Multidisciplinary Design Optimization, application to aerospace vehicle design - MAP 554

Part 4: Bayesian Optimization

M. Balesdent, L. Brevault

Contacts : mathieu.balesdent@onera.fr, loic.brevault@onera.fr

Outline

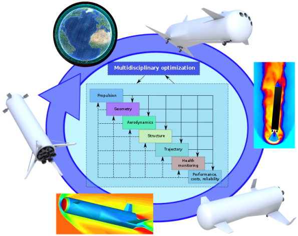
Context

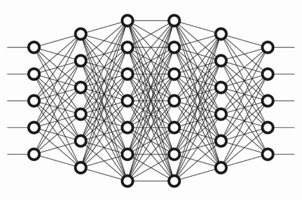
Optimization of problems characterized by:

Context

Optimization of problems characterized by:

- Black box and computationally expensive functions,





Context

Optimization of problems characterized by:

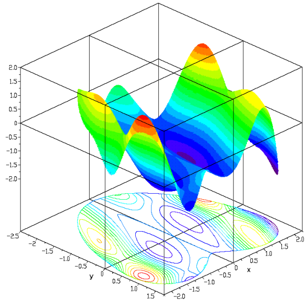
- Black box and computationally expensive functions,

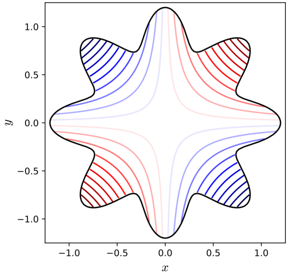
✗ Gradient based optimization approaches	✗ Classic evolutionary algorithms	✓ Bayesian Optimization

Context

Optimization of problems characterized by:

- Black box and computationally expensive functions,
- Constrained problems,

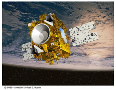




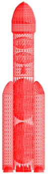
Context

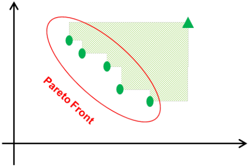
Optimization of problems characterized by:

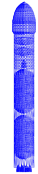
- Black box and computationally expensive functions,
- Multi-objective problems.



Maximize f_1: The payload value.





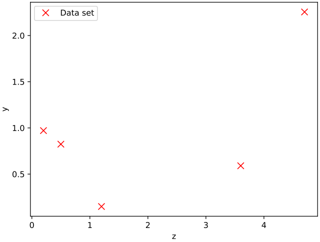


Minimize f_2: The gross lift off-weight value.

Intuitive solution

Single objective optimization problem with computationally intensive black-box

$$\begin{aligned} \min \quad & f(z) \\ \text{w.r.t.} \quad & z \\ & z_{min} \le z \le z_{max} \end{aligned}$$

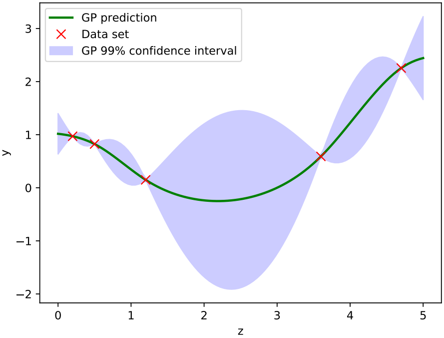


Where is the minimum of $f(\cdot)$?
Where should be the next evaluation ?

Intuitive solution II

Use of Bayesian Optimization (BO):

- the objective function is model with a Gaussian Process (GP),
- an infill criterion is used to decide where should be the next evaluation



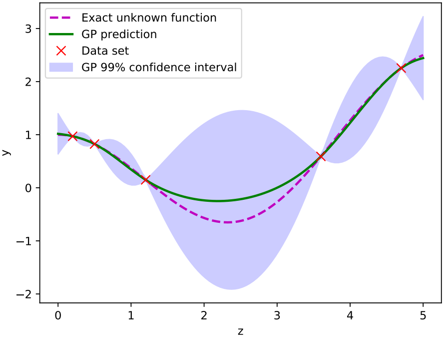
Where is the minimum of $f(\cdot)$?

Where should be the next evaluation ?

Intuitive solution II

Use of Bayesian Optimization (BO):

- the objective function is model with a Gaussian Process (GP),
- an infill criterion is used to decide where should be the next evaluation

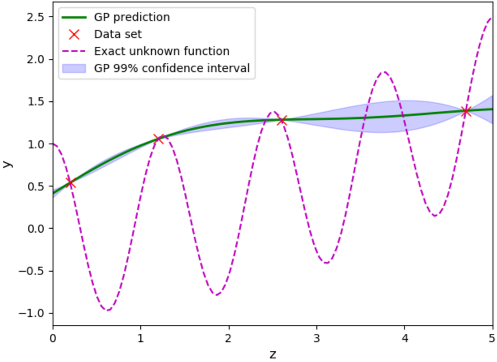


Where is the minimum of $f(\cdot)$?
Where should be the next evaluation ?

Intuitive solution II

Use of Bayesian Optimization (BO):

- the objective function is model with a Gaussian Process (GP),
- an infill criterion is used to decide where should be the next evaluation



Where is the minimum of $f(\cdot)$?

Where should be the next evaluation ?

Outline

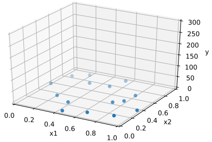
BO framework

Bayesian Optimization (BO) [Jones et al., 1998]

BO framework

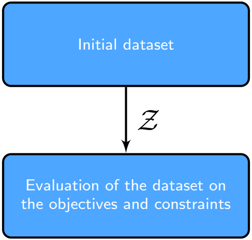
Bayesian Optimization (BO) [Jones et al., 1998]

Initial dataset

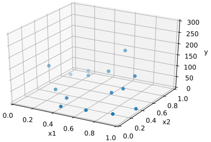


Design of Experiment depending on the dimension and the nature of the problem

BO framework



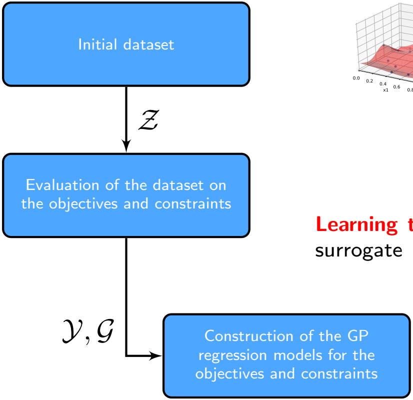
Bayesian Optimization (BO) [Jones et al., 1998]

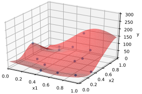


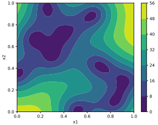
Evaluations of the **expensive** black-box functions

BO framework

Bayesian Optimization (BO) [Jones et al., 1998]



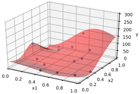


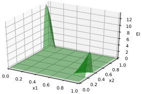


Learning the hyperparameters of each surrogate model

BO framework

Bayesian Optimization (BO) [Jones et al., 1998]





Initial dataset

$\mathcal{Z}$

Evaluation of the dataset on the objectives and constraints

$\mathcal{Y}, \mathcal{G}$

Construction of the GP regression models for the objectives and constraints

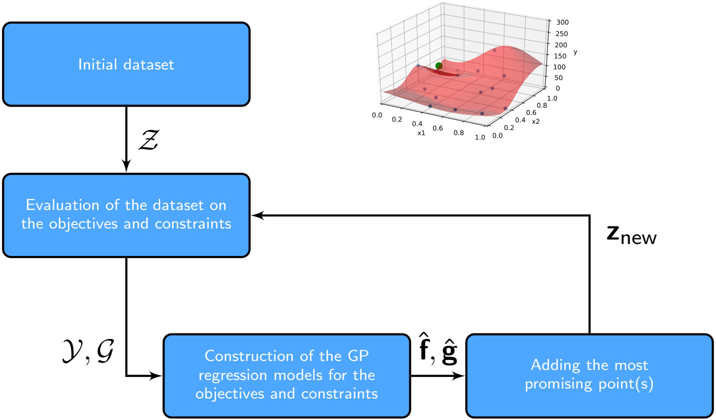
$\hat{\mathbf{f}}, \hat{\mathbf{g}}$

Adding the most promising point(s)

