An Approach to Minimize the Number of Function Evaluations in Global Optimization

Tomi Haanpää
Department of Mathematical Information Technology
University of Jyväskylä
8.3.2010

Thesis

- Working title: Approximating the Pareto Front with Meta Models
- Supervisors: Kaisa Miettinen and Jussi Hakanen
- Monograph, 2012
- Contribution: To approximate the Pareto Front so that the decision maker can “move” on it.
- Relation to this presentation: The approach will be modified to approximate holes on the front.

Content

- Global Optimization
- Lipschitzian
- Radial Basis Functions
- The Approach
- Numerical Examples
- References

Global Optimization (1/3)

- A considered Global Optimization Problem (GOP) is

$$\min_{x \in S} f(x),$$

where

  - $S \subset \mathbb{R}^n$ is compact,
  - f is a computationally expensive continuous function from S into $\mathbb{R}$.

- In other words, the problem is to find out $x^* \in S$ so that

$$f(x^*) \leq f(x) \text{ for all } x \in S$$

with a small number of function evaluations.

Global Optimization (2/3)

Solving techniques for computationally expensive objective functions, based on a given data

$$\mathcal{D} = \{(x^1, y^1), \ldots, (x^i, y^i = f(x^i)), \ldots, (x^k, y^k)\} \subset S \times \mathbb{R},$$

- Support vector regression [Vapnik, 1998]
- A radial basis function method [Gutmann, 2001]
- The EGO algorithm [Jones et al., 1998]

However, each of the above methods requires a global optimization sub problem to be solved.

Global Optimization (3/3)

The goal and motivation of this study:

- Let us assume that we have an algorithm, which produce a solution candidate x^c to the GOP.
- If x^c is close to a vector x^j for which the corresponding y^j is far from the so-far best known minimum of f, that is,

$$y^* = \min_{i=1,\ldots,k} y^i,$$

  then $f(x^c)$ is close to y^j, because f has been assumed to be continuous.
- We can save computation time by rejecting the evaluation of $f(x^c)$.
- We want a computationally efficient approach to decide whether to evaluate $f(x^c)$ or not.

Lipschitzian (1/3)

Function f is said to be *Lipschitzian*, if there exists a *Lipschitz* constant $l > 0$ so that

$$|f(x) - f(y)| \leq l\|x - y\| \text{ for all } x, y \in S,$$

where $\|\cdot\|$ is a norm on $\mathbb{R}^n$ (for example *Euclidean* norm).

- Function f from $\mathbb{R}$ into $\mathbb{R}$, $f(x) = 2x$, is Lipschitzian.

$$|f(x) - f(y)| = |2x - 2y| \leq 2|x - y|.$$

- A differentiable function f from $\mathbb{R}^n$ into $\mathbb{R}$ is Lipschitzian, if there exists $l > 0$ so that $\|\nabla f(z)\| \leq l$ for all $z \in \mathbb{R}^n$.

$$|f(x) - f(y)| \leq \sup_{z \in \mathbb{R}^n} \|\nabla f(z)\| \|x - y\| \leq l\|x - y\|.$$

Lipschitzian (2/3)

