



**HAL**  
open science

# Optimization under uncertainties: an overview with a focus on Gaussian processes

Rodolphe Le Riche

► **To cite this version:**

Rodolphe Le Riche. Optimization under uncertainties: an overview with a focus on Gaussian processes. Doctoral. Porquerolles, France. 2019. cel-02285533

**HAL Id: cel-02285533**

**<https://hal.science/cel-02285533v1>**

Submitted on 12 Sep 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Optimization under uncertainties: an overview with a focus on Gaussian processes

Rodolphe Le Riche

CNRS LIMOS at Mines Saint-Etienne, France

September 2019

Modeling and Numerical Methods for Uncertainty Quantification  
(MNMUQ 2019)

French-German Summer School / Ecole Thématique CNRS  
Porquerolles

## Optimization under uncertainties: an overview with a focus on Gaussian processes

- 1 Problem formulation
  - Motivation and the two classes of variables
  - Formulations as an optimization of statistical measures
- 2 Methods
  - A pointer to methods without metamodels
    - Approximation methods
    - Approaches based on stochastic gradients
  - Methods with metamodels
    - Overall scheme
    - Choice of  $x$
    - Choice of  $u$  knowing  $x$
- 3 Bibliography

# Goals of the class

- Optimization under uncertainty is a huge field: 3480 article have “optimization” and “uncertainty” in their title (google scholar).
- We focus on **uncertainties within an optimization problem**.  
There are two aspects:
  - ① How to formulate the problem.
  - ② How to solve it ← we will mainly cover approaches involving Gaussian processes (as they are the most integrated), and mention connexions.
- We will not cover statistical estimation (in particular simulation methods for reliability), generic adaptive design of experiments for kriging (in particular for constraints) beyond the strict minimum needed.

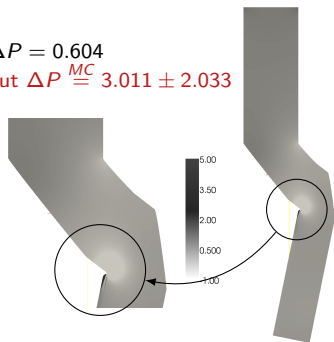
# A motivating example: robust air duct design

Minimize the pressure loss by changing the parameterized shape [Janusevskis and Le Riche, 2011]. Uncertain bottom position.

deterministic design

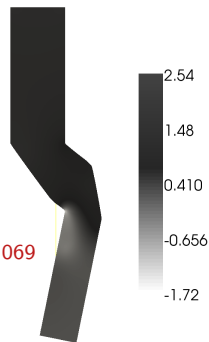
$$\Delta P = 0.604$$

$$\text{but } \Delta P^{MC} = 3.011 \pm 2.033$$



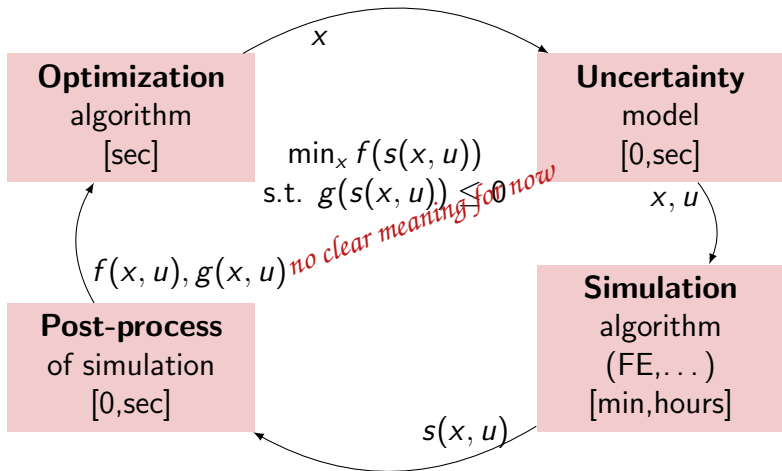
robust design

$$\Delta P^{MC} = 1.198 \pm 0.069$$



The deterministic optimization exploits mesh flaws. The robust optimization (mean of  $\Delta P$ ) has two advantages: the final solution accounts for uncertainties in production; the numerical model flaws are avoided.

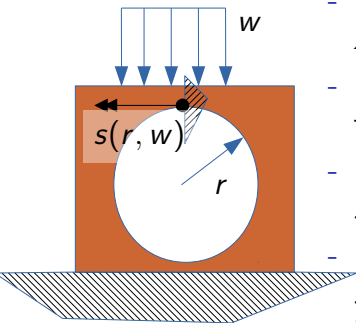
Besides this specific example, uncertainties affect simulation models (lack of knowledge, epistemic) and their working conditions (aleatory) and must be accounted for when the simulation is used within an optimization problem.



# The double $(x, U)$ parameterization

- $x \in \mathcal{X}$ , vector of deterministic, controlled variables over which the optimization is carried out.
- $U : \Omega \rightarrow \mathcal{A}$ , vector of random variables of pdf  $p_U(\cdot)$ .  
 $\Rightarrow s(x, U)$ ,  $f(x, U)$ ,  $g(x, U)$  are dependent random variables.
- This double parameterization, underlying Taguchi's methods in the 80's, is general. It is also called "augmented space" or "hybrid space". Cf. [Beyer and Sendhoff, 2007], [Pujol et al., 2009].

- $x \in \mathcal{X}$ , vector of deterministic, controlled variables over which the optimization is carried out.  $U : \Omega \rightarrow \mathcal{A}$ , vector of random variables of pdf  $p_U(\cdot)$ .
- Expl with a structure,  $s(\cdot, \cdot)$  is the stress and/or displacement:



- Noise homogeneous to  $x$ :  $r = x + U$ , both  $x$  and  $U$  are length,  $s(x + U, w)$ .
- Noise exogenous to  $x$ :  $U$  perturbation to the load  $w$ ,  $s(r, w + U)$ .
- Noise as an error model for the simulation:  $s(r, w) + U$ .
- Noise controlled by  $x$ :  $x$  tolerance class,  $s(r + xU, w)$  (any parameter of  $p_U(\cdot)$  can be  $x$ 's).



## Optimization under uncertainties: an overview with a focus on Gaussian processes

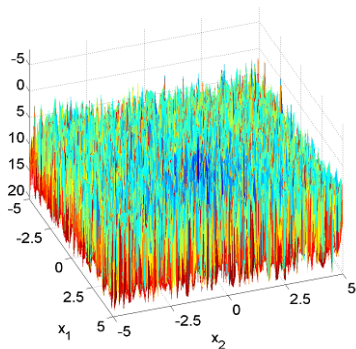
- 1 Problem formulation
  - Motivation and the two classes of variables
  - Formulations as an optimization of statistical measures
- 2 Methods
  - A pointer to methods without metamodels
    - Approximation methods
    - Approaches based on stochastic gradients
  - Methods with metamodels
    - Overall scheme
    - Choice of  $x$
    - Choice of  $u$  knowing  $x$
- 3 Bibliography

# Problem formulation: the noisy case

Let's not do anything about the uncertainties i.e., solve

$$\min_x f(x, U)$$

such that  $g(x, U) \leq 0$



It does not look good : gradients are not defined, what is the result of the optimization ?

