Numerical Methods I

MATH-GA 2010.001/CSCI-GA 2420.001

Benjamin Peherstorfer
Courant Institute, NYU

Today

Last time

- Numerical integration
- Newton-Cotes and Gauss quadrature

Today

- Function approximation in higher dimensions
- Quadrature in higher dimensions

Announcements

- Homework 7 is due Mon, Dec 12 before class (one week!)

There are many opportunities for asking questions

Mon, Dec 5	office hour
Thu, Dec 8	office hour
Mon, Dec 12	recap of important topics of class
Mon, Dec 12	office hour
Thu, Dec 15	office hour (all HWs graded by Dec 15; no HW re-grading after Dec 17)
Mon, Dec 19	final exam

Final exam

- In-person
- Pen & paper
- Closed book
- There can be coding questions
- There can be multiple choice/short answer questions (see HW7)
- Quietly leave the room when you are done! Due to a conflict, several students will start late and then work longer than 7.00pm.

Sparse grids

Curse of dimensionality

- The curse of dimensionality is a term coined by Bellmann (1961) that refers to an exponential increase of costs with the dimension of a problem

- For example, consider an approximation with a prescribed accuracy $\epsilon > 0$, let the costs of achieving this approximation scale as $\mathcal{O}(\epsilon^{-d})$ in d dimensions $\rightsquigarrow$ exponential increase of the costs as we increase d

- Consider a simple uniform grid over the domain $\Omega = [0, 1]^d$. To have a mesh with mesh width $h = 1/9$ in $d = 1$, we need $N = 10$ grid points. In $d = 2$ dimensions, need $N^2 = 100$ grid points. In $d = 5$ dimensions, need $N^5 = 10^5$ $\rightsquigarrow$ exponential growth of cost and storage requirements as we increase dimension d while keeping mesh width h ("accuracy") fixed

Curse of dimensionality

- The curse of dimensionality is a term coined by Bellmann (1961) that refers to an exponential increase of costs with the dimension of a problem
- For example, consider an approximation with a prescribed accuracy $\epsilon > 0$, let the costs of achieving this approximation scale as $\mathcal{O}(\epsilon^{-d})$ in d dimensions $\rightsquigarrow$ exponential increase of the costs as we increase d
- Consider a simple uniform grid over the domain $\Omega = [0,1]^d$. To have a mesh with mesh width $h = 1/9$ in $d = 1$, we need $N = 10$ grid points. In $d = 2$ dimensions, need $N^2 = 100$ grid points. In $d = 5$ dimensions, need $N^5 = 10^5$ $\rightsquigarrow$ exponential growth of cost and storage requirements as we increase dimension d while keeping mesh width h ("accuracy") fixed
- The curse can be circumvented (to some extent) if?

Curse of dimensionality

- The curse of dimensionality is a term coined by Bellmann (1961) that refers to an exponential increase of costs with the dimension of a problem

- For example, consider an approximation with a prescribed accuracy $\epsilon > 0$, let the costs of achieving this approximation scale as $\mathcal{O}(\epsilon^{-d})$ in d dimensions $\rightsquigarrow$ exponential increase of the costs as we increase d

- Consider a simple uniform grid over the domain $\Omega = [0, 1]^d$. To have a mesh with mesh width $h = 1/9$ in $d = 1$, we need $N = 10$ grid points. In $d = 2$ dimensions, need $N^2 = 100$ grid points. In $d = 5$ dimensions, need $N^5 = 10^5$ $\rightsquigarrow$ exponential growth of cost and storage requirements as we increase dimension d while keeping mesh width h ("accuracy") fixed

- The curse can be circumvented (to some extent) if? if we make stronger assumptions on the functions to approximate $\rightsquigarrow$ topic of today

Sparse grids*

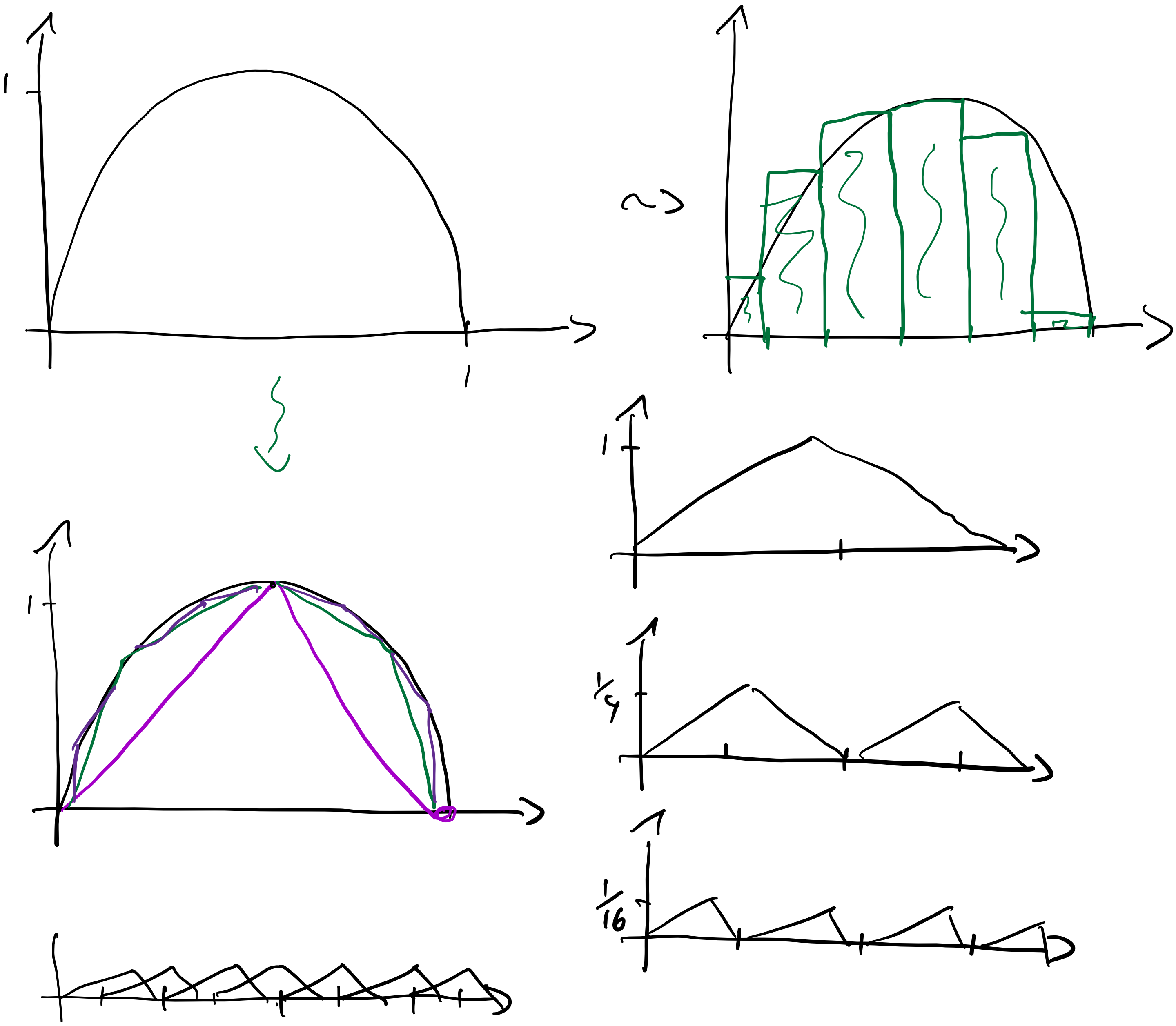
*Follows lecture by H.-J. Bungartz. See also, Bungartz, Griebel, *Sparse grids*, Acta Numerica, 2004

