Dealing with asynchronicity in parallel Gaussian Process based global optimization

David Ginsbourger, Janis Janusevskis, Rodolphe Le Riche

Dealing with asynchronicity in parallel Gaussian Process based global optimization *

David Ginsbourger†‡
Joint Work with Janis Janusevkis §and Rodolphe Le Riche ¶

July 30, 2010

Abstract

During the last decade, Kriging-based sequential algorithms like EGO and its variants have become reference optimization methods in computer experiments. Such algorithms rely on the iterative maximization of a sampling criterion, the expected improvement (EI), which takes advantage of Kriging conditional distributions to make an explicit trade-off between promizing and uncertain search space points. We have recently worked on a multipoints EI criterion meant to simultaneously choose several points, which is useful for instance in synchronous parallel computation. Here we propose extensions of these works to asynchronous parallel optimization and focus on a variant of EI, EEI, for the case where some new evaluation(s) have to be done while the reponses of previously simulations are not all known yet. In particular, different issues regarding EEI's maximization are addressed, and a proxy strategy is proposed.

*Deliverable 2.1.1-A of the OMD^2 ANR project (Optimisation Multidisciplinaire Distribuée)
†Senior Assistant, Department of Mathematics and Statistics, University of Bern, Switzerland
‡Scientific collaborator, Department of Hydrogeology, University of Neuchâtel, Switzerland
§Postdoctoral Researcher, LSTI/CROCUS team, Ecole Nat. Sup. des Mines de Saint-Etienne
¶CNRS UMR 5146 & LSTI/CROCUS team, Ecole Nat. Sup. des Mines de Saint-Etienne

Contents

1 Introduction

Sampling criteria such as the *Expected Improvement* (EI) allow sequentially choosing evaluation points in Gaussian Process (GP) based optimization algorithms. A major strength of using EI in a sequential procedure is that it progressively integrates available observations (points and responses), and thus avoids resampling at already explored points, at least in the case of noise-free simulations.

Recent works deal with the adaptation of EI for parallel Kriging-based optimization. In particular, a multipoints version of EI, called q-EI (or simply EI when there is no ambiguity), has been proposed for synchronous parallel Kriging-based optimization. q-EI measures the joint potential of a given additional q-points design of experiments. So, a natural parallelization of EI algorithms for q synchronous processors is to maximize q-EI at each iteration, with update of the Kriging metamodel (including possible re-estimation of covariance parameters) whenever new observations are assimilated. However, maximizing q-EI is in most cases unaffordable, all the more so that the dimension of inputs d and the number of points q increase. Furthermore, evaluating q-EI when $q \geq 3$ bases on a Monte Carlo method relying on GP conditional simulations, which would make this intermediate optimization problem noisy and potentially very time consuming. Heuristic strategies such as the *Constant Liar* (CL) and the *Kriging Believer* (KB) have been proposed in [34] to circumvent the computational difficulties associated with q-EI maximization, by turning it into a stepwise maximization with virtual intermediate observation values.

In this report, we focus on the case of asynchronous parallel Kriging-based optimization, i.e. where sampling decisions have to be made before all previously started simulations have produced their result. Hence we have to distinguish between already visited points, currently visited (or "busy") points, and candidate points for forthcoming simulations. Having to take *busy points* into account within GP modeling sends back to similar problems encountered in the heuristic strategies mentioned above, so that we take advantage of previous work on that topic, and revisit those heuristic strategies in a more general manner.

After some necessary definitions concerning the EI criterion with (partially or completely) enriched information, a variant of EI is proposed for the case where a processor did not give its result yet. Relying on conditional simulations, EEI proves to be a sensible criterion, but practically not straightforward to maximize. A special variant of CL with random lie is then investigated to generate a finite set of candidate EEI maximizers, and gives surprisingly excellent results on the basis of a first 1-dimensional toy example.

Some research perspectives and a rich bibliography close this report, and set a framework for additional testing and developements to be made in forthcoming project phases.

2 Injecting incomplete information in EI criteria

2.1 EI with (partially) enriched information

The regular EI is defined as follows

$$\mathrm{EI}(\mathbf{x}) = \mathbb{E}\left[(\min(Y(\mathbf{X})) - Y(\mathbf{x}))^+ | A\right], \tag{1}$$

where $(.)^+ = \max(0, .)$, and $A = \{Y(\mathbf{X}) = \mathbf{y}\}$ is the event summing up all currently available observation points and corresponding responses.

Let us now assume that an evaluation of y at point $\mathbf{x}^{\mathrm{busy}} \in D$ (say at an EI global maximizer, assumed unique for convenience) has been started, that the result is not known yet, and that one wishes to determine the next most promizing point (in terms of EI) where to perform an evaluation using a newly available processor.

Ignoring that an evaluation is going on at $\mathbf{x}^{\mathrm{busy}}$ may cause loosing crucial information for the next evaluation(s). However, it is not possible to add the event $\left\{Y(\mathbf{x}^{\mathrm{busy}}) = y(\mathbf{x}^{\mathrm{busy}})\right\}$ to A since $y(\mathbf{x}^{\mathrm{busy}})$ is not kwown yet.

Question 1 *How is it possible to inject the partial knowledge that an evaluation is going on at $\mathbf{x}^{busy}$ in the EI criterion without knowing $y(\mathbf{x}^{busy})$?*

Before giving a probabilistic answer to that question, let us introduce some necessary additional concepts and notations.

Definition 1 *Given $\mathbf{X}$, $\mathbf{y}$, A as above, and an additional point $\mathbf{x}^{add} \in D$ together with its corresponding response y_{add}, we denote by* ***EI with enriched information*** *the function*

$$\mathbf{x} \in D \longrightarrow EI(\mathbf{x}; \mathbf{x}^{add}, y_{add}) := \mathbb{E}\left[(\min(Y(\mathbf{X}), y_{add}) - Y(\mathbf{x}))^+ | A, Y(\mathbf{x}^{add}) = y_{add}\right] \tag{2}$$

This modification of the EI criterion basically amounts to updating the underlying conditioning event A to $\left\{Y(\mathbf{X}) = \mathbf{y}, Y(\mathbf{x}^{\mathrm{add}}) = y(\mathbf{x}^{\mathrm{add}})\right\}$. As a consequence, the EI with enriched information is 0 at $\mathbf{x}^{\mathrm{add}}$ whatever the observation value y_{add}. This property will be useful in the sequel since we will actually often be in situations where the value of y_{add} is unkown.

A probabilistic answer to question 1 is to continue treating the unknown $y(\mathbf{x}^{\mathrm{busy}})$ as a random variable $Y(\mathbf{x}^{\mathrm{busy}})$, but to insert it in on the side of conditioning events by plugging it in the previously defined EI with enriched information.

Definition 2 *Let us define the* ***EI with partially enriched information*** *as*

$$\begin{aligned} EI(\mathbf{x}; \mathbf{x}^{busy}) : &= EI(\mathbf{x}; \mathbf{x}^{busy}, Y(\mathbf{x}^{busy})) \\ &= \mathbb{E}\left[\left(\min(Y(\mathbf{X}), Y(\mathbf{x}^{busy})) - Y(\mathbf{x})\right)^+ | A, Y(\mathbf{x}^{busy})\right] \end{aligned} \tag{3}$$

For any given fixed $\mathbf{x} \in D$, $\mathrm{EI}(\mathbf{x};\mathbf{x}^{\mathrm{busy}})$ is hence a function of the random variable $Y(\mathbf{x}^{\mathrm{busy}})$, thus inheriting from its randomness. Indeed, the classical EI formula can be developed, and the randomness of the obtained expression appears clearly through the instances of $Y(\mathbf{x}^{\mathrm{busy}})$ in the current minimum and in the Kriging mean.

Since the function $\mathrm{EI}(.;\mathbf{x}^{\mathrm{busy}})$ is consequently a random function, it does not make sense to maximize it in order to derive the next most promizing point $\mathbf{x}^{\mathrm{new}}$.

