Convergence of sparse grid Gaussian convolution approximation for multi-dimensional periodic functions

Simon Hubbert[*], Janin Jäger[†] and Jeremy Levesley[‡]

October 21, 2022

Abstract

We consider the problem of approximating $[0,1]^d$-periodic functions by convolution with a scaled Gaussian kernel. We start by establishing convergence rates to functions from periodic Sobolev spaces and we show that the saturation rate is $\mathcal{O}(h^2)$, where h is the scale of the Gaussian kernel. Taken from a discrete point of view, this result can be interpreted as the accuracy that can be achieved on the uniform grid with spacing h. In the discrete setting, the curse of dimensionality would place severe restrictions on the computation of the approximation. For instance, a spacing of 2^{-n} would provide an approximation converging at a rate of $\mathcal{O}(2^{-2n})$ but would require $(2^n+1)^d$ grid points. To overcome this we introduce a sparse grid version of Gaussian convolution approximation, where substantially fewer grid points are required (from $\mathcal{O}(2^{nd})$ on the full grid to just $\mathcal{O}(2^n n^{d-1})$ on the sparse grid) and show that the sparse grid version delivers a saturation rate of $O(n^{d-1}2^{-2n})$. This rate is in line with what one would expect in the sparse grid setting (where the full grid error only deteriorates by a factor of order n^{d-1}) however the analysis that leads to the result is novel in that it draws on results from the theory of special functions and key observations regarding the form of certain weighted geometric sums.

MSC Class: 42B05; 65D40; 65D15

Keywords: Sparse grids; Gaussian convolution, Approximation of periodic functions

1 Introduction

Many methods that are designed to deliver approximations are based on the convolution of a kernel function with the function being approximated. The general approach involves selecting a suitable integrable function $K : \mathbb{R}^d \to \mathbb{R}$ (*the convolution kernel*) satisfying

$$\int_{\mathbb{R}^d} K(\mathbf{x})d\mathbf{x} = \int_{\mathbb{R}^d} K(x_1,\ldots,x_d)dx_1\cdots dx_d = 1.$$

A scaling vector $\mathbf{h} = (h_1,\ldots,h_d)^T \in \mathbb{R}^d$, with $h_i > 0$ $(1 \le i \le d)$ is then used to define a parameterised family of convolution kernels by

$$K_{\mathbf{h}}(\mathbf{x}) = \frac{1}{h_1\cdots h_d} K\left(\frac{x_1}{h_1},\ldots,\frac{x_d}{h_d}\right).$$

[*]Department of Economics, Mathematics and Statistics, Birkbeck, University of London, WC1H 7HX, UK. s.hubbert@bbk.ac.uk

[†]Corresponding author: Lehrstuhl Numerische Mathematik, Justus-Liebig University, 35392 Giessen, Germany. janin.jaeger@math.uni-giessen.de

[‡]Department of Mathematics, University of Leicester, LE1 7RH, UK. jl1@le.ac.uk

The convolution approximation to a function $f \in L_1(\mathbb{R}^d)$ is defined by

$$\mathcal{C}_{\mathbf{h}}(f)(\mathbf{x}) = (f * K_{\mathbf{h}})(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{y}) K_{\mathbf{h}}(\mathbf{x}-\mathbf{y}) d\mathbf{y}. \tag{1.1}$$

The convolution kernel described above is anisotropic as each direction is allowed to be scaled by its own factor. This is practically useful because a typical data sample of a function will show variety along different directions and so a well designed anisotropic scaling can efficiently capture these features. However, from a theoretical perspective, most convergence results relate to the isotropic scale where each direction is scaled by the same factor $h > 0$. In this case the scaled kernel is $K_h(\mathbf{x}) = h^{-d} K\left(\frac{\mathbf{x}}{h}\right)$ and the corresponding convolution approximation

$$\mathcal{C}_h(f)(\mathbf{x}) = \int_{\mathbb{R}^d} f(\mathbf{y}) K_h(\mathbf{x}-\mathbf{y}) d\mathbf{y} = \frac{1}{h^d} \int_{\mathbb{R}^d} f(\mathbf{y}) K\left(\frac{\mathbf{x}-\mathbf{y}}{h}\right) d\mathbf{y}, \tag{1.2}$$

can be shown to converge to f as $h \to 0$, the convergence being uniform on compact sets, [6] chapter 20, theorem 2. The rate of convergence depends upon the smoothness of f and the polynomial reproduction properties of the underlying kernel. The convolution approximation can be viewed as the continuous counterpart of quasi-interpolation; a discrete method which generates an approximation over the whole of $\mathbb{R}^d$ by linearly combining the values of f sampled at the scaled integer lattice $h\mathbf{Z}^d$ together with the appropriately shifted and scaled kernel function. The classical construction, as for example described in [3], takes the form

$$Q_h(f)(\mathbf{x}) = \sum_{\mathbf{k} \in \mathbf{Z}^d} f(\mathbf{k}h) K\left(\frac{\mathbf{x}}{h} - \mathbf{k}\right), \quad \mathbf{x} \in \mathbb{R}^d, \quad h > 0. \tag{1.3}$$

Following [11] the connection between continuous convolution and discrete quasi-interpolation can be seen if we write

$$\mathcal{C}_h(f)(\mathbf{x}) = \frac{1}{h^d} \sum_{\mathbf{k} \in \mathbf{Z}^d} \int_{h \cdot (\mathbf{k} + [-\frac{1}{2}, \frac{1}{2}]^d)} f(\mathbf{y}) K\left(\frac{\mathbf{x}-\mathbf{y}}{h}\right) d\mathbf{y}.$$

The integrals above are taken over appropriately shifted and scaled versions of the cube $[-\frac{1}{2}, \frac{1}{2}]^d$. If we approximate each integrand by its value at the midpoint of the cube we get

$$\int_{h \cdot (\mathbf{k} + [-\frac{1}{2}, \frac{1}{2}]^d)} f(\mathbf{y}) K\left(\frac{\mathbf{x}-\mathbf{y}}{h}\right) d\mathbf{y} \approx h^d f(\mathbf{k}h) K\left(\frac{\mathbf{x}}{h} - \mathbf{k}\right),$$

and so we have that $\mathcal{C}_h(f)(\mathbf{x}) \approx Q_h(f)(\mathbf{x})$. Quasi-interpolation using Gaussians in one dimension was described in [10].

In this paper we will examine the approximation of $[0,1]^d-$periodic functions by convolution with the multi-dimensional Gaussian kernel. Given the close connection of continuous convolution to quasi-interpolation the results we establish in the continuous setting will serve as a baseline for what should be expected in the discrete case.

We begin in Section 2 by deriving the formula for the Fourier expansion of the pointwise error using the anisotropic scaling of the Gaussian; this result allows us to deduce that convolution approximation is only able to reproduce the constant function. We then analyse the isotropic case in some detail. In this setting we demonstrate that the convergence has a saturation rate of $\mathcal{O}(h^2)$.

In Section 3 we consider the practical issues of employing the discrete (quasi-interpolation) analogue of continuous convolution in high dimensions. Such a recasting involves constructing a full grid in $[0,1]^d$ with an isotropic spacing of $h = \frac{1}{2^n}$, where n is a positive integer. In this setting, the convolution approximation will converge to f at a rate of $\frac{1}{2^{2n}}$, provided f is sufficiently smooth. However, in the discrete setting we are restricted by the curse of dimensionality since the construction of the quasi-interpolant would require $(2^n+1)^d$ evaluations and this is prohibitively large as n grows. In order to overcome this we consider replacing the full-grid approximation with a sparse grid version which is built from a certain linear combination of smaller full grid approximations. Numerical experiments on closely connected methods have been published in [14]. To analyse this theoretically we mimic the approach of Section 2, i.e., we first derive the formula for the Fourier expansion of the pointwise error using the sparse grid convolution approximation. We then investigate the Fourier coefficients of the error expansion and we state the main theorem of the paper, concerning the decay rate of the coefficients. We then establish that, provided f is sufficiently smooth, the sparse grid convolution approximation will converge to f at a rate of $\frac{n^{d-1}}{2^{2n}}$. Section 4 is devoted to the proof of the aforementioned main theorem of the paper.

2 Gaussian convolution approximation

Our choice of convolution kernel is the multi-dimensional Gaussian

$$\Psi(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{1}{2}\mathbf{x}^T\mathbf{x}\right) = \frac{1}{(2\pi)^{\frac{d}{2}}} \exp\left(-\frac{1}{2}\left(\sum_{i=1}^{d} x_i^2\right)\right) = \prod_{i=1}^{d} \psi(x_i),$$

where $\psi : \mathbb{R} \to \mathbb{R}$ is the univariate Gaussian $\psi(x) = (2\pi)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}x^2\right)$. Fourier theory will play an important role in our analysis and we recall that if we let $e_x(z) = \exp(2\pi i x \cdot z)$ then the univariate Fourier transform of ψ is

$$\widehat{\psi}(z) := \int_{-\infty}^{\infty} \psi(x) e_{-x}(z) dx = \exp(-2\pi^2 z^2).$$

Our general aim is to approximate a $[0,1]^d$-periodic function

$$f(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbf{Z}^d} \widehat{f}(\mathbf{k}) e_{\mathbf{k}}(\mathbf{x}) \quad \text{where} \quad e_{\mathbf{y}}(\mathbf{x}) := \exp(2\pi i \mathbf{y}^T\mathbf{x}) = \prod_{i=1}^{d} e_{y_i}(x_i),$$

by the continuous multi-variable convolution

$$\mathcal{C}_{\mathbf{h}}(f) = \int_{\mathbb{R}^d} f(\mathbf{z})\Psi_{\mathbf{h}}(\mathbf{x}-\mathbf{z})d\mathbf{z} = \sum_{\mathbf{j}\in\mathbf{Z}^d} \int_{[0,1]^d} f(\mathbf{z})\Psi_{\mathbf{h}}(\mathbf{x}-\mathbf{z}-\mathbf{j})d\mathbf{z} = \int_{[0,1]^d} f(\mathbf{z})\Phi_{\mathbf{h}}(\mathbf{x}-\mathbf{z})d\mathbf{z},$$

where

$$\Phi_{\mathbf{h}}(\mathbf{x}) = \sum_{\mathbf{j}\in\mathbf{Z}^d} \Psi_{\mathbf{h}}(\mathbf{x}-\mathbf{j}).$$