But sometimes (often!?) one still wants to do it:  $f()$  and  $g()$  are inexpensive, large base of  $u^i$ 's available ← application scope of stochastic gradient methods.

# Problem formulation: statistical measures

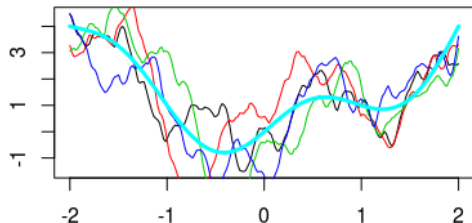
Remove the uncertainty (mathematically) by formulating the problem with statistical risk measures, either as a constrained optimization problem,

$$\begin{aligned} \min_x \rho^f(x) \quad , \quad \rho^f(x) &= \rho^f(f(x, U)) \\ \text{such that } \rho^g(x) &\leq 0 \quad , \quad \rho^g(x) = \rho^g(g(x, U)) \end{aligned}$$

or through an aggregation of statistical measures (remember  $f(\cdot, U)$  and  $g(\cdot, U)$  are dependent),

$$\min_x \rho(x) \quad , \quad \rho(x) = \rho(f(x, U), g(x, U))$$

or a combination thereof,  $\min_x \rho(x)$  s.t.  $\rho^g(x) \leq 0$ .



$f(x, u^i)$  in black, blue, red, green.

$\rho^f(x) = \mathbb{E}f(x, U)$  in thick light blue does not explicitly depend on the  $u$ 's (marginalized, implicit dependence through  $\rho_U(\cdot)$ )

# Risk measures for the objective function (1/2)

The “robust” formulations, single criteria:

- $\rho^f$   $\equiv$  expectation ( $\mathbb{E}$ ) as average performance.
- $\equiv$  variance ( $\mathbb{V}$ ) as performance dispersion.
- $\equiv$  quantile ( $Q_\alpha$ ) as guaranteed performance since  $\alpha\%$  of the realizations will be better.  $Q_{50}$ , the median, as representative performance.  $Q_{100}$  is the worst-case formulation.
- $\equiv$  a quantile difference ( $Q_\alpha - Q_{1-\alpha}$ ) as a performance dispersion.
- $\equiv$  super-quantile as guaranteed performance with an account for extreme values.

## Risk measures for the objective function (2/2)

The “robust” formulations, multiple criteria:

- $\rho^f$   $\equiv$  a multi-objective formulation accounting for an average performance and a deviation measure [Park et al., 2006], with various subsequent resolution approaches (goal programming, ordering, full Pareto, aggregations) and in particular
- $\equiv$  a linear combination of average and dispersion measures, typically  $\mathbb{E}f(x, U) + \alpha\sqrt{\mathbb{V}f(x, U)}$ : minimize average cost penalized by dispersion.  
For  $f(x, U) \sim \mathcal{N}$ , it is equivalent to the quantile (e.g.  $\alpha = 1.645$  for  $\mathbb{Q}_{90}$ ).

# Risk measures for the constraints

The “reliable” formulations:

$\rho^g \equiv \alpha - \mathbb{P}(g(x, U) \leq 0) \leq 0$ , as at least  $\alpha\%$  chances of satisfying the constraints.

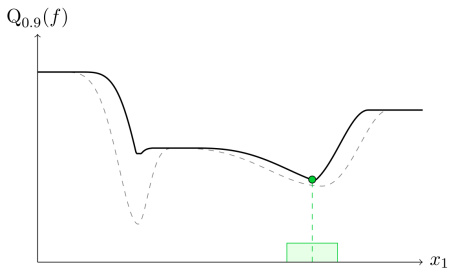
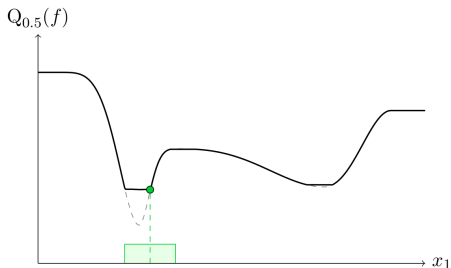
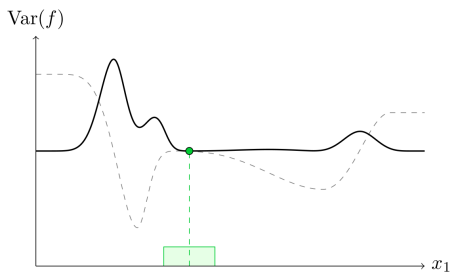
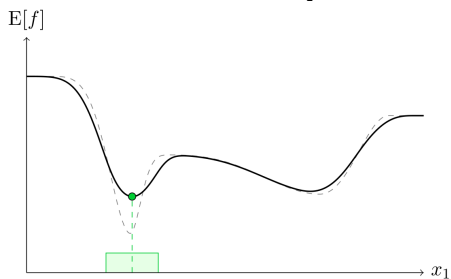
$$\begin{aligned}\mathbb{P}(g(x, U) \leq 0) &= \int_{g(x,u) \leq 0} p_U(u) du = \int_{\mathcal{A}} \mathbb{1}_{g(x,u) \leq 0} p_U(u) du \\ &= \mathbb{E}(\mathbb{1}_{g(x,u) \leq 0})\end{aligned}$$

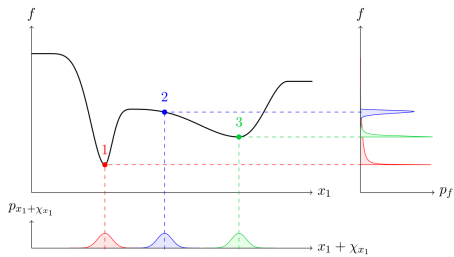
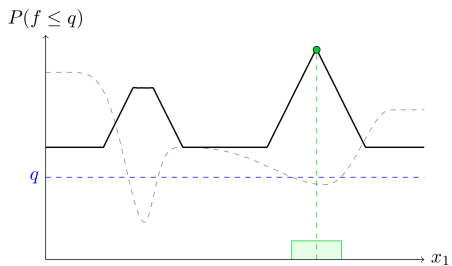
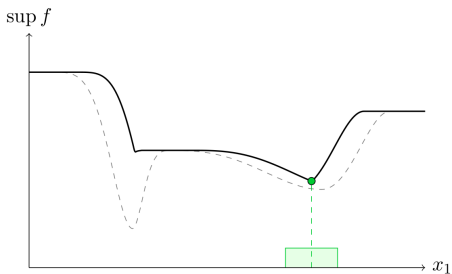
$\equiv$  Equivalently,  $\mathbb{Q}_\alpha(g(x, U)) \leq 0$ .

**but** other formulations involving  $\mathbb{E}g(x, U)$  and  $\mathbb{V}g(x, U)$  (common in the litterature) do not match the meaning of what a constraint is. What counts is the risk of not satisfying the constraint, not the average quantity by which it is violated.