Question 2 *Indeed, what would it mean to maximize a random function?*

Several alternatives are possible, such as maximizing quantiles of it conditional on the current information, or basically focus on the following conditional expectation function

Definition 3 *The expectation of $EI(.;\mathbf{x}^{busy})$ conditional on A is denoted by*

$$EEI(.;\mathbf{x}^{busy}) := \mathbb{E}[EI(.;\mathbf{x}^{busy})|A] \tag{4}$$

Denoting by $f_{Y(\mathbf{x}^{\mathrm{busy}})|A}(.)$ the probability density function (pdf) of $Y(\mathbf{x}^{\mathrm{busy}})$ conditional on A, which is deduced from the standard gaussian pdf ϕ by an affine change of variable,

$$f_{Y(\mathbf{x}^{\mathrm{busy}})|A}(y_{\mathrm{busy}}) = \frac{1}{s(\mathbf{x}^{\mathrm{busy}})}\phi\left(\frac{y_{\mathrm{busy}} - m(\mathbf{x}^{\mathrm{busy}})}{s(\mathbf{x}^{\mathrm{busy}})}\right) \text{ (whenever } s(\mathbf{x}^{\mathrm{busy}}) \neq 0 \text{)} \tag{5}$$

we may in fact explicitely derive the latter conditional expectation in integral form:

$$\boxed{\mathrm{EEI}(.;\mathbf{x}^{\mathrm{busy}}) = \int \mathrm{EI}(.;\mathbf{x}^{\mathrm{busy}}, y_{\mathrm{busy}}) f_{Y(\mathbf{x}^{\mathrm{busy}})|A}(y_{\mathrm{busy}}) dy_{\mathrm{busy}}} \tag{6}$$

Now, $\mathrm{EI}(.;\mathbf{x}^{\mathrm{busy}}, y_{\mathrm{busy}})$ depends on y_{busy} in a quite complicated non-linear way, and there is no hope of getting an analytical expression for EEI, even though $f_{Y(\mathbf{x}^{\mathrm{busy}})|A}$ is known and rather simple. However, coming back to its probabilistic definition, EEI can be approximated by averaging out the $\mathrm{EI}(.;\mathbf{x}^{\mathrm{busy}}, y_{\mathrm{busy}})$'s obtained when drawing a sufficiently high number of y_{busy} candidate values generated by using the conditional law of $Y(\mathbf{x}^{\mathrm{busy}})|A$.

Algorithm 1 Estimate EEI using MC simulations

1: **for** $i \leftarrow 1, n_{sim}$ **do**
2: $\quad Y_i^{sim} \sim \mathcal{L}(Y(\mathbf{x}^{\mathrm{busy}})|A)$
3: $\quad EI_i(\mathbf{x}) = EI(\mathbf{x};\mathbf{x}^{\mathrm{busy}}, Y_i^{sim})$
4: **end for**
5: $\widehat{\mathrm{EEI}(\mathbf{x})} = \frac{1}{n_{sim}} \sum_i^{n_{sim}} \mathrm{EI}_i(\mathbf{x})$

Thus, a straightforward way of getting statistical estimates of $\text{EEI}(\mathbf{x}; \mathbf{x}^{\text{busy}})$ (pointwise, for given $\mathbf{x} \in D$, or of the whole function $\text{EEI}(.; \mathbf{x}^{\text{busy}})$) is to base on Monte Carlo drawings, as in alg. 2.1. Obviously, the reliability of such a Monte Carlo approximation of EEI will depend on the invested number of drawings n_{sim}. To be more precise, on can state that for any fixed $\mathbf{x} \in D$, the estimator $\widehat{\text{EEI}(\mathbf{x})}$ obtained above is unbiased, and has a variance given by a formula of a well-known kind:

$$\text{Var}\left[\widehat{\text{EEI}(\mathbf{x})}\right] = \frac{1}{n_{sim}} \int \Big(\text{EI}(.; \mathbf{x}^{\text{busy}}, y_{\text{busy}}) - \text{EEI}(\mathbf{x})\Big)^2 f_{Y(\mathbf{x}^{\text{busy}})|A}(y_{\text{busy}}) dy_{\text{busy}} \tag{7}$$

2.2 Main example

Let us consider a deterministic one-dimensional function,

$$y : \mathbf{x} \in [0,1] \longrightarrow y(\mathbf{x}) = \frac{sin(10x+1)}{(1+x)} + 2cos(5x) \times x^4 \in \mathbb{R}, \tag{8}$$

and first approximate it using a Kriging model based on y's observation at a three-points design $\mathbf{X} = \{0, 0.475, 0.95\}$. The chosen Simple Kriging model is centered, has a Matérn covariance structure with regularity coefficient $\nu = \frac{3}{2}$, unit variance, and range parameter $\theta = \frac{0.5}{\sqrt{3}}$. The objective function, the Kriging mean predictor, and the corresponding Expected Improvement function are represented on figure 2.1. This very simple example will enable us to get a first illustration of the previously discussed notions, and to put the finger on some related problems, in a surprisingly rather general way (despite the low-dimensional framework). Let us now detail the example. First, we consider the Kriging predictor and the EI function of figure 2.1, both in blue. Assume that a first processor is vacant, and that one wishes to use it to perform a fourth simulation. It is then natural to perform this additional simulation at the point maximizing the current version of EI, that is to say at the point 0.698 magnified by an orange dashed line on figure 2.1.

Let us now pretend that the evaluation of y at the latter point is getting started at a currently available processor, say $proc_1$, and set $\mathbf{x}^{\text{busy}} = 0.698$. Assume further that a second processor, $proc_2$, is becoming available while $proc_1$ is still busy with the computation of $y(\mathbf{x}^{\text{busy}})$. The main issue here is obviously to choose $\mathbf{x}^{\text{new}}$, the point at which to evaluate y with $proc_2$, in such a way that the all available information is optimally used.

Computing the EEI criterion of eq. 6 by means of a procedure such as alg. 2.1 will lead us to makes conditional simulations of $Y(\mathbf{x}^{\text{new}})$ knowing the past observations. Before averaging out a whole collection of EI curves obtained by this means and maximizing the obtained EEI estimate, let us first focus on a single conditional simulation and detail the underlying mechanism. Figure 2.2 illustrates the effect of enriching EI with a simulated observation. The latter is represented by a green point on the upper graph, as well as the new (virtual) Kriging mean predictor. In the same color, the curve on the lower graph stands for the corresponding EI with enriched information. As mentioned earlier, this updated EI is null

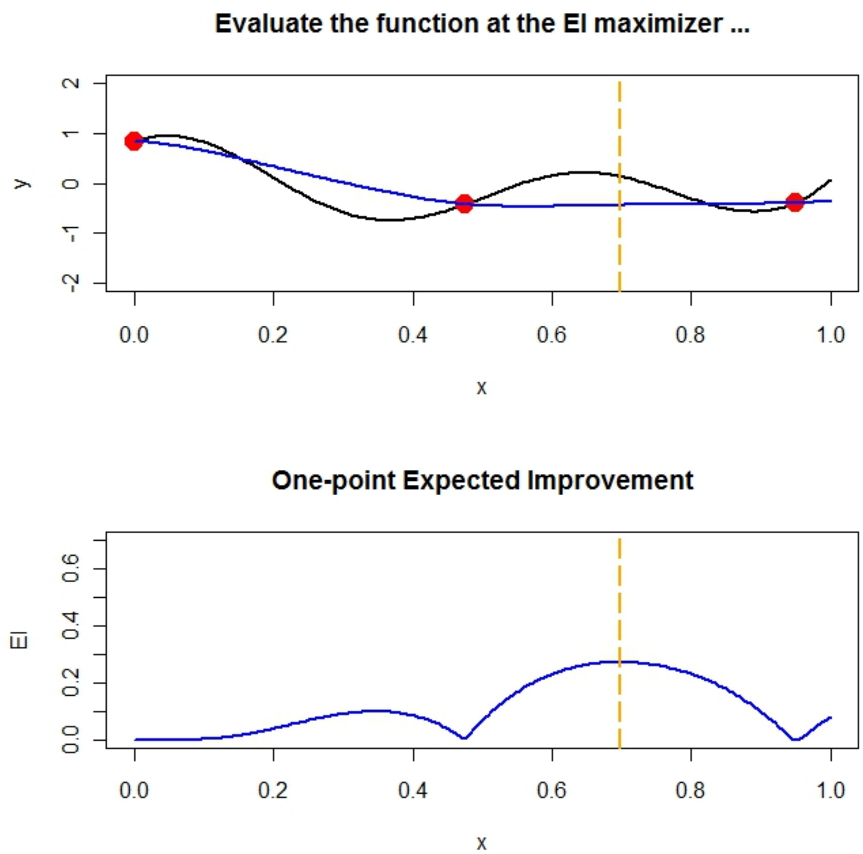

Figure 2.1: A first Kriging model, with maximization of the Expected Improvement.

