Functions of bounded variation, signed measures, and a general Koksma–Hlawka inequality

by

Christoph Aistleitner (Kobe) and Josef Dick (Sydney)

1. Introduction. Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be a set of points in the d-dimensional unit cube $[0,1]^d$. The *star-discrepancy* D_N^* of this point set is defined as

$$(1) \qquad D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N) = \sup_{A \in \mathcal{A}^*} \left| \frac{1}{N} \sum_{n=1}^N \mathbb{1}_A(\mathbf{x}_n) - \lambda(A) \right|.$$

Hereafter, we write $\mathbb{1}_A$ for the indicator function of the set A, λ for the (d-dimensional) Lebesgue measure and $\mathcal{A}^*$ for the class of all closed axis-parallel boxes contained in $[0,1]^d$ which have one vertex at the origin. We generally write vectors in bold font. For vectors $\mathbf{a}, \mathbf{b}$ we write $\mathbf{a} \leq \mathbf{b}$ and $\mathbf{a} < \mathbf{b}$ if the respective inequalities hold in each coordinate, and we write $[\mathbf{a}, \mathbf{b}]$ for the set $\{\mathbf{x} : \mathbf{a} \leq \mathbf{x} \leq \mathbf{b}\}$. We write $\mathbf{0}$ and $\mathbf{1}$ for the d-dimensional vectors $(0, \ldots, 0)$ and $(1, \ldots, 1)$, respectively. Since the star-discrepancy D_N^* is the only discrepancy mentioned in this paper, we will use the word "discrepancy" synonymously with "star-discrepancy".

The *Koksma–Hlawka inequality* states that for any function f on $[0,1]^d$ which has bounded variation in the sense of Hardy and Krause and any point set $\mathbf{x}_1, \ldots, \mathbf{x}_N \in [0,1]^d$ we have

$$(2) \qquad \left| \frac{1}{N} \sum_{n=1}^N f(\mathbf{x}_n) - \int_{[0,1]^d} f(\mathbf{x}) \, d\mathbf{x} \right| \leq (\mathrm{Var}_{\mathrm{HK}} f) D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N).$$

A definition of the variation in the sense of Hardy and Krause (denoted by $\mathrm{Var}_{\mathrm{HK}}$ and abbreviated as *HK-variation* in this paper) is given in Section 2 below. More precisely, by $\mathrm{Var}_{\mathrm{HK}}$ and HK-variation we mean the variation in the sense of Hardy and Krause *anchored at* $\mathbf{1}$ (later the HK-variation anchored at $\mathbf{0}$ will also play a role; however, the HK-variation anchored at $\mathbf{1}$

---

is the "usual" HK-variation). The one-dimensional version of the inequality (2) was first proved by Koksma [24] in 1942, and the multidimensional generalization by Hlawka [19] in 1961.

The Koksma–Hlawka inequality suggests that point sets having small discrepancy can be used for the approximation of the integral of a multivariate function—this observation is one of the cornerstones of the *Quasi-Monte Carlo method* (QMC method) for numerical integration, which uses cleverly designed deterministic points as sampling points of a quadrature rule (as opposed to the *Monte Carlo method*, where randomly generated points are used). Since there exist several constructions of point sets $\mathbf{x}_1, \ldots, \mathbf{x}_N$ in $[0,1]^d$ which achieve a discrepancy bounded by

$$(3) \qquad D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N) \le c_d (\log N)^{d-1} N^{-1},$$

for large N the error estimates in QMC integration can be much better than the (randomized) error of the *Monte Carlo method* (MC method), which is of order $N^{-1/2}$. More information on discrepancy in the context of the previous paragraphs can be found in the monographs [13, 14, 25]. Discrepancy theory in a more general context (geometric, combinatorial, etc.) is described in [11, 29]. A comparison between MC and QMC methods can be found in [26].

The QMC method is widely applied to numerical integration problems, for example to the problem of option pricing in financial mathematics. The general idea is that the problem of calculating the expected value of a multidimensional random variable or the problem of calculating an integral over a general domain $\Omega \subset \mathbb{R}^d$ with respect to a general measure μ can be transformed into an integration problem with respect to the uniform measure on $[0,1]^d$. However, this transformation process is not without its pitfalls (see [48, 49]). This is a particularly critical issue as the Koksma–Hlawka inequality is extremely sensitive with respect to the smallest changes of the function f, since the slightest deformation can turn a function having small HK-variation into a function of infinite HK-variation. For example, if a smooth function f on a general integration domain $\Omega \subset [0,1]^d$ is simply embedded into $[0,1]^d$ by defining $f(\mathbf{x}) = 0$ on $[0,1]^d \setminus \Omega$, then clearly $\int_\Omega f(\mathbf{x})\,d\mathbf{x} = \int_{[0,1]^d} f(\mathbf{x})\,d\mathbf{x}$ but the Koksma–Hlawka inequality is not applicable since the extended function is in general not of bounded HK-variation (unless, roughly speaking, Ω itself is an axis-parallel box). Similar problems appear if one tries to switch from an integral with respect to a general measure to an integral with respect to the uniform measure. Consequently, it is desirable to find a variant of QMC integration which is directly applicable to integration with respect to general measures (this includes the case of general domains $\Omega \subset [0,1]^d$, by taking a measure which is only supported on Ω). Another motivation for studying such general problems comes from

the fact that they are closely related to importance sampling for QMC; see Corollary 1 below.

Let μ be a normalized Borel measure $(^1)$ on $[0,1]^d$. The star-discrepancy with respect to μ of a point set $\mathbf{x}_1, \ldots, \mathbf{x}_N \in [0,1]^d$ is defined as

$$
(4) \qquad D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N; \mu) = \sup_{A \in \mathcal{A}^*} \left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_A(\mathbf{x}_n) - \mu(A) \right|.
$$

Improving results of Beck [4], the authors of the present paper recently showed that for any μ and any N there exists a point set $\mathbf{x}_1, \ldots, \mathbf{x}_N \in [0,1]^d$ for which

$$
(5) \qquad D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N; \mu) \leq c_d (\log N)^{(3d+1)/2} N^{-1}
$$

(see [2]). There is a gap between this upper bound and that for the uniform measure in (3), and it is an interesting open problem whether the smallest possible discrepancy with respect to general measures μ is asymptotically of the same order as the smallest possible discrepancy with respect to the uniform measure. It should be noted that it is also unknown whether (3) is optimal or if the exponent of the logarithmic term can be further reduced; this is known as the *Grand Open Problem* of discrepancy theory (see [5, 6]).

To show that QMC integration is in principle also possible with respect to general measures, the estimate (5) is not sufficient. Additionally one needs a generalized Koksma–Hlawka inequality for non-uniform measures, which is given in Theorem 1 below.

THEOREM 1. *Let f be a measurable $(^2)$ function on $[0,1]^d$ which has bounded HK-variation. Furthermore, let μ be a normalized Borel measure on $[0,1]^d$, and let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be a set of points in $[0,1]^d$. Then*

$$
(6) \qquad \left| \frac{1}{N} \sum_{n=1}^{N} f(\mathbf{x}_n) - \int_{[0,1]^d} f(\mathbf{x}) \, d\mu(\mathbf{x}) \right| \leq (\mathrm{Var}_{\mathrm{HK}} f) D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N; \mu).
$$

Theorem 1 directly implies the following result, which shows that the method of importance sampling can be used for Quasi-Monte Carlo integration.

$(^1)$ Throughout this paper we understand that a *measure* is always non-negative, while a *signed measure* may also have negative values.

$(^2)$ We use the word "measurable" in the sense of "Borel-measurable", that is, in the sense of "measurable with respect to Borel sets". It is possible that a function which has bounded HK-variation is always Borel-measurable as well, and that the assumption of f being measurable can be omitted in the statement of the theorem. However, we have not found in the literature any evidence of the assertion that bounded HK-variation implies Borel-measurability.

COROLLARY 1. *Let f be a measurable function on $[0,1]^d$, and let g be the density of a normalized Borel measure μ_g on $[0,1]^d$. Assume further that f/g has bounded HK-variation, and that $g(\mathbf{x}) > 0$ for all $\mathbf{x} \in [0,1]^d$. Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be a set of points in $[0,1]^d$. Then*

$$(7) \qquad \left| \frac{1}{N} \sum_{n=1}^{N} \frac{f(\mathbf{x}_n)}{g(\mathbf{x}_n)} - \int_{[0,1]^d} f(\mathbf{x})\, d\mathbf{x} \right| \leq \left(\mathrm{Var}_{\mathrm{HK}} \frac{f}{g} \right) D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N; \mu_g).$$

The idea of importance sampling is to find a function g for which the HK-variation of f/g is significantly smaller than that of f; in this case the error bound in (7) can be much better than that in the standard Koksma–Hlawka inequality (2).

Corollary 1 was obtained by Chelson [12] in his PhD thesis, which was published in 1976; it is stated there with the incorrect conclusion that on the right-hand side of (7) one can take the discrepancy with respect to the *uniform* measure of a point set which is related to $\mathbf{x}_1, \ldots, \mathbf{x}_N$ by a simple transformation, instead of the discrepancy of $\mathbf{x}_1, \ldots, \mathbf{x}_N$ with respect to the measure induced by g. Chelson's result and its correct and incorrect parts are described in detail in Section 5 below. Chelson's result is formulated in the language of Corollary 1, which can only be sensibly considered with the assumption that g is the density of a measure; accordingly, a variant of Theorem 1 can only be deduced from Chelson's formulation with the additional assumption that μ has a density, and not for general μ.

