Learning Overlap Optimization for Domain Decomposition Methods

17th Pacific-Asia Conference on Knowledge Discovery and Data Mining

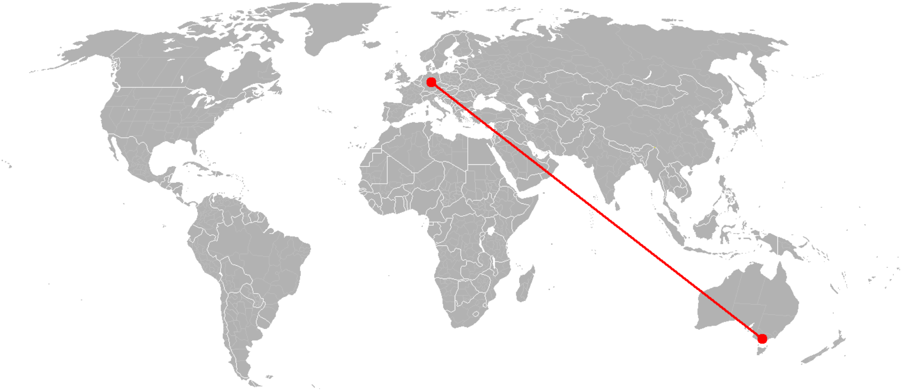
Steven Burrows† Jörg Frochte‡ Michael Völske†
Ana Belén Martina Tores‡ Benno Stein†

† Bauhaus-Universität Weimar
‡ Bochum University of Applied Science

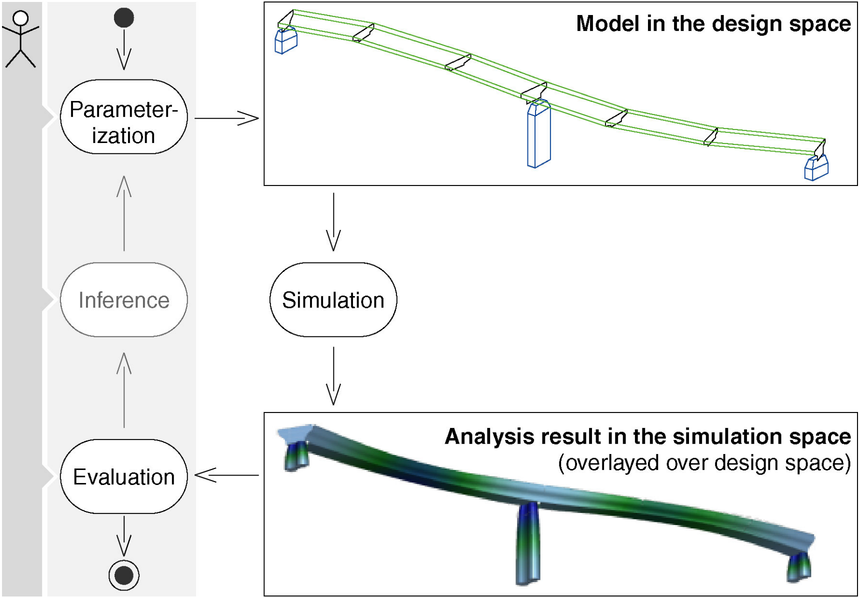
14–17 April 2013

About Me

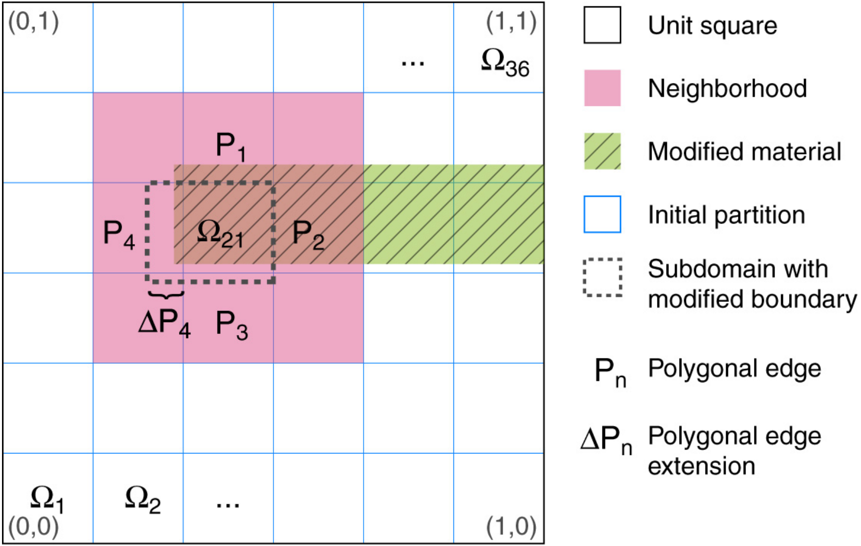
- RMIT University: Undergrad, Honours, and PhD up to 2010.
- Bauhaus-Universität Weimar: PostDoc from 2011–2012.
  - Research in Digital Engineering and Simulation Data Mining.
- German Institute for International Educational Research: Research Scientist from 2013 to current.