SG: Motivating example

Approximate the integral

$$\int_0^1 4x(1-x)\mathrm{d}x = \frac{2}{3}$$

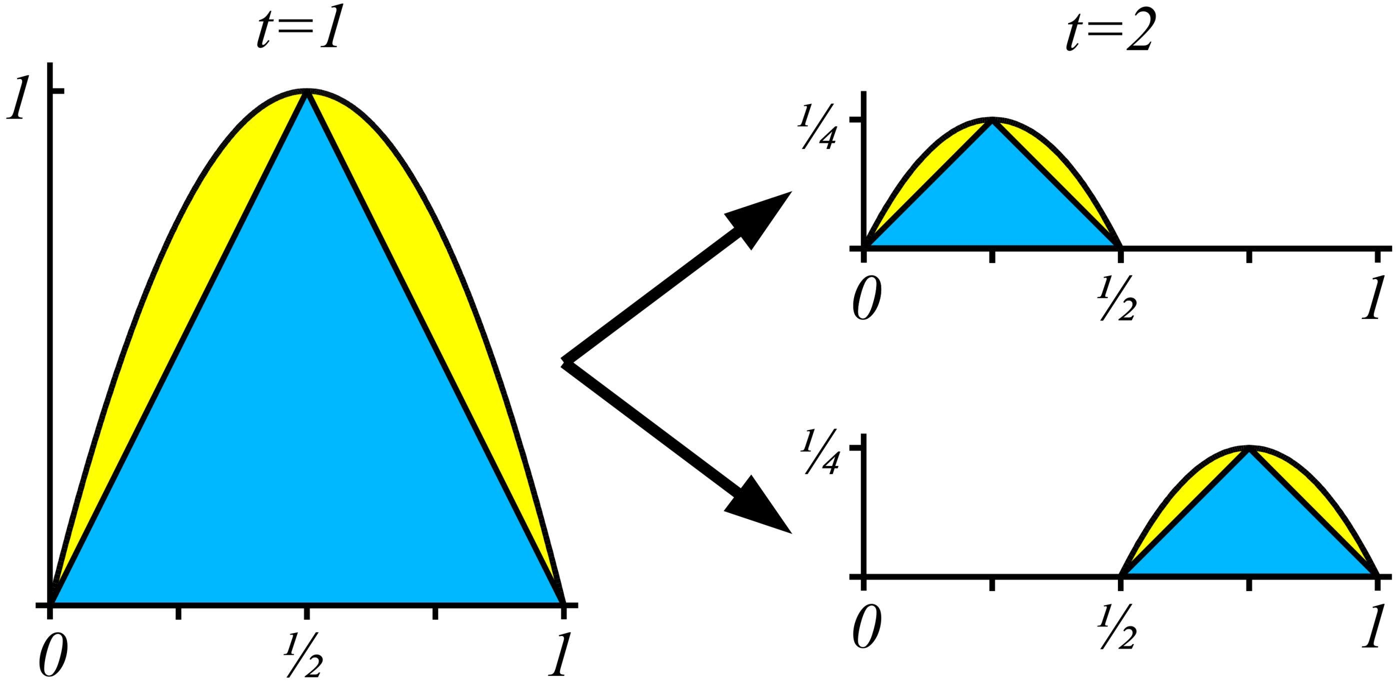
Board!


1
1
1
1
1
1/4
1/16

SG: Hierarchical decomposition 1D

Archimedes Quadrature

$$\int_0^1 4x(1-x)\mathrm{d}x = \frac{2}{3}$$



depth	1	2	3	4	...	t
interval length h	1/2	1/4	1/8	1/16	...	2^{-t}
number of triangles	1	2	4	8	...	$\frac{1}{2}2^{t}$
surplus v	1	1/4	1/16	1/64	...	$4 \cdot 2^{-2t}$
triangle area D_1	1/2	1/16	1/128	1/1024	...	$4 \cdot 2^{-3t}$
sum (of this t)	1/2	1/8	1/32	1/128	...	$2 \cdot 2^{-2t}$
overall sum ($\leq t$)	1/2	5/8	21/32	85/128	...	$\frac{2}{3}\left(1 - 2^{-2t}\right)$
error	1/6	1/24	1/96	1/384	...	$\frac{2}{3}2^{-2t}$

What do we observe with respect to depth t?

depth	1	2	3	4	...	t
interval length h	1/2	1/4	1/8	1/16	...	2^{-t}
number of triangles	1	2	4	8	...	$\frac{1}{2}2^t$
surplus v	1	1/4	1/16	1/64	...	$4 \cdot 2^{-2t}$
triangle area D_1	1/2	1/16	1/128	1/1024	...	$4 \cdot 2^{-3t}$
sum (of this t)	1/2	1/8	1/32	1/128	...	$2 \cdot 2^{-2t}$
overall sum ($\leq t$)	1/2	5/8	21/32	85/128	...	$\frac{2}{3}\left(1 - 2^{-2t}\right)$
error	1/6	1/24	1/96	1/384	...	$\frac{2}{3}2^{-2t}$

What do we observe with respect to depth t? Contribution (e.g., "sum (of this t)") goes down with rate 2^{-2t}

SG: Approximation of functions

Analyze this approach in more general context.

Let $\phi_1, \ldots, \phi_n$ be basis functions and represent

$$u(x) = \sum_{i=1}^{n} \alpha_i \phi_i(x)$$

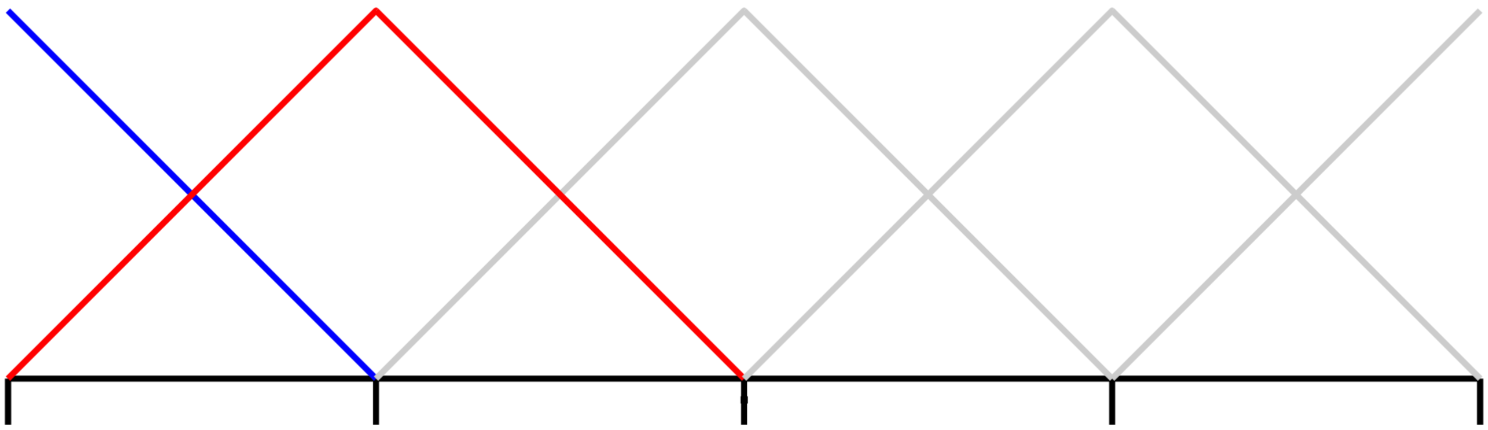
Obtain

$$\int_a^b u(x)\mathrm{d}x = \sum_{i=1}^{n} \alpha_i \int_a^b \phi_i(x)\mathrm{d}x \,,$$

i.e., a quadrature rule as a weighted sum of the coefficients $\alpha_1, \ldots, \alpha_n$

SG: Approximation of functions (cond't)

Represent a continuous, piecewise linear u in nodal point basis



The coefficient $\alpha_1, \ldots, \alpha_n$ are the function values of u at the nodal points

$$u(x) = \sum_{i=1}^{n} \alpha_i \phi_i(x)$$

$\Rightarrow$ instead of nodal point basis, consider a **hierarchical** basis

SG: Piecewise linear functions

Consider only functions $u : [0,1] \to \mathbb{R}$ with $u(0) = u(1) = 0$ in the following.

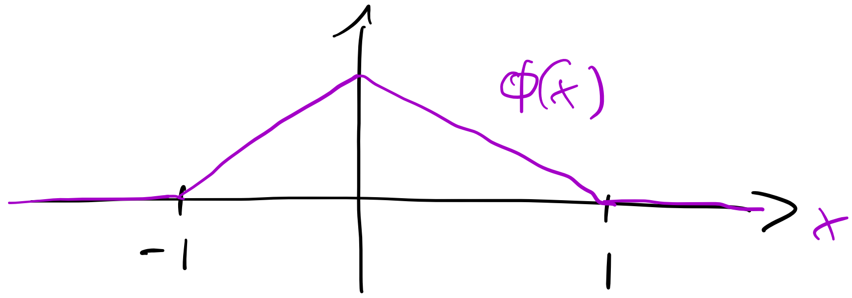
Need the following quantities

- mesh width $h_l = 2^{-l}$
- grid points $x_{l,i} = ih_l = i2^{-l}$
- Basis function

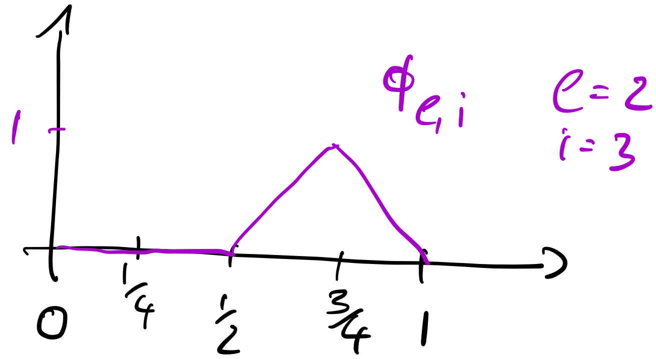
$$\phi_{l,i}(x) = \phi\left(\frac{x - x_{l,i}}{h_l}\right), \qquad \phi(x) = \max\{1 - |x|, 0\}$$

- Nodal point basis $\Phi_l = \{\phi_{l,i} : 1 \leq i < 2^l\}$

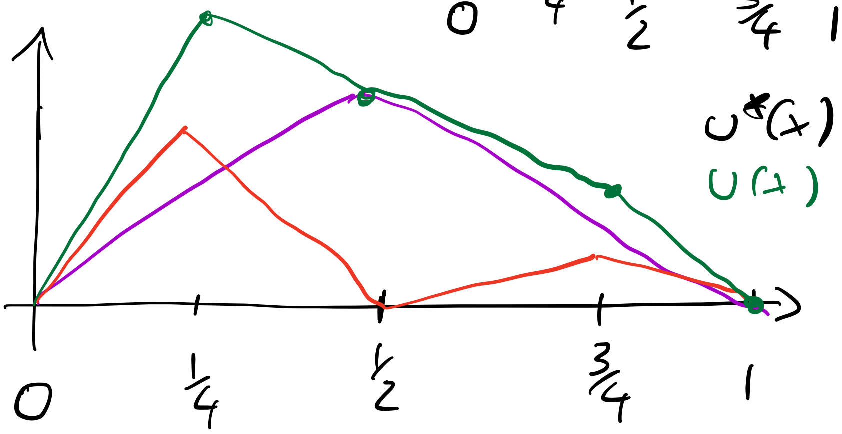
Visualize these basis functions on the board!



