Uncertainty functionals and the greedy reduction of uncertainty

Julien Bect [a]

Joint work with François Bachoc [b] and David Ginsbourger [c,d]

(a) L2S (Gif-sur-Yvette, France)
(b) IMT (Toulouse, France)
(c) Idiap (Martigny, Switzerland)
(d) IMSV (Bern, Switzerland)

Workshop on "Design of Experiments: New Challenges"
CIRM, Luminy, Marseille, May 2018

My co-authors

This talk is based on our arXiv:1608.01118 paper:

A supermartingale approach to Gaussian process based sequential design of experiments.

(v2 arXiv-ed last July; v3 soon, email me if you're interested)

Early example I: QUEST and variants

This picture comes from Watson & Pelli (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, 33, 113–120.

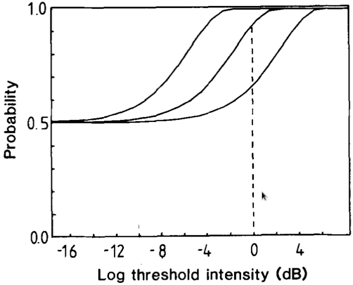
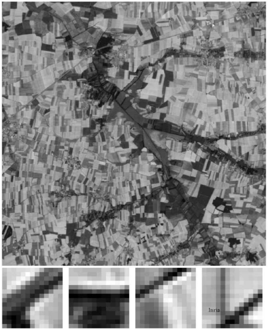


Figure 1. Three examples of a Weibull psychometric function. Center curve is the canonical form, flanking curves are for thresholds of −4 and 4 dB. The parameters of all functions are: $\beta = 3.5$, $\gamma = .5$, and $\delta = .01$.

King-Smith (1984), Pelli (1987): tbmk, first examples of (parametric) SUR

Early example II: active testing

Geman & Jedynak, *Shape recognition and Twenty Questions*, INRIA RR-2155, 1993



Method named "entropy strategy", "entropy testing", "stepwise entropy reduction" and later "stepwise uncertainty reduction" (Fleuret & G., 1999)

The origins of SUR ($\leq$ 2009): more refs / less pictures

- Versions of the idea have appeared in various places in the 80's / 90's
  - psychometry: King-Smith (1984), Pelli (1987)
  - Geman and co-authors: shape recognition, image retrieval, etc.
  - Active learning: MacKay (1992), Cohn et al. (1996), ...

The origins of SUR ($\leq$ 2009): more refs / less pictures

- Versions of the idea have appeared in various places in the 80's / 90's
  - psychometry: King-Smith (1984), Pelli (1987)
  - Geman and co-authors: shape recognition, image retrieval, etc.
  - Active learning: MacKay (1992), Cohn et al. (1996), ...
- For the sequential design of numerical experiments, the idea was first proposed by E. Vazquez and co-authors around 2006–2009
  - optimization (IAGO): Villemonteix (2008), Villemonteix et al (2009), ...
  - reliability: Vazquez & Piera-Martinez (2007), Vazquez & B. (2009), ...

The origins of SUR ($\leq$ 2009): more refs / less pictures

- Versions of the idea have appeared in various places in the 80's / 90's
  - psychometry: King-Smith (1984), Pelli (1987)
  - Geman and co-authors: shape recognition, image retrieval, etc.
  - Active learning: MacKay (1992), Cohn et al. (1996), ...
- For the sequential design of numerical experiments, the idea was first proposed by E. Vazquez and co-authors around 2006–2009
  - optimization (IAGO): Villemonteix (2008), Villemonteix et al (2009), ...
  - reliability: Vazquez & Piera-Martinez (2007), Vazquez & B. (2009), ...
- Some methods based on utility maximization *can be seen* as stepwise uncertainty reduction strategies
  - the "Bayesian method for seeking the extremum" (Mockus et al., 1978)
  - the "expected improvement" method (Jones et al, 1998)
  - the "knowledge gradient policy" (Frazier et al, 2008)

Setting: computer experiments / nonparametric regression

- $f : \mathbb{X} \to \mathbb{R}$ is a computer model for
  - a system to be designed (engineering),
  - a physical or biological phenomenon,
  - ...
- "Computer experiment"
  - 1 experiment = run the program for some $x \in \mathbb{X}$
  - assumed to be time-consuming

Setting: computer experiments / nonparametric regression

- $f : \mathbb{X} \to \mathbb{R}$ is a computer model for
  - a system to be designed (engineering),
  - a physical or biological phenomenon,
  - ...
- "Computer experiment"
  - 1 experiment = run the program for some $x \in \mathbb{X}$
  - assumed to be time-consuming
- Observation model in this work:

$$Z_i = f(X_i) + \varepsilon_i$$

with $\varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \tau(X_i))$. Special case $\tau \equiv 0$ allowed.

General principle of SUR strategies (1/2)

1. Adopt a Bayesian framework: choose a prior
   - Suppose $f \in \mathbb{S} = \mathcal{C}(\mathbb{X})$, with $\mathbb{X}$ compact metric, say, $\mathbb{X} = [0;1]^d$
   - Prior $=$ probability distribution on $\mathbb{S}$

General principle of SUR strategies (1/2)

1. Adopt a Bayesian framework: choose a prior
    - Suppose $f \in \mathbb{S} = \mathcal{C}(\mathbb{X})$, with $\mathbb{X}$ compact metric, say, $\mathbb{X} = [0;1]^d$
    - Prior = probability distribution on $\mathbb{S}$
    - unknown function $f \rightarrow$ stochastic process ξ

General principle of SUR strategies (1/2)

1. Adopt a Bayesian framework: choose a prior
   - Suppose $f \in \mathbb{S} = \mathcal{C}(\mathbb{X})$, with $\mathbb{X}$ compact metric, say, $\mathbb{X} = [0;1]^d$
   - Prior = probability distribution on $\mathbb{S}$
   - unknown function $f \rightarrow$ stochastic process ξ

