On generalized sparse grids

Michael Griebel

Outline

1. High-dimensional problems and curse of dimensionality
2. Dimension decomposition of functions
3. Sparse grids
4. Energy-norm based sparse grids
5. Adaptive sparse grids
6. Applications

High(er) dimensional problems

- Classical physics: most problems in 3d space+time, compl. geometry
- Higher dimensional problems ?
  - PDEs from mathematical modelling, stochastics
    - diffusion equation, Fokker-Planck equation,
      - diffusion approximation of discrete processes, networks (Mitzlaff,Dai)
      - viscoelasticity in polymer fluids (Rousse), reaction mechanisms in biology and chemistry (Sjoeberg, Loetstedt, Hegland),, option pricing,
      - homogenization with multiple scales (Cioranescu,, Hoang, Matache, Schwab)
    - quantum mechanics, Schrödinger equation (Yserentant, Flad)
    - data analysis, statistical learning (Garcke, Hegland)
    - stochastic PDEs (Todor,Schwab,Matthies)
  - Domain simple, product structure
    - $[0,1]^d$, $[-a,a]^d$, hypersphere S_d, R^d with decay for $x_i \to \pm\infty$

Curse of dimension

- $f : \Omega^{(d)} \to \Re, \quad f \in V^{(r)},$ r isotropic smoothness
- Bellmann ´61: curse of dimension

$$\| f - f_N \|_{H^s} = C(d) \cdot N^{-r/d} \, | f |_{H^{s+r}} = O(N^{-r/d})$$

- Find situations where curse can be broken ?
- Trivial: restrict to $r = O(d)$

$$\| f - f_N \| = O(N^{-cd/d}) = O(N^{-c})$$

but practically not very relevant

Curse of dimension

- Consider class of functions of $\Re^d$ with $\nabla f \in FL_1$ where FL_1 class of functions with Fourier transform in L_1
  => expect $\| f - f_N \| = O(N^{-1/d})$
- But Barron ´93 showed $\| f - f_N \| = O(N^{-1/2})$
- Meanwhile other function classes known
  - Radial basis schemes, Gaussian bumps, (Y. Meyer)
    Niyogi, Girosi ´98: ball in Besov space $B^d_{1,1}(\Re^d) \Rightarrow r = d$
  - Stochastic sampling techniques, MC
  - Spaces with bounded mixed derivatives
- In any case: some smoothness changes with d

Concentration of measure

- What means smoothness for $d \to \infty$ anyway?
- Concentration of measure: (Milman´88, Talagrand´95, Gromov ´99)

f Lipschitz with constant L on d-sphere,

P normalized Lebesgue measure,

X uniformly distributed

Then: $P(|f(X) - Ef(X)| > t) \leq c_1 \exp(-c_2 t^2 / L^2)$

=> every Lipschitz function on sufficiently high-dimensional domain is well approximated by constant function ! (Hegland, Pozzi ´05)

Lemma of Kolmogorov

- Kolmogorov ´56:

  ex. 2d+1 cont. strictly increasing functions $\varphi_i : (0,1) \to (0,1)$
  ex. d constants $\lambda_i, \sum \lambda_i \leq 1$

$$f(x_1,\dots,x_d) = \sum_{i=1}^{2d+1} g\Big(\sum_{j=1}^{d} \lambda_j \varphi_i(x_j)\Big)$$

  for some (non-smooth) $g \in C(0,1)$ dependent on f

  but: non-constructive result,

  G.,Braun 2009: recent constructive proof in Constructive Approximation

- IBC, weighted RK Hilbert spaces, Wozniakowski, Sloan

=>There is hope for high-dimensional problems

Approach

- Basic principles:
  - 1dim series expansion with decay
  - d-dim product construction
  - Trunctation of resulting multivariate expansion
- Effect:
  - reduction of cost complexity
  - nearly same accuracy as „full“ product
  - necessary: certain smoothness requirements

Introductory examples

- Napier´s multiplication (John Napier (1550 –1617)
- Archimedes´ approach for pi and Cavalieri´s/Fubini´s theorem

- Sparse grids for integration, approximation of functions and PDEs, etc.

Summary

Classical approach: $d = 1,..,3/4$

curse of dimension and intractability

$$\| f - f_N \|_{H^s} = c(d) \cdot N^{-r/d} \, | f |_{H^{s+r}} = O(N^{-r/d})$$

Stronger regularity/norms

curse only wrt log-terms

$$\| f - f_N \|_{H^s} = c(d) \cdot N^{-r} (\log(N))^{(d-1)/2} \, | f |_{H^{s+r}_{mix}}$$

or no curse at all

$$\| f - f_N \|_{H^s} = c(d) \cdot N^{-r} \, | f |_{H^{s+r}_{mix}}$$

but still not tractable,
constant grows exponentially

$d = 1,..,10/12$

Lower effective dimension and lower-dim. manifolds

no curse due to effective dimension

$$\| f - f_N \|_{H^s} = c(d^{eff}) \cdot N^{-r/d^{eff}} \, | f |_{H^{s+r}}$$

and constant grows exponentially only wrt effective dimension

$d = 1,...100..$

$d^{eff} = 1,..,10$

Function decomposition

- splitting of one-dimensional space $V = C \oplus W$
  projection $P : V \to C$ onto subspace of constants
- splitting of d-dim. space by tensor product

$$
\begin{aligned}
V^{(d)} = \bigotimes_{i=1}^{d} (C^{(i)} \oplus W^{(i)}) = \; & C^{(1)} \times \cdots \times C^{(d)} + \\
& \sum_{i=1}^{d} C^{(1)} \times \cdots \times W^{(i)} \times \cdots \times C^{(d)} + \\
& \sum_{i=1}^{d} \sum_{i<j} C^{(1)} \times \cdots \times W^{(i)} \times \cdots \times W^{(j)} \times \cdots \times C^{(d)} + \\
& \sum_{i=1}^{d} \sum_{i<j} \sum_{j<k} C^{(1)} \times \cdots \times W^{(i)} \times \cdots \times W^{(j)} \times \cdots \times W^{(k)} \times C^{(d)} \\
& \vdots \\
& W^{(1)} \times \cdots \times W^{(d)}
\end{aligned}
$$

Function decomposition

- splitting of associated d-dim function

$$f(x_1,\ldots,x_d) = \sum_{u\subseteq\{1,\ldots,d\}} f_u(x_u)$$

$$= f_0 + \sum_{i=1}^{d} f_i(x_i) + \sum_{i=1}^{d}\sum_{i<j} f_{i,j}(x_i,x_j) +$$

$$\sum_{i=1}^{d}\sum_{i<j}\sum_{j<k} f_{i,j,k}(x_i,x_j,x_k) + \cdots + f_{1,\ldots,d}(x_1,\ldots,x_d)$$

- 2^d subspaces, 2^d terms
- decomposition into correlations, clusters
- Choice of one-dimensional projector P ?
  - integral mean => ANOVA decomposition, induces decomposition of variance of f (Efron, Stein, Wahba, Owen, Hickernell)
  - evaluation at one fixed point => Anchor ANOVA

Function decompositions

- 2d Example:

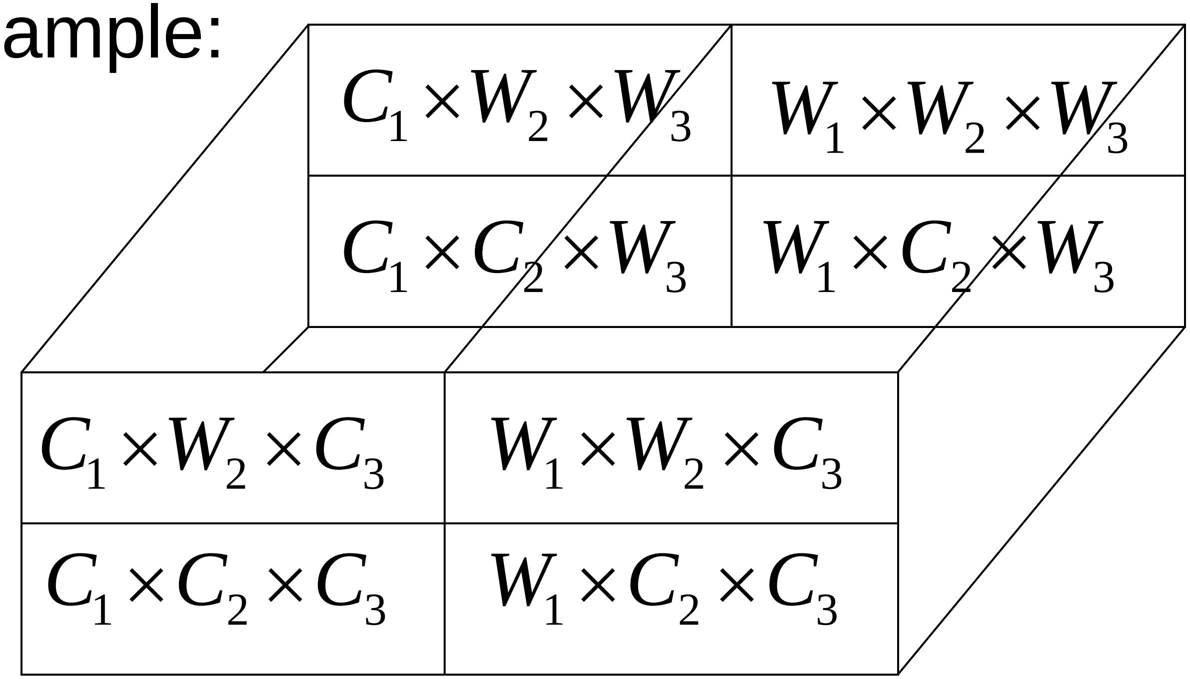
$$V^{(2)} =$$

$C_1 \times W_2$	$W_1 \times W_2$
$C_1 \times C_2$	$W_1 \times C_2$

$$f(x_1, x_2) = f_0 + f_1(x_1) + f_2(x_2) + f_{1,2}(x_1, x_2)$$

- 3d Example:

$$V^{(3)} =$$



$$f(x_1, x_2, x_3) = f_0 + f_1(x_1) + f_2(x_2) + f_3(x_3) + f_{1,2}(x_1, x_2) + f_{1,3}(x_1, x_3) + f_{2,3}(x_2, x_3) + f_{1,2,3}(x_1, x_2, x_3)$$

Function decompositions

- approximation by truncation after q-order terms

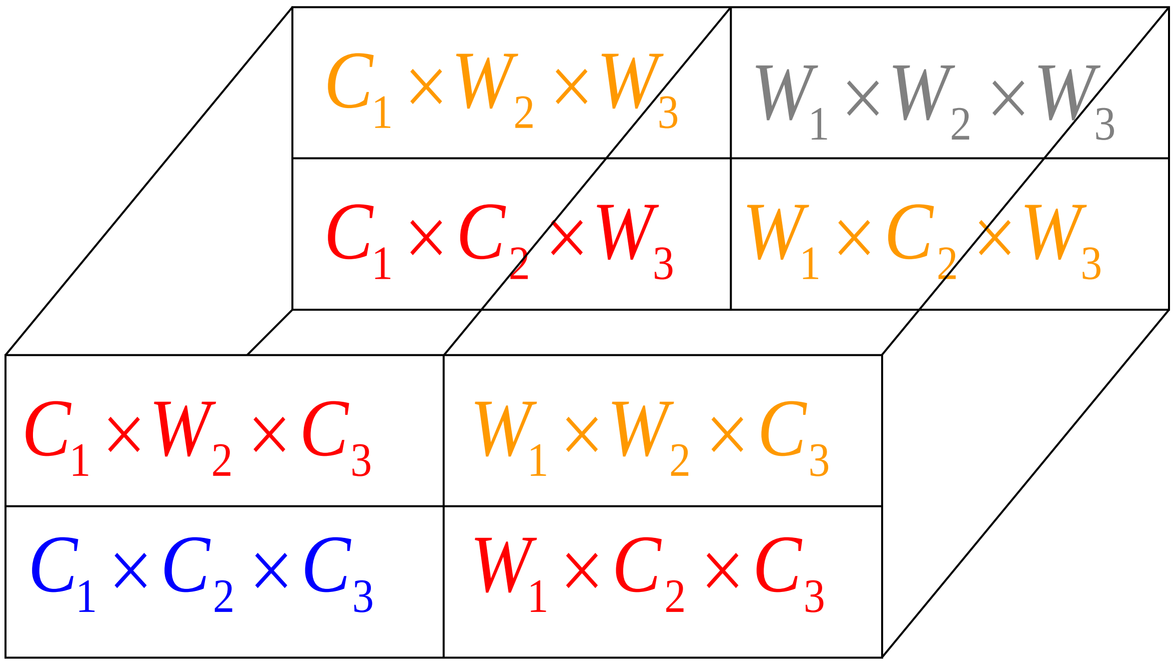
$$f(x_1,\ldots,x_d) \approx \sum_{u\subseteq\{1,\ldots,d\},|u|\le q} f_u(x_u)$$

$q = 0$

$q = 1$

$q = 2$

$q = 2$



$$f(x_1,x_2,x_3) = f_0 + f_1(x_1) + f_2(x_2) + f_3(x_3) + f_{1,2}(x_1,x_2) + f_{1,3}(x_1,x_3) + f_{2,3}(x_2,x_3) + f_{1,2,3}(x_1,x_2,x_3)$$

Function decompositions

- Fast decay of series or even finite order q << d?
- surely not in general, but: consider f as input-output model

$$(x_1,\ldots,x_d) \rightarrow f(x_1,\ldots,x_d)$$

- „Correlated“ effects of the input variables ?
- Many body expansion of potential energy surface of molecular systems: mostly only two-, three- or four-body potentials (i.e. q=4) for physical reasons
- Cluster expansions in statistical mechanics
- Statistics: second order, covariances but i.g. not more
- Data-mining: MARS, only up to q=5,..,7 for real data

Truncation

- Truncation after q terms introduces a modelling error
- The remaining subspaces needs to be finitely represented => discretization error
- After truncation after q terms no more balancing of modelling error and subsequent discretization error possible.
- Unnatural distinction between modelling error and subsequent discretization error
- better: relate it somehow

Further decomposition of W: Sparse grids

- decompose 1d subspace W further $$W = \sum_{l=1}^{\infty} W_l$$
- tensor product and subsequent truncation

  => sparse grid representation
- Fourier series or polynomials (global)

  => Korobov-spaces, hyperbolic cross approximation
- piecewise polynomials (local)

  hierarchical basis, interpolets, wavelets, multilevel basis

  => sparse grid finite element spaces

History of Sparse Grids

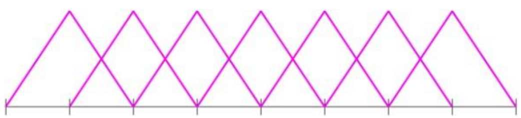
re-invented several times:

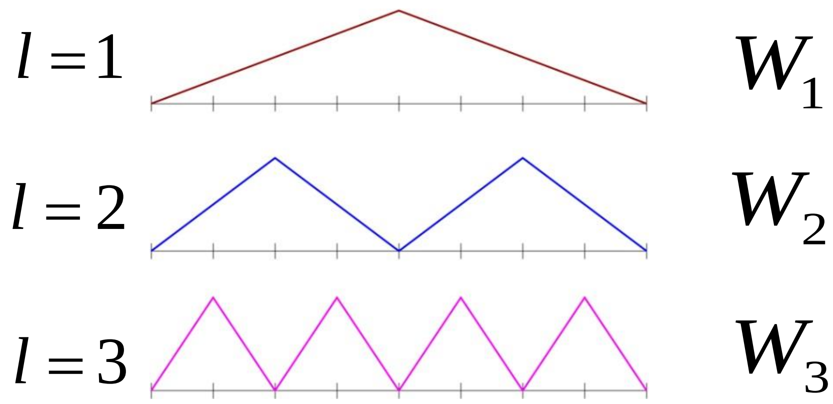
1957	Korobov, Babenko	hyperbolic cross points
1963	Smolyak	
1971	Gordon	blending method
1980	Delvos, Posdorf	Boolean interpolation
1990	Zenger, G.	sparse grids
1998	Stromberg, deVore	hyperbolic wavelets

application areas include:

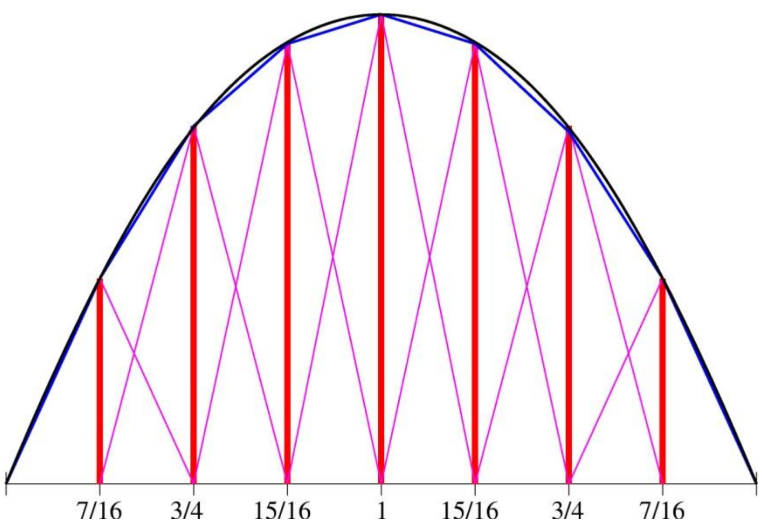
- quadrature (Novak, Ritter)
- interpolation
- data compression
- solution of PDEs
- integral equations
- eigenvalue problems

Example: Hierarchical basis

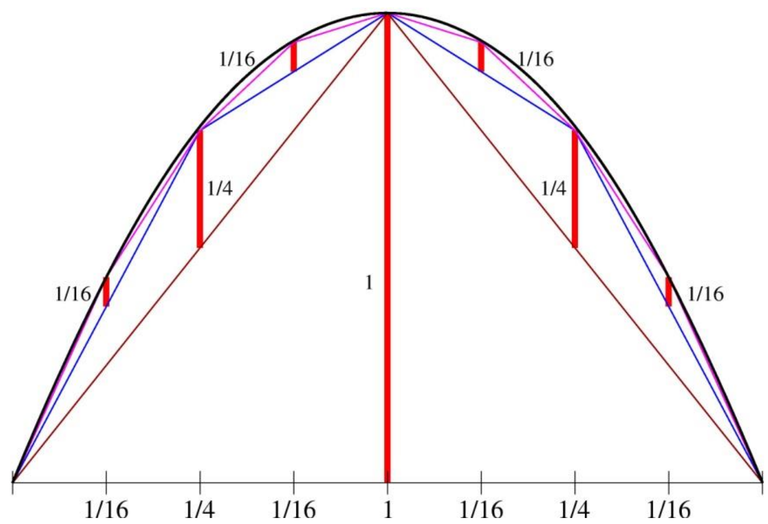




parabola $f(x) = -(x-1)(x+1)$ in [-1,1]



conventional coefficients
no decay from level to level



hierarchical coefficients
decay by ¼ from level to level

Tensor product hierarchical basis

Generalization to higher dimension by tensor product

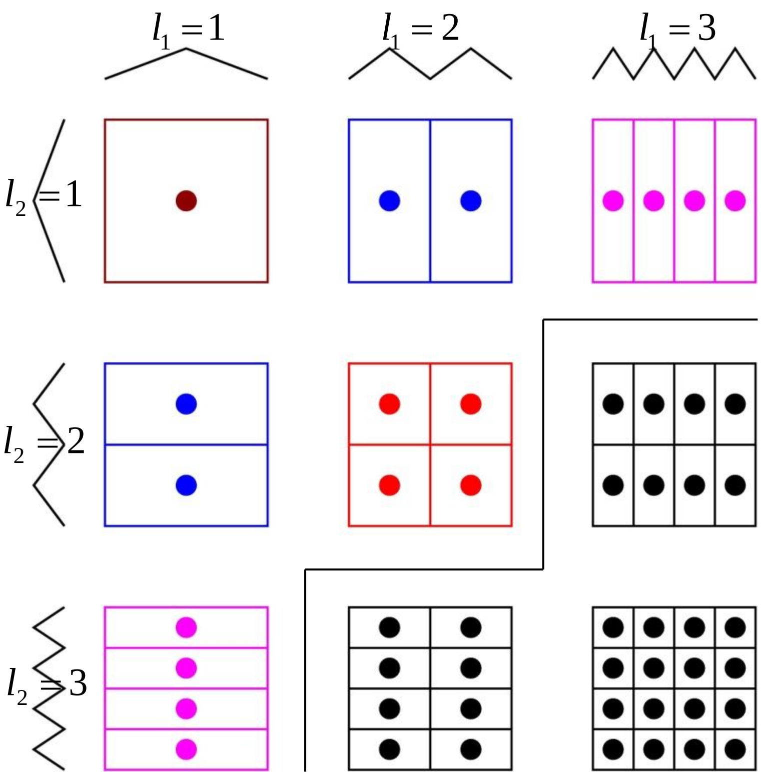


Table of subspaces $W_{l_1 l_2}$

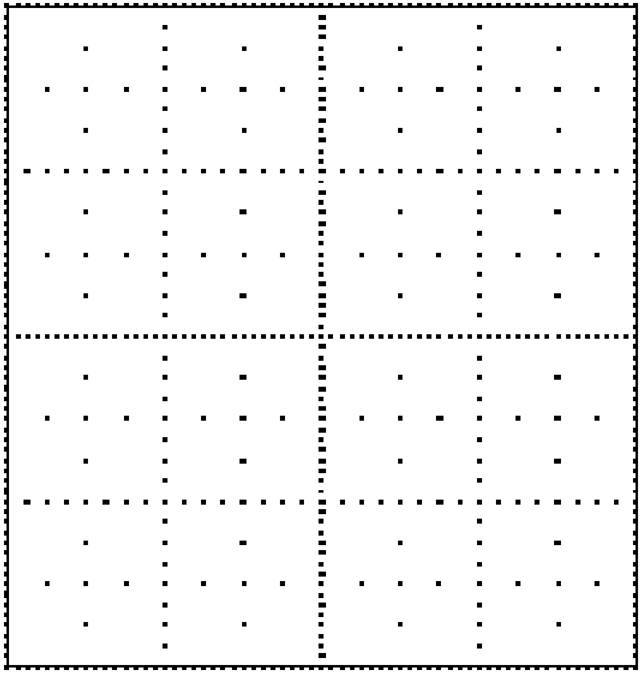
$\frac{1}{256}$	$\frac{1}{64}$	$\frac{1}{256}$	$\frac{1}{16}$	$\frac{1}{256}$	$\frac{1}{64}$	$\frac{1}{256}$
$\frac{1}{64}$	$\frac{1}{16}$	$\frac{1}{64}$	$\frac{1}{4}$	$\frac{1}{64}$	$\frac{1}{16}$	$\frac{1}{64}$
$\frac{1}{256}$	$\frac{1}{64}$	$\frac{1}{256}$	$\frac{1}{16}$	$\frac{1}{256}$	$\frac{1}{64}$	$\frac{1}{256}$
$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{16}$	1	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{1}{16}$
$\frac{1}{256}$	$\frac{1}{64}$	$\frac{1}{256}$	$\frac{1}{16}$	$\frac{1}{256}$	$\frac{1}{64}$	$\frac{1}{256}$
$\frac{1}{64}$	$\frac{1}{16}$	$\frac{1}{64}$	$\frac{1}{4}$	$\frac{1}{64}$	$\frac{1}{16}$	$\frac{1}{64}$
$\frac{1}{256}$	$\frac{1}{64}$	$\frac{1}{256}$	$\frac{1}{16}$	$\frac{1}{256}$	$\frac{1}{64}$	$\frac{1}{256}$

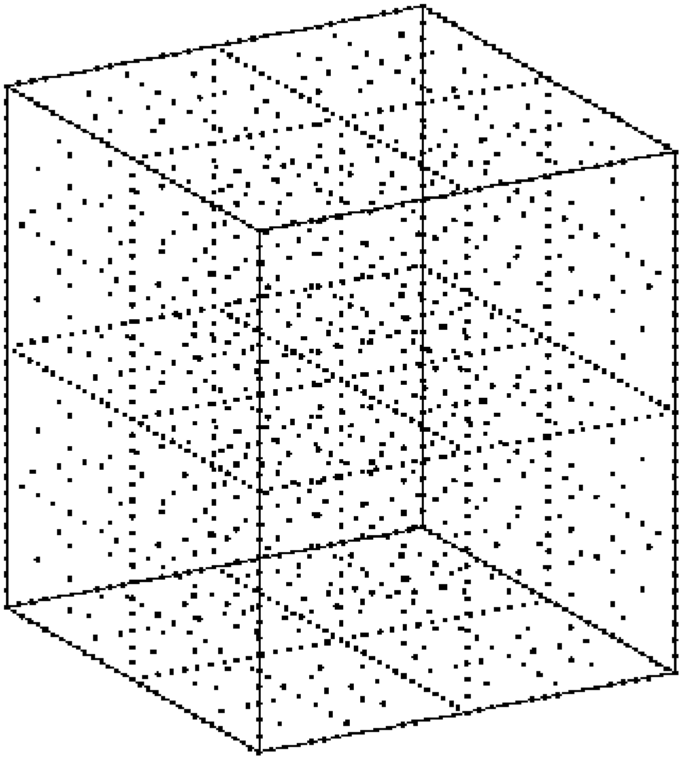
decay in x- and y-direction by 1/4

decay in diagonal direction by 1/16

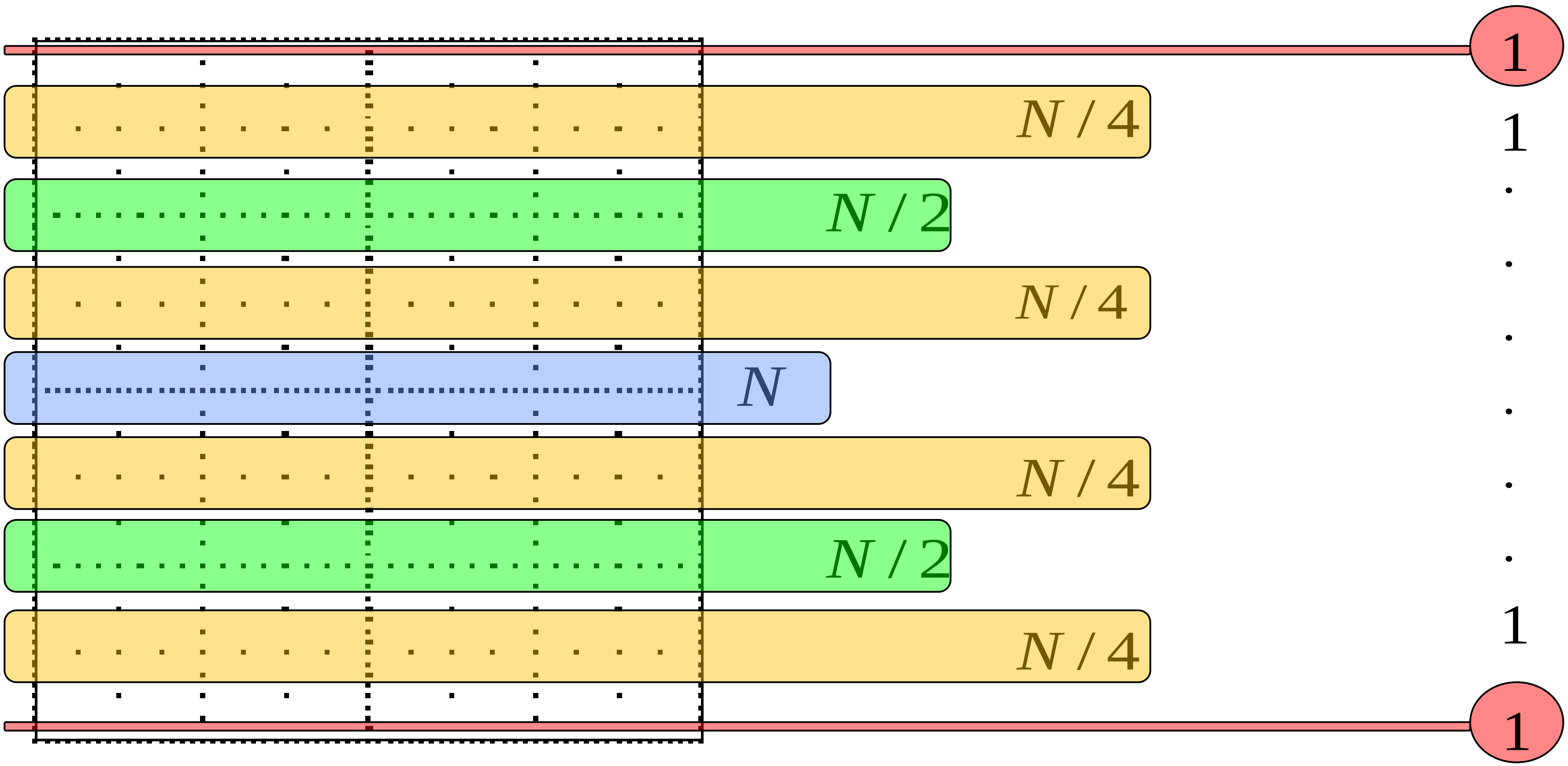
Idea:
Omit points with small associated hierarchial coefficient values