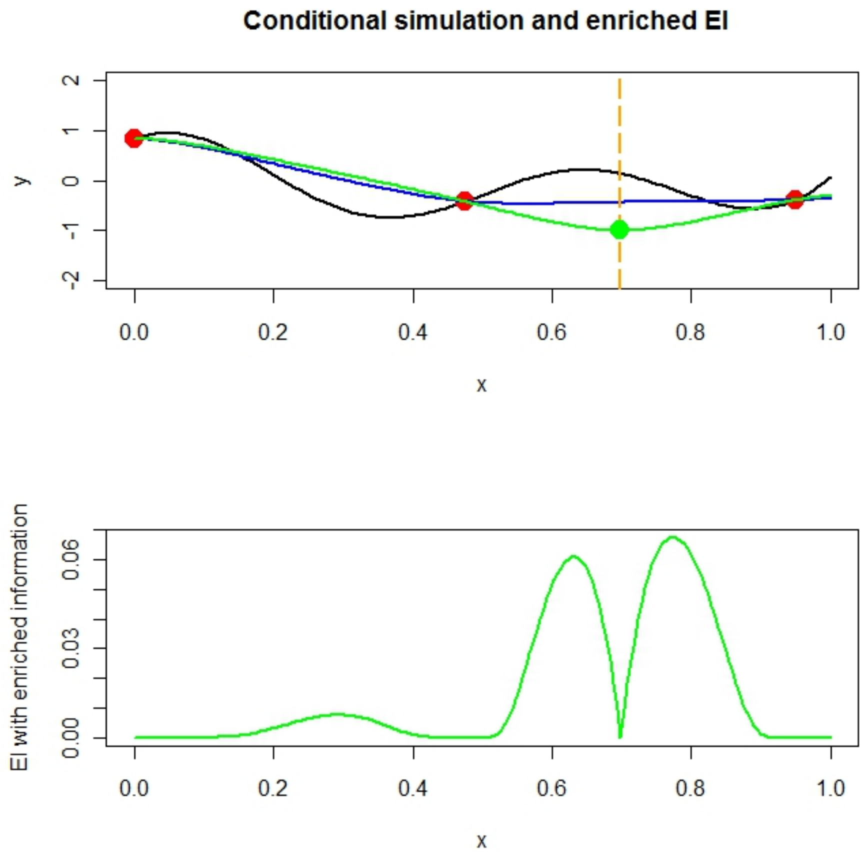

Figure 2.2: Evolution of EI afer one conditional simulation at $\mathbf{x}^{\text{busy}}$.

at $\mathbf{x}^{\text{busy}}$. Furthermore, since the simulated value of $Y(\mathbf{x}^{\text{new}})$ is particularly low, and hence appealing in our minimization context, EI is now pointing to values of $\mathbf{x}$ close to $\mathbf{x}^{\text{busy}}$ (note the EI values of smaller magnitude than at the previous step). However, choosing the global maximizer of this EI for $\mathbf{x}^{\text{new}}$ does not completely make sense. Indeed, the simulated value of $Y(\mathbf{x}^{\text{new}})$ is arbitrary, and another simulated value would maybe lead to discard this point and go for a candidate $\mathbf{x}^{\text{new}}$ in a different zone of the input space.

It seems therefore necessary to propagate the uncertainty concerning the new EI function by integrating it over the possible scenarii concerning the value of $Y(\mathbf{x}^{\text{new}})$, or at least to circumvent the problems associated to this unkown value and its variability by using some adapted heuristic. Here we propose at first to use brute force: simulating 100 values of $Y(\mathbf{x}^{\text{new}})$ conditional on the past, calculate the 100 associated EI with enriched information,

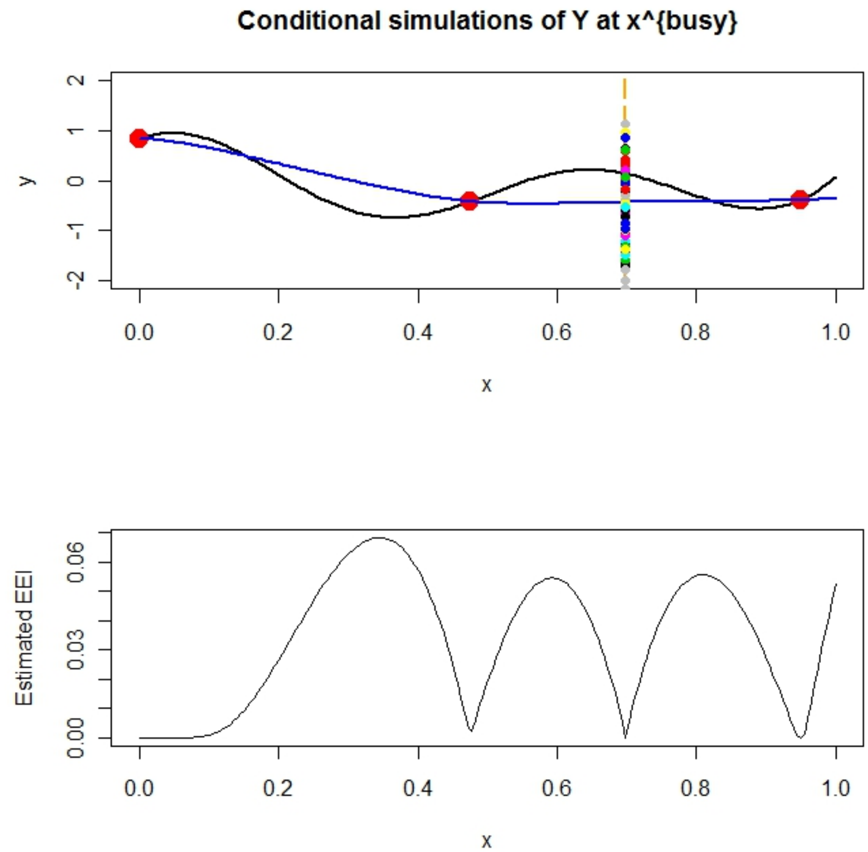

Figure 2.3: 100 conditional simulations at $\mathbf{x}^{\text{busy}}$ and the corresponding EEI estimate.

and take the arithmetic mean of them as proxy for EEI.

The lower graph of figure 2.3 above represents the estimate of EEI obtained by averaging out the 100 EI's with enriched information obtained by conditional simulation of $Y(\mathbf{x}^{\text{new}})$ (maximum reached at $\mathbf{x} \approx 0.347$). The correponding values are materialized by colored points on the upper graph of figure 2.3. The 100 EI curves are printed on figure 2.4, with colors matching the one previously used on figure 2.3 for the simulated responses at $\mathbf{x}^{\text{new}}$.

A natural surrogate for EEI maximization is to maximize its Monte Carlo estimate $\widehat{\text{EEI}}$, such as represented on figure 2.3 (lower graph, see maximization result above). However, n_{sim} has to be chosen sufficiently large in order to guarantee a good approximation of EEI, which may cause high computational costs. Furthermore, based on figure 2.4 above, one can imagine that a small set of carefully selected EI's might lead the user to a simplified

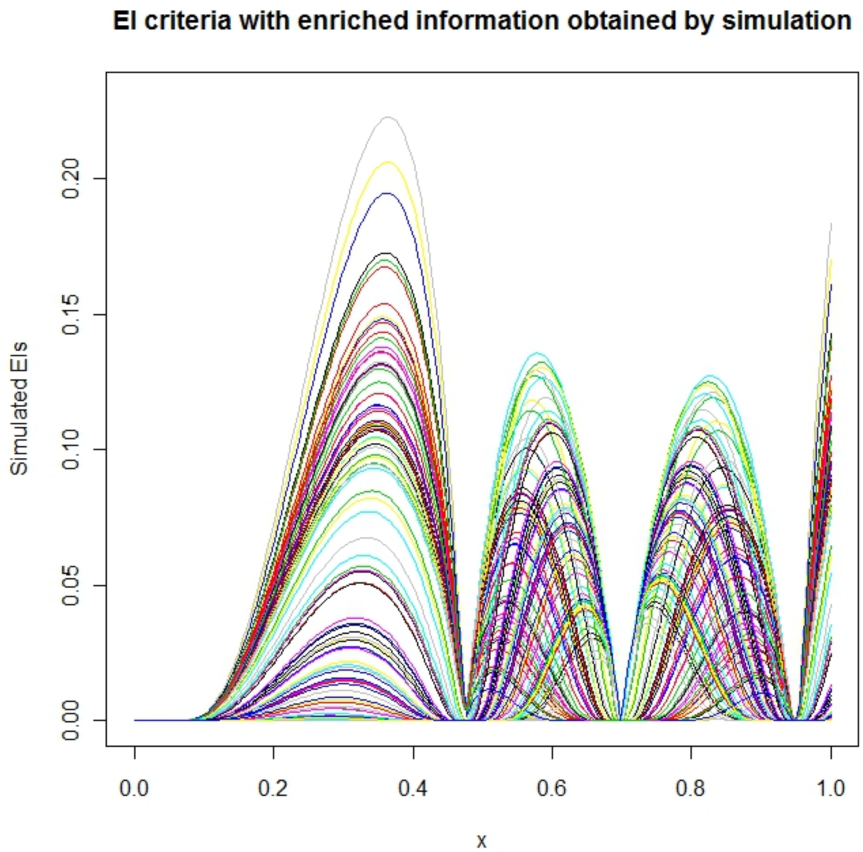

Figure 2.4: 100 simulations of the Expected Improvement with enriched information, based on the simulated $Y(\mathbf{x}^{\mathrm{new}})$ of figure 2.3.

description of this family of curves, yielding an efficient proxy for EEI maximization. Let us now explore these questions in some more detail.

3 Maximization of the EEI and related problems

3.1 Direct approach: maximizing the mean of a random function

3.1.1 A bit of theory and bibliography

Maximizing the previously introduced *EEI* criterion is in fact a special case of a classic stochastic optimization problem, i.e. the maximization of an expectation function such as

$$\mathbf{x} \in D \longrightarrow \mathbb{E}[Q(\mathbf{x}, \zeta)], \tag{9}$$

where Q is a function depending on $\mathbf{x}$ and on the scalar- or vector-valued random variable ζ), and the distribution of ζ is assumed known or at least reproducible by simulation.