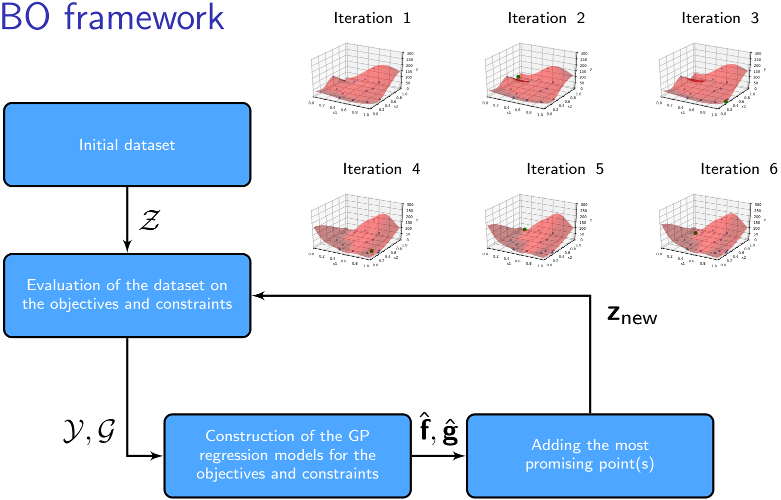
Optimization of a chosen **infill sampling criterion**

BO framework

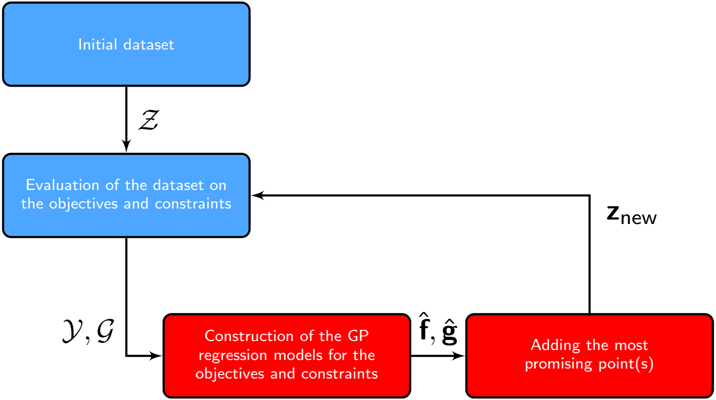
Bayesian Optimization (BO) [Jones et al., 1998]



BO framework



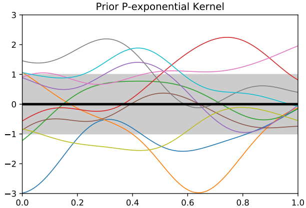
BO framework

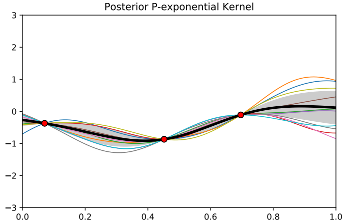


Gaussian Process Regression

Gaussian process [Rasmussen and Williams, 2006]

A Gaussian Process is used to describe a distribution over function. It is a collection of infinite random variables, **any finite number of which have a joint Gaussian distribution**.
It is defined by its mean function and covariance function (Kernel):
$F(\cdot) \sim \mathcal{GP}(\mu(\cdot), k^{\Theta}(\cdot, \cdot))$





Outline

Infill Criteria

This step of BO consists in performing a simultaneous search for the problem optimum and refinement of the GP modeling the objective function by determining the location in the design space at which the problem optimum is most likely to be found according to a given acquisition function.

The infill criteria should do the following operations:

- Exploration for more accurate learning $\rightarrow$ reduce the uncertainty in the model,
- Exploitation/optimization for better results $\rightarrow$ minimize the loss in a sequence $z_1, \ldots, z_n$.

A first naive criterion

Use of mean prediction of GP:

$$m(\mathbf{z}) = -\hat{f}(\mathbf{z})$$

- does not exploit the uncertainty prediction given by GP,
- exploitation only.

Let's see in the following three alternative criteria:

- Upper (lower) confidence band,
- Probability of improvement (PI),
- Expected improvement (EI).

Upper (lower) Confidence Band

Considering the objective function modeled by a GP such that:
$F(\cdot) \sim \mathcal{GP}(\mu(\cdot), k^{\Theta}(\cdot, \cdot))$.

$$\text{LCB}(\mathbf{z}) = -\hat{f}(\mathbf{z}) + \beta \hat{s}(\mathbf{z})$$

$\widehat{f}(\mathrm{z})$ the mean GP prediction and $\widehat{s}(\mathrm{z})$ the standard deviation of the objective function prediction and β a fixed parameter.

To identify the next data point that should be evaluated on the exact objective function and added to the current DoE, the following optimization problem is solved:

$$\max_{\mathrm{z}} \ \mathrm{LCB}(\mathrm{z}) \tag{1}$$

This problem is computationally non-expensive as it involves only the GP of the objective function.
Analysis of this infill criterion:

- allows to take into account the uncertainty in a static way.

Probability of Improvement (PI)

The improvement from the current samples $y_{\min} = \min\{y_{(1)}, \ldots, y_{(n)}\}$ is defined as follows [Forrester et al., 2008]:

$$I(\mathrm{z}) = \max(y_{\min} - F(\mathrm{z}), 0) = \begin{cases} y_{\min} - F(\mathrm{z}), & F(\mathrm{z}) < y_{\min} \\ 0, & \text{otherwise} \end{cases} .$$

The Probability of Improvement (PI) is defined as the probability that $F(\mathrm{z}) \leq y_{\min}$,

$$\mathrm{PI}(\mathrm{z}) = \mathbb{P}[F(\mathrm{z}) \leq y_{\min}] = \frac{1}{\widehat{s}(\mathrm{z})\sqrt{2\pi}} \int_{-\infty}^{y_{min}} \exp\left(\frac{-\left(t - \widehat{f}(\mathrm{z})\right)^2}{2\widehat{s}(\mathrm{z})^2} \right) \mathrm{d}t$$

$$\mathrm{PI}(\mathrm{z}) = \Phi\left(\frac{y_{\min} - \widehat{f}(\mathrm{z})}{\widehat{s}(\mathrm{z})} \right),$$

with $\Phi(\cdot)$ the Cumulative Density Function of the Normal distribution, $\widehat{f}(\mathrm{z})$ the mean GP prediction and $\widehat{s}(\mathrm{z})$ the standard deviation of the objective function prediction.

Probability of Improvement (PI)

The Probability of Improvement (PI) is defined as the probability that $F(\mathrm{z}) \leq y_{\min}$,

$$\mathrm{PI}(\mathrm{z}) \quad = \Phi\left(\frac{y_{\min} - \widehat{f}(\mathrm{z})}{\widehat{s}(\mathrm{z})}\right),$$

To identify the next data point that should be evaluated on the exact objective function and added to the current DoE, the following optimization problem is solved:

$$\max_{\mathrm{z}} \ \mathrm{PI}(\mathrm{z}) \tag{2}$$

This problem is computationally non-expensive as it involves only the GP of the objective function.

Analysis of this infill criterion:

- data points proposed by PI are often located near the current best point where the probability is high,
- not relevant criterion for global exploration.

Expected Improvement [Jones et al., 1998]

Improvement:

$$I(z) = \max(y_{\min} - F(z), 0.)$$

Mathematical expectation of Improvement is given by:

$$\mathrm{EI}(z) = \mathbb{E}[\max(y_{\min} - F(z), 0.)]$$
$$\mathrm{EI}(z) = \left(y_{\min} - \widehat{f}(z)\right) \Phi\left(\frac{y_{\min} - \widehat{f}(z)}{\widehat{s}(z)}\right) + \widehat{s}(z)\phi\left(\frac{y_{\min} - \widehat{f}(z)}{\widehat{s}(z)}\right)$$

Two parts in the sum:

- first: exploitation of zone in the search space in which there is potential improvement of the function (scaled by discrepancy between the prediction and the current minimum).
- second: exploration of the zones for which the uncertainty of the model is high ($\phi(\cdot)$ the PDF of the Normal distribution).

