SIAM J. NUMER. ANAL.
Vol. 57, No. 2, pp. 547–562

EFFECTIVE DIMENSION OF SOME WEIGHTED PRE-SOBOLEV SPACES WITH DOMINATING MIXED PARTIAL DERIVATIVES*

ART OWEN†

Abstract. This paper considers two notions of effective dimension for quadrature in weighted pre-Sobolev spaces with dominating mixed partial derivatives. We begin by finding a ball in those spaces just barely large enough to contain a function with unit variance. If no function in that ball has more than ε of its variance from analysis of variance (ANOVA) components involving interactions of order s or more, then the space has effective dimension at most s in the superposition sense. A similar truncation sense notion replaces the cardinality of an ANOVA component by the largest index it contains. These effective dimension definitions for the integration problem coincide with some of the definitions in information-based complexity for the function approximation problem. Some Poincaré-type inequalities are used to bound variance components by multiples of these space's squared norm and those in turn provide bounds on effective dimension. Very low effective dimension in the superposition sense holds for some spaces defined by product weights in which quadrature is strongly tractable. The superposition dimension is $O(\log(1/\varepsilon)/\log(\log(1/\varepsilon)))$ just like the superposition dimension used in the multivariate decomposition method. Surprisingly, even spaces where all subset weights are equal, regardless of their cardinality or included indices, have low superposition dimension in this sense. This paper does not require periodicity of the integrands.

Key words. Poincaré inequalities, quasi-Monte Carlo, tractability

AMS subject classifications. 41A55, 65D30, 45E35

DOI. 10.1137/17M1158975

1. Introduction.

This paper gives upper bounds for the effective dimension of certain weighted pre-Sobolev spaces of functions on $[0,1]^d$. They are pre-Sobolev spaces because the requisite derivatives must exist as continuous functions and not just in the distributional sense. The notion of effective dimension used here is the one in the technical report [37]. That work was based on a Fourier expansion and it required a periodicity condition on certain partial derivatives of the functions f. In this paper, no periodicity assumption is required.

The weighted spaces we consider are used as models for certain numerical integration problems. There we seek $\mu = \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$, where $1 \leqslant d < \infty$, and d might be large. For large d, integration becomes quite hard for worst case integrands, even smooth ones, yielding a curse of dimensionality described by Bakhvalov [2]. Quasi-Monte Carlo (QMC) integration [30, 6], often succeeds in high dimensions despite the curse. This can be explained by the functions having less than full d-dimensional complexity. Weighted Sobolev spaces [11, 45] provide one model for such reduced complexity. This paper translates the weights defining those spaces into bounds on certain L^2 norms quantifying the notion that those spaces of integrands are "effectively" s dimensional where s might be much less than d.

The functions we consider have a partial derivative taken once with respect to each of d coordinates and, moreover, that partial derivative is a continuous function

*Received by the editors November 28, 2017; accepted for publication (in revised form) January 2, 2019; published electronically March 14, 2019.
http://www.siam.org/journals/sinum/57-2/M115897.html
Funding: The work of the author was funded by the National Science Foundation under grants DMS-1521145 and DMS-1407397. Travel support by SAMSI (NSF DMS-1638521) is also gratefully acknowledged.
†Statistics, Stanford University, Stanford, CA 94305 (owen@stanford.edu).

on $[0,1]^d$. That is sufficient smoothness to place them in the weighted Sobolev spaces mentioned above. Such weighted spaces have been used to model settings in which higher order interactions [11], or successive dimensions, or both [45] are less important. When these high order or high index components decay quickly enough, the result is a set of functions that evades the curse of dimensionality established by [2].

The contribution of various high order or high index parts of an integrand to the Monte Carlo variance can be quantified through the analysis of variance (ANOVA) decomposition defined below. Suppose that an integrand f has less than one percent of its variance coming from high dimensional components, and that another quadrature method proves to be far more accurate than plain Monte Carlo. That improvement cannot be attributed to better handling of the high dimensional parts, because they caused at most one percent of the squared error for Monte Carlo. The improvement must have come from superior handing of the low dimensional aspects of f.

In this paper we investigate some senses in which a whole space of functions is of low effective dimension. We look at some weighted Sobolev norms (and some seminorms) and define measures of the extent to which balls in those normed spaces are dominated by their low dimensional parts with respect to an ANOVA decomposition. We select a ball in which the worst case Monte Carlo variance is unity, and then consider whether *any* integrand in that ball has meaningfully large variance coming from its high dimensional components. If not, then the space itself is said to have low effective dimension. This way of formulating an effective dimension for the *integration* problem gives an equivalent definition to a notion of effective dimension for the *approximation* problem, from the information based complexity (IBC) literature.

An outline of the paper is as follows. Section 2 gives our notation, introducing weighted pre-Sobolev spaces with some conditions on the weights and some decompositions (ANOVA and anchored) of $L^2\big([0,1]^d\big)$. Section 3 presents a survey of the notion of effective dimension, dating back to 1951. It also makes an explicit connection to the IBC literature. Section 4 introduces some Poincaré-type inequalities that we use to lower bound the Sobolev squared norms in terms of ANOVA components. Section 5 defines what it means for a ball in a space of functions to have a given effective dimension, i.e., to lack meaningfully large higher dimensional structure in any of its functions. Section 6 gives upper bounds on the effective dimension of a space in terms of its weights under monotonicity conditions that give smaller weights to higher order and higher indexed subsets of variables. Explicit effective dimension bounds are worked out and tabulated. One surprise in this work is that giving every subset equal weight, regardless of its cardinality or the size of the indices it contains, still leads to modest superposition dimension. Section 7 has some conclusions.

To finish this section, we mention some related work in addition to that covered in section 3. Poincaré inequalities have been used in the global sensitivity analysis (GSA) literature to bound Sobol' indices. See, for instance, [49, 50] and [24]. Roustant et al. [41] use Poincaré inequalities based on mixed partial derivatives to bound some superset variable importance measures in GSA.

2. Notation.

The indices of $\boldsymbol{x} \in [0,1]^d$ are $j \in \{1,2,\ldots,d\} \equiv 1{:}d$. For $u \subseteq 1{:}d$, we use $|u|$ for its cardinality and $-u$ for its complement with respect to $1{:}d$. For $u \subseteq 1{:}d$, the point $\boldsymbol{x}_u \in [0,1]^{|u|}$ consists of the components x_j for $j \in u$. For $i = 1,\ldots,n$, the point $\boldsymbol{x}_{i,u} \in [0,1]^{|u|}$ has those components from the point $\boldsymbol{x}_i \in [0,1]^d$. The hybrid point $\boldsymbol{y} = \boldsymbol{x}_u{:}\boldsymbol{z}_{-u}$ has $y_j = x_j$ for $j \in u$ and $y_j = z_j$ for $j \notin u$.

The differential $\mathrm{d}\boldsymbol{x}_u$ is $\prod_{j\in u}\mathrm{d}x_j$. Similarly, $\partial^u f$ denotes $\partial^{|u|}f/\prod_{j\in u}\partial x_j$ and, by convention $\partial^{\varnothing} f$ is f. The functions we consider belong to

$$\mathcal{F} = \{f : [0,1]^d \to \mathbb{R} \mid \partial^{1:d} f \text{ is continuous on } [0,1]^d\}. \tag{1}$$

Continuity here allows the partial derivatives to be taken in any order and it allows some interchanges of order between differentiation and integration. Our space is smaller than the Sobolev spaces usually studied, so we call it a pre-Sobolev space. A Sobolev space also includes functions with derivatives in the sense of distributions which then implies completeness [28]. We need the somewhat stronger condition (1) in order to apply a Poincaré inequality. Many applications have an integrand that satisfies (1).

