Mixture Surrogate Models Based on Dempster-Shafer Theory for Global Optimization

Juliane Müller juliane.mueller@tut.fi

Department of Mathematics, Tampere University of Technology, Finland

Motivation and Objective

Optimization is needed in many areas of life:

- engineering (aerospace, automobile industry, medical imaging)
- economics (utility maximization/expenditure minimization)
- operations research (transportation, scheduling, environmental applications)

Common for all problems is that a certain objective should be fulfilled. Considered are problems of the form

$$\min f(\mathbf{x}), \quad \mathbf{x} \in \mathcal{D} \subset \mathbb{R}^d.$$

Difficulties arising during the optimization process:

- objective function evaluation may become extremely time consuming due to simulations
- "black-box" - no gradient information available
- local optimality

Goal: the development of efficient algorithms for finding the global optimal solution, i.e. finding the optimum while using as few function evaluations as possible in order to achieve low computational complexity.

Surrogate Models

Surrogate models [3] are simplifications of simulation models

$$f(\mathbf{x}) = s(\mathbf{x}) + \epsilon$$

- $f(\mathbf{x})$ output of simulation model at point $\mathbf{x}$
- $s(\mathbf{x})$ output of surrogate model at point $\mathbf{x}$
- ϵ error

Surrogate models can be

- Interpolating
  - Radial basis function interpolant
  - Kriging
- Non-interpolating
  - Polynomial regression models
  - Multivariate adaptive regression splines

Which surrogate model should be used is problem dependent and in general unknown.

Additionally, mixture surrogate models have been introduced:

$$\hat{y}(\mathbf{x}) = \sum_{i=1}^{N_m} w_i \hat{y}_i(\mathbf{x}), \quad \sum_{i=1}^{N_m} w_i = 1 \qquad (1)$$

- $\hat{y}_i(\mathbf{x})$ prediction of ith contributing model
- $w_i \geq 0$ weight of ith contributing model
- N_m number of contributing models

The purpose of mixture models is to emphasize and restrict good and bad model characteristics, respectively, by adjusting the weights w_i.

Problem:

How should the weights w_i be adjusted?

Solution:

Dempster-Shafer theory

Dempster-Shafer Theory

Dempster-Shafer Theory (DST) [1, 2] is a mathematical theory of evidence. It allows to combine (conflicting) information from different sources. Conflicts can be redistributed amongst the information sources using different rules:

- Dempster's rule
- Yager's rule
- Inagaki's rule
- Proportional conflict redistribution rule (PCR5)

Functions that are generally used as decision criteria:

- Belief function BEL(A)
- Plausibility function PL(A)
- Pignistic probability BetP(A)

The interval [Bel(A), PL(A)] contains the precise probability of the set of interest A.

Proposed Algorithm

1. Construct initial experimental design (e.g. Latin hypercube sampling)
2. Evaluate costly objective function
3. Build different surrogate models
4. Use cross validation to obtain information about the models:
   - correlation coefficients
   - root mean squared errors
   - median absolute deviation
   - maximal absolute errors
5. Apply DST for calculating BetP of each model based on information from 4
6. Use BetP to determine the weights w_i of the models contributing to mixture models
7. Apply DST to determine BEL and PL of each (mixture) model
8. Choose model with highest BEL and build response surface
9. Find minimum of response surface
10. Evaluate costly objective function at the minimum site
11. Repeat 3 to 10 until stopping criterion met

Experimental Results

Experiments have been conducted on six global optimization benchmark problems with two to six variables. The numerical results lead to the following conclusions

\+ thorough examination of variable domain (global minima could be detected)

\+ approximation of global minima in most cases with error $< 1\%$

\+ only limited number of function evaluations needed

\- difficulties approximating very steep minima

• the choice of the conflict redistribution rule influences the results

• algorithm's computation time can be considered negligible compared to time needed for function evaluation

• mixture models prove better for higher dimensional problems

Consider the Branin function as example. This two-dimensional problem has three global minima. The surface and contour plots are illustrated in Figure 1. The approximated contour and surface plots at various stages of the algorithm are illustrated in Figure 2 (red asterisks denote the true minima, black dots denote the samples taken).

