SIAM J. NUMER. ANAL.
Vol. 54, No. 3, pp. 1946–1966

TRANSFORMATIONS AND HARDY–KRAUSE VARIATION*

KINJAL BASU† AND ART B. OWEN‡

Abstract. Using a multivariable Faa di Bruno formula we give conditions on transformations $\tau : [0,1]^m \to \mathcal{X}$ where $\mathcal{X}$ is a closed and bounded subset of $\mathbb{R}^d$ such that $f \circ \tau$ is of bounded variation in the sense of Hardy and Krause for all $f \in C^m(\mathcal{X})$. We give similar conditions for $f \circ \tau$ to be smooth enough for scrambled net sampling to attain $O(n^{-3/2+\epsilon})$ accuracy. Some popular symmetric transformations to the simplex and sphere are shown to satisfy neither condition. Some other transformations due to Fang and Wang (Monogr. Statist. Appl. Probab. 51, CRC Press, Boca Raton, FL, 1993) satisfy the first but not the second condition. We provide transformations for the simplex that makes $f \circ \tau$ smooth enough to fully benefit from scrambled net sampling for all f in a class of generalized polynomials. We also find sufficient conditions for conditional inversion in $\mathbb{R}^2$ and for importance sampling to be of bounded variation in the sense of Hardy and Krause.

Key words. importance sampling, quasi–Monte Carlo, randomized quasi–Monte Carlo

AMS subject classifications. 65D30, 41A55

DOI. 10.1137/15M1052184

1. Introduction.

Quasi–Monte Carlo (QMC) sampling is usually applied to integration problems over the domain $[0,1]^d$. Other domains, such as triangles, disks, simplices, spheres, or balls, are also of importance in applications. Monte Carlo (MC) sampling over such a domain $\mathcal{X} \subset \mathbb{R}^d$ is commonly done by finding a uniformity preserving transformation $\tau : [0,1]^m \to \mathcal{X}$. Such transformations yield $\boldsymbol{x} = \tau(\boldsymbol{u}) \sim \mathbf{U}(\mathcal{X})$ when $\boldsymbol{u} \sim \mathbf{U}[0,1]^m$ so that

$$\frac{1}{\mathbf{vol}(\mathcal{X})} \int_{\mathcal{X}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = \int_{[0,1]^m} f(\tau(\boldsymbol{u}))\,\mathrm{d}\boldsymbol{u}. \tag{1}$$

Then we estimate $\mu = \int_{\mathcal{X}} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x}$ by $(\mathbf{vol}(\mathcal{X})/n)\sum_{i=1}^n f(\tau(\boldsymbol{u}_i))$ for $\boldsymbol{u}_i \overset{\text{iid}}{\sim} \mathbf{U}[0,1]^d$. We will often take $\mathbf{vol}(\mathcal{X}) = 1$ for simplicity.

A very standard approach to QMC sampling of such domains is to employ the same transformation τ as in MC but to replace independent random $\boldsymbol{u}_i$ by QMC or randomized QMC (RQMC) points. The uniformity-preserving transformation τ satisfies (1) when $m = d$ and τ has a Jacobian determinant everywhere equal to $1/\mathbf{vol}(\mathcal{X}) = 1$. It also holds when that Jacobian determinant is piecewise constant and equal to ± 1 at all $\boldsymbol{u}$. Equation (1) does not require $m = d$. For instance, in section 5.1, we study a logarithmic mapping from $[0,1]^3$ to a two-dimensional equilateral triangle which satisfies (1).

When the function $f \circ \tau$ is of bounded variation in the sense of Hardy and Krause (BVHK), then the Koksma–Hlawka inequality applies and QMC can attain the convergence rate $O(n^{-1+\epsilon})$. Under additional smoothness conditions on $f \circ \tau$, certain RQMC methods (scrambled nets) have a root mean squared error (RMSE) of $O(n^{-3/2+\epsilon})$.

*Received by the editors December 10, 2015; accepted for publication (in revised form) April 18, 2016; published electronically June 23, 2016. This research was supported by the NSF under grant DMS-1407397.
http://www.siam.org/journals/sinum/54-3/M105218.html

†Corresponding author. Statistics, Stanford University, Stanford, CA 94305 (kinjal@stanford.edu, http://web.stanford.edu/~kinjal).
‡Statistics, Stanford University, Stanford, CA 94305 (owen@stanford.edu, http://statweb.stanford.edu/~owen).

The paper proceeds as follows. Section 2 gives a sufficient condition for a function f on $[0, 1]^d$ to be in BVHK and a stronger condition for that function to be integrable with RMSE $O(n^{-3/2+\epsilon})$ by scrambled nets. Those conditions are expressed in terms of certain partial derivatives of f. Section 3 considers how to apply the conditions from section 2 to compositions $f \circ \tau$. There are good sufficient conditions for compositions of single variable functions to be in BVHK, but the multivariate setting is more complicated, as shown by a counterexample there. Then we specialize a multivariable Faa di Bruno formula from Constantine and Savits (1996) to the mixed partial derivatives required for QMC. Section 4 gives sufficient conditions for $f \circ \tau$ to be in BVHK and also for it to be smooth enough for scrambled nets to improve on the QMC rate. There is also a discussion of necessary conditions. We stipulate there that at least the components of τ should themselves be in BVHK. Section 5 considers two widely used transformations τ that are symmetric operations on d input variables to yield uniform points in the $d-1$ dimensional simplex and sphere, respectively. Unfortunately some components of τ fail to be in BVHK for these transformations. Section 6 shows that some classic mappings to the simplex, sphere, and ball from Fang and Wang (1993) are in BVHK, although they are not smooth enough to benefit from RQMC. Section 7 considers nonuniform transformations including importance sampling, sequential inversion of the Rosenblatt transformation, and a nonuniform transformation on the unit simplex that yields the customary RQMC convergence rate for a class of functions including all polynomials on the simplex. We finish this section by mentioning some additional literature.

Several works have previously considered applying transformations to QMC points. Sampling of the simplex and especially the triangle was considered by Pillards and Cools (2005). They looked at six different transformations. Most of the transformations that they study involve nonsmooth operations on $[0, 1]^d$ such as sorting coordinates, deleting values, or making cuts and folds to the space. Such transformations are not amenable to derivative-based methods that we consider here. One exception is their root transformation, which is the same as the Fang and Wang (1993) transformation that we consider in section 6.1.

QMC sampling on the sphere was studied by Brauchart and Dick (2012), Brauchart et al. (2014), and many others. There one commonly works with spherical cap discrepancy. Brauchart and Dick (2012) present a transformation that yields spherical cap discrepancy $O((\log(n)/n)^{1/2})$ that empirically seems to attain the rate $n^{-3/4}$.

Another approach to QMC sampling on noncubical domains involves recursive partitioning of such domains in a way that mimicks the van der Corput construction for $[0, 1]$. The van der Corput construction in Basu and Owen (2015a) attains the rate $O(n^{-1/2})$ for the discrepancy of Brandolini et al. (2013) in the triangle, while a Kronecker sequence construction attains the rate $O(\log(n)/n)$. Basu and Owen (2015b) generalize triangular van der Corput construction to digital net sampling of Cartesian products of triangles and other compact sets.

While finishing up this paper we noticed that Cambou, Hofert, and Lemieux (2015) have also applied the Faa di Bruno formula in a QMC application, though they apply it to a different set of problems. They use it to give sufficient conditions for some integrands with respect to copulas to be in BVHK. They extend Hlawka and Mück (1972) for inverse CDF sampling to some copulas with mixed partial derivatives that are singular on the boundaries of the unit cube. They closely study the Marshall–Olkin algorithm which generates points from a d dimensional Archimedean copula from a point in $[0, 1]^{d+1}$ and give conditions for quadrature errors to be bounded by a multiple of the $d + 1$ dimensional discrepancy and weaker conditions for a bound $\log(n)$ times as large as that.

2. Smoothness conditions. QMC sampling attains an error rate of $O(n^{-1}(\log n)^{d-1})$ if the function $f \in \mathrm{BVHK}$. Here we give a simply checked sufficient condition for $f \in \mathrm{BVHK}$. We use $\mathrm{V}_{\mathrm{HK}}(f)$ for the total variation of f in the sense of Hardy and Krause and $\mathrm{V}_{\mathrm{IT}}(f)$ for the total variation of f in the sense of Vitali.

Let $1{:}d = \{1, 2, \ldots, d\}$. For a set $u \subseteq 1{:}d$, let $|u|$ denote the cardinality of u and $-u = 1{:}d \setminus u$ its complement. Let $\partial^u f$ denote the partial derivative of f taken once with respect to each variable $j \in u$. By convention $\partial^{\varnothing} f = f$. For $\boldsymbol{x} \in [0,1]^d$ and $u \subseteq 1{:}d$ let $\boldsymbol{x}_u{:}\mathbf{1}_{-u}$ be the point $\boldsymbol{y} \in [0,1]^d$ with $y_j = x_j$ for $j \in u$ and $y_j = 1$ for $j \notin u$.

If the mixed partial derivative $\partial^{1:d} f$ exists, then

$$\mathrm{V}_{\mathrm{IT}}(f) \leqslant \int_{[0,1]^d} |\partial^{1:d} f(\boldsymbol{x})| \,\mathrm{d}\boldsymbol{x} \quad \text{and} \tag{2}$$

$$\mathrm{V}_{\mathrm{HK}}(f) \leqslant \sum_{u \neq \varnothing} \int_{[0,1]^{|u|}} |\partial^u f(\boldsymbol{x}_u{:}\mathbf{1}_{-u})| \,\mathrm{d}\boldsymbol{x}_u. \tag{3}$$

These and related results are presented in Owen (2005). Fréchet (1910) shows that the Vitali bound (2) becomes an equality if $\partial^{1:d} f$ is continuous on $[0,1]^d$. The Hardy–Krause variation is a sum of Vitali variations for which (3) arises by applying (2) term by term.

For scrambled nets, a kind of RQMC, to attain an RMSE of order $O(n^{-3/2}(\log n)^{(d-1)/2})$ the function f must be smooth in the following sense:

$$\|\partial^u f\|_2^2 \equiv \int (\partial^u f(\boldsymbol{x}))^2 \,\mathrm{d}\boldsymbol{x} < \infty \quad \forall u \subseteq 1{:}d. \tag{4}$$

For a description of digital nets including scramblings of them, see Dick and Pillichshammer (2010). Two scramblings with RMSE of $O(n^{-3/2+\epsilon})$ are the nested uniform scramble in Owen (1995) and the nested linear scramble of Matoušek (1998). Geometric nets and scrambled geometric nets have been introduced in Basu and Owen (2015b) for sampling uniformly on $\mathcal{X}^s$ where $\mathcal{X}$ is a closed and bounded subset of $\mathbb{R}^d$. Scrambled geometric nets attain an RMSE of $O(n^{-1/2-1/d}(\log n)^{(s-1)/2})$ for certain smooth functions defined on $\mathcal{X}^s$. The construction of scrambled geometric nets is based on the recursive partitions like those used by Basu and Owen (2015a) to sample the triangle.

We will study transformations by considering which combinations of f and τ give $\mathrm{V}_{\mathrm{HK}}(f \circ \tau) < \infty$. For such combinations, plain QMC will be asymptotically better than geometric nets when $s = 1$ and $d \geqslant 3$. Similarly, if $\partial^u(f \circ \tau) \in L^2$ for all $u \subseteq 1{:}d$, then scrambled nets are asymptotically better than geometric nets for $d \geqslant 2$.

Higher order digital nets (Dick, 2009) achieve even better rates of convergence than plain (R)QMC does, but they require even stronger smoothness conditions. Their randomized versions (Dick, 2011) further increase accuracy (in root mean square) under yet stronger smoothness conditions.

