Multiobjective optimization using Gaussian process emulators via stepwise uncertainty reduction

Victor Picheny

Multiobjective optimization using Gaussian process emulators via stepwise uncertainty reduction

Victor Picheny
INRA, 31326 Castanet Tolosan, France
Tel.: +33-5-61 28 54 39
victor.picheny@toulouse.inra.fr

October 2, 2013

Abstract

Optimization of expensive computer models with the help of Gaussian process emulators in now commonplace. However, when several (competing) objectives are considered, choosing an appropriate sampling strategy remains an open question. We present here a new algorithm based on stepwise uncertainty reduction principles to address this issue. Optimization is seen as a sequential reduction of the volume of the excursion sets below the current best solutions, and our sampling strategy chooses the points that give the highest expected reduction. Closed-form formulae are provided to compute the sampling criterion, avoiding the use of cumbersome simulations. We test our method on numerical examples, showing that it provides an efficient trade-off between exploration and intensification.

keywords Kriging; EGO; Pareto front; Excursion sets

1 Introduction

We consider the problem of simultaneous optimization of several objective functions over a design region $\mathbb{X} \subset \mathbb{R}^d$:

$$\min \quad y^{(1)}(\mathbf{x}), \ldots, y^{(q)}(\mathbf{x}),$$

where $y^{(i)} : \mathbb{X} \to \mathbb{R}$ are outputs of a complex computer code. The objectives being typically conflicting, there exists no unique minimizer, and the goal is to identify the set of optimal solutions, called Pareto front (Collette and Siarry, 2003). Defining that a point dominates another if *all* his objectives are better, the Pareto front $\mathbb{X}^*$ is the subset of the non-dominated points in $\mathbb{X}$:

$$\forall \mathbf{x}^* \in \mathbb{X}^*, \forall \mathbf{x} \in \mathbb{X}, \exists k \in \{1, \ldots, q\} \text{ such that } y^{(k)}(\mathbf{x}^*) \leq y^{(k)}(\mathbf{x}).$$

When the computational cost of a single model evaluation is high, a well-established practice consists of using Gaussian process (GP) emulators to approximate the model outputs and guide the optimization process. Following the seminal article of Jones et al. (1998) and its Efficient Global Optimization (EGO) algorithm for single objective optimization, several strategies have been proposed in the past few years to address the multi-objective problem (Knowles, 2006; Keane, 2006; Ponweiser et al., 2008; Wagner et al., 2010). They consist in evaluating sequentially the computer model at the set of inputs that maximizes a so-called *infill criterion*, derived from the GP emulator, that expresses a trade-off between exploration of unsampled areas and sampling intensification in promising regions. While the single objective case has been extensively discussed (Jones, 2001; Wang and Shan, 2007), finding efficient and statistically consistent infill criteria for the multi-objective case remains an open question.

Alternatively to the EGO paradigm, *stepwise uncertainty reduction* (SUR) strategies aim at reducing, by sequential sampling, an uncertainty measure about a quantity of interest. In a single objective optimization context, Villemonteix et al. (2009) defined the Shannon entropy of the maximizer (computed using the GP model) as an uncertainty measure: a smaller entropy implies that the maximizer is well-identified. They show that their approach outperforms the EGO strategy on a series of problems. Another example in a reliability assessment context can be found in Bect et al. (2012). In general, SUR approaches allow to define policies rigorously with respect to a given objective, resulting in very good performances. However, they are often challenging to use in practice, as they rely on very expensive GP simulations.

We propose here a novel SUR strategy to address the multi-objective problem. It is based on a measure of uncertainty of the current identification of the Pareto front $\mathbb{X}^*$, hence avoiding some of the drawbacks of the existing criteria (hierarchy between objectives, difficult-to-tune parameters, etc.). Following Chevalier et al. (2012), explicit formulae for the expected uncertainty reduction are provided, avoiding the need to rely on simulations.

The paper is organized as follows: section 2 presents the GP model and the basics of GP-based optimization. Then, we describe our SUR strategy for a single objective in section 3 and for several objectives in section 4. We provide some numerical experiments in section 5 and compare our method to the state-of-the-art. Finally, advantages and drawbacks of the method are discussed in section 6.

2 Some concepts of Gaussian-process-based optimization

2.1 Gaussian process emulation

We consider first the emulation of a single computer response y. The response is modelled as

$$Y(.) = \mathbf{f}(.)^T \boldsymbol{\beta} + Z(.), \tag{1}$$

where $\mathbf{f}(.)^T = (f_1(.), \ldots, f_p(.))$ is a vector of trend functions, $\boldsymbol{\beta}$ a vector of (unknown) coefficients and $Z(.)$ is a Gaussian process (GP) Z with zero mean and known covariance kernel k (Cressie, 1993; Rasmussen and Williams, 2006). Let us call $\mathcal{A}_n$ the event:

$$\{Y(\mathbf{x}_1) = y_1, \ldots, Y(\mathbf{x}_n) = y_n\};$$

conditionally on $\mathcal{A}_n$, the mean and covariance of Y are given by:

$$\begin{aligned}
m_n(\mathbf{x}) &= \mathbb{E}\left(Y(\mathbf{x})\middle|\mathcal{A}_n\right) = \\
&= \mathbf{f}(\mathbf{x})^T\hat{\boldsymbol{\beta}} + \mathbf{k}_n(\mathbf{x})^T\mathbf{K}_n^{-1}(\mathbf{y}_n - \mathbf{F}_n\hat{\boldsymbol{\beta}}), \\
c_n(\mathbf{x},\mathbf{x}') &= cov\left(Y(\mathbf{x}), Y(\mathbf{x}')\middle|\mathcal{A}_n\right) \\
&= k(\mathbf{x},\mathbf{x}') - \mathbf{k}_n(\mathbf{x})^T\mathbf{K}_n^{-1}\mathbf{k}_n(\mathbf{x}') \\
&+ \left(\mathbf{f}(\mathbf{x})^T - \mathbf{k}_n(\mathbf{x})^T\mathbf{K}_n^{-1}\mathbf{F}_n\right)^T\left(\mathbf{F}_n^T\mathbf{K}_n^{-1}\mathbf{F}_n\right)^{-1} \\
&\quad \left(\mathbf{f}(\mathbf{x}')^T - \mathbf{k}_n(\mathbf{x}')^T\mathbf{K}_n^{-1}\mathbf{F}_n\right),
\end{aligned}$$

where

- $\mathbf{y}_n = (y_1, \ldots, y_n)^T$ are the observations,
- $\mathbf{K}_n = (k(\mathbf{x}_i, \mathbf{x}_j))_{1 \le i,j \le n}$ is the observation covariance matrix,
- $\mathbf{k}_n(\mathbf{x})^T = (k(\mathbf{x}, \mathbf{x}_1), \ldots, k(\mathbf{x}, \mathbf{x}_n))$,
- $\mathbf{F}_n = \left(\mathbf{f}(\mathbf{x}_1)^T, \ldots, \mathbf{f}(\mathbf{x}_n)^T\right)^T$, and
- $\hat{\boldsymbol{\beta}} = \left(\mathbf{F}_n^T\mathbf{K}_n^{-1}\mathbf{F}_n\right)^{-1}\mathbf{F}_n^T\mathbf{K}_n^{-1}\mathbf{y}_n$ is the best linear unbiased estimate of $\boldsymbol{\beta}$.

In addition, the *prediction variance* is defined as

$$s_n^2(\mathbf{x}) = c_n(\mathbf{x}, \mathbf{x}).$$

The covariance kernel depends on parameters that are usually unknown and must be estimated from an initial set of responses. Typically, maximum likelihood estimates are obtained by numerical optimization and used as face value, the estimates being updated when new observations are added to the model. The reader can refer to Stein (1999) (chapter 6), Rasmussen and Williams (2006) (chapter 5) or Roustant et al. (2012) for detailed calculations and implementation issues.

When several functions $y^{(1)}, \ldots, y^{(q)}$ are predicted simultaneously, it is possible to take their dependency into account (Kennedy and O'Hagan, 2001; Craig et al., 2001). However, in this work we consider all the processes $Y^{(i)}$ independent, hence modelled as above, which is in line with current practice.

2.2 Gaussian-process-based optimization with a single objective

The EGO strategy, as well as most of its modifications, is based on the following scheme. An initial set of observations is generated, from which the GP model is constructed and validated. Then, new observations are obtained sequentially, at the point in the design space that maximizes the infill criterion, and the model is updated every time a new observation is added to the training set. The two later steps are repeated until a stopping criterion is met.

The expected improvement criterion (EI) used in EGO relies on the idea that progress is achieved by performing an evaluation at step n if the $(n+1)^{\text{th}}$ design has a lower objective function value than any of the n previous designs. Hence, the *improvement* is defined as the difference between the current observed minimum and the new function value if it is positive, or zero otherwise, and EI is its conditional expectation under the GP model:

$$EI(\mathbf{x}) = \mathbb{E}\left[\max\left(0, y_n^{\min} - Y(\mathbf{x})\right)|\mathcal{A}_n\right],$$

where $y_n^{\min}$ denotes the current minimum of y found at step n: $y_n^{\min} = \min(y_1, \ldots, y_n)$.

EGO is the one-step optimal strategy (in expectation) regarding improvement: at step n, the new measurement is chosen as

$$\mathbf{x}_{n+1} = \arg\max_{\mathbf{x}\in\mathbb{X}} EI(\mathbf{x}),$$

which is in practice done by running an optimization algorithm. It has been shown in Jones (2001) that EGO provides, among numerous alternatives, an efficient solution for global optimization.

2.3 Gaussian-process-based optimization with several objectives

Several adaptations of EGO to the multi-objective framework have been proposed; a review can be found in Ponweiser et al. (2008). The main difficulty is that the concept of *improvement* cannot be transferred directly, as the current best point is here a set, and the gain is measured on several objectives simultaneously. In Knowles (2006), the objectives are aggregated in a single function using random weights, which allows using the standard EGO. Keane (2006) derived an EI with respect to multiple objectives. Ponweiser et al. (2008) proposed an hypervolume-based infill criterion, where the improvement is measured in terms of hypervolume increase.

3 Single objective optimization by stepwise uncertainty reduction

We consider first the case of a problem with a single objective y to minimize. In this section, we propose a new strategy in a form similar to EGO that uses an alternative infill criterion based on stepwise uncertainty reduction principles. The adaptation of this criterion to the multi-objective case is presented in Section 4.

3.1 Definition of an uncertainty measure for optimization

The EGO strategy focuses on progress in terms of objective function value. It does not account (or only indirectly) for the knowledge improvement that a new measurement would provide to the GP model, nor for the discrepancy between the location of the current best design found and the actual minimizer (which is actually most users' objective).