# Risk measures: illustrations

Illustrations taken from [Baudoui, 2012]





The choice of the risk measure changes the optimum and the mathematical properties of the problem!



# Risk measures for both objective and constraints

The robust and reliable formulations:

- They are based on the feasible trajectories,  $[f(x, U) \mid g(x, U) \leq 0]$ , used within the  $\rho^f$  measures, e.g.,  $\mathbb{E}(f(x, U) \mid g(x, U))$ ,
- associated to a constraint risk measure  $\rho^g$ .

An ideal series formulation:

- $G(x)$  the random event “all constraints are satisfied at  $x$ ”,  
 $G(x) \equiv \bigcap_i [g_i(x, U) \leq 0] \leftarrow$  now several constraints  $g_i(\cdot)$ ,

$$\begin{aligned} & \min_{x \in \mathcal{X}} \mathbb{Q}_\alpha(f(x, U) \mid G(x)) \\ & \text{such that } \alpha' - \mathbb{P}(G(x)) \leq 0 \end{aligned}$$

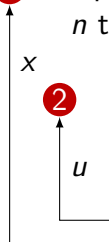
(to the best of my knowledge, no contribution to solve this problem: it cumulates the difficulties of quantile estimation and dependent criteria)

# The double loop issue

The cumulation of the search on  $x$  with the estimation of the risk measures creates a double loop that takes too long to calculate in the context of expensive functions.

(this, in addition with the importance of taking uncertainties into account, explains the number of articles dealing with both uncertainties and optimization.)

1 Optimization loop: propose new  $x$  based on past  $[x, p(x)]$ ,  
 $n$  times



2 Risk estimation loop: at a given  $x$  loop on  $u$ 's to estimate  $p(x)$ .

For example, crude Monte Carlo and mean,

$$\mathbb{E}f(x, U) \approx \frac{1}{m} \sum_{i=1}^m f(s(x, u^i))$$

for a multiplicative cost of  $n \times m \Rightarrow$  methods to avoid the double loop.

## Optimization under uncertainties: an overview with a focus on Gaussian processes

- 1 Problem formulation
  - Motivation and the two classes of variables
  - Formulations as an optimization of statistical measures
- 2 Methods
  - A pointer to methods without metamodels
    - Approximation methods
    - Approaches based on stochastic gradients
  - Methods with metamodels
    - Overall scheme
    - Choice of  $x$
    - Choice of  $u$  knowing  $x$
- 3 Bibliography

# Approximation methods

The original reliability optimization problem (RBDO) is transformed into a problem expressed in terms of *reliability indices*

$$\begin{array}{ll} \min_{x \in \mathcal{X}} f(x) & \\ \text{such that } \mathbb{P}(g(x, U) \leq 0) \geq \alpha & \Rightarrow \end{array} \quad \begin{array}{ll} \min_{x \in \mathcal{X}} f(x) & \\ \text{such that } \beta(x) \geq \beta^{\text{tol}} & \end{array}$$

where  $\beta^{\text{tol}} = \Phi^{-1}(\alpha)$  and the reliability index is solution of

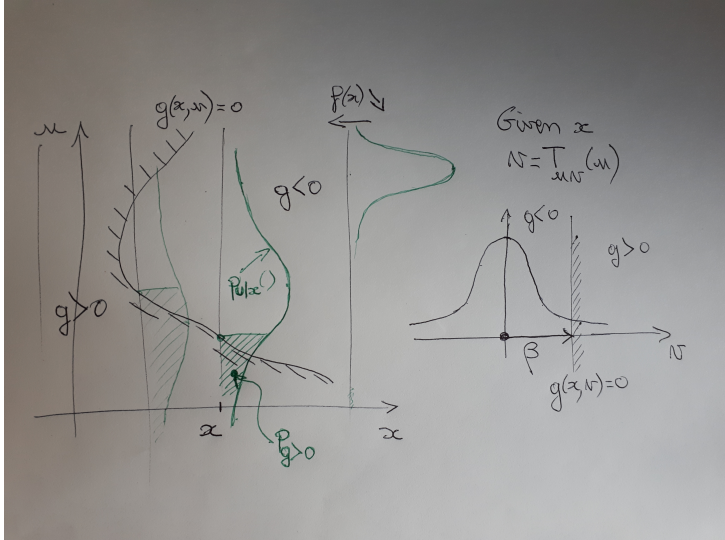
$$\beta(x) = \arg \min_v \|v\| \text{ such that } g(x, T_{vu}(v)) > 0$$

and  $T_{vu}()$  is a transformation from a vector of random variables  $V$  that are standard normal to  $U$ , typically the inverse of  $v = \Phi^{-1}(\text{CDF}(u)) \leftarrow$  trade the cost of an integral for an optimization problem.

See I. Papaioannou's class for transformation,

[Hasofer and Lind, 1974] for the reliability indices,

[Valdebenito and Schuëller, 2010] for a review of FORM / SORM in optimization



If  $g(x, T_{vu}(\cdot))$  is linear in  $v$ , there is an equivalence with the probabilities,  $\mathbb{P}(g(x, u) \leq 0) = \Phi(\beta(x))$ . The equivalence exists also with quadratic  $g(x, T_{vu}(\cdot))$ , but in general it is an approximation.

# Approximation methods: 2, 1, 0 loop

## 2 loops

- 1 Optimization algorithm: proposes  $x^{t+1}$
- 2 Calculate  $\beta(x^{t+1})$  (reliability analysis, sub-optimization with iteration on  $u$  through  $T_{vu}(v)$ ) and  $f(x^{t+1})$   
Stop or go back to 1

## 1 loop

- 1 Optimization algorithm: proposes  $x^{t+1}$  and approximate reliability analyses, e.g.,  $\hat{\beta}(x^{t+1}) = \beta(x^t) + \nabla_x \beta(x^t)(x^{t+1} - x^t)$   
Stop or sometimes update approximation  $\hat{\beta}(\cdot)$  and go back to 1

## “decoupled”

(find worst uncertainty)

$$\max_v g(x^t, T_{vu}(v))$$

such that  $\|v\| = \beta^{\text{tol}}$

$v^t$

$\rightleftharpoons$

$x^t$

(optim with fixed uncertainty)

$$\min_x f(x)$$

such that  $g(x, T_{vu}(v^t))$

# Optimization under uncertainty with stochastic gradients

Many risk measures can be written as sums over  $u$  samples:

$$\frac{1}{N} \sum_{i=1}^N f(x, u^i) \xrightarrow{N \nearrow} \mathbb{E}f(x, U)$$
$$\frac{1}{N^2} \sum_{i < j}^N (f(x, u^i) - f(x, u^j))^2 \xrightarrow{N \nearrow} \mathbb{V}f(x, U)$$
$$\frac{1}{N} \sum_{i=1}^N \mathbb{1}_{g(x, u^i) \leq 0} \xrightarrow{N \nearrow} \mathbb{P}(g(x, U) \leq 0)$$

which makes them appropriate for stochastic gradients. Parallel with machine learning methods where the  $u_i$ 's are data samples.