Regular sparse grids





cost complexity (d=2,interior points)



$$N + 2*N/2 + 4*N/4 + \ldots$$

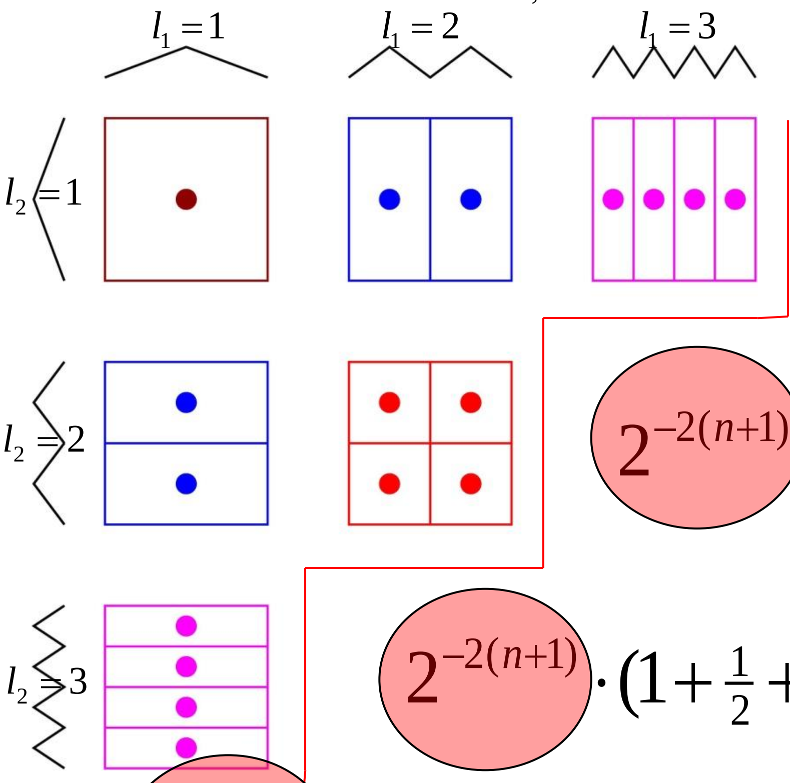
$$\cong \log(N)*N$$

Accuracy: $f(x_1,x_2)=\sum_{l_1,l_2} f_{l_1,l_2}(x_1,x_2)$ $\quad f_{l_1,l_2}(x_1,x_2)=\sum_{i_1,i_2} f_{l_1,i_1,l_2,i_2}\varphi_{l_1,i_1}(x_1)\varphi_{l_2,i_2}(x_2)$

Table of subspaces $W_{l_1 l_2}$

Contribution $\quad \| f_{l_1,l_2} \|_2 \leq 3^{-2}\cdot 2^{-2(l_1+l_2)}\cdot | f |_{\mathbf{2},mix}$

truncate at level n



$2^{-2(n+1)}\cdot(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots)\cdot | f |_{\mathbf{2},mix}\cdot 3^{-2}$

$2^{-2(n+1)}\cdot(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots)\cdot | f |_{\mathbf{2},mix}\cdot 3^{-2}$

$2^{-2(n+1)}\cdot(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots)\cdot | f |_{\mathbf{2},mix}\cdot 3^{-2}$

$2^{-2(n+1)}\cdot(1+\frac{1}{2}+\frac{1}{4}+\frac{1}{8}+\ldots)\cdot | f |_{\mathbf{2},mix}\cdot 3^{-2}$

$2^{-2(n+2)}\cdot\ldots$

$2^{-2(n+3)}\cdot\ldots$

further summation results in

$$\| f - f_n^{SG} \|_0 \leq c_2 \cdot n \cdot 2^n \, | f |_{\mathbf{2},mix}$$

Properties of sparse grids

	Sparse grids		Full grids	
Cost:	$O(N(\log N)^{d-1})$	instead of	$O(N^d)$	
Accuracy:	$O(N^{-2}(\log N)^{d-1})$		$O(N^{-2})$	L_2-norm
	$O(N^{-1})$		$O(N^{-1})$	energy-norm
Smoothness:	$\left\lvert \frac{\partial^{2d} f}{\partial x_1^2 \ldots \partial x_d^2} \right\rvert \leq c$		$\left\lvert \sum_{i=1}^{d} \frac{\partial^2 f}{\partial x_i^2} \right\rvert \leq c$	
Space and seminorm:	$H^2_{mix}, \lvert f \rvert_2$		$H^2, \lvert f \rvert_2$	

breaks curse of dimension of conventional full grids at least to some extent

Note: higher regularity in mixed derivative, r~d

L^2 norm-based sparse grids

- For orthogonal wavelets and general stable multiscale systems we can even obtain

$$O\left(N^{-2}\left(\log N\right)^{(d-1)/2}\right)$$

  - Hint: estimate directly for squared error.
- Complexities with boundary terms:
  - Cost: same order but additional factor of 3^d
  - Accuracy same order
  - Smoothness assumptions related to variation of Hardy and Krause
  - Start multiscale series with constant then linear etc.

L^2 norm-based sparse grids

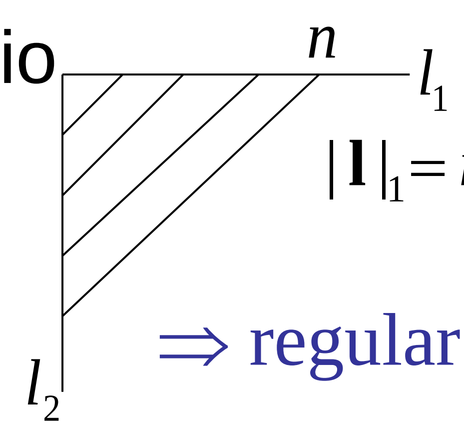
- Representation $\quad f(\mathbf{x}) = \sum_{\mathbf{l}} f_{\mathbf{l}}(\mathbf{x}) \qquad f_{\mathbf{l}}(\mathbf{x}) \in W_{\mathbf{l}}$

  $\mathbf{x} = (x_1,..,x_d) \qquad \mathbf{l} = (l_1,..,l_d)$

- cost per subspace $\quad \dim(W_{\mathbf{l}}) = 2^{|\mathbf{l}-\mathbf{1}|_1}$
- benefit for accuracy $\quad \| f_{\mathbf{l}} \|_2 \leq 3^{-d} \cdot 2^{-2|\mathbf{l}|_1} \cdot | f |_{\mathbf{2}} \; = O(2^{-2|\mathbf{l}|_1})$
- choice of best subspaces ?

  => restricted global optimization problem,

  => local benefit²/cost ratio



$|\mathbf{l}|_1 = n + d - 1$ isoline

$$V_n^{(d,opt)} = \bigoplus_{|\mathbf{l}|_1 = n+d-1} W_{\mathbf{l}}$$

$\Rightarrow$ regular sparse grid space

Energy-norm based sparse grids

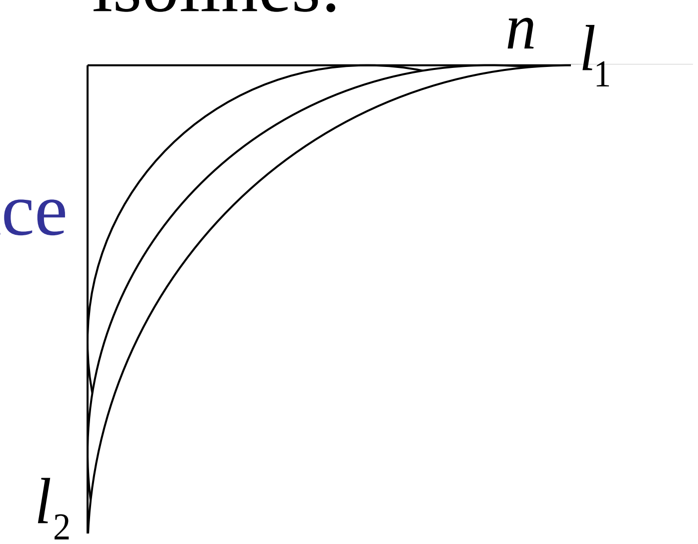
- energy norm $\| f \|_E := \left(\int_\Omega \sum_{j=1}^{d} \left(\frac{\partial f(\mathbf{x})}{\partial \mathbf{x}_j} \right)^2 d\mathbf{x} \right)^{1/2}$
- benefit for accuracy

$$\| f_{\mathbf{l}} \|_E \leq \frac{1}{2 \cdot 12^{(d-1)/2}} \cdot 2^{-2|\mathbf{l}|_1} \cdot \left(\sum_{j=1}^{d} 2^{2l_j} \right) \cdot | f |_2 = O(2^{-2\cdot|\mathbf{l}|_1} \cdot (\sum_{j=1}^{d} 2^{2\cdot l_j})^{1/2})$$

- Now benefit/cost ratio

$$b^2(\mathbf{l}) / c(\mathbf{l}) \approx \frac{2^{-4\cdot|\mathbf{l}|_1}}{2^{-|\mathbf{l}|_1}} \cdot \sum_{j=1}^{d} 2^{2\cdot l_j}$$

isolines:



$\Rightarrow$ energy-norm based sparse grid space

$$V_n^{(d,E)} = \bigoplus_{|\mathbf{l}|_1 - 1/5 \cdot \log_2(\sum_{j=1}^{d} 4^{l_j}) \leq n + d - 1 - 1/5 \cdot \log_2(4^n + 4d - 4)} W_{\mathbf{l}}$$

Energy-norm based sparse grids

- Properties: complexities now independent of d

$$\dim(V_n^{d,E}) = O(2^n) \qquad \| f - f_n^E \|_E = O(2^{-n})$$

- What about the constants ?

$$\dim(V_n^{d,E}) \le \tfrac{d}{2}(1-2^{-2/3})^{-d} \cdot 2^n \le \overbrace{\tfrac{d}{2} e^d}^{c_2} \cdot 2^n$$

$$\| f - f_n^E \|_E \le \underbrace{\frac{d}{3^{(d-1)/2} \cdot 4^{d-1}} \cdot \left(\tfrac{1}{2} + \left(\tfrac{5}{2}\right)^{d-1}\right)}_{c_1} \cdot | f |_2 \cdot 2^{-n}$$

Thus:

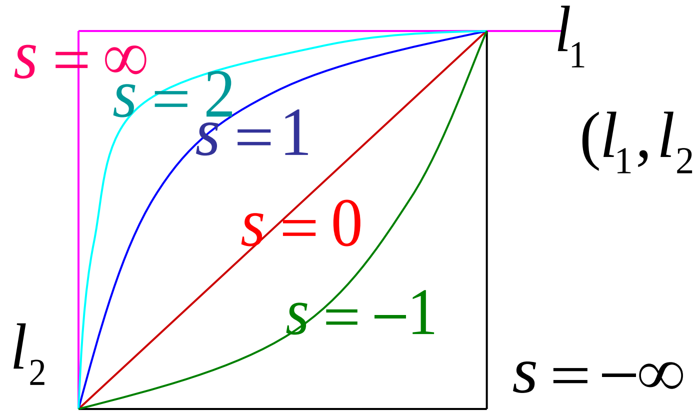
$$\| f - f_n^E \|_E \le c_1 \cdot 2^{-n} | f |_2 \le c_1 \cdot \underbrace{c_2 \cdot | f |_2 \cdot \dim(V_n^{d,E})^{-1}}_{2^{-n} \le c_2 \cdot \dim(V_n^{d,E})^{-1}}$$

with constant $c_1 \cdot c_2 = O(d^2 \cdot 0.97515^d)$

Further generalizations

- H^s-norm based optimal sparse grid spaces (G., Knapek)

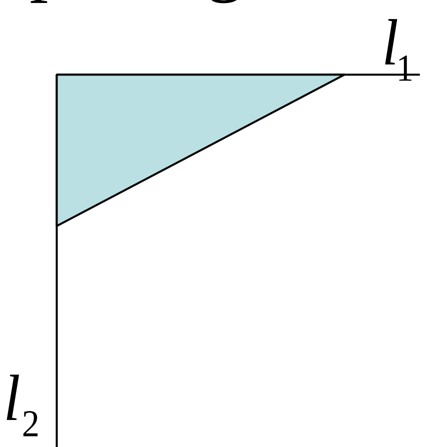
$$s \in (-\infty, \infty)$$



(l_1, l_2) isolines

- More general subspace patterns

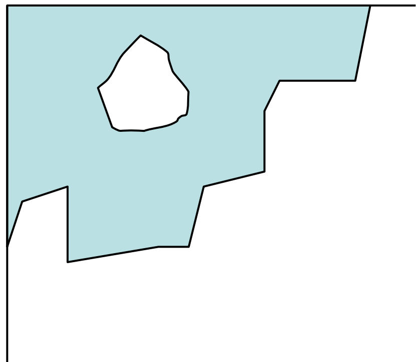
Anisotropic sparse grids



general subset of subspaces

$$V_{\mathfrak{I}}^{(d)} = \bigoplus_{\mathbf{l} \in \mathfrak{I}} W_{\mathbf{l}}$$

$\mathfrak{I} =$ set of indices



Further generalizations

- $H_{mix}^{t,l}$ spaces and regularity assumption (G., Knapek)

$$H_{mix}^{t,l}(I^d) := H^{t\mathbf{1}+l\mathbf{e_1}}(I^d) \cap \ldots \cap H^{t\mathbf{1}+l\mathbf{e_d}}(I^d)$$

$$H_{mix}^{\mathbf{k}}(I^d) := H^{k_1}(I) \otimes \ldots \otimes H^{k_d}(I)$$

Mixture of the standard Sobolev space

$$H^s(I^d) = H_{mix}^{0,s}(I^d)$$

and the space of dominating mixed derivative

$$H_{mix}^t(I^d) = H_{mix}^{t,0}(I^d)$$

Norm equivalency (for stable decompositions, wavelets)

$$\| f \|_{H_{mix}^{t,l}}^2 \approx \sum_{\mathbf{l}} 2^{2t|\mathbf{l}|_1 + 2l|\mathbf{l}|_\infty} \| f_{\mathbf{l}} \|_0^2$$

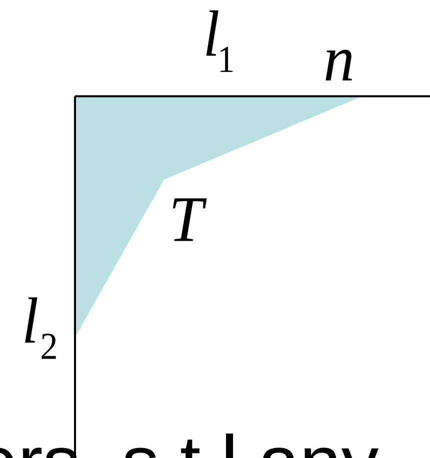
Further generalizations

- Otimization allows again to determine the best sparse grid spaces $V_n^T = \bigoplus_{\mathbf{l} \in I_n^T} W_{\mathbf{l}}$

$$I_n^T = \{\mathbf{l} \in N^d : -|\mathbf{l}|_1 + \tfrac{s}{t}|\mathbf{l}|_\infty \geq -(d+n-1) + \tfrac{s}{t} n\}$$

  with approximations like

$$\inf_{v \in V_n^T} \| f - v \|_{H^s}^2 \leq c 2^{-2(l+t-s)n} \| u \|_{H_{mix}^{t,l}}^2$$



- For a large range of smoothness parameters s,t,l any log-term is avoided in the cost and accuracy estimates
- But the constants may depend strongly on d

- BTW: The solution of Schrödinger´s equation lives in

$$H_{mix}^{1,1}((\Re^3)^d) \qquad H_{mix}^{3/4-\varepsilon,1}((\Re^3)^d)$$

Dimension-adapted sparse grids

- So far: function class known, and a-priori choice of best subspaces by optimization
- Size of benefit/cost ratio indicated if subspace is active => patterns for $\mathfrak{I}$
- Now: for single given function adaptively build up a set $\mathfrak{I}$ of active indices
- Needed:
  - „local“ error indicator for subspace $W_{\mathbf{l}}$
  - refinement strategy to build new index set
  - global stopping criterion

Dimension-adaptive methods

- A proper adaptive algorithm then
  - uses lower resolution in less important dimensions and correlations

  and thus automatically detects
  - important dimensions
  - important correlations between the dimensions
- large reduction of cost if important dimensions are few (small effective dimension, finite order weight spaces), curse of dimensionality broken
- But: no need to know function class a-priori

