Parallel optimization based on FANOVA graph decomposition

Momchil Ivanov

July 1st, MASCOT-SAMO 2013 Meeting

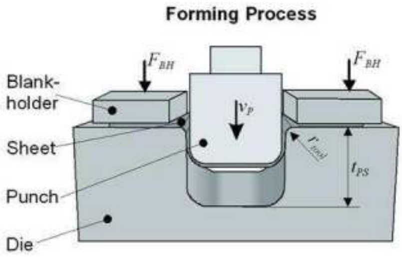
Institut für Mathematische Statistik und industrielle Anwendungen

technische universität dortmund

Overview

Motivation

Sheet metal forming process analysis



Motivation

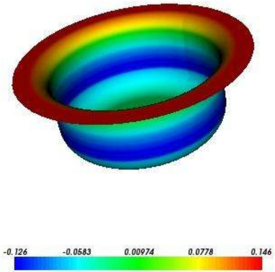
Forming press at TU Dortmund University

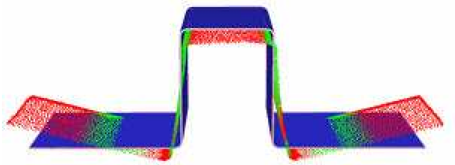
Problems: experiments expensive/not yet possible

→ Simulation by finite element methods (computer experiment)

Motivation

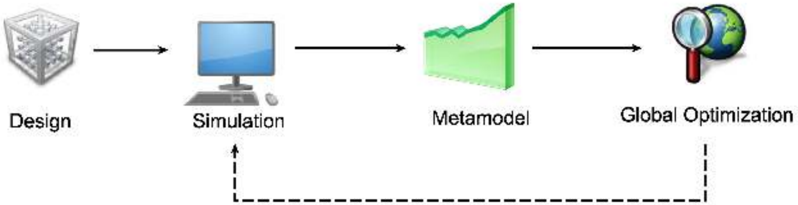
SFB 708 project C3:
Strategies for compensation of tearing, wrinkling or springback related shape deviation with the help of computer experiments





Overview

Sequential Black-Box Optimization



Standard Kriging Approach

Standard Kriging Model:

- Let $\mathcal{D} \subset \mathbb{R}^d$, and $D_0 = \{\mathbf{x_1}, \cdots, \mathbf{x_n}\} \subset \mathcal{D}$ initial design
- Output Y at $\mathbf{x} \in \mathcal{D}$:

$$\boxed{Y(\mathbf{x}) = \mu + Z(\mathbf{x})}$$

  μ- (constant) mean, $Z(\cdot)$- ***stationary*** Gauss process with mean 0 and variance σ^2, Covariance: $Cov(Z(\mathbf{x_i}), Z(\mathbf{x_j})) = \gamma(\mathbf{x_i} - \mathbf{x_j}) := R_{i,j}$

- BLUP of $Y(.)$: $\widehat{Y}(\mathbf{x}) = \widehat{\mu} + r^T(\mathbf{x})R^{-1}(\mathbf{y} - \mathbf{1}\widehat{\mu})$,
  $r(\mathbf{x}) = (\gamma(\mathbf{x} - \mathbf{x_1}), \cdots, \gamma(\mathbf{x} - \mathbf{x_n}))^T$
- **Uncertainty estimator** is a function in the estimated parameters, upper bound is σ^2

Introduction to EGO

Efficient Global Optimization (EGO):

- Optimization of black box functions
- Symbiosis with the prominent Kriging method
- Expected Improvement (EI) criterion to find potentially optimal points
- Critically relies on the ability of Kriging to assess its own uncertainty

EGO algorithm outline

1. Fit metamodel (Kriging) to the data
2. Calculate the point with the highest expected improvement
3. Evaluate the black box function at the calculated location
4. Update the model with the new information
5. Iterate until stopping criterion is reached

Jones et al. (1998)

Expected Improvement

- Let $y(\mathbf{x_i}) := y^{(i)}$ be the corresponding output to $\mathbf{x_i}$
- For $\mathbf{x} \in \mathcal{D}$ treat the unknown output $y(\mathbf{x})$ as a realization of a random variable $Y \sim \mathcal{N}(\widehat{y}(\mathbf{x}), s^2(\mathbf{x}))$, $\widehat{y}(\mathbf{x})$ and $s(\mathbf{x})$ - Kriging predictor and uncertainty measure in $\mathbf{x}$
- Denote with $f_{min} := \{y^{(1)}, \cdots, y^{(n)}\}$ the current best function value
- Define the improvement at point $\mathbf{x}$: $I = max(f_{min} - Y, 0)$

