www.oeaw.ac.at

Johann Radon Institute for
Computational and Applied Mathematics
Austrian Academy of Sciences (ÖAW)

On the dependence structure of scrambled (t, m, s)-nets

J. Wiart, C. Lemieux

RICAM-Report 2019-11

www.ricam.oeaw.ac.at

On the dependence structure of scrambled (t, m, s)-nets

Jaspar Wiart[1,2] and Christiane Lemieux[1]

[1]Department of Statistics and Actuarial Science
University of Waterloo,
200 University Avenue West
Ontario, Canada, N2L 3G1
email: clemieux@uwaterloo.ca
[2]Johann Radon Institute for Computational and Applied Mathematics (RICAM)
Austrian Academy of Sciences
Altenbergerstr. 69
4040 Linz, Austria
jaspar.wiart@ricam.oeaw.ac.at

Abstract

In this paper, we study the dependence structure of scrambled (t, m, s)-nets. We show that they have a negative lower/upper orthant dependence structure if and only if $t = 0$. This study allows us to gain a deeper understanding about the classes of functions for which the variance of estimators based on scrambled $(0, m, s)$-nets can be proved to be no larger than that of a Monte Carlo estimator.

1 Introduction

Quasi-Monte Carlo methods rely on low-discrepancy point sets and sequences to construct estimates for multi-dimensional integrals over the unit hypercube. In this context, the notion of discrepancy refers to the distance between the uniform distribution and the empirical distribution induced by a point set. This measure of non-uniformity is particularly suitable for deterministic point sets, for which a number of results exist that provide asymptotic results on the discrepancy of various constructions, including digital (t, m, s)-nets [3, 9].

In recent years, the use of randomized quasi-Monte Carlo methods has gained in popularity. By introducing randomness in a low-discrepancy point set, one gains not only access to probabilistic error estimates, but also in some cases to an improvement in the uniformity of the point set. In particular, the scrambled digital nets introduced by Owen in 1995 [10] have been used in different applications in practice. A number of results studying the variance of the corresponding estimators have been proved: see, for example, [5, 14]. For smooth enough functions, results in show a much better convergence rate for the variance of these scrambled net estimators than the Monte Carlo equivalent [12]. Other results give bounds holding for all square-integrable functions, where the scrambled net variance is shown to be no larger than a constant (larger than one, and possibly quite large depending on the net) times the Monte Carlo variance [11, 13].

In [7], a new approach to study scrambled $(0, m, s)$-nets was introduced. It is based on the concept of negative lower/upper orthant dependence, and how it can be used to study the covariance term that differentiates the variance of Monte Carlo sampling-based estimators from that of scrambled $(0, m, s)$-nets. To study this covariance term, a new representation result was used. It is based on multivariate integration by parts, which allows to decompose the covariance term in a part that assesses the underlying point set—via its dependence structure—and a part that depends on the function. It is worth noting that a larger class of functions than those of bounded variation in the sense of Hardy and Krause [9] can be studied via this decomposition. In the same paper, it was proved that two-dimensional scrambled $(0, m, 2)$-nets have a variance no larger than a Monte Carlo estimator for functions that are monotone in each variable. This result was obtained by first establishing that scrambled $(0, m, 2)$-nets are negatively lower orthant dependent.

In this paper, we significantly generalize these results as follows. We prove that scrambled (t, m, s)-nets are negative lower orthant dependent (NLOD) and negative upper orthant dependent (NUOD) if and only if $t = 0$. This dependence property turns out to be equivalent to a certain continuous linear functional on $\ell^1(\mathbb{N})$ restricted to a particular subset having norm one. We prove this is the case by way of a convexity argument. This result also yields a new quality measure for digital nets that can be seen as a measure of how far from being NLOD/NUOD a net is. This measure also captures more information about the equidistribution properties of the net than the well-known quality parameter t [9] does. Then, having shown the existence of NLOD and NUOD sampling schemes for dimensions higher than two, we explore what are the functions for which the variance of scrambled $(0, m, s)$-nets in base b is no larger than that of a Monte Carlo estimator with b^m points. One such class of functions are those that are quasi-monotone and possibly unbounded on the boundary.

This paper is organized as follows. In Section 2 we review some background information on scrambled nets and dependence concepts, and prove a few key properties of scrambled nets that are relevant when studying their dependence structure. In Section 3 we study the joint probability density function (pdf) of pairs of distinct points in a scrambled net, which allows us to prove, in Section 4, that a scrambled (t, m, s)-net is an NLOD sampling scheme if and only if $t = 0$. In Section 5 we use the properties of this pdf to explore the class of functions for which scrambled nets reduce the variance compared to Monte Carlo. Concluding comments and ideas for future work are presented in Section 6.

2 Preliminaries

We start by recalling key properties of scrambled nets, as presented in [7].

A *digital net in base b* (for b prime) [3, 9] is a point set $P_n = \{\mathbf{V}_1, \ldots, \mathbf{V}_n\} \subseteq [0, 1]^s$ with $n = b^m$ that is constructed via s generating matrices $C_1, \ldots, C_s$ of size $m \times m$ with entries in $\mathbb{F}_b$, in the following way: for $0 \leq i < b^m$ we write $i = \sum_{r=0}^{m-1} i_r b^r$, then $\mathbf{V}_i = (V_{i,1}, \ldots, V_{i,s})$ is obtained as $V_{i,\ell} = \sum_{r=1}^{m} V_{i,\ell,r} b^{-(r+1)}$, and $V_{i,\ell,r} = \sum_{p=1}^{m} C_{\ell,r,p} i_{p-1}$, where $C_{\ell,r,p}$ is the element on the rth row and pth column of C_ℓ.

To assess the uniformity of the net, the concept of $(q_1, \ldots, q_s)$-equidistribution is used. More precisely, we say that P_n with $n = b^m$ is *$(q_1, \ldots, q_s)$-equidistributed in base b* if every *elementary $(q_1, \ldots, q_s)-$interval* of the form

$$\prod_{\ell=1}^{s} \left[\frac{a_\ell}{b^{q_\ell}}, \frac{a_\ell + 1}{b^{q_\ell}} \right)$$

for $0 \leq a_\ell < b^{q_\ell}$ contains exactly $b^{m-q_1-\cdots-q_s}$ points from P_n, assuming $m \geq q_1 + \ldots + q_s$. We say that a digital net in base b has a *quality parameter t* if P_n is $(q_1, \ldots, q_s)$-equidistributed for all s-dimensional vectors of non-negative integers $(q_1, \ldots, q_s)$ such that $q_1 + \ldots + q_s \leq m - t$. We then refer to P_n as a digital (t, m, s)-net in base b. So the lower is t, the more uniform P_n is [9]. The construction proposed by Faure in [4] provides $(0, m, s)$-nets in base $b \geq s$. The widely used Sobol' sequences [15] provide (t, m, s)-nets in base 2 with $t = 0$ when $s = 2$ and $t > 0$ otherwise. Information on newer constructions can be found in [2, 3].

A *scrambled digital net in base b* is a randomized point set $\tilde{P}_n = \{\mathbf{U}_1, \ldots, \mathbf{U}_n\}$ with $n = b^m$ which we assume has the following two properties [8, 14]. Let $U_{i,\ell} = \sum_{r=1}^{\infty} U_{i,\ell,r} b^{-(r+1)}$, that is, $U_{i,\ell,r}$ represents the rth digit in the base b expansion of the ℓth coordinate of the ith point $\mathbf{U}_i$. Then we must have:

1. Each $\mathbf{U}_i \sim U([0, 1]^s)$;
2. For two distinct points $\mathbf{U}_i, \mathbf{U}_j$ and in each dimension ℓ, if the two deterministic points $\mathbf{V}_i, \mathbf{V}_j$ (before scrambling is applied) have the same first r digits and differ on the $(r + 1)$th digit, then (i) the scrambled

points $(U_{i,\ell}, U_{j,\ell})$ also have the same first r digits, and the pair $(U_{i,\ell,r+1}, U_{j,\ell,r+1})$ is uniformly distributed over $\{(k_1, k_2), 0 \leq k_1 \neq k_2 < b\}$; (ii) the pairs $(U_{i,\ell,v}, U_{j,\ell,v})$ for $v > r+1$ are independent and uniformly distributed over $\{(k_1, k_2), 0 \leq k_1, k_2 < b\}$.

One way to scramble a digital net P_n so that the scrambled net $\tilde{P}_n$ has these properties is to multiply from the left each generating matrix C_ℓ by a randomly chosen NLT matrix S_ℓ (i.e., with entries on the diagonal uniformly chosen in $\{1, \ldots, b-1\}$, and entries below the diagonal uniformly chosen in $\{0, \ldots, b-1\}$, with the other entries set to 0), and then add a digital shift in base b [6]. Also note that if P_n is a (t, m, s)-net then the scrambled net $\tilde{P}_n$ is a (t, m, s)-net as well. We refer the reader to [14] for further information on scrambling methods for digital nets.

Next, we introduce dependence concepts from [7] that will be used throughout this paper.

A vector $\mathbf{X} = (X_1, \ldots, X_s)$ of random variables is NLOD if

$$P(X_1 \leq x_1, \ldots, X_s \leq x_s) \leq \prod_{\ell=1}^{s} P(X_\ell \leq x_\ell),$$

and it is NUOD if

$$P(X_1 \geq x_1, \ldots, X_s \geq x_s) \leq \prod_{\ell=1}^{s} P(X_\ell \geq x_\ell).$$

Note that when $s = 2$, the NLOD and NUOD properties are equivalent (and both correspond to NQD), but it is not necessarily the case when $s \geq 3$.

Consider a sampling scheme $\tilde{P}_n = \{\mathbf{U}_1, \ldots, \mathbf{U}_n\}$ designed to construct an unbiased estimator of the form

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} f(\mathbf{U}_i)$$

for

$$I(f) = \int_{[0,1)^s} f(\mathbf{x}) d\mathbf{x},$$

where we assume each $\mathbf{U}_i$ is uniformly distributed over $[0,1]^s$ with a possible dependence structure between the $\mathbf{U}_i$'s. To assess this dependence, a key quantity of interest is

$$H(\mathbf{x}, \mathbf{y}; \tilde{P}_n) := \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j>i} P(\mathbf{U}_i \leq \mathbf{x}, \mathbf{U}_j \leq \mathbf{y}). \tag{1}$$

We can think of $H(\mathbf{x}, \mathbf{y}; \tilde{P}_n)$ as the joint distribution of a pair of (distinct) points $(\mathbf{U}_I, \mathbf{U}_J)$ randomly chosen in $\tilde{P}_n$. (Here, we use capital letters for the indices I and J to make it clear the points are randomly selected.)

If $H(\mathbf{x}, \mathbf{y}; \tilde{P}_n) \leq \prod_{\ell=1}^{s} x_\ell y_\ell$ for all $0 \leq x_\ell, y_\ell \leq 1$, $\ell = 1, \ldots, s$, then we say $\tilde{P}_n$ is an *NLOD sampling scheme*.

We are also interested in the quantity

$$T(\mathbf{x}, \mathbf{y}; \tilde{P}_n) := \frac{2}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j>i} P(\mathbf{U}_i \geq \mathbf{x}, \mathbf{U}_j \geq \mathbf{y}), \tag{2}$$

and say that $\tilde{P}_n$ is an *NUOD sampling scheme* if $T(\mathbf{x}, \mathbf{y}; \tilde{P}_n) \geq \prod_{\ell=1}^{s} (1 - x_\ell)(1 - y_\ell)$ for all $0 \leq x_\ell, y_\ell \leq 1$, $\ell = 1, \ldots, s$.

