User preferences in Bayesian multi-objective optimization: the expected weighted hypervolume improvement criterion

Paul Feliot, Julien Bect, Emmanuel Vazquez

User preferences in Bayesian multi-objective optimization: the expected weighted hypervolume improvement criterion

Paul Feliot[1], Julien Bect[2], and Emmanuel Vazquez[2]

[1] Safran Aircraft Engines, paul.feliot@safrangroup.com
[2] Laboratoire des Signaux et Systèmes (L2S), Centrale-Supélec, firstname.lastname@centralesupelec.fr

Abstract. In this article, we present a framework for taking into account user preferences in multi-objective Bayesian optimization in the case where the objectives are expensive-to-evaluate black-box functions. A novel *expected improvement* criterion to be used within Bayesian optimization algorithms is introduced. This criterion, which we call the *expected weighted hypervolume improvement* (EWHI) criterion, is a generalization of the popular *expected hypervolume improvement* to the case where the hypervolume of the dominated region is defined using an absolutely continuous measure instead of the Lebesgue measure. The EWHI criterion takes the form of an integral for which no closed form expression exists in the general case. To deal with its computation, we propose an importance sampling approximation method. A sampling density that is optimal for the computation of the EWHI for a predefined set of points is crafted and a sequential Monte-Carlo (SMC) approach is used to obtain a sample approximately distributed from this density. The ability of the criterion to produce optimization strategies oriented by user preferences is demonstrated on a simple bi-objective test problem in the cases of a preference for one objective and of a preference for certain regions of the Pareto front.

Keywords: Bayesian optimization · Multi-objective optimization · User preferences · Importance sampling · Sequential Monte-Carlo.

1 Introduction

In this article, we present a Bayesian framework for taking into account user preferences in multi-objective optimization when evaluation results for the functions of the problem are obtained using a computationally intensive computer program. Such a setting is representative of engineering problems where finite elements analysis or fluid dynamics are used. The number of runs of the computer program that can be afforded is limited and the objective is to build a sequence of observation points that rapidly provides a "good" approximation of the set of *Pareto optimal solutions*, where "good" is measured using some user-defined loss function.

To this end, we formulate an *expected improvement* (EI) criterion (see, e.g., [20]) to be used within the BMOO algorithm of [17] that uses the *weighted hypervolume indicator* (WHI) introduced by [29] as a loss function. This new criterion, which we call the *expected weighted hypervolume improvement* (EWHI) criterion, can be viewed as a generalization of the *expected hypervolume improvement* (EHVI) criterion of [14] that enables practitionners to tailor optimization strategies according to user preferences.

The article is structured as follows. First, we recall in Section 2 the framework of Bayesian optimization. Then, we detail in Section 3 the construction of the EWHI criterion and discuss computational aspects. The ability of the criterion to produce optimization strategies according to user preferences is then demonstrated on a simple bi-objective test problem in the cases of a preference for one objective and of a preference for certain regions of the Pareto front in Section 4. Finally, conclusions and perspectives are drawn in Section 5.

2 Bayesian optimization

2.1 The Bayesian approach to optimization

Consider a continuous optimization problem $\mathcal{P}$ defined over a search space $\mathbb{X} \subset \mathbb{R}^d$ and let $\underline{X} = (X_1, X_2, X_3 \ldots)$ be a sequence of observation points in $\mathbb{X}$. The problem $\mathcal{P}$ can be, for example, an unconstrained single-objective optimization problem or a constrained multi-objective problem. The quality at time $n > 0$ of the sequence $\underline{X}$ viewed as an approximate solution to the optimization problem $\mathcal{P}$ can be measured using a positive loss function

$$\varepsilon_n : \underline{X} \mapsto \mathbb{R}^+ , \tag{1}$$

such that $\varepsilon_n(\underline{X}) = 0$ if and only if the set $(X_1, \ldots, X_n)$ solves $\mathcal{P}$ and, given two optimization strategies $\underline{X}_1$ and $\underline{X}_2$, $\varepsilon_n(\underline{X}_1) < \varepsilon_n(\underline{X}_2)$ if and only if $\underline{X}_1$ offers a better solution to $\mathcal{P}$ than $\underline{X}_2$ at time n. Under this framework, one can formulate the notion of improvement as a measure of the loss reduction yielded by the observation of a new point X_{n+1}:

$$I_{n+1} = \varepsilon_n(\underline{X}) - \varepsilon_{n+1}(\underline{X}) , \, n \geq 0 . \tag{2}$$

The improvement is positive if X_{n+1} improves the quality of the solution at time $n + 1$ and zero otherwise.