Expected Improvement

$$E\left[I(\mathbf{x})\right] = E\left[max(f_{min} - Y, 0)\right] = (f_{min} - \widehat{y})\Phi\left(\frac{f_{min} - \widehat{y}}{s}\right) + s\phi\left(\frac{f_{min} - \widehat{y}}{s}\right)$$

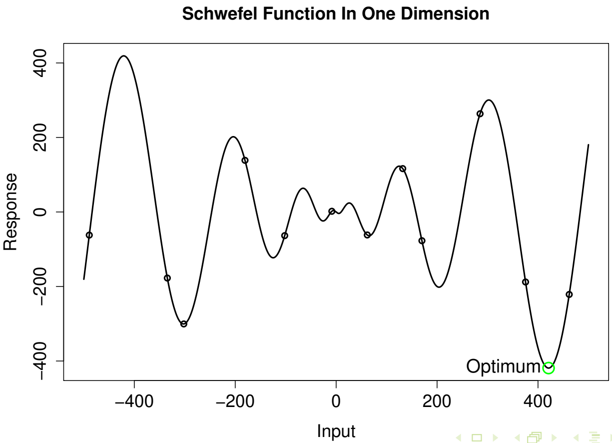
Where ϕ and Φ denote standard normal density and distribution function

Schwefel Function Example

$f_s(\mathbf{x}) = \sum_{i=1}^{n} -x_i \cdot \sin(\sqrt{abs(x_i)})$; $x_i \in [-500, 500]$; $n \in \mathbb{N}$, $i \in \{1, \cdots, n\}$

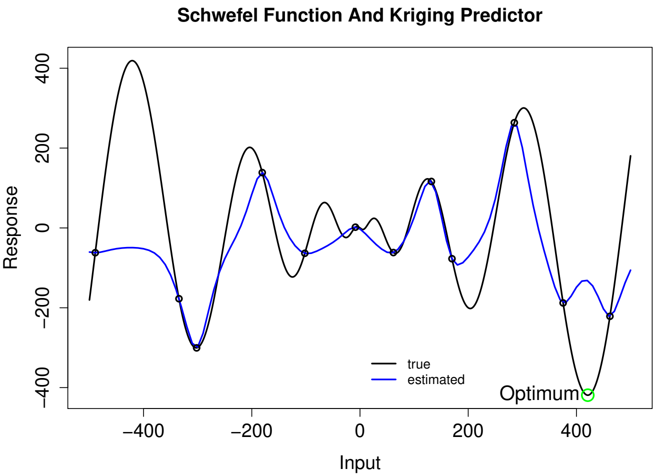
One global minimum: $f_s(\mathbf{x}^*) = -418.9829$, $\mathbf{x}^* = [420.9687, \cdots, 420.9687]^T$

