Sparse grid methods

Vesa Kaarnioja (kaarnioja@uni-mannheim.de)
Claudia Schillings (c.schillings@uni-mannheim.de)

October 7, 2019
University of Mannheim

Course overview

Mondays 8.30 – 11.45 in B6, 23-25 A304.

Course completion by active participation and giving a presentation ($\sim$ 45 min).

Timetable:

7.10.	First lecture and assignment of topics
14.10.	Second lecture/programming sparse grids
21.10.	Student talks
28.10.	Student talks
4.11.–	Student talks (organized by Claudia Schillings)

What is multivariate integration?

In many applications, we are interested in computing statistical quantities such as expectations and variances. In these situations, high dimensional integrals such as

$$\mathcal{I}f = \int_{[0,1]^d} f(x)\,\mathrm{d}P(x)$$

may need to be approximated using quadrature rules of the form

$$\mathcal{Q}f = \sum_{i=1}^{n} w_i f(x_i) \approx \mathcal{I}f,$$

where $(w_i)_{i=1}^n$ are weights and $(x_i)_{i=1}^n$ are nodes in $[0,1]^d$.

Example (Uncertainty quantification)

Consider the following elliptic PDE with a random diffusion coefficient A:

$$-\nabla \cdot (A(x,\omega)\nabla u(x,\omega)) = f(x), \quad x \in D,\ \omega \in \Gamma,$$
$$u(\cdot,\omega)|_{\partial D} = 0,$$

where $(\Gamma, \mathcal{A}, P)$ is a probability space. Natural quantities of interest are the response statistics of the solution u; typically $\mathbb{E}[u]$ and $\mathrm{Var}[u]$.

Multidimensional Monte Carlo integration

Let $\Omega \subset \mathbb{R}^d$ be a bounded domain. We wish to integrate

$$\int_\Omega f(x)\,\mathrm{d}x.$$

Idea: $\mathrm{d}x/\mathrm{vol}(\Omega)$ is the probability density of the uniform distribution; hence

$$\int_\Omega f(x)\,\mathrm{d}x \approx \frac{\mathrm{vol}(\Omega)}{n}\sum_{i=1}^{n} f(x_i), \quad \mathrm{vol}(\Omega) = \int_\Omega \mathrm{d}x,$$

where $(x_i)_{i=1}^n$ is a random sample of points in Ω.

Convergence rate: $\mathcal{O}(1/\sqrt{n})$ according to the Central Limit Theorem.
Independent of dimension d, but extremely slow nonetheless!

Multidimensional Quasi-Monte Carlo integration

Quasi-Monte Carlo (QMC) methods leverage smoothness of the integrand in order to *deterministically* select a sequence of quadrature nodes $(x_i)_{i=1}^n$ such that

$$\int_{[0,1]^d} f(x)\,\mathrm{d}x \approx \frac{1}{n}\sum_{i=1}^{n} f(x_i),$$

where the approximation to the integral converges at a faster-than-Monte-Carlo rate.

Convergence rate depends on the function space and QMC method, and can range from linear convergence to higher order convergence.

Multidimensional integration over hypercubes $[a, b]^d$

Let $((w_i, x_i))_{i=1}^n$ be the weights and nodes of your favorite univariate quadrature rule:

$$\int_a^b f(x)\,\mathrm{d}x \approx \sum_{i=1}^n w_i f(x_i).$$

Let $g\colon [a, b]^d \to \mathbb{R}$ be a d-variate function. Then you can integrate over the hypercube $[a, b]^d$ by composing your favorite quadrature rule over all axes:

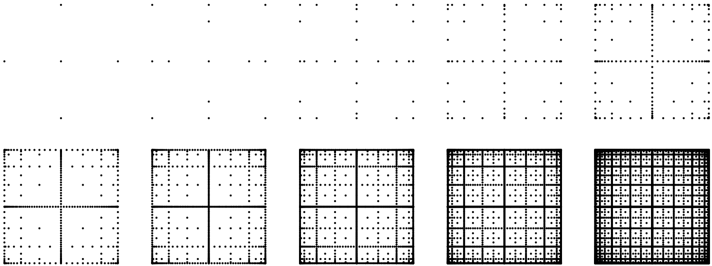
$$\int_{[a,b]^d} g(x_1, \ldots, x_d)\,\mathrm{d}x_1 \cdots \mathrm{d}x_d$$

$$\approx \sum_{i_1=1}^n \cdots \sum_{i_d=1}^n w_{i_1} \cdots w_{i_d} g(x_{i_1}, \ldots, x_{i_d}).$$

Cost: n^d function evaluations!

Sparse grids are, by design, certain subsets of the tensor product quadrature on the previous slide which retain many of the good approximation properties of tensor grids (e.g., polynomial exactness) while lessening the number of evaluation points.

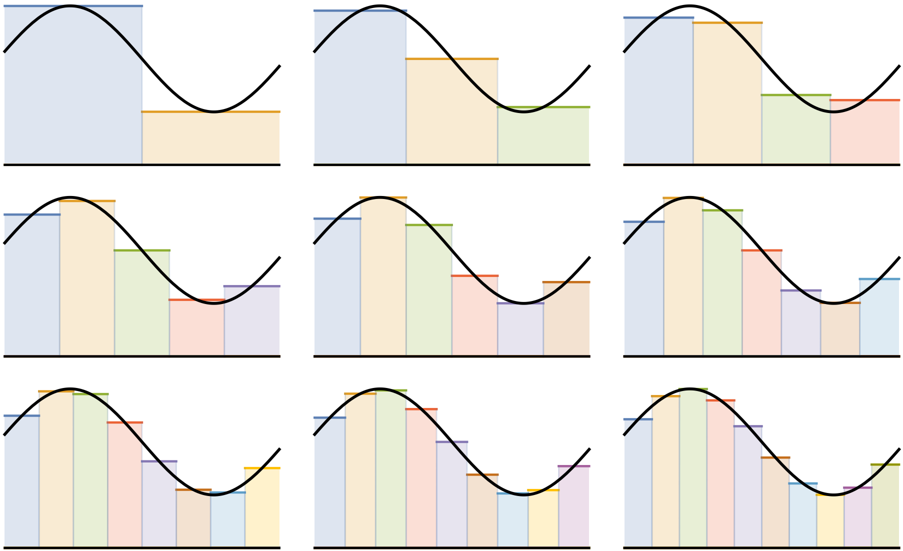
For sparse grid methods and QMC methods, the smoothness of the integrand is leveraged in order to obtain a convergence rate faster than Monte Carlo.