3. Function composition. We would like a condition under which the composition $f \circ \tau : [0,1]^m \to \mathbb{R}^d \to \mathbb{R}$ is in BVHK. For the case $d = m = 1$, BVHK for $f \circ \tau$ reduces to ordinary BV. Josephy (1981) gives a very complete characterization of when compositions of one-dimensional functions are in BV.

Let f and τ be functions of bounded variation from $[0,1]$ to $[0,1]$. Theorem 4 of Josephy (1981) shows that $f \circ \tau \in \mathrm{BV}$ holds for all $\tau \in \mathrm{BV}$ if and only if f is Lipschitz. The statement on τ is a bit more complicated. His Theorem 3 shows that $f \circ \tau \in \mathrm{BV}$

for all $f \in \mathrm{BV}$ if and only if τ belongs to a special subset of BV, in which preimages of intervals are unions of a finite set of intervals.

3.1. A counterexample. No such comprehensive characterization is available for bounded variation in the sense of Hardy and Krause in higher dimensions. Here we present functions $f : \mathbb{R}^2 \to \mathbb{R}$ and $\tau : \mathbb{R}^2 \to \mathbb{R}^2$ such that f is Lipschitz and $\tau \in \mathrm{BVHK}$ but $f \circ \tau \notin \mathrm{BVHK}$. We take τ to be the identity map on $[0,1]^2$ so that both τ_1 and τ_2 are in BVHK. Then we construct a Lipschitz function $f : [0,1]^2 \to \mathbb{R}$ with $f \circ \tau = f \notin \mathrm{BVHK}$.

We define the function f in a recursive way using a Sierpinsky gasket type splitting of the unit square. Let A be the square $(x_1, x_1 + \ell) \times (x_2, x_2 + \ell) \subset [0,1]^2$ for some $\ell > 0$. Then for $\boldsymbol{x}' \in [0,1]^2$, define the pyramid function

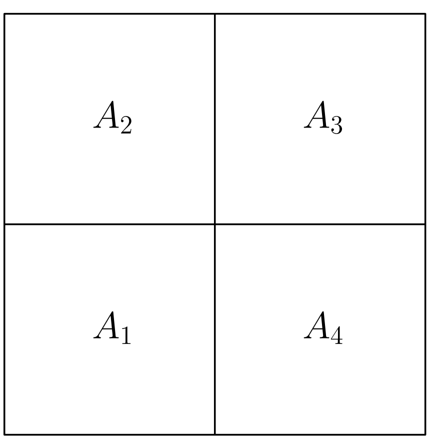
$$f_A(\boldsymbol{x}') = \max\Bigl(0, \frac{\ell}{2} - \max_{j=1,2} |x'_j - (x_j + \ell/2)|\Bigr).$$

This function is 0 for $\boldsymbol{x}' \notin A$ and inside A it defines the upper surface of a square based pyramid of height $\ell/2$ centered over the midpoint of A. For an illustration, see the lower right-hand corner of the second panel in Figure 1. For any A the function f_A is Lipschitz continuous with Lipschitz constant 1.

We construct f as follows. First we split $[0,1]^2$ into four congruent sub-squares as shown in the left panel of Figure 1. Then we select one of those sub-squares, say, A_4 and initially set $f = f_{A_4}$. Next, we partition each of the remaining three sub-squares $A_1, \ldots, A_3$ into four congruent sub-sub-squares A_{ij} for $i = 1, \ldots, 3$ and $j = 1, \ldots, 4$. Then we add $f_{A_{1,4}} + f_{A_{2,4}} + f_{A_{3,4}}$ to f. This construction is carried out recursively summing 3^k pyramidal functions at level $k = 0, 1, 2, \ldots$ over squares of side 2^{-k-1}, as depicted in the right panel of Figure 1.

Lemma 1. *The function f described above has Lipschitz constant one and has infinite Vitali variation and hence infinite variation in the sense of Hardy and Krause.*

Proof. Let $\boldsymbol{x}, \boldsymbol{y} \in [0,1]^2$. Consider the function $g(t) = f(\boldsymbol{x} + t(\boldsymbol{y} - \boldsymbol{x}))$ on $0 \leqslant t \leqslant 1$. This function is continuous and piecewise linear with absolute slope at most 1. Thus $|f(\boldsymbol{x}) - f(\boldsymbol{y})| \leqslant \|\boldsymbol{x} - \boldsymbol{y}\|$ and so f is Lipschitz with constant 1.



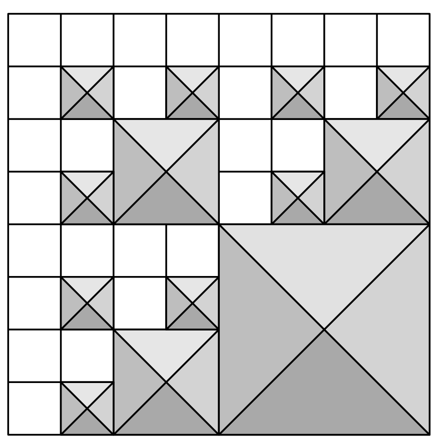


Fig. 1. *The plot on the left shows the square partition $\mathcal{P}$ which is repeated in a recursive manner. The right figure shows the function as a two-dimensional projection for $k = 3$. Each such pyramidal structure has a height of half the length of its base square.*

Now we turn to variation using definitions and results from Owen (2005). The Vitali variation of f_A equals the Vitali variation of f_A over the square A. By considering a 3×3 grid covering the edges and center of A we find that $\mathrm{V_{IT}}(f_A) \geqslant 2\ell$. In fact, $\mathrm{V_{IT}}(f_A) = 2\ell$ but we only need the lower bound.

The Vitali variation of f is the sum of its Vitali variations over a square subpartition. As a result, $\mathrm{V_{IT}}(f)$ is the sum of $\mathrm{V_{IT}}(f_A)$ for all the sets A in our recursive construction. For $k = 0, 1, 2, \ldots$ there are 3^k terms f_A with $\ell = 2^{-k-1}$. Then $2\ell = 2^{-k}$ and the Vitali variation of f is at least $\sum_{k=0}^{\infty} 3^k 2^{-k} = \infty$. □

This counterexample applies to any $d \geqslant 2$ simply by constructing a function that equals the above constructed function f applied to two of its input variables. Such a function has infinite Hardy–Krause variation arising from the Vitali variation in those two variables. As a result, even if f is in Lipschitz and τ is BVHK along with every component, we might still have $f \circ \tau \notin$ BVHK.

3.2. Faa di Bruno formulas. We will study variation via a mixed partial derivative of the composition of the integrand on $\mathcal{X}$ with a transformation from the unit cube to $\mathcal{X}$. We need partial derivatives of order up to the dimension of the unit cube. High order derivatives of a composition become awkward even in the case with $d = m = 1$, which was solved by Faa di Bruno (1855). We will use a multivariable Faa di Bruno formula from Constantine and Savits (1996).

To remain consistent with the notation in Constantine and Savits (1996) we will consider functions $h = f(g(\cdot))$ here. After obtaining the formulas we need, we will revert to $f(\tau(\cdot))$, which is more suitable for the MC and QMC context.

To illustrate Faa di Bruno, suppose first that $d = m = 1$. Then let g be defined on an open set containing x_0 and have derivatives up to order n at x_0. Let f be defined on an open set containing $y_0 = g(x_0)$ and have derivatives up to order n at y_0. For $0 \leqslant k \leqslant n$ define derivatives $f_k = \mathrm{d}^k f(y_0)/\mathrm{d}y^k$, $g_k = \mathrm{d}^k g(x_0)/\mathrm{d}x^k$, and $h_k = \mathrm{d}^k h(x_0)/\mathrm{d}x^k$. From the chain rule we can easily find that

$$h_4 = f_4 g_1^4 + 6 f_3 g_1^2 g_2 + 3 f_2 g_2^2 + 4 f_2 g_1 g_3 + f_1 g_4. \tag{5}$$

The derivative f_k appears in h_n in as many terms as there are distinct ways of finding k positive integers that sum to n. That number of terms is known as the Stirling number of the second kind (Graham, Knuth, and Patashnik, 1989). These Stirling numbers sum to the nth Bell number which grows rapidly with n. We omit the $m = d = 1$ Faa di Bruno formula for arbitrary n and present instead the generalization due to Constantine and Savits (1996).

In the multivariate setting, $h(\boldsymbol{x}) = f(\boldsymbol{g}(\boldsymbol{x}))$, where $\boldsymbol{x} \in \mathbb{R}^m$ and $\boldsymbol{y} = \boldsymbol{g}(\boldsymbol{x}) \in \mathbb{R}^d$. In our applications $\boldsymbol{x} \in [0,1]^m$. We write $\boldsymbol{g}(\boldsymbol{x}) = (g^{(1)}(\boldsymbol{x}), \ldots, g^{(d)}(\boldsymbol{x}))$. The multivariate Faa di Bruno formula gives an arbitrary mixed partial derivative of h with respect to components of $\boldsymbol{x}$ in terms of partial derivatives of f and $g^{(i)}$. The formula requires that the needed derivatives exist.

The formula uses some multi-index notation. We use $\mathbb{N}_0$ for the set of nonnegative integers. Let $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_m) \in \mathbb{N}_0^m$. Then $h_{\boldsymbol{\nu}}$ is the derivative of h taken ν_i times with respect to x_i. Similarly, f and $g^{(i)}$ subscripted by tuples of d and m nonnegative integers, respectively, are the corresponding partial derivatives. When the subscript is all zeros, the result is the function itself, undifferentiated.

For a multi-index $\boldsymbol{\nu} \in \mathbb{N}_0^m$ we write $|\boldsymbol{\nu}| = \sum_{i=1}^m \nu_i$ and $\boldsymbol{\nu}! = \prod_{i=1}^m (\nu_i!)$. For $\boldsymbol{z} \in \mathbb{R}^m$ and $\boldsymbol{\nu} \in \mathbb{N}_0^m$ we write $\boldsymbol{z}^{\boldsymbol{\nu}}$ for $\prod_{i=1}^m z_i^{\nu_i}$. We use an ordering $\prec$ on $\mathbb{N}_0^m$, where $\boldsymbol{\mu} \prec \boldsymbol{\nu}$ means that either $|\boldsymbol{\mu}| < |\boldsymbol{\nu}|$ or $|\boldsymbol{\mu}| = |\boldsymbol{\nu}|$ holds along with $\mu_i < \nu_i$ at the smallest i where $\mu_i \neq \nu_i$. Multi-indices in $\mathbb{N}_0^d$ are treated the same way. The quantity $\boldsymbol{g}_{\boldsymbol{\ell}}$ is the vector $(g_{\boldsymbol{\ell}}^{(1)}, \dots, g_{\boldsymbol{\ell}}^{(d)})$.