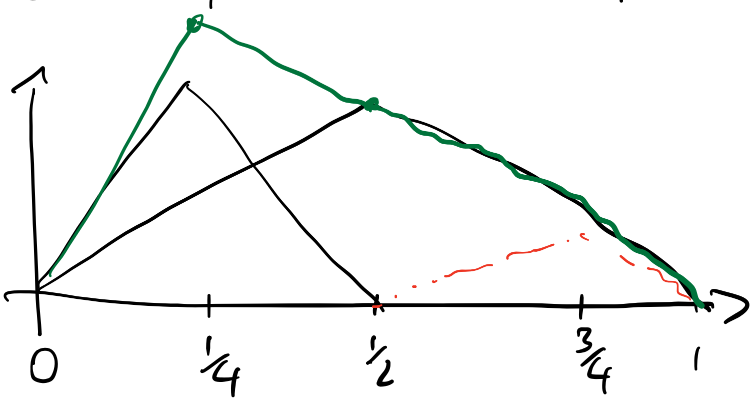
φ(x)
-1
1
x



1
φ_{l,i}
l = 2
i = 3
0
1/4
1/2
3/4
1



u*(x)
u(x)
0
1/4
1/2
3/4
1



0
1/4
1/2
3/4
1

SG: Piecewise linear functions

Consider only functions $u : [0,1] \to \mathbb{R}$ with $u(0) = u(1) = 0$ in the following.

Need the following quantities

- mesh width $h_l = 2^{-l}$
- grid points $x_{l,i} = ih_l = i2^{-l}$
- Basis function

$$\phi_{l,i}(x) = \phi\left(\frac{x - x_{l,i}}{h_l}\right), \qquad \phi(x) = \max\{1 - |x|, 0\}$$

- Nodal point basis $\Phi_l = \{\phi_{l,i} \,:\, 1 \leq i < 2^l\}$

Visualize these basis functions on the board!

The space $V_l = \mathrm{span}(\Phi_l)$ is the space of piecewise linear, continuous functions with respect to the grid points $x_{l,i}$ for $i = 1, \ldots, 2^l - 1$

SG: Hierarchical basis

For the hierarchical representation, we consider the hierarchical increment W_l, spanned by the basis functions $\phi_{l,i}$ such that

$$V_l = V_{l-1} \oplus W_l$$

is a direct sum (each $u_l \in V_l$ can be uniquely decomposed as $u_l = u_{l-1} + w_l$ with $u_{l-1} \in V_{l-1}$ and $w_l \in W_l$ $\rightarrow$ remember the triangles in approach by Archimedes)

Because $\dim(V_l) = 2^l - 1$ and $\dim(V_l) = \dim(V_{l-1}) + \dim(W_l)$ we need $\dim(W_l) = 2^{l-1}$. These are given by $\phi_{l,i}$ with

$$i \in I_l = \{j \,:\, 1 \leq j < 2^l \,, \quad j \text{ odd}\}$$

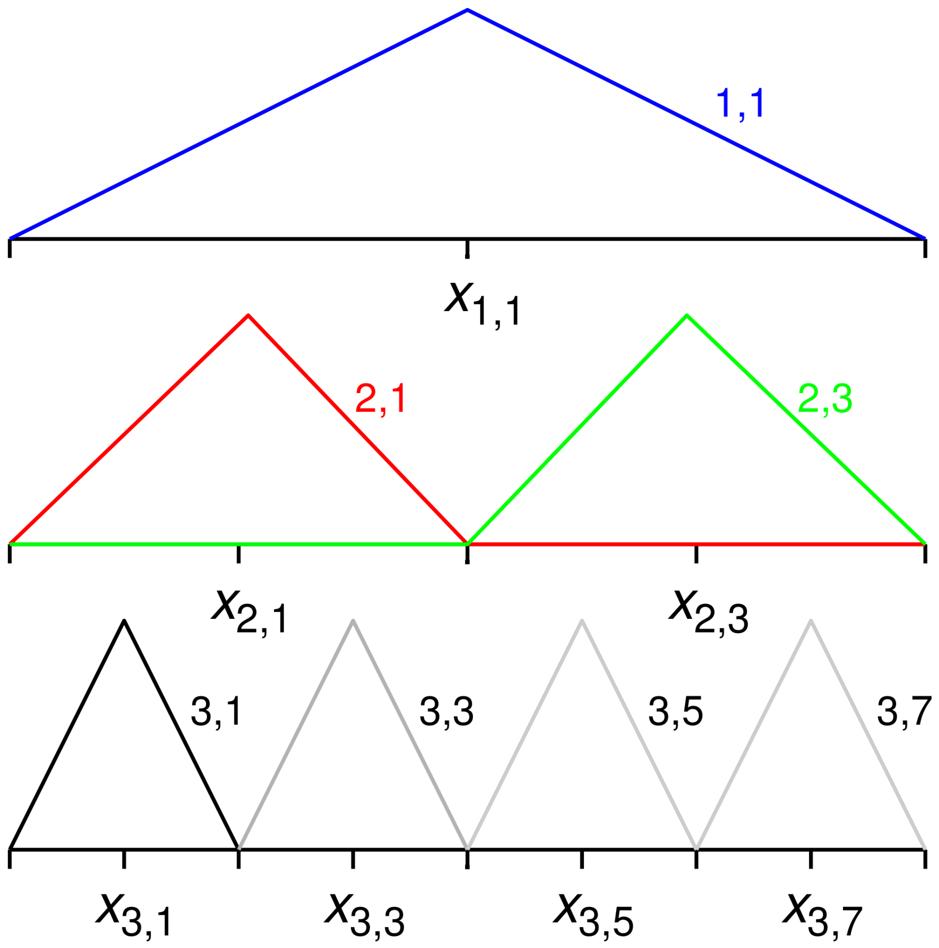
Then

$$W_l = \operatorname{span}\{\phi_{l,i}\,,\, i \in I_l\}$$

Note that $W_1 = V_1$

SG: Hierarchical basis cont'd

The bases for spaces W_1, W_2 and W_3



SG: Hierarchical basis cont'd

Obtain

$$V_n = \bigoplus_{l=1}^{n} W_l \,,$$

so that there is a unique representation for each $u \in V_n$ as

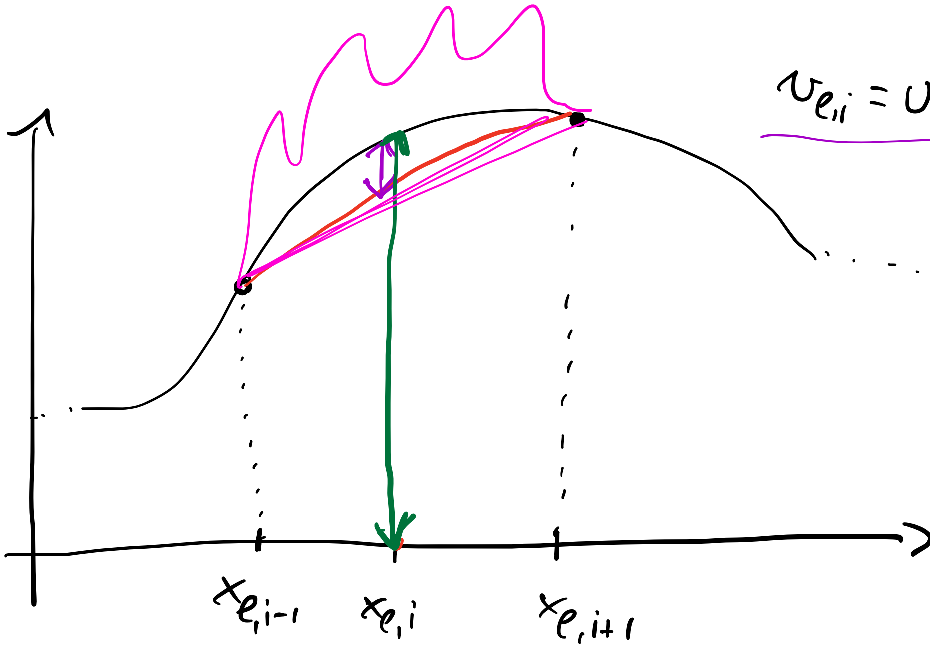
$$u = \sum_{l=1}^{n} w_l = \sum_{l=1}^{n} \sum_{i \in I_l} v_{l,i} \phi_{l,i}$$

Coefficients $v_{l,i}$ in case of interpolation $\rightsquigarrow$ visualize on board

The coefficients $v_{l,i}$ are hierarchical differences

$$v_{l,i} = u^*(x_{l,i}) - \frac{u(x_{l,i-1}) + u(x_{l,i+1})}{2}$$

where u^* is the function to be interpolated



$$v_{e,i} = U^*(x_{e,i}) - \frac{U(x_{e,i-1}) + U(x_{e,i+1})}{2}$$

SG: Analysis of hierarchical decomposition

We now analyze the hierarchical decomposition

$$u_n = \sum_{l=1}^{n} w_l = \sum_{l=1}^{n} \sum_{i \in I_l} v_{l,i} \phi_{l,i}$$

when interpolating a u. What would we like to obtain?