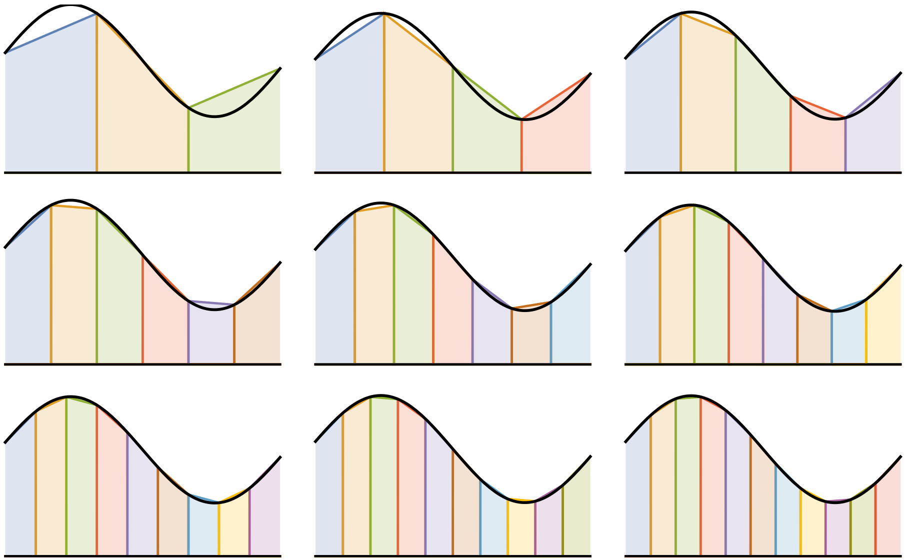
Univariate integration

Newton–Cotes rules

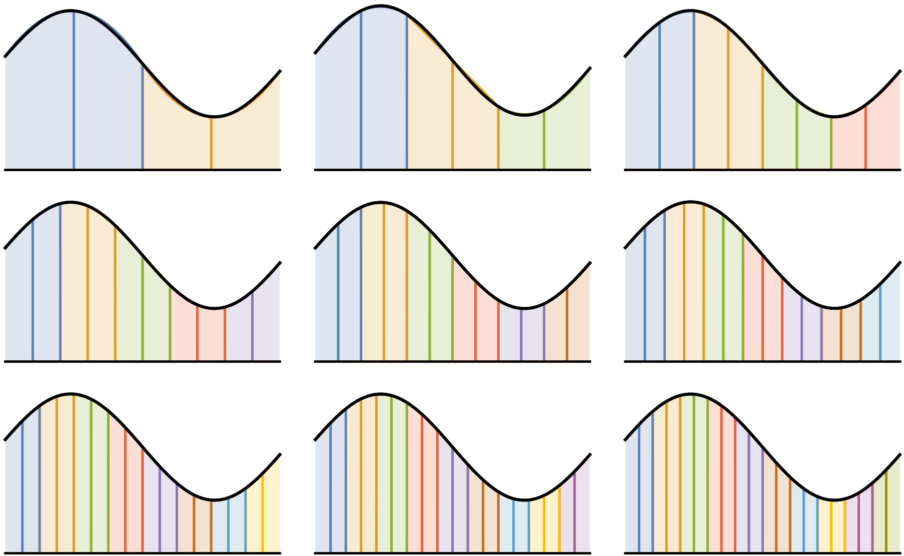
Midpoint rule: $U_n f = \sum_{i=1}^{n} (x_i - x_{i-1}) f\left(\frac{x_{i-1}+x_i}{2}\right) \approx \int_a^b f(x)\,\mathrm{d}x$



Trapezoidal rule: $U_n f = \sum_{i=1}^{n} (x_i - x_{i-1}) \frac{f(x_{i-1}) + f(x_i)}{2} \approx \int_a^b f(x)\,\mathrm{d}x$



Simpson's rule: $U_n f = \sum_{i=1}^{n} (x_i - x_{i-1}) \frac{f(x_{i-1}) + 4f(x_i) + f(x_{i+1})}{3} \approx \int_a^b f(x)\,\mathrm{d}x$



Gaussian quadratures

Gaussian quadratures are used to evaluate integrals of the form

$$\int_{\mathcal{I}} f(x)W(x)\,\mathrm{d}x \approx \sum_{i=1}^{n} w_i f(x_i). \tag{1}$$

Let $(p_k)_{k=0}^{\infty}$ be a family of orthogonal polynomials with respect to the inner product $\langle p, q\rangle_W := \int_{\mathcal{I}} p(x)q(x)W(x)\,\mathrm{d}x$.[1] It turns out that by taking the roots of p_n as the quadratures nodes $(x_i)_{i=1}^{n}$, the weights $(w_i)_{i=1}^{n}$ can be chosen s.t. equality in (1) holds for polynomials with degree $\leq 2n-1$.

It can be shown that any orthogonal polynomial p_n w.r.t. $\langle\cdot,\cdot\rangle_W$ has n real, distinct roots that lie in the interval $\mathcal{I}$.

- The interval $\mathcal{I} = [-1, 1]$ and weight $W(x) = 1$ is associated with the Legendre polynomials P_n.
- The interval $\mathcal{I} = \mathbb{R}$ and weight $W(x) = \mathrm{e}^{-x^2}$ is associated with the Hermite polynomials H_n.

[1] The weight function is assumed to be non-negative a.e. such that $\langle x^k, 1\rangle_W < \infty\ \forall k$.