THEOREM 2. *Let $g_{\boldsymbol{\mu}}^{(i)}$ be continuous in a neighborhood of $\boldsymbol{x}_0 \in \mathbb{R}^m$ for all $|\boldsymbol{\mu}| \leqslant |\boldsymbol{\nu}|$ and all $i = 1, \dots, d$, where $\boldsymbol{\mu}, \boldsymbol{\nu} \in \mathbb{N}_0^m - \{\mathbf{0}\}$. Similarly, let $f_{\boldsymbol{\lambda}}$ be continuous in a neighborhood of $\boldsymbol{y}_0 = \boldsymbol{g}(\boldsymbol{x}_0) \in \mathbb{R}^d$ for all $|\boldsymbol{\lambda}| \leqslant |\boldsymbol{\nu}|$. Let $h = f \circ g$. Then in a neighborhood of $\boldsymbol{x}_0$,*

$$h_{\boldsymbol{\nu}} = \boldsymbol{\nu}! \sum_{1 \leqslant |\boldsymbol{\lambda}| \leqslant |\boldsymbol{\nu}|} f_{\boldsymbol{\lambda}} \sum_{s=1}^{|\boldsymbol{\nu}|} \sum_{\mathrm{KL}(s,\boldsymbol{\nu},\boldsymbol{\lambda})} \prod_{r=1}^{s} \frac{[\boldsymbol{g}_{\boldsymbol{\ell}_r}]^{\boldsymbol{k}_r}}{(\boldsymbol{k}_r!)[\boldsymbol{\ell}_r!]^{|\boldsymbol{k}_r|}}, \tag{6}$$

where

$$\begin{aligned} \mathrm{KL}(s,\boldsymbol{\nu},\boldsymbol{\lambda}) = \Big\{ &(\boldsymbol{k}_1, \dots, \boldsymbol{k}_s, \boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_s) \in \left(\mathbb{N}_0^d - \{\mathbf{0}\}\right)^s \times \left(\mathbb{N}_0^m - \{\mathbf{0}\}\right)^s \mid \\ &\boldsymbol{\ell}_1 \prec \boldsymbol{\ell}_2 \prec \cdots \prec \boldsymbol{\ell}_s,\ \sum_{r=1}^{s} \boldsymbol{k}_r = \boldsymbol{\lambda},\ \text{and}\ \sum_{r=1}^{s} |\boldsymbol{k}_r| \boldsymbol{\ell}_r = \boldsymbol{\nu} \Big\}. \end{aligned} \tag{7}$$

Proof. See Constantine and Savits (1996, Theorem 1). ☐

For our study of (R)QMC, we only need $h_{\boldsymbol{\nu}}$ with $\boldsymbol{\nu} \in \{0,1\}^m - \{\mathbf{0}\}$. Equations (6) and (7) simplify considerably for this case.

LEMMA 3. *For $m \geqslant 1$ let $\boldsymbol{\nu} \in \{0,1\}^m - \{\mathbf{0}\}$. If $1 \leqslant |\boldsymbol{\lambda}| \leqslant |\boldsymbol{\nu}|$ and $1 \leqslant s \leqslant |\boldsymbol{\nu}|$ and $(\boldsymbol{k}_1, \dots, \boldsymbol{k}_s, \boldsymbol{\ell}_1, \dots, \boldsymbol{\ell}_s) \in \mathrm{KL} = \mathrm{KL}(s, \boldsymbol{\nu}, \boldsymbol{\lambda})$, then for $r = 1, \dots, s$, $\boldsymbol{k}_r \in \{0,1\}^d - \{\mathbf{0}\}$, $\boldsymbol{\ell}_r \in \{0,1\}^m - \{\mathbf{0}\}$, and $|\boldsymbol{k}_r| = 1$. Also $\boldsymbol{\nu}! = \boldsymbol{k}_r! = \boldsymbol{\ell}_r! = 1$.*

Proof. Definition (7) of KL includes the condition $\sum_{r=1}^{s} |\boldsymbol{k}_r| \boldsymbol{\ell}_r = \boldsymbol{\nu}$. Because $\boldsymbol{\nu}$ is a binary vector and $|\boldsymbol{k}_r| \geqslant 1$, no component of $\boldsymbol{\ell}_r$ can be larger than 1. Therefore $\boldsymbol{\ell}_r \in \{0,1\}^m - \{\mathbf{0}\}$. Similarly, $\boldsymbol{\ell}_r$ has at least one nonzero component and so $|\boldsymbol{k}_r| \leqslant 1$. Because $\boldsymbol{k}_r \neq \mathbf{0}$ we now have $|\boldsymbol{k}_r| = 1$. Finally, the factorial of any binary vector is 1. ☐

It follows from Lemma 3 that for $\boldsymbol{\nu} \in \{0,1\}^m - \{\mathbf{0}\}$,

$$h_{\boldsymbol{\nu}} = \sum_{1 \leqslant |\boldsymbol{\lambda}| \leqslant |\boldsymbol{\nu}|} f_{\boldsymbol{\lambda}} \sum_{s=1}^{|\boldsymbol{\nu}|} \sum_{\mathrm{KL}(s,\boldsymbol{\nu},\boldsymbol{\lambda})} \prod_{r=1}^{s} [\boldsymbol{g}_{\boldsymbol{\ell}_r}]^{\boldsymbol{k}_r}. \tag{8}$$

Next, we use Lemma 3 to simplify the derivatives of g. Because $\boldsymbol{\nu}$ is a nonzero binary vector, we can identify it with a nonempty subset v of $1{:}m$. Specifically, $j \in v$ if and only if $\nu_j = 1$. Similarly we may identify the binary vector $\boldsymbol{\ell}_r \in \{0,1\}^m - \{\mathbf{0}\}$ with the set $\ell_r \subseteq 1{:}m$. The nonzero binary vector $\boldsymbol{k}_r \in \{0,1\}^d$ corresponds to a singleton set. We can therefore identify it with an integer in $1{:}d$. We identify $\boldsymbol{k}_r$ with the integer k_r such that $\boldsymbol{k}_{ri} = 1$ if and only if $i = k_r$.

With this identification,

$$[\boldsymbol{g}_{\boldsymbol{\ell}_r}]^{\boldsymbol{k}_r} = \prod_{i=1}^{d} \left(\frac{\partial^{|\boldsymbol{\ell}_r|} g^{(i)}}{\prod_{j=1}^{m} \partial x_j^{\ell_{rj}}} \right)^{k_{ri}} = \prod_{i=1}^{d} \left(\partial^{\ell_r} g^{(i)} \right)^{k_{ri}} = \partial^{\ell_r} g^{(k_r)}. \tag{9}$$

Now switching from $g^{(k)}$ back to τ_k we get a Faa di Bruno formula for mixed partial derivatives taken at most once with respect to every index:

$$\partial^{v}(f \circ \tau) = \sum_{\substack{\boldsymbol{\lambda} \in \mathbb{N}_0^d \\ 1 \leqslant |\boldsymbol{\lambda}| \leqslant |v|}} f_{\boldsymbol{\lambda}} \sum_{s=1}^{|v|} \sum_{(\ell_r, k_r) \in \widetilde{\mathrm{KL}}(s,v,\boldsymbol{\lambda})} \prod_{r=1}^{s} \partial^{\ell_r} \tau_{k_r}, \tag{10}$$

where $\widetilde{\mathrm{KL}}(s,v,\boldsymbol{\lambda})$ equals

$$\begin{aligned}(11)\qquad &\Big\{(\ell_r,k_r),\ r=1,\dots,s,\ \Big|\ \ell_r\subseteq 1{:}m,\ k_r\in 1{:}d,\ \cup_{r=1}^s\ell_r=v,\\ &\quad \ell_r\cap\ell_{r'}=\varnothing \text{ for } r\neq r' \text{ and } |\{j\in 1{:}s\mid k_j=i\}|=\lambda_i\Big\}.\end{aligned}$$

4. Necessary and sufficient conditions. Equation (10) allows us to find sufficient conditions on f and τ so that $f\circ\tau\in\mathrm{BVHK}$ for all $f\in C^m(\mathcal{X})$. Similarly, we find conditions under which $f\circ\tau$ is smooth in the sense of (4) for all $f\in C^m(\mathcal{X})$.

THEOREM 4. *Let $\tau(\boldsymbol{u})=(\tau_1(\boldsymbol{u}),\dots,\tau_d(\boldsymbol{u}))$ be a map from $[0,1]^m$ to the closed and bounded set $\mathcal{X}\subset\mathbb{R}^d$ such that $\partial^{1:m}\tau_j$ is continuous for all $j=1,\dots,d$. Assume that*

$$(12)\qquad \int_{[0,1]^{|v|}}\prod_{r=1}^s\big|\partial^{\ell_r}\tau_{k_r}(\boldsymbol{u}_v{:}\mathbf{1}_{-v})\big|\,\mathrm{d}\boldsymbol{u}_v<\infty$$

holds for all nonempty $v\subseteq 1{:}m$, for all $s\in 1{:}|v|$, for all disjoint ℓ_r with union v, and for all distinct $k_r\in 1{:}d$. Then $f\circ\tau\in\mathrm{BVHK}$ for all $f\in C^m(\mathcal{X})$.

Proof. By specializing the bound on V_{HK} in (3) to $f\circ\tau$ and using (10), it suffices to show that

$$(13)\qquad \int_{[0,1]^{|v|}}\Big|f_{\boldsymbol{\lambda}}(\tau(\boldsymbol{u}_v{:}\mathbf{1}_{-v}))\prod_{r=1}^s\partial^{\ell_r}\tau_{k_r}(\boldsymbol{x}_u{:}\mathbf{1}_{-u})\Big|\,\mathrm{d}\boldsymbol{x}_u<\infty$$

holds, for all $\boldsymbol{\lambda}\in\mathbb{N}_0^d$ with $1\leqslant|\boldsymbol{\lambda}|\leqslant|v|$, and the v, s, ℓ_r, and k_r of the theorem statement, and where λ_i of the k_r are equal to i for $i\in 1{:}d$. The condition on f makes each of the $f_{\boldsymbol{\lambda}}$ in (10) a bounded function. Then (12) suffices to establish (13). ∎

Very often $m=d$, but sometimes, as in section 5, $m>d$. In such cases, Theorem 4 imposes a more stringent smoothness requirement on f.

We will use a generalized Hölder inequality (Bogachev, 2007, p. 141) to get sufficient conditions for Theorem 4. For a positive integer s, suppose that $f_r\in L^{p_r}(\mu)$ for $r=1,\dots,s$, for some nonnegative measure μ, and that $\sum_{r=1}^s 1/p_r=1/p$. Then $(\int|\prod_{r=1}^s f_r|^p\,\mathrm{d}\mu)^{1/p}\leqslant\prod_{r=1}^s(\int|f_r|^{p_r}\,\mathrm{d}\mu)^{1/p_r}$ and so $\prod_{r=1}^s f_r\in L^p(\mu)$.

COROLLARY 5. *Let $\tau(\boldsymbol{u})=(\tau_1(\boldsymbol{u}),\dots,\tau_d(\boldsymbol{u}))$ be a map from $[0,1]^m$ to the closed and bounded set $\mathcal{X}\subset\mathbb{R}^d$ such that $\partial^{1:m}\tau_j$ is continuous for all $j=1,\dots,d$. Assume that $\partial^v\tau_j(\boldsymbol{u}_v{:}\mathbf{1}_{-v})\in L^{p_j}([0,1]^{|v|})$ for all $j=1,\dots,d$ and for all nonempty $v\subseteq 1{:}m$, where $p_j\in[1,\infty]$. If $\sum_{j=1}^d 1/p_j\leqslant 1$, then $f\circ\tau\in\mathrm{BVHK}$ for all $f\in C^m(\mathcal{X})$.*

Proof. The L^{p_j} conditions on the derivatives of τ_j combined with the generalized Hölder condition establish (12). ∎

As a special case, suppose that all τ_j satisfy the same moment condition: $p_j=p_1$, $j=1,\dots,d$. Then Corollary 5 requires $p_j\geqslant d$. That is, the moment conditions on τ in Corollary 5 become more stringent as the dimension d of $\mathcal{X}$ increases. For their QMC analysis on spheres, Brauchart et al. (2014) have a similar finding that greater smoothness is required in higher dimensions to control a worst-case quadrature error.