Very complex function, but ***purely additive***

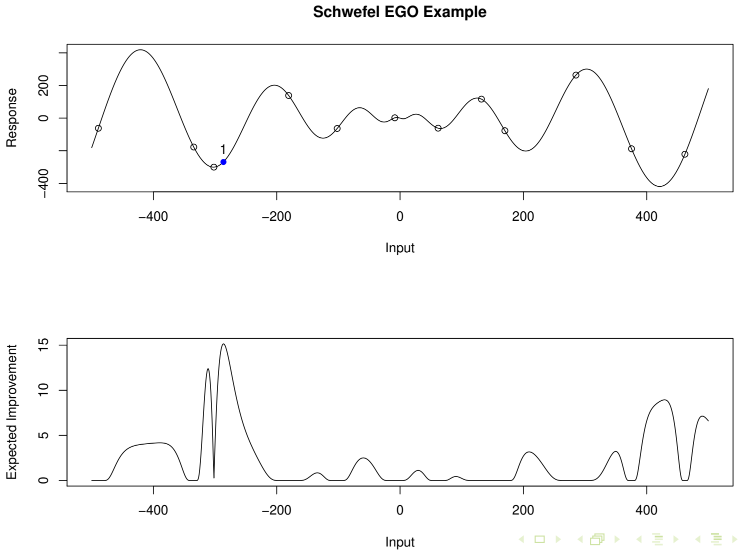


Schwefel Function in One Dimension

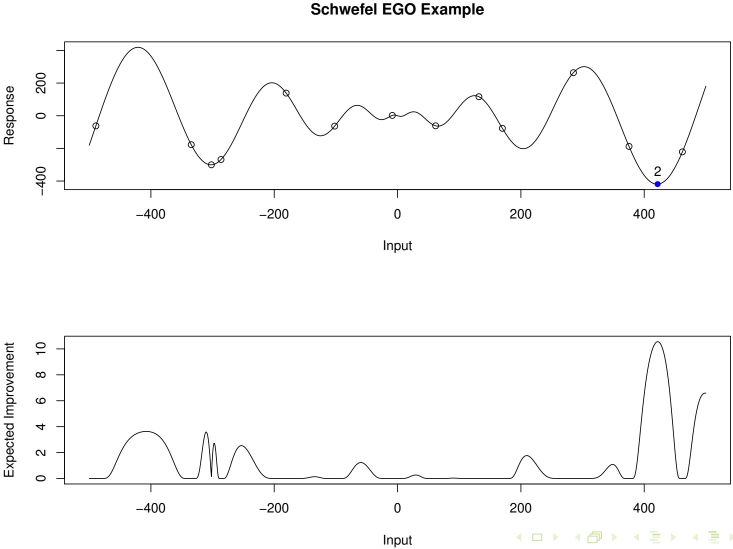
Initial design and Kriging prediction:



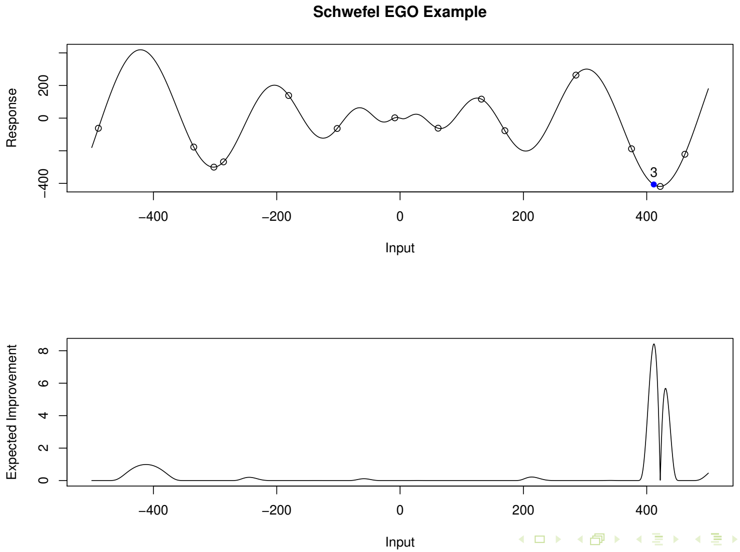
EGO Optimization Example



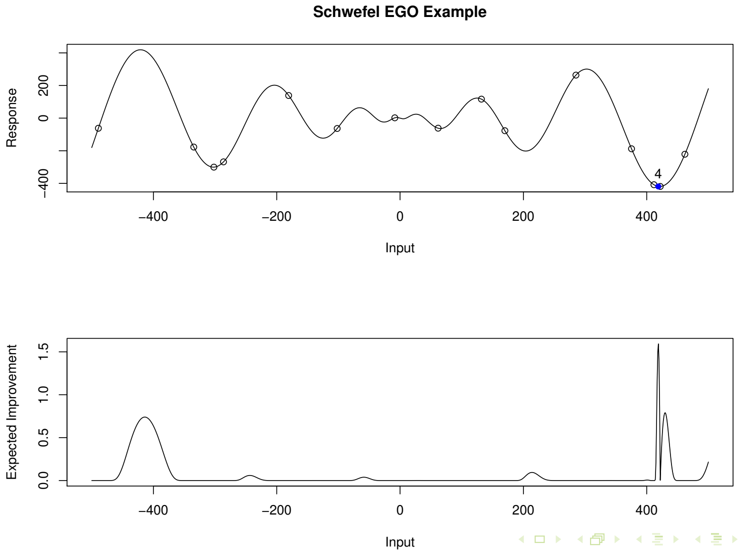
EGO Optimization Example



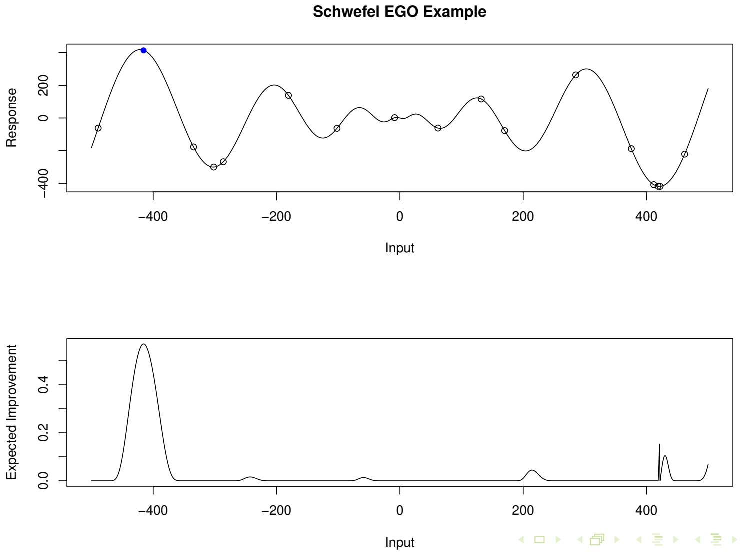
EGO Optimization Example



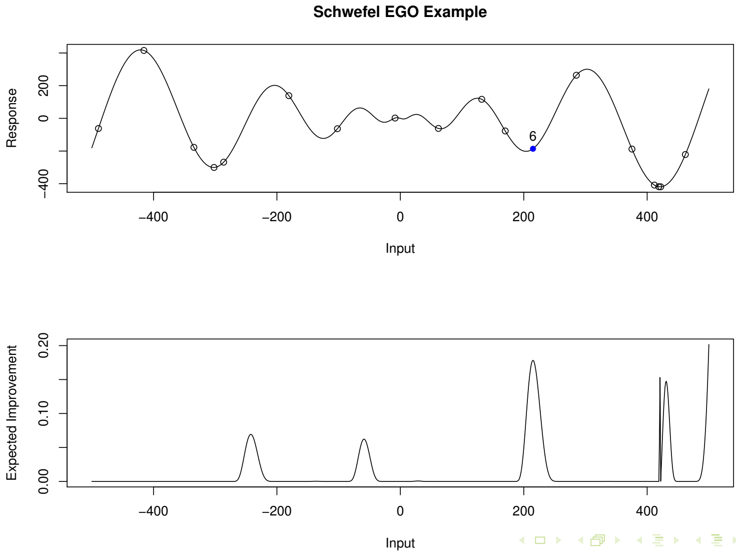
EGO Optimization Example



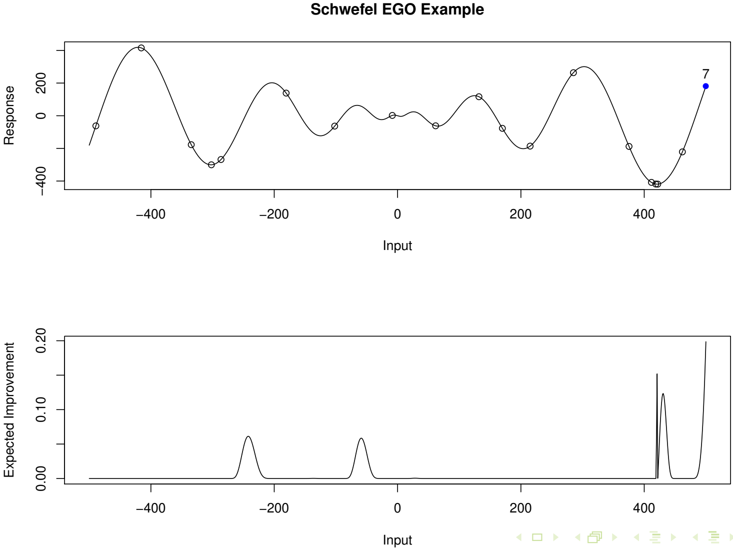
EGO Optimization Example



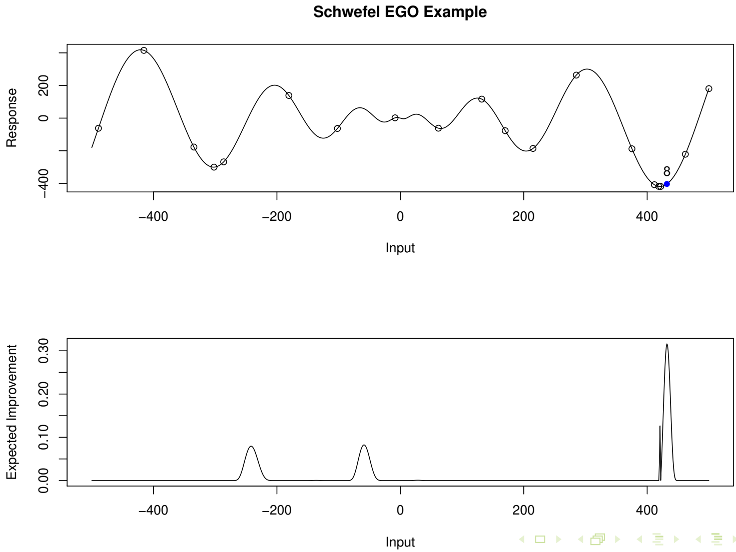
EGO Optimization Example



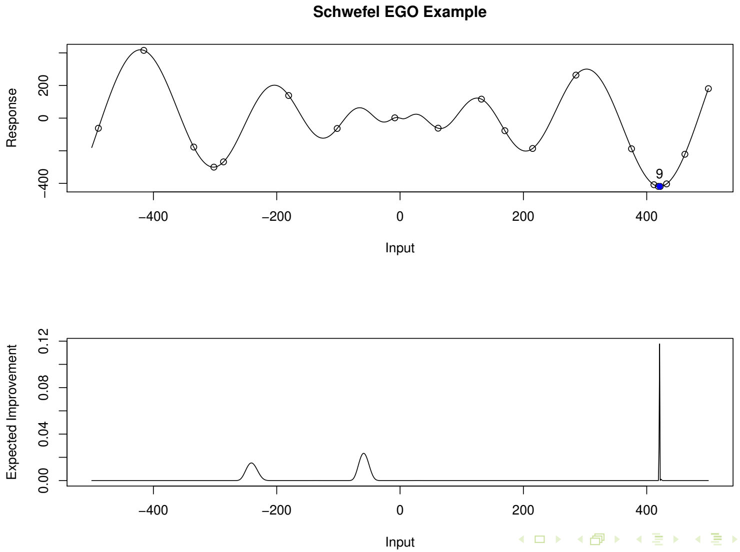
EGO Optimization Example



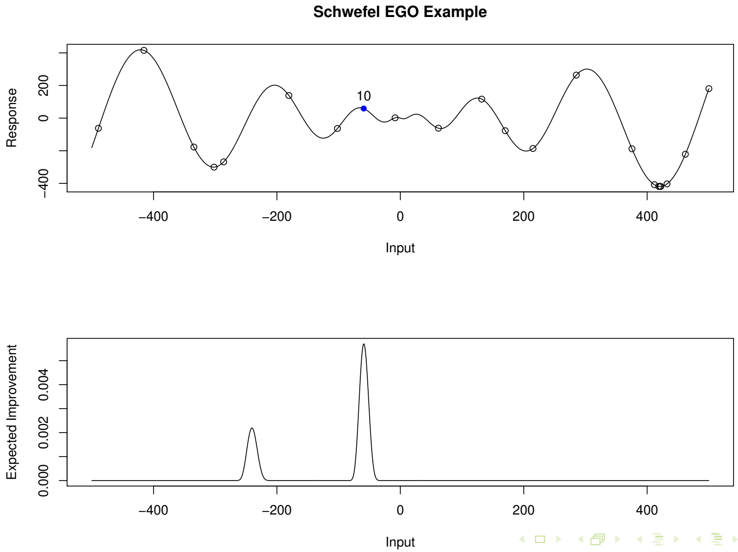
EGO Optimization Example



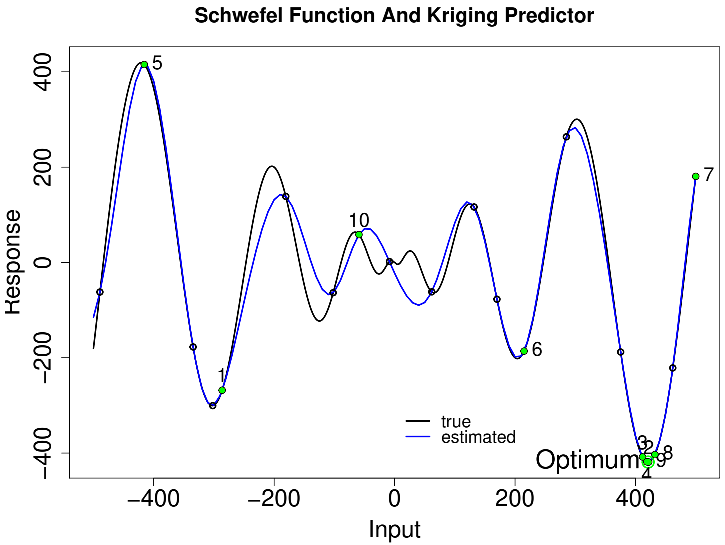
EGO Optimization Example



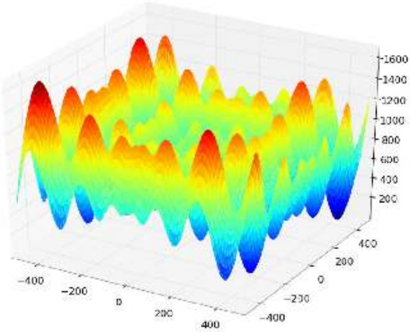
EGO Optimization Example



EGO Optimization Example



