Sequential approaches to reliability estimation and optimization based on kriging

Victor Picheny1, Rodolphe Le Riche[2,3]
and Olivier Roustant[3]
[1] CERFACS ; [2] CNRS ; [3] Ecole des Mines de Saint-Etienne

GDR Mascot Num 2012, Paris

Intro : uncertainties and optimization

Models and working conditions are partially unknown → uncertainties. These uncertainties need to be taken into account during design.

Ex : a +/- 1mm dispersion in the manufacturing of the air admission line can degrade the engine's performance (g CO_2/km) by +20% (worst case).

Working assumption : uncertainties can be described by random parameters of the models.

Optimization terminology (1)

Model or « simulator », y , (analytical, finite elements, coupled sub-models …) of the object you need to optimize.

Formulation of the optimization problem

$$\min_{x \in S} f(y(x))$$
$$g(y(x)) \leqslant 0$$

x : optimization variables
f : objective functions
g : optimization constraints
f , g : optimization (performance) criteria

Optimization terminology (2) : the double (x,U) parameterization for uncertainties

x is a vector of deterministic optimization (controlled) variables.
x in *S*, the search space.
Without loss of generality, introduce *U*, a vector of uncertain (random) parameters that affect the simulator *y*. $U : (\Omega, C, P) \rightarrow S_U$

y(x) → *y(x,U)* , therefore *f(x)* → *f(y(x,U)) = f(x,U)*
and *g(x)* → *g(y(x,U)) = g(x,U)*

U used to describe

- noise (as in identification with noise measurement)
- model error (epistemic uncertainty)
- uncertainties on the values of some parameters of *y*.

G. Pujol, R. Le Riche, O. Roustant and X. Bay, *L'incertitude en conception: formalisation, estimation,* Chapter 3 of the book *Optimisation Multidisciplinaire en Mécaniques : réduction de modèles, robustesse, fiabilité, réalisations logicielles*, Hermes, 2009.

Ideal formulation of optimization with uncertainties

Replace the noisy optimization criteria by statistical measures

$OK(x)$ is the random event "all constraints are satisfied" ,

$$OK(x) = \bigcap_i \{ g_i(x, U) \leqslant 0 \}$$

$\min_{x \in S} \; q_\alpha^c(x)$ (conditional α-quantile)

such that $P\big(OK(x)\big) \geqslant 1 - \varepsilon$

where $P\big(f(x, U) \leqslant q_\alpha^c(x) \mid OK(x)\big) = \alpha$

$\varepsilon > 0$, small

Outline of the talk

Two sub-problems addressed in this talk.

1. Estimation of reliability constraints

Failure probability : $P_f = Prob\left(g(U) > T \right)$

(notation here : x fixed so $g(x,U) \leq 0 \rightarrow g(U) \leq T$)

2. Optimization with uncertainties

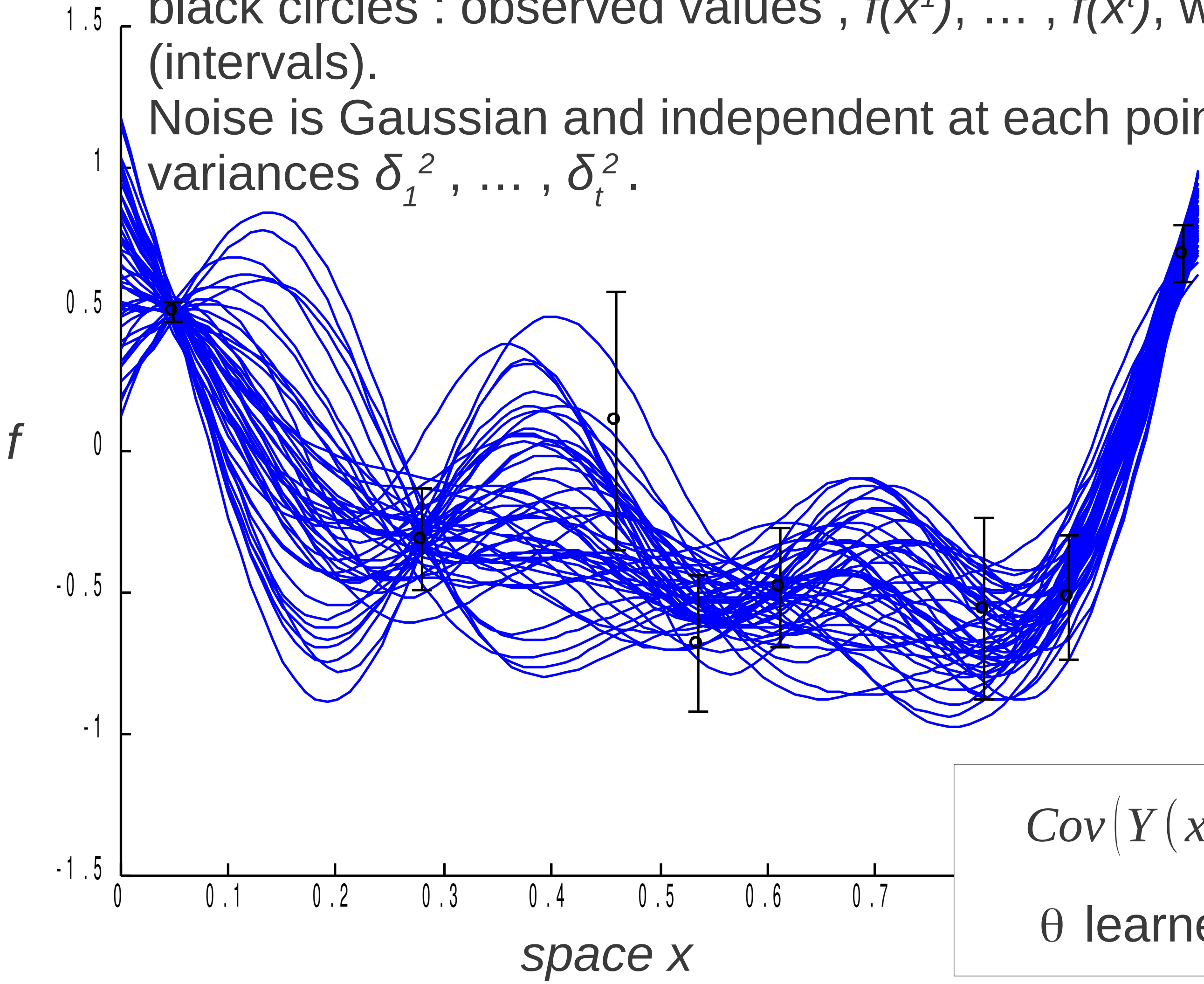
Average minimization : $\min_{x \in S \subset \mathbb{R}^n} E_U\left(f(x,U)\right)$

Unifying thread : iterative kriging strategies. Start by introducing kriging.

Kriging : quick intro (1)

(presentation only with *f(x)* , but generalizes to *f(x,u)* and *g(x)* or *g(x,u)* when needed)

black circles : observed values , $f(x^1)$, … , $f(x^t)$, with heterogeneous noise (intervals).
Noise is Gaussian and independent at each point (nugget effect), variances δ_1^2 , … , δ_t^2 .



Assume : the blue curves are possible underlying true functions.
They are instances of stationary Gaussian processes *Y(x)* → fully characterized by their average *μ* and their covariance,

$$Cov\left(Y(x),Y(x')\right)=\text{Fct}_\theta\left(\text{dist}\left(x,x'\right)\right)$$

θ learned from data , $\left(x^i, f\left(x^i\right)\right)$

Kriging : quick intro (2)

$f(x)$ represented by $Y^t(x) = [\ Y(x) \,|\, f(x^1), \ldots, f(x^t)\ ]$

$Y^t(x) \sim N\left(m_K(x), s_K^2(x)\right)$ (simple kriging)

Kriging average : $m_K(x) = \mu + \boldsymbol{c}^T(x)\boldsymbol{C}_\Delta^{-1}(\boldsymbol{f} - \mu\mathbf{1})$

Kriging variance : $s_K^2(x) = \sigma^2 - \boldsymbol{c}^T(x)\boldsymbol{C}_\Delta^{-1}\boldsymbol{c}(x)$

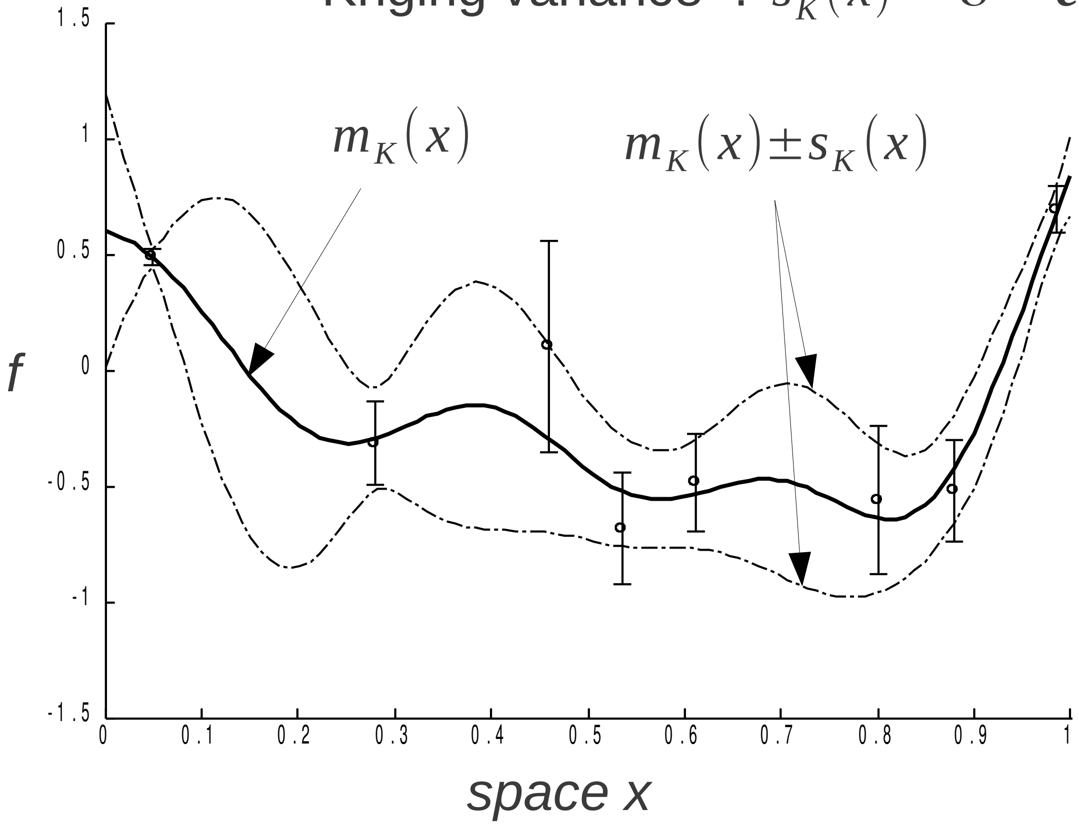
where

$\boldsymbol{c}(x) = \left[Cov\left(Y(x), Y(x^i)\right)\right]_{i=1,t}$

$\boldsymbol{C}_\Delta = \boldsymbol{C} + \boldsymbol{\Delta}$

$\boldsymbol{C} = \left[Cov\left(Y(x^i), Y(x^j)\right)\right]_{i,j}$

$\boldsymbol{\Delta} = diag\left[\delta_1^2, \ldots, \delta_t^2\right]$



Iterative sampling based on kriging for reliability estimation

Goal : estimate the failure probability , $P_f = Prob\left(g\left(U\right) > T\right)$

while sparingly calling $g\left(.\right)$ (g is expensive)

1. Build a kriging-based approximation to $g\left(.\right) \rightarrow m_K\left(.\right)$

2. Use it in a MC procedure : $\widehat{P_f} = \frac{1}{N}\sum_{i=1}^{N} I\left(m_K\left(u^i\right) > T\right)$,

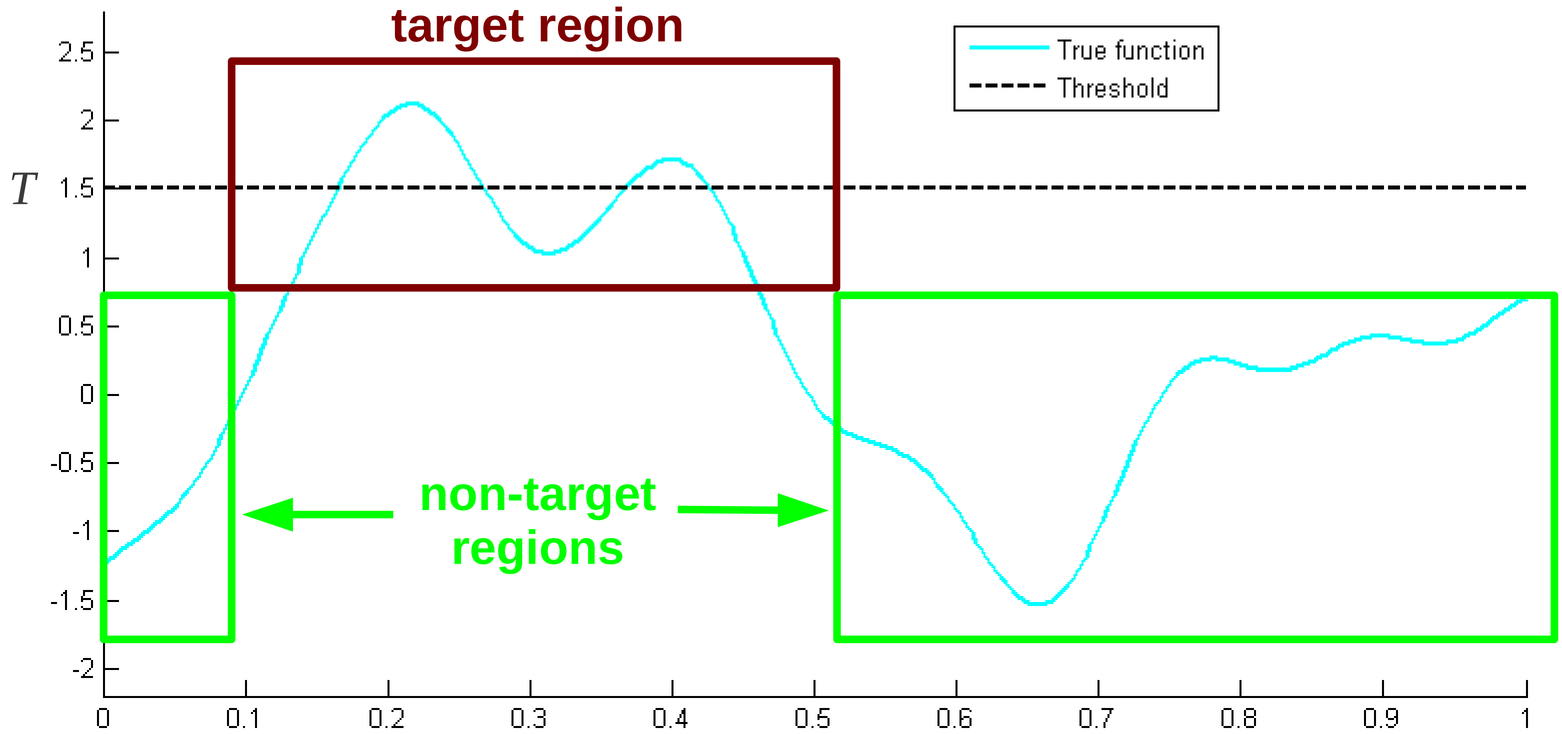
where $u^i \sim$ pdf of U

Question : how to choose a small number of u's to approximate well P_f ?