Now, $\Phi_{\mathbf{h}}(\mathbf{x})$ is $[0,1]^d-$periodic and so has a multi-dimensional Fourier series

$$\Phi_{\mathbf{h}}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbf{Z}^d} \widehat{\Phi_{\mathbf{h}}}(\mathbf{k}) e_{\mathbf{k}}(\mathbf{x}),$$

where

$$\begin{aligned}
\widehat{\Phi_{\mathbf{h}}}(\mathbf{k}) &= \int_{[0,1]^d} \Phi_{\mathbf{h}}(\mathbf{x}) e_{-\mathbf{k}}(\mathbf{x}) d\mathbf{x} \\
&= \int_{[0,1]^d} \left(\sum_{\mathbf{j}\in\mathbf{Z}^d} \Psi_{\mathbf{h}}(\mathbf{x}-\mathbf{j}) \right) e_{-\mathbf{k}}(\mathbf{x}) d\mathbf{x} \\
&= \sum_{\mathbf{j}\in\mathbf{Z}^d} \int_{[0,1]^d} \Psi_{\mathbf{h}}(\mathbf{x}-\mathbf{j}) e_{-\mathbf{k}}(\mathbf{x}) d\mathbf{x} \\
&= \int_{\mathbb{R}^d} \Psi_{\mathbf{h}}(\mathbf{x}) e_{-\mathbf{k}}(\mathbf{x}) d\mathbf{x} = \prod_{i=1}^{d} \int_{-\infty}^{\infty} \psi_{h_i}(x_i) e_{-k_i}(x_i) dx_i = \prod_{i=1}^{d} \widehat{\psi}(h_i k_i).
\end{aligned}$$

Applying the d-dimensional convolution formula for $[0,1]^d-$periodic functions we have:

$$\mathcal{C}_{\mathbf{h}}(f)(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbf{Z}^d} \widehat{f}(\mathbf{k}) \widehat{\Phi_{\mathbf{h}}}(\mathbf{k}) e_{\mathbf{k}}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbf{Z}^d} \widehat{f}(\mathbf{k}) \left(\prod_{i=1}^{d} \widehat{\psi}(h_i k_i) \right) e_{\mathbf{k}}(\mathbf{x}). \tag{2.1}$$

Thus the error in the convolution approximation is

$$E_{\mathbf{h}}(f)(\mathbf{x}) = f(\mathbf{x}) - \mathcal{C}_{\mathbf{h}}(f)(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbf{Z}^d} \widehat{f}(\mathbf{k}) \left(1 - \left(\prod_{i=1}^{d} \widehat{\psi}(h_i k_i) \right) \right) e_{\mathbf{k}}(\mathbf{x}). \tag{2.2}$$

We note that, since $\widehat{\psi}(0) = 1$, the above error representation immediately shows that the convolution reproduces the constant but not any other trigonometric polynomial.

2.1 Convergence with Isotropic scaling

If we consider the isotropic case where the same scale factor h is applied to all coordinate directions then (2.2) can be written as

$$E_h(f)(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbf{Z}^d\setminus\{\mathbf{0}\}} \widehat{f}(\mathbf{k}) \left(1 - e^{-2\pi^2 h^2 \|\mathbf{k}\|^2} \right) e_{\mathbf{k}}(\mathbf{x}). \tag{2.3}$$

The functions we wish to approximate are taken from a periodic Sobolev space

$$\mathcal{N}_\beta = \left\{ f = \sum_{\mathbf{k}\in\mathbf{Z}^d} \widehat{f}(\mathbf{k}) e_{\mathbf{k}} : \|f\|_\beta = \left(\sum_{\mathbf{k}\in\mathbf{Z}^d\setminus\{\mathbf{0}\}} \|\mathbf{k}\|^{2\beta} |\widehat{f}(\mathbf{k})|^2 \right)^{1/2} < \infty \right\}.$$

The Sobolev embedding theorem [2] ensures that if $\beta > \frac{d}{2}$ then all functions in $\mathcal{N}_\beta$ will be continuous. The following result gives error bounds for Gaussian convolution approximation of such functions.

Proposition 2.1. *Let $f \in \mathcal{N}_\beta$, where $\beta > \frac{d}{2}$. Then*

$$\|E_h f\|_\infty \le \|f\|_\beta \cdot \begin{cases} C_1 h^2 & \text{for } \beta > \frac{d}{2}+2; \\ h^2 \left(C_2 \sqrt{\ln\left(\frac{1}{h}\right)} + C_3 \right) & \text{for } \beta = \frac{d}{2}+2; \\ C_4 h^{\beta-\frac{d}{2}} & \text{for } \frac{d}{2} < \beta < \frac{d}{2}+2, \end{cases}$$

where C_i $i = 1,2,3,4$, are positive constants independent of h.

Proof. Using (2.3) together with the elementary bound $1 - e^{-x} < x$ (for $x > 0$), we can deduce that

$$\|E_h f\|_\infty \leq 2\pi^2 h^2 \sum_{\mathbf{k}\in\mathbf{Z}^d\setminus\{\mathbf{0}\}} \|\mathbf{k}\|^2 |\widehat{f}(\mathbf{k})|. \tag{2.4}$$

Suppose that $\beta = \frac{d}{2} + 2 + \alpha$, where $\alpha > 0$, then an application of the Cauchy Schwarz inequality yields

$$\begin{aligned}
\|E_h f\|_\infty &\leq 2\pi^2 h^2 \sum_{\mathbf{k}\in\mathbf{Z}^d\setminus\{\mathbf{0}\}} \frac{1}{\|\mathbf{k}\|^{\frac{d}{2}+\alpha}} \cdot \|\mathbf{k}\|^{\frac{d}{2}+\alpha+2} |\widehat{f}(\mathbf{k})| \\
&\leq 2\pi^2 h^2 \left(\sum_{\mathbf{k}\in\mathbf{Z}^d\setminus\{\mathbf{0}\}} \frac{1}{\|\mathbf{k}\|^{d+2\alpha}} \right)^{\frac{1}{2}} \cdot \left(\sum_{\mathbf{k}\in\mathbf{Z}^d\setminus\{\mathbf{0}\}} \left(\|\mathbf{k}\|^2\right)^{\frac{d}{2}+2+\alpha} |\widehat{f}(\mathbf{k})|^2 \right)^{\frac{1}{2}} \\
&\leq 2\pi^2 C h^2 \|f\|_\beta.
\end{aligned}$$

Now assume that $\beta = \frac{d}{2} + 2 - \alpha$ where $0 < \alpha < 2$. In the following development we will work with a partition of the punctured integer lattice

$$\mathbf{Z}^d \setminus \{\mathbf{0}\} = \underbrace{\left\{ \mathbf{k} \in \mathbf{Z}^d \setminus \{\mathbf{0}\} : 1 \leq \|\mathbf{k}\| < \frac{1}{h} \right\}}_{=\Theta_h} \cup \underbrace{\left\{ \mathbf{k} \in \mathbf{Z}^d : \|\mathbf{k}\| \geq \frac{1}{h} \right\}}_{=\Gamma_h}.$$

Using this we bound the error in two parts as follows

$$\|E_h f\|_\infty \leq \sum_{\mathbf{k}\in\Theta_h} |\widehat{f}(\mathbf{k})| \cdot |1 - e^{-2\pi^2 h^2 \|\mathbf{k}\|^2}| + \sum_{\mathbf{k}\in\Gamma_h} |\widehat{f}(\mathbf{k})| \cdot |1 - e^{-2\pi^2 h^2 \|\mathbf{k}\|^2}|.$$

For the sum over Θ_h we again employ $1 - e^{-x} < x$ and, bounding as before, we conclude that

$$\begin{aligned}
&\sum_{\mathbf{k}\in\Theta_h} |\widehat{f}(\mathbf{k})| \cdot |1 - e^{-2\pi^2 h^2 \|\mathbf{k}\|^2}| \leq 2\pi^2 h^2 \sum_{\mathbf{k}\in\Theta_h} |\widehat{f}(\mathbf{k})| \cdot \|\mathbf{k}\|^2 \\
&= 2\pi^2 h^2 \sum_{\mathbf{k}\in\Theta_h} \|\mathbf{k}\|^{\alpha - \frac{d}{2}} \|\mathbf{k}\|^{\frac{d}{2}+2-\alpha} |\widehat{f}(\mathbf{k})| \\
&\leq 2\pi^2 h^2 \left(\sum_{\mathbf{k}\in\Theta_h} \|\mathbf{k}\|^{2\alpha-d} \right)^{\frac{1}{2}} \left(\sum_{\mathbf{k}\in\Theta_h} \left(\|\mathbf{k}\|^2\right)^{\frac{d}{2}+2-\alpha} |\widehat{f}(\mathbf{k})|^2 \right)^{\frac{1}{2}} \\
&\leq 2\pi^2 h^2 \left(\sum_{\mathbf{k}\in\Theta_h} \left(\frac{1}{h}\right)^{2\alpha-d} \right)^{\frac{1}{2}} \|f\|_{\frac{d}{2}+2-\alpha} \\
&\leq C \cdot 2\pi^2 h^2 \left[\left(\frac{1}{h}\right)^d \left(\frac{1}{h}\right)^{2\alpha-d} \right]^{\frac{1}{2}} \|f\|_{\frac{d}{2}+2-\alpha} \\
&\leq C \cdot 2\pi^2 h^{2-\alpha} \|f\|_{\frac{d}{2}+2-\alpha} = C \cdot 2\pi^2 h^{\beta - \frac{d}{2}} \|f\|_\beta.
\end{aligned}$$

We note that for the case where $\alpha = 0$, corresponding to $\beta = \frac{d}{2} + 2$, the above development can be traced to the third line to yield

$$\sum_{\mathbf{k}\in\Theta_h} |\widehat{f}(\mathbf{k})| \cdot |1 - e^{-2\pi^2 h^2 \|\mathbf{k}\|^2}| \leq 2\pi^2 h^2 \left(\sum_{\mathbf{k}\in\Theta_h} \frac{1}{\|\mathbf{k}\|^d} \right)^{\frac{1}{2}} \|f\|_\beta.$$

Applying the integral test, with a change to polar coordinates, we have

$$\sum_{\mathbf{k}\in\Theta_h} \frac{1}{\|\mathbf{k}\|^d} \leq \int_{1\leq\|\mathbf{x}\|\leq\frac{1}{h}} \frac{d\mathbf{x}}{\|\mathbf{x}\|^d} = C_d \int_1^{\frac{1}{h}} \frac{dr}{r} = C_d \ln\left(\frac{1}{h}\right).$$

In summary, for this part of the sum we can conclude that

$$\sum_{\mathbf{k}\in\Theta_h} |\widehat{f}(\mathbf{k})| \cdot |1 - e^{-2\pi^2 h^2 \|\mathbf{k}\|^2}| \leq \|f\|_\beta \begin{cases} Ch^2\sqrt{\ln\left(\frac{1}{h}\right)} & \text{for} \quad \beta = \frac{d}{2}+2; \\ Ch^{\beta-\frac{d}{2}} & \text{for} \quad \frac{d}{2} < \beta < \frac{d}{2}+2. \end{cases} \tag{2.5}$$

