Surrogate Model for Mixed-Variables Evolutionary Optimization based on GLM and RBF Networks

Lukáš Bajer[1,2] Martin Holeňa[2]

[1]Faculty of Mathematics and Physics, Charles University, and
[2]Institute of Computer Science, Czech Academy of Sciences

Prague, Czech Republic

SOFSEM, January 2013

Contents

Contents

Optimization of empirical functions

optimization:

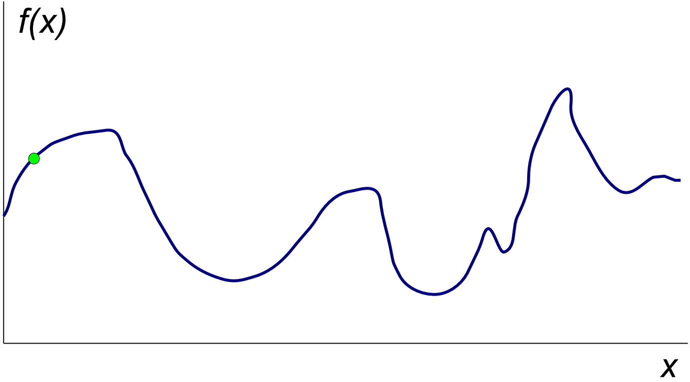
- S – set of feasible solutions, $f(x)$ – objective function
- **optimization** (minimization) is finding such $x^\star \in S$ that

$$f(x^\star) = \min_{\forall x \in S} f(x)$$

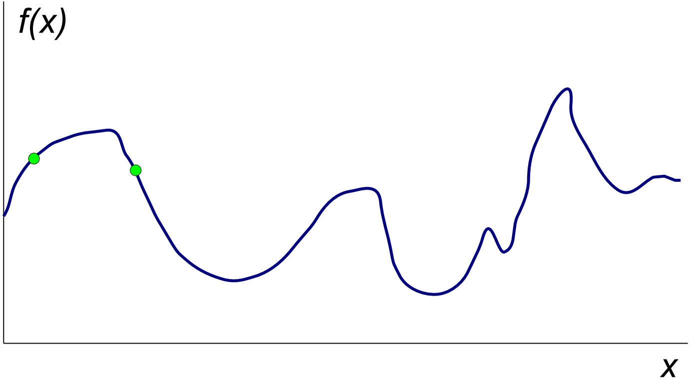
empirical objective functions:

- costly and/or time demanding evaluation of suggested solutions
- **cost** of the whole optimization measured by the resources spent on evaluations (time, money,...)
- discrete and continuous variables expected: $S = (\mathbf{S}^{(\mathrm{D})}, \mathbf{S}^{(\mathrm{C})})$