Approximation of a target region

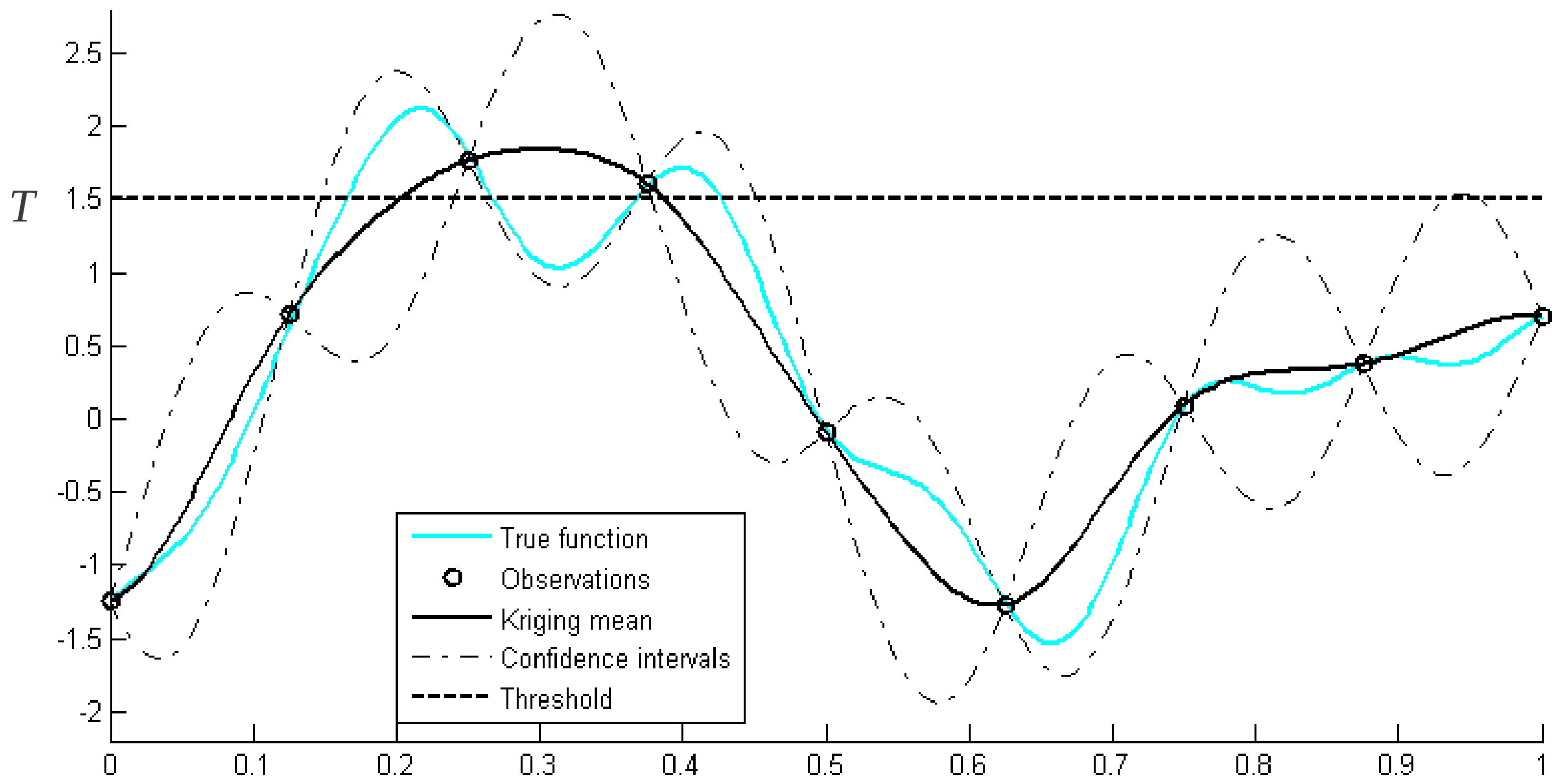
Idea : a global accuracy of the metamodel, $m_k()$, is not needed.
It needs to be accurate when $g(u) \approx T$

[V. Picheny, D. Ginsbourger, O. Roustant, R.T. Haftka and N.-H. Kim, Adaptive designs of experiments for accurate approximation of a target region », Journal of Mechanical Design, 2010.]



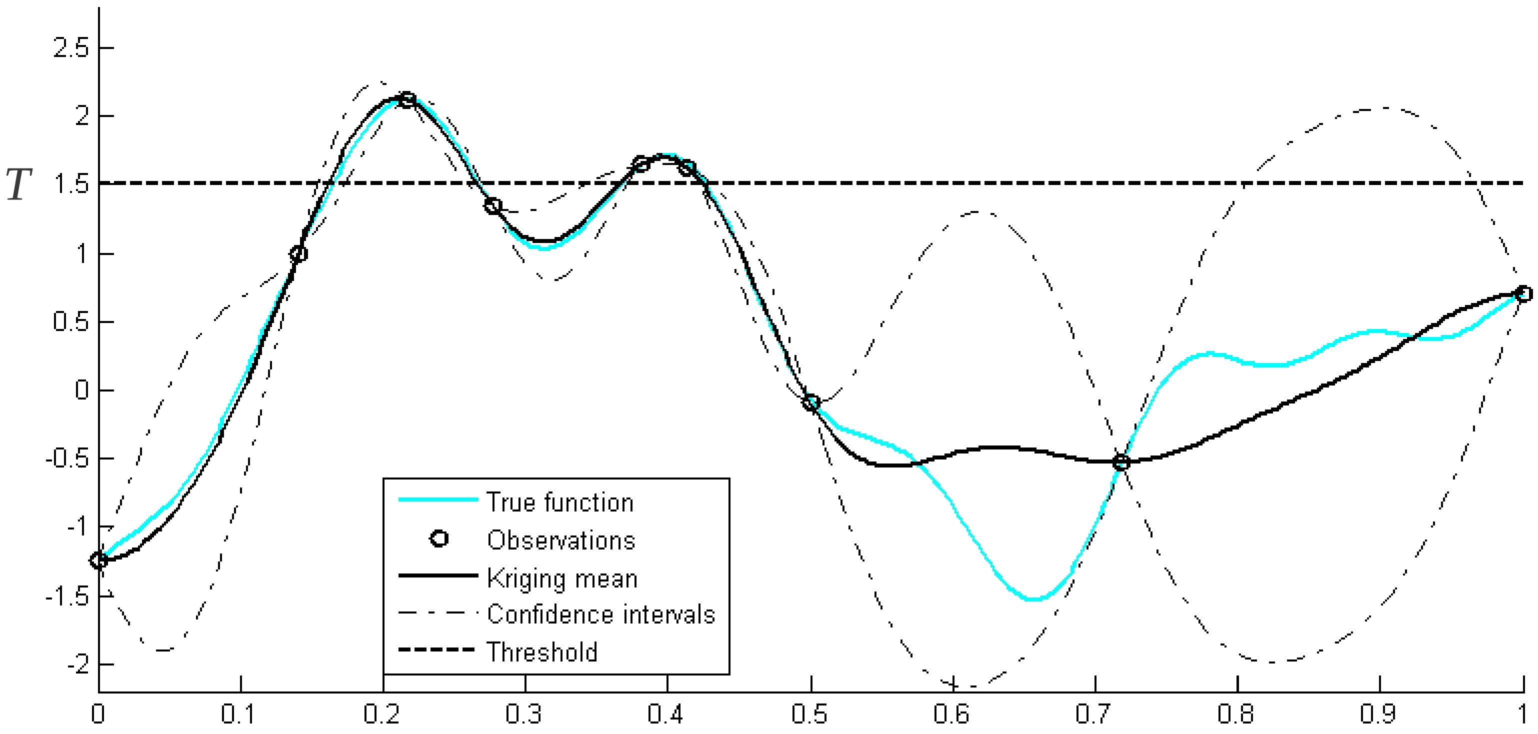
Approximation of a target region – Example (1/2)

Kriging based on a uniform design of experiments :
• reasonable variance everywhere,
• large errors in the target region.



Approximation of a target region – Example (2/2)

Customized design
- large variance in non target regions,
- good accuracy in the target region.

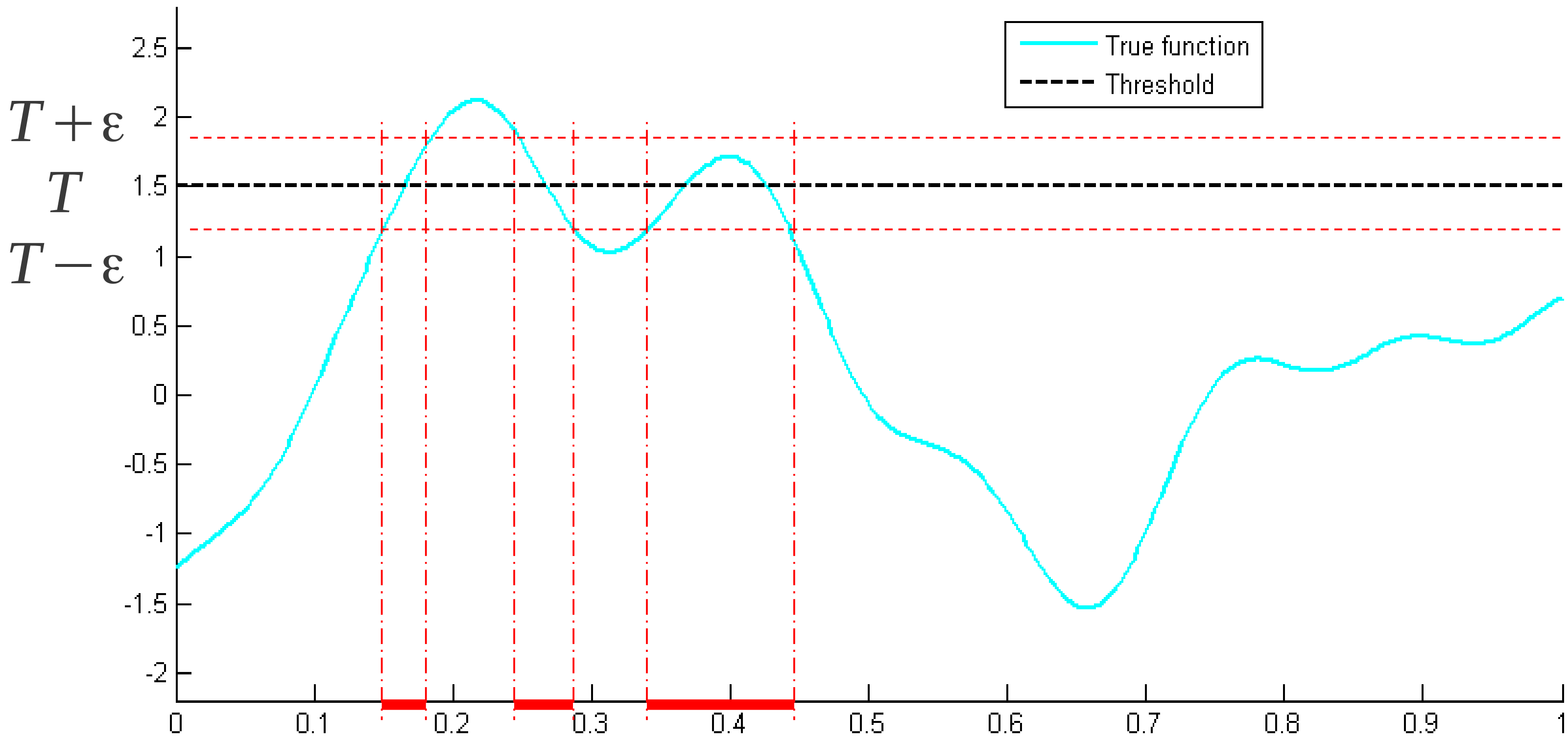


Approximation of a target region – Criterion (1/2)

Target region : $\boldsymbol{U_T} = \left\{ u \in S_U \text{ s. t. } \left| g(u) - T \right| \le \varepsilon \right\}$

Ideal criterion : integration of variance over U_T only,

$$\int s_K^2(u) I(U_T) du$$



BUT $\boldsymbol{U_T}$ is unknown !

Approximation of a target region – Criterion (2/2)

Replace $I(U_T)$ by $E(I(U_T)) = Prob(u \in U_T) = Prob(|Y(u) - T| \leq \varepsilon)$
where $Y(.)$ is the conditional Gaussian process.

The weight in the integral becomes

$$W_\varepsilon(u) \equiv Prob(u \in U_T) = \Phi\left(\frac{T + \varepsilon - m_K(u)}{s_K(u)}\right) - \Phi\left(\frac{T - \varepsilon - m_K(u)}{s_K(u)}\right)$$

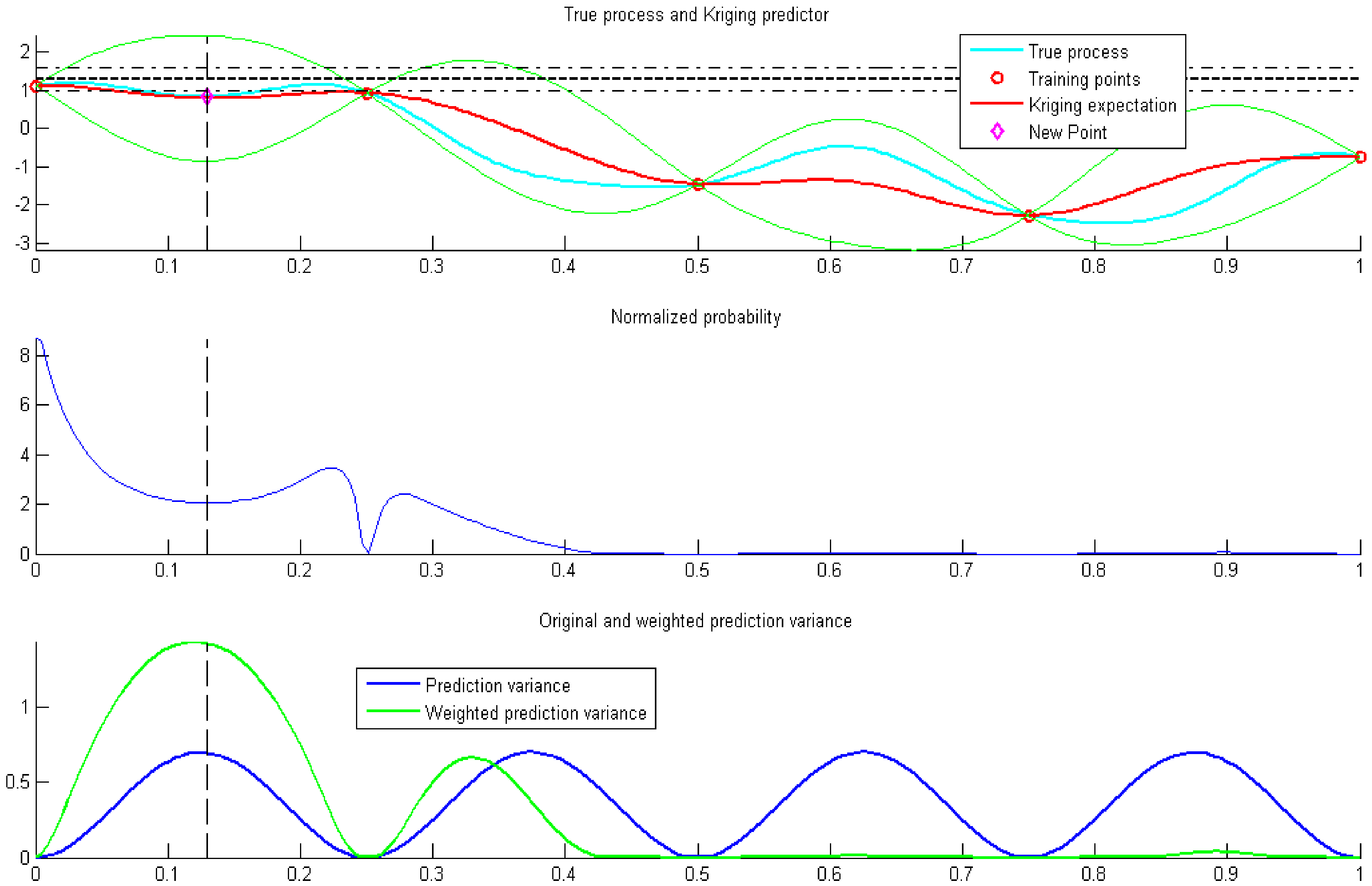
The criterion :

$$IMSE_T = \frac{1}{2\varepsilon} \int s_K^2(u) W_\varepsilon(u) du$$

AN : $W(x) \equiv \lim_{\varepsilon \to 0} \frac{W_\varepsilon(u)}{2\varepsilon} = d_{N(m_K(u), s_K^2(u))}(T)$, the kriging density.