2. Choose a measure of uncertainty about the QoI:

$$H_n = \mathcal{H}(\mathrm{P}_n^\xi)$$

   where

   - P_n^ξ is the posterior of ξ given $\mathcal{F}_n = \sigma(X_1, Z_1, \ldots, X_n, Z_n)$

General principle of SUR strategies (1/2)

1. Adopt a Bayesian framework: choose a prior
   - Suppose $f \in \mathbb{S} = \mathcal{C}(\mathbb{X})$, with $\mathbb{X}$ compact metric, say, $\mathbb{X} = [0;1]^d$
   - Prior = probability distribution on $\mathbb{S}$
   - unknown function $f \rightarrow$ stochastic process ξ

2. Choose a measure of uncertainty about the QoI:

$$H_n = \mathcal{H}(\mathrm{P}_n^{\xi})$$

where

- P_n^{ξ} is the posterior of ξ given $\mathcal{F}_n = \sigma(X_1, Z_1, \ldots, X_n, Z_n)$
- $\mathcal{H} : \mathbb{M} \rightarrow [0, +\infty)$ is our uncertainty functional

General principle of SUR strategies (1/2)

1. Adopt a Bayesian framework: choose a prior
   - Suppose $f \in \mathbb{S} = \mathcal{C}(\mathbb{X})$, with $\mathbb{X}$ compact metric, say, $\mathbb{X} = [0;1]^d$
   - Prior = probability distribution on $\mathbb{S}$
   - unknown function $f \to$ stochastic process ξ

2. Choose a measure of uncertainty about the QoI:

$$H_n = \mathcal{H}(\mathrm{P}_n^\xi)$$

where

- P_n^ξ is the posterior of ξ given $\mathcal{F}_n = \sigma(X_1, Z_1, \ldots, X_n, Z_n)$
- $\mathcal{H} : \mathbb{M} \to [0, +\infty)$ is our uncertainty functional
- $\mathbb{M}$ is a set of measures on $\mathbb{S}$ that contains P_0^ξ, P_1^ξ, …

General principle of SUR strategies (2/2)

3. Select design points in a greedy ("myopic", or "one-step look-ahead") way to minimize, at each step, the expected uncertainty at the next step.

General principle of SUR strategies (2/2)

3. Select design points in a greedy ("myopic", or "one-step look-ahead") way to minimize, at each step, the expected uncertainty at the next step.

   - Deduce a SUR sampling criterion:

$$J_n(x) = \mathrm{E}_{n,x}\left(H_{n+1}\right)$$

where $\mathrm{E}_n = \mathrm{E}\left(\cdot \mid \mathcal{F}_n\right)$ with $X_{n+1} = x$.

General principle of SUR strategies (2/2)

3. Select design points in a greedy ("myopic", or "one-step look-ahead") way to minimize, at each step, the expected uncertainty at the next step.

   - Deduce a SUR sampling criterion:

     $$J_n(x) = \mathsf{E}_{n,x}\left(H_{n+1}\right)$$

     where $\mathsf{E}_n = \mathsf{E}\left(\cdot \mid \mathcal{F}_n\right)$ with $X_{n+1} = x$.

   - Define the corresponding SUR strategy (sequential design):

     $$X_{n+1} = \operatorname{argmin}_{x \in \mathbb{X}} J_n(x), \quad n \geq n_0,$$

     where n_0 is the size of the initial "exploratory" design.

Gaussian process priors

- From now on, we assume a Gaussian process prior on f.

Gaussian process priors

- From now on, we assume a Gaussian process prior on f.
- As a consequence of our observation model

$$Z_i = f(X_i) + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \tau(X_i)),$$

we have the conjugacy property

$$\xi \mid \mathcal{F}_n \sim \mathcal{GP}\left(\hat{\xi}_n, k_n\right)$$

(even for sequential designs !)

Gaussian process priors

- From now on, we assume a Gaussian process prior on f.
- As a consequence of our observation model

$$Z_i = f(X_i) + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \tau(X_i)),$$

we have the conjugacy property

$$\xi \mid \mathcal{F}_n \sim \mathcal{GP}\left(\hat{\xi}_n, k_n\right)$$

(even for sequential designs !)

- We can take $\mathbb{M} = \{\text{all Gaussian measures on } \mathbb{S} = \mathcal{C}\left(\mathbb{X}\right)\}$

Uncertainty functionals: a few examples in optimization

- Design problem
  - single-objective, box-constrained optimization
  - QoI: $M(\xi) = \max \xi$ and/or $X^* = \operatorname{argmax} \xi$

Uncertainty functionals: a few examples in optimization

- Design problem
  - single-objective, box-constrained optimization
  - QoI: $M(\xi) = \max \xi$ and/or $X^* = \operatorname{argmax} \xi$

- Examples of uncertainty measures for this problem
  - $H_n = \mathsf{E}_n\left(\max \xi - \widehat{M}_n\right) = \mathsf{E}_n(\max \xi) - \widehat{M}_n$
    with $\widehat{M}_n = \max \hat{\xi}_n$ (Mockus et al, 1978; Frazier et al, 2008)
  - $H_n = \mathsf{E}_n(\max \xi - M_n) = \mathsf{E}_n(\max \xi) - M_n$
    with $M_n = \max_{i \le n} \xi(X_i)$ (Jones et al, 1998) (noiseless only!)
  - $H_n = -\sum_{x \in \mathbb{X}} \pi_n^x \log(\pi_n^x)$
    with $\pi_n^x = \mathsf{P}_n(X^* = x)$ (Villemonteix et al, 2009) (discrete $\mathbb{X}$)

Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

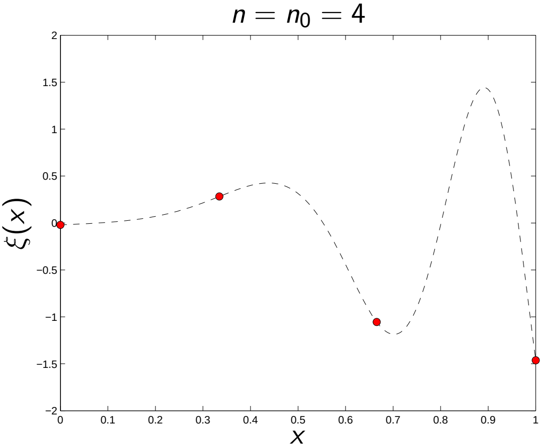


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

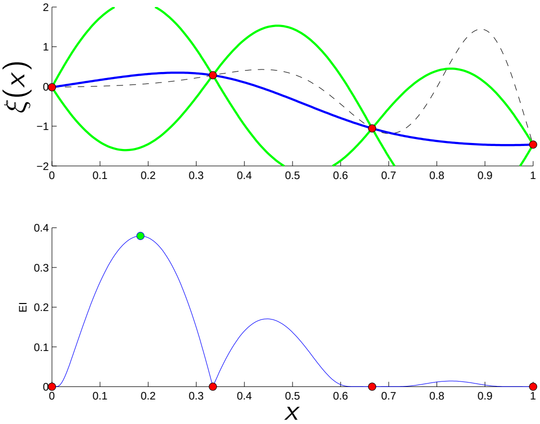


Illustration of the EI criterion (“EGO algorithm”)

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

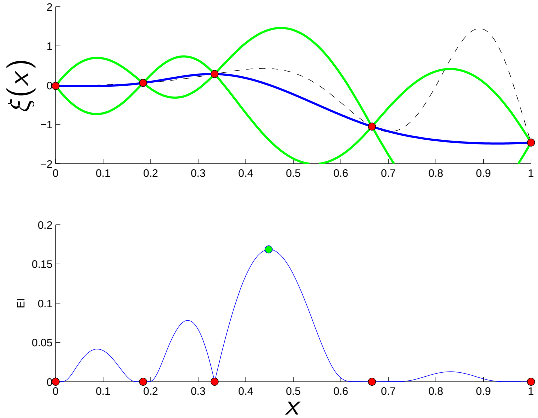
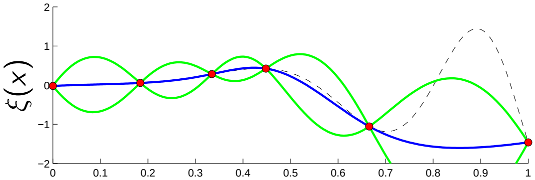


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)



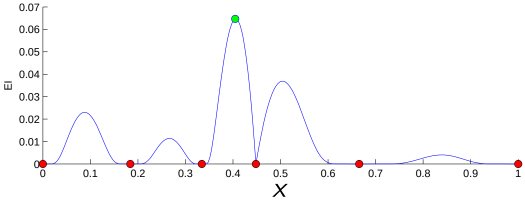


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

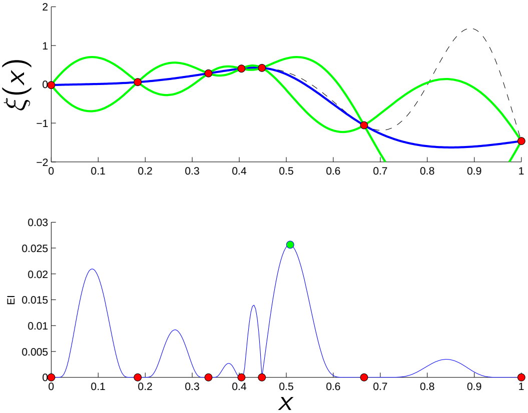


Illustration of the EI criterion (“EGO algorithm”)

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

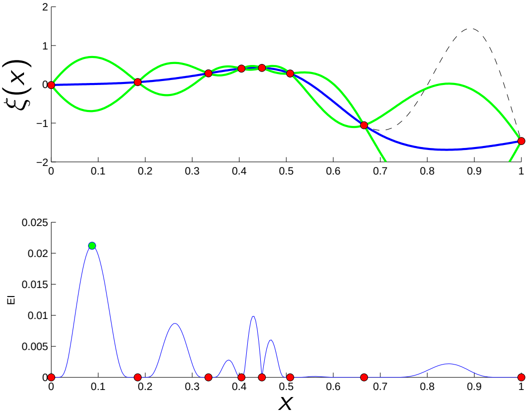


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

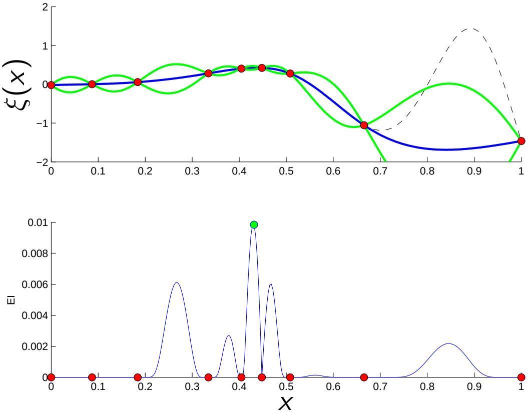


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

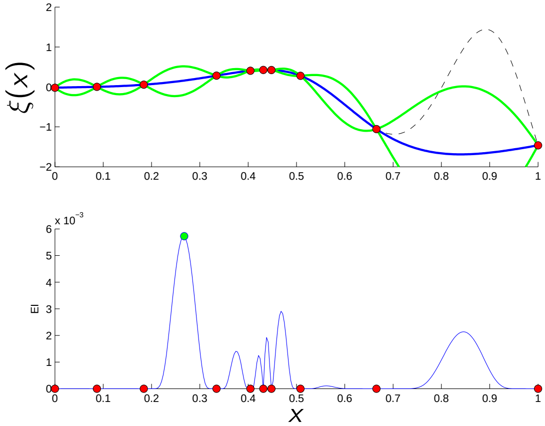


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

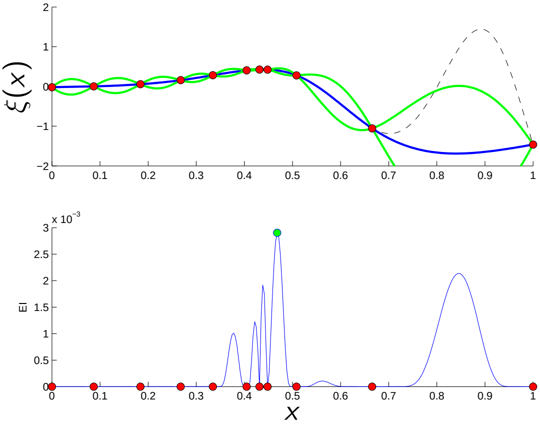


Illustration of the EI criterion (“EGO algorithm”)