Because of the issues mentioned in the previous paragraph, the main impulse for writing this paper was to discuss Chelson's result and method, and to give a correct proof of Theorem 1 without assuming that the measure μ has a density. However, when the present manuscript was almost finished, we coincidentally found out that such a result had already been obtained by Götz [15]. Götz's paper was published in 2002, but apparently it was almost completely overlooked until now; we only found it casually noted in a short survey article of Niederreiter [30]. Apparently Götz did not know of Chelson's result. It should be noted that despite the remarks concerning Chelson's result in the previous paragraph, his method of proof is in principle correct, and could be modified to give a correct proof of Corollary 1. Chelson's and Götz's results are both proved using the same method which is usually used for proving the standard Koksma–Hlawka inequality, namely Abel partial summation. Our proof for Theorem 1 is simpler; it can be seen as an application of a partial integration formula for the Stieltjes integral, which, however, is nothing other than the continuous analogue of the Abel summation formula. The proof is based on a correspondence principle between functions of bounded HK-variation and signed measures of finite total variation, which is of some interest in its own right (see Theorem 3 below).

Together with recent new results on the existence of low-discrepancy point sets with respect to general measures μ, Theorem 1 implies strong convergence results for QMC integration with respect to general measures (see Corollaries 2 and 3 below).

A result somewhat similar to Theorem 1 in the case when μ is the uniform measure on the unit simplex was obtained in [3, 37, 38]. A result similar to Theorem 1 in the special case when μ is the uniform measure on a set $\Omega \subset [0, 1]^d$ (or, more generally, for bounded $\Omega \subset \mathbb{R}^d$) has been obtained recently by Brandolini et al. [10]; however, their error estimate contains multiplicative factors which depend exponentially on the dimension, and which accordingly spoil all tractability results (the case of star-discrepancy and HK-variation is the special case $p = 1$ and $q = \infty$ of the more general result in their paper). Another result of Brandolini et al. [9] gives a general Koksma–Hlawka inequality for the uniform measure on compact parallelepipeds or simplices.

The following theorem establishes the existence of the Jordan decomposition of a multivariate function of bounded HK-variation. It is a generalization of the well-known Jordan decomposition theorem for functions of bounded variation in the one-dimensional case (see for example [51, §12, Section III]). The key ingredient in its proof is a decomposition theorem of Leonov [27, Theorem 3]. The statement of the theorem uses the notion of a *completely monotone function*, which is defined in Section 2 below. It also uses the notion of the *HK-variation anchored at* $\mathbf{0}$, defined in Section 2, and denoted by *HK$\mathbf{0}$-variation* and $\text{Var}_{\text{HK}\mathbf{0}}$ in this paper. The HK$\mathbf{0}$-variation of a function is in general different from the "usual" HK-variation anchored at $\mathbf{1}$. However, a function which has bounded HK-variation also has bounded HK$\mathbf{0}$-variation, and vice versa (see Lemma 2 below). Thus in the assumptions of Theorems 1–3 and throughout this paper the phrase "f has bounded HK-variation" could be replaced by "f has bounded HK$\mathbf{0}$-variation". However, the variations Var_{HK} and $\text{Var}_{\text{HK}\mathbf{0}}$ must *not* be simply exchanged in the conclusions of the respective theorems.

THEOREM 2. *Let f be a function on $[0, 1]^d$ which has bounded HK-variation. Then there exist two uniquely determined completely monotone functions f^+ and f^- on $[0, 1]^d$ such that $f^+(\mathbf{0}) = f^-(\mathbf{0}) = 0$ and*

$$f(\mathbf{x}) = f(\mathbf{0}) + f^+(\mathbf{x}) - f^-(\mathbf{x}), \quad \mathbf{x} \in [0, 1]^d,$$

and

(8) $$\text{Var}_{\text{HK}\mathbf{0}} f = \text{Var}_{\text{HK}\mathbf{0}} f^+ + \text{Var}_{\text{HK}\mathbf{0}} f^-.$$

We call the unique decomposition $f = f^+ - f^-$ having the properties described in Theorem 2 the *Jordan decomposition* of the function f. We note that using relation (20) below it is easy to obtain a variant of Theorem 2 for the HK-variation instead of the HK$\mathbf{0}$-variation.

The following theorem shows, simply speaking, that any right-continuous function of bounded HK-variation defines a finite signed measure and vice versa, and that the HK**0**-variation of the function and the total variation of the signed measure coincide. The Jordan decomposition of a signed measure and the total variation of a signed measure, denoted by $\mathrm{Var}_{\mathrm{total}}$, are defined in Section 2.

Theorem 3.

(a) *Let f be a right-continuous (^{3}) function on $[0,1]^d$ which has bounded HK-variation. Then there exists a unique signed Borel measure ν on $[0,1]^d$ for which*

(9) $$f(\mathbf{x}) = \nu([\mathbf{0}, \mathbf{x}]), \qquad \mathbf{x} \in [0,1]^d.$$

In this case,

(10) $$\mathrm{Var}_{\mathrm{total}}\, \nu = \mathrm{Var}_{\mathrm{HK\mathbf{0}}}\, f + |f(\mathbf{0})|.$$

Furthermore, if $f(\mathbf{x}) = f(\mathbf{0}) + f^+(\mathbf{x}) - f^-(\mathbf{x})$ is the Jordan decomposition of f and $\nu = \nu^+ - \nu^-$ is the Jordan decomposition (^{4}) of ν, then

(11) $$f^+(\mathbf{x}) = \nu^+([\mathbf{0}, \mathbf{x}] \setminus \{\mathbf{0}\}), \quad f^-(\mathbf{x}) = \nu^-([\mathbf{0}, \mathbf{x}] \setminus \{\mathbf{0}\}), \quad \mathbf{x} \in [0,1]^d.$$

(b) *Let ν be a finite signed Borel measure on $[0,1]^d$. Then there exists a unique right-continuous function f of bounded HK-variation on $[0,1]^d$ for which (9) and (10) hold. Furthermore, if $f(\mathbf{x}) = f(\mathbf{0}) + f^+(\mathbf{x}) - f^-(\mathbf{x})$ is the Jordan decomposition of f and $\nu = \nu^+ - \nu^-$ is the Jordan decomposition of ν, then (11) holds.*

This connection between functions of bounded HK-variation and finite signed measures is quite natural, and has probably been observed and used in a less specific form before. For example, this is the same mechanism by which a multidimensional additive set-function of bounded variation defines a finite signed measure and a corresponding multidimensional Stieltjes-integral, as described in Chapter 3 of Stanisław Saks' classical monograph [43] on the *Theory of the Integral*. In essence, this is also the same principle which has been used by Zaremba [52] for his proof of the Koksma–Hlawka inequality by multidimensional Abel summation, which is now the standard method.

(^{3}) We call a multivariate function right-continuous if it is coordinatewise right-continuous in each coordinate, at every point.

(^{4}) Note that in contrast to the Jordan decomposition of a multivariate function of bounded variation, which we have not found in the literature and whose definition is given and existence established by Theorem 2, the Jordan decomposition of a signed measure is a well-known concept in measure theory; in particular, the Jordan decomposition of a signed measure always exists and is unique. See Section 2 for details.

However, the decomposition in Theorem 2 and the correspondence between functions of bounded variation and finite signed measures in Theorem 3, and in particular the relation between the HK**0**-variation of the function and the total variation of the corresponding signed measure, are far from being self-evident, and we have not found any explicit statement like Theorem 2 or Theorem 3 in the literature.

A somewhat vague connection between functions of bounded HK-variation and signed measures is casually noted in [7]. A new notion of bounded variation of a function (which has later been called *bounded variation in the measure sense*), defined in terms of a signed measure corresponding to a function, is introduced in [8], and is used there to prove a general Koksma–Hlawka inequality; however, it is not stated which functions are of bounded variation in this sense. A connection between functions of bounded HK-variation and functions of bounded variation in the measure sense (significantly weaker than our Theorem 3) is mentioned as a "Proposition" (without proof) in [50], with reference to an unpublished manuscript. The same result is stated as a "Theorem" (without proof) in [34], and then, by the same author and several years later, in [35] as an "unproven conjecture". As our Theorem 3 shows, the notion of bounded variation in the measure sense is superfluous, since it coincides with the notion of bounded HK-variation (aside from continuity issues).

Finally, we want to state two consequences of Theorem 1 and the recent results obtained in [2]. The first result shows that QMC integration with respect to general measures is possible with a convergence rate for the error which is almost the same as in the case of the uniform measure; the second is a *tractability* result, which states that QMC integration is possible with a moderate number of sampling points in comparison with the dimension, just as in the case of the uniform measure.

COROLLARY 2. *For any normalized Borel measure μ on $[0,1]^d$ and any $N \geq 1$ there exist points $\mathbf{x}_1, \ldots, \mathbf{x}_N \in [0,1]^d$ for which*

$$\sup_f \left| \frac{1}{N} \sum_{n=1}^N f(\mathbf{x}_n) - \int_{[0,1]^d} f(\mathbf{x}) \, d\mu(\mathbf{x}) \right| \leq 63\sqrt{d} \, \frac{(2 + \log_2 N)^{(3d+1)/2}}{N},$$

where the supremum is taken over all measurable functions f on $[0,1]^d$ which satisfy $\mathrm{Var}_{\mathrm{HK}} \, f \leq 1$.

COROLLARY 3. *Let $\varepsilon > 0$ be given. Then for any normalized Borel measure μ on $[0,1]^d$ there exists a point set $\mathbf{x}_1, \ldots, \mathbf{x}_N \in [0,1]^d$ such that*

(12) $$N \leq 2^{26} d \varepsilon^{-2}$$

and

$$(13) \qquad \sup_f \left| \frac{1}{N} \sum_{n=1}^{N} f(\mathbf{x}_n) - \int_{[0,1]^d} f(\mathbf{x}) \, d\mu(\mathbf{x}) \right| \le \varepsilon,$$

where the supremum is taken over all measurable functions f on $[0,1]^d$ which satisfy $\mathrm{Var}_{\mathrm{HK}} f \le 1$.