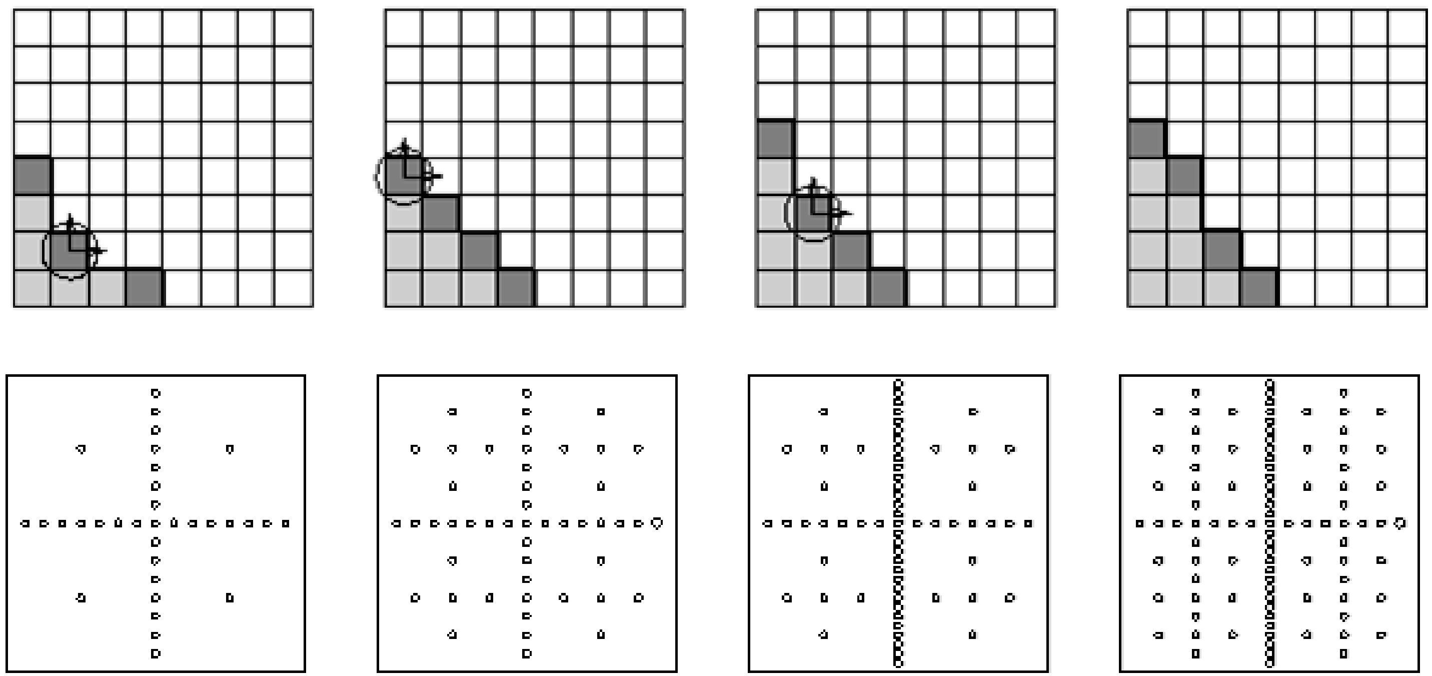
(Hegland ´01, Gerstner,G. ´03, Garcke ´04)

Example (Index Sets)

Evolution of the algorithm:

index sets:

corresponding grids:



Special data structures for the bookkeeping of the different index sets required. => $O(d^2)$

Error Estimation

- differential integral for index ***l***

$$d_{\mathbf{l}} = \| f_{\mathbf{l}} \|$$

can be used as ***local error estimate***

- problem: too early stopping (no saturation)
- solution: consider also ***involved work***

$$n_{\mathbf{l}} = | W_{\mathbf{l}} |$$

and use as estimate

$$\max\left\{ w\frac{d_{\mathbf{l}}}{d_{\mathbf{1}}}, (1-w)\frac{n_{\mathbf{1}}}{n_{\mathbf{l}}} \right\}$$

with weight $w \in [0,1]$.

(Gerstner, G. ´03)

A simple example

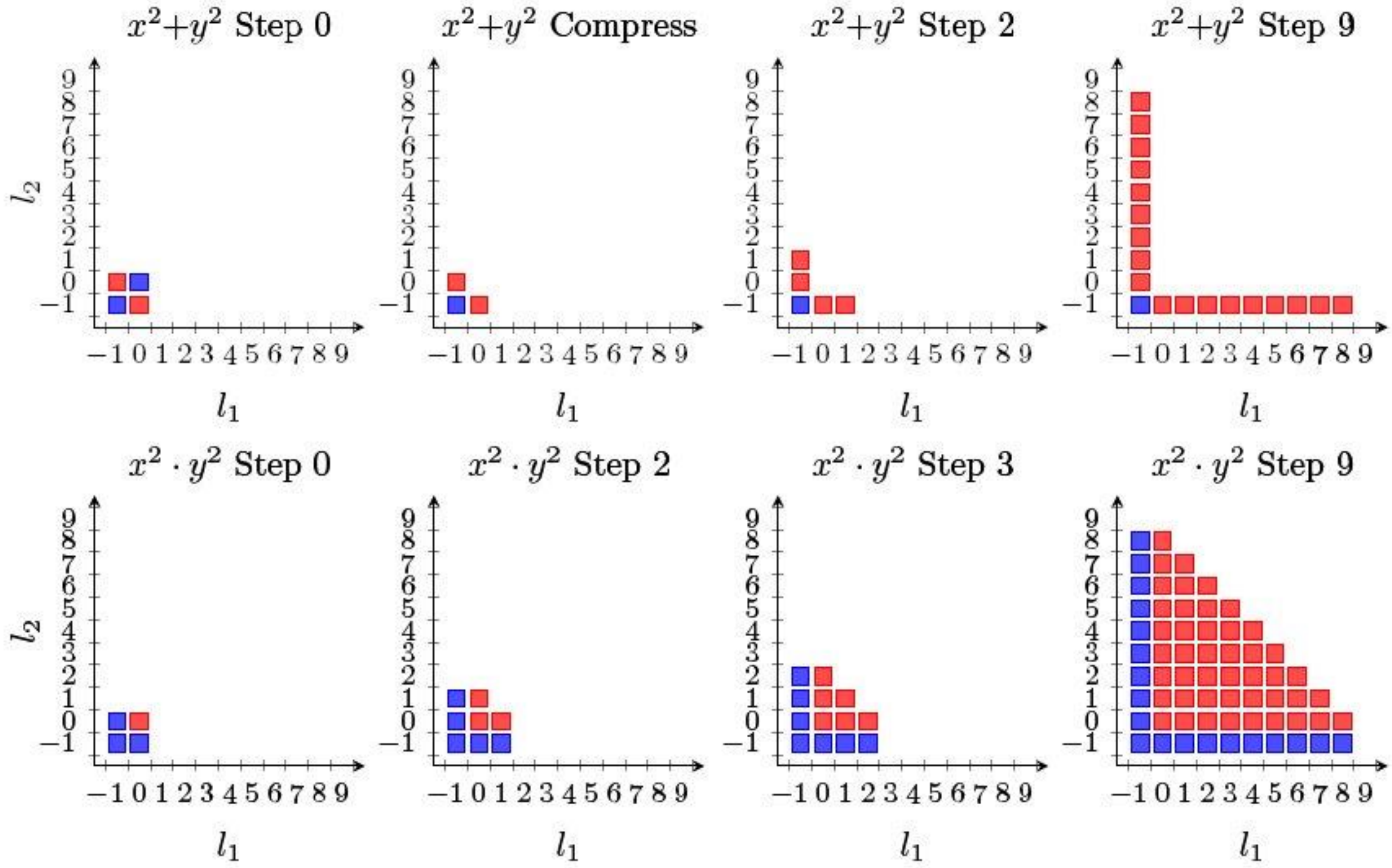


Figure 4.14: Dimension adaptive refinement with the new ANOVA based admissibility criterion for $f(x, y) = x^2 + y^2$ (top row) and $f(x, y) = x^2 \cdot y^2$ (bottom row). Here, ■ (blue) indicates a vanishing contribution $f_l \equiv 0$ whereas ■ (red) indicates a non–vanishing contribution $f_l \not\equiv 0$.

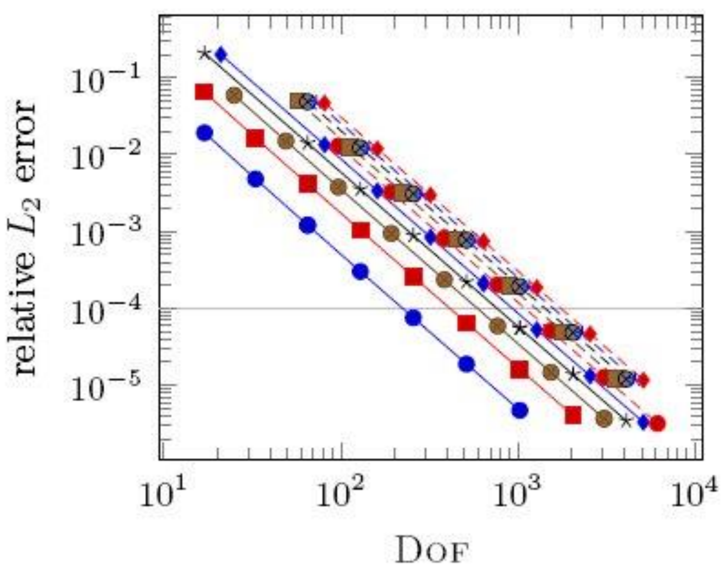
High nominal but low effective dimension

- Model problem $$f(\mathbf{x}) = \sum_{\substack{u \subseteq \{1,\dots,d\} \\ |u|=q}} \prod_{j \in u} g(x_j)$$

$$g(x) = B(\alpha,\beta)^{-1} x^{\alpha-1}(1-x)^{\beta-1}, \text{here with } \alpha = 2, \beta = 5$$

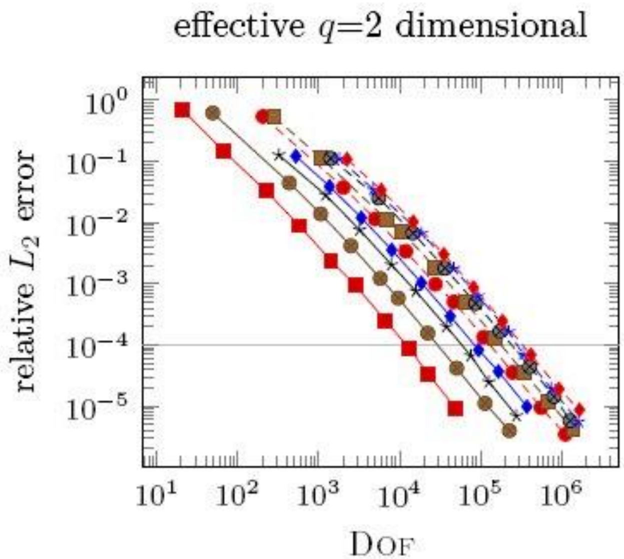
- We expect a behavior of the method as for a smooth q-dimensional function and cost $O\left(\binom{d}{q} N_q\right)$ with N_q cost for one smooth q-dimensional problem

effective q=1 dimensional

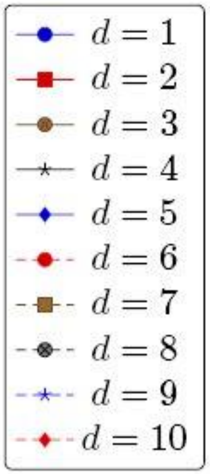


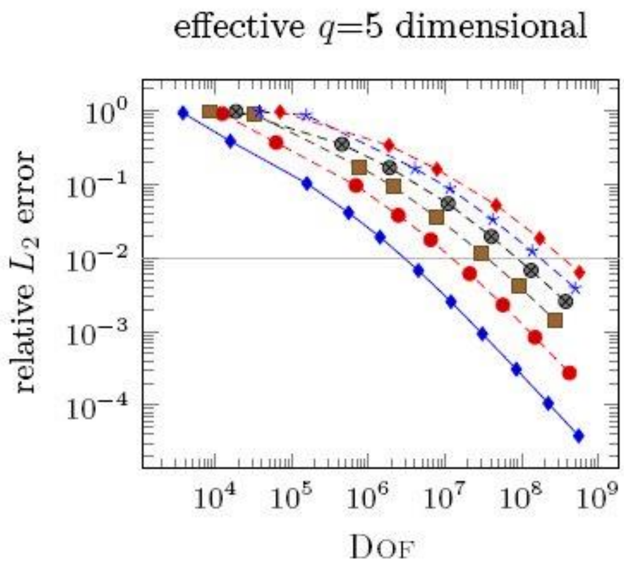
d	N_d^{1-4}	$N_d^{\bar\epsilon}/N_1^{\bar\epsilon}$	$\binom{d}{1}$
1	223	1.0	1
2	413	1.8	2
3	590	2.6	3
4	762	3.4	4
5	930	4.2	5
6	1,096	4.9	6
7	1,260	5.6	7
8	1,424	6.4	8
9	1,586	7.1	9
10	1,748	7.8	10

High nominal but low effective dimension



d	N_d^{1-4}	$N_d^{\bar{\epsilon}}/N_2^{\bar{\epsilon}}$	$\binom{d}{2}$
2	12,137	1.0	1
3	29,355	2.4	3
4	58,543	4.8	6
5	83,981	6.9	10
6	130,670	10.8	15
7	179,231	14.8	21
8	241,512	19.9	28
9	283,755	23.4	36
10	330,196	27.2	45





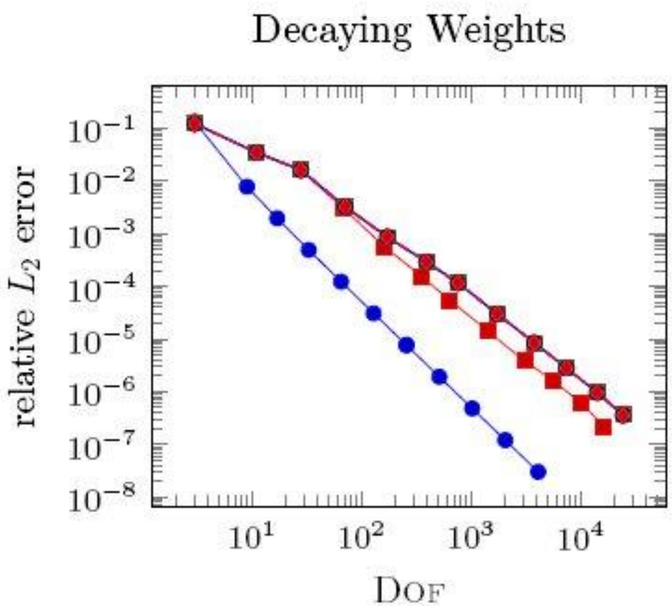
d	N_d^{1-2}	$N_d^{\bar{\epsilon}}/N_5^{\bar{\epsilon}}$	$\binom{d}{5}$
5	2,904,750	1.0	1
6	12,153,200	4.2	6
7	35,269,600	12.1	21
8	85,576,500	29.5	56
9	173,730,000	59.8	126
10	342,084,000	117.8	252

Cost increase factors for a fixed error 0.0001, right cloumn

Decay of importance of the dimensions

- Weighted model problem $$f(\mathbf{x}) = \sum_{\substack{u \subseteq \{1,\ldots,d\} \\ |u|=q}} \prod_{j \in u} w_j g(x_j)$$

$$g(x) = \frac{1}{\gamma\pi}\left(1+\left(\frac{x-x_0}{\gamma}\right)^2\right)^{-1} \text{with } \gamma = 1/2,\ \mathrm{x}_0 = 0.8,\ w_j = 2\cdot 2^{3/2(j-1)}$$



u	γ_u	u	γ_u	u	γ_u
(0)	1_{0}	(2,3)	2_{-9}	(0,1,2,4)	2_{-22}
(1)	3_{-1}	(0,2,3)	2_{-9}	(3,4)	3_{-26}
(0,1)	3_{-1}	(1,2,3)	7_{-10}	(0,3,4)	3_{-26}
(2)	8_{-3}	(0,1,2,3)	7_{-10}	(1,3,4)	9_{-27}
(0,2)	8_{-3}	(4)	1_{-19}	(0,1,3,4)	9_{-27}
(1,2)	2_{-3}	(0,4)	1_{-19}	(2,3,4)	3_{-28}
(0,1,2)	2_{-3}	(1,4)	3_{-20}	(0,2,3,4)	3_{-28}
(3)	3_{-7}	(0,1,4)	3_{-20}	(1,2,3,4)	7_{-29}
(0,3)	3_{-7}	(2,4)	8_{-22}	(0,1,2,3,4)	7_{-29}
(1,3)	9_{-8}	(0,2,4)	8_{-22}		
(0,1,3)	9_{-8}	(1,2,4)	2_{-22}		

Figure 4.18: Dimension adaptive interpolation of the weighted superposition (4.118) for dimensions $d = 1, 2, \ldots, 10$ and the associated weights for the case $d = 5$ (ordered by magnitude). See Figure 4.19 for dimension adaptive index sets.

