Chair for Scientific Computing in Computer Science, Technische Universität München

# Efficient Regression for Big Data using Adaptive Sparse Grids

Michael Lettrich, m.lettrich@mytum.de

## Problem Statement

In Regression we describe an unknown, dependent variable by independent variables using a (multivariate-) function.

Challenges:

- Independent variables are given indirectly as a set of samples which are usually obtained empirically
- Samples from empirical data are usually perturbed
- Complex models with many variables are high dimensional

Objective:

- Find a robust approximation for the underlying function
- Design algorithms that can deal with high dimensional datasets

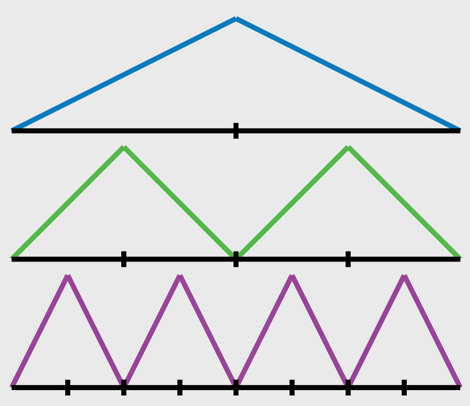
## Sparse Grids

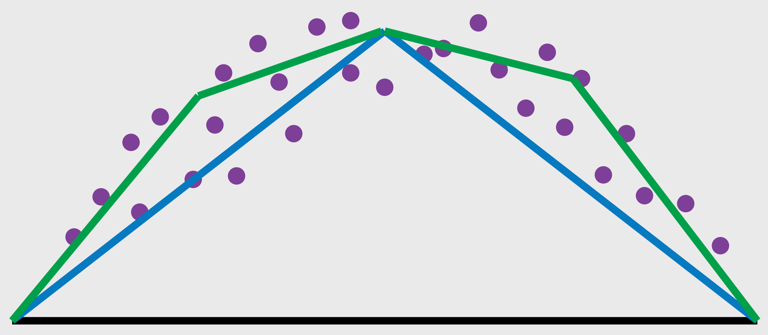
Curse of Dimensionality:

- Regular grids with *n* points per dimension require $n^d$ grid points
- Even for small dimensionality computationally unfeasible

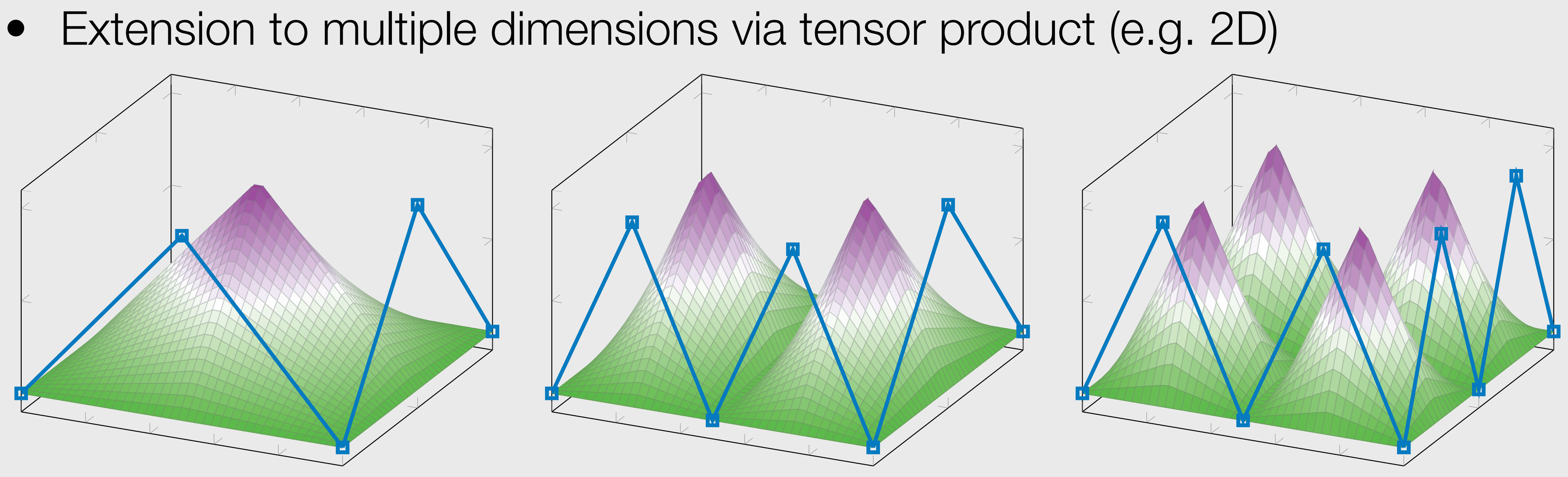
Sparse Grids:

- Use multi-scale approximation based on sets of hierarchical basis functions

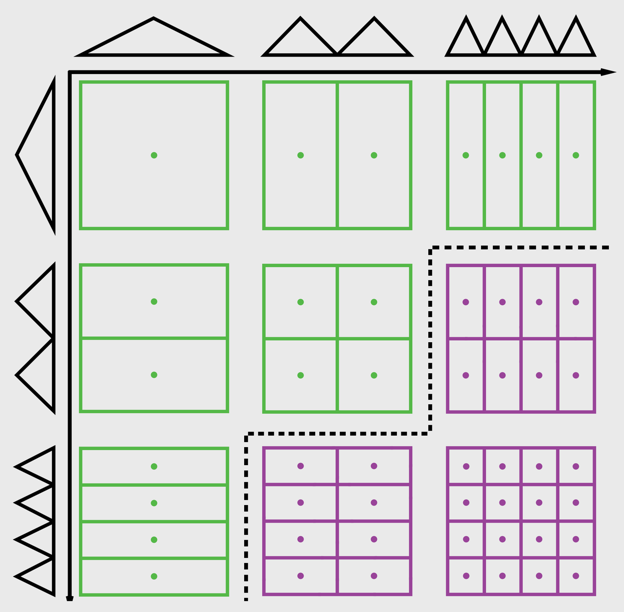


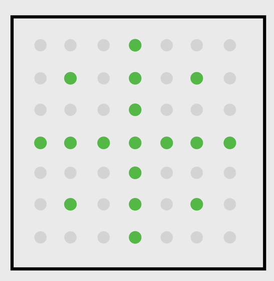


- Extension to multiple dimensions via tensor product (e.g. 2D)



- Omitting basis functions without a big impact on accuracy is possible, reducing amount of grid-points to $n (\log n)^d$, postpone curse of dimensionality





- Multi-scale approach is fully adaptive and well suited for regression

## Regression using Adaptive Sparse Grids

Regression as an Optimization Problem:

- Approximate regression function by a linear combination of sparse grid basis functions

$$f(\vec{x}) \approx u(\vec{x}) = \sum_{i=1}^{m} \alpha_i \varphi_i(\vec{x})$$

- Find optimal linear combination for a fixed set of basis functions by finding optimal weights for each basis function using least squares

$$\underset{u \in V_n^{(d)\text{sparse}}}{\operatorname{argmin}} \left( \frac{1}{m} \sum_{i=1}^{m} (y_i - u(\vec{x}_i))^2 + \lambda \mathcal{C}(u) \right).$$

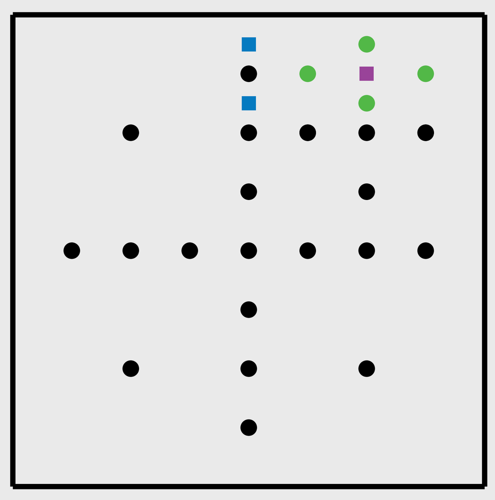
- Perform regularization to ensure smoothness
- Optimization problem only grows linearly with the amount of samples