The corollaries follow directly from Theorem 1 together with [2, Theorem 1] and [2, Corollary 1], respectively. Corollary 3 implies that the problem of integrating d-dimensional functions whose HK-variation is uniformly bounded with respect to *any* normalized measure is polynomially tractable. For more information on tractability see the three volumes on *Tractability of Multivariate Problems* by Novak and Woźniakowski [31, 32, 33]. In the case of the function class being restricted to indicator functions of sets $A \in \mathcal{A}^*$ (that is, in the case of the left-hand side of (13) being the star-discrepancy with respect to μ) Corollary 3 was proved in [18] (without an effective value for the constant in (12)). It should be noted that the results in [2] are pure existence results, and that the problem of constructing point sets satisfying the conclusions of Corollaries 2 and 3 is completely open. This issue is briefly addressed in Section 5.

The outline of the remaining part of this paper is as follows. Section 2 contains all necessary definitions, and some basic properties of the concepts needed for our proofs. In Section 3 the proof of Theorems 2 and 3 is given, as well as the proof of a lemma establishing a connection between the HK-variation and the HK0-variation (see Lemma 2 below). In Section 4 we deduce Theorem 1 from Theorem 3. In Section 5 we discuss Chelson's result, which is formulated together with a transformation process which supposedly transforms low-discrepancy point sets with respect to λ into low-discrepancy point sets with respect to a general measure μ. We show why this transformation does not have the alleged properties, and that it does actually work when the measure μ is of product type.

2. Definitions and basic properties. To define the variation in the sense of Hardy and Krause, we follow the exposition in [27]; however, additionally we pay attention to the fact that we can put the anchor either to the lower left corner $\mathbf{0}$ (as in [27]) or to the upper right corner $\mathbf{1}$ (as is usually done in the context of discrepancy theory; see for example [25]). Subsequently, we introduce completely monotone functions, following [27, Section 3]. We note that a comprehensive survey on HK-variation and its properties can be found in [36].

Let $f(\mathbf{x})$ be a function on $[0,1]^d$. Let $\mathbf{a} = (a_1, \ldots, a_d)$ and $\mathbf{b} = (b_1, \ldots, b_d)$ be elements of $[0,1]^d$ such that $\mathbf{a} < \mathbf{b}$. We introduce the d-dimensional

difference operator $\Delta^{(d)}$, which assigns to the axis-parallel box $A = [\mathbf{a}, \mathbf{b}]$ a d-dimensional *quasi-volume*

(14)

$$\Delta^{(d)}(f; A) = \sum_{j_1=0}^{1} \cdots \sum_{j_d=0}^{1} (-1)^{j_1 + \cdots + j_d} f(b_1 + j_1(a_1 - b_1), \ldots, b_d + j_d(a_d - b_d)).$$

For $s = 1, \ldots, d$, let

$$0 = x_0^{(s)} < \cdots < x_{m_s}^{(s)} = 1$$

be a partition of $[0, 1]$, and let $\mathcal{P}$ be the partition of $[0, 1]^d$ which is given by

(15)

$$\mathcal{P} = \{[x_{l_1}^{(1)}, x_{l_1+1}^{(1)}] \times \cdots \times [x_{l_d}^{(d)}, x_{l_d+1}^{(d)}] : l_s = 0, \ldots, m_s - 1, \ s = 1, \ldots, d\}.$$

Then the *variation of f on $[0, 1]^d$ in the sense of Vitali* is given by

(16)
$$V^{(d)}(f; [0, 1]^d) = \sup_{\mathcal{P}} \sum_{A \in \mathcal{P}} |\Delta^{(d)}(f; A)|,$$

where the supremum is extended over all partitions of $[0, 1]^d$ into axis-parallel boxes generated by d one-dimensional partitions of $[0, 1]$, as in (15). For $1 \leq s \leq d$ and $1 \leq i_1 < \cdots < i_s \leq d$, let $V^{(s)}(f; i_1, \ldots, i_s; [0, 1]^d)$ denote the s-dimensional variation in the sense of Vitali of the restriction of f to the face

(17) $\qquad U_d^{(i_1, \ldots, i_s)} = \{(x_1, \ldots, x_d) \in [0, 1]^d : x_j = 1 \text{ for all } j \neq i_1, \ldots, i_s\}$

of $[0, 1]^d$. Then the *variation of f on $[0, 1]^d$ in the sense of Hardy and Krause anchored at* $\mathbf{1}$, abbreviated to *HK-variation*, is given by

(18) $\qquad \mathrm{Var}_{\mathrm{HK}}(f; [0, 1]^d) = \sum_{s=1}^{d} \sum_{1 \leq i_1 < \cdots < i_s \leq d} V^{(s)}(f; i_1, \ldots, i_s; [0, 1]^d).$

Note that for the definition of the HK-variation in (18), we add the d-dimensional variation in the sense of Vitali plus the variation in the sense of Vitali on all lower-dimensional faces of $[0, 1]^d$ which are adjacent to $\mathbf{1}$. For the HK**0**-variation, we take instead the sum over those lower-dimensional faces which are adjacent to $\mathbf{0}$. More precisely, let $V^{(s;\mathbf{0})}(f; i_1, \ldots, i_s; [0, 1]^d)$ denote the s-dimensional variation in the sense of Vitali of the restriction of f to the face

$$W_d^{(i_1, \ldots, i_s)} = \{(x_1, \ldots, x_d) \in [0, 1]^d : x_j = 0 \text{ for all } j \neq i_1, \ldots, i_s\}.$$

Then the *variation of f on $[0, 1]^d$ in the sense of Hardy and Krause anchored*

at **0**, abbreviated to *HK***0**-*variation*, is given by

$$(19) \qquad \operatorname{Var}_{\mathrm{HK}\mathbf{0}}(f; [0,1]^d) = \sum_{s=1}^{d} \sum_{1 \le i_1 < \cdots < i_s \le d} V^{(s;\mathbf{0})}(f; i_1, \ldots, i_s; [0,1]^d).$$

For any $\mathbf{a} \in [0,1]^d$, $\mathbf{a} \ne \mathbf{0}$, we can define the variation in the sense of Vitali and the HK**0**-variation of f on $[\mathbf{0}, \mathbf{a}]$ in a similar way to the above, by considering decompositions of $[\mathbf{0}, \mathbf{a}]$ into axis-parallel boxes instead of decompositions of $[0,1]^d$, and again taking a sum over all lower-dimensional faces adjacent to $\mathbf{0}$ as in (19). For notational convenience we also define $\operatorname{Var}_{\mathrm{HK}\mathbf{0}}(f; [\mathbf{0}, \mathbf{0}]) = 0$. Throughout this paper, we simply write $\operatorname{Var}_{\mathrm{HK}} f$ and $\operatorname{Var}_{\mathrm{HK}\mathbf{0}} f$ for $\operatorname{Var}_{\mathrm{HK}}(f; [0,1]^d)$ and $\operatorname{Var}_{\mathrm{HK}\mathbf{0}}(f; [0,1]^d)$, respectively.

The HK-variation and the HK**0**-variation of a function are in general different; for example, the indicator function f of the closed axis-parallel box stretching from the point $(1/2, \ldots, 1/2)$ to $\mathbf{1}$ has HK-variation $2^d - 1$, but HK**0**-variation only 1. This difference reflects the fact that on the one hand the function f can be written as the sum/difference of no less than $2^d - 1$ indicator functions of axis-parallel boxes which have one vertex at the origin (which affects the error term in the Koksma–Hlawka inequality for f in Theorem 1), but on the other hand f is the distribution function of a measure whose total mass is only 1 (namely the Dirac measure centered at $(1/2, \ldots, 1/2)$; consequently this version of the variation appears in Theorem 3). This example represents the "worst case": we will show in Lemma 2 that we always have $\operatorname{Var}_{\mathrm{HK}} \le (2^d - 1) \operatorname{Var}_{\mathrm{HK}\mathbf{0}}$ and vice versa.

We note that by a simple mirroring argument, for any function of bounded HK-variation f we have

$$(20) \qquad \operatorname{Var}_{\mathrm{HK}} f = \operatorname{Var}_{\mathrm{HK}\mathbf{0}} g, \qquad \text{where } g(\mathbf{x}) = f(\mathbf{1} - \mathbf{x}), \ \mathbf{x} \in [0,1]^d;$$

this relation will be needed in the proof of Theorem 1, and explains why the HK**0**-variation turns into the HK-variation when Theorem 3 is used to prove Theorem 1 (see Section 4).

We will also need the following lemma.

