An introduction to some EGO-like algorithms for constrained / multi-objective / noisy problems

Julien Bect

Laboratoire des signaux et systèmes, Gif-sur-Yvette, France

École-chercheurs Mexico: “Analyse de sensibilité, métamodélisation et optimisation de modèles complexes”

La Rochelle, 30 mars 2018

Where do you think you're going ? (M. Knopfler, 1979)

In this lecture we will...

- introduce some "EGO like" (Bayesian optimization) algorithms for other types of optimization problems,
- present some (toy) examples in R and Matlab/Octave

Where do you think you're going ? (M. Knopfler, 1979)

In this lecture we will. . .

- introduce some "EGO like" (Bayesian optimization) algorithms for other types of optimization problems,
- present some (toy) examples in R and Matlab/Octave

Types of problems that we will consider (if time permits. . .)

- deterministic simulators
  - inequality-constrained problems
  - multi-objective problems

Where do you think you're going ? (M. Knopfler, 1979)

In this lecture we will...

- introduce some "EGO like" (Bayesian optimization) algorithms for other types of optimization problems,
- present some (toy) examples in R and Matlab/Octave

Types of problems that we will consider (if time permits...)

- deterministic simulators
  - inequality-constrained problems
  - multi-objective problems
- stochastic simulators
  - optimization of the mean response

Disclaimers

- This is just the tip of the iceberg. . .
  - gradients (e.g., with adjoint codes)
  - "hidden constraints": handling simulation failures
  - equality constraints
  - robust optimization
  - estimating extreme level sets / probabilities / quantiles
  - . . .

Disclaimers

- This is just the tip of the iceberg...
  - gradients (e.g., with adjoint codes)
  - "hidden constraints": handling simulation failures
  - equality constraints
  - robust optimization
  - estimating extreme level sets / probabilities / quantiles
  - ...
- Very (biased) selective view of the subject...
  - Chosen algo. are available in R or Matlab/Octave packages
  - I like when things fit in a nice generic framework ☺
- A (random ?) sample of references is given in each section

SUR : a generic principle to create new algorithms

Inequality-constrained problems

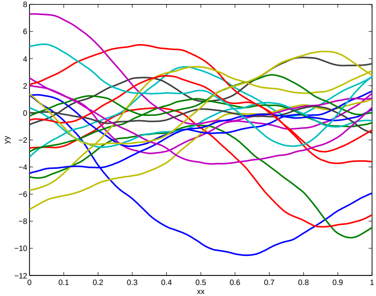
Multi-objective optimization

Noisy optimization

References

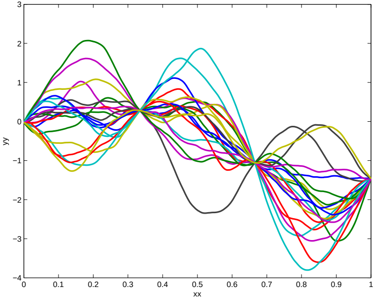
Recap: Bayesian framework, GP priors/posteriors

prior



$Z \sim \mathcal{GP}(m_0, k_0)$

posterior $(n = 4)$

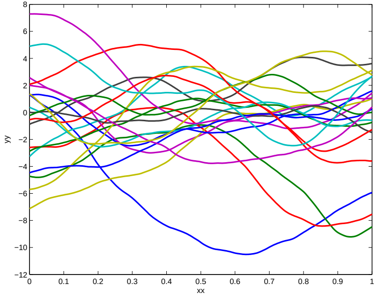


$Z \mid \mathrm{Obs}_n \sim \mathcal{GP}(m_n, k_n)$

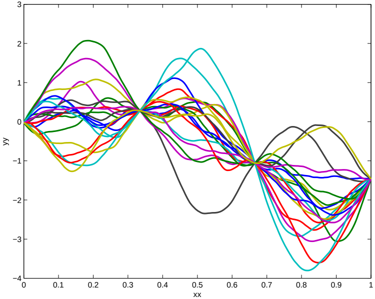
where $\mathrm{Obs}_n = ((x_1, Z_1), \ldots, (x_n, Z_n))$

Recap: Bayesian framework, GP priors/posteriors

prior



posterior $(n = 4)$



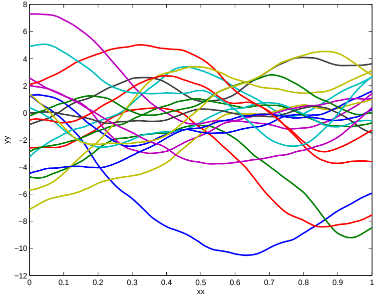
$Z \sim \mathcal{GP}(m_0, k_0)$

$Z \mid \mathrm{Obs}_n \sim \mathcal{GP}(m_n, k_n)$

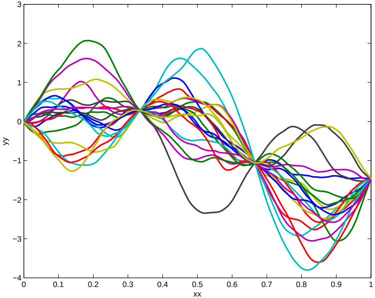
where $\mathrm{Obs}_n = ((x_1, Z_1), \ldots, (x_n, Z_n))$

Extends to noisy observation (for Gaussian noise)

Recap: Bayesian framework, GP priors/posteriors



$$Z \sim \mathcal{GP}(m_0, k_0)$$



$$Z \mid \mathrm{Obs}_n \sim \mathcal{GP}(m_n, k_n)$$

where $\mathrm{Obs}_n = ((x_1, Z_1), \ldots, (x_n, Z_n))$

Extends to noisy observation (for Gaussian noise)

In practice: estimation of hyperparam. + validation (e.g., LOO-CV)

Recap: the Expected Improvement (EI) criterion

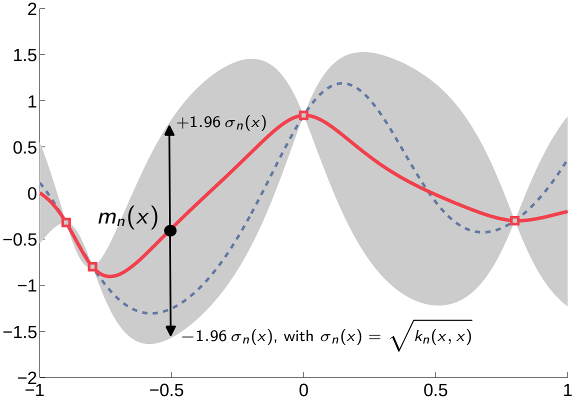


Illustration borrowed from slides by Emmanuel Vazquez, Ecole d'été CEA-EDF-INRIA, July 2017

Recap: the Expected Improvement (EI) criterion

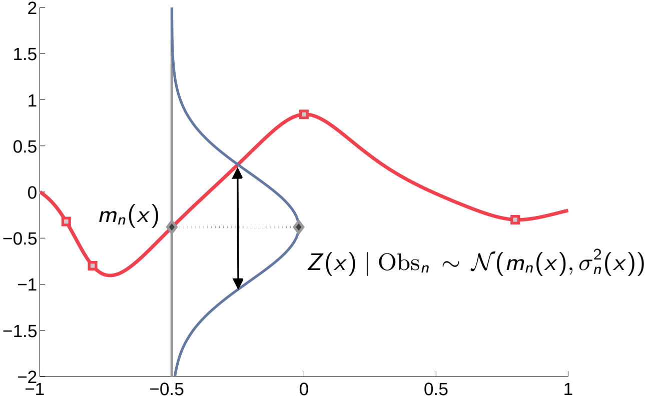
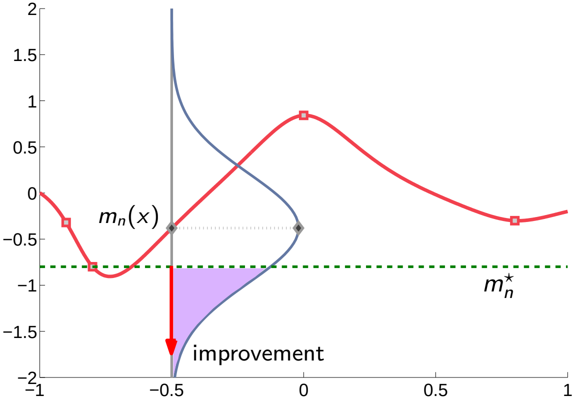