For the sum over Γ_h we employ $1 - e^{-x} < 1$ and develop the bound as follows:

$$\begin{aligned} &\sum_{\mathbf{k}\in\Gamma_h} |\widehat{f}(\mathbf{k})| \cdot |1 - e^{-2\pi^2 h^2 \|\mathbf{k}\|^2}| \leq \sum_{\mathbf{k}\in\Gamma_h} |\widehat{f}(\mathbf{k})| = \sum_{\mathbf{k}\in\Gamma_h} |\|\mathbf{k}\|^{\alpha-2-\frac{d}{2}} \cdot \|\mathbf{k}\|^{\frac{d}{2}+2-\alpha} |\widehat{f}(\mathbf{k})| \\ &\leq \left(\sum_{\mathbf{k}\in\Gamma_h} \|\mathbf{k}\|^{2\alpha-4-d} \right)^{\frac{1}{2}} \left(\sum_{\mathbf{k}\in\Gamma_h} \left(\|\mathbf{k}\|^2\right)^{\frac{d}{2}+2-\alpha} |\widehat{f}(\mathbf{k})|^2 \right)^{\frac{1}{2}} \\ &\leq \left(\sum_{\mathbf{k}\in\Gamma_h} \|\mathbf{k}\|^{2\alpha-4-d} \right)^{\frac{1}{2}} \|f\|_{\frac{d}{2}+2-\alpha} = \left(\sum_{\mathbf{k}\in\Gamma_h} \|\mathbf{k}\|^{-2\beta} \right)^{\frac{1}{2}} \|f\|_\beta. \end{aligned}$$

Keeping in mind that $\frac{d}{2} < \beta \leq \frac{d}{2}+2$, the integral comparison test yields

$$\sum_{\mathbf{k}\in\Gamma_h} \|\mathbf{k}\|^{-2\beta} \leq \int_{\|\mathbf{x}\|\geq\frac{1}{h}} \|\mathbf{x}\|^{-2\beta} d\mathbf{x} = C_d \int_{\frac{1}{h}}^{\infty} r^{d-1-2\beta} dr \leq Ch^{2\beta-d},$$

and, for this parameter range, we can deduce that

$$\sum_{\mathbf{k}\in\Gamma_h} |\widehat{f}(\mathbf{k})| \cdot |1 - e^{-2\pi^2 h^2 \|\mathbf{k}\|^2}| \leq Ch^{\beta-\frac{d}{2}}, \ \text{for} \ \frac{d}{2} < \beta \leq \frac{d}{2}+2. \tag{2.6}$$

Combining this with (2.5) provides the bounds stated in the proposition.

□

3 Gaussian convolution approximation on sparse grids

The convergence results of the previous section are of theoretical interest, however, from a practical perspective, the implementation of the discrete (quasi-interpolation) analogue in high dimensions is restricted by the curse of dimensionality. A direct recasting of the continuous case to discrete setting would require that we sample values on a full grid in $[0,1]^d$, thus for $h = 1/2^n$ this would amount to $(2^n+1)^d$ evaluations. One remedy that can be used to alleviate the curse of dimensionality, at least for moderately high dimensions, is to approximate on a carefully chosen subset of the full grid, where substantially fewer points are needed to achieve an acceptable level of accuracy. To describe our approach we let $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_d)$ denote a general multi-index where $\ell_i \geq 1$ for $i = 1, \ldots, d$ then we define $\mathcal{X}_{\boldsymbol{\ell}}$ to be the anisotropic (directionally uniform) grid in $[0,1]^d$ where $h_i = 1/2^{\ell_i}$ denotes the spacing in the i^{th} coordinate direction. The number of nodes in $\mathcal{X}_{\boldsymbol{\ell}}$ is then given by

$$|\mathcal{X}_{\boldsymbol{\ell}}| = \prod_{i=1}^{d} (2^{\ell_i} + 1).$$

We let $\mathcal{X}_{n,d} = \mathcal{X}_{2^{-n}\mathbf{1}}$ denote the full isotropic grid with a uniform spacing of $h = \frac{1}{2^n}$. As a starting point we can appeal to Proposition 2.1 to conclude that the approximation error for the continuous convolution approximation to any given $f \in \mathcal{N}_\beta$ ($\beta > \frac{d}{2} + 2$) on the full grid $\mathcal{X}_{n,d}$ satisfies

$$\|E_{\frac{1}{2^n}} f\|_\infty = \|f - \mathcal{C}_{\frac{1}{2^n}}(f)\|_\infty = \mathcal{O}\left(\frac{1}{2^{2n}}\right). \tag{3.1}$$

In what follows we will consider an approach to convolution approximation on sparse grid subsets of $\mathcal{X}_{n,d}$. To be more precise, we consider the following subset of $\mathcal{X}_{n,d}$,

$$\mathcal{S}_{n,d} = \bigcup_{|\boldsymbol{\ell}|_1 = n+d-1} \mathcal{X}_{\boldsymbol{\ell}}, \tag{3.2}$$

with $|\boldsymbol{\ell}|_1 = \ell_1 + \cdots + \ell_d$, which will be referred to as the sparse grid at level n in d dimensions. We note that there is some redundancy in this definition; the sparse grid is represented as a combination of sub-grids and some grid points are included in more than one sub-grid, this is nicely illustrated, for the 2 dimensional case, in Figure 1.

Figure 1: The sparse grid $\mathcal{S}_{4,2}$ constructed via (3.2)

An effort to reduce this redundancy is possible by employing the Boolean sum representation of Delvos [7], specifically one can express the sparse grid as

$$\mathcal{S}_{n,d} = \sum_{q=0}^{d-1} (-1)^q \binom{d-1}{q} \sum_{|\boldsymbol{\ell}|_1 = n+(d-1)-q} \mathcal{X}_{\boldsymbol{\ell}}, \tag{3.3}$$

where we interpret the positive contributions as the inclusion of points and the negative contributions as their removal, this approach is nicely illustrated for the 2 dimensional case, in Figure 2. A precise formula for $|\mathcal{S}_{n,d}|$, the number of points in the sparse grid $\mathcal{S}_{n,d}$ is provided in Lemma 7 of [5], specifically it shown there that

$$|\mathcal{S}_{n,d}| = \sum_{i=0}^{n-1} 2^i \binom{d-1+i}{d-1} = \mathcal{O}(2^n n^{d-1}), \tag{3.4}$$

which is a significant reduction from $\mathcal{O}(2^{dn})$ for a full grid. Later in the paper we will show that the accuracy of the sparse grid approximation only deteriorates by a factor of n^{d-1} in comparison to the full grid approximation. This is typical of what is to be expected from sparse grid methods (see [4]) and this advantage makes them an attractive solution.

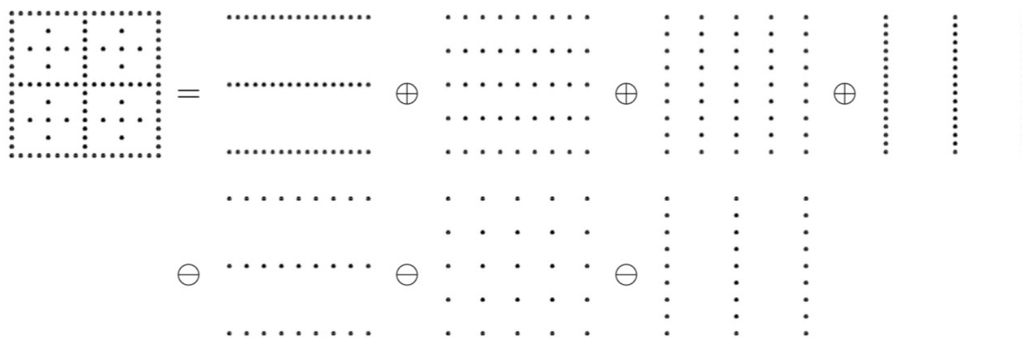

Figure 2: The sparse grid $\mathcal{S}_{4,2}$ constructed via (3.3)

Following (2.1) we represent the anisotropic convolution approximation on $\mathcal{X}_\ell$ as

$$\mathcal{C}_{\boldsymbol{\ell}}(f)(\mathbf{x}) = f * \Psi_{\left(\frac{1}{2^{\ell_1}},\ldots,\frac{1}{2^{\ell_d}}\right)}(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbf{Z}^d} \widehat{f}(\mathbf{k}) \left(\prod_{i=1}^{d} \widehat{\psi}\left(\frac{k_i}{2^{\ell_i}}\right)\right) e_{\mathbf{k}}(\mathbf{x}). \tag{3.5}$$

The convolution approximation on the sparse grid $\mathcal{S}_{n,d}$ is formed via the so-called *combination technique* [9]. This technique takes anisotropic convolution approximations $\mathcal{C}_{\boldsymbol{\ell}}(f)(\mathbf{x})$ (3.5) on the coarser grids $\mathcal{X}_\ell$ used in the formation $\mathcal{S}_{n,d}$ and then linearly combines these according to the Boolean decomposition (3.3), it is defined as

$$\mathcal{C}_{n,d}(f)(\mathbf{x}) = \sum_{q=0}^{d-1} (-1)^q \binom{d-1}{q} \sum_{|\boldsymbol{\ell}|_1 = n+(d-1)-q} \mathcal{C}_{\boldsymbol{\ell}}(f)(\mathbf{x}). \tag{3.6}$$

Substituting (3.5) into the above one can show that

$$\mathcal{C}_{n,d}(f)(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbf{Z}^d} \widehat{f}(\mathbf{k}) \widehat{\mathcal{C}_{n,d}}(\mathbf{k}) e_{\mathbf{k}}(\mathbf{x})$$

with

$$
\begin{aligned}
\widehat{\mathcal{C}_{n,d}}(\mathbf{k}) &= \sum_{q=0}^{d-1}(-1)^q \binom{d-1}{q} \sum_{|\boldsymbol{\ell}|_1=n+(d-1)-q} \prod_{i=1}^{d} \widehat{\psi}\left(\frac{k_i}{2^{\ell_i}}\right) \\
&= (-1)^{d-1} \sum_{q=0}^{d-1}(-1)^q \binom{d-1}{q} \sum_{|\boldsymbol{\ell}|_1=n+q} e^{-\sum_{i=1}^{d} \frac{2\pi^2 k_i^2}{2^{2\ell_i}}},
\end{aligned}
\tag{3.7}
$$

where, in the final line, the order of the terms in the original outer sum are reversed.

