Unif. Distrib. Theory
?? (????), no.?, 1–28

ON THE ASYMPTOTIC AVERAGE DIAMETER OF BLOCKS OF UNIFORMLY DISTRIBUTED SEQUENCES AND RELATED RESULTS

Sebastian Heintze[1] — Wolfgang Trutschnig[2]

[1]University of Salzburg, Salzburg, Austria

[2]University of Salzburg, Salzburg, Austria

Abstract. This paper was triggered by recent results on the maximal 'average distance between consecutive points' of uniformly distributed sequences (u.d.s.). Here we address a generalized version of this question, consider pairwise maximal/minimal/total distances in blocks/segments of $d \geq 2$ consecutive points of u.d.s., and derive sharp upper bounds for all three aggregations. Our main idea of proof consists in, firstly, adding degrees of freedom, secondly, translating the resulting problem to a solvable optimization problem over the compact family of d-stochastic measures, and, thirdly, showing that the obtained bounds are also sharp bounds for the original problem.

Communicated by ???

1. Introduction

A sequence $(x_n)_{n\in\mathbb{N}}$ in the unit interval $[0,1]$ is called uniformly distributed if for every interval $[a,b] \subseteq [0,1]$ we have that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[a,b]}(x_i) = b - a = \lambda([a,b])$$

© ???? Mathematical Institute, Slovak Academy of Sciences.
2020 Mathematics Subject Classification: 11K06, 11J71.
Keywords: Uniformly distributed sequences, copula, d-stochastic measure.
The authors gratefully acknowledge the support of the project 'IDA Lab Salzburg' (20102/F2300464-KZP; 20204-WISS/225/348/3-2023).
Licensed under the Creative Commons Attribution-NC-ND 4.0 International Public License.

holds, where $\mathbf{1}_{[a,b]}$ denotes the indicator function of $[a,b]$. Translating to the stochastic perspective, $(x_n)_{n\in\mathbb{N}}$ is uniformly distributed if and only if the empirical measure ϑ_n, given by

$$\vartheta_n = \frac{1}{n}\sum_{i=1}^{n} \delta_{x_i} \in \mathcal{P}([0,1]),$$

where δ_a denotes the Dirac measure at a, converges weakly to the Lebesgue measure λ for $n \to \infty$. In what follows, we will let $\mathcal{U}$ denote the family of all uniformly distributed sequences (u.d.s., for short). For extensive background on u.d.s. we refer to the textbooks [4, 11] and the references therein.

Our contribution is motivated by the following question tackled by Pillichshammer and Steinerberger in [13], where the authors studied, how large the asymptotic 'average distance between consecutive points' of $(x_n)_{n\in\mathbb{N}} \in \mathcal{U}$ can possibly be. Among other things, they proved that

$$\sup_{(x_n)_{n\in\mathbb{N}}\in\mathcal{U}} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} |x_{i+1} - x_i| \leq \frac{1}{2} \tag{1}$$

holds and that the upper bound $\frac{1}{2}$ is attainable, i.e., that there exist u.d.s. such that equation (1) becomes an identity (one example being the van der Corput sequence in base 2). Here we address a generalized version of this problem, consider pairwise maximal/minimal/total distances in blocks/segments of $d \geq 2$ consecutive points, and work with the three (obviously continuous) functions $f_{\min}, f_{\max}, f_{\sum} \colon [0,1]^d \to [0,1]$, defined by

$$f_{\min}(\mathbf{x}) \coloneqq \min_{i,j\in[d],\ i\neq j} |x_i - x_j|\,, \tag{2}$$

$$f_{\max}(\mathbf{x}) \coloneqq \max_{i,j\in[d],\ i\neq j} |x_i - x_j|\,, \tag{3}$$

$$f_{\sum}(\mathbf{x}) \coloneqq \sum_{i,j\in[d]} |x_i - x_j|\,, \tag{4}$$

with $\mathbf{x} = (x_1, \ldots, x_d)$ and $[d] \coloneqq \{1, \ldots, d\}$. Notice that for $d = 2$, up to the factor 2 all three functions boil down to the pairwise distance of consecutive points as considered in inequality (1). Our main objective is to determine

$$B_f^d \coloneqq \sup_{(x_n)_{n\in\mathbb{N}}\in\mathcal{U}} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(x_i, x_{i+1}, \ldots, x_{i+d-1})$$

for each of the three functions $f_{\min}, f_{\max}, f_{\sum}$. In fact, we will prove that

$$B_{f_{\min}}^d = \frac{1}{d}, \quad B_{f_{\max}}^d = \frac{d-1}{d}, \quad B_{f_{\sum}}^d = \frac{d^2-1}{3}$$

for every $d \geq 2$. Doing so, our proof strategy can be subsumed as follows:

i) Considering that for every u.d.s. $(x_n)_{n\in\mathbb{N}}$ and every $j \in \mathbb{N}_0$ the shifted sequence and $(x_{j+n})_{n\in\mathbb{N}}$ is a u.d.s. too, we obviously have that the upper bound B_f^d fulfills

$$B_f^d \leq \sup_{\mathbf{x}^1,\mathbf{x}^2,\ldots,\mathbf{x}^d\in\mathcal{U}} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(x_i^1, x_i^2, \ldots, x_i^d) =: C_f^d. \tag{5}$$

ii) We show that C_f^d can be expressed as

$$C_f^d = \sup_{\vartheta\in\mathcal{P}_{\mathrm{ds}}([0,1]^d)} \int_{[0,1]^d} f \,\mathrm{d}\vartheta, \tag{6}$$

i.e., the maximization problem can be translated to a problem of maximizing integrals over the family $\mathcal{P}_{\mathrm{ds}}([0,1]^d)$ of all d-stochastic measures (see Section 2).

iii) For determining the right-hand side of (6) it suffices to work with (handy) subfamilies of $\mathcal{P}_{\mathrm{ds}}([0,1]^d)$ which are dense (with respect to the topology induced by weak convergence).

iv) We show (the surprising fact) that $B_f^d = C_f^d$ holds.

The remainder of this paper is organized as follows: Section 2 gathers notation and relevant background on d-stochastic measures and copulas that will be used in the sequel. Section 3 justifies the afore-mentioned translation to the d-stochastic setting, proves equation (6) as main result, and then establishes denseness for the family of all so-called shuffles of M. Main objective of Section 4 is to determine C_f^d for all three functions $f_{\min}, f_{\max}, f_{\sum}$. Finally, Section 5 proves $B_f^d = C_f^d$ by constructing u.d.s. either attaining the upper bound B_f^d or approximating B_f^d arbitrarily well.

2. Notation and preliminaries

For every metric space (S, ρ) the Borel σ-field on S will be denoted by $\mathcal{B}(S)$. Throughout the whole paper $d \geq 2$ will denote the dimension, bold symbols will denote vectors or sequences and we will write either $\mathbf{x} = (x_1, \ldots, x_d)$ or $\mathbf{x} = (x_1, x_2, \ldots)$, respectively; from the context it will be clear, whether $\mathbf{x}$ is a vector or a sequence. $[N]$ will denote the set $\{1, \ldots, N\}$ for every $N \in \mathbb{N}$, and $\pi_i : [0,1]^d \to [0,1]$ the projection on the i-th coordinate, i.e., $\pi_i(x_1, \ldots, x_d) = x_i$.

$\mathcal{P}([0,1]^d)$ will denote the family of all probability measures on $\mathcal{B}([0,1]^d)$, $\vartheta^{\pi_i} \in \mathcal{P}([0,1])$ the i-th marginal of ϑ, i.e., the push-forward of ϑ via π_i. A

measure $\vartheta \in \mathcal{P}([0,1]^d)$ is called d-stochastic, if all univariate marginals of ϑ coincide with the Lebesgue measure λ on $[0,1]$; $\mathcal{P}_{\mathrm{ds}}([0,1]^d)$ will denote the family of all d-stochastic measures on $[0,1]^d$. Moreover, $\mathcal{C}_d$ denotes the family of all d-dimensional copulas, i.e., the distribution functions of d-stochastic measures restricted to $[0,1]^d$. Considering

$$C(x_1,\ldots,x_d) = \vartheta\left(\bigtimes_{i=1}^{d}[0,x_i]\right), \qquad (x_1,\ldots,x_d)\in[0,1]^d$$

establishes a one-to-one correspondence between $\mathcal{C}_d$ and $\mathcal{P}_{\mathrm{ds}}([0,1]^d)$. It is well known (see, e.g., [2,12]) that endowing $\mathcal{P}([0,1]^d)$ with the Hutchinson / Wasserstein / Kantorovich-Rubinstein metric ρ_W, defined by

$$\rho_W(\mu,\nu) = \sup\left\{\left|\int_{[0,1]^d} f\,\mathrm{d}\mu - \int_{[0,1]^d} f\,\mathrm{d}\nu\right| : \quad f\colon [0,1]^d\to\mathbb{R} \text{ is 1-Lipschitz}\right\},$$

yields a compact metric space $(\mathcal{P}([0,1]^d),\rho_W)$, which has $\mathcal{P}_{\mathrm{ds}}([0,1]^d)$ as closed (hence compact) convex subset. Moreover (again see [2,12]), ρ_W is a metrization of weak convergence in $\mathcal{P}([0,1]^d)$, i.e., a sequence $(\vartheta_n)_{n\in\mathbb{N}}$ in $\mathcal{P}([0,1]^d)$ converges weakly to ϑ if and only if for every continuous function $f\colon[0,1]^d\to\mathbb{R}$ we have that

$$\lim_{n\to\infty}\int_{[0,1]^d} f(\mathbf{x})\,\mathrm{d}\vartheta_n(\mathbf{x}) = \int_{[0,1]^d} f(\mathbf{x})\,\mathrm{d}\vartheta(\mathbf{x}).$$