Assume a statistical model with a vector-valued stochastic process model ξ with probability measure $\mathbb{P}_0$ representing prior knowledge over the functions involved in the optimization problem $\mathcal{P}$. Under the Bayesian paradigm, optimization algorithms are crafted to achieve, on average, a small value of $\varepsilon_n(\underline{X})$ when n increases; where the average is taken with respect to ξ. In this framework, the choice of the observation points X_i is a sequential decision problem. The associated Bayesian-optimal strategy for a finite budget of N observations is, however, not tractable in the general case for N larger than a few units. To

circumvent this difficulty, a common approach is to consider one-step look-ahead strategies (also referred to as myopic strategies, see, e.g., [22, 24] and [8, 18] for discussions about two-step look-ahead strategies) where observation points are chosen one at a time to minimize the conditional expectation of the future loss given past observations:

$$\begin{aligned} X_{n+1} &= \operatorname{argmin}_{x\in\mathbb{X}} \mathbb{E}_n\big(\varepsilon_{n+1}(\underline{X}) \mid X_{n+1} = x\big) \\ &= \operatorname{argmax}_{x\in\mathbb{X}} \mathbb{E}_n\big(\varepsilon_n(\underline{X}) - \varepsilon_{n+1}(\underline{X}) \mid X_{n+1} = x\big) \\ &= \operatorname{argmax}_{x\in\mathbb{X}} \mathbb{E}_n\big(I_{n+1}(\underline{X}) \mid X_{n+1} = x\big)\,,\ n \geq 0\,, \end{aligned} \tag{3}$$

where $\mathbb{E}_n$ stands for the conditional expectation with respect to X_1, $\xi(X_1)$, …, X_n, $\xi(X_n)$. The function

$$\rho_n : x \mapsto \mathbb{E}_n\big(I_{n+1}(\underline{X}) \mid X_{n+1} = x\big)\,,\ n \geq 0\,, \tag{4}$$

is called the *expected improvement* (EI). It is a popular sampling criterion in the Bayesian optimization literature for designing optimization algorithms (see, e.g., [20, 26] for applications to constrained and unconstrained global optimization problems).

2.2 Multi-objective Bayesian optimization

We focus in this work on unconstrained multi-objective optimization problems. Given a set of objective functions $f_j : \mathbb{X} \to \mathbb{R}$, $j = 1, \ldots, p$, to be minimized, the objective is to build an approximation of the Pareto front and of the set of corresponding solutions

$$\Gamma = \{x \in \mathbb{X} : \nexists\, x' \in \mathbb{X} \text{ such that } \ f(x') \prec f(x)\}\,, \tag{5}$$

where $\prec$ stands for the Pareto domination rule defined on $\mathbb{R}^p$ by

$$y = (y_1, \ldots, y_p) \prec z = (z_1, \ldots, z_p) \Longleftrightarrow \begin{cases} \forall i \leq p,\ \ y_i \leq z_i\,, \\ \exists j \leq p,\ \ y_j < z_j\,. \end{cases} \tag{6}$$

In this setting, it is common practice to measure the quality of optimization strategies using the hypervolume loss function (see, e.g., [21, 23, 30]) defined by

$$\varepsilon_n(\underline{X}) = |H \setminus H_n|\,, \tag{7}$$

where $|\cdot|$ denotes the usual (Lebesgue) volume measure in $\mathbb{R}^p$ and where, given an upper-bounded set $\mathbb{B}$ of the form $\mathbb{B} = \{y \in \mathbb{R}^p;\ y \leq y^{\mathrm{upp}}\}$ for some $y^{\mathrm{upp}} \in \mathbb{R}^p$, the subsets

$$H = \{y \in \mathbb{B}\,;\, \exists x \in \mathbb{X}\,,\, f(x) \prec y\}\,, \tag{8}$$

and

$$H_n = \{y \in \mathbb{B}\,;\, \exists i \leq n\,,\, f(X_i) \prec y\}\,, \tag{9}$$

denote respectively the subset of points of $\mathbb{B}$ dominated by the points of the Pareto front and the subset of points of $\mathbb{B}$ dominated by $(f(X_1), \dots, f(X_n))$. The set $\mathbb{B}$ is introduced to ensure that the volumes of H and H_n are finite.

Using the loss function (7), the improvement function (2) takes the form

$$I_{n+1}\left(\underline{X}\right) = |H \setminus H_n| - |H \setminus H_{n+1}| = |H_{n+1} \setminus H_n| \,, \tag{10}$$

and an expected improvement criterion can be formulated as

$$\begin{aligned} \rho_n(x) &= \mathbb{E}_n\big(I_{n+1}(\underline{X}) \mid X_{n+1} = x\big) \\ &= \mathbb{E}_n\left(\int_{\mathbb{B}\setminus H_n} \mathbb{1}_{\xi(x)\prec y}\,\mathrm{d}y\right) \\ &= \int_{\mathbb{B}\setminus H_n} \mathbb{E}_n\left(\mathbb{1}_{\xi(x)\prec y}\right)\,\mathrm{d}y \\ &= \int_{\mathbb{B}\setminus H_n} \mathbb{P}_n\left(\xi(x) \prec y\right)\,\mathrm{d}y\,, \end{aligned} \tag{11}$$

where $\mathbb{P}_n$ stands for the probability $\mathbb{P}_0$ conditioned on $X_1, \xi(X_1), \dots, X_n, \xi(X_n)$. The multi-objective sampling criterion (11) is called the *expected hypervolume improvement* (EHVI) criterion. It has been proposed and studied by Emmerich and coworkers [12, 14, 15].