Using this representation the pointwise error formula is given by

$$
E_{n,d}(f)(\mathbf{x}) = f(\mathbf{x}) - \mathcal{C}_{n,d}(f)(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbf{Z}^d} \widehat{f}(\mathbf{k}) \widehat{\mathcal{E}_{n,d}}(\mathbf{k}) e_{\mathbf{k}}(\mathbf{x}),
\tag{3.8}
$$

where

$$
\widehat{\mathcal{E}_{n,d}}(\mathbf{k}) = 1 - \widehat{\mathcal{C}_{n,d}}(\mathbf{k}) = 1 - (-1)^{d-1} \sum_{q=0}^{d-1}(-1)^q \binom{d-1}{q} \sum_{|\boldsymbol{\ell}|_1=n+q} e^{-\sum_{i=1}^{d} \frac{2\pi^2 k_i^2}{2^{2\ell_i}}}.
\tag{3.9}
$$

We notice that when $\mathbf{k} = \mathbf{0}$ we have that

$$
\begin{aligned}
\widehat{\mathcal{E}_{n,d}}(\mathbf{0}) &= 1 - (-1)^{d-1} \sum_{q=0}^{d-1}(-1)^q \binom{d-1}{q} \sum_{|\boldsymbol{\ell}|_1=n+q} 1 \\
&= 1 - (-1)^{d-1} \sum_{q=0}^{d-1}(-1)^q \binom{d-1}{q} \binom{n+q-1}{d-1},
\end{aligned}
\tag{3.10}
$$

where we have used the fact that that the number of ways to write s as the sum of r positive integers is $\binom{s-1}{r-1}$. The following identity, which is taken from [12] Formula 4.2.5.47, is valid for non-negative integers r and s such that $0 \leq r \leq s$

$$
\sum_{q=0}^{s}(-1)^q \binom{s}{q} \binom{a+bq}{r} = (-1)^s b^s \delta_{r,s}.
$$

Applying this with $r = s = d-1$, $a = n-1$ and $b = 1$ we can conclude that the sum in the expression above equates to $(-1)^{d-1}$ and hence we have that $\widehat{\mathcal{E}_{n,d}}(\mathbf{0}) = 0$. Thus, as with the plain convolution approximation, the combination convolution approximation on the sparse grid also reproduces the constant function. At this point in the paper it is pertinent to compare the two error representations for convolution approximation that we have developed so far, in the continuous (full grid) setting we have

$$
E_{n,d}(f)(\mathbf{x}) = f(\mathbf{x}) - \mathcal{C}_{\frac{1}{2^n}}(f)(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbf{Z}^d\setminus\{\mathbf{0}\}} \widehat{f}(\mathbf{k}) \left(1 - e^{-\frac{2\pi^2 \|\mathbf{k}\|^2}{2^{2n}}}\right) e_{\mathbf{k}}(\mathbf{x})
$$

and in the sparse grid case we have

$$
E_{n,d}(f)(\mathbf{x}) = f(\mathbf{x}) - \mathcal{C}_{n,d}(f)(\mathbf{x}) = \sum_{\mathbf{k}\in\mathbf{Z}^d\setminus\{\mathbf{0}\}} \widehat{f}(\mathbf{k}) \left(1 - \sum_{q=0}^{d-1}(-1)^{q+d-1} \binom{d-1}{q} \sum_{|\boldsymbol{\ell}|_1=n+q} e^{-\sum_{i=1}^{d} \frac{2\pi^2 k_i^2}{2^{2\ell_i}}}\right) e_{\mathbf{k}}(\mathbf{x}).
$$

In Section 2 we found that error bounds, for sufficiently smooth functions, in the full grid case are easy to access by applying the simple inequality $1 - e^{-x} \leq x$. The situation for the sparse grid case is clearly not as straightforward and this leads us to embark on a thorough investigation of the coefficients (3.9). To this end we will begin with a detailed examination of the 2−dimensional case. The results of this investigation will form the base case of an inductive proof which we will use to establish convergence in higher dimensions.

3.1 Convergence in two dimension

In two dimensions the sparse grid convolution coefficients of the point-wise error formula (3.9) have the form

$$\widehat{\mathcal{E}_{n,2}}(\mathbf{k}) = 1 + \sum_{i+j=n} e^{-2\pi^2\left(\frac{k_1^2}{2^{2i}}+\frac{k_2^2}{2^{2j}}\right)} - \sum_{i+j=n+1} e^{-2\pi^2\left(\frac{k_1^2}{2^{2i}}+\frac{k_2^2}{2^{2j}}\right)}. \tag{3.11}$$

Let us develop the general term in the above expression using the full series expansion of the exponential function. Specifically, we consider

$$\sum_{i+j=m} e^{-2\pi^2\left(\frac{k_1^2}{2^{2i}}+\frac{k_2^2}{2^{2j}}\right)} = \sum_{i+j=m}\sum_{p=0}^{\infty}(-1)^p\frac{2^p\pi^{2p}}{p!}\left(\frac{k_1^2}{2^{2i}}+\frac{k_2^2}{2^{2j}}\right)^p = \sum_{p=0}^{\infty}(-1)^p\frac{2^p\pi^{2p}}{p!}\sigma_{2,p}(m,\mathbf{k}), \tag{3.12}$$

where

$$\sigma_{2,p}(m,\mathbf{k}) = \sum_{i+j=m}\left(\frac{k_1^2}{2^{2i}}+\frac{k_2^2}{2^{2j}}\right)^p = \begin{cases} m-1 & \text{if } p=0; \\ \sum_{j=1}^{m-1}\left(\frac{k_1^2}{2^{2(m-j)}}+\frac{k_2^2}{2^{2j}}\right)^p & \text{if } p\geq 1. \end{cases} \tag{3.13}$$

For $p \geq 1$ we can apply the binomial theorem to yield

$$\sigma_{2,p}(m,\mathbf{k}) = \sum_{j=1}^{m-1}\sum_{r=0}^{p}\binom{p}{r}\frac{k_1^{2r}}{2^{2(m-j)r}}\frac{k_2^{2(p-r)}}{2^{2j(p-r)}} = \sum_{r=0}^{p}\binom{p}{r}\frac{k_1^{2r}k_2^{2(p-r)}}{2^{2mr}}\sum_{j=1}^{m-1}2^{2(2r-p)j}. \tag{3.14}$$

The inner term of the above expression is a geometric sum of the form $\sum_{j=1}^{n} x^j$ with $n := m-1$ and either $x := 2^{2(2r-p)}$, if $2r \neq p$, or $x = 1$, if $2r = p$. To help distinguish between these cases we define

$$\Delta_{p,2} = \begin{cases} 0 & \text{if } p \text{ is odd}; \\ 1 & \text{otherwise}, \end{cases} \tag{3.15}$$

then applying the geometric sum formula

$$\sum_{j=1}^{n} x^j = \begin{cases} \frac{1}{x^{-1}-1}(1-x^n) & \text{if } x\neq 1; \\ n & \text{if } x=1, \end{cases} \tag{3.16}$$

we find that

$$
\begin{aligned}
\sigma_{2,p}(m,\mathbf{k}) &= \sum_{\substack{r=0\\ r\neq\frac{p}{2}}}^{p}\binom{p}{r}\frac{k_1^{2r}k_2^{2(p-r)}}{2^{2mr}}\frac{1-2^{2(2r-p)(m-1)}}{2^{2(p-2r)}-1}+\Delta_{p,2}\binom{p}{\frac{p}{2}}\frac{k_1^pk_2^p}{2^{mp}}(m-1)\\
&= \sum_{\substack{r=0\\ r\neq\frac{p}{2}}}^{p}\binom{p}{r}\frac{k_1^{2r}k_2^{2(p-r)}}{2^{2mr}(2^{2(p-2r)}-1)}-\sum_{\substack{r=0\\ r\neq\frac{p}{2}}}^{p}\binom{p}{r}\frac{k_1^{2r}k_2^{2(p-r)}2^{2(2r-p)(m-1)}}{2^{2mr}(2^{2(p-2r)}-1)}+\Delta_{p,2}\binom{p}{\frac{p}{2}}\frac{k_1^pk_2^p}{2^{mp}}(m-1)\\
&= \sum_{\substack{r=0\\ r\neq\frac{p}{2}}}^{p}\binom{p}{r}\frac{k_1^{2r}k_2^{2(p-r)}}{2^{2mr}(2^{2(p-2r)}-1)}-\sum_{\substack{r=0\\ r\neq\frac{p}{2}}}^{p}\binom{p}{r}\frac{k_1^{2(p-r)}k_2^{2r}2^{2(p-2r)(m-1)}}{2^{2m(p-r)}(2^{2(2r-p)}-1)}+\Delta_{p,2}\binom{p}{\frac{p}{2}}\frac{k_1^pk_2^p}{2^{mp}}(m-1)\\
&= \sum_{\substack{r=0\\ r\neq\frac{p}{2}}}^{p}\binom{p}{r}\frac{k_1^{2r}k_2^{2(p-r)}}{2^{2mr}(2^{2(p-2r)}-1)}-\sum_{\substack{r=0\\ r\neq\frac{p}{2}}}^{p}\binom{p}{r}\frac{k_1^{2(p-r)}k_2^{2r}}{2^{2mr}(1-2^{2(p-2r)})}+\Delta_{p,2}\binom{p}{\frac{p}{2}}\frac{k_1^pk_2^p}{2^{mp}}(m-1)\\
&= \sum_{\substack{r=0\\ r\neq\frac{p}{2}}}^{p}\binom{p}{r}\frac{k_1^{2r}k_2^{2(p-r)}+k_1^{2(p-r)}k_2^{2r}}{2^{2mr}(2^{2(p-2r)}-1)}+\Delta_{p,2}\binom{p}{\frac{p}{2}}\frac{k_1^pk_2^p}{2^{mp}}(m-1),
\end{aligned}
\tag{3.17}
$$

where the third line follows from reversing the index of the second sum and the penultimate line follows from a cancellation in the summands of the second sum. The following result captures the terms that dominate the behaviour of $\sigma_{2,p}(m,\mathbf{k})$ for large m.

Proposition 3.1. *Let $\boldsymbol{k}=(k_1,k_2)^T\in\mathbf{Z}$. Then, for $m\geq 2$ and $p=1$ we have*

$$\sigma_{2,1}(m,\boldsymbol{k})=\|\boldsymbol{k}\|^2\cdot\left(C_0-\frac{4}{3}\cdot\frac{1}{2^{2m}}\right)\quad\text{where}\quad C_0=\frac{1}{3}. \tag{3.18}$$

Furthermore, for $p\geq 2$ we have that

$$\sigma_{2,p}(m,\boldsymbol{k})=C_0^{\boldsymbol{k},p}+\delta_{p,2}2k_1^2k_2^2\frac{m}{2^{2m}}+\frac{C(\boldsymbol{k},p,2)}{2^{2m}}+O\left(\frac{m}{2^{4m}}\right) \tag{3.19}$$

where $\delta_{p,2}$ denotes the Kronecker delta function and

$$C_0^{\boldsymbol{k},p}=\frac{k_1^{2p}+k_2^{2p}}{2^{2p}-1}\quad\text{and}\quad C(\boldsymbol{k},p,2)=\frac{(1-\delta_{p,2})p\left(k_1^{2(p-1)}k_2^2+k_1^2k_2^{2(p-1)}\right)}{2^{2(p-2)}-1}-\delta_{p,2}2k_1^2k_2^2. \tag{3.20}$$

Proof. The first equality follows from setting $p=1$ in (3.17). The second equality arises from isolating the dominant terms of (3.17) while ignoring those which decay faster than $\frac{1}{2^{2m}}$. □