In the sequel we will also use the fact that weak convergence in $\mathcal{P}_{\mathrm{ds}}([0,1]^d)$ is equivalent to uniform convergence of the corresponding copulas. In other words, letting $C, C_1, C_2, \ldots$ denote the d-dimensional copulas corresponding to the d-stochastic measures $\vartheta, \vartheta_1, \vartheta_2, \ldots$, the following two assertions are equivalent (see [5]):

i) The sequence $(\vartheta_n)_{n\in\mathbb{N}}$ converges weakly to ϑ.

ii) $\lim\limits_{n\to\infty} d_\infty(C_n, C) = \lim\limits_{n\to\infty} \max\left\{|C_n(\mathbf{x}) - C(\mathbf{x})| : \mathbf{x}\in[0,1]^d\right\} = 0.$

A *Markov kernel* from $\mathbb{R}$ to $\mathcal{B}(\mathbb{R}^{d-1})$ is a mapping $K\colon \mathbb{R}\times\mathcal{B}(\mathbb{R}^{d-1})\to[0,1]$ such that $x\mapsto K(x,B)$ is measurable for every fixed $B\in\mathcal{B}(\mathbb{R}^{d-1})$ and $B\mapsto K(x,B)$ is a probability measure for every fixed $x\in\mathbb{R}$. Suppose that $\mathbf{Y}\colon\Omega\to\mathbb{R}^{d-1}$ and $X\colon\Omega\to\mathbb{R}$ are random vectors/variables on a joint probability space $(\Omega,\mathcal{A},\mathbb{P})$. Then a Markov kernel $K\colon\mathbb{R}\times\mathcal{B}(\mathbb{R}^{d-1})\to[0,1]$ is called regular conditional distribution of $\mathbf{Y}$ given X if for every $B\in\mathcal{B}(\mathbb{R}^{d-1})$

$$K(X(\omega),B) = \mathbb{E}(\mathbf{1}_B\circ\mathbf{Y}\mid X)(\omega)$$

holds $\mathbb{P}$-a.s.. It is well known that for each random vector $(X,\mathbf{Y})$ a regular conditional distribution $K(\cdot,\cdot)$ of $\mathbf{Y}$ given X exists, that $K(\cdot,\cdot)$ is unique $\mathbb{P}^X$-a.s. (i.e.,

unique for $\mathbb{P}^X$-almost all $x \in \mathbb{R}$) and that $K(\cdot,\cdot)$ only depends on the distribution $\mathbb{P}^{(X,\mathbf{Y})}$ of the vector $(X,\mathbf{Y})$. Hence, if $(X,\mathbf{Y})$ has distribution $\vartheta \in \mathcal{P}_{\mathrm{ds}}([0,1]^d)$, then we will denote (a version of) the regular conditional distribution of $\mathbf{Y}$ given X by $K_\vartheta(\cdot,\cdot)$, view it directly as function from $[0,1] \times \mathcal{B}([0,1]^{d-1})$ to $[0,1]$, and refer to $K_\vartheta(\cdot,\cdot)$ simply as regular conditional distribution or Markov kernel of ϑ. Note that for every $\vartheta \in \mathcal{P}_{\mathrm{ds}}([0,1]^d)$, its conditional regular distribution $K_\vartheta(\cdot,\cdot)$, and a Borel set $F \in \mathcal{B}([0,1]^d)$ we have (with $F_x = \{\mathbf{y} \in [0,1]^{d-1} : (x,\mathbf{y}) \in F\}$ denoting the x-cut of F for every $x \in [0,1]$)

$$\int_{[0,1]} K_\vartheta(x, F_x)\,\mathrm{d}\lambda(x) = \vartheta(F).$$

In particular,

$$\int_{[0,1]} K_\vartheta(x, E)\,\mathrm{d}\lambda(x) = \lambda(E_{i_0})$$

holds for $E = \bigtimes_{i=1}^{d-1} E_i$ with $E_i = [0,1]$ for all $i \neq i_0$. On the other hand, every Markov kernel $K\colon [0,1] \times \mathcal{B}([0,1]^{d-1}) \to [0,1]$ fulfilling equation (2) is the regular conditional distribution of a d-stochastic measure $\vartheta \in \mathcal{P}_{\mathrm{ds}}([0,1]^d)$. More generally, for every continuous or ϑ-integrable function $f\colon [0,1]^d \to \mathbb{R}$ the following identity (to which we will loosely refer as 'disintegration' in the sequel) holds:

$$\int_{[0,1]^d} f(\mathbf{x})\,\mathrm{d}\vartheta(\mathbf{x}) = \int_{[0,1]} \int_{[0,1]^{d-1}} f(x_1, x_2, \ldots, x_d) K_\vartheta(x_1, \mathrm{d}(x_2, \ldots, x_d))\,\mathrm{d}\lambda(x_1)$$

For more details and properties of conditional expectation, regular conditional distributions and disintegration we refer to [9, 10] as well as to [1, 3].

3. Translating the maximization to the d-stochastic setting

The bivariate version of the following general result was already used in [13]. For the sake of completeness and clarity we state the general d-dimensional version and prove it using properties of weak convergence of measures and compactness (x_i^j denotes the i-th element of the sequence $\mathbf{x}^j$).

Lemma 3.1. *Suppose that $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^d$ are u.d.s. and that $f\colon [0,1]^d \to \mathbb{R}$ is continuous. Then for every accumulation point a of the sequence $(a_n)_{n\in\mathbb{N}}$, given by*

$$a_n \coloneqq \frac{1}{n} \sum_{i=1}^{n} f(x_i^1, x_i^2, \ldots, x_i^d),$$

there exists some d-stochastic measure $\vartheta \in \mathcal{P}_{ds}([0,1]^d)$ *with*

$$a = \int_{[0,1]^d} f \,\mathrm{d}\vartheta.$$

P r o o f. Continuity of f and compactness of $[0,1]^d$ imply that all accumulation points of the sequence $(a_n)_{n\in\mathbb{N}}$ are contained in a bounded interval. Suppose that

$$a = \lim_{j\to\infty} \frac{1}{n_j} \sum_{i=1}^{n_j} f(x_i^1, x_i^2, \ldots, x_i^d)$$

for some strictly increasing sequence $(n_j)_{j\in\mathbb{N}}$. Let $\delta_{x_i^1, x_i^2, \ldots, x_i^d}$ denote the Dirac measure at $(x_i^1, x_i^2, \ldots, x_i^d)$ and set

$$\vartheta_{n_j} := \frac{1}{n_j} \sum_{i=1}^{n_j} \delta_{x_i^1, x_i^2, \ldots, x_i^d} \in \mathcal{P}([0,1]^d), \qquad j \in \mathbb{N}.$$

