

# A note on the convergence analysis of a sparse grid multivariate probability density estimator

Stephen G. Roberts<sup>1</sup>

Sarah Bolt<sup>2</sup>

(Received 16 August 2008; revised 25 February 2009)

## Abstract

With the recent growth in volume and complexity of available data has come a renewed interest in the problem of estimating multivariate probability density functions. However, traditional methods encounter the curse of dimensionality (complexity grows exponentially with dimension). Here we provide an outline of a convergence analysis of a sparse grid based probability density estimation, which supports the use of the method for moderately complex (up to 15 dimensions) data sets, as has already been demonstrated for sparse grid quadrature and interpolation.

## Contents

<b>1 Introduction</b>	<b>C859</b>
<b>2 Sparse grid function spaces</b>	<b>C860</b>

---

<http://anziamj.austms.org.au/ojs/index.php/ANZIAMJ/article/view/1472>

gives this article, © Austral. Mathematical Soc. 2009. Published March 14, 2009. ISSN 1446-8735. (Print two pages per sheet of paper.)

<i>1</i>	<i>Introduction</i>	C859
<b>3</b>	<b>Estimation of probability density functions</b>	<b>C864</b>
3.1	Outline of proof . . . . .	C866
3.2	The empirical error . . . . .	C867
	<b>References</b>	<b>C869</b>

# 1 Introduction

A standard method for estimating probability density functions is the simple histogram. This method and various extensions work well for one to three dimensional data, as demonstrated by Scott [5]. This method is based on discretising the various dimensions and forming bins into which the data is associated. Another way of looking at this method is to search for a function from a discretisation space (in this case the piecewise constant functions associated with the bins) which matches the distribution of the data in some averaged sense. For high dimensional problems this leads to impractical methods requiring an astronomical number of bins or equivalently an astronomical number of basis functions in the discretisation space (for a given approximation error). For instance if each of the  $d$  dimensions is discretised into sub-intervals with sides of length  $h$  then the number of bins is  $O(h^{-d})$ . For these piecewise constant spaces, the approximation order of the space is  $O(h)$ . For a 15 dimensional problem, suppose we divide each dimension into just four sub-interval intervals ( $h = 1/4$ ) then the number of bins is  $4^{15} = 1,073,741,824$  (a large number of bins!). The approximation error for such a coarse space is  $O(1/4)$ . Other methods based on kernel estimation can deal with high dimensions, (the size of the discretisation space is the same as the number of data records) but leads to algorithms which depend at least quadratically on the number of data records, at least for high dimensional problems.

In this article we describe a probability density estimation method which is somewhat like the histogram method, in that a grid based discrete approx-

imation space is used, but which can be applied to higher dimensions and also scales linearly with data size, both in storage and execution time. In this direction we investigate generalizations of histograms, in which we use an  $L^2$  projection onto a discrete approximation space to estimate probability density functions. When the discrete space is the space of tensor products of piecewise constant functions we obtain the classical histogram, and if we use sparse grid spaces as described by Zenger [6] and Garcke et al. [4], then we arrive at the sparse grid histograms described in this article.

The outline of our analysis shows, for smooth enough underlying probability density functions, that the approximation error of our sparse grid histogram is  $O(|\log h|^{d-1}h^2 + |\log h|^d n^{-1/2}h^{-1/2})$  when using a classic piece-wise linear sparse grid space of dimension  $O(|\log h|^{d-1}h^{-1})$ . The method has a complexity which scales linearly with  $n$ , the number of data records.

Our analysis supports the use of the method for moderately complex data sets with up to 15 dimensions and millions of data points. This complements the theoretical results on sparse grid quadrature and interpolation as reviewed by Bungartz and Griebel [2]. Practical implementation of our method is an area of current research and will be reported in a subsequent article.

## 2 Sparse grid function spaces

This section provides an introduction to sparse grid functions and, in particular, demonstrates some of the approximation and complexity properties that make them useful for high dimensional approximation. The online tutorial by Garcke [3] and the extensive review by Bungartz and Griebel [2] provides details of the results noted in this section. The results of this section provide a basis for the description of our probability estimation problem in section 3.

