3.5 The Smolyak method

The Smolyak method is a deterministic method that has been used for various applications (under different names), not only for numerical cubature. The idea is starting from a 1D-method to construct an efficient nD-method using a clever setting of the grid points. In this regard, the Smolyak method can be seen as a special case of a QMC method. This is also known as *sparse grids* which is also a well-known method used for the numerical solution of partial differential equations.

Definition 3.5.1 *Let* $L_i[f] := \sum_{\nu=1}^{n_1} c_{\nu,i} f(x_{\nu,i}),\ x_{\nu,i} \in \mathbb{R}\ ,\ i = 1, \ldots, d,$ *be a linear functional, then*

$$(L_1 \otimes \ldots \otimes L_d)[f] := \sum_{\nu_1=1}^{n_1} \cdots \sum_{\nu_d=1}^{n_d} c_{\nu_1,1} \ldots c_{\nu_d,d}\ f(x_{\nu_1,1}, \ldots, x_{\nu_d,d})$$

is called tensor product *of the operators* $L_1, \ldots, L_d$.

Obviously, the standard product in formula (3.1.1) is of this form, i.e.,

$$Q^{[d]}_{n^d}[F] = (\underbrace{Q_n \otimes \ldots \otimes Q_n}_{d-\text{times}})[F]\ .$$

Remark 3.5.2 *As already mentioned above, the concept of Smolyak can be applied to many problems having a product-structure. Here we only look at the particular application to quadrature, resp. cubature.*

Definition 3.5.3 *Let* $Q^{(1)}, Q^{(2)}, \ldots,$ *be a sequence of quadrature-formulae with* n_i *quadrature points and* $Q^{(0)}[f] = 0$ *(i.e.* $n_0 = 0$*) and set*

$$\Delta^{(i)} := Q^{(i+1)} - Q^{(i)}\ . \tag{3.5.1}$$

Then

$$Q(k,d) := \sum_{|i| \le k} \Delta^{(i_1)} \otimes \cdots \otimes \Delta^{(i_d)}, \qquad i = (i_1, \ldots, i_d), \tag{3.5.2}$$

is called k-th Smolyak-Quadrature-Formula*, where, as usual, we set*

$$|i| := \sum_{\nu=1}^{d} i_\nu.$$

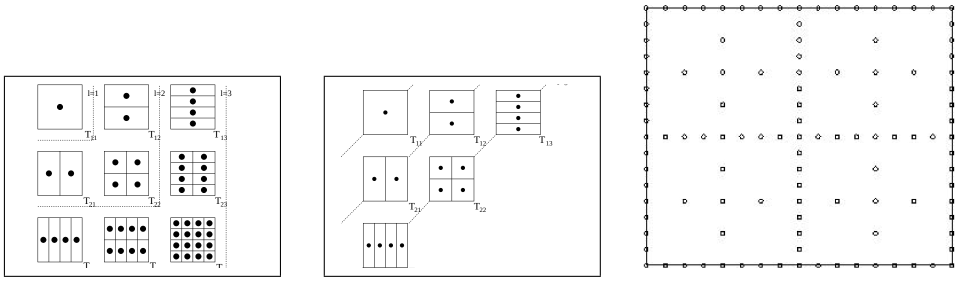

Figure 3.3: Idea for the building of a sparse grid (first two figures) and one particular example of a sparse grid.

This is shown in Figure 3.3 above.