Interactive Bridge Design in Civil Engineering



Parallel Simulation with Domain Decomposition



Problem Definition

Problem: Poisson's Equation

- A second-order elliptic partial-differential equation (PDE).
- Has application in modeling stationary heat.
- Additional applications in Newtonian gravity and electrostatics.
- Transferable results. E.g: Stress modeling in engineering science.

The Maths

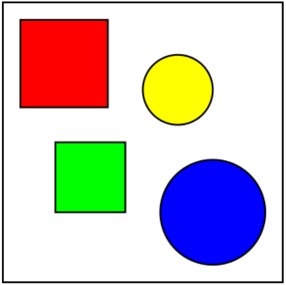
- $-\varepsilon(x)\nabla^2 u = f(x)$ on Ω ; $u = g(x)$ on $\partial\Omega$
- Ω: geometry (i.e. a bar). $f(x) \geq 0$: heat sources. $\varepsilon(x)$: material property. $g(x)$: temperatures on the boundary $\partial\Omega$ of the domain Ω.

Numerical Method: Finite Element Method

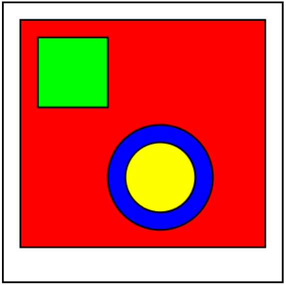
- A standard method in most engineering software solutions.
- Applied to the unit square. $\Omega = [0,1] \times [0,1]$.
- Checkerboard partitioning for a restricted problem space.

Generating Diffusion Specifications

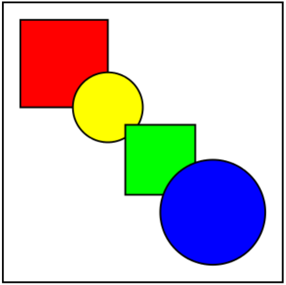
Diffusion specification: A unique set of material values within the unit square to solve Poisson's equation.



Isolated



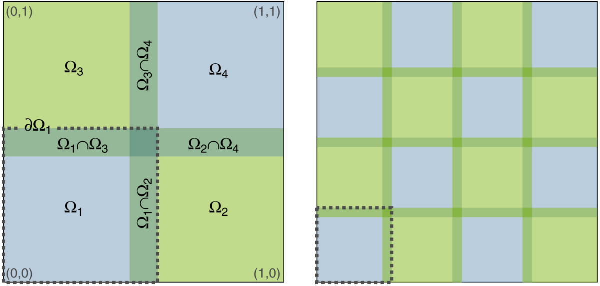
Nested



Sequence

Shapes and sizes are based on a deterministic pseudo random number generator.

Generating Domain Specifications

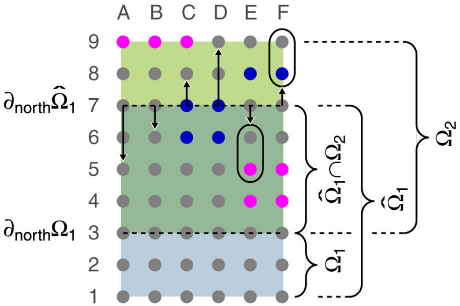


- Assuming 0.4% global overlap on a 4 $\times$ 4 checkerboard, and
- Three adjustments per sub-domain (-0.2%, +0.0%, +0.2%), gives us
- 48 domain specifications per diffusion specification.

Extracting Features from Neighborhoods

Example. Material settings are:

- Pink: $\epsilon = 10\,000$. Gray: $\epsilon = 1\,000$. Blue: $\epsilon = 100$.



Region	Max Value in Region	Min Value in Region	Max Diff in Region	Max Diff to Boundary	Min Diff to Boundary
A	10 000	1 000	9 000	9 000	0
B	1 000	100	900	0	0
C	1 000	100	900	900	0
D	10 000	1 000	9 000	9 900	0
E	10 000	100	9 900	9 000	0
F	10 000	100	9 900	9 900	0

- Feature sets: FINE (A–D), COARSE (E–F), and COMBINED (A–F).
- 120 features can be extracted in total.

The FPO Evaluation Measure

Motivation:

- Need a theoretical and architecture independent measure.
- We propose “FPO” (floating point operations).

Notation:

- Assume a hardware architecture with s computation nodes.
- s: also number of sub-domains.
- n_i: number of unknowns in a sub-domain.
- I: number of domain decomposition iterations.
- $FPO \approx \sum_{i=1}^{s} \frac{n_i^3}{3} + I \cdot n_i^2$.

