Efficient Global Optimization (EGO)

Introduction to Optimization with Surrogates

- Each optimization cycle consists of sampling design points by simulations, fitting surrogates to simulations and then optimizing an objective.
  - Construct surrogate, optimize original objective, refine region and surrogate.
  - Typically small number of cycles with large number of simulations in each cycle.
- Adaptive sampling
  - Construct surrogate, add points by taking into account not only surrogate prediction but also uncertainty in prediction.
  - Most popular, Jones's EGO (Efficient Global Optimization).
  - Easiest with one added sample at a time.

Background: Surrogate Modeling

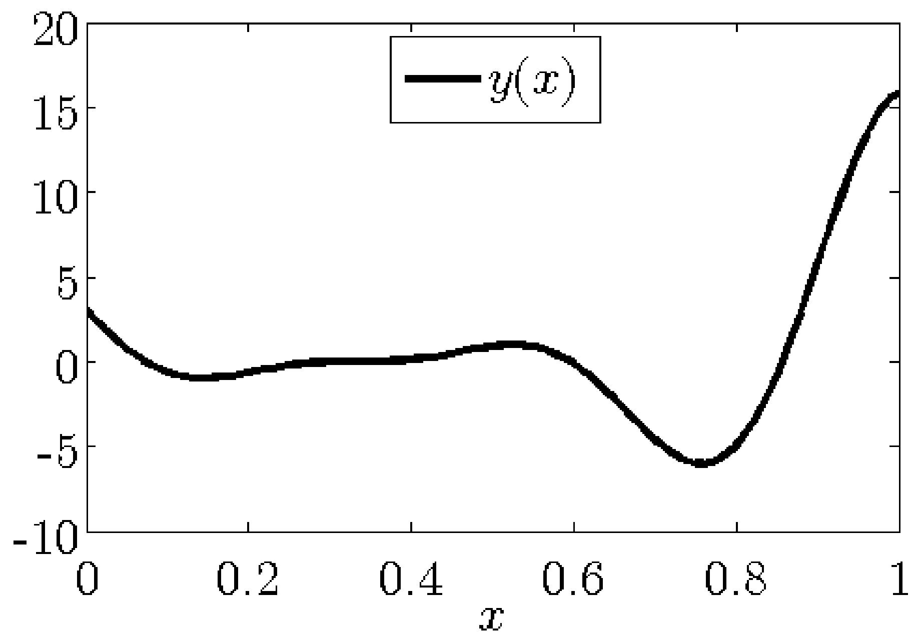
Surrogates **replace expensive simulations** by simple algebraic expressions fit to data.

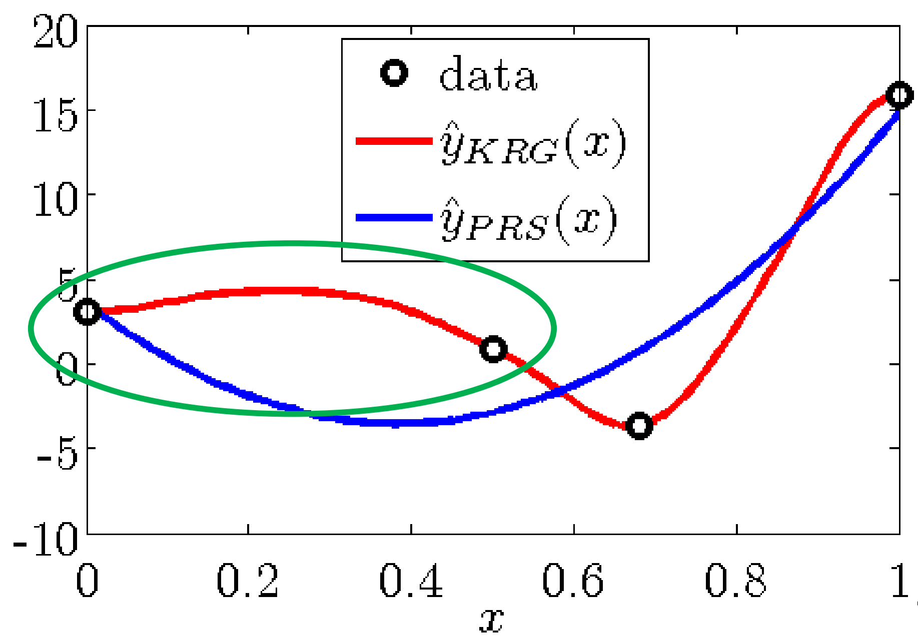
$\hat{y}(x)$ is an estimate of $y(x)$.

Example:

- Kriging (KRG)
- Polynomial response surface (PRS)
- Support vector regression
- Radial basis neural networks

- Differences are larger in regions of low point density.



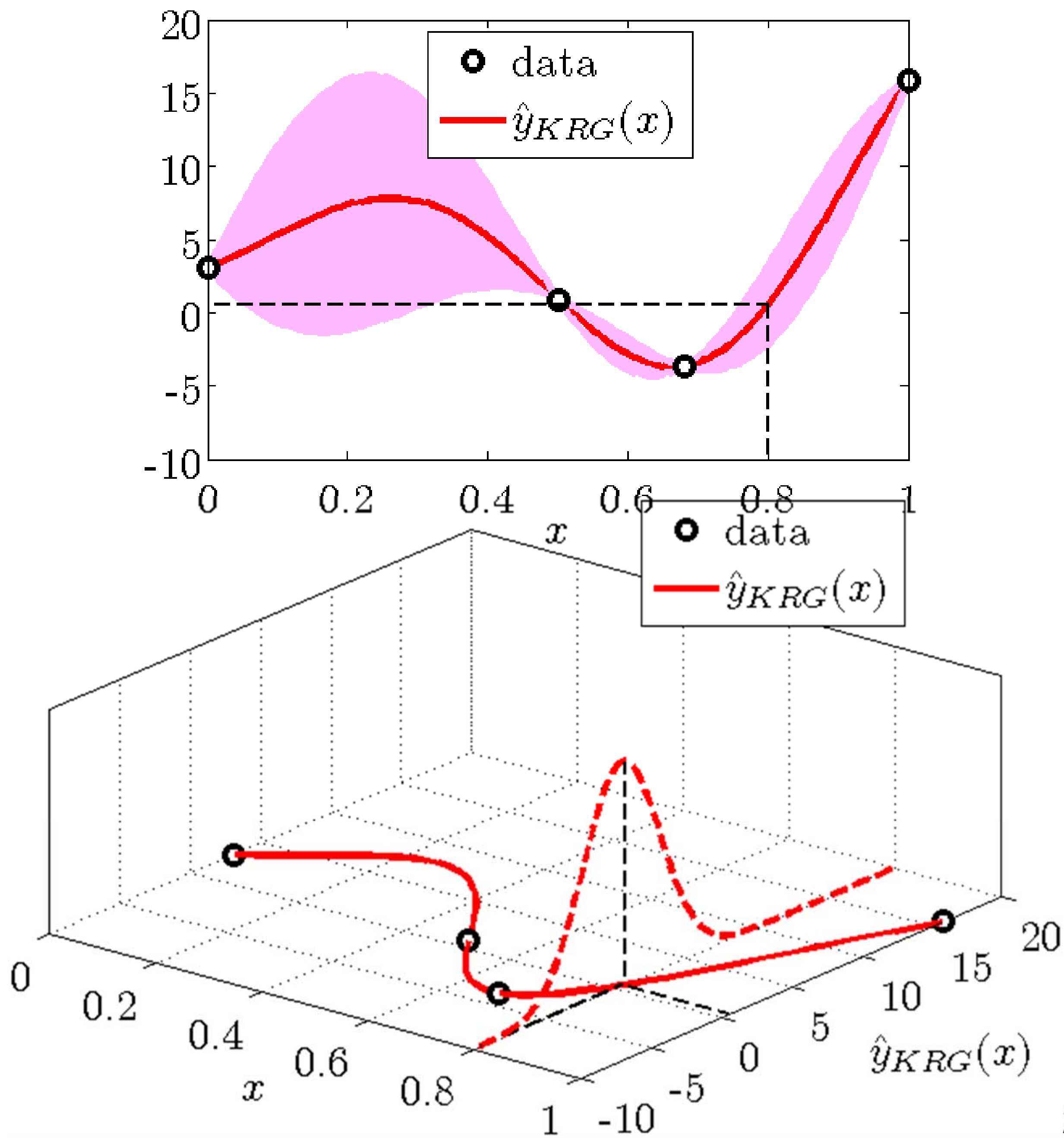


Background: Uncertainty

Some surrogates also provide an **uncertainty estimate**: **standard error, $s(x)$**.