SG: Analysis of hierarchical decomposition

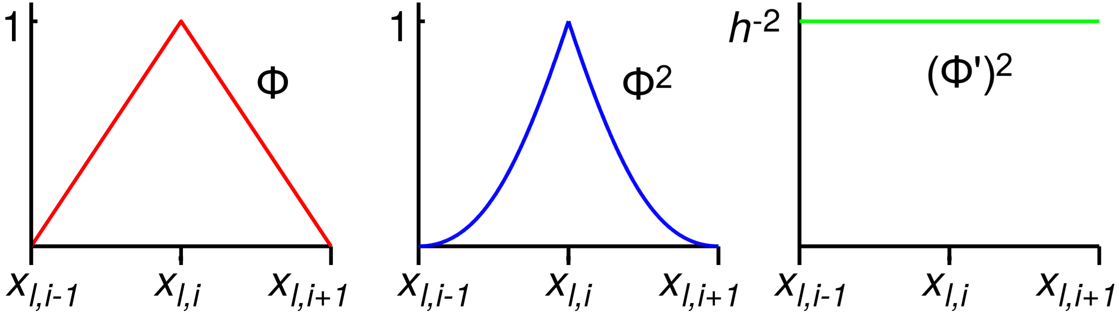
We now analyze the hierarchical decomposition

$$u_n = \sum_{l=1}^{n} w_l = \sum_{l=1}^{n} \sum_{i \in I_l} v_{l,i} \phi_{l,i}$$

when interpolating a u. What would we like to obtain?

For our decomposition, we first calculate the norms of the basis functions

$$\|\phi_{l,i}\|_\infty = 1\,, \qquad \|\phi_{l,i}\|_2 = \sqrt{\frac{2h_l}{3}}$$



Recall

$$u_n = \sum_{l=1}^{n} w_l = \sum_{l=1}^{n} \sum_{i \in I_l} v_{l,i} \phi_{l,i}$$

If u is twice differentiable, the $\| \cdot \|_2$ norm of the increments

$$w_l = \sum_{i \in I_l} v_{l,i} \phi_{l,i}$$

decays as $\mathcal{O}(h_l^2) \rightsquigarrow$ board

$$w_e = \sum_{i \in I_e} v_{e,i} \phi_{e,i}$$

Represent

$$v_{e,i} = \int_0^1 \psi_{e,i}(x) \frac{\partial^2}{\partial x} u(x) dx$$

↳ $\psi_{e,i}(x) = -\frac{h_e}{2} \phi_{e,i}$

↳ Proof: BG, Lemma 3.2

twice differentiable u

$$|v_{e,i}| \leq \frac{h_e}{2} \underbrace{\|\phi_{e,i}\|_2}_{\leq \sqrt{\frac{2h_e}{3}}} \left\|\frac{\partial^2}{\partial x^2} u_{|T_i}\right\|_2$$

$$\left(T_i = [x_{e,i-1}, x_{e,i+1}] \right)$$

$$\leq \sqrt{\frac{h_e^3}{6}} \left\|\frac{\partial^2}{\partial x^2} u_{|T_i}\right\|_2$$

$$\mu_2(u|T_i) = \left\|\frac{\partial^2}{\partial x^2} u_{|T_i}\right\|_2$$

$$w_\ell = \sum_{i \in I_\ell} v_{\ell,i} \, \phi_{\ell,i}$$

$$\|w_\ell\|_2^2 = \int_0^1 w_\ell^2(x)\,dx = \int_0^1 \Big(\sum_{i \in I_\ell} v_{\ell,i}\, \phi_{\ell,i}(x)\Big)^2 dx$$

$$\leq \sum_{i \in I_\ell} v_{\ell,i}^2 \, \|\phi_{\ell,i}\|_2^2$$

$$\leq \frac{h_\ell^3}{6} \, \frac{2h_\ell}{3} \sum_{i \in I_\ell} \mu_2(u|T_i)^2 =$$

$$= \frac{h_\ell^4}{9} \mu_2(u)^2$$

$\Rightarrow \|w_\ell\|_2 \in \mathcal{O}(h_\ell^2)$

$\Rightarrow$ increments decay quad. with level ℓ

<u>if</u> u has bounded 2nd. deriv

$$u = \sum_{\ell=1}^{\infty} w_\ell$$

Recall

$$u_n = \sum_{l=1}^{n} w_l = \sum_{l=1}^{n} \sum_{i \in I_l} v_{l,i} \phi_{l,i}$$

If u is twice differentiable, the $\|\cdot\|_2$ norm of the increments

$$w_l = \sum_{i \in I_l} v_{l,i} \phi_{l,i}$$

decays as $\mathcal{O}(h_l^2) \rightsquigarrow$ board

This means, we can write a twice differentiable u as a series

$$u = \sum_{l=1}^{\infty} w_l \,,$$

that converges because $\|w_l\|_2 \in \mathcal{O}(h_l^2)$. In particular, obtain

$$u - u_n = \sum_{l=n+1}^{\infty} w_l$$

$\rightsquigarrow$ decay of $\|w_l\|_2$ helps us to understand error $u - u_n$

In the 1-dimensional case, the benefit of the hierarchical basis is limited:

In the 1-dimensional case, the benefit of the hierarchical basis is limited:

- We can leave out nodes and still get a reasonable approximation in contrast to a nodal-point basis

In the 1-dimensional case, the benefit of the hierarchical basis is limited:

- We can leave out nodes and still get a reasonable approximation in contrast to a nodal-point basis

- Adding a new level requires us to compute the coefficients of the new level only but keeps the coefficients at the previous levels unchanged

What will “can remove nodes and still get reasonable results” amount to in higher dimension?

In the 1-dimensional case, the benefit of the hierarchical basis is limited:

- We can leave out nodes and still get a reasonable approximation in contrast to a nodal-point basis

- Adding a new level requires us to compute the coefficients of the new level only but keeps the coefficients at the previous levels unchanged

What will "can remove nodes and still get reasonable results" amount to in higher dimension? In the multi-dimensional case, we will now see that many of the hierarchical increments have high costs and low benefit in terms of error $\rightarrow$ we will then remove those hierarchical increments and obtain *sparse grids*

SG: Hierarchical basis in multivariate case

Let now $\boldsymbol{x} = [x_1, \ldots, x_d]$ with $d \in \mathbb{N}$ and $d > 1$.

Again consider the domain $\Omega = [0,1]^d$ and functions $u|_{\partial\Omega} = 0$

