(1):25–45, 2012.
27. A. Waldock and D. Corne. Multiple objective optimisation applied to route planning. In *Proceedings of the 13th Annual Conference on Genetic and Evolutionary Computation, Dublin, Ireland*, pages 1827–1834. ACM, 2011.
28. K. Deb. *Multi-objective optimization using evolutionary algorithms*. John Wiley & Sons, Chichester, 2001.
29. C. Audet, G. Savard, and W. Zghal. A mesh adaptive direct search algorithm for multiobjective optimization. *European Journal of Operational Research*, 204(3):545–556, 2010.