All orthogonal polynomials admit to a *three-term recurrence relation*

$$\begin{aligned} p_0(x) &= 1, \\ p_1(x) &= (x - \alpha_1) p_0(x), \\ p_{k+1}(x) &= (x - \alpha_{k+1}) p_k(x) - \beta_{k+1} p_{k-1}(x), \end{aligned}$$

where

$$\alpha_{k+1} = \frac{\langle x p_k, p_k \rangle_W}{\langle p_k, p_k \rangle_W} \quad \text{and} \quad \beta_{k+1} = \frac{\langle p_k, p_k \rangle_W}{\langle p_{k-1}, p_{k-1} \rangle_W}.$$

Example (Legendre polynomials)

$$\alpha_k = 0 \ \forall k \quad \text{and} \quad \beta_k = \frac{(k-1)^2}{4k^2 - 8k + 3}, \quad k \geq 2.$$

Example (Hermite polynomials)

$$\alpha_k = 0 \ \forall k \quad \text{and} \quad \beta_k = \frac{k-1}{2}, \quad k \geq 2.$$

Similar formulae exist for most known families of orthogonal polynomials.

Algorithm (Golub–Welsch)

(i) *Construct the tridiagonal $n \times n$ matrix*

$$A = \begin{bmatrix} \alpha_1 & \sqrt{\beta_2} & & & \\ \sqrt{\beta_2} & \alpha_2 & \sqrt{\beta_3} & & \\ & \sqrt{\beta_3} & \alpha_3 & \ddots & \\ & & \ddots & \ddots & \sqrt{\beta_n} \\ & & & \sqrt{\beta_n} & \alpha_n \end{bmatrix}.$$

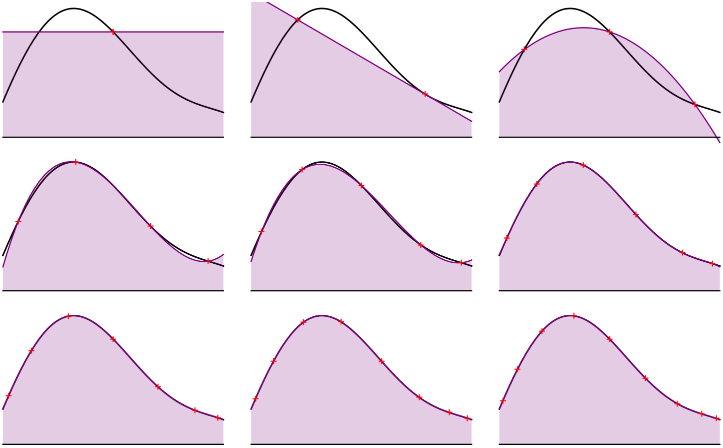
(ii) **Fact:** *The eigenvalues $x_1, \ldots, x_n$ of A are precisely the roots of p_n.*

(iii) **Fact:** *Let $\mathbf{q}_j = [q_{1,j}, \ldots, q_{n,j}]^{\mathrm{T}}$ be the normalized eigenvector corresponding to eigenvalue x_j. Then $w_j = q_{1,j}^2 \int_{\mathcal{I}} W(x)\,\mathrm{d}x$.*

(iv) *Compute the Gaussian quadrature*

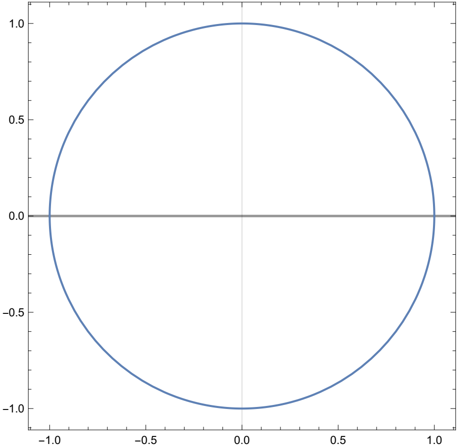
$$\int_{\mathcal{I}} W(x) f(x)\,\mathrm{d}x \approx \sum_{i=1}^{n} w_i f(x_i).$$

Gauss–Legendre quadrature in $[-1, 1]$ based on roots of the Legendre polynomial P_n for $n \in \{1, \dots, 9\}$.