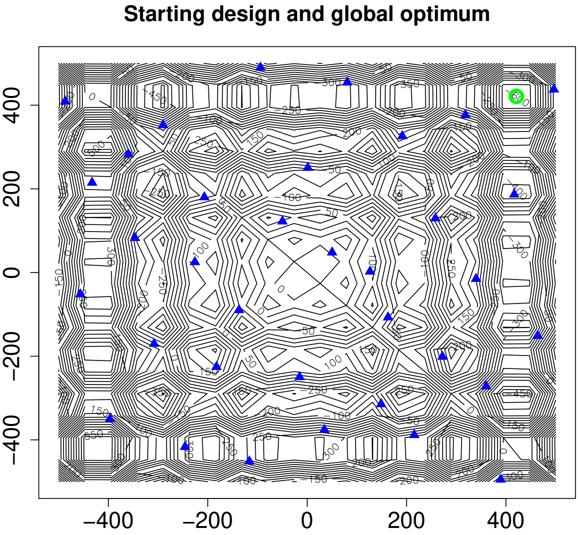
Schwefel Function in Two Dimensions: Plot



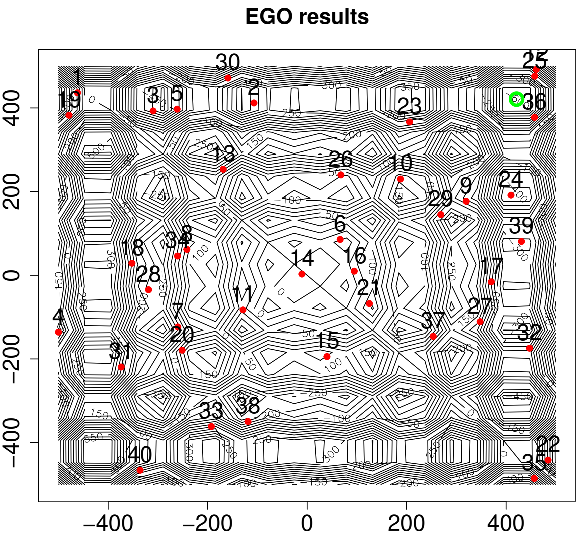
Schwefel Function

- Many local minima
- Only one global minimum

Schwefel Function in Two Dimensions: Contourplot



Schwefel Function in Two Dimensions: Optimization Results



Motivation: Optimization Complexity

"The curse of dimensionality":

- Modelling complexity grows exponentially
- Optimization complexity does not grow linearly
- Computing capabilites are limited
- ...

Possible solution: *reduce the dimensionality*

Dimensionality Reduction

Ideas:

- Use additive structure to decompose and optimize
- Optimize the independent parts parallel on several machines