Next we consider the kind of smoothness that allows scrambled nets to improve upon the QMC rate. In this setting we require mixed partial derivatives in L^2 but we do not have to pay special attention to components of $\boldsymbol{u}$ that equal 1.

THEOREM 6. *Let $\tau(\boldsymbol{u}) = (\tau_1(\boldsymbol{u}), \dots, \tau_d(\boldsymbol{u}))$ be a map from $[0,1]^m$ to the closed and bounded set $\mathcal{X} \subset \mathbb{R}^d$ such that $\partial^{1:m}\tau_j$ exists for all $j = 1, \dots, d$. Assume that*

$$\int_{[0,1]^d} \prod_{r=1}^{s} \left|\partial^{\ell_r} \tau_{k_r}(\boldsymbol{u})\right|^2 \mathrm{d}\boldsymbol{u} < \infty \tag{14}$$

holds for all nonempty $v \subseteq 1{:}m$, for all $s \in 1{:}|v|$, for all disjoint ℓ_r with union v, and for all distinct $k_r \in 1{:}d$. Then $f \circ \tau$ is smooth in the sense of (4) *for all $f \in C^m(\mathcal{X})$.*

Proof. The same argument used in Theorem 4 applies here. ◻

COROLLARY 7. *Let $\tau(\boldsymbol{u}) = (\tau_1(\boldsymbol{u}), \dots, \tau_d(\boldsymbol{u}))$ be a map from $[0,1]^m$ to the closed and bounded set $\mathcal{X} \subset \mathbb{R}^d$ such that $\partial^{1:m}\tau_j$ exists for all $j = 1, \dots, d$. Assume that $\partial^v \tau_j \in L^{p_j}([0,1]^m)$ for all $j = 1, \dots, d$ and for all nonempty $v \subseteq 1{:}m$, where $p_j \in [1, \infty]$. If $\sum_{j=1}^d 1/p_j \leqslant 1/2$, then $f \circ \tau$ is smooth in the sense of* (4) *for all $f \in C^m(\mathcal{X})$.*

Proof. The argument from Corollary 5 applies here. ◻

Necessary conditions are more subtle. To take an extreme case, τ could fail to be in BVHK or to be smooth, but f could repair that problem by being constant everywhere, or just in a region outside of which τ is well behaved. Our working definition is that we consider a transformation τ to be unsuitable for QMC when one or more of the components τ_j has $\partial^u \tau_j(\cdot : \mathbf{1}_{-v}) \notin L_1$ for some $v \subset 1{:}m$. In that case even the coordinate function $\tau_j \notin$ BVHK. Similarly, if $\partial^v \tau_j \notin L^2$ for any j and v, then the transformation τ is one that does not lead to the improved rate for scrambled nets even for integration of τ, much less $f \circ \tau$ for all $f \in C^m(\mathcal{X})$.

It is possible to weaken the condition on f in Corollary 5 while strengthening the conditions on τ_j. We could require only that $(f_{\boldsymbol{\lambda}} \circ \tau)(\cdot{:}\mathbf{1}_{-v})$ is in $L^{p_0}([0,1]^{|v|})$ whenever $|\boldsymbol{\lambda}| \leqslant m$ and then require $\sum_{j=0}^d 1/p_j \leqslant 1$. Similarly for Corollary 7, we could require $f_{\boldsymbol{\lambda}} \circ \tau \in L^{p_0}$ whenever $|\boldsymbol{\lambda}| \leqslant m$, where $\sum_{j=0}^d 1/p_j \leqslant 1/2$.

5. Counterexamples.

In this section we give two common transformations for which some $\tau_j \notin$ BVHK, which means they do not satisfy the conditions of Theorem 4. Thus unless f is somehow specially suited to counter the infinite variation coming from its argument, we will have $\mathrm{V}_{\mathrm{HK}}(f \circ \tau) = \infty$.

5.1. Map from [0, 1]^3 to an equilateral triangle.

Let $T^2 = \{\boldsymbol{x} \in [0,1]^3 \mid \sum_{j=1}^3 x_j = 1\}$, an equilateral triangle. Consider the map $\tau : [0,1]^3 \to T^2$ defined by

$$\tau_j(\boldsymbol{u}) = \frac{\log u_j}{\sum_{i=1}^3 \log u_i}, \quad j = 1, 2, 3. \tag{15}$$

It is well known that $\tau(\boldsymbol{u}) \sim \mathbf{U}(T^2)$ when $\boldsymbol{u} \sim \mathbf{U}([0,1]^3)$. The mapping in (15) is well defined for $\boldsymbol{u} \in (0,1)^3$. There are various reasonable ways to extend it to problematic boundary points with either some $u_j = 0$ or with all $u_j = 1$. We will show that none of those extensions can yield $\tau_j \in$ BVHK.

First we find that

$$\iint_{(0,1)^2} \left|\frac{\partial^2 \tau_1}{\partial u_1 \partial u_2}\right|_{u_3=1} \mathrm{d}u_1\,\mathrm{d}u_2 = \iint_{(0,1)^2} \left|\frac{\log u_1 - \log u_2}{u_1 u_2 (\sum_{i=1}^2 \log u_i)^3}\right| \mathrm{d}u_1\,\mathrm{d}u_2.$$

After a change of variable to $x_1 = \log u_1$ and $x_2 = \log u_2$ the integral becomes

$$\begin{aligned}
&\int_{-\infty}^0 \int_{-\infty}^0 \left|\frac{x_1 - x_2}{(x_1+x_2)^3}\right| \mathrm{d}x_1\,\mathrm{d}x_2 \\
&= \int_{-\infty}^0 \int_{-\infty}^{x_1} \frac{x_1 - x_2}{(x_1+x_2)^3} \mathrm{d}x_2\,\mathrm{d}x_1 + \int_{-\infty}^0 \int_{x_1}^0 \frac{x_2 - x_1}{(x_1+x_2)^3} \mathrm{d}x_2\,\mathrm{d}x_1 \\
&= \int_{-\infty}^0 \frac{1}{2x_1} \mathrm{d}x_1 = \infty.
\end{aligned}$$

Thus $\tau \notin$ BVHK. There is no extension from $(0,1)^3$ to $[0,1]^3$ that would yield $\tau \in$ BVHK. The same argument applies if $\tau_j(\boldsymbol{u}) = \log(u_j)/\sum_{i=1}^m \log(u_i)$ for any $m \geqslant 3$, mapping $[0,1]^m$ to a $d = m-1$ dimensional simplex. We can set $u_i = 1$ for $i \geqslant 3$ and integrate as before.

5.2. Inverse Gaussian map to the hypersphere. A very convenient way to sample uniformly from the sphere $\mathbb{S}^{d-1} = \{\boldsymbol{x} \in \mathbb{R}^d \mid \|\boldsymbol{x}\| = 1\}$ is to generate d independent $\mathcal{N}(0,1)$ random variables and standardize them. We write φ and Φ for the probability density function and cumulative distribution function, respectively, of $\mathcal{N}(0,1)$. The mapping from $[0,1]^d$ to $\mathcal{X} = \mathbb{S}^{d-1}$ is

$$\tau_j(\boldsymbol{u}) = \frac{\Phi^{-1}(u_j)}{\sqrt{\sum_{i=1}^d \Phi^{-1}(u_i)^2}}.$$

We will use the double factorial function $n!! = n(n-2)(n-4)\cdots 1$ for odd n and set $g(n) = (2n-1)!!$. For $j \in 1{:}d$ and $v \subset 1{:}d$ with $j \notin v$ we find that

$$\partial^v \tau_j = \frac{(-1)^{|v|} g(|v|) \Phi^{-1}(u_j)}{\left(\sum_{i=1}^d \Phi^{-1}(u_i)^2\right)^{|v|+1/2}} \times \prod_{i\in v} \frac{\Phi^{-1}(u_i)}{\varphi(\Phi^{-1}(u_i))}.$$

Now if $j \in v$, we can write after some algebra,

$$\begin{aligned}
\partial^v \tau_j &= \frac{\partial}{\partial u_j} \partial^{v-\{j\}} \tau_j \\
&= \frac{(-1)^{|v|} g(|v|) \prod_{i\in v-\{j\}} \Phi^{-1}(u_i) \left[2(|v|-1)\Phi^{-1}(u_j)^2 - \sum_{i\neq j} \Phi^{-1}(u_i)^2\right]}{\left(\sum_{i=1}^d \Phi^{-1}(u_i)^2\right)^{|v|+1/2} \prod_{i\in v} \varphi(\Phi^{-1}(u_i))}.
\end{aligned}$$

We choose $v = 1{:}d$ and integrate $|\partial^{1:d}\tau_j|$ over $[0,1]^d$. The integration is done with a change of variable $x_i = \Phi^{-1}(u_i)$ so $\mathrm{d}u_i = \varphi(x_i)\,\mathrm{d}x_i$. Because $g(d-1) \geqslant 1$,

$$\begin{aligned}\int_{[0,1]^d} \bigl|\partial^{1:d}\tau_j\bigr|\,\mathrm{d}u &\geqslant \int_{\mathbb{R}^d} \left|\frac{\prod_{i\neq j} x_i\bigl[2(d-1)x_j^2 - \sum_{i\neq j} x_i^2\bigr]}{\bigl(\sum_{i=1}^d x_i^2\bigr)^{d+1/2}}\right| \mathrm{d}\boldsymbol{x} \\ &\geqslant \int_{[0,\infty)^{d-1}} \int_0^\infty \frac{\bigl|\prod_{i\neq j} x_i\bigl[2(d-1)x_j^2 - \sum_{i\neq j} x_i^2\bigr]\bigr|}{\bigl(\sum_{i=1}^d x_i^2\bigr)^{d+1/2}}\,\mathrm{d}x_j\,\mathrm{d}\boldsymbol{x}_{-j} \\ &\geqslant \int_{[0,\infty)^{d-1}} \int_0^{\left(\frac{\sum_{i\neq j} x_i^2}{2(d-1)}\right)^{1/2}} \frac{\prod_{i\neq j} x_i\bigl[\sum_{i\neq j} x_i^2 - 2(d-1)x_j^2\bigr]}{\bigl(\sum_{i=1}^d x_i^2\bigr)^{d+1/2}}\,\mathrm{d}x_j\,\mathrm{d}\boldsymbol{x}_{-j} \\ &\geqslant \int_{[0,\infty)^{d-1}} \prod_{i\neq j} x_i \int_0^{\left(\frac{\sum_{i\neq j} x_i^2}{2(d-1)}\right)^{1/2}} \frac{\bigl[\sum_{i\neq j} x_i^2 - 2(d-1)x_j^2\bigr]}{\bigl(\frac{2d-1}{2d-2}\sum_{i\neq j}^d x_i^2\bigr)^{d+1/2}}\,\mathrm{d}x_j\,\mathrm{d}\boldsymbol{x}_{-j} \\ &= \tilde{K} \int_{[0,\infty)^{d-1}} \frac{\prod_{i\neq j} x_i}{\bigl(\sum_{i\neq j} x_i^2\bigr)^{d-1}}\,\mathrm{d}\boldsymbol{x}_{-j},\end{aligned}$$

where $\tilde{K} = ((2d-2)/(2d-1))^{d+1/2}$.

Now we integrate this one at a time for each $i \neq j$. Note that for $k < d-1$,

$$\int_0^\infty \frac{x}{(x^2+z)^{d-k}}\,\mathrm{d}x = c_k \frac{1}{z^{d-k-1}}, \tag{16}$$

where $c_k = 1/(2(d-k-1))$. Applying (16) repeatedly for $k = 1$ to $k = d-2$, we get

$$\int_{[0,1]^d} \bigl|\partial^{1:d}\tau_j\bigr|\,\mathrm{d}\boldsymbol{u} \geqslant \left(\tilde{K}\prod_{k=1}^{d-2} c_k\right) \int_0^\infty \frac{1}{x_j}\,\mathrm{d}x_j = \infty$$

and so $\tau_j \notin$ BVHK for all $j \in 1{:}d$.