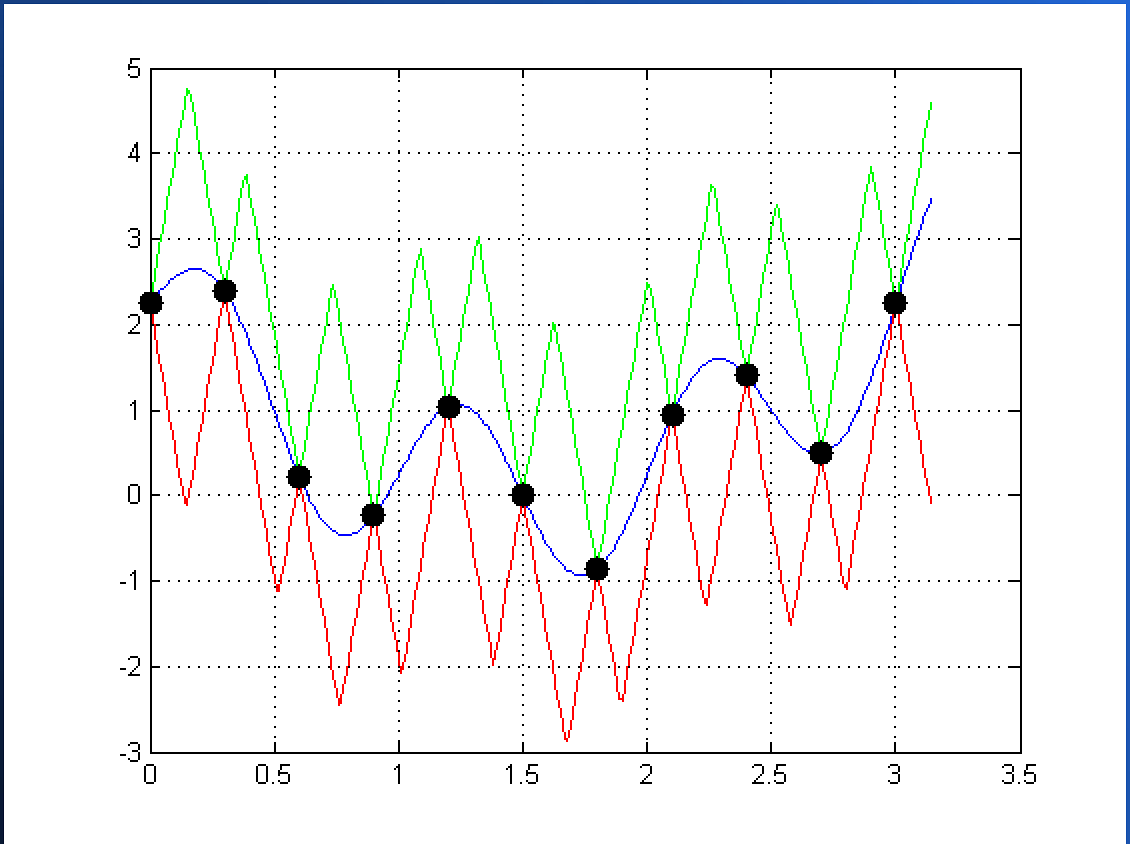
Global optimization solving algorithms based on the Lipschitzian property

- The DIRECT algorithm [Jones et al., 1993]
- Partition methods [Pintér, 1996]
  - Lower and upper approximations based on l and $\mathcal{D}$

$$f_l(x) = \max_{i=1,\ldots,k} \left(y^i - l\|x - x^i\| \right),$$
$$f_u(x) = \min_{i=1,\ldots,k} \left(y^i + l\|x - x^i\| \right).$$

  - However, the Lipschitz constant l must be known or approximated.

Lipschitzian (3/3)



Radial Basis Functions (1/3)

A basic Radial Basis Function [Buhmann, 2003] (RBF) g from S into $\mathbb{R}$, based on $\mathcal{D}$, is

$$g(x) = \sum_{i=1}^{k} \lambda_i \phi(\|x - x^i\|),$$

where

- coefficients λ_i are solution of a system of linear equations

$$\begin{bmatrix} \phi(\|x^1 - x^1\|) & \cdots & \phi(\|x^1 - x^k\|) \\ \vdots & \ddots & \vdots \\ \phi(\|x^k - x^1\|) & \cdots & \phi(\|x^k - x^k\|) \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_k \end{bmatrix} = \begin{bmatrix} y^1 \\ \vdots \\ y^k \end{bmatrix},$$

- function ϕ from $\mathbb{R}$ into $\mathbb{R}$ is a given *basis function*.

Radial Basis Functions (2/3)

