Sparse grid quadrature on products of spheres

Paul Leopardi

Mathematical Sciences Institute, Australian National University.
For presentation at ANU 2012
Joint work with Markus Hegland, ANU.

21 May 2012

AUSTRALIAN RESEARCH COUNCIL
Centre of Excellence for Mathematics
and Statistics of Complex Systems

ARC

ANU
THE AUSTRALIAN NATIONAL UNIVERSITY

Topics

- Applications of integration on products of spheres

- Quadrature on spaces of functions

- Weighted tensor product spaces

- Dimension adaptive sparse grid quadrature

- Lattice-constrained knapsack problems

- Numerical results

Applications of integration on products of spheres

Applications include quantum mechanics, and transport and multiple scattering problems in various areas, including acoustics, optical scattering, and neutron transport.

One prototypical problem is scattering by a sequence of objects, with some decay after each object.

This can be modelled using a multiple integral of a function on a product of spheres.

(Zakowicz at al. 2003; Sato 1988; Altmann 1988; Kaplan et al. 2001; Vineyard 1954)

An abstract look at quadrature

We start with some normed linear space $\mathcal{F}$ of real valued functions defined on some domain $\mathcal{D}$.

To make things simpler and more concrete, suppose $\mathcal{D}$ is a compact subset of some $\mathbb{R}^{(s+1)}$, with a probability measure μ, meaning $\mu(\mathcal{D}) = 1$.

The integral Int is a linear functional on $\mathcal{F}$, that is, an element of the dual space $\mathcal{F}^*$.

For $f \in \mathcal{F}$,

$$\langle \mathrm{Int}, f \rangle := \int_{\mathcal{D}} f(x) \, d\mu(x).$$

An abstract look at quadrature (cont.)

In this framework, a quadrature rule is a finite linear combination of point evaluation functionals, which approximates the integral $\mathbf{Int}$.

For $f \in \mathcal{F}$,

$$\langle Q, f \rangle := \sum_{i=1}^{n} w_i f(x_i).$$

For this to make sense, each point evaluation functional must be well defined. This constrains our choice of the space $\mathcal{F}$.

The theory turns out to be simpler if we choose $\mathcal{F}$ to be a *reproducing kernel Hilbert space* (RKHS) of functions on $\mathcal{D}$.

An RKHS on $\mathcal{D}$

$\mathcal{F}$ is a reproducing kernel Hilbert space of functions $f : \mathcal{D} \to \mathbb{R}$ with inner product $\langle \cdot, \cdot \rangle_{\mathcal{F}}$ if $\mathcal{F}$ is a Hilbert space with this inner product *and* there is some kernel $K : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ such that for all $x \in \mathcal{D}$, the function k_x defined by $k_x(y) := K(x, y)$ satisfies

$$k_x \in \mathcal{F}, \quad \text{and, for all } f \in \mathcal{F}, \quad \langle k_x, f \rangle_{\mathcal{F}} = f(x).$$

Thus each k_x is the representer in the sense of Riesz of the point evaluation functional $\langle K_x, f \rangle := f(x)$, for every point of $\mathcal{F}$, and each of these functionals is bounded and therefore continuous.

An RKHS on $\mathcal{D}$ of functions with mean zero

Let $\mathcal{D} \subset \mathbb{R}^{s+1}$ be a compact s-dimensional manifold with probability measure μ, and let $\mathcal{H}$ be a reproducing kernel Hilbert space (RKHS) of functions $f : \mathcal{D} \to \mathbb{R}$, such that

$$\int_{\mathcal{D}} f(x) \, d\mu(x) = 0 \text{ for all } f \in \mathcal{H},$$

with kernel $\mathcal{K} : \mathcal{D} \times \mathcal{D} \to \mathbb{R}$ such that for all $x \in \mathcal{D}$, the function k_x defined by $k_x(y) := \mathcal{K}(x, y)$ satisfies

$$k_x \in \mathcal{H}, \quad \text{and, for all } f \in \mathcal{H}, \quad \langle k_x, f \rangle_{\mathcal{H}} = f(x).$$