Example with the mean minimization, the Arrow-Hurwicz iteration reads

$$x^{i+1} = \text{Proj}_{\mathcal{X}} (x^i - \varepsilon_i \nabla_x f(x^i, u^i)) \quad , \quad u^i \sim p_U(\cdot) \quad , \\ \sum_{i \in \mathbb{N}} \varepsilon_i = +\infty \quad , \quad \sum_{i \in \mathbb{N}} \varepsilon_i^2 < +\infty$$

If  $\nabla_x f(x^i, u^i)$  is known and an unbiased noisy sample of  $\nabla_x \mathbb{E}f(x, U)$ , the algorithm converges to a local minimizer of the mean.

Cf. [Andrieu et al., 2011] for RBDO formulation & theory (like what do you do with the gradient of  $\mathbb{1}_{g(x, u_i) \leq 0}$ ).



## Optimization under uncertainties: an overview with a focus on Gaussian processes

- 1 Problem formulation
  - Motivation and the two classes of variables
  - Formulations as an optimization of statistical measures
- 2 Methods
  - A pointer to methods without metamodels
    - Approximation methods
    - Approaches based on stochastic gradients
  - Methods with metamodels
    - Overall scheme
    - Choice of  $x$
    - Choice of  $u$  knowing  $x$
- 3 Bibliography

# Methods with metamodels (kriging)

- 1 Create an initial Design of Experiments (DoE),  $(x^i, u^i, f \text{ or } g(x^i, u^i))$  and use it to initialize 1/many Gaussian Processes (GPs) (in  $\mathcal{X}$  or in augmented  $\mathcal{X}, \mathcal{A}$ ).

Then, there are 2 steps

- 2 Use the GP(s) to choose the next  $x^{t+1}$
- 3 Choose the next  $u^{t+1}$  knowing  $x^{t+1}$
- 4 Evaluate  $f(x^{t+1}, u^{t+1})$  and  $g(x^{t+1}, u^{t+1})$  (i.e.,  $s(x^{t+1}, u^{t+1})$ ), update the GPs, stop or return to 2.

# Methods with metamodels (kriging)

Ideally,  $x^{t+1}$  and  $u^{t+1}$  should be such that

- $x^{t+1}$  guides the search towards good regions (low risk estimators): global optimization.
- $u^{t+1}$  helps improve the risk estimators at  $x^{t+1}$ : statistical estimation.
- both  $x^{t+1}$  and  $u^{t+1}$  once evaluated, contributed to an improved GP in important parts of the design space.

It is sometimes possible to build the metamodel in the lower dimensional space of real variables  $x^{\text{real}}(x, u)$ , for example the earlier radius where  $x^{\text{real}} = r = x + u$ ,  $Y(x, u) = Z(x + u)$ .

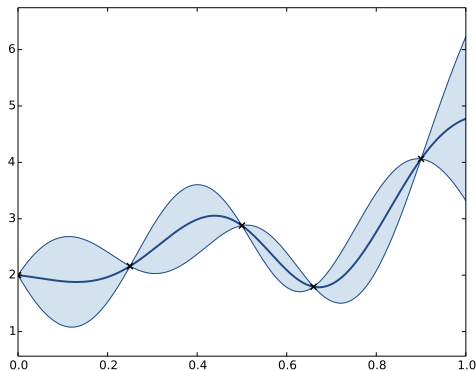
Batch versions: many  $x^{t+1}$ 's, many  $u^{t+1}$ 's. Ex: brute force Monte Carlo (MC), parallel EGO [Ginsbourger et al., 2010].



**Global optimization** methods are a trade-off between

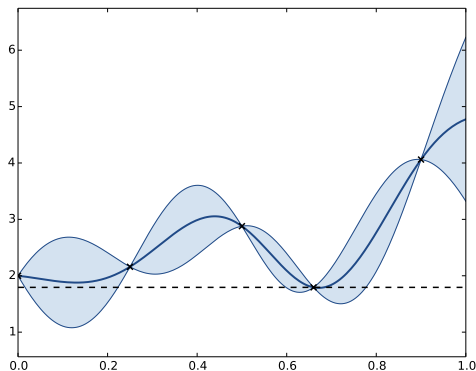
- Intensification in known good regions
- Exploration of new regions

**How can GPR models be helpful?**



(EGO figures from [Durrande and Le Riche, 2017])

In our example, the best observed value is 1.79



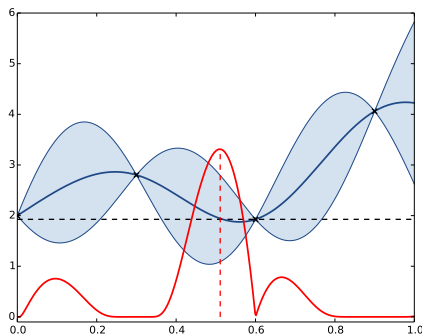
We need a criterion that uses the GP and seeks a compromise between exploration and intensification: the expected improvement

...

# The Expected Improvement

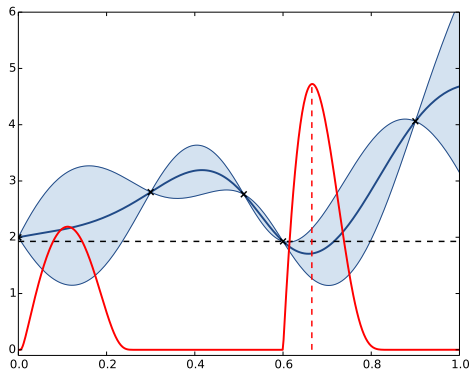
$$\text{EI}(x) = \int_{-\infty}^{+\infty} \max(0, (\min(F) - Y(x))) dy(x) = \dots = \sqrt{c(x, x)} [w(x)\text{cdf}_{\mathcal{N}}(w(x)) + \text{pdf}_{\mathcal{N}}(w(x))]$$