LEMMA 1. *Let f and g be functions on $[0,1]^d$ which have bounded HK***0**-*variation. Then for any $\mathbf{a}, \mathbf{b} \in [0,1]^d$, $\mathbf{a} \le \mathbf{b}$, we have*

$$\operatorname{Var}_{\mathrm{HK}\mathbf{0}}(f + g; [\mathbf{0}, \mathbf{b}]) - \operatorname{Var}_{\mathrm{HK}\mathbf{0}}(f + g; [\mathbf{0}, \mathbf{a}])$$
$$\le \operatorname{Var}_{\mathrm{HK}\mathbf{0}}(f; [\mathbf{0}, \mathbf{b}]) + \operatorname{Var}_{\mathrm{HK}\mathbf{0}}(g; [\mathbf{0}, \mathbf{b}])$$
$$- \operatorname{Var}_{\mathrm{HK}\mathbf{0}}(f; [\mathbf{0}, \mathbf{a}]) - \operatorname{Var}_{\mathrm{HK}\mathbf{0}}(g; [\mathbf{0}, \mathbf{a}]).$$

Note that as a special case of the lemma (for $\mathbf{a} = \mathbf{0}$) we get

$$(21) \qquad \operatorname{Var}_{\mathrm{HK}\mathbf{0}}(f + g; [\mathbf{0}, \mathbf{b}]) \le \operatorname{Var}_{\mathrm{HK}\mathbf{0}}(f; [\mathbf{0}, \mathbf{b}]) + \operatorname{Var}_{\mathrm{HK}\mathbf{0}}(g; [\mathbf{0}, \mathbf{b}]),$$

which is just the triangle inequality for the HK**0**-variation. In a similar way

we could prove the (well-known) triangle inequality for the HK-variation:

$$(22) \qquad \mathrm{Var}_{\mathrm{HK}}(f + g) \leq \mathrm{Var}_{\mathrm{HK}} f + \mathrm{Var}_{\mathrm{HK}} g.$$

Proof of Lemma 1. There exist a number m and axis-parallel boxes

$$[\mathbf{a}_i, \mathbf{b}_i], \quad i = 1, \ldots, m,$$

such that $[\mathbf{a}_1, \mathbf{b}_1] = [\mathbf{0}, \mathbf{a}]$ and such that

$$\bigcup_{i=1}^{m} [\mathbf{a}_i, \mathbf{b}_i] = [\mathbf{0}, \mathbf{b}] \quad \text{and} \quad [\mathbf{a}_i, \mathbf{b}_i) \cap [\mathbf{a}_j, \mathbf{b}_j) = \emptyset \quad \text{for } i \neq j.$$

The system $[\mathbf{a}_i, \mathbf{b}_i]$, $1 \leq i \leq m$, is called a *split* of $[\mathbf{0}, \mathbf{b}]$, and for any function h which has bounded variation on $[0, 1]^d$ we have

$$(23) \qquad V^{(d)}(h; [\mathbf{0}, \mathbf{b}]) = \sum_{i=1}^{m} V^{(d)}(h; [\mathbf{a}_i, \mathbf{b}_i])$$

(this is stated, for example, in [36, Lemma 1]). Thus by the triangle inequality for the variation in the sense of Vitali (which follows directly from the ordinary triangle inequality for real numbers) we have

$$V^{(d)}(f + g; [\mathbf{0}, \mathbf{b}]) - V^{(d)}(f + g; [\mathbf{0}, \mathbf{a}]) = \sum_{i=2}^{m} V^{(d)}(f + g; [\mathbf{a}_i, \mathbf{b}_i])$$

$$\leq \sum_{i=2}^{m} \left(V^{(d)}(f; [\mathbf{a}_i, \mathbf{b}_i]) + V^{(d)}(g; [\mathbf{a}_i, \mathbf{b}_i]) \right)$$

$$= V^{(d)}(f; [\mathbf{0}, \mathbf{b}]) - V^{(d)}(f; [\mathbf{0}, \mathbf{a}]) + V^{(d)}(g; [\mathbf{0}, \mathbf{b}]) - V^{(d)}(g; [\mathbf{0}, \mathbf{a}]).$$

Similar inequalities hold for the variation in the sense of Vitali on all lower-dimensional faces of $[\mathbf{0}, \mathbf{b}]$ adjacent to $\mathbf{0}$. Since the HK**0**-variation is defined as the sum over these variations in the sense of Vitali, and since the required inequality holds in each summand, we obtain the conclusion of the lemma. ∎

The following lemma establishes the connection between HK-variation and HK**0**-variation, which was already announced before the statement of Theorem 2. Its proof relies upon Theorem 3, and will be given at the end of Section 3.

LEMMA 2. *Let f be a function on $[0, 1]^d$ which has bounded HK**0**-variation. Then f has bounded HK-variation as well, and*

$$\mathrm{Var}_{\mathrm{HK}} f \leq (2^d - 1) \mathrm{Var}_{\mathrm{HK}\mathbf{0}} f.$$

*The same statement holds if the roles of HK**0**-variation and HK-variation are interchanged.*

A function h on $[0, 1]^d$ is called *completely monotone* if for any closed axis-parallel box $A \subset [0, 1]^d$ of arbitrary dimension s (where $1 \leq s \leq d$)

its s-dimensional quasi-volume $\Delta^{(s)}$ generated by the function h is non-negative (other terms which are used for this property are *quasi-monotone*, *monotonely monotone* and *entirely monotone*). A function of bounded HK-variation can be split into the difference of two completely monotone functions; for the two-dimensional case this is mentioned in [1], where it is attributed to Hobson [22]. The following result of Leonov [27] shows a way for obtaining such a decomposition.

LEMMA 3 ([27, Theorem 3]). *Let $f(\mathbf{x})$ be a function on $[0,1]^d$ which has bounded HK$\mathbf{0}$-variation. Then the functions*

$$(24) \qquad f_1(\mathbf{x}) = \mathrm{Var}_{\mathrm{HK}\mathbf{0}}(f; [\mathbf{0}, \mathbf{x}]) \quad and \quad f_2(\mathbf{x}) = f_1(\mathbf{x}) - f(\mathbf{x})$$

are completely monotone, and

$$f(\mathbf{x}) = f_1(\mathbf{x}) - f_2(\mathbf{x}), \qquad \mathbf{x} \in [0,1]^d.$$

Note that the decomposition into two completely monotone functions is not unique. If f has bounded HK0-variation and g, h are completely monotone functions such that $f = g - h$, then for any completely monotone function r the two functions $f + r$ and $g + r$ also form a decomposition of f as the difference of two completely monotone functions. Thus the decomposition given in Lemma 3 is just one of many possible decompositions of f; in particular, it is *not* the outstanding decomposition which is mentioned in Theorem 2.

For a completely monotone function h, $\mathrm{Var}_{\mathrm{HK}\mathbf{0}}(h; [\mathbf{0}, \mathbf{x}]) = h(\mathbf{x}) - h(\mathbf{0})$ (as noted after [27, Definition 2], this follows from equations (6), (7) and Theorem 1 of [27]). Consequently, for the functions f_1, f_2 in Lemma 3 we have $\mathrm{Var}_{\mathrm{HK}\mathbf{0}} f_1 = \mathrm{Var}_{\mathrm{HK}\mathbf{0}} f$ and $\mathrm{Var}_{\mathrm{HK}\mathbf{0}} f_2 = \mathrm{Var}_{\mathrm{HK}\mathbf{0}} f - f(\mathbf{1}) + f(\mathbf{0})$, which implies that both f_1, f_2 are of bounded HK0-variation (and thus, by Lemma 2, also of bounded HK-variation).

A *signed measure* is a measure which is also allowed to have negative values. A formal definition can be found for example in [51, Chapter 10]. By the Jordan Decomposition Theorem (see for example [51, Theorem 10.21]) any signed measure ν on a measurable space $(\Omega, \mathscr{A})$ can be uniquely decomposed into a "positive" and a "negative" part which are orthogonal to each other. More precisely, there exist measures ν^+ and ν^- such that $\nu^+ \perp \nu^-$ and $\nu = \nu^+ - \nu^-$. Here $\nu^+ \perp \nu^-$ means that there exist two sets $C^+, C^- \in \mathscr{A}$ such that $\Omega = C^+ \cup C^-$ and $C^+ \cap C^- = \emptyset$, and such that $\nu^-(C^+) = 0$ and $\nu^+(C^-) = 0$. Furthermore, at least one of the measures ν^+, ν^- is finite; as a consequence, if ν is finite, then both ν^+ and ν^- must be finite. The pair (ν^+, ν^-) is called the *Jordan decomposition* of ν. The measure $|\nu| = \nu^+ + \nu^-$ is called the *variation measure* of ν, and the quantity $\mathrm{Var}_{\mathrm{total}} \nu = |\nu|(\Omega) = \nu^+(\Omega) + \nu^-(\Omega)$ is called the *total variation* of ν.

3. Functions of bounded variation and signed measures: Proof of Theorems 2 and 3 and Lemma 2. In the proofs of Theorems 2 and 3 below, only the variation anchored at $\mathbf{0}$ (that is, the HK0-variation) plays a role. Subsequently, in Lemma 2, the connection between HK0-variation and HK-variation is established.

Proof of Theorem 2. Let f be a function on $[0,1]^d$ which has bounded HK0-variation, and let f_1, f_2 be the functions defined in Lemma 3. These functions do not provide the desired decomposition; in fact,

$$\mathrm{Var}_{\mathrm{HK0}}\, f_1 + \mathrm{Var}_{\mathrm{HK0}}\, f_2 = f_1(\mathbf{1}) - f_1(\mathbf{0}) + f_2(\mathbf{1}) - f_2(\mathbf{0})$$
$$= \mathrm{Var}_{\mathrm{HK0}}\, f + \mathrm{Var}_{\mathrm{HK0}}\, f - f(\mathbf{1}) + f(\mathbf{0})$$
$$= 2\,\mathrm{Var}_{\mathrm{HK0}}\, f - f(\mathbf{1}) + f(\mathbf{0}),$$

and while for any function f of bounded HK0-variation we have

$$(25) \qquad\qquad f(\mathbf{1}) - f(\mathbf{0}) \le \mathrm{Var}_{\mathrm{HK0}}\, f,$$

there is in general no equality in (25). Consequently, the sum of the variations of the functions f_1 and f_2 from Lemma 3 is in general larger than the variation of f.