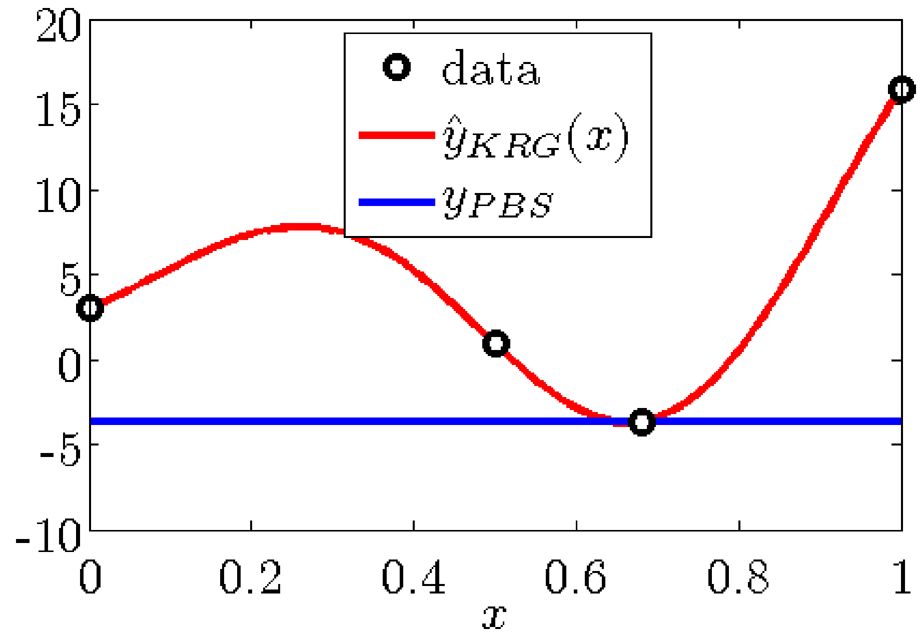
Example: **kriging** and polynomial response surface.

Both of these are used in EGO.

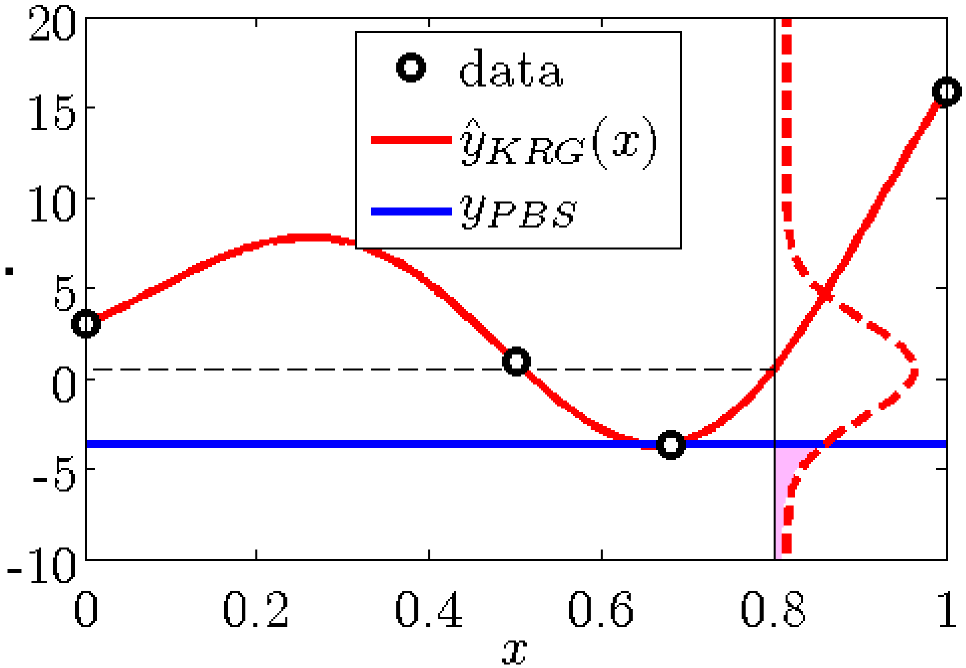


Kriging Fit and Defining Improvement

- First, we sample the function and fit a Kriging model.
- We note the present best solution (PBS)
- At every x there is some chance of improving on the PBS.
- Then we ask: Assuming an improvement over the PBS, where is it likely to be largest?



Consider the point x=0.8, and the random variable Y, which is the possible values of the function there. Its mean is the kriging prediction, which is slightly above zero.



$$E[I(x)] \equiv E[\max(y_{PBS} - Y, 0)]$$

$$E[I(x)] = (y_{PBS} - \hat{y}(x))\Phi\left(\frac{y_{PBS} - \hat{y}(x)}{s(x)}\right) + s(x)\phi\left(\frac{y_{PBS} - \hat{y}(x)}{s(x)}\right)$$

y_{PBS}: Present Best Solution
$\hat{y}(x)$: Kriging Prediction
$s(x)$: Prediction standard deviation $\left(\sqrt{Prediction\ Variance}\right)$
Φ: Cumulative density function of standard normal distribution
ϕ: Probability density function of standard normal distribution

Expected Improvement (EI)

- Idea can be found in Mockus's work as far back as 1978.
- Expectation of improving beyond the Present Best Solution (PBS) or current best function value, yPBS.

$$EI(x) = \underbrace{\underbrace{\left(y_{PBS} - \hat{y}(x)\right)}_{\text{Predicted difference between current minimum and prediction at } x} \underbrace{\Phi\left(\frac{y_{PBS} - \hat{y}(x)}{s(x)}\right)}_{\text{Penalized by area under the curve}}}_{\text{Large when } \hat{y} \text{ is small with respect to } y_{PBS}\text{, promotes exploitation}} + \underbrace{s(x)\phi\left(\frac{y_{PBS} - \hat{y}(x)}{s(x)}\right)}_{\text{Large when s(x) is large, promotes exploration}}$$

Balance between seeking promising areas of design space and the uncertainty in the model.

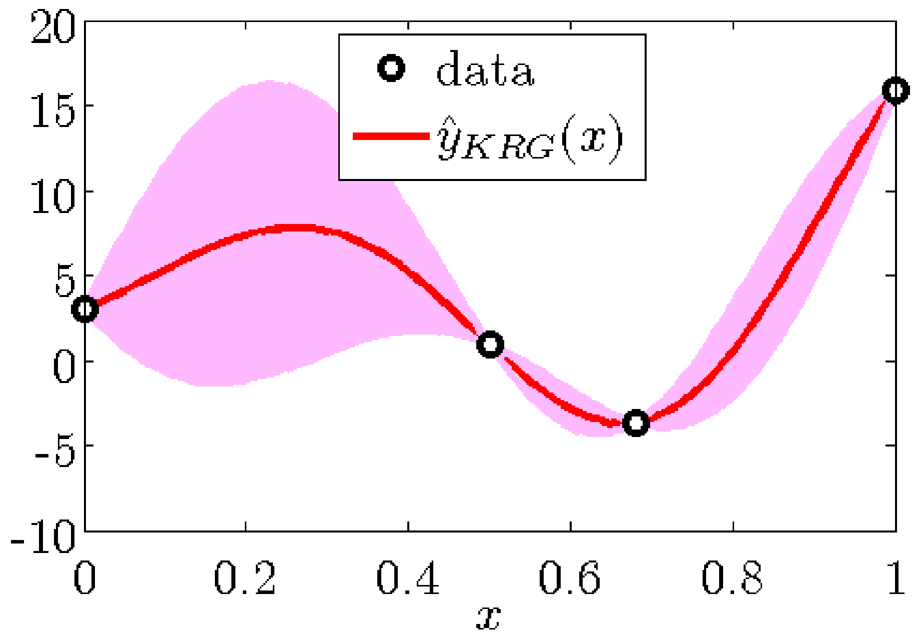
Mockus, J., Tiesis, V. and Zilinskas, A. (1978), The application of Bayesian methods for seeking the extremum, in L.C.W. Dixon and G.P. Szego (eds.), *Towards Global Optimisation*, Vol.2, pp. 117–129. North Holland, Amsterdam.