Then $(\vartheta_{n_j})_{j\in\mathbb{N}}$ is a sequence in $\mathcal{P}([0,1]^d)$, which by compactness has a weakly convergent subsequence $(\vartheta_{n_{j_l}})_{l\in\mathbb{N}}$ with weak limit $\vartheta \in \mathcal{P}([0,1]^d)$. This directly yields

$$\begin{aligned} a &= \lim_{l\to\infty} \frac{1}{n_{j_l}} \sum_{i=1}^{n_{j_l}} f(x_i^1, x_i^2, \ldots, x_i^d) \\ &= \lim_{l\to\infty} \int_{[0,1]^d} f(z_1, \ldots, z_d) \,\mathrm{d}\vartheta_{n_{j_l}}(z_1, \ldots, z_d) \\ &= \int_{[0,1]^d} f(z_1, \ldots, z_d) \,\mathrm{d}\vartheta(z_1, \ldots, z_d), \end{aligned}$$

so it remains to show that ϑ is d-stochastic. Applying Portmanteau theorem (see [2]) for the rectangles $E_1 = [0,x] \times [0,1]^{d-1}$ and $E_2 = [0, x+\Delta) \times [0,1]^{d-1}$ with $x \in [0,1)$ and $\Delta \in (0,1]$ fulfilling $x + \Delta \in [0,1]$ yields

$$\begin{aligned} \vartheta(E_1) &\geq \limsup_{l\to\infty} \vartheta_{n_{j_l}}([0,x] \times [0,1]^{d-1}) \\ &= \limsup_{l\to\infty} \frac{1}{n_{j_l}} \sum_{i=1}^{n_{j_l}} \delta_{x_i^1}([0,x]) = \lambda([0,x]) = x \end{aligned}$$

as well as

$$\begin{aligned} \vartheta(E_2) &\leq \liminf_{l\to\infty} \vartheta_{n_{j_l}}([0, x+\Delta) \times [0,1]^{d-1}) \\ &= \liminf_{l\to\infty} \frac{1}{n_{j_l}} \sum_{i=1}^{n_{j_l}} \delta_{x_i^1}([0, x+\Delta)) \end{aligned}$$

$$= \lambda([0, x + \Delta)) = x + \Delta.$$

This implies

$$x \leq \vartheta\left([0, x] \times [0, 1]^{d-1}\right) \leq \vartheta\left([0, x + \Delta) \times [0, 1]^{d-1}\right) \leq x + \Delta,$$

so, using the fact that probability measures are continuous from above and considering $\Delta \to 0$, we conclude that $x = \vartheta\left([0, x] \times [0, 1]^{d-1}\right)$. As $x \in [0, 1)$ was arbitrary, it follows that the first marginal ϑ^{π_1} of ϑ coincides with λ. Since the previous arguments also work for any other coordinate, it altogether follows that ϑ is d-stochastic. □

Lemma 3.2. *For every $d \geq 2$ the following identity holds:*

$$C_f^d = \sup_{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^d \in \mathcal{U}} \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i^1, x_i^2, \ldots, x_i^d) = \sup_{\vartheta \in \mathcal{P}_{ds}([0,1]^d)} \int_{[0,1]^d} f \, \mathrm{d}\vartheta.$$

P r o o f. As a direct consequence of the previous lemma we already have that

$$\sup_{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^d \in \mathcal{U}} \limsup_{n \to \infty} \frac{1}{n} \sum_{i=1}^{n} f(x_i^1, x_i^2, \ldots, x_i^d) \leq \sup_{\vartheta \in \mathcal{P}_{\mathrm{ds}}([0,1]^d)} \int_{[0,1]^d} f \, \mathrm{d}\vartheta.$$

In order to show the reverse inequality we proceed as follows: Suppose that the d-dimensional random vector $\mathbf{X}$ has distribution ϑ and suppose that $\mathbf{X}^1, \mathbf{X}^2, \ldots$ is an independent sample from $\mathbf{X}$. Then according to the (multivariate) Glivenko-Cantelli theorem (see, e.g., [14]) the sequence $(\vartheta_n)_{n \in \mathbb{N}}$ of empirical measures

$$\vartheta_n \coloneqq \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i^1, X_i^2, \ldots, X_i^d} \in \mathcal{P}([0, 1]^d)$$

converges weakly to ϑ with probability one. Hence, considering

$$\int_{[0,1]^d} f \, \mathrm{d}\vartheta_n = \frac{1}{n} \sum_{i=1}^{n} f(X_i^1, X_i^2, \ldots, X_i^d)$$

directly yields the desired result. □

Using the fact that for continuous f the mapping $\vartheta \mapsto \int_{[0,1]^d} f \, \mathrm{d}\vartheta$ is continuous on $\mathcal{P}_{\mathrm{ds}}([0, 1]^d)$ with respect to weak convergence, for determining the supremum in C_f^d it suffices to calculate the supremum over dense subfamilies of $\mathcal{P}_{\mathrm{ds}}([0, 1]^d)$. One of the most handy dense subsets of $\mathcal{P}_{\mathrm{ds}}([0, 1]^d)$ is that of equidistant multivariate shuffles of M (see [5] and the references therein). We will first recall the concept of equidistant shuffles of M and then formally prove denseness of this class. Doing so, we will write $I_i^N \coloneqq [\frac{i-1}{N}, \frac{i}{N})$ for every $N \in \mathbb{N}$ and every $i \in [N]$.

Definition 3.3. An element $\vartheta \in \mathcal{P}_{\mathrm{ds}}([0,1]^d)$ is called an equidistant shuffle of M with resolution $N \in \mathbb{N}$ if and only if there exist $d-1$ permutations $\sigma_2, \ldots, \sigma_d$ of $[N]$ such that for the $d-1$ transformations $h_2, \ldots, h_d \colon [0,1] \to [0,1]$, defined by

$$h_i(x) \coloneqq \sum_{j=1}^{N} \left(x - \frac{j-1}{N} + \frac{\sigma_i(j)-1}{N} \right) \mathbf{1}_{I_j^N}(x),$$

the following property holds: ϑ is the distribution of the random vector

$$(X_1, h_2(X_1), h_3(X_1), \ldots, h_d(X_1)) \tag{7}$$

where X_1 is a uniformly $[0,1]$-distributed random variable.

Notice that each of the transformations h_i preserves the Lebesgue measure λ, i.e., the push-forward λ^{h_i} of λ via h_i coincides with λ. It is therefore straightforward to verify that the distribution of the vector (7) is a d-stochastic measure. Moreover, it might seem natural to also transform the variable X_1 in equation (7) via a λ-preserving transformation h_1 – since $h_1 \circ X_1$, however, is uniformly distributed on $[0,1]$ one would still end up with a shuffle of M according to Definition 3.3. For generalized shuffles of copulas (via general λ-preserving transformations) we refer to [5, 6] and the references therein.

In the sequel we will let $\mathcal{S}_d^N$ denote the family of all shuffles of M with resolution N and refer to

$$\mathcal{S}_d^\infty \coloneqq \bigcup_{N=1}^{\infty} \mathcal{S}_d^N \subseteq \mathcal{P}_{\mathrm{ds}}([0,1]^d)$$

as the family of all equidistant shuffles of M.

Remark 3.4. Notice that for every $N_0 \in \mathbb{N}$ we have

$$\mathcal{S}_d^\infty = \bigcup_{N=N_0}^{\infty} \mathcal{S}_d^N$$

since obviously every element of $\mathcal{S}_d^N$ also is an element of $\mathcal{S}_d^{kN}$ for every $k \geq 2$. We will use this fact in the next section and only work with the case $N \geq d$ or even $N = kd$ for some $k \in \mathbb{N}$.

Lemma 3.5. *$\mathcal{S}_d^\infty$ is dense in the compact metric space $(\mathcal{P}_{ds}([0,1]^d), \rho_W)$.*

Proof. It has already been mentioned that weak convergence in $\mathcal{P}_{\mathrm{ds}}([0,1]^d)$ is equivalent to uniform convergence of the corresponding copulas. Let $\vartheta \in \mathcal{P}_{\mathrm{ds}}([0,1]^d)$ be arbitrary but fixed and let C_ϑ denote the corresponding copula. Then, according to [8] (also see [7, Lemma 6.1]), the so-called empirical (multilinear) copula $\hat{C}_n$ (i.e., the copula induced by a sample from C_ϑ via pseudo-ranks

and multilinear interpolation) converges uniformly to C_ϑ with probability one. For every n, however, it is straightforward to verify that there exists a copula A_n corresponding to a d-stochastic measure $\nu_n \in \mathcal{S}_d^n$ fulfilling $d_\infty(A_n, \hat{C}_n) \leq \frac{1}{n^{d-1}}$ (in fact, one may simply consider the M-interpolation instead of the multilinear one corresponding to Π). This directly yields that the sequence $(\nu_n)_{n\in\mathbb{N}}$ in $\mathcal{S}_d^\infty$ converges weakly to ϑ and the proof is complete. □

4. Calculating the upper bounds C_f^d for the d-stochastic setting

In this section we will derive our main results regarding the d-stochastic setting and determine

$$C_f^d = \sup_{\vartheta\in\mathcal{P}_{\mathrm{ds}}([0,1]^d)} \int_{[0,1]^d} f \,\mathrm{d}\vartheta$$

for each of the three functions $f_{\min}, f_{\max}, f_{\sum}$. We start with the first two functions and work with equidistant shuffles.