Illustration borrowed from slides by Emmanuel Vazquez, Ecole d'été CEA-EDF-INRIA, July 2017

Recap: the Expected Improvement (EI) criterion



$$\mathrm{EI}_n(x) = \mathsf{E}_n\left((m_n^\star - Z(x))_+ \right)$$

Illustration borrowed from slides by Emmanuel Vazquez, Ecole d'été CEA-EDF-INRIA, July 2017

Another look at the expected improvement (1/3)

- We have denoted the current minimum as

$$m_n^\star = \min_{i \le n} Z_i.$$

- The improvement at x is defined as

$$(m_n^\star - Z(x))_+ = \begin{cases} 0 & \text{if } Z(x) \ge m_n^\star \\ m_n^\star - Z(x) & \text{otherwise,} \end{cases}$$

Another look at the expected improvement (1/3)

- We have denoted the current minimum as

$$m_n^\star = \min_{i \le n} Z_i.$$

- The improvement at x is defined as

$$(m_n^\star - Z(x))_+ = \begin{cases} 0 & \text{if } Z(x) \ge m_n^\star \\ m_n^\star - Z(x) & \text{otherwise,} \end{cases}$$

- If we select a particular x as x_{n+1}, we have $Z_{n+1} = Z(x)$ and

$$(m_n^\star - Z(x))_+ = m_n^\star - m_{n+1}^\star.$$

Another look at the expected improvement (2/3)

- Now set $Z^\star = \min_x Z(x)$ and rewrite (again with $x_{n+1} = x$)

$$(m_n^\star - Z(x))_+ = (m_n^\star - Z^\star) - (m_{n+1}^\star - Z^\star) \tag{1}$$

Another look at the expected improvement (2/3)

- Now set $Z^\star = \min_x Z(x)$ and rewrite (again with $x_{n+1} = x$)

$$(m_n^\star - Z(x))_+ = (m_n^\star - Z^\star) - \left(m_{n+1}^\star - Z^\star\right) \qquad (1)$$

- Taking the expectation at time n, we get

$$\mathrm{EI}_n(x) = \mathsf{E}_n\left(m_n^\star - Z^\star\right) \; - \; \mathsf{E}_n\left(m_{n+1}^\star - Z^\star \mid x_{n+1} = x\right)$$

Another look at the expected improvement (2/3)

- Now set $Z^\star = \min_x Z(x)$ and rewrite (again with $x_{n+1} = x$)

$$(m_n^\star - Z(x))_+ = (m_n^\star - Z^\star) - \left(m_{n+1}^\star - Z^\star\right) \qquad (1)$$

- Taking the expectation at time n, we get

$$\mathrm{EI}_n(x) = \underbrace{\mathsf{E}_n\left(m_n^\star - Z^\star\right)}_{\text{call this } H_n} - \mathsf{E}_n\left(m_{n+1}^\star - Z^\star \mid x_{n+1} = x\right)$$

- $H_n \geq 0$ and, when H_n is small, then $Z^\star \approx m_n^\star$ with high proba.

Interpretation

H_n can be seen as a measure of uncertainty about $Z^\star$, at time n

Another look at the expected improvement (2/3)

- One last effort... recall the law of total expectation:

$$\mathsf{E}_n(U) = \mathsf{E}_n(\mathsf{E}_{n+1}(U)), \qquad \text{for any r.v. } U.$$

Another look at the expected improvement (2/3)

- One last effort... recall the law of total expectation:

$$\mathsf{E}_n(U) = \mathsf{E}_n(\mathsf{E}_{n+1}(U)), \qquad \text{for any r.v. } U.$$

- We can thus rewrite the EI criterion as

$$\mathrm{EI}_n(x) = H_n \; - \; \mathsf{E}_n(H_{n+1} \mid x_{n+1} = x)$$

Another look at the expected improvement (2/3)

- One last effort... recall the law of total expectation:

$$\mathsf{E}_n(U) = \mathsf{E}_n(\mathsf{E}_{n+1}(U)), \qquad \text{for any r.v. } U.$$

- We can thus rewrite the EI criterion as

$$\mathrm{EI}_n(x) = H_n \; - \; \mathsf{E}_n\left(H_{n+1} \mid x_{n+1} = x\right)$$

Interpretation

- The EGO algorithm minimizes greedily, at each step, the expected uncertainty at the next step.
- This is a particular case of a **Stepwise Uncertainty Reduction** (SUR) algorithm.

Stepwise Uncertainty Reduction (SUR)

- Assume that you want to estimate a certain QoI
  - QoI = Quantity of Interest
  - Example: $x^\star$ or $f(x^\star)$ in a single-objective optim problem

Stepwise Uncertainty Reduction (SUR)

- Assume that you want to estimate a certain QoI
  - QoI = Quantity of Interest
  - Example: $x^\star$ or $f(x^\star)$ in a single-objective optim problem

- Then the SUR approach consists of the following steps

SUR approach

1. Choose a prior: f is seen as a samplepath from a GP Z

Stepwise Uncertainty Reduction (SUR)

- Assume that you want to estimate a certain QoI
  - QoI = Quantity of Interest
  - Example: $x^\star$ or $f(x^\star)$ in a single-objective optim problem

- Then the SUR approach consists of the following steps

SUR approach

1. Choose a prior: f is seen as a samplepath from a GP Z
2. Choose a "measure of uncertainty" H_n

Stepwise Uncertainty Reduction (SUR)

- Assume that you want to estimate a certain QoI
  - QoI = Quantity of Interest
  - Example: $x^\star$ or $f(x^\star)$ in a single-objective optim problem
- Then the SUR approach consists of the following steps

SUR approach

1. Choose a prior: f is seen as a samplepath from a GP Z
2. Choose a "measure of uncertainty" H_n
3. Iterate (possibly after some exploratory initial design)

$$x_{n+1} = \operatorname{argmin}_x \mathsf{E}_n \left(H_{n+1} \mid x_{n+1} = x \right)$$

or, equivalently, $x_{n+1} = \operatorname{argmax}_x H_n - \mathsf{E}_n \left(H_{n+1} \mid x_{n+1} = x \right)$.

Illustration of the EI criterion ("EGO algorithm")

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

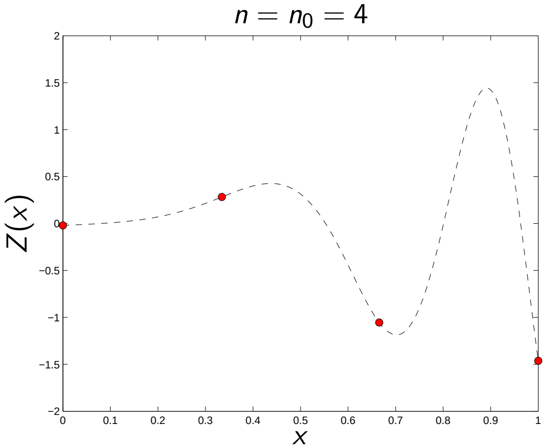


Illustration of the EI criterion (“EGO algorithm”)

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

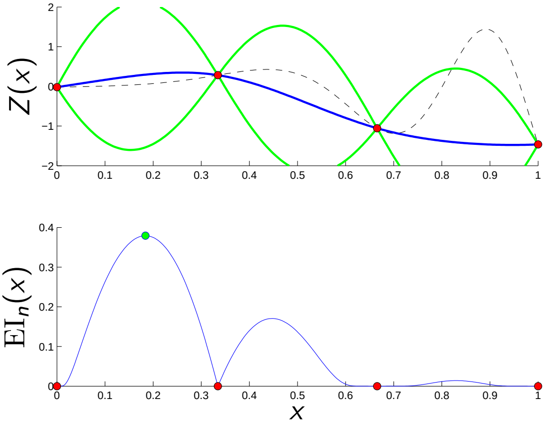


Illustration of the EI criterion ("EGO algorithm")