Decay of importance of the dimensions

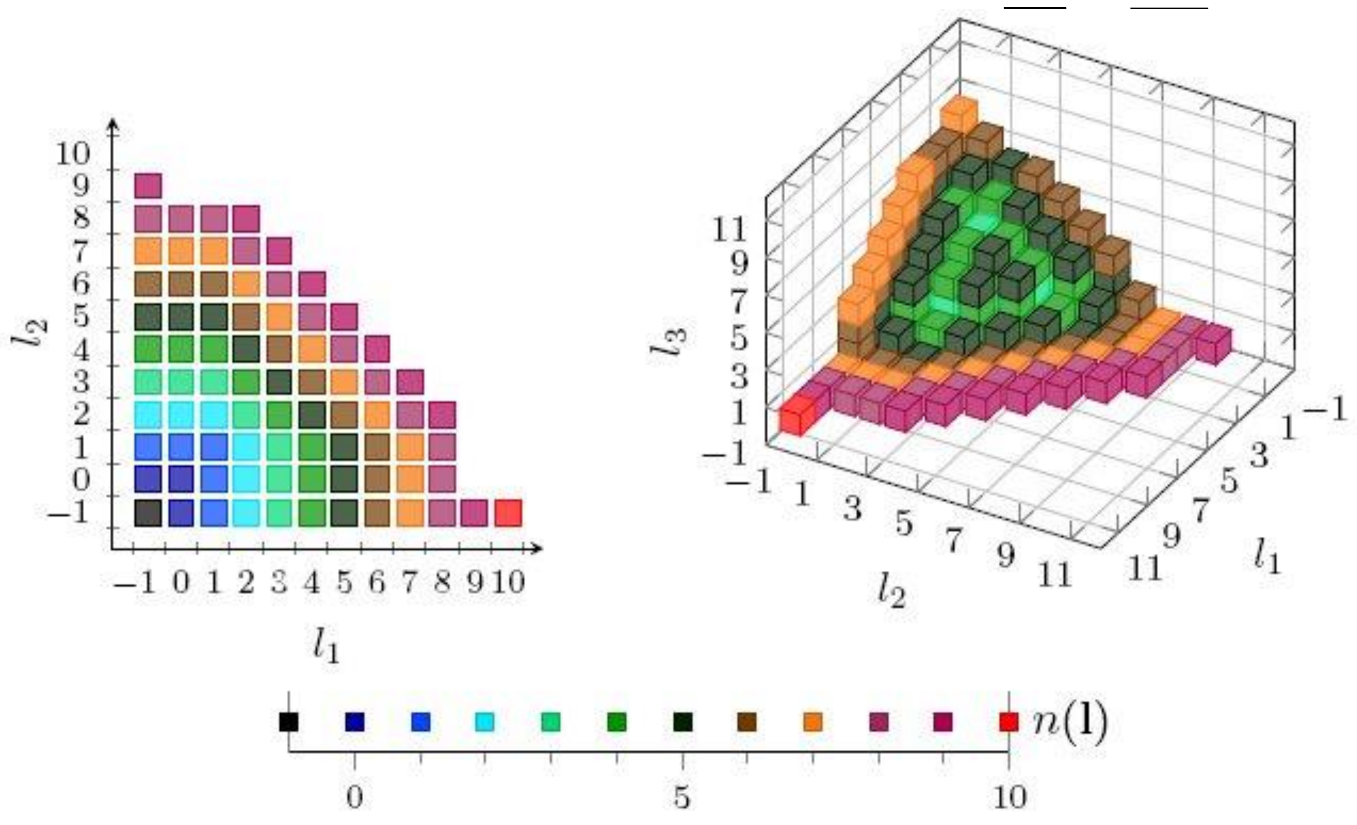


Figure 4.19: Dimension adaptive index sets for the experiment of Figure 4.18 in dimensions $d = 2$, $d = 3$. Each level $n(\mathbf{l})$ has the same color.

PDE solver

- Problem

$$-\Delta u = h \text{ in } [0,1]^d$$

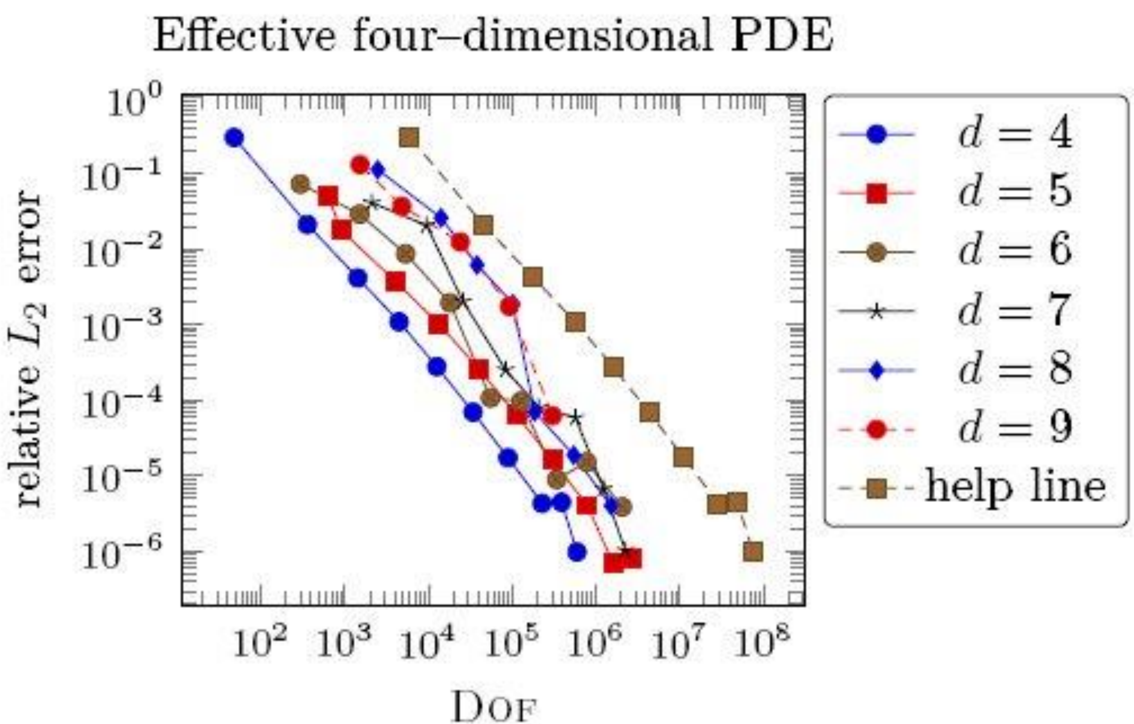
$$\frac{\partial}{\partial n} f = \text{ on } \partial\, [0,1]^d$$

$$f(0,\ldots,0) = f_0$$

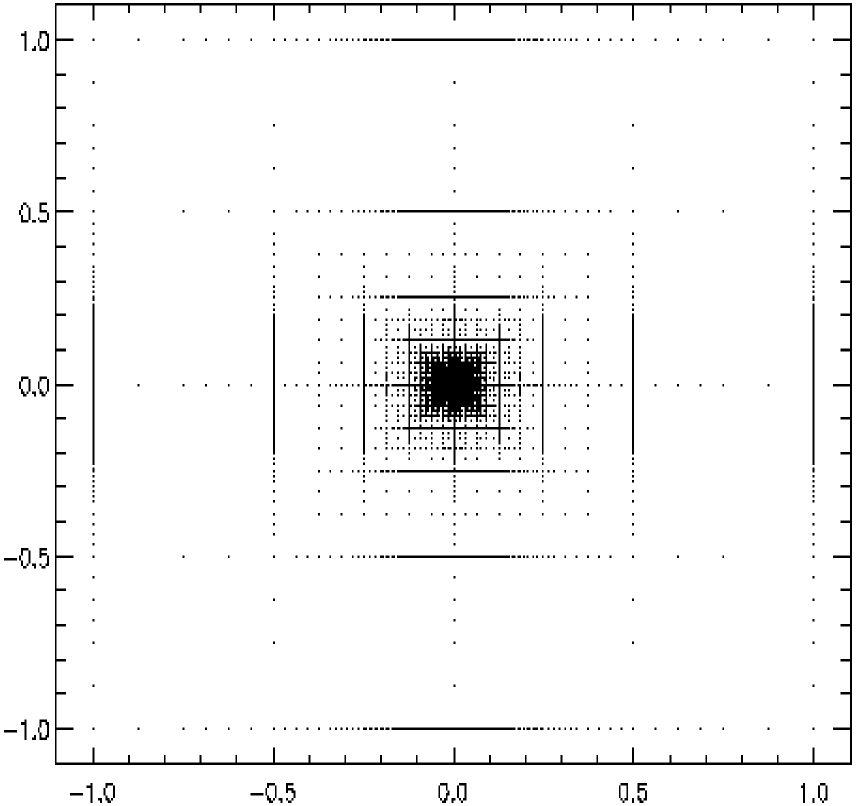
right hand side of finite order q=4

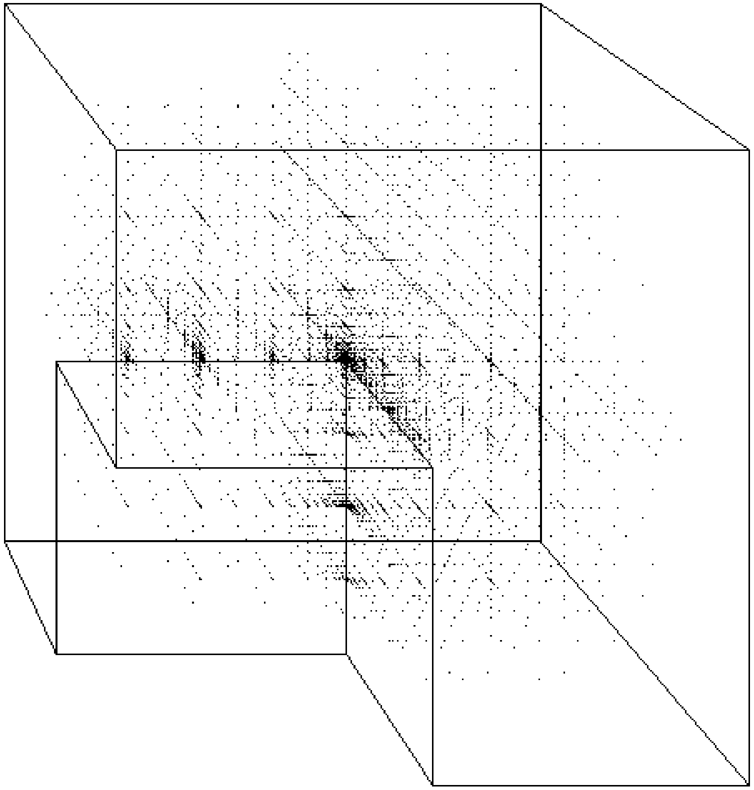
$$f(\mathbf{x}) = \sum_{\substack{u \subseteq \{1,\ldots,d\} \\ |u| = q}} \prod_{j \in u} g(x_j)$$

$$g(x) = \frac{1}{\gamma\pi}(1 + (\frac{x - x_0}{\gamma})^2)^{-1} \text{with } \gamma = 1/2,\ \mathrm{x}_0 = 0.8$$



