Parallel optimization based on FANOVA graph decomposition

Momchil Ivanov

July 1st, MASCOT-SAMO 2013 Meeting
Overview

1. Introduction
2. Black-Box Optimization
3. FANOVA Graph Decomposition
4. Application and Results
5. Discussion
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

1. Introduction

2. Black-Box Optimization

3. FANOVA Graph Decomposition

4. Application and Results

5. Discussion
Sequential Black-Box Optimization
Standard Kriging Approach

**Standard Kriging Model:**

- Let $\mathcal{D} \subset \mathbb{R}^d$, and $D_0 = \{x_1, \cdots, x_n\} \subset \mathcal{D}$ initial design
- Output $Y$ at $x \in \mathcal{D}$:
  \[ Y(x) = \mu + Z(x) \]

  - $\mu$- (constant) mean, $Z(\cdot)$- **stationary** Gauss process with mean 0 and variance $\sigma^2$, Covariance: $\text{Cov}(Z(x_i), Z(x_j)) = \gamma(x_i - x_j) := R_{i,j}$
  - BLUP of $Y(\cdot)$: $\hat{Y}(x) = \hat{\mu} + r^T(x)R^{-1}(y - 1\hat{\mu})$,
    \[ r(x) = (\gamma(x - x_1), \cdots, \gamma(x - x_n))^T \]

  - **Uncertainty estimator** is a function in the estimated parameters, upper bound is $\sigma^2$
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)
Let $y(x_i) := y^{(i)}$ be the corresponding output to $x_i$

For $x \in D$ treat the unknown output $y(x)$ as a realization of a random variable $Y \sim \mathcal{N}(\hat{y}(x), s^2(x))$, $\hat{y}(x)$ and $s(x)$ - Kriging predictor and uncertainty measure in $x$

Denote with $f_{min} := \{y^{(1)}, \ldots, y^{(n)}\}$ the current best function value

Define the improvement at point $x$: $I = max(f_{min} - Y, 0)$

**Expected Improvement**

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

Where $\phi$ and $\Phi$ denote standard normal density and distribution function
Schwefel Function Example

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

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

Very complex function, but purely additive

Schwefel Function In One Dimension

![Schwefel Function Graph](image-url)
Schwefel Function in One Dimension

Initial design and Kriging prediction:

![Graph showing the Schwefel Function and Kriging Predictor](image-url)
EGO Optimization Example

Schwefel EGO Example

![Schwefel EGO Example](image)

Expected Improvement

![Expected Improvement](image)

TU Dortmund  Momchil Ivanov
EGO Optimization Example

Schwefel EGO Example

Response vs. Input

Expected Improvement vs. Input
EGO Optimization Example

Schwefel EGO Example

Response

Input

Expected Improvement

Input
EGO Optimization Example

Schwefel EGO Example

Response

Input

Expected Improvement

Input
EGO Optimization Example

Schwefel EGO Example

Response

Input

Expected Improvement

Input
EGO Optimization Example

Schwefel EGO Example

Response

Input

Expected Improvement

Input
EGO Optimization Example

Schwefel EGO Example

Response vs Input

Expected Improvement vs Input
EGO Optimization Example

Schwefel EGO Example

![Schwefel EGO Example Graph]

Expected Improvement

![Expected Improvement Graph]
EGO Optimization Example

Schwefel EGO Example

Input

Response

Expected Improvement

Input
EGO Optimization Example

Schwefel EGO Example

Response vs. Input

Expected Improvement vs. Input
EGO Optimization Example

Schwefel Function And Kriging Predictor

Input

Response

true

estimated

Optimum

TU Dortmund

Momchil Ivanov
Schwefel Function in Two Dimensions: Plot

Many local minima
Only one global minimum
Schwefel Function in Two Dimensions: Contourplot