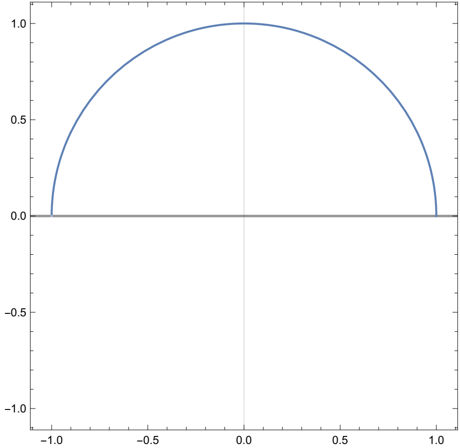


Clenshaw–Curtis quadrature and other nested quadrature rules

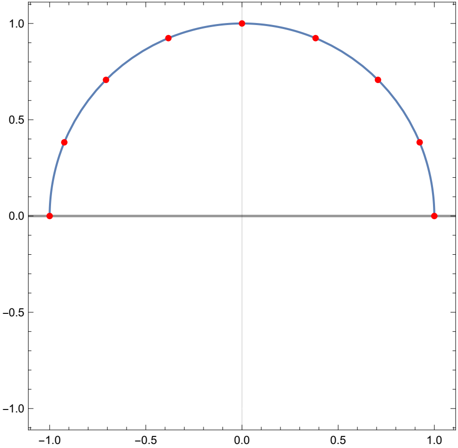
Clenshaw–Curtis nodes in 1d



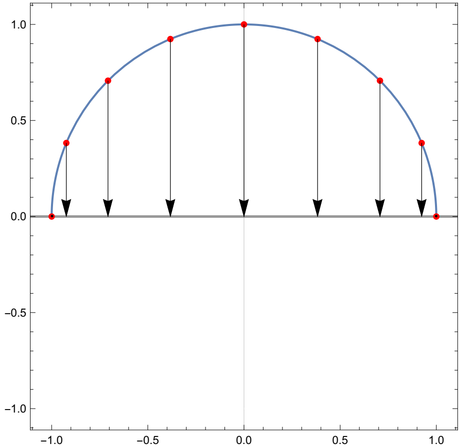
Clenshaw–Curtis nodes in 1d



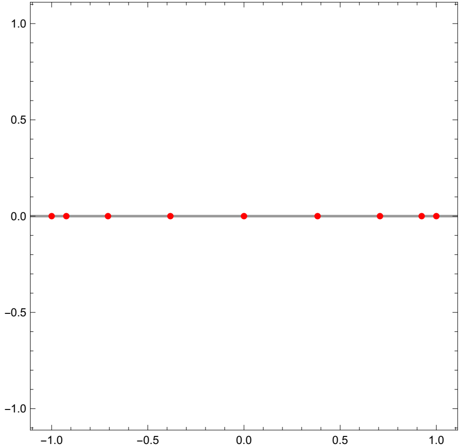
Clenshaw–Curtis nodes in 1d



Clenshaw–Curtis nodes in 1d

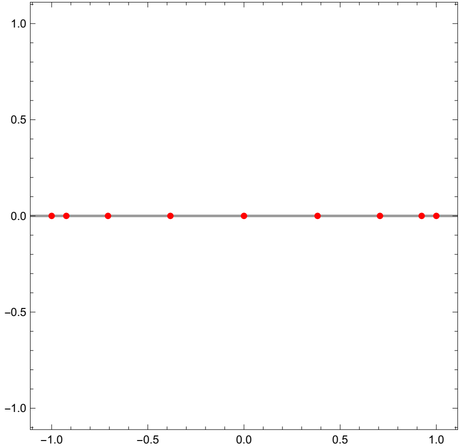


Clenshaw–Curtis nodes in 1d



9-point Clenshaw–Curtis rule on $[-1, 1]$

Clenshaw–Curtis nodes in 1d



9-point Clenshaw–Curtis rule on $[-1, 1]$

- The sequence of $1, 3, 5, \ldots, 2^{n-1} + 1$ -point CC nodes is important (a nested sequence of nodes).

The Clenshaw–Curtis quadrature rule is an interpolatory (but not Gaussian) quadrature rule in the interval $\mathcal{I} = [-1, 1]$ with the weight $W(x) = 1$.

- An n-point CC rule integrates exactly polynomials up to degree $n - 1$.
- For non-polynomials, CC has comparable accuracy to Gaussian quadratures for the same number of points (it is enough that the Chebyshev approximation of the integrand is rapidly converging).

It is convenient to choose a sequence of CC rules with $(m_i)_{i=1}^{\infty} = (1, 3, 5, 9, 13, \ldots)$ points to make the subsequent CC rules nested. Here, $m_1 = 1$ and $m_i = 2^{i-1} + 1$, $i > 1$. Then the nodes $(x_j^i)_{j=1}^{m_i}$ and weights $(w_j^i)_{j=1}^{m_i}$ of an m_i-point CC rule have explicit formulae

$$x_j^i = -\cos\left(\frac{\pi(j-1)}{m_i - 1}\right), \quad j \in \{1, \ldots, m_i\}$$

$$w_j^i = w_{m_i+1-j}^i = \begin{cases} \frac{1}{m_i(m_i-2)}, & j \in \{1, m_i\} \\ \frac{2}{m_i-1}\left(1 - \frac{\cos(\pi(j-1))}{m_i(m_i-2)} - 2\sum_{k=1}^{(m_i-3)/2} \frac{1}{4k^2-1}\cos\left(\frac{2\pi k(j-1)}{m_i-1}\right)\right) & \text{otherwise.} \end{cases}$$

An $\mathcal{O}(n \log n)$ FFT-based implementation available, e.g., at `https://www.mathworks.com/matlabcentral/fileexchange/6911-fast-clenshaw-curtis-quadrature?focused=5058689&tab=function`.

Miscellaneous nested quadrature rules:

- The Gauss–Patterson rules are a nested variant of the Gauss–Legendre rule with $\mathcal{I} = [-1, 1]$ and $W(x) = 1$. The degree of exactness is between $n - 1$ and $2n - 1$. Cf., e.g., `https://people.sc.fsu.edu/~jburkardt/f_src/patterson_rule/patterson_rule.html`.
- The Genz–Keister rules are a nested variant of the Gauss–Hermite rule with $\mathcal{I} = \mathbb{R}$ and $W(x) = \mathrm{e}^{-x^2}$. The degree of exactness is between $n - 1$ and $2n - 1$. Cf., e.g., `https://people.sc.fsu.edu/~jburkardt/f_src/sandia_rules/sandia_rules.html`.

Generating these rules is, in general, very difficult and in practice one needs to resort to using mathematical tables to obtain their nodes and weights.

In many cases, it is desirable to use nested univariate rules in order to generate efficient sparse grid quadrature rules.

Tensor products

Let E be an arbitrary normed space equipped with the norm $\|\cdot\|_E$ and let $T\colon E \to \mathbb{R}$ be a bounded, linear functional. The *operator norm* of T is defined by

$$\|T\| := \sup_{\|x\|_E \leq 1} |Tx|.$$

Multi-index notation:
We use the convention $0 \in \mathbb{N}$. If $\alpha \in \mathbb{N}^d$, then we refer to its j^{th} coordinate universally as α_j. Let $\beta \in \mathbb{N}^d$. We write $\alpha \geq \beta$ if $\alpha_j \geq \beta_j$ for all $j = 1, ..., d$. We define additionally the shorthand $\mathbf{1} = (1, ..., 1) \in \mathbb{N}^d$. We define the following multi-index norms

$$|\alpha|_1 = \sum_{i=1}^{d} \alpha_i \quad \text{and} \quad |\alpha|_\infty = \max_{1 \leq i \leq d} \alpha_i$$

and introduce the following convention for the mixed derivative operator:

$$\frac{\partial^\alpha}{\partial x^\alpha} = \frac{\partial^{|\alpha|_1}}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}, \quad x = (x_1, ..., x_d) \in \mathbb{R}^d.$$