In [7], this quantity arises in the analysis of $\mathrm{Cov}(f(\mathbf{U}_I), f(\mathbf{U}_J))$, the covariance term that differentiates the variance of $\hat{\mu}_n$–when $\tilde{P}_n$ is a dependent sampling scheme–from that of a Monte Carlo estimator with the same number of points n. More precisely, $\mathrm{Cov}(f(\mathbf{U}_I), f(\mathbf{U}_J))$ is such that

$$\mathrm{Var}(\hat{\mu}_n) = \frac{\sigma^2}{n} + \frac{n-1}{n} \mathrm{Cov}(f(\mathbf{U}_I), f(\mathbf{U}_J)),$$

where $\sigma^2 = \mathrm{Var}(f(\mathbf{U}))$. In the present work, rather than using the expression developed in [7] to write this covariance in terms of the survival function $T(\mathbf{x}, \mathbf{y}; \tilde{P}_n)$, we instead work with the direct representation

$$\sigma_{I,J} := \mathrm{Cov}(f(\mathbf{U}_I), f(\mathbf{U}_J)) = \int_{[0,1]^{2s}} f(\mathbf{x})f(\mathbf{y})\psi(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} - \int_{[0,1]^{2s}} f(\mathbf{x})f(\mathbf{y})d\mathbf{x}d\mathbf{y}.$$

where $\psi(\mathbf{x}, \mathbf{y})$ is the joint pdf of $(\mathbf{U}_I, \mathbf{U}_J)$ evaluated at $(\mathbf{x}, \mathbf{y})$. In particular, this means we can also write

$$H(\mathbf{x}, \mathbf{y}; \tilde{P}_n) = \int_{R(\mathbf{x},\mathbf{y})} \psi(\mathbf{u}, \mathbf{v})d\mathbf{u}d\mathbf{v},$$

where $R(\mathbf{x}, \mathbf{y}) = \{(\mathbf{u}, \mathbf{v}) \in [0, 1]^{2s} : u_j \le x_j, v_j \le y_j, j = 1, \ldots, s\}$.

The joint pdf $\psi(\mathbf{x}, \mathbf{y})$ corresponding to a scrambled digital (t, m, s)-net in base b is the topic of Section 3. The rest of this section develops tools to analyze this joint pdf.

Remark 2.1. In the previous two equations we used different letters within the integrals. This is to emphasis the function of interest. In the first of the two equations we were interested in $f(\mathbf{x})$, $f(\mathbf{y})$, and $\psi(\mathbf{x}, \mathbf{y})$, while in the second we were interested in $H(\mathbf{x}, \mathbf{y}; \tilde{P}_n)$.

Definition 2.2. For $x, y \in [0, 1]$, let $\gamma_b(x, y)$ be the exact number of initial digits shared by x and y in their base b expansion, i.e. the smallest number i such that

$$\lfloor b^i x\rfloor = \lfloor b^i y\rfloor \quad \text{but} \quad \lfloor b^{i+1}x\rfloor \ne \lfloor b^{i+1}y\rfloor.$$

For $\mathbf{x}, \mathbf{y} \in [0, 1]^s$, we define

$$\boldsymbol{\gamma}_b^s(\mathbf{x}, \mathbf{y}) = (\gamma_b(x_1, y_1), \ldots, \gamma_b(x_s, y_s)) \text{ and } \gamma_b^s(\mathbf{x}, \mathbf{y}) = \sum_{j=1}^{s} \gamma_b(x_j, y_j).$$

Given $\boldsymbol{i}, \boldsymbol{k} \in \mathbb{N}^s$, we say that $\boldsymbol{k} \le \boldsymbol{i}$ if $k_j \le i_j$ for all $j = 1, \ldots, s$. To simplify notation, whenever we use i or k in the same formula as $\boldsymbol{i}$ or $\boldsymbol{k}$, they will denote the sum of the coordinates: $i = i_1 + \ldots + i_s$ and $k = k_1 + \ldots + k_s$. For each $\boldsymbol{k}, \boldsymbol{i} \in (\mathbb{N} \cup \{\infty\})^s$ we define $C^s_{\boldsymbol{k}}, D^s_{\boldsymbol{i}} \subseteq [0, 1]^{2s}$ to be the subsets

$$\begin{aligned} C^s_{\boldsymbol{k}} &= \{(\mathbf{x}, \mathbf{y}) \in [0, 1]^{2s} : \boldsymbol{k} \le \boldsymbol{\gamma}_b^s(\mathbf{x}, \mathbf{y})\} \qquad \text{and} \\ D^s_{\boldsymbol{i}} &= \{(\mathbf{x}, \mathbf{y}) \in [0, 1]^{2s} : \boldsymbol{\gamma}_b^s(\mathbf{x}, \mathbf{y}) = \boldsymbol{i}\}. \end{aligned}$$

In the special case $s = 1$ we denote these sets by C_k and D_i respectively. It is clear that $C^s_{\boldsymbol{k}} = \cup_{\boldsymbol{k}\le\boldsymbol{i}} D^s_{\boldsymbol{i}}$ and that the $D^s_{\boldsymbol{i}}$'s partition $[0, 1]^{2s}$. One can easily verify that

$$C_k = \bigcup_{a=0}^{b^k-1} \left[\frac{a}{b^k}, \frac{a+1}{b^k}\right] \text{ and } C^s_{\boldsymbol{k}} = \prod_{j=1}^{s} C_{k_j}$$

from which it follows that $\mathrm{Vol}(C_k) = b^{-k}$ and $\mathrm{Vol}(C^s_{\boldsymbol{k}}) = b^{-k}$. Finally, since $D_i = C_i \backslash C_{i+1}$ and since $D^s_{\boldsymbol{i}} = \prod_{j=1}^{s} D_{i_j}$ we have

$$\mathrm{Vol}(D^s_{\boldsymbol{i}}) = \frac{(b-1)^s}{b^{s+i}}$$

because $\mathrm{Vol}(D_i) = (b - 1)/b^{i+1}$.

In Theorem 3.6 we see that the joint pdf of a scrambled digital net in base b is constant on each $D^s_{\boldsymbol{i}}$ (and is zero on those $D^s_{\boldsymbol{i}}$ for which $i = \infty$). To prove the NLOD/NUOD property one must show that

$$\int_{R(\mathbf{x},\mathbf{y})} \psi(\mathbf{u}, \mathbf{v})d\mathbf{u}d\mathbf{v} \le \mathbf{x}\mathbf{y} = \prod_{j=1}^{s} x_j y_j$$

holds for all $\mathbf{x}, \mathbf{y} \in [0,1]^{2s}$. This integral may be written as

$$\int_{R(\mathbf{x},\mathbf{y})} \psi(\mathbf{u},\mathbf{v}) d\mathbf{u} d\mathbf{v} = \sum_{i=0}^{\infty} V_{\boldsymbol{i}}^{s}(\mathbf{x},\mathbf{y}) \psi_{\boldsymbol{i}},$$

where

$$V_{\boldsymbol{i}}^{s}(\mathbf{x},\mathbf{y}) = \int_{R(\mathbf{x},\mathbf{y})} 1_{D_{\boldsymbol{i}}^{s}}(\mathbf{u},\mathbf{v}) d\mathbf{u} d\mathbf{v} = \mathrm{Vol}(R(\mathbf{x},\mathbf{y}) \cap D_{\boldsymbol{i}}^{s})$$

and $\psi_{\boldsymbol{i}}$ is the value of ψ on $D_{\boldsymbol{i}}^{s}$. As before, in the special case $s = 1$ we use the notation $V_i(x,y)$ or simply V_i when x and y are fixed. We will use the fact that

$$V_{\boldsymbol{i}}^{s}(\mathbf{x},\mathbf{y}) = \int_{D_{i_1}} \cdots \int_{D_{i_s}} \prod_{j=1}^{s} 1_{R(x_j,y_j)}(u_j,v_j) du_s dv_s \ldots du_1 dv_1 = \prod_{j=1}^{s} V_{i_j}(x_j,y_j)$$

together with the following lemma (stating a result that appears in the proof of [7, Proposition 7]) to simplify the calculation of $V_{\boldsymbol{i}}(\mathbf{x},\mathbf{y})$.

Lemma 2.3. *Let $V_i = V_i(x,y)$ where $x, y \in [0,1]$. Then we have*

$$V_i = \begin{cases} \frac{b-1}{b}\frac{\min(x,y)}{b^i} & \text{if } \gamma_b(x,y) < i, \\ xy - k_i(x+y-k_i-b^{-i}) - \frac{\min(x,y)}{b^{i+1}} & \text{if } \gamma_b(x,y) = i, \\ k_{i+1}(x+y-k_{i+1}-b^{-i-1}) - k_i(x+y-k_i-b^{-i}) & \text{if } \gamma_b(x,y) > i, \end{cases}$$

where $k_i = \lfloor b^i \min(x,y) \rfloor b^{-i}$.

We only need the above formula to prove the following technical lemma, which gives us a critical relation between V_i and V_{i+1}. While its proof is rather tedious, it is not hard, we simply use the above formula and carefully work through the cases.

Lemma 2.4. *Let $x, y \in [0,1]$ be given and $V_i = V_i(x,y)$ be defined as above. Then we have $bV_i - V_{i-1} \geq 0$ for all $i \geq 1$.*

Proof. Let

$$x = \sum_{k=0}^{\infty} \frac{x_k}{b^k}, \text{ and } y = \sum_{k=0}^{\infty} \frac{y_k}{b^k}$$

be the base b digital expansion of x and y chosen so that only finitely many digits are non-zero whenever possible. When $\gamma_b(x,y) \geq i$ let

$$k_i = \sum_{k=0}^{i} \frac{x_k}{b^k} = \sum_{k=0}^{i} \frac{y_k}{b^k},$$

$r_x^i = x - k_i$, and $r_y^i = y - k_i$. Without loss of generality we assume $x \leq y$. There are four cases.

Case 1: $(\gamma_b(x,y) < i-1)$

In this case

$$bV_i - V_{i-1} = b\frac{x}{b^i} - \frac{x}{b^{i-1}} = 0.$$

Case 2: $(\gamma_b(x,y) = i-1)$

In this case, $bV_i - V_{i-1}$ becomes

$$\begin{aligned} &\frac{x}{b^{i-1}} - xy + k_{i-1}\Big(x + y - k_{i-1} - \frac{1}{b^{i-1}}\Big) = \\ &= \frac{k_{i-1} + r_x^{i-1}}{b^{i-1}} - (k_{i-1} + r_x^{i-1})(k_{i-1} + r_y^{i-1}) + k_{i-1}\Big(k_{i-1} + r_x^{i-1} + r_y^{i-1} - \frac{1}{b^{i-1}}\Big) \\ &= \frac{k_{i-1} + r_x^{i-1}}{b^{i-1}} - r_x^{i-1} r_y^{i-1} - \frac{k_{i-1}}{b^{i-1}} \geq \frac{k_{i-1} + r_x^{i-1}}{b^{i-1}} - \frac{r_x^{i-1}}{b^{i-1}} - \frac{k_{i-1}}{b^{i-1}} = 0 \end{aligned}$$

because $r_y^{i-1} \le 1/b^{i-1}$.

Case 3: $(\gamma_b(x,y)=i)$

We use the calculation in case 2 and the identities $r_x^{i-1} = x_i/b^i + r_x^i$ and $r_y^{i-1} = x_i/b^i + r_y^i$ to simplify $bV_i - V_{i-1}$:

$$\begin{aligned}
&b\Big(xy - \frac{x}{b^{i+1}} - k_i\Big(x+y-k_i-\frac{1}{b^i}\Big)\Big)\\
&-\Big(k_i\Big(x+y-k_i-\frac{1}{b^i}\Big) - k_{i-1}\Big(x+y-k_{i-1}-\frac{1}{b^{i-1}}\Big)\Big)\\
&=(b+1)\Big(xy-\frac{x}{b^i}-k_i\Big(x+y-k_i-\frac{1}{b^i}\Big)\Big)\\
&-\Big(xy-\frac{x}{b^{i-1}}-k_{i-1}\Big(x+y-k_{i-1}-\frac{1}{b^{i-1}}\Big)\Big)\\
&=(b+1)\Big(r_x^i r_y^i - \frac{r_x^i}{b^i}\Big) - \Big(r_x^{i-1}r_y^{i-1} - \frac{r_x^{i-1}}{b^{i-1}}\Big)\\
&= b r_x^i r_y^i - \frac{r_x^i}{b^i} - \frac{x_i^2}{b^{2i}} - \frac{x_i(r_x^i + r_y^i)}{b^i} + \frac{x_i}{b^{2i-1}}.
\end{aligned}$$

Multiply by b^i to get

$$b^{i+1} r_x^i r_y^i - r_x^i - \frac{x_i^2}{b^i} - x_i r_x^i - x_i r_y^i + \frac{x_i}{b^{i-1}}$$

which will be shown to be non-negative. Note that by assumption $x<y$ and since their base b expansion differ for the first time at the $(i+1)^{\text{th}}$ digit we always have $x_{i+1} < y_{i+1}$.