3 Expected weighted hypervolume improvement (EWHI)

3.1 Formulation of the criterion

To measure the quality of Pareto approximation sets according to user preferences, Zitzler et al. (2007) proposed to use a user-defined continuous measure in the definition of the hypervolume indicator[3] instead of the Lebesgue measure (see [29]):

$$\varepsilon_n(\underline{X}) = \mu(H \setminus H_n)\,, \tag{12}$$

where the measure μ is defined by $\mu(\mathrm{d}y) = \omega(y)\,\mathrm{d}y$ using a positive weight function $\omega : \mathbb{R}^p \to \mathbb{R}^+$. The value $\omega(y)$ for some $y \in \mathbb{R}^p$ can be seen as a reward for dominating y that the user may specify. Optimization strategies crafted using the loss function (12) have been studied by [3, 4, 13, 29].

Observe that, as discussed by [13], assuming that μ possesses the bounded improper integral property, (12) is well defined and upper-bounding values are no longer required in the definition of the sets H and H_n, which can be redefined as:

$$\begin{cases} H &= \{y \in \mathbb{R}^p\,;\, \exists x \in \mathbb{X}\,,\, f(x) \prec y\}\,, \\ H_n &= \{y \in \mathbb{R}^p\,;\, \exists i \leq n\,,\, f(X_i) \prec y\}\,. \end{cases} \tag{13}$$

[3] In the original definition, the authors introduce additional terms to weight the axis. In this work, one of our objective is to get rid of the bounding set $\mathbb{B}$, as proposed by [13]. Therefore we do not consider these terms.

Similarly to (7), the improvement function associated to the loss function (12) takes the form

$$I_{n+1}\left(\underline{X}\right) = \mu(H \setminus H_n) - \mu(H \setminus H_{n+1}) = \mu(H_{n+1} \setminus H_n)\,, \tag{14}$$

and an expected improvement criterion can be formulated as:

$$\begin{aligned}\rho_n(x) &= \mathbb{E}_n\big(I_{n+1}(\underline{X}) \mid X_{n+1} = x\big) \\ &= \mathbb{E}_n\left(\int_{H_n^{\mathrm{c}}} \mathbb{1}_{\xi(x) \prec y}\, \mu(\mathrm{d}y)\right) \\ &= \int_{H_n^{\mathrm{c}}} \mathbb{P}_n\left(\xi(x) \prec y\right) \omega(y)\, \mathrm{d}y\,, \end{aligned} \tag{15}$$

where H_n^{c} denotes the complementary of H_n in $\mathbb{R}^p$. By analogy with the EHVI criterion, we call the expected improvement criterion (15) the *expected weighted hypervolume improvement* (EWHI) criterion.

3.2 Computation of the criterion

Under the assumption that the components ξ_i of ξ are mutually independent stationary Gaussian processes, which is a common modeling assumption in the Bayesian optimization literature (see, e.g., [25]), the term $\mathbb{P}_n\left(\xi(x) \prec y\right)$ in the expression (15) of the EWHI can be expressed in closed form: for all $x \in \mathbb{X}$ and $y \in H_n^{\mathrm{c}}$,

$$\mathbb{P}_n\left(\xi(x) \prec y\right) = \prod_{i=1}^{p} \Phi\left(\frac{y_i - \widehat{\xi}_{i,n}(x)}{\sigma_{i,n}(x)}\right)\,, \tag{16}$$

where Φ denotes the Gaussian cumulative distribution function and $\widehat{\xi}_{i,n}(x)$ and $\sigma^2_{i,n}(x)$ denote respectively the kriging mean and variance at x for the i^{th} component of ξ (see, e.g., [25, 28]).

The integration of (16) over H_n^{c} on the other hand, is a non-trivial problem. Besides, it has to be done several times to solve the optimization problem (3) and choose X_{n+1}. To address this issue, we propose to choose X_{n+1} among a set of predefined candidate points obtained using sequential Monte-Carlo techniques as in [17], and derive a method to compute approximations of (15) with arbitrary weight functions ω for this set.

Let then $\mathcal{X}_n = (x_{n,k})_{1 \leq k \leq m_x} \in \mathbb{X}^{m_x}$ be a set of m_x points where ρ_n is to be evaluated and denote

$$\rho_{n,k} = \rho_n(x_{n,k}) = \int_{H_n^{\mathrm{c}}} \omega(y)\, \mathbb{P}_n\left(\xi(x_{n,k}) \prec y\right)\, \mathrm{d}y\,, \quad 1 \leq k \leq m_x\,. \tag{17}$$

Using a sample $\mathcal{Y}_n = (y_{n,i})_{1 \leq i \leq m_y}$ of m_y points obtained from a density π_n on H_n^{c} with un-normalized density γ_n and with normalizing constant

$$Z_n = \int_{H_n^{\mathrm{c}}} \gamma_n(y)\, \mathrm{d}y\,, \tag{18}$$

an importance sampling approximation of the $(\rho_{n,k})_{1\leq k\leq m_x}$ can be written as

$$\widehat{\rho}_{n,k} = \frac{Z_n}{m_y}\sum_{i=1}^{m_y}\frac{\omega(y_{n,i})\,\mathbb{P}_n\left(\xi(x_{n,k}) \prec y_{n,i}\right)}{\gamma_n(y_{n,i})}\,, \quad 1\leq k\leq m_x\,. \tag{19}$$