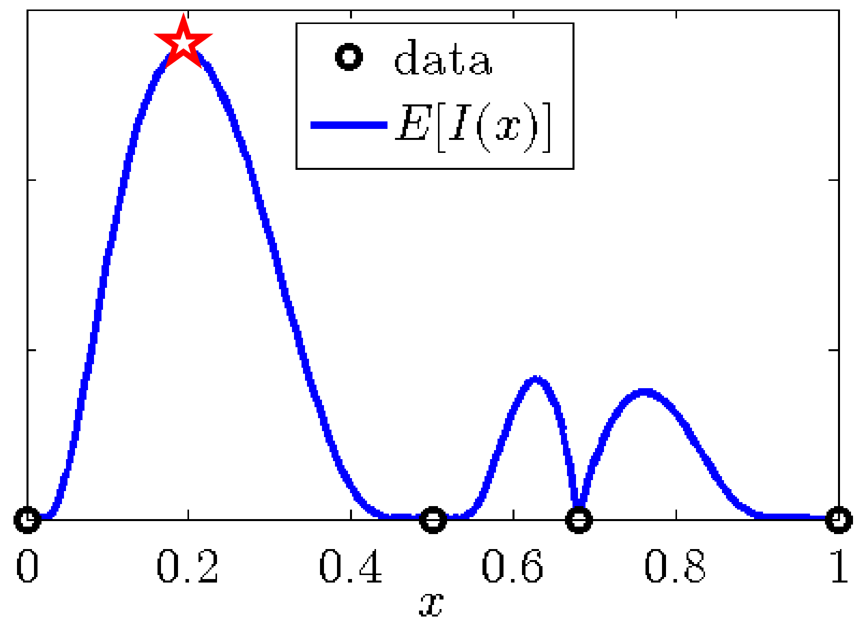
Exploration and Exploitation

EGO maximizes $E[I(x)]$ to find the next point to be sampled.

- The expected improvement balances exploration and exploitation because it can be high either because of high uncertainty or low surrogate prediction.

- When can we say that the next point is "exploration?"





Problem Expected Improvement

- What is unusual in the way the expected improvement is calculated?
- If the kriging fit shown on the previous slide is very close to the true function, give the coordinates of four sampling points that would lead to the next step being exploitation.

EGO with Probability of Target Improvement (PI)

- EGO with EI is expensive to find multiple samples at a time
- EGO-PI uses the probability of improvement beyond a given target as the selection criterion
  - Maximizing PI can balance local and global searches
- Performance can be sensitive to the target value
  - If the target is too ambitious, the search is excessively global and slow to focus on promising areas
  - If the target is too modest, there is exhaustive search around the PBS before moving to global search
- EGO-AT: adapts the target for each cycle according to the success of meeting the target in the previous cycle

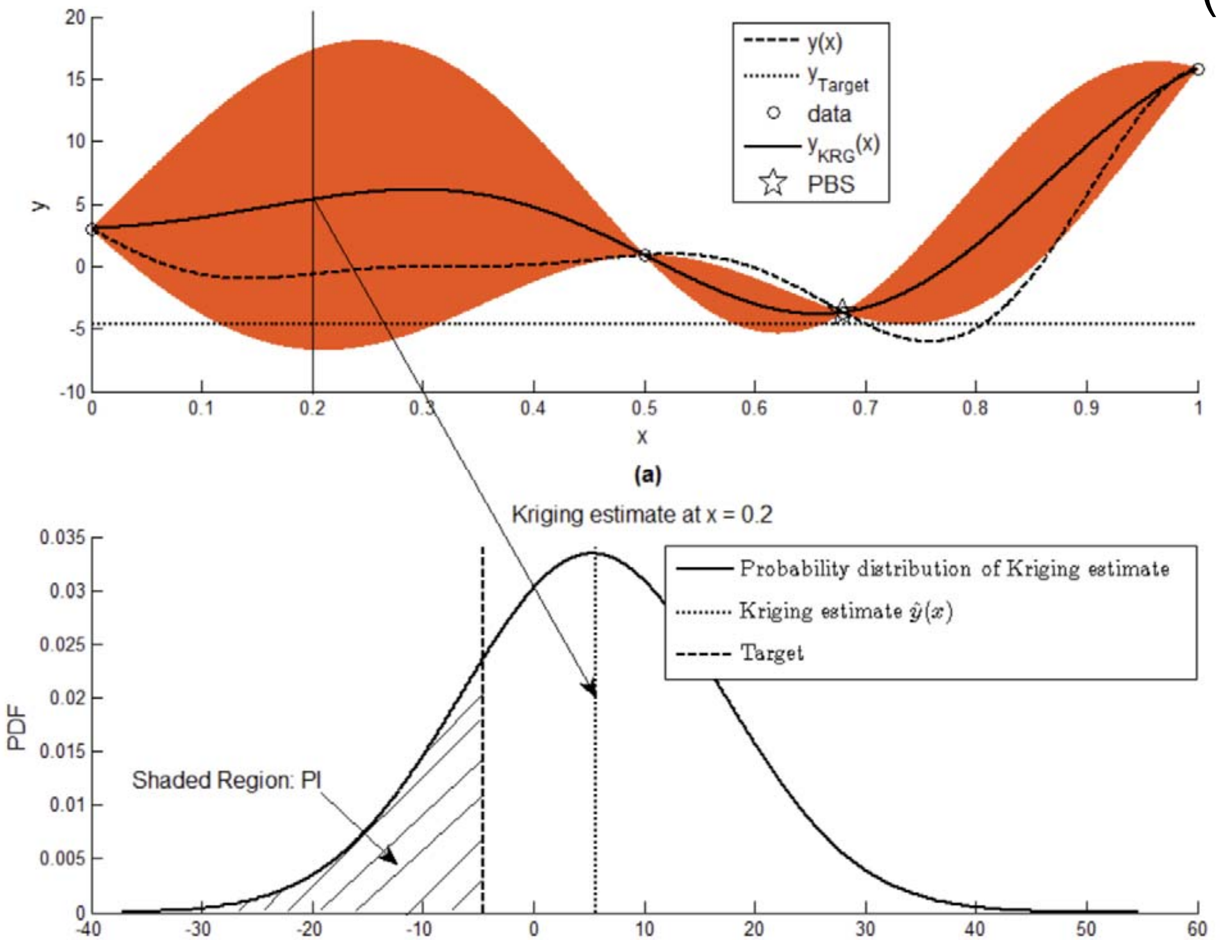
Probability of Target Improvement (PI)