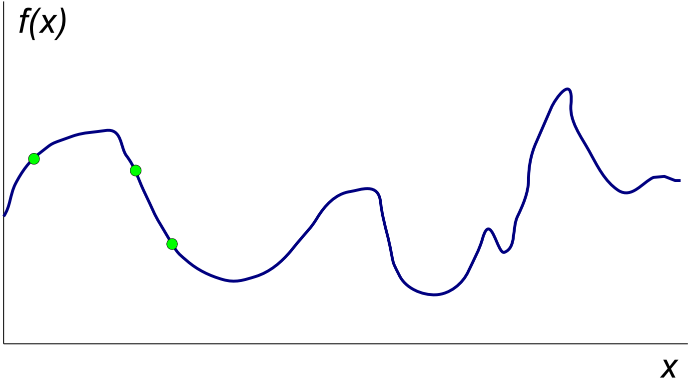
Optimization of empirical functions



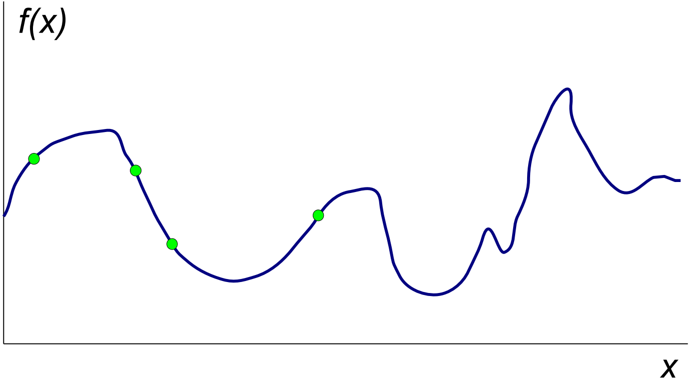
Optimization of empirical functions



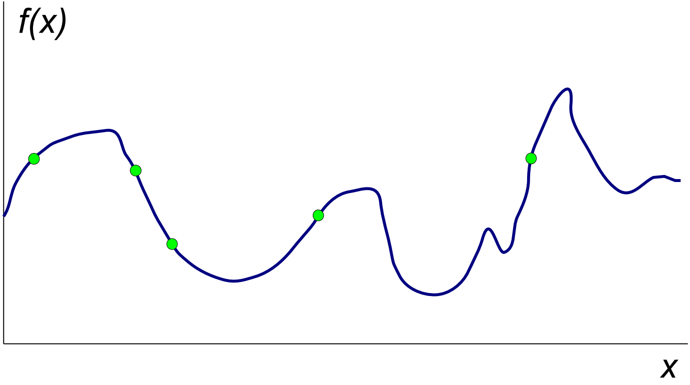
Optimization of empirical functions



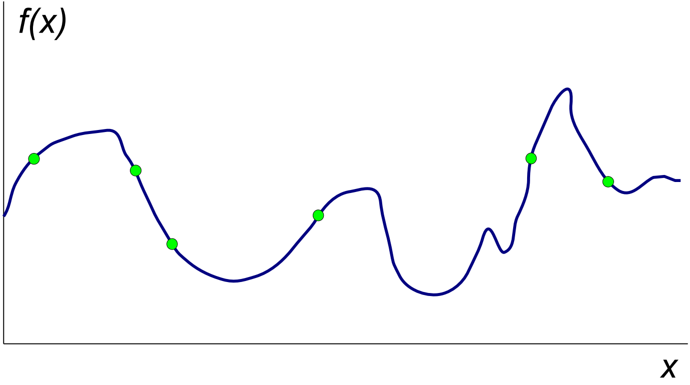
Optimization of empirical functions



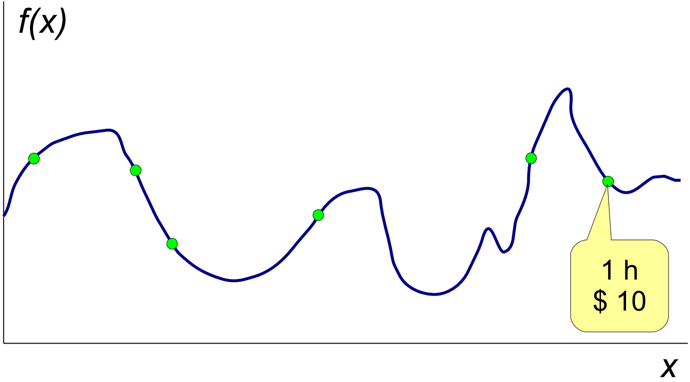
Optimization of empirical functions



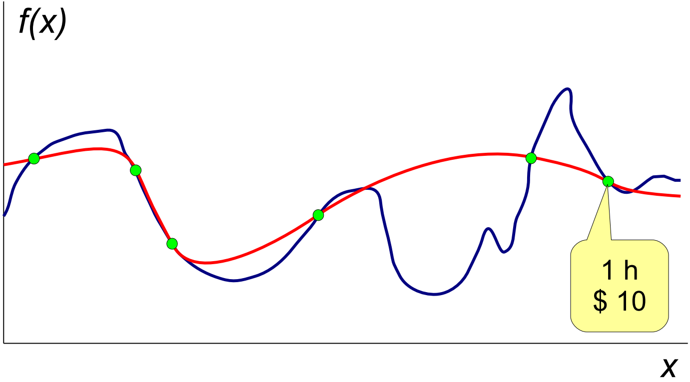
Optimization of empirical functions



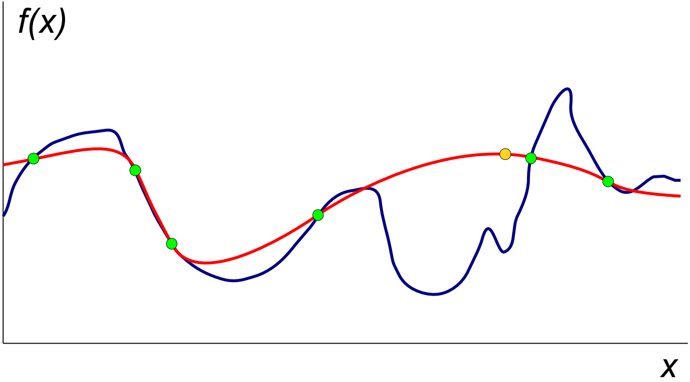
Optimization of empirical functions



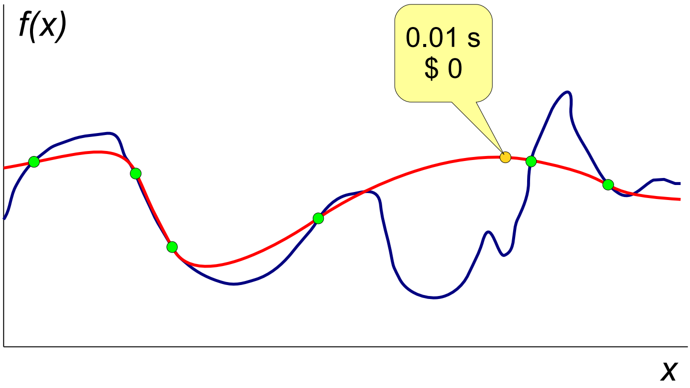
Optimization of empirical functions



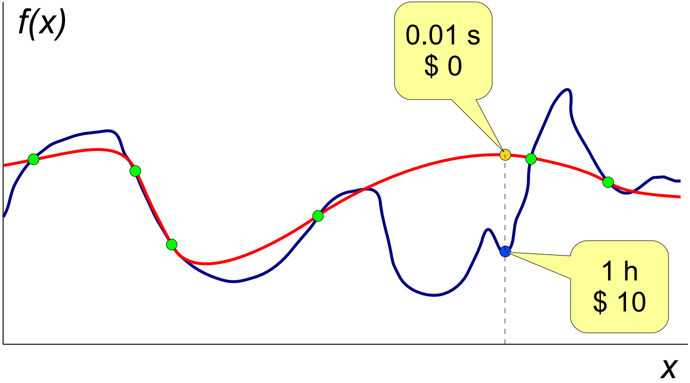
Optimization of empirical functions



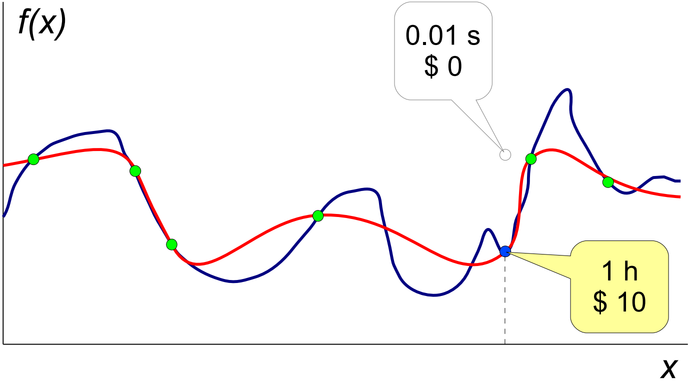