Such problems are commonly solved in practice thanks to approximations of $\mathbb{E}[Q(\mathbf{x}, \zeta)]$ by finite averages of the form $\frac{1}{N}\sum_{i=1}^{N} Q(\mathbf{x}, \zeta^{(i)})$, where N and the $\zeta^{(i)}$ are chosen in order to get a good trade-off between the sample size (N, wished small) and a good approximation of the distribution of ζ by a uniform discrete law over the N-sample $\{\zeta^{(1)}, \dots, \zeta^{(N)}\}$.

One natural idea is to generate $\{\zeta^{(1)}, \dots, \zeta^{(N)}\}$ by Monte Carlo, i.e. to draw by simulation a N-sample of variables independently following the distribution of ζ. This is what Shapiro calls *SAA* (*Sample Average Approximation*). To quote him,

> " The term sample average approximation method was coined in [47], although this approach was used long before that paper under various names. Statistical properties of the SAA method are discussed in [74] and complexity of two and multi-stage stochastic programming in [75] ... Rates of convergence of Monte Carlo and Quasi-Monte Carlo estimates of the expected values are discussed in [58]. Numerical experiments with the SAA method are reported in [50, 51, 83], for example. "

Such an approach makes sense, and without going into difficult detail concerning the different kinds of stochastic convergence, the law of large numbers ensures (under specific conditions) that the Sample Average Approximation will resemble the expectation objective function $\mathbb{E}[Q(\mathbf{x}, \zeta)]$ always more as N grows, and that so will do their respctive maxima and maximizers. See e.g. [81] for a discussion on sufficient conditions for the sequence of proxy maximizers to converge to the actual maximizer in a suitable way.

Anyway, even under conditions where nice asymptotic properties might be guaranteed, the major question to answer in practice is the number of Monte Carlo simulations N needed to ensure that $\arg\min \frac{1}{N}\sum_{i=1}^{N} Q(\mathbf{x}, \zeta^{(i)})$ is a reasonable approximation of $\arg\min \mathbb{E}[Q(\mathbf{x}, \zeta)]$. A theoretical study of this would certainly be very demanding, and is somehow out of scope here. However, we will come back in the next section to our previous example, and propose some illustrations related to these convergence questions.

3.1.2 Back to the main example

Before focusing on the estimation of EEI and on the maximization of the obtained estimate, let us brievly study the variability of the maximum and maximizer of EI with enriched information when y_{busy} randomly varies following the distribution defined by density 5. The y_{busy} values used here are the same as on figures 2.3 and 2.4. The variability shown by the upper left histogram of fig. 3.1 illustrates the fact that the value of y_{busy} significantly influences the EI with enriched information and its maximum. Indeed, a very low value of y_{busy} changes the current minimum, and makes it harder to exceed. On the other hand, a very high value of y_{busy} does not affect the current minimum but may locally attract the Kriging predictor, and prevents from any improvement in the close neighbourhood of $\mathbf{x}^{\text{busy}}$. However, depending on the rigidity of the covariance kernel used, such high y_{busy} values might cause some low overshooting and could magnify the value of EI in farther areas. Finally, medium values of y_{busy}, in particular slightly above the current minimum, are susceptible of keeping the EI maximum at the same order of magnitude as earlier.

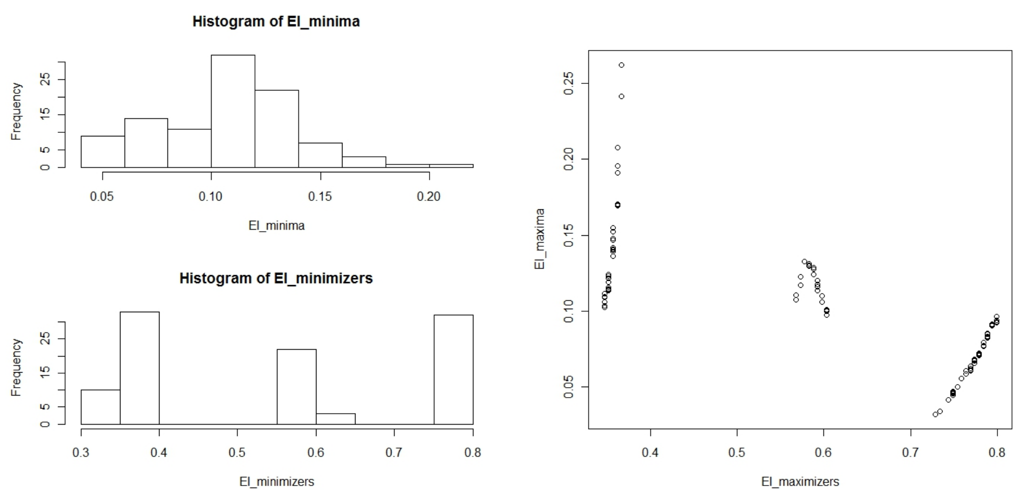

Figure 3.1: Left: histograms of the minima and minimizers of the 100 EI with simulated enriched information represented on figure 2.4. Right: joint distribution.

The lower left histogram of fig. 3.1 shows how the maximizer of the EI with enriched information behaves when y_{busy} varies. We clearly distinguish three modes. The first one, around 0.35, roughly matches with the left local maximum of EI at the previous step. As can be seen on the left graph of fig. 3.2, this mode corresponds to relatively high values of y_{busy}. The second mode, around 0.6, gathers a smaller number of points with medium y_{busy} values. A clear linear trend appears in the corresponding cluster of points on figure 3.2, higher values of y_{busy} tendencially matching with minimizers on the left, while smaller y_{busy} seem pushing the maximizer of the enriched EI closer to $\mathbf{x}^{\text{busy}}$. Note that a reverse trend is similarly visible on the first cluster, which may presumably be imputed to some kind of rigidity effect, as evocated earlier: a very high y_{busy} value would drastically lower EI in the close neighbourhood of $\mathbf{x}^{\text{busy}}$, but the associated curvature in the Kriging predictor would also create a steeper hill, moving the maximizer of the enriched $\mathbf{x}^{\text{busy}}$ to the right of the left cluster. Finally, the right cluster corresponds to the lowest simulated values of y_{busy}, with a similar slope to the one of the first cluster. A moderatly small value of y_{busy} causes a slight decrease of the current minimum response, and sends the maximizer of the enriched EI to a zone with high variance, the right of the third cluser. As y_{busy} decreases further, however, the influence of this very appealing value at $\mathbf{x}^{\text{busy}}$ reaches the third cluster, progressively moving the maximizer of the enriched EI to the left of the cluster.

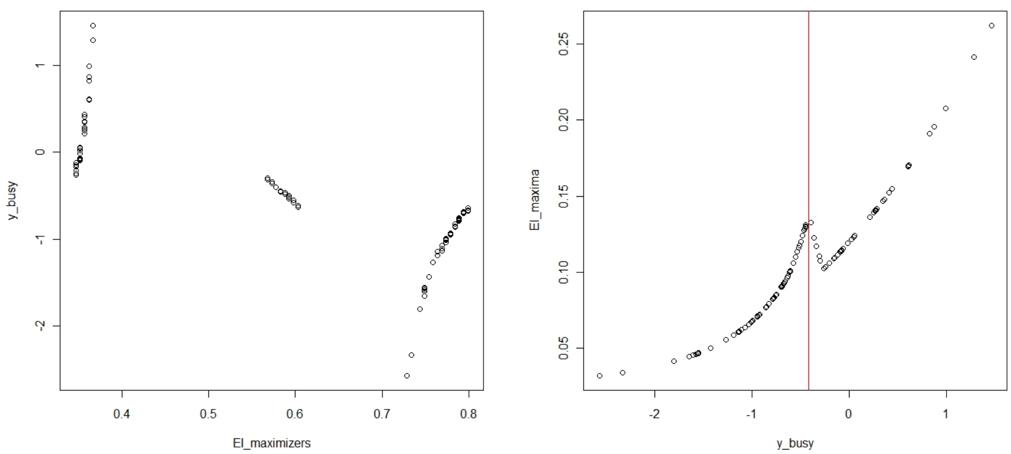

Figure 3.2: Left: Dependence bewteen the location of the maximizer of the enriched EI and y_{busy}. Right: maximum of the enriched EI as a function of y_{busy}.

The right graph of fig. 3.2 represents the evolution of the maximum of the enriched EI as y_{busy} increases. Contrarily to the previous graph, this scatter plot seems to stem from a continuous function, which makes sense from a mathematical point of view. Indeed, the enriched EI is a continuous function of y_{busy}, and hence absolutely continuous over any compact set. One can show that the maximum of EI with enriched information is consequently a continuous function of y_{busy}. As could already be felt on the basis of the previous graphs, only the range of medium values of y_{busy} breaks the monotonicty of this function. Indeed, for very small values of y_{busy}, there isn't much room for improvement. Then, the maximum increases until y_{busy} reaches the actual current minimum (red vertical line on the right graph of fig. 3.2). At this stage, the maximum of the enriched EI starts decreasing with increasing y_{busy}: higher y_{busy} now do not change anything to the minimum of reference, but do pull the response surface up in the neighbourhood of the maximizer of the enriched EI, hence affecting its maximum negatively through the increasing Kriging mean predictor. At a certain point, however, the maximizer of the enriched EI jumps to the left cluster, and the associated maximum starts again to increase with y_{busy}. This last effect seems due to the previously evocated rigidity artifact, a higher observation value at $\mathbf{x}^{\text{busy}}$ causing deeper valleys in the surrounding clusters.