$$\text{with } w(x) = \frac{\min(F) - m(x)}{\sqrt{(c(x, x))}}.$$



# Expected Improvement

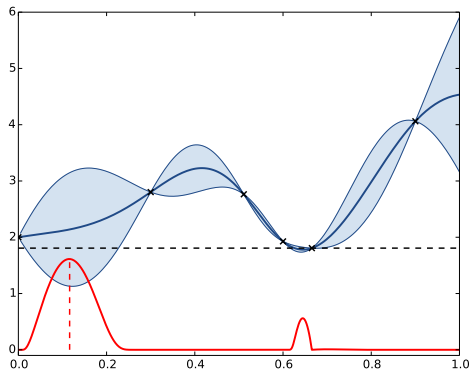
Let's see how it works... iteration 1





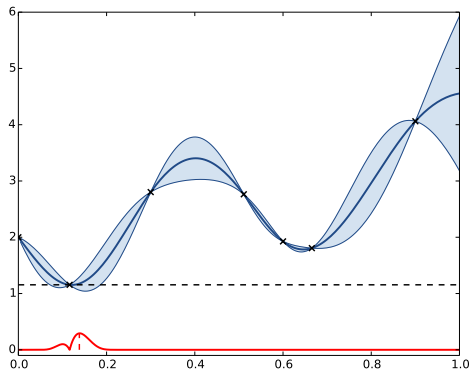
# Expected Improvement

Let's see how it works... iteration 2



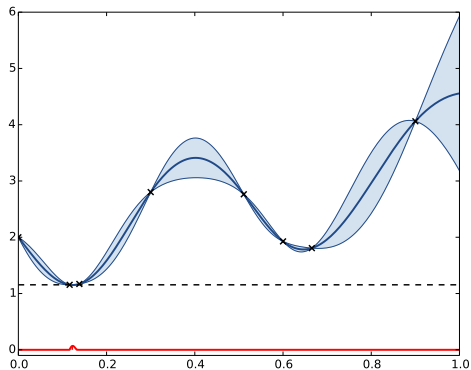
# Expected Improvement

Let's see how it works... iteration 3



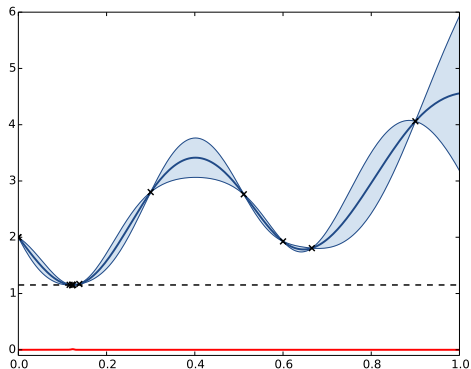
# Expected Improvement

Let's see how it works... iteration 4



# Expected Improvement

Let's see how it works... iteration 5



This algorithm is called **Efficient Global Optimization** (EGO, [Jones et al., 1998]):

- 1 make an initial design of experiments  $X$  and calculate the associated  $F$ ,  $t = \text{length}(F)$
  - 2 built a GP from  $(X, F)$  (max. log-likelihood on  $\sigma$  and  $\theta_i$ 's)
  - 3  $X_{t+1} = \arg \max_x EI(x)$  (with another optimizer, e.g. CMA-ES [Hansen and Ostermeier, 2001])
  - 4 calculate  $F_{t+1} = f(X_{t+1})$ , increment  $t$
  - 5 stop ( $t > t^{\max}$ ) or go to 2.
- + EGO provides a good trade-off between intensification and exploration without arbitrary parameters.
  - + It requires few function observations to get close to optima.
  - × EGO does not converge in the traditional sense: it creates dense samples in the volume of  $S$ . The efficiency comes from the order in which points are sampled.



## Optimization under uncertainties: an overview with a focus on Gaussian processes

- 1 Problem formulation
  - Motivation and the two classes of variables
  - Formulations as an optimization of statistical measures
- 2 Methods
  - A pointer to methods without metamodels
    - Approximation methods
    - Approaches based on stochastic gradients
  - Methods with metamodels
    - Overall scheme
    - Choice of  $x$
    - Choice of  $u$  knowing  $x$
- 3 Bibliography

# EGO for uncertain responses

Consider an optimization problem under uncertainty and its implementation with an estimator

$$\min_x \rho^f(x) \quad \xrightarrow{\text{implemented as}} \quad \min_x \hat{\rho}^f(x)$$

Example:  $\rho^f(x) = \mathbb{E}f(x, U)$ ,  $\hat{\rho}^f(x) = \overline{f(x)} = \frac{1}{N} \sum_{i=1}^N f(x, u^i)$ ,  
 $u^i \sim p_U()$ .



The estimator of the risk, the objective function, is either noisy or biased.

### Example of noisy estimator:

Crude MC estimation of mean is noisy and has variance

$$\mathbb{V} \left( \overline{f(x)} \right) = \frac{1}{N(N-1)} \left( f(x, u^i) - \overline{f(x)} \right)^2$$

**Example of biased estimator:** the **common random numbers** strategy

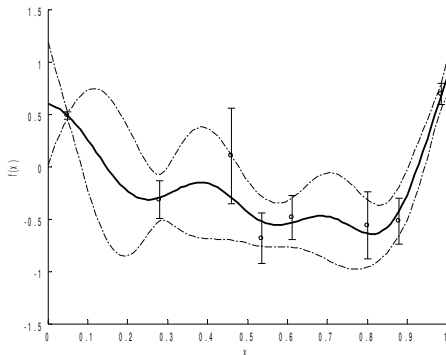
Choose a unique set  $u_{\text{CRN}}^1, \dots, u_{\text{CRN}}^N$ , following  $p_U()$ , for all  $x$ 's. Use your favorite MC estimator with these  $u$ 's, e.g.,  $\overline{f(x)}_{\text{CRN}} = \frac{1}{N} \sum_{i=1}^N f(x, u_{\text{CRN}}^i)$ .

In both cases, don't trust the response, consider it noisy.

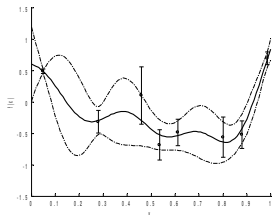
# GP to guide the optimization of risk estimates

Let's start putting things together.

EGO can be applied to the response  $\hat{p}^f(x)$  but 2 changes are needed to account for the noise:



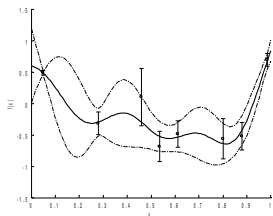
2 additions to EGO to optimize the estimator  $\hat{p}^f(x)$ :



1. Use kriging with noise:  $c(X, X) \rightarrow c(X, X) + N(X, X)$  (cf. course [Le Riche and Durrande, 2019])

- Homogeneous nugget,  $N(X, X) = \text{diag}(\tau^2)$ , and  $\tau$  is estimated with the other GP parameters in the likelihood maximization.
- Heterogeneous nugget,  $N(X, X) = \text{diag}(\tau_i^2)$ , the  $\tau_i^2$  are the variances of the risk estimator at the  $x^i$ 's (example with a quantile in [Le Riche et al., 2009]). Cannot be estimated because not enough data points (ill-posed max likelihood).

2 additions to EGO to optimize the estimator  $\hat{\rho}^f(x)$ :



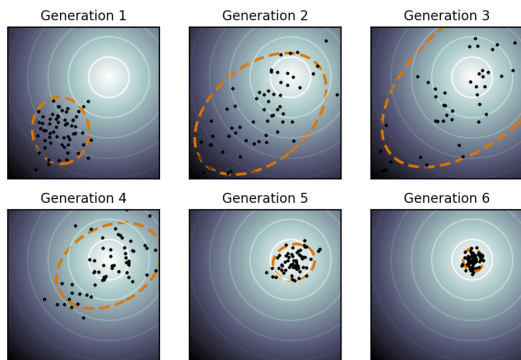
## 2. Don't trust the best observation:

- Let noisy kriging filter out the noise: replace  $\min(F)$  in the EI formula by  $\min(m(X))$ .
- 2 levels kriging: do a normal EGO on the filtered observations  $m(F)$  (i.e., 2 GP models built)

# Evolutionary algorithms for uncertain responses

It has been known for some time that population based stochastic optimizers can locate the region of the optimum of noisy functions.