(Hickernell and Woźniakowski 2001; Sloan and Woźniakowski 2001; Kuo and Sloan, 2005)

The weighted space $\mathcal{H}^\gamma$

For $0 < \gamma \leqslant 1$, extend $\mathcal{H}$ into the space $\mathcal{H}^\gamma$ of all functions of the form

$$g = a1 + f,$$

where $1(x) := 1$, $a \in \mathbb{R}$, and $f \in \mathcal{H}$, with norm

$$\|g\|_{\mathcal{H}^\gamma}^2 := |a|^2 + \frac{1}{\gamma} \|f\|_{\mathcal{H}}^2.$$

$\mathcal{H}^\gamma$ is an RKHS with reproducing kernel

$$\mathcal{K}^\gamma(x, y) = 1 + \gamma \mathcal{K}(x, y),$$

where $\mathcal{K}$ is the reproducing kernel of $\mathcal{H}$.

(Hickernell and Woźniakowski 2001; Sloan and Woźniakowski 2001; Kuo and Sloan, 2005)

The weighted tensor product space $\mathcal{H}^{d,\gamma}$

Let $\gamma := (\gamma_1, \ldots, \gamma_d)$, with $1 \geqslant \gamma_1 \geqslant \ldots \geqslant \gamma_d > 0$.

On $\mathcal{D}^d$ define the tensor product RKHS

$$\mathcal{H}^{d,\gamma} := \bigotimes_{h=1}^{d} \mathcal{H}^{\gamma_h}.$$

The reproducing kernel of $\mathcal{H}^{d,\gamma}$ is

$$\mathcal{K}^{d,\gamma}(x,y) := \prod_{h=1}^{d} \mathcal{K}^{\gamma_h}(x_h, y_h).$$

(Hickernell and Woźniakowski 2001; Sloan and Woźniakowski 2001; Kuo and Sloan, 2005)

Quadrature rules on $\mathcal{H}^{d,\gamma}$

For $\{x_1, \ldots, x_n\} \subset \mathcal{D}^d$, the quadrature rule

$$\langle Q, f \rangle := \sum_{i=1}^{n} w_i f(x_i)$$

is a continuous linear functional on $\mathcal{H}^{d,\gamma}$, satisfying

$$\langle Q, f \rangle = \langle q, f \rangle_{\mathcal{H}^{d,\gamma}},$$

where

$$q := \sum_{i=1}^{n} w_i k_{x_i}^{d,\gamma}, \quad k_{x_i}^{d,\gamma}(y) := \mathcal{K}^{d,\gamma}(x_i, y).$$

(Wasilkowski and Woźniakowski 1999; Hickernell and Woźniakowski 2001)

Optimal quadrature weights on $\mathcal{H}^{d,\gamma}$

The worst case error

$$e(q) := \sup_{\|f\|_{\mathcal{H}^{d,\gamma}} \leqslant 1} \left| \int_{\mathcal{D}^d} f(x) d\mu_d(x) - \langle q, f \rangle_{\mathcal{H}^{d,\gamma}} \right|$$
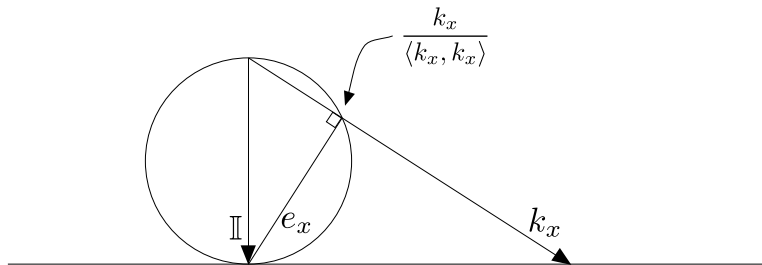
satisfies

$$e(q)^2 = \|1 - q\|_{\mathcal{H}^{d,\gamma}}^2 = \langle 1 - q, 1 - q \rangle_{\mathcal{H}^{d,\gamma}}$$

$$= 1 - 2 \sum_{i=1}^{n} w_i + w^T G w, \quad \text{where}$$

$$G_{i,j} := \langle k_{x_i}^{d,\gamma}, k_{x_j}^{d,\gamma} \rangle_{\mathcal{H}^{d,\gamma}} = \mathcal{K}^{d,\gamma}(x_i, x_j).$$