To obtain a good approximation for all $\widehat{\rho}_{n,k}$ using a single sample $\mathcal{Y}_n$, the un-normalized density γ_n can be chosen to minimize the average sum of squared approximation errors:

$$\begin{aligned}
&\mathbb{E}\left(\sum_{k=1}^{m_x}\left(\widehat{\rho}_{n,k}-\rho_{n,k}\right)^2\right)\\
&=\frac{1}{m_y}\sum_{k=1}^{m_x}\left(Z_n\int_{H_n^{\mathrm{c}}}\frac{\omega(y)^2\,\mathbb{P}_n\left(\xi(x_{n,k})\prec y\right)^2}{\gamma_n(y)^2}\gamma_n(y)\,\mathrm{d}y-\rho_{n,k}^2\right)\\
&=\frac{1}{m_y}\left(Z_n\int_{H_n^{\mathrm{c}}}\frac{\sum_{k=1}^{m_x}\omega(y)^2\,\mathbb{P}_n\left(\xi(x_{n,k})\prec y\right)^2}{\gamma_n(y)^2}\gamma_n(y)\,\mathrm{d}y-\sum_{k=1}^{m_x}\rho_{n,k}^2\right)\,.
\end{aligned} \tag{20}$$

This leads, using the Cauchy-Schwarz inequality (see, e.g., [7]), to the definition of the following density on H_n^{c}:

$$L_2^{\mathrm{opt}}(y)\propto\gamma_n(y)=\sqrt{\sum_{k=1}^{m_x}\omega(y)^2\,\mathbb{P}_n\left(\xi(x_{n,k})\prec y\right)^2}\,. \tag{21}$$

To obtain a sample distributed from the L_2^{opt} density and carry out the approximate computation of the EWHI using (19), we resort to sequential Monte-Carlo techniques as well (see, e.g., [2, 11, 17]). The algorithm that we use is not detailed here for the sake of brevity. The reader is referred to Section 4 of [17] for a discussion about this aspect. Details about the computation of the normalizing constant Z_n and about the variance of the proposed estimator are given in Appendix A.

4 Numerical experiments

In our experiments, we illustrate the operation of the EWHI criterion on the bi-objective BNH problem as defined in [10] for the following two weight functions adapted from [29]:

$$\begin{cases}
\omega_1(y_1,y_2)=\dfrac{1}{15}e^{-\frac{y_1}{15}}\cdot\dfrac{\mathbb{1}_{[0,150]}(y_1)}{150}\cdot\dfrac{\mathbb{1}_{[0,60]}(y_2)}{60}\,,\\
\omega_2(y_1,y_2)=\dfrac{1}{2}\left(\varphi\left(y,\mu_1,C\right)+\varphi\left(y,\mu_2,C\right)\right)\,,
\end{cases} \tag{22}$$

where $\varphi(y,\mu,C)$ denotes the Gaussian probability density function with mean μ and covariance matrix C, evaluated at y. The ω_1 weight function is based

on an exponential distribution and encodes preference for the minimization of the first objective. The ω_2 weight function is a sum of two bivariate Gaussian distributions and encodes preference for improving upon two reference points μ_1 and μ_2, chosen as $\mu_1 = (80, 20)$ and $\mu_2 = (30, 40)$ with $C = RS(RS)^T$, where

$$R = \begin{bmatrix} \cos\left(\frac{\pi}{4}\right) & -\sin\left(\frac{\pi}{4}\right) \\ \sin\left(\frac{\pi}{4}\right) & \cos\left(\frac{\pi}{4}\right) \end{bmatrix} \text{ and } S = \begin{bmatrix} 20 & 0 \\ 0 & 3 \end{bmatrix}. \tag{23}$$

To carry out the experiments, we use the BMOO algorithm of [16] with $m_x = m_y = 1000$ particles for both SMC algorithms. The functions of the problem are modeled using stationnary Gaussian processes with a constant mean and an anisotropic Matérn covariance kernel. A log-normal prior distribution is placed on the parameters of the kernel and these are updated at each iteration of the algorithm using maximum a posteriori substition (see, e.g., [5]). The algorithm is initialized with a pseudo-maximin latin hypercube design of $N = 10$ experiments and is iterated over 20 iterations. To handle the constraints of the BNH problem, the EWHI criterion is multiplied by the probability of feasibility, as is common practice in the Bayesian optimization litterature (see, e.g., [26]).