Optimization of empirical functions



Optimization of empirical functions



Optimization of empirical functions



Surrogate models of objective functions

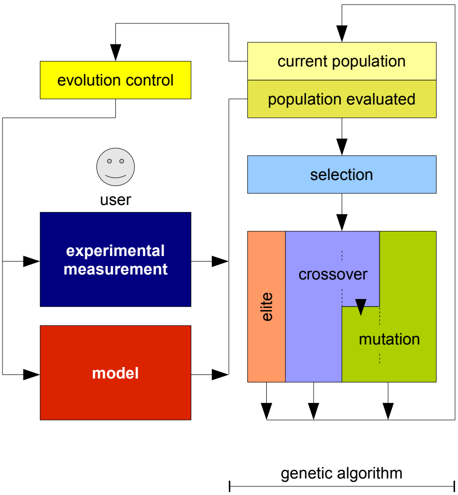
General idea

- usually evaluate suggested solutions with a cheap regression model
- use the original measurements to focus the search and ensure the true evaluation of the best found solution

- $(-)$ surrogate model is an approximation of the original fitness – almost always more or less **inaccurate**
- $(+)$ very fast and **cheap**
- **learning** from new measured results of experiments

Genetic algorithm and surrogate model

- genetic algorithm used as an optimization technique
- two kinds of fitness evaluation
  - experimental measurement (blue)
  - model fitness (red)