The given basis function can be, for example,

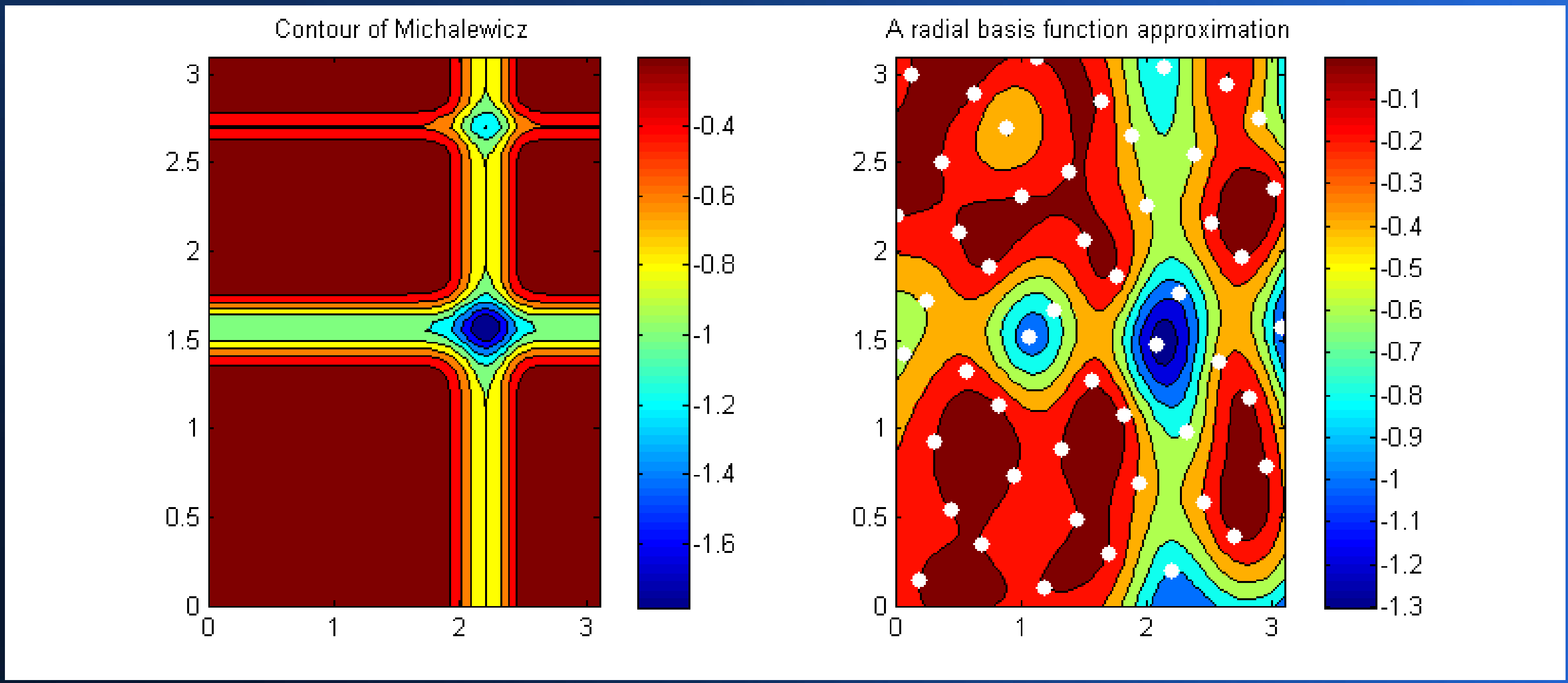
- a polyharmonic spline: $\phi(r) = r^d$, $d = 1, 3, 5, \ldots$
- thin plate spline: $\phi(r) = r^d \ln r$, $d = 2, 4, 6, \ldots$
- multiquadric: $\phi(r) = \sqrt{r^2 + \varepsilon^2}$, $\varepsilon > 0$
- Gaussian: $\phi(r) = e^{-\varepsilon r^2}$, $\varepsilon > 0$

We consider an RBF with a polyharmonic spline of third degree ($d = 3$), because it is smooth and Lipschitzian on S. Let such an RBF be denoted by f_{RBF} and let the Lipschitz constant be l_{RBF}.

- We have an approximation l_{RBF} for l.
- We have lower and upper approximations for f.

Radial Basis Functions (3/3)

- Michalewicz function: $f(x) = -\sum_{i=1}^{2} \sin(x_i) \left(\sin \left(\frac{i x_i^2}{\pi} \right) \right)^{20}$
- f_{RBF} with 50 *Hammersley* points [Kalagnanam and Diwekar, 1997]



The Approach (1/3)