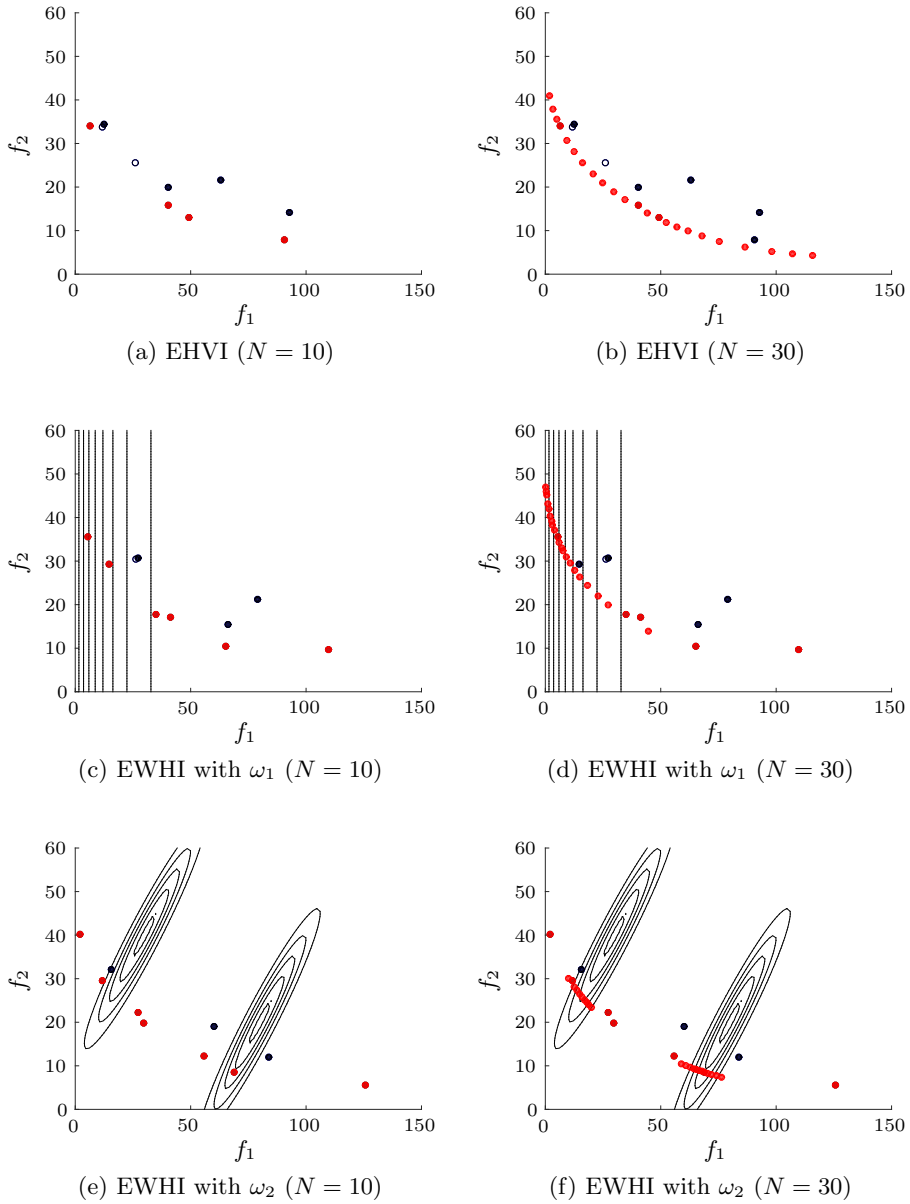
In Figure 1, results obtained by the algorithm using the weight functions ω_1 and ω_2 in the EWHI definition are compared to results obtained by the same algorithm using the EHVI criterion. Observe in Figures 1(d) and 1(f) that observations are concentrated in regions of the Pareto front that correspond to high ω values, whereas observations are spread along the front in Figure 1(b) where the EHVI is used. In practice, this means that less iterations would have been required to satisfyingly populate the interesting regions of the Pareto front.

5 Conclusions and perspectives

It is shown in this paper how user-defined weight functions can be leveraged by a Bayesian framework to produce optimization strategies that focus on preferred regions of the Pareto front of multi-objective optimization problems. Two example weight functions from [29] which encode respectively a preference for one objective and a preference toward specific regions of the Pareto front are used, and the demonstration of the effectiveness of the proposed approach is carried out on a simple bi-objective optimization problem.

On more practical problems, crafting sensible weight functions can be a difficult task, especially when one has no prior knowledge about the approximate location of the Pareto front. The use of desirability functions (see, e.g. [13, 19, 27]) or utility functions (see, e.g., [1]) might provide useful insights on that issue and shall be the object of future investigations to provide a more principled approach.

In the presented framework, optimization strategies are built sequentially using an expected improvement sampling criterion called the expected weighted hypervolume improvement (EWHI) criterion. The exact computation of the criterion being intractable in general, an approximate computation prodecure using



(a) EHVI ($N = 10$) (b) EHVI ($N = 30$)

(c) EWHI with ω_1 ($N = 10$) (d) EWHI with ω_1 ($N = 30$)

(e) EWHI with ω_2 ($N = 10$) (f) EWHI with ω_2 ($N = 30$)

Fig. 1. Distributions obtained after 20 iterations of the optimization algorithm on the BNH problem when the weight functions ω_1 and ω_2 are used. The results obtained using the EHVI criterion are shown for reference. The contours of the weight functions are represented as black lines and the non-dominated solutions as red disks. Black disks indicate feasible dominated solutions and black circles indicate non-feasible solutions.

importance sampling is proposed. A sampling density that is optimal for the simultaneous computation of the criterion for a set of candidate points is crafted and a sequential Monte-Carlo algorithm is used to produce samples from this density.