$W(.)$ is large when 1) Y is near the target region;
2) $s_K()$ is large.

Approximation of a target region – Illustration



Approximation of a target region – Algorithm

- Create an initial design (u's), compute the associated g's.
- Do
  - Estimate the kriging parameters,
  - Find the next iterate, minimizer of the one step ahead IMSE$_T$

$$u^* = arg \min_{v \in S_U} \int s_k^2(u \mid v, \mathbf{u}) \, W(u \mid \mathbf{u}, g(\mathbf{u})) \, du$$

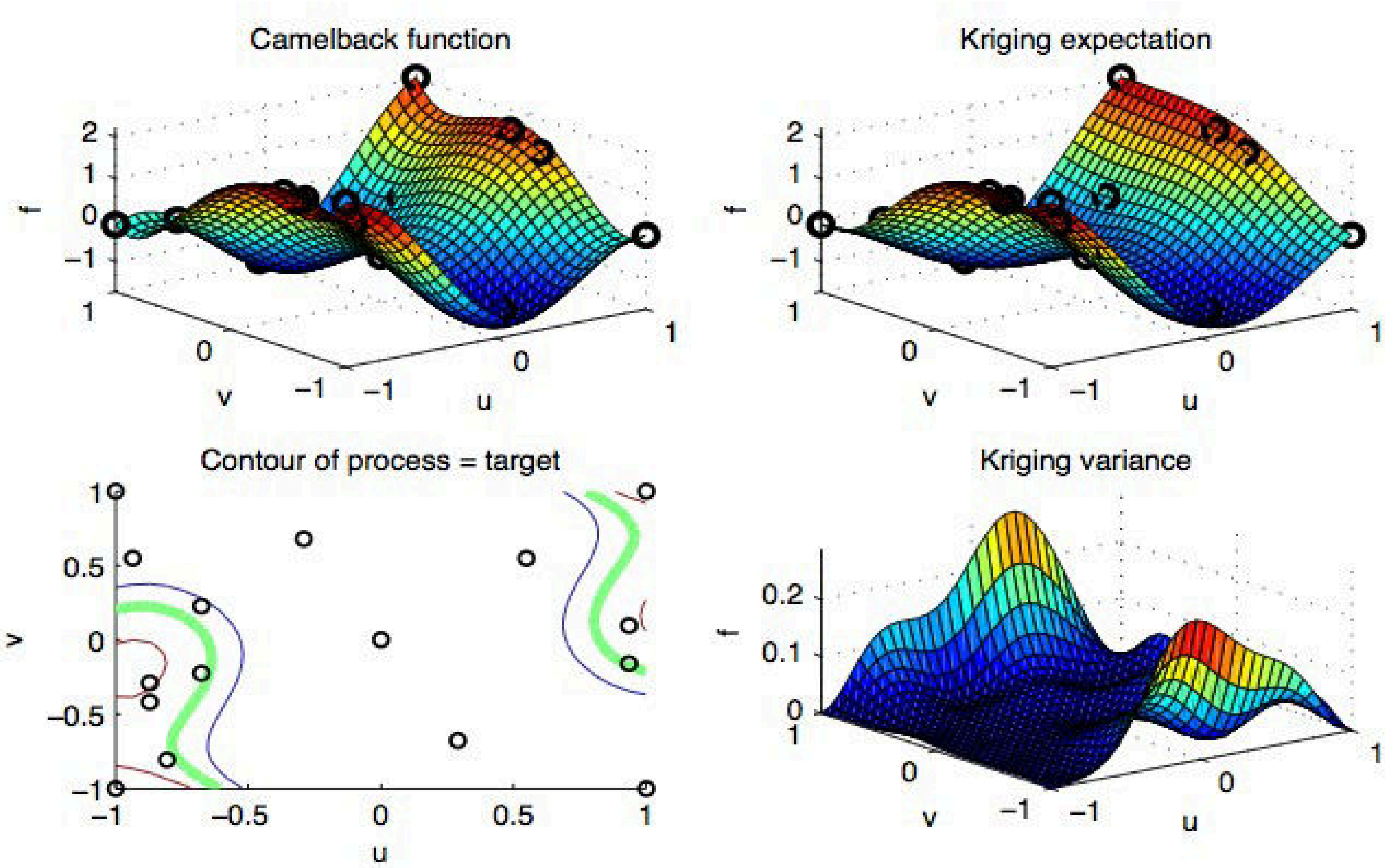
  - Calculate $g(u^*)$, add $(u^*, g(u^*))$ to $(\mathbf{u}, g(\mathbf{u}))$.
- Until max iterations reached

Approximation of a target region – 2D example

A 2D example (Camelback function)

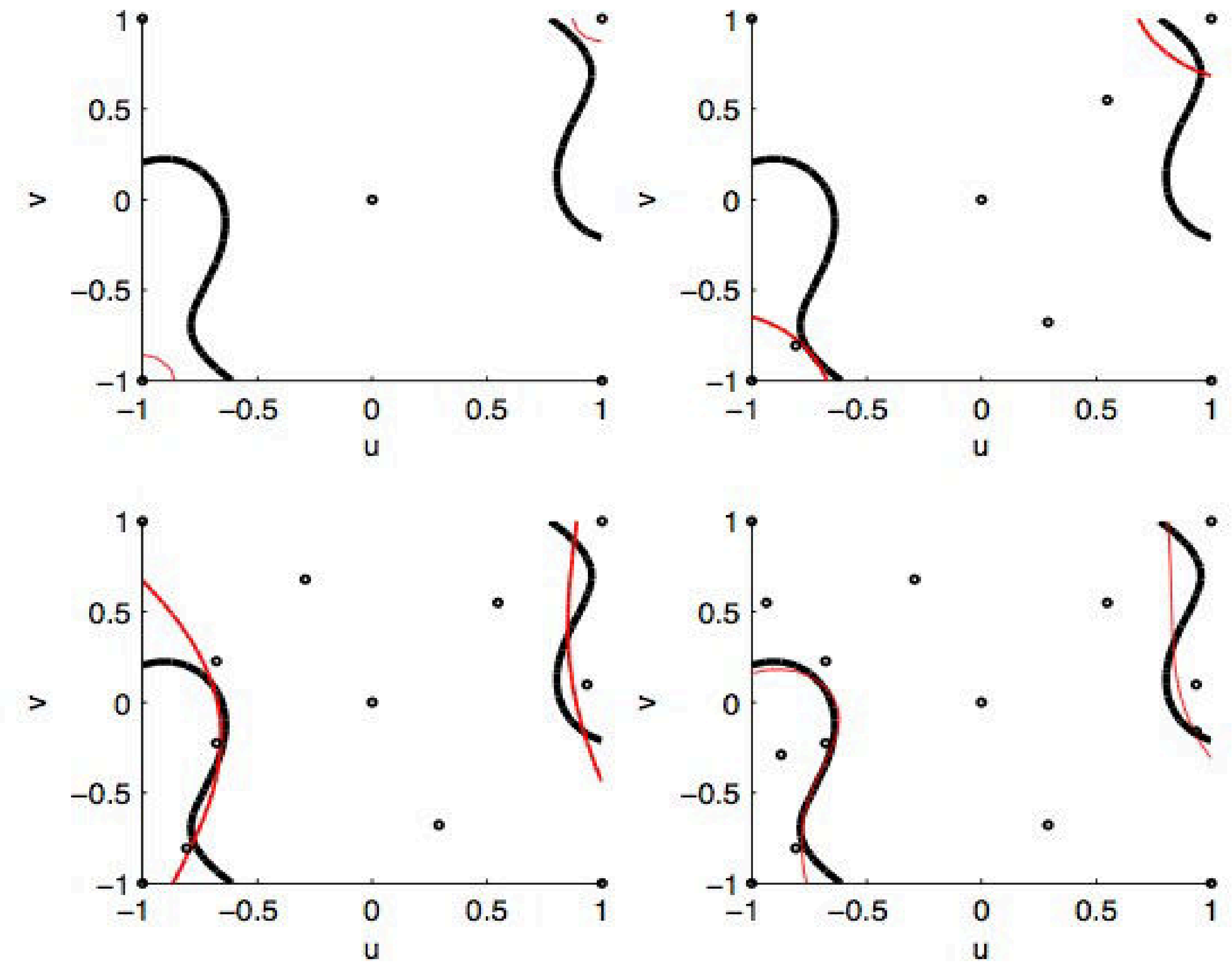
Target region $g(u_1,u_2) = 1.3$

Look at the DoE after 11 iterations



Approximation of a target region – 2D example

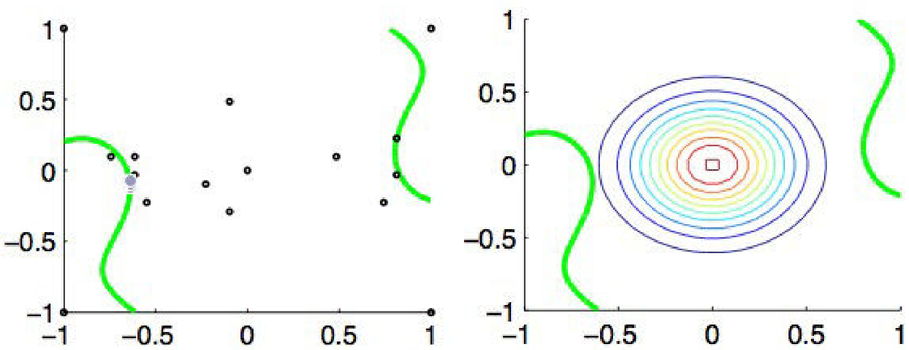
Evolution of kriging target contour lines



Approximation of a target region – 2D example

Application to the estimation of a failure probability :

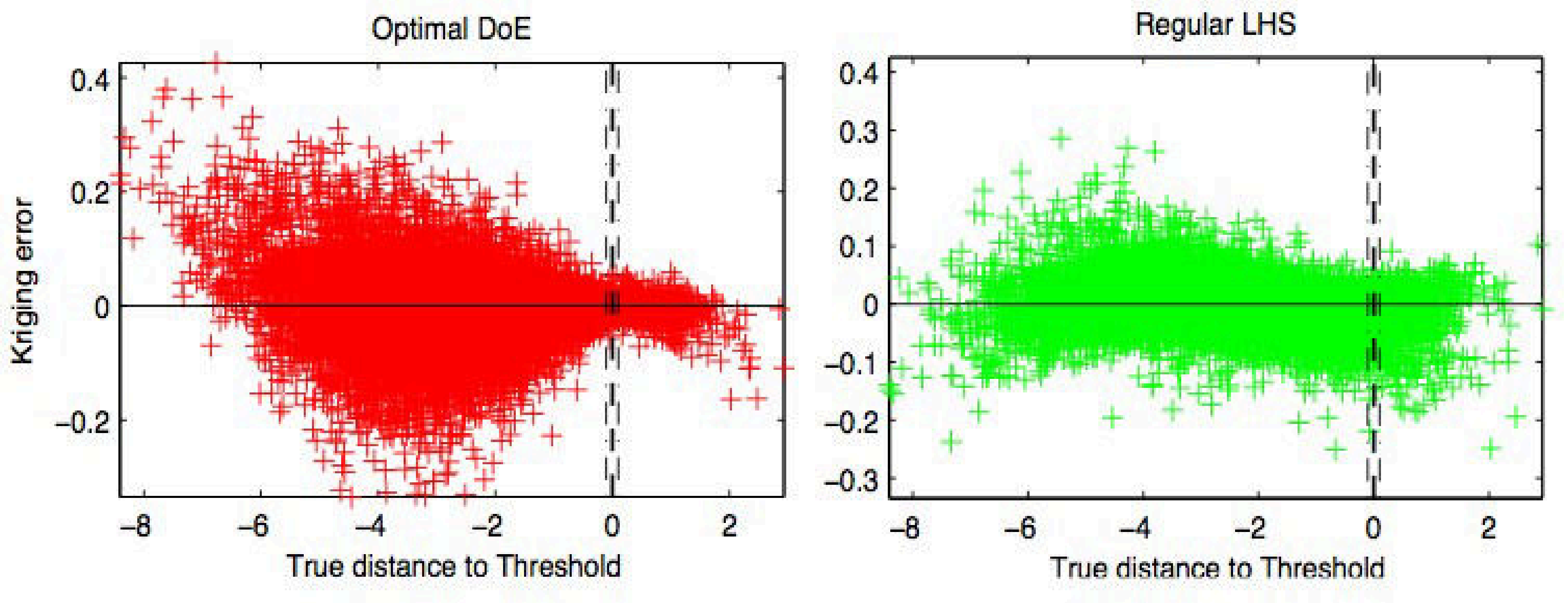
$$Prob\left(g\left(u_1, u_2\right) > 1.3\right) \quad \text{with} \quad u_1, u_2 \text{ i.i.d. } N\left(0, 0.028^2\right)$$



DoE	Full Factorial	Optimal without input distribution	Optimal with input distribution	Probability estimate based on 10^7 MCS
Probability of failure (%)	0.17	0.70	0.77	0.75
Relative error	77 %	7 %	3 %	

Approximation of a target region – other results

A 6D example (GP sample with linear trend and Gaussian covariance)



Good results in a numerical comparison of 4 methods (Ling Li, UCM 2010, Sheffield) along with Stepwise Uncertainty Reduction method.