  In the case of the Schwefel function:

$$\begin{aligned}\mathbb{R}^n \ni \mathbf{x}^* &= \arg\min_{\mathbf{x}\in\mathbb{R}^n} \sum_{i=1}^{n} -x_i \cdot \sin(\sqrt{abs(x_i)}) \\ &\equiv \left(\min_{x_1\in\mathbb{R}} -x_1 \cdot \sin(\sqrt{abs(x_1)}), \ldots, \min_{x_n\in\mathbb{R}} -x_n \cdot \sin(\sqrt{abs(x_n)}) \right)\end{aligned}$$

Problem to solve:

How to detect additive structure in black-box functions?

Overview

Total Interaction Index

Definition

For any pair of variables (X_i, X_j) the sum of sobol sensitivity indices D_J

$$\mathfrak{D}_{ij} := \sum_{J \supset \{i,j\}} D_J, \quad i,j \in \{1, \ldots, d\}$$

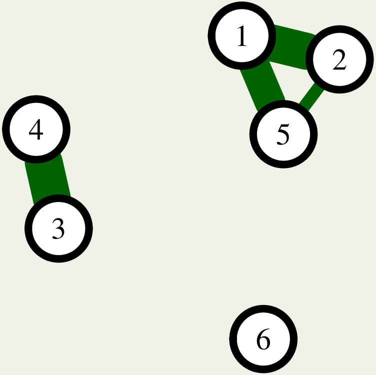
is called **total interaction index** of X_i and X_j.