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

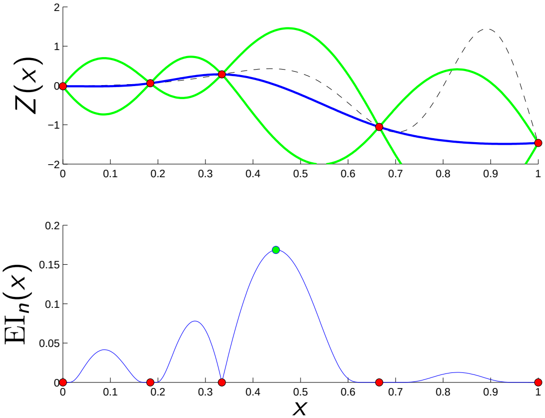
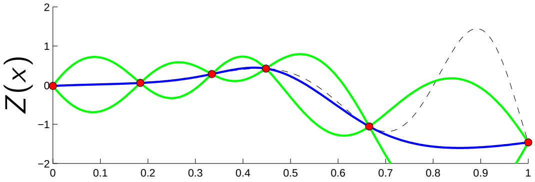


Illustration of the EI criterion ("EGO algorithm")

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



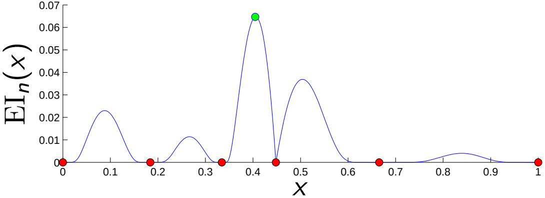


Illustration of the EI criterion ("EGO algorithm")

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

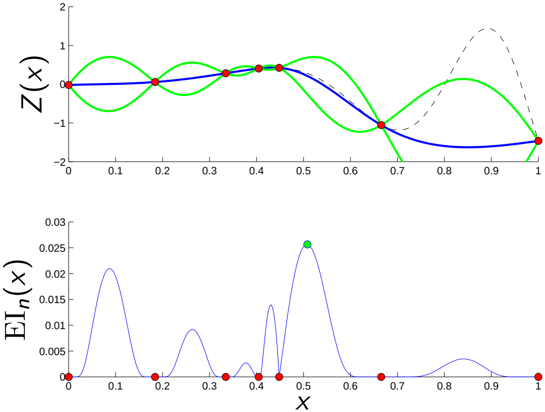


Illustration of the EI criterion ("EGO algorithm")

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

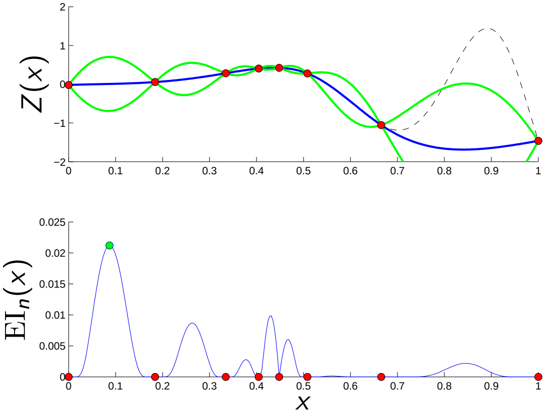


Illustration of the EI criterion (“EGO algorithm”)

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

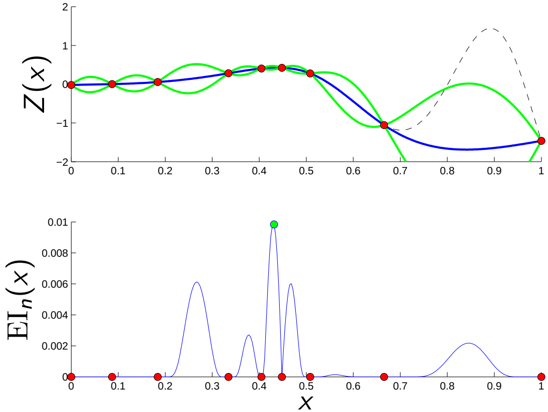


Illustration of the EI criterion ("EGO algorithm")

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

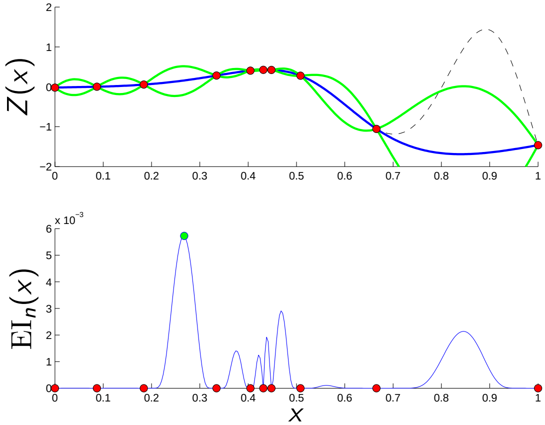


Illustration of the EI criterion (“EGO algorithm”)

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

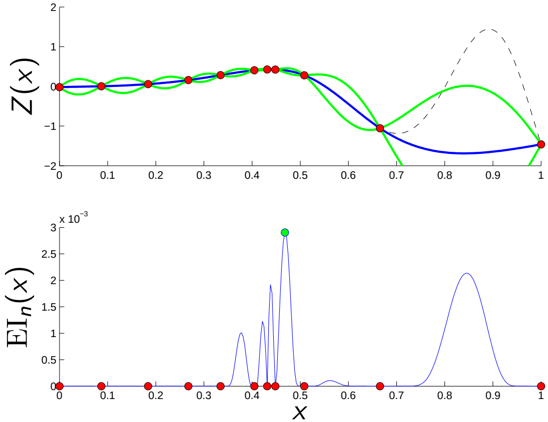


Illustration of the EI criterion ("EGO algorithm")

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

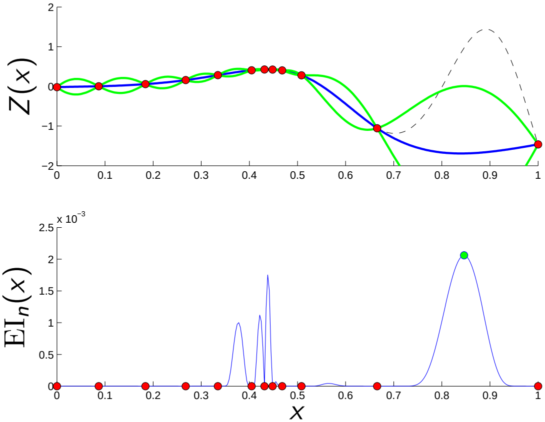


Illustration of the EI criterion ("EGO algorithm")

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

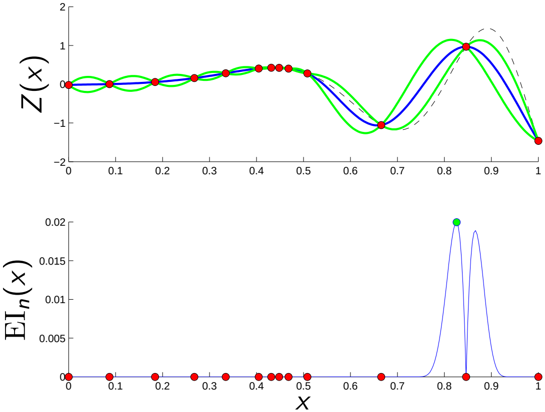


Illustration of the EI criterion ("EGO algorithm")

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

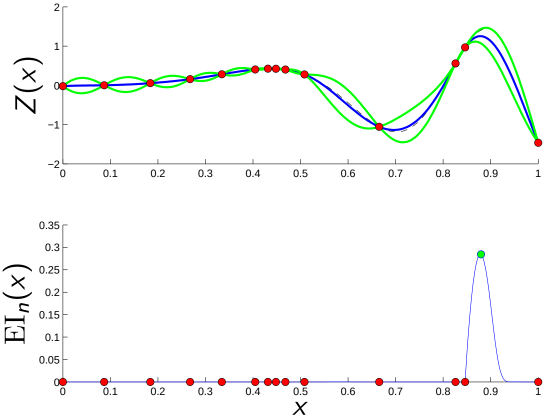


Illustration of the EI criterion ("EGO algorithm")