Instead, we define the functions

$$(26) \qquad\quad f^+(\mathbf{x}) = \tfrac{1}{2}(\mathrm{Var}_{\mathrm{HK0}}(f; [\mathbf{0}, \mathbf{x}]) + f(\mathbf{x}) - f(\mathbf{0})),$$
$$(27) \qquad\quad f^-(\mathbf{x}) = \tfrac{1}{2}(\mathrm{Var}_{\mathrm{HK0}}(f; [\mathbf{0}, \mathbf{x}]) - f(\mathbf{x}) + f(\mathbf{0})),$$

for $\mathbf{x} \in [0,1]^d$. Then obviously $f(\mathbf{x}) = f(\mathbf{0}) + f^+(\mathbf{x}) - f^-(\mathbf{x})$ for every $\mathbf{x}$. Furthermore, since the function f_2 in Lemma 3 is completely monotone, the same is true for the function f^- in (27). Now set $g = -f$. Then it is easily seen that for any $\mathbf{x}$ we have $\mathrm{Var}_{\mathrm{HK0}}(g; [\mathbf{0}, \mathbf{x}]) = \mathrm{Var}_{\mathrm{HK0}}(f; [\mathbf{0}, \mathbf{x}])$. Applying Lemma 3 to g we see that the function

$$\mathrm{Var}_{\mathrm{HK0}}(g; [\mathbf{0}, \mathbf{x}]) - g(\mathbf{x}) = \mathrm{Var}_{\mathrm{HK0}}(f; [\mathbf{0}, \mathbf{x}]) + f(\mathbf{x}) = 2f^+(\mathbf{x}) + f(\mathbf{0})$$

is completely monotone, which implies that so is f^+. Thus both f^+ and f^- are completely monotone, and we have

$$\mathrm{Var}_{\mathrm{HK0}}\, f^+ = f^+(\mathbf{1}) - f^+(\mathbf{0}) = \tfrac{1}{2}(\mathrm{Var}_{\mathrm{HK0}}\, f + f(\mathbf{1}) - f(\mathbf{0}))$$

and similarly

$$\mathrm{Var}_{\mathrm{HK0}}\, f^- = \tfrac{1}{2}(\mathrm{Var}_{\mathrm{HK0}}\, f - f(\mathbf{1}) + f(\mathbf{0})),$$

which proves that

$$\mathrm{Var}_{\mathrm{HK0}}\, f = \mathrm{Var}_{\mathrm{HK0}}\, f^+ + \mathrm{Var}_{\mathrm{HK0}}\, f^-.$$

It is easily seen that $f^+(\mathbf{0}) = f^-(\mathbf{0}) = 0$, and thus the functions f^+ and f^- from (26) and (27) have the properties required in Theorem 2.

It remains to show that this decomposition is the only one satisfying the statement of Theorem 2. Thus, suppose that g^+ and g^- are two completely

monotone functions such that $f(\mathbf{x}) = f(\mathbf{0}) + g^+(\mathbf{x}) - g^-(\mathbf{x})$ for every $\mathbf{x}$ and $g^+(\mathbf{0}) = g^-(\mathbf{0}) = 0$. Then for every $\mathbf{x}$ we have

$$
(28) \quad
\begin{aligned}
g^+(\mathbf{x}) + g^-(\mathbf{x}) &= \mathrm{Var}_{\mathrm{HK0}}(g^+; [\mathbf{0}, \mathbf{x}]) + \mathrm{Var}_{\mathrm{HK0}}(g^-; [\mathbf{0}, \mathbf{x}]) \\
&\geq \mathrm{Var}_{\mathrm{HK0}}(f; [\mathbf{0}, \mathbf{x}]) = f^+(\mathbf{x}) + f^-(\mathbf{x}),
\end{aligned}
$$

where (28) follows from the triangle inequality for the HK0-variation, that is, from (21). By adding $f^+ - f^- = g^+ - g^-$ to each line of this inequality we obtain

$$(29) \qquad\qquad g^+(\mathbf{x}) \geq f^+(\mathbf{x}), \quad \mathbf{x} \in [0,1]^d,$$

and by subtracting the same quantity from each line we similarly get

$$(30) \qquad\qquad g^-(\mathbf{x}) \geq f^-(\mathbf{x}), \quad \mathbf{x} \in [0,1]^d.$$

Suppose that there exists a point $\mathbf{x} \in [0,1]^d$ such that $g^+(\mathbf{x}) \neq f^+(\mathbf{x})$. By (29) this implies $g^+(\mathbf{x}) > f^+(\mathbf{x})$, which together with (30) and the assumption that $g^+(\mathbf{0}) = g^-(\mathbf{0}) = 0$ also implies

$$
(31) \quad
\begin{aligned}
\mathrm{Var}_{\mathrm{HK0}}(g^+; [\mathbf{0}, \mathbf{x}]) + \mathrm{Var}_{\mathrm{HK0}}(g^-; [\mathbf{0}, \mathbf{x}]) &= g^+(\mathbf{x}) + g^-(\mathbf{x}) \\
&> \mathrm{Var}_{\mathrm{HK0}}(f; [\mathbf{0}, \mathbf{x}]).
\end{aligned}
$$

By Lemma 1,

$$
\begin{aligned}
\mathrm{Var}_{\mathrm{HK0}}(g^+; [\mathbf{0}, \mathbf{1}]) &+ \mathrm{Var}_{\mathrm{HK0}}(g^-; [\mathbf{0}, \mathbf{1}]) \\
&- \mathrm{Var}_{\mathrm{HK0}}(g^+; [\mathbf{0}, \mathbf{x}]) - \mathrm{Var}_{\mathrm{HK0}}(g^-; [\mathbf{0}, \mathbf{x}]) \\
&\geq \mathrm{Var}_{\mathrm{HK0}}(f; [\mathbf{0}, \mathbf{1}]) - \mathrm{Var}_{\mathrm{HK0}}(f; [\mathbf{0}, \mathbf{x}]).
\end{aligned}
$$

Combining this with (31) we get

$$\mathrm{Var}_{\mathrm{HK0}}\, g^+ + \mathrm{Var}_{\mathrm{HK0}}\, g^- > \mathrm{Var}_{\mathrm{HK0}}\, f.$$

Thus the decomposition of f into g^+ and g^- violates (8), which means that it does not have the properties required in Theorem 2. Consequently, the decomposition of f described in Theorem 2 is unique. ∎

We note that the functions f^+ and f^- are the *positive variation* and the *negative variation* of f, respectively. They could also be defined by taking into consideration only the positive or only the negative contributions in (16), respectively, instead of taking absolute values. However, this aspect is not important for our paper, so we do not pursue it any further.

Proof of Theorem 3(a). Assume that f is a right-continuous function on $[0,1]^d$ which has bounded HK0-variation. Let f^+ and f^- be the functions in the Jordan decomposition of f as in Theorem 2. As noted, both f^+ and f^- are completely monotone.

Now we will show that f^+ and f^- are right-continuous. We define functions $\tilde{f}^+$ and $\tilde{f}^-$ by setting

$$\tilde{f}^+(\mathbf{x}) = \lim_{\varepsilon \searrow 0} f^+(x_1 + \varepsilon, x_2, \ldots, x_d),$$

(32)

$$\tilde{f}^-(\mathbf{x}) = \lim_{\varepsilon \searrow 0} f^-(x_1 + \varepsilon, x_2, \ldots, x_d),$$

for $\mathbf{x} = (x_1, \ldots, x_d) \in [0,1)^d$, and $\tilde{f}^+(\mathbf{x}) = f^+(\mathbf{x})$ and $\tilde{f}^-(\mathbf{x}) = f^-(\mathbf{x})$ for $\mathbf{x} \in [0,1]^d \setminus [0,1)^d$. Note that the limits in (32) exist since f^+ and f^- are monotone in every coordinate and bounded. By construction, the functions $\tilde{f}^+$ and $\tilde{f}^-$ are right-continuous in the first coordinate, at every point. Also, both $\tilde{f}^+$ and $\tilde{f}^-$ are completely monotone (this property is inherited from f^+ and f^-, respectively). Furthermore $\mathrm{Var_{HK0}}\, \tilde{f}^+ = \tilde{f}^+(\mathbf{1}) - \tilde{f}^+(\mathbf{0}) \leq f^+(\mathbf{1}) - f^+(\mathbf{0}) = \mathrm{Var_{HK0}}\, f^+$, and a similar inequality holds for $\tilde{f}^-$. We also have

$$\tilde{f}^+(\mathbf{x}) - \tilde{f}^-(\mathbf{x}) = \lim_{\varepsilon \searrow 0} \left(f^+(x_1 + \varepsilon, x_2, \ldots, x_d) - f^-(x_1 + \varepsilon, x_2, \ldots, x_d) \right)$$

(33)

$$= \lim_{\varepsilon \searrow 0} f(x_1 + \varepsilon, x_2, \ldots, x_d) - f(\mathbf{0})$$

(34)

$$= f(\mathbf{x}) - f(\mathbf{0}),$$

where we used the fact that f is right-continuous to get from (33) to (34). Thus by (21) we must actually have

(35) $\qquad \mathrm{Var_{HK0}}\, \tilde{f}^+ = \mathrm{Var_{HK0}}\, f^+ \quad \text{and} \quad \mathrm{Var_{HK0}}\, \tilde{f}^- = \mathrm{Var_{HK0}}\, f^-,$

and

$$\mathrm{Var_{HK0}}\, \tilde{f}^+ + \mathrm{Var_{HK0}}\, \tilde{f}^- = \mathrm{Var_{HK0}}\, f.$$

Since by construction $\tilde{f}^+(\mathbf{1}) = f^+(\mathbf{1})$ and $\tilde{f}^-(\mathbf{1}) = f^-(\mathbf{1})$, and since the functions $\tilde{f}^+$ and $\tilde{f}^-$ are completely monotone, by (35) we have

$$\tilde{f}^+(\mathbf{0}) = \tilde{f}^+(\mathbf{1}) - \mathrm{Var_{HK0}}\, \tilde{f}^+ = f^+(\mathbf{1}) - \mathrm{Var_{HK0}}\, f^+ = f^+(\mathbf{0}) = 0,$$

and similarly $\tilde{f}^-(\mathbf{0}) = 0$. Overall, the functions $\tilde{f}^+$ and $\tilde{f}^-$ have all the properties required from the decomposition in Theorem 2. However, since this decomposition is unique, we must have $\tilde{f}^+ = f^+$ and $\tilde{f}^- = f^-$. In other words, the functions f^+ and f^- must already be right-continuous in the first coordinate. The same argument can be applied to show that f^+ and f^- must be right-continuous in all other coordinates as well.