2.1. Weighted spaces. Let $\gamma_u > 0$ for all $u \subseteq 1{:}d$ and let $\boldsymbol{\gamma}$ comprise all of those choices for γ_u. We use an unanchored norm defined by

$$\|f\|_{\boldsymbol{\gamma}}^2 = \sum_{u\subseteq 1:d} \frac{1}{\gamma_u} \int_{[0,1]^{|u|}} \left(\int_{[0,1]^{d-|u|}} \partial^u f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_{-u}\right)^2 \mathrm{d}\boldsymbol{x}_u. \tag{2}$$

See [5] for background on this and a related anchored norm. This norm is finite for every choice of $\boldsymbol{\gamma}$ and all $f \in \mathcal{F}$. Some of our results allow a seminorm instead found by dropping $u = \varnothing$ from the sum or, equivalently, taking $\gamma_\varnothing = \infty$. More generally, the $u = \varnothing$ term is $\mu^2/\gamma_\varnothing$, where $\mu = \int f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$. There are now many efficient methods of constructing QMC point sets for weighted spaces. See [43, 32, 33].

Numerous choices of weights are given in the survey [5], and a comprehensive treatment is available in [31]. Sloan and Woźniakowski [45] use product weights $\gamma_u = \prod_{j\in u}\gamma_j$ including $\gamma_\varnothing = 1$. Typically $1 = \gamma_1 \geqslant \gamma_2 \geqslant \cdots \geqslant \gamma_j \geqslant \gamma_{j+1} \cdots > 0$. Hickernell [11] uses weights $\gamma_u = \gamma^{|u|}$ for some $\gamma \in (0,1)$. Such weights are commonly called order weights. The more general order weights in [7] take the form $\gamma_u = \Gamma_{|u|}$, where $\Gamma_r \geqslant 0$ is a nonincreasing function of r. The case with $\Gamma_r = 0$ for all $r \geqslant r_0$, known as finite-order weights, was studied in [44]. However, Sloan [42] points out a danger from this choice. We will suppose that $\Gamma_r > 0$. Dick et al. [7] also consider completely general weights γ_u, but such generality sharply raises the costs of using those weights to design an algorithm [5]. Product and order weights, known as POD weights, defined by $\gamma_u = \Gamma_{|u|}\prod_{j\in u}\gamma_j$, have proved useful in QMC-based algorithms for solving PDEs with random coefficients [22]. One notable choice has $\gamma_u = (|u|!)^\alpha \times \prod_{j\in u} j^{-\beta}$, where $0 < \alpha < \beta$. Note that with this choice, $\Gamma_{|u|} = (|u|!)^\alpha$ is increasing in $|u|$.

Higher weights are placed on the more important subsets. Other things being equal, subsets with fewer components are usually considered more important as are subsets containing components with lower indices.

We partially order subsets by $|u|$ and we use a parallel notation $\lceil u\rceil = \max\{j \mid j \in u\}$ to order subsets by their largest element. By convention, $\lceil\varnothing\rceil = 0$. Two frequently satisfied conditions on the weights are

$$|u| \geqslant s \implies \gamma_u \leqslant \gamma_{1:s} \quad \text{and} \tag{3}$$

$$\lceil u\rceil \geqslant s \implies \gamma_u \leqslant \gamma_{\{s\}}. \tag{4}$$

For instance, (3) holds if the highest weighted subset of cardinality s is $1{:}s$ and the associated weight $\gamma_{1:s}$ is nonincreasing in s. Similarly, (3) holds if the highest weighted subset with maximal element s is $\{s\}$ and the associated weights $\gamma_{\{s\}}$ are nonincreasing in s.

Condition (4) implies that $\gamma_u \leqslant \gamma_{\{1\}}$ for all $u \neq \varnothing$. This also holds for the POD weights described above. Regarding $1/\gamma_u$ as a multiplicative penalty factor, the singleton $\{1\}$ is the "least penalized" index subset, though it may not be uniquely least penalized. Many but not all weights in use satisfy $\gamma_u \geqslant \gamma_v$ when $u \subseteq v$.

Most of the widely studied weights satisfy both (3) and (4). The POD weights $\gamma_u = (|u|!)^\alpha \times \prod_{j \in u} j^{-\beta}$ are an exception. For those weights, $\gamma_{u \cup \{1\}} > \gamma_u$ whenever $1 \notin u$.

2.2. Function decompositions. This section gives notation for two function decompositions. The first is the ANOVA decomposition of $L^2\bigl([0,1]^d\bigr)$. The second is an anchored decomposition.

The ANOVA decomposition was introduced independently in [15] and [47]. The decomposition extends naturally to any mean square integrable function of d independent random inputs, and $d = \infty$ is allowed [34]. There is additional background in [36]. In this decomposition $f(\boldsymbol{x}) = \sum_{u \subseteq 1:d} f_u(\boldsymbol{x})$, where f_u depends on $\boldsymbol{x}$ only through $\boldsymbol{x}_u$. The ANOVA uniquely satisfies $\int_0^1 f_u(\boldsymbol{x})\,\mathrm{d}x_j = 0$ whenever $j \in u$. For $u \neq \varnothing$, define the variance component $\sigma_u^2 = \int_{[0,1]^d} f_u(\boldsymbol{x})^2\,\mathrm{d}\boldsymbol{x}$, and take $\sigma_\varnothing^2 = 0$. Then the variance of f decomposes as

$$\sigma^2 \equiv \int_{[0,1]^d} (f(\boldsymbol{x}) - \mu)^2\,\mathrm{d}\boldsymbol{x} = \sum_{u \subseteq 1:d} \sigma_u^2.$$

A useful alternative to the ANOVA is the anchored decomposition, $f(\boldsymbol{x}) = \sum_u f_u^*(\boldsymbol{x})$, from [47]. See [21] for a unified presentation of this and other decompositions. One picks an anchor point $\boldsymbol{a} \in [0,1]^d$ and then $f_u^*(\boldsymbol{x})$ depends on $\boldsymbol{x}$ only through the values of $f(\boldsymbol{x}_v{:}\boldsymbol{a}_{-v})$ for $v \subseteq u$, and if $x_j = a_j$ for any $j \in u$ then $f_u^*(\boldsymbol{x}) = 0$. For instance $f_\varnothing^*(\boldsymbol{x}) = f(\boldsymbol{a})$, while for $j \in 1{:}d$, $f_{\{j\}}^*(\boldsymbol{x}) = f(\boldsymbol{x}_{\{j\}}{:}\boldsymbol{a}_{-\{j\}}) - f(\boldsymbol{a})$ and for $1 \leqslant j < k \leqslant d$,

$$f_{\{j,k\}}^*(\boldsymbol{x}) = f(\boldsymbol{x}_{\{j,k\}}{:}\boldsymbol{a}_{-\{j,k\}}) - f(\boldsymbol{x}_{\{j\}}{:}\boldsymbol{a}_{-\{j\}}) - f(\boldsymbol{x}_{\{k\}}{:}\boldsymbol{a}_{-\{k\}}) + f(\boldsymbol{a}).$$

The generalization to larger $|u|$ is given in [21]. The anchored decomposition is especially useful for integrands with large or infinite d where any function evaluations can only change finitely many inputs from a default value which then serves as a natural anchor. The function f_u^* can be evaluated through $O(2^{|u|})$ evaluations of f. In some cost models, the cost to evaluate $f(\boldsymbol{x})$ increases with the number of components x_j that differ from a_j.

3. Literature on effective dimension. This section is a brief survey of the literature on effective dimension. Effective dimension is used to explain how QMC can be superior to Monte Carlo in high dimensional integration problems. It can be used as a postmortem to explain why QMC did or did not bring an improvement in a problem. Also, other things being equal, reducing the effective dimension using methods such as those in [1, 16] is expected to improve QMC performance. Finally, the approach from [23] that is now called the multivariate decomposition method (MDM) uses notions of effective dimension to devise algorithms for integration and approximation.