We turn our attention to the function spaces

$$H^r(\Omega)=\left\{f:\Omega\to\mathbb{R};\ \frac{\partial^\alpha f(x)}{\partial x^\alpha} \text{ exists and is bounded in } \Omega \text{ for all } |\alpha|_\infty\leq r\right\}$$

for a fixed region $\varnothing\neq\Omega\subseteq\mathbb{R}^d$. We call r the *regularity* of functions in $H^r(\Omega)$ and accompany these function spaces with the respective norms

$$||f||_{H^r(\Omega)}=\max_{\substack{\alpha\in\mathbb{N}^d\\|\alpha|_\infty\leq r}}\sup\left\{\left|\frac{\partial^\alpha f(x)}{\partial x^\alpha}\right|;\ x\in\Omega\right\}.$$

Suppose that $\varnothing \neq \Omega \subseteq \mathbb{R}^{d_1}$ and $\varnothing \neq \Xi \subseteq \mathbb{R}^{d_2}$ and let $S\colon H^r(\Omega) \to \mathbb{R}$ and $T\colon H^r(\Xi) \to \mathbb{R}$ be functionals. Suppose additionally that they admit to representations

$$Sf = \sum_{i=1}^{m} a_i f(x_i) \quad \text{and} \quad T\tilde{f} = \sum_{i=1}^{n} b_i \tilde{f}(y_i)$$

for a selection of positive weights $(a_i)_{i=1}^m$ and $(b_i)_{i=1}^n$ and vectors $(x_i)_{i=1}^m$ and $(y_i)_{i=1}^n$ in the domains Ω and Ξ. Now $\Omega \times \Xi \subseteq \mathbb{R}^{d_1+d_2}$ and the *tensor product* of S and T is the linear functional $S \otimes T\colon H^r(\Omega \times \Xi) \to \mathbb{R}$ defined by setting

$$(S \otimes T)f = \sum_{i=1}^{m} \sum_{j=1}^{n} a_i b_j f(x_i, y_j).$$

Let $\Omega_j \neq \varnothing$ be regions in Euclidean spaces $\mathbb{R}^{d_j}$ and let $T_j \colon H^r(\Omega_j) \to \mathbb{R}$ be functionals for $j = 1, 2, 3, \ldots$ such that

$$T_j f = \sum_{i=1}^{m_j} w_i^{(j)} f(x_i^{(j)}),$$

where $(w_i^{(j)})_{i=1}^{m_j}$ are positive weights and $(x_i^{(j)})_{i=1}^{m_j}$ is a sequence of vectors in Ω_j.

We define the following shorthand notation:

$$\bigotimes_{i=1}^{1} T_i := T_1 \quad \text{and} \quad \bigotimes_{i=1}^{n} T_i := \left(\bigotimes_{i=1}^{n-1} T_i \right) \otimes T_n \quad \text{for } n = 2, 3, 4, \ldots .$$

By induction with respect to n it is easy to see that $T_1 \otimes \cdots \otimes T_n$ defines a linear functional $H^r(\Omega_1 \times \cdots \times \Omega_n) \to \mathbb{R}$ such that

$$\bigotimes_{i=1}^{n} T_i f = \sum_{i_1=1}^{m_1} \cdots \sum_{i_n=1}^{m_n} w_{i_1}^{(1)} \cdots w_{i_n}^{(n)} f(x_{i_1}^{(1)}, \ldots, x_{i_n}^{(n)}) \quad \text{for } n = 1, 2, 3, \ldots .$$

Properties of tensor products of linear functionals

- Noncommutative: generally $S \otimes T \neq T \otimes S$.
- Associative: $(S \otimes T) \otimes R = S \otimes (T \otimes R)$.
- Distributive: $(S + T) \otimes R = S \otimes R + T \otimes R$.

Theorem

Let T_i be quadrature operators. Then

$$\left\| \bigotimes_{i=1}^{n} T_i \right\| = \prod_{i=1}^{n} \| T_i \|$$

in their respective operator norms.

Example: tensor product quadrature

Consider the problem of approximating

$$\mathcal{I}^d f := \int_{[0,1]^d} f(y_1, \ldots, y_d)\, \mathrm{d}y_1 \cdots \mathrm{d}y_d.$$

Suppose that we have a “good” univariate quadrature rule that satisfies

$$\mathcal{Q}^1 f := \sum_{i=1}^{n} w_i f(x_i) \approx \int_0^1 f(x)\, \mathrm{d}x.$$

Applying this good univariate quadrature rule component-wise to this multidimensional integral yields

$$
\begin{aligned}
\mathcal{I}^d f &= \int_{[0,1]^d} f(y_1, \ldots, y_d)\,\mathrm{d}y_1 \cdots \mathrm{d}y_d \\
&\approx \int_{[0,1]^{d-1}} \sum_{i_d=1}^{n} w_{i_d} f(y_1, \ldots, y_{d-1}, x_{i_d})\,\mathrm{d}y_1 \cdots \mathrm{d}y_{d-1} \\
&\approx \int_{[0,1]^{d-2}} \sum_{i_{d-1}=1}^{n} \sum_{i_d=1}^{n} w_{i_{d-1}} w_{i_d} f(y_1, \ldots, y_{d-2}, x_{i_{d-1}} x_{i_d})\,\mathrm{d}y_1 \cdots \mathrm{d}y_{d-2} \\
&\;\;\vdots \\
&\approx \sum_{i_1=1}^{n} \cdots \sum_{i_d=1}^{n} w_{i_1} \cdots w_{i_d} f(x_{i_1}, \ldots, x_{i_d}) = \bigotimes_{i=1}^{d} \mathcal{Q}^1 f.
\end{aligned}
$$

The tensor product quadrature rule has good approximation properties, but it requires evaluating the integrand f over the point set

$$\{(x_{i_1}, \dots, x_{i_d}) \mid 1 \le i_1, \dots, i_d \le n\}$$

which contains n^d elements. Curse of dimensionality.