6. Mappings from Fang and Wang (1993). Fang and Wang (1993) provide mappings from the unit cube to other important spaces for quadrature problems. Their mappings are more complicated than the elegant symmetric ones in section 5. Instead of symmetry, their mappings are designed to use a unit cube of exactly the same dimension as the space they map to. The domains that they consider, and their nomenclature for them, are

$$\begin{aligned} A_d &= \{(x_1,\dots,x_d) : 0 \leqslant x_1 \leqslant \cdots \leqslant x_d \leqslant 1\}, \\ B_d &= \{(x_1,\dots,x_d) : x_1^2 + \cdots + x_d^2 \leqslant 1\}, \\ U_d &= \{(x_1,\dots,x_d) : x_1^2 + \cdots + x_d^2 = 1\}, \\ V_d &= \{(x_1,\dots,x_d) \in \mathbb{R}_+^d : x_1 + \cdots + x_d \leqslant 1\}, \quad \text{and} \\ T_d &= \{(x_1,\dots,x_d) \in \mathbb{R}_+^d : x_1 + \cdots + x_d = 1\}. \end{aligned} \tag{17}$$

Spaces A_d, V_d, and T_{d+1} are all simplices of dimension d, B_d is a ball, and U_d is the $d-1$ dimensional hypersphere.

We show next that all of their mappings have components τ in BVHK and none of them have all mixed partial derivatives in L^2. They are thus better suited to QMC than the symmetric mappings are but they are not able to take advantage of the improved rate for RQMC versus QMC. The transformations have a separable character that lets us study them directly without recourse to the generalized Hölder inequality.

6.1. Mapping from $[0,\,1]^d$ to A_d. The map $\tau = (\tau_1, \dots, \tau_d)$ is given by $\tau_j(\boldsymbol{u}) = \prod_{i=j}^d u_i^{1/i}$ for $j = 1, \dots, d$. We find that

$$(\partial^{1:d}\tau_1)^2 = \prod_{i=1}^d \frac{1}{i^2} u_i^{2/i-2},$$

which diverges on integrating with respect to u_2. Thus $\partial^{1:d}\tau_1$ is not in L^2, outside the trivial case $d = 1$.

Next we show that τ satisfies the BVHK conditions of Theorem 4. Pick any nonempty $\ell \subseteq 1{:}d$. If there exists $i \in \ell$ with $i < j$, then $\partial^\ell \tau_j = 0$, so we may assume that $\ell \subseteq j{:}d$. For such an ℓ, we have $\partial^\ell \tau_j = \prod_{i\in\ell} i^{-1} u_i^{1/i-1} \prod_{i \in j:d-\ell} u_i^{1/i}$. The exponents are all above -1 and so this quantity is integrable. The integrand in (12) is a product where no two of the factors are differentiated with respect to the same variable. Thus all the powers of any u_i are above -1 and so this τ satisfies the conditions of Theorem 4.

6.2. Mapping from $[0,\,1]^d$ to B_d. The mapping involves the inverse transform of a distribution function on B_d. Define

$$F_j(x) = \begin{cases} x^d & \text{if } j = 1, \\ \frac{\pi}{B\left(\frac{1}{2}, \frac{d-j+1}{2}\right)} \int_0^x (\sin \pi t)^{d-j}\,\mathrm{d}t & \text{if } j = 2, \dots, d, \end{cases}$$

where $B(\cdot,\cdot)$ is the Beta function. Next define intermediate variables

$$b_1 = u_1^{1/d} \quad \text{and} \quad b_i = F_i^{-1}(u_i) \qquad \text{for } i = 2, \dots, d.$$

Their mappings are then

$$\begin{aligned} \tau_j &= b_1 \prod_{i=2}^j \sin(\pi b_i)\cos(\pi b_{j+1}) \quad \text{for } 1 \leqslant j \leqslant d-2, \\ \tau_{d-1} &= b_1 \prod_{i=2}^{d-1} \sin(\pi b_i)\cos(2\pi b_d), \quad \text{and} \\ \tau_d &= b_1 \prod_{i=2}^{d-1} \sin(\pi b_i)\sin(2\pi b_d). \end{aligned}$$

For $d = 2$ we get $F_2(x) = x$ and so $\tau_2 = u_1^{1/2}\sin(2\pi u_2)$. Therefore $\partial^{1:2}\tau_2 = \pi\cos(2\pi u_2)/\sqrt{u_1}$ which is not in L^2. For general $d > 2$, we have

$$\partial^{1:d}\tau_d = \frac{1}{d} u_1^{1/d-1} \left(\prod_{i=2}^{d-1} \pi \cos(\pi b_i) \frac{\partial b_i}{\partial u_i} \right) 2\pi\cos(2\pi u_d), \tag{18}$$

which is also not in L^2 because of the factor $u_1^{1/d-1}$.

For later use with the transformation to U_d, we also consider the factor for $i = d-1$ in (18). The definition of b_{d-1} simplifies to $b_{d-1} = \cos^{-1}(1 - 2u_{d-1})/\pi$ and so

$$\frac{\partial b_{d-1}}{\partial u_{d-1}} = \frac{2}{\pi \sin(\pi b_{d-1})}.$$

This simplifies the above mixed partial to

$$\partial^{1:d}\tau_d = \frac{1}{d}u_1^{1/d-1}\left(\prod_{i=2}^{d-2}\pi\cos(\pi b_i)\frac{\partial b_i}{\partial u_i}\right)\left(2\frac{(1-2u_{d-1})}{\sqrt{1-(1-2u_{d-1})^2}}\right)2\pi\cos(2\pi u_d).$$

Now

$$\int_0^1 \frac{(1-2u_{d-1})^2}{1-(1-2u_{d-1})^2}du_{d-1} = \frac{1}{4}\left(\log x - \log(1-x) - 4x\right)\Bigg|_0^1 = \infty. \tag{19}$$

To show that this transformation is in BVHK we again use Theorem 4. We must show that

$$\int_{[0,1]^{|v|}}\Big|\prod_{r=1}^{s}\partial^{\ell_r}\tau_{k_r}(\boldsymbol{u}_v{:}\mathbf{1}_{-v})\Big|\,\mathrm{d}\boldsymbol{u}_v < \infty$$

for $v \subseteq 1{:}m$, $s \leqslant |v|$ distinct $k_r \in 1{:}d$ and disjoint ℓ_r with union v. Thus, we differentiate at most once with respect to any original variable and we get

$$\int_{[0,1]^{|v|}}\Big|\prod_{r=1}^{s}\partial^{\ell_r}\tau_{k_r}(\boldsymbol{u}_v{:}\mathbf{1}_{-v})\Big|\,\mathrm{d}\boldsymbol{u}_v \leqslant \int_{[0,1]^{|v|}}\Big|\prod_{i\in v}\frac{\partial b_i}{\partial u_i}\Big|\,\mathrm{d}\boldsymbol{u}_v \leqslant \prod_{i\in v}\int_{[0,1]}\Big|\frac{\partial b_i}{\partial u_i}\Big|\,\mathrm{d}u_i.$$

Now

$$\frac{\partial b_i}{\partial u_i} = \frac{B\left(\frac{1}{2},\frac{d-j+1}{2}\right)}{\pi\left[\sin\left(\pi F_i^{-1}(u_i)\right)\right]^{d-i}}.$$

Note that $F_i^{-1}(u_i) \in [0,1]$ for all $u_i \in [0,1]$. Thus we have

$$\int_{[0,1]}\Big|\frac{\partial b_i}{\partial u_i}\Big|\,\mathrm{d}u_i = \int_{[0,1]}\frac{\partial b_i}{\partial u_i}\,\mathrm{d}u_i = 1,$$

as required.

6.3. Mapping from $[0,1]^{d-1}$ to U_d. This mapping is similar to one in B_d, with the densities being different. Here we have

$$f_j(u) = \frac{\pi}{B\left(\frac{1}{2},\frac{d-j}{2}\right)}(\sin(\pi u))^{d-j-1}$$

and $b_i = F_i^{-1}(u_i)$ for $i = 1,\dots,d-1$. The explicit transformation can be written as

$$\tau_j = \prod_{i=1}^{j-1}\sin(\pi b_i)\cos(\pi b_j) \qquad \text{for } j = 1,\dots,d-2,$$

$$\tau_{d-1} = \prod_{i=1}^{d-2}\sin(\pi b_i)\cos(2\pi b_{d-1}), \quad \text{and}$$

$$\tau_d = \prod_{i=1}^{d-2}\sin(\pi b_i)\sin(2\pi b_{d-1}).$$

We first consider the case $d = 2$. It is easy to see that

$$\tau_1 = \cos(2\pi u_1) \quad \text{and} \quad \tau_2 = \sin(2\pi u_1).$$

Note that in this case, $\partial^v \tau_j \in L^2$ for each $v \subseteq \{1, 2\}$ and $j = 1, 2$. This is the only case with this property, but then the set U_d is intrinsically one-dimensional. For $d \geqslant 3$, consider the $(d-2)$th term in the expansion of $\partial^{1:d-1}\tau_d$,

$$\partial^{1:d-1}\tau_d = \left(\prod_{i=1}^{d-3} \pi \cos(\pi bi) \frac{\partial b_i}{\partial u_i}\right) \left(2 \frac{(1-2u_{d-2})}{\sqrt{1-(1-2u_{d-2})^2}}\right) 2\pi \cos(2\pi u_{d-1}).$$

Using (18) as in the previous case, this proves that $\partial^{1:d-1}\tau_d \notin L^2$. Furthermore, following the same argument as in section 6.2, we may show that this transformation satisfies Theorem 4 yielding $f \circ \tau \in$ BVHK for $f \in C^{d-1}$.

6.4. Mapping from $[0, 1]^d$ to V_d. Assume that $d \geqslant 2$. Then we have

$$\tau_i = u_1^{1/d} \prod_{j=2}^{i} u_j^{\frac{1}{d-j+1}} \left(1 - u_{i+1}^{\frac{1}{d-i}}\right) \qquad \text{for } i = 1, \ldots, d-1 \quad \text{and}$$

$$\tau_d = u_1^{1/d} \prod_{j=2}^{d} u_j^{\frac{1}{d-j+1}}.$$

Considering the mixed partial, $\partial^{1:d}\tau_d$ we have

$$\partial^{1:d}\tau_d = \frac{1}{d} u_1^{\frac{1}{d}-1} \prod_{j=2}^{d} \frac{1}{d-j+1} u_j^{\frac{1}{d-j+1}-1} = \frac{1}{d!} \frac{1}{u_1^{\frac{d-1}{d}} u_2^{\frac{d-2}{d-1}} \ldots u_{d-1}^{1/2}}.$$

Observing the integral with respect to u_{d-1} it is clear that $\partial^{1:d}\tau_d \notin L^2$. Furthermore, following the same argument as in section 6.1, we may show that this transformation satisfies Theorem 4.

6.5. Mapping from $[0, 1]^{d-1}$ to T_d. Similar to V_d, and assuming that $d \geqslant 3$ to get a dimension of at least 2, we have

$$\tau_i = \prod_{j=1}^{i-1} u_j^{\frac{1}{d-j}} \left(1 - u_i^{\frac{1}{d-i}}\right) \qquad \text{for } i = 1, \ldots, d-1,$$

$$\tau_d = \prod_{j=1}^{d-1} u_j^{\frac{1}{d-j}}.$$

It is thus clear from

$$\partial^{1:(d-1)}\tau_d = \frac{1}{(s-1)!} \frac{1}{u_1^{\frac{d-2}{d-1}} u_2^{\frac{d-3}{d-2}} \ldots u_{d-2}^{1/2} u_{d-1}}$$

that $\partial^{1:(d-1)}\tau_d \notin L^2$. Following the argument in section 6.1, we may show that this transformation satisfies Theorem 4.