The notion of effective dimension seems to be as old as QMC itself. QMC as we know it was introduced in 1951 by Richtmyer [40] working at Los Alamos. He used what we now call Kronecker sequences to study neutron chain reactions. Here is what he said about effective dimension on page 13:

> "The peculiarity of the integrands in question is that although k may be large (e.g., > 20) and, in fact, is really indefinite, as noted above, the effective number of dimensions is smaller (say, 4 or 5). (The effective number of dimensions may be defined as follows: at point $(x_1, x_2, \ldots, x_k)$ let s be the smallest integer such that $f(x_1, x_2, \ldots, x_s, x'_{s+1}, \ldots, x'_k)$ is independent of $x'_{s+1}, \ldots, x'_k$. Then the average of s over the unit cube is the effective number of dimensions for the integrand f.)"

His k is the nominal dimension. Inputs after x_s are completely ignored by f, although that index s depends on $\boldsymbol{x}$. This notion is close to the modern notion of truncation dimension described below. In Richtmyer's application, one could tell for a given sequence $x_1, \ldots, x_s$ that subsequent inputs could not make any difference, so his measure can be estimated by sampling. Incidentally, Richtmyer's abstract is pessimistic about the prospects for QMC to improve upon Monte Carlo, based on his numerical experience. We can now see that his integrands were not smooth enough for him to benefit greatly from QMC, despite their low effective dimension.

Paskov [38] used QMC points to estimate the expected payouts of ten tranches in a collateralized mortgage obligation (CMO). The integrands had a nominal dimension of 360, corresponding to one random interest rate value per month in a 30 year model. He noticed that only one of the ten integrands seemed to use all 360 inputs. The least number used by any tranche was 77. The number of inputs used is much like Richtmyer's definition, except that Paskov is interested in the number used anywhere in the cube while Richtmyer is averaging the number used over the cube.

Using d for nominal and k for effective dimension, Paskov goes on to define the effective dimension to be the smallest k for which

$$\left|\int_{[0,1]^k} f(x_1, \ldots, x_k, 0, \ldots, 0)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k - \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \leqslant \varepsilon \left|\int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right| \tag{5}$$

holds. This definition allows for an x_{k+1} to be used by f, but with negligible impact. It is also a form of truncation dimension. He used $\varepsilon = 0.001$. He estimated the effective dimension for the ten tranches, getting values between 42 and 338. These values are not especially small. The choice of trailing zeros inside f in (5) appears odd given that the CMO used Gaussian random variables. Perhaps trailing 1/2's were actually used within $[0,1]^d$, corresponding to trailing zeros for the resulting Gaussian variables.

By comparing error variances from Latin hypercube sampling [27, 52] and plain Monte Carlo, Caflisch, Morokoff and Owen [3] infer that about 99.96% of the variance for an integrand in a CMO model comes from a sum of 360 functions of just one variable each. They introduced two notions of effective dimension in order to explain the success of QMC on high dimensional problems from financial valuation. Let f have ANOVA effects f_u with variances σ_u^2 for $u \subseteq 1{:}d$. Then f has effective dimension s in the truncation sense if s is the smallest integer with

$$\sum_{u:\lceil u\rceil \leqslant s} \sigma_u^2 \geqslant 0.99\sigma^2. \tag{6}$$

Like the notions of Richtmyer and of Paskov, this quantity describes the importance of the first s compnents of $\boldsymbol{x}$. It has effective dimension s in the superposition sense if s is the smallest integer with

$$\sum_{u:|u|\leqslant s} \sigma_u^2 \geqslant 0.99\sigma^2. \tag{7}$$

The CMO integrand mentioned above has effective dimension 1 in the superposition sense.

The arbitrary constant 0.99 may be explained as follows. If we use plain Monte Carlo with n observations then our variance is $\sigma^2/n = \sum_u \sigma_u^2/n$. If more than 99% of the variance comes from some subset of effects f_u then a method such as QMC which can integrate them at a better rate of convergence has a possibility of attaining a 100-fold improvement over Monte Carlo, even if it is no better than Monte Carlo for the other effects, for a given sample size n. Users can reasonably ignore a method that is say twice as fast, due to trade-offs in implementation difficulty or even familiarity. When a 100-fold improvement is available, it should be harder to ignore.

By writing $f(\boldsymbol{x}) = \sum_{u\subseteq 1:d} f_u(\boldsymbol{x})$, where f_u depends on $\boldsymbol{x}$ only through $\boldsymbol{x}_u$, we can bound the QMC error by applying the Koksma–Hlawka inequality [30] term by term. The integration error is at most

$$\sum_{u\subseteq 1:d, u\neq\varnothing} D_n^*(\boldsymbol{x}_{1,u},\dots,\boldsymbol{x}_{n,u}) V_{\mathrm{HK}}(f_u), \tag{8}$$

where V_{HK} is the Hardy–Krause variation of f_u and D_n^* is the star discrepancy of some $|u|$-dimensional points. Equation (8) shows how the notion of superposition dimension can explain the success of QMC in high dimensional problems. It is common that QMC points have very small values of D_n^* above when $|u|$ is small. Then all we need is for $V_{\mathrm{HK}}(f_u)$ to be small for large $|u|$, that is, for f to be dominated by its low dimensional parts.

The Hardy–Krause variations in (8) have no practical estimates. The superposition dimension (7) is essentially using variance components σ_u^2 as a proxy. While $V_{\mathrm{HK}}(f_u)$ and σ_u both measure the magnitude of f_u, only V_{HK} captures the smoothness that QMC exploits.

The mean dimension [35] of a (nonconstant) function in the superposition sense is $\sum_u |u|\sigma_u^2/\sigma^2$. This quantity varies continuously, and so it can distinguish between integrands that have the same superposition dimension (7). It is easier to estimate than the effective dimension because [26] shows that it is a simple sum of Sobol' indices [48].

The above definitions of effective dimension apply to integrands taken one at a time. We are also interested in notions of effective dimension that apply to spaces of functions. The first such notion was due to Hickernell [12]. Letting f be randomly drawn from a space of functions he says that those functions are proportion p (such as $p = 0.99$) of truncation dimension s if $\sum_{\lceil u\rceil\leqslant s} \mathbb{E}(\sigma_u^2(f)) = p\mathbb{E}(\sigma^2(f))$, where the expectation $\mathbb{E}(\cdot)$ is with respect to random f. They are proportion p of superposition dimension s if $\sum_{|u|\leqslant s} \mathbb{E}(\sigma_u^2(f)) = p\mathbb{E}(\sigma^2(f))$. He then develops expressions for these proportions in terms of a certain shift-invariant kernel derived from the covariance kernel of his random functions.

Wang and Fang [54] defined a different notion of effective dimension for spaces. They consider Korobov spaces of periodic functions on $[0,1)^d$. Those spaces are reproducing kernel Hilbert spaces. Letting the kernel be $K(\boldsymbol{x},\boldsymbol{x}')$ for $\boldsymbol{x},\boldsymbol{x}'\in[0,1)^d$, they define the "typical functions" in that space to be $f_{\boldsymbol{x}'}(\boldsymbol{x}) = K(\boldsymbol{x},\boldsymbol{x}')$ for some $\boldsymbol{x}'\in[0,1)^d$. For product weights, the effective dimension of these typical functions, in either the superposition or truncation senses, is independent of $\boldsymbol{x}'$. They adopt that common effective dimension as the effective dimension of the space.

The bound in (8) applies also for the anchored decomposition $f = \sum_u f_u^*$ or, indeed, for any of the decompositions in [21]. The series of papers [23, 39, 55] introduced an ε-superposition dimension $d(\varepsilon)$ based on the anchored decomposition. Their MDM is an algorithm that integrates only some of the terms in the anchored decomposition. The algorithm can attain a worst case error of at most ε for any f in the unit ball of a weighted Sobolev space, and it does so using only f_u^* for $|u| \leqslant d(\varepsilon)$. Lemma 1 of [39] shows that $d(\varepsilon) = O(\log(1/\varepsilon)/\log(\log(1/\varepsilon)))$ as $\varepsilon \to 0$ for product weights with sufficiently fast decay. Gilbert and Wasilkowski [8] develop some constructions to identify the sets u needed in approximation. In so doing they also show that only $O(\varepsilon^{-1})$ such sets are needed.