2nd part of the talk

1. Estimation of reliability constraints

Failure probability : $P_f = Prob(\, g(U) > T\,)$

2. Optimization with uncertainties

Average minimization : $\min_{x \in S \subset \mathbb{R}^n} E_U(f(x,U))$

2.1. Preamble : expected improvement, ...
2.2. Estimation of the average
2.3. Simultaneous optimization and sampling

Expected Improvement criterion (1)

A sampling criterion for global optimization without noise :

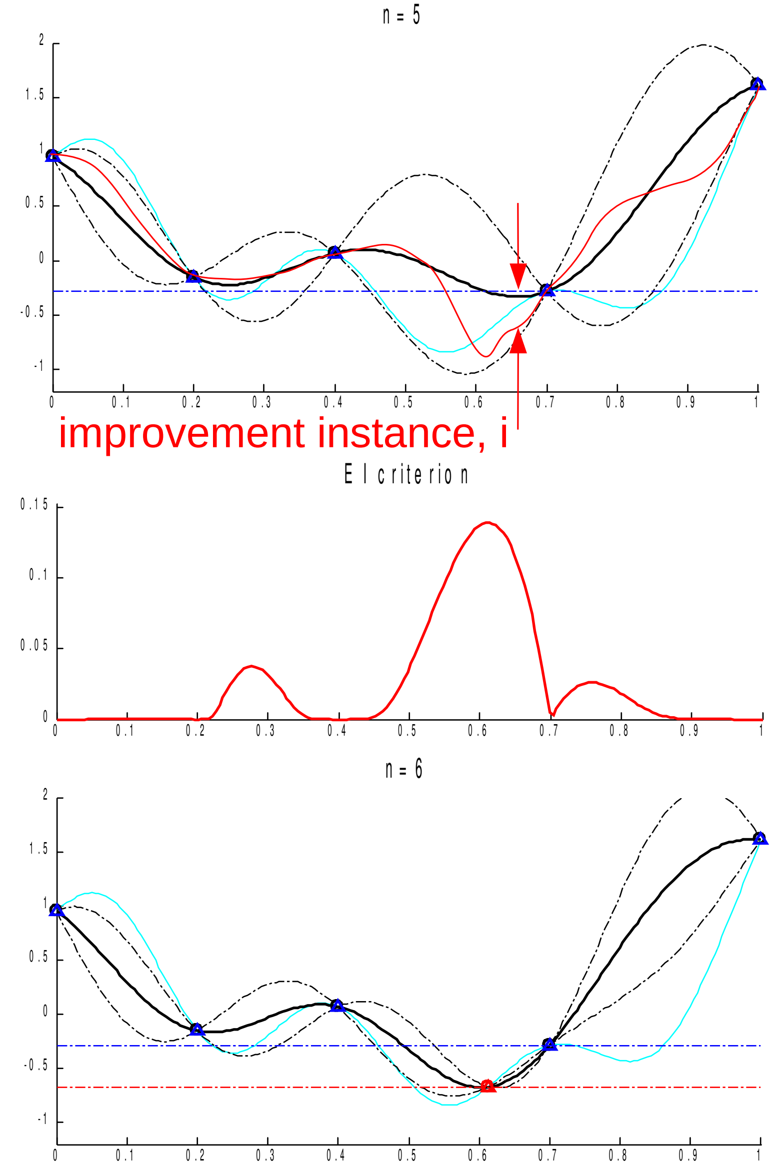
Improvement at x , $I(x)=max(y_{min}-Y(x),0)$

The expected improvement, *EI(x)* , can be analytically calculated.

$$EI(x) = s(x)\left[a(x)\Phi(a(x)) + \phi(a(x)) \right] ,$$

$$a(x)=\frac{y_{min}-m_K(x)}{s_K(x)}$$

EI increases when m_K decreases and when s_K increases. *EI(x)* quantifies the exploration-exploitation compromise of global optimization.



Expected Improvement criterion (2)

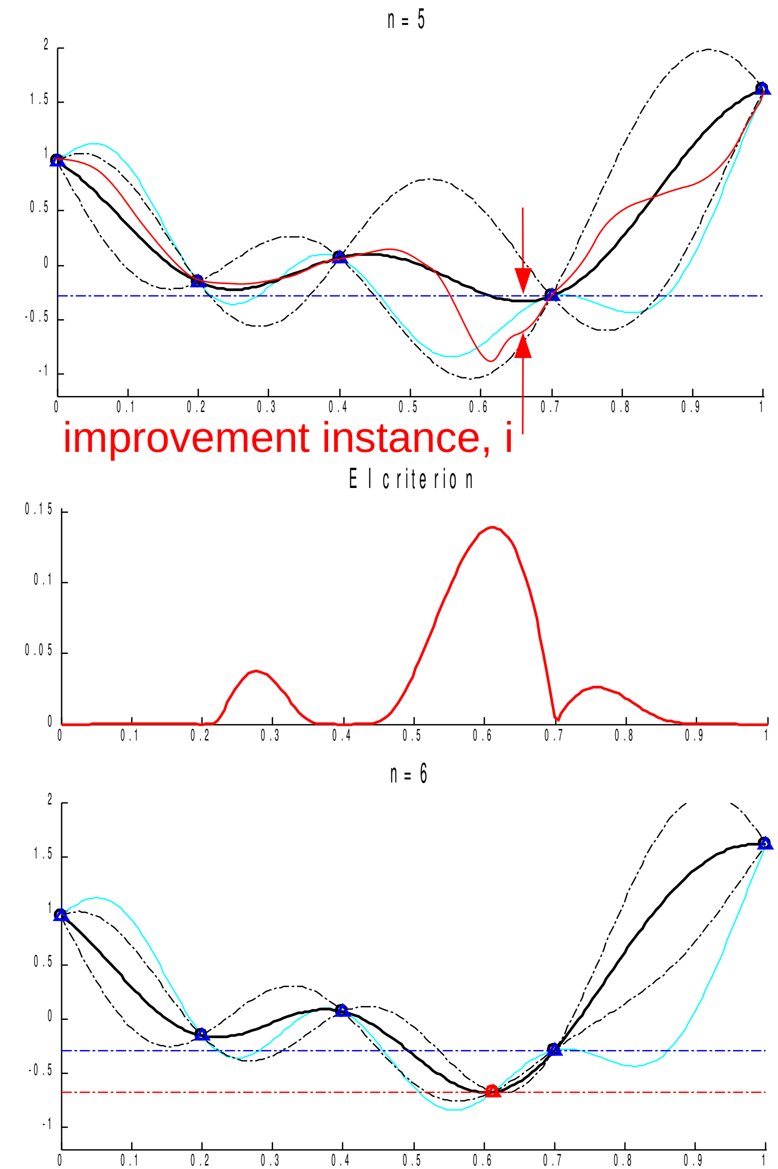
EGO algorithm (Efficient Global Optimization), D. Jones, 1998 :

while computation budget not exhausted

Next iterate : $x^{t+1} = max_x EI(x)$

Update kriging with x^{t+1}

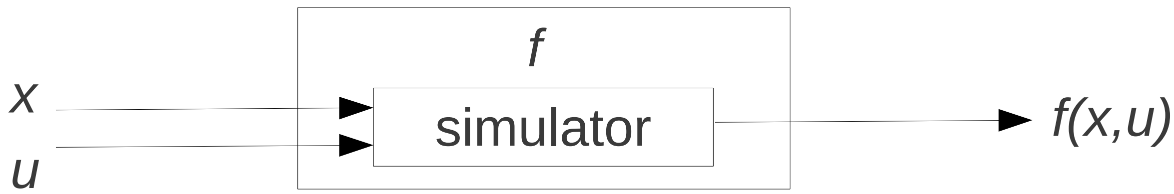
(cannot be applied directly to noisy functions)



2.1. kriging-based average minimization : preamble

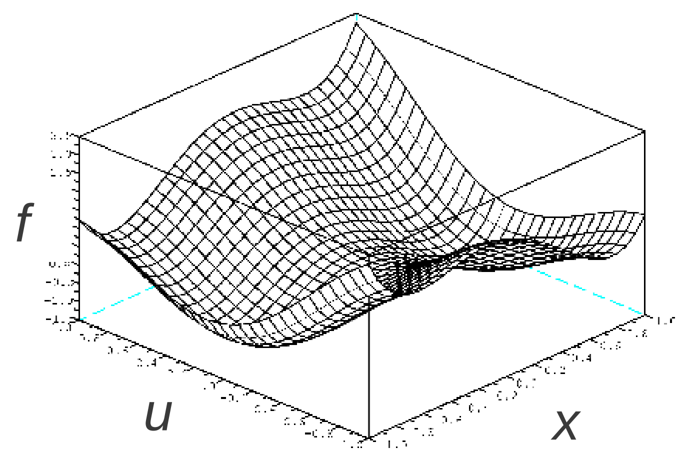
the direct approach to average minimization

x and *u* can be chosen before calling the simulator and calculating the objective function. This is the general case.



Direct approaches to optimization with uncertainties have a double loop : propagate uncertainties on *U*, optimize on *x*.

- Optimization : loop on *x*
  - Estimation of the performance (average, std dev, percentile of *f(x,U)*) : loop on *u* , Monte Carlo



Such a double loop is very costly (more than only propagating uncertainties or optimizing, which are already considered as costly) !

bibliography

Based on

J. Janusevskis and R. Le Riche, *Simultaneous kriging-based estimation and optimization of mean response*, J. of Global Optimization, Springer, published online in Jan. 2012

Other related works :

D. Jones, M. Schonlau and W. J. Welch, *Efficient global optimization of expensive functions*, J. of Global Optimization, 1998.

Dubourg, V., Sudret, B. and Bourinet, J.-M., *Reliability-based design optimization using kriging and subset simulation,* Struct. Multidisc. Optim, accepted for publication, 2011.

E. Vazquez, J. Villemonteix, M. Sidorkiewicz and E. Walter, *Global optimization based on noisy evaluations: an empirical study of two statistical approaches*, 6th Int. Conf. on Inverse Problems in Engineering, 2010.

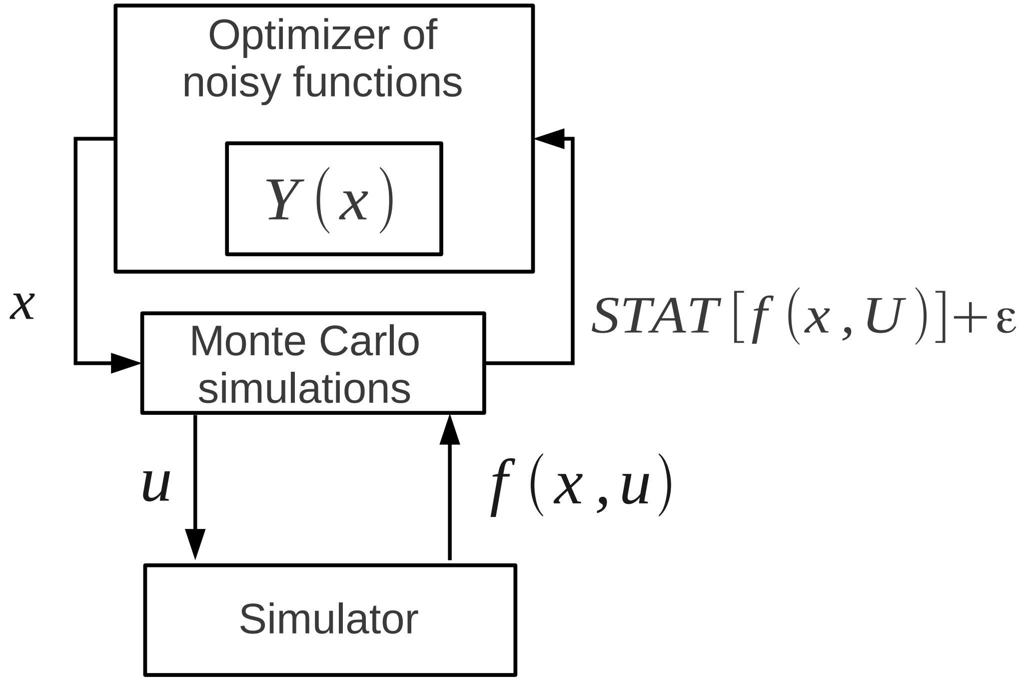
J. Bect, *IAGO for global optimization with noisy evaluations,* workshop on noisy kriging-based optimization (NKO), Bern, 2010.

Avoiding the double loop scenario

Assumptions : *x* and *U* controlled

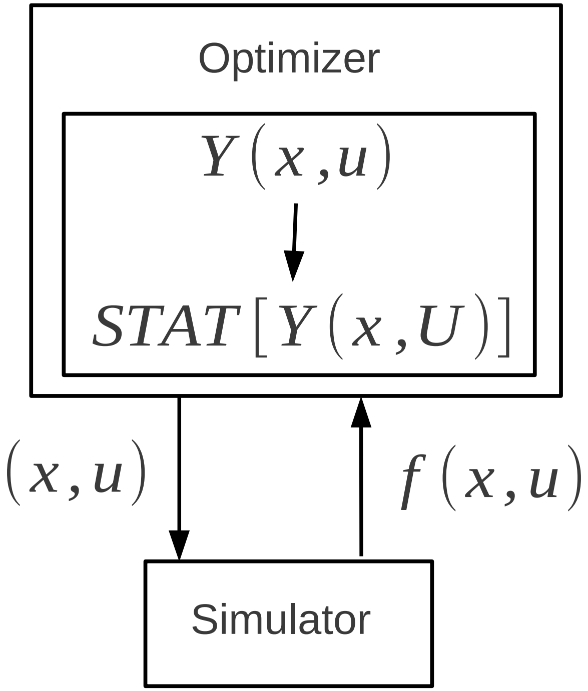
Y : surrogate model

Direct approach



Multiplicative cost of two loops involving *f*

(x,u) surrogate based approach

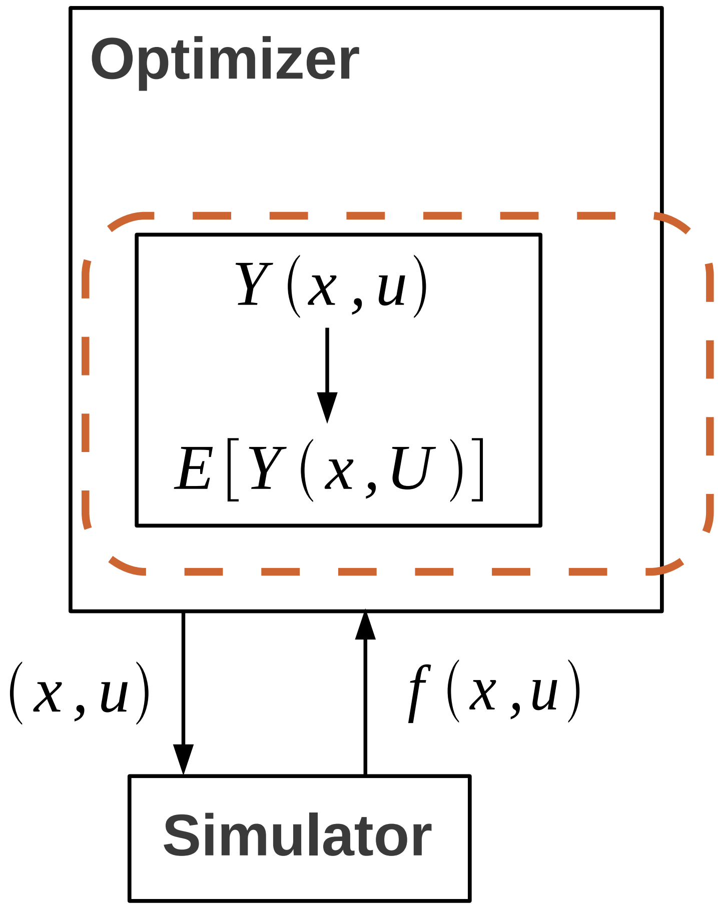


Only one loop of *f*

Estimation of the average

Assumptions : *x* and *U* controlled, *U* normal.