Starting design and global optimum
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 capabilities 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:

\[
\mathbb{R}^n \ni \mathbf{x}^* = \arg\min_{\mathbf{x} \in \mathbb{R}^n} \sum_{i=1}^{n} -x_i \cdot \sin(\sqrt{\text{abs}(x_i)})
\]

\[
\equiv \left( \min_{x_1 \in \mathbb{R}} -x_1 \cdot \sin(\sqrt{\text{abs}(x_1)}), \ldots, \min_{x_n \in \mathbb{R}} -x_n \cdot \sin(\sqrt{\text{abs}(x_n)}) \right)
\]

Problem to solve:
*How to detect additive structure in black-box functions?*
Overview

1. Introduction

2. Black-Box Optimization

3. FANOVA Graph Decomposition

4. Application and Results

5. Discussion
Total Interaction Index

Definition

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

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

is called total interaction index of \(X_i\) and \(X_j\).

Fruth et al. (2012)
Visualization of interaction structure

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

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

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

where \(\bigcap_{I \in B} I = \emptyset\)
Applications

Detection of a block-additive structure \( \equiv \) disjoint subgraphs

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

where \( \bigcap_{I \in \mathcal{B}} I = \emptyset \)
Detection of a block-additive structure

\[ f(x_1, \ldots, x_d) = \sum_{l \in \mathcal{B}} f_l(x_l), \quad \mathcal{B} \subseteq \mathcal{P} \left( \{x_1, \ldots, x_d\} \right), \quad \bigcap_{l \in \mathcal{B}} l = \emptyset \]

Block-additive structure example 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

\[ f(x_1, \ldots, x_d) = \sum_{I \in \mathcal{B}} f_I(x_I), \quad \mathcal{B} \subseteq \mathcal{P}(\{x_1, \ldots, x_d\}), \quad \bigcap_{I \in \mathcal{B}} I = \emptyset \]
Applications

Simplify and parallelize optimization problems by separation

\[ f(x_1, \ldots, x_d) = \sum_{I \in B} f_I(x_I), \quad B \subseteq \mathcal{P}(\{x_1, \ldots, x_d\}), \quad \bigcap_{I \in B} I = \emptyset \]

\[
\min_{x \in \mathbb{R}^n} f(x) = \sum_{I \in B} \min_{x_I} f_I(x_I) = f\left( \pi_{id}\left( \bigcup_{I \in B} \arg\min_{x_I} f(x_I, c_I) \right) \right), \quad c_I \in \mathbb{R}^{n-\text{dim}(I)}
\]

c_I - a fix constant, not being optimized
Applications

Simplify and parallelize **optimization** problems by separation

\[
f(x_1, \ldots, x_d) = \sum_{I \in \mathcal{B}} f_I(x_I), \quad \mathcal{B} \subseteq \mathcal{P}(\{x_1, \ldots, x_d\}), \quad \bigcap_{I \in \mathcal{B}} I = \emptyset
\]

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

\[
\min_{x_1, \ldots, x_6} f(x_1, \ldots, x_6) = \min_{x_1, x_3, x_5} f_1(x_1, x_3, x_5) + \min_{x_2, x_4, x_6} f_2(x_2, x_4, x_6) =
\]

\[
f\left(\pi_{id}\left(\arg \min_{x_1, x_2, x_5} f(x_1, c^1, x_3, c^1, x_5, c^1), \min_{x_2, x_4, x_6} f(c^2, x_2, c^2, x_4, c^2, x_6)\right)\right)
\]
Notes on Application

- The total interaction indices $\mathcal{D}_{ij}$ 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

<table>
<thead>
<tr>
<th>Interaction</th>
<th>totalInt</th>
</tr>
</thead>
<tbody>
<tr>
<td>X1*X2</td>
<td>1.476550e-07</td>
</tr>
<tr>
<td>X1*X3</td>
<td>1.457002e-07</td>
</tr>
<tr>
<td>X2*X3</td>
<td>1.645334e-07</td>
</tr>
</tbody>
</table>
Overview

1. Introduction
2. Black-Box Optimization
3. FANOVA Graph Decomposition
4. Application and Results
5. Discussion
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: $8.78\% \Rightarrow$ 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

1. Introduction

2. Black-Box Optimization

3. FANOVA Graph Decomposition

4. Application and Results

5. Discussion
Discussion and Outlook

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