First, though, we introduce some standard function space notation and definitions. Following common usage, the space of square integrable functions on  $I^d$ , the  $d$  dimensional unit cube, is denoted  $L^2$ , with associated inner

product denoted  $(\mathbf{u}, \mathbf{v})_{L^2}$  and norm  $\|\mathbf{u}\|_{L^2}$ . The standard Sobolov spaces are

$$H^m = \left\{ \mathbf{u} \in L^2 : \sum_{|\alpha| \leq m} \|D^\alpha \mathbf{u}\|_{L^2} < \infty \right\}$$

where  $\alpha = (\alpha_1, \dots, \alpha_d) \in \mathbb{N}_0^d$  is a  $d$  multi-index,  $D^\alpha$  is the differential operator  $D^\alpha = \partial_1^{\alpha_1} \partial_2^{\alpha_2} \dots \partial_d^{\alpha_d}$  and  $|\alpha| = \sum_{s=1}^d \alpha_s$ . The associated  $H^m$  inner product and norm are  $(\mathbf{u}, \mathbf{v})_{H^m} = \sum_{|\alpha| \leq m} (D^\alpha \mathbf{u}, D^\alpha \mathbf{v})_{L^2}$  and  $\|\mathbf{u}\|_{H^m}^2 = (\mathbf{u}, \mathbf{u})_{H^m}$ .

When working with tensor product spaces, and sparse grid spaces in particular, it is convenient to work with the mixed derivative spaces

$$H_{\text{mix}}^m = \left\{ \mathbf{u} \in L^2 : \sum_{|\alpha|_\infty \leq m} \|D^\alpha \mathbf{u}\|_{L^2} < \infty \right\}$$

where  $|\alpha|_\infty = \max_s \alpha_s$ . Note that the sum is now over all multi-indices such that  $|\alpha|_\infty \leq m$ . The inner product and norm on the mixed derivative space are  $(\mathbf{u}, \mathbf{v})_{H_{\text{mix}}^m} = \sum_{|\alpha|_\infty \leq m} (D^\alpha \mathbf{u}, D^\alpha \mathbf{v})_{L^2}$  and  $\|\mathbf{u}\|_{H_{\text{mix}}^m}^2 = (\mathbf{u}, \mathbf{u})_{H_{\text{mix}}^m}$ .

The one dimensional uniform grid space is generated by sums of translations and dilations of the standard triangular ‘‘hat’’ function

$$\phi(x) = \begin{cases} 1 - |x|, & \text{if } x \in (-1, 1), \\ 0, & \text{otherwise.} \end{cases}$$

We denote translated and dilated versions of the standard hat function by

$$\phi_{j,i}(x) = \phi\left(\frac{x - ih_j}{h_j}\right)$$

where  $h_j = 2^{-j}$ . These functions are naturally associated with a grid on  $[0, 1]$  with a grid size of  $h_j$ . The parameter  $j$  denotes the level.

Consider the index set  $R_j = \{\text{all integers between } 1 \text{ and } 2^j - 1\}$ . The standard homogeneous piece-wise linear finite element space on the interval  $I =$

$[0, 1]$  is  $\mathbb{V}_j = \text{span}\{\phi_{j,i} : i \in \mathbb{R}_j\}$  where the functions are restricted to the unit interval.

Now consider the index sets consisting of the odd indices in  $\mathbb{R}_j$ , that is  $\mathbb{Q}_j = \{\text{all odd integers between } 1 \text{ and } 2^j - 1\}$ . An alternative basis for  $\mathbb{V}_j$ , the hierarchical basis, is the set of basis functions  $\{\phi_{k,i} : k \leq j, i \in \mathbb{Q}_k\}$ . Note that for this basis we have basis functions ranging in level from 1 to  $j$ .