Using the notation introduced above we can write (3.11) as

$$
\begin{aligned}
\widehat{\mathcal{E}_{n,2}}(\mathbf{k}) =& 1+\sum_{p=0}^{\infty}\frac{(-1)^p2^p\pi^{2p}}{p!}\left(\sigma_{2,p}(n,\mathbf{k})-\sigma_{2,p}(n+1,\mathbf{k})\right)\\
=& -2\pi^2\left(\sigma_{2,1}(n,\mathbf{k})-\sigma_{2,1}(n+1,\mathbf{k})\right)+\sum_{p=2}^{\infty}\frac{(-1)^p2^p\pi^{2p}}{p!}\left(\sigma_{2,p}(n,\mathbf{k})-\sigma_{2,p}(n+1,\mathbf{k})\right),
\end{aligned}
\tag{3.21}
$$

the final equality being true since $\sigma_{2,0}(n,\mathbf{k}) = n-1$. Using (3.18) and(3.19) we make note of the following

$$\sigma_{2,1}(n,\mathbf{k}) - \sigma_{2,1}(n+1,\mathbf{k}) = -\frac{\|\mathbf{k}\|^2}{2^{2n}}$$
$$\sigma_{2,p}(n,\mathbf{k}) - \sigma_{2,p}(n+1,\mathbf{k}) = \frac{1}{2^{2n}}\left[\delta_{p,2}\left(\frac{3n-1}{2}\right)k_1^2k_2^2 + \frac{3}{4}C(\mathbf{k},p,2) + O\left(\frac{n}{2^{2n}}\right)\right].$$

This allows us to write

$$\widehat{\mathcal{E}_{n,2}}(\mathbf{k}) = \frac{1}{2^{2n}}\left(2\pi^2\|\mathbf{k}\|^2 + \sum_{p=2}^{\infty}\frac{(-1)^p 2^p \pi^{2p}}{p!}\left[\delta_{p,2}\left(\frac{3n-1}{2}\right)k_1^2k_2^2 + \frac{3}{4}C(\mathbf{k},p,2) + O\left(\frac{n}{2^{2n}}\right)\right]\right) \tag{3.22}$$

Examining the above, we observe that the term dominating the asymptotic rate of decay corresponds to the first ($p=2$) term of the infinite sum, and hence we can deduce that

$$\widehat{\mathcal{E}_{n,2}}(\mathbf{k}) \sim \frac{2^2\pi^4}{2!}\cdot\frac{3}{2}\frac{n}{2^{2n}}k_1^2k_2^2 = 3\pi^4 k_1^2k_2^2\frac{n}{2^{2n}}. \tag{3.23}$$

Employing this result in (3.8) we can deduce

$$E_{n,2}(f)(\mathbf{x}) \sim 3\pi^4\frac{n}{2^{2n}}\sum_{\mathbf{k}\in Z^2}\widehat{f}(\mathbf{k})k_1^2k_2^2 e_{\mathbf{k}}(\mathbf{x}). \tag{3.24}$$

3.2 Convergence in $d > 2$ dimensions

In this part we will embark on the $d-$dimensional analogue of the approach from the previous subsection. We begin by defining the $d-$dimensional version of (3.13)

$$\sigma_{d,p}(m,\mathbf{k}) = \begin{cases}\binom{m-1}{d-1} & \text{if } p=0;\\ \sum_{|\boldsymbol{\ell}|_1=m}\left(\sum_{i=1}^{d}\frac{k_i^2}{2^{2\ell_i}}\right)^p & \text{if } p\geq 1.\end{cases} \tag{3.25}$$

Then, using this notation in the expansion of the exponential function, the error coefficients (3.9) can be represented as

$$\begin{aligned}\widehat{\mathcal{E}_{n,d}}(\mathbf{k}) =& 1-(-1)^{d-1}\sum_{p=0}^{\infty}\frac{(-1)^p2^p\pi^{2p}}{p!}\sum_{q=0}^{d-1}(-1)^q\binom{d-1}{q}\sigma_{d,p}(n+q,\mathbf{k})\\ =&1-(-1)^{d-1}\sum_{q=0}^{d-1}(-1)^q\binom{d-1}{q}\binom{n+q-1}{d-1}\\ &-\sum_{p=1}^{\infty}\frac{(-1)^p2^p\pi^{2p}}{p!}\sum_{q=0}^{d-1}(-1)^{d-1-q}\binom{d-1}{q}\sigma_{d,p}(n+q,\mathbf{k}).\end{aligned}$$

We note that the penultimate line above coincides with $\mathcal{E}_{n,d}(\mathbf{0})$ (3.10) which we have shown to be zero. To simplify the notation we recall the forward divided difference functional of order k is defined by

$$\Delta^k f = \sum_{q=0}^{k}(-1)^{k-q}\binom{k}{q}f(q).$$

Taking $k = d-1$ we can express the $d-$dimensional sparse grid convolution error coefficients as

$$\widehat{\mathcal{E}_{n,d}}(\mathbf{k}) = 2\pi^2\Delta^{d-1}\sigma_{d,1}(n+\cdot,\mathbf{k}) - \sum_{p=2}^{\infty}\frac{(-1)^p 2^p \pi^{2p}}{p!}\Delta^{d-1}\sigma_{d,p}(n+\cdot,\mathbf{k}). \tag{3.26}$$

The above result is the $d-$dimensional equivalent of (3.21). In order to complete the investigation we need to derive a representation of $\sigma_{d,p}(m,\mathbf{k})$, analogous to (3.17) in two dimensions, that describes the rate at which the error coefficients decay. The following result provides the insight we need.

Theorem 3.2. *Let $d \geq 2$, $\boldsymbol{\ell} = (\ell_1,\ldots,\ell_d)^T \in \mathbb{N}^d$ and $\boldsymbol{k} = (k_1,\ldots,k_d)^T \in \mathbf{Z}^d$. Then, for $m \geq d$, we have*

$$\sigma_{d,1}(m,\boldsymbol{k}) = \|\boldsymbol{k}\|^2\left(\pi_{d-2}(m) + (-1)^{d-1}\left(\frac{4}{3}\right)^{d-1}\frac{1}{2^{2m}}\right), \tag{3.27}$$

where $\pi_{d-2}(m)$ is a polynomial in m of degree $d-2$. Furthermore, for $p \geq 2$ we also have

$$\sigma_{d,p}(m,\boldsymbol{k}) = \pi_{d-2}^{\boldsymbol{k},p}(m) + \delta_{p,d}\frac{dk_1^2\cdots k_d^2\, m^{d-1}}{2^{2m}} + \frac{C(\boldsymbol{k},p,d)m^{d-2}}{2^{2m}} + O\left(\frac{m^{d-3}}{2^{2m}}\right), \tag{3.28}$$

where $\pi_{d-2}^{\boldsymbol{k},p}(m)$ is a polynomial in m of degree $d-2$ whose coefficients depend upon $\boldsymbol{k}$ and p.

We observe that Proposition 3.1 verifies the $d=2$ case of the theorem, where

$$\pi_0(m) = C_0 \text{ (see (3.18))}, \quad \text{and} \quad \pi_0^{\mathbf{k},p}(m) = C_0^{\mathbf{k},p} \text{ (see (3.19), (3.20))}.$$

and thus serves as the base case for a proof by induction.

The full proof of this theorem relies on some rather technical machinery and this, together with the proof, is provided in the final section of the paper. We close this subsection in two parts. First, we will use the results of Theorem 3.2 to establish the asymptotic decay rate of the error coefficents (3.26), and hence deliver approximation rates for sparse grid Gaussian convolution approximation. Second, we will present some numerical results to demonstrate how well the asymptotic formula tracks the values of the coefficients as n becomes large.

The key insight from Theorem 3.2 is that, for each p, the $\sigma_{d,p}(m,\mathbf{k})$ function can be expressed as a polynomial in m of degree $d-2$ plus either a constant multiple of $m^{d-2}/2^{2m}$ (when $p \neq d$) or $dk_1^2\cdots k_d^2 m^{d-1}/2^{2m}$ (when $p=d$) followed by higher order terms (i.e., those decaying at a faster rate as m grows). Given that the forward divided difference functional annihilates polynomials of degree $d-2$ we can, after ignoring the higher order terms, deduce that

$$\begin{aligned}\widehat{\mathcal{E}_{n,d}}(\mathbf{k}) = \frac{1}{2^{2n}}\Bigg[&(-1)^d 2\pi^2\|\mathbf{k}\|^2\sum_{q=0}^{d-1}(-1)^{d-1-q}\binom{d-1}{q}\frac{1}{2^{2q}}\\ &-\sum_{\substack{p=2\\p\neq d}}^{\infty}\frac{(-1)^p 2^p\pi^{2p}C(\mathbf{k},p,d)}{p!}\sum_{q=0}^{d-1}(-1)^{d-1-q}\binom{d-1}{q}\frac{(n+q)^{d-2}}{2^{2q}}\\ &-\frac{(-1)^d 2^d\pi^{2d}dk_1^2\cdots k_d^2}{d!}\sum_{q=0}^{d-1}(-1)^{d-1-q}\binom{d-1}{q}\frac{(n+q)^{d-1}}{2^{2q}} + O(n^{d-3})\Bigg].\end{aligned}$$

Examining the above sum we see that the asymptotic decay of the coefficients is dominated by the n^{d-1} weight arising from the application of the forward divided difference operator to $(n+\cdot)^{d-1}$ in the final sum. Thus, using the binomial identity

$$\sum_{q=0}^{d-1}(-1)^{d-1-q}\binom{d-1}{q}x^q = (-1)^{d-1}(1-x)^{d-1},$$

with $x = 1/4$ we may deduce that

$$\widehat{\mathcal{E}_{n,d}}(\mathbf{k}) \sim \frac{2\pi^{2d}k_1^2\cdots k_d^2}{(d-1)!}\left(\frac{3}{2}\right)^{d-1}\frac{n^{d-1}}{2^{2n}}. \tag{3.29}$$

Employing this result in (3.8) we can deduce

$$E_{n,d}(f)(\mathbf{x}) \sim \frac{2\pi^{2d}}{(d-1)!}\left(\frac{3}{2}\right)^{d-1}\frac{n^{d-1}}{2^{2n}}\sum_{\mathbf{k}\in Z^d}\widehat{f}(\mathbf{k})k_1^2\cdots k_d^2 e_{\mathbf{k}}(\mathbf{x}), \tag{3.30}$$

which we observe is the $d-$dimensional analogue of (3.24). Furthermore, by mirroring the proof of Proposition 2.1, we can conclude the following.

Corollary 3.3. *Let $f \in \mathcal{N}_\beta$, where $\beta > \frac{d}{2}+2$. Let $\mathcal{C}_{\frac{1}{2^n}}(f)$ denote the plain Gaussian convolution approximation (2.1) to f on the full isotropic grid $\mathcal{X}_{n,d}$ with spacing $1/2^n$ and $\mathcal{C}_{n,d}(f)$ denote the combined convolution approximation to f (3.6) on the sparse grid $\mathcal{S}_{n,d}$. Then*

$$\|\mathcal{C}_{\frac{1}{2^n}}(f)-f\|_\infty \le \frac{C_d}{2^{2n}}\|f\|_\beta \quad \text{and} \quad \|\mathcal{C}_{n,d}(f)-f\|_\infty \le \frac{C_d n^{d-1}}{2^{2n}}\|f\|_\beta,$$

where C_d denotes a generic dimension dependent constant.