Equation (10) of [39] gives an expression for $d(\varepsilon)$. Definitions of $d(\varepsilon)$ are given in equation (9) of [23] (see also (25)) and equation (14) of [39]. The precise definitions depend on numerous additional constants describing error bounds for integration with respect to $\boldsymbol{x}_u$ and how to trade off between two sources of error that contribute to an error bound.

The technical report [37] is a predecessor of the present paper. It investigates some weighted spaces where much better performance than Monte Carlo could be attained despite high nominal dimension. It showed that none of the integrands involved could have large ANOVA contributions from high dimensional parts. We postpone statements of these conditions and results to section 5. A serious weakness in that paper was that the results only apply to periodic functions. The present paper removes that weakness.

Kritzer, Pillichshammer, and Wasilkowski [17] define a truncation dimension counterpart to the $d(\varepsilon)$ quantity given above, using the anchored decomposition. In their definition, the space has effective dimension s if the difference between an estimate computed using just s inputs and a computation using all inputs is below some small multiple ε of the norm of a ball in that weighted space. The bound still holds when the "all inputs" computation is hypothetical, as it could be for infinite dimensional problems.

The IBC literature defines effective dimension through active sets. Equation (4) of [8] (see also [14]) defines the set $\mathcal{A} \subseteq 1{:}d$ to be an active set if

$$\left|\mathcal{S}\left(\sum_{u \notin \mathcal{A}(\varepsilon)} f_u^*\right)\right| \leqslant \varepsilon \left\|\sum_{u \notin \mathcal{A}(\varepsilon)} f_u^*\right\|_{\mathcal{F}} \quad \forall f \in \mathcal{F}, \tag{9}$$

where $\mathcal{S}$ is the operator of interest, such as integration over the unit cube. They define the ε-superposition dimension of f to be

$$d^{\mathrm{SPR}}(\varepsilon) = \min\bigl\{\#\mathcal{A} \mid \mathcal{A} \text{ satisfies } (9)\bigr\}. \tag{10}$$

These definitions can be used on any decomposition [58, 56] and they apply to very general operators. They do however have a surprising consequence when applied to integration for the ANOVA decomposition. Because $\int f_u(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = 0$ for $u \neq \varnothing$ we find that $\mathcal{A} = \{\varnothing\}$ is an active set for any $\varepsilon > 0$ and then we always get $d^{\mathrm{SPR}}(\varepsilon) = 0$. Intuitively, if we could actually compute the truncated ANOVA decomposition, we would have the exact answer from $f_\varnothing$ and get superposition or truncation dimension zero. Effective dimension for the approximation operator (discussed below) does not vanish like this.

Recently, Sobol' and Shukhman [51] look at some particle transport problems, similar to what Richtmyer studied. They show that those problems typically have low mean dimension.

Our focus here is on integration and an L^2 approach that allows comparisons to Monte Carlo. Effective dimension has also been studied for approximation problems. See [57, 19] and references therein. The latter reference includes a truncation dimension for approximation. See also [14], which considered general linear operators. There has been some work outside of L^2 on Banach spaces [18]. The results we present are for unanchored spaces. Many of the papers in IBC work instead with anchored spaces. Hefter and Ritter [10] establish bounds on the norms of embeddings between anchored spaces and the unanchored spaces we consider here. Such embeddings allow error bounds in unanchored spaces to be translated into bounds on anchored spaces, although one then has to keep track of the norms of those embedding operators.

In the IBC literature on L^2 approximation, the operator $\mathcal{S}$ is the identity $\mathcal{S}(f) = f$. In this context (9) becomes, when using the ANOVA decomposition and the orthogonality of ANOVA terms,

$$\Big\| \sum_{u \notin \mathcal{A}} f_u \Big\|_{L^2} = \Big(\sum_{u \notin \mathcal{A}} \sigma_u^2 \Big)^{1/2} \leqslant \varepsilon \Big\| \sum_{u \notin \mathcal{A}} f_u \Big\|_{\mathcal{F}} \quad \forall f \in \mathcal{F}. \tag{11}$$

Later we will see that this notion of effective dimension for L^2 approximation coincides with the ones in this paper for integration after a minor reparameterization of ε.

The literature has a third sense in which functions can have a low effective dimension. L'Ecuyer and Lemieux [25] give a "successive dimensions" sense in which the scale of nonempty $u \subseteq 1{:}d$ is given by $1 + \max\{j \mid j \in u\} - \min\{j \mid j \in u\}$ and $\varnothing$ has scale 0. A reviewer points out that effective dimension in a generalized sense can be defined through a nested sequence of $M + 1$ sets of subsets of $1{:}d$ given by $\varnothing = U_0 \subset U_1 \subset U_2 \subset \cdots \subset U_M = 2^{1:d}$. Then the general notion of effective dimension is

$$\min\Big\{ s \mid \sum_{u \notin U_s} \sigma_u^2(f) < \varepsilon \sigma^2(f) \Big\}.$$

4. Some Poincaré-type inequalities. Our main tool will be bounds on L^2 norms based on integrated squared derivatives. This section presents them with some history. These are generally known as Poincaré inequalities. Poincaré worked with integrals over more general domains than $[0, 1]$ as well as more general differential operators than those used here.

THEOREM 4.1. *Let f be differentiable on the finite interval (a, b) and satisfy $\int_a^b f(x)\,\mathrm{d}x = 0$. Then*

$$\int_a^b f'(x)^2\,\mathrm{d}x \geqslant \Big(\frac{\pi}{b-a}\Big)^2 \int_a^b f(x)^2\,\mathrm{d}x, \tag{12}$$

and equality is attained for some nonzero f.

Proof. The proof is on pages 295–296 of Stekloff [53]. ∎

The constant π in (12) will appear throughout our formulas. Things would be different if we were to work on the interval $[0, \pi]$ or $[0, 2\pi]$. We retain a focus on $[0, 1]$ because weighted spaces are almost always defined over $[0, 1]^d$.

Theorem 4.1 is commonly known as Wirtinger's theorem, though Stekloff's work was earlier. Sobol' [46] has a different proof than Stekloff, based on the calculus of variations. This theorem is often given with an additional condition that $f(a) = f(b)$ (periodicity), though such a condition is not necessary.

This inequality could be much older than 1901; see [29, Chapter 2]. The condition $\int_a^b f(x)\,\mathrm{d}x = 0$ can be removed if $f(a) = f(b) = 0$. Equality is attained if $f(x)$ is a multiple of $\sin(\pi(x-m)/s)$ for $m = (a+b)/2$ and $s = b-a$.