We use tensor products to generalise to multi-dimensions. Consider the multi-index,  $\mathbf{j} = (j_1, \dots, j_d)$ . Note that we use bold to denote multi-indices. We define  $\mathbf{k} \leq \mathbf{j}$  if and only if  $k_s \leq j_s$  for all  $s = 1, \dots, d$ , and in particular  $\mathbf{k} \leq \mathbf{j}\mathbf{1}$  if and only if  $k_s \leq j$  for all  $s = 1, \dots, d$ .

We consider a multi-dimensional grid with varying grid sizes  $h_{j_s}$  in each dimension  $s = 1, \dots, d$ . The associated space of piece-wise multi-linear grid functions is

$$\mathbb{V}_{\mathbf{j}} = \bigotimes_{s=1}^d \mathbb{V}_{j_s}.$$

The space  $\mathbb{V}_{\mathbf{j}}$  is uniform in each dimension, but each dimension can have a different mesh size. A basis for  $\mathbb{V}_{\mathbf{j}}$  is constructed from tensor products of the one dimensional basis functions  $\phi_{j,i}$ . Let

$$\phi_{\mathbf{j}, \mathbf{i}}(\mathbf{x}) = \prod_{s=1}^d \phi_{j_s, i_s}(x_s).$$

Introducing the multi-dimensional index set

$$\mathbb{R}_{\mathbf{j}} = \mathbb{R}_{j_1} \times \dots \times \mathbb{R}_{j_d},$$

then a basis for  $\mathbb{V}_{\mathbf{j}}$  is  $\{\phi_{\mathbf{j}, \mathbf{i}} : \mathbf{i} \in \mathbb{R}_{\mathbf{j}}\}$ . Note that the standard uniform grid space is  $\mathbb{V}_{\mathbf{j}\mathbf{1}}$ .

For the hierarchical decomposition we consider the multi-dimensional index sets

$$\mathbb{Q}_{\mathbf{k}} = \mathbb{Q}_{k_1} \times \dots \times \mathbb{Q}_{k_d}.$$

Then the multi-dimensional hierarchical subspaces are defined by  $\mathbb{W}_{\mathbf{k}} = \text{span}\{\phi_{\mathbf{k},\mathbf{i}} : \mathbf{i} \in Q_{\mathbf{k}}\}$ . The uniform spaces have the decomposition

$$\mathbb{V}_j = \bigoplus_{\mathbf{k} \leq j} \mathbb{W}_{\mathbf{k}}$$

and a corresponding hierarchical basis for  $\mathbb{V}_j$  is  $\{\phi_{\mathbf{k},\mathbf{i}} : \mathbf{k} \leq j, \mathbf{i} \in Q_{\mathbf{k}}\}$ . Now each function  $\mathbf{u} \in \mathbb{V}_j$  is represented by a hierarchical subspace decomposition

$$\mathbf{u} = I_{\mathbb{V}_j} \mathbf{u} = \sum_{\mathbf{k} \leq j} I_{\mathbb{W}_{\mathbf{k}}} \mathbf{u}$$

where the  $I_{\mathbb{W}_{\mathbf{k}}} \mathbf{u}$  are components of  $\mathbf{u}$  in the hierarchial subspace  $\mathbb{W}_{\mathbf{k}}$ . Indeed for general functions  $\mathbf{u} \in H_{\text{mix}}^2$ ,

$$\mathbf{u} = \sum_{\mathbf{k}} I_{\mathbb{W}_{\mathbf{k}}} \mathbf{u}$$

where  $I_{\mathbb{W}_{\mathbf{k}}} \mathbf{u} \in \mathbb{W}_{\mathbf{k}}$  and

$$\|I_{\mathbb{W}_{\mathbf{k}}} \mathbf{u}\|_{L^2} \leq 3^{-d} 2^{-2|\mathbf{k}|} |\mathbf{u}|_{H_{\text{mix}}^2}. \tag{1}$$

Hence, for a function in  $H_{\text{mix}}^2$ , all components  $I_{\mathbb{W}_{\mathbf{k}}} \mathbf{u}$ , with  $|\mathbf{k}| = j$ , are comparable in size. It is then natural to consider the reduced space

$$\mathbb{S}_j = \bigoplus_{|\mathbf{k}| \leq j} \mathbb{W}_{\mathbf{k}}.$$

This is the classic sparse grid function space.