We now continue with the maximization of the estimated EEI, and with an investigation concerning the number of y_{busy} simulations needed here to achieve a good result.

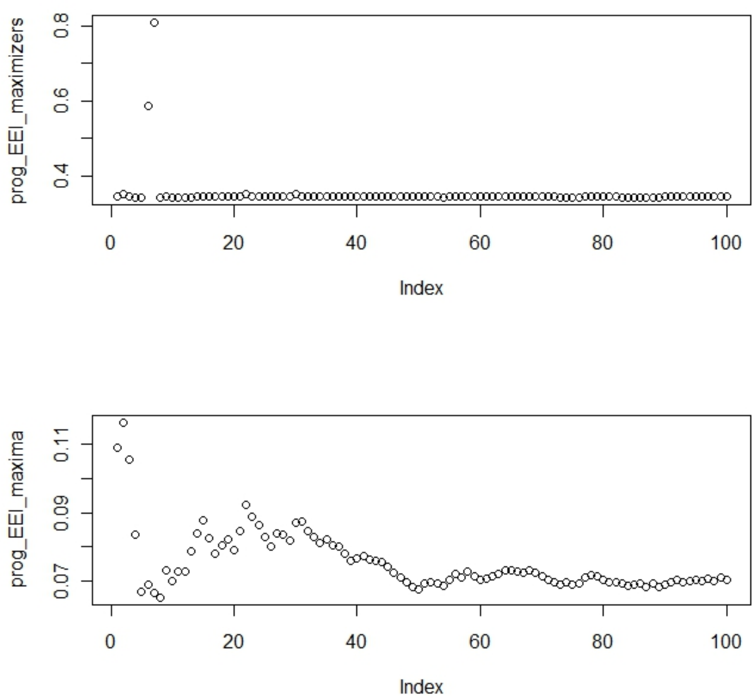

Figure 3.3: Evolution of EEI maxima and maximizers with progressively growing sample

Fig. 3.3 shows how the maximum and the global maximizer of the estimated EEI evolves, when the sample size of y_{busy} values used for the estimation of EEI varies. Here, only the 100 values previously simulated are used, and the increasing number of y_{busy} values means that nested subsets of that 100 values are progressively considered (the first one, the two first ones, and so on ...). It is quite impressive to observe how fast the maximizer of the estimated EEI converges to the actual EEI maximizer. Here, after an unstable phase during the 8 first iterations (averaging with less or equal than the 8 first simulated y_{busy} values), the maximizer of the estimated EEI stabilizes to a unique point. Meanwhile, the estimated EEI maximum seems to converge with a classical rate of convergence of $n_{\text{sim}}^{-\frac{1}{2}}$.

In order to see in what measure this early stability is due to the particular sample drawn in the previous example, let us generate other samples at random following the same distribution and observe the variability of the new estimated EEI maxima and maximizers. We follow for each simulated sample the same nesting scheme as above.

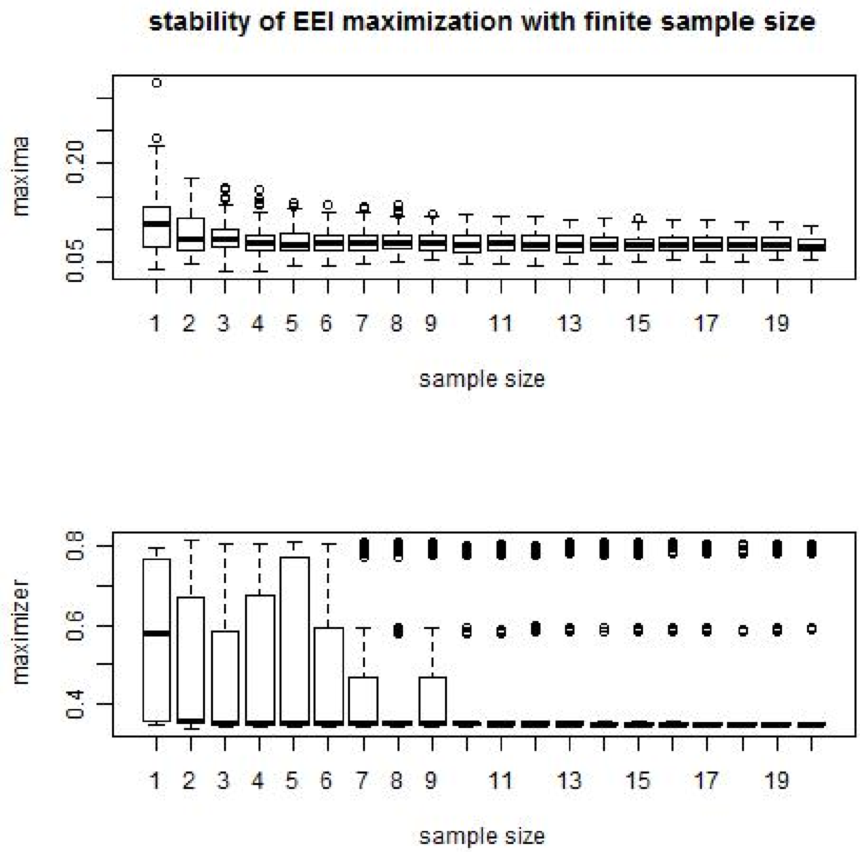

Figure 3.4: Boxplots retracing the evolution of EEI maxima and maximizers with progressively growing sample (1 to 20 here), when the sample is drawn 100 times at random.

Despite the particularly good results obtained in the previous example (stability for a sample of size greater than 8), the boxplots of fig. 3.4 are here to signify that this first attempt was rather lucky, and that significantly larger samples were nedeed in order to guarantee a sucessful approximate EEI maximization with high probability.

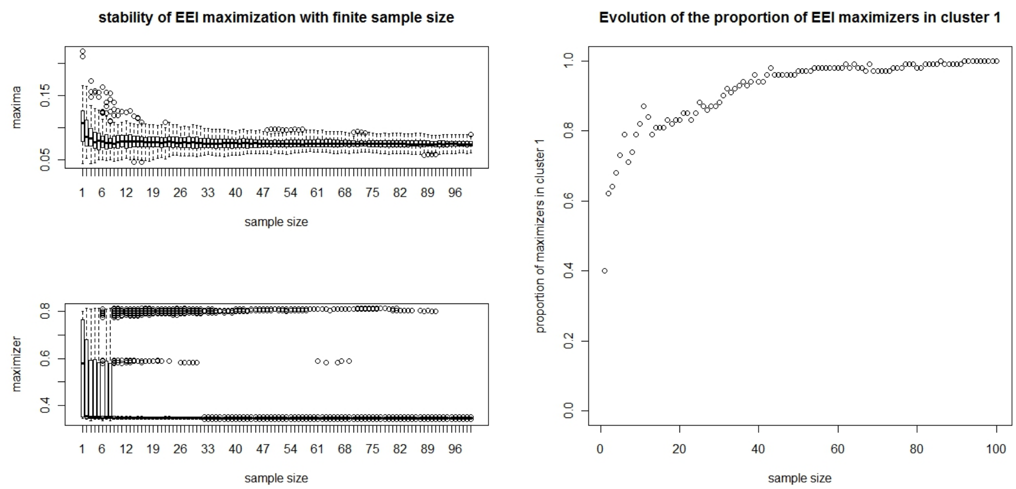

Figure 3.5: Left: boxplots retracing the evolution of EEI maxima and maximizers with progressively growing sample (1 to 100 here, 100 sample realizations). Right: proportion of approximate EEI maximizers in the first cluster in function of the sample size.

Let us finally see what happens if we replicate 100 times the complete previous experiment, i.e. every time with 100 simulated y_{busy} values. The results are presented on fig. 3.5. On the left graph, the boxplots confirm again that we had been lucky in the previous experiment: here, we need to invest about 100 y_{busy} simulations to get rid of approximate EEI maximizers landing in the right cluster (upper points in lower graph). As we can see on the right graph of figure 3.5, investing 20 conditional simulations leads to the good cluster (i.e. near to the actual EEI maximizer) in about 80% of the cases. This leaves 20% of misleading approximate optimization results, which a reasonable user may choose not to accept. Always on the same graph, increasing the sample size to 50 leads to a probability of about 95% to get an approximate optimizer in the good cluster, while the 100% probability level seems to be reached for a sample of 100 simulated y_{busy} values. Working with an average of so many EI functions can be computationally very demanding, so that reliable and efficient methods involving a limited number of y_{busy} values might be appealing. We address this issue in the following section, where encouraging results are obtained in the modest present framework case of a unique busy processor.