Adaptivity:

- Iteratively improve approximation quality of sparse grid by adding new basis functions

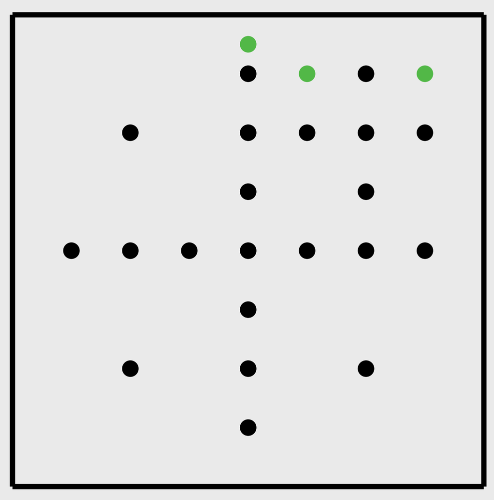
Local Adaptive Refinement (state of the art)

- Identify high error neighborhoods based on the current approximation and add basis functions in all dimensions
- Maintain hierarchical structure by adding missing hierarchical ancestors



Local and Dimension Adaptive Refinement (new)

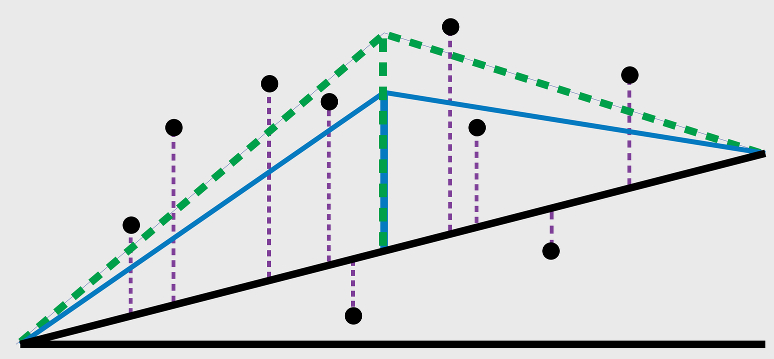
- Add individual basis functions which promise highest possible error reduction according to heuristic
- Hierarchical structure remains unharmed



Heuristic for Local and Dimension Adaptive Refinement

- Use mean weighted local approximation errors (at least *O(d)* samples required) to estimate yet unknown weights of new basis functions

$$\hat{\alpha} = \frac{1}{m} \sum_{i=1}^{m} \varphi(x_i)(y_i - u_{old}(x_i))$$



Why Local and Dimension Adaptive Refinement?

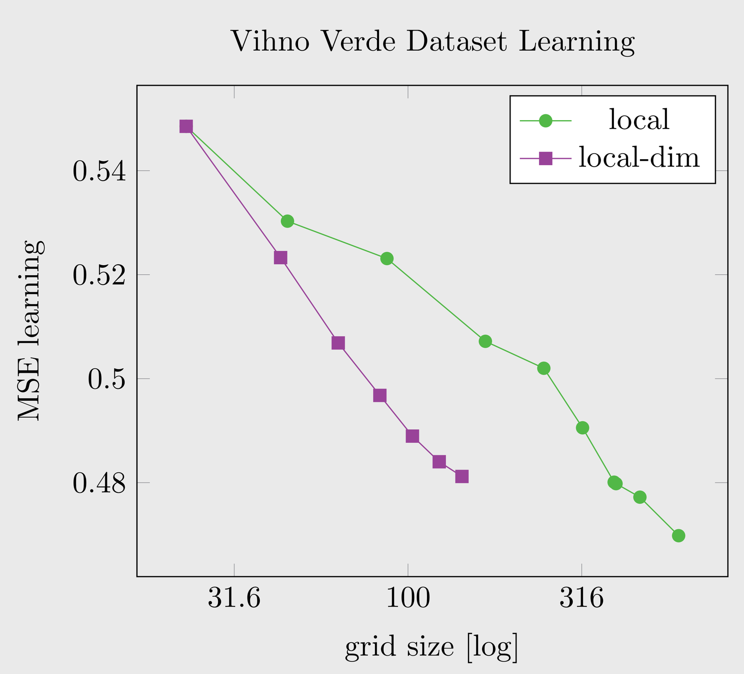
- Smaller grids
- Force faster decay of approximation error