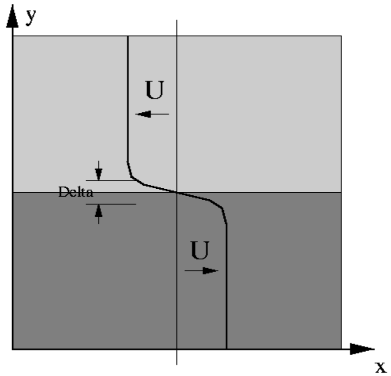
Locally adaptive sparse grids for PDEs



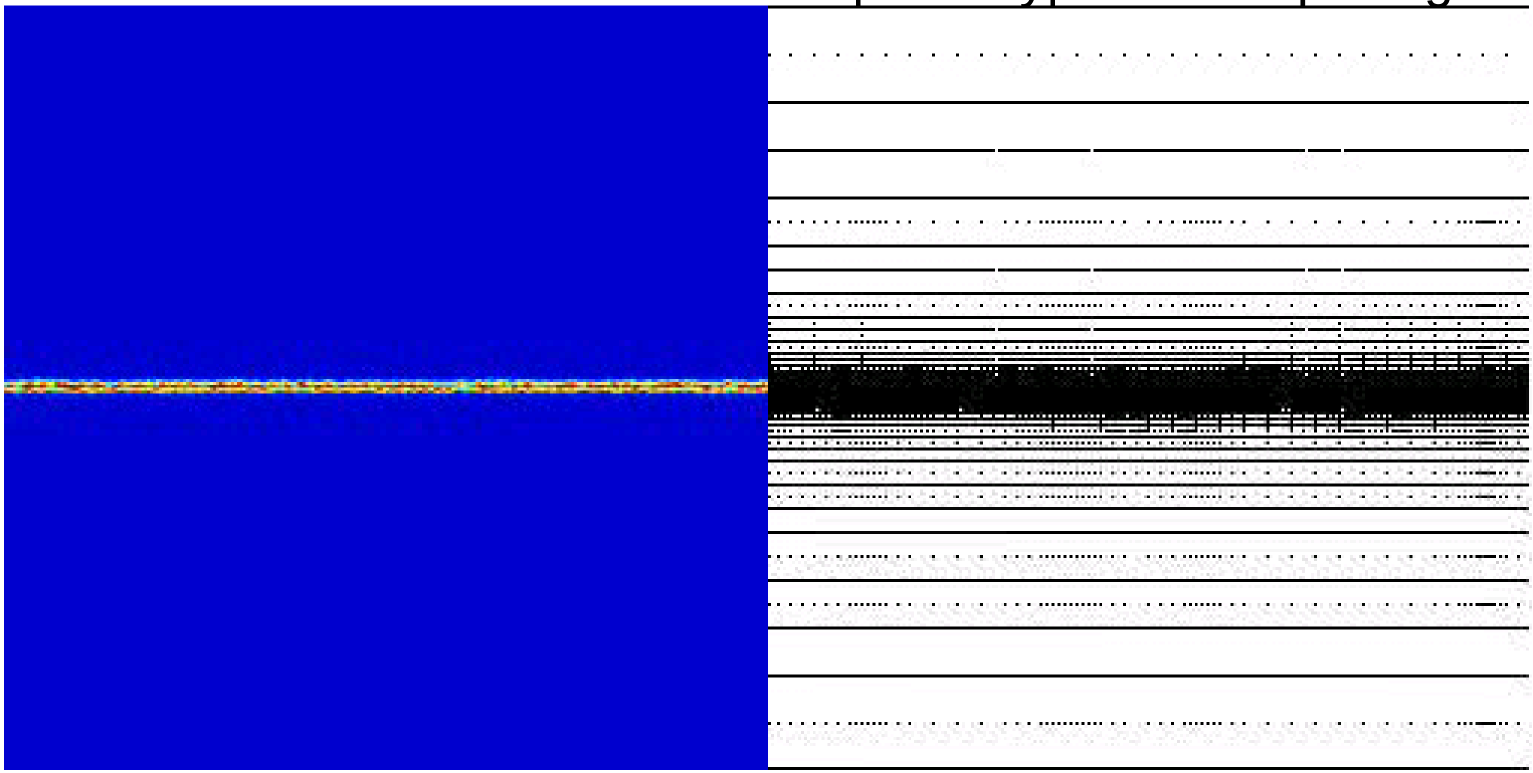


- principle: refine near points with large hierarchical coefficient
  nonlinear N-term approximation
- for Besov spaces: same rates as isotropic nonlinear refinement schemes (wavelets, adaptive finite elements) (Nitsche, Schwab)
- line/face singularities aligned with coordinate axes are cheap to resolve

2D Navier-Stokes equation

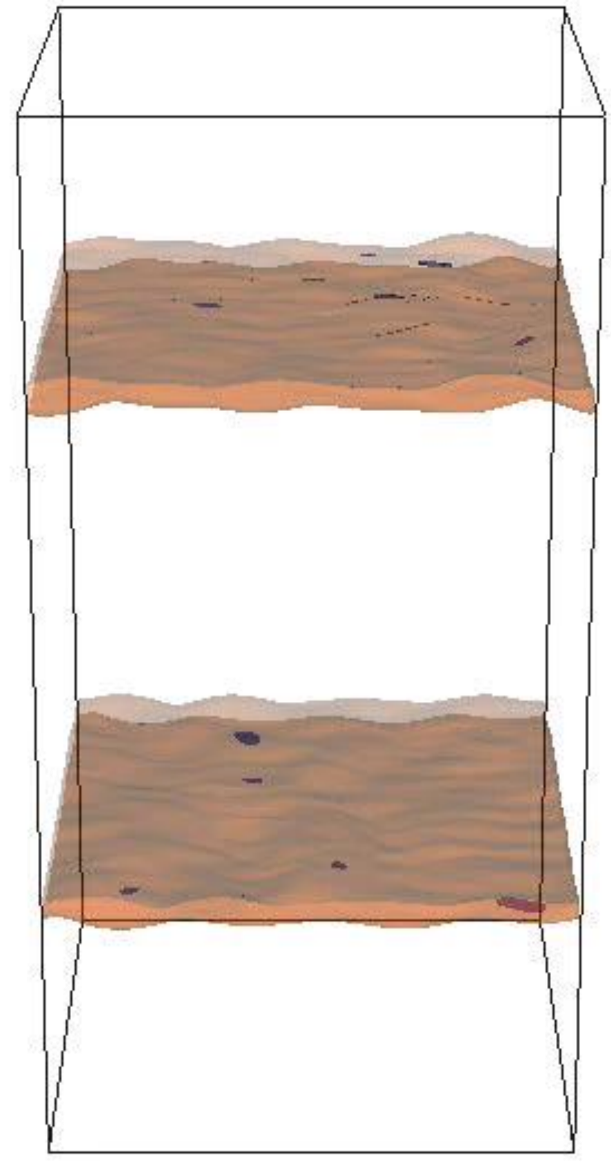


- 2D mixing layer
- Chorin projection scheme, incompressible flow
- Re=U/ν = 16000
- pertubations for initial condition
- evolution of |vorticity| and adaptive grids



3D Navier-Stokes equations

- 3D Mixing layer
- initial conditions analogous to 2D
- Re=4000
- discretization as before
- number of DOF between 1 ... 2 million
- three different isosurfaces of vorticity



Implementation for higher dimensional PDEs

- naive implementation of sparse grids for PDEs:

  work count $O(d^2 2^d \tilde{N})$

  storage $O(d\tilde{N})$

  $\tilde{N} = O(dof)$

- new data structures and multigrid algorithms, use of unidirectional principle, hash techniques

- separable, non-constant coefficient functions now:

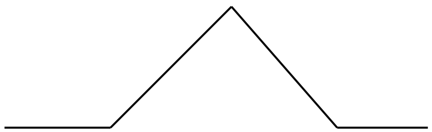
  work count $O(d^2 \tilde{N})$

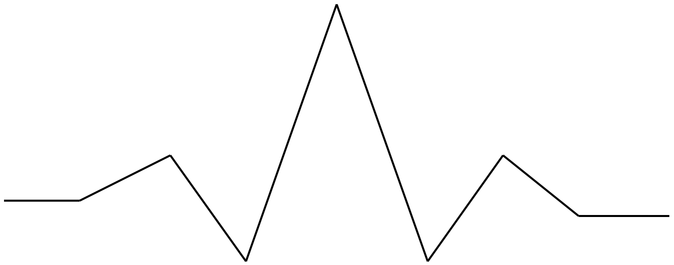
  storage $O(d\tilde{N})$

- elliptic PDEs possible with up to 120 dimensions with homogeneous bc and product-type right hand side (Feuersänger).

Implementation for higher dimensional PDEs

- implementation uses semi-orthogonal prewavelets instead of piecewise linear hat functions





- orthogonality between levels simplifies mass matrix contributions and results in improved complexity w.r.t dimension d
- Full orthogonal wavelets would reduce complexity to just $O(d\tilde{N})$ but are more difficult to work with

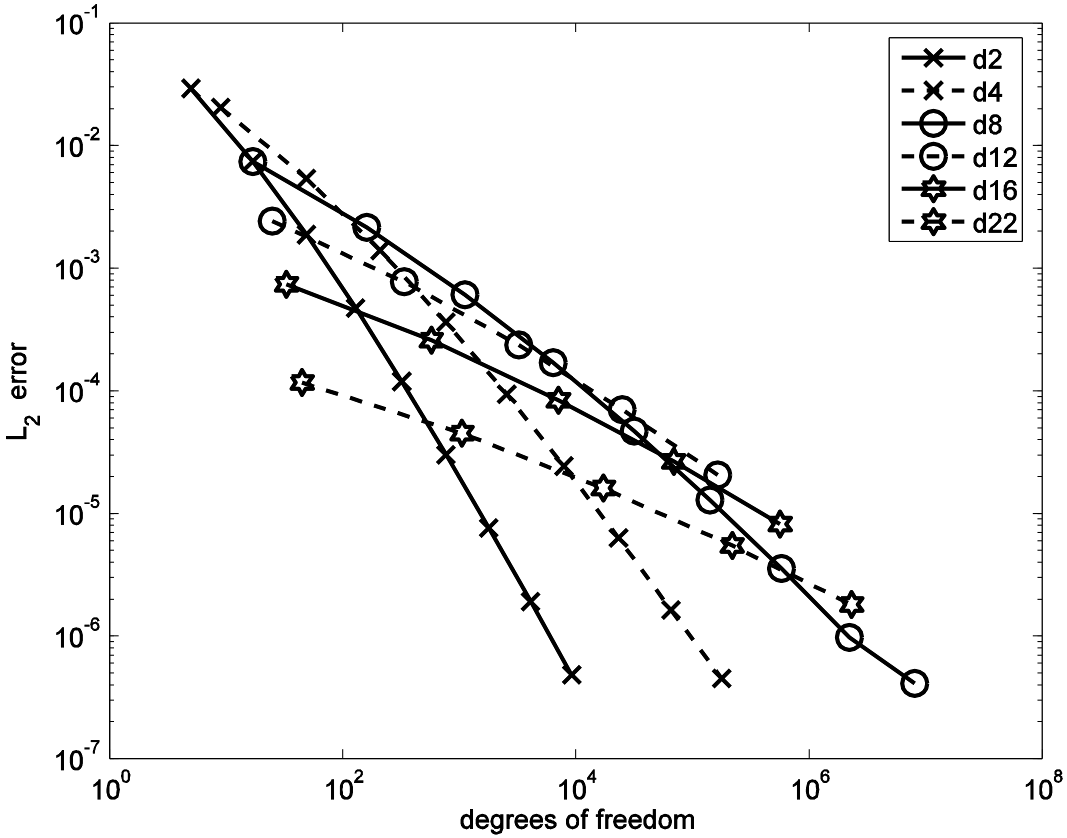
Example: 2nd order PDE

$$-\Delta f(x) + \sum_{k=1}^{d} b_k(x)\partial^k f(x) + c(x) f(x) = r(x), \;\; x \in \Omega = [0,1]^d$$

$$f(x) = 0, \;\; x \in \partial\Omega$$

Here: regular L^2 norm based sparse grid

L^2 convergene rates with up to $d = 22$



We see the influence of the $\log(N)^{(d-1)}$ -terms

an energy norm approach and adaptivity is necessary

Caveat

- the regularity term $| f |_{2,mix}$ might cause problems and can postpone the onset of convergence
- Example 1:

$$f(x_1 \ldots x_d) = \prod_{j=1}^{d} \sin(2\pi \; k_j x_j)$$

$$D^{(2,..2)} f(x_1 \ldots x_d) = (-1)^d \prod_{j=1}^{d} (2\pi k_j)^2 \prod_{j=1}^{d} \sin(2\pi \; k_j x_j)$$

$$| f |_{2,mix} = (2\pi)^{2d} \prod_{j=1}^{d} k_j^2$$

- Thus at most 15-18 dimensions treatable in practice

Gaussian

- Example: $f(x_1 \ldots x_d) = \frac{1}{(2\pi\sigma)^{d/2}} \exp\left(-x^T x / \sigma\right)$

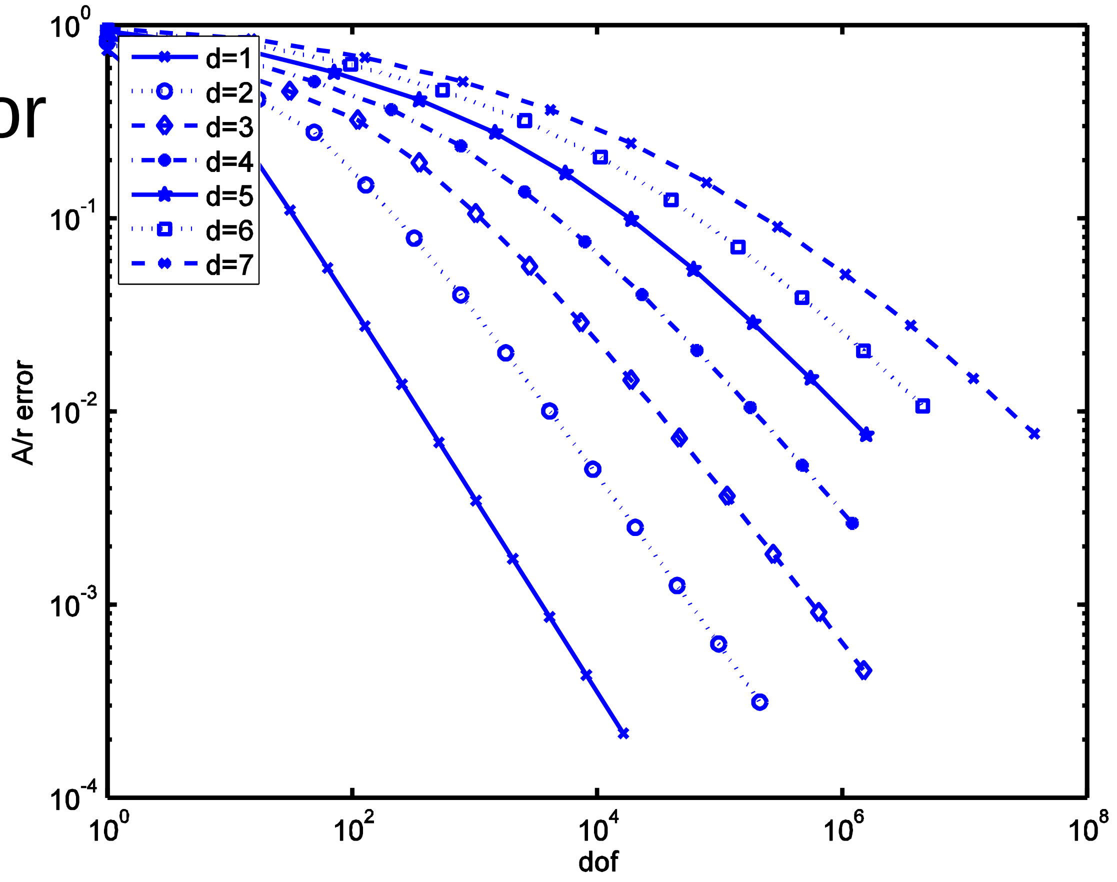
- Energy-norm based sparse grid in $\Omega = [-5\sigma,\ 5\sigma]^d$

- Rate for relative error in energy norm is asymptotically

$$\approx 2^{d/2} 2^{-n} = 2^{-(n-d/2)}$$

- Necessary:

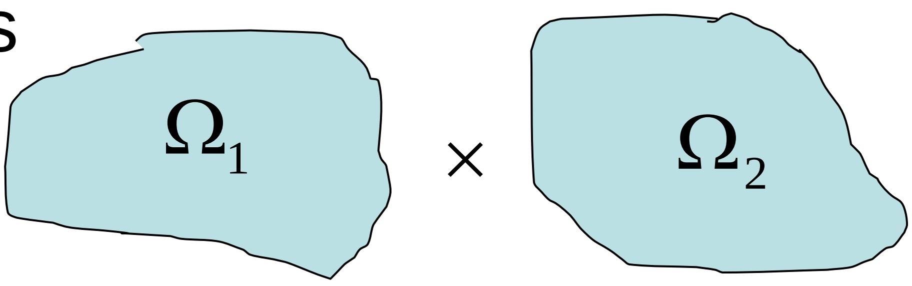
$$n - d/2 > 0$$



Tensor product sparse grids

- So far: one-dimensional domain, multiscale basis, d-fold tensor product, proper truncation
- Now: E.g. two general domains Ω_1, Ω_2, each with
  - its dimension d_1, d_2 and its smoothness s_1, s_2
  - its isotropic multilevel basis (one level index)
  - tensor product between the two domains and multiscale bases

$$\Omega = \Omega_1 \times \Omega_2$$



  - Mixed regularity $H^{s_1 s_2}_{mix}(\Omega) := H^{s_1}(\Omega_1) \times H^{s_2}(\Omega_2)$