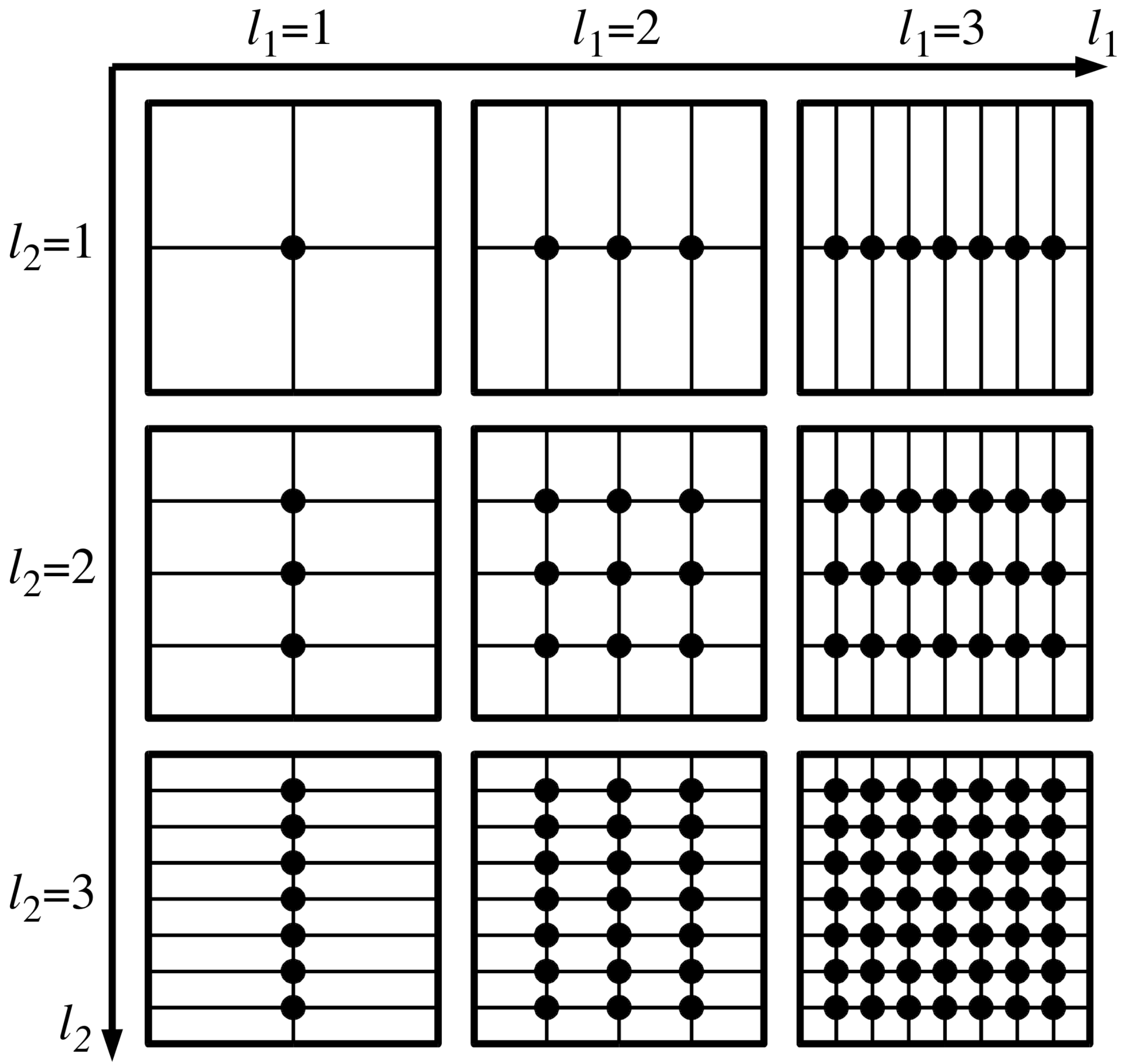
Multidimensional level $\boldsymbol{l} = [l_1, \ldots, l_d] \in \mathbb{N}^d$

Multidimensional mesh width $h_{\boldsymbol{l}} = [h_1, \ldots, h_d] = [2^{-l_1}, \ldots, 2^{-l_d}]$. Note that different mesh width in different dimensions allowed

Define $|\boldsymbol{l}|_1 = l_1 + \cdots + l_d$ and $|\boldsymbol{l}|_\infty = \max\{l_1, \ldots, l_d\}$

Grid points are $\boldsymbol{x}_{\boldsymbol{l},\boldsymbol{i}} = [i_1 h_1, \ldots, i_d h_d]$

SG: Notation

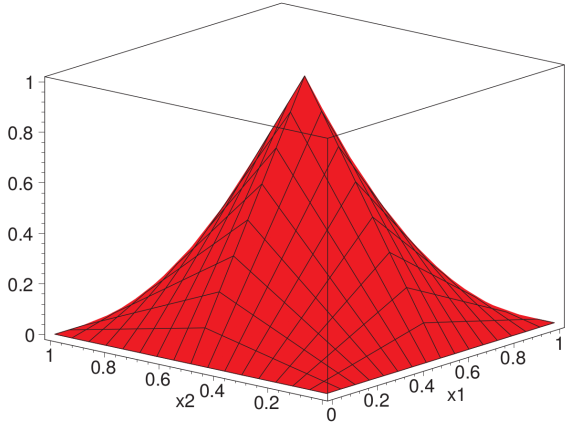


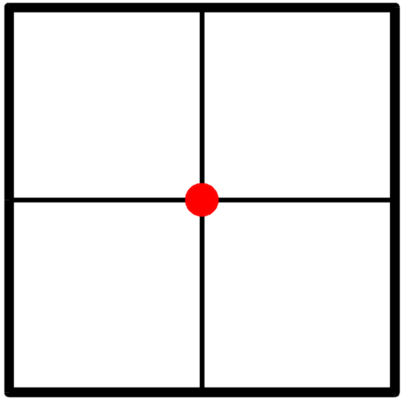
SG: Piecewise d-linear functions

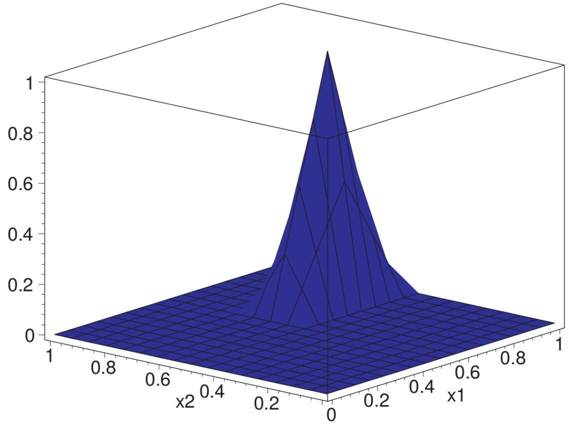
Generalize continuous, piecewise linear functions to continuous, piecewise d-linear functions with respect to $h_{\boldsymbol{l}}$:

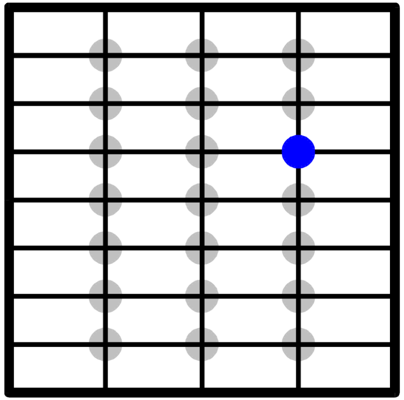
$$\phi_{\boldsymbol{l},\boldsymbol{i}}(\boldsymbol{x}) = \prod_{j=1}^{d} \phi_{l_j,i_j}(x_j)$$

For $d = 2$, the functions $\phi_{[1,1],[1,1]}$ and $\phi_{[2,3],[3,5]}$ are plotted on the right









SG: Spaces

Consider

$$\Phi_{\boldsymbol{l}} = \{\phi_{l,i} \; : \; 1 \leq \boldsymbol{i} < 2^{\boldsymbol{l}}\},$$

where the $\leq$ is to be read component-wise: each i_j must be at least 1 and at most $2^{l_j} - 1$

The space of piecewise d-linear functions is

$$V_{\boldsymbol{l}} = \text{span}\{\Phi_{\boldsymbol{l}}\}$$

with dimension

$$\dim(V_{\boldsymbol{l}}) = (2^{l_1} - 1) \cdots (2^{l_d} - 1) \in \mathcal{O}(2^{|\boldsymbol{l}|_1})$$

Special case $l_1 = \cdots = l_d$ set $V_n = V_{[n,\ldots,n]}$

SG: Hierarchical increments

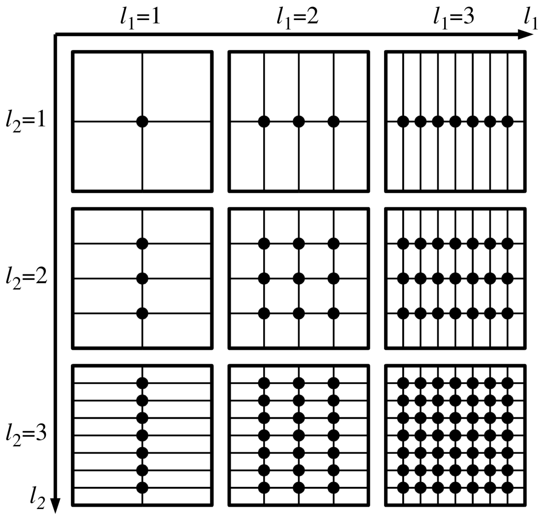
Define the hierarchical increment $W_{\boldsymbol{l}}$ as

$$W_{\boldsymbol{l}} = \mathrm{span}\{\phi_{\boldsymbol{l},\boldsymbol{i}} \,:\, \boldsymbol{i} \in I_{\boldsymbol{l}}\}\,,$$

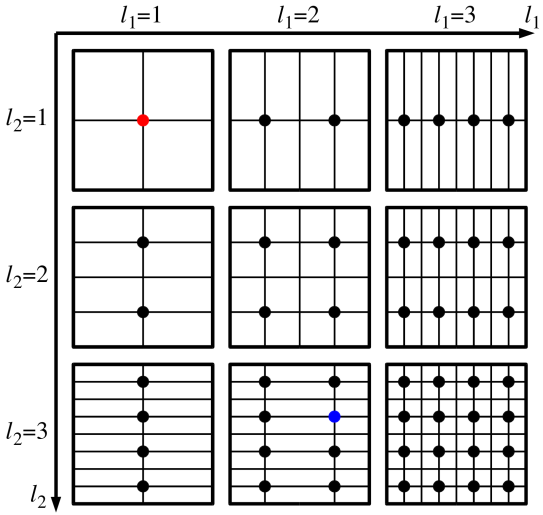
where $\boldsymbol{I}_{\boldsymbol{l}} = \{\boldsymbol{i} \,:\, 1 \leq \boldsymbol{i} < 2^{\boldsymbol{l}}\,,\ \text{all } i_j \text{ odd}\}$