This choice triggers an immediate question: What is the sample size m_y required by the algorithm? In fact, the problem is not so much to obtain a precise approximation of ρ_n for all $x \in \mathcal{X}_n$, which would require a large sample size to distinguish very close points, but to deal with the optimization problem (3) and to identify with good confidence the points of $\mathcal{X}_n$ that correspond to high values of ρ_n. A first step toward a solution to this problem is to compute an approximation of the variance of $\widehat{\rho}_n$, as carried out in Appendix A. Further investigations on this issue are left for future work.

A Approximate variance of the EI estimator

We derive in this appendix the variance of the SMC estimator for ρ_n. In the SMC procedure that we consider, the particles $(y_{n,i})_{1\leq i\leq m}$ are obtained from a sequence of densities $(\pi_{n,t})_{0\leq t\leq T}$, where $\pi_{n,0}$ is an easy-to-sample initial density and $\pi_{n,T} = \pi_n$ is the target density. Let $(\gamma_{n,t})_{0\leq t\leq T}$ and $(Z_{n,t})_{0\leq t\leq T}$ denote the corresponding sequences of un-normalized densities and normalizing constants.

First, observe that, for $1 \leq t \leq T$,

$$\begin{aligned} Z_{n,t} &= \int_{H_n^c} \gamma_{n,t}(y)\,\mathrm{d}y \\ &= Z_{n,t-1} \int_{G_n} \frac{\gamma_{n,t}(y)}{\gamma_{n,t-1}(y)}\,\pi_{n,t-1}(y)\,\mathrm{d}y\,. \end{aligned} \tag{24}$$

Thus, we can derive a sequence of approximations $\widehat{Z}_{n,t}$ of $Z_{n,t}$, $t \geq 1$, using the following recursion formula:

$$\begin{cases} \widehat{Z}_{n,0} = Z_{n,0} = \int_{G_n} \gamma_{n,0}(y)\,\mathrm{d}y, \\ \widehat{Z}_{n,t} = \widehat{Z}_{n,t-1} \left(\frac{1}{m} \sum_{i=1}^m \frac{\gamma_{n,t}(y_{n,t-1,i})}{\gamma_{n,t-1}(y_{n,t-1,i})} \right)\,, \end{cases} \tag{25}$$

where the particles $(y_{n,t-1,i})_{1\leq i\leq m} \sim \pi_{n,t-1}$ are obtained using an SMC procedure (see, e.g., [6]). The estimator of $\rho_n(x)$ that we actually consider is then

$$\widehat{\rho}_n(x) = \frac{\widehat{Z}_n}{m} \sum_{i=1}^m \frac{\omega(y)\,\mathbb{P}_n\left(\xi(x) \prec y_{n,i}\right)}{\gamma_n(y_{n,i})} = \widehat{Z}_n \widehat{\alpha}_n(x) \tag{26}$$

where

$$\widehat{\alpha}_n(x) = \frac{1}{m} \sum_{i=1}^m \frac{\omega(y)\,\mathbb{P}_n\left(\xi(x) \prec y_{n,i}\right)}{\gamma_n(y_{n,i})}\,, \tag{27}$$

and

$$\widehat{Z}_n = \widehat{Z}_{n,T} = Z_{n,0} \prod_{u=1}^T \widehat{\theta}_{n,\,u}\,, \tag{28}$$

with

$$\widehat{\theta}_{n,t} = \frac{1}{m}\sum_{i=1}^{m} \frac{\gamma_{n,t}(y_{n,t-1,i})}{\gamma_{n,t-1}(y_{n,t-1,i})}\,. \tag{29}$$

Now, assume the idealized setting, as usual in the SMC literature (see, e.g., [9]), where

(i) $y_{n,t,i} \overset{\text{i.i.d}}{\sim} \pi_{n,t}$, $1 \le i \le m$,
(ii) the samples $\mathcal{Y}_{n,t} = (y_{n,t,i})_{1\le i\le m}$ are independent, $0 \le t \le T$.

Observe from (19) and (24) that under (i), $\widehat{\alpha}_n(x)$ is an unbiased estimator of $\alpha_n(x) = \frac{\rho_n(x)}{Z_n}$, and $\widehat{\theta}_{n,t}$ is an unbiased estimator of $\theta_{n,t} = \frac{Z_{n,t}}{Z_{n,t-1}}$, $1 \le t \le T$. Moreover, under (ii), $\widehat{\alpha}_n(x)$ and the $(\widehat{\theta}_{n,t})_{1\le t\le T}$ are independent. Thus,

$$\begin{aligned}
&\operatorname{Var}\widehat{\rho}_n(x) = \mathbb{E}\big(\widehat{\alpha}_n^2\big)\mathbb{E}\big(\widehat{Z}_n^2\big) - \mathbb{E}\big(\widehat{\alpha}_n(x)\big)^2\mathbb{E}\big(\widehat{Z}_n\big)^2\\
&= \big(\operatorname{Var}\widehat{\alpha}_n(x) + \alpha_n(x)^2\big)\big(\operatorname{Var}\widehat{Z}_n + Z_n^2\big) - \alpha_n(x)^2 Z_n^2\\
&= \operatorname{Var}\widehat{\alpha}_n(x)\operatorname{Var}\widehat{Z}_n + \alpha_n(x)^2\operatorname{Var}\widehat{Z}_n + Z_n^2\operatorname{Var}\widehat{\alpha}_n(x)\,.
\end{aligned}$$