Lemma 4.2. *Let $f(\boldsymbol{x})$ defined on $[0,1]^d$ satisfy $\int_0^1 f(\boldsymbol{x})\,\mathrm{d}x_j = 0$ for $j = 1,\dots,d$. If $d \geqslant r \geqslant 0$ and $\partial^{1:r} f$ exists then*

$$\int_{[0,1]^d} \Bigl(\frac{\partial^r}{\partial x_1 \cdots \partial x_r} f(\boldsymbol{x})\Bigr)^2 \,\mathrm{d}\boldsymbol{x} \geqslant \pi^{2r} \int_{[0,1]^d} f(\boldsymbol{x})^2 \,\mathrm{d}\boldsymbol{x}. \tag{13}$$

Proof. This holds for $r = 0$ by convention. It holds for $d = r = 1$ by Theorem 4.1. It extends to $d \geqslant r \geqslant 1$ by induction. ∎

Theorem 4.3. *Let f be defined on $[0,1]^d$ with $\partial^{1:d} f$ continuous. Furthermore let f have ANOVA effects f_v for $v \subseteq 1{:}d$ with variance components σ_v^2. Then*

$$\|f\|_{\boldsymbol{\gamma}}^2 \geqslant \gamma_\varnothing^{-1}\mu^2 + \sum_{u\neq\varnothing} \gamma_u^{-1}\pi^{2|u|}\sigma_u^2. \tag{14}$$

Proof. Let f have ANOVA effects f_v for $v \subseteq 1{:}d$. Then for $u \neq \varnothing$,

$$\begin{aligned}\int_{[0,1]^{d-|u|}} \partial^u \sum_{v\subseteq 1:d} f_v(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_{-u} &= \partial^u \int_{[0,1]^{d-|u|}} \sum_{v\supseteq u} f_v(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_{-u}\\ &= \partial^u \int_{[0,1]^{d-|u|}} f_u(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}_{-u} = \partial^u f_u(\boldsymbol{x}_u{:}\boldsymbol{0}_{-u}),\end{aligned}$$

because f_u does not depend on $\boldsymbol{x}_{-u}$. Next

$$\int_{[0,1]^{|u|}} (\partial^u f_u(\boldsymbol{x}_u{:}\boldsymbol{0}_{-u}))^2\,\mathrm{d}\boldsymbol{x}_u = \int_{[0,1]^d} (\partial^u f_u(\boldsymbol{x}))^2\,\mathrm{d}\boldsymbol{x} \geqslant \pi^{2|u|}\sigma_u^2$$

by Lemma 4.2. The result follows from (2). ∎

5. Effective dimension of a space. We begin with a ball of real-valued functions on $[0,1]^d$ given by

$$\mathcal{B}(\boldsymbol{\gamma},\rho) = \{f \in \mathcal{F} \mid \|f\|_{\boldsymbol{\gamma}} \leqslant \rho\},$$

where $\rho > 0$. We pick the radius ρ of the ball to make it just large enough to contain a function of unit variance. That radius is

$$\rho^* = \rho^*(\boldsymbol{\gamma}) = \inf\bigl\{\rho > 0 \mid \exists f \in \mathcal{B}(\boldsymbol{\gamma},\rho) \text{ with } \sigma^2(f) = 1\bigr\}. \tag{15}$$

Monte Carlo sampling of a function $f \in \mathcal{B}(\boldsymbol{\gamma},\rho^*)$ can have a root mean squared error as high as $1/\sqrt{n}$.

The ball is said to be of effective dimension s_T in the truncation sense at level $1-\varepsilon$ (such as $\varepsilon = 0.01$) if the sum of σ_u^2 for all sets u that include any index $j > s_T$ is below ε for every $f \in \mathcal{B}(\boldsymbol{\gamma},\rho^*)$, and if the same is not true of $s_T - 1$. That is,

$$s_T = \min\Biggl\{s \in \mathbb{N} \mid \sup_{f\in\mathcal{B}(\boldsymbol{\gamma},\rho^*)} \sum_{u:\lceil u\rceil > s} \sigma_u^2(f) < \varepsilon\Biggr\}, \tag{16}$$

or, equivalently,

$$\sup_{f \in \mathcal{B}(\boldsymbol{\gamma},\rho^*)} \sum_{u:\lceil u \rceil > s_T} \sigma_u^2(f) < \varepsilon \leqslant \sup_{f \in \mathcal{B}(\boldsymbol{\gamma},\rho^*)} \sum_{u:\lceil u \rceil \geqslant s_T} \sigma_u^2(f). \tag{17}$$

For superposition dimension, we replace the largest index by the cardinality. The ball $\mathcal{B}(\boldsymbol{\gamma}, \rho^*)$ is of effective dimension s_S in the superposition sense at level $1-\varepsilon$ for

$$s_S = \left\{ \min s \in \mathbb{N} \mid \sup_{f \in \mathcal{B}(\boldsymbol{\gamma},\rho^*)} \sum_{u:|u|>s} \sigma_u^2(f) < \varepsilon \right\}. \tag{18}$$

If the ball is of low effective dimension then no function in it contains any non-negligible high order components, where order is quantified by $|u|$ or $\lceil u \rceil$. Here $\varepsilon = 0.01$ corresponds to the usual choice, but we will want to see how effective dimension depends on ε.

Normalizing to variance one is interpretable, but not really necessary. We can work with ratios of norms. An equivalent condition to (16) is that s_T is the smallest integer s for which

$$\sup_{f:\, 0<\|f\|_{\boldsymbol{\gamma}}<\infty} \frac{\sum_{u:\lceil u \rceil \geqslant s+1} \sigma_u^2(f)}{\|f\|_{\boldsymbol{\gamma}}^2} < \frac{\varepsilon}{\rho^{*2}}.$$

This definition matches the d^{SPR} in the IBC literature (see (10) and (11)) modulo a reparameterization of ε.

PROPOSITION 5.1. *Let weights γ_u be such that $\gamma_{\{1\}}\pi^{-2} \geqslant \gamma_u \pi^{-2|u|}$ for all nonempty $u \subseteq 1{:}d$. Then the smallest ρ for which $\mathcal{B}(\boldsymbol{\gamma}, \rho)$ contains a function of variance 1 is $\rho^*(\boldsymbol{\gamma}) = \pi(\gamma_{\{1\}})^{-1/2}$.*

Proof. Let f have $\sigma^2(f) = 1$ and suppose that $\rho < \rho^*$. Then by Theorem 4.3,

$$\|f\|_{\boldsymbol{\gamma}}^2 \geqslant \mu^2 \gamma_{\varnothing}^{-1} + \min_{u \neq \varnothing} \gamma_u^{-1} \pi^{2|u|} \geqslant \gamma_{\{1\}}^{-1} \pi^2 = (\rho^*)^2 > \rho^2,$$

so $f \notin \mathcal{B}(\boldsymbol{\gamma}, \rho)$. If $f(\boldsymbol{x}) = \sqrt{2}\sin(\pi(x_1 - 1/2))$ then $\sigma^2(f) = 1$ and $f \in \mathcal{B}(\boldsymbol{\gamma}, \rho^*)$. ∎

Weights with $\gamma_u \leqslant \gamma_{\{1\}}$ for $u \neq \varnothing$ automatically satisfy the condition in Proposition 5.1.

PROPOSITION 5.2. *For weights $\boldsymbol{\gamma}$, let $\rho^*(\boldsymbol{\gamma})$ be defined by* (15)*, let $u \subseteq 1{:}d$ be nonempty, and let $f \in \mathcal{B}(\boldsymbol{\gamma}, \rho^*(\boldsymbol{\gamma}))$. Then*

$$\sigma_u^2(f) \leqslant \rho^*(\boldsymbol{\gamma})^2 \gamma_u \pi^{-2|u|}.$$

If also $\gamma_{\{1\}}\pi^{-2} \geqslant \gamma_u \pi^{-2|u|}$ then

$$\sigma_u^2(f) \leqslant \pi^{-2(|u|-1)} \gamma_u / \gamma_{\{1\}}.$$

Proof. From Theorem 4.3, $\rho^{*2} \geqslant \|f\|_{\boldsymbol{\gamma}}^2 \geqslant \gamma_u^{-1}\pi^{2|u|}\sigma_u^2$, establishing the first claim. Then the second one follows from Proposition 5.1. ∎

6. Bounds on effective dimension. In this section we derive bounds on effective dimension under assumption (3) where any set u of cardinality s or larger has $\gamma_u \leqslant \gamma_{1:s}$. We also consider assumption (4) where any set u containing an index s or larger has $\gamma_u \leqslant \gamma_{\{s\}}$.