Tensor product sparse grids

- Examples:
  - space $\times$ time, $d_1 = 3, d_2 = 1,$ parabolic problems
  - space $\times$ angle $d_1 = 3, d_2 = 2,$ radiosity
  - space $\times$ parameters $d_1 = 3, d_2 = 10 - 20$
    but smooth in parameter variables
  - space $\times$ stochastics $d_1 = 3, d_2 = \infty$
    but analytic in stochastic variables
- Main result: curse of dimension only w.r.t. the larger dimension and/or the lower smoothness
- Time comes for free, angle space comes for free, parametrization/stochastics comes for free, just space dimension matters

Optimized general sparse grid space

- Multiscale analyses on $\Omega_i, i = 1,2$ with associated approximation order r_i

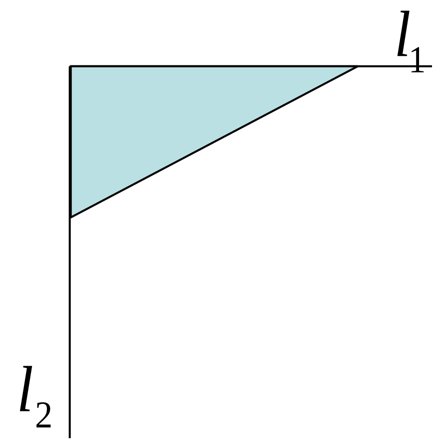
$$V_0^{(i)} \subset V_1^{(i)} \subset V_2^{(i)} \subset \ldots \subset L^2(\Omega_i)$$

- Complementary spaces:

$$W_{l_i}^{(i)} \qquad\qquad V_{l_i}^{(i)} = W_{l_i}^{(i)} \oplus V_{l_i-1}^{(i)}$$

- Anisotropic sparse grid space:

$$\widehat{V}_n^{\sigma} = \bigoplus_{l_1\sigma + l_2/\sigma \le n} W_{l_1}^{(1)} \oplus W_{l_2}^{(2)}$$



- Parameter σ now allows to optimize with respect to dimensions d_1, d_2 and smoothness s_1, s_2

Properties (G.+Harbrecht 2011)

- The sparse grid space $\widehat{V}_J^{\sigma}$ possesses

$$\widehat{V}_n^{\sigma} \sim \begin{cases} 2^{n\max\{d_1/\sigma, n_2\sigma\}} & \text{if } d_1/\sigma \neq d_2\sigma \\ n \cdot 2^{nd_2\sigma} & \text{if } d_1/\sigma = d_2\sigma \end{cases}$$

degrees of freedom.

- For a given $f \in H_{mix}^{s_1,s_2}(\Omega_1 \times \Omega_2)$ with $0 < s_1 \leq r_1,\ 0 < s_2 \leq r_2$ we have for the accuracy

$$\inf_{\widehat{f}_n \in \widehat{V}_n^{\sigma}} \| f - \widehat{f}_n \|_{L^2(\Omega_1 \times \Omega_2)} \leq \begin{cases} 2^{-n\min\{s_1/\sigma, s_2\sigma\}} \| f \|_{H_{mix}^{s_1,s_2}(\Omega_1 \times \Omega_2)} & \text{if } s_1/\sigma \neq s_2\sigma \\ \sqrt{n} \cdot 2^{-ns_1/\sigma} \| f \|_{H_{mix}^{s_1,s_2}(\Omega_1 \times \Omega_2)} & \text{if } s_1/\sigma = s_2\sigma \end{cases}$$

- No log-terms in many situations
- Analogous results by simple shift for other error norms like $H_{mix}^{q_1,q_2}(\Omega_1 \times \Omega_2)$ than just for $L^2(\Omega_1 \times \Omega_2)$

Space-time sparse grids

- Approximation error and necessary regularity

$$\inf_{u_n \in V_n^0} \|u-u_n\|_{H^1(\Omega) \otimes L^2(0,T)} \le c\, 2^{-n} \|u\|_{H^2(\Omega) \otimes H^2((0,T))}$$

- How realistic are these regularity assumptions?
  - $\partial_{\vec{x}}^2 \partial_t^2 u$ is also needed for error estimates of conventional discretization methods
  - classical regularity theory shows (Ladyzenskaja, Wloka)

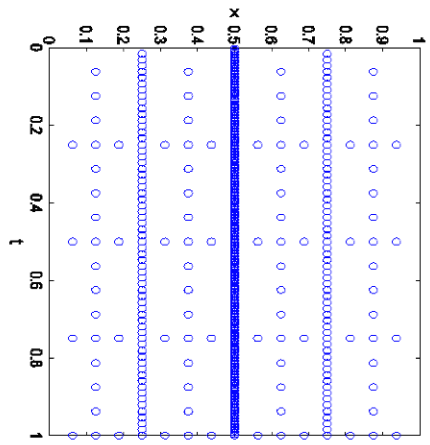
$$u \in H^2(\Omega) \otimes H^2((0,T))$$

- Space-time sparse grids possess the same approximation rate as conventional full space-time grids but only the cost complexity of space problem

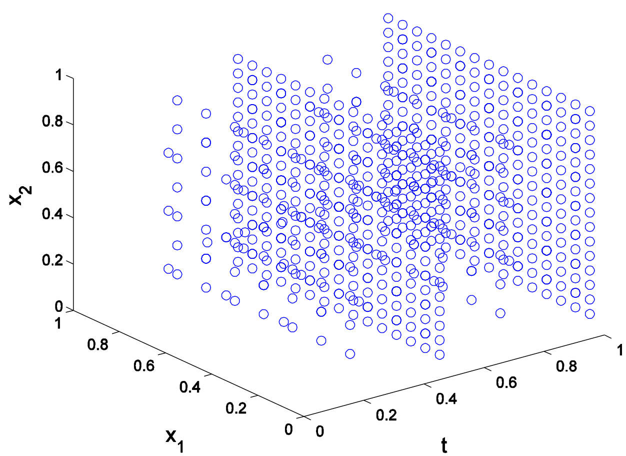
=> time coordinate comes for free

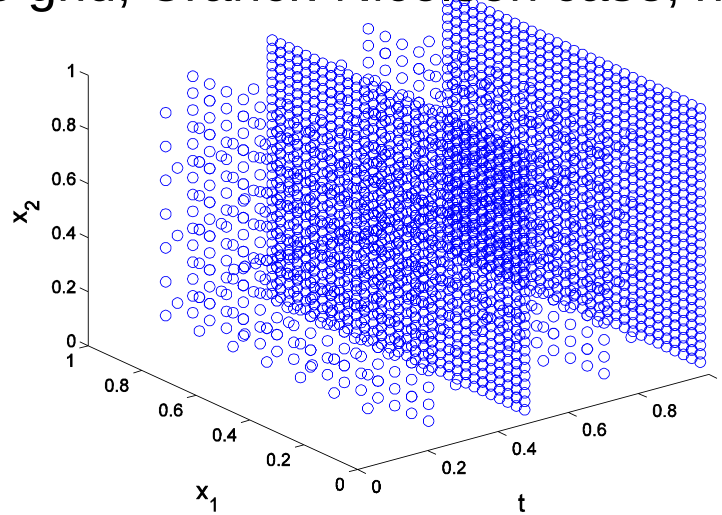
Examples of space time sparse grids

space dimension 1, space-time sparse grid, Euler case



space dimension 2, space-time sparse grid, Cranck-Nicolson case, n=4,5:

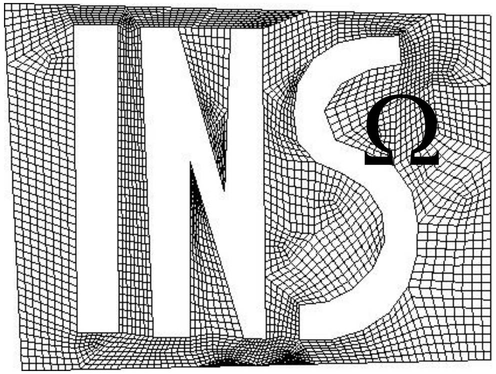




in each time slice there is a conventional full grid

Instationary distributed control problems

space dimension 2, adaptive space time grids



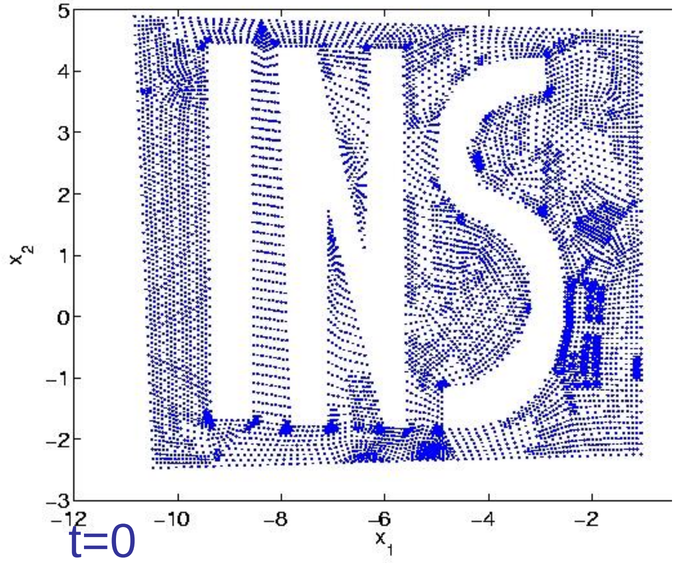
Problem:

$$\partial_t y - \Delta y - 10^3 p = 0 \text{ in } \Omega \times (0,1],$$

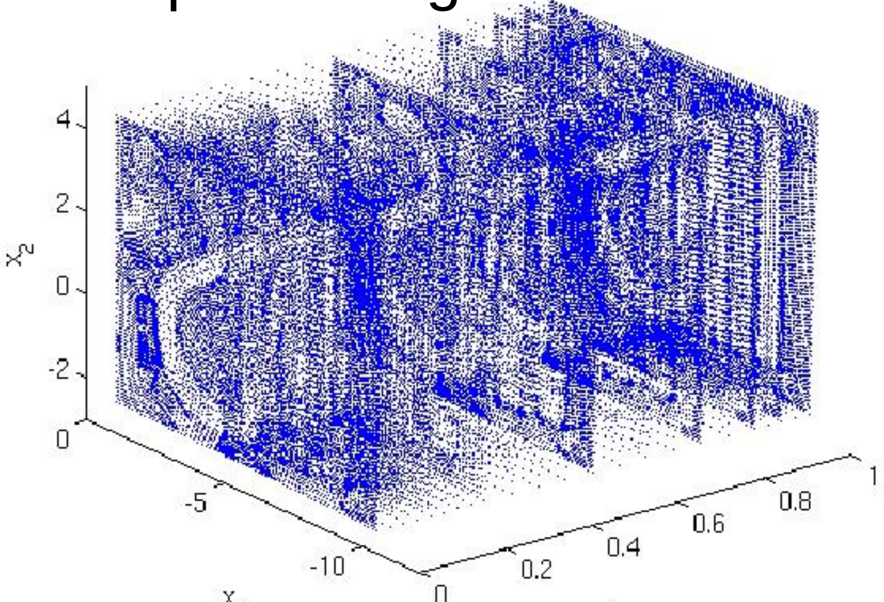
$$-\partial_t p - \Delta p + y = 1 \text{ in } \Omega \times [0,1),$$

with homogeneous initial/end and boundary conditions

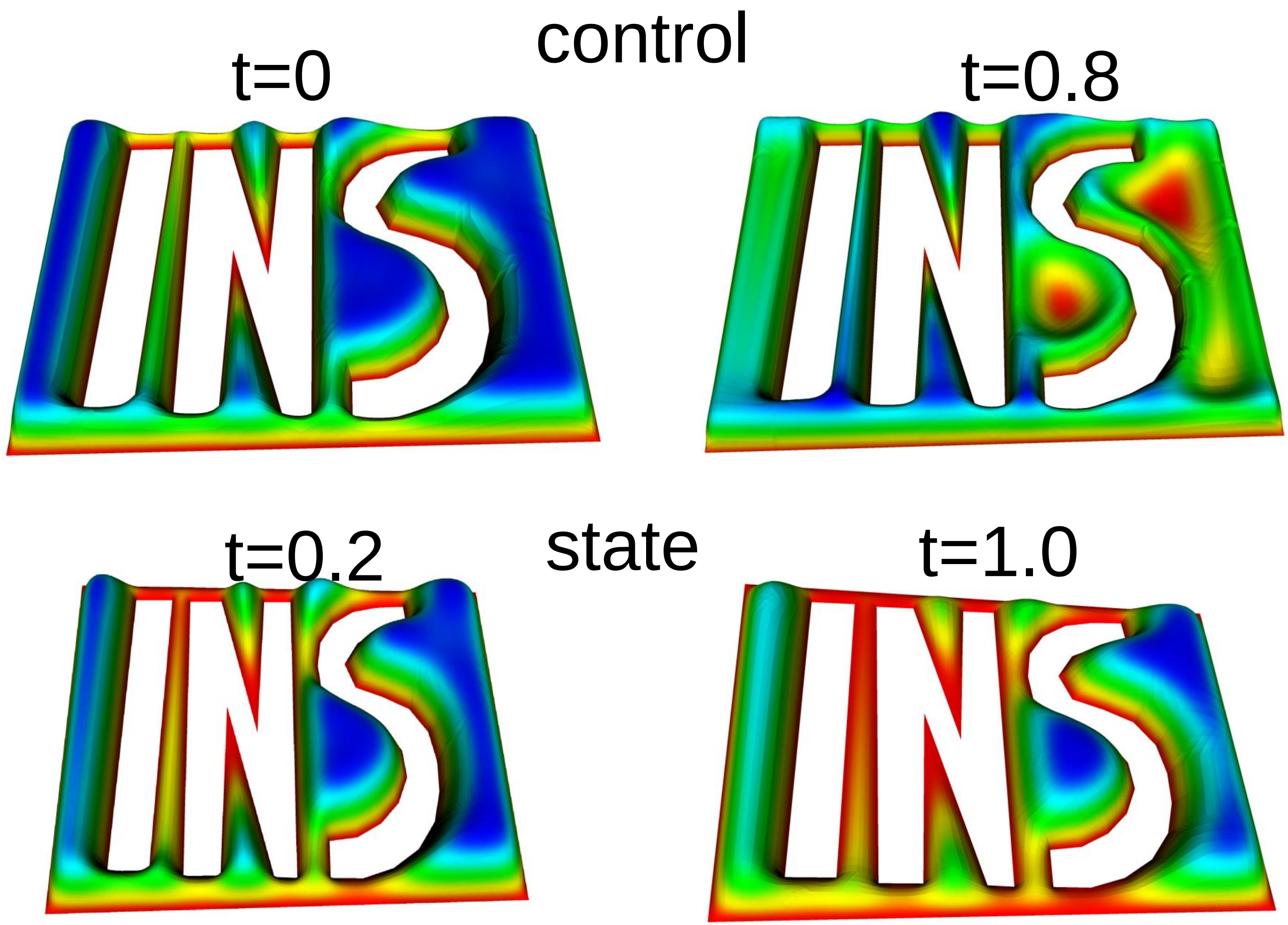
Adaptivity with 5 refinement steps starting at level 3



t=0



Instationary distributed control problems



Instationary distributed control problems

space dimension=3, adaptive space time grids

Problem: $\partial_t y - \Delta y - 10^3 p = 0 \text{ in } \Omega \times (0,1],$ $\quad \Omega = (-1,1)^3 \setminus (-1,0)^3$