Case 3a: $(x_i \le x_{i+1} < y_{i+1})$

The assumption implies to $0 \le b^{i+1} r_x^i - x_i$ and $(x_i+1)/b^{i+1} \le r_y^i$. We estimate

$$\begin{aligned}
&(b^{i+1}r_x^i - x_i)r_y^i - r_x^i - \frac{x_i^2}{b^i} - x_i r_x^i + \frac{x_i}{b^{i-1}}\\
&\ge (b^{i+1}r_x^i - x_i)\frac{x_i+1}{b^{i+1}} - r_x^i - \frac{x_i^2}{b^i} - x_i r_x^i + \frac{x_i}{b^{i-1}}\\
&= \frac{x_i}{b^{i+1}}(b^2 - (b+1)x_i - 1) \ge \frac{x_i}{b^{i+1}}(b^2-(b+1)(b-1)-1) = 0.
\end{aligned}$$

Case 3b: $(x_{i+1} < x_i < y_{i+1})$

The assumption implies $r_x^i \le x_i/b^{i+1}$ and $(b^{i+1}r_y^i - x_i - 1) \le 0$. We estimate

$$\begin{aligned}
&\frac{x_i}{b^{i-1}} - \frac{x_i^2}{b^i} - x_i r_y^i + (b^{i+1}r_y^i - x_i - 1)r_x^i\\
&\ge \frac{x_i}{b^{i-1}} - \frac{x_i^2}{b^i} - x_i r_y^i + (b^{i+1}r_y^i - x_i - 1)\frac{x_i}{b^{i+1}}\\
&= \frac{x_i}{b^{i+1}}(b^2-(b+1)x_i-1)) \ge \frac{x_i}{b^{i+1}}(b^2-(b+1)(b-1)-1)) = 0.
\end{aligned}$$

Case 3c: $(x_{i+1} < y_{i+1} \le x_i)$

The assumption is implies $0 \le (b^{i+1}r_y^i - x_i - 1)$. We estimate

$$\begin{aligned}
&b^{i+1}r_x^i r_y^i - r_x^i - \frac{x_i^2}{b^i} - x_i r_x^i - x_i r_y^i + \frac{x_i}{b^{i-1}} =\\
&= \frac{x_i}{b^{i-1}} - \frac{x_i^2}{b^i} - x_i r_y^i + (b^{i+1}r_y^i - x_i - 1)r_x^i\\
&\ge \frac{x_i}{b^{i-1}} - \frac{x_i^2}{b^i} - x_i r_y^i \ge x_i\Big(\frac{1}{b^{i-1}} - \frac{b-1}{b^i} - \frac{1}{b^i}\Big) = 0.
\end{aligned}$$

Case 4: $(\gamma_b(x,y) > i)$

In this case we need to show that

$$bk_{i+1}(x+y-k_{i+1}-1/b^{i+1}) - (b+1)k_i(x+y-k_i-1/b^i) + k_{i-1}(x+y-k_{i-1}-1/b^{i-1}) \tag{3}$$

is greater than or equal to zero. Using the identities $k_{i+1} = k_{i-1} + x_i/b^i + x_{i+1}/b^{i+1}$, $k_i = k_{i-1} + x_i/b^i$, $x = k_{i-1} + x_i/b^i + x_{i+1}/b^{i+1} + r_x^{i+1}$, and $y = k + x_i/b^i + x_{i+1}/b^{i+1} + r_y^{i+1}$ write

$$\begin{aligned}
&k_{i+1}\Big(x+y-k_{i+1}-\frac{1}{b^{i+1}}\Big) = \\
&= \Big(k_{i-1}+\frac{x_i}{b^i}+\frac{x_{i+1}}{b^{i+1}}\Big)\Big(k_{i-1}+\frac{x_i}{b^i}+\frac{x_{i+1}}{b^{i+1}}+r_x^{i+1}+r_y^{i+1}-\frac{1}{b^{i+1}}\Big) \\
&= k_{i-1}^2+\frac{x_i^2}{b^{2i}}+\frac{x_{i+1}^2}{b^{2i+2}}+\frac{2k_{i-1}x_i}{b^i}+\frac{2k_{i-1}x_{i+1}}{b^{i+1}}+\frac{2x_ix_{i+1}}{b^{2i+1}}+k_{i-1}(r_x^{i+1}+r_y^{i+1}) \\
&+\frac{x_i(r_x^{i+1}+r_y^{i+1})}{b^i}+\frac{x_{i+1}(r_x^{i+1}+r_y^{i+1})}{b^{i+1}}-\frac{k_{i-1}}{b^{i+1}}-\frac{x_i}{b^{2i+1}}-\frac{x_{i+1}}{b^{2i+2}},
\end{aligned}$$

and

$$\begin{aligned}
&k_i\Big(x+y-k_i-\frac{1}{b^i}\Big) = \\
&= \Big(k_{i-1}+\frac{x_i}{b^i}\Big)\Big(k_{i-1}+\frac{x_i}{b^i}+\frac{2x_{i+1}}{b^{i+1}}+r_x^{i+1}+r_y^{i+1}-\frac{1}{b^i}\Big) \\
&= k_{i-1}^2+\frac{x_i^2}{b^{2i}}+\frac{2k_{i-1}x_i}{b^i}+\frac{2k_{i-1}x_{i+1}}{b^{i+1}}+\frac{2x_ix_{i+1}}{b^{2i+1}} \\
&+k_{i-1}(r_x^{i+1}+r_y^{i+1})+\frac{x_i(r_x^{i+1}+r_y^{i+1})}{b^i}-\frac{k_{i-1}}{b^i}-\frac{x_i}{b^{2i}},
\end{aligned}$$

and

$$\begin{aligned}
&k_{i-1}\Big(x+y-k_{i-1}-\frac{1}{b^{i-1}}\Big) = \\
&= k_{i-1}\Big(k_{i-1}+\frac{2x_i}{b^i}+\frac{2x_{i+1}}{b^{i+1}}+r_x^{i+1}+r_y^{i+1}-\frac{1}{b^{i-1}}\Big) \\
&= k_{i-1}^2+\frac{2k_{i-1}x_i}{b^i}+\frac{2k_{i-1}x_{i+1}}{b^{i+1}}+k_{i-1}(r_x^{i+1}+r_y^{i+1})-\frac{k_{i-1}}{b^{i-1}}.
\end{aligned}$$

Now substituting into (3) and combining like terms we get

$$\begin{aligned}
bV_{i+1}-V_i &= 0\Big(k_{i-1}^2+\frac{2k_{i-1}x_i}{b^i}+\frac{2k_{i-1}x_{i+1}}{b^{i+1}}+k_{i-1}(r_x^{i+1}+r_y^{i+1})\Big) \\
&-\Big(\frac{x_i^2}{b^{2i}}+\frac{2x_ix_{i+1}}{b^{2i+1}}+\frac{x_i(r_x^{i+1}+r_y^{i+1})}{b^i}\Big)+b\Big(\frac{x_{i+1}^2}{b^{2i+2}}+\frac{x_{i+1}(r_x^{i+1}+r_y^{i+1})}{b^{i+1}}\Big) \\
&-k_{i-1}\Big(\frac{b}{b^{i+1}}-\frac{b+1}{b^i}+\frac{1}{b^{i-1}}\Big)-x_i\Big(\frac{b}{b^{2i+1}}-\frac{b+1}{b^{2i}}\Big)-\frac{bx_{i+1}}{b^{2i+2}} \\
&= \frac{x_{i+1}^2}{b^{2i+1}}+\frac{x_i}{b^{2i-1}}-\frac{x_i^2}{b^{2i}}-\frac{2x_ix_{i+1}}{b^{2i+1}}-\frac{x_{i+1}}{b^{2i+1}}+\frac{(x_{i+1}-x_i)(r_x^{i+1}+r_y^{i+1})}{b^i}.
\end{aligned}$$

By multiplying the above by b^{2i+1} we see that to finish the proof we need to show that

$$x_{i+1}^2+b^2x_i-bx_i^2-2x_ix_{i+1}-x_{i+1}+b^{i+1}(x_{i+1}-x_i)(r_x^{i+1}+r_y^{i+1})$$

is non-negative.

Case 4a: $(x_i < x_{i+1})$

We have

$$\begin{aligned}
&x_{i+1}^2 + b^2 x_i - b x_i^2 - 2x_i x_{i+1} - x_{i+1} + b^{i+1}(x_{i+1} - x_i)(r_x^{i+1} + r_y^{i+1}) \\
&\geq x_{i+1}^2 + b^2 x_i - b x_i^2 - 2x_i x_{i+1} - x_{i+1} \\
&\geq x_{i+1}^2 + b x_i (x_{i+1} + 1) - b x_i x_{i+1} - 2x_i x_{i+1} - x_{i+1} \\
&= x_{i+1}^2 + b x_i - 2x_i x_{i+1} - x_{i+1} \\
&\geq x_{i+1}^2 + (x_{i+1} + 1) x_i - 2x_i x_{i+1} - x_{i+1} \\
&\geq x_{i+1}^2 - x_i x_{i+1} - x_{i+1} + x_i \\
&\geq x_{i+1}(x_{i+1} - x_i - 1) \\
&\geq 0.
\end{aligned}$$

Case 4b: $(x_{i+1} \leq x_i)$

Since $r_x^{i+1}, r_y^{i+1} \leq 1/b^{i+1}$, we have

$$\begin{aligned}
&x_{i+1}^2 + b^2 x_i - b x_i^2 - 2x_i x_{i+1} - x_{i+1} + b^{i+1}(x_{i+1} - x_i)(r_x^{i+1} + r_y^{i+1}) \\
&\geq x_{i+1}^2 + b^2 x_i - b x_i^2 - 2x_i x_{i+1} + x_{i+1} - 2x_i \\
&= (x_{i+1} - x_i)^2 + b^2 x_i - (b+1) x_i^2 - 2x_i + x_{i+1} \\
&\geq (x_{i+1} - x_i)^2 + x_i(b^2 - (b+1)(b-1) - 2) + x_{i+1} \\
&= (x_{i+1} - x_i)^2 + x_{i+1} - x_i \\
&= (x_{i+1} - x_i)(x_{i+1} - x_i + 1) \\
&= (x_i - x_{i+1})(x_i - x_{i+1} - 1) \geq 0.
\end{aligned}$$

□

3 The joint pdf of scrambled digital (t, m, s)-nets

We must first introduce some notation that will be helpful for important counting arguments that are needed to derive the joint pdf of scrambled digital (t, m, s)-nets.

Definition 3.1. Let $\tilde{P}_n = \{\mathbf{U}_1, \ldots, \mathbf{U}_n\}$ be a scrambled digital (t, m, s)-net in base b with generating matrices $\mathbf{C} = (C_1, \ldots, C_s)$, where each C_j is assumed to have m columns, at least m rows, and with elements in $\mathbb{F}_b$.

1. Let $m(\boldsymbol{k}; \tilde{P}_n, \mathbf{U}_l)$ be the number of points $\mathbf{U}_j \in \tilde{P}_n$ satisfying $\boldsymbol{k} \leq \boldsymbol{\gamma}_b^s(\mathbf{U}_l, \mathbf{U}_j)$. If these numbers are constant then we write $m(\boldsymbol{k}; \tilde{P}_n, \mathbf{U}_l) = m(\boldsymbol{k}; \tilde{P}_n)$ or simply $m(\boldsymbol{k})$.
2. We define $n(\boldsymbol{i}; \tilde{P}_n, \mathbf{U}_l)$ to be the number of points $\mathbf{U}_j \in \tilde{P}_n$ satisfying $\boldsymbol{\gamma}_b^s(\mathbf{U}_l, \mathbf{U}_j) = \boldsymbol{i}$. If these numbers are constant we write $n(\boldsymbol{i}; \tilde{P}_n, \mathbf{U}_l) = n(\boldsymbol{i}; \tilde{P}_n)$ or simply $n(\boldsymbol{i})$.
3. Let $r(\boldsymbol{k})$ be the rank of the matrix formed by the first k_j rows of C_j, $j = 1, \ldots, s$. (If $\boldsymbol{k} = \mathbf{0}$ then we set $r(\boldsymbol{k}) = 0$.)