Contents

Regression model

requirement: both for continuous ($\mathbf{S}^{(\mathrm{D})}$) and discrete ($\mathbf{S}^{(\mathrm{D})}$) variables

- **continuous variables**: RBF network(s)
- **discrete variables**: GLM

RBF networks

RBF network

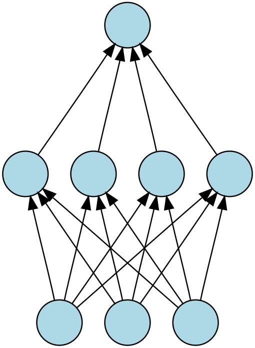
$$f(\mathbf{x}) = \sum_{i=1}^{g} \pi_i f_i(||\mathbf{x} - \mathbf{c}_i||)$$

composed of

- g Gaussian radial components $f_1, \ldots, f_g$ with variable widths w_i

$$f_i(||\mathbf{x} - \mathbf{c}_i||) = \mathrm{e}^{||w_i(\mathbf{x} - \mathbf{c}_i)||}$$

- variable centers $\mathbf{c}_i \in \mathbb{R}^n$
- components' weights $\pi_1, \ldots, \pi_g$



(Generalized) Linear Models

GLM

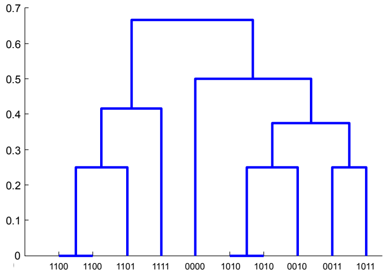
$$\hat{y} = g^{-1}\left(\sum_{j=0}^{d} x_j \beta_j\right)$$

- independent observed values Y;
  assumed from an exponential-family distribution
- explanatory variables $(x_1, x_2, \ldots, x_d)$ transformed through a linear model with parameters $\beta_1, \ldots, \beta_d$
- link function g connects the random and systematic component, depends on the distribution

Model construction – (1) clustering

clustering of the discrete data (in $\mathbf{S}^{(\mathrm{D})}$)

- similar discrete combinations are grouped into the same cluster
- hierarchical clustering with guaranteed minimal size
- m – the number of clusters



Model construction – (2) RBF networks

RBF networks

- trained on continuous parts of each such a cluster
- for $k = 1, \ldots, m$

$$\hat{f}^{(k)}_{\mathrm{DSCL}} : \mathbf{S}^{(\mathrm{C})} \to \mathbf{R}$$

- the best number of components $g^\star$ chosen by cross-validation
- for evaluation on a new $\mathbf{x}$
  - the nearest cluster $k^\star$ is identified (from discrete part)
  - its RBF network's value is computed (from continuous part):

$$\hat{f}^{(k^\star)}_{\mathrm{DSCL}}(\mathbf{x}^{(\mathrm{C})}) = \sum_{i=1}^{g^\star} \pi_{k^\star,i} f_{k^\star,i}(||\mathbf{x}^{(\mathrm{C})} - \mathbf{c}_{k^\star,i}||)$$

Model construction – (3) GLM model

GLM model

- model of the residuals

$$\hat{f}_{\mathrm{GLM}} : \mathbf{x}^{(\mathrm{D})} \to (y - \hat{f}_{\mathrm{DSCL}}(\mathbf{x}^{(\mathrm{C})}))$$

- trained on the discrete variables (explanatory variables)

$$\hat{f}_{\mathrm{GLM}} : \mathbf{S}^{(\mathrm{D})} \to \mathbf{R}$$

- dummy or integer coding of discrete variables
- residuals assumed from Gaussian, inverse Gaussian or gamma distribution – the best option is chosen via cross-validation

Model construction

final model

$$\hat{f} : (\mathbf{S}^{(\mathrm{D})}, \mathbf{S}^{(\mathrm{C})}) \to \mathbf{R}$$

- sum of the responses of the two previous models

$$\hat{f}(\mathbf{x}) = \hat{f}^{(k)}_{\mathrm{DSCL}}(\mathbf{x}^{(\mathrm{C})}) + \hat{f}_{\mathrm{GLM}}(\mathbf{x}^{(\mathrm{D})}).$$

evaluation of the final model:

1. identify the nearest cluster $k^\star$
2. compute response of the corresp. $k^\star$-th RBF network $\hat{y}_{\mathrm{DSCL}}$
3. get the value of the GLM model $\hat{y}_{\mathrm{GLM}}$
4. sum these two values $\hat{y} = \hat{y}_{\mathrm{DSCL}} + \hat{y}_{\mathrm{GLM}}$