Smolyak's sparse grids

Definition (Smolyak quadrature rule)

Let $(U_i)_{i=1}^{\infty}$ be a sequence of univariate quadrature rules in the interval $\varnothing \neq I \subseteq \mathbb{R}$. We introduce the *difference operators* by setting

$$\Delta_0 = 0, \quad \Delta_1 = U_1 \quad \text{and} \quad \Delta_{i+1} = U_{i+1} - U_i \quad \text{for } i = 1, 2, 3, \ldots .$$

The *Smolyak quadrature rule* of order k in the hyperrectangle $I^d = I \times \cdots \times I$ is the operator

$$\mathcal{Q}_k^d = \sum_{\substack{|\alpha|_1 \leq k \\ \alpha \in \mathbb{N}^d}} \bigotimes_{i=1}^{d} \Delta_{\alpha_i}. \tag{2}$$

The tensor product $\Delta_{\alpha_1} \otimes \cdots \otimes \Delta_{\alpha_d}$ in the summand of (2) vanishes whenever $\alpha_i = 0$ for some index i. In the sequel we always assume that $\alpha \geq \mathbf{1}$ and hence $k \geq d$.

Example

This example illustrates an important point about sparse grids: they generally contain negative weights as well as positive weights!

$$
\begin{aligned}
\mathcal{Q}_3^2 &= \sum_{\substack{|\alpha|_1 \le 3 \\ \alpha \in \mathbb{N}^2}} \bigotimes_{i=1}^{2} \Delta_{\alpha_i} \\
&= \Delta_1 \otimes \Delta_1 + \Delta_2 \otimes \Delta_1 + \Delta_1 \otimes \Delta_2 \\
&= U_1 \otimes U_1 + (U_2 - U_1) \otimes U_1 + U_1 \otimes (U_2 - U_1) \\
&= U_2 \otimes U_1 + U_1 \otimes U_2 - U_1 \otimes U_1.
\end{aligned}
$$

If we choose the midpoint rule in $[0,1]$ as the basis of the Smolyak quadrature, i.e., $U_1 f = f\left(\frac{1}{2}\right)$ and $U_2 f = \frac{1}{2} f\left(\frac{1}{4}\right) + \frac{1}{2} f\left(\frac{3}{4}\right)$, then

$$
\mathcal{Q}_3^2 = \frac{1}{2} f\left(\frac{1}{4}, \frac{1}{2}\right) + \frac{1}{2} f\left(\frac{3}{4}, \frac{1}{2}\right) + \frac{1}{2} f\left(\frac{1}{2}, \frac{1}{4}\right) + \frac{1}{2} f\left(\frac{1}{2}, \frac{3}{2}\right) - f\left(\frac{1}{2}, \frac{1}{2}\right).
$$

Important properties

In the case $d = 1$, we obtain

$$\mathcal{Q}_k^1 = \sum_{i=1}^{k} \Delta_i = U_1 + (U_2 - U_1) + \ldots + (U_k - U_{k-1}) = U_k \quad \forall k \geq 1.$$

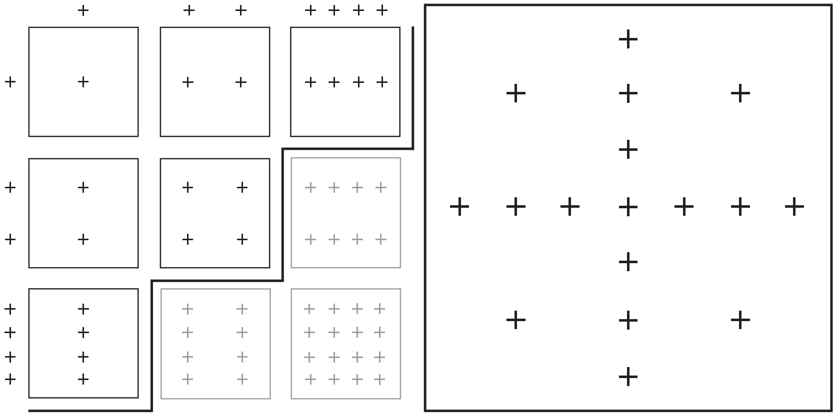
We can directly apply properties of univariate quadrature rules to the Smolyak rule in the one-dimensional case. This makes properties of the Smolyak rule easy to prove by dimension-wise induction.

Using the difference operators defined above, we can write the tensor product quadrature operator $U_k \otimes \cdots \otimes U_k$ of order k in the form

$$\bigotimes_{i=1}^{d} U_k = \left(\sum_{\alpha_1=0}^{k} \Delta_{\alpha_1}\right) \otimes \cdots \otimes \left(\sum_{\alpha_d=0}^{k} \Delta_{\alpha_d}\right) = \sum_{\alpha_1=0}^{k} \cdots \sum_{\alpha_d=0}^{k} \bigotimes_{i=1}^{d} \Delta_{\alpha_i}$$

$$= \sum_{\substack{|\alpha|_\infty \le k \\ \alpha \in \mathbb{N}^d}} \bigotimes_{i=1}^{d} \Delta_{\alpha_i}.$$

The Smolyak quadrature rule can be considered as a delayed sum of the ordinary tensor product operator $U_k \otimes \cdots \otimes U_k$.



Left: Product grids $X_{i_1} \times X_{i_2}$ such that $\#X_k = 2^{k-1}$ and $|(i_1, i_2)|_\infty \leq 3$.
Right: The grid corresponding to rule $\mathcal{Q}_4^2$ is the set
$\bigcup\{X_{i_1} \times X_{i_2};\ |(i_1, i_2)|_1 \leq 4\}$

Characterizations and the combination method

Dimension recursion is a quintessential property of the Smolyak rule. Dimensionally recursive formulations allow us to perform the dimension-wise induction step and prove properties of the Smolyak rule.