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

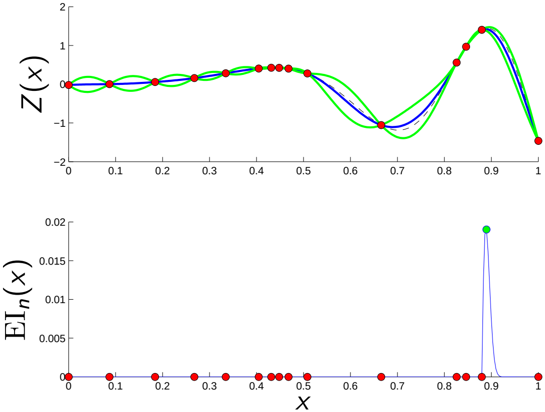


Illustration of the EI criterion ("EGO algorithm")

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

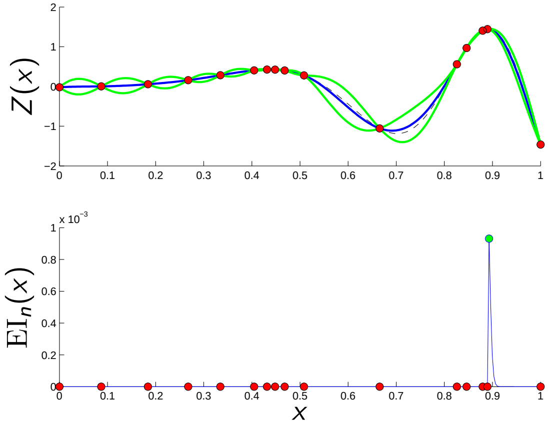


Illustration of the EI criterion ("EGO algorithm")

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

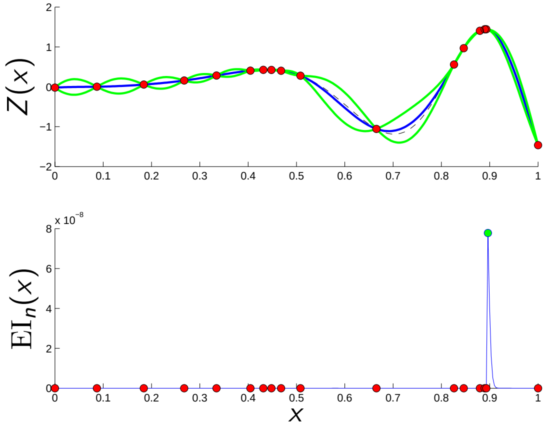
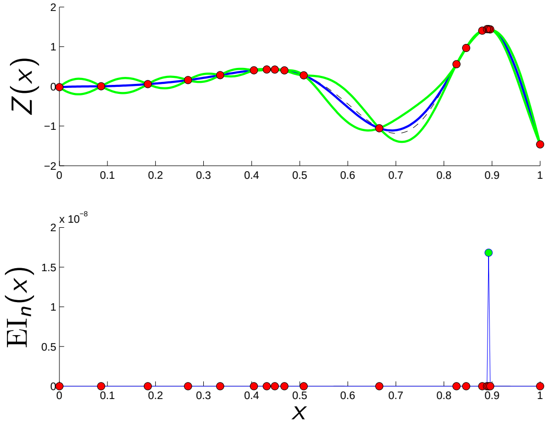


Illustration of the EI criterion ("EGO algorithm")

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



References for the SUR approach

- Oldest papers that I know of
  - Psychophysics: King-Smith (1984), Pelli (1987)
  - Sequential DoE to estimate “activation thresholds”

References for the SUR approach

- Oldest papers that I know of
  - Psychophysics: King-Smith (1984), Pelli (1987)
  - Sequential DoE to estimate "activation thresholds"
- References from the nineties
  - Computer vision: Geman and Jedynak [1996]
  - Active learning: MacKay [1992], Cohn et al. [1996]

References for the SUR approach

- Oldest papers that I know of
  - Psychophysics: King-Smith (1984), Pelli (1987)
  - Sequential DoE to estimate "activation thresholds"
- References from the nineties
  - Computer vision: Geman and Jedynak [1996]
  - Active learning: MacKay [1992], Cohn et al. [1996]
- More recent works, dealing with GP models
  - IAGO: Villemonteix [2008], Villemonteix et al. [2009]
  - Reliability: Picheny et al. [2010], Bect et al. [2012], Chevalier et al. [2014]
  - A little bit of theory: Bect et al. [2017]

Inequality-constrained problems

- Consider a single-objective, inequality-contrained problem:
  - minimize $f(x)$
  - under the constraints $x \in \mathbb{X}$ and $g_j(x) \leq 0$, $1 \leq j \leq q$
- where
  - $\mathbb{X}$ is a nice, known compact subset of $\mathbb{R}^d$ (e.g., $\mathbb{X} = [0;1]^d$)
  - $f, g_1, \ldots, g_q : \mathbb{X} \to \mathbb{R}$
  - both f and the g_j's are expensive to evaluate

Inequality-constrained problems

- Consider a single-objective, inequality-contrained problem:
  - minimize $f(x)$
  - under the constraints $x \in \mathbb{X}$ and $g_j(x) \leq 0$, $1 \leq j \leq q$
- where
  - $\mathbb{X}$ is a nice, known compact subset of $\mathbb{R}^d$ (e.g., $\mathbb{X} = [0;1]^d$)
  - $f, g_1, \ldots, g_q : \mathbb{X} \to \mathbb{R}$
  - both f and the g_j's are expensive to evaluate
- Important: now there are several "unknown" functions !
  - We will need to use $q + 1$ GP models
  - Usually assumed independent $\Rightarrow$ not really harder than one GP

Inequality-constrained problems

- Consider a single-objective, inequality-contrained problem:
  - minimize $f(x)$
  - under the constraints $x \in \mathbb{X}$ and $g_j(x) \leq 0$, $1 \leq j \leq q$
- where
  - $\mathbb{X}$ is a nice, known compact subset of $\mathbb{R}^d$ (e.g., $\mathbb{X} = [0;1]^d$)
  - $f, g_1, \ldots, g_q : \mathbb{X} \to \mathbb{R}$
  - both f and the g_j's are expensive to evaluate
- Important: now there are several "unknown" functions !
  - We will need to use $q + 1$ GP models
  - Usually assumed independent $\Rightarrow$ not really harder than one GP
  - Notations: $f \mapsto Z_{\mathrm{o}}$ and $g_j \mapsto Z_{\mathrm{c},j}$, $1 \leq j \leq q$

Measure of uncertainty ?

- Recall the measure of uncertainty used in the EI case:

$$H_n = \mathsf{E}_n\left(m_n^\star - Z^\star\right)$$

- How do we adapt this to inequality contraints ?

Measure of uncertainty ?

- Recall the measure of uncertainty used in the EI case:

$$H_n = \mathsf{E}_n\left(m_n^\star - Z^\star\right)$$

- How do we adapt this to inequality contraints ?