Fruth et al. (2012)

Applications

Visualization of interaction structure

Example function in $\mathbb{R}^6$



Applications

Detection of a **block-additive structure** $\equiv$ *disjoint subgraphs*

$$f(x_1, \ldots, x_d) = \sum_{I \in \mathcal{B}} f_I(x_I), \ \mathcal{B} \subseteq \mathcal{P}(\{x_1, \ldots, x_d\})$$

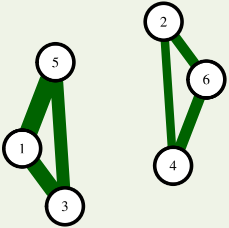
$$\text{where} \bigcap_{I \in \mathcal{B}} I = \emptyset$$

Applications

Detection of a **block-additive structure** $\equiv$ *disjoint subgraphs*

$$f(x_1, \dots, x_d) = \sum_{I \in \mathcal{B}} f_I(x_I),\ \mathcal{B} \subseteq \mathcal{P}(\{x_1, \dots, x_d\})$$

$$\text{where } \bigcap_{I \in \mathcal{B}} I = \emptyset$$



Applications

Detection of a **block-additive structure**

$$f(x_1, \ldots, x_d) = \sum_{I \in \mathcal{B}} f_I(x_I), \ \mathcal{B} \subseteq \mathcal{P}\left(\{x_1, \ldots, x_d\}\right), \ \bigcap_{I \in \mathcal{B}} I = \emptyset$$

Block-additive structure exaple for a $\mathbb{R}^6$ function:

$$f(x_1, \ldots, x_6) = f_1(x_1, x_3, x_5) + f_2(x_2, x_4, x_6)$$

Applications

Simplify and parallelize **optimization** problems by separation

$$f(x_1, \ldots, x_d) = \sum_{I \in \mathcal{B}} f_I(x_I), \; \mathcal{B} \subseteq \mathcal{P}\left(\{x_1, \ldots, x_d\}\right), \; \bigcap_{I \in \mathcal{B}} I = \emptyset$$

Applications

Simplify and parallelize **optimization** problems by separation

$$f(x_1, \dots, x_d) = \sum_{I \in \mathcal{B}} f_I(x_I), \; \mathcal{B} \subseteq \mathcal{P}(\{x_1, \dots, x_d\}), \; \bigcap_{I \in \mathcal{B}} I = \emptyset$$

$$\min_{x \in \mathbb{R}^n} f(x) = \sum_{I \in \mathcal{B}} \min_{x_I} f_I(x_I) = \mathbf{f}\left(\pi_{\mathbf{id}}\left(\bigcup_{\mathbf{I} \in \mathcal{B}} \arg\min_{\mathbf{x_I}} \mathbf{f}(\mathbf{x_I}, \mathbf{c_I})\right)\right), \; c_I \in \mathbb{R}^{n-dim(I)}$$

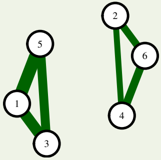
c_I - a fix constant, not being optimized

Applications

Simplify and parallelize optimization problems by separation

$$f(x_1,\dots,x_d)=\sum_{I\in\mathcal{B}} f_I(x_I),\ \mathcal{B}\subseteq\mathcal{P}\left(\{x_1,\dots,x_d\}\right),\ \bigcap_{I\in\mathcal{B}} I=\emptyset$$

Example function in $\mathbb{R}^6$



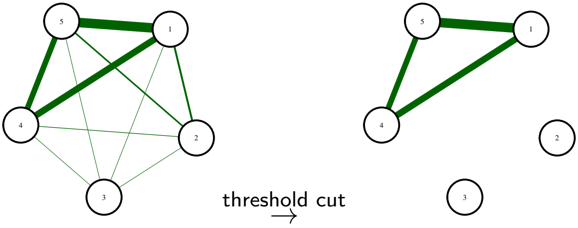
$$\min_{x_1,\dots,x_6} f(x_1,\dots x_6)=\min_{x_1,x_3,x_5} f_1(x_1,x_3,x_5)+\min_{x_2,x_4,x_6} f_2(x_2,x_4,x_6)=$$

$$f\left(\pi_{id}\left(\arg\min_{x_1,x_2,x_5} f(x_1,c^1,x_3,c^1,x_5,c^1),\min_{x_2,x_4,x_6} f(c^2,x_2,c^2,x_4,c^2,x_6)\right)\right)$$

Notes on Application

- The total interaction indices $\mathfrak{D}_{ij}$ have to be estimated with the help of a metamodel (from the data)
- Inactive (in the true model) interactions could be marked as significant

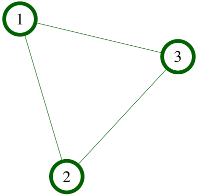
$\Rightarrow$ Filter "phantom" interactions by *thresholding*



Reproducing all active interactions is crucial for optimization

Schwefel Function in Three Dimensions

Schwefel function is completely additive (no interactions)