$\rightarrow$ **The expected improvement allows to balance the part relative to exploration and the part dedicated to exploitation in the refinement criterion. It is the most used infill criterion in the literature**.

Which criterion?

Let consider the following function:

$$f(z) = 0.4\sin(10z) + 0.5z^2 - 0.2z + 0.4z^3 + 0.4$$

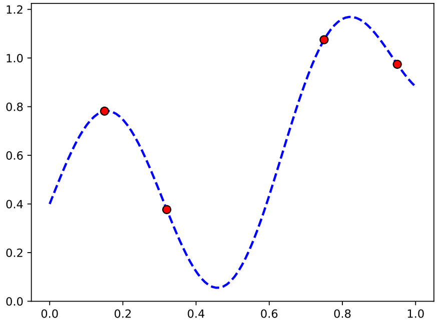
defined in $[0., 1.]$. Only 4 samples of this function are available at $z = 0.15$, $z = 0.32$, $z = 0.75$ and $z = 0.95$.

Which criterion?

Let consider the following function:

$$f(z) = 0.4\sin(10z) + 0.5z^2 - 0.2z + 0.4z^3 + 0.4$$

defined in $[0., 1.]$. Only 4 samples of this function are available at $z = 0.15$, $z = 0.32$, $z = 0.75$ and $z = 0.95$.

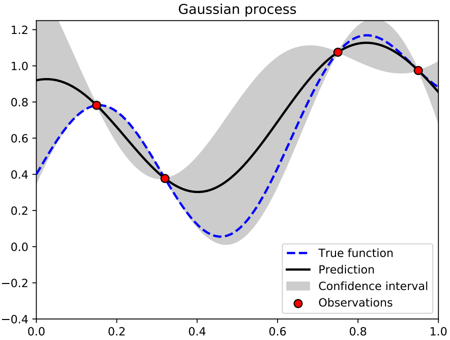


Which criterion?

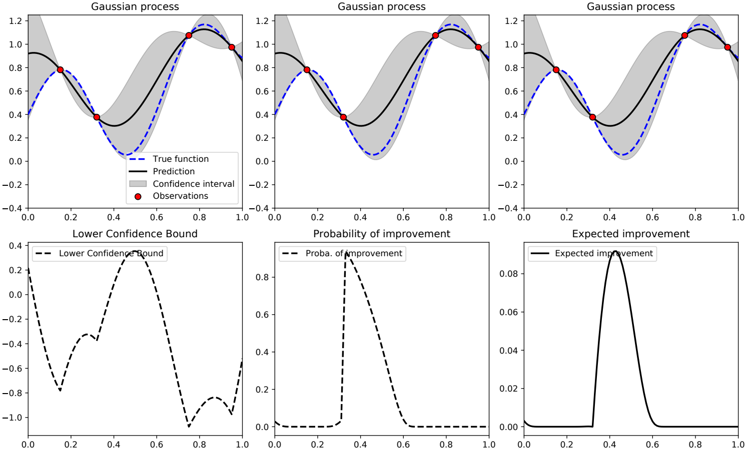
Let consider the following function:

$$f(z) = 0.4\sin(10z) + 0.5z^2 - 0.2z + 0.4z^3 + 0.4$$

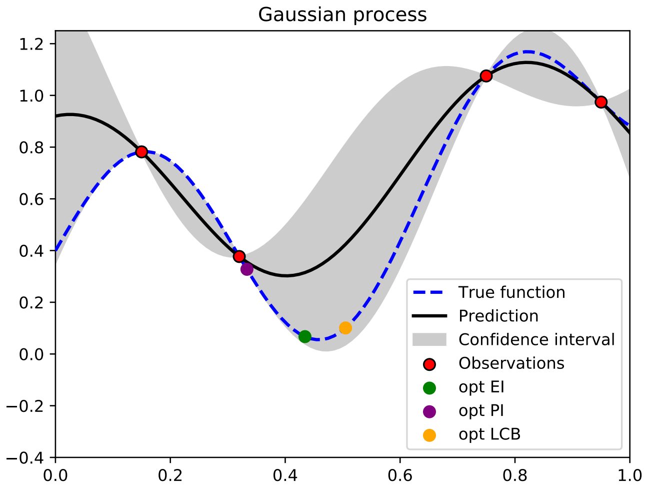
defined in $[0., 1.]$. Only 4 samples of this function are available at $z = 0.15$, $z = 0.32$, $z = 0.75$ and $z = 0.95$.



Which criterion?



Which criterion?

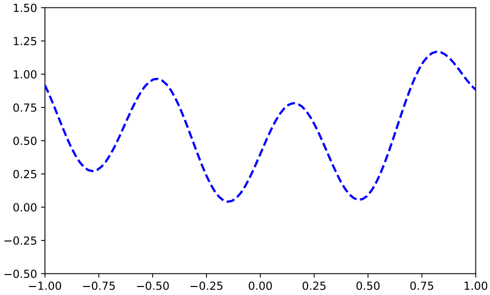


1-D example with EI

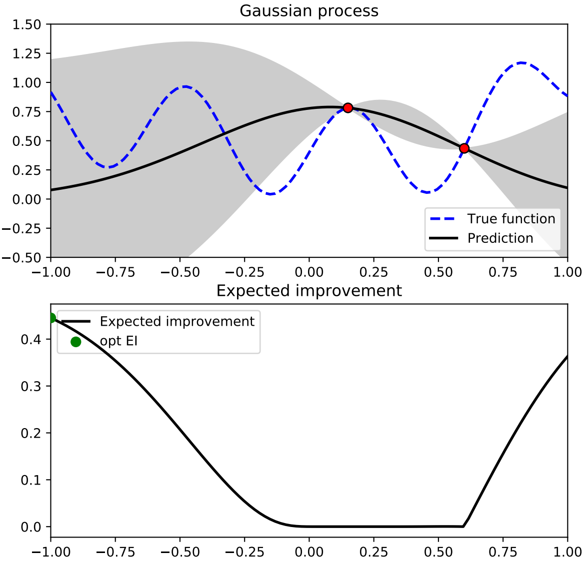
Let consider the following function:

$$f(z) = 0.4\sin(10z) + 0.5z^2 - 0.2z + 0.4z^3 + 0.4$$

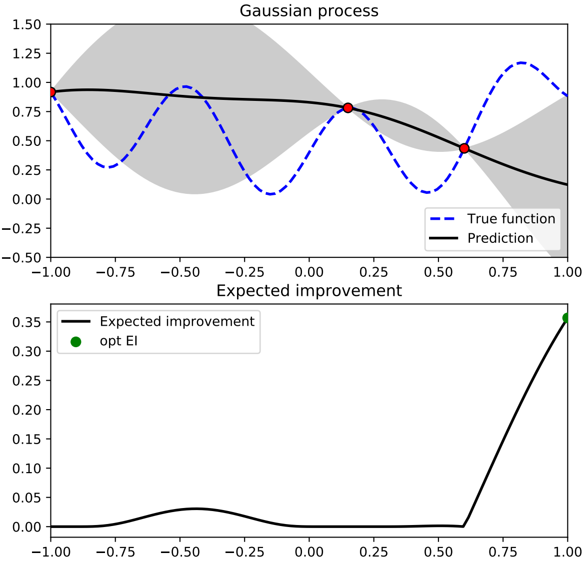
defined in $[-1., 1.]$. Only 2 samples of this function are available at $z = 0.15$ and $z = 0.6$.