3.2 Towards reducing to a limited number of scenarii

3.2.1 Theory: per un pugno di quantili ...

In the last section, we proposed to approximate EEI by averaging several EI criteria with enriched information, where the unkown response y_{busy} was randomly simulated following the "Kriging conditional distribution" of $Y(\mathbf{x}^{\text{busy}})|A$.

Here we depart from this Monte Carlo approach by deterministically choosing a representative subset of y_{busy} values, and restricting the EEI maximization to the maximizers of the different EI enriched with the latter y_{busy} values.

The approach proposed here to get a subset of y_{busy} values representing well the Kriging conditional distribution is to select quantiles of it. We focus on quantiles of level α, α varying regularly between 0.05 and 0.95 with a finesse depending on the chosen number of subset points. Denoting n_{quant} this number of points, the proposed protocole reads:

1. Calculate n_{quant} quantiles of $Y(\mathbf{x}^{\text{busy}})|A$
2. For y_{busy} equal to each of these quantiles, maximize the enriched EI
3. For each of the obtained maximizers, estimate EEI
   (possibly with the same n_{quant} y_{busy} values as earlier)
4. Choose the enriched EI maximizer with highest estimated EEI

Such a technique should a priori not be expected to work well. In fact, in a general stochastic optimization framework, nothing guarantees that the maximum of an expected function will be reached at the maximizer of some of the underlying realizations. It is even easy to build counter-examples: consider for instance the piecewise linear function f_1 over $[-1, 1]$ with slope 2 over $[-1, 0]$ and slope 1 over $[-1, 0]$, f_2 similarly defined with slopes -1 and -2, and f being a random function equal to f_1 or f_2 with probability one half. It is then clear that f_1 takes its global maximum (1) at 1, f_2 takes its global maximum (1) at -1, but none of these points is a maximizer of the expectation of f. indeed, the expectation value of f at these two points is -1, while the latter takes the value 0 (its maximum) at 0, which is not a maximum of f_1 nor f_2. Note that a variant of this counter-example with positive functions may be constructed by a simple vertical translation.

Coming back to our EI functions with enriched information, things are going much better than in the counter-example above. Indeed, it is quite unlikely that a point being a EI maximizer for some value of y_{busy} becomes dramatically bad for another y_{busy} value. Furthermore, it is natural to suspect that a point maximizing EI with a high EI value for some y_{busy} is a good one for EEI. If however the high potential of a point would be due to an artifact and rely on the precise y_{busy} value considered in the enriched EI, then we could expect that the fact of averaging enriched EI values over several y_{busy} candidates would smooth this outlier out and cancel the dominantion of this point. Even if we are not convinced that this method should work in every situation, the example below shows quite encouraging results. Before coming to any general conclusions however, the latter have to be questioned and complemented by further investigations.

3.2.2 Back to the main example

Let us first choose $n_{\text{quant}} = 10$ and detail review the protocole above in the framework of our main example. We recall that the EEI maximizer obtained by Monte Carlo with a 100-sample over a 200-elements grid was 0.3467337. Now, we replace the Monte-Carlo sample by the 10 quantiles of $Y(\mathbf{x}^{\text{busy}})|A$ with respective levels $0.05, 0.1, \ldots, 0.9, 0.95$:

```
> quantiles
 [1] -1.52060808 -1.11769068 -0.87799880 -0.68650068 -0.51454523 -0.34811045
 [7] -0.17615500  0.01534313  0.25503501  0.65795240
```

Maximizing the 10 obtained enriched EI criteria, we get the following candidates:

```
> EI_maximizers
 [1] 0.7487437 0.7688442 0.7788945 0.7939698 0.5929648 0.5728643 0.3467337
 [8] 0.3517588 0.3567839 0.3618090
```

We notice that the seventh is exactly the previous EEI maximizer, and that the three following candidates are in its immediate neighbourhood. These values are plotted against the corresponding enriched EI values on the left graph of fig. 3.6. We can observe that the three clusters are well represented, even if the associated enriched EI estimates do not respect the order observed in fig 3.2. This is due to the the approximation of these enriched EI on the basis of 10 y_{busy} values only.

We now restrict our attention to the 10 previously derived maximizers, and compute EEI estimates obtained at these points, enriched EI averages over the same 10 y_{busy} values. The right graph of fig 3.2 represents the 10 enriched EI values obtained for each one of the 10 points. The four last points seem to clearly dominate the others in terms of median EI, which would not be necessarily the case for other metrics penalizing uncertainty. Here we give the estimated EEI values obtained by the "quantile approximation":

```
> EEI_values
 [1] 0.03858103 0.04777052 0.05104971 0.05436474 0.05516403 0.05399162
 [7] 0.07446641 0.07434650 0.07404384 0.07355171
```

And the point maximizing this rough EEI estimate among the 10 candidate points is the 7^{th}, i.e. 0.3467337. This excellent result might seem relatively surprising, in particular its precision. But one may not forget that the search (enriched EI maximization) has

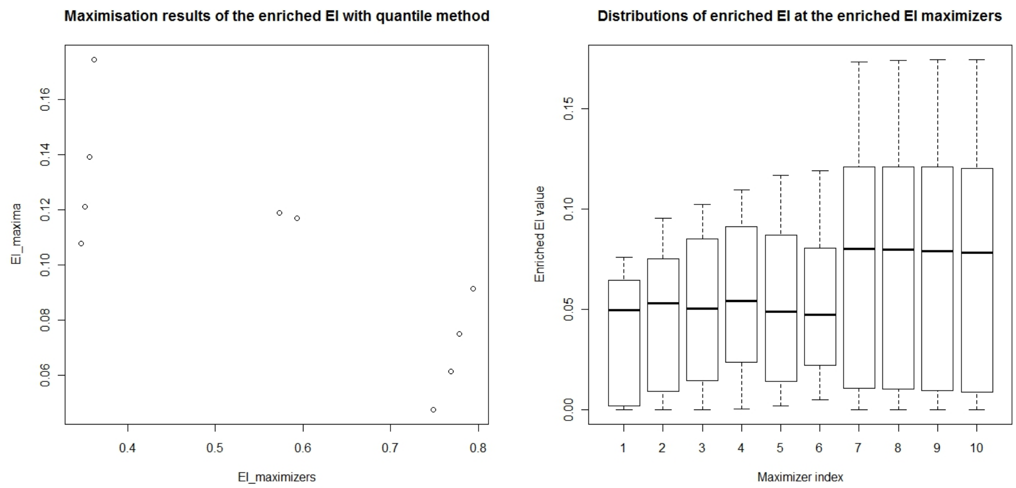

Figure 3.6: Left: maximizers and maxima of the enriched EI criteria associated with the 10 quantile-based y_{busy} values mentioned above. Right: enriched EI

been restricted in all cases to 200 points. Furthermore, this good finding might be highly depending on the number of quantiles, and even be non monotonic in n_{quant}, which could lead to serious practical issues in the *a priori* choice of n_{quant}. So let us go one step deeper, and investigate the influence of n_{quant} on the approximate EEI maximizer obtained by the proposed method. Here below are the results obtained by repeating the quantile approximation and optimization method for n_{quant} varying from 1 to 30:

```
> maximizers
 [1] 0.7487437 0.3618090 0.3618090 0.3467337 0.3517588 0.3517588 0.3467337
 [8] 0.3517588 0.3467337 0.3467337 0.3467337 0.3467337 0.3467337 0.3467337
[15] 0.3467337 0.3467337 0.3467337 0.3467337 0.3467337 0.3467337 0.3467337
[22] 0.3467337 0.3467337 0.3467337 0.3467337 0.3467337 0.3467337 0.3467337
[29] 0.3467337 0.3467337
```

We observe that after a first point (0.7487437) in the third cluster, all followers belong to the first cluster, and are close neighbours of 0.3467337. Starting from $n_{\text{quant}} = 9$, they are even all equal to 0.3467337. Compared to the variability observed on fig. 3.4 for the case of Monte Carlo EEI estimation, the "quantile method" provides a deterministic result, free of any variability, and with very good performances on this example in terms of locating the EEI correctly based on a small number of scenarii. Indeed, 2 or 3 quantile values are necessary to locate the EEI maximizer in the right zone ("first cluster"), while $n_{\text{quant}} \geq 9$ guarantees the same result as found by Monte Carlo with a 100-elements sample (search restricted to a 200-points grid). Despite the impressively good results obtained on this example, one has also to notice and remember that $n_{\text{quant}} = 1$ does lead to a sub-optimal cluster, which means in essence that plugging in the Kriging mean prediction (i.e. the KB strategy) as a proxy to the unknown y_{busy} response is clearly not the best thing to do.

4 Conclusion and research perspectives

A variant of the EI criterion, EEI, has been proposed for the case of asynchronous Kriging-based parallel optimization. Here we have restricted our attention to the canonical situation where one processor is free while another one is still busy with the evaluation of the objective function at a candidate point $\mathbf{x}^{\text{busy}}$. The unkown response at $\mathbf{x}^{\text{busy}}$ can be simulated conditional on the available information, hence providing a means to estimate EEI.