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

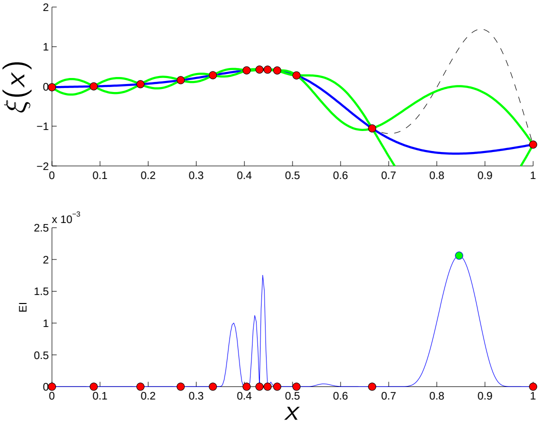


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

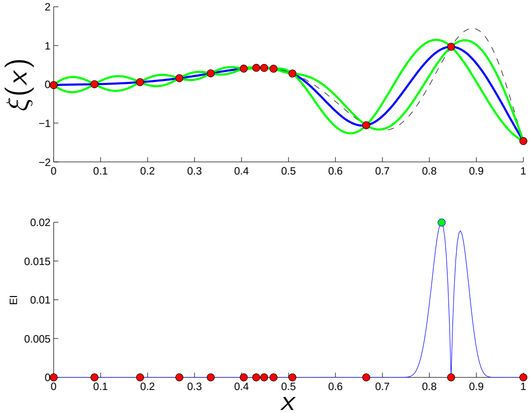


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

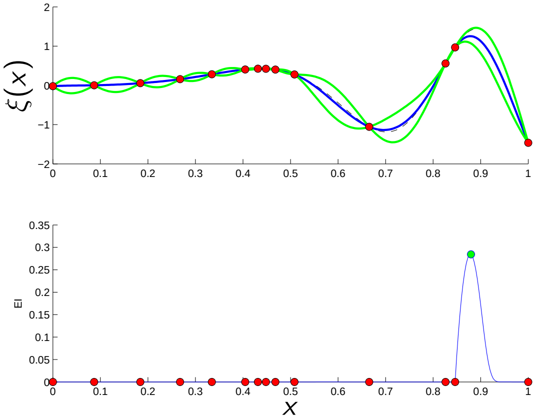


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

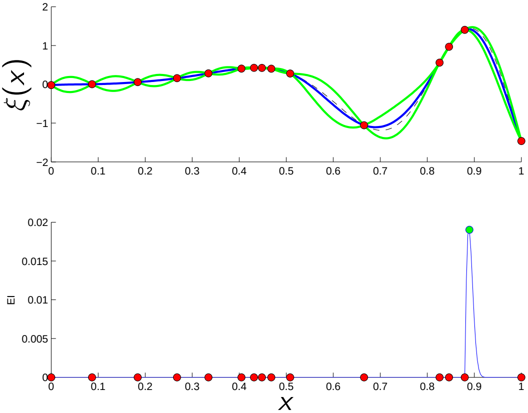


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

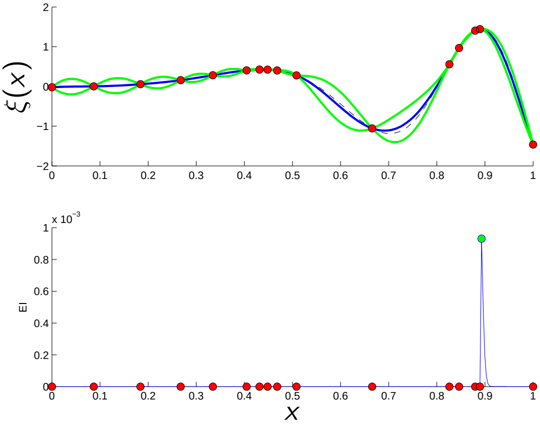


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)

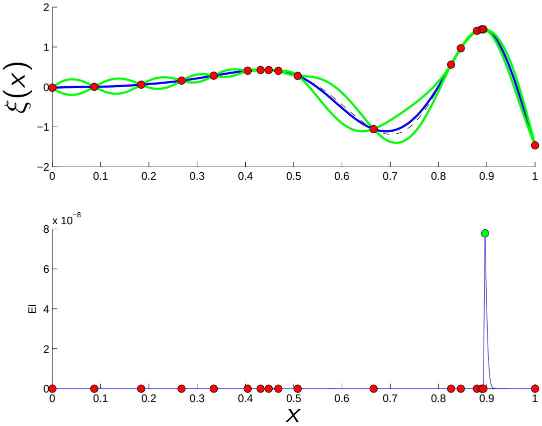
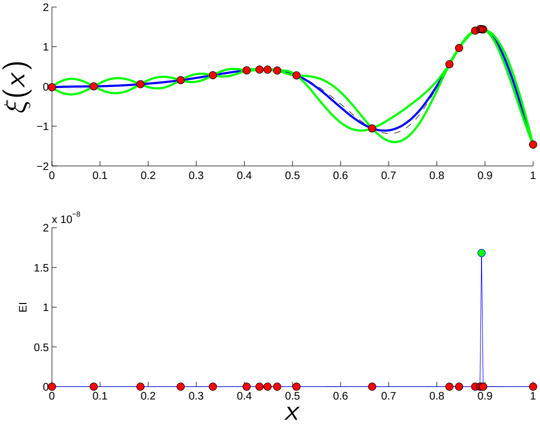