THEOREM 6.1. *Let the weights $\boldsymbol{\gamma}$ satisfy* (3). *Then the corresponding pre-Sobolev space has effective dimension in the superposition sense at level* $1-\varepsilon$ *no larger than*

$$\max\{s \geqslant 1 \mid \gamma_{1:s} \geqslant \pi^{2(s-1)}\gamma_{\{1\}}\varepsilon\}. \tag{19}$$

If the weights $\boldsymbol{\gamma}$ satisfy (4) *then the corresponding pre-Sobolev space has effective dimension in the truncation sense at level* $1-\varepsilon$ *no larger than*

$$\max\{s \geqslant 1 \mid \gamma_{\{s\}} \geqslant \gamma_{\{1\}}\varepsilon\}. \tag{20}$$

Proof. Let $f \in \mathcal{B}(\boldsymbol{\gamma}, \rho^*)$, where ρ^* is given by (15). Let f have ANOVA variance components σ_u^2. Choose an integer $s > 0$. From Proposition 5.1, Theorem 4.3, and condition (3),

$$\gamma_{\{1\}}^{-1}\pi^2 = \rho^{*2} \geqslant \|f\|_{\boldsymbol{\gamma}}^2 \geqslant \mu^2\gamma_{\varnothing}^{-1} + \sum_{u \neq \varnothing} \gamma_u^{-1}\pi^{2|u|}\sigma_u^2 \geqslant \gamma_{1:s}^{-1}\pi^{2s}\sum_{|u|\geqslant s}\sigma_u^2.$$

Therefore $\sum_{|u|\geqslant s}\sigma_u^2 \leqslant \gamma_{1:s}\pi^{-2(s-1)}/\gamma_{\{1\}}$. If $\gamma_{1:s} < \gamma_{\{1\}}\pi^{2(s-1)}\varepsilon$ then $\sum_{|u|\geqslant s}\sigma_u^2 < \varepsilon$ and the effective dimension cannot be as large as s. This establishes the superposition bound (19).

For the truncation dimension, we find that $\gamma_{\{1\}}^{-1}\pi^2 \geqslant \gamma_{\{s\}}^{-1}\pi^2\sum_{\lceil u\rceil \geqslant s}\sigma_u^2$ by a similar argument to the one used above, this time using (4). Then (20) follows just as (19) did. $\square$

6.1. Tractability and product weights. Here we look at product weights of the form $\gamma_u = \prod_{j\in u}\gamma_j$ for monotone values $\gamma_j \geqslant \gamma_{j+1}$ for $j \geqslant 1$, including $\gamma_\varnothing = 1$. Sloan and Woźniakowski [45] give conditions on the weights for high dimensional quadrature to be tractable, which we define next. They consider a sequence of d-dimensional settings in which $d \to \infty$. We will look at their weights restricted to $u \subseteq 1{:}d$. We draw on the summary of tractability results given by [20].

Suppose that $f \in \mathcal{B}(\boldsymbol{\gamma}, \rho)$ for $0 < \rho < \infty$. If we had to pick an $n = 0$ point rule for functions in $\mathcal{B}(\boldsymbol{\gamma}, \rho)$ it would be a constant and we minimize the worst case error by taking that constant to be 0. Our initial error is then $\sup_{f\in\mathcal{B}(\boldsymbol{\gamma},\rho)}\left|\int f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}\right|$. Now let $n = n_{\boldsymbol{\gamma}}(\varepsilon, d)$ be the smallest integer for which some QMC rule reduces the initial error by a factor of ε. This n does not depend on our choice ρ.

The problem of quadrature is defined to be tractable if there exist points $\boldsymbol{x}_i$ with $n_{\boldsymbol{\gamma}}(\varepsilon, d) \leqslant Cd^q\varepsilon^{-p}$ for nonnegative constants C, p, and q. If $q = 0$ is possible then the cost n can be taken independent of dimension d, and the problem is then said to be strongly tractable. The problem of quadrature is strongly tractable if $\sum_{j=1}^{\infty}\gamma_j < \infty$ [45]. That result was nonconstructive and it had $p = 2$, comparable to plain Monte Carlo. Hickernell and Woźniakowski [13] gave an improved nonconstructive proof showing errors $n^{-1+\delta}$ are possible (so $p = 1/(1-\delta)$) if

$$\sum_{j=1}^{\infty}\gamma_j^{1/2} < \infty. \tag{21}$$

Constructions attaining those rates have been found [43, 32, 33].

Now suppose that an error of $O(n^{-1+\delta})$ is attainable by QMC in a ball $\mathcal{B}$ where Monte Carlo would have root mean square error $O(n^{-1/2})$. If that ball contains functions with significantly large high dimensional interactions then the QMC method must be successfully integrating some high dimensional functions. If the ball has no functions with high dimensional components then the success of QMC is attributable to its performance on low dimensional functions.

Table 1
Upper bounds on effective dimension for product weights defined by $\gamma_j = j^{-\eta}$.

	Truncation			Superposition		
ε	$\eta = 2$	$\eta = 1$	$\eta = 0$	$\eta = 2$	$\eta = 1$	$\eta = 0$
0.1	3	10	∞	1	1	2
0.01	9	100	∞	2	2	3
0.001	31	1000	∞	2	3	4
0.0001	99	10000	∞	3	3	5

6.2. Effective dimension. We consider weight factors of the form $\gamma_j = j^{-\eta}$ for various η. Some "phase transitions" happen at special values of η. For any $\eta > 2$, (21) holds and then there exist QMC points with worst case errors that decrease at the rate $O(n^{-1+\delta})$ for any $\delta > 0$. For any $\eta > 1$, strong tractability holds but not at a better rate than Monte Carlo provides. These weight choices always have $\gamma_1 = 1$ as [45] does. Taking $\gamma_1 > 1$ would yield $\gamma_{\{1\}} > \gamma_\varnothing$ which is incompatible with the customary rule that smaller sets get higher weight.

From Theorem 6.1 we see that the truncation dimension satisfies

$$s_T(\varepsilon) \leqslant \max\{s \geqslant 1 \mid s^{-\eta} \geqslant \varepsilon\} \tag{22}$$

and the superposition dimension satisfies

$$s_S(\varepsilon) \leqslant \max\{s \geqslant 1 \mid (s!)^{-\eta} \geqslant \pi^{2(s-1)}\varepsilon\}. \tag{23}$$

Both of these effective dimensions are nonincreasing with η.

Lemma 6.2. *For $\eta > 0$ the dimensions given in* (22) *and* (23) *satisfy $s_T \leqslant \varepsilon^{-1/\eta}$ and $s_S = O(\log(1/\varepsilon)/\log(\log(1/\varepsilon)))$ as $\varepsilon \to 0$.*

Proof. The first claim is immediate because $s_T^{-\eta} \geqslant \varepsilon$. For the second claim, s_S satisfies $(s!)^{-\eta} \geqslant \varepsilon$. For any $\lambda \in (0,1)$ and very small ε we get s large enough that $s! \geqslant s^{\lambda s}$. Then $\varepsilon^{-1/\eta} \geqslant s^{\lambda s} = e^{\lambda s \log(s)}$ and so $s\log(s) \leqslant \log(\varepsilon^{-1})/(\lambda\eta)$. Then $w \leqslant \log(A)/W(A)$, where $A = \log(\varepsilon^{-1})/(\lambda\eta)$ and W is the principal branch of the Lambert function that solves $W(x)e^{W(x)} = x$. To complete the proof, we note that $W(x)$ is asymptotic to $\log(x)$ as $x \to \infty$ from [4]. ∎

The convergence rate in Lemma 6.2 is the same as in Lemma 1 of [39] for $d(\varepsilon)$ in MDM. From the proof of Lemma 6.2 we find that

$$\lim_{\varepsilon\to 0} s_S(\varepsilon) \Big/ \frac{\log(1/\varepsilon)/(\eta\lambda)}{\log(\log(1/\varepsilon)/(\eta\lambda))} \leqslant 1$$

for any $\lambda \in (0,1)$.

The effective dimensions in both the truncation and superposition senses for product weights with $\eta \in \{0,1,2\}$ are given in Table 1. These values are identical to those in [37] under a periodicity constraint except that the truncation dimension for $\varepsilon = 10^{-4}$ given there is 101 instead of 100. In this case, equality holds in (20) for $\eta = 2$ and $s = 100$.