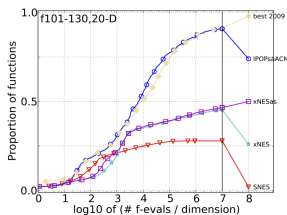
Principle of a CMA-ES algorithm [Hansen and Ostermeier, 2001]:



(from wikipedia I guess)

The two ingredients are large enough populations and mutations [Beyer and Sendhoff, 2007],[Hansen et al., 2008]. But this goes against cost (number of function calls).

Versions of these algorithms that perform well on noisy functions have cost control mechanisms added:



summary of results on the 2012 noisy  
testbed of COCO

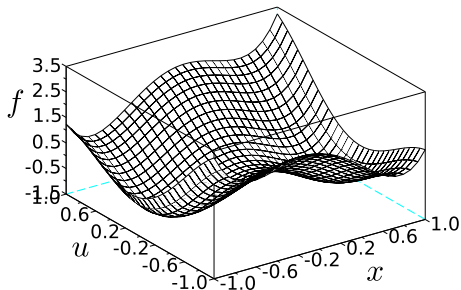
- A (1,4)-ES with mirrored sampling and sequential selection [Brockhoff et al., 2010].
- The IPOPsaACM = CMA algorithm with increasing population (IPOP) and a surrogate [Loshchilov et al., 2012].

# $x^{t+1}$ from kriging in the augmented space

An example of what can be done in optimization under uncertainty with a GP built in the augmented  $(\mathcal{X}, \mathcal{A})$ -space of  $(x, u)$  variables.

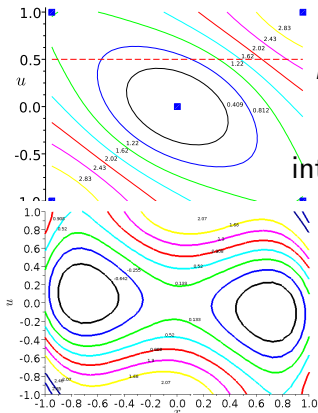
Objective

$$\min_{x \in \mathcal{X}} \mathbb{E}f(x, U)$$



Cf. [Janusevskis and Le Riche, 2012]

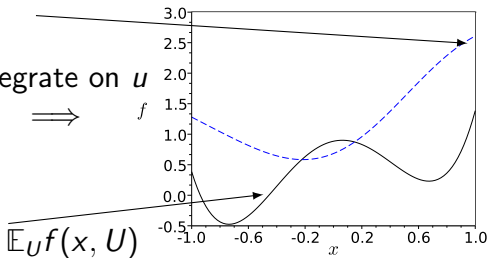
- $f(x, u)$  is approximated by  $Y_{\omega}^t(x, u)$   
 $Y_{\omega}^t(x, u)$  the GP conditioned by the DoE at time  $t$ ,  
 $[(x^1, u^1), f((x^1, u^1)), \dots, (x^t, u^t), f((x^t, u^t))]$
- $\mathbb{E}_U f(x, U)$  is approximated by  $Z_{\omega}^t(x) := \mathbb{E}_U Y_{\omega}^t(x, U)$   
 $Z_{\omega}^t(x)$  the integrated process.



$$m_Z(x) = \mathbb{E}_{\omega} \mathbb{E}_U Y_{\omega}(x, U)$$

integrate on  $u$

$\Rightarrow f$





The integrated process

$$Z_{\omega}^t(x) := \mathbb{E}_U Y_{\omega}^t(x, U) = \int_{\mathcal{A}} Y_{\omega}^t(x, u) p_U(u) du$$

is a linear transformation of the Gaussian process  $Y^t(\cdot, \cdot) \Rightarrow$  it is Gaussian and fully defined by its mean and covariance

$$m_Z(x) = \int_{\mathcal{A}} m(x, u) p_U(u) du$$
$$c_Z(x, x') = \int_{\mathcal{A}} \int_{\mathcal{A}} c((x, u), (x', u')) p_U(u) p_U(u') du du'$$

(analytical expressions given in [Janusevskis and Le Riche, 2012] for  $U$  Gaussian, otherwise the integrations needs to be done numerically)

$Z_{\omega}^t(x)$  is just another GP so, back to the minimization of the average, the next  $x$  can be found by maximization of the expected improvement,

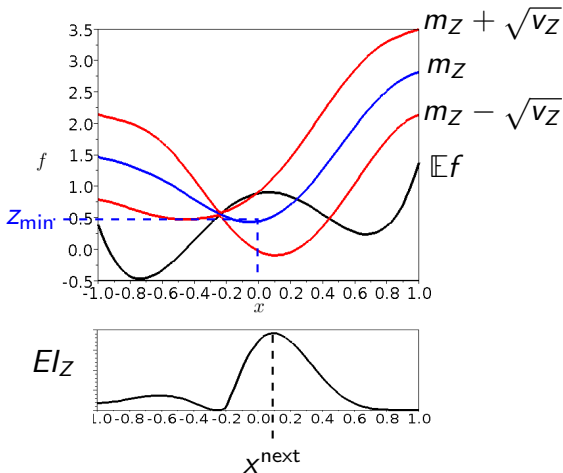
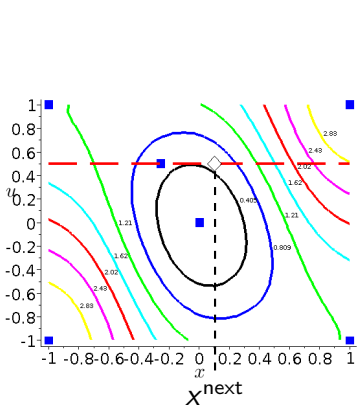
$$x^{\text{next}} = \max_{x \in \mathcal{X}} \text{El}_Z(x)$$

where, remember,

$$\text{El}_Z(x) = \sqrt{c(x, x)} [w(x) \text{cdf}_{\mathcal{N}}(w(x)) + \text{pdf}_{\mathcal{N}}(w(x))]$$

$$\text{with } w(x) = \frac{z_{\min} - m_Z(x)}{\sqrt{(c_Z(x, x))}}.$$

But  $Z(x)$  is not observed so define  $z_{\min} := \min(m_Z(X))$ .  
Start EGO on  $Z(x)$  ...



**x ok. What about  $u$ , which we need to call the simulator?**

## Optimization under uncertainties: an overview with a focus on Gaussian processes

- 1 Problem formulation
  - Motivation and the two classes of variables
  - Formulations as an optimization of statistical measures
- 2 Methods
  - A pointer to methods without metamodels
    - Approximation methods
    - Approaches based on stochastic gradients
  - Methods with metamodels
    - Overall scheme
    - Choice of  $x$
    - Choice of  $u$  knowing  $x$
- 3 Bibliography

Remember optimizers with uncertainty:

- 1 Create an initial Design of Experiments (DoE),  $(x^i, u^i, f \text{ or } g(x^i, u^i))$  and use it to initialize the GP(s) (in  $\mathcal{X}$  or in augmented  $\mathcal{X}, \mathcal{A}$ ).