- Probability of improving the target beyond y_{Target} at x

$$PI(x) = \Phi\left(\frac{y_{\text{Target}} - \hat{y}(x)}{s(x)}\right)$$

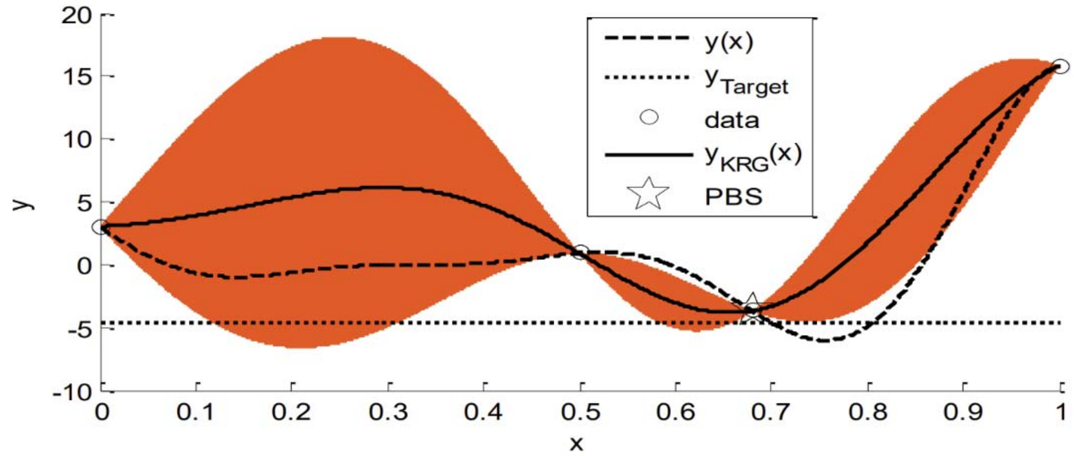
$y_{\text{Target}} = y_{PBS} - TI$

TI: Target of improvement (normally 10%)

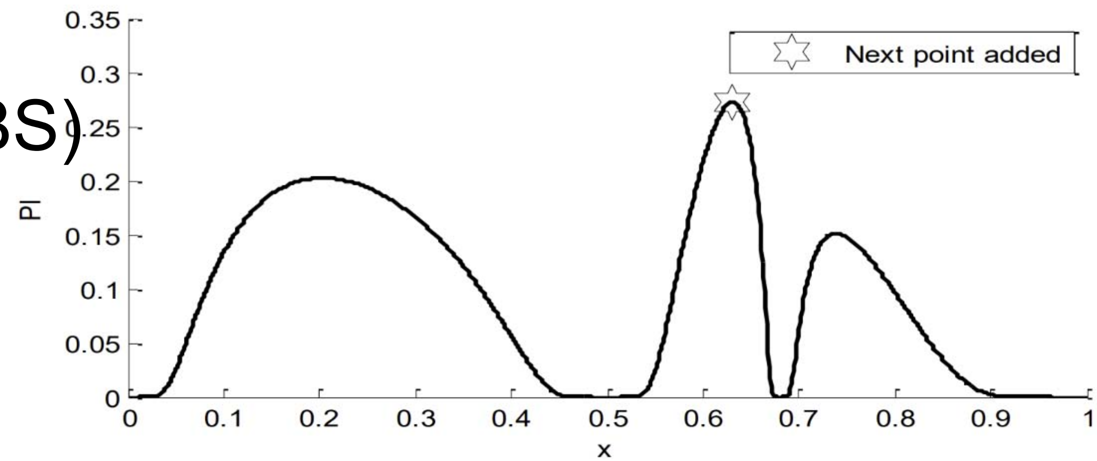


Maximum PI for Adding a Sample

- True function: $y(x) = (6x-2)^2 \sin(12x-4)$
- Initial sample : x = [0 0.5 0.68 1]
- Uncertainty 2*s(x)



- Maximum PI at x=0.62 (exploitation of better PBS)



Adding Multiple Samples Simultaneously

- Good for parallel processing
- Find n_s local minima with different initial designs and add samples there
- Prevent too-close samples by adding eps at every sample

$$eps = 0.1(x_{\max} - x_{\min})\sqrt{n_{dim}}$$

- Repeat until n_s samples are added
- Pick multiple competing optima by putting an exclusion radius

EGO with Adaptive Target (EGO-AT)

- Percentage-based TI: Difficult near the optimum
- Adaptive target: $y_{Target_k} = y_{PBS_k} - TI_k$
- Target of improvement:

$$TI_{k+1} = \begin{cases} 1.5TI_k, & \text{if } \eta_k > 2 \\ 0.5TI_k(\eta_k + 1), & \text{if } 0.05 \le \eta_k \le 2 \\ 0.5TI_k, & \text{if } \eta_k < 0.05 \end{cases}$$

- Improvement ratio

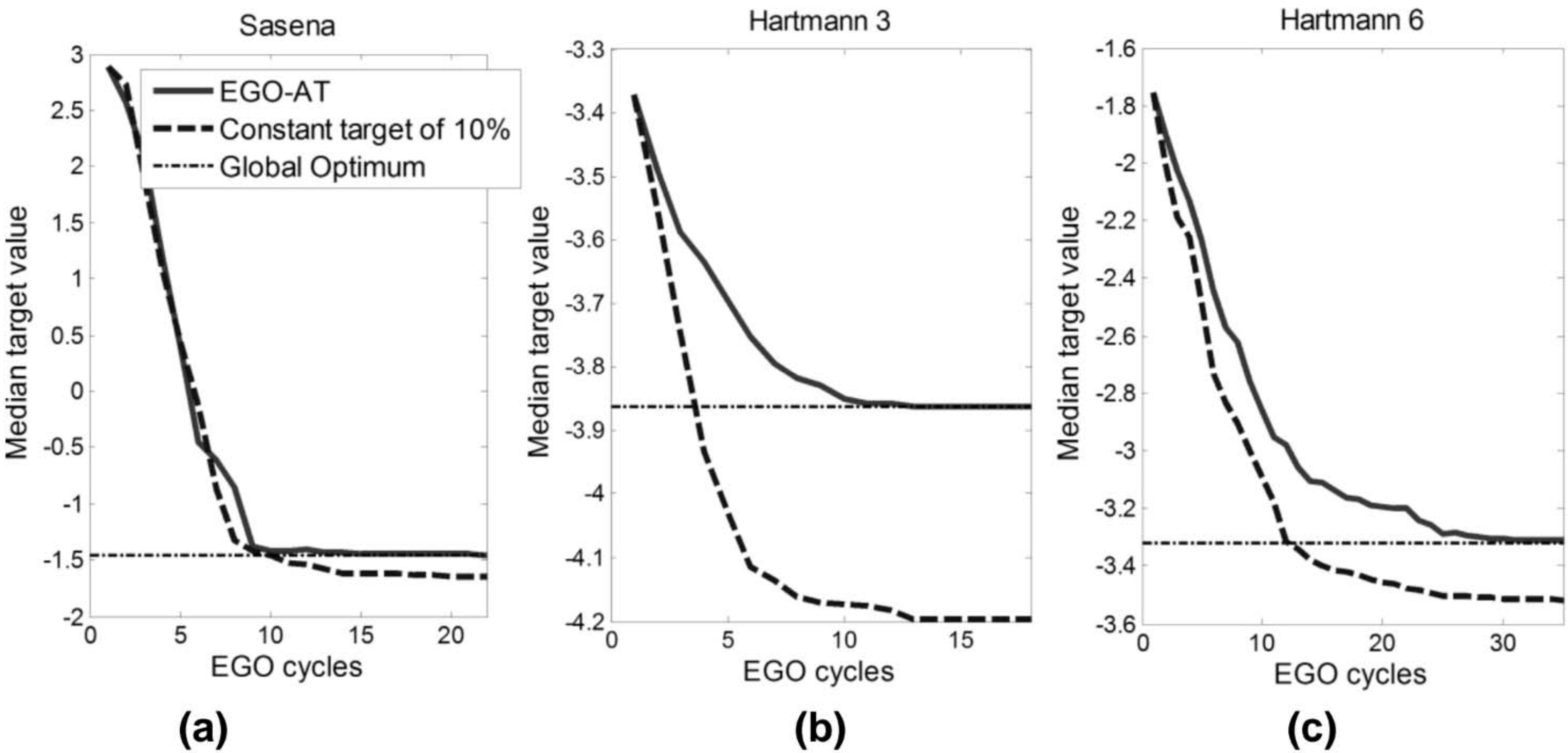
$$\boxed{\eta_k = \frac{y_{BS_k} - y_{PBS_k}}{y_{Target_k} - y_{PBS_k}}}$$