Theorem 6.1 gives an upper bound on effective dimension. It is then possible that the bounds in Table 1 are too high. In the case of the superposition dimensions there is not much room to lower them, so the bounds in the theorem could well be giving exact values for some ε.

For weights given by $\eta = 2$ and using $\varepsilon = 0.01$ we find that the truncation dimension is at most 9 and the superposition dimension is at most 2. When $\eta = 0$, the truncation dimension is unbounded. For instance, an integrand of variance one depending only on x_j would be inside the unit ball. Since there is no a priori upper bound on j we get ∞ for the truncation dimension. It is surprising that the superposition dimension is not very large for $\eta = 0$. For $\eta = 0$, all of the weights are $\gamma_u = 1$. Decreasing weights are commonly motivated by the reduced importance that they give to subsets of higher cardinality, but here we see modest truncation dimension even with constant weights.

7. Conclusions. The effectiveness of QMC sampling on nominally high dimensional integrands can be explained in part by those specific integrands having a low effective dimension [3] as measured by ANOVA. It is only partial because ANOVA components are not necessarily smooth enough for QMC to be beneficial. It is however common for the process by which ANOVA components are defined to make the low order components smooth [9].

This paper considers effective dimension of weighted Sobolev spaces without requiring periodicity of the integrands. Some weighted Sobolev spaces describe families of integrands over which QMC has uniformly good performance. A ball in such a space just barely large enough to contain an integrand of unit variance, will contain no integrands with meaningfully large high dimensional or high index variance components. Thus algorithms for integration in these settings ought to focus on certain low dimensional aspects of the input space.

It is surprising that weighted spaces with all $\gamma_u = 1$ (which does not allow tractability) leads to spaces with modest superposition dimension. In this sense, $\gamma_u = 1$ is not a model for a situation where all subsets of variables are equally important.

The proof techniques in this paper use simple bounds rooted in a multidimensional Poincaré inequality. The MDM literature uses different techniques and different quantities. In the case of product weights, both methods have to discount high order subsets. The MDM literature (e.g., [23]) uses a tail quantity $\sum_{j>d}\gamma_j$. The superposition dimension here uses $\prod_{j=1}^{d+1}\gamma_j$ while the truncation dimension uses γ_{d+1}. Theorem 2 of [14] gives a truncation dimension below

$$\min\left\{k \in 1{:}s \;\Big|\; \left(\prod_{j=1}^{s}(1+C_1\gamma_j^2)\Big(1-e^{-C_1^2\sum_{j=k+1}^{s}\gamma_j^2}\Big)\right)^{1/2} \leqslant \varepsilon\right\},$$

where some norms of the parts of the function decomposition obey a growth condition parameterized by C_1.

Up to this point, we have emphasized functions of low effective dimension. It is important to consider what happens if the integrand f at hand is not dominated by its low dimensional components. If f is in the unit ball in one of the weighted Sobolev spaces then a modestly large n can be found which will yield an integration error smaller than ε for this f and all other integrands in that ball. Even if f is not in that ball, $\bar{f} = f/\|f\|_{\boldsymbol{\gamma}}$ is in that ball, and we can be sure of an error below ε for $\bar{f}$. Using Monte Carlo and QMC methods, this means we have an error below $\varepsilon\|f\|_{\boldsymbol{\gamma}}$ for f. If $\|f\|_{\boldsymbol{\gamma}}$ is very large then we need a very small ε to compensate. A function such as $f_d = \prod_{j=1}^{d}(x_j - 1/2)$ makes a good test case. For product weights $\|f_d\| = (d!)^{\eta/2}$, and so for good results we would need ε to be comparable to $(d!)^{-\eta/2}$, which then requires n to be a power of $d!$.

Acknowledgments. I thank the following people for helpful discussions: Sergei Kucherenko, Josef Dick, and Fred Hickernell. Greg Wasilkowski made a good suggestion about formulating the definition of subspace dimension that was incorporated into [37]. Frances Kuo shared an early version of [20] while I was at MCQMC 2012. I thank Jiangming Xiang for catching an error in a previous version of this paper. I also would like to acknowledge SAMSI, whose QMC workshop came as I was completing this document. While there, I had additional informative conversations about tractability, initial error, and normalization in the Poincaré inequailty with Ian Sloan, Fred Hickernell, Greg Wasilkowski and Henryk Woźniakowski. Finally, I thank the reviewers of this paper for many helpful comments pointing out connections to the IBC literature.

REFERENCES

[1] P. Acworth, M. Broadie, and P. Glasserman, *A comparison of some Monte Carlo techniques for option pricing*, in Monte Carlo and Quasi-Monte Carlo Methods '96, H. Niederreiter, P. Hellekalek, G. Larcher, and P. Zinterhof, eds., Springer, New York, 1997, pp. 1–18.

[2] N. S. Bakhvalov, *On approximate calculation of multiple integrals*, Vestnik Moskov. Univ., Ser. Mat. Mekhaniki Astron. Fiz. khimii, 4 (1959), pp. 3–18 (in Russian).

[3] R. E. Caflisch, W. Morokoff, and A. B. Owen, *Valuation of mortgage backed securities using Brownian bridges to reduce effective dimension*, J. Comput. Finance, 1 (1997), pp. 27–46.

[4] R. M. Corless, G. H. Gonnet, D. E. G. Hare, D. J. Jeffrey, and D. E. Knuth, *On the Lambert W function*, Adv. Comput. Math., 5 (1996), pp. 329–359.

[5] J. Dick, F. Y. Kuo, and I. H. Sloan, *High-dimensional integration: The quasi-Monte Carlo way*, Acta Numer., 22 (2013), pp. 133–288.

[6] J. Dick and F. Pillichshammer, *Digital Sequences, Discrepancy and Quasi-Monte Carlo Integration*, Cambridge University Press, Cambridge, 2010.

[7] J. Dick, I. H. Sloan, X. Wang, and H. Woźniakowski, *Good lattice rules in weighted Korobov spaces with general weights*, Numer. Math., 103 (2006), pp. 63–97.

[8] A. D. Gilbert and G. W. Wasilkowski, *Small superposition dimension and active set construction for multivariate integration under modest error demand*, J. Complexity, 42 (2017), pp. 94–109.

[9] M. Griebel, F. Y. Kuo, and I. H. Sloan, *The smoothing effect of the ANOVA decomposition*, J. Complexity, 26 (2010), pp. 523–551.

[10] M. Hefter and K. Ritter, *On embeddings of weighted tensor product Hilbert spaces*, J. Complexity, 31 (2015), pp. 405–423.

[11] F. J. Hickernell, *Quadrature error bounds with applications to lattice rules*, SIAM J. Numer. Anal., 33 (1996), pp. 1995–2016.

[12] F. J. Hickernell, *Lattice rules: How well do they measure up?*, in Random and Quasi-Random Point Sets, Lect. Notes Stat. 138, Springer, New York, 1998.

[13] F. J. Hickernell and H. Woźniakowski, *Integration and approximation in arbitrary dimensions*, Adv. Comput. Math., 12 (2000), pp. 25–58.

[14] A. Hinrichs, P. Kritzer, F. Pillichshammer, and G. W. Wasilkowski, *Truncation dimension for linear problems on multivariate function spaces*, Numer. Algorithms, 80 (2018), pp. 661–685.

[15] W. Hoeffding, *A class of statistics with asymptotically normal distribution*, Ann. Math. Statist., 19 (1948), pp. 293–325.

[16] J. Imai and K. S. Tan, *Enhanced quasi-Monte Carlo methods with dimension reduction*, in Proceedings of the 2002 Winter Simulation Conference, E. Yücesan, C.-H. Chen, J. L. Snowdon, and J. M. Charnes, eds., IEEE, New York, 2002, pp. 1502–1510.