Alternative sampling criteria have been proposed to account for these two aspects. In Villemonteix et al. (2009), the IAGO stategy chooses the point that minimizes the posterior Shannon entropy of the minimizer: the interest of performing a new observation is measured in gain of information about the location of the minimizer. Unfortunately, it relies on expensive GP simulations, which makes its use challenging in practice. Gramacy and Lee (2011) proposed an *integrated expected conditional improvement* to measure a global informational gain of an observation. In the noisy case, Scott et al. (2011) proposed a somehow similar *knowledge gradient policy* that also measures global information gain. However, as both criteria rely on notions of improvement, it makes them difficult to adapt to the multiobjective case. The criterion we propose below address this issue.

Consider that n measurements have been performed. As a measure of performance regarding the optimization problem, we consider the expected volume of excursion set below the current minimum $y_n^{\min}$:

$$ev_n = \mathbb{E}_{\mathbb{X}}\left[\mathbb{P}\left(Y(\mathbf{x}) \leq y_n^{\min} | \mathcal{A}_n\right)\right]. \quad (2)$$

Similarly to the Shannon entropy measure in IAGO, a large volume indicates that the optimum is not yet precisely located (see Figure 1); on the contrary, a small volume indicates that very little can be gained by pursuing the optimization process. Following the stepwise uncertainty reduction paradigm, this volume is an uncertainty measure related to our objective (finding the minimizer of y); minimizing the uncertainty amounts to solving the optimization problem.

The probability $p_n(\mathbf{x}, y_n^{\min}) := \mathbb{P}\left(Y(\mathbf{x}) \leq y_n^{\min} | \mathcal{A}_n\right)$, which is often referred to as *probability of improvement* (Jones, 2001), can be expressed in closed form, and Eq. (2) writes:

$$ev_n = \int_{\mathbb{X}} p_n(\mathbf{x}, y_n^{\min}) d\mathbf{x} = \int_{\mathbb{X}} \Phi\left(\frac{y_n^{\min} - m_n(\mathbf{x})}{s_n(\mathbf{x})}\right) d\mathbf{x}, \quad (3)$$

where $\Phi(.)$ is the cumulative distribution function (CDF) of the standard Gaussian distribution. Hypothesizing that a measurement y_{n+1} is performed at a point $\mathbf{x}_{n+1}$, its benefit can be measured by the reduction of the expected volume of excursion set $\Delta = ev_n - ev_{n+1}$, with:

$$\begin{aligned} ev_{n+1} &= \int_{\mathbb{X}} p_{n+1}(\mathbf{x}, \min\left(y_n^{\min}, y_{n+1}\right)) d\mathbf{x} \\ &= \int_{\mathbb{X}} \Phi\left(\frac{\min\left(y_n^{\min}, y_{n+1}\right) - m_{n+1}(\mathbf{x})}{s_{n+1}(\mathbf{x})}\right) d\mathbf{x}. \end{aligned}$$

Of course, ev_{n+1} cannot be known exactly without evaluating y_{n+1}. However, we show in the following that its expectation can be calculated in closed form, leading to a suitable infill criterion. To do so, we first formulate a series of propositions in the next subsection.

3.2 Probabilities updates

An interesting property of the GP model is that, when a new observation $y_{n+1} = y(\mathbf{x}_{n+1})$ is added to the training set, its new predictive distribution can be expressed simply as a function of the old one (Emery, 2009):

$$\begin{aligned} m_{n+1}(\mathbf{x}) &= m_n(\mathbf{x}) + \frac{c_n(\mathbf{x}, \mathbf{x}_{n+1})}{c_n(\mathbf{x}_{n+1}, \mathbf{x}_{n+1})}\left(y_{n+1} - m_n\left(\mathbf{x}_{n+1}\right)\right); \\ s_{n+1}^2(\mathbf{x}) &= s_n^2(\mathbf{x}) - \frac{c_n(\mathbf{x}, \mathbf{x}_{n+1})^2}{s_n^2(\mathbf{x}_{n+1})}. \end{aligned} \quad (4)$$

Note that only $m_{n+1}(\mathbf{x})$ depends on the value of the new observation y_{n+1}. Now, conditionally on the n first observations, Y_{n+1} is a random variable (as the new observation has not yet been performed) with its moments given by the GP model:

$$Y_{n+1} \sim \mathcal{N}\left(m_n(\mathbf{x}_{n+1}), s_n^2(\mathbf{x}_{n+1})\right).$$

We can then define the future expectation $M_{n+1}(\mathbf{x})$ (or any quantity depending on it) as a random variable conditionally on $\mathcal{A}_n$ and on the fact that the next observation will be at $\mathbf{x}_{n+1}$. This applies to any quantity depending on Y_{n+1} or $M_{n+1}(\mathbf{x})$, for instance, the probability of being below a threshold $a \in \mathbb{R}$:

$$P_{n+1}(\mathbf{x}, a) = \Phi\left(\frac{a - M_{n+1}(\mathbf{x})}{s_{n+1}(\mathbf{x})}\right).$$

Proposition 3.1. *Without any restriction on the value of Y_{n+1}, the expectation of the future probability of being below the threshold is equal to the current probability:*

$$\begin{aligned} \mathbb{P}\left(Y(\mathbf{x}) \leq a | \mathcal{A}_n, Y(\mathbf{x}_{n+1}) = Y_{n+1}\right) &= \mathbb{E}\left[P_{n+1}(\mathbf{x}, a) \middle| \mathcal{A}_n\right] \\ &= p_n(\mathbf{x}, a). \end{aligned}$$

Proposition 3.2. *Conditioning further by $Y_{n+1} \leq b$, the probability expectation writes in simple form using the Gaussian bivariate CDF:*

$$\begin{aligned} q(\mathbf{x}, b, a) &:= \mathbb{P}\left[Y(\mathbf{x}) \leq a \middle| \mathcal{A}_n, Y(\mathbf{x}_{n+1}) = Y_{n+1}, Y_{n+1} \leq b\right] \\ &\times \mathbb{P}\left[Y_{n+1} \leq b \middle| \mathcal{A}_n\right] \\ &= \mathbb{E}\left[P_{n+1}(\mathbf{x}, a) \times 1_{Y_{n+1} \leq b} \middle| \mathcal{A}_n\right] \\ &= \mathbf{\Phi}_\rho\left(\bar{b}, \tilde{a}\right), \end{aligned} \quad (5)$$

where $\mathbf{\Phi}_\rho$ is the Gaussian bivariate CDF with zero mean and covariance $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, $\bar{b} = \frac{b - m_n(\mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1})}$, $\tilde{a} = \frac{a - m_n(\mathbf{x})}{s_n(\mathbf{x})}$ and $\rho = \frac{c_n(\mathbf{x}, \mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1}) s_n(\mathbf{x})}$.

Corollary 3.3. *Similarly, conditioning by $Y_{n+1} \geq b$ leads to:*

$$\begin{aligned} r(\mathbf{x}, b, a) &:= \mathbb{P}\left[Y(\mathbf{x}) \leq a \middle| \mathcal{A}_n, Y(\mathbf{x}_{n+1}) = Y_{n+1}, Y_{n+1} \geq b\right] \\ &\times \mathbb{P}\left[Y_{n+1} \geq b \middle| \mathcal{A}_n\right] \\ &= \mathbf{\Phi}_{-\rho}\left(-\bar{b}, \tilde{a}\right). \end{aligned} \quad (6)$$

The final proposition resembles Proposition 3.2, but the fixed threshold a is here replaced by Y_{n+1}:

Proposition 3.4. *The expectation of the probability that $Y(\mathbf{x})$ is smaller than Y_{n+1}, conditionally on $Y_{n+1} \leq b$, is given by:*

$$\begin{aligned} h(\mathbf{x}, b) &:= \mathbb{P}\big[Y(\mathbf{x}) \leq Y_{n+1} | \mathcal{A}_n, Y(\mathbf{x}_{n+1}) \\ &= Y_{n+1}, Y_{n+1} \leq b\big]\mathbb{P}\big[Y_{n+1} \leq b \big| \mathcal{A}_n\big] \\ &= \mathbb{E}\left[P_{n+1}(\mathbf{x}, Y_{n+1}) \times 1_{Y_{n+1} \leq b} \big| \mathcal{A}_n\right] \\ &= \mathbf{\Phi}_\nu\left(\bar{b}, \eta\right), \end{aligned} \tag{7}$$

with:

$$\eta = \frac{m_n(\mathbf{x}_{n+1}) - m_n(\mathbf{x})}{\sqrt{s_n^2(\mathbf{x}) + s_n^2(\mathbf{x}_{n+1}) - 2c_n(\mathbf{x}, \mathbf{x}_{n+1})}} \text{ and}$$

$$\nu = \frac{c_n(\mathbf{x}, \mathbf{x}_{n+1}) - s_n^2(\mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1})\sqrt{s_n^2(\mathbf{x}) + s_n^2(\mathbf{x}_{n+1}) - 2c_n(\mathbf{x}, \mathbf{x}_{n+1})}}.$$

All the proofs are reported in Appendix A.

3.3 A Stepwise uncertainty reduction criterion

Coming back to the SUR criterion, at step n the future volume of excursion set EV_{n+1} is a random variable, and its expectation is:

$$\begin{aligned} EEV(\mathbf{x}_{n+1}) &:= \mathbb{E}\left(EV_{n+1} \big| \mathcal{A}_n, Y(\mathbf{x}_{n+1}) = Y_{n+1}\right) \\ &= \int_{\mathbb{X}} \mathbb{E}\Big[\Phi\left(\frac{\min\left(y_n^{\min}, Y_{n+1}\right) - M_{n+1}(\mathbf{x})}{s_{n+1}(\mathbf{x})}\right) \\ &\quad \big| \mathcal{A}_n, Y(\mathbf{x}_{n+1}) = Y_{n+1}\Big] d\mathbf{x}. \end{aligned}$$

Let $\varphi\left(y_{n+1}\right)$ be the probability density function (PDF) of Y_{n+1} conditionally on $\mathcal{A}_n$. We have:

$$\begin{aligned} & EEV(\mathbf{x}_{n+1}) \\ = & \int_{\mathbb{X}} \int_{\mathbb{R}} \Phi\left(\frac{\min\left(y_n^{\min}, y_{n+1}\right) - m_{n+1}(\mathbf{x})}{s_{n+1}(\mathbf{x})}\right) d\varphi\left(y_{n+1}\right) d\mathbf{x} \\ = & \int_{\mathbb{X}} \Big[\int_{-\infty}^{y_n^{\min}} \Phi\left(\frac{y_{n+1} - m_{n+1}(\mathbf{x})}{s_{n+1}(\mathbf{x})}\right) \\ + & \int_{y_n^{\min}}^{+\infty} \Phi\left(\frac{y_n^{\min} - m_{n+1}(\mathbf{x})}{s_{n+1}(\mathbf{x})}\right) d\varphi\left(y_{n+1}\right) \Big] d\mathbf{x} \\ = & \int_{\mathbb{X}} \left[h(\mathbf{x}, y_n^{\min}) + r(\mathbf{x}, y_n^{\min}, y_n^{\min})\right] d\mathbf{x}. \end{aligned}$$