y_{BS_k}: Best value among added sample at k

y_{PBS_k}: Best value before adding samples at k

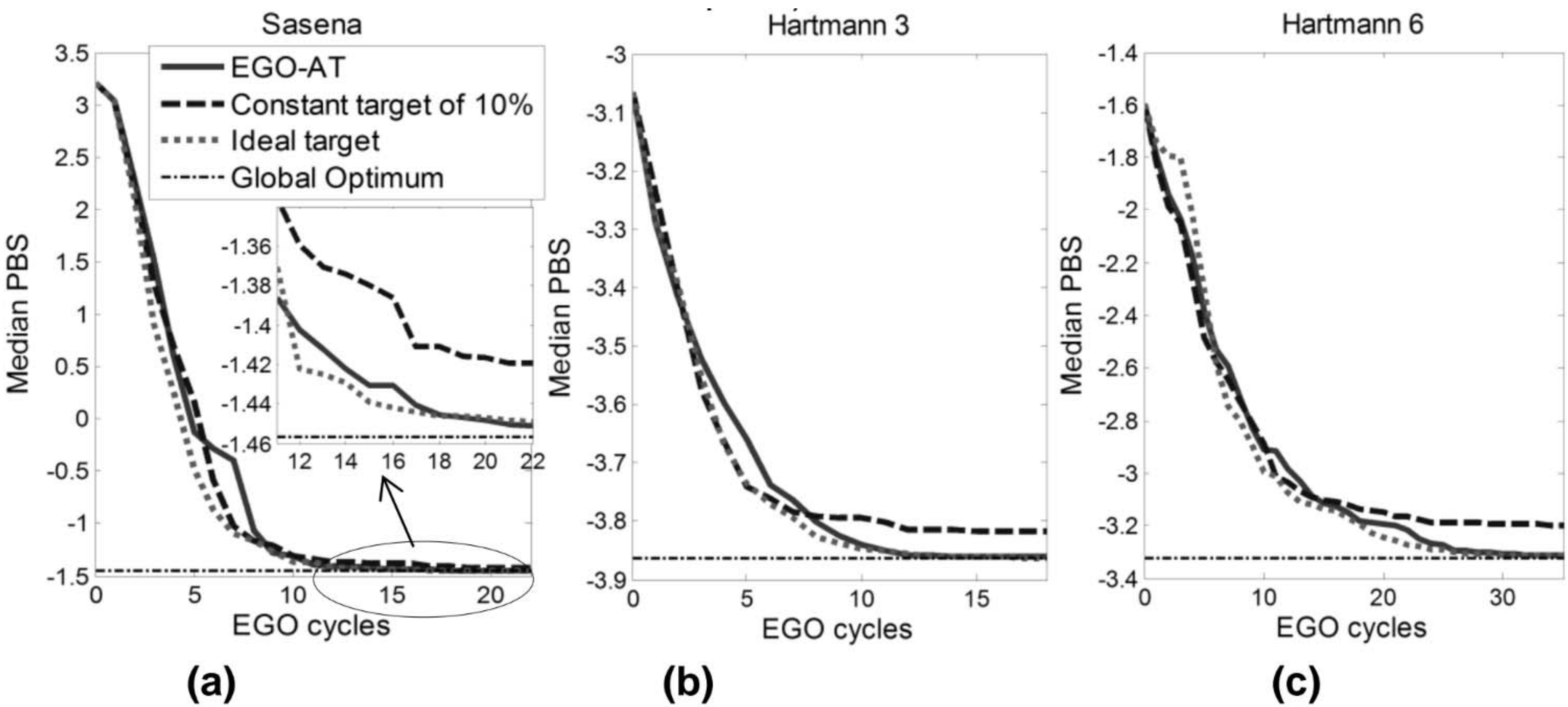
Accuracy of Target for EGO-AT

- y_{Target} of EGO-AT converges to the global optimum
- y_{Target} with constant TI overshoots the global optimum



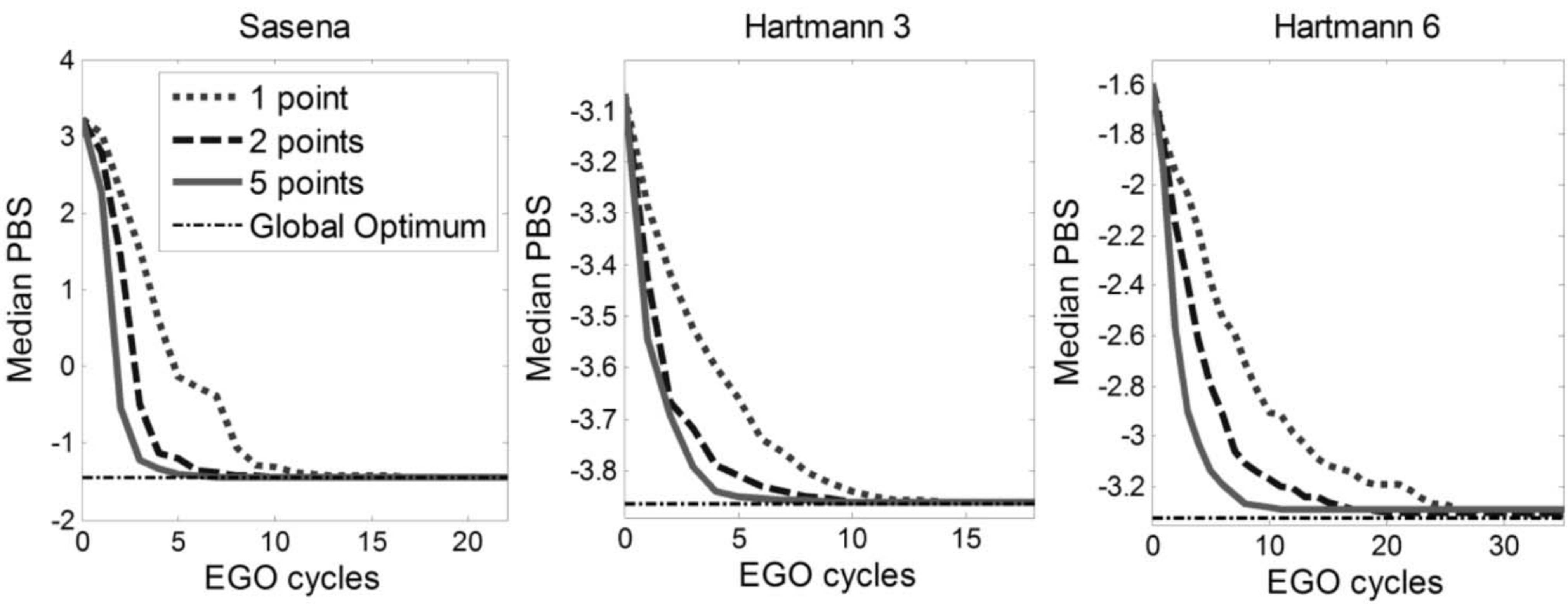
Convergence of EGO-AT

- EGO-AT performs the same as the ideal target case
- EGO with constant TI consistently under-performed



Performance of EGO-AT with Adding Multiple Samples

- Adding multiple samples reduces the EGO cycles, but increases the total number of samples