Definition 3.5.4 *A sequence $Q^{(1)}, Q^{(2)}, \ldots$ of quadrature formulae with respect to the quadrature points $X^{(i)}$ is called* nested, *if $X^{(i)} \subseteq X^{(i+1)}$ holds for the sampling points $X^{(i)}, i = 1, 2, \ldots$.*

To analyze the error of a Smolyak formula, we define the following space of functions of mixed maximal smoothness order $r \in \mathbb{N}$ on $\mathbb{R}^d$

$$\mathcal{F}_d^r := \left\{ g : \mathbb{R}^d \to \mathbb{R} : \|g\|_r := \left\| \frac{\partial^{|i|} g}{\partial x_1^{i_1} \ldots \partial x_d^{i_d}} \right\| < \infty \ , \quad \text{if } i_\nu \leq r \right\}.$$

For a linear functional $L \in L(\mathcal{F}_d^r, \mathbb{R})$, consider the standard operator norm, namely

$$\|L\|_r := \sup_{0 \neq f \in \mathcal{F}_d^r} \frac{|L[f]|}{\|f\|_r} \ .$$

Now we start with 1D quadrature formulae satisfying the following error estimate

$$|R_n^{(i)}[f]| \leq \frac{c_r}{n_i^r} \|f\|_r \ , \ f \in \mathcal{F}_1^r \ , \tag{3.5.3}$$

see Theorem 3.1.3.

Theorem 3.5.5 *Let $Q^{(1)}, Q^{(2)}, \ldots$ be nested such that (3.5.3) holds and the number of knots can be bounded by*

$$a 2^i \leq n_i \leq A 2^i \ . \tag{3.5.4}$$

Then,

$$|R(k,d)[f]| \le C_{d,r} \frac{(\log n(k,d))^{(d-1)(r+1)}}{n(k,d)^r} \|f\|_r \ , \tag{3.5.5}$$

holds for all $f \in \mathcal{F}_d^r$, where $n(k,d)$ is the number of quadrature points of $Q(k,d)$ and the constant $C_{d,r}$ does not *depend on f.*

For the proof we again need some preparations:

Lemma 3.5.6 *Let L_i be linear functionals on $\mathcal{F}_1^r$, then*

$$\|L_1 \otimes \cdots \otimes L_d\|_r = \|L_1\|_r \cdots \|L_d\|_r. \tag{3.5.6}$$

Proof: We proceed by induction over d. For $d = 1$ nothing has to be done. For $d \ge 2$ and $f \in \mathcal{F}_d^r$, we consider the following function

$$g := (L_2 \otimes \cdots \otimes L_d)[f] \in \mathcal{F}_1^r,$$

since one component is left free.
If the functions have the representation

$$L_i h = \sum_{\nu_i=1}^{n_i} C_{\nu_i,i} \ h(x_{\nu_i,i}),$$

we easily obtain

$$\begin{aligned}
\frac{\partial}{\partial x_1} g(x_1) &= \sum_{\nu_2=1}^{n_2} \cdots \sum_{\nu_d=1}^{n_d} c_{\nu_2,2} \dots c_{\nu_d,d} \ \frac{\partial}{\partial x_1} f(x_1, x_{\nu_2,2}, \dots, x_{\nu_d,d}) \\
&= (L_2 \otimes \cdots \otimes L_d) \left[\frac{\partial}{\partial x_1} f(x_1, \cdot, \dots, \cdot) \right] (x_{\nu_2,2}, \dots, x_{\nu_d,d}) \ ,
\end{aligned}$$

where $h = \frac{\partial}{\partial x_1} f(x_1, \cdot, \dots, \cdot) \in \mathcal{F}_{d-1}^r$ with $\|h\|_r \le \|f\|_r$. Then, using the induction hypothesis, we get

$$\|g\|_r \le \|(L_2 \otimes \cdots \otimes L_d)\|_r \|f\|_r = \|L_2\|_r \cdots \|L_d\|_r \|f\|_r.$$

Now, we get

$$|(L_1 \otimes \cdots \otimes L_d)[f]| = |L_1[g]| \le \|L_1\|_r \ \|g\|_r \le (\|L_1\|_r \cdots \|L_d\|_r) \ \|f\|_r \ .$$

This finally implies

$$\|L_1 \otimes \cdots \otimes L_d\|_r \leq \|L_1\|_r \cdots \|L_d\|_r \ . \tag{3.5.7}$$