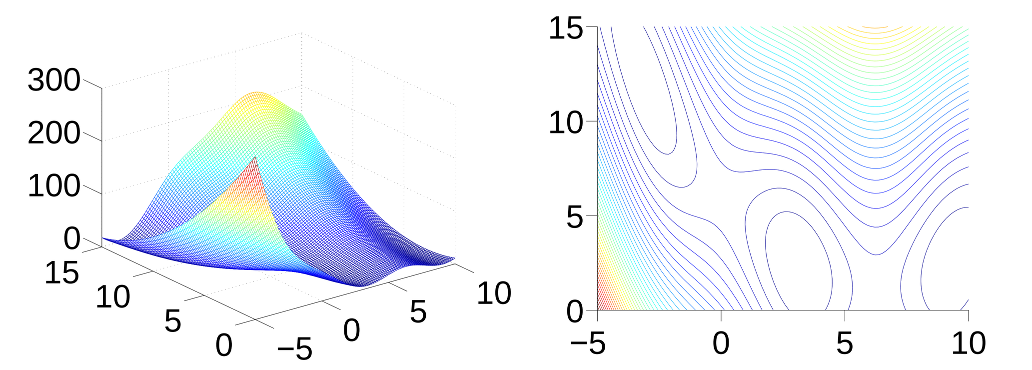

Figure 1: Original surface and contour plots

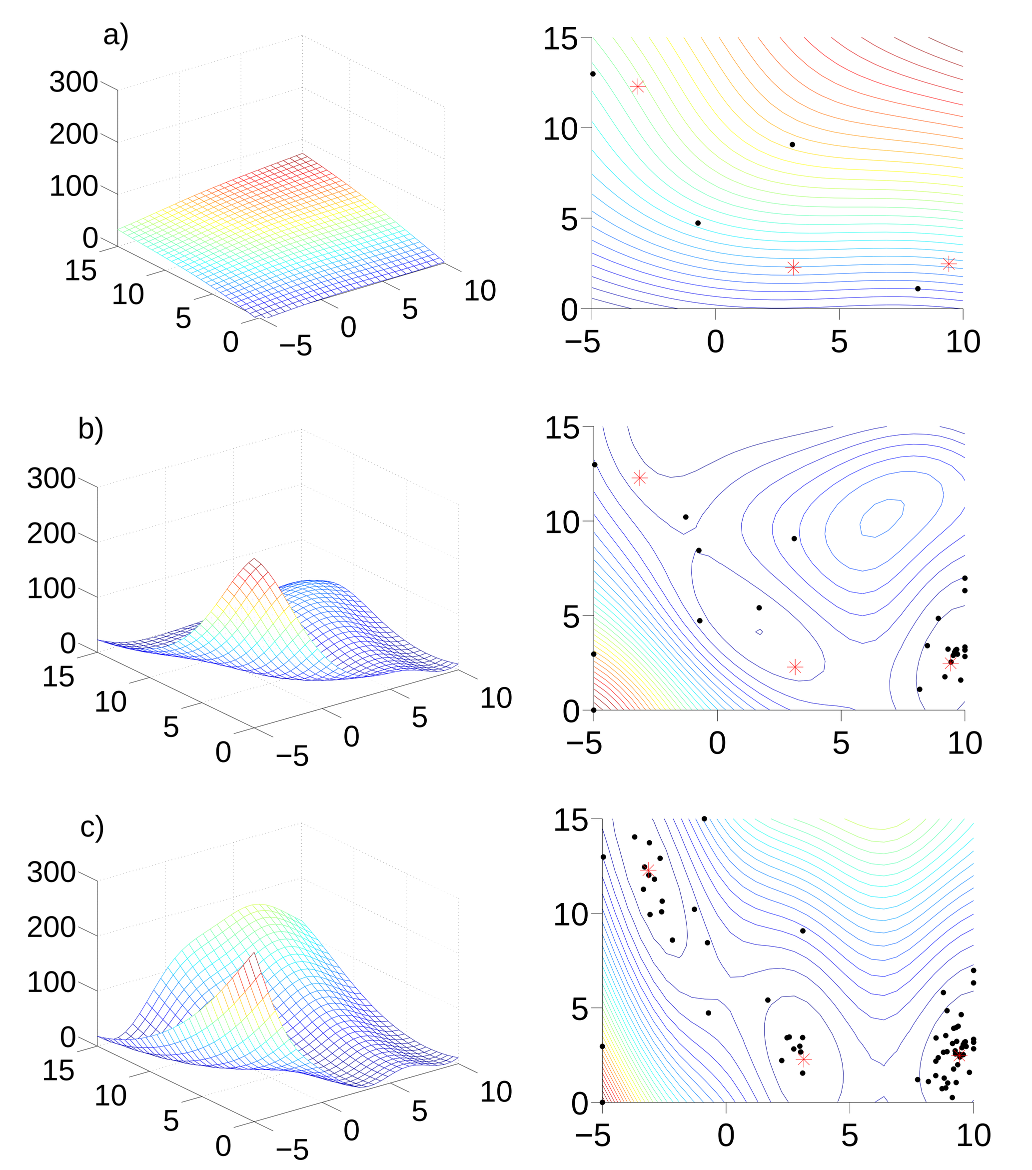

Figure 2: Approximation of surface and contours after a) 4, b) 29 and c) 73 function evaluations, respectively

This example shows that the vicinities of all three global minima could be detected and that the algorithm repeatedly samples close to the true optima.

Conclusions

The problem of finding the global optimum of a given optimization task has been tackled by using surrogate models. Appropriate (mixture) models have been selected using Dempster-Shafer theory. The results showed that in most cases the global optima could be approximated with high accuracy. Extensions of the algorithm to handle realistic engineering problems with more complicated constraints and the application of Bayesian model choice and model averaging will be considered in future research work.

References

[1] A.P. Dempster, 1968. A Generalization of Bayesian Framework *J. Roy. Statistical Society* B 30: 105-247

[2] G. Shafer, 1976. A Mathematical Theory of Evidence Princeton University Press

[3] D.R. Jones, 2001. A Taxonomy of Global Optimization Methods Based on Response Surfaces *J. Global Optim.* 21:345-383

TAMPERE UNIVERSITY OF TECHNOLOGY

Department of Mathematics

http://math.tut.fi/