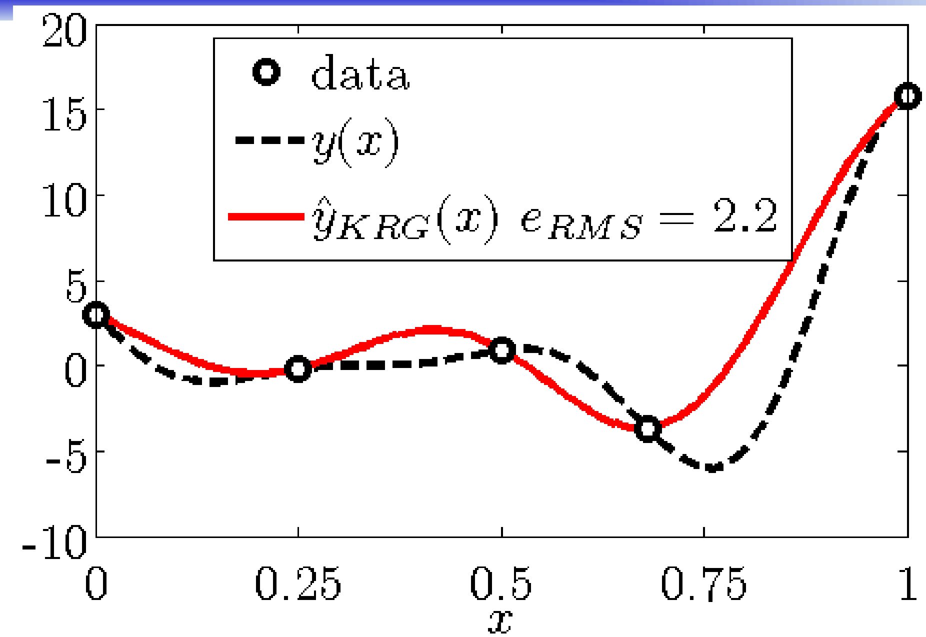
(a) (b) (c)

Expanding EGO to Surrogates Other Than Kriging

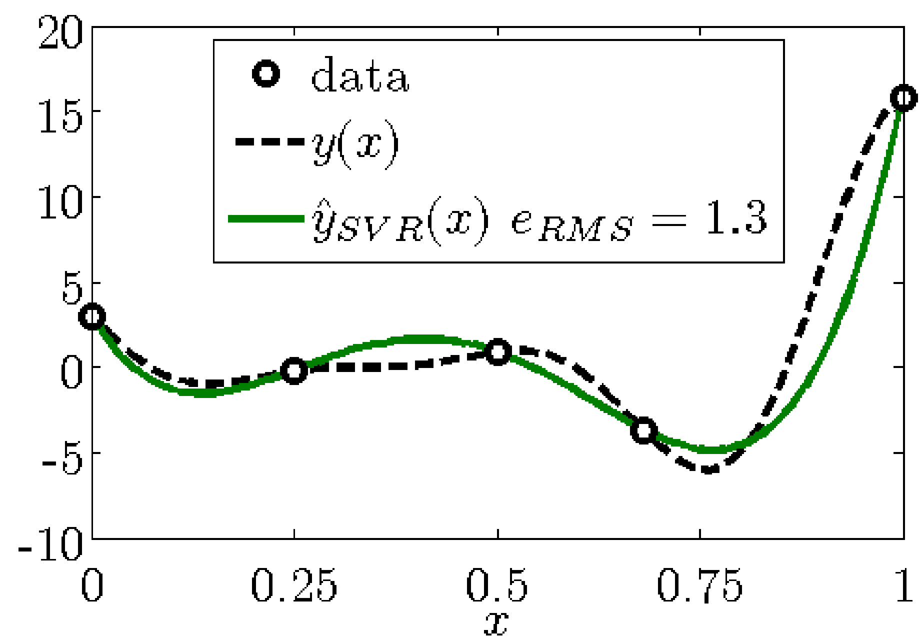
Considering the root mean square error, e_{RMS} :

$$e_{RMS} = \sqrt{\int_0^1 |e(x)|^2 \, dx}$$

We want to run EGO with the most accurate surrogate. **But we have no uncertainty model for SVR**

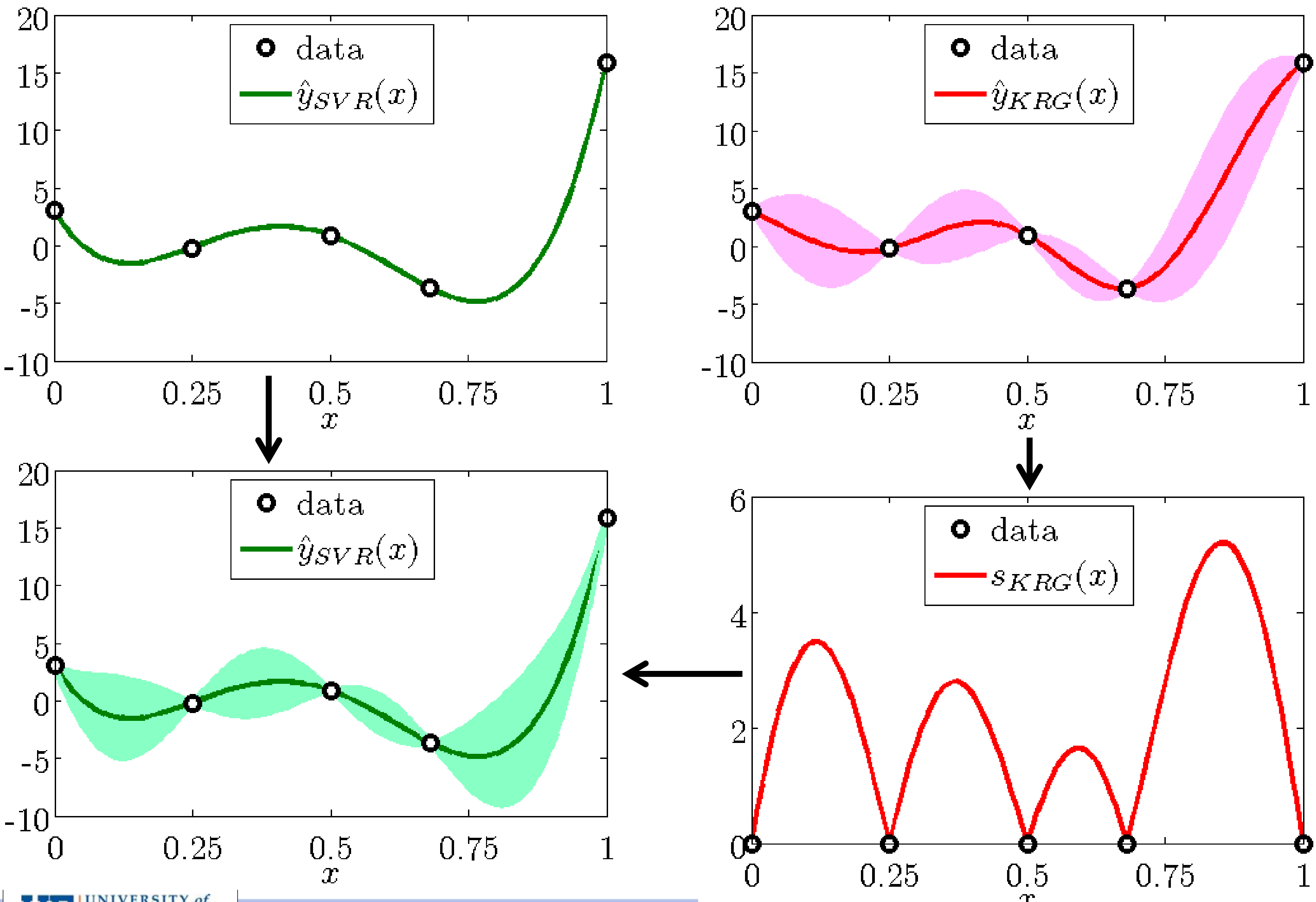


(a) Kriging



(b) Support vector regression

Importation of Uncertainty Model



Hartmann-3 Example

Hartmann 3 function (initially fitted with 20 points):

$$y(\mathbf{x}) = -\sum_{i=1}^{4} a_i \exp\left[-\sum_{j=1}^{3} B_{ij}\left(x_j - D_{ij}\right)^2\right],$$