The first term of the integrand is given by Eq. (7) in Proposition 3.4, with $b = y_n^{\min}$, and the second term is given by Eq. (6) in Corrolary 3.3, with $a = b = y_n^{\min}$, hence:

$$EEV(\mathbf{x}_{n+1}) = \int_{\mathbb{X}} \left[\mathbf{\Phi}_\nu\left(\overline{y}_n^{\min}, \eta\right) + \mathbf{\Phi}_\rho\left(-\overline{y}_n^{\min}, \widetilde{y}_n^{\min}\right)\right] d\mathbf{x}, \tag{8}$$

with:

$$\overline{y}_n^{\min} = (y_n^{\min} - m_n(\mathbf{x}_{n+1}))/s_n(\mathbf{x}_{n+1})$$

and

$$\widetilde{y}_n^{\min} = (y_n^{\min} - m_n(\mathbf{x}))/s_n(\mathbf{x}).$$

The SUR optimization strategy consists in adding the experiment that minimizes the expected volume of excursion set (or maximizes the difference), that is, the one-step optimal policy in terms of reduction of the uncertainty on the objective function minimizer:

$$\mathbf{x}_{n+1} = \arg\min_{\mathbf{x}^+ \in \mathbb{X}} EEV(\mathbf{x}^+) \tag{9}$$

Remark In general, the probability of improvement $p_n(\mathbf{x}, y_n^{\min})$ is high where the prediction mean $m_n(.)$ is low and/or the prediction variance $s_n^2(.)$ is high. Simply choosing points that maximize $p_n(\mathbf{x}, y_n^{\min})$ is known to be inefficient (Jones, 2001), as it does not consider the amplitude of the gain in the objective function. Here, $EEV(\mathbf{x}^+)$ strongly depends on the potential gain amplitude. Indeed, minimizing the expected volume relies on two mecanisms: reducing the local uncertainty and lowering the current minimum value ($y_n^{\min}$). The first is achieved by adding measurements in unsampled regions (high $s_n^2(.)$), the second in regions where this potential reduction is high. Hence, the EEV criterion can be seen as a mixed measure of uncertainty on the current minimum location and of potential gain in the objective function.

3.4 Illustration

Figure 1 illustrates the concept of reduction of volume of excursion on a toy example. A GP model is built on a six-point training set, from which the probability of improvement $p_6(\mathbf{x}, y_6^{\min})$ is computed for every point in $\mathbb{X} = [0, 1]$. We see that it can be intepreted as an indicator of the uncertainty we have about the location of the actual minimizer $\mathbf{x}^* = 0.47$, as the model can only predict that $\mathbf{x}^*$ is likely to be between 0.4 and 0.6. Then, we consider two candidate points ($\mathbf{x}^+ = 0.2$ and $\mathbf{x}^+ = 0.5$) and compute, for each, the expected new probability (integrand in Eq. (8)). We see that the probability is likely to remain mostly unchanged by adding the measurement at $\mathbf{x}^+ = 0.2$ (which is, indeed, a region with high response value), while it would be considerably reduced by adding a measurement at $\mathbf{x}^+ = 0.5$. In terms of volume of excursion set, we have $EEV(0.2) \approx ev_6$ (no reduction), while $EEV(0.5) \approx ev_6/3$ (large reduction): the EEV criterion clearly indicates $\mathbf{x}^+ = 0.5$ as a better sampling location.

4 Multi-objective optimization by stepwise uncertainty reduction

4.1 Volume of excursion behind the Pareto front

Let $\mathbf{y}(\mathbf{x}) = \left(y^{(1)}(\mathbf{x}), \ldots, y^{(q)}(\mathbf{x})\right)$ be the vector of objective functions to minimize. A point $\mathbf{x}$ dominates another point $\mathbf{x}'$ if $y^{(k)}(\mathbf{x}) \leq y^{(k)}(\mathbf{x}')$ for all k in $\{1, \ldots, q\}$, which we denote by $\mathbf{x}' \prec \mathbf{x}$ in the following. At step n, $\mathbf{X}_n = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$ is the current experimental set and $\mathbf{Y}_n = \{\mathbf{y}(\mathbf{x}_1), \ldots, \mathbf{y}(\mathbf{x}_n)\}$ the corresponding set of measures. The non-dominated subset $\mathbf{X}_n^*$ of $\mathbf{X}_n$ constitutes the *current Pareto front* (of size $m \leq n$). In the objective

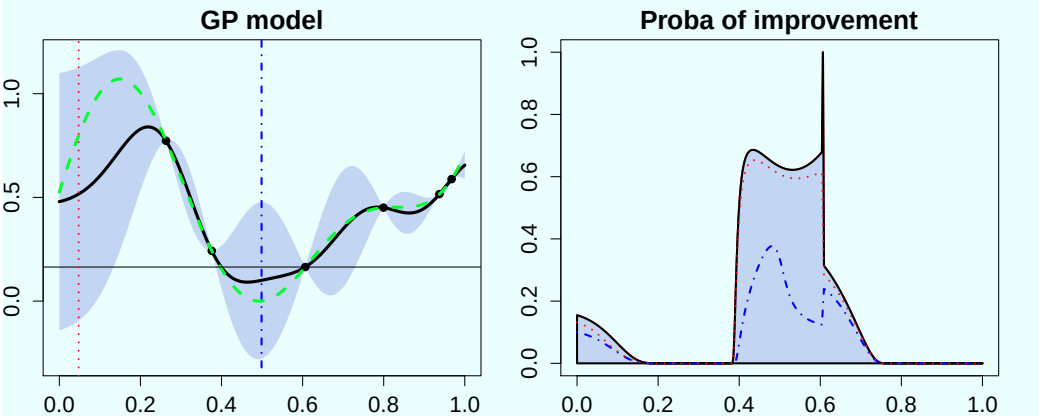

Figure 1: Illustration of the effect of a new observation on the EV criterion. Left: actual objective function (dotted line), GP model (depicted by its mean in black plain line and 95% confidence interval in grey) based on six observations (black circles). The horizontal line shows the current minimum; the vertical bars are placed at two candidate locations. Right: probability of improvement given by the current model and expected updated probability for each candidate. Adding a point at $x^+ = 0.5$ (mixed line) is likely to reduce substantially the probability, while adding a point at $x^+ = 0.05$ (dotted line) has little expected effect.

space, the corresponding subset $\mathbf{Y}_n^*$ separates the regions dominated and not dominated by the experimental set.

Then, we decompose the objective space plane using a tesselation $\{\Omega_i\}_{i\in\{1,\dots,I\}}$ of size $I = (m+1)^q$ ($\cup_{i\in I}\Omega_i = \mathbb{R}^q$ and $\cap_{i\in I}\Omega_i = \emptyset$), each cell being a hyperrectangle defined as:

$$\Omega_i = \{\mathbf{y} \in \mathbb{R}^q | y_{i-}^{(k)} \leq y^{(k)} < y_{i+}^{(k)}, k \in \{1,\dots,q\}\}.$$

Each couple $(y_{i-}^{(k)}, y_{i+}^{(k)})$ consists of two consecutive values of the vector $\left[-\infty, y^{(k)}(\mathbf{x}_1^*), \dots, y^{(k)}(\mathbf{x}_m^*), +\infty\right]$. An illustration is given in Figure 2.

A cell Ω_i *dominates* another cell Ω_j ($\Omega_j \prec \Omega_i$) if any point in Ω_i dominates any point in Ω_j, and it *partially dominates* Ω_j if there exists a point in Ω_j that is dominated by any point in Ω_i. Otherwise, we say that Ω_j is not dominated by Ω_i ($\Omega_j \not\prec \Omega_i$).

We denote by I^* the indices of all the non-dominated cells at step n, that is, the cells that are not dominated by any point of $\mathbf{X}_n^*$. In two dimensions, the non-dominated cells are located in the bottom left half of the plane (Figure 2).

Now, let us assume that GP models are fitted to each objective $y^{(k)}$. At step n, the probability that $\mathbf{Y}(\mathbf{x})$ belongs to the cell Ω_i is:

$$\begin{aligned} p_n^i(\mathbf{x}) &= \mathbb{P}\left[\mathbf{Y}(\mathbf{x}) \in \Omega_i \middle| \mathcal{A}_n\right] \\ &= \prod_{k=1}^{q} \Phi\left(\frac{y_{i+}^{(k)} - m_n^{(k)}(\mathbf{x})}{s_n^{(k)}(\mathbf{x})}\right) - \Phi\left(\frac{y_{i-}^{(k)} - m_n^{(k)}(\mathbf{x})}{s_n^{(k)}(\mathbf{x})}\right) \\ &:= \prod_{k=1}^{q} p_n^{i(k)} \end{aligned}$$

by pairwise independence of $Y^{(1)}, \dots, Y^{(q)}$. The probability that $\mathbf{x}$ is not dominated by any point of $\mathbf{X}_n$ is then the probability that $\mathbf{Y}(\mathbf{x})$ belongs to one of the non-dominated parts of the objective space. As the Ω_i's are disjoint, it is equal to:

$$\mathbb{P}(\mathbf{x} \not\prec \mathbf{X}_n \middle| \mathcal{A}_n) = \sum_{i\in I^*} p_n^i(\mathbf{x}). \qquad (10)$$

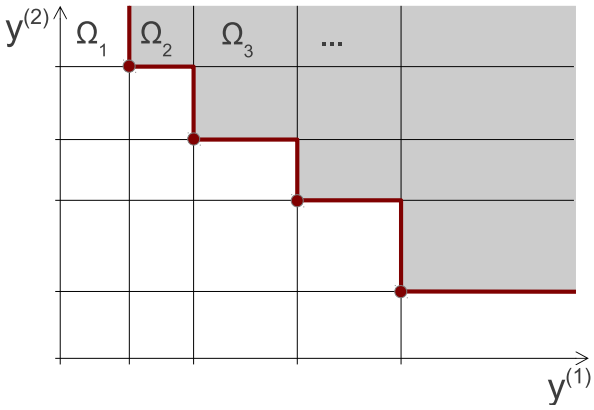

Figure 2: Example of Pareto front generated by four points (circles), and associated tesselation. The grey area corresponds to the dominated cells.

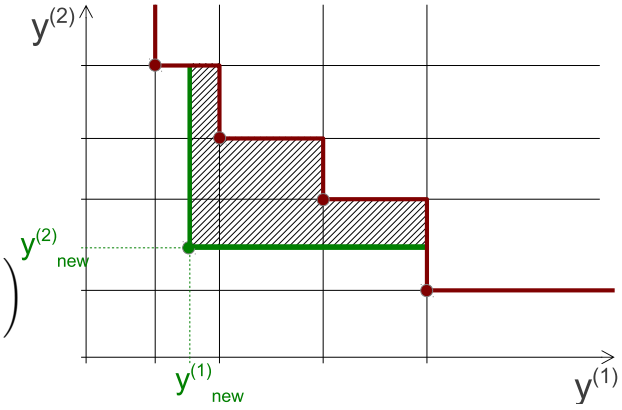

Figure 3: Example of Pareto front modification due to a new measurement. Two points are removed from the Pareto front while the new point is added. The hatched area represents the additional dominated region.