6.6. Efficient mapping from $[0, 1]^{d-1}$ to U_d. Fang and Wang (1993) gave another mapping to U_d which avoids computing the incomplete beta function that was used in section 6.3. Once again the transformation fails to have all partial derivatives in L^2. We assume that $d \geqslant 3$ and we deal with the case of d being even and odd differently.

Even case: $\mathbf{d = 2m}$. Here we have $(u_1, \dots, u_{d-1}) \in [0,1]^{d-1}$. Define $g_m = 1$ and $g_0 = 0$. For j from $m-1$ down to 1, let $g_j = g_{j+1} u_j^{1/j}$. Put $d_l = \sqrt{g_l - g_{l-1}}$. Now for $l = 1, \dots, m$, define

$$\tau_{2l-1} = d_l \cos(2\pi u_{m+l-1}) \quad \text{and} \quad \tau_{2l} = d_l \sin(2\pi u_{m+l-1}).$$

It is easy to see that

$$\tau_1 = d_1 \cos(2\pi u_m) = \prod_{j=1}^{m-1} u_j^{1/2j} \cos(2\pi u_m)$$

and so

$$|\partial^{1:m}\tau_1| = \left| \prod_{j=1}^{m-1} \frac{1}{2j} u_j^{1/2j-1} 2\pi \sin(2\pi u_m) \right| = \left| \frac{1}{2^{m-1}(m-1)!} \frac{2\pi \sin(2\pi u_m)}{u_1^{\frac{1}{2}} u_2^{\frac{3}{4}} \dots u_{m-1}^{\frac{2m-3}{2m-2}}} \right|.$$

Integrating with respect to u_1 shows that $\partial^{1:m}\tau_1 \notin L^2$.

Odd case: $\boldsymbol{d = 2m+1}$. Again we begin with $(u_1, \dots, u_{d-1}) \in [0,1]^{d-1}$. Define $g_m = 1$ and $g_0 = 0$. For $j = m-1$ down to $j = 1$, let $g_j = g_{j+1} u_j^{2/(2j+1)}$. As for the even case, put $d_l = \sqrt{g_l - g_{l-1}}$. Now for $l = 2, \dots, m$, define

$$\begin{aligned}
\tau_1 &= d_1(1 - 2u_m), \\
\tau_2 &= d_1 \sqrt{u_m(1-u_m)} \cos(2\pi u_{m+1}), \\
\tau_3 &= d_1 \sqrt{u_m(1-u_m)} \sin(2\pi u_{m+1}), \quad \text{and then} \\
\tau_{2l} &= d_l \cos(2\pi u_{2l}), \quad \text{and} \quad \tau_{2l+1} = d_l \sin(2\pi u_{2l}).
\end{aligned}$$

Simplifying d_1 we see that

$$\tau_2 = u_1^{\frac{1}{3}} u_2^{\frac{1}{5}} \dots u_{m-1}^{\frac{1}{2m-1}} \sqrt{u_m(1-u_m)} \cos(2\pi u_{m+1}).$$

Thus

$$\left| \frac{\partial \tau_2}{\partial u_m} \right|_{u_j = 1, j \neq m} = \frac{1 - 2u_m}{2\sqrt{u_m(1-u_m)}}.$$

Applying (19) we see that $\partial^{u_m}\tau_2 \notin L^2$.

All of the τ_j are in BVHK. This follows from the fact that each component of the transformation is a product of functions of only one of the original variables and hence it is completely separable.

7. Nonuniform transformations. Here we consider transformations that are not uniformity preserving. Section 7.1 considers importance sampling methods for integrals with respect to a nonuniform measure on $\mathcal{X}$. Section 7.2 gives conditions for sequential inversion to yield an integrand in BVHK. Section 7.3 shows that some importance sampling transformations lead to the $O(n^{-3/2+\epsilon})$ rate for RMSE on the simplex for a class of functions including polynomials.

Suppose that $\mu = \int_{\mathcal{X}} f(\boldsymbol{x}) p(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$ for a nonuniform density p. Instead of sampling $\boldsymbol{x}_i$ from the uniform distribution on $\mathcal{X}$ and averaging fp a Monte Carlo strategy can be to sample $\boldsymbol{x}_i \sim p$ and average f or, if $q(\boldsymbol{x}) > 0$ whenever $f(\boldsymbol{x})p(\boldsymbol{x}) \neq 0$, sample

$\boldsymbol{x}_i \sim q$ and average fp/q. This latter technique is known as importance sampling. We consider it in section 7.1.

Aistleitner and Dick (2015) show that if f is a measurable function on $[0,1]^d$ which is BVHK and P is a normalized Borel measure on $[0,1]^d$, then for $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$ in $[0,1]^d$,

$$\left|\frac{1}{n}\sum_{i=1}^n f(\boldsymbol{x}_i) - \int_{[0,1]^d} f(\boldsymbol{x})\,\mathrm{d}P(\boldsymbol{x})\right| \leqslant \mathrm{V}_{\mathrm{HK}}(f) D_n^*(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n;P), \tag{20}$$

where

$$D_n^*(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n;P) = \sup_{A\in\mathcal{A}^*}\left|\frac{1}{n}\sum_{i=1}^n \mathbf{1}_A(\boldsymbol{x}_i) - P(A)\right|$$

and $\mathcal{A}^*$ is the class of all closed axis-parallel boxes contained in $[0,1]^d$. Aistleitner and Dick (2013) prove that for any measure P and any n there exist points $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$ such that

$$D_n^*(\boldsymbol{x}_1,\dots,\boldsymbol{x}_n;P) \leqslant c_d(\log n)^{(3d+1)/2} n^{-1}.$$

They do not, however, give an explicit construction. Instead of using (20) one might use the original Koksma–Hlawka inequality by using an appropriate nonmeasure preserving transformation τ. Below we give a corollary to Theorem 4 stating sufficient conditions for attaining the QMC convergence rate when using importance sampling.

7.1. Importance sampling. We suppose that the measure P has a probability density p. We then use a transformation τ on $[0,1]^m$ which yields $\boldsymbol{x} = \tau(\boldsymbol{u})$ with probability density function q on $\mathcal{X}$ when $\boldsymbol{u} \sim \mathbf{U}[0,1]^m$. We estimate μ by

$$\hat{\mu}_q^n = \frac{1}{n}\sum_{i=1}^n \frac{f(\tau(\boldsymbol{u}_i))p(\tau(\boldsymbol{u}_i))}{q(\tau(\boldsymbol{u}_i))} = \frac{1}{n}\sum_{i=1}^n \left(\frac{fp}{q}\circ\tau\right)(\boldsymbol{u}_i). \tag{21}$$

If $q(\boldsymbol{x}) > 0$ whenever $f(\boldsymbol{x})p(\boldsymbol{x}) \neq 0$ (and if μ exists), then $\mathbb{E}(\hat{\mu}_q^n) = \mu$. Thus, to apply the Koksma–Hlawka inequality we only need $(fp/q)\circ\tau \in \mathrm{BVHK}[0,1]^m$. Following Theorem 4 we can now state sufficient conditions for the above to hold.

COROLLARY 8. *Let p and q denote densities on the closed and bounded set $\mathcal{X} \subset \mathbb{R}^d$ with $q(\boldsymbol{x}) > 0$ whenever $p(\boldsymbol{x}) > 0$. Let τ be a map from $[0,1]^m$ to $\mathcal{X}$ for which $\boldsymbol{u} \sim \mathbf{U}[0,1]^m$ implies $\boldsymbol{x} = \tau(\boldsymbol{u}) \sim q$. Moreover, assume that τ satisfies the conditions of Theorem 4 and that $fp/q \in C^m(\mathcal{X})$. Then, for a low-discrepancy point set $\boldsymbol{u}_1,\dots,\boldsymbol{u}_n$ in $[0,1]^m$,*

$$\left|\int_{\mathcal{X}} f(\boldsymbol{x})p(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} - \frac{1}{n}\sum_{i=1}^n\left(\frac{fp}{q}\circ\tau\right)(\boldsymbol{u}_i)\right| = O\left(\frac{(\log n)^{d-1}}{n}\right).$$

Proof. The result follows from Theorem 4 and the Koksma–Hlawka inequality. □

There is a similar counterpart to Theorem 6. When τ satisfies the conditions there, $fp/q \in C^m$, and $\boldsymbol{u}_1,\dots,\boldsymbol{u}_n$ are a digital net with a nested uniform or linear scramble, then the RMSE of $\hat{\mu}_q^n$ is $O(n^{-3/2+\epsilon})$. In both cases it is clearly advantageous to have p/q bounded above, as is often recommended for importance sampling in Monte Carlo. See, for instance, Geweke (1989).

Importance sampling is often used for quantities with long tails and for quantities dominated by narrow modes. The first sort of problem usually arises on an unbounded domain. Equation (21) yields an unbiased estimate of μ when $\boldsymbol{u}_i$ are MC or RQMC points, even if $\mathcal{X}$ is not bounded. But Corollary 8 requires boundedness in order to apply Theorem 4. So the most interesting setting for Corollary 8 is in problems involving "spiky" integrands on compact sets $\mathcal{X}$.

If $f \in C^m$, then $fp/q \in C^m$ too so long as $p/q \in C^m$. A common way to generate importance sampling distributions is to use exponential tilting (Asmussen and Glynn, 2007), taking $q(\boldsymbol{x}) \propto p(\boldsymbol{x})\exp(\theta^\mathsf{T}\boldsymbol{x})$ for a parameter $\theta \in \mathbb{R}^d$. Then $p/q \in C^m(\mathcal{X})$ when $\mathcal{X}$ is bounded.

An early contribution on QMC with importance sampling is in the thesis of Chelson (1976). He gives a Koksma–Hlawka inequality for QMC with importance sampling but uses the discrepancy of points with respect to the uniform measure. Aistleitner and Dick (2015) correct that result using the discrepancy with respect to P.

7.2. Sequential inversion. For $d = 1$, a standard way to generate a non-uniform random variable is to invert the CDF at a uniformly distributed point. The multivariate version of this procedure can be used to sample from an arbitrary distribution provided we can invert all the necessary conditional CDFs.

Let F be the target distribution on the compact set $\mathcal{X} \subset \mathbb{R}^d$. Further let F_1 be the marginal distribution of X_1, and for $j = 2, \dots, d$, let $F_{j|1:(j-1)}(\,\cdot \mid \boldsymbol{x}_{1:(j-1)})$ be the conditional CDF of X_j given $X_1, \dots, X_{j-1}$. The transformation τ of $\boldsymbol{u} \in [0,1]^d$ is given by $\boldsymbol{x} = \tau(\boldsymbol{u}) \in \mathbb{R}^d$, where

$$x_1 = F_1^{-1}(u_1) \quad \text{and} \quad x_j = F_{j|1:(j-1)}^{-1}(u_j \mid \boldsymbol{x}_{1:(j-1)}) \quad \text{for } j \geqslant 2. \tag{22}$$

Conditional sampling via (22) goes back at least to Shreider (1966). The inverse transformation, from $\boldsymbol{x}$ to $\boldsymbol{u}$, was studied by Rosenblatt (1952). Hlawka and Mück (1972) studied the use of this transformation for generating low discrepancy points. Under conditions on F they show that the resulting points have a discrepancy with respect to F of order $D_n^{1/d}$ where D_n is the discrepancy of points $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ that it uses. Because discrepancy can at best be $O(n^{-1}\log(n)^{d-1})$, that rate has a severe deterioration with respect to dimension d.