We obtain the coefficient of variation of $\widehat{\rho}_n(x)$

$$\frac{\operatorname{Var}\widehat{\rho}_n(x)}{\rho_n(x)^2} = \Lambda_n(x)^2 + \big(1 + \Lambda_n(x)^2\big)\,\Delta_{n,T}^2, \tag{30}$$

where $\Lambda_n(x)^2 = \frac{\operatorname{Var}\widehat{\alpha}_n(x)}{\alpha_n(x)^2}$ and $\Delta_{n,t}^2 = \frac{\operatorname{Var}\widehat{Z}_{n,t}}{Z_{n,t}^2}$ are the coefficients of variation of $\widehat{\alpha}_n(x)$ and $\widehat{Z}_{n,t}$ respectively.

Using the same ideas as above, we have

$$\Delta_{n,t}^2 = \delta_{n,t}^2 + \big(1 + \delta_{n,t}^2\big)\,\Delta_{n,t-1}^2, \tag{31}$$

where $\delta_{n,t}^2 = \frac{\operatorname{Var}\widehat{\theta}_{n,t}}{\theta_{n,t}^2}$ is the coefficient of variation of $\widehat{\theta}_{n,t}$.

Estimators of $\Lambda_n(x)^2$, $\Delta_{n,t}^2$ and $\delta_{n,t}^2$ can be derived under (ii). For instance, observe that

$$\delta_{n,t}^2 = \frac{1}{m}\,\frac{\operatorname{Var}\left(\frac{\gamma_{n,t}(y_{n,t-1,1})}{\gamma_{n,t-1}(y_{n,t-1,1})}\right)}{\mathbb{E}\left(\frac{\gamma_{n,t}(y_{n,t-1,1})}{\gamma_{n,t-1}(y_{n,t-1,1})}\right)^2}\,. \tag{32}$$

Thus, an estimator of $\delta_{n,t}^2$ is

$$\widehat{\delta}_{n,t}^2 = \frac{\sum_{i=1}^m \frac{\gamma_{n,t}(y_{n,t-1,i})^2}{\gamma_{n,t-1}(y_{n,t-1,i})^2}}{\left(\sum_{i=1}^m \frac{\gamma_{n,t}(y_{n,t-1,i})}{\gamma_{n,t-1}(y_{n,t-1,i})}\right)^2} - \frac{1}{m}. \tag{33}$$

Plugging (33) in (31), we obtain an estimator of $\Delta_{n,t}^2$:

$$\widehat{\Delta}_{n,t}^2 = \widehat{\delta}_{n,t}^2 + \left(1 + \widehat{\delta}_{n,t}^2\right)\cdot\widehat{\Delta}_{n,t-1}^2. \tag{34}$$

Similarly, an estimator of $\Lambda_n(x)^2$ is

$$\widehat{\Lambda}_n(x)^2 = \frac{\sum_{i=1}^{m} \frac{\omega(y)^2 \, \mathbb{P}_n(\xi(x) \prec y_{n,i})^2}{\gamma_n(y_{n,i})^2}}{\left(\sum_{i=1}^{m} \frac{\omega(y) \, \mathbb{P}_n(\xi(x) \prec y_{n,i})}{\gamma_n(y_{n,i})}\right)^2} - \frac{1}{m}. \tag{35}$$

As a result, we obtain the following numerically tractable approximation of the variance of $\widehat{\rho}_n(x)$:

$$\mathrm{Var}\left(\widehat{\rho}_n(x)\right) \approx \widehat{\rho}_n(x)^2 \cdot \left(\widehat{\Lambda}_n(x)^2 + \left(1 + \widehat{\Lambda}_n(x)^2\right) \cdot \widehat{\Delta}_{n,t}^2\right), \tag{36}$$

where $\widehat{Z}_{n,t}$ and $\widehat{\Delta}_{n,t}^2$ are obtained recursively using (25) and (34), $\widehat{\Lambda}_n(x)^2$ is computed using (35) and $\widehat{\rho}_n(x)$ is computed using (26).

