Sparse grid quadrature as a knapsack problem

Paul Leopardi

Mathematical Sciences Institute, Australian National University.
For presentation at HDA 2011, Bonn.
Joint work with Markus Hegland, ANU.

30 June 2011

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

ARC

ANU
THE AUSTRALIAN NATIONAL UNIVERSITY

Topics

- Weighted tensor product spaces
- Dimension adaptive sparse grid quadrature
- Lattice-constrained knapsack problems

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

$$g = a\mathbf{1} + f,$$

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

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

$\mathcal{H}^\gamma$ 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

$$Qf := \sum_{i=1}^{n} w_i f(x_i)$$

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

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

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

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

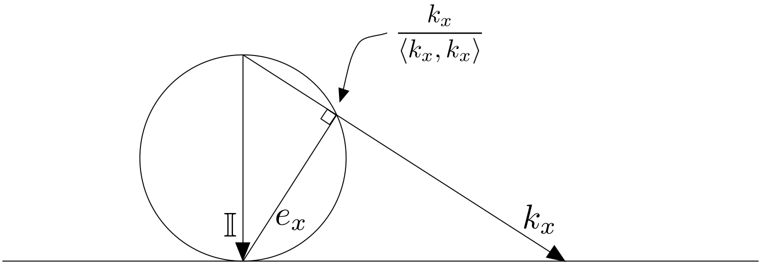
$$= 1 - 2 \sum_{i=1}^{n} w_i + w^T G w, \quad \text{where}$$

$$G_{i,j} := \langle k^{d,\gamma}_{x_i}, k^{d,\gamma}_{x_j} \rangle = \mathcal{K}^{d,\gamma}(x_i, x_j).$$

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

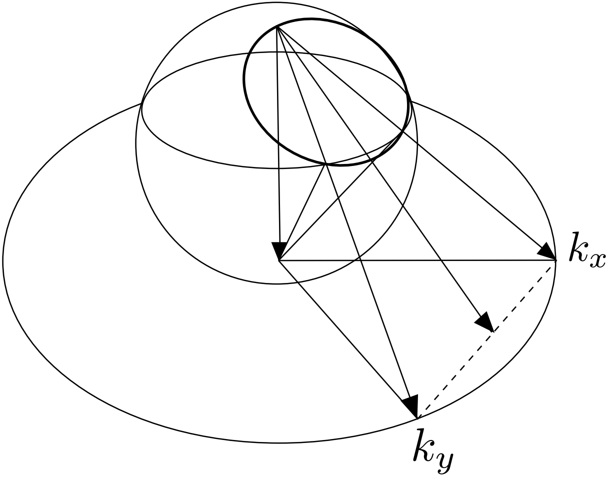
(Wasilkowski and Woźniakowski 1999)

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

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

Define the pair-wise orthogonal spaces U_j^γ by $U_0^\gamma = V_0^\gamma$, and 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 $X \subset \mathbb{J}$, let $\downarrow X$ denote the down-set of X, defined by $\downarrow X := \bigcup_{i \in X} \downarrow i$.

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

(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 $I \subset \mathbb{J} = \mathbb{N}^d$.

(Gerstner and Griebel 1998; Wasilkowski and Woźniakowski 1999)

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

From the orthogonal decomposition $V_j^\gamma = \bigoplus_{i=1}^{j} U_i^\gamma$
one derives the multidimensional orthogonal decomposition

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

An optimal $q \in V_I$ is obtained as

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

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

(Gerstner and Griebel 1998; Hegland 2003)

A dimension adaptive algorithm to choose I

Here, $m^{(h)}_{j_h} := \dim U^{\gamma_h}_{j_h}$, $\delta^{(h)}_{j_h} := \delta^{\gamma_h}_{j_h}$,
$n_j := \prod_{h=1}^{d} m^{(h)}_{j_h}$, $\Delta_j := \bigotimes_{h=1}^{d} \delta^{(h)}_{j_h}$.