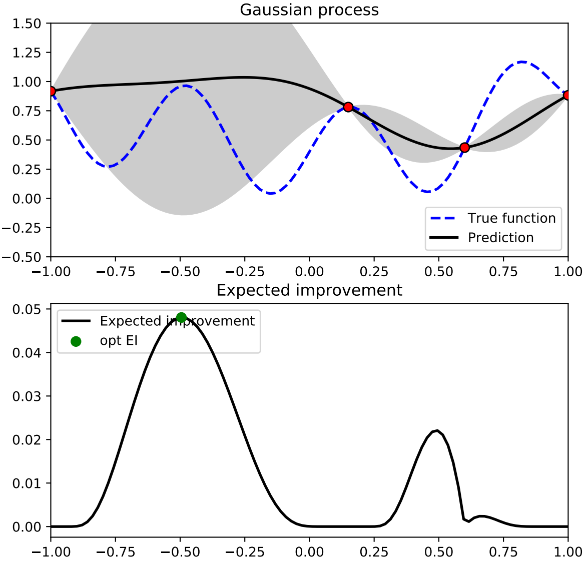
1-D example with EI



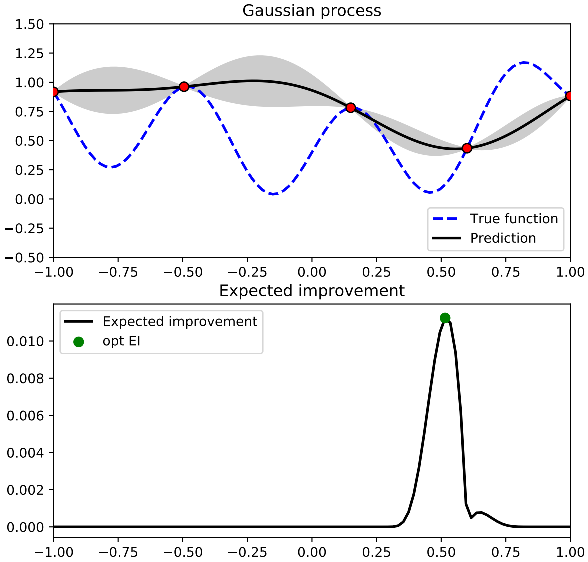
1-D example with EI



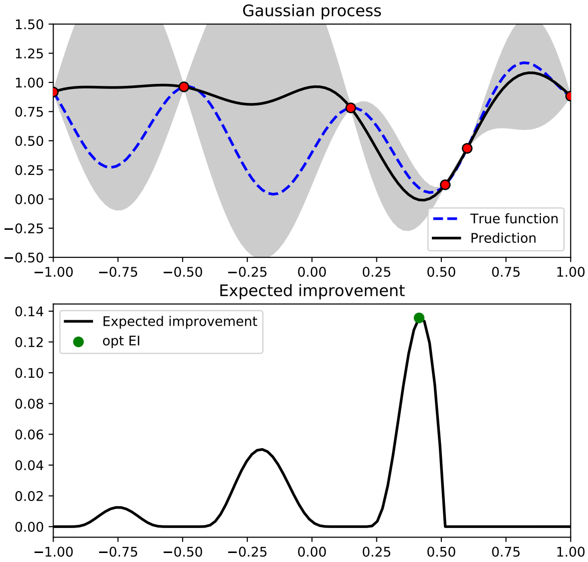
1-D example with EI



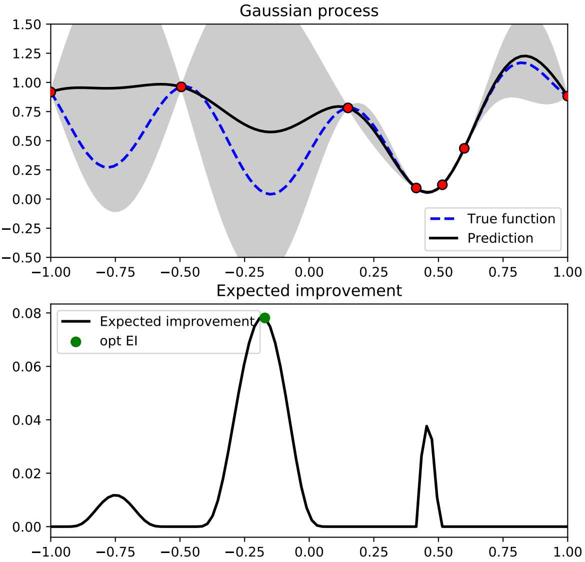
1-D example with EI



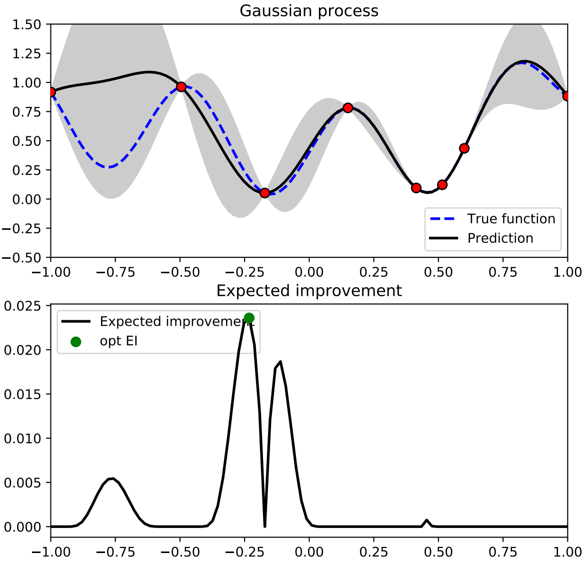
1-D example with EI



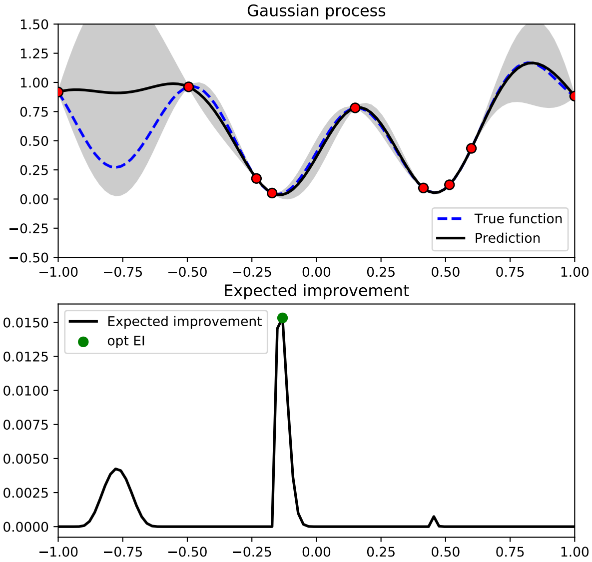
1-D example with EI



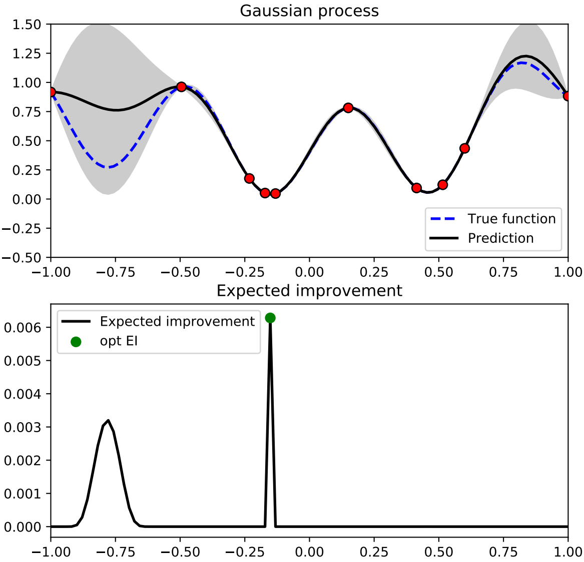
1-D example with EI



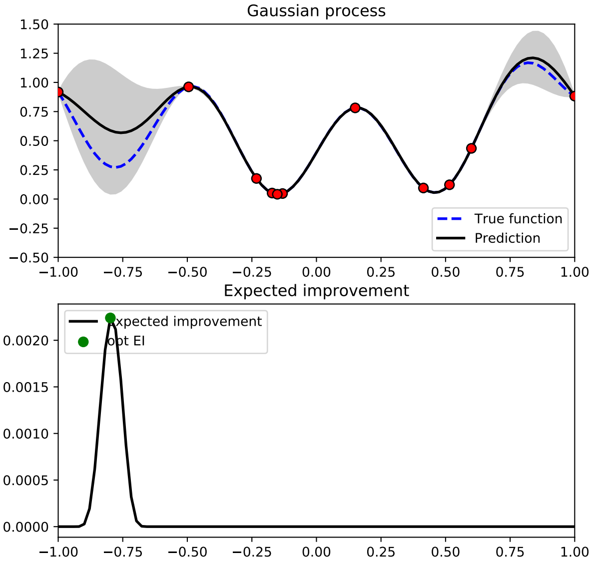
1-D example with EI



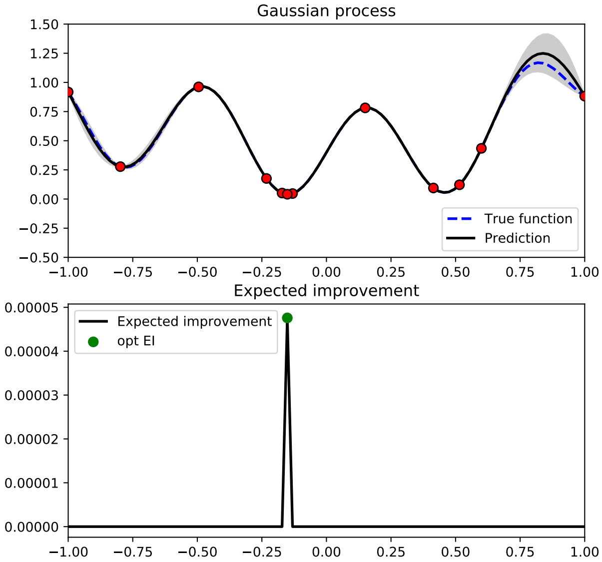
1-D example with EI



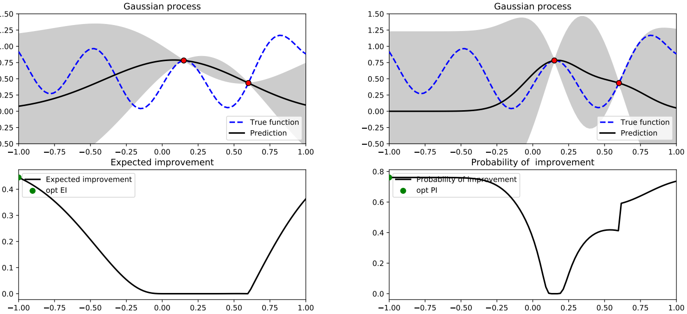
1-D example with EI



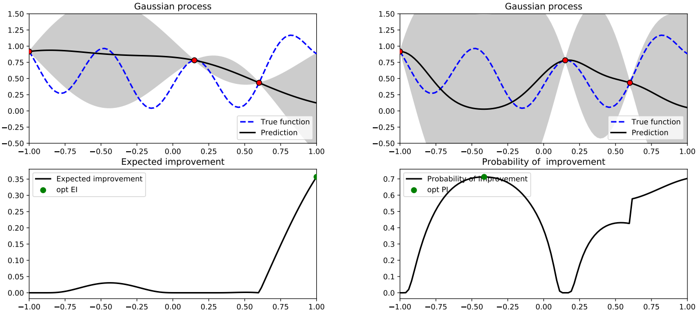
1-D example with EI



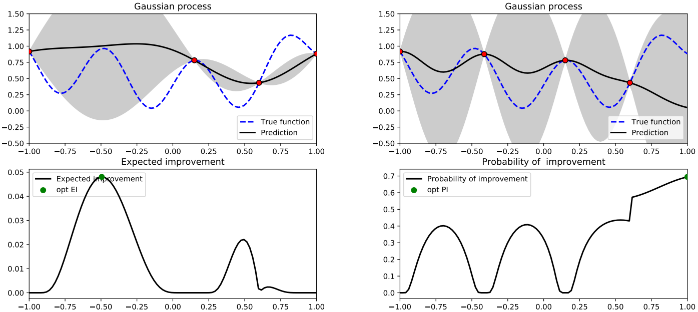
1-D example with EI versus PI



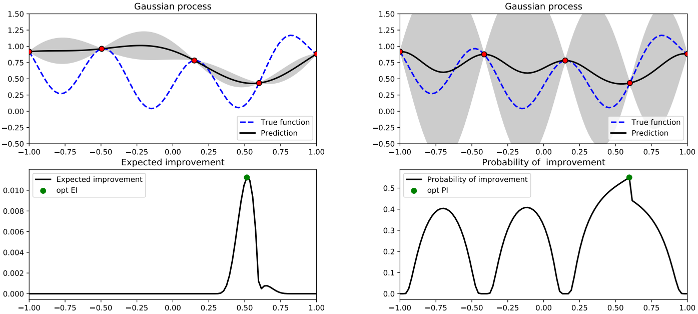
1-D example with EI versus PI



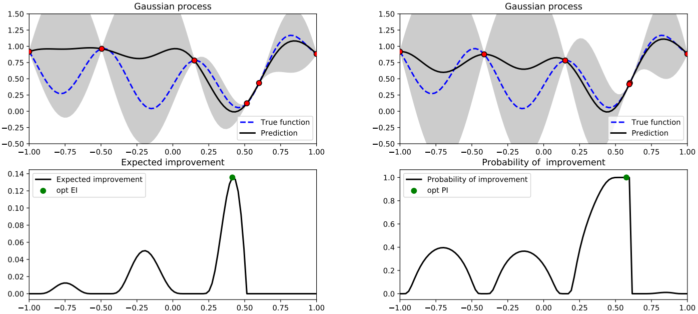
1-D example with EI versus PI



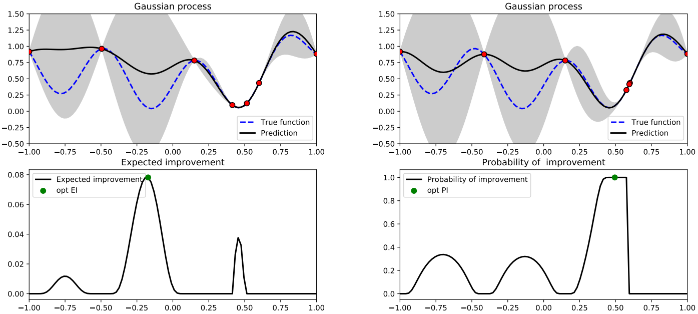
1-D example with EI versus PI



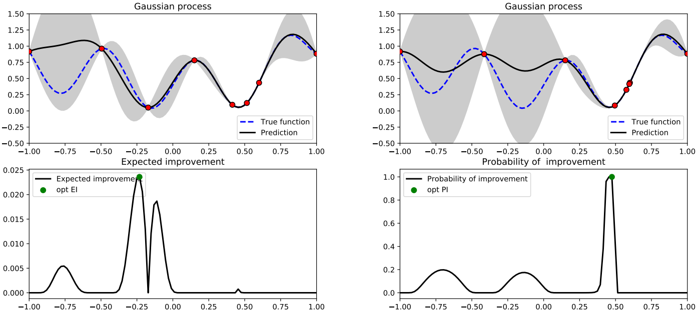
1-D example with EI versus PI



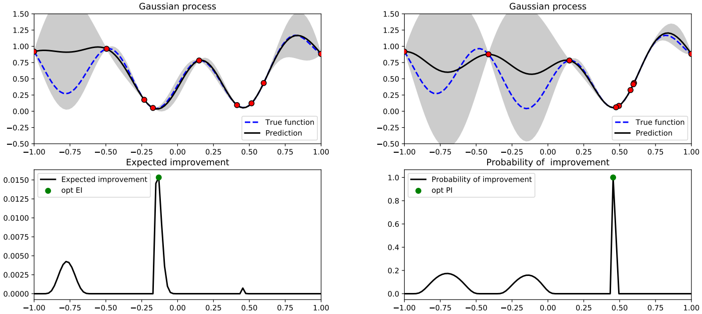
1-D example with EI versus PI



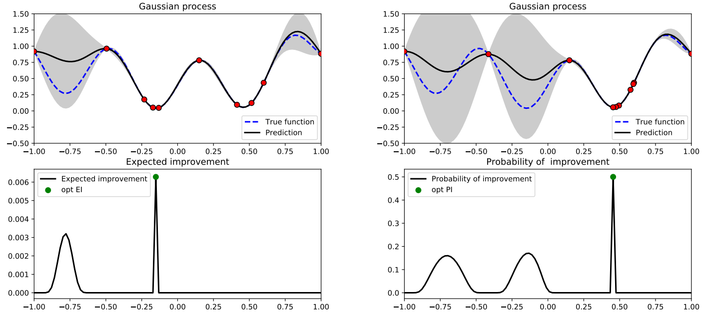
1-D example with EI versus PI



1-D example with EI versus PI



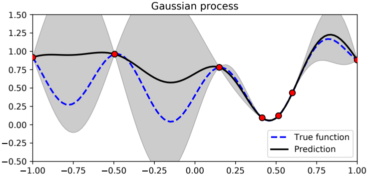
1-D example with EI versus PI

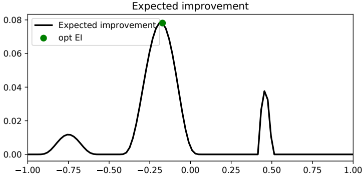


Optimization of infill criterion - bottlenecks

- In the same way as the likelihood in the training phase of GP, infill criteria often present a lot of local minima (especially in multi dimensional case)
- Infill criteria present large areas with null value.

$\rightarrow$ Need to use either multi-start gradient based optimization or global exploration algorithms (*e.g.* CMA-ES, Genetic algorithms, see lecture 1).





Outline

Constrained case

$$
\begin{aligned}
\min \quad & f(\mathrm{z}) \\
\text{w.r.t.} \quad & \mathrm{z} \\
\text{s.t.} \quad & g_i(\mathrm{z}) \leq 0 \;\; i = 1, \dots, n_c \\
& \mathrm{z}_{\min} \leq \mathrm{z} \leq \mathrm{z}_{\max}