Illustration of the EI criterion ("EGO algorithm")

Model: GP with Matérn covariance ($\sigma^2 = 9$, $\nu = 2$, $\rho = 0.5$)



More uncertainty functionals: probability of failure

- Problem data
  - a threshold $T \in \mathbb{R}$
  - a probability distribution $\mathsf{P}_{\mathbb{X}}$ on $\mathbb{X}$

- QoI: $\theta = \int_{\mathbb{X}} \mathbb{1}_{\xi \geq T} \, \mathrm{d}\mathsf{P}_{\mathbb{X}}$ or $\Gamma = \{x \in \mathbb{X} : \; \xi(x) \geq T\}$

More uncertainty functionals: probability of failure

- Problem data
  - a threshold $T \in \mathbb{R}$
  - a probability distribution $\mathrm{P}_{\mathbb{X}}$ on $\mathbb{X}$
- QoI: $\theta = \int_{\mathbb{X}} \mathbb{1}_{\xi \geq T} \, \mathrm{dP}_{\mathbb{X}}$ or $\Gamma = \{x \in \mathbb{X} : \; \xi(x) \geq T\}$
- Lots of measures of uncertainty have been proposed...

  (Vazquez & B. 2009; Picheny et al. 2010; B. et al 2012;

  Chevalier et al 2014; Azzimonti et al 2016; Walter 2016...)
- Two examples
  - $H_n = \mathrm{var}_n(\theta)$
  - $H_n = \mathrm{E}_n\left(\|\mathbb{1}_\Gamma - p_n\|^2_{\mathrm{P}_{\mathbb{X}}}\right) = \int p_n (1 - p_n) \, \mathrm{dP}_{\mathbb{X}}$

    with $p_n(x) = \mathrm{P}_n(\xi(x) \geq T)$.

Theoretical properties ?

- Very little is known about SUR strategies in general. . .
  - see our arXiv paper for references concerning special cases

Theoretical properties ?

- Very little is known about SUR strategies in general...
  - see our arXiv paper for references concerning special cases
- Assume that $\mathcal{H} = 0$ corresponds to no uncertainty

Theoretical properties ?

- Very little is known about SUR strategies in general...
  - see our arXiv paper for references concerning special cases
- Assume that $\mathcal{H} = 0$ corresponds to no uncertainty
- Basic questions of interest are thus
  - consistency: do we have $H_n = \mathcal{H}\left(\mathrm{P}_n^{\xi}\right) \to 0$?
  - convergence rate: if it is so, how fast ?

Theoretical properties ?

- Very little is known about SUR strategies in general…
  - see our arXiv paper for references concerning special cases

- Assume that $\mathcal{H} = 0$ corresponds to no uncertainty
- Basic questions of interest are thus
  - consistency: do we have $H_n = \mathcal{H}(\mathrm{P}_n^\xi) \to 0$?
  - ~~convergence rate: if it is so, how fast ?~~ (not there yet, sorry about that…)

Problem discussed in this talk

Under which condition on $\mathcal{H}$ can we guarantee that, using SUR,

$$H_n = \mathcal{H}(\mathrm{P}_n^\xi) \to 0 \qquad \mathrm{P}_0\text{-almost surely,}$$

for any prior $\mathrm{P}_0^\xi \in \mathbb{M}$?

Definition

- Recall that
  - $\mathcal{H} : \mathbb{M} \to [0, +\infty)$ and $H_n = \mathcal{H}\left(\mathsf{P}_n^{\xi}\right)$,
  - $\mathbb{M}$ is the set of all Gaussian measures on $\mathbb{S}$ (more generally: conjugate priors. . .)

Definition

- Recall that
  - $\mathcal{H} : \mathbb{M} \to [0, +\infty)$ and $H_n = \mathcal{H}\left(\mathrm{P}_n^\xi\right)$,
  - $\mathbb{M}$ is the set of all Gaussian measures on $\mathbb{S}$ (more generally: conjugate priors. . .)

- Key property

Supermartingale property (SMP)

$\mathcal{H}$ is said to have the supermartingale property if, for any prior $\mathrm{P}_0^\xi \in \mathbb{M}$ and any sequential design X_1, X_2, . . . , the sequence (H_n) is an $(\mathcal{F}_n)$-supermartingale, i.e.,

$$\mathrm{E}\left(H_{n+1}\right) \le H_n, \qquad \forall n \ge 0.$$

$$\mathcal{F}_n = \sigma\left(X_1, Z_1, \ldots, X_n, Z_n\right)$$

A simple example

- Consider any variance functional: $\forall \nu \in \mathbb{M}$,

$$\mathcal{H}(\nu) = \mathsf{var}_\nu(\theta) \;=\; \mathsf{E}_\nu\Big(\underbrace{(\theta - \mathsf{E}_\nu \theta)^2}_{\text{not decreasing (in general)}} \Big)$$

where $\theta = \varphi(\xi)$ is a scalar QoI (e.g., $\theta = \int_{\mathbb{X}} \mathbb{1}_{\xi \geq T} \mathrm{d}\mathsf{P}_{\mathbb{X}}$)

A simple example

- Consider any variance functional: $\forall \nu \in \mathbb{M}$,

$$\mathcal{H}(\nu) = \mathsf{var}_\nu(\theta) \;=\; \mathsf{E}_\nu\Big(\underbrace{(\theta - \mathsf{E}_\nu \theta)^2}_{\text{not decreasing (in general)}} \Big)$$

where $\theta = \varphi(\xi)$ is a scalar QoI (e.g., $\theta = \int_{\mathbb{X}} \mathbb{1}_{\xi \geq T} \mathrm{d}\mathrm{P}_{\mathbb{X}}$)