[17] P. Kritzer, F. Pillichshammer, and G. W. Wasilkowski, *Very low truncation dimension for high dimensional integration under modest error demand*, J. Complexity, 35 (2016), pp. 63–85.

[18] P. Kritzer, F. Pillichshammer, and G. W. Wasilkowski, *A note on equivalence of anchored and ANOVA spaces; lower bounds*, J. Complexity, 38 (2017), pp. 31–38.

[19] P. Kritzer, F. Pillichshammer, and G. W. Wasilkowski, *Truncation dimension for function approximation*, in Contemporary Computational Mathematics - A Celebration of the 80th Birthday of Ian Sloan, Springer, Cham, Switzerland, 2018, pp. 771–792.

[20] F. Kuo, C. Schwab, and I. H. Sloan, *Quasi-Monte Carlo methods for high-dimensional integration — the standard (weighted Hilbert space) setting and beyond*, ANZIAM J., 53 (2012), pp. 1–37.
[21] F. Kuo, I. Sloan, G. Wasilkowski, and H. Woźniakowski, *On decompositions of multivariate functions*, Math. Comp., 79 (2010), pp. 953–966.
[22] F. Y. Kuo, C. Schwab, and I. H. Sloan, *Quasi-Monte Carlo finite element methods for a class of elliptic partial differential equations with random coefficients*, SIAM J. Numer. Anal., 50 (2012), pp. 3351–3374.
[23] F. Y. Kuo, I. H. Sloan, G. W. Wasilkowski, and H. Woźniakowski, *Liberating the dimension*, J. Complexity, 26 (2010), pp. 422–454.
[24] M. Lamboni, B. Iooss, A.-L. Popelin, and F. Gamboa, *Derivative-based Global Sensitivity Measures: General Links with Sobol' Indices and Numerical Tests*, Technical report, Université Paris Descartes, Paris, 2012.
[25] P. L'Ecuyer and C. Lemieux, *Variance reduction via lattice rules*, Manag. Sci., 46 (2000), pp. 1214–1235.
[26] R. Liu and A. B. Owen, *Estimating mean dimensionality of analysis of variance decompositions*, J. Amer. Statist. Assoc., 101 (2006), pp. 712–721.
[27] M. D. McKay, R. J. Beckman, and W. J. Conover, *A comparison of three methods for selecting values of input variables in the analysis of output from a computer code*, Technometrics, 21 (1979), pp. 239–245.
[28] N. G. Meyers and J. Serrin, $H{=}W$, Proc. Natl. Acad. Sci. USA, 51 (1964), pp. 1055–1056.
[29] D. S. Mitrinovic, J. Pecaric, and A. M. Fink, *Inequalities Involving Functions and Their Integrals and Derivatives*, Math. Appl. East Eur. Ser. 53, Kluwer, Dordrecht, The Netherlands, 1991.
[30] H. Niederreiter, *Random Number Generation and Quasi-Monte Carlo Methods*, CBMS-NSF Regional Conf. Ser. in Appl. Math. 63, SIAM, Philadelphia, 1992.
[31] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems: Standard Information for Functionals*, European Mathematical Society, Zurich, 2010.
[32] D. Nuyens and R. Cools, *Fast algorithms for component-by-component construction of rank-1 lattice rules in shift-invariant reproducing kernel Hilbert spaces*, Math. Comp., 75 (2006), pp. 903–920.
[33] D. Nuyens and R. Cools, *Fast component-by-component construction of rank-*1 *lattice rules with a non-prime number of points*, J. Complexity, 22 (2006), pp. 4–28.
[34] A. B. Owen, *Latin supercube sampling for very high dimensional simulations*, ACM Trans. Model. Comput. Simul., 8 (1998), pp. 71–102.
[35] A. B. Owen, *The dimension distribution and quadrature test functions*, Statist. Sinica, 13 (2003), pp. 1–17.
[36] A. B. Owen, *Variance components and generalized Sobol' indices*, SIAM/ASA J. Uncertain. Quantif., 1 (2013), pp. 19–41.
[37] A. B. Owen, *Effective Dimension for Weighted Function Spaces*, Technical report, Stanford University, Palo Alto, CA, 2014.
[38] S. H. Paskov, *Computing High Dimensional Integrals with Applications to Finance*, Technical report CUCS-023-94, Columbia University, New York, 1994.
[39] L. Plaskota and G. W. Wasilkowski, *Tractability of infinite-dimensional integration in the worst case and randomized settings*, J. Complexity, 27 (2011), pp. 505–518.
[40] R. D. Richtmyer, *The Evaluation of Definite Integrals, and a Quasi-Monte Carlo Method Based on the Properties of Algebraic Numbers*, Technical report LA–1342, Los Alamos, NM, 1951.
[41] O. Roustant, J. Fruth, B. Iooss, and S. Kuhnt, *Crossed-derivative based sensitivity measures for interaction screening*, Math. Comput. Simul. 105 (2014), pp. 105–118.
[42] I. H. Sloan, *Finite-order integration weights can be dangerous*, Comput. Methods Appl. Math., 7 (2007), pp. 239–254.
[43] I. H. Sloan, F. Y. Kuo, and S. Joe, *Constructing randomly shifted lattice rules in weighted Sobolev spaces*, SIAM J. Numer. Anal., 40 (2002), pp. 1650–1665.
[44] I. H. Sloan, X. Wang, and H. Woźniakowski, *Finite-order weights imply tractability of multivariate integration*, J. Complexity, 20 (2004), pp. 46–74.
[45] I. H. Sloan and H. Woźniakowski, *When are quasi-Monte Carlo algorithms efficient for high dimensional integration?*, J. Complexity, 14 (1998), pp. 1–33.
[46] I. M. Sobol', *The use of the ω^2-distribution for error estimation in the calculation of integrals by the Monte Carlo method*, USSR Comput. Math. Phys., 2 (1963), pp. 808–816.
[47] I. M. Sobol', *Multidimensional Quadrature Formulas and Haar Functions*, Nauka, Moscow, 1969 (in Russian).

[48] I. M. Sobol', *Sensitivity estimates for nonlinear mathematical models*, Math. Model. Comput. Exp., 1 (1993), pp. 407–414.

[49] I. M. Sobol' and S. Kucherenko, *Derivative based global sensitivity measures and their link with global sensitivity indices*, Math. Comput. Simul., 10 (2009), pp. 3009–3017.

[50] I. M. Sobol' and S. Kucherenko, *A new derivative based importance criterion for groups of variables and its link with global sensitivity indices*, Comput. Phys. Commun., 181 (2010), pp. 1212–1217.

[51] I. M. Sobol' and B. V. Shukhman, *On average dimensions of particle transport estimators*, Monte Carlo Methods Appl., 24 (2018), pp. 147–151.

[52] M. Stein, *Large sample properties of simulations using Latin hypercube sampling*, Technometrics, 29 (1987), pp. 143–51.

[53] W. Stekloff, *Sur le problème de refroidissement d'une barre héterogène*, Ann. Fac. Sci. Toulouse, 3 (1901), pp. 281–313.

[54] X. Wang and K.-T. Fang, *The effective dimension and quasi-Monte Carlo integration*, J. Complexity, 19 (2003), pp. 101–124.

[55] G. W. Wasilkowski, *Liberating the dimension for function approximation and integration*, in Monte Carlo and Quasi-Monte Carlo Methods 2010, L. Plaskota and H. Woźniakowski, eds., Springer, Berlin, 2012, pp. 211–231.

[56] G. W. Wasilkowski, *On tractability of linear tensor product problems for* 8*-variate classes of functions*, J. Complexity, 29 (2013), pp. 351–369.

[57] G. W. Wasilkowski, *Tractability of approximation of* 8*-variate functions with bounded mixed partial derivatives*, J. Complexity, 30 (2014), pp. 325–346.

[58] G. W. Wasilkowski and H. Woźniakowski, *Liberating the dimension for function approximation*, J. Complexity, 27 (2011), pp. 86–110.