Estimating EEI by conditional simulations amounts to averaging several enriched EI criteria obtained with the simulated response values at $\mathbf{x}^{\text{busy}}$, i.e. applying several times a *Liar* strategy with a random lie generated by Monte Carlo. Maximizing the estimated EEI may then become quite computationally demanding, all the more so as the number of simulations gets large when it comes to approximating the actual EEI well. Here we relate that to a classical *stochastic optimization* problem, referred to as *Sample Average Approximation* in the litterature, and propose to simplify it by summarizing the conditional distribution of the unkown reponse(s) by a small set of representative scenarii.

The idea of reducing to a few scenarii is not new, and what is proposed here is just a first draft of what could be developed for EEI maximization. In our main example, summarizing the unkown reponse distribution by a few quantiles has given sensible results, and a good picture of the full Monte Carlo results (with 100 drawings) with a lower computationnal cost (10 evaluations). Moreover, restricting EEI maximization to the maximizers of the enriched EI criteria obtained by the "quantile method" has delivered the same results as by maximizing the 100-drawings Monte Carlo EEI estimate over a 200-points grid. This is a quite unexpected outcome, to be urgently investigated with further test configurations. Apart from a bigger scale testing, and the generalization to multiple busy and vacant processors, supplementary research could be done on several aspects including:

- a fine study of the impact of y_{busy} on the enriched EI and its maximizer(s),
- applying curve classification to the EI functions obtained with many y_{busy} values,
- distinguishing regions of low EI variation, in particular in higher dimensions.

The final report will include further tests and theoretical considerations related to the questions above. A further study of several articles listed in the following bibliography —and with the help of [82]— is expected to shed some more light on existing tricks in Monte Carlo methods and scenarii selection for stochastic optimization.

References

[1] N. E. Abboud, M. Y. Jaber, and N. A. Noueihed. Economic lot sizing with the consideration of random machine unavailability time. *Comput. Oper. Res.*, 27(4):335–351, 2000.

[2] P. Abel. Decisions in stochastic linear programming models under partial information. *Z. Angew. Math. Mech. 73, No.7-8, T 737-T 738*, 1993.

[3] Shabbir Ahmed and Alexander Shapiro. The sample average approximation method for stochastic programs with integer recourse. Optimization Online, `http://www.optimization-online.org`, 2002.

[4] Susanne Albers, Rolf H. Möhring, Georg Ch. Pflug, and Rüdiger Schultz. 05031 Summary – Algorithms for Optimization with Incomplete Information. In S. Albers, R.H. Möhring, G.Ch. Pflug, and R. Schultz, editors, *Dagstuhl Seminar 05031: Algorithms for Optimization with Incomplete Information*, `http://www.dagstuhl.de/05031`, 2005.

[5] H. Arsham. Techniques for Monte Carlo optimizing. *Monte Carlo Methods Appl.*, 4(3):181–229, 1998.

[6] F. Azadivar and J. Talavage. Optimization of stochastic simulation models. *Math. Comput. Simulation 22, 231-241*, 1980.

[7] C.A. Baker, L. T. Watson, B. Grossman, W. H. Mason, and R. T. Haftka. Parallel global aircraft configuration design space exploration. *Practical parallel computing*, pages 79–96, 2001.

[8] Russell R. Barton. Minimization algorithms for functions with random noise. *Amer. J. Math. Management Sci.*, 4(1-2):109–138, 1984. Statistics and optimization : the interface.

[9] R. Bellman. *Dynamic Programming.* Princeton University Press, Princeton, 1957.

[10] L.S. Belyaev. *The solution of complex optimization problems under uncertainty conditions. (Reshenie slozhnykh optimizatsionnykh zadach v usloviyakh neopredelennosti).* Akademiya Nauk SSSR, Sibirskoe Otdelenie, Sibirskij Ehnergeticheskij Institut. Novosibirsk: Izdatel'stvo "Nauka", Sibirskoe Otdelenie., 1978.

[11] Aharon Ben-Tal, Tamar Margalit, and Arkadi Nemirovski. Robust modeling of multistage portfolio problems. In *High performance optimization*, pages 303–328. Kluwer Acad. Publ., Dordrecht, 2000.

[12] D.P. Bertsekas. *Dynamic Programming and Optimal Control, Vol. 1*. Athena Scientific, Belmont, MA, 2007.

[13] Ruud Brekelmans, Lonneke Driessen, Herbert Hamers, and Dick den Hertog. Constrained optimization involving expensive function evaluations: A sequential approach. *European Journal of Operational Research*, 160(1):121 – 138, 2005. Applications of Mathematical Programming Models.

[14] E. Brochu, M. Cora, and N. de Freitas. A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning. Technical report, Department of Computer Science, University of British Columbia, 2009.

[15] Michael S. Casey and Suvrajeet Sen. The scenario generation algorithm for multistage stochastic linear programming. *Math. Oper. Res.*, 30(3):615–631, 2005.

[16] Yair Censor and Stavros A. Zenios. *Parallel Optimization: Theory, Algorithms and Applications*. Oxford University Press, Oxford, UK, 1997.

[17] Louis Anthony Cox, Jr. and Djangir A. Babayev. Optimization under uncertainty via random sampling of scenarios. I. *Appl. Comput. Math.*, 3(2):95–106, 2004.

[18] N.A.C. Cressie. *Statistics for spatial data*. Wiley series in probability and mathematical statistics, 1993.

[19] I. Deák. Computing probabilities of rectangles in case of multinormal distribution. *J. Statist. Comput. Simulation*, 26(1-2):101–114, 1986.

[20] I. Deák. Multidimensional integration and stochastic programming. In *Numerical techniques for stochastic optimization*, volume 10 of *Springer Ser. Comput. Math.*, pages 187–200. Springer, Berlin, 1988.

[21] István Deák. Monte Carlo methods for computing probabilities of sets in higher-dimensional spaces in case of normal distribution. *Alkalmaz. Mat. Lapok*, 4(1-2):35–94 (1979), 1978.

[22] István Deák. Two-stage stochastic problems with correlated normal variables: computational experiences. *Ann. Oper. Res.*, 142:79–97, 2006.

[23] Morris H. DeGroot. Optimization with an uncertain objective. *Stochastics*, 21(2):97–112, 1987.

[24] G. Dellino, P. Lino, C. Meloni, and A. Rizzo. Kriging metamodel management in the design optimization of a cng injection system. *Mathematics and Computers in Simulation*, 79(8):2345 – 2360, 2009. Nonstandard Applications of Computer Algebra;

Applied Scientific Computing VI: Numerical Grid Generation, Approximation and Visualization; Modelling and Managing Financial Risk.

[25] Bernard Delyon and Anatoli Juditsky. Stochastic optimization with averaging of trajectories. *Stochastics Stochastics Rep.*, 39(2-3):107–118, 1992.

[26] Nico Di Domenica, George Birbilis, Gautam Mitra, and Patrick Valente. Stochastic programming and scenario generation within a simualtion framework. Stochastic Programming E-Print Series, `http://www.speps.org`, 2004.

[27] L. C. W. Dixon and G. P. Szego. *Towards Global Optimisation 2*, chapter The global optimisation problem: an introduction, pages 1–15. North Holland, Amsterdam., 1978.

[28] José H. Dulá. An upper bound on the expectation of simplicial functions of multivariate random variables. *Math. Programming*, 55(1, Ser. A):69–80, 1992.

[29] C. A. Floudas and C. E. Gounaris. A review of recent advances in global optimization. *Journal of Global Optimization*, 45:3–38, 2009.

[30] C.A. Floudas, I.G. Akrotirianakis, S. Caratzoulas, C.A. Meyer, and J. Kallrath. Global optimization in the 21st century: Advances and challenges. *Computers & Chemical Engineering*, 29(6):1185 – 1202, 2005. Selected Papers Presented at the 14th European Symposium on Computer Aided Process Engineering.

[31] A. I. J. Forrester, N. W. Bressloff, and A. J Keane. Optimization using surrogate models and partially converged computational fluid dynamics simulations. *Proc. R. Soc. A*, 462:2177–2204, 2006.

[32] A. I. J. Forrester and A. J. Keane. Recent advances in surrogate-based optimization. *Progress in Aerospace Sciences*, 45(1-3):50 – 79, 2009.

[33] D. Ginsbourger, C. Helbert, and L. Carraro. Discrete mixtures of kernels for kriging-based optimization. *Quality and Reliability Eng. Int.*, 24(6):681–691, August 2008.

[34] D. Ginsbourger, R. Le Riche, and L. Carraro. *Computational Intelligence in Expensive Optimization Problems*, chapter "Kriging is well-suited to parallelize optimization". Studies in Evolutionary Learning and Optimization. Springer-Verlag, 2010.

[35] B. Glaz, P. P. Friedmann, and L Liu. Surrogate based optimization of helicopter rotor blades for vibration reduction in forward flight. *Journal of Global Optimization*, 35:341–363, 2007.