Algorithm 1: A dimension adaptive algorithm.

Data: accuracy ϵ, incremental rules Δ_j and costs n_j for $j \in \mathbb{J}$
Result: ϵ approximation q and index set I
$I := \{0\};\quad q := \Delta_0$;
while $\|1 - q\| > \epsilon$ **do**
  $i := \operatorname{argmax}_j \{\|\Delta_j\|^2 / n_j \mid I \cup \{j\}$ is a down-set$\}$;
  $I := I \cup \{i\};\quad q := q + \Delta_i$;

(Hegland 2003; Gerstner and Griebel 2003)

Our optimization problem

Our optimization problem is to maximize

$$p(X) := \sum_{i \in X} p_i,$$

subject to

$$n(\downarrow X) := \sum_{i \in \downarrow X} n_i \leqslant N, \qquad (1)$$

where $p_i := \|\Delta_i\|^2 \in \mathbb{R}_+$ and n_i and N are in $\mathbb{N}_+$, that is, the p_i are positive real numbers and the n_i and N are positive integers.

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

The admissibility condition of problem (1)

The solution of the optimisation problem (1) satisfies an admissibility condition:

Proposition 1

If $\boldsymbol{X}$ *is a solution of the optimisation problem* (1) *then*

$$\boldsymbol{X} = \downarrow \boldsymbol{X}. \tag{2}$$

(Gerstner and Griebel 1998; 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 maximize

$$p(X) = \sum_{i \in X} p_i,$$

subject to

$$n(X) = \sum_{i \in X} n_i \leqslant N, \tag{3}$$

where $p_i \in \mathbb{R}_+$ and n_i and N are in $\mathbb{N}$.

(Dantzig 1957)

A lattice-constrained knapsack problem

We can now formulate a converse of Proposition 1.

Proposition 2

If $\boldsymbol{X}$ *is a solution of the knapsack problem* (3), *and satisfies the admissibility condition* $\boldsymbol{X} = \downarrow \boldsymbol{X}$, *then it is a solution of the optimization problem* (1).

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

Monotonicity

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

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

The definitions of “monotonically increasing” and “strictly increasing” are similar.

Monotonicity implies admissibility

The following proposition holds.

Proposition 3

If $\boldsymbol{p} \in \mathbb{R}_+^{\mathbb{J}}$ *is monotonically decreasing and* $\boldsymbol{n} \in \mathbb{N}_+^{\mathbb{J}}$ *is monotonically increasing, there exists a solution of the knapsack problem* (3) *which also solves the optimization problem* (1).

If $\boldsymbol{p}$ *is strictly decreasing, then any solution of* (3) *is a solution of* (1).

One can therefore use any method to solve the knapsack problem (3), check admissibility (2), and then swap multi-indices to get a solution of problem (1).

Enumeration by decreasing efficiency

The algorithm we adapt is based on efficiency $r_i := p_i/n_i$, and generates the initial values of an enumeration $i^{(t)}$ of $\mathbb{J}$, $t \in \mathbb{N}_+$, satisfying

$$r_{i^{(t)}} \geqslant r_{i^{(t+1)}}.$$

The algorithm recursively generates $i^{(t+1)}$ from $i^{(t)}$, until for some T the condition

$$n(X_{(T)}) \leqslant N < n(X_{(T+1)})$$

holds, where

$$X_{(t)} := \bigcup_{s=1}^{t} i^{(s)}.$$

(Dantzig 1957; Bungartz and Griebel 1999)

Enumeration by decreasing efficiency (cont.)

One then gets

Proposition 4

The construction of $\boldsymbol{i}^{(t)}$ *terminates for some* $t = T$.

Also, if $\boldsymbol{p}$ *is strictly decreasing,* $\boldsymbol{n}$ *is monotonically increasing, and* $\boldsymbol{n}(\boldsymbol{X}_{(T)}) = \boldsymbol{N}$, *then* $\boldsymbol{X}_{(T)}$ *is a solution of problem* (1).