References

1. Astudillo, R., Frazier, P.: Multi-attribute bayesian optimization under utility uncertainty. Proceedings of the NIPS Workshop on Bayesian Optimization, December 2017, Long Beach, USA (To appear)
2. Au, S.K., Beck, J.L.: Estimation of small failure probabilities in high dimensions by subset simulation. Probabilistic Engineering Mechanics **16**(4), 263–277 (2001)
3. Auger, A., Bader, J., Brockhoff, D., Zitzler, E.: Articulating user preferences in many-objective problems by sampling the weighted hypervolume. In: Proceedings of the 11th Annual conference on Genetic and evolutionary computation. pp. 555–562. ACM (2009)
4. Auger, A., Bader, J., Brockhoff, D., Zitzler, E.: Investigating and exploiting the bias of the weighted hypervolume to articulate user preferences. In: Proceedings of the 11th Annual conference on Genetic and evolutionary computation. pp. 563–570. ACM (2009)
5. Bect, J., Ginsbourger, D., Li, L., Picheny, V., Vazquez, E.: Sequential design of computer experiments for the estimation of a probability of failure. Statistics and Computing **22**(3), 773–793 (2012)
6. Bect, J., Li, L., Vazquez, E.: Bayesian subset simulation. SIAM/ASA Journal on Uncertainty Quantification **5**(1), 762–786 (2017)
7. Bect, J., Sueur, R., Gérossier, A., Mongellaz, L., Petit, S., Vazquez, E.: Échantillonnage préférentiel et méta-modèles: méthodes bayésiennes optimale et défensive. In: 47èmes Journées de Statistique de la SFdS-JdS 2015 (2015)
8. Benassi, R.: Nouvel algorithme d'optimisation bayésien utilisant une approche Monte-Carlo séquentielle. Ph.D. thesis, Supélec (2013)
9. Cérou, F., Del Moral, P., Furon, T., Guyader, A.: Sequential Monte Carlo for rare event estimation. Statistics and Computing **22**(3), 795–808 (2012)
10. Chafekar, D., Xuan, J., Rasheed, K.: Constrained multi-objective optimization using steady state genetic algorithms. In: Genetic and Evolutionary Computation-GECCO 2003. pp. 813–824. Springer (2003)
11. Del Moral, P., Doucet, A., Jasra, A.: Sequential monte carlo samplers. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **68**(3), 411–436 (2006)

12. Emmerich, M.: Single- and multiobjective evolutionary design optimization assisted by Gaussian random field metamodels. Ph.D. thesis, Technical University Dortmund (2005)
13. Emmerich, M., Deutz, A.H., Yevseyeva, I.: On reference point free weighted hypervolume indicators based on desirability functions and their probabilistic interpretation. Procedia Technology **16**, 532–541 (2014)
14. Emmerich, M., Giannakoglou, K.C., Naujoks, B.: Single- and multi-objective evolutionary optimization assisted by Gaussian random field metamodels. IEEE Transactions on Evolutionary Computation **10**(4), 421–439 (2006)
15. Emmerich, M., Klinkenberg, J.W.: The computation of the expected improvement in dominated hypervolume of Pareto front approximations. Technical report, Leiden University (2008)
16. Feliot, P.: A Bayesian approach to constrained multi-objective optimization. Ph.D. thesis, IRT SystemX and Centrale Supélec (2017)
17. Feliot, P., Bect, J., Vazquez, E.: A Bayesian approach to constrained single- and multi-objective optimization. Journal of Global Optimization **67**(1-2), 97–133 (2017)
18. Ginsbourger, D., Le Riche, R.: Towards Gaussian process-based optimization with finite time horizon. In: mODa 9 – Advances in Model-Oriented Design and Analysis, pp. 89–96. Springer (2010)
19. Harrington, E.C.: The desirability function. Industrial quality control **21**(10), 494–498 (1965)
20. Jones, D.R., Schonlau, M., Welch, W.J.: Efficient global optimization of expensive black-box functions. Journal of Global Optimization **13**(4), 455–492 (1998)
21. Knowles, J., Corne, D.: On metrics for comparing nondominated sets. In: Proceedings of the 2002 Congress on Evolutionary Computation, 2002. CEC'02. vol. 1, pp. 711–716. IEEE (2002)
22. Kushner, H.J.: 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)
23. Laumanns, M., Rudolph, G., Schwefel, H.P.: Approximating the pareto set: Concepts, diversity issues, and performance assessment. Secretary of the SFB 531 (1999)
24. Mockus, J., Tiesis, V., Žilinskas, A.: The application of Bayesian methods for seeking the extremum. In: Dixon, L.C.W., Szegö, G.P. (eds.) Towards Global Optimization. vol. 2, pp. 117–129. North Holland, New York (1978)
25. Santner, T.J., Williams, B.J., Notz, W.: The design and analysis of computer experiments. Springer (2003)
26. Schonlau, M., Welch, W.J., Jones, D.R.: Global versus local search in constrained optimization of computer models. In: New Developments and Applications in Experimental Design: Selected Proceedings of a 1997 Joint AMS-IMS-SIAM Summer Conference. IMS Lecture Notes-Monographs Series, vol. 34, pp. 11–25. Institute of Mathematical Statistics (1998)
27. Wagner, T., Trautmann, H.: Integration of preferences in hypervolume-based multiobjective evolutionary algorithms by means of desirability functions. IEEE Transactions on Evolutionary Computation **14**(5), 688–701 (2010)
28. Williams, C.K.I., Rasmussen, C.: Gaussian processes for machine learning. the MIT Press **2**(3), 4 (2006)
29. Zitzler, E., Brockhoff, D., Thiele, L.: The hypervolume indicator revisited: On the design of pareto-compliant indicators via weighted integration. In: Evolutionary multi-criterion optimization. pp. 862–876. Springer (2007)

30. Zitzler, E., Thiele, L.: Multiobjective optimization using evolutionary algorithms–a comparative case study. In: Parallel problem solving from nature–PPSN V. pp. 292–301. Springer (1998)