- Law of total variance ("Eve's law")

$$\mathsf{var}_n(\theta) \;=\; \mathsf{E}_n\left(\mathsf{var}_{n+1}(\theta)\right) + \mathsf{var}_n\left(\mathsf{E}_{n+1}(\theta)\right)$$

A simple example

- Consider any variance functional: $\forall \nu \in \mathbb{M}$,

$$\mathcal{H}(\nu) = \mathsf{var}_{\nu}(\theta) = \mathsf{E}_{\nu}\Big(\underbrace{(\theta - \mathsf{E}_{\nu}\theta)^2}_{\text{not decreasing (in general)}} \Big)$$

where $\theta = \varphi(\xi)$ is a scalar QoI (e.g., $\theta = \int_{\mathbb{X}} \mathbb{1}_{\xi \geq T} \mathrm{dP}_{\mathbb{X}}$)

- Law of total variance ("Eve's law")

$$\begin{aligned} \mathsf{var}_n(\theta) &= \mathsf{E}_n(\mathsf{var}_{n+1}(\theta)) + \mathsf{var}_n(\mathsf{E}_{n+1}(\theta)) \\ &\geq \mathsf{E}_n(\mathsf{var}_{n+1}(\theta)) \end{aligned}$$

- therefore $H_n \geq \mathsf{E}_n(H_{n+1})$, i.e., (H_n) is a supermartingale

A more general explanation

- Consider any "risk-like" uncertainty functional:

$$\mathcal{H}(\nu) = \inf_{d \in \mathbb{D}} \mathsf{E}_\nu \left(L(\xi, d) \right)$$

where $\mathbb{D}$ is a certain "decision space".

A more general explanation

- Consider any "risk-like" uncertainty functional:

$$\mathcal{H}(\nu) = \inf_{d \in \mathbb{D}} \mathsf{E}_\nu \left(L(\xi, d) \right)$$

  where $\mathbb{D}$ is a certain "decision space".

- Assume for simplicity that, $\forall n$, the infimum is attained at $d_n \in \mathbb{D}$

- Then we have:

$$H_{n+1} = \inf_{d \in \mathbb{D}} \mathsf{E}_{n+1} \left(L(\xi, d) \right) \le \mathsf{E}_{n+1} \left(L(\xi, d_n) \right),$$

  and thus

$$\mathsf{E}_n \left(H_{n+1} \right) \le \mathsf{E}_n \left(\mathsf{E}_{n+1} \left(L(\xi, d_n) \right) \right) = H_n.$$

Relationship with concavity (1/2)

- Connection pointed out by L. Pronzato with DeGroot (1962). Uncertainty, Information, and Sequential Experiments. Annals Math. Stat., 33(2):404–419.

Relationship with concavity (1/2)

- Connection pointed out by L. Pronzato with DeGroot (1962). Uncertainty, Information, and Sequential Experiments. Annals Math. Stat., 33(2):404–419.
- Assume for simplicity that $\mathcal{H}$ is defined on $\mathbb{M}_1(\mathbb{S})$
  - $\mathbb{M}_1(\mathbb{S}) = \{$all probability measures on $\mathbb{S}\}$
  - remark: $\mathbb{M}$ itself is not convex...

Relationship with concavity (1/2)

- Connection pointed out by L. Pronzato with DeGroot (1962). Uncertainty, Information, and Sequential Experiments. Annals Math. Stat., 33(2):404–419.
- Assume for simplicity that $\mathcal{H}$ is defined on $\mathbb{M}_1(\mathbb{S})$
  - $\mathbb{M}_1(\mathbb{S}) = \{\text{all probability measures on } \mathbb{S}\}$
  - remark: $\mathbb{M}$ itself is not convex...
- If $\mathcal{H}$ satisfies Jensen's inequality:

$$\mathsf{E}\left(\mathcal{H}(\boldsymbol{\nu})\right) \leq \mathcal{H}(\mathsf{E}(\boldsymbol{\nu}))$$

for all random element $\boldsymbol{\nu}$ in $\mathbb{M}$ (random measure), then

$\mathcal{H}$ has the SMP.

Relationship with concavity (2/2)

- The condition (Jensen's inequality) is sufficient, but not necessary

Relationship with concavity (2/2)

- The condition (Jensen's inequality) is sufficient, but not necessary
- Example: $\mathcal{H}(\nu) = \max_{x \in \mathbb{X}} \sigma_\nu^2(x)$
  - $\mathcal{H}$ has the SMP, and even more: $H_n = \mathcal{H}\left(\mathrm{P}_n^\xi\right)$ is decreasing

Relationship with concavity (2/2)

- The condition (Jensen's inequality) is sufficient, but not necessary
- Example: $\mathcal{H}(\nu) = \max_{x \in \mathbb{X}} \sigma_\nu^2(x)$
  - $\mathcal{H}$ has the SMP, and even more: $H_n = \mathcal{H}\left(\mathsf{P}_n^\xi\right)$ is decreasing
  - but it is not even concave on $\mathrm{co}(\mathbb{M})$
- To see that it is not concave:
  - Pick ν_1, $\nu_2 \in \mathbb{M}$ corresponding to two zero-mean GPs
  - Consider the mixture: $\nu = \frac{1}{2}\nu_1 + \frac{1}{2}\nu_2$ (no longer the distrib. of a GP !)
  - Note that $\sigma_\nu^2(x) = \frac{1}{2}\sigma_{\nu_1}^2(x) + \frac{1}{2}\sigma_{\nu_1}^2(x)$ because of zero-mean
  - Construct variance functions s.t. $\max \sigma_{\nu_1}^2 = \max \sigma_{\nu_2}^2 > \max \sigma_\nu^2$.