Tables 1 and 2 contain numerical results to show how closely the Fourier coefficients of the sparse grid convolution approximation track the asymptotic formula (3.29). Specifically, Table 1 covers the two dimensional case where we compare the Fourier coefficients of the error function (3.9) given by

$$\widehat{\mathcal{E}_{n,2}}(\mathbf{k}) = 1 + \sum_{i=1}^{n-1} e^{-2\pi^2\left(\frac{k_1^2}{2^{2i}}+\frac{k_2^2}{2^{2(n-i)}}\right)} - \sum_{i=1}^{n} e^{-2\pi^2\left(\frac{k_1^2}{2^{2i}}+\frac{k_2^2}{2^{2(n+1-i)}}\right)}$$

against their predicted asymptotic decay rate (3.29) of $3\pi^4k_1^2k_2^2\cdot n2^{-2n}$. Table 2 covers the three dimensional case where we compare

$$\begin{aligned}\widehat{\mathcal{E}_{n,3}}(\mathbf{k}) = 1 &- \sum_{i=1}^{n-2}\sum_{j=1}^{n-i-1} e^{-2\pi^2\left(\frac{k_1^2}{2^{2i}}+\frac{k_2^2}{2^{2j}}+\frac{k_3^2}{2^{2(n-i-j)}}\right)} + 2\sum_{i=1}^{n-1}\sum_{j=1}^{n-i} e^{-2\pi^2\left(\frac{k_1^2}{2^{2i}}+\frac{k_2^2}{2^{2j}}+\frac{k_3^2}{2^{2(n+1-i-j)}}\right)} \\ &- \sum_{i=1}^{n}\sum_{j=1}^{n-i+1} e^{-2\pi^2\left(\frac{k_1^2}{2^{2i}}+\frac{k_2^2}{2^{2j}}+\frac{k_3^2}{2^{2(n+2-i-j)}}\right)},\end{aligned}$$

against the predicted asymptotic decay rate of $\left(\frac{3}{2}\right)^2\pi^6k_1^2k_2^2k_3^2\cdot n^2 2^{-2n}$; in both cases the values were computed using Wolfram Mathematica v12.

Table 1: Comparison of numerically computed 2−d expansion coefficients $\widehat{\mathcal{E}_{n,2}}(\mathbf{k})$ against their predicted asymptotic formula (3.29), with $\mathbf{k} = (1, 1)$ (left) and $\mathbf{k} = (500, 700)$ (right)

n	$\widehat{\mathcal{E}_{n,2}}(\mathbf{k})$	formula	$\widehat{\mathcal{E}_{n,2}}(\mathbf{k})$	formula
40	8.69 (-21)	9.67 (-21)	5.19 (-10)	1.18 (-9)
80	1.52 (-44)	1.60 (-44)	1.41 (-33)	1.99 (-33)
160	2.13 (-92)	2.19 (-92)	2.31 (-81)	2.68 (-81)
320	2.02 (-188)	2.05 (-188)	2.33 (-177)	2.51 (-177)
640	8.93 (-381)	8.98 (-381)	1.06 (-369)	1.10 (-369)

Table 2: Comparison of numerically computed 3−d expansion coefficients $\widehat{\mathcal{E}_{n,3}}(\mathbf{k})$ against their predicted asymptotic formula (3.29), with $\mathbf{k} = (1, 1, 1)$ (left) and $\mathbf{k} = (500, 700, 900)$ (right)

n	$\widehat{\mathcal{E}_{n,3}}(\mathbf{k})$	formula	$\widehat{\mathcal{E}_{n,3}}(\mathbf{k})$	formula
40	1.72 (-18)	2.86 (-18)	3.54 (-2)	2.84 (-1)
80	6.65 (-42)	9.47 (-42)	4.62 (-25)	9.40 (-25)
160	1.95 (-89)	2.59 (-89)	1.69 (-72)	2.57 (-72)
320	3.80 (-185)	4.85 (-185)	3.54 (-168)	4.82 (-168)
640	3.39 (-377)	4.26 (-377)	3.27 (-360)	4.22 (-360)

4 Proof of Main Theorem

In order to convey the main results stated in Theorem 3.2 it is instructive to revisit the $2-d$ case of subsection 3.1. In this setting we applied the binomial theorem to represent $\sigma_{p,2}$ as (3.14), an expression containing an inner finite geometric sum. We observed there that in the case where p is even this collapses to $\sum_{j=1}^{m-1} 1 = m-1$, thus introducing a linear term in m. If one were to carefully examine the $3-d$ case with an application of the trinomial theorem then one would arrive upon an expression for $\sigma_{m,3}$ containing nested geometrics sums. In the cases where p is not a multiple of 3, these sums lead to terms that are linear in m and in the cases where p is a multiple of 3, the sums collapse to $\sum_{j=1}^{m-2} j = (m-1)(m-2)/2$, and hence introduce a quadratic term in m. This pattern continues into higher dimensions where an application of the multinomial expansion leads to an expression for $\sigma_{p,d}$ involving linear combinations of nested finite geometric sums and these introduce polynomial terms in m. In the case where p is is a multiple of d these nested sums collapse to $\sum_{j=1}^{m-2} j^{d-2}$ and so introduce the highest degree polynomial in m of order $d-1$, and hence the form of (3.28).

The formal proof of the result is made difficult due, in part, to the notational complexity that is involved. We will establish the theorem in two parts. First we will directly establish identity (3.27); a surprisingly neat representation for the $p = 1$ case. Here we will require an intermediate result concerning the evaluation of certain geometric sum, this will then be used as a tool to prove (3.27). We will then move on to establish (3.28), here we will proceed via induction (we have already established the result to be true for $d = 2$). As alluded to in the previous paragraph, the polynomial terms that appear in (3.28) arise from the nested geometric sums that follow from an application of the multinomial theorem and, to aid the proof, we will present background results on such sums before we use them to verify (3.28).

4.1 Towards the proof of (3.27)

We begin with the following lemma which sheds some insight on a particular finite sum.

Lemma 4.1. *Let $d \geq 2$ be a positive integer and $m > d$. Then, for a positive integer r, we have*

$$\sum_{j=1}^{m-(d-1)} \frac{\binom{m-j-1}{d-2}}{2^{2jr}} = p_{d-2}(m) + \left(\frac{2^{2r}}{1-2^{2r}}\right)^{d-1} \frac{1}{2^{2rm}},$$

where $p^r_{d-2}(m)$ is a polynomial in m of degree $d-2$, whose coefficients depend on the value of r.

Proof. Recall the Gauss hypergeometric function (see [1], 15.1.1) is defined by

$$ {}_2F_1(a,b;c;z) := \sum_{j=0}^{\infty} \frac{(a)_j (b)_j}{(c)_j} \frac{z^j}{j!}, \tag{4.1}$$

where

$$(x)_j := x(x+1)\cdots(x+j-1) \quad j \geq 1 \tag{4.2}$$

denotes the Pochhammer symbol, with $(x)_0 = 1$. If n is a positive integer we have

$$(n)_j = \frac{(n+j-1)!}{(n-1)!} \quad \text{and} \quad (-n)_j = \frac{(-1)^j n!}{(n-j)!}.$$

Using the above it is straight forward to verify that

$$\sum_{j=1}^{m-(d-1)} \frac{\binom{m-j-1}{d-2}}{2^{2jp}} = \frac{1}{2^{2p}} \binom{m-2}{d-2} {}_2F_1\left(1, -(m-d); -(m-2); \frac{1}{2^{2p}}\right). \tag{4.3}$$

The following identity, see [13] Formula 7.3.1.178, is valid for non-negative integers p and q

$${}_2F_1(1,-p;-q;z) = \frac{q+1}{p+1} \sum_{k=0}^{q-p} \frac{(p-q)_k}{(p+2)_k} (1-z)^{-(k+1)} + \frac{(-1)^p p!}{(-q)_p} z^{q+1} (z-1)^{p-q-1}.$$

Applying this with $p = m-d$, $q = m-2$ and $z = \frac{1}{2^{2r}}$, we find that

$$\begin{aligned}
&{}_2F_1\left(1, -(m-d); -(m-2); \frac{1}{2^{2r}}\right) = \\
&\frac{m-1}{m-(d-1)} \sum_{k=0}^{d-2} \frac{(-(d-2))_k \left(\frac{2^{2r}}{2^{2r}-1}\right)^{k+1}}{(m-(d-2))_k} + \frac{(-1)^{m-1}(m-d)! \left(\frac{2^{2r}}{2^{2r}-1}\right)^{d-1}}{(-(m-2))_{m-d}} \frac{2^{2r}}{2^{2rm}} \\
&= (m-1) \sum_{k=0}^{d-2} \frac{(-1)^k (d-2)!(m-d)! \left(\frac{2^{2r}}{2^{2r}-1}\right)^{k+1}}{(d-2-k)!(m+k-(d-1))!} + \frac{(-1)^{d-1} 2^{2r} \left(\frac{2^{2r}}{2^{2r}-1}\right)^{d-1}}{\binom{m-2}{d-2} 2^{2rm}}.
\end{aligned}$$

In view of (4.3) we now multiply this by $\frac{1}{2^{2r}}\binom{m-2}{d-2}$ and, following some elementary simplifications, we have the following expression

$$\sum_{j=1}^{m-(d-1)} \frac{\binom{m-j-1}{d-2}}{2^{2rj}} = \sum_{k=0}^{d-2} \frac{(-1)^k \binom{m-1}{m+k-(d-1)} \left(\frac{2^{2r}}{2^{2r}-1}\right)^k}{2^{2r}-1} + \frac{(-1)^{d-1}\left(\frac{2^{2r}}{2^{2r}-1}\right)^{d-1}}{2^{2rm}}$$
$$= p^r_{d-2}(m) + \left(\frac{2^{2r}}{1-2^{2r}}\right)^{d-1} \frac{1}{2^{2rm}},$$

where $p^r_{d-2}(m)$, which represents the sum appearing above, is a polynomial in m of degree $d-2$. □

4.1.1 Proof of identity (3.27)

We know from (3.15) that

$$\sigma_{d,1}(m,\mathbf{k}) = \sum_{|\boldsymbol{\ell}|_1=m} \frac{k_1^2}{2^{2\ell_1}} + \cdots + \frac{k_d^2}{2^{2\ell_d}} = \sum_{i=1}^{d} k_i^2 \sum_{|\boldsymbol{\ell}|_1=m} \frac{1}{2^{2\ell_i}}.$$

The inner sum of the above expression concerns the set of $d-$dimensional multi-indices $\boldsymbol{\ell}$ satisfying $|\boldsymbol{\ell}|_1 = m$. A typical component ℓ_i of $\boldsymbol{\ell}$ can, theoretically, take on any value between 1 and $m-(d-1)$ included (in the latter case the remaining $d-1$ components are all set to 1). The number of times ℓ_i takes on a certain value $j \in \{1, 2, \dots, m-(d-1)\}$ is precisely the number of ways in which the remaining $d-1$ components of $\boldsymbol{\ell}$ sum to $m-j$ and this is given by $\binom{m-j-1}{d-2}$. Since the last sum in the above expression only depends on the value ℓ_i and not on i we have that

$$\sigma_{d,1}(m,\mathbf{k}) = \sum_{i=1}^{d} k_i^2 \sum_{j=1}^{m-(d-1)} \binom{m-j-1}{d-2} \frac{1}{2^{2j}} = \|\mathbf{k}\|^2 \left(p_{d-2}(m) + \left(\frac{2^2}{1-2^2}\right)^{d-1} \frac{1}{2^{2m}}\right),$$

where the last equation follows from Lemma 4.1, with $r=1$, and the proof of (3.27) is complete.