We consider the case of $d = 2$ dimensions. Then $\tau(\boldsymbol{u}) = (\tau_1(\boldsymbol{u}), \tau_2(\boldsymbol{u}))$, where

$$\tau_1(u_1, u_2) = F_1^{-1}(u_1) \quad \text{and} \quad \tau_2(u_1, u_2) = F_{2|1}^{-1}(u_2 \mid F_1^{-1}(u_1)). \tag{23}$$

Let $f_{1,2}$ be the joint density, f_1 be the marginal density of X_1 corresponding to F_1, and $f_{2|1}(\cdot|x_1)$ be the conditional density of X_2 given $X_1 = x_1$. Then

$$\begin{aligned}
\frac{\partial \tau_1}{\partial u_1} &= \frac{1}{f_1(F_1^{-1}(u_1))} = \frac{1}{f_1(\tau_1(\boldsymbol{u}))} \quad \text{and}\\
\frac{\partial \tau_2}{\partial u_2} &= \frac{1}{f_{2|1}(F_{2|1}^{-1}(u_2 \mid F_1^{-1}(u_1)) | F_1^{-1}(u_1))}\\
&= \frac{f_1(F_1^{-1}(u_1))}{f_{1,2}(F_1^{-1}(u_1), F_{2|1}^{-1}(u_2 \mid F_1^{-1}(u_1)))} = \frac{f_1(\tau_1(\boldsymbol{u}))}{f_{1,2}(\tau(\boldsymbol{u}))}.
\end{aligned}$$

Now $\tau_1 \in \text{BVHK}$ because the support of X_1 is finite by compactness of $\mathcal{X}$. Let $a = F_1^{-1}(0)$ and $b = F_1^{-1}(1)$ define the support $[a, b]$ of X_1. Similarly, define $a_2(x_1) = F_{2|1}^{-1}(0 \mid x_1)$ and $b_2(x_1) = F_{2|1}^{-1}(1 \mid x_1)$ which are both finite by compactness of $\mathcal{X}$. Then X_2 has support $[a_2(x_1), b_2(x_1)]$ given that $X_1 = x_1$.

We now consider sufficient conditions for τ_2 to be in BVHK. Assuming that $f_{1,2}$ is strictly positive, we have

$$\int_0^1 \left|\frac{\partial \tau_2}{\partial u_2}\right|_{u_1=1} \mathrm{d}u_2 = \int_0^1 \frac{1}{f_{2|1}(F_{2|1}^{-1}(u_2|b)|b)}\,\mathrm{d}u_2$$
$$= F_{2|1}^{-1}(1|b) - F_{2|1}^{-1}(0|b) = b_2(b) - a_2(b).$$

Recall that $f^{\{k\}}$ denotes partial differentiation of f with respect to x_k for $k = 1, 2$. From the implicit function theorem, we obtain

$$\frac{\partial \tau_2(u_1,u_2)}{\partial u_1} = \frac{\frac{-\partial F_{2|1}}{\partial x_1}\Big|_{(F_1^{-1}(u_1),\tau_2(u_1,u_2))}\frac{1}{f_1(F_1^{-1}(u_1))}}{\frac{\partial F_{2|1}}{\partial x_2}\Big|_{(F_1^{-1}(u_1),\tau_2(u_1,u_2))}} = \frac{\frac{-\partial F_{2|1}}{\partial x_1}\Big|_{\tau(\boldsymbol{u})}}{f_1(\tau_1(\boldsymbol{u}))\times\frac{\partial F_{2|1}}{\partial x_2}\Big|_{\tau(\boldsymbol{u})}},$$

where

$$\frac{\partial F_{2|1}(x_1,x_2)}{\partial x_1} = \int_{F_{2|1}^{-1}(1|x_1)}^{x_2} \frac{f_{1,2}^{\{1\}}(x_1,t)f_1(x_1) - f_{1,2}(x_1,t)f_1^{\{1\}}(x_1)}{f_1(x_1)^2}\,\mathrm{d}t \tag{24}$$

and

$$\frac{\partial F_{2|1}(x_1,x_2)}{\partial x_2} = f_{2|1}(x_2|x_1). \tag{25}$$

Now we get, using a change of variable,

$$\int_0^1 \left|\frac{\partial \tau_2}{\partial u_1}\right|_{u_2=1} \mathrm{d}u_1 \leqslant \int_a^b \left|\frac{\partial F_{2|1}}{\partial x_1}\Big/\frac{\partial F_{2|1}}{\partial x_2}\right|_{(x_1,b(x_1))} \mathrm{d}x_1.$$

Finally to evaluate the complete mixed partial, differentiating τ_2 with respect to u_1 and u_2 we have

$$\frac{\partial^2 \tau_2}{\partial u_1 \partial u_2} = \frac{1}{f_{1,2}^2}\left(f_{1,2}\frac{f_1^{\{1\}}}{f_1} - f_{1,2}^{\{1\}} - f_{1,2}^{\{2\}} f_1 \frac{\partial \tau_2}{\partial u_1}\right).$$

This gives us

$$\int_0^1\int_0^1 \left|\frac{\partial^2 \tau_2}{\partial u_1 \partial u_2}\right| \mathrm{d}u_1\,\mathrm{d}u_2 = \int_0^1\int_0^1 \frac{1}{f_{1,2}^2}\left|f_{1,2}\frac{f_1^{\{1\}}}{f_1} - f_{1,2}^{\{1\}} - f_{1,2}^{\{2\}} f_1 \frac{\partial \tau_2}{\partial u_1}\right| \mathrm{d}u_1\,\mathrm{d}u_2.$$

Again by a change of variables via $x = F_1^{-1}(u_1)$ and $y = F_{2|1}^{-1}(u_2 \mid F_1^{-1}(u_1))$ we have

$$\int_0^1\int_0^1 \left|\frac{\partial^2 \tau_2}{\partial u_1 \partial u_2}\right| \mathrm{d}u_1\,\mathrm{d}u_2$$
$$= \int_a^b\int_{a_2(x)}^{b_2(x)} \left|\frac{f_1^{\{1\}}(x)}{f_1(x)} - \frac{1}{f_{1,2}(x,y)}\left(f_{1,2}^{\{1\}}(x,y) + f_{1,2}^{\{2\}}(x,y)\frac{\mathrm{d}y}{\mathrm{d}x}\right)\right| \mathrm{d}y\,\mathrm{d}x.$$

Let us write Δ for the total derivative operator:

$$\Delta f = \sum_{k=1}^d f^{\{k\}}(x_1,\ldots,x_d)\,\mathrm{d}x_k \quad \text{and} \quad \frac{\Delta f}{\mathrm{d}x_j} = \sum_{k=1}^d f^{\{k\}}(x_1,\ldots,x_d)\frac{\mathrm{d}x_k}{\mathrm{d}x_j}.$$

This allows us to write the integral above as

$$\int_0^1\int_0^1\left|\frac{\partial^2\tau_2}{\partial u_1\partial u_2}\right|\,\mathrm{d}u_1\,\mathrm{d}u_2=\int_a^b\int_{a_2(x)}^{b_2(x)}\left|\frac{\Delta\log(f_1)}{\mathrm{d}x}-\frac{\Delta\log(f_{1,2})}{\mathrm{d}x}\right|\mathrm{d}y\,\mathrm{d}x.$$

Combining these, we now have a sufficient condition for the inverse Rosenblatt transformation in two dimensions to be in BVHK, which we summarize in Lemma 9.

LEMMA 9. *Let X_1 be supported on the finite interval $[a,b]$ and, given $X_1=x_1$, let X_2 be supported on the finite interval $[a_2(x_1),b_2(x_1)]$. Let f_1 and $f_{1,2}$ be the densities of X_1 and (X_1,X_2), respectively. Then the inverse Rosenblatt transformation* (23) *is of bounded variation in the sense of Hardy and Krause if for each $k=1,2$,*

$$\int_a^b\int_{a_2(x_1)}^{b_2(x_1)}\left|\frac{\Delta\log(f_{1,\cdots,k})}{\mathrm{d}x_1}\right|\mathrm{d}x_2\,\mathrm{d}x_1<\infty\quad\textit{and}\tag{26}$$

$$\int_a^b\left|\frac{\partial F_{2|1}}{\partial x_1}\Big/\frac{\partial F_{2|1}}{\partial x_2}\right|_{(x_1,b_2(x_1))}\mathrm{d}x_1<\infty,\tag{27}$$

where $\partial F_{2|1}/\partial x_1$ and $\partial F_{2|1}/\partial x_2$ are given by (24) *and* (25)*, respectively.*

From condition (26) we see that the densities f_1 and $f_{1,2}$ can be problematic as they approach zero on $\mathcal{X}$, for then $\Delta\log f$ will become large. Thus we anticipate better results when these densities are uniformly bounded away from zero on $\mathcal{X}$. Condition (27) involves an integral over the upper boundary of $\mathcal{X}$. If that upper boundary is flat, that is, $b_2(x_1)$ is constant on $a\leqslant x_1\leqslant b$, then the partial derivative in the numerator there vanishes. It is possible to generalize Lemma 9 to $d>2$ but the resulting quantities become very difficult to interpret.

7.3. Importance sampling QMC for the simplex. We map $\boldsymbol{u}=(u_1,u_2,\ldots,u_d)\in[0,1]^d$ to $\boldsymbol{x}=\tau(\boldsymbol{u})$ in the simplex

$$A_d=\{(x_1,\ldots,x_d)\in[0,1]^d\mid x_1\leqslant x_2\leqslant\cdots\leqslant x_d\}.$$

The mapping is given by

$$x_j=\tau_j(\boldsymbol{u})=\prod_{k\geqslant j}u_k^{a_k}$$

for constants $a_k>0$. Fang and Wang's (1993) uniformity preserving mapping from section 6.1 uses $a_k=1/k$.

The Jacobian matrix for this transformation is upper triangular and hence the Jacobian determinant is

$$J(\boldsymbol{u})=\prod_{j=1}^d\frac{\partial x_j}{\partial u_j}=\prod_{j=1}^d a_j u_j^{a_j-1}\prod_{k>j}u_k^{a_k}=A\prod_{j=1}^d u_j^{ja_j-1},$$

where $A=\prod_j a_j$. The average of $J(\boldsymbol{u})$ is $1/\mathbf{vol}(A_d)=1/d!$ and $0\leqslant J(\boldsymbol{u})\leqslant A$. The choice of Fang and Wang (1993) gives $J=1/d!$ for all $\boldsymbol{u}$. It is desirable to have J be nearly constant. If $A\gg 1/d!$, then $J(\boldsymbol{u})$ is a very spiky function and that will tend to defeat the purpose of low discrepancy sampling.

The RQMC estimate of

$$\mu = d!\int_{A_d} f(\boldsymbol{x})\,\mathrm{d}\boldsymbol{x} = d!\int_{[0,1]^d} f(\tau(\boldsymbol{u}))J(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u} \quad \text{is}$$

$$\hat{\mu} = \frac{d!}{n}\sum_{i=1}^{n} f(\tau(\boldsymbol{u}_i))J(\boldsymbol{u}_i).$$

Suppose that $f \in C^d$. Ignoring the $d!$ factor, the integrand on $[0,1]^d$ is now $\widetilde{f}(\boldsymbol{u}) = f(\tau(\boldsymbol{u}))J(\boldsymbol{u})$, and $\partial^v \widetilde{f} = \sum_{w\subseteq v} \partial^w (f\circ\tau)\times\partial^{v-w}J$. The definition of τ_j in this case makes it convenient to work with a simple function class consisting of integrands of the form $\prod_{j=1}^d x_j^{q_j}$ for real values $q_j \geqslant 0$.