Theorem 4.1. *For the two functions $f_{\min}$ and $f_{\max}$, defined according to equations* (2) *and* (3)*, respectively, and for every $d \geq 2$ we have the identities*

$$C_{f_{\min}}^d = \frac{1}{d} \qquad \text{and} \qquad C_{f_{\max}}^d = \frac{d-1}{d}.$$

Proof. Before going into details we sketch the structure of the proof. (i) Our first step consists in proving that the stated maximal value is an upper bound for the integral. We achieve this goal by using denseness of the family of shuffles, which, in turn, transforms the maximization problem to the realm of permutations. We will proceed with step by step maximization and illustrate the arguments/steps for a small example graphically. (ii) The second step is to provide an example attaining the bound.

(i) By Lemma 3.5, it suffices to prove the upper bounds for every $\vartheta \in \mathcal{S}_d^\infty$. Therefore, let $d < N \in \mathbb{N}$ as well as $\vartheta \in \mathcal{S}_d^N$ be arbitrary but fixed and denote the (piecewise linear) λ-preserving transformations according to Definition 3.3 by $h_2, \ldots, h_d$. Using disintegration we have

$$\begin{aligned}\int_{[0,1]^d} f(\mathbf{x}) \,\mathrm{d}\vartheta(\mathbf{x}) &= \int_{[0,1]} \int_{[0,1]^{d-1}} f(\mathbf{x})\, K_\vartheta(x_1, \mathrm{d}(x_2, \ldots, x_d)) \,\mathrm{d}\lambda(x_1) \\ &= \int_{[0,1]} f(x_1, h_2(x_1), \ldots, h_d(x_1)) \,\mathrm{d}\lambda(x_1).\end{aligned}$$

The latter expression can be further simplified for $f_{\min}$ and $f_{\max}$: Each transformation h_i with $i \geq 2$ corresponds to a unique permutation σ_i of the set $[N]$ and vice versa. In order to simplify notation we will set $h_1 = \mathrm{id}_{[0,1]}$ as well as $\sigma_1 = \mathrm{id}_{[N]}$. Returning to the integral, for $\mathrm{symb} \in \{\min, \max\}$ this yields

$$\begin{aligned}\int_{[0,1]^d} f_{\mathrm{symb}}(\mathbf{x})\,\mathrm{d}\vartheta(\mathbf{x}) &= \int_{[0,1]} f_{\mathrm{symb}}(x_1, h_2(x_1), \ldots, h_d(x_1))\,\mathrm{d}\lambda(x_1)\\ &= \int_{[0,1]} \operatorname*{symb}_{i,j\in[d],\ i\neq j} |h_i(x_1) - h_j(x_1)|\,\mathrm{d}\lambda(x_1)\\ &= \frac{1}{N}\sum_{n=1}^{N} \operatorname*{symb}_{i,j\in[d],\ i\neq j} \frac{|\sigma_i(n) - \sigma_j(n)|}{N}\\ &= \frac{1}{N^2}\sum_{n=1}^{N} \operatorname*{symb}_{i,j\in[d],\ i\neq j} |\sigma_i(n) - \sigma_j(n)|\,.\end{aligned}$$

(i-1) We now consider $f_{\max}$ in detail and derive an upper bound for

$$\sum_{n=1}^{N} \max_{i,j\in[d],\ i\neq j} |\sigma_i(n) - \sigma_j(n)|\,. \tag{8}$$

Making it easier to follow the subsequent arguments, one may view permutations of $[N]$ as checkerboards with exactly one black field in each row and each column, respectively (see Figure 1).

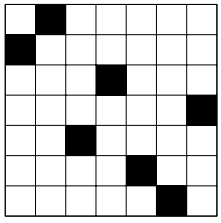

FIGURE 1. Example of a (checkerboard representation of a) permutation of $[N]$ with $N = 7$.

Equation (8) involves exactly d permutations. When overlaying them in one picture, each row can contain at most d different black squares (note that they could overlap and therefore we have 'at most' instead of 'exactly'). We are interested in maximizing the sum over the distances of the highest and lowest black block of each column. For addressing this, it is sufficient to depict only

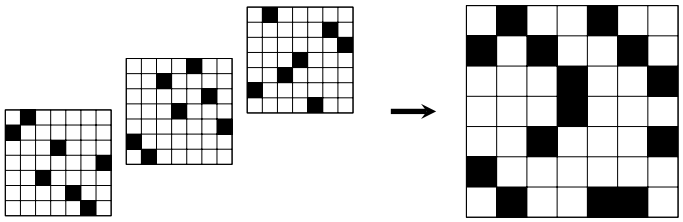

FIGURE 2. Exemplary aggregation of d permutations of $[N]$ for the case $d = 3, N = 7$.

the highest and lowest (black) block within each column. We will illustrate the subsequent arguments graphically for $d = 3$ and $N = 7$.

We start with our d arbitrary permutations (see Figure 2) and first look at the highest row. If there are less than d different black squares in this row, then we move the highest black block of some column(s) not having a black square in the highest row up to the highest row until there are exactly d different black squares in the highest row or until there is no highest black block in a lower row remaining which could be moved up (see Figure 3). Note that there cannot be more than d different black blocks in the highest row and thus we would never move highest blocks down instead of up.

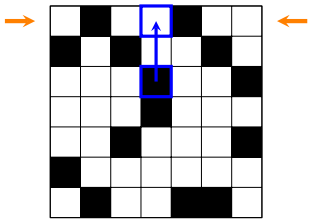

FIGURE 3. Establishing exactly d squares in the highest row.

This procedure of moving black blocks upwards only increases the sum of distances between the highest and lowest black block in each column and therefore only increases the sum of interest (8). Next we take a look at the second to highest row. If there are less than d different black columns in this row, then we move the highest black block of some column(s), where it currently is located in a lower row, up to the currently handled row until there are exactly d different black columns in the currently handled row or until there is no highest black block in a lower row remaining which could be moved up (see Figure 4).

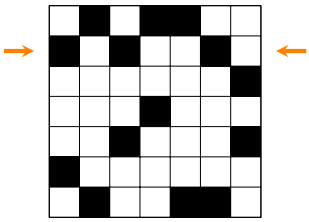

FIGURE 4. Exactly $d = 3$ black squares in the second row.

Again this procedure of moving black blocks only increases the sum of distances between the highest and lowest black block in each column. We proceed in this row-wise manner until there are no highest black blocks in a row below the one currently handled left (see Figure 5).

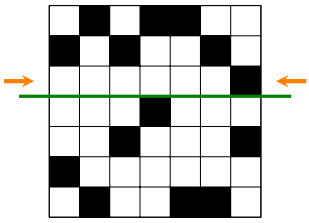

FIGURE 5. Finishing the rearrangment of the highest black blocks.

At this point we stop looking at the highest block in the columns and switch to looking at the lowest black block within each column. Here we proceed analogously to the previous steps but move the blocks down instead of up. Starting with the lowest row and jumping up line by line, we check for the current row, whether there are less than d different black blocks. In that case we move the lowest block of some columns where it currently is located in a higher row down to the currently handled row until there are exactly d different black columns in the currently handled row or until there is no lowest black block in a higher row remaining which could be moved down. The procedure ends when there are no lowest black blocks in a row above the currently handled one left. As before, this procedure of moving black blocks only increases the sum of distances between the highest and lowest black block in each column (see Figure 6).

This procedure yields the situation in which the row containing the lowest one of the highest black blocks of each column is *not below* the row containing the highest one of the lowest black blocks of each column; depending on d and

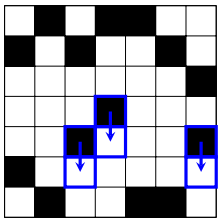

FIGURE 6. Optimize the lowest black blocks.

N, it is either above or they are equal. For seeing this, note that the highest (or lowest, respectively) blocks of the columns, after the before described tuning, will range over $\lceil \frac{N}{d} \rceil$ rows (see Figure 7).

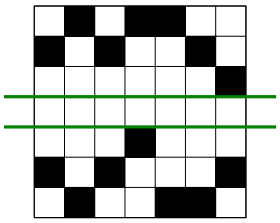

FIGURE 7. Separation of highest and lowest black blocks.

Starting with a constellation of the afore-mentioned type we now calculate the sum of distances between the highest and lowest black block in each column, which yields an upper bound for the sum in expression (8). If in this configuration we exchange the highest black blocks from two columns, then the sum of distances between the highest and lowest black block in each column will stay constant. Namely, the distance in one column increases by exactly the same ammount as the distance in the other column decreases. Recall that we only consider the highest and lowest black block of each column and that they are somewhat separated. Moreover, this exchanging property analogously holds for the lowest black blocks. Together they justify to sort the blocks in such a way that the highest black blocks are in decreasing, and the lowest black blocks are in increasing order (see Figure 8).