Y : kriging model



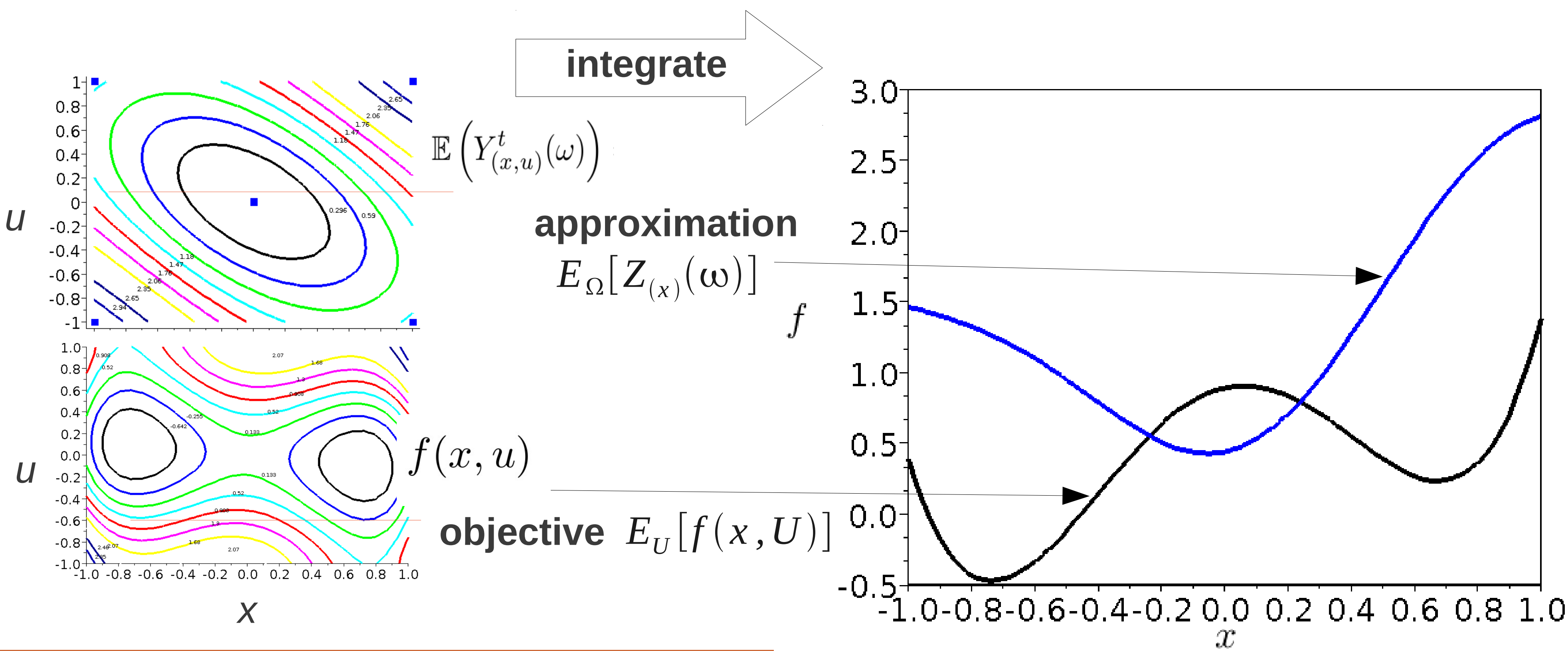
1. Building internal representation of the objective (mean performance) by «integrated» kriging.

Integrated kriging (1)

$\min_x \mathbb{E}_U[f(x,U)]$ **: objective**

$Y^t_{(x,u)}(\omega)$ **: kriging approximation to deterministic** $f(x,u)$

$Z^t_{(x)}(\omega) = \mathbb{E}_U[Y^t_{(x,U)}(\omega)]$ **: integrated process approximation to** $\mathbb{E}_U[f(x,U)]$



Integrated kriging (2)

The integrated process over *U* is defined as

$$Z_{(x)}(\omega) = \mathbb{E}_U[Y^t_{(x,U)}(\omega)] = \int_{\mathbb{R}^m} Y^t_{(x,u)}(\omega) d_\mu(u)$$

$d\mu(u)$ -probability measure on *U*

Because it is a linear transformation of a Gaussian process, it is Gaussian, and fully described by its mean and covariance

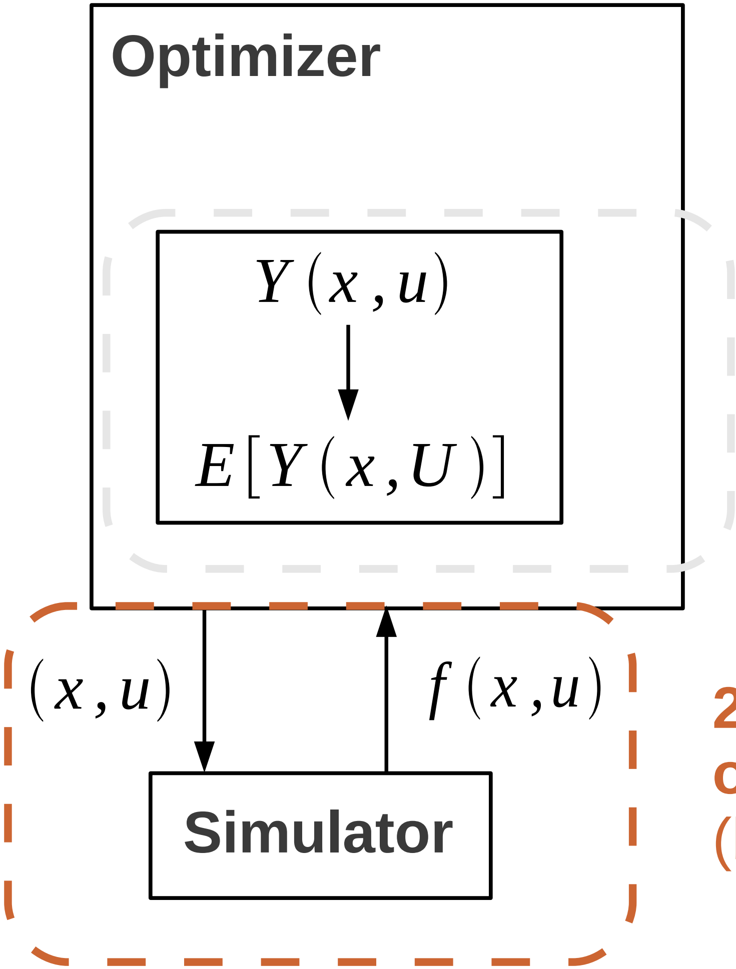
$$m_Z(x) = \int_{\mathbb{R}^m} m_Y(x,u) d\mu(u)$$

$$cov_Z(x;x') = \int_{\mathbb{R}^m}\int_{\mathbb{R}^m} cov_Y(x,u;x'u') d\mu(u) d\mu(u')$$

Analytical expressions of m_z and cov_z for Gaussian *U*'s are given in

J. Janusevskis, R. Le Riche. Simultaneous kriging-based sampling for optimization and uncertainty propagation, HAL report: hal-00506957

Simultaneous optimization and sampling (2)



1. Building internal representation of the objective (mean performance) by «projected» kriging.

2. Simultaneous sampling and optimization criterion for x and u
(both needed by the simulator to calculate f)

EI on the integrated process (1)

Z is a process approximating the objective function $\mathbb{E}_U[f(x,U)]$

Optimize with an Expected Improvement criterion,

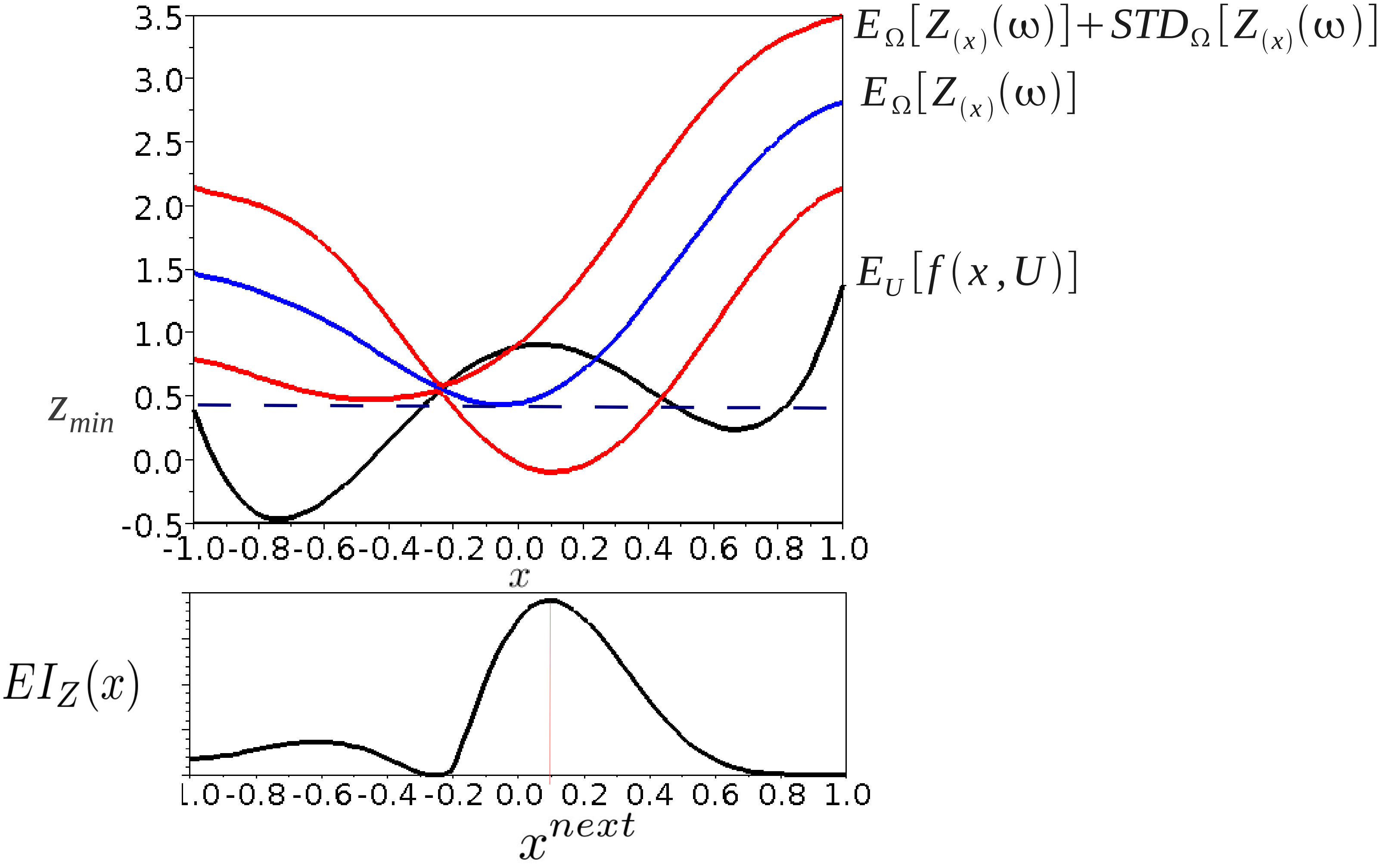
$$x^{next} = \arg\max_x EI_Z(x)$$

where,

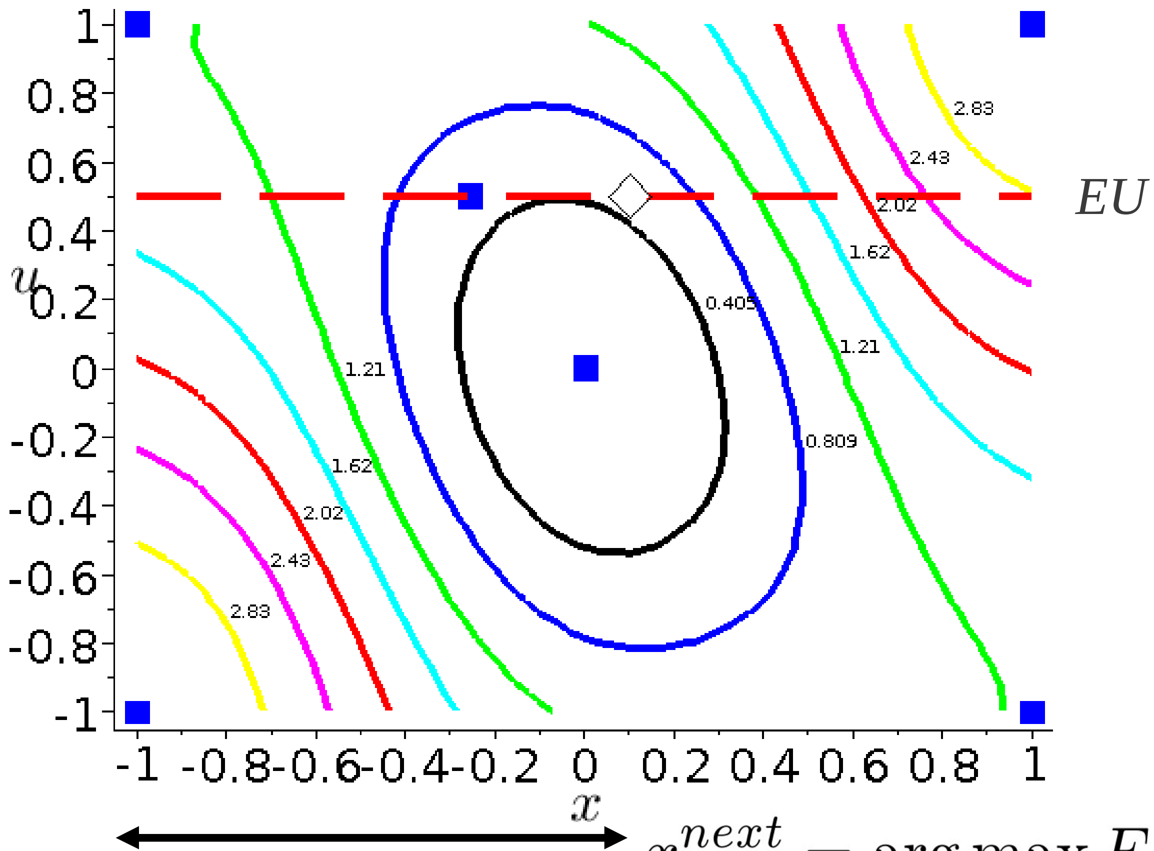
$I_Z(x) = max(z_{min} - Z(x), 0)$, but z_{min} not observed (in integrated space).

$\Rightarrow$ Define $z_{min} = \min_{x^1, \dots, x^t} E(Z(x))$

EI on the integrated process (2)



EI on the integrated process (3)



$$x^{next} = \arg\max_x EI_Z^t(x)$$

x ok. What about *u* ? (which we need to call the simulator)

Method

x^{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 → one choice of (x,u).

Choose (x^{t+1},u^{t+1}) that provides the most information, i.e., which minimizes the variance of the integrated process at x^{next}

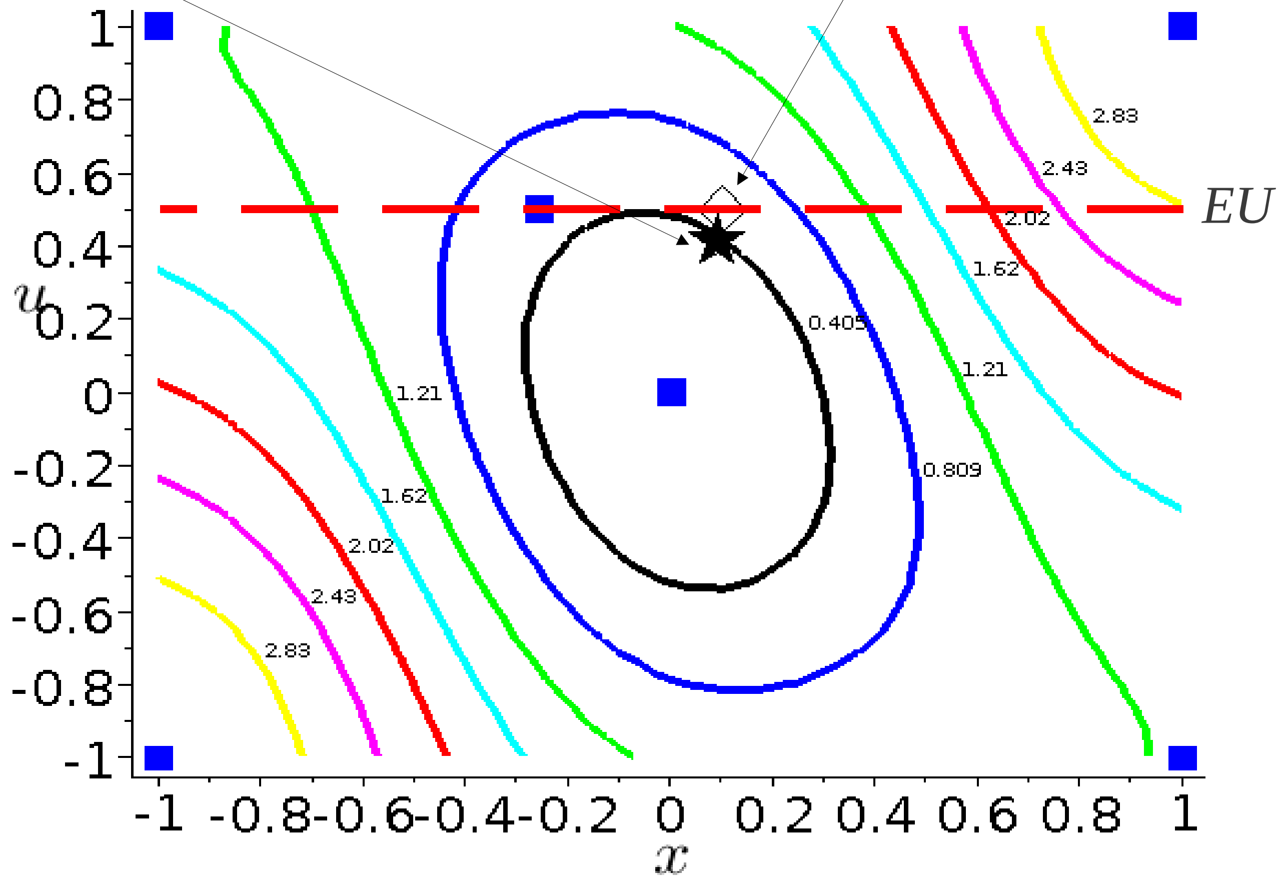
$$(x^{t+1}, u^{t+1}) = \arg\min_{x,u} \mathbb{VAR}\ Z^{t+1}_{(x^{next})}(\omega)$$

(no calculation details, cf. article. Note that VAR of a Gaussian process does not depend on *f* values but only on *x*'s).