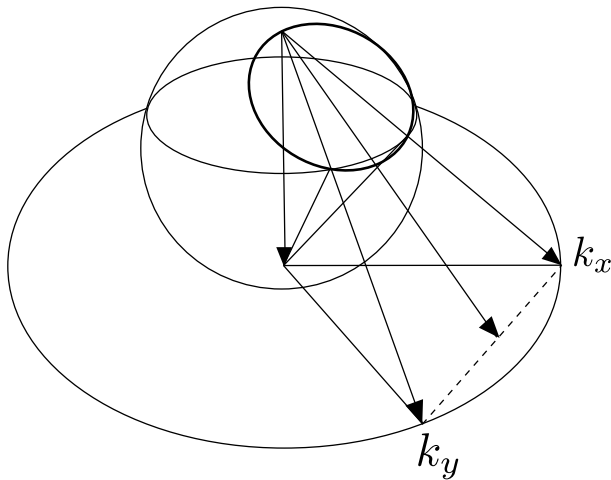The weights w are optimal when $Gw = [1, \ldots, 1]^T$.

(Wasilkowski and Woźniakowski 1999)

Optimal weight for one quadrature point

(Illustration by Osborn, 2009)

Optimal weights for two quadrature points

(Illustration by Osborn, 2009)

Optimal quadrature in $\mathcal{H}^\gamma$

Consider a sequence of quadrature points $x_1, x_2, \ldots \in \mathcal{D}$, and a sequence of positive integers $m_0 < m_1 < \ldots$

For $j \geqslant 0$, let q_j^γ denote the optimal quadrature rule in

$$V_j^\gamma := \mathrm{span}\{k_{x_1}^\gamma, \ldots, k_{x_{m_j}}^\gamma\} \subset \mathcal{H}^\gamma.$$

Define the pair-wise orthogonal spaces U_j^γ by $U_0^\gamma = V_0^\gamma$, by the orthogonal decomposition $V_{j+1}^\gamma = V_j^\gamma \oplus U_{j+1}^\gamma$.

Since the q_j^γ are optimal,

$$\delta_{j+1}^\gamma := q_{j+1}^\gamma - q_j^\gamma \in U_{j+1}^\gamma, \quad \text{and}$$
$$\delta_0^\gamma := q_0^\gamma \in U_0^\gamma = V_0^\gamma.$$

(Gerstner and Griebel 1998; Wasilkowski and Woźniakowski 1999)

Multi-indices and down-sets

Elements of $\mathbb{J} := \mathbb{N}^d$ are treated as multi-indices, with a partial order such that for $i, j \in \mathbb{J}$, $i \leqslant j$ if and only if $i_h \leqslant j_h$ for all h from 1 to d.

For a multi-index $i \in \mathbb{J}$, let $\downarrow i$ denote the down-set of i, defined by $\downarrow i := \{j \in \mathbb{J} \mid j \leqslant i\}$.

Subsets of $\mathbb{J}$ are partially ordered by set inclusion.
For a subset $I \subset \mathbb{J}$, let $\downarrow I$ denote the down-set of I, defined by $\downarrow I := \bigcup_{i \in I} \downarrow i$.

Then $\downarrow I$ is the smallest set $Y \supseteq I$ such that if $i \in Y$ and $j \leqslant i$ then $j \in Y$. Thus $\downarrow \downarrow I = \downarrow I$.