Contains just those functions from $\boldsymbol{V}_{\boldsymbol{l}}$ that vanish at all points of coarser grid

full grid



hierarchical increments



Multi-dimensional

Represent $u \in V_\ell$ as

$$u = \sum_{\ell' \leq \ell} w_{\ell'} \quad , \quad w_{\ell'} \in W_{\ell'}$$

$\Rightarrow$ estimate decay of $\|w_\ell\|_2$

Hierarchical coefficient

$$v_{\ell,i} = \int_{[0,1]^d} \psi_{\ell,i} \, \partial^{2d} u \, dx$$

$$\psi_{\ell,i} = 2^{-|\ell|_1 - d} \, \phi_{\ell,i}$$

$$\partial^{2d} u = \frac{\partial^{2d} u}{\partial x_1^2 \cdots \partial x_d^2}$$

mixed
2d-fold derivative

$\Rightarrow$ strong assumption

$$\|w_\ell\|_2 \leq 3^{-d} \underbrace{2^{-2|\ell|_1}}_{\rightarrow \text{fast}} \|\partial^{2d} u\|_2$$

SG: Hierarchical subspace decomposition

Obtain unique representation of $u_{\boldsymbol{l}} \in V_{\boldsymbol{l}}$ for $\boldsymbol{l} \in \mathbb{N}^d$ as

$$u_{\boldsymbol{l}} = \sum_{\boldsymbol{l}' \leq \boldsymbol{l}} w_{\boldsymbol{l}'} \,,$$

with $w_{\boldsymbol{l}'} \in W_{\boldsymbol{l}'}$

Now it will be worthwhile to estimate the norm of $w_{\boldsymbol{l}}$ to understand which contribute most to the accuracy of the representation

Details in [Bungartz et al., 2004]: Hierarchical coefficient is now

$$v_{\boldsymbol{l},\boldsymbol{i}} = \int_{[0,1]^d} \psi_{\boldsymbol{l},\boldsymbol{i}} \partial^{2d} u \mathrm{d}\boldsymbol{x} \,,$$

which now depends on **mixed** 2d-fold derivative

$$\partial^{2d} u = \frac{\partial^{2d}}{\partial x_1^2 \cdots \partial x_d^2}$$

and $\psi_{\boldsymbol{l},\boldsymbol{i}} = 2^{-|\boldsymbol{l}|_1 - d} \phi_{\boldsymbol{l},\boldsymbol{i}}$

It is very important to note that the following holds *only* for functions with (L_2-)bounded $\partial^{2d} u \Rightarrow$ strong assumption on function (smoothness)

Obtain

$$\|w_{\boldsymbol{l}}\|_2 \leq 3^{-d} 2^{-2|\boldsymbol{l}|_1} \|\partial^{2d} u\|_2$$

Now we select those subspaces from the subspace scheme that minimize cost and maximize benefit for approximation function $u : [0,1]^d \to \mathbb{R}$ with sufficient smoothness

How should we measure costs?

Now we select those subspaces from the subspace scheme that minimize cost and maximize benefit for approximation function $u : [0, 1]^d \rightarrow \mathbb{R}$ with sufficient smoothness

How should we measure costs? We measure cost via the number of grid points

$$c(\boldsymbol{l}) = 2^{|\boldsymbol{l}|_1 - d}$$

How should we measure benefit?

Now we select those subspaces from the subspace scheme that minimize cost and maximize benefit for approximation function $u : [0,1]^d \to \mathbb{R}$ with sufficient smoothness

How should we measure costs? We measure cost via the number of grid points

$$c(\boldsymbol{l}) = 2^{|\boldsymbol{l}|_1 - d}$$

How should we measure benefit? Measure benefit of subspace selection via introduced error if left out. Let $L \subset \mathbb{N}^d$ of levels that are selected, then obtain

$$u_L = \sum_{\boldsymbol{l} \in L} w_{\boldsymbol{l}}$$

and

$$u - u_L = \sum_{\boldsymbol{l} \notin L} w_{\boldsymbol{l}}$$

For each component $w_{\boldsymbol{l}}$ have derived bounds of the form

$$\|w_{\boldsymbol{l}}\| \leq s(\boldsymbol{l})\mu(u)$$

where $\mu(u)$ was typically $\|\partial^{2d} u\|_2$. Then obtain

$$\|u - u_L\| \leq \sum_{\boldsymbol{l} \notin L} \|w_{\boldsymbol{l}}\|_2 \leq \left(\sum_{\boldsymbol{l} \notin L} s(\boldsymbol{l}) \right) \mu(u)$$

$$= \left[\left(\sum_{\boldsymbol{l} \in \mathbb{N}^d} s(\boldsymbol{l}) \right) - \left(\sum_{\boldsymbol{l} \in L} s(\boldsymbol{l}) \right) \right] \mu(u)$$

Thus, if we select $\boldsymbol{l}$ with $s(\boldsymbol{l})$ big, then the error $\|u - u_L\|$ is reduced $\rightarrow$ use $s(\boldsymbol{l})$ as the benefit of subspace $W_{\boldsymbol{l}}$

SG: Quality of full-grid space

With this new tool in hand, let us analyze the cost/benefit of a "full grid", i.e., a grid corresponding to the selection

$$L_n = \{\boldsymbol{l} : |\boldsymbol{l}|_\infty \leq n\}$$

In the L_2 norm, we have bounds of the order

$$s(\boldsymbol{l}) = 2^{-2|\boldsymbol{l}|_1}$$

Then, calculations show that

$$\sum_{\boldsymbol{l} \in L_n} s(\boldsymbol{l}) \geq \left(\frac{1}{3}\right)^d (1 - d2^{-2n})$$

and for $n \to \infty$

$$\sum_{\boldsymbol{l} \in \mathbb{N}^d} s(\boldsymbol{l}) = \left(\frac{1}{3}\right)^d$$

Thus

$$\sum_{\boldsymbol{l} \in \mathbb{N}^d} s(\boldsymbol{l}) - \sum_{\boldsymbol{l} \in L_n} s(\boldsymbol{l}) \leq \left(\frac{1}{3}\right)^d - \left(\frac{1}{3}\right)^d (1 - d2^{-2n}) \leq \frac{d}{3^d} 2^{-2n}$$

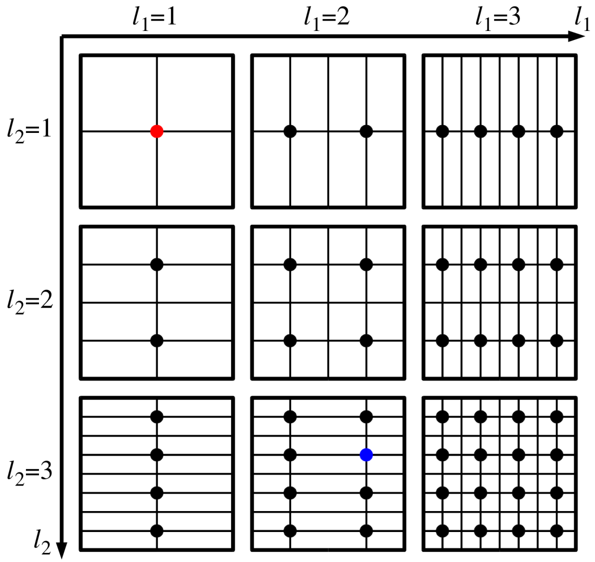
SG: Approximation quality of full-grid space

Obtain

$$\|u - u_{L_n}\|_2 \leq C \sum_{\boldsymbol{l} \notin L_n} s(\boldsymbol{l}) \leq \frac{Cd}{3^d} 2^{-2n} \in \mathcal{O}(h_n^2)$$

This is what we expect from a piecewise linear approximation

What additional insights have we obtained?



SG: Approximation quality of full-grid space

Obtain

$$\|u - u_{L_n}\|_2 \leq C \sum_{\boldsymbol{l} \notin L_n} s(\boldsymbol{l}) \leq \frac{Cd}{3^d} 2^{-2n} \in \mathcal{O}(h_n^2)$$

This is what we expect from a piecewise linear approximation

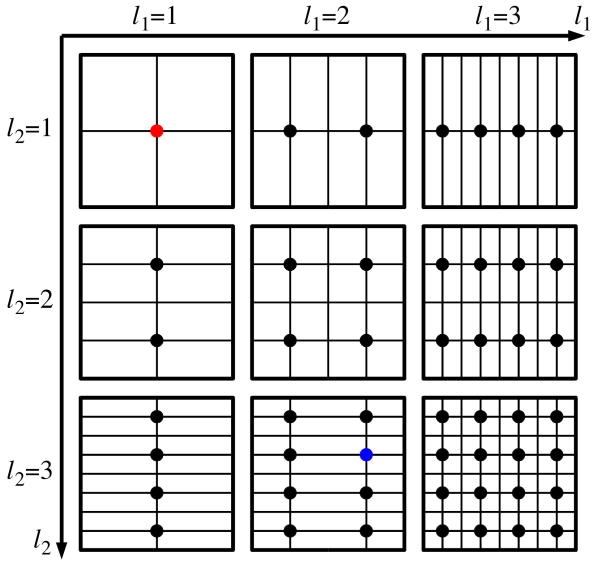
What additional insights have we obtained?