Using the definition of the operator norm, we get that for any $\varepsilon > 0$, there exists a function $f_i = f_i(\varepsilon)$ such that

$$L_i[f_i] \geq \|L_i\|_r \ \|f_i\|_r(1-\varepsilon).$$

Using these functions f_i for all i,, we set

$$f(u_1, \ldots, u_d) := f_1(u_1) \cdots f_d(u_d) = (f_1 \otimes \cdots \otimes f_d)(u_1, \ldots, u_d),$$

thus

$$\|f\|_r = \|f_1\|_r \cdots \|f_d\|_r$$

and

$$\begin{aligned} |(L_1 \otimes \cdots \otimes L_d)[f]| &= |L_1[f_1] \cdots L_d[f_d]| \\ &\geq \|L_1\|_r \cdots \|L_d\|_r \|f_1\|_r \cdots \|f_d\|_r (1-\varepsilon)^d \ , \end{aligned}$$

which yields the assertion with (3.5.7) if we consider $\varepsilon \to 0+$. $\square$

The next result will be needed in order to estimate the number of cubature points which in turns is required to analyze the rate of convergence.

Lemma 3.5.7 *Under the hypothesis of Theorem 3.5.5 we have*

$$n(k,d) \leq A^d \, 2^{k+d} \binom{d+k-1}{d-1}.$$

Proof: Let $X(k,d)$ be the cubature points of $Q(k,d)$, $X^{(i)}$ the cubature points of $Q^{(i)}$ as before and $Y^{(i)}$ the cubature points of $\Delta^{(i)} = Q^{(i+1)} - Q^{(i)}$ in (3.5.1). Because of the nestedness we have

$$\#Y^{(i)} = \#\Delta^{(i)} = \#(Q^{(i+1)} - Q^{(i)}) = \#X^{(i+1)} \ ,$$

thus by the bound on n_i in (3.5.4) and the nestedness of the cubature points

$$\begin{aligned} \#X(k,d) &= \# \left(\sum_{|i|=k} Y^{(i_1)} \otimes \cdots \otimes Y^{(i_d)} \right) \\ &= \sum_{|i|=k} (\#X^{(i_1+1)}) \cdots (\#X^{(i_d+1)}) \\ &= \sum_{|i|=k} n_{i_1+1} \cdots n_{i_d+1} \\ &\leq \sum_{|i|=k} A \cdot 2^{i_1+1} \cdot \cdots \cdot A 2^{i_d+1} = \sum_{|i|=k} A^d 2^{|i|+d}. \end{aligned}$$

Because of

$$\#\{(i_1,\ldots,i_d)\in\mathbb{N}^d : |i|=k\} = \binom{d+k-1}{d-1} = \binom{d+k-1}{k} \tag{3.5.8}$$

(for a proof see below) we conclude

$$\#X(k,d) \le A^d 2^{k+d}\binom{d+k-1}{d-1}.$$

It remains to prove (3.5.8). We first show that

$$\sum_{n=0}^{k}\binom{n}{d} = \binom{k+1}{d+1} \tag{3.5.9}$$

for all $d \ge 1$ by induction. For $k=0$ the claim holds because of $\binom{0}{d} = \binom{1}{d+1} = 0$ for all $d \ge 1$. For $k \ge 1$, we conclude by the induction hypothesis

$$\sum_{n=0}^{k+1}\binom{n}{d} = \binom{k+1}{d} + \sum_{n=0}^{k}\binom{n}{d} = \binom{k+1}{d} + \binom{k+1}{d+1} = \binom{k+2}{d+1}$$

so that (3.5.9) is shown.

Now let $N_k^d := \#\{(i_1,\ldots,i_d)\in\mathbb{N}^d : |i|=k\}$, then obviously we have

$$N_k^1 = \#\{(k)\} = 1 = \binom{1+k-1}{1-1} = \binom{k}{0} = 1$$

and again by induction

$$N_k^d = \sum_{m=0}^{k} N_{k-m}^{d-1} = \sum_{m=0}^{k} N_m^{d-1} = \sum_{m=0}^{k}\binom{d-1+m-1}{d-2} = \binom{d+k-1}{d-1},$$

which proves (3.5.8) in view of (3.5.9). $\square$

Now one final auxiliary result in preparation for the proof of the main result.