We can define a set-function ν_f^+ on the elements of $\mathcal{A}^*$ by setting

$$\nu_f^+([\mathbf{0}, \mathbf{x}]) = f^+(\mathbf{x}) \quad \text{for } \mathbf{x} \in [0,1]^d.$$

This function can be extended in a unique way into a countably additive set-function on the algebra of all finite unions of elements of $\mathcal{A}^*$ (here it is important that f^+ is right-continuous, to ensure that the resulting set-function is countably additive). Finally, by the Carathéodory extension theorem, this countably additive set-function can be extended in a unique way into a measure ν_f^+ on the sigma-field generated by $\mathcal{A}^*$. Since this sigma-field

consists of the Borel sets on $[0, 1]^d$, the measure ν_f^+ is a Borel measure. All the necessary extension theorems for this construction are contained in [51, Chapter 5]; however, this is just the standard construction of how a multivariate distribution function (namely f^+) defines a measure, as described for example in [23, Chapter 3]. In the same way, we can construct a measure ν_f^- from f^-. Note that both ν_f^+ and ν_f^- are finite.

Let $\delta_{f(\mathbf{0})}$ be the signed Borel measure on $[0, 1]^d$ for which

$$\delta_{f(\mathbf{0})}(A) = \begin{cases} f(\mathbf{0}) & \text{if } \mathbf{0} \in A, \\ 0 & \text{otherwise.} \end{cases}$$

We define

(36) $$\nu_f = \nu_f^+ - \nu_f^- + \delta_{f(\mathbf{0})}.$$

Then ν_f is a finite signed Borel measure, and

$$\nu_f([\mathbf{0}, \mathbf{x}]) = f^+(\mathbf{x}) - f^-(\mathbf{x}) + f(\mathbf{0}) = f(\mathbf{x}) \quad \text{for } \mathbf{x} \in [0, 1]^d.$$

By the Jordan Decomposition Theorem (see the end of the previous section) there exist measures ν^+ and ν^- such that $\nu^+ \perp \nu^-$ and $\nu_f = \nu^+ - \nu^-$. Furthermore there exist two (Borel) sets C^+, C^- such that $[0, 1]^d = C^+ \cup C^-$ and $C^+ \cap C^- = \emptyset$, and such that $\nu^-(C^+) = 0$ and $\nu^+(C^-) = 0$. It is not a priori clear that the Jordan decomposition of the measure ν_f is in direct correspondence with the Jordan decomposition of the function f, that $\nu_f^+ \perp \nu_f^-$, and that (36) already gives the Jordan representation of $\nu_f - \delta_{f(\mathbf{0})}$. However, we will now show that this is actually the case.

Let (ν^+, ν^-) be the Jordan decomposition of ν_f. We define functions $g^+(\mathbf{x})$ and $g^-(\mathbf{x})$ on $[0, 1]^d$ by setting

$$g^+(\mathbf{x}) = \nu^+([\mathbf{0}, \mathbf{x}] \setminus \{\mathbf{0}\}), \quad g^-(\mathbf{x}) = \nu^-([\mathbf{0}, \mathbf{x}] \setminus \{\mathbf{0}\})), \quad \mathbf{x} \in [0, 1]^d.$$

Then it is easily seen that g^+ and g^- are completely monotone, and that $f(\mathbf{x}) = f(\mathbf{0}) + g^+(\mathbf{x}) - g^-(\mathbf{x})$ for all $\mathbf{x} \in [0, 1]^d$. Furthermore $g^+(\mathbf{0}) = g^-(\mathbf{0}) = 0$. Let C^+, C^- be the two sets from above. Then

$$\text{Var}_{\text{HK}\mathbf{0}}\, g^+ = g^+(\mathbf{1}) - \underbrace{g^+(\mathbf{0})}_{=0} = \nu^+([\mathbf{0}, \mathbf{1}] \setminus \{\mathbf{0}\}))$$

$$= \nu^+(C^+ \setminus \{\mathbf{0}\}) = \nu_f(C^+ \setminus \{\mathbf{0}\}) = \nu_f^+(C^+) - \nu_f^-(C^+)$$

$$\leq \nu_f^+([\mathbf{0}, \mathbf{1}]) = f^+(\mathbf{1}) = \text{Var}_{\text{HK}\mathbf{0}}\, f^+.$$

Similarly we obtain $\text{Var}_{\text{HK}\mathbf{0}}\, g^- \leq \text{Var}_{\text{HK}\mathbf{0}}\, f^-$, which implies

(37) $$\text{Var}_{\text{HK}\mathbf{0}}\, g^+ + \text{Var}_{\text{HK}\mathbf{0}}\, g^- \leq \text{Var}_{\text{HK}\mathbf{0}}\, f^+ + \text{Var}_{\text{HK}\mathbf{0}}\, f^- = \text{Var}_{\text{HK}\mathbf{0}}\, f.$$

By (21) the inequality sign in (37) must actually be an equality sign. Consequently, $f(\mathbf{x}) = f(\mathbf{0}) + g^+(\mathbf{x}) - g^-(\mathbf{x})$ is a decomposition having all the properties described in Theorem 2. Now since the decomposition in Theorem 2 is unique, this implies that $f^+ = g^+$ and $f^- = g^-$. As a consequence,

we have

$$\text{Var}_{\text{total}}\, \nu_f = \nu^+([0,1]^d) + \nu^-([0,1]^d) = f^+(\mathbf{1}) + f^-(\mathbf{1}) + |f(\mathbf{0})|$$
$$= \text{Var}_{\text{HK}\mathbf{0}}\, f + |f(\mathbf{0})|.$$

This proves part (a) of the theorem. ∎

Proof of Theorem 3(b). Assume that a finite signed measure ν on $[0,1]^d$ is given. Then by the Jordan Decomposition Theorem there exist two finite measures ν^+ and ν^- such that $\nu = \nu^+ - \nu^-$, and such that $\nu^+ \perp \nu^-$. We define two functions f^+, f^- on $[0,1]^d$ by setting

$$f^+(\mathbf{x}) = \nu^+([\mathbf{0},\mathbf{x}] \setminus \{\mathbf{0}\}), \quad f^-(\mathbf{x}) = \nu^-([\mathbf{0},\mathbf{x}] \setminus \{\mathbf{0}\}), \quad \mathbf{x} \in [0,1]^d.$$

Then f^+ and f^- are two finite, right-continuous, completely monotone functions. Furthermore, $f(\mathbf{x}) = f^+(\mathbf{x}) - f^-(\mathbf{0}) + \nu(\{\mathbf{0}\})$, $\mathbf{x} \in [0,1]^d$, is a right-continuous function of bounded HK$\mathbf{0}$-variation (since it can be written as the difference of two finite, completely monotone functions; see [27, Corollary 3]). As explained in part (a) of this proof, the function f defines measures ν_f^+ and ν_f^- and a finite signed measure ν_f. It is then easily seen that ν_f coincides with ν, that the pair (ν_f^+, ν_f^-) is the unique Jordan decomposition of $\nu_f - \delta_{f(\mathbf{0})}$, and that consequently f^+ and f^- are the Jordan decomposition of the function f. As a consequence,

$$\text{Var}_{\text{HK}\mathbf{0}}\, f = \text{Var}_{\text{HK}\mathbf{0}}\, f^+ + \text{Var}_{\text{HK}\mathbf{0}}\, f^- = \nu^+([0,1]^d \setminus \{\mathbf{0}\}) + \nu^-([0,1]^d \setminus \{\mathbf{0}\})$$
$$= \text{Var}_{\text{total}}\, \nu - \underbrace{|\nu(\{\mathbf{0}\})|}_{=|f(\mathbf{0})|},$$

which proves the theorem. ∎

Proof of Lemma 2. Let a function f on $[0,1]^d$ be given, and assume that f has bounded HK$\mathbf{0}$-variation. In the first step we assume that f is right-continuous. Then by Theorems 2 and 3 there exist completely monotone functions f^+ and f^- such that (8) holds, and there exists a signed measure ν which, together with its Jordan decomposition $\nu = \nu^+ - \nu^-$, satisfies (9)–(11).

To calculate the HK-variation of f, we have to calculate its variation in the sense of Vitali on faces of the form (17). However, the situation becomes much easier if we separately calculate the variations in the sense of Vitali of f^+ and f^- instead. For a completely monotone function h we have

(38) $$V^{(d)}(h; [0,1]^d) = \Delta^{(d)}(h; [0,1]^d).$$

This equality follows from the fact that for a completely monotone function all summands in the sum in (16) are non-negative, and consequently this sum is a sort of telescoping sum, no matter which partition $\mathcal{P}$ is chosen (alternatively, this equality may be deduced from the fact that it trivially

holds in the one-dimensional case, and that the d-dimensional difference operator $\Delta^{(d)}$ actually is the composition of d one-dimensional difference operators). For the same reason, the same equality as (38) holds for the s-dimensional variation in the sense of Vitali $V^{(s)}$ on every s-dimensional face of $[0,1]^d$, for $1 \leq s \leq d$. In particular,

$$(39) \qquad V^{(s)}(h; U_d^{(i_1,\ldots,i_s)}) = \Delta^{(s)}(h; U_d^{(i_1,\ldots,i_s)}),$$

provided h is completely monotone.