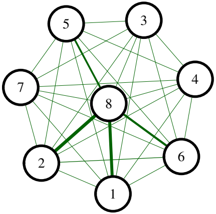


fanovaGraph(R-package):

```
Scaled total interaction indices
          totalInt
X1*X2 1.476550e-07
X1*X3 1.457002e-07
X2*X3 1.645334e-07
```

Overview

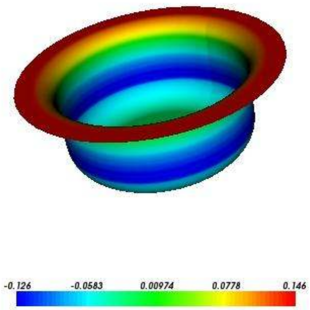
Application to Sheet Metal Forming



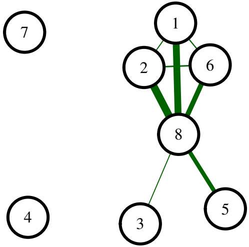
1	yield stress
2	thickness
3	blankholder force
4	friction1
5	friction2
6	friction3
7	strain hardening
8	cutting
Objective	thickness reduction

Optimization with EGO

- Standard approach ⇒ No additive decomposition
- Simulations budget:
  - Starting design: standard latin hypercube, 50 training runs
  - Additional 50 EGO steps made
  - Overall total: 100 Simulations
- Best solution reached, after $50 + 43 = 93$ Simulations: 8.78% ⇒ good result



Additive Decomposition



1	yield stress
2	thickness
3	blankholder force
4	friction1
5	friction2
6	friction3
7	strain hardening
8	cutting

Note: A "conservative" threshold cut was chosen

Parallel Optimization

- Parallel optimization of the additive decomposition
- Simulation budget:
  - Estimate the FANOVA Graph 50 Simulations (not used for optimization)
  - Training the separate models: 60 Simulations
  - Total number of additional runs: 50
  - Overall total: 160 Simulations
- Best solution reached, after $50 + 60 + 37 = 147$ Simulations: 7.6%

Notes:

- More simulations needed as for the standard approach
- Apart from the initial 50 runs for estimation, all other simulation runs were done in parallel
- A better solution was found

Overview

Discussion and Outlook

- The shown procedure is flexible:
  independent of the metamodel and the optimization method

Discussion and Outlook

- The shown procedure is flexible:
  independent of the metamodel and the optimization method
  - ⇒ Application of other metamodels and optimization schemes

Discussion and Outlook

- The shown procedure is flexible:
  independent of the metamodel and the optimization method
  - ⇒ Application of other metamodels and optimization schemes
- It allows to search for the "real" interactions structure:
  finding and optimal cut

Discussion and Outlook

- The shown procedure is flexible:
  independent of the metamodel and the optimization method
  - $\Rightarrow$ Application of other metamodels and optimization schemes
- It allows to search for the “real” interactions structure:
  finding and optimal cut
  - $\Rightarrow$ Optimal thresholding

References

- Jones, D. R.; Schonlau, M.; Welch, W. J. (1998): Efficient global optimization of expensive black-box functions; Journal of Global optimization 13 (4), 455-492.
- Fang, K. T.; Li, R.; Sudjianto, A. (2006): Design and modeling for computer experiments. London: Chapman & Hall.
- Ginsbourger, D.; Riche, R.; Carraro, L. (2010): Kriging is well-suited to parallelize optimization; Y. Tenne und C.-K. Goh (Hg.): Computational intelligence in expensive optimization problems. 1. Aufl. Berlin, New York: Springer, 131-162.
- Fruth, J.; Roustant, O.; Kuhnt, S. (2012): Total interaction index: A variance-based sensitivity index for interaction screening. HAL version available.
- Mühlenstädt, T.; Roustant, O.; Carraro, L.; Kuhnt, S. (2011): Data-driven kriging models based on FANOVA-decomposition. *Statistics and Computing* **22**, 723-738.

References

- Liu, R.; Owen, A. B. (2006): Estimating mean dimensionality of analysis of variance decompositions. *Journal of the American Statistical Association* **101** (474), 712-721.
- Saltelli, A.; Chan, K.; Scott, E. M. (2000): *Sensitivity analysis*, Wiley, Chichester.
- Sobol', I. M. (1993): Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment* **1**, 407-414.
- Efron, B.; Stein, C. (1981): The jackknife estimate of variance. *The Annals of Statistics* **9** (3), 586-596.
- Mühlenstädt, T.; Kuhnt, S. (2011): Kernel interpolation; Computational Statistics & Data Analysis 55, S. 2962-2974.
- Jones, D. R. (2001): A taxonomy of global optimization methods based on response surfaces; Journal of Global optimization 21 (4), 345-383.