For functions in  $H_{\text{mix}}^2$ , the sparse grid space approximation error is essentially second order,  $\mathcal{O}(h_j^2)$ . Indeed, summing up error terms bounded by equation (1) implies that

$$\|\mathbf{u} - I_{\mathbb{S}_j} \mathbf{u}\|_{L^2} = \mathcal{O}(j^{d-1} 2^{-2j}) |\mathbf{u}|_{H_{\text{mix}}^2} = \mathcal{O}(|\log h_j|^{d-1} h_j^2) |\mathbf{u}|_{H_{\text{mix}}^2}. \tag{2}$$

The details of these results are described by Bungartz and Griebel [1]. For function in  $H^2$ , the approximation error on a uniform grid satisfies the classical second order result,

$$\| \mathbf{u} - \mathbf{I}_{\mathbb{V}_{j1}} \mathbf{u} \|_{L_2} = O(h_j^2) |u|_{H^2}. \tag{3}$$

The dimension of the corresponding full and sparse grid spaces are very different,

$$\dim\{\mathbb{V}_{j1}\} = O(2^{j(d-1)}) = O(h_j^{-d}) \tag{4}$$

and

$$\dim\{\mathbb{S}_j\} = O(j^{d-1}2^j) = O(|\log h_j|^{d-1} h_j^{-1}). \tag{5}$$

Sparse grids provide very similar approximation order as the uniform grid, but the dimension of the sparse grid space (as a finite dimensional vector space) is geometrically smaller.

An alternative view of the sparse grid is that of a union of uniform grids. Sparse grid functions can be represented as a combination of functions on these uniform grids. This representation is known as the combination formula. To be explicit, consider a class of projections  $P_{\mathbb{V}_k}$  onto  $\mathbb{V}_k$  for  $|\mathbf{k}| \leq j$ , such that  $P_{\mathbb{V}_k} P_{\mathbb{V}_m} = P_{\mathbb{V}_k \cap \mathbb{V}_m}$ . Then the associated projection  $P_{\mathbb{S}_j}$  onto the sparse grid  $\mathbb{S}_j$ , is given by the combination formula (described by Garcke [3])

$$P_{\mathbb{S}_j} = \sum_{j-d+1 \leq |\mathbf{k}| \leq j} (-1)^{j-|\mathbf{k}|} \binom{d-1}{j-|\mathbf{k}|} P_{\mathbb{V}_k}. \tag{6}$$

In particular, the projections can be  $L^2$  projections, and indeed in the sequel we will consider  $P_{\mathbb{V}_k}$  to be the  $L^2$  projection onto  $\mathbb{V}_k$ .

### 3 Estimation of probability density functions

Given a set of  $n$  data points  $X_i \in \mathbb{R}^d$  distributed according to a probability density function  $p$ , we want to look for an approximation  $p_j \in \mathbb{S}_j$  which

satisfies the following projection problem in the finite dimensional space  $\mathbb{S}_j$ .

**Problem 1** Find a sparse grid function  $\mathbf{p}_j \in \mathbb{S}_j$  such that

$$(\mathbf{p}_j, \mathbf{v})_{L^2} = \frac{1}{n} \sum_{i=1}^n \mathbf{v}(X_i) \quad \text{for all } \mathbf{v} \in \mathbb{S}_j.$$

The expected value of the right hand side is simply

$$\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n \mathbf{v}(X^i) \right] = (\mathbf{p}, \mathbf{v})_{L^2}.$$

So

$$(\mathbb{E}\mathbf{p}_j, \mathbf{v})_{L^2} = (\mathbf{p}, \mathbf{v})_{L^2}$$

for all  $\mathbf{v} \in \mathbb{S}_j$  and so we expect  $\mathbb{E}\mathbf{p}_j$  to be close to  $\mathbf{p}$ .