Maximal expected uncertainty reduction

- Assume that $\mathcal{H}$ has the SMP

Maximal expected uncertainty reduction

- Assume that $\mathcal{H}$ has the SMP
- Recall the SUR sampling criterion:

$$J_n(x) = \mathsf{E}_{n,x}\left(\mathcal{H}\left(\mathsf{P}^{\xi}_{n+1}\right)\right) \leq H_n$$

and a corresponding SUR sequential design:

$$X_{n+1} \in \operatorname{argmin}_{x\in\mathbb{X}} J_n(x), \quad n \geq n_0.$$

Maximal expected uncertainty reduction

- Assume that $\mathcal{H}$ has the SMP
- Recall the SUR sampling criterion:

$$J_n(x) = \mathsf{E}_{n,x}\left(\mathcal{H}\left(\mathsf{P}^{\xi}_{n+1}\right)\right) \le H_n$$

and a corresponding SUR sequential design:

$$X_{n+1} \in \operatorname{argmin}_{x \in \mathbb{X}} J_n(x), \quad n \ge n_0.$$

- We introduce another functional, denoted by $\mathcal{G}$, s.t.

$$\mathcal{G}\left(\mathsf{P}^{\xi}_{n}\right) = H_n - \min J_n \ge 0$$

measures the maximal expected uncertainty reduction in one step

Consequence of the SMP

Theorem (part 1)

Let $\mathcal{H}$ denote a measurable uncertainty functional with the supermartingale property. If

i) (X_n) is SUR sequential design for $\mathcal{H}$.

then $\mathcal{G}(\mathrm{P}_n^\xi) \to 0$ almost surely.

The SMP in itself is not enough to prove consistency

Zeros of $\mathcal{H}$ and $\mathcal{G}$

- Let us define

$$\mathbb{Z}_{\mathcal{H}} = \{\nu \in \mathbb{M} : \mathcal{H}(\nu) = 0\}\,,$$
$$\mathbb{Z}_{\mathcal{G}} = \{\nu \in \mathbb{M} : \mathcal{G}(\nu) = 0\}\,.$$

Zeros of $\mathcal{H}$ and $\mathcal{G}$

- Let us define

$$\mathbb{Z}_{\mathcal{H}} = \{\nu \in \mathbb{M} : \mathcal{H}(\nu) = 0\}\,,$$
$$\mathbb{Z}_{\mathcal{G}} = \{\nu \in \mathbb{M} : \mathcal{G}(\nu) = 0\}\,.$$

- It is always true that $\mathbb{Z}_{\mathcal{H}} \subset \mathbb{Z}_{\mathcal{G}}$.

Zeros of $\mathcal{H}$ and $\mathcal{G}$

- Let us define

$$\mathbb{Z}_{\mathcal{H}} = \{\nu \in \mathbb{M} : \mathcal{H}(\nu) = 0\},$$
$$\mathbb{Z}_{\mathcal{G}} = \{\nu \in \mathbb{M} : \mathcal{G}(\nu) = 0\}.$$

- It is always true that $\mathbb{Z}_{\mathcal{H}} \subset \mathbb{Z}_{\mathcal{G}}$.
- The converse, however, is not true in general
  - but was found to be true for all the examples that we studied

Example: $\mathcal{H}(\nu) = \mathsf{var}_\nu(\theta)$ with $\theta = \int_{\mathbb{X}} \mathbb{1}_{\xi \geq T} \mathrm{dP}_{\mathbb{X}}$

Example: $\mathcal{H}(\nu) = \text{var}_\nu(\theta)$ with $\theta = \int_{\mathbb{X}} \mathbb{1}_{\xi \geq T} \mathrm{dP}_{\mathbb{X}}$

- Let $\xi \sim \nu \in \mathbb{Z}_{\mathcal{G}}$. Then we can prove (see the paper) that

$$\theta(\xi) - \mathrm{E}(\theta(\xi)) \perp L^2(\xi(x)), \quad \forall x \in \mathbb{X}.$$

Example: $\mathcal{H}(\nu) = \mathrm{var}_\nu(\theta)$ with $\theta = \int_{\mathbb{X}} \mathbb{1}_{\xi \geq T} \mathrm{d}\mathrm{P}_{\mathbb{X}}$

- Let $\xi \sim \nu \in \mathbb{Z}_{\mathcal{G}}$. Then we can prove (see the paper) that

$$\theta(\xi) - \mathrm{E}(\theta(\xi)) \perp L^2(\xi(x)), \quad \forall x \in \mathbb{X}.$$

- In particular, $\theta(\xi) - \mathrm{E}(\theta(\xi)) \perp \mathbb{1}_{\xi(x) \leq T}$, for all $x \in \mathbb{X}$, and thus

$$\mathrm{var}\left(\theta(\xi)\right) \;=\; \int \mathrm{cov}\left(\theta(\xi), \mathbb{1}_{\xi(x) \leq T}\right) \mu(\mathrm{d}x) \;=\; 0.$$

Example: $\mathcal{H}(\nu) = \mathrm{var}_\nu(\theta)$ with $\theta = \int_{\mathbb{X}} \mathbb{1}_{\xi \geq T} \mathrm{d}\mathrm{P}_{\mathbb{X}}$

- Let $\xi \sim \nu \in \mathbb{Z}_{\mathcal{G}}$. Then we can prove (see the paper) that

$$\theta(\xi) - \mathrm{E}(\theta(\xi)) \perp L^2(\xi(x)), \quad \forall x \in \mathbb{X}.$$

- In particular, $\theta(\xi) - \mathrm{E}(\theta(\xi)) \perp \mathbb{1}_{\xi(x) \leq T}$, for all $x \in \mathbb{X}$, and thus

$$\mathrm{var}\left(\theta(\xi)\right) = \int \mathrm{cov}\left(\theta(\xi), \mathbb{1}_{\xi(x) \leq T}\right) \mu(\mathrm{d}x) = 0.$$

- thus $\mathbb{Z}_{\mathcal{G}} \subset \mathbb{Z}_{\mathcal{H}}$, and therefore $\mathbb{Z}_{\mathcal{G}} = \mathbb{Z}_{\mathcal{H}}$.