Remark 3.2. Our analysis of the joint pdf employs the generating matrices that are used in the digital method of construction of a (t, m, s)-net. There are only small changes necessary if one wanted to study scrambled (t, m, s)-nets that do not arise from the digital method. We sketch this in Remarks 3.7 and 4.11.

Lemma 3.3. *Let $\tilde{P}_n$ be a scrambled digital (t, m, s)-net in base b. Then*

$$m(\boldsymbol{k}; \tilde{P}_n, \mathbf{U}_l) = m(\boldsymbol{k}; \tilde{P}_n) = b^{m - r(\boldsymbol{k})}.$$

In the special case $t = 0$ this formula becomes $m(\boldsymbol{k}; \tilde{P}_n) = \max(b^{m-k}, 1)$.

Proof. Consider the partition $I_{\boldsymbol{k}}(\mathbf{a})$ of $[0,1]^s$ into elementary $\boldsymbol{k}-$intervals and let $\mathbf{a}_l$ denote the k-dimensional integer vector corresponding to the elementary interval $I^l_{\boldsymbol{k}}$ from this partition in which $\mathbf{U}_l$ lies. Let $T_{\boldsymbol{k}}$ be the linear transformation from $\mathbb{F}_b^m$ to $\mathbb{F}_b^k$ determined by the k rows formed by the first k_j rows of C_j, for $j = 1, \dots, s$. Given that $\mathbf{U}_l$ lies in $I^l_{\boldsymbol{k}}$, it means $\mathbf{a}_l$ is in the image of $T_{\boldsymbol{k}}$. The dimension of the null space of $T_{\boldsymbol{k}}$ is $m - r(\boldsymbol{k})$, and thus the number of points in $I^l_{\boldsymbol{k}}$ (including $\mathbf{U}_l$) is given by $b^{m-r(\boldsymbol{k})}$.

When $t = 0$, we know that $r(\boldsymbol{k}) = k$ for $k \leq m$ and if $k > m$, then each elementary interval either has 0 or 1 point. □

Lemma 3.4. *If $\tilde{P}_n$ is a scrambled (t, m, s)-net in base b, then*

$$n(\boldsymbol{i}; \tilde{P}_n) = \sum_{\boldsymbol{k} \in \{0,1\}^s} (-1)^k m(\boldsymbol{i} + \boldsymbol{k}; \tilde{P}_n) = \sum_{\boldsymbol{k} \in \{0,1\}^s} (-1)^k b^{m - r(\boldsymbol{i}+\boldsymbol{k})}.$$

In the special case $t = 0$ this formula becomes

$$n(\boldsymbol{i}; \tilde{P}_n) = \sum_{k=0}^{s} (-1)^k \binom{s}{k} \max(b^{m-i-k}, 1).$$

Proof. Fix $\mathbf{U}_l \in \tilde{P}_n$ and for each $\boldsymbol{k} \in \mathbb{N}^s$ let $I^l_{\boldsymbol{k}}$ denote the elementary $\boldsymbol{k}-$interval that contains $\mathbf{U}_l$. Since a point $\mathbf{U}_j \in \tilde{P}_n \cap I^l_{\boldsymbol{i}}$ satisfies $\boldsymbol{\gamma}^s_b(\mathbf{U}_l, \mathbf{U}_j) = \boldsymbol{i}$ if and only if $\mathbf{U}_j$ is not in any $I^l_{\boldsymbol{i}+\boldsymbol{k}}$, where $\boldsymbol{k} \in \{0,1\}^s$ with $k = 1$, it holds that $n(\boldsymbol{i}; \tilde{P}_n, \mathbf{U}_l)$ counts the number of points from $\tilde{P}_n$ that are in the set

$$I^l_{\boldsymbol{i}} \backslash \left(\bigcup_{\boldsymbol{k} \in \{0,1\}^s,\, k=1} I^l_{\boldsymbol{i}+\boldsymbol{k}} \right).$$

To apply the Principle of Inclusion-Exclusion we observe that the intersection of any n distinct elementary intervals in the above union is an elementary interval of the form $I^l_{\boldsymbol{i}+\boldsymbol{k}}$, where $\boldsymbol{k} \in \{0,1\}^s$ with $k = n$, and that $m(\boldsymbol{i} + \boldsymbol{k}; \tilde{P}_n, \mathbf{U}_l)$ counts the number of points in $I^l_{\boldsymbol{i}+\boldsymbol{k}}$. Thus by the Principle of Inclusion-Exclusion we have

$$\begin{aligned} n(\boldsymbol{i}; \tilde{P}_n, \mathbf{U}_l) &= \sum_{n=0}^{s} \sum_{\substack{\boldsymbol{k} \in \{0,1\}^s \\ k=n}} (-1)^n m(\boldsymbol{i} + \boldsymbol{k}; \tilde{P}_n, \mathbf{U}_l) \\ &= \sum_{\boldsymbol{k} \in \{0,1\}^s} (-1)^k m(\boldsymbol{i} + \boldsymbol{k}; \tilde{P}_n) = \sum_{\boldsymbol{k} \in \{0,1\}^s} (-1)^k b^{m-r(\boldsymbol{i}+\boldsymbol{k})}, \end{aligned}$$

where the second and third equality follow from the previous lemma, which implies that for fixed $\boldsymbol{i}'$, $m(\boldsymbol{i}'; \tilde{P}_n, \mathbf{U}_l)$ is the same for all $\mathbf{U}_l$. In the special case $t = 0$ we use the fact that there are exactly $\binom{s}{n}$ vectors $\boldsymbol{k} \in \{0,1\}^s$ with $k = n$, together with the previous lemma to simplify

$$n(\boldsymbol{i}; \tilde{P}_n) = \sum_{n=0}^{s} \sum_{\substack{\boldsymbol{k} \in \{0,1\}^s \\ k=n}} (-1)^k \max(b^{m-i-k}, 1) = \sum_{k=0}^{s} (-1)^k \binom{s}{k} \max(b^{m-i-k}, 1).$$ □

Remark 3.5. In the remainder of this paper, we make the assumption that we are working with digital (t, m, s)-nets whose one-dimensional projections have $t = 0$. Equivalently, this means we assume all s generating matrices have rank m. Doing so avoids the case where we have two points with equal coordinates in one or more dimension, which would in turn lead to a joint pdf with non-zero value on the diagonal of the unit hypercube, i.e., on $D^s_{\boldsymbol{i}}$ with $i = \infty$.

Theorem 3.6. *Let $\tilde{P}_n = \{\mathbf{U}_1, \dots, \mathbf{U}_n\}$ be a scrambled digital net in base b whose one-dimensional projections are scrambled $(0, m, 1)$-nets. Then the joint pdf $\psi(\mathbf{x}, \mathbf{y})$ of two distinct points $(\mathbf{U}_I, \mathbf{U}_J)$ randomly chosen from $\tilde{P}_n$ is given by*

$$\psi(\mathbf{x}, \mathbf{y}) = \begin{cases} \frac{n(\boldsymbol{i})}{b^m - 1} \frac{b^{s+i}}{(b-1)^s} & \text{if } i < \infty, \\ 0 & \text{if } i = \infty, \end{cases}$$

where $\boldsymbol{i} = \boldsymbol{\gamma}^s_b(\mathbf{x}, \mathbf{y})$ and $i = \gamma^s_b(\mathbf{x}, \mathbf{y})$.

Proof. Because the one-dimensional projections of $\tilde{P}_n$ are scrambled $(0, m, 1)$-nets, no distinct pair $(\mathbf{U}_l, \mathbf{U}_j)$ lies in $D_{\boldsymbol{i}}^s$ with $i = \infty$, therefore $\psi(\mathbf{x}, \mathbf{y})$ is 0 on these regions. Next, examining parts 1 and 2 in the definition of a scrambled digital net in base b we see that the only thing we can say about the distribution of the randomized pair $(\mathbf{U}_l, \mathbf{U}_j)$ is that $\boldsymbol{\gamma}_b^s(\mathbf{U}_l, \mathbf{U}_j) = \boldsymbol{\gamma}_b^s(\mathbf{V}_l, \mathbf{V}_j)$ (where $(\mathbf{V}_l, \mathbf{V}_j)$ is the corresponding deterministic pair) and that the pair is otherwise uniformly dsitributed. In other words, $(\mathbf{U}_l, \mathbf{U}_j) \sim U(D_{\boldsymbol{i}}^s)$ where $\boldsymbol{i} = \boldsymbol{\gamma}(\mathbf{U}_l, \mathbf{U}_j)$. It follows that $\psi(\mathbf{x}, \mathbf{y})$ is constant on $D_{\boldsymbol{i}}^s$. The value $\psi_{\boldsymbol{i}}$ of $\psi(\mathbf{x}, \mathbf{y})$ on $D_{\boldsymbol{i}}^s$ $(i \neq \infty)$ can be found by observing that the integral of $\psi(\mathbf{x}, \mathbf{y})$ over $D_{\boldsymbol{i}}^s$ is equal to the probability that a random pair of distinct points from $\tilde{P}_n$ lie in $D_{\boldsymbol{i}}^s$ and then solving

$$\frac{b^m n(\boldsymbol{i})}{b^m(b^m-1)} = \int_{D_{\boldsymbol{i}}^s} \psi(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} = \psi_{\boldsymbol{i}} \operatorname{Vol}(D_{\boldsymbol{i}}^s) = \psi_{\boldsymbol{i}} \frac{(b-1)^s}{b^{s+i}}. \qquad \square$$

Remark 3.7. A few observations are in order:

1. The joint pdf is a simple function because if $\gamma_b(x_j, y_j) \geq m$, then $\psi(\mathbf{x}, \mathbf{y}) = 0$.
2. The joint pdf of a $(0, m, s)$-net depends only on the sum $\gamma_b^s(\mathbf{x}, \mathbf{y})$ and not on the vector $\boldsymbol{\gamma}_b^s(\mathbf{x}, \mathbf{y})$.
3. If one used the equidistribution property to define a scrambled (t, m, s)-net (rather than the digital method), then the value of the joint pdf on $D_{\boldsymbol{i}}^s$ would be

$$\frac{N(\boldsymbol{i})}{b^m(b^m-1)} \frac{b^{s+i}}{(b-1)^s}$$

   where $N(\boldsymbol{i}) = \sum_{l=1}^n n(\boldsymbol{i}; \tilde{P}_n, \mathbf{U}_l)$ is the number of pairs of points in $D_{\boldsymbol{i}}^s$.
4. Using Lemma 2.3 and its preceding discussion along with the formulas in this section, one can compute $H(\mathbf{x}, \mathbf{y}; \tilde{P}_n)$ exactly.

4 Dependence structure of scrambled (t, m, s)-nets

By the end of this section we will have shown that the only srambled (t, m, s)-nets that are NLOD/NUOD are those for which $t = 0$. We have seen that the joint pdf of such nets is a function that is constant on $D_{\boldsymbol{i}}^s$ when $\boldsymbol{i} \in \mathbb{N}^s$ and is otherwise zero. (Note that in this paper, we assume that $\mathbb{N}$ includes 0.) For most of this section we will work towards finding an exact value for the minimum constant $C \geq 1$ that satisfies

$$G(\mathbf{x}, \mathbf{y}) := \int_{R(\mathbf{x}, \mathbf{y})} \psi(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v} \leq C \int_{R(\mathbf{x}, \mathbf{y})} d\mathbf{u} d\mathbf{v} = C\mathbf{x}\mathbf{y} = C \prod_{j=1}^s x_j y_j$$

for all $\mathbf{x}, \mathbf{y} \in [0,1]^s$. As we will see, this can be stated and solved using standard techniques from functional analysis.