- Assume that at least one feasible solution is known
- Then the same H_n can be used, with

$$m_n^\star = \{Z_{\rm o}(x_i) \mid i \le n \ \text{ s.t. } \ \forall j \le q,\ Z_{{\rm c},j}(x_i) \le 0\}$$
$$Z^\star = \{Z_{\rm o}(x) \mid x \in \mathbb{X} \ \text{ s.t. } \ \forall j \le q,\ Z_{{\rm c},j}(x) \le 0\}$$

Sampling criterion

- The corresponding "expected improvement" (to be maximized) is

$$\mathrm{EFI}_n(x) = H_n - \mathsf{E}_n\left(H_{n+1} \mid x_{n+1} = x\right)$$
$$= \mathsf{E}_n\Big(\underbrace{\left(m_n^\star - Z_\mathrm{o}(x)\right)_+}_{\text{improvement}} \cdot \underbrace{\mathbb{1}_{\forall j \le q,\, Z_{\mathrm{c},j}(x) \le 0}}_{\text{feasibility}} \Big)$$

- EFI = Expected Feasible Improvement (cf. DiceOptim)
  - WARNING: there is another ~~Skywalker~~ SUR crit. in DiceOptim

Sampling criterion

- The corresponding "expected improvement" (to be maximized) is

$$\mathrm{EFI}_n(x) = H_n - \mathsf{E}_n\left(H_{n+1} \mid x_{n+1} = x\right)$$
$$= \mathsf{E}_n\Big(\underbrace{\left(m_n^\star - Z_\mathrm{o}(x)\right)_+}_{\text{improvement}} \cdot \underbrace{\mathbb{1}_{\forall j \le q,\; Z_{\mathrm{c},j}(x) \le 0}}_{\text{feasibility}} \Big)$$

- EFI = Expected Feasible Improvement (cf. DiceOptim)
  - WARNING: there is another ~~Skywalker~~ SUR crit. in DiceOptim

- Assuming independent GPs, this simplifies to

$$\mathrm{EFI}_n(x) = \underbrace{\mathsf{E}_n\Big(\left(m_n^\star - Z_\mathrm{o}(x)\right)_+\Big)}_{\approx \text{ the "usual" EI}} \cdot \underbrace{\Pi_{j=1}^{q}\, \mathsf{P}_n\left(Z_{\mathrm{c},j}(x) \le 0\right)}_{\text{probability of feasibility}}.$$

Example

DiceOptim demo (`easyEGO.cst`)

... single-objective optimization

with $q = 2$ inequality constraints ...

References

- Original work by M. Schonlau and co-authors
  - Schonlau et al. [1998]
  - see also Schonlau's PhD thesis

References

- Original work by M. Schonlau and co-authors
  - Schonlau et al. [1998]
  - see also Schonlau's PhD thesis
- Other approaches, extension, comparisons. . .
  - Sasena [2002], Sasena et al. [2002]
  - Parr [2013], Parr et al. [2012]
  - PESC: Gelbart [2015], Hernández-Lobato et al. [2015]
  - A different SUR: Picheny [2014a]

References

- Original work by M. Schonlau and co-authors
  - Schonlau et al. [1998]
  - see also Schonlau's PhD thesis
- Other approaches, extension, comparisons...
  - Sasena [2002], Sasena et al. [2002]
  - Parr [2013], Parr et al. [2012]
  - PESC: Gelbart [2015], Hernández-Lobato et al. [2015]
  - A different SUR: Picheny [2014a]
- Equality constraints
  - ALBO: Picheny et al. [2016]

SUR : a generic principle to create new algorithms

Inequality-constrained problems

Multi-objective optimization

Noisy optimization

References

Multi-objective optimization

- Several objective functions to be minimized: $\underline{f} = (f_1, \ldots, f_p)$
  - $f_j : \mathbb{X} \to \mathbb{R}$, $1 \leq j \leq p$

Multi-objective optimization

- Several objective functions to be minimized: $\underline{f} = (f_1, \ldots, f_p)$
  - $f_j : \mathbb{X} \to \mathbb{R}$, $1 \leq j \leq p$

Pareto domination relation

$$\underline{z} \prec \underline{z}' \quad \text{if (def)} \quad \begin{cases} z_j \leq z'_j & \text{for all } j \leq p, \\ z_j < z'_j & \text{for at least one } j \leq p. \end{cases}$$

Multi-objective optimization

- Several objective functions to be minimized: $\underline{f} = (f_1, \ldots, f_p)$
  - $f_j : \mathbb{X} \to \mathbb{R}$, $1 \leq j \leq p$

Pareto domination relation

$$\underline{z} \prec \underline{z}' \quad \text{if (def)} \quad \begin{cases} z_j \leq z_j' & \text{for all } j \leq p, \\ z_j < z_j' & \text{for at least one } j \leq p. \end{cases}$$

- The "quantities of interest" here are
  - the Pareto set $\mathbb{P} = \{x \in \mathbb{X} : \nexists x' \in \mathbb{X}, \underline{f}(x') \prec \underline{f}(x)\}$ (a.k.a. set of Pareto-efficient solutions)

Multi-objective optimization

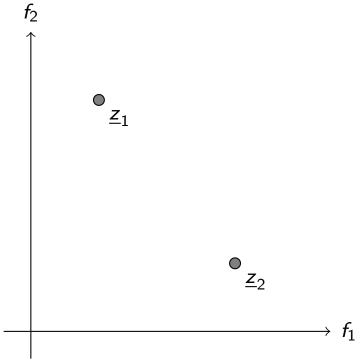
- Several objective functions to be minimized: $\underline{f} = (f_1, \ldots, f_p)$
  - $f_j : \mathbb{X} \to \mathbb{R}$, $1 \leq j \leq p$

Pareto domination relation

$$\underline{z} \prec \underline{z}' \quad \text{if (def)} \quad \begin{cases} z_j \leq z'_j & \text{for all } j \leq p, \\ z_j < z'_j & \text{for at least one } j \leq p. \end{cases}$$

- The "quantities of interest" here are
  - the Pareto set $\mathbb{P} = \{x \in \mathbb{X} : \nexists x' \in \mathbb{X}, \underline{f}(x') \prec \underline{f}(x)\}$ (a.k.a. set of Pareto-efficient solutions)
  - and/or the Pareto front $\{\underline{z} \in \mathbb{R}^p : \exists x \in \mathbb{P}, \underline{z} = \underline{f}(x)\}$ (a.k.a Pareto frontier, Pareto boundary...)

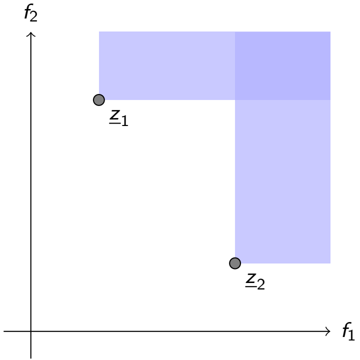
Illustration in the objective space



Bi-objective illustration ($p = 2$)

$\underline{z}_i = (z_{i,1}, z_{i,2})$

Illustration in the objective space

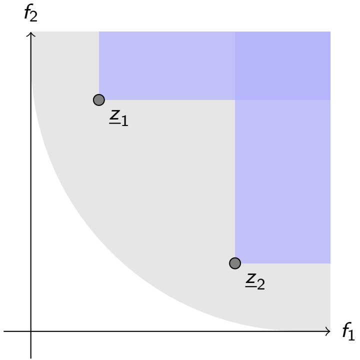


Bi-objective illustration ($p = 2$)

$\underline{z}_i = (z_{i,1}, z_{i,2})$

Blue: region dominated by the current observations

Illustration in the objective space



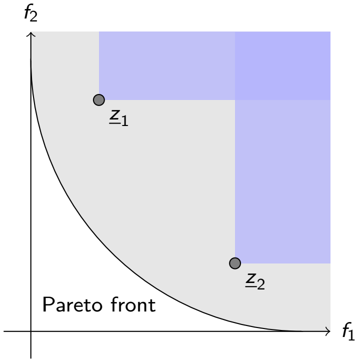
Bi-objective illustration ($p = 2$)

$\underline{z}_i = (z_{i,1}, z_{i,2})$

Blue: region dominated by the current observations

Grey: all feasible solutions (i.e., $\underline{f}(x)$ for some $x \in \mathbb{X}$)

Illustration in the objective space



Bi-objective illustration ($p = 2$)

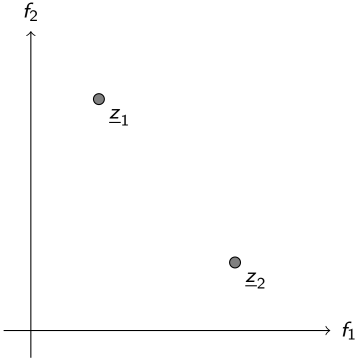
$\underline{z}_i = (z_{i,1}, z_{i,2})$

Blue: region dominated by the current observations

Grey: all feasible solutions (i.e., $\underline{f}(x)$ for some $x \in \mathbb{X}$)