A general consistency result

Theorem (part 1)

Let $\mathcal{H}$ denote a measurable uncertainty functional with the supermartingale property. If

i) (X_n) is SUR sequential design for $\mathcal{H}$.

then $\mathcal{G}(\mathrm{P}_n^{\xi}) \to 0$ almost surely.

A general consistency result

Theorem

Let $\mathcal{H}$ denote a measurable uncertainty functional with the supermartingale property. If

i) (X_n) is SUR sequential design for $\mathcal{H}$.

then $\mathcal{G}(\mathrm{P}_n^{\xi}) \to 0$ almost surely. If, moreover,

i) $\mathbb{Z}_{\mathcal{H}} = \mathbb{Z}_{\mathcal{G}}$;

A general consistency result

Theorem

Let $\mathcal{H}$ denote a measurable uncertainty functional with the supermartingale property. If

i) (X_n) is SUR sequential design for $\mathcal{H}$.

then $\mathcal{G}(\mathrm{P}_n^\xi) \to 0$ almost surely. If, moreover,

i) $\mathbb{Z}_\mathcal{H} = \mathbb{Z}_\mathcal{G}$;

ii) $H_n = \mathcal{H}(\mathrm{P}_n^\xi) \to \mathcal{H}(\mathrm{P}_\infty^\xi)$ almost surely,

iii) $\mathcal{G}(\mathrm{P}_n^\xi) \to \mathcal{G}(\mathrm{P}_\infty^\xi)$ almost surely;

A general consistency result

Theorem

Let $\mathcal{H}$ denote a measurable uncertainty functional with the supermartingale property. If

i) (X_n) is SUR sequential design for $\mathcal{H}$.

then $\mathcal{G}(\mathrm{P}_n^\xi) \to 0$ almost surely. If, moreover,

i) $\mathbb{Z}_{\mathcal{H}} = \mathbb{Z}_{\mathcal{G}}$;

ii) $H_n = \mathcal{H}(\mathrm{P}_n^\xi) \to \mathcal{H}(\mathrm{P}_\infty^\xi)$ almost surely,

iii) $\mathcal{G}(\mathrm{P}_n^\xi) \to \mathcal{G}(\mathrm{P}_\infty^\xi)$ almost surely;

then $H_n = \mathcal{H}(\mathrm{P}_n^\xi) \to 0$ almost surely.

A general consistency result

Theorem

Let $\mathcal{H}$ denote a measurable uncertainty functional with the supermartingale property. If

i) (X_n) is SUR sequential design for $\mathcal{H}$.

then $\mathcal{G}(\mathrm{P}_n^{\xi}) \to 0$ almost surely. If, moreover,

i) $\mathbb{Z}_{\mathcal{H}} = \mathbb{Z}_{\mathcal{G}}$;

ii) $H_n = \mathcal{H}(\mathrm{P}_n^{\xi}) \to \mathcal{H}(\mathrm{P}_\infty^{\xi})$ almost surely,

iii) $\mathcal{G}(\mathrm{P}_n^{\xi}) \to \mathcal{G}(\mathrm{P}_\infty^{\xi})$ almost surely;

then $H_n = \mathcal{H}(\mathrm{P}_n^{\xi}) \to 0$ almost surely.

Remark about ii) and iii): the functionals are not continuous in general, for any reasonable topology on $\mathbb{M}$... (see the paper for more on this issue)

Conclusion / perspectives

- A general approach to establish the almost sure consistency of SUR strategies is now available

Conclusion / perspectives

- A general approach to establish the almost sure consistency of SUR strategies is now available
- Four examples are covered in detail in the paper
  - optimization: EI and "knowledge gradient"
  - reliablity: $H_n = \mathrm{var}_n(\theta)$ and $H_n = \int p_n(1 - p_n)\, \mathrm{dP}_{\mathbb{X}}$
- Remark about the EI case: holds for any continuous GP, unlike previous result by Vazquez & B. (2010).

Conclusion / perspectives

- A general approach to establish the almost sure consistency of SUR strategies is now available
- Four examples are covered in detail in the paper
  - optimization: EI and “knowledge gradient”
  - reliablity: $H_n = \mathrm{var}_n(\theta)$ and $H_n = \int p_n(1 - p_n)\,\mathrm{d}\mathrm{P}_{\mathbb{X}}$
- Remark about the EI case: holds for any continuous GP, unlike previous result by Vazquez & B. (2010).
- Perspective: convergence rate ? (even the most stupid design can be consistent. . .)

Convergence rate: where do we go from here ?

- Parametric models ?
    - Paninski (2006) claims very strong results (entropy, variance)
    - Proofs are very sketchy. . .

Convergence rate: where do we go from here ?

- Parametric models ?
  - Paninski (2006) claims very strong results (entropy, variance)
  - Proofs are very sketchy...
- Back to GPs: the "Expected improvement" strategy
  - Bull (2011): disappointing upper-bound for $f \in \mathrm{RKHS}(k)$
  - Yarotsky (2013): exponentially fast convergence for the squared-exponential covariance, under some very strong conditions on f
  - These results are not satisfactory from a Bayesian point of view

Convergence rate: where do we go from here ?

- Parametric models ?
  - Paninski (2006) claims very strong results (entropy, variance)
  - Proofs are very sketchy. . .
- Back to GPs: the "Expected improvement" strategy
  - Bull (2011): disappointing upper-bound for $f \in \mathrm{RKHS}(k)$
  - Yarotsky (2013): exponentially fast convergence for the squared-exponential covariance, under some very strong conditions on f
  - These results are not satisfactory from a Bayesian point of view
- The linear / Gaussian case
  - Assume $\xi \sim \mathcal{GP}$, $\theta = \int_{\mathbb{X}} \xi \mathrm{d}\mu$ and $\mathcal{H}(\nu) = \mathrm{var}_\nu(\theta)$
  - Then SUR reduces to Orthogonal Matching Pursuit (OMP)
  - Much can be learned from (greedy) approximation theory

Thank you

Everything is in the title