\end{aligned}
$$

How to take into account the constraint functions in addition to the objective function in a Bayesian Optimization framework ?

General ideas:

- Create a GP surrogate model for each constraint function in addition to the GP of the objective function,
- Incorporate the information of the GPs of the constraints into the infill criteria

In the following, two infill criterion adapted to the presence of constraint are introduced.

Direct approach - constrained EI

$$
\begin{aligned}
\min \quad & f(\mathrm{z}) \\
\text{w.r.t.} \quad & \mathrm{z} \\
\text{s.t.} \quad & \hat{g}_i(\mathrm{z}) \leq 0 \;\; i = 1, \ldots, n_c \\
& \mathrm{z}_{\min} \leq \mathrm{z} \leq \mathrm{z}_{\max}
\end{aligned}
$$

where $\hat{g}_i(\cdot)$ is the mean approximation of constraint i.

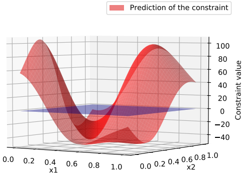
It is a naive approach as the information on the uncertainty associated to the constraint is not taken into account.

Probability of feasibility

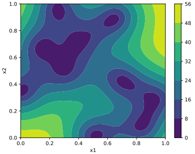
The probability of feasibility is defined as the probability that all the constraints are satisfied.

$P_f(\mathbf{z}) = \prod_{i=1}^{n_c} \mathbb{P}(G_i(\mathbf{z}) \leq 0)$
with $G_i \sim \mathcal{N}(\hat{g}_i, \hat{s}_i)$

$$\mathbb{P}(G_i(\mathbf{z}) \leq 0) = \Phi\left(-\frac{\hat{g}_i(\mathbf{z})}{\hat{s}_i(\mathbf{z})}\right)$$



Prediction



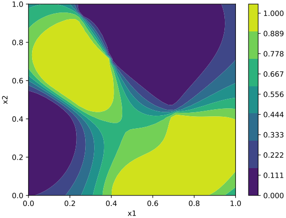
Variance of the prediction

Probability of feasibility

The probability of feasibility is defined as the probability that all the constraints are satisfied.

$P_f(\mathbf{z}) = \prod_{i=1}^{n_c} \mathbb{P}(G_i(\mathbf{z}) \leq 0)$
with $G_i \sim \mathcal{N}(\hat{g}_i, \hat{s}_i)$

$$\mathbb{P}(G_i(\mathbf{z}) \leq 0) = \Phi\left(-\frac{\hat{g}_i(\mathbf{z})}{\hat{s}_i(\mathbf{z})}\right)$$



Probability of feasibility

Probability of feasibility

In order to take into account the probability of feasibility into the infill criteria, the following infill optimization problem is solved (using the EI criteria for the objective function) :