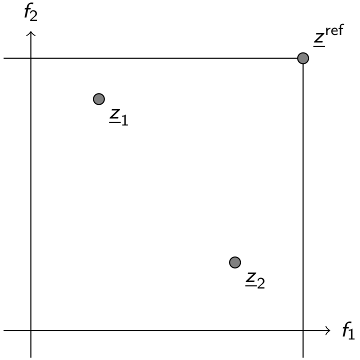
Expected HyperVolume Improvement (EHVI)

- GP models $Z_1, \ldots, Z_p$ for $f_1, \ldots, f_p$
- EHVI: a natural extension of EI to multi-objective problems



Expected HyperVolume Improvement (EHVI)

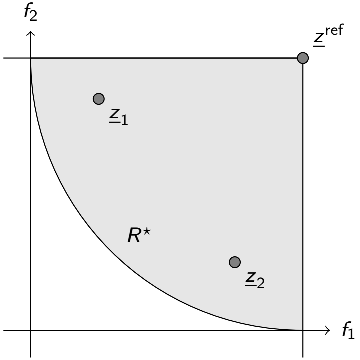
- GP models $Z_1, \ldots, Z_p$ for $f_1, \ldots, f_p$
- EHVI: a natural extension of EI to multi-objective problems



$\mathbb{B} = \Pi_{j=1}^{p} \left]-\infty; z_j^{\text{ref}}\right]$: bounding box

Expected HyperVolume Improvement (EHVI)

- GP models $Z_1, \ldots, Z_p$ for $f_1, \ldots, f_p$
- EHVI: a natural extension of EI to multi-objective problems



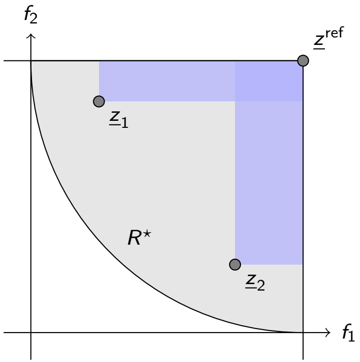
$\mathbb{B} = \Pi_{j=1}^{p} \left]-\infty; z_j^{\text{ref}}\right]$: bounding box

Maximal dominated region for $\underline{Z}$:

$R^\star = \{\underline{z} \in \mathbb{B},\ \exists x \in \mathbb{X},\ \underline{Z}(x) \preceq z\}$

Expected HyperVolume Improvement (EHVI)

- GP models $Z_1, \ldots, Z_p$ for $f_1, \ldots, f_p$
- EHVI: a natural extension of EI to multi-objective problems



$\mathbb{B} = \Pi_{j=1}^{p} \left]-\infty; z_j^{\text{ref}}\right]$: bounding box

Maximal dominated region for $\underline{Z}$:

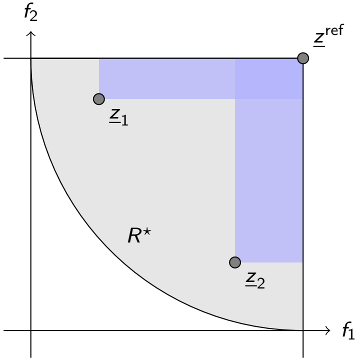
$$R^\star = \{\underline{z} \in \mathbb{B},\ \exists x \in \mathbb{X},\ \underline{Z}(x) \preceq z\}$$

Current dominated region:

$$R_n = \{\underline{z} \in \mathbb{B},\ \exists i \leq n,\ \underline{Z}(x_i) \preceq \underline{z}\}$$

Expected HyperVolume Improvement (EHVI)

- GP models $Z_1, \ldots, Z_p$ for $f_1, \ldots, f_p$
- EHVI: a natural extension of EI to multi-objective problems



$\mathbb{B} = \Pi_{j=1}^{p} \left]-\infty; z_j^{\text{ref}}\right]$: bounding box

Maximal dominated region for $\underline{Z}$:

$$R^\star = \{\underline{z} \in \mathbb{B},\ \exists x \in \mathbb{X},\ \underline{Z}(x) \preceq z\}$$

Current dominated region:

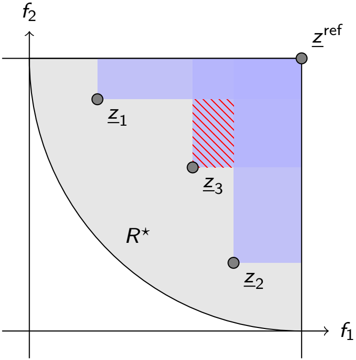
$$R_n = \{\underline{z} \in \mathbb{B},\ \exists i \leq n,\ \underline{Z}(x_i) \preceq \underline{z}\}$$

Measure of uncertainty:

$$H_n = \mathsf{E}_n\left(\mathrm{vol}_p\left(R^\star \setminus R_n\right)\right)$$

Expected HyperVolume Improvement (EHVI)

- GP models $Z_1, \ldots, Z_p$ for $f_1, \ldots, f_p$
- EHVI: a natural extension of EI to multi-objective problems



$\mathbb{B} = \Pi_{j=1}^{p} \left]-\infty; z_j^{\mathrm{ref}}\right]$: bounding box

Maximal dominated region for $\underline{Z}$:

$$R^\star = \{\underline{z} \in \mathbb{B},\ \exists x \in \mathbb{X},\ \underline{Z}(x) \preceq z\}$$

Current dominated region:

$$R_n = \{\underline{z} \in \mathbb{B},\ \exists i \leq n,\ \underline{Z}(x_i) \preceq \underline{z}\}$$

Measure of uncertainty:

$$H_n = \mathsf{E}_n\left(\mathrm{vol}_p\left(R^\star \setminus R_n\right)\right)$$

$$\mathrm{EHVI}_n(x) = \mathsf{E}_n\left(\mathrm{vol}_p\left(R_{n+1} \setminus R_n\right) \mid x_{n+1} = x\right)$$

Multi-objective problems

- Implementation
  - Exactly computable for independent GP priors, $2 \leq p \lesssim 5$
  - Implemented in GPareto (R), STK (Matlab/Octave)...
  - Dependent priors, larger p: Monte Carlo approx.

Multi-objective problems

- Implementation
  - Exactly computable for independent GP priors, $2 \leq p \lesssim 5$
  - Implemented in GPareto (R), STK (Matlab/Octave)...
  - Dependent priors, larger p: Monte Carlo approx.

Multi-objective problems

- Implementation
  - Exactly computable for independent GP priors, $2 \leq p \lesssim 5$
  - Implemented in GPareto (R), STK (Matlab/Octave)...
  - Dependent priors, larger p: Monte Carlo approx.

STK demo

... bi-objective optimization with the EHVI criterion ...

code by Etienne Leloup, Guillaume Maistre-Bazin, Lucain Pouget
CentraleSupelec final year project for CEA DIF

References

- Original work on the EHVI by M. Emmerich and co-authors
  - Emmerich [2005]
  - Emmerich et al. [2006]
  - Wagner et al. [2010]

References

- Original work on the EHVI by M. Emmerich and co-authors
  - Emmerich [2005]
  - Emmerich et al. [2006]
  - Wagner et al. [2010]
- Fast computation of the EHVI criterion
  - Emmerich and Klinkenberg [2008]
  - Hupkens et al. [2014]

References

- Original work on the EHVI by M. Emmerich and co-authors
  - Emmerich [2005]
  - Emmerich et al. [2006]
  - Wagner et al. [2010]
- Fast computation of the EHVI criterion
  - Emmerich and Klinkenberg [2008]
  - Hupkens et al. [2014]
- Extensions & other approaches (far from exhaustive)
  - Constraints: Feliot et al. [2017]
  - ParEGO: Knowles [2006]
  - A different SUR: Picheny [2014b]
  - ... more refs in Feliot et al. [2017], section 2.2