In order to apply tools from functional analysis, we first associate the joint pdf ψ with the vector $\psi = (\psi_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathbb{N}^s} \in \ell^\infty(\mathbb{N}^s)$ where $\psi_{\boldsymbol{i}}$ is the value assumed by ψ on $D_{\boldsymbol{i}}^s$. This *value vector* induces a continuous linear functional $\widehat{\psi} : \ell^1(\mathbb{N}^s) \to \mathbb{C}$ via the formula

$$\widehat{\psi}(\eta) := \sum_{\boldsymbol{i} \in \mathbb{N}^s} \eta_{\boldsymbol{i}} \psi_{\boldsymbol{i}}.$$

Next, for each $\mathbf{x}, \mathbf{y} \in [0,1]^s$ we define

$$V^s(\mathbf{x}, \mathbf{y}) := (V_{\boldsymbol{i}}^s(\mathbf{x}, \mathbf{y}))_{\boldsymbol{i} \in \mathbb{N}^s} \in \ell^1(\mathbb{N}^s)$$

to be the *volume vector* of the region $R(\mathbf{x}, \mathbf{y})$, and observe that

$$\|V^s(\mathbf{x}, \mathbf{y})\|_1 = \operatorname{Vol}(R(\mathbf{x}, \mathbf{y})) = \mathbf{x}\mathbf{y}.$$

With this notation $G(\mathbf{x},\mathbf{y}) = \widehat{\psi}(V^s(\mathbf{x},\mathbf{y}))$. As usual, in the special case $s=1$ we drop the exponent and write

$$V(x,y) := (V_i(x,y))_{i=0}^{\infty} \in \ell^1(\mathbb{N}).$$

By letting

$$\mathcal{C}^s := \left\{ \frac{V^s(\mathbf{x},\mathbf{y})}{\mathbf{xy}} : \mathbf{x},\mathbf{y} \in [0,1]^s \right\} \subseteq \ell^1(\mathbb{N}^s)$$

be the set of normalized volume vectors and denoting the norm of $\widehat{\psi}$ over $\mathcal{C}^s$ by

$$\|\widehat{\psi}\|_{\mathcal{C}^s} := \sup_{\eta \in \mathcal{C}^s} \widehat{\psi}(\eta)$$

we get

$$G(\mathbf{x},\mathbf{y}) = \widehat{\psi}(V^s(\mathbf{x},\mathbf{y})) \leq \mathbf{xy}\|\widehat{\psi}\|_{\mathcal{C}^s},$$

which holds for all $\mathbf{x},\mathbf{y} \in [0,1]^s$. Thus, in the language of functional analysis our goal is to estimate $\|\widehat{\psi}\|_{\mathcal{C}^s}$, which we will do by way of a convexity argument.

Definition 4.1. We have the two following definitions.

1. For each $\boldsymbol{k} \in \mathbb{N}^s$ we define $S^{\boldsymbol{k}} : \ell^1(\mathbb{N}^s) \to \ell^1(\mathbb{N}^s)$ to be the bounded linear operator that acts on the standard Schauder basis $\{e_{\boldsymbol{i}}\}_{\boldsymbol{i}\in\mathbb{N}^s}$ according the the rule $S^{\boldsymbol{k}} e_{\boldsymbol{i}} = e_{\boldsymbol{i}+\boldsymbol{k}}$.
2. Given $\mathbf{k} = (k_1,\ldots,k_s) \in \mathbb{N}^s$ and $\mathbf{x} \in [0,1]^s$ we define

$$b^{-\mathbf{k}}\mathbf{x} := (b^{-k_1}x_1,\ldots,b^{-k_s}x_s).$$

Lemma 4.2. *Let $\mathbf{x},\mathbf{y} \in [0,1]^s$ and $\boldsymbol{k} \in \mathbb{N}^s$. Then*

$$S^{\boldsymbol{k}} V^s(\mathbf{x},\mathbf{y}) = b^{2k} V^s(b^{-\mathbf{k}}\mathbf{x}, b^{-\mathbf{k}}\mathbf{y}).$$

In particular $S^{\boldsymbol{k}}\mathcal{C}^s \subseteq \mathcal{C}^s$.

Proof. We start by observing $\gamma_b(b^{-k}u, b^{-k}v) = \gamma_b(u,v) + k$, for all $k \in \mathbb{N}$ and $u,v \in [0,1]$. From this it follows that given $x,y \in [0,1]$, the region $R(b^{-k}x, b^{-k}y) \cap D_{i+k}$ can be obtained by scaling the region $R(x,y) \cap D_i$ by a factor of b^{-k} in each coordinate, indeed

$$R(b^{-k}x, b^{-k}y) \cap D_{i+k} = \{(b^{-k}u, b^{-k}v) : (u,v) \in R(x,y) \cap D_i\}.$$

Since no pair $(u,v) \in R(b^{-k}x, b^{-k}y)$ can have less than k initial common digits, we can write

$$b^{2k} V_i(b^{-k}x, b^{-k}y) = \begin{cases} V_{i-k}(x,y) & \text{if } k \leq i, \\ 0 & \text{otherwise.} \end{cases}$$

Now,

$$\begin{aligned} S^{\boldsymbol{k}} V^s(\mathbf{x},\mathbf{y}) &= \sum_{\boldsymbol{i}\in\mathbb{N}^s} V^s_{\boldsymbol{i}}(\mathbf{x},\mathbf{y}) e_{\boldsymbol{i}+\boldsymbol{k}} = \sum_{\boldsymbol{k}\leq\boldsymbol{i}\in\mathbb{N}^s} V^s_{\boldsymbol{i}-\boldsymbol{k}}(\mathbf{x},\mathbf{y}) e_{\boldsymbol{i}} \\ &= \sum_{\boldsymbol{k}\leq\boldsymbol{i}\in\mathbb{N}^s} \left(\prod_{j=1}^{s} V_{i_j-k_j}(x_j,y_j) \right) e_{\boldsymbol{i}} \\ &= \sum_{\boldsymbol{i}\in\mathbb{N}^s} \left(\prod_{j=1}^{s} b^{2k_j} V_{i_j}(b^{-k_j}x_j, b^{-k_j}y_j) \right) e_{\boldsymbol{i}} \\ &= b^{2k} \sum_{\boldsymbol{i}\in\mathbb{N}^s} V_{\boldsymbol{i}}(b^{-\boldsymbol{k}}\mathbf{x}, b^{-\boldsymbol{k}}\mathbf{y}) e_{\boldsymbol{i}} = b^{2k} V^s(b^{-\boldsymbol{k}}\mathbf{x}, b^{-\boldsymbol{k}}\mathbf{y}). \end{aligned}$$

□

Definition 4.3. Let $\xi \in \ell^1(\mathbb{N})$ and $\xi^s \in \ell^1(\mathbb{N}^s)$ be defined as

$$\xi := V(1,1) = (\mathrm{Vol}(D_i))_{i=0}^{\infty} = (\tfrac{b-1}{b^{i+1}})_{i=0}^{\infty} \text{ and}$$
$$\xi^s := V^s(\mathbf{1},\mathbf{1}) = (\mathrm{Vol}(D_{\boldsymbol{i}}^s))_{\boldsymbol{i}\in\mathbb{N}^s} = (\tfrac{(b-1)^s}{b^{s+i}})_{\boldsymbol{i}\in\mathbb{N}^s}.$$

By applying the shift operators $S^{\boldsymbol{k}}$ to these vectors, we generate Schauder bases

$$\{S^k\xi : k \in \mathbb{N}\} \text{ and } \{S^{\boldsymbol{k}}\xi^s : \boldsymbol{k} \in \mathbb{N}^s\}.$$

This means that every element of $\ell^1(\mathbb{N}^s)$ can be written uniquely as $\sum_{\boldsymbol{k}\in\mathbb{N}^s} t_{\boldsymbol{k}} S^{\boldsymbol{k}}\xi^s$, where $t_{\boldsymbol{k}} \in \mathbb{C}$. By Lemma 4.2 we see that if $\boldsymbol{k} = (k_1, \dots, k_s)$, then

$$S^{\boldsymbol{k}}\xi^s = b^{2k} V^s(b^{-\boldsymbol{k}}\mathbf{1}, b^{-\boldsymbol{k}}\mathbf{1}) \in \mathcal{C}^s.$$

The next two lemmas show, in particular, that the elements $\eta \in \mathcal{C}^s$ are convex combinations of $\{S^{\boldsymbol{k}}\xi^s : \boldsymbol{k} \in \mathbb{N}^s\}$, i.e.

$$\eta = \sum_{\boldsymbol{k}\in\mathbb{N}^s} t_{\boldsymbol{k}} S^{\boldsymbol{k}}\xi^s \text{ where } t_{\boldsymbol{k}} \geq 0 \text{ and } \sum_{\boldsymbol{k}\in\mathbb{N}^s} t_{\boldsymbol{k}} = 1.$$

As usual the general case follows from the one-dimensional case.

Remark 4.4. The reason why there is no mention of convexity in the next two lemmas is that to do so would mean having to deal with $V(x,y)/xy$ and $V^s(\mathbf{x},\mathbf{y})/\mathbf{xy}$ instead of $V(x,y)$ and $V^s(\mathbf{x},\mathbf{y})$.

Lemma 4.5. *For all $x, y \in [0,1]$ there exists a sequence $t_k \geq 0$, $k \in \mathbb{N}$, such that*

$$V(x,y) = \sum_{k=0}^{\infty} t_k S^k \xi.$$

Proof. Because $\{S^k\xi : k \in \mathbb{N}\}$ is a Schauder basis, we can write $V(x,y) = \sum_{k=0}^{\infty} t_k S^k\xi$ for some scalars t_k. To prove that $t_k \geq 0$ we will find an explicit formula for t_k. To that end, we define the sequence $\eta_n \in \ell^1(\mathbb{N})$ by

$$\eta_n = V(x,y) - \sum_{k=0}^{n} t_k S^k\xi = \sum_{k=n+1}^{\infty} t_k S^k\xi,$$

which satisfies $t_n = \eta_{n-1,n}/\xi_0$ for $n \geq 1$, and claim that

$$\eta_{n,i} = \begin{cases} 0 & \text{if } i \leq n, \\ V_i - b^{n-i}V_n & \text{if } i > n, \end{cases}$$

where $V(x,y) = (V_0, V_1, \dots)$ and $\eta_n = (\eta_{n,0}, \eta_{n,1}, \dots)$ for $n \geq 0$.

Proceeding by induction, we observe that $t_0 = V_0/\xi_0$ and $\xi_i/\xi_j = b^{j-i}$ to get

$$\eta_0 = (0, V_1 - b^{-1}V_0, \dots, V_i - b^{-i}V_0, \dots).$$

Next, we assume that the coordinates of η_{n-1} satisfy the formula and note that $\eta_n = \eta_{n-1} - t_n S^n\xi$ for $n \geq 1$. It is easy to see that $\eta_{n,i} = 0$ when $i \leq n$, and for $i > n$ we compute

$$\begin{aligned} \eta_{n,i} &= \eta_{n-1,i} - (\eta_{n-1,n}/\xi_0)\xi_{i-n} \\ &= V_{\boldsymbol{i}} - b^{n-1-i}V_{n-1} - (V_n - b^{-1}V_{n-1})\xi_{i-n}/\xi_0 \\ &= V_{\boldsymbol{i}} - b^{n-1-i}V_{n-1} - b^{n-i}V_n + b^{n-i-1}V_{n-1} \\ &= V_{\boldsymbol{i}} - b^{n-i}V_n, \end{aligned}$$

which proves the claim. Finally, for $k \geq 1$

$$t_k = \frac{bV_k - V_{k-1}}{b-1} \geq 0$$

by Lemma 2.4. $\square$

Lemma 4.6. *For all* $\mathbf{x}, \mathbf{y} \in [0,1]^s$ *there exist scalars* $t_{\boldsymbol{k}} \geq 0$, $\boldsymbol{k} \in \mathbb{N}^s$, *such that*

$$V^s(\mathbf{x}, \mathbf{y}) = \sum_{\boldsymbol{k} \in \mathbb{N}^s} t_{\boldsymbol{k}} S^{\boldsymbol{k}} \xi^s \text{ and } \sum_{\boldsymbol{k} \in \mathbb{N}^s} t_{\boldsymbol{k}} = \mathbf{xy}.$$