**Theorem 1 (Error estimate)** If the  $n$  data points  $X_i \in \mathbb{R}^d$  are distributed according to a probability density function  $\mathbf{p} \in H_{\text{mix}}^2$ , then the solution  $\mathbf{p}_j \in \mathbb{S}_j$  of Problem 1 satisfies

$$\mathbb{E} \|\mathbf{p}_j - \mathbf{p}\|_{L^2} \leq C \left( |\log h_j|^{d-1} h_j^2 \|\mathbf{p}\|_{H_{\text{mix}}^2} + |\log h_j|^d h_j^{-1/2} n^{-1/2} \right).$$

Hence, the expected squared error is bounded by the squared bias plus the variance.

The first term measures the error when using a sparse grid function to approximate an  $H_{\text{mix}}^2$  function and the second term measures the error when using the empirical distribution in place of using the exact distribution to calculate the expected values of the test functions.

For a given number of data points  $n$  we obtain a nearly optimal choice of sparse grid size, by equating the two error terms. Up to a log term, the optimal choice of grid size  $h_j$  is  $n^{-1/5}$  which provides an error bound

$$\mathbb{E} \|\mathbf{p}_j - \mathbf{p}\|_{L^2} \leq C |\log n|^d n^{-2/5} \|\mathbf{p}\|_{H_{\text{mix}}^2}.$$

**Remark 2** For uniform spaces we expect

$$\mathbb{E}\|p_j - p\|_{L^2} \leq C(h_j^2 + h_j^{-d/2}n^{-1/2})\|p\|_{H^2}$$

with an optimal choice of grid size  $h_j$  being  $n^{-1/(4+d)}$ , implying

$$\mathbb{E}\|p_j - p\|_{L^2} \leq Cn^{-2/(4+d)}\|p\|_{H^2}.$$

Note the dependence on dimension, for higher dimensions, the approximation grows progressively worse. This is independent of the fact that the dimension of the uniform grid space is much larger than the corresponding sparse grid space. Of course, this does depend on the much stronger assumption on the smoothness of the underlying density function.

### 3.1 Outline of proof

The rest of the article is an outline of the convergence analysis of the method. The details will appear in a longer article which is in preparation.

We want to estimate

$$\|p_j - p\|_{L^2}^2 = (p_j - p, p_j - p)_{L^2}.$$

As is typical in finite element analysis, we split the expression

$$(p_j - p, p_j - p)_{L^2} = (p_j - p, p_j - P_{S_j}p)_{L^2} + (p_j - p, P_{S_j}p - p)_{L^2}. \quad (7)$$

The first term is bounded by the empirical error presented in subsection 3.2, namely

$$(p_j - p, v)_{L^2} \leq Z2^d |\log h_j|^d h_j^{-1/2} \|v\|_{L^2}$$

for all  $v \in S_j$ , where  $\mathbb{E}Z \leq n^{-1/2}$ . So we obtain

$$\begin{aligned} & |(p_j - p, p_j - P_{S_j}p)_{L^2}| \\ & \leq Z2^d |\log h_j|^d h_j^{-1/2} \|p_j - P_{S_j}p\|_{L^2} \end{aligned}$$

$$\begin{aligned} &\leq CZ^2|\log h_j|^{2d}h_j^{-1} + 1/8\|p_j - P_{S_j}p\|_{L^2}^2 \\ &\leq CZ^2|\log h_j|^{2d}h_j^{-1} + 1/4\|p_j - p\|_{L^2}^2 + 1/4\|p - P_{S_j}p\|_{L^2}^2. \end{aligned}$$

The second term is bounded by the approximation result, equation (2):

$$\begin{aligned} |(p_j - p, P_{S_j}p - p)_{L^2}| &\leq \|p_j - p\|_{L^2}\|p - P_{S_j}p\|_{L^2} \\ &\leq 1/4\|p_j - p\|_{L^2}^2 + \|p - P_{S_j}p\|_{L^2}^2 \\ &\leq 1/4\|p_j - p\|_{L^2}^2 + C|\log h_j|^{2(d-1)}h_j^4\|p\|_{H_{\text{mix}}^2}^2. \end{aligned}$$