Then, there are 2 steps

- 2 Use the GP(s) to choose the next  $x^{t+1}$
- 3 Choose the next  $u^{t+1}$  knowing  $x^{t+1}$
- 4 Evaluate  $f(x^{t+1}, u^{t+1})$  and  $g(x^{t+1}, u^{t+1})$  (i.e.,  $s(x^{t+1}, u^{t+1})$ ), update the GP(s), stop or return to 2.

## Choosing $u$ knowing $x^{\text{next}}$

- **Brute force MC**: calculate all  $[f \text{ and } g](x^{\text{next}}, u^i)$  needed by the risk estimator. Ex:  $\hat{p}^f = 1/N \sum_{i=1}^N f(x^{\text{next}}, u^i)$ ,  $u_i \sim p_U()$ . Costs  $N$ .
- **MC-kriging**: replace calls to  $f$  or  $g$  by calls to the metamodel (GP mean). Ex:  $\hat{p}^f = 1/N \sum_{i=1}^N m(x^{\text{next}}, u^i)$ ,  $u^i \sim p_U()$ . Still need to choose  $u^{t+1}$  where to call the true simulator. Simplest: sample it,  $u^{t+1} \sim p_U()$ . Costs 1. But the augmented GP is not used in the choice. ☹

We now explain two approaches taking advantage of the GP in augmented space:

- **Minimize the average function**:  $\min_x f(x, U)$  continued.
- **Reliability problem**:  $\min_x f(x)$  such that  $\mathbb{P}(g(x, U) \leq 0) > \alpha$ .

## Back to average minimization: $u$ knowing $x^{\text{next}}$

continued, from [Janusevskis and Le Riche, 2012]

$x^{\text{next}}$  gives a region of interest from an optimization of the expected  $f$  point of view.

One simulation will be run to improve our knowledge of this region of interest  $\rightarrow$  one choice of  $(x, u)$ .

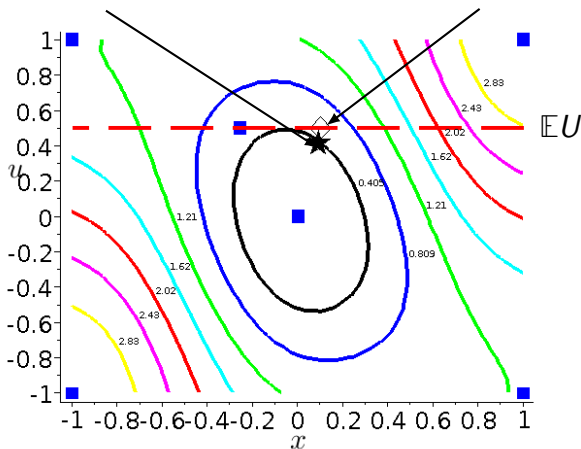
Choose  $(x^{t+1}, u^{t+1})$  that provides the most information, i.e., which minimizes the one-step-ahead variance<sup>1</sup> of the integrated process at  $x^{\text{next}}$ :

$$(x^{t+1}, u^{t+1}) = \arg \min_{(x, u) \in \mathcal{X} \times \mathcal{A}} \mathbb{V}[Z^{t+1}(x^{\text{next}})]$$

<sup>1</sup>See Stepwise Uncertainty Reduction methods (SUR). 

$$x^{t+1} = \arg \min_{(x,u)} \mathbb{V}[Z^{t+1}(x^{\text{next}})]$$

$$x^{\text{next}} = \arg \max_x \mathbb{E}U$$





$\min_x \mathbb{E}f(x, U)$  with EGO in augmented space

Putting it together:

1 Create an initial Design of Experiments (DoE),  $(x^i, u^i, f(x^i, u^i))$

**While** budget no exhausted or other stopping criterion

Update the conditional GP  $Y^t(x, u)$  from last DoE. The updated  $Z^t(x)$  stems from it.

2 Maximize EI of  $Z^t(x) \rightarrow x^{\text{next}}$

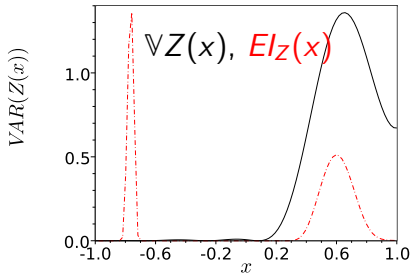
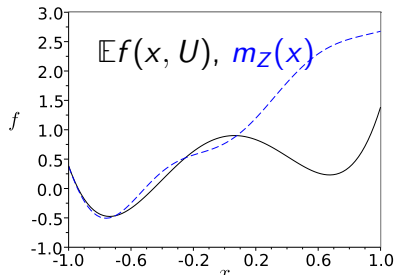
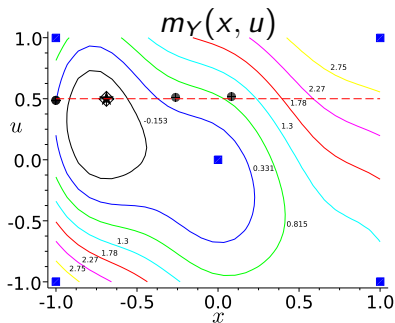
2' & 3 Minimize  $\mathbb{V}[Z^{t+1}(x^{\text{next}})] \rightarrow x^{t+1} \& u^{t+1}$

Calculate  $f(x^{t+1}, u^{t+1})$ , update DoE,  $t \leftarrow t + 1$

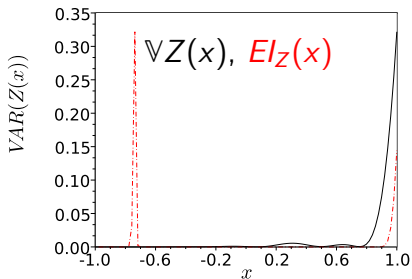
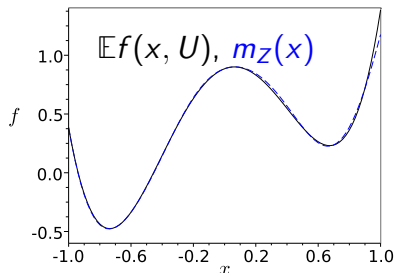
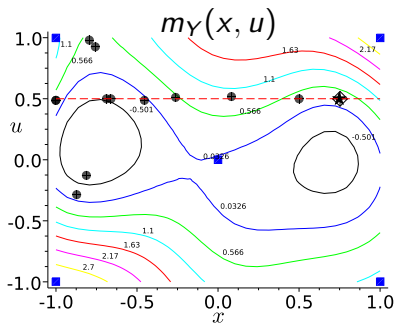
**End while**

3 (nonlinear, multimodal) sub-optimizations involved but they do not call the expensive  $f$ . Use a global optimizer (e.g., CMA-ES or restarted BFGS).