- The sum of local benefits

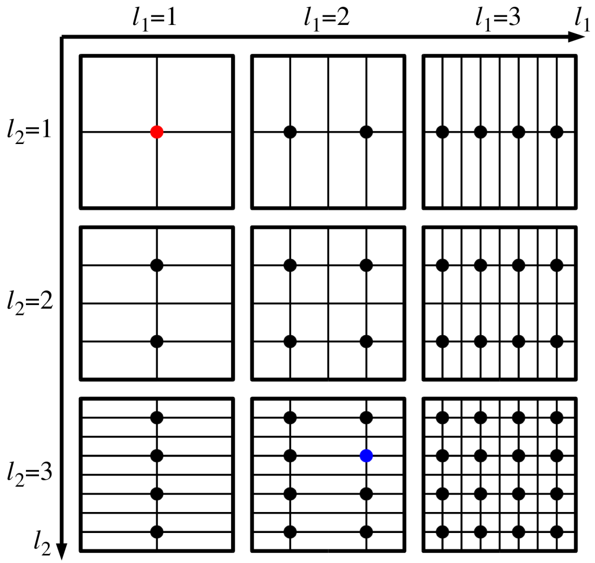
$$\sum_{\boldsymbol{l} \in L_n} 2^{-2|\boldsymbol{l}|_1} ,$$

  means that subspace on diagonals (i.e., with constant $|\boldsymbol{l}|_1$) have the same benefit.

- Also, as we move further to the bottom right, the benefit gets less and less



If we now look at the costs $c(\boldsymbol{l}) = 2^{|\boldsymbol{l}|_1 - d}$ (which is the number of grid points), then we see that the costs are constant on diagonals as well

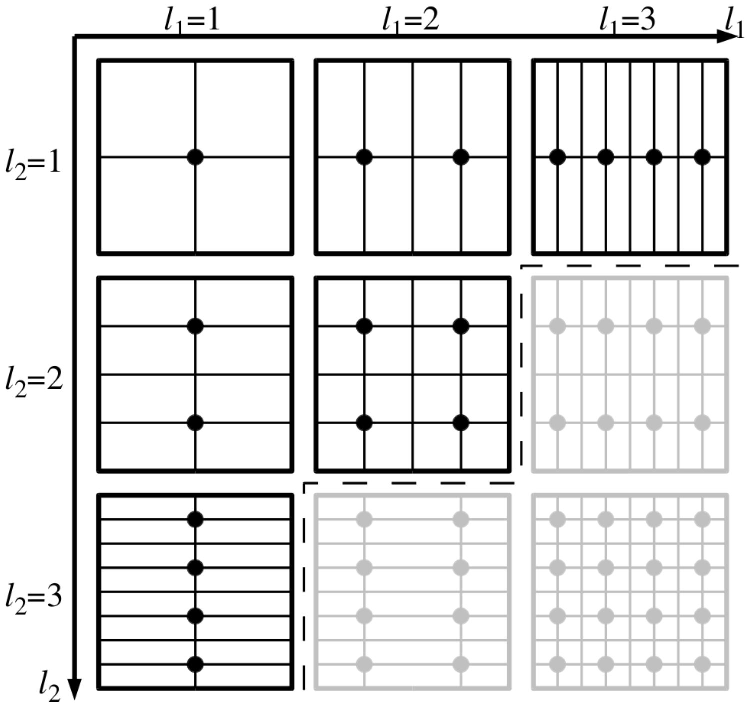


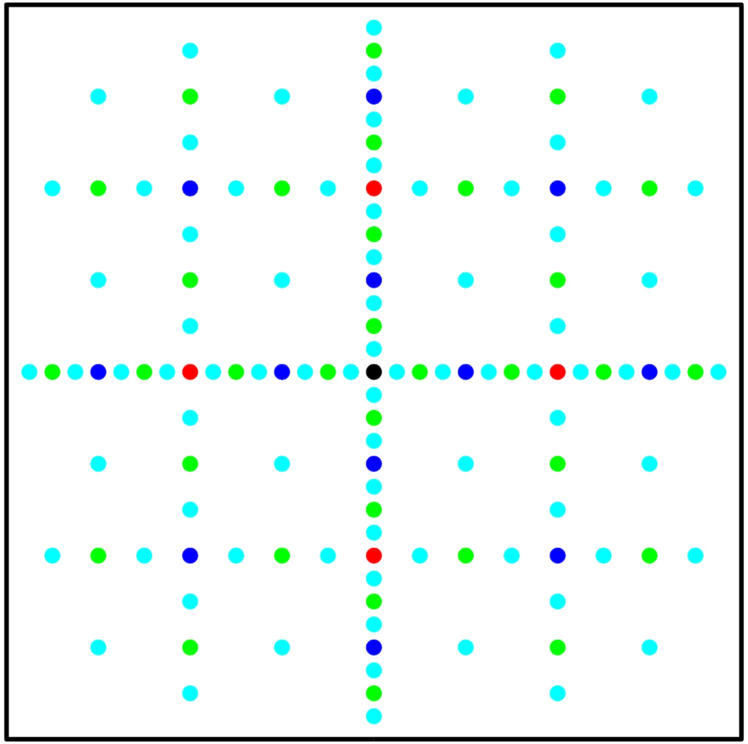
Thus, the cost/benefit ratio $c(\boldsymbol{l})/s(\boldsymbol{l})$ is *constant* on diagonals. In particular, for lower-triangular diagonals, we add subspaces with worse and worse cost/benefit ratios $\Rightarrow$ let's truncate them

Consider the diagonal cut $L_n^{(1)} = \{\boldsymbol{l} : |\boldsymbol{l}|_1 \leq n + d - 1\}$ and the sparse grid space

$$\boldsymbol{V}_n^{(1)} = \bigoplus_{|\boldsymbol{l}|_1 \leq n+d-1} W_{\boldsymbol{l}}$$

Here is an example of a sparse grid





SG: Properties of sparse grids

The number of grid points of a sparse grid grows as $\mathcal{O}(2^n n^{d-1})$ in contrast to $\mathcal{O}(2^{nd})$ of a full grid

If u has (L_2-)bounded mixed derivatives up to order $2d$, then

$$\|u - u_n^{(1)}\|_2 \in \mathcal{O}(2^{-2n} n^{d-1})\,,$$

whereas a full-grid space achieves

$$\|u - u_n\|_2 \in \mathcal{O}(2^{-2n})$$

Sparse-grid spaces achieve slightly worse error than full-grid space but drastically reduced points in higher dimensions d

Comparing the number of grid points corresponding to full-grid and sparse-grid spaces:

Dimension $d = 2$:

n	1	2	3	4	5	...	10
$\dim V_n = (2^n - 1)^2$	1	9	49	225	961	...	1 046 529
$\dim V_n^1 = 2^n(n-1)+1$	1	5	17	49	129	...	9 217

Dimension $d = 3$:

n	1	2	3	4	...	10
$\dim V_n = (2^n - 1)^3$	1	27	343	3 375	...	1 070 590 167
$\dim V_n^1 = 2^n\left(\frac{n^2}{2} - \frac{n}{2} + 1\right) - 1$	1	7	31	111	...	47 103

Why does this work?

Why does this work?

Exploit the additional smoothness given by the assumption on the mixed derivatives of function u

The hierarchical basis is a key ingredient:

Why does this work?

Exploit the additional smoothness given by the assumption on the mixed derivatives of function u

The hierarchical basis is a key ingredient:

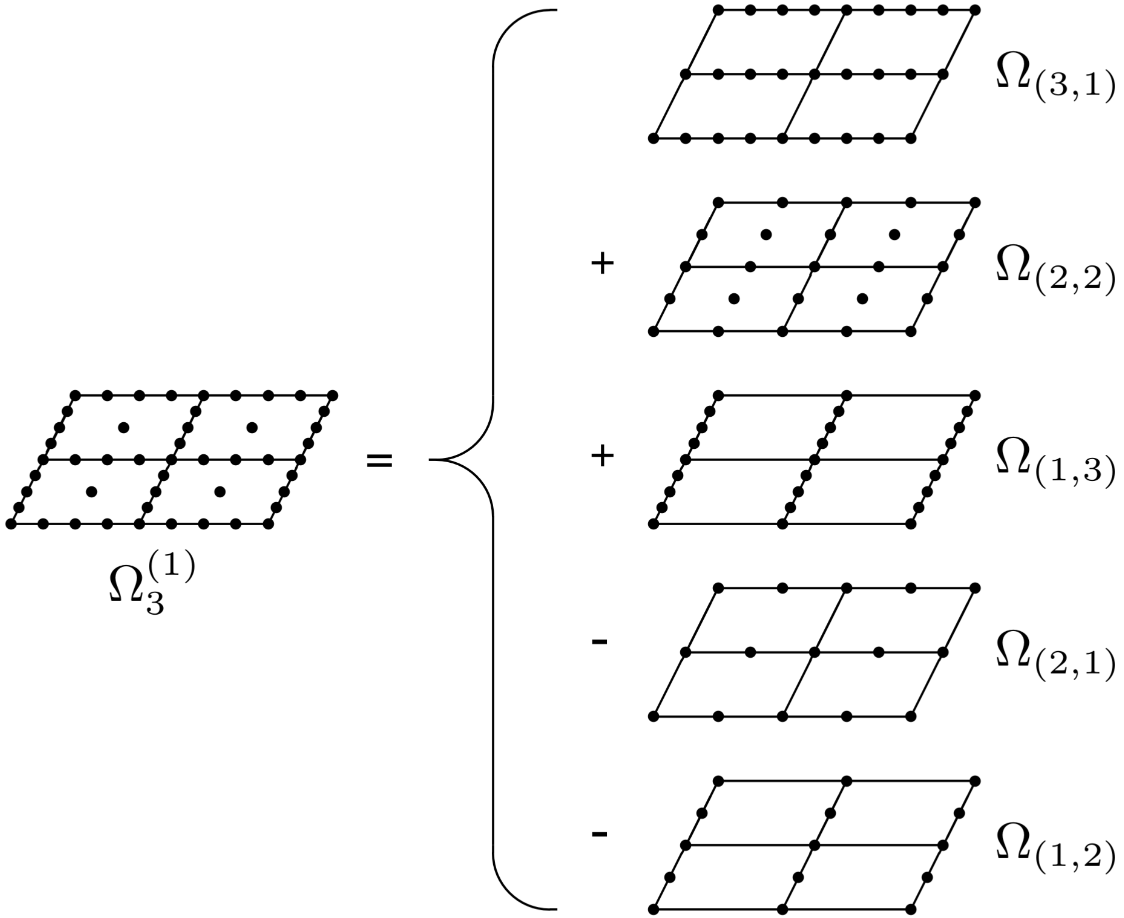
- Exploits smoothness by having "semi-global" support (i.e., function is smoother, so we can reach far over the domain and know it won't change too much),
- Introduced a hierarchy/multilevel and the coefficients in this hierarchy/multilevel basis decay fast ($\rightarrow$ multigrid, multilevel Monte Carlo)

(Logarithmic dependence can be avoided if measure error in energy norm)

Details: Bungartz, Griebel, Acta Numerica, 2004

SG: Combination technique

Formally, sparse grids are superpositions of coarser full grids



SG: Combination technique (cont'd)

This works for grids and also for functions (in certain situations)

- interpolation
- quadrature
- solutions of partial differential equations $\rightarrow$ limited

$\Rightarrow$ we are interested in quadrature

For quadrature, Smolyak has developed a related approach already in 1963

SG: Smolyak quadrature - 1D

Set $D = [-1, 1]$ and consider a one-dimensional function $f : D \to \mathbb{R}$ and we are interested in

$$If = \int_D f(x)\mathrm{d}x$$

One-dimensional quadrature rule

$$Q_l^1 f = \sum_{i=1}^{n_l} w_i f(x_i)$$

with weights $w_1, \ldots, w_{n_l}$ and points $x_1, \ldots, x_{n_l}$ and

$$X_l = \{x_i \, : \, 1 \leq i \leq n_l\}$$

Quadrature rules are nested if $X_l \subset X_{l+1}$

SG: Smolyak quadrature

Define the difference formula

$$\Delta_k f = (Q^1_k - Q^1_{k-1})f \, ,$$

with $Q^1_0 f = 0$. What does the difference formula remind you of?

SG: Smolyak quadrature

Define the difference formula

$$\Delta_k f = (Q_k^1 - Q_{k-1}^1) f \, ,$$

with $Q_0^1 f = 0$. What does the difference formula remind you of? hierarchical coefficients

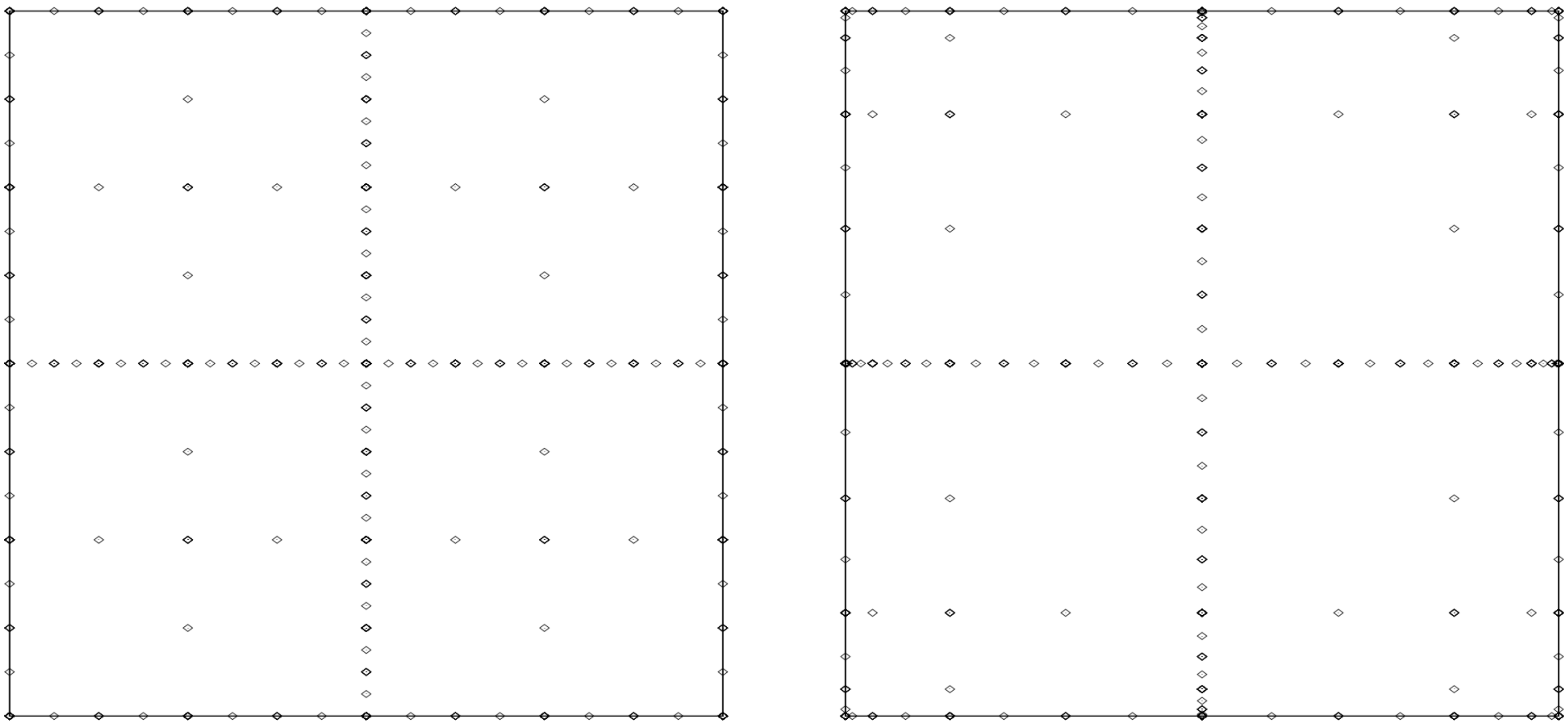
Smolyak's quadrature rule is

$$Q_l^d f = \sum_{|\boldsymbol{k}|_1 \leq n+d-1} (\Delta_{k_1}^1 \otimes \cdots \otimes \Delta_{k_d}^1) f \, ,$$

where the tensor product of quadrature rules is

$$(\Delta_{k_1}^1 \otimes \cdots \otimes \Delta_{k_d}^1) f = \sum_{i_1=1}^{n_{k_1}} \cdots \sum_{i_d=1}^{n_{k_d}} w_{k_1,i_1} \cdots w_{k_d,i_d} f(x_{k_1,i_1}, \ldots, x_{k_d,i_d})$$

SG: Smolyak grid w.r.t. Clenshaw-Curtis rule



Left: sparse grid w.r.t. trapezoidal rule (piecewise constant), right: sparse grid obtained with Clenshaw-Curtis rule

SG: Alternative representations of Smolyak

A non-hierarchical representation is

$$Q_l^d f = \sum_{n \le |\boldsymbol{l}|_1 \le n+d-1} (-1)^{n+d-1-|\boldsymbol{l}|_1} \binom{d-1}{|\boldsymbol{l}|_1 - n} (Q_{l_1}^1 \otimes \cdots \otimes Q_{l_d}^1) f$$

- Non-hierarchical: Can work on regular grids as with combination technique
- Simple to implement
- (Equivalence to hierarchical representation not obvious.)

Conclusions

- Computations in higher dimensions are typically affected by the curse of dimensionality, which means that computational costs become exponentially more expensive as the dimension is increased.
- We have two options in high dimension. The first option is using randomized methods and Monte Carlo which can circumvent the curse
- The other option is exploiting additional structure in the problem that can help to circumvent the curse to some extent.
- In case of sparse grids, we can circumvent the curse by assuming additional smoothness of the function to be interpolated.
- Sparse grids are very useful for quadrature in moderately high dimensions (10 up to few 100s of dimensions). Quadrature based on sparse grids is sometimes called Smolyak quadrature.