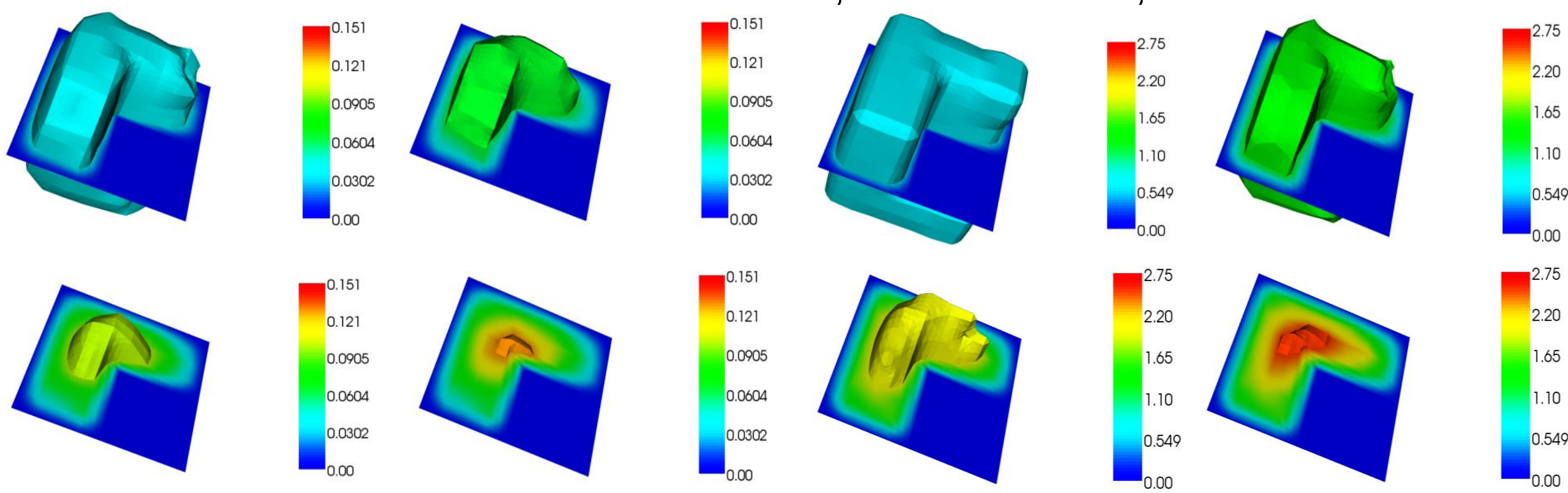
$$-\partial_t p - \Delta p + y = 1 \text{ in } \Omega \times [0,1),$$

with homogeneous initial/end and boundary conditions

Adaptivity with 4 refinement steps starting at level 3

t=0, state variable, four isosurfaces

t=1, control variable, four isosurfaces



Stochastic and parametric PDEs

- Solutions $f(\mathbf{x}_1, \mathbf{x}_2)$ of stochastic/parametric PDEs

$$-\nabla \cdot A(\mathbf{x}_2)\nabla f(\mathbf{x}_1, \mathbf{x}_2) = r(\mathbf{x}_1, \mathbf{x}_2)$$

  live on product $(\mathbf{x}_1, \mathbf{x}_2) \in \Omega_1 \times \Omega_2$
  - of spatial domain Ω_1 with $d_1 = 1,2,3$
  - and stochastic/parametric domain Ω_2 with d_2 large or even infinity.
- Often: Very high smoothness in $\mathbf{x}_2$ -part

  Here: especially weighted analyticity for the different coordinates due to decay in covariance
- Therefore, even infinite-dimensional Ω_2 become treatable independently of d_2

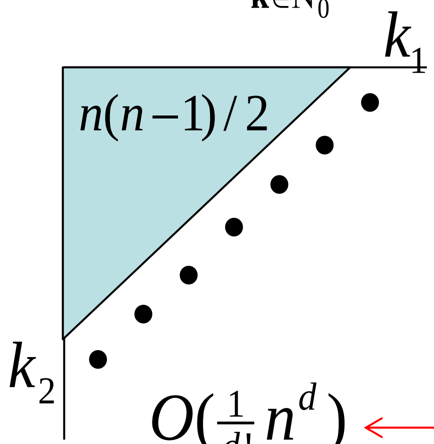
Stochastic and parametric PDEs

- Sparse grids methods can be used for Ω_2 to cope with its high dimensionality !
  - The stochastic part is smooth or even analytic
    - sparse grids with spectral, polynomial bases
  - The stochastic coordinates are not equally important
    - weights/decay of the different coordinate directions related to the eigenvalues of covariance of parameters, algebraic or exponential!
    - Then, anisotropic sparse grids (with spectral, polynomial bases) and dimension-adaptive sparse grids are successfully used
    - The decay kills the curse, the sparse grid approach then reduces the dimension-dependency of the constant
- Moreover: The two-variate sparse grid product approach works fine between the spatial domain and the parametric/ stochastic domain.

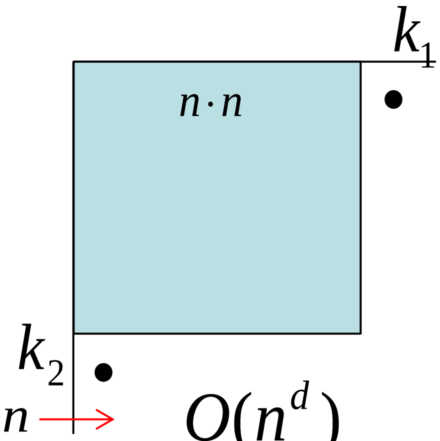
Sparse grids and analytic functions

- What may happen if the function is too smooth ?
  - 1-D: L_2-orthogonal polynomial basis $\{\phi_k(x_i)\}_{k\in\mathbb{N}0}$
  - d-D: Product of polynomials $\phi_{\mathbf{k}}(\mathbf{x}) = \phi_{k_1}(x_1)\cdots\phi_{k_d}(x_2)$
- Here: the isotropic case:
  - Representation $f(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbb{N}_0^d} f_{\mathbf{k}} \cdot \phi_{\mathbf{k}}(\mathbf{x})$ $\quad |f_{\mathbf{k}}|^2 \le c \cdot 2^{-2(k_1+\ldots+k_d)}$

regular sparse grid



full grid



  - cost: $O(\frac{1}{d!}n^d)$ ← same order in n → $O(n^d)$
  - accuracy[2] : $O(n^d \cdot 2^{-2n})$ ← additional log term $O(d\cdot 2^{-2n})$
- Anisotropic case: no curse, but still d-dependent constants

Stochastic and parametric PDEs

- Weighted analytic approximation space for Ω_2

  Let be given an ordered real sequence $\mathbf{a} = (a_i)_{i \in \mathrm{N}}$ with

  $1 = a_1 \geq a_2 \geq a_3 \geq \ldots.$ and a fixed base $b > 1$

$$A_{d_2,\mathbf{a}}(\Omega_2) = \left\{ f \in L^2(\Omega_2) : \sum_{\mathbf{k} \in \mathrm{N}_0^{d_2}} b^{2\sum_{i=1}^{d_2} k_i / a_i} |f_{\mathbf{k}}|^2 \leq C < \infty \right\}$$

- Characterization of $\mathbf{a}$ -weighted analytic functions

$$f(\mathbf{x}) = \sum_{\mathbf{k} \in \mathrm{N}_0^{d_2}} f_{\mathbf{k}} \cdot \phi_{\mathbf{k}}(\mathbf{x}) \qquad |f_{\mathbf{k}}| \leq c \cdot b^{-\sum_{i=1}^{d_2} k_i / a_i}$$

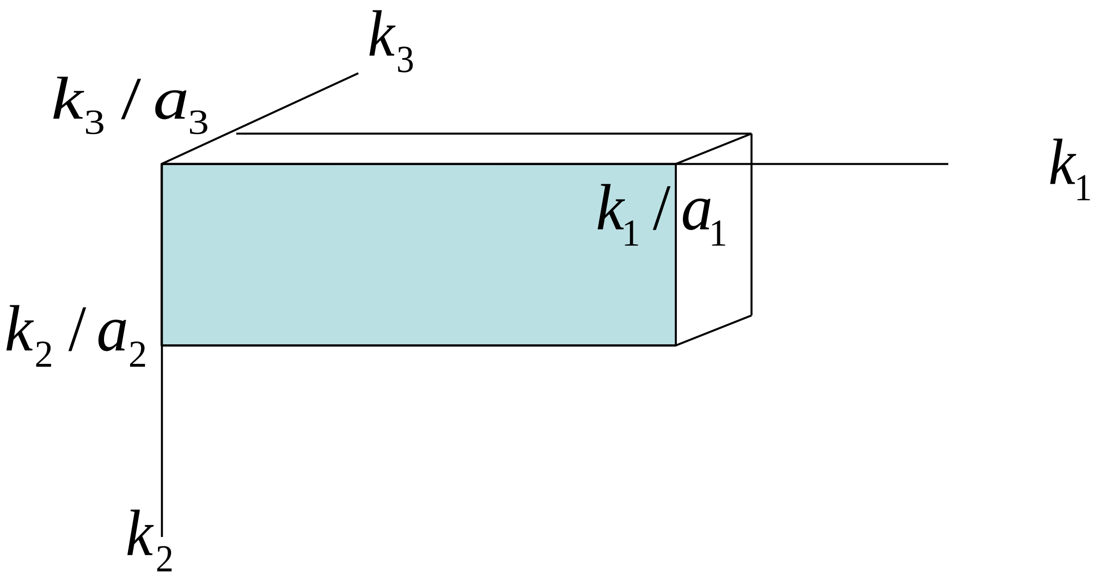
Discrete anisotropic, regular grid subspace

- Index set, brick-type with successively smaller size

$$I_{d_2,\mathbf{a}}(n) := \{\mathbf{k} \in \mathrm{N}_0^{d_2} : k_i / a_i \le n \text{ for all } 1 \le i \le d_2\}$$

- Corresponding subspace

$$V_{d_2,\mathbf{a},n} := \text{span}\,\{\phi_{\mathbf{k}}(\mathbf{x}),\ \mathbf{k} \in I_{d_2,\mathbf{a}}(n)\}$$



Discrete regular grid subspace

- Degrees of freedom

$$\dim(V_{d_2,\mathbf{a},n}) = \prod_{i=1}^{d_2}(1+\lfloor n\cdot a_i \rfloor) \le \exp(n\cdot\sum_i a_i)$$

- With the summability condition $\sum_i a_i \le A < \infty$

  we get, independently of d_2,

$$\dim(V_{d_2,\mathbf{a},n}) \le \exp(n\cdot A)$$

- Accuracy: about linear in n, mainly independent of d_2

$$\| f - f_{V_{d_2,\mathbf{a},n}} \|_{L^2(\Omega_2)} \lesssim \begin{cases} \sqrt{d_2}\cdot b^{-n} & \text{in any case} \\ b^{-\mu(n)} & \text{for} \quad \mu(n) := \min_r\{\lfloor n\cdot a_r \rfloor = 0\} \\ b^{-(1-\varepsilon)\cdot n} & \text{if} \quad 1/a_r - 1/a_s \ge (r-s)\delta \end{cases}$$

Stochastic and parametric PDEs

- With the sequence

$$V_{d_2,\mathbf{a},0} \subset V_{d_2,\mathbf{a},1} \subset V_{d_2,\mathbf{a},2} \subset \ldots \subset \mathrm{L}^2(\Omega_2)$$

and associated sequence of complimentary spaces

$$W_{d_2,\mathbf{a},j} \qquad V_{d_2,\mathbf{a}.j} = W_{d_2\mathbf{a},j} \oplus V_{d_2,\mathbf{a},j-1}$$

we get, together with the usual sequence of complementary spaces on Ω_1, a two-variate sparse grid construction on $\Omega_1 \times \Omega_2$, which is independent of the dimension d_2 (even if $d_2 = \infty$).

- The sparse grid product approach works fine between the spatial and the parametric/ stochastic domains.

Summary

Classical approach: $d = 1, .., 3 \text{ or } 4$

curse of dimension and intractability

$$\| f - f_N \|_{H^s} = c(d) \cdot N^{-r/d} \mid f \mid_{H^{s+r}} = O(N^{-r/d})$$

Stronger regularity/norms

curse only wrt log-terms

$$\| f - f_N \|_{H^s} = c(d) \cdot N^{-r} (\log(N))^{(d-1)/2} \mid f \mid_{H^{s+r}_{mix}}$$

or no curse at all

$$\| f - f_N \|_{H^s} = c(d) \cdot N^{-r} \mid f \mid_{H^{s+r}_{mix}}$$

but still not tractable,

constant grows exponentially

$$d = 1, .., 10 \text{ to } 12$$

Lower effective dimension and lower-dim. manifolds

no curse due to effective dimension

$$\| f - f_N \|_{H^s} = c(d^{eff}) \cdot N^{-r/d^{eff}} \mid f \mid_{H^{s+r}}$$

and constant grows exponentially only w.r.t. effective dimension

$$d = 1, ..., 100 \qquad d^{eff} = 1, .., 10$$

Literature

- M. Griebel, Sparse tensor product spaces for stochastic and parametric approximations, INS Report, in preparation, to appear 2013.
- M. Griebel, H. Harbrecht, On the construction of sparse tensor product spaces, Math. Comp., 2012.
- M. Griebel and H. Harbrecht. A note on the construction of L-fold sparse tensor product spaces. Constructive Approximation, 2012.
- M. Griebel, H. Harbrecht, Approximation of two-variate functions: singular value decomposition versus sparse grids, IMA Journal of Numerical Analysis, 2012.
- M. Griebel, S. Knapek, Optimized general sparse grid approximation spaces for operator equations, Math. Comp. 78, 2223-2257, 2009.
- M. Griebel and D. Oeltz. A sparse grid space-time discretization scheme for parabolic problems. *Computing*, 81(1):1-34, 2007.
- D. Oeltz. *Ein Raum-Zeit Dünngitterverfahren zur Diskretisierung parabolischer Differentialgleichungen*. Dissertation, INS, Universität Bonn, 2006.
- M. Griebel. Sparse grids and related approximation schemes for higher dimensional problems. In L. Pardo, A. Pinkus, E. Suli, and M. Todd, editors, *Foundations of Computational Mathematics (FoCM05), Santander*, pages 106-161. Cambridge University Press, 2006.
- H.-J. Bungartz and M. Griebel. Sparse grids. *Acta Numerica*, 13:1-123, 2004.