Finally, the volume of the excursion sets behind the Pareto front is equal to the integral of this probability over $\mathbb{X}$:

$$\begin{aligned} ev_n &= \int_{\mathbb{X}} \mathbb{P}(\mathbf{x} \nprec \mathbf{X}_n \big| \mathcal{A}_n) d\mathbf{x} \\ &= \int_{\mathbb{X}} \sum_{i \in I^*} p_n^i(\mathbf{x}) d\mathbf{x} = \sum_{i \in I^*} \int_{\mathbb{X}} p_n^i(\mathbf{x}) d\mathbf{x}. \quad (11) \end{aligned}$$

When ev_n is high, a large proportion of the design space is likely to be better than the current Pareto set; inversely, when $\mathbf{X}_n^*$ approaches the actual Pareto set $\mathbb{X}^*$, the volume tends to zero. Hence, it defines naturally an uncertainty indicator for a SUR strategy.

4.2 SUR criterion derivation

Now, let us consider that a measurement $\mathbf{y}_{n+1}$ is performed at a point $\mathbf{x}_{n+1}$. Compared to step n, the volume ev is modified by two means. First, the new measurement will modify the quantities $m_n^{(k)}(\mathbf{x})$, $s_n^{(k)}(\mathbf{x})$ ($k \in \{1, \ldots, q\}$), hence, the probabilities $p_n^i(\mathbf{x})$. Second, if the new measurement is not dominated by the current Pareto set, it modifies the Pareto optimal front, as the new value $\mathbf{y}(\mathbf{x}_{n+1})$ is added to $\mathbf{Y}^*$ and the values of $\mathbf{Y}^*$ dominated by $\mathbf{y}(\mathbf{x}_{n+1})$ (if they exist) are removed. An example of such update is given in Figure 3.

Focusing on the probability that a point remains non-dominated (Eq. (10)), accounting for the modifications of the models is relatively easy (that is, computing the quantity $p_{n+1}^i(.)$), but accounting for modifications in the Pareto front is complex, as both the number of elements and their values might change. To address this issue, we consider that the updated probability $\mathbb{P}(\mathbf{x} \nprec \mathbf{X}_{n+1} \big| \mathcal{A}_{n+1})$ can be computed using the same sum as for $\mathbb{P}(\mathbf{x} \nprec \mathbf{X}_n \big| \mathcal{A}_n)$ (Eq. (10)) by modifying its elements $p_n^i(\mathbf{x})$:

$$\mathbb{P}(\mathbf{x} \nprec \mathbf{X}_{n+1} \big| \mathcal{A}_{n+1}) = \sum_{i \in I^*} \tilde{p}_{n+1}^i(\mathbf{x}),$$

with

$$\tilde{p}_{n+1}^i(\mathbf{x}) = \mathbb{P}\Big(\mathbf{x} \nprec \mathbf{X}_{n+1} \cap \mathbf{Y}(\mathbf{x}) \in \Omega_i \Big| \mathcal{A}_n, \mathbf{y}(\mathbf{x}_{n+1}) = \mathbf{y}_{n+1}\Big)$$

and the Ω_i's defined using $\mathbf{Y}_n^*$ (not $\mathbf{Y}_{n+1}^*$).

Seing from step n, the $\tilde{P}_{n+1}^j(\mathbf{x})$ are random, as

$$Y(\mathbf{x}_{n+1})^{(k)} \sim \mathcal{N}\left(m_n^{(k)}(\mathbf{x}_{n+1}), s_n^{(k)2}(\mathbf{x}_{n+1})\right),$$

$\forall k \in \{1, \ldots, q\}$.

The expectation of the new volume is then:

$$\begin{aligned} EEV(\mathbf{x}_{n+1}) &= \mathbb{E}\left[\int_{\mathbb{X}} \sum_{j \in I^*} \tilde{P}_{n+1}^j(\mathbf{x}) d\mathbf{x}\right] \\ &= \sum_{j \in I^*} \int_{\mathbb{X}} \mathbb{E}\left[\tilde{P}_{n+1}^j(\mathbf{x})\right] d\mathbf{x}. \end{aligned}$$

This expression can be decomposed by conditioning on the range values of the new observation (using the fact that the Ω_i's are disjoint):

$$\begin{aligned} &\mathbb{E}\left[\tilde{P}_{n+1}^j(\mathbf{x})\right] \\ &= \sum_{i \in I} \mathbb{P}_{n+1}\Big[\quad \mathbf{x} \quad \nprec \mathbf{X}_{n+1} \cap \mathbf{Y}(\mathbf{x}) \in \Omega_j \Big| \mathbf{Y}_{n+1} \in \Omega_i\Big] \\ &\qquad\qquad \times \quad \mathbb{P}_{n+1}\Big[\mathbf{Y}_{n+1} \in \Omega_i\Big] \\ &:= \sum_{i \in I} p_{ij}(\mathbf{x}), \end{aligned}$$

where $\mathbb{P}_{n+1}$ is the probability conditional on $\Big(\mathcal{A}_n, \mathbf{Y}(\mathbf{x}_{n+1}) = \mathbf{Y}_{n+1}\Big)$.

We first note from Proposition 3.1 that

$$\mathbb{P}_{n+1}\left[\mathbf{Y}_{n+1} \in \Omega_i\right] = p_n^i(\mathbf{x}_{n+1}).$$

Then, leaving aside non-domination, the probability that $\mathbf{Y}(\mathbf{x})$ belongs to Ω_j knowing that $\mathbf{Y}_{n+1}$ belongs to Ω_i is given by:

$$\mathbb{P}_{n+1}\Big[\mathbf{Y}(\mathbf{x}) \in \Omega_j \Big| \mathbf{Y}_{n+1} \in \Omega_i\Big] \times p_n^i(\mathbf{x}_{n+1}) = \prod_{k=1}^{q} b_{ij}^{(k)}(\mathbf{x}),$$

with:

$$\begin{aligned} b_{ij}^{(k)}(\mathbf{x}) &= \mathbb{P}_{n+1}\Big[y_{j-}^{(k)} \leq Y^{(k)}(\mathbf{x}) < y_{j+}^{(k)} \Big| y_{i-}^{(k)} \leq Y_{n+1}^{(k)} < y_{i+}^{(k)}\Big] \\ &\times \quad p_n^{i(k)}(\mathbf{x}_{n+1}), \end{aligned}$$

$$p_n^{i(k)}(\mathbf{x}_{n+1}) = \mathbb{P}_n\left[y_{i-}^{(k)} \leq Y_{n+1}^{(k)} < y_{i+}^{(k)}\right],$$

by pairwise independence of $Y^{(1)}, \ldots, Y^{(q)}$. We show in Appendix B that $b_{ij}^{(k)}(\mathbf{x})$ can be expressed in closed form as:

$$\begin{aligned} b_{ij}^{(k)}(\mathbf{x}) &= \mathbf{\Phi}_\rho^{(k)}\left(\overline{y_{i+}^{(k)}}, \widetilde{y_{j+}^{(k)}}\right) - \mathbf{\Phi}_\rho^{(k)}\left(\overline{y_{i+}^{(k)}}, \widetilde{y_{j-}^{(k)}}\right) \\ &- \mathbf{\Phi}_\rho^{(k)}\left(\overline{y_{i-}^{(k)}}, \widetilde{y_{j+}^{(k)}}\right) + \mathbf{\Phi}_\rho^{(k)}\left(\overline{y_{i-}^{(k)}}, \widetilde{y_{j-}^{(k)}}\right) \end{aligned}$$

with the notations introduced in Section 3.2.

Now, we define:

$$\begin{aligned} d_{ij}^{(k)}(\mathbf{x}) &= \mathbb{P}_{n+1}\Big[y_{j-}^{(k)} \leq Y^{(k)}(\mathbf{x}) < y_{j+}^{(k)} \cap Y_{n+1}^{(k)} \leq Y^{(k)}(\mathbf{x}) \\ &\Big| \quad y_{i-}^{(k)} \leq Y_{n+1}^{(k)} < y_{i+}^{(k)}\Big] \times p_n^{i(k)}(\mathbf{x}_{n+1}), \end{aligned}$$

which is identical to $b_{ij}^{(k)}(\mathbf{x})$ with the additional condition $Y_{n+1}^{(k)} \leq Y^{(k)}(\mathbf{x})$. This condition is met when the k-th component of the new observation dominates the k-th component of $\mathbf{Y}(\mathbf{x})$. We have $\mathbf{x} \prec \mathbf{x}_{n+1}$ only if the condition $Y_{n+1}^{(k)} \leq Y^{(k)}(\mathbf{x})$ is met for all components, hence, with probability of occurence $\prod_{k=1}^{p} d_{ij}^{(k)}(\mathbf{x})$. Three cases arise:

- $y_{i-}^{(k)} \geq y_{j+}^{(k)}$: the component cannot be dominated, which implies $d_{ij}^{(k)}(\mathbf{x}) = 0$;
- $y_{i+}^{(k)} \leq y_{j-}^{(k)}$: the component is always dominated, which implies $d_{ij}^{(k)}(\mathbf{x}) = b_{ij}^{(k)}(\mathbf{x})$;

- $y_{i+}^{(k)} = y_{j+}^{(k)}$ (and $y_{i-}^{(k)} = y_{j-}^{(k)}$): $Y^{(k)}(\mathbf{x})$ and $Y_{n+1}^{(k)}$ share the same interval of variation, and:

$$\begin{aligned} d_{ij}^{(k)}(\mathbf{x}) &= \mathbb{P}_{n+1}\left[F_{n+1}^{(k)} \leq Y^{(k)}(\mathbf{x}) < y_{i+}^{(k)} \middle| y_{i-}^{(k)} \leq Y_{n+1}^{(k)} < y_{i+}^{(k)}\right] \\ &\times \; p_n^{i(k)}(\mathbf{x}_{n+1}), \end{aligned}$$

which is equal (as shown in Appendix B) to:

$$\begin{aligned} d_{ij}^{(k)}(\mathbf{x}) &= \mathbf{\Phi}_{\rho}^{(k)}\left(\overline{y_{i+}^{(k)}}, \widetilde{y_{j+}^{(k)}}\right) - \mathbf{\Phi}_{\nu}^{(k)}\left(\overline{y_{i+}^{(k)}}, \eta^{(k)}\right) \\ &+ \mathbf{\Phi}_{\nu}^{(k)}\left(\overline{y_{i-}^{(k)}}, \eta^{(k)}\right) - \mathbf{\Phi}_{\rho}^{(k)}\left(\overline{y_{i-}^{(k)}}, \widetilde{y_{j+}^{(k)}}\right). \end{aligned}$$

The probability of $\mathbf{Y}(\mathbf{x})$ being non-dominated while in Ω_j (and $\mathbf{Y}_{n+1}$ being in Ω_i) is then:

$$p_{ij}(\mathbf{x}) = \prod_{k=1}^{q} b_{ij}^{(k)}(\mathbf{x}) - \prod_{k=1}^{q} d_{ij}^{(k)}(\mathbf{x}).$$

If $\Omega_j \prec \Omega_i$, the new observation dominates any point in Ω_j, hence $d_{ij}^{(k)}(\mathbf{x}) = b_{ij}^{(k)}(\mathbf{x})$ for all k, which gives $p_{ij}(\mathbf{x}) = 0$. Inversely, if $\Omega_j \not\prec \Omega_i$, the new observation cannot dominate any point in the cell Ω_j. We have $d_{ij}^{(k)}(\mathbf{x}) = 0$ for at least one value of k, and $p_{ij}(\mathbf{x})$ is the probability that $\mathbf{Y}(\mathbf{x})$ belongs to Ω_j: $p_{ij}(\mathbf{x}) = \prod_{k=1}^{q} b_{ij}^{(k)}(\mathbf{x})$.