$$
\begin{array}{ll}
\max & EI(\mathrm{z}) \times \prod_{i=1}^{n_c} \mathbb{P}(G_i(\mathbf{z}) < 0) \\
\text{w.r.t.} & \mathrm{z} \\
 & \mathrm{z}_{\min} \leq \mathrm{z} \leq \mathrm{z}_{\max}
\end{array}
$$

Probability of feasibility

In order to take into account the probability of feasibility into the infill criteria, the following infill optimization problem is solved (using the EI criteria for the objective function) :

$$
\begin{aligned}
\max \quad & EI(\mathrm{z}) \times \prod_{i=1}^{n_c} \mathbb{P}(G_i(\mathbf{z}) < 0) \\
\text{w.r.t.} \quad & \mathrm{z} \\
& \mathrm{z}_{\min} \leq \mathrm{z} \leq \mathrm{z}_{\max}
\end{aligned}
$$

- the infill criterion is close to zero in the areas where there is a very low probability of feasibility,
- if the number of constraints increases, the probability of feasibility is null in the entire space that makes difficult the optimization problem solving,
- uses of non constrained optimization algorithms.

Expected violation

- The expected violation expresses the mathematical expectation of the constraint violation,
- Similar to EI adapted to the constraint.

$$\begin{aligned}
\mathrm{EV}_i(\mathrm{z}) &= \mathbb{E}\left[\max(G_i(\mathrm{z}) - 0, 0)\right] \\
&= (\widehat{g}_i(\mathrm{z}) - 0)\Phi\left(\frac{\widehat{g}_i(\mathrm{z}) - 0}{\widehat{s}_{g_i}(\mathrm{z})}\right) + \widehat{s}_{g_i}(\mathrm{z})\phi\left(\frac{\widehat{g}_i(\mathrm{z}) - 0}{\widehat{s}_{g_i}(\mathrm{z})}\right)
\end{aligned}$$

Infill optimization problem with expected violation

Constrained infill optimization problem :

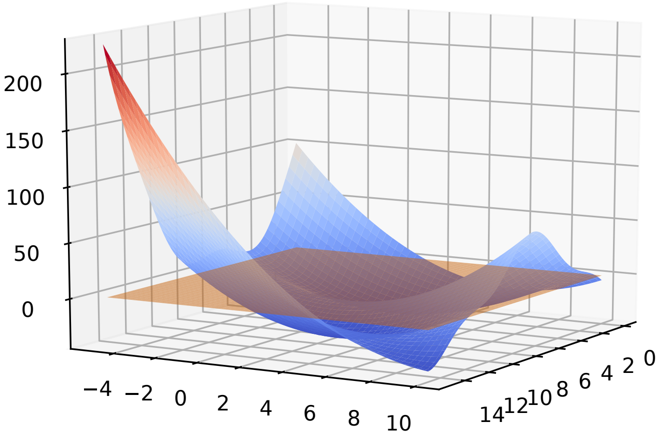
$$
\begin{aligned}
\max \quad & EI(\mathrm{z}) \\
\text{w.r.t.} \quad & \mathrm{z} \\
\text{s.t.} \quad & EV_i(\mathrm{z}) \leq t_i \quad \text{for} \quad i = 1, \ldots, n_c \\
& \mathrm{z}_{\min} \leq \mathrm{z} \leq \mathrm{z}_{\max}
\end{aligned}
$$

- Multi modal + constrained optimization problem
  $\rightarrow$ Need for dedicated optimization algorithms.