$$0 \le x_j \le 1, \; j = 1, 2, 3.$$

$$\mathbf{B} = \begin{bmatrix} 3.0 & 10.0 & 30.0 \\ 0.1 & 10.0 & 35.0 \\ 3.0 & 10.0 & 30.0 \\ 0.1 & 10.0 & 35.0 \end{bmatrix} \quad \mathbf{D} = \begin{bmatrix} 0.3689 & 0.1170 & 0.2673 \\ 0.4699 & 0.4387 & 0.7470 \\ 0.1091 & 0.8732 & 0.5547 \\ 0.03815 & 0.5743 & 0.8828 \end{bmatrix}$$

$$PRESS_{RMS}^{KRG} = 0.89$$

$$e_{RMS}^{KRG} = 0.50$$

$$PRESS_{RMS}^{RBNN} = 0.66$$

$$e_{RMS}^{RBNN} = 0.48$$

After 20 iterations (i.e., total of 40 points), improvement (I) over initial best sample (IBS):

$$\boxed{I = \frac{y_{IBS} - y_{optim}}{|y_{IBS}|}}$$

y_{IBS}: initial best sample
y_{optim}: best sample after 20 EGO cycles

$$I_{KRG} = 3.5\% \qquad I_{RBNN} = 27.2\%$$

Two other Design of Experiments

FIRST:

$$PRESS_{RMS}^{KRG} = 0.98 \qquad PRESS_{RMS}^{RBNN} = 0.84$$

$$e_{RMS}^{KRG} = 0.64 \qquad e_{RMS}^{RBNN} = 0.47$$

$$I_{KRG} = 9.8\% \qquad I_{RBNN} = 18.2\%$$

SECOND:

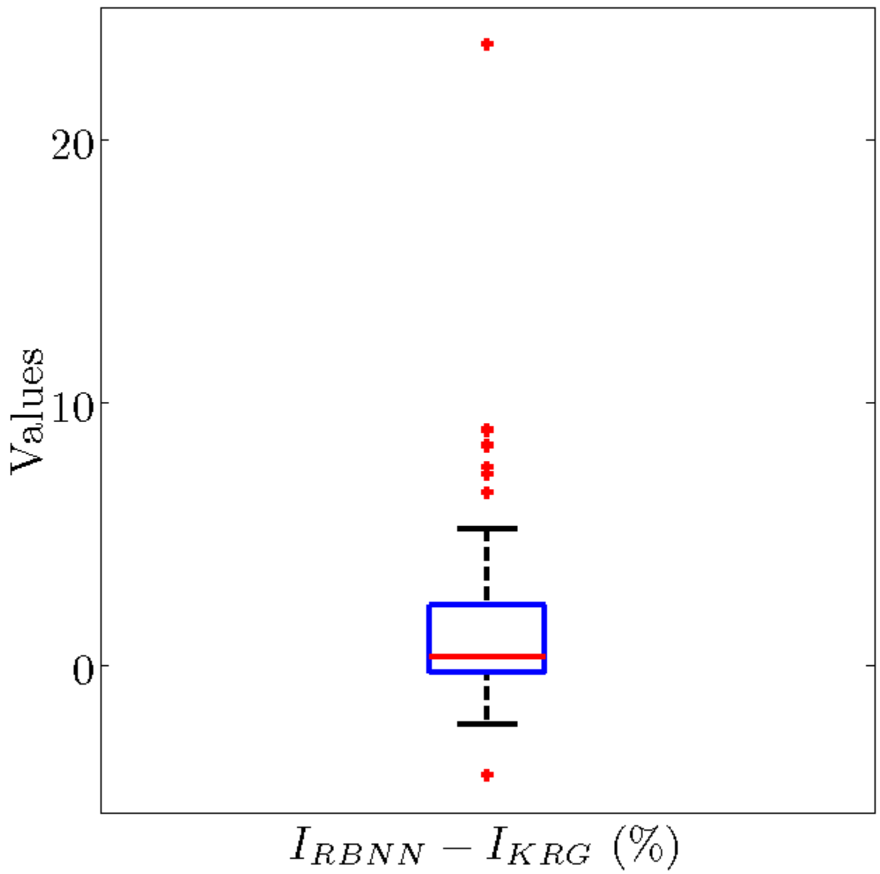
$$PRESS_{RMS}^{KRG} = 0.67 \qquad PRESS_{RMS}^{RBNN} = 0.61$$

$$e_{RMS}^{KRG} = 0.65 \qquad e_{RMS}^{RBNN} = 0.53$$

$$I_{KRG} = 32.3\% \qquad I_{RBNN} = 41.3\%$$

Summary: Hartmann-3 example

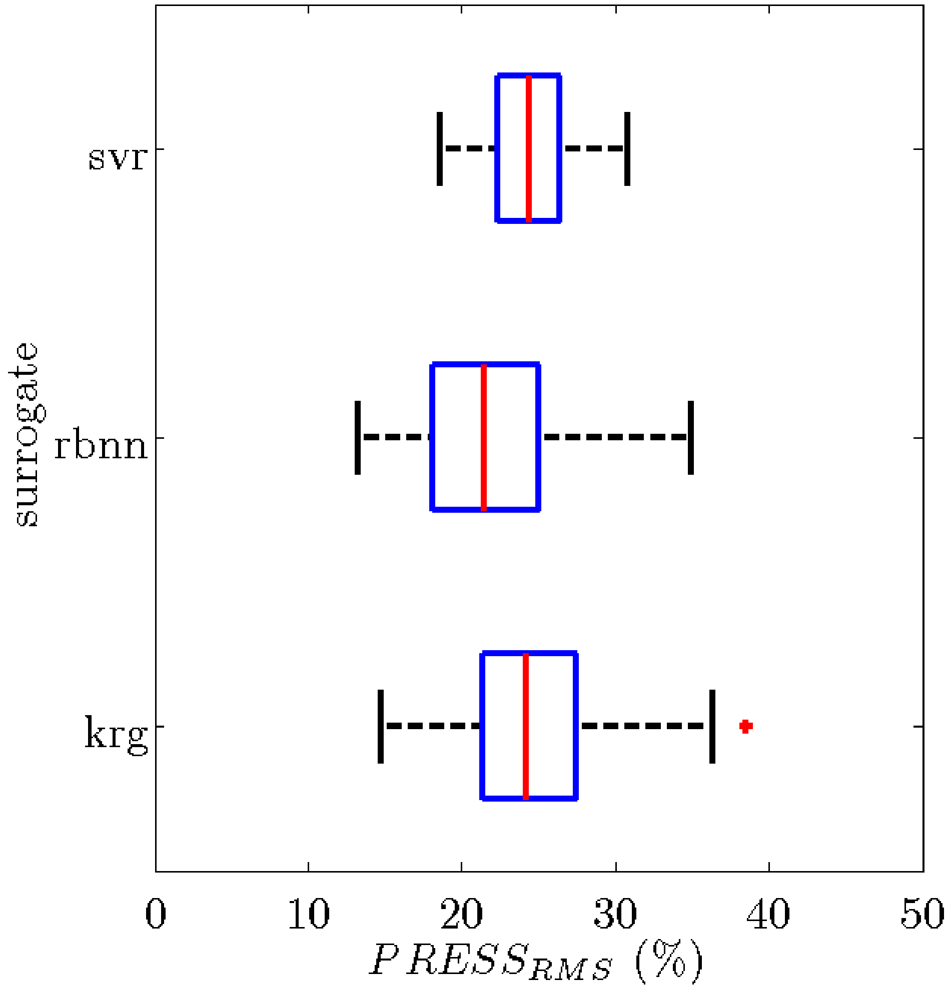
Box plot of the difference between improvement offered by different surrogates after 20 iterations (out of 100 DOEs)



In 34 DOEs (out of 100) KRG outperforms RBNN (in those cases, the difference between the improvements has mean of only 0.8%).