In this situation, however, it is now straightforward to compute the sum S of distances, which yields an upper bound as desired. Writing $N = dK + R$ with

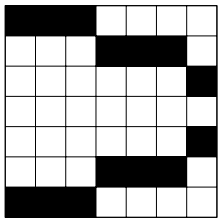

FIGURE 8. Reordering of the highest and lowest black blocks.

$0 \leq R < d$, we get

$$\begin{aligned}
\sum_{n=1}^{N} \max_{i,j\in[d],\ i\neq j} |\sigma_i(n) - \sigma_j(n)| &\leq S \\
&= d\sum_{\ell=1}^{K}(N-(2\ell-1)) + R(N-(2K+1)) \\
&= dKN - d\sum_{\ell=1}^{K}(2\ell-1) + RN - R(2K+1) \\
&= N^2 - dK^2 - KR - KR - R \\
&= N^2 - KN - KR - R \\
&= N((d-1)K + R) - KR - R \\
&= N\cdot\frac{d-1}{d}N + N\cdot\frac{1}{d}R - KR - R \\
&= \frac{d-1}{d}N^2 - R\left(1-\frac{R}{d}\right) \\
&\leq \frac{d-1}{d}N^2.
\end{aligned}$$

Returning to the original integral, this altogether yields

$$\int_{[0,1]^d} f_{\max}(\mathbf{x})\,\mathrm{d}\vartheta(\mathbf{x}) = \frac{1}{N^2}\sum_{n=1}^{N} \max_{i,j\in[d],\ i\neq j} |\sigma_i(n) - \sigma_j(n)| \leq \frac{d-1}{d} \tag{9}$$

which completes the proof for the upper bound concerning $f_{\max}$.

(i-2) Having established this upper bound helps us to determine an upper bound for the corresponding integral over the function $f_{\min}$. We now aim at

bounding

$$\sum_{n=1}^{N} \min_{i,j\in[d],\ i\neq j} |\sigma_i(n) - \sigma_j(n)|$$

from above. As before, we will argue with shifting 'black blocks'. As soon as the highest and lowest black block for a column is fixed, the minimum distance of two blocks in this column is at most $\frac{1}{d-1}$ times the distance of the highest and lowest black block in this column (equidistant placing). Even if this cannot always be realized, it still yields a numerical upper bound. Summing over all columns yields

$$\sum_{n=1}^{N} \min_{i,j\in[d],\ i\neq j} |\sigma_i(n) - \sigma_j(n)| \leq \frac{1}{d-1} \sum_{n=1}^{N} \max_{i,j\in[d],\ i\neq j} |\sigma_i(n) - \sigma_j(n)| \,.$$

Therefore, using (9), we end with

$$\int_{[0,1]^d} f_{\min}(\mathbf{x}) \,\mathrm{d}\vartheta(\mathbf{x}) = \frac{1}{N^2} \sum_{n=1}^{N} \min_{i,j\in[d],\ i\neq j} |\sigma_i(n) - \sigma_j(n)| \leq \frac{1}{d}$$

which completes the proof of the upper bound for $f_{\min}$.

(ii) The last step consists in providing examples showing that the established bounds are already sharp. For this we consider (the rotations)

$$h_r(t) \coloneqq t + \frac{r-1}{d} - \left\lfloor t + \frac{r-1}{d} \right\rfloor \tag{10}$$

for $r \in \{1, \ldots, d\}$. Plotting these functions jointly in one coordinate system clarifies the reason for choosing shuffles of this type – the bundle of functions h_r describes equidistant shifting/rotating (see Figure 9). Letting $\vartheta^* \in \mathcal{P}_{\mathrm{ds}}([0,1]^d)$

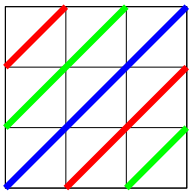

FIGURE 9. The rotations considered in step (ii) for the special case $d = 3$.

denote the corresponding d-stochastic measure, a straightforward calculation yields

$$\int_{[0,1]^d} f_{\min}(\mathbf{x}) \,\mathrm{d}\vartheta^*(\mathbf{x}) = \int_{[0,1]} f_{\min}(x_1, h_2(x_1), \ldots, h_d(x_1)) \,\mathrm{d}\lambda(x_1) = \frac{1}{d}$$

as well as

$$\int_{[0,1]^d} f_{\max}(\mathbf{x})\,\mathrm{d}\vartheta^*(\mathbf{x}) = \int_{[0,1]} f_{\max}(x_1, h_2(x_1), \ldots, h_d(x_1))\,\mathrm{d}\lambda(x_1) = \frac{d-1}{d}.$$

Thus, the upper bounds are attainable and the proof is complete. □

Figure 10 depicts a sample from ϑ^* from the previous proof as well as its bivariate marginal samples.

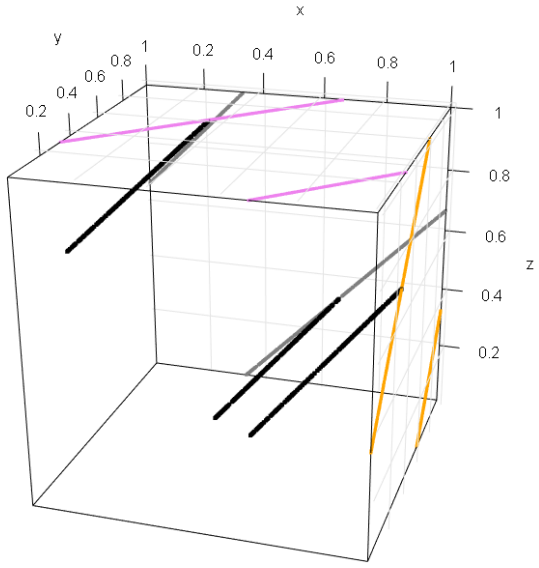

FIGURE 10. Sample (black) of size 1.000 of the 3-stochastic measure ϑ^* attaining the upper bound $\frac{1}{3}$ for $f_{\min}$ and $\frac{2}{3}$ for $f_{\max}$ (as considered in step (ii) in the proof of Theorem 4.1). The gray/orange/pink points denote the bivariate marginal samples.

Notice that the above proof also shows that the maximal value for the integral can in case of equidistant shuffles only be attained if the resolution N is a multiple of the dimension d. When proving the following theorem concerning the function $f_{\sum}$, we have to use a slightly different proof strategy due to higher complexity of the involved function.

THEOREM 4.2. *For the function $f_{\sum}$ as defined in (4) and every $d \geq 2$ we have*

$$C^d_{f_{\sum}} = \frac{d^2-1}{3}.$$

P r o o f. As before we start with sketching the proof structure: (i) Similar to the proof of Theorem 4.1 we will use denseness of shuffles to translate the problem to the language of permutations. Here, once more we will step by step maximize the function of interest to establish an upper bound. (ii) The obtained maximal configuration then turns out to be induced by a well known shuffle, which completes the proof.

(i) Using continuity of $f_{\sum}$, according to Lemma 3.5 it suffices to prove the upper bound for every $\vartheta \in \mathcal{S}_d^\infty$. According to Remark 3.4, without loss of generality we may consider the case $N = d\ell$ for some integer ℓ. Let $\vartheta \in \mathcal{S}_d^N$ be arbitrary but fixed and again let $h_2, \dots, h_d$ denote the corresponding λ-preserving transformations. Analogously to the proof of Theorem 4.1, we then get

$$\begin{aligned}\int_{[0,1]^d} f_{\sum}(\mathbf{x})\,\mathrm{d}\vartheta(\mathbf{x}) &= \int_{[0,1]} f_{\sum}(x_1, h_2(x_1), \dots, h_d(x_1))\,\mathrm{d}\lambda(x_1)\\ &= \frac{1}{N^2}\sum_{n=1}^{N}\sum_{i,j\in[d]} |\sigma_i(n) - \sigma_j(n)|\,,\end{aligned}$$

where the permutations σ_i as above correspond to the transformations h_i (to keep notation simple we will again write $h_1 = \mathrm{id}_{[0,1]}$ as well as $\sigma_1 = \mathrm{id}_{[N]}$). We want to derive an upper bound for

$$\sum_{n=1}^{N}\sum_{i,j\in[d]} |\sigma_i(n) - \sigma_j(n)|$$

and proceed as follows again working with 'black blocks'. For the illustration of our arguments we consider the special case $d = 4$ and $\ell = 2$. In the same way as for the function $f_{\max}$ inside the proof of Theorem 4.1 we first overlay all d permutations in one joint picture. Contrary to before, however, we are now interested in all differences, and therefore keep all the blocks, not just the highest and lowest one. It is likely that there are some overlaps. To indicate overlaps we write the corresponding numbers into the black blocks (see Figure 11).