Proof. In this proof we identify elements $(\eta_i)_{i=0}^{\infty} \in \ell^1(\mathbb{N})$ and $(\eta_{\boldsymbol{i}})_{\boldsymbol{i} \in \mathbb{N}^s} \in \ell^1(\mathbb{N}^s)$ with the power series

$$\sum_{i=0}^{\infty} \eta_i z^i \text{ and } \sum_{\boldsymbol{i} \in \mathbb{N}^s} \eta_{\boldsymbol{i}} \mathbf{z}^{\boldsymbol{i}}$$

where $\mathbf{z} = \prod_{j=1}^{s} z_j^{i_j}$. In particular, we define

$$\begin{aligned} f_j(z) &= \sum_{i=0}^{\infty} V_i(x_j, y_j) z^i, & f^s(z) &= \sum_{\boldsymbol{i} \in \mathbb{N}^s}^{\infty} V_{\boldsymbol{i}}^s(\mathbf{x}, \mathbf{y}) \mathbf{z}^{\boldsymbol{i}}, \\ g(z) &= \sum_{i=0}^{\infty} V_i(1,1) z^i, \text{ and} & g^s(z) &= \sum_{\boldsymbol{i} \in \mathbb{N}^s}^{\infty} V_{\boldsymbol{i}}(\mathbf{1}, \mathbf{1}) \mathbf{z}^{\boldsymbol{i}}, \end{aligned}$$

and we observe that $S^k \xi$ and $S^{\boldsymbol{k}} \xi^s$ correspond with $z^k g(z)$ and $\mathbf{z}^{\boldsymbol{k}} g^s(\mathbf{z})$ respectively. With this notation, the conclusion of Lemma 4.5 becomes $f_j(z) = g(z) h_j(z)$, where $h_j(z) = \sum_{k=0} t_{j,k} z^k$ and $t_{j,k} \geq 0$. Since $f^s(\mathbf{z}) = \prod_{j=1}^{s} f_j(z_j)$ and $g^s(\mathbf{z}) = \prod_{j=1}^{s} g(z_j)$, we have

$$f^s(\mathbf{z}) = \prod_{j=1}^{s} g(z_j) h_j(z_j) = g^s(\mathbf{z}) h^s(\mathbf{z}),$$

where $h^s(\mathbf{z}) = \sum_{\boldsymbol{k} \in \mathbb{N}^s} t_{\boldsymbol{k}} \mathbf{z}^{\boldsymbol{k}}$ is the product of the $h_j(z_j)$'s. Finally, evaluating $f^s(\mathbf{z})$ and $g^s(\mathbf{z})$ at $\mathbf{z} = \mathbf{1}$ yields

$$f^s(\mathbf{1}) = \sum_{\boldsymbol{i} \in \mathbb{N}^s} V_{\boldsymbol{i}}^s(\mathbf{x}, \mathbf{y}) = \|V^s(\mathbf{x}, \mathbf{y})\|_1 = \mathbf{xy}$$

and $g^s(\mathbf{1}) = \mathbf{1}$, thus

$$\sum_{\boldsymbol{k} \in \mathbb{N}^s} t_{\boldsymbol{k}} = g^s(\mathbf{1}) h^s(\mathbf{1}) = f^s(\mathbf{1}) = \mathbf{xy}. \qquad \square$$

Remark 4.7. The convexity property in the previous lemma directly implies that $\|\widehat{\zeta}\|_{\mathcal{C}^s} = \sup |\widehat{\zeta}(S^{\boldsymbol{k}} \xi^s)|$ for all $\zeta \in \ell^{\infty}(\mathbb{N}^s)$.

Theorem 4.8. *Let* $\tilde{P}_n = \{\mathbf{U}_1, \ldots, \mathbf{U}_n\}$ *be a scrambled digital* (t, m, s)*-net in base* b *whose one-dimensional projections are scrambled* $(0, m, 1)$*-nets and let* $\psi(\mathbf{x}, \mathbf{y})$ *be the joint pdf of two distinct points* $(\mathbf{U}_I, \mathbf{U}_J)$ *randomly chosen from* $\tilde{P}_n$. *Then*

$$\int_{R(\mathbf{x}, \mathbf{y})} \psi(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v} \leq C \mathbf{xy} \text{ where } C = \max_{\boldsymbol{k} \in \mathbb{N}^s} \frac{b^k (m(\boldsymbol{k}) - 1)}{b^m - 1} = \max_{\boldsymbol{k} \in \mathbb{N}^s} \frac{b^k (b^{m - r(\boldsymbol{k})} - 1)}{b^m - 1} \geq 1$$

holds for all $\mathbf{x}, \mathbf{y} \in [0,1]^s$. *Moreover* C *is the smallest constant that satisfies the inequality.*

Proof. By Lemma 4.6, given $\mathbf{x}, \mathbf{y} \in [0,1]^s$ there exists scalars $t_{\boldsymbol{k}} \geq 0$ with $\sum_{\boldsymbol{k} \in \mathbb{N}^s} t_{\boldsymbol{k}} = \mathbf{xy}$ such that $V^s(\mathbf{x}, \mathbf{y}) = \sum_{\boldsymbol{k} \in \mathbb{N}^s} t_{\boldsymbol{k}} S^{\boldsymbol{k}} \xi^s$. Now

$$\begin{aligned} \int_{R(\mathbf{x}, \mathbf{y})} \psi(\mathbf{u}, \mathbf{v}) d\mathbf{u} d\mathbf{v} &= \widehat{\psi}(V^s(\mathbf{x}, \mathbf{y})) = \widehat{\psi} \left(\sum_{\boldsymbol{k} \in \mathbb{N}^s} t_{\boldsymbol{k}} S^{\boldsymbol{k}} \xi^s \right) \\ &\leq \sum_{\boldsymbol{i} \in \mathbb{N}^s} t_i \sup_{\boldsymbol{k} \in \mathbb{N}^s} \widehat{\psi}(S^{\boldsymbol{k}} \xi^s) = \mathbf{xy} \cdot \sup_{\boldsymbol{k} \in \mathbb{N}^s} \widehat{\psi}(S^{\boldsymbol{k}} \xi^s). \end{aligned}$$

Recalling the form that the joint pdf takes in Theorem 3.6, for each $\boldsymbol{k} \in \mathbb{N}^s$ we have

$$\begin{aligned}
\widehat{\psi}(S^{\boldsymbol{k}}\xi^s) &= \widehat{\psi}\left(\sum_{\boldsymbol{i}\in\mathbb{N}^s} \frac{(b-1)^s}{b^{s+i}} e_{\boldsymbol{i}+\boldsymbol{k}}\right) = \sum_{\boldsymbol{i}\in\mathbb{N}^s} \psi_{\boldsymbol{i}+\boldsymbol{k}} \frac{(b-1)^s}{b^{s+i}} \\
&= \sum_{\boldsymbol{i}\in\mathbb{N}^s} \frac{n(\boldsymbol{i}+\boldsymbol{k})}{b^m-1} \frac{b^{s+i+k}}{(b-1)^s} \frac{(b-1)^s}{b^{s+i}} = \sum_{\boldsymbol{i}\in\mathbb{N}^s:\boldsymbol{i}\geq\boldsymbol{k}} \frac{b^k n(\boldsymbol{i})}{b^m-1} \\
&= \frac{b^k(m(\boldsymbol{k})-1)}{b^m-1} = \frac{b^k(b^{m-r(\boldsymbol{k})}-1)}{b^m-1}
\end{aligned}$$

because we do not count any $\boldsymbol{i}$ with $i=\infty$. To see that only finitely many values in the supremum are non-zero, we suppose that $\boldsymbol{k}\in\mathbb{N}^s$ has $k_j \geq m$ and denote by $\tilde{P}_n^j$ the orthogonal projection of $\tilde{P}_n$ onto the jth coordinate. Then $1 \leq m(\boldsymbol{k};\tilde{P}_n) \leq m(k_j;\tilde{P}_n^j) = 1$ by the equidistribution properties of $(0,m,1)$-nets, and thus $\widehat{\psi}(S^{\boldsymbol{k}}\xi^s)=0$.

To see that C is the minimum possible value for each $\boldsymbol{k}\in\mathbb{N}^s$, we apply Lemma 4.2 to $\mathbf{x} = b^{-\boldsymbol{k}}\mathbf{1}$:

$$\int_{R(\mathbf{x},\mathbf{x})} \psi(\mathbf{u},\mathbf{v})d\mathbf{u}d\mathbf{v} = b^{-2\boldsymbol{k}}\widehat{\psi}(S^{\boldsymbol{k}}\xi^s) = \widehat{\psi}(S^{\boldsymbol{k}}\xi^s)\mathbf{x}\mathbf{x}. \qquad \square$$

Remark 4.9. The quantity C defined in the statement of Theorem 4.8 can be thought of as a measure of how close a sampling scheme is to having the NLOD/NUOD property (corresponding to having $C=1$), which we argue is desirable. Since two nets with the same value of t, m, and s might have different values for C—as C further depends on the generating matrices through the quantity $r(\boldsymbol{k})$—it might turn out that

$$\max_{\boldsymbol{k}\in\mathbb{N}^s} \frac{b^k(b^{m-r(\boldsymbol{k})}-1)}{b^m-1}$$

is a good way to decide which of the two nets is better. Of even more interest would be to define a vector $\mathbf{C} = (C_{\boldsymbol{k}})_{\boldsymbol{k}\geq\mathbf{0}}$ where $C_{\boldsymbol{k}} = (b^k(b^{m-r(\boldsymbol{k})}-1))/(b^m-1)$, and consider, say, a weighted average of its components to assess the quality of a net, rather than looking at the largest component. We certainly plan to study this vector $\mathbf{C}$ further in future work. Also note that $\mathbf{C}$ can be used to assess the quality of a deterministic (t,m,s)-net, since it is defined through the $m(\boldsymbol{k})$'s, which are derived from properties of the unscrambled version of the net.

Corollary 4.10. *Let $\tilde{P}_n = \{\mathbf{U}_1,\ldots,\mathbf{U}_n\}$ be a scrambled digital (t,m,s)-net in base b whose one-dimensional projections are scrambled $(0,m,1)$-nets. Then $\tilde{P}_n$ is an NUOD/NLOD sampling scheme if and only if $t=0$.*

Proof. The joint pdf $\psi(\mathbf{x},\mathbf{y})$ of two distinct points $(\mathbf{U}_I,\mathbf{U}_J)$ randomly chosen from $\tilde{P}_n$ is invariant under the transformation $\mathbf{x},\mathbf{y} \mapsto (\mathbf{1}-\mathbf{x},\mathbf{1}-\mathbf{y})$ because $\gamma_b^s(\mathbf{u},\mathbf{v})$ and therefore $D_{\boldsymbol{i}}^s$ are invariant under that transformation. It follows that the NLOD and NUOD properties are equivalent for scrambled (t,m,s)-nets.

When $t=0$ we apply Lemma 3.3 to get

$$C = \max_{\boldsymbol{k}\in\mathbb{N}^s} \frac{b^k(m(\boldsymbol{k})-1)}{b^m-1} = \max_{\boldsymbol{k}\in\mathbb{N}^s} \frac{b^k(\max(b^{m-k},1)-1)}{b^m-1} = \max_{\boldsymbol{k}\in\mathbb{N}^s, k\leq m} \frac{b^m-b^k}{b^m-1} = 1.$$

Thus by Theorem 4.8 all scrambled $(0,m,s)$-nets are NLOD/NUOD.

Since, when $t>0$, a scrambled (t,m,s)-net is not a $(t-1,m,s)$-net (i.e., we assume the value of the parameter t is exact), there is some $\boldsymbol{k}\in\mathbb{N}^s$ with $k=m-t+1$ such that $r(\boldsymbol{k})=k-1=m-t$. Hence we get

$$\begin{aligned}
\frac{b^k(b^{m-r(\boldsymbol{k})}-1)}{b^m-1} &= \frac{b^{m-t+1}(b^t-1)}{b^m-1} = \frac{b^m(b-b^{1-t})}{b^m-1} \\
&\geq \frac{b^m(b-1)}{b^m-1} > 1,
\end{aligned}$$

which by Theorem 3.6 means that a scrambled (t,m,s)-net is not NLOD when $t>0$. $\square$

Remark 4.11. Once we have the formula for C in Theorem 4.8 we can directly apply the equidistribution property to prove that scrambled $(0, m, s)$-nets are NLOD. Although we use $r(\boldsymbol{k})$ to prove that scrambled (t, m, s)-nets are not NLOD when $t > 0$ we could have appealed directly to the equidistribution property.