Finally, for a given point $\mathbf{x} \in \mathbb{X}$, we compute the probability that it is non-dominated at step $n+1$ using:

$$\mathbb{P}_{n+1}(\mathbf{x} \not\prec \mathbf{X}_{n+1}) = \sum_{i \in I} \sum_{j \in I^*} p_{ij}(\mathbf{x}),$$

and the SUR criterion is:

$$EEV(\mathbf{x}_{n+1}) = \sum_{i \in I} \sum_{j \in I^*} \int_{\mathbb{X}} p_{ij}(\mathbf{x}) d\mathbf{x}, \tag{12}$$

with:

$$p_{ij}(\mathbf{x}) = \begin{cases} 0 & \text{if } \Omega_j \prec \Omega_i \\ \prod_{k=1}^{q} b_{ij}^{(k)}(\mathbf{x}) & \text{if } \Omega_j \not\prec \Omega_i \\ \prod_{k=1}^{q} b_{ij}^{(k)}(\mathbf{x}) - \prod_{k=1}^{q} d_{ij}^{(k)}(\mathbf{x}) & \text{otherwise} \end{cases} \tag{13}$$

The first sum in Eq. (12) accounts for $\mathbf{Y}_{n+1}$ potentially being in any cell Ω_i; the second sum accounts for $\mathbf{Y}(\mathbf{x})$ potentially being in a non-dominated cell Ω_j.

4.3 Computation

Evaluating the criterion as in Eq. (12) is a non-trivial task; besides, a relatively fast computation is needed, as it may be embedded in an optimization loop to search for the best new observation (Eq. (9)). We provide here some technical solutions to ease its computation. Some of these issues have also been experienced with SUR criteria for inversion, as reported in Chevalier et al. (2012, 2013).

Firstly, as no closed form exists for the integration over the design domain $\mathbb{X}$ in Eq. (12), one may rely on Monte-Carlo integration, with approximations of the form:

$$\int_{\mathbb{X}} p_{ij}(\mathbf{x}) d\mathbf{x} \approx \frac{1}{L} \sum_{l=1}^{L} w^l p_{ij}(\mathbf{x}^l),$$

where the $\mathbf{x}^l$'s and w^l's are integration points and weights, respectively. One solution to alleviate the computational cost is to use a fixed set of integration points while searching for the best new observation. Then, many quantities that do not depend on $\mathbf{x}_{n+1}$ can be precalculated only once beforehand outside the optimization loop, as suggested in Chevalier et al. (2012).

Secondly, the criterion relies on the bivariate normal distribution, which also must be computed numerically. Very efficient programs can be found, such as the R package `pbivnorm` (Kenkel, 2012), which makes this task relatively inexpensive.

Thirdly, the tesselation used in the previous section has $I = (m+1)^q$ elements, with $I^* = I/2$ non-dominated elements, making the computation of the double sum in Eq. (12) intensive. As detailed in Section 4.4 for the two dimensional case, the number of elements can be very substantially reduced by grouping cells together. Note however that such decomposition may not be straightforward in high dimension.

Finally, as the optimization progresses, it is likely that the Pareto set grows, making the criterion more expensive to compute as more cells are to be considered. This problem is shared by all GP-based strategies, and some solutions have been proposed to filter the Pareto set and retain a small representative set (Wagner et al., 2010). Such types of strategies may be applicable to our criterion, as some small cells would contribute to a very small part of the volume of excursion sets and could be neglected without introducing a critical error, and would reduce substantially the computational cost, especially when the number of observations is high.

4.4 Efficient formulas in the two-objective case

We consider here the two-objective case, for which the *EEV* criterion can be expressed in a compact and computationally efficent way. With two objectives, the Pareto set can be ordered as follows (the first and second objective functions in ascending and descending order, respectively): $y_1^{(1)*} \leq \ldots \leq y_m^{(1)*}$ and $y_1^{(2)*} \geq \ldots \geq y_m^{(2)*}$.

The non-dominated part of the objective space can be divided in $m+1$ cells. Then, given a non-dominated cell Ω_j, only four cases arise for Ω_i (the cell of the new observation), as shown in Figure 4, for which the quantities $p_{ij}(\mathbf{x})$ need to be computed.

Hence, the criterion can be expressed as a sum of at most $(m+1) \times 4$ terms. As many terms can be factorized, we finally obtain:

$$EEV(\mathbf{x}_{n+1}) = \sum_{j=0}^{m} \int_{\mathbb{X}} \alpha_j(\mathbf{x}),$$

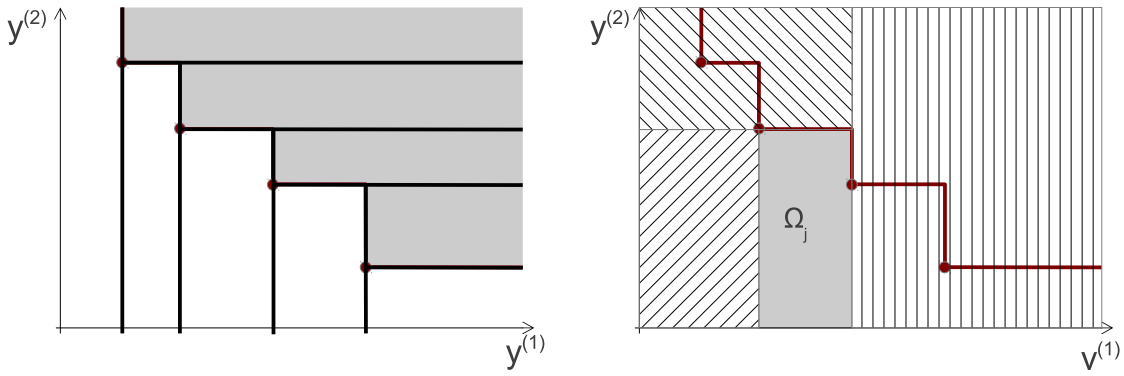

Figure 4: Left: the $m+1$ non-dominated cells (in white). Right: the four cases for Ω_i given Ω_j: three are represented by the different hatched regions, the fourth corresponds to $\Omega_j = \Omega_i$.

with:

$$\begin{aligned}
\alpha_0(\mathbf{x}) &= \left[\mathbf{\Phi}_\rho^{(1)}\left(\bar{y}_1^{(1)*}, \tilde{y}_1^{(1)*}\right) - \mathbf{\Phi}_\nu^{(1)}\left(\bar{y}_1^{(1)*}, \eta^{(1)}\right)\right] \\
&\times \left[\Phi\left(\eta^{(2)}\right) - 1\right] + \Phi\left(\tilde{y}_1^{(1)*}\right), \\
\alpha_j(\mathbf{x}) &= \left[\mathbf{\Phi}_\rho^{(1)}\left(\bar{y}_{j+1}^{(1)*}, \tilde{y}_{j+1}^{(1)*}\right) - \mathbf{\Phi}_\nu^{(1)}\left(\bar{y}_{j+1}^{(1)*}, \eta^{(1)}\right)\right. \\
&+ \left.\mathbf{\Phi}_\nu^{(1)}\left(\bar{y}_j^{(1)*}, \eta^{(1)}\right) - \mathbf{\Phi}_\rho^{(1)}\left(\bar{y}_j^{(1)*}, \tilde{y}_j^{(1)*}\right)\right] \\
&\times \left[\mathbf{\Phi}_\nu^{(2)}\left(\bar{y}_j^{(2)*}, \eta^{(2)}\right) - \mathbf{\Phi}_\rho^{(2)}\left(\bar{y}_j^{(2)*}, y_j^{(2)*}\right)\right] \\
&+ \left[\Phi\left(\tilde{y}_{j+1}^{(1)*}\right) - \Phi\left(\tilde{y}_j^{(1)*}\right)\right] \Phi\left(\tilde{y}_j^{(2)*}\right), \\
\forall j &\in \{1, \ldots, m-1\},
\end{aligned}$$

and:

$$\begin{aligned}
\alpha_m(\mathbf{x}) &= \left[1 - \Phi\left(\eta^{(1)}\right) + \mathbf{\Phi}_\nu^{(1)}\left(\bar{y}_m^{(1)*}, \eta^{(1)}\right)\right. \\
&- \left.\mathbf{\Phi}_\rho^{(1)}\left(\bar{y}_m^{(1)*}, \tilde{y}_m^{(1)*}\right)\right] \\
&\times \left[\mathbf{\Phi}_\nu^{(2)}\left(\bar{y}_m^{(2)*}, \eta^{(2)}\right) - \mathbf{\Phi}_\rho^{(2)}\left(\bar{y}_m^{(2)*}, \tilde{y}_m^{(2)*}\right)\right] \\
&+ \left[1 - \Phi\left(\bar{y}_m^{(1)*}\right)\right] \Phi\left(\bar{y}_m^{(2)*}\right).
\end{aligned}$$

Calculations are not detailed, as they are straightforward developments of Eq. (13). The two extremal terms ($j = 0$ and $j = m+1$) correspond to special cases of Ω_j (first and last cells in Figure 4, right).

5 Numerical experiments

5.1 One-dimensional, bi-objective problem

In this section, we apply the method to the following bi-objective problem: $F^{(1)}$ and $F^{(2)}$ are independent realizations of one-dimensional GPs, indexed by a 300-point regular grid on $[0, 1]$, with a stationary Matern covariance with regularity parameter $\nu = 3/2$ (Rasmussen and Williams, 2006, chapter 4). The variance and range parameters are taken as one and $1/5$, respectively.

Now, two GP models are built based on four randomly chosen observations. The covariance function is considered as known. Figure 5 shows the initial models and Pareto front. Here, a single point dominates the three others. After building the tesselation as described in Section 4.1, we compute the volume of the excursion sets corresponding to each cell (Eq. (11)). As there are only four observations, the probability to belong to a non-dominated cell is relatively high (Figure 5, bottom right). Then, the criterion is computed for each point in the grid (Figure 5, bottom right). Its maximum is obtained in a region with high uncertainty and low expected values for the two functions.

After 10 iterations (Figure 6), the Pareto front is well-approximated. The models are accurate in the optimal regions and have high prediction variances in the other regions, which indicates a good allocation of the computational resources.