We partition the grid into d groups of ℓ adjacent rows, i.e., the first ℓ rows form a group, the next ℓ rows another one, and so on. Next we introduce the notion of the 'level' of a black block. All blocks in the highest group as well as those in the lowest group have level 1. Blocks located in the second to highest or second to lowest group have level 2, and so on. Rows that already have a (lower) level number are not releveled (see Figure 12).

Initially, no group is locked. We start with optimizing/shifting the level 1 black blocks and start with those located in the upper group. If there is exactly one of them in each column, then we have nothing to modify und consider this

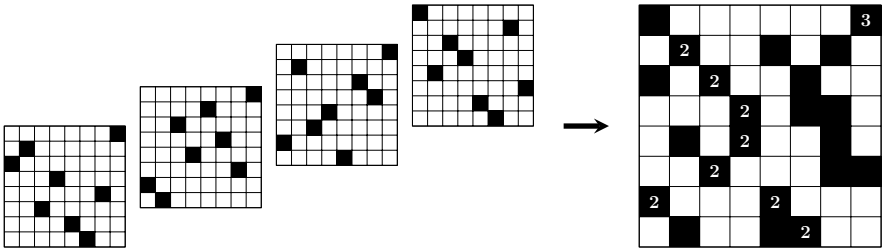

FIGURE 11. Aggregation of $d = 4$ permutations with some overlaps.

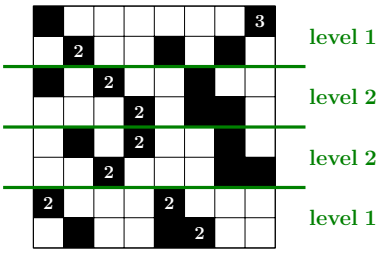

FIGURE 12. Levels of black blocks.

group as locked. Otherwise there is an overlapping or there is one column in which black blocks are at least in two different rows. In case of an overlapping, we move a black block from a field with multiple blocks to the same row in a column where currently no black block appears in the upper level 1 group. In case of a column with black blocks in different rows of the upper level 1 group, we move one of them to the same row in a column where currently no black block appears in the upper level 1 group. In order to keep the number of blocks per column constant at the value d, we have to move back (i.e., from the column whose block number was just increased to the column whose block number was just reduced) a black block in another row. For doing this take into account that temporarily we created a column with more than d black blocks. Hence there must be another, no already locked group, with a black field in this column; take the highest such black block for moving back (not the just added one). This procedure cannot decrease the sum of distances due to the following fact: Inside the column where we first added a block, in fact the highest block was moved up. Thus the distances from this highest block to each of the $d - 1$ other blocks in this column increased by δ, the row difference between the new and old position of the highest block. All other distances in this column remain

unchanged. Inside the other modified column each of the $d-1$ distances involving the modified block decrease at most by δ (due to the triangle inequality), whereas the other ones remain unchanged. Thus the sum over all cannot decrease. We repeat the previous steps until there is exactly one black block in each column of the upper level 1 group (see Figure 13). Then we lock this group.

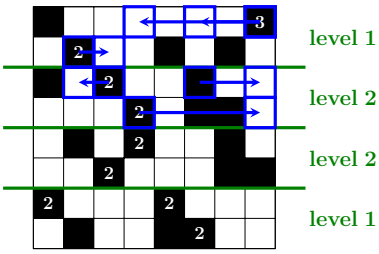

FIGURE 13. Optimizing the upper level 1 group.

We proceed with considering the level 1 black blocks in the lower group. If there is exactly one of them in each column, then we have nothing to modify and consider this group as locked. Otherwise there is again an overlapping or there is one column with black blocks in at least two different rows. In case of an overlapping, we move a black block from a field with multiple blocks to the same row in a column where currently no black block appears in the lower level 1 group. In case of a column with black blocks in different rows of the lower level 1 group we move one of them to the same row in a column where currently no black block appears in the lower level 1 group. In order to keep the number of blocks per column constant at the value d, we again have to move back a black block to another row. As before, we temporarily created a column with more than d black blocks. Hence there must be another group, which is not already locked, with a black field in this column; take the lowest such black block for moving back (not the just added one). This procedure cannot decrease the sum of distances because of the following behaviour. Inside the column where we first added a block, in fact the lowest block was moved down. Thus the distances from this lowest block to each of the $d-1$ other blocks in this column increased by δ, the row difference between the new and old position of the lowest block. All other distances in this column remain unchanged. Inside the other modified column each of the $d-1$ distances involving the modified block decreases at most by δ (triangle inequality), whereas the other ones remain unchanged as well. Thus the sum over all cannot decrease. We repeat the previous steps until there is exactly one black block in each column of the lower level 1 group (see Figure 14). Then we lock this group.

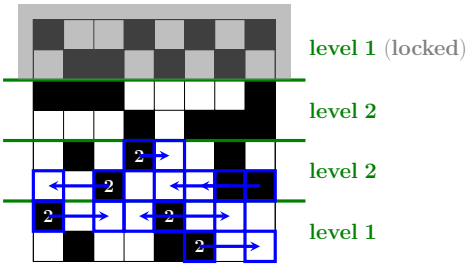

FIGURE 14. Optimizing the lower level 1 group.

Now all level 1 blocks are locked. Note the very important fact that for the current configuration the exact position of level 1 blocks is not relevant for moving blocks of higher level up and down (not entering the locked area) since the sum of the distance to the upper level 1 block and the distance to the lower level 1 block is the same for all possible block positions in between, i.e., for all higher level positions. Therefore, all level 1 blocks can be ignored in the optimizing of higher level blocks. We proceed level-wise and perform the same steps as before for a not locked group with the lowest level. In fact, it does not matter, whether we start with the upper or lower group of that concrete level (see Figure 15).

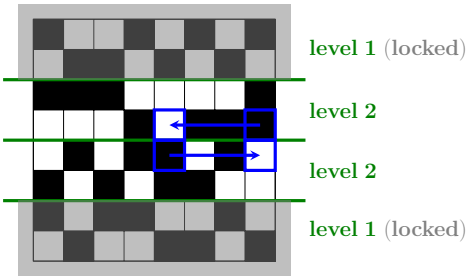

FIGURE 15. Optimizing the next level.

For even dimension d the algorithm stops with locking the two groups of highest level. For odd dimension d the algorithm stops with locking the single group of highest level. Note that if there is only one unlocked group remaining, then this group automatically satisfies that each column contains exactly one black block. In fact, each locked group has exactly one black block per column

and at the moment, where a group gets locked, each column contains exactly d black blocks.

After finitely many steps, all groups are locked. We now ignore the previous levels and enumerate the d groups top-down (see Figure 16).

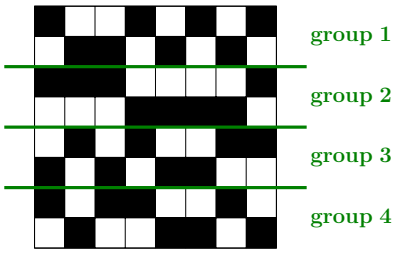

FIGURE 16. Groups with exactly one black block per column.

The previous steps yield a constellation, in which each group contains exactly one black block per column. Hence, if inside a group two columns are exchanged, then the sum of distances within one of the columns increases by the same amount as the sum of distances in the other column decreases. In other words: reordering the columns group-wisely does not change the total sum of differences. This allows us to sort the columns in each group such that they form d strictly increasing 'lines' of length ℓ (see Figure 17).

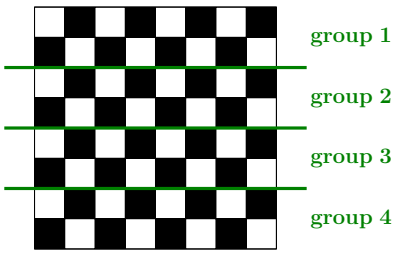

FIGURE 17. Groups with reordered columns.