4.2 Towards the proof of (3.28)

In this subsection we outline some key results on the representation of the kinds of weighted geometric series that are encountered if one applies the appropriate multinomial expansion in order to examine the sums $\sigma_{d,p}(m,\mathbf{k})$ (3.25) for $p \geq 2$. We begin by differentiating the plain geometric sum formula, followed by multiplication by x to deduce that

$$\sum_{j=1}^{n} jx^j = x\frac{d}{dx}\sum_{j=1}^{n} x^j = \begin{cases} \frac{x}{(1-x)^2}\left(1 - x^n\left((1+n) - nx\right)\right) & \text{if } x \neq 1; \\ \sum_{k=1}^{n} k = \frac{n(n+1)}{2} & \text{if } x = 1. \end{cases} \tag{4.4}$$

Let us consider the more general weighted geometric sum

$$G_i^{(n)}(x) = \sum_{j=1}^{n} j^i x^j.$$

We note in the case where $x = 1$ we have the sum of the i^{th} powers of the first n positive integers which, due to Faulhaber's formula, see [8] Formula 0.121, is a polynomial in n of degree $i+1$,

$$G_i^{(n)}(1) = \sum_{j=1}^{n} j^i = \frac{n^{i+1}}{i+1} + q_i(n), \tag{4.5}$$

where $q_i(n)$ is a polynomial in n of degree i. For the more general case ($x \neq 1$) we observe that

$$G_{i+1}^{(n)}(x) = x\frac{d}{dx}G_i^{(n)}(x) \tag{4.6}$$

and this allows us to establish the following.

Lemma 4.2. *Let j denote a non-negative integer and $x \neq 1$, then*

$$G_i^{(n)}(x) = \frac{x}{(1-x)^{i+1}}\left(q_{i-1}(x) - x^n p_i(x,n)\right), \tag{4.7}$$

where $q_{i-1}(x)$ is a polynomial of degree $i-1$ in x and $p_i(x,n)$ is a polynomial of degree i in both x and n.

Proof. We establish the result via induction. Appealing to (4.4) we see that the result is true for $i = 1$ with $q_0(x) = 1$ and $p_1(x,n) = 1 + n - nx$. Assume the result is true for i and consider the following development, using (4.6), for the case $i+1$.

$$\begin{aligned}
G_{i+1}^{(n)}(x) = x\frac{d}{dx}G_i^{(n)}(x) &= x\frac{d}{dx}\left(\frac{x}{(1-x)^{i+1}}\left(q_{i-1}(x) - x^n p_i(x,n)\right)\right)\\
&= x\Big[\frac{x}{(1-x)^{i+1}}\left(\frac{d}{dx}q_{i-1}(x) - x^n\frac{d}{dx}p_i(x,n) - nx^{n-1}p_i(x,n)\right)\\
&+ \left(\frac{1}{(1-x)^{i+1}} + \frac{ix+x}{(1-x)^{i+2}}\right)\left(q_{i-1}(x) - x^n p_i(x,n)\right)\Big]\\
&= \frac{x}{(1-x)^{i+2}}\Big[x(1-x)\left(\frac{d}{dx}q_{i-1}(x) - x^n\frac{d}{dx}p_i(x,n) - nx^{n-1}p_i(x,n)\right)\\
&+ (1+ix)\left(q_{i-1}(x) - x^n p_i(x,n)\right)\Big]\\
&= \frac{x}{(1-x)^{i+2}}\Big[q_i(x) - x^n p_{i+1}(x,n)\Big],
\end{aligned}$$

where

$$q_i(x) = (1+ix)q_{i-1}(x) + x(1-x)\frac{d}{dx}q_{i-1}(x)$$

is clearly a polynomial in x of degree i and, likewise, where

$$p_{i+1}(x,n) = (1+ix+n-nx)p_i(x,n) + x(1-x)\frac{d}{dx}p_i(x,n)$$

is clearly a polynomial of degree $i+1$ in both x and n.

□

In order to prepare for how the above result will be used, we let d be the fixed spatial dimension and $m > d$ a positive integer. In what follows we will evaluate various sums and, in each case, we will ignore terms that decay faster than $m^{d-2}2^{-2m}$ for large m. In each case we consider a fixed integer parameter $t \geq 1$ and, where appropriate, we will also consider specific cases of $t = 0$ and $t = -1$. We begin with a straightforward geometric sum for $t \geq 1$

$$\sum_{j=1}^{m-d} \frac{1}{2^{2jt}} = \frac{1}{2^{2t}-1}\left(1 - \frac{2^{2dt}}{2^{2mt}}\right) = \text{constant} + O(2^{-2m}). \tag{4.8}$$

For the following sum with $t > 1$ we can directly use (4.7) in its evaluation:

$$\begin{aligned} \sum_{j=1}^{m-d} \frac{(m-j)^i}{2^{2tj}} &= \frac{1}{2^{2mt}} \sum_{j=d}^{m-1} 2^{2tj} j^i = \frac{1}{2^{2mt}}\left(G_i^{(m-1)}(2^{2t}) - G_i^{(d-1)}(2^{2t})\right) \\ &= \frac{1}{(1-2^{2t})^{i+1}}\left[\frac{2^{2td}}{2^{2tm}} p_i(2^{2t}, d-1) - p_i(2^{2t}, m-1)\right] \\ &= \frac{P_i(m)}{(1-2^{2t})^{i+1}} + O(2^{-2m}), \end{aligned} \tag{4.9}$$

where $P_i(m)$ is a polynomial in m of degree i whose coefficients depend on 2^{2t}. In the case where $t = 0$ the above collapses to

$$\begin{aligned} \sum_{j=1}^{m-d} (m-j)^i &= \sum_{j=d}^{m-d} j^i = G_i^{m-1}(1) - G_i^{d-1}(1) \\ &= \frac{(m-1)^{i+1}}{i+1} - \frac{(d-1)^{i+1}}{i+1} + q_i(0, d-1) - q_i(0, m-1) \\ &= \frac{m^{i+1}}{i+1} + \pi_i(m), \end{aligned} \tag{4.10}$$

where $\pi_i(m)$ is a polynomial in m of degree i. In the case where $t = -1$ we have

$$\sum_{j=1}^{m-d} 2^{2j}(m-j)^i = C_{d,i} 2^{2m} + \left(\frac{4}{3}\right)^{i+1} P_i^*(m) \tag{4.11}$$

where $P_i^*(m)$ is a polynomial in m of degree i and $C_{d,i}$ is a constant depending on d and i.

4.2.1 Proof of (3.28)

We know, from subsection 3.1, that (3.28) holds for $d = 2$ and any value p, let us also assume that it is true for $d_0 < d$ and any value of p, we will now proceed to show, by induction, that the same statement is true for d and any value of p. First we establish a recurrence relation for $\sigma_{d,p}$ using

$$\begin{aligned} \boldsymbol{r} &= (r_1, \ldots, r_d)^T,\ \hat{\boldsymbol{r}} = (r_1, \ldots, r_{d-1})^T \\ \boldsymbol{\ell} &= (\ell_1, \ldots, \ell_d)^T,\ \hat{\boldsymbol{\ell}} = (\ell_1, \ldots, \ell_{d-1})^T \\ \boldsymbol{k} &= (k_1, \ldots, k_d)^T,\ \hat{\boldsymbol{k}} = (k_1, \ldots, k_{d-1})^T \end{aligned}$$

and the multinomial theorem to find

$$
\begin{aligned}
\sigma_{d,p}(m,\mathbf{k}) &= \sum_{|\boldsymbol{\ell}|_1=m}\left(\sum_{i=1}^{d}\frac{k_i^2}{2^{2\ell_i}}\right)^p \\
&= \sum_{|\boldsymbol{\ell}|_1=m}\sum_{|r|=p}\binom{p}{r_1\ \ldots\ r_d}\prod_{i=1}^{d}\left(\frac{k_i^2}{2^{2\ell_i}}\right)^{r_i} \quad \text{(multinomial theorem)} \\
&= \sum_{|\boldsymbol{r}|=p}\binom{p}{r_1\ \ldots\ r_d}\prod_{i=1}^{d}k_i^{2r_i}\sum_{|\boldsymbol{\ell}|_1=m}2^{-2\sum_{i=1}^{d}\ell_i r_i} \\
&= \sum_{r_d=0}^{p}\frac{p!k_d^{2r_d}}{(p-r_d)!r_d!}\sum_{\ell_d=1}^{m-d}2^{-2\ell_d r_d}\sum_{|\hat{\boldsymbol{r}}|=p-r_d}\binom{p-r_d}{r_1\ \ldots\ r_{d-1}}\prod_{i=1}^{d-1}k_i^{2r_i}\sum_{|\hat{\boldsymbol{\ell}}|_1=m-\ell_d}2^{-2\sum_{i=1}^{d}\ell_i r_i} \\
&= \sum_{r_d=0}^{p}\frac{p!k_d^{2r_d}}{(p-r_d)!r_d!}\sum_{\ell_d=1}^{m-d}\frac{\sigma_{d-1,p-r_d}(m-\ell_d,\hat{\boldsymbol{k}})}{2^{2\ell_d r_d}}.
\end{aligned}
$$

Applying the inductive hypothesis (3.28) we have

$$
\begin{aligned}
\sigma_{d-1,p-r_d}(m-\ell_d,\hat{\boldsymbol{k}}) =&\pi_{d-3}^{\hat{\boldsymbol{k}},p-r_d}(m-\ell_d)+\delta_{p-r_d,d-1}\frac{(d-1)k_1^2\cdots k_{d-1}^2(m-\ell_d)^{d-2}}{2^{2(m-\ell_d)}} \\
&+C(\hat{\boldsymbol{k}},p-r_d,d-1)\frac{(m-\ell_d)^{d-3}}{2^{2(m-\ell_d)}}+O\left(\frac{m^{d-4}}{2^{2m}}\right).
\end{aligned}
$$