Next, we compare these results to a state-of-the-art method, called SMS-EGO (Ponweiser et al., 2008), which has been shown to outperform significantly non-GP based methods (such as NSGA-II), in particular when only a limited budget of evaluation is available. As measuring performances is non-trivial in multi-criteria optimization, we use a series of three indicators: hypervolume, epsilon and R_2 indicators (Zitzler et al., 2003; Hansen and Jaszkiewicz, 1998), all available in the R package EMOA (Mersmann, 2012). They provide different measures of distance to the actual Pareto set and coverage of the objective space. Results are reported in Figure 7. The Pareto front obtained with SMS-EGO shows that the algorithm only detected one of the two Pareto optimal regions. As a consequence, the Pareto front is locally more accurate than the one obtained with the SUR strategy, but the indicators are much poorer.

5.2 Six-dimensional, bi-objective problem

Here, the objectives functions are realizations of six-dimensional GPs indexed by a 2000-point Sobol sequence on $[0, 1]^6$, with a stationary Matern covariance with regularity parameter $\nu = 5/2$. The variance and range parameters are taken as one and $\sqrt{6}/6$, respectively. The initial experimental set consists of 10 points randomly chosen, and 40 points are added iteratively using the SUR and SMS-EGO strategies. The results are given in Figure 8. Again, the SUR strategy shows better performances compared to SMS-EGO.

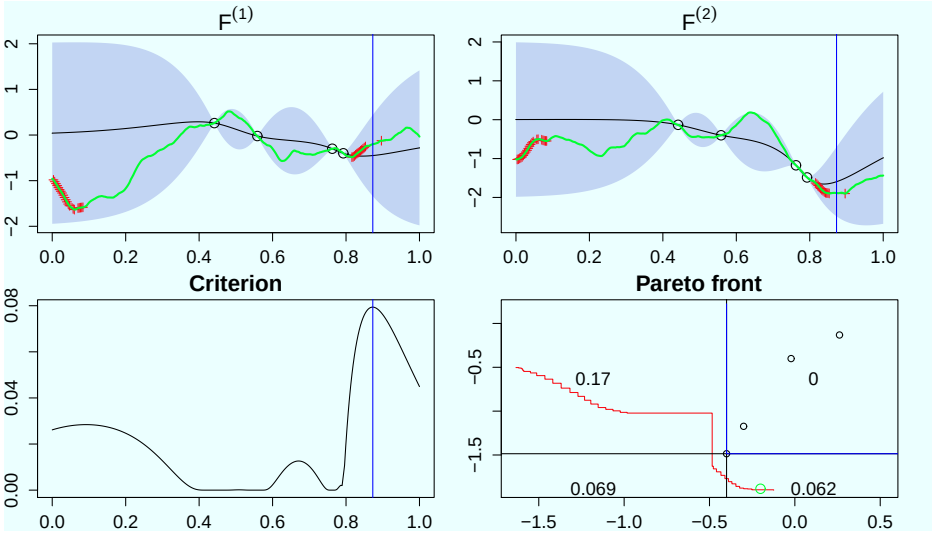

Figure 5: Top graphs: initial models (same representation as Figure 1); the actual Pareto-optimal points are represented by red crosses. Bottom right: observations (black circles) represented in the objective space, actual Pareto front (red) and current front (blue). Bottom left: criterion value as a function of $\mathbf{x}$. The vertical bars show the new observation location; the green circle is the new observation.

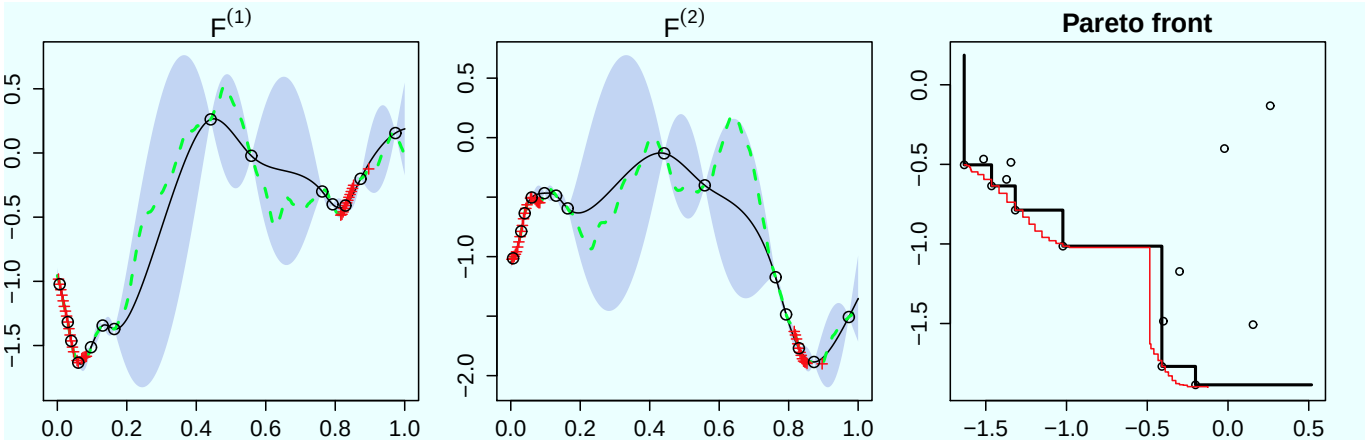

Figure 6: Models and Pareto front after 10 iterations.

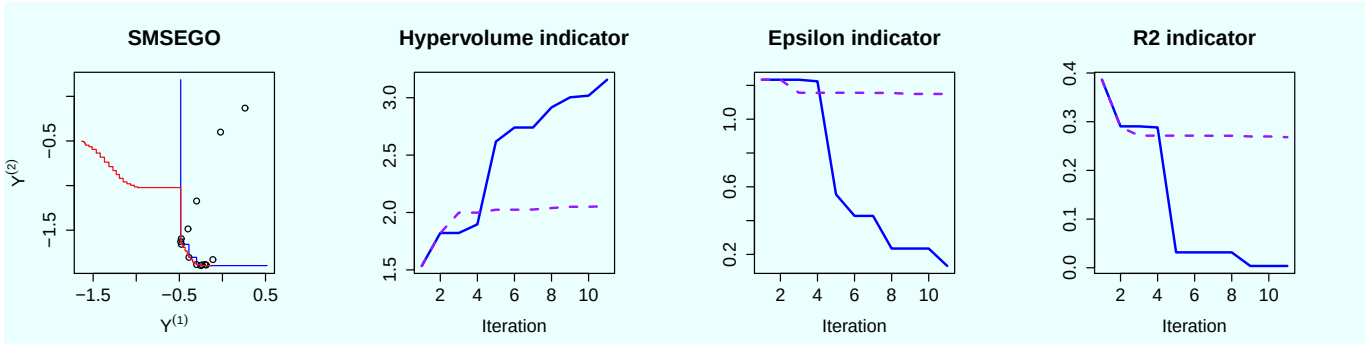

Figure 7: SMS-EGO Pareto front after 10 iterations (left) and performance comparison between SUR (plain line) and SMS-EGO (dotted line) on a one-dimensional problem. Hypervolume indicator: higher is better; other indicators should tend to zero.

6 Discussion

We have proposed a new sequential sampling strategy, based on stepwise uncertainty reduction principles, for multi-objective optimization. Closed-form expressions were provided for the infill criterion. Numerical experiments showed promising performances of our strategy compared to a state-of-the-art method. We point here some strengths and weaknesses of our approach.

First of all, as it is based on Gaussian process modeling,

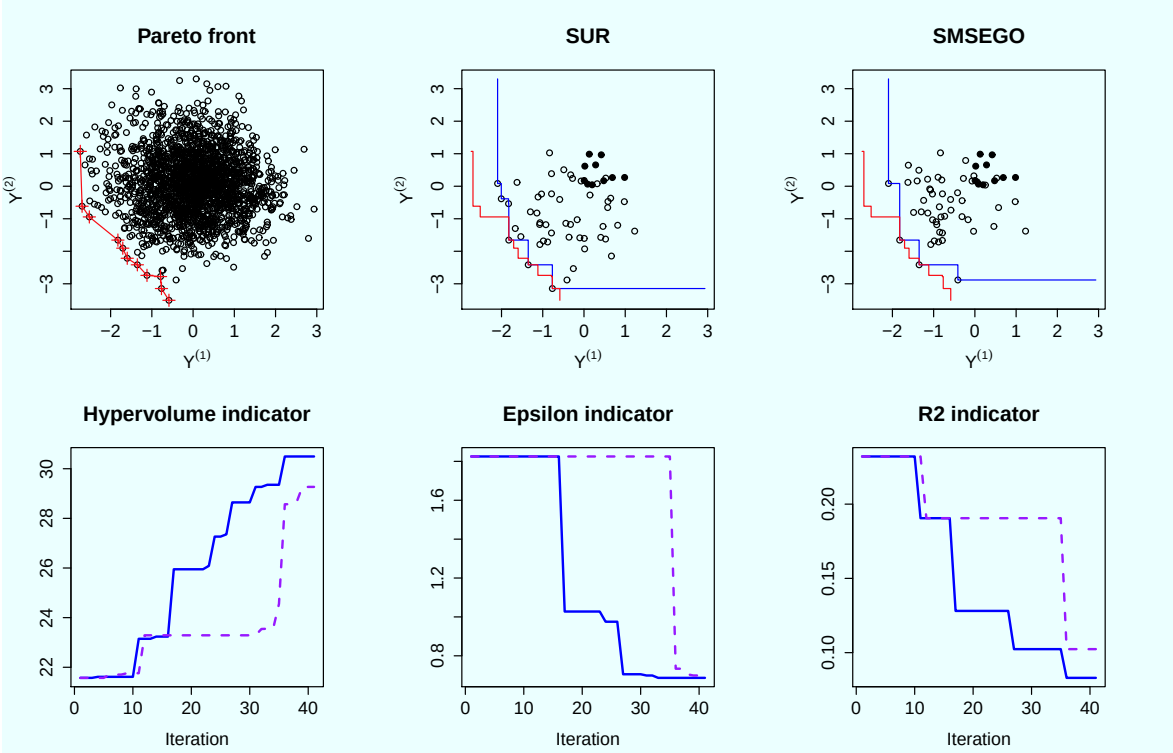

Figure 8: Performance comparison between SUR (plain line) and SMS-EGO (dotted line) on a six-dimensional problem. Top left: all 2000 points in the objective space; Pareto-optimal points are the red crosses. Top middle and right: Pareto fronts after 40 iterations.

it shares the limits inherents to the model. In particular, it is well-known that classical GP models cannot cope with large datasets (> 1000) or high-dimensional spaces (> 100). Most models also have restrictive conditions on the approximated function (typically, stationarity), and the strategy efficiency may be greatly penalized by important inadequations between the model hypothesis and the actual function characteristics. Using the proposed strategy on more complex GP models (Gramacy and Lee, 2008; Banerjee et al., 2013) may help mitigate these issues.