Together we have

$$\frac{1}{2}(p_j - p, p_j - p)_{L^2} \leq C\left(|\log h_j|^{2(d-1)}h_j^4\|p\|_{H_{\text{mix}}^2}^2 + Z^2|\log h_j|^{2d}h_j^{-1}\right)$$

which implies that

$$\|p_j - p\|_{L^2} \leq C\left(|\log h_j|^{(d-1)}h_j^2\|p\|_{H_{\text{mix}}^2} + Z|\log h_j|^d h_j^{-1/2}\right).$$

Taking the expected value, provides the result.

### 3.2 The empirical error

Now we estimate  $(p_j - p, v)_{L^2}$  for  $v \in S_j$ . Consider the decomposition of  $v$  using the combination formula, equation (6). Let us write the combination formula as

$$v = \sum_{|\mathbf{k}| \leq j} \beta_{\mathbf{k}} P_{V_{\mathbf{k}}} v. \tag{8}$$

We write each projection as a linear combination of the standard basis functions  $P_{V_{\mathbf{k}}} v = \sum_{\mathbf{i} \leq \mathbf{k}} \gamma_{\mathbf{k}, \mathbf{i}} \phi_{\mathbf{k}, \mathbf{i}}$ . Now

$$\|P_{V_{\mathbf{k}}} v\|_{L^2} \leq \|v\|_{L^2}$$

as  $\mathbf{P}_{\mathbf{V}_k}$  is taken as the  $L^2$  projection. We need to bound the  $L^2$  norm in terms of the basis coefficients. Let  $\boldsymbol{\gamma}_k$  be the vector of coefficients  $\gamma_{k,i}$ ,  $\mathbf{i} \in \mathbf{R}_k$ . Then

$$\|\mathbf{P}_{\mathbf{V}_k} \mathbf{v}\|_{L^2}^2 = \boldsymbol{\gamma}_k^T \mathbf{M}_k \boldsymbol{\gamma}_k$$

where  $\mathbf{M}_k$  is the standard finite element mass matrix for a uniform grid. It is obtained as a tensor product of the corresponding one dimensional matrix  $\mathbf{M}_k$ . The eigenvalues of the one dimension matrix  $\mathbf{M}_k$  lie between  $\frac{1}{4}h_k$  and  $h_k$ , so

$$\left( \sum_{\mathbf{i} \in \mathbf{R}_k} \gamma_{k,i}^2 \right)^{1/2} \leq 2^d h_k^{-1/2} \|\mathbf{P}_{\mathbf{V}_k} \mathbf{v}\|_{L^2} \leq 2^d h_k^{-1/2} \|\mathbf{v}\|_{L^2}. \quad (9)$$

Here we have used  $h_k$  to represent the product  $h_{k_1} \cdots h_{k_d}$ . Now

$$\begin{aligned} (\mathbf{p}_j - \mathbf{p}, \mathbf{P}_{\mathbf{V}_k} \mathbf{v})_{L^2} &= \frac{1}{n} \sum_{i=1}^n \mathbf{P}_{\mathbf{V}_k} \mathbf{v}(\mathbf{X}_i) - (\mathbf{p}, \mathbf{P}_{\mathbf{V}_k} \mathbf{v})_{L^2} \\ &\leq \left( \sum_{\mathbf{i} \in \mathbf{R}_k} \gamma_{k,i}^2 \right)^{1/2} \left( \sum_{\mathbf{i} \in \mathbf{R}_k} \left[ \frac{1}{n} \sum_{i=1}^n \phi_{k,i}(\mathbf{X}_i) - \mathbb{E} \phi_{k,i} \right]^2 \right)^{1/2}. \end{aligned}$$