[36] R. B. Gramacy and H. K. H. Lee. Adaptive design and analysis of supercomputer experiments. *Technometrics*, 51:130–145, 2009.

[37] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning*. Springer, 2001.

[38] A. F. Hernandez and M. A. Grover. Stochastic dynamic predictions using kriging for nanoparticle synthesis. In Caludio Augusto Oller do Nascimento Rita Maria de Brito Alves and Jr. Evaristo Chalbaud Biscaia, editors, *10th International Symposium on Process Systems Engineering: Part A*, volume 27 of *Computer Aided Chemical Engineering*, pages 357 – 362. Elsevier, 2009.

[39] D. Huang, T.T. Allen, W. Notz, and R.A. Miller. Sequential kriging optimization using multiple fidelity evaluations. *Sructural and Multidisciplinary Optimization*, 32:pp. 369–382 (14), 2006.

[40] D. Huang, T.T. Allen, W. Notz, and N. Zheng. Global optimization of stochastic black-box systems via sequential kriging meta-models. *Journal of Global Optimization*, 34:441–466, 2006.

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

[42] D.R. Jones, M. Schonlau, and W.J. Welch. Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13:455–492, 1998.

[43] I. Kaymaz. Application of kriging method to structural reliability problems. *Structural Safety*, 27(2):133 – 151, 2005.

[44] A. J. Keane. Statistical improvement criteria for use in multiobjective design optimization. *AIAA Journal*, 44, 2006.

[45] M.C. Kennedy and A. O'Hagan. Predicting the output from a complex computer code when fast approximations are available. *Biometrika*, 87:1–13, 2000.

[46] J. P.C. Kleijnen, W. van Beers, and I. van Nieuwenhuyse. Constrained optimization in expensive simulation: Novel approach. *European Journal of Operational Research*, 202(1):164 – 174, 2010.

[47] A. J. Kleywegt, A. Shapiro, and T. Homem-de Mello. The sample average approximation method for stochastic discrete optimization. *SIAM J. Optimization*, 12:479–502, 2001.

[48] D.G. Krige. A statistical approach to some basic mine valuation problems on the witwatersrand. *J. of the Chem., Metal. and Mining Soc. of South Africa*, 52 (6):119139, 1951.

[49] S. J. Leary, A. Bhaskar, and Keane. A. J. A constraint mapping approach to the structural optimization of an expensive model using surrogates. *Optimization and Engineering*, 2:385–398, 2004.

[50] J. Linderoth, A. Shapiro, and S. Wright. The empirical behavior of sampling methods for stochastic programming. *Annals of Operations Research*, 142:215–241, 2006.

[51] W.K. Mak, D.P. Morton, and R.K. Wood. Monte carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24:4756, 1999.

[52] F.B. Manolache and S. Costiner. Parallel processing approaches for multi disciplinary optimization algorithms. *Sci. Ann. Cuza Univ. 11*, pages 69–79, 2002.

[53] J.D. Martin and T.W. Simpson. A monte carlo simulation of the kriging model. In *10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, NY, August 30 - September 2, AIAA, AIAA-2004-4483.*, 2004.

[54] J.D. Martin and T.W. Simpson. Use of kriging models to approximate deterministic computer models. *AIAA Journal*, 43 (4):853–863, 2005.

[55] R. Martinez-Cantin, N. de Freitas, E. Brochu, J. Castellanos, and A. Doucet. A bayesian exploration-exploitation approach for optimal online sensing and planning with a visually guided mobile robot. *Autonomous Robots*, 27:93–103, 2009.

[56] J. Mockus. *Bayesian Approach to Global Optimization.* Kluwer academic publishers, 1988.

[57] H. Nakayama, K. Inoue, and Y. Yoshimori. Approximate optimization using computaional intelligence and its application to reinforcement of cable-stayed bridges. In *Frontiers in Artificial Intelligence and Applications; Vol. 149 Proceeding of the 2006 conference on Integrated Intelligent Systems for Engineering Design*, 2006.

[58] H. Niederreiter. *Random Number Generation and Quasi-Monte Carlo Methods.* SIAM, Philadelphia, 1992.

[59] A. O'Hagan. Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering and System Safety*, 91(91):1290–1300, 2006.

[60] M. A. Osborne, R. Garnett, and Roberts S. J. Gaussian processes for global optimization. In *3rd International Conference on Learning and Intelligent Optimization (LION3)*, 2009.

[61] P. M. Pardalos, H. E Romeijn, and H. Tuy. Recent developments and trends in global optimization. *Journal of Computational and Applied Mathematics*, 124(1-2):209 – 228, 2000.

[62] W.B. Powell. *Approximate Dynamic Programming. Solving the Curses of Dimensionality.* Wiley, 2007.

[63] N. V. Queipo, R.T. Haftka, W. Shyy, T. Goel, R. Vaidyanathan, and P.K. Tucker. Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41:1–28, 2005.

[64] R Development Core Team. R: A language and environment for statistical computing, 2006.

[65] C.E. Rasmussen and K.I. Williams. *Gaussian Processes for Machine Learning.* M.I.T. Press, 2006.

[66] B.D. Ripley. *Stochastic Simulation.* John Wiley and Sons, New York, 1987.

[67] J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design and analysis of computer experiments. *Statistical Science*, (4):409–435, 1989.

[68] T.J. Santner, B.J. Williams, and W.J. Notz. *The Design and Analysis of Computer Experiments.* Springer, 2003.

[69] M. Sasena, P. Papalambros, and P. Goovaerts. Global optimization of problems with disconnected feasible regions via surrogate modeling. In *AIAA/ISSMO Symposium on Multidisciplinary Analysis and Optimi*, 2002.

[70] M. J. Sasena, P. Papalambros, and P. Goovaerts. Exploration of metamodeling sampling criteria for constrained global optimization. *Journal of Engineering Optimization*, 2002.

[71] M. J. Sasena, P. Papalambros, and P. Goovaerts. Exploration of metamodeling sampling criteria for constrained global optimization. *Engineering Optimization*, 34:263–278, 2002.

[72] M.J. Sasena. *Flexibility and Efficiency Enhancements for Constrained Global Design Optimization with Kriging Approximations.* PhD thesis, University of Michigan, 2002.

[73] M. Schonlau. *Computer Experiments and Global Optimization.* PhD thesis, University of Waterloo, 1997.

[74] A. Shapiro. *Stochastic Programming, Handbook in OR & MS, Vol. 10,*, chapter Monte Carlo sampling methods. North-Holland Publishing Company, Amsterdam, 2003.

[75] A. Shapiro and A. Nemirovski. *Continuous Optimization: Current Trends and Applications*, chapter On complexity of stochastic programming problems, pages 111–144. Springer, 2005.

[76] A. Sóbester, S.J. Leary, and A.J. Keane. A parallel updating scheme for approximating and optimizing high fidelity computer simulations. *Structural and Multidisciplinary Optimization*, 27:371–383, 2004.

[77] N.Z. Šor and M.B. Ščepakin. Multistage stochastic programming problems in parametric form. In *Theory of optimal solutions (Proc. Sem., Kiev, 1968), No. 3 (Russian)*, pages 64–73, Kiev, 1969. Akad. Nauk Ukrain. SSR.

[78] M.L. Stein. *Interpolation of Spatial Data, Some Theory for Kriging.* Springer, 1999.

[79] E. Stinstra and D. den Hertog. Robust optimization using computer experiments. *European Journal of Operational Research*, 191(3):816 – 837, 2008.

[80] M. A. Taddy, H. K. H. Lee, G. A. Gray, and J. D. Griffin. Bayesian guided pattern search for robust local optimization. *Technometrics*, 51:389–401, 2009.

[81] A.W. Van der Vaart. *Asymptotic Statistics.* Cambridge Series in Statistical and Probabilistic Mathematics, 1998.

[82] Maarten H. van der Vlerk. Stochastic programming bibliography. World Wide Web, `http://mally.eco.rug.nl/spbib.html`, 1996-2007.

[83] B. Verweij, S. Ahmed, A.J. Kleywegt, G. Nemhauser, and A. Shapiro. The sample average approximation method applied to stochastic routing problems: a computational study. *Computational Optimization and Applications*, 24:289333, 2003.

[84] J. Villemonteix. *Optimisation de fonctions coûteuses: Modèles gaussiens pour une utilisation efficace du budget dévaluations : théorie et pratique industrielle.* PhD thesis, Université Paris-sud XI, Facult des Sciences dOrsay, 2008.

[85] D. Williams. *Probability with Martingales.* Cambridge University Press, 1991.

[86] A.I. Ždanok. Relaxation of stochastic optimization algorithms in the presence of noise. *Problemy Sluchaĭn. Poiska*, 7:123–134, 318, 1978. Adaptation problems in technical systems.

[87] A Zhigljavsky and A. Zilinskas. *Stochastic global optimization.* Springer, 2007.

[88] A. Žilinskas. A review of statistical models for global optimization. *J. Global Optim.*, 2(2):145–153, 1992.