To do this we would use the formula for the joint pdf that appears in Remark 3.7 to get

$$C = \max_{\boldsymbol{k}\in\mathbb{N}^s} \frac{b^k \sum_{l=1}^{n} \left[m(\boldsymbol{k}; \tilde{P}_n, \mathbf{U}_l) - 1\right]}{b^m - 1}.$$

Next a Pigeon-Hole-Principle argument can be used to show that there is some elementary $\boldsymbol{k}$−interval with $k = m$ with at least two points. Thus $C \geq b^m/(b^m - 1) > 1$. This shows that Corollary 4.10 is independent of the construction method for the scrambled net.

5 Using scrambled $(0, m, s)$-nets for numerical integration

Let $\mathcal{F}^s$ be the collection of functions $f : [0, 1]^s \to \mathbb{R}$ for which the variance of the estimator based on any scrambled $(0, m, s)$-net in any base b is no larger than the variance of the Monte Carlo estimator based on the same number of points. The goal of this section is to obtain a partial description of $\mathcal{F}^s$ that allows us to say more about that collection of functions than what can be inferred from the fact that scrambled $(0, m, s)$-nets are NLOD/NUOD. The outline is as follows. In Section 5.1 we use the NLOD/NUOD property to find an initial set of functions in $\mathcal{F}^s$. In Section 5.2 we expand the set found in Section 5.1 by analyzing which operations may be preformed on f so that the inequality

$$\int_{[0,1]^s} \int_{[0,1]^s} f(\mathbf{x})f(\mathbf{y})\psi(\mathbf{x}, \mathbf{y})d\mathbf{x}d\mathbf{y} \leq \int_{[0,1]^s} \int_{[0,1]^s} f(\mathbf{x})f(\mathbf{y})d\mathbf{x}d\mathbf{y}$$

is maintained when ψ is the joint pdf of two distinct points from a scrambled $(0, m, s)$-net in some base b. Finally, in Section 5.3 we give an example of a function in $\mathcal{F}^s$ that satisfies the results in Sections 5.1 and 5.2 and an example showing that $\mathcal{F}^s$ does not contain all functions of bounded variation in the sense of Hardy and Krause.

5.1 Exploiting the NLOD/NUOD property

The $s = 1$ case is special so we handle it separately. It turns out that $\mathcal{F}^1$ contains all integrable functions.

Proposition 5.1. *Let ψ be the joint pdf of two distinct points (U_I, U_J) randomly chosen from a scrambled $(0, m, 1)$-net. Then $\mathcal{F}^1$ contains all integrable functions $f : [0, 1] \to \mathbb{R}$.*

Proof. Recall that the joint pdf of two distinct points (U_I, U_J) randomly chosen from a scrambled $(0, m, 1)$-net is

$$\psi(x, y) = \begin{cases} \frac{b^m}{b^m - 1} & \text{if } \gamma_b(x, y) < m, \\ 0 & \text{if } \gamma_b(x, y) \geq m. \end{cases}$$

If $I_i = [(i - 1)/b^m, i/b^m]$, then $\psi(x, y)$ is non-zero exactly when $(x, y) \in I_i \times I_j$ for some $i \neq j$. Letting

$$\alpha_i = \int_{I_i} f(x)dx,$$

we estimate

$$\begin{aligned}
&(b^m-1)\Big(\int_{[0,1]^s}\int_{[0,1]^s} f(x)f(y)dxdy - \int_{[0,1]}\int_{[0,1]} f(x)f(y)\psi(x,y)dxdy\Big)\\
&= (b^m-1)\sum_{i=1}^{b^m}\sum_{j=1}^{b^m}\alpha_i\alpha_j - b^m\sum_{i=1}^{b^m}\sum_{j\neq i}\alpha_i\alpha_j = (b^m-1)\sum_{i=1}^{b^m}\alpha_i^2 - 2\sum_{i=1}^{b^m}\sum_{j<i}\alpha_i\alpha_j\\
&= \sum_{i=1}^{b^m}\sum_{j<i}\alpha_i^2 - 2\alpha_i\alpha_j + \alpha_j^2 = \sum_{i=1}^{b^m}\sum_{j<i}(\alpha_i-\alpha_j)^2 > 0.
\end{aligned}$$

□

For the case $s \geq 2$ we may simply recall the result in [7, Prop. 3 & Remark 8] which implies that functions that are quasi-monotone (see Definition 5.5) are in $\mathcal{F}^s$. Alternatively, the following result uses the NLOD property directly and the reader may find the set of functions that it describes easier to understand than the quasi-monotone property. It is unclear if the collection of functions obtained by applying the results from Section 5.2 to quasi-monotone functions yields more functions than applying those results to the functions in the next proposition. It may be that both collections yields the same functions. The functions in the following result also have a connection with the concept of $\mathcal{R}^*$-variation studied in [1]. This connection is something we wish to study in the near future.

Proposition 5.2. *The set $\mathcal{F}^s$ contains all functions of the form $\sum_{i=0}^n \alpha_i 1_{[\mathbf{0},\mathbf{a}_i]}$ with $\alpha_i > 0$.*

Proof. Consider $f(\mathbf{x}) = \sum_{i=0}^n \alpha_i 1_{[\mathbf{0},\mathbf{a}_i]}(\mathbf{x})$ and let $A_i = [\mathbf{0}, \mathbf{a}_i]$. Then we have that

$$\begin{aligned}
\int_{[0,1]^s}\int_{[0,1]^s} f(\mathbf{x})f(\mathbf{y})\psi(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} &= \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\int_{A_i}\int_{A_j}\psi(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y}\\
&\leq \sum_{i=1}^{n}\sum_{j=1}^{n}\alpha_i\alpha_j\int_{A_i}\int_{A_j}d\mathbf{x}d\mathbf{y}\\
&= \int_{[0,1]^s}\int_{[0,1]^s} f(\mathbf{x})f(\mathbf{y})d\mathbf{x}d\mathbf{y},
\end{aligned}$$

which shows that $f \in \mathcal{F}^s$. □

5.2 Expanding $\mathcal{F}^s$

Next, we take a closer look at $\mathcal{F}^s$ and show that it is closed under scalar multiplication and translation. Additionally we can say that it is closed under $\|\cdot\|_1$-limits.

Lemma 5.3. *The set $\mathcal{F}^s$ is closed under translation, scalar multiplication, and taking $\|\cdot\|_1-$limits.*

Proof. Let $\psi(\mathbf{x},\mathbf{y})$ be the joint pdf of a scrambled digital $(0,m,s)$-net. As mentioned above, the variance of the estimator $\hat{\mu}_n$ for $I(f)$ is no larger than the variance of the MC estimator if and only if

$$\int_{[0,1]^s}\int_{[0,1]^s} f(\mathbf{x})f(\mathbf{y})\psi(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} \leq \int_{[0,1]^s}\int_{[0,1]^s} f(\mathbf{x})f(\mathbf{y})d\mathbf{x}d\mathbf{y}. \tag{4}$$

First, by properties of the covariance, it is clear that $\mathcal{F}^s$ is closed under translation and scalar multiplication. Next, in the inequality (4), we are testing $L^1([0,1]^{2s})$ functions against continuous linear functionals. It follows that if $g_n \in L^1([0,1]^{2s})$ is a sequence that converges in norm to $g \in L^1([0,1]^{2s})$ and each g_n satisfies the inequality

$$\int_{[0,1]^s}\int_{[0,1]^s} g_n(\mathbf{x},\mathbf{y})\psi(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y} \leq \int_{[0,1]^s}\int_{[0,1]^s} g_n(\mathbf{x},\mathbf{y})d\mathbf{x}d\mathbf{y},$$

then so does g. The result follows from the observation that if $f_n \in L^1([0,1]^s)$ is a sequence that converges in norm to $f \in L^1([0,1]^s)$ then $f_n(x)f_n(y)$ converges in $L^1([0,1]^{2s})$ to $f(x)f(y)$. □

Next we recall the definition of quasi-monotone functions.

Definition 5.4. Consider a function f defined over $[0,1]^s$. Its *quasi-volume* or *increment* over an interval of the form $A = [\mathbf{a}, \mathbf{b}] = \prod_{j=1}^{s} [a_j, b_j] \subseteq [0,1]^s$ is given by

$$\Delta^{(s)}(f; A) = \sum_{\mathcal{I} \subseteq \{1,\dots,s\}} (-1)^{|\mathcal{I}|} f(\mathbf{a}^{\mathcal{I}}; \mathbf{b}^{-\mathcal{I}}).$$

Definition 5.5. If $\Delta^{(d)}(f; A) \geq 0$ for all closed axis-parallel boxes $A = [\mathbf{a}, \mathbf{b}] \subseteq [0,1]^s$ of arbitrary dimension $1 \leq d \leq s$, then f is said to be *quasi-monotone* or *completely monotone*. (The dimension d of A refers to the value $d = \sum_{j=1}^{s} 1_{a_j < b_j}$.)

As an example of how the previous lemma can be applied, we use [7, Prop.3 & Rmk 8] together with the lemma to prove the following result.

Proposition 5.6. *The set $\mathcal{F}^s$ contains all integrable functions that are*

(i) bounded and

(ii) quasi-monotone

on all closed axis parallel boxes $[\mathbf{a}, \mathbf{b}] \subseteq (0,1)^s$.

Proof. Let $f : [0,1]^s \to \mathbb{R}$ be an integrable function that is bounded and quasi-monotone on every $[\mathbf{a}, \mathbf{b}] \subseteq (0,1)^s$. Applying Lemma 5.3 we only need to show that f is an $L^1([0,1]^s)$ limit of bounded quasi-monotone functions. For each n let $g_n : [0,1]^s \to [1/n, (n-1)/n]^s$ be the function defined on the coordinates of $\mathbf{x} \in [0,1]^s$ by

$$x_j \mapsto \begin{cases} \frac{1}{n} & \text{if } x_j < \frac{1}{n}, \\ x_j & \text{if } \frac{1}{n} \leq x_j \leq \frac{n-1}{n}, \\ \frac{n-1}{n} & \text{if } \frac{n-1}{n} < x_j, \end{cases}$$

and define $f_n : [0,1]^s \to \mathbb{R}$ to be the composition $f \circ g_n$. It is clear that $f_n \to f$ point-wise and f_n is bounded and quasi-monotone on $[0,1]^s$ because f is bounded and quasi-monotone on $[1/n, (n-1)/n]^s$. Now $f_n \to f$ in $L^1([0,1]^s)$ because each f_n and f are integrable. □

Notice that the functions considered in the previous proposition could be unbounded on the boundary of $[0,1]^s$.

The following proposition uses invariance properties of scrambled $(0, m, s)$-nets to further expand $\mathcal{F}^s$.

Proposition 5.7. *The set $\mathcal{F}^s$ contains all functions that can be obtained by precomposing some subset of $\{g_1, \dots, g_s\}$, where*

$$g_j(\mathbf{x}) = (x_1, \dots, x_{j-1}, 1 - x_j, x_{j+1}, \dots, x_s),$$

with a function that is known to be in $\mathcal{F}^s$.