Illustration

$$x^{next} = \arg\max_{x} EI_Z^t(x)$$

$$(x^{t+1}, u^{t+1}) = \arg\min_{x,u} \mathbb{VAR}\ Z^{t+1}_{(x^{next})}(\omega)$$



Algorithm

Create initial DOE in (x,u) space;

While stopping criterion is not met:

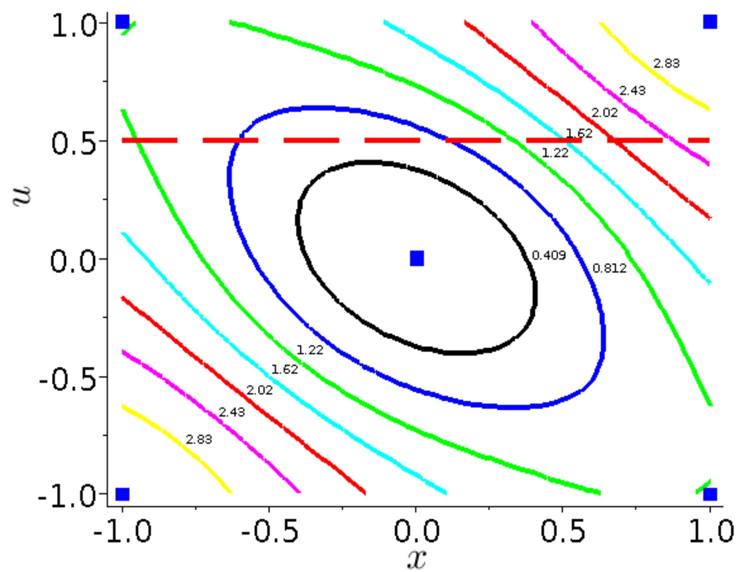
- **Create kriging approximation *Y* in the joint space** (x,u)
- **Calculate the mean and covariance of *Z* from those of *Y***
- **Minimize EI of Z to choose** (x^{next})
- **Minimize** $VAR(Z(x^{next}))$ **to obtain the next point** (x^{t+1}, u^{t+1}) **for simulation**
- **Calculate simulator response at the next point** $f(x^{t+1}, u^{t+1})$

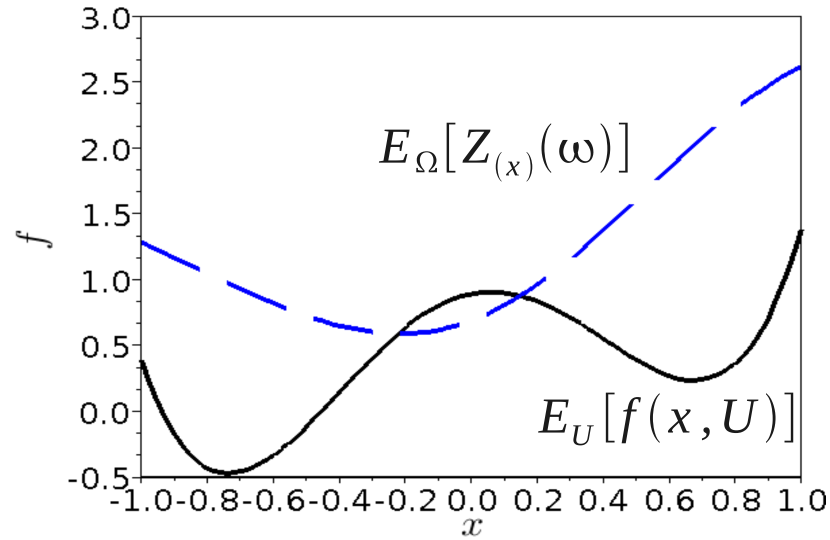
(4 sub-optimizations, solved with CMA-ES)

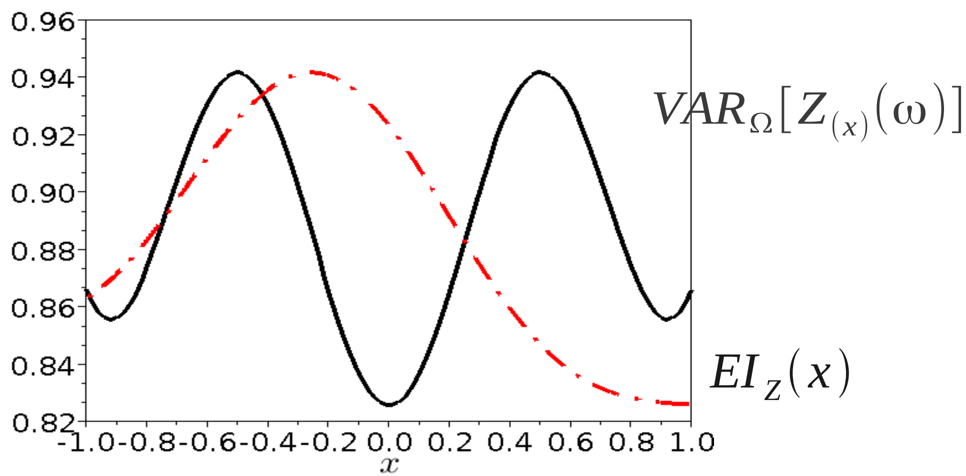
2.3. kriging-based average minimization : simult. opt. and sampling

2D Example

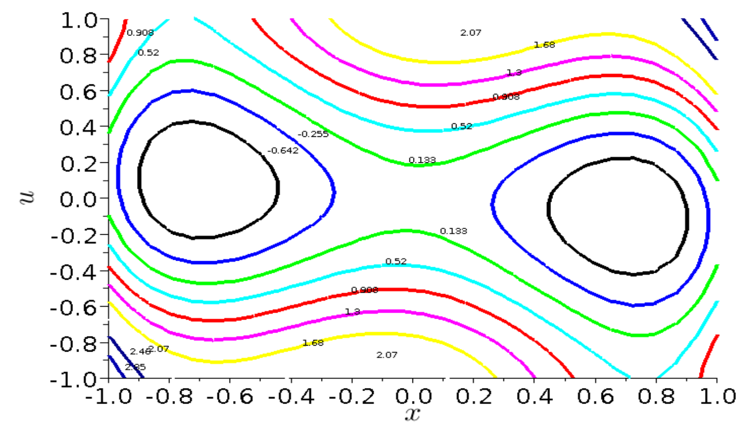
DOE and $E[Y(x,u)]$



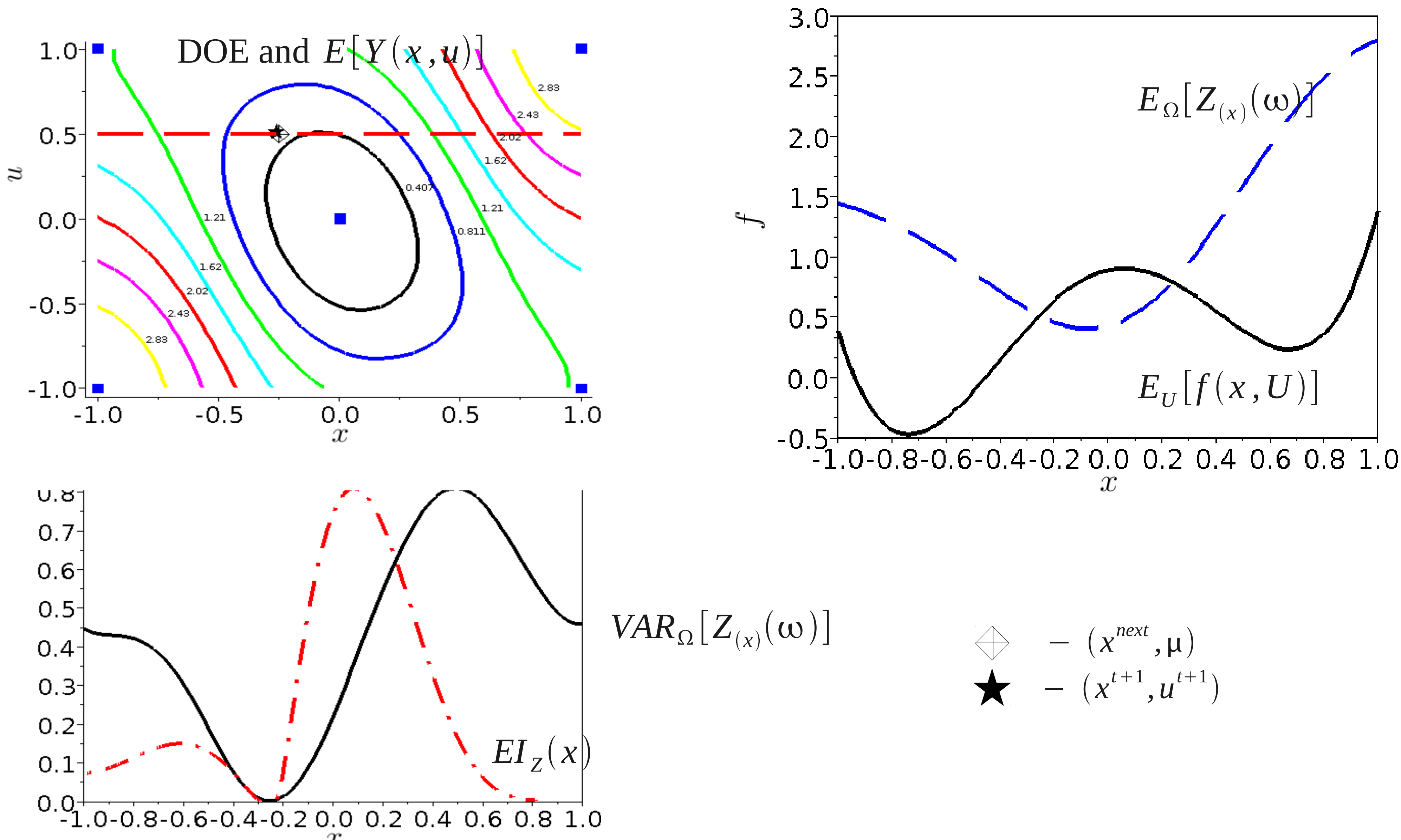




test function

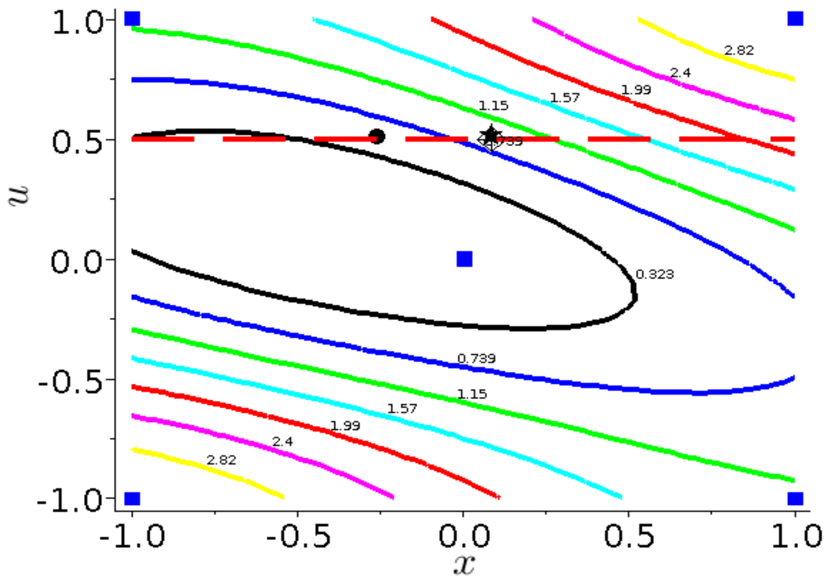


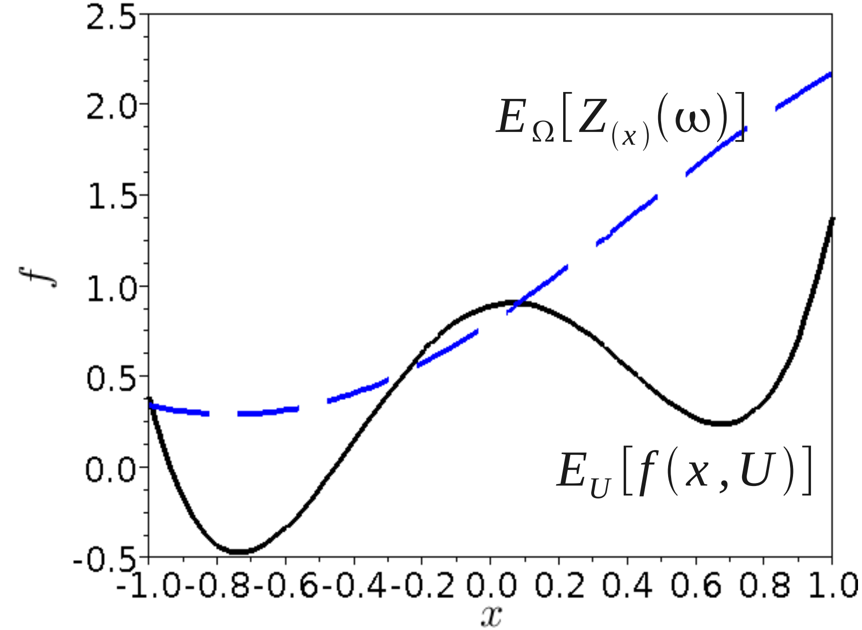
2.3. kriging-based average minimization : simult. opt. and sampling
1st iteration