The first term is bounded by equation (9). The second term is independent of the function  $\mathbf{v}$  and is analyzed as a sum of independent random variables. Let

$$\mathbf{Z}_k = \left( \sum_{\mathbf{i} \in \mathbf{R}_k} \left[ \frac{1}{n} \sum_{i=1}^n \phi_{k,i}(\mathbf{X}_i) - \mathbb{E} \phi_{k,i} \right]^2 \right)^{1/2}.$$

Together we conclude that

$$(\mathbf{p}_j - \mathbf{p}, \mathbf{P}_{\mathbf{V}_k} \mathbf{v})_{L^2} \leq \mathbf{Z}_k 2^d h_k^{-1/2} \|\mathbf{v}\|_{L^2}.$$

Since  $\mathbf{Z}_k^2$  is a sum of the squares of basis functions which are bounded by 1, it follows that

$$\mathbb{E} \mathbf{Z}_k^2 \leq \frac{1}{n} \quad \text{and hence} \quad \mathbb{E} \mathbf{Z}_k \leq \frac{1}{n^{1/2}}.$$

Now we need to add all the contributions from the combination formula. For  $v \in \mathbb{S}_j$ , using equation (8),

$$\begin{aligned} (p_j - p, v)_{L^2} &= (p_j - p, \sum_{|\mathbf{k}| \leq j} \beta_{\mathbf{k}} P_{V_{\mathbf{k}}})_{L^2} \\ &= \sum_{|\mathbf{k}| \leq j} \beta_{\mathbf{k}} (p_j - p, P_{V_{\mathbf{k}}})_{L^2} \\ &\leq 2^d h_j^{-1/2} \|v\|_{L^2} \sum_{|\mathbf{k}| \leq j} Z_{\mathbf{k}} |\beta_{\mathbf{k}}|. \end{aligned}$$

The expected value of  $\sum_{|\mathbf{k}| \leq j} Z_{\mathbf{k}} |\beta_{\mathbf{k}}|$  is bounded by a sum of binomial terms which are bounded by

$$O(j^d n^{-1/2}) = O(|\log h_j|^d n^{-1/2}).$$

Together this leads to

$$\mathbb{E}(p_j - p, v)_{L^2} \leq 2^d |\log h_j|^d h_j^{-1/2} n^{-1/2} \|v\|_{L^2}$$

which concludes the outline of the convergence analysis for our sparse grid probability density estimator.

## References

- [1] Hans-Joachim Bungartz and Michael Griebel. A note on the complexity of solving Poisson's equation for spaces of bounded mixed derivatives. *J. Complexity*, 15(2):167–199, 1999.  
<http://dx.doi.org/10.1006/jcom.1999.0499> C864
- [2] Hans-Joachim Bungartz and Michael Griebel. Sparse grids. *Acta Numerica*, 13:147–269, 2004.  
<http://dx.doi.org/10.1017/S0962492904000182> C860

- [3] J. Garcke. Sparse grid tutorial. <http://www.math.tu-berlin.de/~garcke/paper/sparseGridTutorial.pdf> [29 January, 2009], 2008. C860, C864
- [4] J. Garcke, M. Griebel, and M. Thess. Data mining with sparse grids. *Computing*, 67:225–253, 2001. <http://dx.doi.org/10.1007/s006070170007> C860
- [5] D. Scott. *Multivariate Density Estimation: Theory, Practice, and Visualization*. Wiley, New York, 1992. C859
- [6] C. Zenger. Sparse grids. In W. Hackbusch, editor, *Parallel Algorithms for Partial Differential Equations: Proceedings of the Sixth GAMM-Seminar, Kiel, 1990*, volume 31 of *Notes on Num. Fluid Mech.*, pages 241–251. Vieweg, 1991. C860

## Author addresses

1. **Stephen G. Roberts**, Mathematical Sciences Institute, Australian National University, Canberra, AUSTRALIA.  
<mailto:stephen.roberts@anu.edu.au>
2. **Sarah Bolt**, Mathematical Sciences Institute, Australian National University, Canberra, AUSTRALIA.