THEOREM 10. *For $\boldsymbol{x} \in A_d$ let $f(\boldsymbol{x}) = \prod_{j=1}^d x_j^{q_j}$ for $q_j \geqslant 0$. For $j = 1,\dots,d$ and $\boldsymbol{u} \in [0,1]^d$, define $x_j = \tau_j(\boldsymbol{u}) = \prod_{k\geqslant j} u_k^{a_k}$ and the Jacobian $J(\boldsymbol{u}) = \prod_{j=1}^d a_j u_j^{ja_j-1}$ where $a_j > 0$. Then $\partial^v f(\boldsymbol{x}(\boldsymbol{u}))J(\boldsymbol{u}) \in L^2[0,1]^d$ for all $v \subseteq 1{:}d$ and all q_j if and only if $a_j > 3/(2j)$ holds for $j = 1,\dots,d$.*

Proof. Let $Q_k = \sum_{j\leqslant k} q_k$ and $A = \prod_j a_j$. Then

$$\widetilde{f}(\boldsymbol{u}) = f(\boldsymbol{\tau}(\boldsymbol{u}))J(\boldsymbol{u}) = A\prod_{k=1}^{d} u_k^{ka_k-1+a_kQ_k}.$$

For $v \subseteq 1{:}d$ we find that $(\partial^v \widetilde{f}(\boldsymbol{u}))^2$ equals

$$A^2\prod_{k\in v}(ka_k - 1 + a_kQ_k)^2 u_k^{2(ka_k-2+a_kQ_k)}\prod_{k\in -v} u_k^{2(ka_k-1+a_kQ_k)}. \tag{28}$$

The coefficient $ka_k - 1 + a_kQ_k$ cannot vanish for all Q_k. Therefore (28) has a finite integral for all q_j if and only if $2(ja_j - 2 + a_jQ_j) > -1$ for all j and all $q_1,\dots,q_d$. This easily holds if all $a_j > 3/(2j)$. Conversely, suppose that $a_j \leqslant 3/(2j)$ for some j. We may choose $Q_j = 0$ and $v = \{j\}$ and see that (28) is not integrable. ∎

From Theorem 10 we see that RQMC can attain the $O(n^{-3/2+\epsilon})$ rate for functions of the form $\prod_{j=1}^d x_j^{q_j}$ on the simplex A_d. That rate extends to linear combinations of finitely many such functions, including polynomials and more. If we choose $a_j = 3/(2j) + \eta$ for some small $\eta > 0$, then for fixed d we have $d!J(\boldsymbol{u}) = (3/2)^d + O(\eta)$. There is thus a dimension effect. The integrand becomes more spiky as d increases. We can expect that the lead constant in the error bound will grow exponentially with d.

For $d = 1$, the simplex reduces to $A_1 = [0,1]$. Theorem 10 requires $a_1 > 3/2$ and then delivers the $O(n^{-3/2+\epsilon})$ for $f(x_1) = x_1^{q_1}$ and any $q_1 \geqslant 0$. Working directly with the integrand $\widetilde{f} \propto x_1^{a_1-1+a_1q_1}$ we find that $\int(\widetilde{f}'(x_1))^2\,\mathrm{d}x_1 < \infty$ when $2(a_1 + q_1) > 3$. Ordinary RQMC in $[0,1]$ corresponds to taking $a_1 = 1$ and so we see that it attains the better rate when $q_1 > 1/2$. Theorem 10 is thus somewhat conservative.

If we work only with polynomials taking only $q_j \in \mathbb{N}$, then the choice $a_k = 1/k$ zeros out (28) when $Q_k = 0$. The smallest nonzero Q_k is then 1 and we would need to impose $2(ka_k - 2 + a_kQ_k) > -1$. That simplifies to $Q_k > k/2$ which can only be ensured for $k = 1$ and hence the Fang and Wang choice $a_k = 1/k$ will not attain the RQMC rate for polynomials when $d \geqslant 2$.

We can extend Theorem 10 to all $f \in C^d$ via Theorem 6, but only for a_k larger than the Fang and Wang choice. We require such large a_j because the generalized Hölder inequality is conservative in this setting.

THEOREM 11. *Let* $f \in C^d(A_d)$*, and define* $x_j = \tau_j(\boldsymbol{u}) = \prod_{k=j}^d u_j^{a_j}$ *for* $a_j > 0$*. Let* $\tilde{f}(\boldsymbol{u}) = f(\tau(\boldsymbol{u}))J(\boldsymbol{u})$ *for the Jacobian* $\prod_{j=1}^d a_j u_j^{j a_j - 1}$*. If* $a_1 > 3/2$ *and* $a_j \geqslant 1$ *for* $2 \leqslant j \leqslant d$*, then* $\partial^v \tilde{f} \in L^2[0,1]^d$ *for all* $v \subseteq 1{:}d$*.*

Proof. Define $\mathcal{X} = A_d \times [0, A] \subset \mathbb{R}^{d+1}$, where $A = \prod_{j=1}^d a_j$ and $\tau_{d+1}(\boldsymbol{u}) = J(\boldsymbol{u})$. Then $\tilde{f}(\tau_1(\boldsymbol{u}), \ldots, \tau_{d+1}(\boldsymbol{u})) = f(\tau_1(\boldsymbol{u}), \ldots, \tau_d(\boldsymbol{u}))\tau_{d+1}(\boldsymbol{u}) \in C^{d+1}(\mathcal{X})$.

For $j \leqslant d$ and $v \subseteq 1{:}d$ we have $\partial^v \tau_j = 0$ unless $v \subseteq j{:}d$, and if $v \subseteq j{:}d$, then $\partial^v \tau_j = \prod_{\ell \in v} a_\ell u_\ell^{a_\ell - 1} \times \prod_{\ell \in j:d-v} u_\ell^{a_\ell}$. Under the conditions of this theorem every $\tau_j \in L^\infty[0,1]^d$. Next we can directly find that under the given conditions $\tau_{d+1} = J \in L^2[0,1]^d$. Then we have $\partial^v \tilde{f} \in L^2$ by Theorem 6. ∎

In MC sampling, the effect of nonuniform importance sampling is sometimes measured via an effective sample size. See Kong, Liu, and Wong (1994). For the Jacobian above the effective sample size is the nominal one multiplied by $(\int J(\boldsymbol{u})\,\mathrm{d}\boldsymbol{u})^2 / \int J(\boldsymbol{u})^2\,\mathrm{d}\boldsymbol{u}$. If we take $a_j = 3/(2j)$ this factor becomes $(8/9)^d$, which corresponds to a mild exponential decay in effectiveness for MC sampling. There seems to be as yet no good measure of effective sample size for RQMC.

Acknowledgment. We thank two anonymous reviewers for very helpful remarks.

REFERENCES

C. Aistleitner and J. Dick (2013), *Low-discrepancy point sets for non-uniform measures*, Acta Arith., 163, pp. 345–369.

C. Aistleitner and J. Dick (2015), *Functions of bounded variation, signed measures, and a general Koksma-Hlawka inequality*, Acta Arith., 167, pp. 143–171.

S. Asmussen and P. Glynn (2007), *Stochastic Simulation: Algorithms and Analysis*, Springer, New York.

K. Basu and A. B. Owen (2015a), *Low discrepancy constructions in the triangle*, SIAM J. Numer. Anal., 53, pp. 743–761.

K. Basu and A. B. Owen (2015b), *Scrambled geometric net integration over general product spaces*, Found. Comput. Math., 16, pp. 1–30.

V. I. Bogachev (2007), *Measure Theory*, Springer, Heidelberg.

L. Brandolini, L. Colzani, G. Gigante, and G. Travaglini (2013), *A Koksma–Hlawka inequality for simplices*, in Trends in Harmonic Analysis, Springer, New York, pp. 33–46.

J. S. Brauchart and J. Dick (2012), *Quasi-Monte Carlo rules for numerical integration over the unit sphere* $\mathbb{S}^2$, Numer. Math., 121, pp. 473–502.

J. S. Brauchart, E. B. Saff, I. H. Sloan, and R. S. Womersley (2014), *QMC designs: optimal order quasi-Monte Carlo integration schemes on the sphere*, Math. comp., 83, pp. 2821–2851.

M. Cambou, M. Hofert, and C. Lemieux (2015), *A Primer on Quasi-Random Numbers for Copula Models*, Technical report, arXiv: 1508.03483.

P. Chelson (1976), *Quasi-Random Techniques for Monte Carlo Methods*, Ph.D. thesis, Claremont Graduate School, Claremont, Calif.

G. Constantine and T. Savits (1996), *A multivariate Faa di Bruno formula with applications*, Trans. Amer. Math. Soc., 348, pp. 503–520.

J. Dick (2009), *On quasi-Monte Carlo rules achieving higher order convergence*, in Monte Carlo and Quasi-Monte Carlo Methods, P. L'Ecuyer and A. B. Owen, eds., Springer, Heidelberg.

J. Dick (2011), *Higher order scrambled digital nets achieve the optimal rate of the root mean square error for smooth integrands*, Ann. Statist., 39, pp. 1372–1398.

J. Dick and F. Pillichshammer (2010), *Digital Sequences, Discrepancy and Quasi-Monte Carlo Integration*, Cambridge University Press, Cambridge, UK.

C. F. Faa di Bruno (1855), *Sullo sviluppo delle funzioni*, Ann. Sci. Mat. Fisiche, 6, pp. 479–480.

K.-T. Fang and Y. Wang (1993), *Number-Theoretic Methods in Statistics*, Monogr. Statist. Appl. Probab. 51, CRC Press, Boca Raton, FL.

M. Fréchet (1910), *Extension au cas des intégrales multiples d'une définition de l'intégrale due à Stieltjes*, Nouv. Ann. Math., 10, pp. 241–256.

J. F. Geweke (1989), *Bayesian inference in econometric models using Monte Carlo integration*, Econometrica, 57, pp. 1317–1340.

R. L. Graham, D. E. Knuth, and O. Patashnik (1989), *Concrete Mathematics*, Addison-Wesley, Reading, MA.

E. Hlawka and R. Mück (1972), *Über eine Transformation von gleichverteilten Folgen*, II, Computing, 9, pp. 127–138.

M. Josephy (1981), *Composing functions of bounded variation*, Proc. Amer. Math. Soc., 83, pp. 354–356.

A. Kong, J. S. Liu, and W. H. Wong (1994), *Sequential imputations and Bayesian missing data problems*, J. Amer. Statist. Assoc., 89, pp. 278–288.

J. Matoušek (1998), *Geometric Discrepancy: An Illustrated Guide*, Springer-Verlag, Heidelberg.

A. B. Owen (1995), *Randomly permuted (t, m, s)-nets and (t, s)-sequences*, in Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing, H. Niederreiter and P. J.-S. Shiue, eds., Springer-Verlag, New York, pp. 299–317.

A. B. Owen (2005), *Multidimensional variation for quasi-Monte Carlo*, in International Conference on Statistics in Honour of Professor Kai-Tai Fang's 65th Birthday, J. Fan and G. Li, eds.

T. Pillards and R. Cools (2005), *Transforming low-discrepancy sequences from a cube to a simplex*, J. Comput. Appl. Math., 174, pp. 29–42.

M. Rosenblatt (1952), *Remarks on a multivariate transformation*, Ann. Math. Statist., pp. 470–472.

Y. A. Shreider (1966), *The Monte Carlo Method: The Method of Statistical Trials*, Pergamon Press, Oxford, UK.