(Davey and Priestley 1990)

Sparse grid quadrature in $\mathcal{H}^{d,\gamma}$

A sparse grid quadrature rule in $\mathcal{H}^{d,\gamma}$ is of the form

$$q \in V_I := \sum_{j \in I} \bigotimes_{h=1}^{d} V_{j_h}^{\gamma_h}$$

for some index set $I \subset \mathbb{J} = \mathbb{N}^d$.

The orthogonal decomposition $V_j^{\gamma} = \bigoplus_{i=1}^{j} U_i^{\gamma}$
yields the multidimensional orthogonal decomposition

$$V_I = \bigoplus_{j \in \downarrow I} \bigotimes_{h=1}^{d} U_{j_h}^{\gamma_h},$$

(Gerstner and Griebel 1998; Wasilkowski and Woźniakowski 1999; Hegland 2003)

Sparse grid quadrature in $\mathcal{H}^{d,\gamma}$ (cont.)

An optimal $q \in V_I$ is

$$q_I = \sum_{j \in \downarrow I} \bigotimes_{h=1}^{d} \delta_{j_h}^{\gamma_h}.$$

Thus both V_I and q_I are obtained in terms of the down-set $\downarrow I$, effectively restricting our choice of I to down-sets.

(Gerstner and Griebel 1998; Hegland 2003)

Our optimization problem

The optimization problem uses the following definitions.

Definition 1

For index $j \in \mathbb{J}$, define

$$\nu_{j_k}^{(k)} := \dim U_{j_k}^{\gamma_{d,k}}, \quad \nu_j := \prod_{k=1}^{d} \nu_{j_k}^{(k)},$$

$$\Delta_j := \bigotimes_{k=1}^{d} \delta_{j_k}^{(k)}, \quad p_j := \|\Delta_j\|^2.$$

Also, define $P := 1 - \epsilon^2$.

Here, j_k is the kth component of the index j.

Our optimization problem (cont.)

Our optimization problem is to

$$\text{Minimize} \quad \nu(I) := \sum_{i \in I} \nu_i$$

$$\text{subject to} \quad I = \downarrow I, \quad p(I) := \sum_{i \in I} p_i \geqslant P, \tag{1}$$

for $P = 1 - \epsilon^2$, where $p_i := \|\Delta_i\|^2 \in \mathbb{R}_+$ and $\nu_i \in \mathbb{N}_+$.

This is a *down-set constrained* binary knapsack problem.

(Gerstner and Griebel 1998; Hegland 2003; Gerstner and Griebel 2003)

A dimension adaptive algorithm to choose I

Algorithm 1: A greedy dimension adaptive algorithm.

Data: accuracy ϵ, incremental rules Δ_j and costs ν_j for $j \in \mathbb{J}$

Result: ϵ approximation q and index set I

$I := \{0\}$;
$q := \Delta_0$;

while $\|1 - q\| > \epsilon$ **do**
    $i := \mathrm{argmax}_j\{\|\Delta_j\|^2 / \nu_j \mid I \cup \{j\} \text{ is a down-set}\}$;
    $I := I \cup \{i\}$; $q := q + \Delta_i$;

(Hegland 2003; Gerstner and Griebel 2003)

The related classical knapsack problem

A widely studied problem in optimisation is the knapsack problem. The knapsack problem related to our problem (1) is to

$$\text{Minimize} \quad \nu(I) := \sum_{i \in I} \nu_i$$

$$\text{subject to} \quad p(I) := \sum_{i \in I} p_i \geqslant P, \tag{2}$$

for $P = 1 - \epsilon^2$, where $p_i := \|\Delta_i\|^2 \in \mathbb{R}_+$ and $\nu_i \in \mathbb{N}_+$.

This is problem (1) without the down-set constraint.

(Dantzig 1957)

Adding the lattice-constraint back in

The following proposition holds.

Proposition 1

If I is a solution of the knapsack problem (2), and satisfies the admissibility condition $I = \downarrow I$, then it is a solution of the optimization problem (1).

This justifies our calling problem (1) a lattice-constrained knapsack problem.