(ii) The resulting pattern is obviously induced by the shuffles defined in equation (10) describing equidistant shifting. Altogether we get

$$\max_{\vartheta\in\mathcal{P}_{\mathrm{ds}}([0,1]^d)} \int_{[0,1]^d} f_{\sum}\,\mathrm{d}\vartheta = \int_{[0,1]} f_{\sum}(x_1, h_2(x_1), \ldots, h_d(x_1))\,\mathrm{d}\lambda(x_1)$$

and it remains to compute the latter integral. A straightforward calculation yields

$$\begin{aligned}\int_{[0,1]} f_{\sum}(x_1, h_2(x_1), \ldots, h_d(x_1))\,\mathrm{d}\lambda(x_1) &= f_{\sum}\left(0, \frac{1}{d}, \frac{2}{d}, \ldots, \frac{d-1}{d}\right)\\ &= 2\sum_{j=1}^{d-1} \frac{j}{d}\cdot(d-j) = \frac{2}{d}\sum_{j=1}^{d-1}\left(jd - j^2\right)\\ &= \frac{2}{d}\left(d\cdot\frac{(d-1)d}{2} - \frac{(d-1)d(2d-1)}{6}\right)\\ &= (d-1)d - \frac{(d-1)(2d-1)}{3} = \frac{d^2-1}{3},\end{aligned}$$

and the proof is complete. □

5. Solving the original maximization problems

Having determined the upper bound C_f^d for each of the three functions $f_{\min}$, $f_{\max}$ and $f_{\sum}$, we now return to the original problem of calculating

$$B_f^d = \sup_{(x_n)_{n\in\mathbb{N}}\in\mathcal{U}} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(x_i, x_{i+1}, \ldots, x_{i+d-1}).$$

According to ineq. (5) we have already proved that $B_f^d \le C_f^d$ holds for all three functions of interest. We now show (the somewhat surprising fact) that B_f^d and C_f^d coincide for all three cases. For $f_{\max}$ and $f_{\sum}$ we will construct a u.d.s. attaining the upper bounds, while for $f_{\min}$ we show the existence of u.d.s. coming arbitrarily close to the upper bound (implying that the upper bound is best possible).

Proposition 5.1. *For every $d \ge 2$ there exists a u.d.s. $(z_n)_{n\in\mathbb{N}} \in \mathcal{U}$ with*

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_{\max}(z_i, z_{i+1}, \ldots, z_{i+d-1}) = \frac{d-1}{d} = C_{f_{\max}}^d$$

and

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f_{\sum}(z_i, z_{i+1}, \ldots, z_{i+d-1}) = \frac{d^2-1}{3} = C_{f_{\sum}}^d.$$

P r o o f. Suppose that $(x_n)_{n\in\mathbb{N}} \in \mathcal{U}$ is a u.d.s.. Then the sequence $(y_n)_{n\in\mathbb{N}}$ defined by $y_n \coloneqq \frac{x_n}{d}$ obviously is uniformly distributed on the interval $[0, \frac{1}{d}]$. Based on $(y_n)_{n\in\mathbb{N}}$ we construct a u.d.s. $(z_n)_{n\in\mathbb{N}}$ as follows. Decomposing the index as $i = kd + r$ with a non-negative integer k and $0 < r \le d$, set

$$z_i \coloneqq y_{k+1} + \frac{r-1}{d}. \tag{11}$$

In other words, $(z_n)_{n\in\mathbb{N}}$ has the form

$$\begin{aligned}\Big(& y_1,\ y_1 + \frac{1}{d},\ y_1 + \frac{2}{d},\ \ldots,\ y_1 + \frac{d-1}{d},\\ & y_2,\ y_2 + \frac{1}{d},\ y_2 + \frac{2}{d},\ \ldots,\ y_2 + \frac{d-1}{d},\\ & y_3, \qquad \ldots \qquad \Big).\end{aligned}$$

It is straightforward to verify that $(z_n)_{n\in\mathbb{N}}$ is a u.d.s..

(i) We first consider $f_{\max}$. Reusing the index decomposition $i = kd + r$ from above directly yields

$$f_{\max}(z_i, \ldots, z_{i+d-1}) = \begin{cases} \frac{d-1}{d} & \text{for } r = 1\\ \frac{d-1}{d} + y_{k+1} - y_{k+2} & \text{for } r \ge 2. \end{cases}$$

Therefore it follows that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} f_{\max}(z_i, \ldots, z_{i+d-1}) = \frac{d-1}{d} + \frac{1}{n}\cdot(d-1)\sum_{k=1}^{\lfloor \frac{n}{d} \rfloor}(y_k - y_{k+1})\\ + \frac{1}{n}\max\Big(0, n - d\Big\lfloor\frac{n}{d}\Big\rfloor - 1\Big)\cdot\Big(y_{\lfloor \frac{n}{d} \rfloor+1} - y_{\lfloor \frac{n}{d} \rfloor+2}\Big)\end{aligned}$$

which, using the telescoping structure of the sum, yields

$$\left|\frac{1}{n}\sum_{i=1}^{n} f_{\max}(z_i, \ldots, z_{i+d-1}) - \frac{d-1}{d}\right| \le \frac{1}{n}\cdot 2\cdot(d-1)\cdot\frac{1}{d} \quad \overset{n\to\infty}{\longrightarrow} \quad 0.$$

Hence we have found a u.d.s. $(z_n)_{n\in\mathbb{N}}$ satisfying

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f_{\max}(z_i, \ldots, z_{i+d-1}) = \frac{d-1}{d},$$

proving the first assertion of our proposition.

(ii) We now turn towards $f_{\sum}$. Using the u.d.s. $(z_n)_{n\in\mathbb{N}}$ from the previous step, we get

$$f_{\sum}(z_i,\dots,z_{i+d-1}) = \frac{d^2-1}{3} + C(d,r)\,(y_{k+1}-y_{k+2})$$

where once again the index decomposition $i = kd+r$ with a non-negative integer k and $0 < r \leq d$ is used; thereby $C(d,r)$ denotes a constant only depending on the dimension d and the 'sequence column' parameter r. In fact, $C(d,r)$ counts, how often the difference $y_{k+1} - y_{k+2}$ has to be considered. Making use of the appearing telescoping structure yields

$$\left|\frac{1}{n}\sum_{i=1}^{n} f_{\sum}(z_i,\dots,z_{i+d-1}) - \frac{d^2-1}{3}\right| \leq \frac{1}{n}\cdot\left(\sum_{j=1}^{d} C(d,j)\right)\cdot 2\sup_{k\in\mathbb{N}} y_k \quad \stackrel{n\to\infty}{\longrightarrow} \quad 0.$$

Thus, the u.d.s. $(z_n)_{n\in\mathbb{N}}$ also satisfies

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f_{\sum}(z_i,\dots,z_{i+d-1}) = \frac{d^2-1}{3},$$

which concludes the proof. □

Finally, we revisit the minimum distance function $f_{\min}$. Here the procedure performed for the other two functions doesn't seem to work in a comparably elegant way, since we do not have sufficient control over what values the difference $y_k - y_{k+1}$ can attain (in particular, they are strongly restricted by the uniform distribution property). It is more likely that we could stay away from the extrem bound in this case as the proof of the bound has shown that maximal minimum distance requires maximal maximum distance which makes it even more restrictive. Furthermore, for $f_{\max}$ and $f_{\sum}$ distance values larger and smaller than the average are having an effect (cf. the used telescoping structure), whereas in the case of $f_{\min}$ it is likely that only smaller ones will contribute. We will prove the following result for $d \geq 3$ (which is no restriction, since for $d = 2$ all three functions are identical up to the constant 2).

Proposition 5.2. *For every $d \geq 3$ there exist some u.d.s. $\big((z_n^L)_{n\in\mathbb{N}}\big)_{L\in\mathbb{N}} \subseteq \mathcal{U}$ such that we have*

$$\lim_{L\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f_{\min}(z_i,z_{i+1},\dots,z_{i+d-1}) = \frac{1}{d} = C_{f_{\min}}^{d}.$$

P r o o f. We reuse the construction of the sequence $(z_n)_{n\in\mathbb{N}}$ from the proof of Proposition 5.1, but now do not start with an arbitrary uniformly distributed sequence $(x_n)_{n\in\mathbb{N}} \in \mathcal{U}$. Allowing for more control, we restrict to sequences $(x_n)_{n\in\mathbb{N}} \in \mathcal{U}$ with special properties and construct $(x_n)_{n\in\mathbb{N}}$ as follows: assume

that $(w_n)_{n\in\mathbb{N}}$ is a u.d.s. and let $L \geq 2$ be an arbitrary, positive integer. Then obviously $(\frac{w_n}{L})_{n\in\mathbb{N}}$ is uniformly distributed on $[0, \frac{1}{L}]$. As in the construction of $(z_n)_{n\in\mathbb{N}}$ before, we split the unit interval into subintervals of the same length. This time we use L intervals of length $\frac{1}{L}$ each and run through these subintervalls in reverse order, i.e., we define $(x_n)_{n\in\mathbb{N}}$ by

$$\begin{aligned}\Big(&\frac{w_1}{L}+\frac{L-1}{L},\ \frac{w_1}{L}+\frac{L-2}{L},\ \frac{w_1}{L}+\frac{L-3}{L},\ \ldots,\ \frac{w_1}{L},\\ &\frac{w_2}{L}+\frac{L-1}{L},\ \frac{w_2}{L}+\frac{L-2}{L},\ \frac{w_2}{L}+\frac{L-3}{L},\ \ldots,\ \frac{w_2}{L},\\ &\frac{w_3}{L}+\frac{L-1}{L},\qquad \cdots\qquad \Big).\end{aligned}$$