SUR : a generic principle to create new algorithms

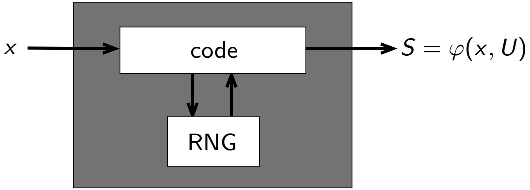
Inequality-constrained problems

Multi-objective optimization

Noisy optimization

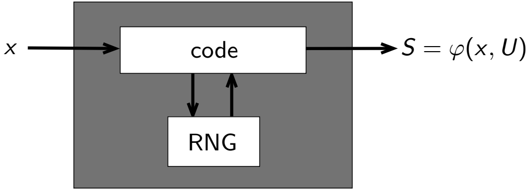
References

Stochastic simulators in a nutshell



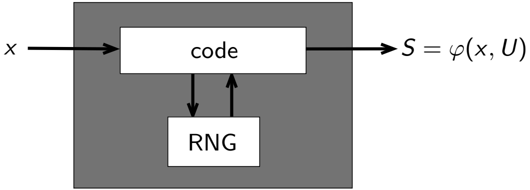
- Inside the box
  - A random vector U is generated during the simulation
  - Remark: most of the time, U is not directly accessible

Stochastic simulators in a nutshell



- Inside the box
  - A random vector U is generated during the simulation
  - Remark: most of the time, U is not directly accessible
- From the outside
  - The response S (assumed scalar here) is a random variable

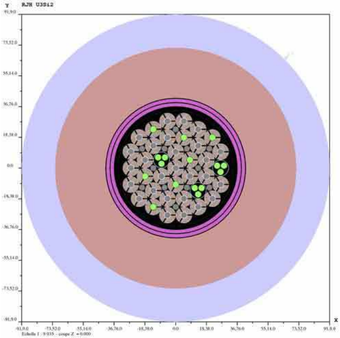
Stochastic simulators in a nutshell



- Inside the box
  - A random vector U is generated during the simulation
  - Remark: most of the time, U is not directly accessible
- From the outside
  - The response S (assumed scalar here) is a random variable
  - The distribution of S depends on x

Example: the MORET code (IRSN)

- MORET simulates neutron transport in fissile materials
- Output = effective multiplication factor (k-eff)
- Uses Monte Carlo methods ⇒ "noisy" estimate of k-eff



Source: MORET website, somewhere on the internet (sorry, no WIFI in my room last night...)

Optimization on stochastic simulators ?

- Assume that we want to "minimize the output" S...
  - what does that even mean for a stochastic simulator ???
  - Recall: the output S is a random variable

Optimization on stochastic simulators ?

- Assume that we want to "minimize the output" S...
  - what does that even mean for a stochastic simulator ???
  - Recall: the output S is a random variable
- Let π_x^S denote the pdf of S at the input point $x \in \mathbb{X}$

Optimization on stochastic simulators ?

- Assume that we want to "minimize the output" S...
  - what does that even mean for a stochastic simulator ???
  - Recall: the output S is a random variable
- Let π_x^S denote the pdf of S at the input point $x \in \mathbb{X}$

The simplest formulation that makes sense (?)

In the following we want to **minimize the expected response**:

Find $x \in \mathbb{X}$ such that $f(x) = \int s \, \pi_x^S(s) \, \mathrm{d}s$ is minimal

Optimization on stochastic simulators ?

- Assume that we want to “minimize the output” S...
  - what does that even mean for a stochastic simulator ???
  - Recall: the output S is a random variable
- Let π_x^S denote the pdf of S at the input point $x \in \mathbb{X}$

The simplest formulation that makes sense (?)

In the following we want to **minimize the expected response**:

$$\text{Find } x \in \mathbb{X} \text{ such that } f(x) = \int s\, \pi_x^S(s)\, \mathrm{d}s \text{ is minimal}$$

- Many other formulations are possible...
  - quantiles, mean/variance, probability constraints, etc.

“Meta-models” (priors) for stochastic simulators ?

- In general GP models can can no longer do the job. . .
  - at each $x \in \mathbb{X}$ we now have a pdf !

“Meta-models” (priors) for stochastic simulators ?

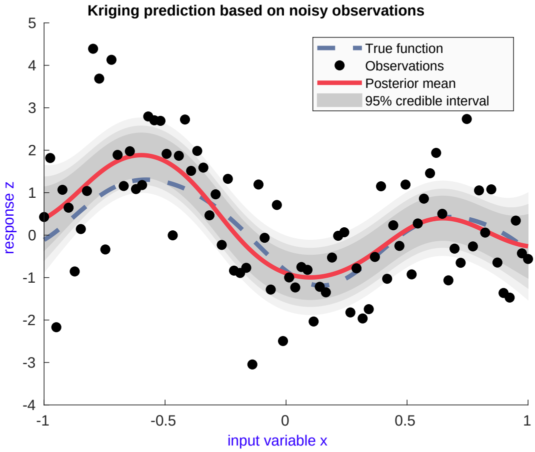
- In general GP models can can no longer do the job. . .
  - at each $x \in \mathbb{X}$ we now have a pdf !
- Simplifying assumption: $\pi_x^S \approx \mathcal{N}\left(f(x), \sigma^2\right)$
  - In this case a GP model Z for f can be used
  - The posterior distribution is still a GP

"Meta-models" (priors) for stochastic simulators ?

- In general GP models can can no longer do the job...
  - at each $x \in \mathbb{X}$ we now have a pdf !
- Simplifying assumption: $\pi_x^S \approx \mathcal{N}\left(f(x), \sigma^2\right)$
  - In this case a GP model Z for f can be used
  - The posterior distribution is still a GP
- This is equivalent to assuming "noisy measurements" of f

$$S_i = f(x_i) + \varepsilon_i, \qquad \varepsilon_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

Illustration: GP model with noisy observations



$\min_{i \leq n} S_i$ is no longer a reasonable estimate of $\min_x Z(x)$ here !

The AKG criterion

- Recall (again) the measure of uncertainty used in the EI case:

$$H_n = \mathsf{E}_n \left(m_n^\star - Z^\star \right)$$

The AKG criterion

- Recall (again) the measure of uncertainty used in the EI case:

$$H_n = \mathsf{E}_n\left(m_n^\star - Z^\star\right)$$

- Here the same H_n can be used (again !) with

$$\color{red}{m_n^\star = \min_x m_n(x)} = \min_x \mathsf{E}_n\left(Z(x)\right)$$

The AKG criterion

- Recall (again) the measure of uncertainty used in the EI case:

$$H_n = \mathsf{E}_n\left(m_n^\star - Z^\star\right)$$

- Here the same H_n can be used (again !) with

$$m_n^\star = \min_x m_n(x) = \min_x \mathsf{E}_n\left(Z(x)\right)$$

Knowledge gradient criterion

$$\mathrm{KG}_n(x) \;=\; m_n^\star - \mathsf{E}\left(m_{n+1}^\star \mid x_{n+1} = x\right)$$

The AKG criterion

- Recall (again) the measure of uncertainty used in the EI case:

$$H_n = \mathsf{E}_n\left(m_n^\star - Z^\star\right)$$

- Here the same H_n can be used (again !) with

$$m_n^\star = \min_x m_n(x) = \min_x \mathsf{E}_n\left(Z(x)\right)$$

Knowledge gradient criterion

$$\mathrm{KG}_n(x) \;=\; m_n^\star - \mathsf{E}\left(m_{n+1}^\star \mid x_{n+1} = x\right)$$

- AKG: approximate min over $\mathbb{X}$ by min over $\{x_1, \ldots, x_n, x\}$.

Example

STK demo (`stk_example_doe05`)

. . . noisy optimization in 1D . . .

For R users: AKG and other noisy optimization sampling criteria also available in DiceOptim ☺

Relaxing the assumptions on the simulator

- Recall our simplifying assumption: $\pi_x^S \approx \mathcal{N}\left(f(x), \sigma^2\right)$
- What if this assumption is too strong for me ?

Relaxing the assumptions on the simulator

- Recall our simplifying assumption: $\pi_x^S \approx \mathcal{N}\left(f(x), \sigma^2\right)$
- What if this assumption is too strong for me ?

- σ^2 depends on x ?
  - still manageable using heteroscedastic GP models

Relaxing the assumptions on the simulator

- Recall our simplifying assumption: $\pi_x^S \approx \mathcal{N}\left(f(x), \sigma^2\right)$
- What if this assumption is too strong for me ?