Monotonicity

The sequence $p \in \mathbb{R}_+^{\mathbb{J}}$ is monotonically decreasing
if $i < j$ implies that $p_i \geqslant p_j$.

If $i < j$ implies that $p_i > p_j$,
then $p \in \mathbb{R}_+^{\mathbb{J}}$ is strictly decreasing.

The definitions of "monotonically increasing" and
"strictly increasing" are similar.

Algorithm 1 is optimal, given monotonicity

The following theorem holds.

Theorem 2

If $p \in \mathbb{R}_+^{\mathbb{J}}$ is strictly decreasing and $\nu \in \mathbb{N}_+^{\mathbb{J}}$ is monotonically increasing, each set I given by Algorithm 1 is a solution of the optimization problem (1) *for $P = p(I)$.*

The proof involves formulating the corresponding greedy algorithm for the knapsack problem (2), then showing that if monotonicity holds, this algorithm generates the same sequence of sets I as Algorithm 1. It is known since Dantzig (1957) that the greedy algorithm for the knapsack problem is optimal when $P = p(I)$.

(Dantzig 1957)

The work of Wasilkowski and Woźniakowski

Wasilkowski and Woźniakowski (1999) define a Weighted Tensor
Product algorithm similar to Algorithm 1 and prove a relationship
between cost and error bounds, given the following conditions on
the difference rules δ_j^γ

$$\left\| \delta_j^\gamma \right\|_{\mathbb{H}_{1,\gamma}^{(r)}} \leqslant \sqrt{\gamma} C D^j, \quad \text{for all } j \geqslant 1. \tag{3}$$

$$(j+1) \, D^{j\rho} \leqslant 1, \quad \text{for all } j \geqslant 1, \tag{4}$$

for some $0 < D < 1$ and some positive C and ρ.

(Wasilkowski and Woźniakowski 1999)

The work of Wasilkowski and Woźniakowski

Denote the sequence of rules generated by the algorithm of
Wasilkowski and Woźniakowski by $q_{\epsilon,d}^{(\mathbf{WW})}$. The following theorem
holds as a corollary to a more complicated bound corresponding to
Theorem 21 of Wasilkowski and Woźniakowski (1999).

Theorem 3

*For every positive δ there exists a positive $c(d, \delta)$ such that the
cost of the quadrature rule $q_{\epsilon,d}^{(\mathbf{WW})}$ is bounded by*

$$\mathrm{cost}(q_{\epsilon,d}^{(\mathbf{WW})}) \leqslant c(d, \delta) \left(\frac{1}{\epsilon}\right)^{\rho+\delta}.$$

*Since the sequence of rules $q_{\epsilon,d}^{(\mathbf{DA})}$ generated by Algorithm 1 is
optimal, these rules satisfy the same bound.*

Specifics for $\mathbb{S}^2$ – spherical designs

We use a sequence of rules on a single sphere $\mathbb{S}^2$, which yields "good enough" worst case quadrature error with optimal weights. We choose a sequence of unions of spherical designs with increasing numbers of points, and non-decreasing strengths.

For the unit sphere $\mathbb{S}^2$, a spherical design of strength t and cardinality m is a set of m points $X = \{x_1, \ldots, x_m\} \subset \mathbb{S}^2$ such that the equal weight quadrature rule

$$\langle Q_X, p \rangle := \frac{1}{m} \sum_{h=1}^{m} p(x_h)$$

is exact for all spherical polynomials p of total degree at most t. Spherical designs do not nest, in general. For the numerical examples, the unions of spherical designs have strength at most 1.

(Delsarte, Goethals and Seidel 1977)

Spherical designs used in numerical example

For the numerical examples, a combination of (approximate) extremal (E) and low cardinality (L) spherical designs are used.

These approximate spherical designs were all provided by Womersley.