(Dantzig 1957; Bungartz and Griebel 1999)

Finite construction

The construction of the enumeration requires sorting an infinite sequence and is thus not feasible in general, but, in the case where $\boldsymbol{p}$ is monotonically decreasing and $\boldsymbol{n}$ is monotonically increasing, the enumeration can be done recursively in finite time.

In this case $\boldsymbol{r}$ is monotonically decreasing. By construction, $\boldsymbol{r}_{\boldsymbol{i}^{(t)}} \geqslant \boldsymbol{r}_{\boldsymbol{i}^{(t+1)}}$, so the enumeration cannot have $\boldsymbol{i}^{(t)} \geqslant \boldsymbol{i}^{(t+1)}$.

It follows that $\boldsymbol{i}^{(1)} = \mathbf{0}$.

(Hegland 2003; Gerstner and Griebel 2003)

Minimal elements

The element $\boldsymbol{i}$ is a minimal element of a subset of $\mathbb{J}$ if there are no elements $\boldsymbol{j} < \boldsymbol{i}$ in that subset. The minimum is thus with respect to the lattice defined by the partial order in $\mathbb{J}$.

Since $\boldsymbol{i}^{(t)}$ is an enumeration of $\mathbb{J}$, no element occurs twice, and so $\boldsymbol{i}^{(t+1)} \in \boldsymbol{X}^{C}_{(t)} := \mathbb{J} \setminus \boldsymbol{X}_{(t)}$.

Any later element $\boldsymbol{i}^{(t+1+s)}$ in the enumeration cannot be smaller than $\boldsymbol{i}^{(t+1)}$, so $\boldsymbol{i}^{(t+1)}$ is a minimal element of $\boldsymbol{X}^{C}_{(t)}$.

The set $\boldsymbol{M}_{(t)}$ of minimal elements of $\boldsymbol{X}^{C}_{(t)}$ is finite.
One can thus find $\boldsymbol{j} = \boldsymbol{i}^{(t+1)}$ with largest $\boldsymbol{r}_{\boldsymbol{j}}$ in this set.

(Hegland 2003; Gerstner and Griebel 2003)

Construction of set of minimal elements $M_{(t)}$

To construct the set of minimal elements of $X^C_{(t)}$, we define $S(i)$, the forward neighbourhood of $i \in \mathbb{J}$, as

$$S(i) := \{j \in \mathbb{J} \mid i < j \text{ and } (i \leqslant \ell < j \Rightarrow \ell = i)\},$$

that is, $S(i)$ is the set of minimal elements of $\{j \in \mathbb{J} \mid i < j\}$.

Let e be the standard basis of $\mathbb{R}^{\mathbb{J}}$.

To construct $M_{(t)}$, start with
$M_{(1)} = S(i^{(1)}) = S(0) = \{e_1, \ldots, e_d\}$.

Then given $M_{(t-1)}$ and $i^{(t)}$, obtain
$M_{(t)} = \left(M_{(t-1)} \setminus \{i^{(t)}\}\right) \cup S(i^{(t)})$.

(Hegland 2003; Gerstner and Griebel 2003)

Review of the dimension adaptive algorithm

Algorithm 2: A dimension adaptive algorithm.

Data: accuracy ϵ, incremental rules $\boldsymbol{\Delta}_j$ and costs n_j for $j \in \mathbb{J}$
Result: ϵ approximation q and index set I
$I := \{0\}; \quad q := \boldsymbol{\Delta}_0$;
while $\|1 - q\| > \epsilon$ **do**
  $i := \operatorname{argmax}_j \{\|\boldsymbol{\Delta}_j\|^2 / n_j \mid I \cup \{j\}$ is a down-set$\}$;
  $I := I \cup \{i\}; \quad q := q + \boldsymbol{\Delta}_i$;