Proof. Since $\gamma_b(x, y) = \gamma_b(1-x, 1-y)$, $\gamma_b^s(\mathbf{x}, \mathbf{y})$ is invariant under precomposition with some subset of $\{g_1 \times g_1, \dots, g_s \times g_s\}$. Therefore if f is a function that satisfies

$$\int_{[0,1]^s} \int_{[0,1]^s} f(\mathbf{x}) f(\mathbf{y}) \psi(\mathbf{x}, \mathbf{y}) d\mathbf{x} d\mathbf{y} \leq \int_{[0,1]^s} \int_{[0,1]^s} f(\mathbf{x}) f(\mathbf{y}) d\mathbf{x} d\mathbf{y}.$$

then so does f precomposed with some subset of $\{g_1, \dots, g_s\}$. □

5.3 Examples

In the context of numerical integration with deterministic digital nets, a common requirement that functions need to fulfill in order for the net to provide a low error is to be of bounded variation in the sense of Hardy and Krause (BVHK) [9]. In short, this refers to functions that are differences of quasi-monotone functions (see [7, Def. 5]). When using scrambled digital nets, this condition does not predict as reliably which functions will lead to estimators of smaller variance than the Monte Carlo method. In this section, we are providing examples to illustrate this point. On one hand we have a first example with an unbounded function for which a scrambled net does better than Monte Carlo; on the other hand for a net of given size m we construct a BVHK function for which the net does worse than Monte Carlo.

First we give an example of a function that satisfies the assumptions of Proposition 5.6.

Example 5.8. When $s \geq 3$ consider

$$f(x_1, \ldots, x_s) = \min\{(1 - x_i)^{-1} : i = 1, \ldots, s\}.$$

This function is constant on regions of the form $\{\mathbf{x} \in [0,1]^s : \min x_i = t\}$ and increases as t increases. It follows that f is the point-wise limit of the sequence of simple functions

$$f_n = \sum_{k=0}^{n} a_{n,k} 1_{[k/n,1]^s},$$

where $a_{n,0} = 1$ and

$$a_{n,k} = f(\tfrac{k}{n}, \ldots, \tfrac{k}{n}) - f(\tfrac{k-1}{n}, \ldots, \tfrac{k-1}{n}) \geq 0,$$

for $1 \leq k \leq n$. Since $1_{[\mathbf{b},\mathbf{1}]}$ is quasi-monotone for all $\mathbf{b} \in [0,1]^s$ and since

(i) $\Delta^{(s)}(g + h; A) = \Delta^{(s)}(g; A) + \Delta^{(s)}(h; A)$ for all functions f and g,

(ii) $\Delta^{(s)}(cg; A) = c\Delta^{(s)}(g; A)$ for all $c \in \mathbb{R}$, and

(iii) $\Delta^{(s)}(\lim_{n\to\infty} f_n; A) = \lim_{n\to\infty} \Delta^{(s)}(f_n; A)$,

we conclude that f is the limit of quasi-monotone functions and is therefore quasi-monotone.

The maximum value of f on $A_n = [\frac{n-1}{n}, 1]^s \setminus [\frac{n}{n+1}, 1]^s$ is $n + 1$, so f is bounded by

$$g(\mathbf{x}) = \sum_{n=1}^{\infty} (n + 1) 1_{A_n}$$

which is integrable because, when $s \geq 3$,

$$\int_{[0,1]^s} g(\mathbf{x}) d\mathbf{x} = \sum_{n=1}^{\infty} (n + 1)\Big(\frac{1}{n^s} - \frac{1}{(n + 1)^s}\Big) < \infty.$$

The following example shows that $\mathcal{F}^s$ is not closed under addition. In particular Proposition 5.2 cannot be extended to include both positive and negative coefficients together. Such functions are closely connected to BVHK functions. More precisely, this example shows that not all BVHK functions are in $\mathcal{F}^s$.

Example 5.9. Fix $s \geq 2$ and $b \geq s$ and let

$$\mathbf{a} = \left(\frac{b^2 + 1}{b^3}, \ldots, \frac{b^2 + 1}{b^3}\right), \ \mathbf{b} = \left(\frac{1}{b}, \ldots, \frac{1}{b}\right), \text{ and } c = \left(\frac{b^2 + 1}{b^2}\right)^s.$$

Using Definition 4.1, set $f_m : [0,1]^s \to \mathbb{R}$ to be

$$f_m = 1_{[\mathbf{0}, b^{-\mathbf{n}}\mathbf{a}]} - c 1_{[\mathbf{0}, b^{-\mathbf{n}}\mathbf{b}]},$$

where $\mathbf{n} = (m-1, 0, \dots, 0)$. Since

$$\int_{[0,1]^s}\int_{[0,1]^s} f_m(\mathbf{x}) f_m(\mathbf{y}) d\mathbf{x} d\mathbf{y} = 0,$$

our goal is to show that

$$\int_{[0,1]^s}\int_{[0,1]^s} f_m(\mathbf{x}) f_m(\mathbf{y}) \psi(\mathbf{x},\mathbf{y}) d\mathbf{x} d\mathbf{y} > 0, \tag{5}$$

where ψ is the joint pdf of a scrambled $(0, m, s)$-net. By defining

$$V_{\boldsymbol{i}}^s(f_m, f_m) = \int_{D_{\boldsymbol{i}}^s} f_m(\mathbf{x}) f_m(\mathbf{y}) d\mathbf{x} d\mathbf{y}, \tag{6}$$

and recalling that the value of the joint pdf ψ on $D_{\boldsymbol{i}}^s$ (which we denote by $\psi_{\boldsymbol{i}}$) is zero when $i \geq m$, the integral in (5) may be written as

$$\sum_{\boldsymbol{i}\in\mathbb{N}^s,\, i<m} V_{\boldsymbol{i}}^s(f_m, f_m)\psi_{\boldsymbol{i}}. \tag{7}$$

Expanding $f_m(\mathbf{x}) f_m(\mathbf{y})$ on the right-hand side of (6) and applying Lemma 4.2 yields the following formula for $V_{\boldsymbol{i}}^s(f_m, f_m)$:

$$\begin{aligned} &V_{\boldsymbol{i}}^s(b^{-\mathbf{n}}\mathbf{a}, b^{-\mathbf{n}}\mathbf{a}) + c^2 V_{\boldsymbol{i}}^s(b^{-\mathbf{n}}\mathbf{b}, b^{-\mathbf{n}}\mathbf{b}) - 2c V_{\boldsymbol{i}}^s(b^{-\mathbf{n}}\mathbf{a}, b^{-\mathbf{n}}\mathbf{b}) \\ =& b^{2-2m}\left(V_{\boldsymbol{i}-\mathbf{n}}^s(\mathbf{a},\mathbf{a}) + c^2 V_{\boldsymbol{i}-\mathbf{n}}^s(\mathbf{b},\mathbf{b}) - 2c V_{\boldsymbol{i}-\mathbf{n}}^s(\mathbf{a},\mathbf{b})\right) \end{aligned}$$

when $\boldsymbol{i} \geq \mathbf{n}$ and $V_{\boldsymbol{i}}^s(f_m, f_m) = 0$ otherwise. It follows that (7) reduces to $b^{2-2m} V_0^s(f_1, f_1)\psi_{\mathbf{n}}$. Since $\psi_{\mathbf{n}} > 0$, (5) will hold if $V_0^s(f_1, f_1) > 0$. Observing that

$$\begin{aligned} R\left(\frac{b^2+1}{b^3}, \frac{b^2+1}{b^3}\right) \cap D_0 &= \left(\left[0, \frac{1}{b}\right) \times \left[\frac{b^2}{b^3}, \frac{b^2+1}{b^3}\right]\right) \cap \left(\left[\frac{b^2}{b^3}, \frac{b^2+1}{b^3}\right] \times \left[0, \frac{1}{b}\right)\right), \\ R\left(\frac{1}{b}, \frac{1}{b}\right) \cap D_0 &= \left(\left[0, \frac{1}{b}\right) \times \left\{\frac{1}{b}\right\}\right) \cup \left(\left\{\frac{1}{b}\right\} \times \left[0, \frac{1}{b}\right)\right), \text{ and} \\ R\left(\frac{b^2+1}{b^3}, \frac{1}{b}\right) \cap D_0 &= \left[\frac{b^2}{b^3}, \frac{b^2+1}{b^3}\right] \times \left[0, \frac{1}{b}\right), \end{aligned}$$

we can bound $V_0^s(f_1, f_1)$ as follows:

$$\begin{aligned} &V_0\left(\frac{b^2+1}{b^3}, \frac{b^2+1}{b^3}\right)^s + c^2 0^s - 2c V_0\left(\frac{b^2+1}{b^3}, \frac{1}{b}\right)^s \\ =&\left(\frac{2}{b^3}\right)^s - 2\left(1 + \frac{1}{b^2}\right)^s \left(\frac{1}{b^3}\right)^s \geq \left(\frac{2}{b^3}\right)^s - 2\left(\frac{5}{4}\right)^s \left(\frac{1}{b^3}\right) \\ =&\frac{8^s - 2\cdot 5^s}{4^s b^3} \geq 0 \end{aligned}$$

when $s \geq 2$.

The previous example describes a BVHK function that is built explicitly to cause $(0, m, s)$-nets of a certain size m to be worse than Monte Carlo. One may think that this is not so problematic because one can show that by using a net of a larger size (with the same base), we can do better than Monte Carlo for this function.

6 Conclusion

In this paper we have proved that scrambled $(0, m, s)$-nets have the property of being NUOD and NLOD and that any scrambled net with $t > 0$ does not have this property. We believe that the tools we have developed to get these results will allow us to explore different paths to generalize these results. In particular, we would like to explore a generalized concept of dependence that considers sets other than the rectangular boxes anchored at the origin or at the opposite corner $(1, \ldots, 1)$ that are used to define the NLOD/NUOD concepts. We would also like to understand better the properties of quasi-monotone functions and provide examples of real-world problems that exhibit this type of behavior. Finally, we also plan to explore how the representation for the covariance term $\mathrm{Cov}(f(\mathbf{U}), f(\mathbf{V}))$ as an integral of the joint pdf associated with a scrambled digital net can be exploited to estimate the variance of estimators based on scrambled nets without having to make use of repeated randomizations.

Acknowledgements

The authors wish to acknowledge the support of the Natural Science and Engineering Research Council (NSERC) of Canada for its financial support via grant # 238959. The first author is also partially supported by the Austrian Science Fund (FWF): Project F5506-N26, which is part of the Special Research Program "Quasi-Monte Carlo Methods: Theory and Applications".

References

[1] C. Aistleitner, F. Pausinger, A.M. Svane, and R.F. Tichy, *On functions of bounded variation*, Mathematical Proceedings of the Cambridge Philosophical Society **162** (2017), 405–418.

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

[3] J. Dick and F. Pillichshammer, *Digital nets and sequences: Discrepancy theory and quasi-monte carlo integration*, Cambridge University Press, UK, 2010.

[4] H. Faure, *Discrépance des suites associées à un système de numération (en dimension s)*, Acta Arithmetica **41** (1982), 337–351.

[5] M. Gerber, *On integration methods based on scrambled nets of arbitrary size*, Journal of Complexity **31** (2015), 798–816.

[6] C. Lemieux, *Monte Carlo and Quasi-Monte Carlo sampling*, Springer Series in Statistics, Springer, New York, 2009.

[7] ———, *Negative dependence, scrambled nets, and variance bounds*, Mathematics of Operations Research **43** (2017), 228–251.

[8] J. Matoušek, *On the L_2-discrepancy for anchored boxes*, Journal of Complexity **14** (1998), 527–556.

[9] H. Niederreiter, *Random number generation and quasi-Monte Carlo methods*, SIAM CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 63, SIAM, Philadelphia, 1992.

[10] A. B. Owen, *Randomly permuted (t, m, s)-nets and (t, s)-sequences*, Monte Carlo and quasi-Monte Carlo methods in scientific computing, 1995, pp. 299–317.

[11] ———, *Monte Carlo variance of scrambled equidistribution quadrature*, SIAM Journal on Numerical Analysis **34** (1997), no. 5, 1884–1910.

[12] ———, *Scrambled net variance for integrals of smooth functions*, Annals of Statistics **25** (1997), no. 4, 1541–1562.

[13] ———, *Scrambling Sobol and Niederreiter-Xing points*, Journal of Complexity **14** (1998), 466–489.

[14] ———, *Variance and discrepancy with alternative scramblings*, ACM Transactions on Modeling and Computer Simulation **13** (2003), 363–378.

[15] I. M. Sobol', *On the distribution of points in a cube and the approximate evaluation of integrals*, USSR Comp. Math. Math. Phys. **7** (1967), 86–112.