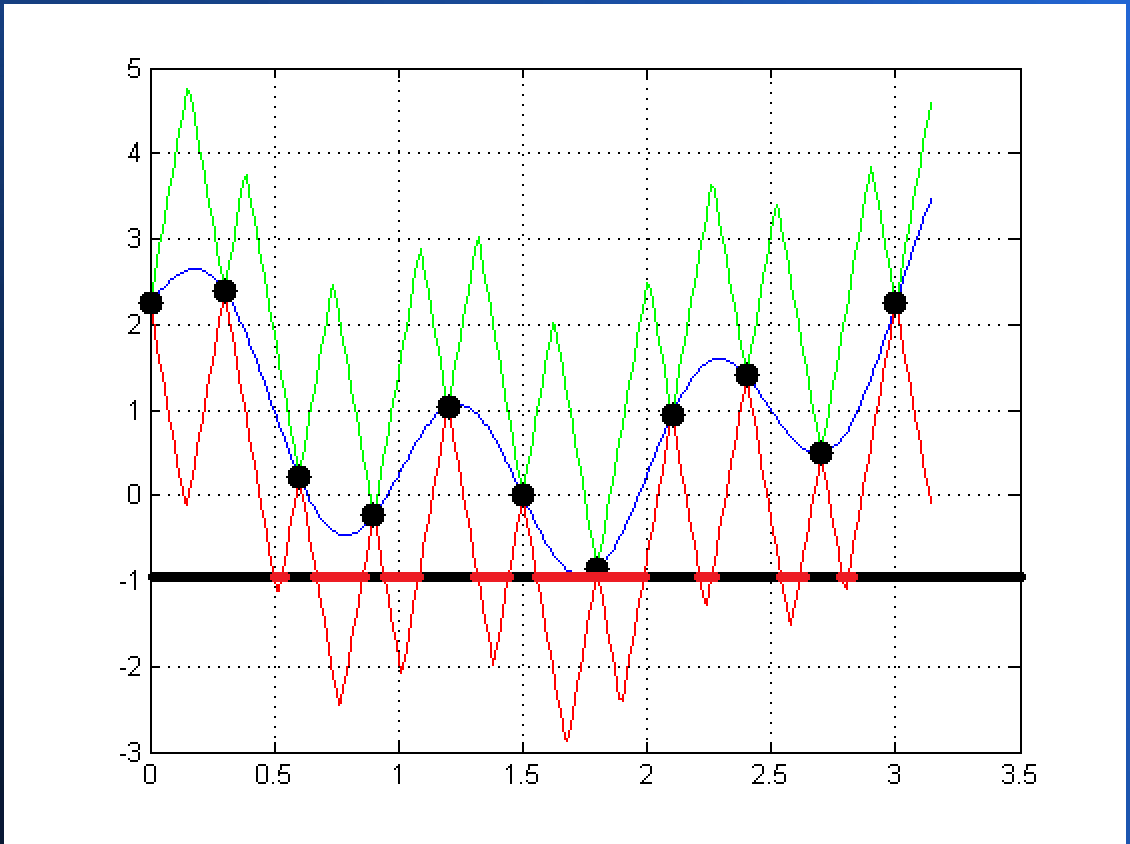
By using the Lipschitzian property (Partition methods), l_{RBF} and $\mathcal{D}$, we can separate S into two regions:

- The most promising region A_1 is a region where x^* of the GOP may exists.
  - $A_1 = \{x \in S : f_l(x) \leq y^*\}$
- The region A_{-1}, which is out of interest, is a region where x^* does not exists.
  - $A_{-1} = \{x \in S : f_l(x) > y^*\}$

In this way, we

- can reject the costly evaluation of $f(x^c)$, in the case that $x^c \in A_{-1}$,
- and try to focus on evaluations on A_1.

The Approach (2/3)



The Approach (3/3)

However,

- the approach does not produce solution candidates
- it only approximates whether to evaluate or not.

This is based on how well f_{RBF} approximates f.

- It is possible that f_{RBF} fails to approximate f, that is, $l_{RBF} < l$ caused by the lack of given information of f in $\mathcal{D}$.
- In this case, the most promising region A_1 may loose the track to the true minimum of the GOP.

Numerical Examples (1/2)

Test functions:

- Rosenbrock:

$$f(x) = 100(x_1^2 - x_2)^2 + (x_1 - 1)^2$$

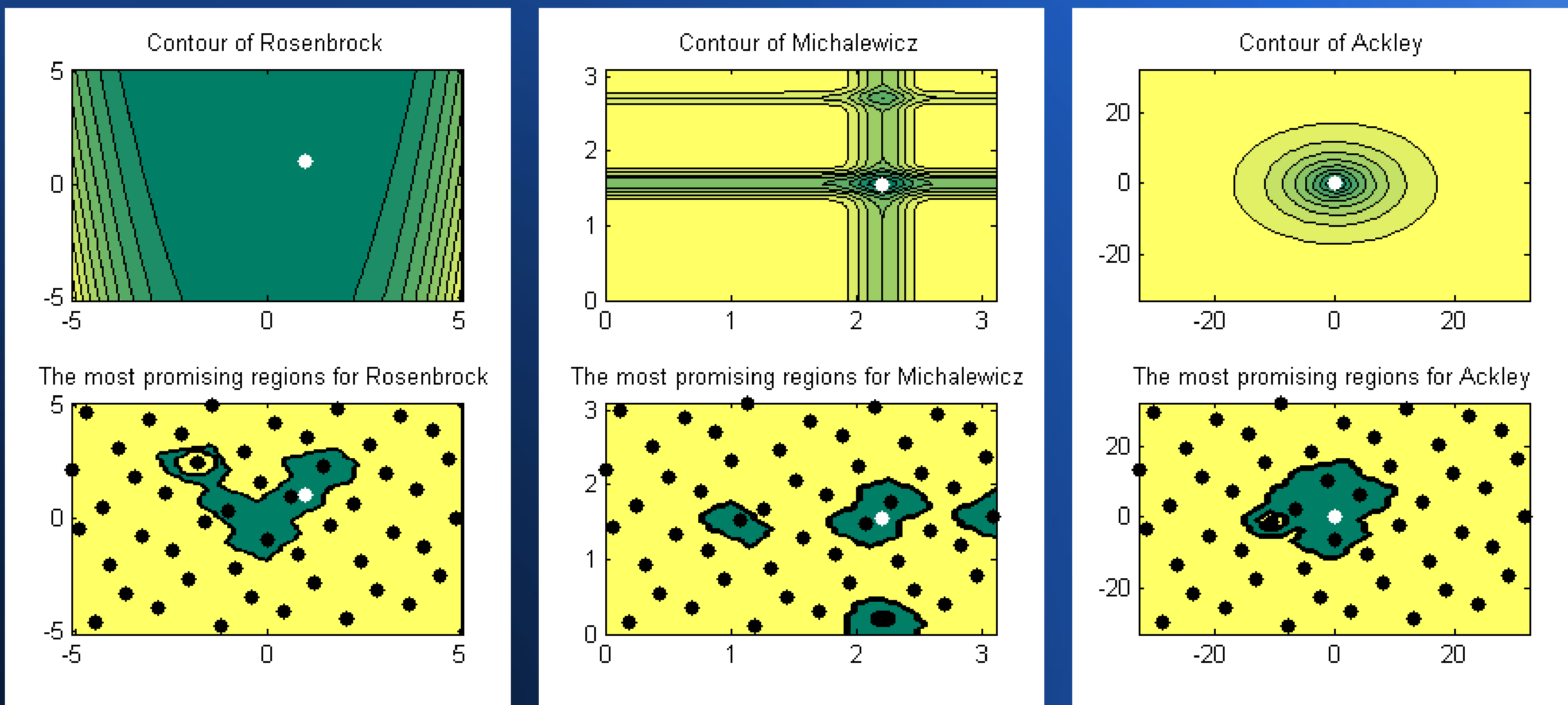
- Michalewicz:

$$f(x) = -\sum_{i=1}^{2} \sin(x_i) \left(\sin \left(\frac{i x_i^2}{\pi} \right) \right)^{20}$$

- Ackley:

$$f(x) = 20 + e - 20 e^{-\sqrt{\frac{1}{5} \sum_{i=1}^{2} x_i^2}} - e^{-\frac{1}{2} \sum_{i=1}^{2} \cos(2\pi x_i)}$$

Numerical Examples (2/2)



Conclusion

- A computationally efficient approach has been constructed to approximate the most promising region where the true minimum of the GOP may exists.
  - The approach is based on lower and upper approximations and a required approximation for the Lipschitz constant is derived from a radial basis function.
- The approach is able to decide whether to evaluate the computationally expensive objective function or not.
  - One can save computation time by rejecting the function evaluation for which an approximated value is not going to be smaller than the known minimum value of GOP.

Thank you!

References

[Buhmann, 2003] Buhmann, M. (2003). *Radial Basis Functions: Theory and Implementations*. Cambridge University Press.

[Gutmann, 2001] Gutmann, H.-M. (2001). A radial basis function method for global optimization. *Journal of Global Optimization*, 19(3):201–227.

[Jones et al., 1993] Jones, D. R., Perttunen, C. D., and Stuckman, B. E. (1993). Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Application*, 79(1):157–181.

[Jones et al., 1998] Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient global optimization of expensive black-box functions. *Journal of Global Optimization*, 13(4):445–492.

[Kalagnanam and Diwekar, 1997] Kalagnanam, J. R. and Diwekar, U. M. (1997). An efficient sampling technique for off-line quality control. *Technometrics*, 39(3):308–319.

[Pintér, 1996] Pintér, J. (1996). *Global Optimization in Action*. Kluwer Academic Publishers.

[Vapnik, 1998] Vapnik, V. N. (1998). *Statistical Learning Theory*. Wiley-Interscience.