Potential of EGO with Multiple Surrogates

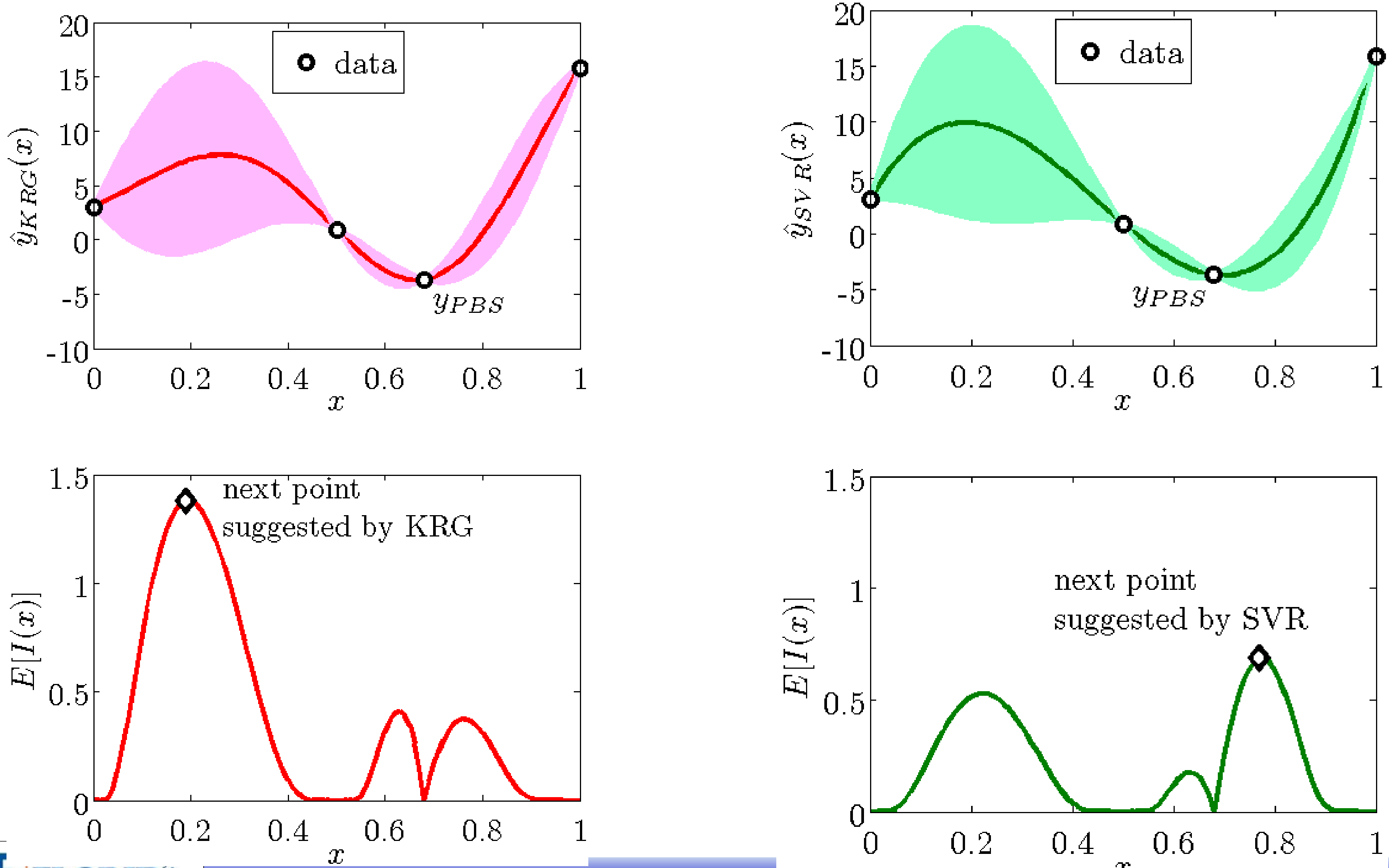
Hartmann 3 function (100 DOEs with 20 points)



Overall, surrogates are comparable in performance.

EGO with Multiple Surrogates (MSEGO)

Traditional EGO uses kriging to generate one point at a time.
We use multiple surrogates to get multiple points.

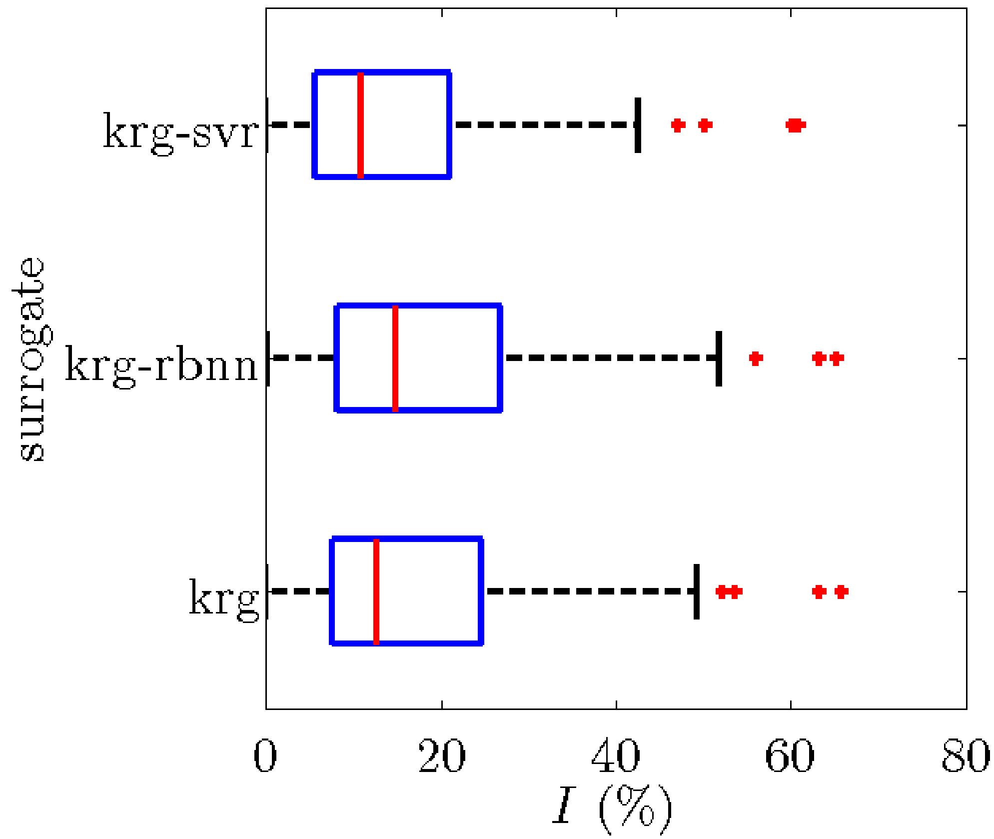


EGO with Multiple Surrogates (MSEGO)

"krg" runs EGO for **20 iterations** adding **one point** at a time.

"krg-svr" and **"krg-rbnn"** run **10 iterations** adding **two points/iteration**.

$$I = \frac{y_{IBS} - y_{optim}}{|y_{IBS}|}$$



Multiple surrogates offer good results in half of the time!!!

EGO with Multiple Surrogates and Multiple Sampling Criteria

Multiple sampling (or infill) criteria (each of them also have the capability of adding multiple points per optimization cycle):

- EI: Expectation of improvement beyond the present best solution.
  - Multipoint EI (q-EI), Kriging Believer, Constant Liar
- PI: Probability of improving beyond a set target.
  - Multiple targets, multipoint PI

Multiple sampling criteria with multiple surrogates has very high potential of:

- Providing insurance against failed optimal designs.
- Providing insurance against inaccurate surrogates.
- Leveraging parallel computation capabilities.

Problem EGO

- We are given a function with four pairs of (x,y): (0,3), (0.5,1), (0.7,3), (1,17), similar to the example used in this lecture.
  - Fit a kriging surrogate and a quadratic polynomial to the data.
  - Plot and compare their standard error as function of x (square root of prediction variance)
  - Plot and compare their expected improvements as functions of x.
  - Where would EGO place the next sampling point for each surrogate?

43