Secondly, we wish to emphasize here that the proposed method has a non-negligible computational cost, as (a) the criterion is evaluated by numerical integration and (b) it is embedded in an optimization loop. Hence, its use may be limited to simulators for which the time to compute an evaluation is much higher than the time to choose the next point to evaluate. However, one may note that the use of closed-form expressions, although relying on the bivariate normal CDF, avoid the need to use conditional simulations (as in Villemonteix et al. (2009)) that would have made the method overly expensive.

On the other hand, moving away from the expected improvement paradigm allowed us to provide a method that does not necessitate any artificial ranking or trade-off between objective functions. It is also scale-invariant, which can be of great advantage when dealing with objectives of different nature. Finally, one advantage of the proposed strategy is that it considers progress in the design space rather than in the objective space, which corresponds to what practitioners are eventually interested in.

Possible extensions of this work are various. Accounting for the uncertainty due to the estimation of the model hyperparameters were left appart here; Bayesian approaches, in the fashion of Kennedy and O'Hagan (2001) or Gramacy and Lee (2008) for instance, may help address this issue. Objective functions were considered as not correlated to ease calculations and allow the use of simple models. As objectives are likely to be negatively correlated in practice, accounting for it while keeping tractable criteria is an important question. Finally, the stepwise uncertainty reduction strategy may be easily adapted to other frameworks, such as constrained or noisy optimization.

A Probabilities update

A.1 Proof of Proposition 3.2

Using the model update equations (4), we note first that:

$$p_{n+1}(\mathbf{x}, a) = \Phi\left[\frac{a - m_n(\mathbf{x}) + \frac{c_n(\mathbf{x},\mathbf{x}_{n+1})}{s_n^2(\mathbf{x}_{n+1})}\left[m_n\left(\mathbf{x}_{n+1}\right) - y_{n+1}\right]}{s_{n+1}(\mathbf{x})}\right]$$

Now, let $\varphi\left(y_{n+1}\right)$ be the PDF of Y_{n+1} (conditional on $\mathcal{A}_n$). We have:

$$\begin{aligned} & q(\mathbf{x}, b, a) \\ &= \int_{-\infty}^{b} p_{n+1}(\mathbf{x}, a) d\varphi\left(y_{n+1}\right) \\ &= \int_{-\infty}^{b} \Phi\left[\frac{a - m_n(\mathbf{x}) + \frac{c_n(\mathbf{x},\mathbf{x}_{n+1})}{s_n^2(\mathbf{x}_{n+1})}\left[m_n\left(\mathbf{x}_{n+1}\right) - y_{n+1}\right]}{s_{n+1}(\mathbf{x})}\right] \\ & d\varphi\left(y_{n+1}\right) \end{aligned}$$

As $Y_{n+1} \sim \mathcal{N}\left(m_n(\mathbf{x}_{n+1}), s_n^2(\mathbf{x}_{n+1})\right)$, we can write (following Chevalier et al. (2012)):

$$Y_{n+1} = m_n(\mathbf{x}_{n+1}) + s_n(\mathbf{x}_{n+1})U$$

with

$$U \sim \mathcal{N}(0,1),$$

which allows to simplify the previous equations to:

$$\begin{aligned} & q(\mathbf{x}, b, a) \\ &= \int_{-\infty}^{\bar{b}} \Phi\left[\frac{a - m_n(\mathbf{x})}{s_{n+1}(\mathbf{x})} - \left(\frac{c_n(\mathbf{x}, \mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1}) s_{n+1}(\mathbf{x})}\right) u\right] d\varphi(u) \\ &= \int_{-\infty}^{\bar{b}} \Phi\left[\hat{a} - \beta u\right] d\varphi(u), \end{aligned} \tag{14}$$

with

$$\begin{aligned} \beta &= (c_n(\mathbf{x}, \mathbf{x}_{n+1}))/(s_n(\mathbf{x}_{n+1}) s_{n+1}(\mathbf{x})), \\ \hat{a} &= (a - m_n(\mathbf{x}))/s_{n+1}(\mathbf{x}) \text{ and} \\ \bar{b} &= (b - m_n(\mathbf{x}_{n+1}))/s_n(\mathbf{x}_{n+1}). \end{aligned}$$

This quantity can be written as a bivariate Gaussian CDF. Indeed:

$$\begin{aligned} & \int_{-\infty}^{\bar{b}} \Phi\left[\hat{a} - \beta u\right] d\varphi(u) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\bar{b}} \Phi\left[\hat{a} - \beta u\right] \exp\left(\frac{-u^2}{2}\right) du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\bar{b}} \int_{-\infty}^{\hat{a} - \beta u} \exp\left[-\frac{1}{2}\left(u^2 + t^2\right)\right] dt du \\ &= \frac{1}{2\pi} \int_{-\infty}^{\bar{b}} \int_{-\infty}^{\hat{a}} \exp\left[-\frac{1}{2}\left(u^2 + [t - \beta u]^2\right)\right] dt du \\ &= \frac{1}{2\pi|\mathbf{\Sigma}_\beta|} \int_{-\infty}^{\bar{b}} \int_{-\infty}^{\hat{a}} \exp\left[-\frac{1}{2}\begin{bmatrix} u & t \end{bmatrix} \mathbf{\Sigma}_\beta^{-1} \begin{bmatrix} u \\ t \end{bmatrix}\right] dt du, \end{aligned}$$

with $\mathbf{\Sigma}_\beta = \begin{bmatrix} 1 & \beta \\ \beta & 1 + \beta^2 \end{bmatrix}$ (noting that $|\mathbf{\Sigma}_\beta| = 1$), which is the standard form of the bivariate Gaussian CDF with zero mean and covariance matrix $\mathbf{\Sigma}_\beta$, hence:

$$q(\mathbf{x}, b, a) = \mathbf{\Phi}_{\mathbf{\Sigma}_\beta}\left(\bar{b}, \hat{a}\right).$$

Finally, applying the normalization $\tilde{a} = \hat{a}/\sqrt{1 + \beta^2}$, we have: $q(\mathbf{x}, b, a) = \mathbf{\Phi}_\rho\left(\bar{b}, \tilde{a}\right)$, with:

$$\rho = \frac{\beta}{\sqrt{1 + \beta^2}} = \frac{c_n(\mathbf{x}, \mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1}) s_n(\mathbf{x})}.$$

A.2 Proof of Proposition 3.1

The result can be obtained directly from Proposition 3.2 with $b \to +\infty$. We have then: $q(\mathbf{x}, b, a) \to \Phi(\tilde{a}) = p_n(\mathbf{x}, a)$.

A.3 Proof of Corollary 3.3

From Eq. (14), we have directly:

$$\begin{aligned} r(\mathbf{x}, b, a) &= \int_{\bar{b}}^{+\infty} \Phi\left[\hat{a} - \beta u\right] d\varphi(u) \\ &= \int_{-\infty}^{-\bar{b}} \Phi\left[\hat{a} + \beta u\right] d\varphi(u) \\ &= \mathbf{\Phi}_{\mathbf{\Sigma}_{-\beta}}\left(-\bar{b}, \hat{a}\right) = \mathbf{\Phi}_{-\rho}\left(-\bar{b}, \tilde{a}\right). \end{aligned}$$

A.4 Proof of Proposition 3.4

The steps of the proof are similar to those of Proposition 3. Using the update equations (4), we have first:

$$\begin{aligned} & \mathbb{P}\left(Y(\mathbf{x}) \leq y_{n+1} | \mathcal{A}_n, y_{n+1} = y(\mathbf{x}_{n+1})\right) \\ &= \Phi\left[\frac{y_{n+1} - m_{n+1}(\mathbf{x})}{s_{n+1}(\mathbf{x})}\right] \\ &= \Phi\left[\frac{-m_n(\mathbf{x}) + \frac{c_n(\mathbf{x}, \mathbf{x}_{n+1}) m_n(\mathbf{x}_{n+1})}{s_n^2(\mathbf{x}_{n+1})} + \left[1 - \frac{c_n(\mathbf{x}, \mathbf{x}_{n+1})}{s_n^2(\mathbf{x}_{n+1})}\right] y_{n+1}}{s_{n+1}(\mathbf{x})}\right] \\ &= \Phi\left[\frac{m_n(\mathbf{x}_{n+1}) - m_n(\mathbf{x})}{s_{n+1}(\mathbf{x})} - \left(\frac{c_n(\mathbf{x}, \mathbf{x}_{n+1}) - s_n^2(\mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1}) s_{n+1}(\mathbf{x})}\right) u\right]. \end{aligned}$$

Now:

$$\begin{aligned} & h(\mathbf{x}, b) \\ &= \int_{-\infty}^{b} \mathbb{P}\left(Y(\mathbf{x}) \leq Y_{n+1} | \mathcal{A}_n, Y_{n+1} = y_{n+1}\right) d\varphi\left(y_{n+1}\right) \\ &= \int_{-\infty}^{\bar{b}} \Phi\Big[\frac{m_n(\mathbf{x}_{n+1}) - m_n(\mathbf{x})}{s_{n+1}(\mathbf{x})} \\ &- \left(\frac{c_n(\mathbf{x}, \mathbf{x}_{n+1}) - s_n^2(\mathbf{x}_{n+1})}{s_n(\mathbf{x}_{n+1}) s_{n+1}(\mathbf{x})}\right) u\Big] d\varphi(u) \\ &= \int_{-\infty}^{\bar{b}} \Phi\left[\mu - \tau u\right] d\varphi(u) \\ &= \mathbf{\Phi}_{\mathbf{\Sigma}_\tau}\left(\bar{b}, \mu\right), \end{aligned}$$

as we get a form similar to Equation 14, with $\mathbf{\Phi}_{\mathbf{\Sigma}_\tau}$ the CDF of the centered bigaussian with covariance $\mathbf{\Sigma}_\tau = \begin{bmatrix} 1 & \tau \\ \tau & 1 + \tau^2 \end{bmatrix}$,

$$\begin{aligned} \mu &= (m_n(\mathbf{x}_{n+1}) - m_n(\mathbf{x}))/s_{n+1}(\mathbf{x}) \\ \tau &= (c_n(\mathbf{x}, \mathbf{x}_{n+1}) - s_n^2(\mathbf{x}_{n+1}))/(s_n(\mathbf{x}_{n+1}) s_{n+1}(\mathbf{x})). \end{aligned}$$

Normalizing $\eta = \mu/\sqrt{1 + \tau^2}$ delivers the final result.