Lemma 3.5.8 *Under the hypotheses of Theorem 3.5.5 we have*

$$\|R(k,d)\|_r \le \tilde{C}_r 2^{-r(k+d)}(1+2^r)^{d-1}\binom{d+k}{d-1}.$$

Proof: Again by induction we obtain for a multi-index $i = (i_1, \dots, i_d)$

$$\begin{aligned} Q(k, d+1) &= \sum_{|i| \le k} \left(\Delta^{(i_1)} \otimes \dots \otimes \Delta^{(i_d)} \otimes \sum_{\nu=0}^{k-|i|} \Delta^{(\nu)} \right) \\ &= \sum_{|i| \le k} \left(\Delta^{(i_1)} \otimes \dots \otimes \Delta^{(i_d)} \otimes Q^{(k+1-|i|)} \right) \end{aligned}$$

thus

$$I_{d+1} - Q(k, d+1) = (I_d - Q(k,d)) \otimes I_1 + \sum_{|i| \le k} \Delta^{(i_1)} \otimes \dots \otimes \Delta^{(i_d)} \otimes (I_1 - Q^{(k+1-|i|)}).$$

Because of Lemma 3.5.6, (3.5.3) and (3.5.4) we have by the triangle inequality

$$\begin{aligned} \|\Delta^{(i_\nu)}\|_r &= \|Q^{(i_\nu+1)} - Q^{(i_\nu)}\|_r \le \|Q^{(i_\nu+1)} - I_1\|_r + \|Q^{(i_\nu)} - I_1\|_r \\ &= \|R^{(i_\nu+1)}_{n_{i_\nu+1}}\|_r + \|R^{(i_\nu)}_{n_{i_\nu}}\|_r \\ &\le \frac{c_r}{n^r_{i_\nu+1}} + \frac{c_r}{n^r_{i_\nu}} \\ &\le \frac{c_r}{a^r}(2^{-r(i_\nu+1)} + 2^{-ri_\nu}) = \frac{c_r}{a^r} 2^{-r(i_\nu+1)}(1+2^r). \end{aligned}$$

Next using $\|I_1\|_r = 1$ we have using the triangle inequality and Lemma 3.5.6

$$\begin{aligned} \|R(k, d+1)\|_r &= \|I_{d+1} - Q(k, d+1)\|_r \\ &\le \|R(k,d)\|_r + \sum_{|i| \le k} \|\Delta^{(i_1)} \otimes \dots \otimes \Delta^{(i_d)} \otimes R(k - |i|, 1)\|_r \\ &\le \|R(k,d)\|_r + \sum_{|i| \le k} \|\Delta^{(i_1)}\|_r \dots \|\Delta^{(i_d)}\|_r \|R(k-|i|, 1)\|_r \\ &\lesssim \|R(k,d)\|_r + \underbrace{\sum_{|i| \le k} (1+2^r)^d \underbrace{2^{-r(|i|+d)} 2^{-r(k+1-|i|)}}_{=2^{-r(k+d+1)}}}_{=(1+2^r)^d 2^{-r(k+d+1)} \binom{d+k}{d}}. \end{aligned}$$

Now, finally

$$
\begin{aligned}
\|R(k,d)\|_r &\lesssim \sum_{m=0}^{d-1} 2^{-r(k+d+1)} \underbrace{(1+2^r)^m}_{\leq (1+2^r)^{d-1}} \binom{m+k}{m} \\
&\leq 2^{-r(k+d+1)} (1+2^r)^{d-1} \underbrace{\sum_{m=0}^{d-1} \binom{m+k}{m}}_{\binom{d+k}{d-1}},
\end{aligned}
$$

which completes the proof. $\square$

Now we have all tools at hand and come to the proof of the main result in this section.