Based on this sequence $(x_n)_{n\in\mathbb{N}}$ now define $(y_n)_{n\in\mathbb{N}}$ and $(z_n)_{n\in\mathbb{N}}$ as in the proof of Proposition 5.1. Again decomposing the index via $i = kd + r$ with a non-negative integer k and $0 < r \leq d$, we get

$$f_{\min}(z_i, \ldots, z_{i+d-1}) = \frac{1}{d} - s(i)$$

where $s(i) = 0$ in the cases $r = 1$ as well as $r \geq 2 \wedge L \nmid k$ and $s(i) \in [\frac{1}{d} - \frac{2}{dL}, \frac{1}{d}]$ for $r \geq 2 \wedge L \mid k$. For our bound this yields

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} f_{\min}(z_i, \ldots, z_{i+d-1}) &= \frac{1}{d} - \frac{1}{n}\sum_{i=1}^{n} s(i)\\ &\geq \frac{1}{d} - \frac{1}{n}\cdot\left\lceil\frac{n}{dL}\right\rceil\cdot(d-1)\cdot\frac{1}{d} \quad \overset{n\to\infty}{\longrightarrow} \quad \frac{1}{d} - \frac{d-1}{d^2L}\end{aligned}$$

which implies

$$\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f_{\min}(z_i, \ldots, z_{i+d-1}) \geq \frac{1}{d} - \frac{d-1}{d^2L}.$$

Letting L increase, using the upper bound from Theorem 4.1, we finally get

$$\lim_{L\to\infty}\limsup_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n} f_{\min}(z_i, \ldots, z_{i+d-1}) = \frac{1}{d}.$$

□

Summing up, we have established the following main result (notice that for $d = 2$ we obtain the result from [13] for all three functions):

Theorem 5.3. *For every $d \geq 2$ the following identities hold and are best-possible:*

$$B^d_{f_{\min}} = \frac{1}{d}, \quad B^d_{f_{\max}} = \frac{d-1}{d}, \quad B^d_{f_\Sigma} = \frac{d^2-1}{3}$$

Remark 5.4. As mentioned in the introduction, our main objective was to determine B^d_f, given by

$$B^d_f = \sup_{(x_n)_{n\in\mathbb{N}}\in\mathcal{U}} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(x_i, x_{i+1}, \ldots, x_{i+d-1})$$

for each of the three functions $f_{\min}, f_{\max}, f_\Sigma$. Instead of tackling the problem directly, we first added degrees of freedom, considered

$$C^d_f = \sup_{\mathbf{x}^1,\mathbf{x}^2,\ldots,\mathbf{x}^d\in\mathcal{U}} \limsup_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} f(x^1_i, x^2_i, \ldots, x^d_i),$$

translated C^d_f to a maximization problem over $\mathcal{P}_{\mathrm{ds}}([0,1]^d)$, and solved the latter. The fact that the (sharp) upper bounds C^d_f obtained for the more general problem coincide with our bounds of interest B^d_f is to a certain extent surprising.

6. Outlook

In this paper we considered the functions $f_{\min}, f_{\max}, f_\Sigma \colon [0,1]^d \to [0,\infty)$ considering all pairwise distances of its arguments. It seems equally interesting to study the related functions

$$\begin{aligned} g_{\min}(\mathbf{x}) &\coloneqq \min_{i=1,\ldots,d-1} |x_i - x_{i+1}|, \\ g_{\max}(\mathbf{x}) &\coloneqq \max_{i=1,\ldots,d-1} |x_i - x_{i+1}|, \\ g_\Sigma(\mathbf{x}) &\coloneqq \sum_{i=1,\ldots,d-1} |x_i - x_{i+1}|, \end{aligned}$$

which only take into account/aggregate distances of consecutive arguments. Considering that changing to the just mentioned functions goes hand in hand with losing symmetry, we conjecture that deriving analogous results might be significantly more difficult. In fact, in this setting it is not enough any more to have a given pool of arguments but beyond this a non-cyclic (!) ordering is needed. In our language of black blocks, which started with highest and lowest one, and then was extended step by step by the number of layers, it would require some kind of timestamp. We leave it as an open problem for future work.

Moreover, as proposed by the referee, it could, firstly, also be interesting to analyze the largest 'gap' between consecutive values, i.e., the maximal distance between adjacent values after ordering $x_i, \dots, x_{i+d-1}$, secondly, to try to generalize our results for $f_{\sum}$ (which corresponds to the 1-norm) to arbitrary p-norms

$$\left(\sum_{i,j \in [d]} |x_i - x_j|^p \right)^{\frac{1}{p}}$$

for $p \in (1, \infty)$, and, thirdly, to replace $[0, 1]$ by the torus $\mathbb{R}/\mathbb{Z}$ and consider the distance to the closest integer instead of $|x_i - x_j|$. From our point of view, none of the just mentioned three points is straightforward, we will therefore try to tackle them in the near future.

Acknowledgement

The authors gratefully acknowledge the support of the project 'IDA Lab Salzburg' (20102/F2300464-KZP; 20204-WISS/225/348/3-2023). Moreover, they are grateful to the referee for carefully reading the manuscript and suggesting some future questions arising from the presented results.

References

[1] H. BAUER: *Wahrscheinlichkeitstheorie*, de Gruyter, Berlin, New York, 2002.
[2] P. BILLINGLEY: *Convergence of probability measures*, Wiley Series in Probability and Statistics: Probability and Statistics John Wiley & Sons Inc., New York, Second edition. 1999.
[3] R. BHATTACHARYA, E. C. WAYMIRE: *A Basic Course in Probability Theory*, Springer-Verlag, New York, 2007.
[4] M. DRMOTA, R. F. TICHY: *Sequences, Discrepancies and Applications*, In: *Lecture Notes in Mathematics 1651*, Springer-Verlag, Berlin, 1997.
[5] F. DURANTE, C. SEMPI: *Copula Theory: An Introduction*, Springer, 2015.
[6] J. FERNÁNDEZ SÁNCHEZ, W. TRUTSCHNIG: *Conditioning based metrics on the space of multivariate copulas and their interrelation with uniform and levelwise convergence and Iterated Function Systems*, Journal of Theoretical Probability **28** (2015), 1311–1336.
[7] F. GRIESSENBERGER, R.R. JUNKER, W. TRUTSCHNIG: *On a multivariate copula-based dependence measure and its estimation*, Electronic Journal of Statistics **16** (2022), 2206–2251.
[8] P. JANSSEN, J. SWANEPOEL, N. VERAVERBEKE: *Large sample behavior of the Bernstein copula estimator*, Journal of Statistical Planning and Inference **142(5)** (2012), 1189–1197.

[9] O. KALLENBERG: *Foundations of modern probability*, Springer-Verlag, New York, Berlin, Heidelberg, 1997.
[10] A. KLENKE: *Probability Theory - A Comprehensive Course*, Springer-Verlag, Berlin, Heidelberg, 2007.
[11] L. KUIPERS, H. NIEDERREITER: *Uniform Distribution of Sequences*, Wiley, New York, 1974; reprint, Dover Publications, Mineola, NY, 2006.
[12] H. KUNZE, D. LA TORRE, F. MENDIVIL, E. R. VRSCAY: *Fractal-Based Methods in Analysis*, Springer Science & Business Media, 2011.
[13] F. PILLICHSHAMMER, S. STEINERBERGER: *Average distance between consecutive points of uniformly distributed sequences*, Unif. Distrib. Theory **4(1)** (2009), 51–67.
[14] A. VAN DER VAART, J.A. WELLNER: *Weak Convergence and Empirical Processes*, Springer Series in Statistics, 1996.

Received month ??, 2026
Accepted month ??, ????

Sebastian Heintze
Department of Artificial Intelligence and Human Interfaces
Faculty of Digital and Analytical Sciences
University of Salzburg
Hellbrunnerstraße 34
5020 Salzburg
Austria
E-mail: sebastian.heintze@plus.ac.at

Wolfgang Trutschnig
Department of Artificial Intelligence and Human Interfaces
Faculty of Digital and Analytical Sciences
University of Salzburg
Hellbrunnerstraße 34
5020 Salzburg
Austria
E-mail: wolfgang@trutschnig.net