Handling of constraints: practical case

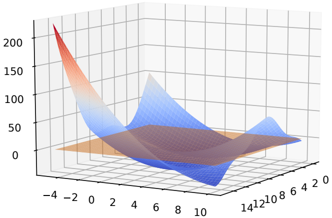
Let consider the modified Branin function in 2D defined as follows:

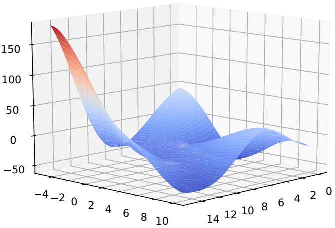
$$g(z_1, z_2) = \left(z_2 - \frac{5.1z_1^2}{4\pi^2} + \frac{5z_1}{\pi} - 6\right)^2 + 10\left(1 - \frac{1}{8\pi}\right)\cos(z_1) + \frac{1}{8\pi} - 30.$$



Handling of constraints: practical case

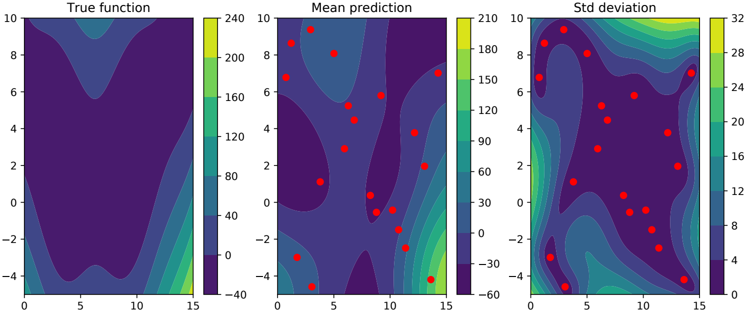
Let consider a GP model of the Branin function from of DoE of 20 points using LHS (see lecture 3).





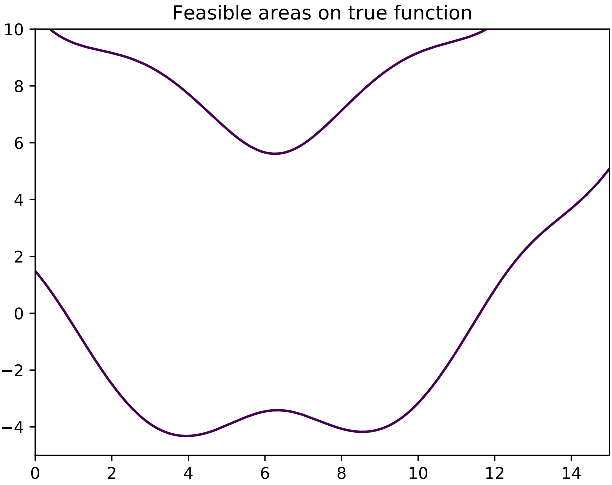
Handling of constraints: practical case

Let consider a GP model of the Branin function from of DoE of 20 points using LHS (see lecture 3).



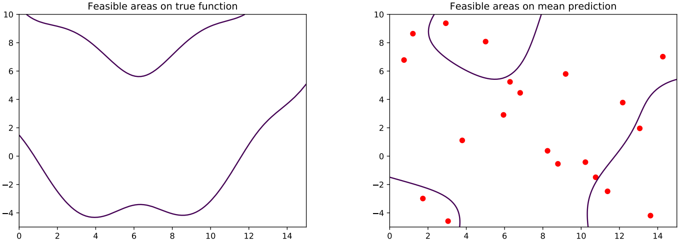
Handling of constraints: practical case

Limit state of the exact constraint function $(z \in \mathbb{R}^2 | g(z) = 0)$.



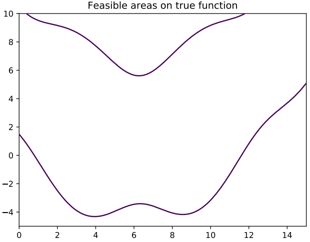
Handling of constraints: practical case

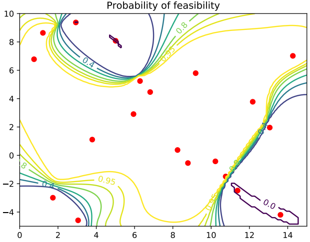
Limit state of the exact constraint function ($z \in \mathbb{R}^2 | g(z) = 0$) and its prediction using GP ($z \in \mathbb{R}^2 | \hat{g}(z) = 0$).



Handling of constraints: practical case

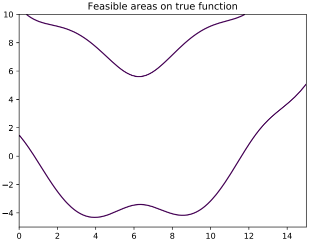
Limit state of the exact constraint function ($z \in \mathbb{R}^2 | g(z) = 0$) vs Probability of Feasibility.

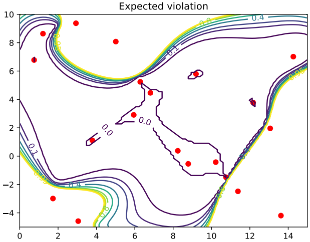




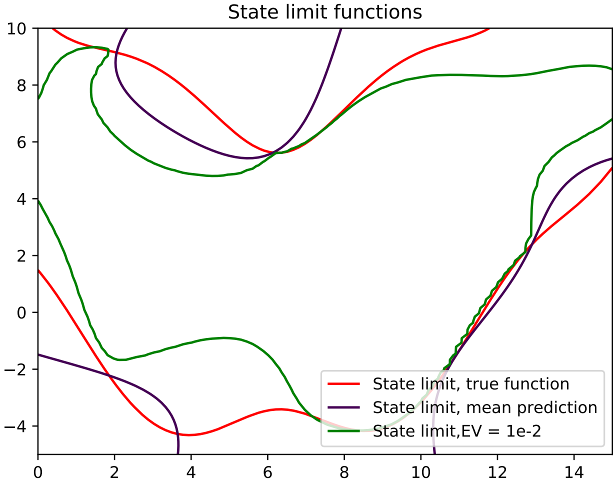
Handling of constraints: practical case

State limit on true function ($z \in \mathbb{R}^2 | g(z) = 0$) vs Expected violation (with $t = 0.01$).





Handling of constraints: practical case



Outline

Multi-objective case

$$
\begin{aligned}
\text{min} \quad & \mathrm{f}(\mathrm{z}) = [f_1(\mathbf{z}), \ldots, f_n(\mathbf{z})] \\
\text{w.r.t.} \quad & \mathrm{z} \\
\text{s.t.} \quad & g_i(\mathrm{z}) \leq 0,\ \ i = 1, \ldots, n_c \\
& \mathrm{z}_{\mathrm{min}} \leq \mathrm{z} \leq \mathrm{z}_{\mathrm{max}}
\end{aligned}
$$

Multi-objective case

$$
\begin{aligned}
\text{min} \quad & \mathrm{f}(\mathrm{z}) = [f_1(\mathbf{z}), \ldots, f_n(\mathbf{z})] \\
\text{w.r.t.} \quad & \mathrm{z} \\
\text{s.t.} \quad & g_i(\mathrm{z}) \leq 0, \;\; i = 1, \ldots, n_c \\
& \mathrm{z}_{\min} \leq \mathrm{z} \leq \mathrm{z}_{\max}
\end{aligned}
$$

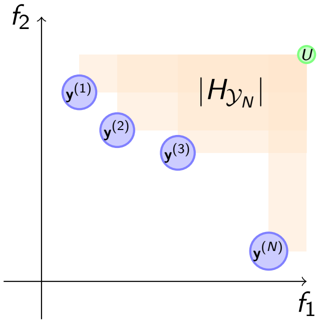
$\rightarrow$ How to build an infill criterion that allows to find design points that are relevant with respect to the Pareto front ?