A little combinatorial reasoning now shows that the representation (11) implies that for any s-dimensional face $U_d^{(i_1,\ldots,i_s)}$ of the form (17), $1 \leq s \leq d$, we have

$$\Delta^{(s)}(f^+; U_d^{(i_1,\ldots,i_s)}) = \nu^+(\{(x_1,\ldots,x_d) \in [0,1]^d : x_{i_1} > 0, \ldots, x_{i_s} > 0\})$$

(this is how a distribution function is used to calculate the measure of a half-open axis-parallel box). Together with (39) this implies

$$(40) \quad V^{(s)}(f^+; U_d^{(i_1,\ldots,i_s)})$$
$$\leq \nu^+(\{(x_1,\ldots,x_d) \in [0,1]^d : x_{i_1} > 0, \ldots, x_{i_s} > 0\})$$
$$\leq \nu^+([0,1]^d \setminus \{\mathbf{0}\}) = f^+(\mathbf{1}) = \mathrm{Var}_{\mathrm{HK0}}\, f^+.$$

We note that the number of summands in the definition of the HK-variation is $2^d - 1$. Since by (40) each of these summands is bounded by $\nu^+([0,1]^d)$, we have

$$\mathrm{Var}_{\mathrm{HK}}\, f^+ \leq (2^d - 1)\, \mathrm{Var}_{\mathrm{HK0}}\, f^+.$$

In a similar way we obtain

$$\mathrm{Var}_{\mathrm{HK}}\, f^- \leq (2^d - 1)\, \mathrm{Var}_{\mathrm{HK0}}\, f^-.$$

Thus (8) and (22) lead to

$$\mathrm{Var}_{\mathrm{HK}}\, f \leq \mathrm{Var}_{\mathrm{HK}}\, f^+ + \mathrm{Var}_{\mathrm{HK}}\, f^- \leq (2^d - 1)\, \mathrm{Var}_{\mathrm{HK0}}\, f,$$

which proves the lemma under the additional assumption that f is right-continuous.

Without assuming that f is right-continuous we can still use Theorem 2 to find a Jordan decomposition $f = f^+ - f^-$ of f, but the functions f^+ and f^- are (in general) not right-continuous and consequently cannot be used to define a measure (as in Theorem 3). However, as (38) and (39) show, the variation in the sense of Vitali (on $[0,1]^d$ as well as on lower-dimensional faces) of a completely monotone function h depends only on the values of h at the *corners* of $[0,1]^d$, and consequently the same must be true for the HK0-variation and the HK-variation of h. We set

$$\overline{f^+}(x_1,\ldots,x_d) = f(\tau(x_1),\ldots,\tau(x_d)) \quad \text{for } \mathbf{x} = (x_1,\ldots,x_d) \in [0,1]^d,$$

where $\tau(y) = 0$ for $0 \leq y < 1$ and $\tau(1) = 1$. Informally speaking, the value of $\overline{f^+}$ at a point $\mathbf{x}$ is the value of f at the maximal corner of $[0,1]^d$ which is $\leq \mathbf{x}$. Trivially $\overline{f^+}$ coincides with f^+ on all corners of $[0,1]^d$, and it is easy to see that $\overline{f^+}$ is also completely monotone. Furthermore, $\overline{f^+}(\mathbf{0}) = 0$ and $\overline{f^+}$ is right-continuous. Thus, since we have already shown the lemma for right-continuous functions, we have

$$\underbrace{\mathrm{Var}_{\mathrm{HK}} \overline{f^+}}_{=\mathrm{Var}_{\mathrm{HK}} f^+} \leq (2^d - 1) \underbrace{\mathrm{Var}_{\mathrm{HK0}} \overline{f^+}}_{=\mathrm{Var}_{\mathrm{HK0}} f^+}.$$

A similar inequality holds for f^-. Together with (8) and (22) this proves the lemma also in the case when f is not right-continuous.

To show that the statement of the lemma also holds when the role of the HK0-variation and the HK-variation are interchanged, we define $g(\mathbf{x}) = f(\mathbf{1} - \mathbf{x})$ for $\mathbf{x} \in [0,1]^d$. Then by (20) and by the version of the lemma which we proved above we have

$$\mathrm{Var}_{\mathrm{HK0}} f = \mathrm{Var}_{\mathrm{HK}} g \leq (2^d - 1) \mathrm{Var}_{\mathrm{HK0}} g = (2^d - 1) \mathrm{Var}_{\mathrm{HK}} f. \blacksquare$$

4. A Koksma–Hlawka inequality for general measures: proof of Theorem 1.

The proof of Theorem 1 follows similar proofs in [8, Theorem C.1.4] and [35, Theorem 3.2].

Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be given. Throughout the proof we may assume without loss of generality that $f(\mathbf{1}) = 0$ (since otherwise we may replace $f(\mathbf{x})$ by $f(\mathbf{x}) - f(\mathbf{1})$, which changes neither the left-hand side nor the right-hand side of (6)).

In a first step, we assume that f is left-continuous. We define the function $g(\mathbf{x}) = f(\mathbf{1} - \mathbf{x})$ for $\mathbf{x} \in [0,1]^d$. Since we assumed that f is left-continuous and $f(\mathbf{1}) = 0$, this clearly implies that g is right-continuous and $g(\mathbf{0}) = 0$. Furthermore, by (20), we have $\mathrm{Var}_{\mathrm{HK0}} g = \mathrm{Var}_{\mathrm{HK}} f$. Now we apply Theorem 3 to g. Let ν be the signed measure from Theorem 3 which is defined by g. Let $\hat{\nu}$ be the "reflected" measure of ν, which satisfies

$$\hat{\nu}(A) = \nu(\mathbf{1} - A)$$

for any Borel set $A \subset [0,1]^d$, where we write $\mathbf{1} - A = \{\mathbf{1} - \mathbf{x} : \mathbf{x} \in A\}$. It is easily verified that the fact that ν is a signed Borel measure implies that $\hat{\nu}$ is a signed Borel measure as well, and that they have the same total variation. Let $|\hat{\nu}|$ be the variation measure of $\hat{\nu}$ (see the end of Section 2); then according to the previous remarks and Theorem 3 we have

$$(41) \qquad \mathrm{Var}_{\mathrm{total}} \hat{\nu} = \mathrm{Var}_{\mathrm{HK0}} g + |g(\mathbf{0})| = \mathrm{Var}_{\mathrm{HK}} f.$$

Now on the one hand,

$$\frac{1}{N} \sum_{n=1}^{N} f(\mathbf{x}_k) = \frac{1}{N} \sum_{n=1}^{N} g(\mathbf{1} - \mathbf{x}_k) = \int_{[0,1]^d} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[\mathbf{0},\mathbf{1}-\mathbf{x}_n]}(\mathbf{y}) \, d\nu(\mathbf{y})$$

$$= \int_{[0,1]^d} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[\mathbf{x_n},\mathbf{1}]}(\mathbf{y}) \, d\hat{\nu}(\mathbf{y}) = \int_{[0,1]^d} \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[\mathbf{0},\mathbf{y}]}(\mathbf{x}_n) \, d\hat{\nu}(\mathbf{y}).$$

On the other hand, in a similar way, by Fubini's theorem we have

$$\int_{[0,1]^d} f(\mathbf{x}) \, d\mu(\mathbf{x}) = \int_{[0,1]^d} \int_{[0,1]^d} \mathbb{1}_{[\mathbf{0},\mathbf{1}-\mathbf{x}]}(\mathbf{y}) \, d\nu(\mathbf{y}) \, d\mu(\mathbf{x})$$

$$= \int_{[0,1]^d} \int_{[0,1]^d} \mathbb{1}_{[\mathbf{x},\mathbf{1}]}(\mathbf{y}) \, d\hat{\nu}(\mathbf{y}) \, d\mu(\mathbf{x})$$

$$= \int_{[0,1]^d} \int_{[0,1]^d} \mathbb{1}_{[\mathbf{0},\mathbf{y}]}(\mathbf{x}) \, d\hat{\nu}(\mathbf{y}) \, d\mu(\mathbf{x})$$

$$= \int_{[0,1]^d} \int_{[0,1]^d} \mathbb{1}_{[\mathbf{0},\mathbf{y}]}(\mathbf{x}) \, d\mu(\mathbf{x}) \, d\hat{\nu}(\mathbf{y}) = \int_{[0,1]^d} \mu([\mathbf{0},\mathbf{y}]) \, d\hat{\nu}(\mathbf{y}).$$

Consequently

$$\left| \frac{1}{N} \sum_{n=1}^{N} f(\mathbf{x}_k) - \int_{[0,1]^d} f(\mathbf{x}) \, d\mu(\mathbf{x}) \right|$$

$$\leq \int_{[0,1]^d} \underbrace{\left| \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{[\mathbf{0},\mathbf{y}]}(\mathbf{x}_n) - \mu([\mathbf{0},\mathbf{y}]) \right|}_{\leq D_N^*(\mathbf{x}_1,\ldots,\mathbf{x}_N;\mu)} \, d|\hat{\nu}|(\mathbf{y}) \leq D_N^*(\mathbf{x}_1,\ldots,\mathbf{x}_N;\mu) \, \mathrm{Var}_{\mathrm{total}} \, \hat{\nu}.$$

Together with (41) this proves the theorem in the case that f is left-continuous.