Index j	0	1	2	3	4	5	6	7	8	9	10	11
Type	L	L	E	L	E	L	E	L	E	L	E	L
Strength t	0	1	1	3	3	7	7	15	15	31	31	63
Card. m	1	2	4	8	16	32	64	129	256	513	1024	2049

For the successive unions of these rules criteria (3) and (4) hold with $D = 2^{-r/2}$, $C \sim 1.453$ as above, and $\rho = 2/r$.

(Chen and Womersley 2006; Womersley 2009)
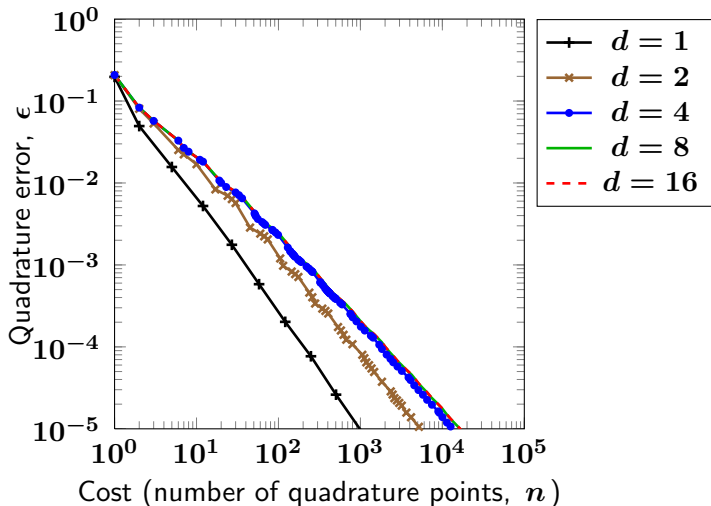
Parameters used in numerical examples

Our numerical examples use $r = 3$ and $\gamma_k = g^k$, for $g = 0.1$, 0.5, and 0.9.

For the DA and WW weighted tensor product algorithms, each program run uses $r = 3$; $g = 0.1$, 0.5, or 0.9; dimension d, from $d = 1$ to 16; maximum 1-norm for indices, typically 20; and maximum number of points, up to $100\,000$.

Typical convergence behaviour

$(\mathbb{S}^2)^4$, $r = 3$, $\gamma_{4,k} = 0.5^k$
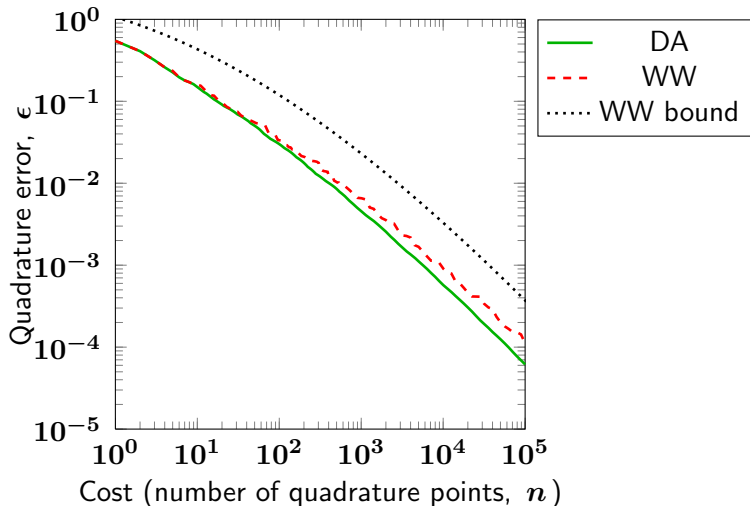
Error of DA rules for $\gamma_{d,k} = 0.1^k$.

$(\mathbb{S}^2)^d$, $d = 1, 2, 4, 8, 16$; $r = 3$

Error of DA rules for $\gamma_{d,k} = 0.9^k$.

$(\mathbb{S}^2)^d$, $d = 1, 2, 4, 8, 16$; $r = 3$

Legend:
- $d = 1$
- $d = 2$
- $d = 4$
- $d = 8$
- $d = 16$

Quadrature error, ϵ

Cost (number of quadrature points, n)