B $b_{ij}^{(k)}(\mathbf{x})$ and $d_{ij}^{(k)}(\mathbf{x})$ computation

Let X and Y be two dependent random variables, and a, b, c and d four real numbers. By direct application of

Bayes formula, we have:

$$\begin{aligned}&\mathbb{P}\left(a \leq X < b | c \leq Y < d\right) \mathbb{P}\left(c \leq Y < d\right)\\ &= \mathbb{P}\left(Y < d\right) \times \left[\mathbb{P}\left(X < b | Y < d\right) - \mathbb{P}\left(X \leq a | Y < d\right)\right]\\ &- \mathbb{P}\left(Y \leq c\right) \times \left[\mathbb{P}\left(X < b | Y \leq c\right) - \mathbb{P}\left(X \leq a | Y \leq c\right)\right]\end{aligned}$$

$$\begin{aligned}&\mathbb{P}\left(Y \leq X < b | a \leq Y < b\right) \mathbb{P}\left(a \leq Y < b\right)\\ &= \mathbb{P}\left(Y < b\right) \times \left[\mathbb{P}\left(X < b | Y < b\right) - \mathbb{P}\left(X \leq Y | Y < b\right)\right]\\ &- \mathbb{P}\left(Y \leq a\right) \times \left[\mathbb{P}\left(X < b | Y \leq a\right) - \mathbb{P}\left(X \leq Y | Y \leq a\right)\right]\end{aligned}$$

Now, by definition, $b_{ij}^{(k)}$ is of the form of the fist equation:

$$\begin{aligned}b_{ij}^{(k)}(\mathbf{x}) &:= \mathbb{P}_{n+1}\left(y_{j-}^{(k)} \leq Y^{(k)}(\mathbf{x}) < y_{j+}^{(k)} | y_{i-}^{(k)} \leq Y_{n+1}^{(k)} < y_{i+}^{(k)}\right)\\ &\times \mathbb{P}_n\left[y_{i-}^{(k)} \leq Y_{n+1}^{(k)} < y_{i+}^{(k)}\right],\end{aligned}$$

hence write as the sum of four terms:

$$\begin{aligned}b_{ij}^{(k)}(\mathbf{x}) &= p_n^{(k)}(\mathbf{x}_{n+1}, y_{i+}^{(k)})\Big(\mathbb{P}_{n+1}\big[Y^{(k)}(\mathbf{x}) \leq y_{j+}^{(k)} | Y_{n+1}^{(k)} \leq y_{i+}^{(k)}\big]\\ &- \mathbb{P}_{n+1}\big[Y^{(k)}(\mathbf{x}) \leq y_{j-}^{(k)} | Y_{n+1}^{(k)} \leq y_{i+}^{(k)}\big]\Big)\\ &- p_n^{(k)}(\mathbf{x}_{n+1}, y_{i-}^{(k)})\Big(\mathbb{P}_{n+1}\big[Y^{(k)}(\mathbf{x}) < y_{j+}^{(k)} | Y_{n+1}^{(k)} \leq y_{i-}^{(k)}\big]\\ &- \mathbb{P}_{n+1}\big[Y^{(k)}(\mathbf{x}) \leq y_{j-}^{(k)} | Y_{n+1}^{(k)} \leq y_{i-}^{(k)}\big]\Big)\\ &= q^{(k)}\left(\mathbf{x}, y_{i+}^{(k)}, y_{j+}^{(k)}\right) - q^{(k)}\left(\mathbf{x}, y_{i+}^{(k)}, y_{j-}^{(k)}\right)\\ &- q^{(k)}\left(\mathbf{x}, y_{i-}^{(k)}, y_{j+}^{(k)}\right) + q^{(k)}\left(\mathbf{x}, y_{i-}^{(k)}, y_{j-}^{(k)}\right),\end{aligned}$$

with $q^{(k)}(\mathbf{x}, b, a)$ given by Eq. (5), thus:

$$\begin{aligned}b_{ij}^{(k)}(\mathbf{x}) &= \mathbf{\Phi}_\rho^{(k)}\left(\overline{y_{i+}^{(k)}}, \widetilde{y_{j+}^{(k)}}\right) - \mathbf{\Phi}_\rho^{(k)}\left(\overline{y_{i+}^{(k)}}, \widetilde{y_{j-}^{(k)}}\right)\\ &- \mathbf{\Phi}_\rho^{(k)}\left(\overline{y_{i-}^{(k)}}, \widetilde{y_{j+}^{(k)}}\right) + \mathbf{\Phi}_\rho^{(k)}\left(\overline{y_{i-}^{(k)}}, \widetilde{y_{j-}^{(k)}}\right).\end{aligned}$$

Similary, $d_{ij}^{(k)}$ is of the form of the second equation: Starting with the definition:

$$\begin{aligned}d_{ij}^{(k)}(\mathbf{x}) &:= \mathbb{P}_{n+1}\left(Y_{n+1}^{(k)} \leq Y^{(k)}(\mathbf{x}) < y_{j+}^{(k)} | y_{i-}^{(k)} \leq Y_{n+1}^{(k)} < y_{i+}^{(k)}\right)\\ &\times \mathbb{P}_n\left[y_{i-}^{(k)} \leq Y_{n+1}^{(k)} < y_{i+}^{(k)}\right],\end{aligned}$$

hence writes:

$$\begin{aligned}d_{ij}^{(k)}(\mathbf{x}) &= p_n^{(k)}(\mathbf{x}_{n+1}, y_{i+}^{(k)})\Big(\mathbb{P}_{n+1}\big[Y^{(k)}(\mathbf{x}) \leq y_{j+}^{(k)} | Y_{n+1}^{(k)} \leq y_{i+}^{(k)}\big]\\ &- \mathbb{P}_{n+1}\big[Y^{(k)}(\mathbf{x}) \leq Y_{n+1}^{(k)} | Y_{n+1}^{(k)} \leq y_{i+}^{(k)}\big]\Big)\\ &- p_n^{(k)}(\mathbf{x}_{n+1}, y_{i-}^{(k)})\Big(\mathbb{P}_{n+1}\big[Y^{(k)}(\mathbf{x}) < y_{j+}^{(k)} | Y_{n+1}^{(k)} \leq y_{i-}^{(k)}\big]\\ &- \mathbb{P}_{n+1}\big[Y^{(k)}(\mathbf{x}) \leq Y_{n+1}^{(k)} | Y_{n+1}^{(k)} \leq y_{i-}^{(k)}\big]\Big)\\ &= q^{(k)}\left(\mathbf{x}, y_{i+}^{(k)}, y_{j+}^{(k)}\right) - h^{(k)}\left(\mathbf{x}, y_{i+}^{(k)}\right)\\ &- h^{(k)}\left(\mathbf{x}, y_{i-}^{(k)}\right) + q^{(k)}\left(\mathbf{x}, y_{i-}^{(k)}, y_{j-}^{(k)}\right),\end{aligned}$$

with $q^{(k)}(\mathbf{x}, b, a)$ given by Eq. (5) and $h^{(k)}(\mathbf{x}, b)$ given by Eq. (7), thus:

$$\begin{aligned}d_{ij}^{(k)}(\mathbf{x}) &= \mathbf{\Phi}_\rho^{(k)}\left(\overline{y_{i+}^{(k)}}, \widetilde{y_{j+}^{(k)}}\right) - \mathbf{\Phi}_\nu^{(k)}\left(\overline{y_{i+}^{(k)}}, \eta^{(k)}\right)\\ &+ \mathbf{\Phi}_\nu^{(k)}\left(\overline{y_{i-}^{(k)}}, \eta^{(k)}\right) - \mathbf{\Phi}_\rho^{(k)}\left(\overline{y_{i-}^{(k)}}, \widetilde{y_{j+}^{(k)}}\right).\end{aligned}$$

References

Banerjee, A., Dunson, D.B., Tokdar, S.T.: Efficient gaussian process regression for large datasets. Biometrika **100**(1), 75–89 (2013)

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)

Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., Richet, Y.: Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. `http://hal.inria.fr/hal-00641108/en` (2012)

Chevalier, C., Picheny, V., Ginsbourger, D.: Kriginv: An efficient and user-friendly implementation of batch-sequential inversion strategies based on kriging. Computational Statistics & Data Analysis (2013)

Collette, Y., Siarry, P.: Multiobjective optimization: principles and case studies. Springer (2003)

Craig, P.S., Goldstein, M., Rougier, J.C., Seheult, A.H.: Bayesian forecasting for complex systems using computer simulators. Journal of the American Statistical Association **96**(454), 717–729 (2001)

Cressie, N.: Statistics for Spatial Data, revised edition, vol. 928. Wiley, New York (1993)

Emery, X.: The kriging update equations and their application to the selection of neighboring data. Computational Geosciences **13**(3), 269–280 (2009)

Gramacy, L., Lee, H.: Optimization under unknown constraints. Bayesian Statistics **9**, 229 (2011)

Gramacy, R.B., Lee, H.K.: Bayesian treed gaussian process models with an application to computer modeling. Journal of the American Statistical Association **103**(483) (2008)

Hansen, M.P., Jaszkiewicz, A.: Evaluating the quality of approximations to the non-dominated set. IMM, Department of Mathematical Modelling, Technical University of Denmark (1998)

Jones, D.R.: A taxonomy of global optimization methods based on response surfaces. Journal of global optimization **21**(4), 345–383 (2001)

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)

Keane, A.J.: Statistical improvement criteria for use in multiobjective design optimization. AIAA journal **44**(4), 879–891 (2006)

Kenkel, B.: pbivnorm: Vectorized Bivariate Normal CDF (2012). URL `http://CRAN.R-project.org/package=pbivnorm`. R package version 0.5-1

Kennedy, M.C., O'Hagan, A.: Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **63**(3), 425–464 (2001)

Knowles, J.: Parego: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. Evolutionary Computation, IEEE Transactions on **10**(1), 50–66 (2006)

Mersmann, O.: emoa: Evolutionary Multiobjective Optimization Algorithms (2012). URL `http://CRAN.R-project.org/package=emoa`. R package version 0.5-0

Ponweiser, W., Wagner, T., Biermann, D., Vincze, M.: Multiobjective optimization on a limited budget of evaluations using model-assisted s-metric selection. In: Parallel Problem Solving from Nature, pp. 784–794. Springer (2008)

Rasmussen, C., Williams, C.: Gaussian processes for machine learning. MIT Press (2006)

Roustant, O., Ginsbourger, D., Deville, Y.: Dicekriging, diceoptim: Two r packages for the analysis of computer experiments by kriging-based metamodeling and optimization. Journal of Statistical Software **51**(1), 1–55 (2012)

Scott, W., Frazier, P., Powell, W.: The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression. SIAM Journal on Optimization **21**(3), 996–1026 (2011)

Stein, M.: Interpolation of spatial data: some theory for kriging. Springer Verlag (1999)

Villemonteix, J., Vazquez, E., Walter, E.: An informational approach to the global optimization of expensive-to-evaluate functions. Journal of Global Optimization **44**(4), 509–534 (2009)

Wagner, T., Emmerich, M., Deutz, A., Ponweiser, W.: On expected-improvement criteria for model-based multi-objective optimization. In: Parallel Problem Solving from Nature, PPSN XI, pp. 718–727. Springer (2010)

Wang, G.G., Shan, S.: Review of metamodeling techniques in support of engineering design optimization. Journal of Mechanical Design **129**, 370 (2007)

Zitzler, E., Thiele, L., Laumanns, M., Fonseca, C.M., Da Fonseca, V.G.: Performance assessment of multiobjective optimizers: An analysis and review. Evolutionary Computation, IEEE Transactions on **7**(2), 117–132 (2003)