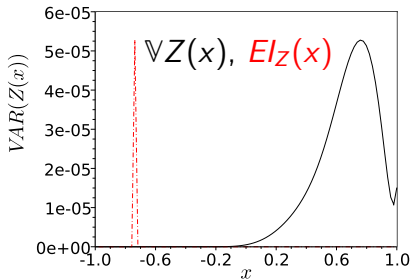
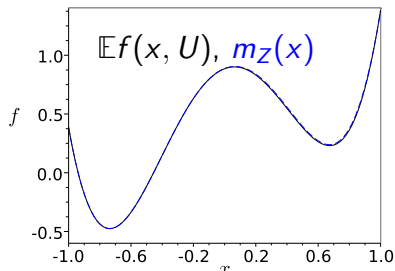
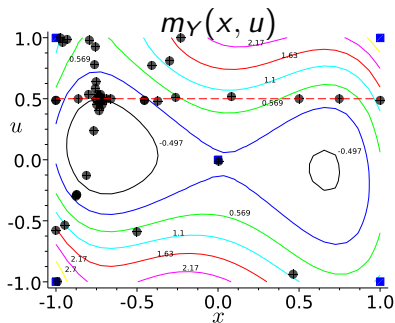
# Example: 2D Camelback function, iteration 6



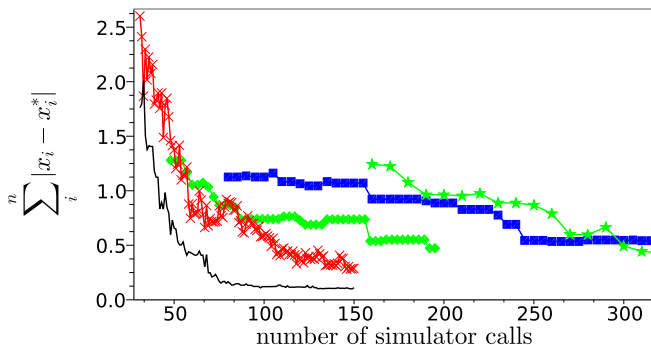
# Example: 2D Camelback function, iteration 14



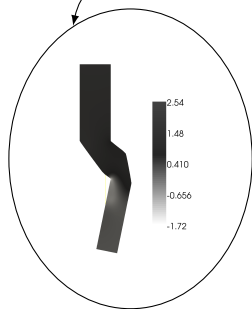
# Example: 2D Camelback function, iteration 50



# Example: 6D Michalewicz function, convergence averaged over 10 runs



designed  
with EI-VAR



EI-VAR, MC-kriging, and brute force MC for  $N = 3$  (diamonds), 5 (squares), 10 (stars) samples

# RBDO in the augmented space

From [Moustapha et al., 2016]

Problem to solve:

$$\min_{x \in \mathcal{X}} f(x) \text{ such that } \mathbb{Q}_\alpha(g(x, U)) \leq 0 \quad (\Leftrightarrow \mathbb{P}(g(x, U) \leq 0) \geq \alpha)$$

Method:

- $x^{t+1}$  given by a 1+1-CMA-ES optimization algorithm
- $u^{t+1}$  comes from a deviation number that is small when the GP is near but inaccurate at the quantile level,

$$u^{t+1} = \arg \min_{u \in \mathcal{A}} \frac{\hat{q}_\alpha(x^{t+1}) - m(x^{t+1}, u)}{v(x^{t+1}, u)}$$

where  $\hat{q}_\alpha(x) = m(x, u^{\lfloor N\alpha \rfloor})$

- Calculate  $f(x^{t+1})$ ,  $g(x^{t+1}, u^{t+1})$ , update and loop as usual.

# References I



Andrieu, L., Cohen, G., and Vázquez-Abad, F. J. (2011).  
Gradient-based simulation optimization under probability constraints.  
*European Journal of Operational Research*, 212(2):345–351.



Baudoui, V. (2012).  
*Optimisation robuste multiobjectifs par modèles de substitution*.  
PhD thesis, Université de Toulouse, France.



Beyer, H.-G. and Sendhoff, B. (2007).  
Robust optimization—a comprehensive survey.  
*Computer methods in applied mechanics and engineering*, 196(33-34):3190–3218.



Brockhoff, D., Auger, A., Hansen, N., Arnold, D. V., and Hohm, T. (2010).  
Mirrored sampling and sequential selection for evolution strategies.  
In *PPSN 2010*, pages 11–21. Springer.



Durrande, N. and Le Riche, R. (2017).  
Introduction to Gaussian Process Surrogate Models.  
Lecture at 4th MDIS form@ter workshop, Clermont-Fd, France.  
HAL report cel-01618068.

# References II



Ginsbourger, D., Le Riche, R., and Carraro, L. (2010).  
Kriging is well-suited to parallelize optimization.  
In *Computational intelligence in expensive optimization problems*, pages 131–162.  
Springer.



Hansen, N., Niederberger, A. S., Guzzella, L., and Koumoutsakos, P. (2008).  
A method for handling uncertainty in evolutionary optimization with an application to  
feedback control of combustion.  
*Transactions on Evolutionary Computation*.



Hansen, N. and Ostermeier, A. (2001).  
Completely derandomized self-adaptation in evolution strategies.  
*Evol. Comput.*, 9(2):159–195.



Hasofer, A. M. and Lind, N. C. (1974).  
Exact and invariant second-moment code format.  
*Journal of the Engineering Mechanics division*, 100(1):111–121.



Janusevskis, J. and Le Riche, R. (2011).  
Robust optimization of a 2D air conditioning duct using kriging.  
Technical report.  
<https://hal.archives-ouvertes.fr/hal-00566285>.



# References III



Janusevskis, J. and Le Riche, R. (2012).  
Simultaneous kriging-based estimation and optimization of mean response.  
*Journal of Global Optimization*.  
DOI 10.1007/s10898-011-9836-5.



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.



Le Riche, R. and Durrande, N. (2019).  
An overview of kriging for researchers.  
Lecture at mnmuq2019 summer school, Porquerolles, France.  
to appear on HAL.



Le Riche, R., Picheny, V., Ginsbourger, D., Meyer, A., and Kim, N.-H. (2009).  
Gears design with shape uncertainties using monte carlo simulations and kriging.  
In *50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*, Palm Springs, USA.  
paper No. AIAA-2009-2257.



Loshchilov, I., Schoenauer, M., and Sebag, M. (2012).  
Self-adaptive surrogate-assisted covariance matrix adaptation evolution strategy.  
In *Proc. of the 14th GECCO*, pages 321–328. ACM.

# References IV



Moustapha, M., Sudret, B., Bourinet, J.-M., and Guillaume, B. (2016).  
Quantile-based optimization under uncertainties using adaptive kriging surrogate models.  
*Structural and multidisciplinary optimization*, 54(6):1403–1421.



Park, G.-J., Hwang, K.-H., Lee, T., and Hee Lee, K. (2006).  
Robust design: An overview.  
*AIAA Journal*, 44:181–191.



Pujol, G., Le Riche, R., Roustant, O., and Bay, X. (2009).  
*Optimisation multidisciplinaire en mécanique: Réduction de modèles, robustesse, fiabilité, réalisations logicielles*, chapter L'incertitude en conception: formalisation, estimation.  
Hermes Science Publications, Paris.



Valdebenito, M. A. and Schuëller, G. I. (2010).  
A survey on approaches for reliability-based optimization.  
*Structural and Multidisciplinary Optimization*, 42(5):645–663.