Proposition (Dimension recursion)

Let $k \geq d \geq 2$. Then

$$\mathcal{Q}_k^d = \sum_{\substack{|\alpha|_1 \leq k-1 \\ \alpha \in \mathbb{N}^{d-1},\ \alpha \geq \mathbf{1}}} \left(\bigotimes_{i=1}^{d-1} \Delta_{\alpha_i} \right) \otimes U_{k-|\alpha|_1} = \sum_{i=d-1}^{k-1} \mathcal{Q}_i^{d-1} \otimes \Delta_{k-i}.$$

Proof. We start by proving the first equality. Let us denote the index set

$$\mathscr{I}(k,d) = \{\alpha \in \mathbb{N}^d;\ |\alpha|_1 \leq k \text{ and } \alpha \geq \mathbf{1}\}.$$

It is easy to check that the following recursion relation is valid for $k \geq d \geq 2$:

$$\mathscr{I}(k,d) = \{(\alpha, j) \in \mathbb{N}^d;\ \alpha \in \mathscr{I}(k-1, d-1) \text{ and } 1 \leq j \leq k - |\alpha|_1\}.$$

The first equality can now be proved by writing the summation index set of the Smolyak rule recursively and utilizing the distributive property of the tensor product. In this way we attain

$$
\begin{aligned}
\mathcal{Q}_k^d &= \sum_{\substack{|\alpha|_1 \leq k \\ \alpha \in \mathbb{N}^d,\ \alpha \geq \mathbf{1}}} \bigotimes_{i=1}^{d-1} \Delta_{\alpha_i} = \sum_{\substack{|\alpha|_1 \leq k-1 \\ \alpha \in \mathbb{N}^{d-1},\ \alpha \geq \mathbf{1}}} \sum_{j=1}^{k-|\alpha|_1} \left(\bigotimes_{i=1}^{d-1} \Delta_{\alpha_i} \right) \otimes \Delta_j \\
&= \sum_{\substack{|\alpha|_1 \leq k-1 \\ \alpha \in \mathbb{N}^{d-1},\ \alpha \geq \mathbf{1}}} \left(\bigotimes_{i=1}^{d-1} \Delta_{\alpha_i} \right) \otimes \sum_{j=1}^{k-|\alpha|_1} \Delta_j \\
&= \sum_{\substack{|\alpha|_1 \leq k-1 \\ \alpha \in \mathbb{N}^{d-1},\ \alpha \geq \mathbf{1}}} \left(\bigotimes_{i=1}^{d-1} \Delta_{\alpha_i} \right) \otimes U_{k-|\alpha|_1},
\end{aligned}
$$

where the last equality follows from the telescoping property.

To prove the second equality, we use the first one to obtain

$$\begin{aligned}
\mathcal{Q}_k^d &= \sum_{\substack{|\alpha|_1 \le k-1 \\ \alpha \in \mathbb{N}^{d-1},\ \alpha \ge \mathbf{1}}} \left(\bigotimes_{i=1}^{d-1} \Delta_{\alpha_i} \right) \otimes U_{k-|\alpha|_1} \\
&= \sum_{j=d-1}^{k-1} \sum_{\substack{|\alpha|_1 = j \\ \alpha \in \mathbb{N}^{d-1},\ \alpha \ge \mathbf{1}}} \left(\bigotimes_{i=1}^{d-1} \Delta_{\alpha_i} \right) \otimes U_{k-j} \\
&= \sum_{j=d-1}^{k-1} \sum_{\substack{|\alpha|_1 = j \\ \alpha \in \mathbb{N}^{d-1},\ \alpha \ge \mathbf{1}}} \left(\bigotimes_{i=1}^{d-1} \Delta_{\alpha_i} \right) \otimes \sum_{\ell=j}^{k-1} \Delta_{k-\ell} \\
&= \sum_{j=d-1}^{k-1} \sum_{\ell=j}^{k-1} \sum_{\substack{|\alpha|_1 = j \\ \alpha \in \mathbb{N}^{d-1},\ \alpha \ge \mathbf{1}}} \left(\bigotimes_{i=1}^{d-1} \Delta_{\alpha_i} \right) \otimes \Delta_{k-\ell}.
\end{aligned}$$

On the previous slide, we obtained

$$\mathcal{Q}_k^d = \sum_{j=d-1}^{k-1} \sum_{\ell=j}^{k-1} \sum_{\substack{|\alpha|_1=j \\ \alpha\in\mathbb{N}^{d-1},\ \alpha\geq\mathbf{1}}} \left(\bigotimes_{i=1}^{d-1} \Delta_{\alpha_i} \right) \otimes \Delta_{k-\ell}.$$

Changing the order of the first two summation signs nets us

$$\mathcal{Q}_k^d = \sum_{\ell=d-1}^{k-1} \left(\sum_{j=d-1}^{\ell} \sum_{\substack{|\alpha|_1=j \\ \alpha\in\mathbb{N}^{d-1},\ \alpha\geq\mathbf{1}}} \bigotimes_{i=1}^{d} \Delta_{\alpha_i} \right) \otimes \Delta_{k-\ell} = \sum_{\ell=d-1}^{k-1} \mathcal{Q}_\ell^{d-1} \otimes \Delta_{k-\ell}$$

proving the claim. □

The next formula is computationally useless, but it plays a large role in the theoretical derivation of the combination method.