Multi-objective infill criterion

Hypervolume indicator

The hypervolume indicator expresses the hypervolume of the objective space dominated by the approximated Pareto set.

$$H_{\mathcal{Y}_N} = \left\{ \mathbf{y} \in \mathbb{B}; \exists i \in \{1, \ldots, N\}, \mathbf{y}^{(i)} \prec \mathbf{y} \right\}$$



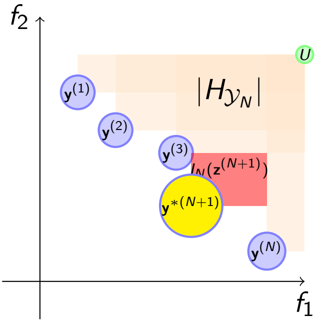
With $\mathbb{B} = \{y \in \mathbb{R}^n; \mathbf{y}^L \leq \mathbf{y} \leq \mathbf{y}^U\}$ where $\mathbf{y}^L$ is the ideal point ($\mathbf{y}^L = [\min f_1(\mathbf{z}), ..., \min f_n(\mathbf{z})]$) and $\mathbf{y}^U$ a chosen upper bound (Nadir point)

Multi-objective infill criterion

Hypervolume improvement [Bradstreet et al., 2006]

The hypervolume improvement is the improvement of the hypervolume by adding a candidate to the data set

$$I_N(\mathbf{z}^{(N+1)}) = |H_{\mathcal{Y}_{N+1}}| - |H_{\mathcal{Y}_N}|$$



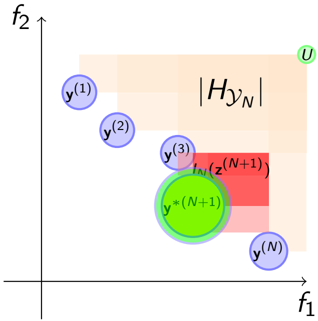
Multi-objective infill criterion

Expected Hypervolume improvement
[Emmerich et al., 2016]

The expected hypervolume improvement is the mathematical expected improvement of the hypervolume by adding a candidate to the sample

$$EHVI_N(\mathbf{z}) = \mathbb{E}(|H_{\mathcal{Y}_{N+1}}| - |H_{\mathcal{Y}_N}|)$$
$$= \int_{\mathbb{B}\setminus H_{\mathcal{Y}_N}} \mathbb{P}(\mathbf{Y}^{*(N+1)} \prec p)\mathrm{d}p$$

with $\mathbf{Y}^{*(N+1)} = \left[Y_1^{*(N+1)}, Y_2^{*(N+1)}\right]$
and $Y_1^{*(N+1)} \sim \mathcal{N}\left(\hat{y}_1^{*(N+1)}, \hat{s}_1^{*(N+1)}\right)$ and $Y_2^{*(N+1)} \sim \mathcal{N}\left(\hat{y}_2^{*(N+1)}, \hat{s}_2^{*(N+1)}\right)$



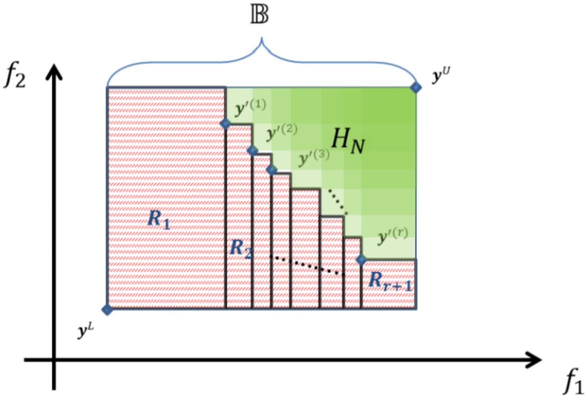
EHVI

How to compute the EHVI in the two objective case (at iteration N):

- Partition the $\mathbb{B} \setminus H_{\mathcal{Y}_N}$ into r rectangles,
- Use this decomposition to facilitate the integration over the points that can be dominated by $Y(\mathrm{z})$.

$$
\begin{aligned}
EHVI(\mathbf{z}) &= \int_{P \in \mathbb{B} \setminus H_N} \mathbb{P}(Y(\mathbf{z}) \prec p) dp \\
&= \int \int_{p=(p_1,p_2) \in \mathbb{B} \setminus H_N} \mathbb{P}(Y_1(\mathbf{z}) \prec p_1) \mathbb{P}(Y_2(\mathbf{z}) \prec p_2) dp_1 dp_2 \\
&= \sum_{t=0}^{r+1} \int_{{y'}_1^{(t-1)}}^{{y'}_1^{(t)}} \mathbb{P}(Y_1(\mathbf{z}) \prec p_1) \int_{{y'}_2^{(0)}}^{{y'}_2^{(t-1)}} \mathbb{P}(Y_2(\mathbf{z}) \prec p_2) dp_1 dp_2 \\
&= \sum_{t=0}^{r+1} \int_{{y'}_1^{(t-1)}}^{{y'}_1^{(t)}} \Phi\left(\frac{p_1 - \hat{y}_1(\mathbf{z})}{\hat{\sigma}_1(\mathbf{z})}\right) \int_{{y'}_2^{(0)}}^{{y'}_2^{(t-1)}} \Phi\left(\frac{p_2 - \hat{y}_2(\mathbf{z})}{\hat{\sigma}_2(\mathbf{z})}\right) dp_1 dp_2
\end{aligned}
$$

EHVI



- Computation of EHVI in high dimension (> 2) is much more difficult.

Constrained multi-objective computation

Expected hypervolume improvement can be associated respectively with either Expected Violation or Probability of Feasibility for constrained optimization problem.

EHVI with EV

$$\begin{array}{rl} \max & EHVI(\mathrm{z}) \\ \text{w.r.t.} & \mathrm{z} \\ \text{s.t.} & EV_i(\mathrm{z}) \leq t_i \quad \text{for} \quad i = 1, \ldots, n_c \\ & \mathrm{z}_{\min} \leq \mathrm{z} \leq \mathrm{z}_{\max} \end{array}$$

EHVI with PoF

$$\begin{array}{rl} \max & EHVI(\mathrm{z}) \times \prod_{i=1}^{n_c} \mathbb{P}(G_i(\mathbf{z}) < 0) \\ \text{w.r.t.} & \mathrm{z} \\ \text{s.t.} & \mathrm{z}_{\min} \leq \mathrm{z} \leq \mathrm{z}_{\max} \end{array}$$

Outline

Conclusions

Bayesian Optimization allows to optimize computationally intensive problems using GP models for the objective and constraint functions.

- Two key elements: the surrogate model and the acquisition function (infill criterion).
- The key is to find a good balance between exploration and exploitation of the minimum in the search.
- The curse of dimension is very important in Bayesian Optimization. In more than 10 dimensions problems, alternative approaches to GP (for instance combining GP and dimension reduction techniques) are more suited.

Toolboxes

A lot of toolboxes implement Bayesian Optimization methods in python:

- Surrogate Toolbox Modeling (SMT) https://smt.readthedocs.io/en/latest/
- scikit-learn https://scikit-learn.org/stable/
- openTURNS http://www.openturns.org/
- GPflowOpt https://github.com/GPflow/GPflowOpt
- BOTorch https://botorch.org/
- GPyOpt https://sheffieldml.github.io/GPyOpt//

References I

Bradstreet, L., Barone, L., and While, L. (2006).
Maximising hypervolume for selection in multi-objective evolutionary algorithms.
In *2006 IEEE International Conference on Evolutionary Computation*, pages 1744–1751. IEEE.

Emmerich, M., Yang, K., Deutz, A., Wang, H., and Fonseca, C. M. (2016).
A multicriteria generalization of bayesian global optimization.
In *Advances in Stochastic and Deterministic Global Optimization*, pages 229–242. Springer.

Forrester, A., Sobester, A., and Keane, A. (2008).
Engineering design via surrogate modelling: a practical guide.
John Wiley & Sons.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998).
Efficient global optimization of expensive black-box functions.
Journal of Global optimization, 13(4):455–492.

Rasmussen, C. E. and Williams, C. K. a. (2006).
Gaussian processes for machine learning, volume 2.
MIT press Cambridge, MA.