Note — FPO is only comparable for solutions with:

- Same number of sub-domains, and
- Same hardware architecture.

Machine Learning Methodology

Training. For each diffusion file:

1. Extract features for all 48 permutations of the neighborhoods.
2. Compute FPO for all 48 permutations with simulation.
3. Record the mapping from the set of input features to FPO.

Testing. For each diffusion file:

1. Extract features for all 48 permutations of the neighborhoods.
2. Predict FPO for all 48 permutations using a regression model.
3. Identify the minimum FPO value for each neighborhood.

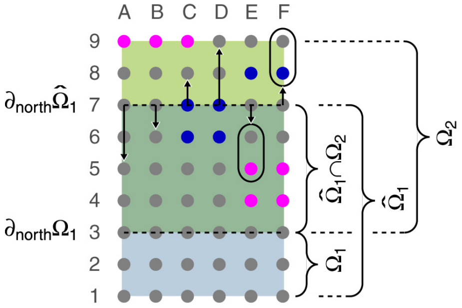
Evaluation. For each diffusion file:

1. Compute FPO on the best combined specification with simulation.
2. Compare the predicted FPO score with that of the baseline.

Baseline Overlap Decision

Global overlap	Total overlap for various grid sizes (% of unknowns) 1×1	2×2	3×3	$\mathbf{4 \times 4}$	5×5	6×6	7×7	8×8
minimum	0.00	0.40	0.80	1.19	1.59	1.99	2.38	2.77
0.2%	0.00	1.19	2.38	3.56	4.73	5.90	7.06	8.21
0.4%	0.00	1.99	3.95	**5.90**	7.82	9.73	11.62	13.48
0.6%	0.00	2.77	5.51	8.21	10.87	13.48	16.06	18.60
0.8%	0.00	3.56	7.06	10.49	13.85	17.16	20.40	23.57

Data Analysis

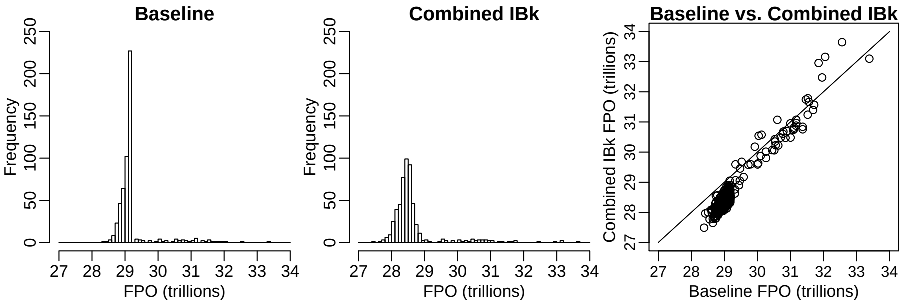


Region	Max Value in Region	Min Value in Region	Max Diff in Region	Max Diff to Boundary	Min Diff to Boundary
A	10 000	1 000	9 000	9 000	0
B	1 000	100	900	0	0
C	1 000	100	900	900	0
D	10 000	1 000	9 000	9 900	0
E	10 000	100	9 900	9 000	0
F	10 000	100	9 900	9 900	0

Region A Feature	Invalid	No Diff	Default	**Other**	Total
Max Value in Region	12 000	0	34 171	1 829	48 000
Min Value in Region	12 000	0	35 990	10	48 000
Max Diff in Region	12 000	34 171	0	**1 829**	48 000
Min Diff to Boundary	12 000	36 000	0	0	48 000
Max Diff to Boundary	12 000	35 361	0	**639**	48 000

Regression Algorithms and Feature Sets

- Algorithms: simple linear, nearest neighbor, decision tree, and SVM.
- Feature sets: COMBINED, FINE, and COARSE.
- Only the nearest neighbor algorithm offered improvement (below).



Evaluation metric	COMBINED	FINE	COARSE
Fraction of baseline	0.9778	0.9791	0.9830
Student's t-test	$p < 2.2 \times 10^{-16}$	$p < 2.2 \times 10^{-16}$	$p < 2.2 \times 10^{-16}$
Cohen's d	$d = 0.85$	$d = 0.79$	$d = 0.62$

Forward Plan

1. Increase the checkerboard size for more precise learning.
2. Increase the training set size with additional diffusion specifications.
3. Apply non-uniform boundary adjustments with sub-domains.
4. Drop the checkerboard constraint in favor of polygonal boundaries.
5. Consider three-dimensional problems later.

Summary

- We have proposed a method for learning overlap optimization.
- New feature sets have been developed.
- The FPO evaluation metric has been developed.
- Results to date are a step in the right direction.

Thankyou!

Steven Burrows
steven.burrows@uni-weimar.de
www.webis.de