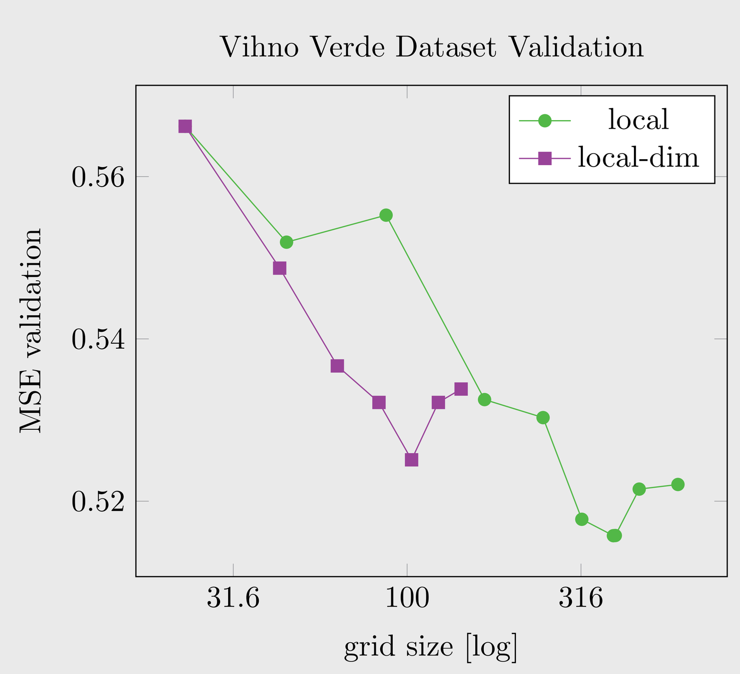
=> Tackle larger problems more efficiently while maintaining good approximation quality of state of the art local adaptive methods

## Experiments and Results

Example — Vino Verde Data-set :

- Regression data-set from the UCI repository donated by Paulo Cortez et al. from the University of Mihno Portugal
- 4898 samples of white vine with 11 independent variables (pH, alcohol, …)
- Quality (dependent variable) rated from 1 (bad) to 10 (excellent)




Test Parameters:

- 5 fold cross-validation, results display median of all folds.
- Best parameter settings for both methods determined empirically

Results:

| Adaptive Technique | Local | Local + Dimensional |
|---|---|---|
| Refine/Add | 1 | 20 |
| Regularization | $10^{-5}$ | $10^{-5}$ |
| Steps Required | 7 | 4 |
| Final Grid Size | 392 | 103 |
| Training Error (MSE) | 0.480 | 0.489 |
| Testing Error (MSE) | 0.516 | 0.525 |

- Decreasing runtime of optimization problem ($O(n^3)$) due to smaller grid
- Local + dimensional algorithm usually requires higher regularization (up to 10x)
- Similar results have been achieved for the UCI data-set *Compression strength of Concrete* and the synthetic *Friedman 1* data-set

References:

- Hans-Joachim Bungartz and Michael Griebel. Sparse grids. Acta Numerica, 13:147– 269, 2004.
- Dirk Pflüger. Spatially Adaptive Sparse Grids for High Dimensional Problems. PhD thesis, Technische Universität München, 02 2010.
- Paulo Cortez et al.. Modeling wine preferences by data mining from physicochemical properties. Decision Support Systems, 47(4):547–533, 2009.