Lemma

Let $\alpha \in \mathbb{N}^d$ and $\alpha \geq \mathbf{1}$. Then

$$\bigotimes_{i=1}^{d} \Delta_{\alpha_i} = \sum_{\substack{\gamma \in \{0,1\}^d \\ \alpha - \gamma \geq \mathbf{1}}} (-1)^{|\gamma|_1} \bigotimes_{i=1}^{d} U_{\alpha_i - \gamma_i}.$$

Proof. By dimension-wise induction. In the base case $d = 1$, we need only to verify the two possible cases:

$$\begin{aligned} \Delta_1 &= U_1 = (-1)^0 U_{1-0}; \\ \Delta_i &= U_i - U_{i-1} = (-1)^0 U_{i-0} + (-1)^1 U_{i-1}, \quad i \geq 2. \end{aligned}$$

Next we suppose that the claim holds for some $d \geq 1$. Let $\alpha \in \mathbb{N}^{d+1}$ and $\alpha \geq \mathbf{1}$. If we first assume that $\alpha_{d+1} \neq 1$, then we get by direct computation

$$\begin{aligned}\sum_{\substack{\gamma\in\{0,1\}^{d+1}\\ \alpha-\gamma\geq\mathbf{1}}} (-1)^{|\gamma|_1} \bigotimes_{i=1}^{d+1} U_{\alpha_i-\gamma_i} &= \sum_{\substack{\gamma\in\{0,1\}^{d}\\ \alpha-\gamma\geq\mathbf{1}}} (-1)^{|\gamma|_1+0} \left(\bigotimes_{i=1}^{d} U_{\alpha_i-\gamma_i}\right) \otimes U_{\alpha_{d+1}-0} \\ &\quad + \sum_{\substack{\gamma\in\{0,1\}^{d}\\ \alpha-\gamma\geq\mathbf{1}}} (-1)^{|\gamma|_1+1} \left(\bigotimes_{i=1}^{d} U_{\alpha_i-\gamma_i}\right) \otimes U_{\alpha_{d+1}-1} \\ &= \sum_{\substack{\gamma\in\{0,1\}^{d}\\ \alpha-\gamma\geq\mathbf{1}}} (-1)^{|\gamma|_1} \left(\bigotimes_{i=1}^{d} U_{\alpha_i-\gamma_i}\right) \otimes \Delta_{\alpha_{d+1}}\end{aligned}$$

since $\Delta_{\alpha_{d+1}} = U_{\alpha_{d+1}} - U_{\alpha_{d+1}-1}$.

On the last slide, we obtained

$$\sum_{\substack{\gamma\in\{0,1\}^{d+1}\\ \alpha-\gamma\geq\mathbf{1}}} (-1)^{|\gamma|_1} \bigotimes_{i=1}^{d+1} U_{\alpha_i-\gamma_i} = \sum_{\substack{\gamma\in\{0,1\}^{d}\\ \alpha-\gamma\geq\mathbf{1}}} (-1)^{|\gamma|_1} \left(\bigotimes_{i=1}^{d} U_{\alpha_i-\gamma_i}\right) \otimes \Delta_{\alpha_{d+1}}$$

under the assumption that $\alpha_{d+1} \neq 1$.

The induction hypothesis implies that

$$\sum_{\substack{\gamma\in\{0,1\}^{d}\\ \alpha-\gamma\geq\mathbf{1}}} (-1)^{|\gamma|_1} \left(\bigotimes_{i=1}^{d} U_{\alpha_i-\gamma_i}\right) = \left(\bigotimes_{i=1}^{d} \Delta_{\alpha_i}\right)$$

as desired.

If $\alpha_{d+1} = 1$, then we substitute $U_{\alpha_{d+1}-1} = 0$ in the computations above and arrive at the same conclusion. This proves the claim. □

A (computationally) useful characterization is given by the following theorem.

Theorem (Combination method)

Let U_i be univariate quadrature rules in the interval $\varnothing \neq I \subseteq \mathbb{R}$ and suppose that $k \geq d$. Then

$$\mathcal{Q}_k^d = \sum_{\substack{\max\{d,k-d+1\}\leq|\alpha|_1\leq k \\ \alpha\in\mathbb{N}^d,\ \alpha\geq\mathbf{1}}} (-1)^{k-|\alpha|_1} \binom{d-1}{k-|\alpha|_1} \bigotimes_{i=1}^{d} U_{\alpha_i}.$$

Proof. The previous lemma immediately yields

$$\mathcal{Q}_k^d = \sum_{\substack{|\alpha|_1 \le k \\ \alpha \in \mathbb{N}^d}} \sum_{\substack{\gamma \in \{0,1\}^d \\ \alpha - \gamma \ge \mathbf{1}}} (-1)^{|\gamma|_1} \bigotimes_{i=1}^d U_{\alpha_i - \gamma_i}.$$

Changing the order of summation we get

$$\mathcal{Q}_k^d = \sum_{\gamma \in \{0,1\}^d} \sum_{\substack{|\alpha|_1 \le k \\ \alpha \in \mathbb{N}^d,\ \alpha - \gamma \ge \mathbf{1}}} (-1)^{|\gamma|_1} \bigotimes_{i=1}^d U_{\alpha_i - \gamma_i}.$$

Change of variable: $\beta = \alpha - \gamma$. The restrictions now are $\beta \ge \mathbf{1}$ and $|\beta|_1 \le k - |\gamma|_1$. Changing the order of summation:

$$\mathcal{Q}_k^d = \sum_{\substack{|\beta|_1 \le k \\ \beta \in \mathbb{N}^d,\ \beta \ge \mathbf{1}}} \sum_{\substack{\gamma \in \{0,1\}^d \\ |\gamma|_1 \le k - |\beta|_1}} (-1)^{|\gamma|_1} \bigotimes_{i=1}^d U_{\beta_i}.$$

The interior sum can be simplified as

$$\sum_{\substack{\gamma\in\{0,1\}^d\\ |\gamma|_1\le k-|\beta|_1}} (-1)^{|\gamma|_1} = \sum_{i=0}^{\min\{d,k-|\beta|_1\}} (-1)^i \sum_{\substack{\gamma\in\{0,1\}^d\\ |\gamma|_1=i}} 1$$

$$= \sum_{i=0}^{\min\{d,k-|\beta|_1\}} (-1)^i \#\{\gamma\in\{0,1\}^d;\ |\gamma|_1=i\}$$

$$= \sum_{i=0}^{\min\{d,k-|\beta|_1\}} (-1)^i \binom{d}{i}$$

$$= (-1)^{k-|\beta|_1}\binom{d-1}{k-|\beta|_1},$$

proving the assertion. □

End of part 1

Next week:

- Convergence rates for sparse grids.
- Dimension adaptive formulations (replacing the index set in the Smolyak construction with a more general formulation).
- Programming sparse grids.