Contents

Experimental results

Task 1: Data from HCN production

- results from a real application of genetic algorithms: finding the best catalyst for production of HCN
- 11 continuous, 1 discrete and 11 binary variables
- 555 different discrete combinations within 696 datapoints
- training of the model only

Experimental results

Task 2: Benchmark function

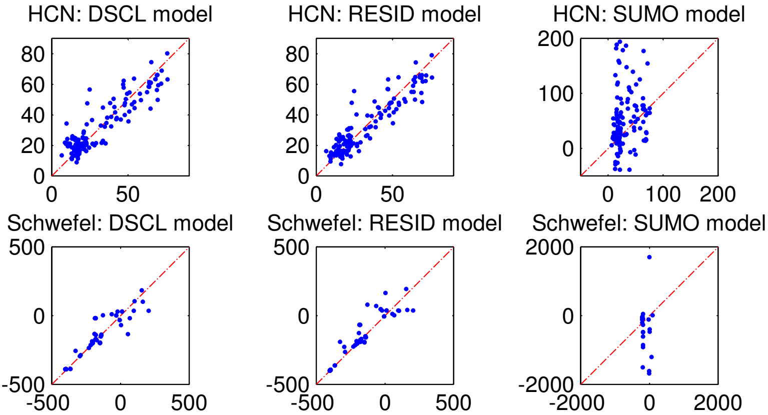
- individuals from the first 10 generations of a run of GA
- modified Schwefel's benchmark function
$$y = \frac{1}{p} \sum_{i=1}^{p} -x_i \sin(\sqrt{|x|})$$
- discrete variables added
$$(x_1^{(\mathrm{D})}, \dots, x_p^{(\mathrm{D})}) \in \{-10, -9, \dots, 10\}$$
- easy to compute for any arbitrary point – possible to test with genetic algorithm

Experimental results

Table: Surrogate-models' regression results on HCN and Schwefel dataset, average results from 50 trainings. RMSE on the testing and training set.

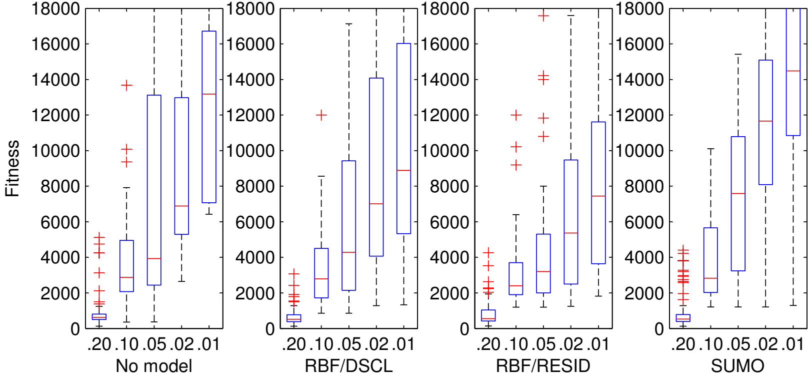
dataset	model	RMSE (test set)	RMSE (train set)
HCN	DSCL	8.2918 ± 0.4373	7.6337
	RESID	$\mathbf{7.6212} \pm \mathbf{0.3187}$	6.6131
	SUMO	115.196 ± 34.1387	**0.5711**
Schwefel	DSCL	62.712 ± 3.4248	83.553
	RESID	$\mathbf{55.605} \pm \mathbf{5.4867}$	73.468
	SUMO	$64576 \pm 1.9e{+}05$	**33.964**

Model fitting



Scatter plots of the DSCL, RESID and SUMO's RBF models on testing data

Benchmark fitness: genetic optimization



The numbers of original evaluations needed to reach 1.2-, 1.1-, 1.05-, 1.02- and 1.01-multiple of minimum, measured on 100 GA runs

Conclusion

Contributions of the work

- detailed description of a surrogate model for genetic optimization
- surrogate model with RBF networks and GLMs, for both continuous and discrete values

Experimental results

- model exhibits significantly lower testing error than competing RBF networks from SUMO toolbox
- utilizing the model on the benchmark fitness saved up to 35 per cent of empirical function evaluations

Conclusion

Thank you for your attention.

bajeluk@matfyz.cz
http://bajeluk.matfyz.cz