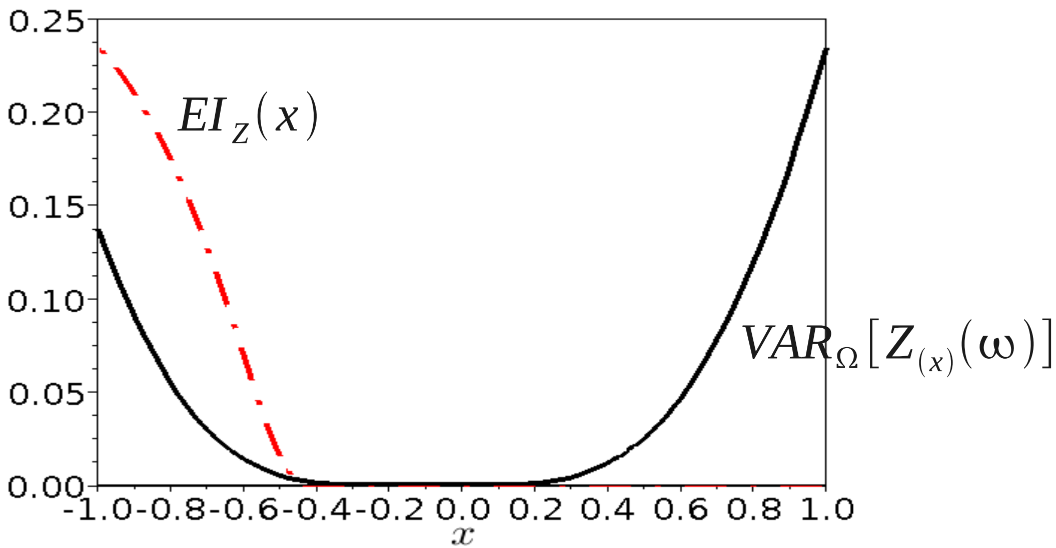


2nd iteration

DOE and $E[Y(x,u)]$





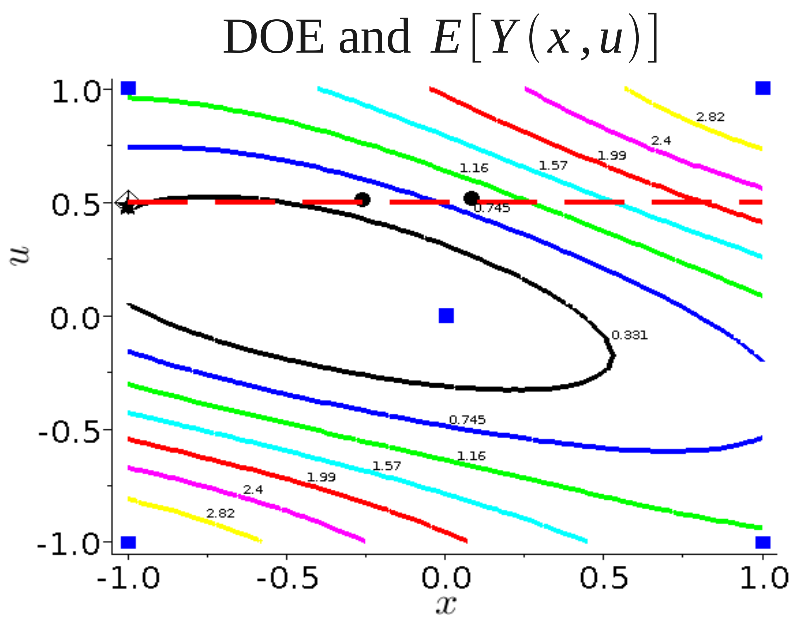


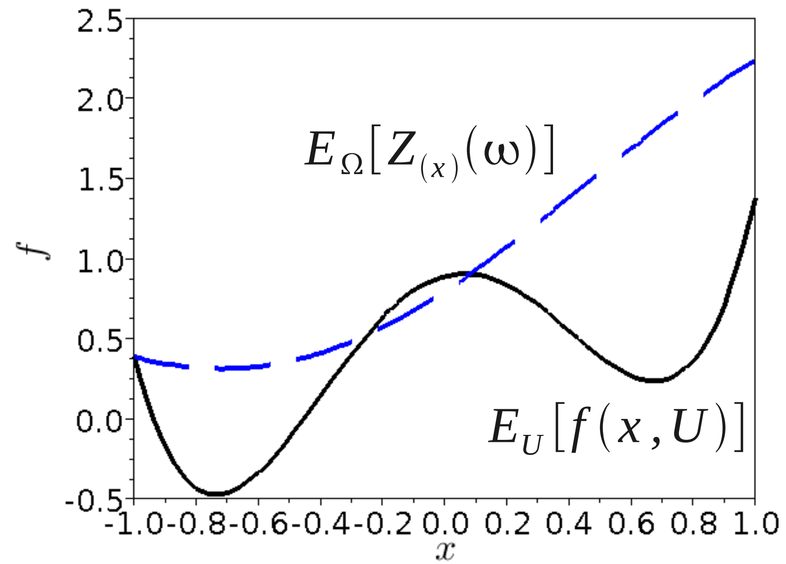
◇ – (x^{next}, μ)

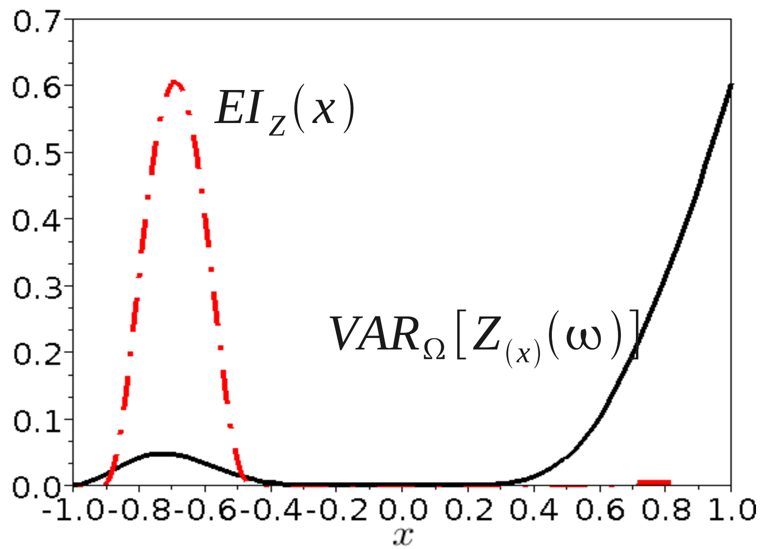
★ – (x^{t+1}, u^{t+1})

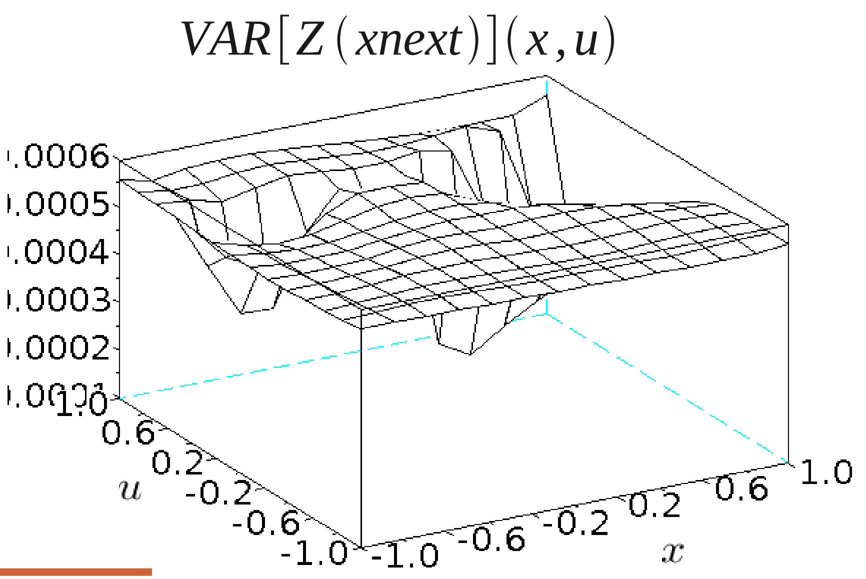
2.3. kriging-based average minimization : simult. opt. and sampling

3rd iteration



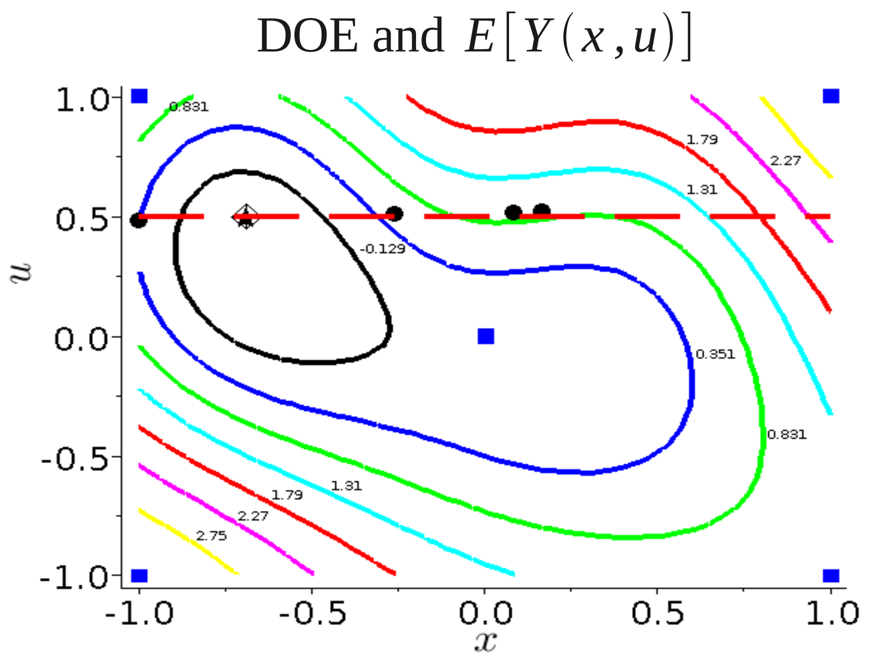


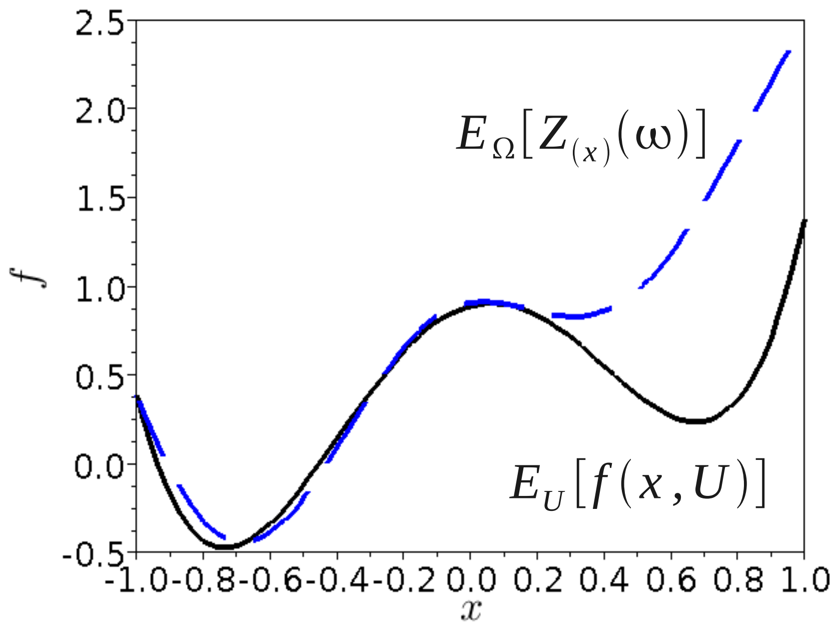


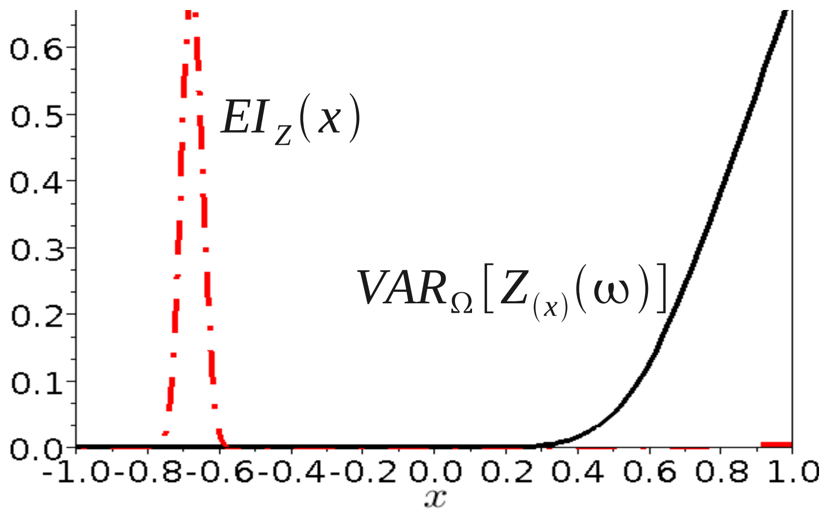


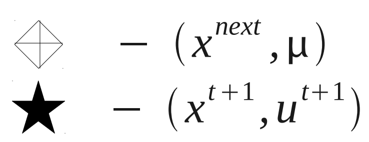
2.3. kriging-based average minimization : simult. opt. and sampling