Inserting this representation into the inner sum of the above computation yields

$$
\sum_{\ell_d=1}^{m-d}\frac{\sigma_{d-1,p-r_d}(m-\ell_d,\hat{\boldsymbol{k}})}{2^{2\ell_d r_d}} = S_1(r_d)+S_2(r_d)+S_3(r_d)+O\left(\frac{m^{d-3}}{2^{2m}}\right), \tag{4.12}
$$

where

$$
\begin{aligned}
S_1(r_d) &= \sum_{\ell_d=1}^{m-d}\frac{\pi_{d-3}^{\hat{\boldsymbol{k}},p-r_d}(m-\ell_d)}{2^{2\ell_d r_d}}, \\
S_2(r_d) &= \delta_{p-r_d,d-1}(d-1)\frac{k_1^2\cdots k_{d-1}^2}{2^{2m}}\sum_{\ell_d=1}^{m-d}\frac{(m-\ell_d)^{d-2}}{2^{2\ell_d(r_d-1)}}, \\
\text{and}\quad S_3(r_d) &= \frac{C(\hat{\boldsymbol{k}},p-r_d,d-1)}{2^{2m}}\sum_{\ell_d=1}^{m-d}\frac{(m-\ell_d)^{d-3}}{2^{2\ell_d(r_d-1)}}.
\end{aligned}
\tag{4.13}
$$

For the final term of (4.12) we have used that the sum consists of less than m terms and each of which is order $\frac{m^{d-4}}{2^{2m}}$. We can use the weighted geometric sums to investigate the three summands above. We begin with $S_1(r_d)$ and in this case we write the polynomial of degree $d-3$ as

$$
\pi_{d-3}^{\hat{\boldsymbol{k}},p-r_d}(m-\ell_d)=\sum_{i=0}^{d-3}a_i^{\hat{\boldsymbol{k}},p-r_d}(m-\ell_d)^i,
$$

and thus we have

$$S_1(r_d) = \sum_{i=1}^{d-3} a_i^{\hat{\boldsymbol{k}},p-r_d} \sum_{\ell_d=1}^{m-d} \frac{(m-\ell_d)^i}{2^{2\ell_d r_d}}.$$

Appealing to (4.9) and (4.10) we have that

$$\sum_{\ell_d=1}^{m-d} \frac{(m-\ell_d)^i}{2^{2\ell_d r_d}} = \begin{cases} \frac{P_i(m)}{(1-2^{2r_d})^{i+1}} + O(2^{-2m}) & \text{if} \quad r_d \neq 0; \\ \frac{m^{i+1}}{i+1} + \pi_i(m) & \text{if} \quad r_d = 0, \end{cases}$$

where $P_i(m)$ and $\pi_i(m)$ are polynomials in m of degree i. This insight allows us to write

$$S_1(r_d) = \begin{cases} P_{d-3}^{\hat{\boldsymbol{k}},p,r_d}(m) + O\left(\frac{1}{2^{2m}}\right) & \text{if} \quad r_d \neq 0; \\ P_{d-2}^{\hat{\boldsymbol{k}},p,0}(m) & \text{if} \quad r_d = 0, \end{cases} \tag{4.14}$$

where

$$P_{d-3}^{\hat{\boldsymbol{k}},p,r_d}(m) = \sum_{i=1}^{d-3} \frac{a_i^{\hat{\boldsymbol{k}},p-r_d}}{(1-2^{2r_d})^{i+1}} P_i(m) \quad \text{and} \quad P_{d-2}^{\hat{\boldsymbol{k}},p,0}(m) = \sum_{i=1}^{d-3} a_i^{\hat{\boldsymbol{k}},p} \left(\frac{m^{i+1}}{i+1} + \pi_i(m)\right) \tag{4.15}$$

are polynomials in m of degree $d-3$ and $d-2$ respectively.

For the second sum $S_2(r_d)$ we can bring the identities (4.9), (4.10) and (4.11) together to give

$$\sum_{\ell_d=1}^{m-d} \frac{(m-\ell_d)^{d-2}}{2^{2\ell_d(r_d-1)}} = \begin{cases} \frac{P_{d-2}(m)}{(1-2^{2(r_d-1)})^{d-1}} + O(2^{-2m}) & \text{if} \quad r_d > 1; \\ \frac{m^{d-1}}{d-1} + \pi_{d-2}(m) & \text{if} \quad r_d = 1; \\ C_{d,d-2}2^{2m} + \left(\frac{4}{3}\right)^{d-1} P_{d-2}^*(m) & \text{if} \quad r_d = 0, \end{cases}$$

and so deduce that

$$S_2(r_d) = \delta_{p-r_d,d-1}(d-1)\left(k_1^2\cdots k_{d-1}^2\right) \begin{cases} \frac{P_{d-2}(m)}{2^{2m}(1-2^{2(r_d-1)})^{d-1}} + O(2^{-4m}) & \text{if} \quad r_d > 1; \\ \frac{m^{d-1}}{2^{2m}(d-1)} + \frac{\pi_{d-2}(m)}{2^{2m}} & \text{if} \quad r_d = 1; \\ C_{d,d-2} + \left(\frac{4}{3}\right)^{d-1} \frac{P_{d-2}^*(m)}{2^{2m}} & \text{if} \quad r_d = 0. \end{cases}$$

Isolating the dominant term from those exhibiting faster decay we have that

$$S_2(r_d) = \delta_{p-r_d,d-1}\left(k_1^2\cdots k_{d-1}^2\right) \begin{cases} \frac{C_d m^{d-2}}{2^{2m}} + O\left(\frac{m^{d-3}}{2^{2m}}\right) & \text{if} \quad r_d > 1; \\ \frac{m^{d-1}}{2^{2m}} + O\left(\frac{m^{d-2}}{2^{2m}}\right) & \text{if} \quad r_d = 1; \\ C_d' + O\left(\frac{m^{d-2}}{2^{2m}}\right) & \text{if} \quad r_d = 0, \end{cases} \tag{4.16}$$

where $C_d' = (d-1)C_{d,d-2}$. For the third sum $S_3(r_d)$ we can use the same approach as above, with $d-3$ replacing $d-2$, to deduce that

$$S_3(r_d) = C(\hat{\boldsymbol{k}}, p-r_d, d-1) \begin{cases} \frac{C_{d-1} m^{d-3}}{2^{2m}} + O\left(\frac{m^{d-4}}{2^{2m}}\right) & \text{if} \quad r_d > 1; \\ \frac{m^{d-2}}{(d-2)2^{2m}} + O\left(\frac{m^{d-3}}{2^{2m}}\right) & \text{if} \quad r_d = 1; \\ C_{d-1}' + O\left(\frac{m^{d-3}}{2^{2m}}\right) & \text{if} \quad r_d = 0. \end{cases} \tag{4.17}$$

We now bring our findings (4.14),(4.16),(4.17) together, where again we isolate the dominant terms from those with faster decay to provide

$$\begin{aligned}\sigma_{d,p}(m,\boldsymbol{k}) &= \sum_{r_d=0}^{p} \frac{p!k_d^{2r_d}}{(p-r_d)!r_d!}\left(S_1(r_d)+S_2(r_d)+S_3(r_d)\right)+O\left(\frac{m^{d-3}}{2^{2m}}\right)\\ &= P_{d-2}^{\hat{\boldsymbol{k}},p,0}(m)+\sum_{r_d=1}^{p}\frac{p!k_d^{2r_d}}{(p-r_d)!r_d!}P_{d-3}^{\hat{\boldsymbol{k}},p,r_d}(m)\\ &+\delta_{p,d-1}\left(k_1^2\cdots k_{d-1}^2\right)C_d'+\delta_{p-1,d-1}\left(k_1^2\cdots k_{d-1}^2\right)\frac{k_d^2p!}{(p-1)!}\frac{m^{d-1}}{2^{2m}}+O\left(\frac{m^{d-2}}{2^{2m}}\right)\\ &+C(\hat{\boldsymbol{k}},p,d-1)C_{d-1}'+C(\hat{\boldsymbol{k}},p-1,d-1)\frac{pk_d^2}{d-2}\frac{m^{d-2}}{2^{2m}}+O\left(\frac{m^{d-3}}{2^{2m}}\right).\end{aligned}$$

By inspection we observe that the above can be expressed as

$$\sigma_{d,p}(m,\mathbf{k}) = \pi_{d-2}^{\mathbf{k},p}(m)+\delta_{p,d}\frac{dk_1^2\cdots k_d^2m^{d-1}}{2^{2m}}+\frac{C(\mathbf{k},p,d)m^{d-2}}{2^{2m}}+O\left(\frac{m^{d-3}}{2^{2m}}\right), \tag{4.18}$$

where $\pi_{d-2}^{\mathbf{k},p}(m)$ is a polynomial in m of degree $d-2$ whose coefficients depend upon $\mathbf{k}$ and p. This completes the proof of Theorem 3.2.

Funding:

The work of Janin Jäger was funded by the Deutsche Forschungsgemeinschaft (DFG - German research foundation) - Projektnummer: 461449252.

References

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions,* Dover Publications, New York (1964). https://doi.org/10.1119/1.15378.

[2] R.A. Adams. *Sobolev Spaces,* Volume 65 of Pure and Applied Mathematics, Academic Press (1975).

[3] M.D. Buhmann and J. Jäger, Quasi-interpolation, Cambridge University Press, (2022), https://doi.org/10.1017/9781139680523.

[4] H. J. Bungartz and M. Griebel, Sparse grids, Acta Numerica, **13** (2004), 147–269.

[5] H.J. Bungartz and M. Griebel, A note on the complexity of solving Poisson's equation for spaces of bounded mixed derivatives, Journal of Complexity, 15(2) (1999), 167–199.

[6] W. Cheney and W. A. Light, A Course in Approximation Theory, Brookes-Cole Publishing Company, 1999.

[7] F.-J. Delvos, d-variate Boolean interpolation, J. Approx. Theory, **34** (1982), 99–114.

[8] I.S. Gradshteyn and I.M. Ryzhik, Table of Integrals, Series, and Products. Academic Press, fourth edition, 1965.

[9] M. Griebel, M. Schneider, and C. Zenger. A combination technique for the solution of sparse grid problems, in Iterative Methods in Linear Algebra. Elsevier, (1992) 263–28.

[10] S. Hubbert and J. Levesley. Uniform Convergence of Multilevel Stationary Gaussian Quasi-Interpolation. arXiv preprint arXiv:1711.10803 (2020).

[11] V. Maz'ya and G. Schmidt, Approximate approximation, Mathematical Surveys and Monographs **41**, AMS, Providence, RI (2007).

[12] A.P. Prudnikov, Yu. A. Brychkov and O. I. Marichev. *Integrals and Series. Volume 1.* Gordon and Breach Science Publishers, (1992).

[13] A.P. Prudnikov., Yu. A. Brychkov and O. I. Marichev. *Integrals and Series. Volume 3.* Gordon and Breach Science Publishers, (1992).

[14] F. Usta and J. Levesley, Multilevel quasi-interpolation on a sparse grid with the Gaussian, Numerical Algorithms 77.3 (2018): 793-808, https://doi.org/10.1007/s11075-017-0340-y