- σ^2 depends on x ?
  - still manageable using heteroscedastic GP models
- Unknown (non-Gaussian) output distribution ?
  - topic of current research !

Relaxing the assumptions on the simulator

- Recall our simplifying assumption: $\pi_x^S \approx \mathcal{N}\left(f(x), \sigma^2\right)$
- What if this assumption is too strong for me ?

- σ^2 depends on x ?
  - still manageable using heteroscedastic GP models
- Unknown (non-Gaussian) output distribution ?
  - topic of current research !
- "batch trick": if $S_1, \ldots, S_m \overset{\text{iid}}{\sim} \pi_x^S$, then by the CLT

$$\frac{1}{m}\sum_{i=1}^{m} S_i \overset{\text{approx}}{\sim} \mathcal{N}\left(f(x), \frac{\sigma^2}{m}\right)$$

References

- In a noiseless setting, KG was invented a long time ago. . .
  - Mockus et al. [1978]
  - (KG reduces to EI for the particular type of GP considered !)

References

- In a noiseless setting, KG was invented a long time ago...
  - Mockus et al. [1978]
  - (KG reduces to EI for the particular type of GP considered !)
- Original work on KG / AKG for noisy optimization
  - KG: Frazier et al. [2009, 2008]
  - AKG: Scott et al. [2011]

References

- In a noiseless setting, KG was invented a long time ago...
  - Mockus et al. [1978]
  - (KG reduces to EI for the particular type of GP considered !)
- Original work on KG / AKG for noisy optimization
  - KG: Frazier et al. [2009, 2008]
  - AKG: Scott et al. [2011]
- A review / comparison of sampling criteria
  - Picheny et al. [2013]

Thank you for your attention ☺

Everything is in the title

References I

J. Bect, D. Ginsbourger, L. Li, V. Picheny, and E. Vazquez. Sequential design of computer experiments for the estimation of a probability of failure. Statistics and Computing, 22 (3):773–793, 2012.

J. Bect, F. Bachoc, and D. Ginsbourger. A supermartingale approach to gaussian process based sequential design of experiments. arXiv:1608.01118v2, 2017.

Clément Chevalier, Julien Bect, David Ginsbourger, Emmanuel Vazquez, Victor Picheny, and Yann Richet. Fast parallel kriging-based stepwise uncertainty reduction with application to the identification of an excursion set. Technometrics, 56(4):455–465, 2014.

D. A. Cohn, Z. Ghahramani, and M. I. Jordan. Active learning with statistical models. Journal of Artificial Intelligence Research, 4:129–145, 1996.

M. Emmerich. Single- and multiobjective evolutionary design optimization assisted by Gaussian random field metamodels. PhD thesis, Technical University Dortmund, 2005.

M. Emmerich and J. W. Klinkenberg. The computation of the expected improvement in dominated hypervolume of Pareto front approximations. Technical report, Leiden University, 2008.

References II

Michael Emmerich, Kyriakos C. Giannakoglou, and Boris Naujoks. Single- and multi-objective evolutionary optimization assisted by Gaussian random field metamodels. IEEE Transactions on Evolutionary Computation, 10(4):421–439, 2006.

Paul Feliot, Julien Bect, and Emmanuel Vazquez. A Bayesian approach to constrained single-and multi-objective optimization. Journal of Global Optimization, 67(1-2): 97–133, 2017.

Peter Frazier, Warren Powell, and Savas Dayanik. The knowledge-gradient policy for correlated normal beliefs. INFORMS journal on Computing, 21(4):599–613, 2009.

Peter I Frazier, Warren B Powell, and Savas Dayanik. A knowledge-gradient policy for sequential information collection. SIAM Journal on Control and Optimization, 47 (5):2410–2439, 2008.

M. A. Gelbart. Constrained Bayesian Optimization and Applications. PhD thesis, Harvard University, Graduate School of Arts and Sciences, 2015.

D. Geman and B. Jedynak. An active testing model for tracking roads in satellite images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(1): 1–14, 1996.

References III

José Miguel Hernández-Lobato, Michael A. Gelbart, Matthew W. Hoffman, Ryan P. Adams, and Zoubin Ghahramani. Predictive entropy search for Bayesian optimization with unknown constraints. In ICML, pages 1699–1707, 2015.

I. Hupkens, M. Emmerich, and A. Deutz. Faster computation of expected hypervolume improvement. arXiv preprint arXiv:1408.7114, 2014.

Joshua Knowles. ParEGO: A hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. IEEE Transactions on Evolutionary Computation, 10(1):50–66, 2006.

D. J. C. MacKay. Information-based objective functions for active data selection. Neural computation, 4(4):590–604, 1992.

Jonas Mockus, Vytautas Tiesis, and Antanas Žilinskas. The application of Bayesian methods for seeking the extremum. In L. C. W. Dixon and G. P. Szegö, editors, Towards Global Optimization, volume 2, pages 117–129, North Holland, New York, 1978.

J. M. Parr. Improvement criteria for constraint handling and multiobjective optimization. PhD thesis, University of Southampton, 2013.

J. M. Parr, A. J. Keane, A. I. J. Forrester, and C. M. E. Holden. Infill sampling criteria for surrogate-based optimization with constraint handling. Engineering Optimization, 44(10):1147–1166, 2012.

References IV

V. Picheny. A stepwise uncertainty reduction approach to constrained global optimization. In Proceedings of the 17th International Conference on Artificial Intelligence and Statistics (AISTATS), 2014, Reykjavik, Iceland., volume 33, pages 787–795. JMLR: W&CP, 2014a.

V. Picheny. Multiobjective optimization using Gaussian process emulators via stepwise uncertainty reduction. Statistics and Computing, DOI:10.1007/s11222-014-9477-x: 1–16, 2014b.

V. Picheny, D. Ginsbourger, O. Roustant, R.T. Haftka, and N.-H. Kim. Adaptive designs of experiments for accurate approximation of target regions. Journal of Mechanical Design, 132(7), July 2010. doi: 10.1115/1.4001873.

Victor Picheny, Tobias Wagner, and David Ginsbourger. A benchmark of kriging-based infill criteria for noisy optimization. Structural and Multidisciplinary Optimization, 48(3):607–626, 2013.

Victor Picheny, Robert B. Gramacy, Stefan Wild, and Sebastien Le Digabel. Bayesian optimization under mixed constraints with a slack-variable augmented lagrangian. In Advances in Neural Information Processing Systems, pages 1435–1443, 2016.

M. J. Sasena. Flexibility and efficiency enhancements for constrained global design optimization with kriging approximations. PhD thesis, University of Michigan, 2002.

References V

M. J. Sasena, P. Papalambros, and P. Goovaerts. Exploration of metamodeling sampling criteria for constrained global optimization. Engineering Optimization, 34 (3):263–278, 2002.

Matthias Schonlau, William J. Welch, and Donald R. Jones. Global versus local search in constrained optimization of computer models. In New Developments and Applications in Experimental Design: Selected Proceedings of a 1997 Joint AMS-IMS-SIAM Summer Conference, volume 34 of IMS Lecture Notes-Monographs Series, pages 11–25, 1998.

Warren Scott, Peter Frazier, and Warren Powell. The correlated knowledge gradient for simulation optimization of continuous parameters using gaussian process regression. SIAM Journal on Optimization, 21(3):996–1026, 2011.

Julien Villemonteix. Optimisation de fonctions coûteuses Modèles gaussiens pour une utilisation efficace du budget d'évaluations: théorie et pratique industrielle. PhD thesis, Université Paris Sud-Paris XI, 2008.

Julien Villemonteix, Emmanuel Vazquez, and Éric Walter. An informational approach to the global optimization of expensive-to-evaluate functions. Journal of Global Optimization, 44(4):509–534, 2009.

References VI

T. Wagner, M. Emmerich, A. Deutz, and W. Ponweiser. On expected-improvement criteria for model-based multi-objective optimization. In Parallel Problem Solving from Nature, PPSN XI. 11th International Conference, Krakov, Poland, September 11-15, 2010, Proceedings, Part I, volume 6238 of Lecture Notes in Computer Science, pages 718–727. Springer, 2010.