Proof of Theorem 3.5.5: Set $q := d+k$. According to Lemma 3.5.7 we have for the number of cubature points

$$
n(k,d) \leq A^d 2^{k+d} \binom{q-1}{d-1} \leq A^d 2^{k+d} \frac{q^{d-1}}{(d-1)!}\ , \tag{3.5.10}
$$

thus in the worst case $q \sim \log n$ and $n \sim 2^{k+d} \frac{q^{d-1}}{(d-1)!}$, or, equivalently $2^{-(k+d)} \sim n^{-1}(\log n)^{d-1}$. Now it results from Lemma 3.5.8:

$$
\begin{aligned}
\|R(k,d)\|_r &\leq \tilde{C}_r (1+2^r)^{d-1} 2^{-r(k+d)} \binom{q}{d-1} \\
&= C_{r,d} \underbrace{2^{-(k+d)(r+1)}}_{\sim \left(\frac{(\log n)^{d-1}}{n}\right)^{r+1}} \underbrace{2^{(k+d)} \binom{q}{d-1}}_{\sim n} \\
&\lesssim C_{r,d} \frac{(\log n)^{(d-1)(r+1)}}{n^r},
\end{aligned}
$$

which proves the desired statement. $\square$

So far, the Smolyak method can be based upon any sequence of univariate quadrature formulae as long as the assumptions of Theorem 3.5.5 are satisfied. That still leaves some freedom to choose particular formulae, i.e., to choose particular sequences of quadrature points (knots). Let us now describe some well-known and widely used sequences.

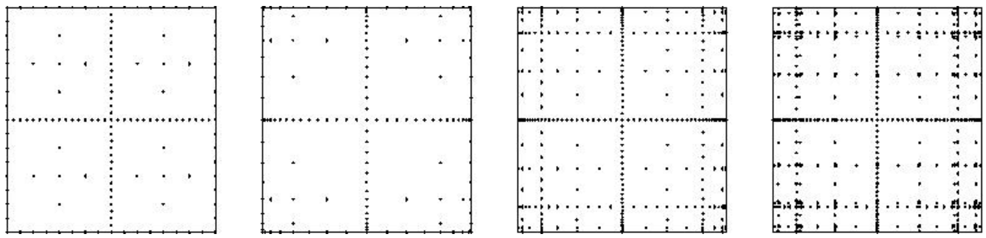

Figure 3.4: Sparse grids based on the trapezoidal, the Clenshaw-Curtis, Patterson and Gauss-Legendre rules, see also Table 3.1. This picture was taken from the homepage of T. Gerstner, Univ. Bonn.

3.5.1 Equidistant Points

Typically, one uses Newton–Cotes formulae in 1D, in the most simple situation, this is just the trapezoidal rule. Subdividing $[0, 1]$ into r_i subintervals results in

$$n_i = r_i + 1$$

knots. The Degree of Exactness (DoE) is then also $n + 1$.

3.5.2 Gauß–Points

As well-known, we obtain the optimal DoE of $2n_i - 1$ with n_i knots.

3.5.3 The Clenshaw-Curtis grids

hese are widely used grids that have been introduced already in 1960. The CC-grids correspond to the settings

$$n_1 = 1 \ , \ n_k = 2^{k-1} + 1 \quad \text{for } k \geq 2,$$

[3]. For the cubature points $X^{CC}(k, 2)$ one typically chooses the roots of the Chebyshev orthogonal polynomials.

One further example is the Konrad-Patterson sequence, which uses the Stieltjes quadrature points. In Figure 3.4, we show the sparse grid points induced by the different constructions and Table 3.1 gives a summary of these.

Name	Abscissas	DoE
Newton-Cotes	equidistant	$n_i + 1$
Chenshow-Curtis	Chebyshev	$n_i - 1$
Patterson (1968)	Stieltjes	$\frac{3}{2}n_i - 1$
Gauss	Legendre	$2n_i - 1$

Table 3.1: Univariate quadrature formulae used for sparse grids construction.