5th iteration



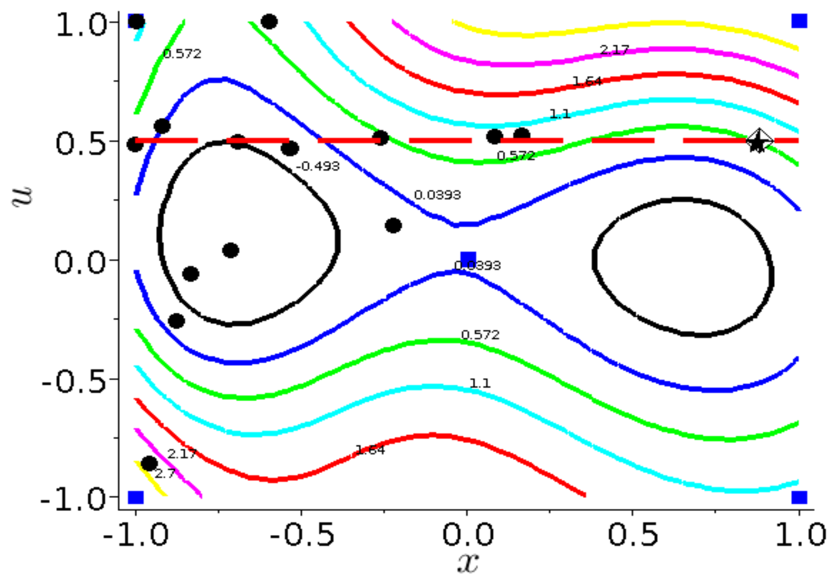


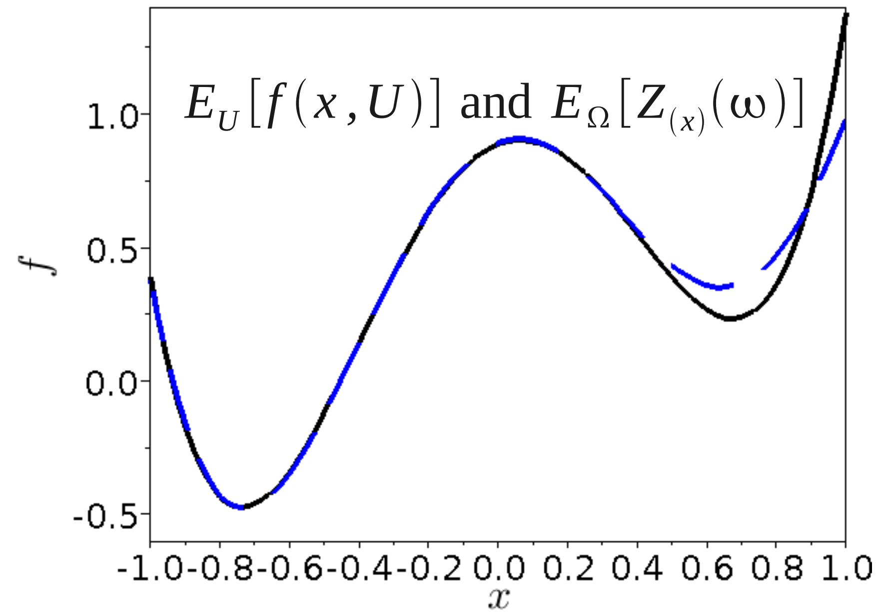


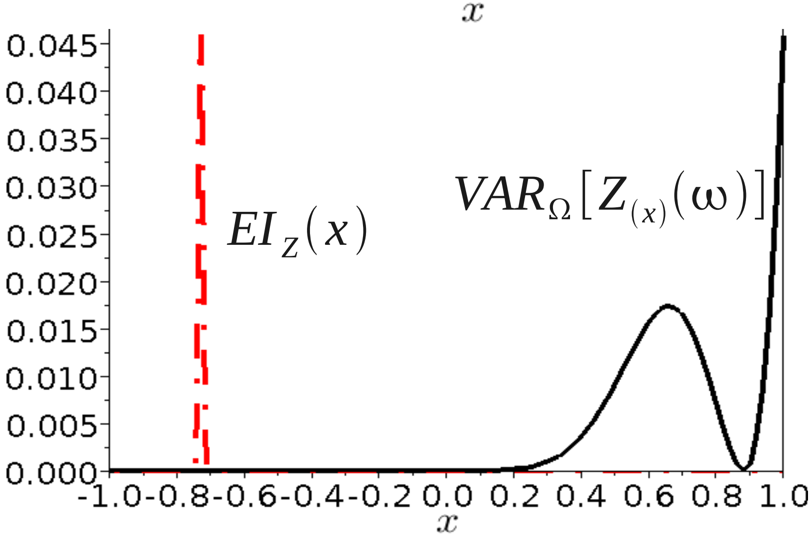


17th iteration

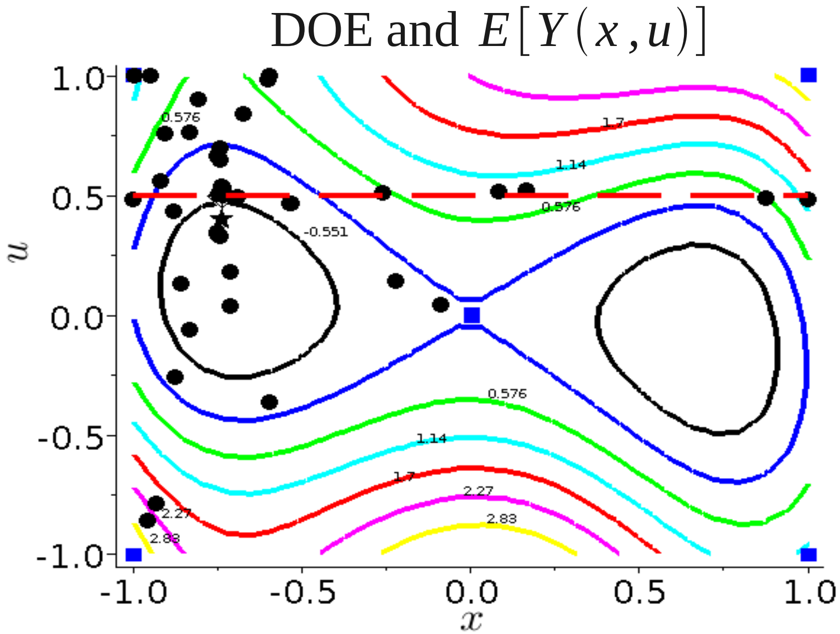
DOE and $E[Y(x,u)]$

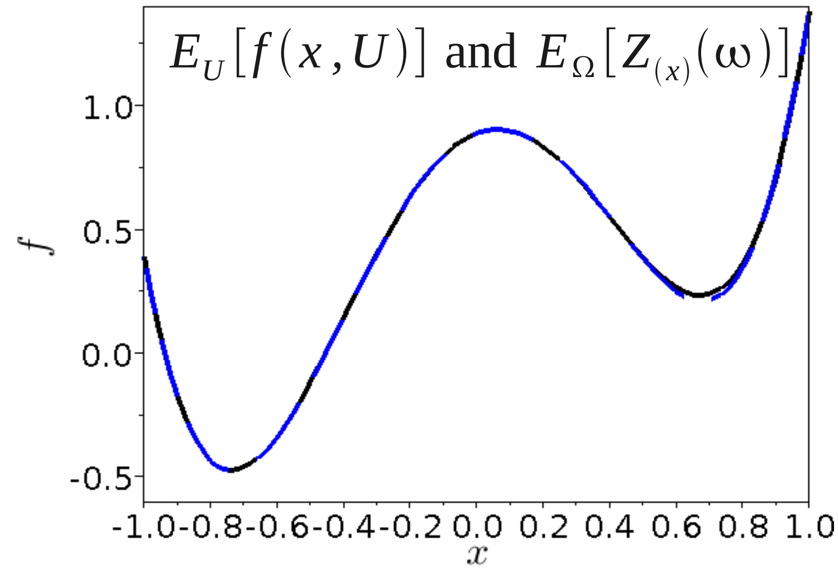


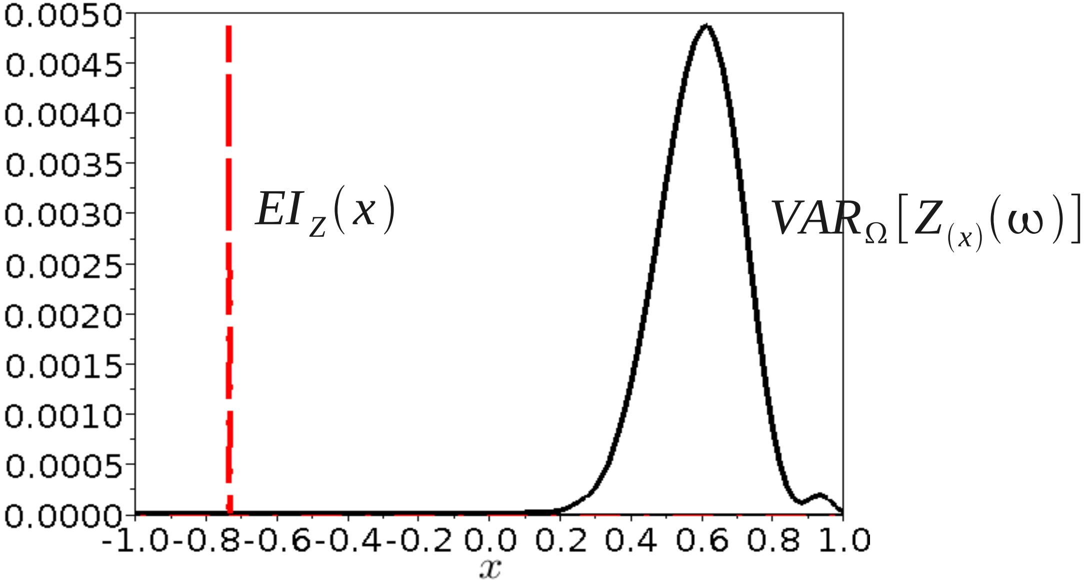




50th iteration







Comparison tests

Compare « simultaneous opt and sampling » method to

1. A direct MC based approach :
   EGO based on MC simulations in *f* with fixed number of runs, *s*.
   Kriging with homogenous nugget to filter noise.

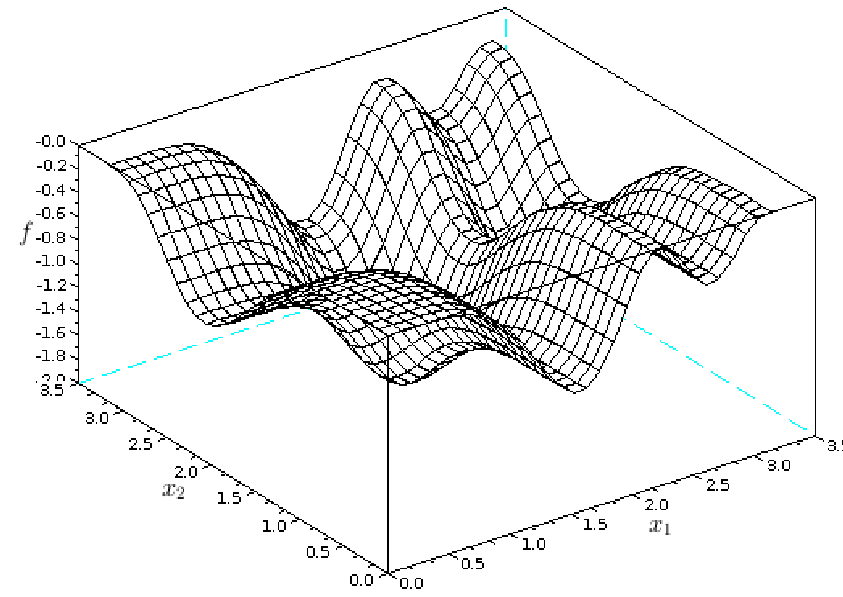
2. An MC-surrogate based approach :
   the MC-kriging algorithm.

Test functions

Test cases based on Michalewicz function

$$f(x) = -\sum_{i=1}^{n} \sin(x_i)\left[\sin\left(i x_i^2/\pi\right)\right]^2$$

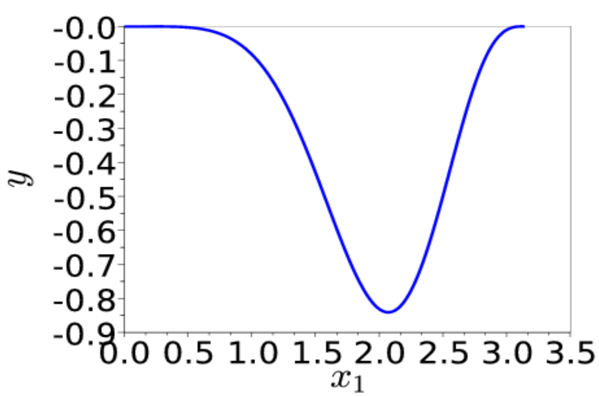
$$f(x,u) = f(x) + f(u)$$

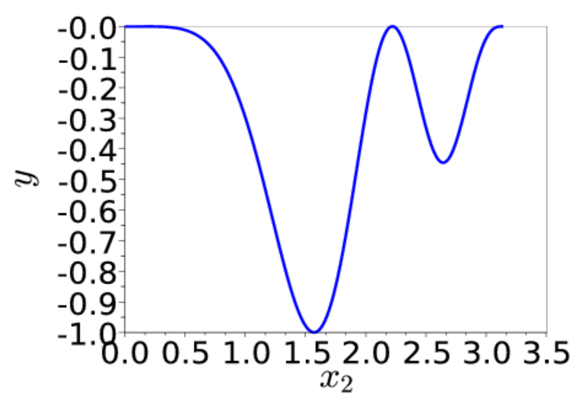


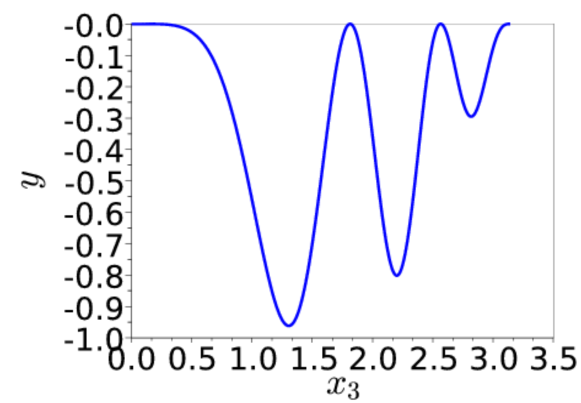
2D: $n_x=1 \quad n_u=1 \quad \mu=1.5 \quad \sigma=0.2$

4D: $n_x=2 \quad n_u=2 \quad \mu=[1.5,\ 2.1] \quad \sigma=[0.2,\ 0.2]$

6D: $n_x=3 \quad n_u=3 \quad \mu=[1.5,\ 2.1,\ 2] \quad \sigma=[0.2,\ 0.2,\ 0.3]$

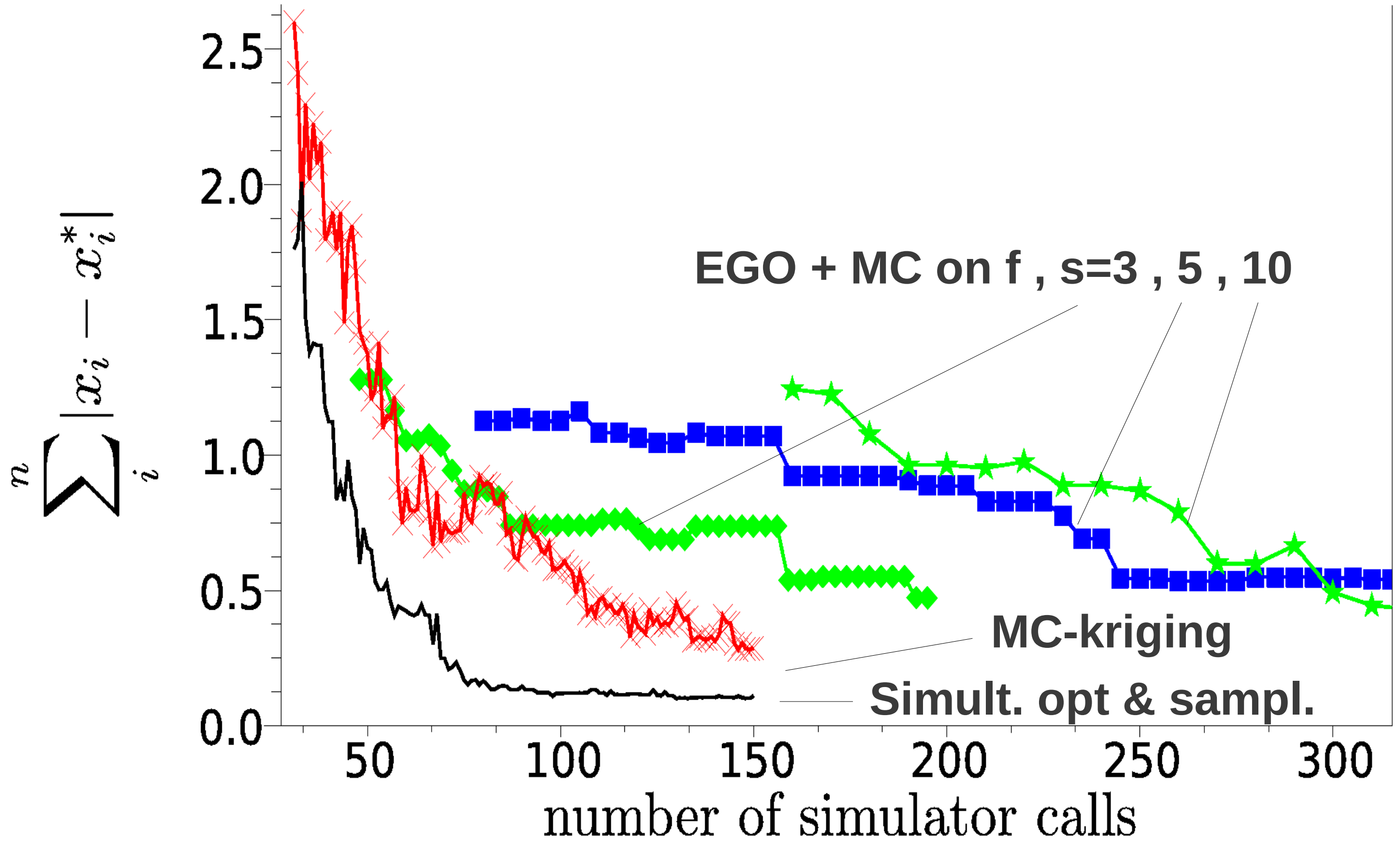






Test results

6D Michalewicz test case, $n_x = 3$, $n_U = 3$.
Initial DOE: RLHS , $m=(n_x+n_U)*5 = (3+3)*5 = 30$;
10 runs for every method.



Sequential kriging-based methods for reliability estimation and optimization : concluding remarks

Kriging provides a rich framework to define iterative sampling strategies for reliability estimation and optimization.

Of course, still a long way to go :

Small step ; properly handle the « ideal formulation » (quantiles and coupled reliability constraints).

Large step ; high dimensions, large number of points, what is the effect of kriging parameters, instationary kriging.