Now we show that we can reduce the general case to that of f being left-continuous. Let f be given. By the strong law of large numbers and by the multidimensional Glivenko–Cantelli theorem (see for example [45, Chapter 26]) for any $\varepsilon > 0$ there exist a number M and points $\mathbf{y}_1, \ldots, \mathbf{y}_M \in [0,1]^d$ such that

$$(42) \qquad \left| \frac{1}{M} \sum_{m=1}^{M} f(\mathbf{y}_m) - \int_{[0,1]^d} f(\mathbf{x}) \, d\mu \right| \leq \varepsilon$$

and

$$(43) \qquad D_N^*(\mathbf{y}_1, \ldots, \mathbf{y}_M; \mu) \leq \varepsilon.$$

Set

$$\mathcal{G} = \{\mathbf{0}, \mathbf{1}, \mathbf{x}_1, \ldots, \mathbf{x}_N, \mathbf{y}_1, \ldots, \mathbf{y}_M\},$$

and let $\mathcal{H}$ be the d-dimensional grid that is generated by the elements of $\mathcal{G}$, that is, the set of all points in $\mathbf{z} \in [0,1]^d$ such that the jth coordinate of $\mathbf{z}$ appears as the jth coordinate of an element of $\mathcal{G}$, for $1 \leq j \leq d$. For $\mathbf{x} \in [0,1]^d$, let $\mathrm{succ}(\mathbf{x})$ denote the uniquely defined element $\mathbf{z}$ of $\mathcal{H}$ for which $\mathbf{x} \leq \mathbf{z}$ and for which $\mathbf{z} \leq \mathbf{y}$ for all $\mathbf{y} \in \mathcal{H}$ such that $\mathbf{x} \leq \mathbf{y}$. Informally speaking, $\mathrm{succ}(\mathbf{x})$ is the smallest element of $\mathcal{H}$ which is $\geq \mathbf{x}$ (that is, $\mathrm{succ}(\mathbf{x})$ is the *successor* of $\mathbf{x}$ within $\mathcal{H}$). We define a function $\tilde{f}$ by setting

$$\tilde{f}(\mathbf{x}) = f(\mathrm{succ}(\mathbf{x})), \quad \mathbf{x} \in [0,1]^d.$$

Note that by construction $\tilde{f}(\mathbf{z}) = f(\mathbf{z})$ for all points $z \in \mathcal{G}$. Furthermore, by construction the function $\tilde{f}$ is left-continuous. Additionally, it is easily seen that $\mathrm{Var}_{\mathrm{HK}}\, \tilde{f} \leq \mathrm{Var}_{\mathrm{HK}}\, f$. Since we have already proved the theorem for left-continuous functions, we get

$$\left| \frac{1}{N} \sum_{n=1}^{N} \underbrace{f(\mathbf{x}_n)}_{=\tilde{f}(\mathbf{x}_n)} - \int_{[0,1]^d} f(\mathbf{x})\, d\mu \right|$$

$$(44) \qquad \leq \left| \frac{1}{N} \sum_{n=1}^{N} \tilde{f}(\mathbf{x}_n) - \int_{[0,1]^d} \tilde{f}(\mathbf{x})\, d\mu \right| + \left| \int_{[0,1]^d} \tilde{f}(\mathbf{x})\, d\mu - \frac{1}{M} \sum_{m=1}^{M} \underbrace{\tilde{f}(\mathbf{y}_m)}_{=f(\mathbf{y}_m)} \right|$$

$$(45) \qquad + \left| \frac{1}{M} \sum_{m=1}^{M} f(\mathbf{y}_m) - \int_{[0,1]^d} f(\mathbf{x})\, d\mu \right|.$$

The first term in (44) is at most $(\mathrm{Var}_{\mathrm{HK}}\, \tilde{f})\, D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N; \mu)$, since $\tilde{f}$ is left-continuous. The second term in (44) is at most $\varepsilon\, \mathrm{Var}_{\mathrm{HK}}\, \tilde{f}$, also since $\tilde{f}$ is left-continuous and by (43). Finally, the term in (45) is at most ε by (42). Since $\mathrm{Var}_{\mathrm{HK}}\, \tilde{f} \leq \mathrm{Var}_{\mathrm{HK}}\, f$ and since $\varepsilon > 0$ was arbitrary, this proves Theorem 1. ∎

5. Transformations of point sets and Chelson's general Koksma–Hlawka inequality. In this section we will present the transformation method proposed by Chelson [12], which supposedly transforms a low-discrepancy point set with respect to the uniform measure into a low-discrepancy point set with respect to a general measure μ. We will show, contrary to what is claimed in [12], that this transformation method generally fails, and only gives the desired result in the case when μ is of product type.

Before turning to Chelson's method, we want to note that the problem of transforming a low-discrepancy sequence with respect to the uniform measure into a low-discrepancy sequence with respect to another measure μ has been considered by several other authors, for example in [16, 20, 21, 44]. Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be a point set in $[0,1]^d$. If μ is of product type, that is, if it is the d-dimensional product measure of d one-dimensional measures, and if μ has

a density (with respect to λ), then in [16, 21] transformation methods are presented which (in a computationally tractable way) generate a sequence $\mathbf{y}_1, \ldots, \mathbf{y}_N \in [0,1]^d$ such that

$$D_N^*(\mathbf{y}_1, \ldots, \mathbf{y}_N; \mu) \leq c(d, \mu) D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N).$$

In the case of more general measures, the known results are much less satisfactory. Even under some technical assumptions on μ, the best known transformation [21] only gives

$$D_N^*(\mathbf{y}_1, \ldots, \mathbf{y}_N; \mu) \leq c(d, \mu)(D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N))^{1/d}.$$

Chelson claims that even in the general case one can reach

$$D_N^*(\mathbf{y}_1, \ldots, \mathbf{y}_N; \mu) = D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N),$$

but as noted his argument is incorrect.

Now we turn to the description of the transformation method suggested in [12]. We change the notation (in such a way that it fits the rest of the paper) and simplify some statements, but our exposition is a truthful renarration of the presentation in [12]. Let $g(y_1, \ldots, y_d)$ be a probability density on $[0,1]^d$, let $G(y_1, \ldots, y_d)$ be its distribution function, and let μ_G be the corresponding probability measure. We require that g is non-zero on $[0,1]^d$. Let g_1 be defined by

$$g_1(y_1) = \underbrace{\int_0^1 \ldots \int_0^1}_{d-1 \text{ integrals}} g(y_1, \ldots, y_d)\, dy_2 \ldots dy_d,$$

that is, g_1 is the marginal density for y_1. Let $G_1(y_1)$ be the distribution function of g_1, and let G_1^{-1} be its inverse (which exists because g is positive). Similarly, for $2 \leq s < d$, let

$$g_{1,\ldots,s}(y_1, \ldots, y_s) = \underbrace{\int_0^1 \ldots \int_0^1}_{d-s \text{ integrals}} g(y_1, \ldots, y_s)\, dy_{s+1} \ldots dy_d$$

be the marginal density for $y_1, \ldots, y_s$. Note that $g_{1,\ldots,d} = g$. Then, for $1 \leq s \leq d$, let

$$g_s(y_s) = g(y_s | y_1, \ldots, y_{s-1}) = \frac{g_{1,\ldots,s}(y_1, \ldots, y_s)}{g_{1,\ldots,s-1}(y_1, \ldots, y_{s-1})}.$$

Furthermore, let $G_s(y_s)$ be the (conditional) distribution function of $g_s(y_s)$, that is,

$$(46) \qquad G_s(y_s) = G_s(y_s | y_1, \ldots, y_{s-1}) = \int_0^{y_s} g(u_s | y_1, \ldots, y_{s-1})\, du_s.$$

Let G_s^{-1} denote the inverse of G_s.

Chelson introduced the transformation as follows:

Let $\mathbf{x} = (x_1, \ldots, x_d) \in [0,1]^d$ be given. First, set $z_1 = G_1^{-1}(x_1)$. Then set $z_2 = G_2^{-1}(x_2)$, and so on, until $z_d = G_d^{-1}(x_d)$. This gives a number $\mathbf{z} = (z_1, \ldots, z_d) \in [0,1]^d$. Note that the values $z_1, z_2, \ldots$ must be calculated sequentially, since each depends on those which are already chosen ([5]). We write T for the transformation which maps $\mathbf{x} \mapsto T\mathbf{x} = \mathbf{z}$ in this way.

It is not mentioned in [12], but this transformation is the *Rosenblatt transformation*, which was introduced in [42].

Chelson proves that if X is a uniformly distributed random variable on $[0,1]^d$, then $Z = TX$ has distribution μ_G. This is true, and was also shown in [42]. However, Chelson also claims the following [12, Theorem 2-5]:

Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ be a sequence in $[0,1]^d$, and let $\mathbf{z}_1 = T\mathbf{x}_1, \ldots, \mathbf{z}_N = T\mathbf{x}_N$ be its image under the transformation T described above. Then

(47) $$D_N^*(\mathbf{z}_1, \ldots, \mathbf{z}_N; \mu_G) = D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N).$$

For the "proof", Chelson [12, p. 29] argues as follows:

We define a vector-valued function $\tilde{G}(\mathbf{z}) : [0,1]^d \mapsto [0,1]^d$ by

(48) $$\tilde{G}(\mathbf{z}) = (G_1(z_1), \ldots, G_d(z_d)).$$

Then, by construction, for any $\mathbf{a} \in [0,1]^d$,

(49) $$\sum_{n=1}^{N} \mathbb{1}_{[\mathbf{0}, \mathbf{a}]}(\mathbf{z}_n) = \sum_{n=1}^{N} \mathbb{1}_{[\mathbf{0}, \tilde{G}(\mathbf{a})]}(\mathbf{x}_n),$$

since by $\mathbf{z} \leq \mathbf{a}$ for $\mathbf{z} = T\mathbf{x}$ we mean

$$z_1 = G_1^{-1}(x_1) \leq a_1,$$
$$\vdots$$
$$z_d = G_d^{-1}(x_d) \leq a_d,$$

and this occurs if and only if $x_s \leq G_s(a_s)$ for $1 \leq s \leq d$.

This looks reasonable, but is actually false. We will give a counterexample below. Chelson continues claiming that

(50) $$\mu_G([\mathbf{0}, \mathbf{a}]) = \lambda([\mathbf{0}, \tilde{G}(\mathbf{a})]).$$

This is also false (again see the counterexample below). The error comes from treating the dependent functions

$$G_1(y_1), G_2(y_2 \mid y_1), \ldots, G_d(y_d \mid y_1, \ldots, y_{d-1})$$

as independent functions

$$G_1(y_1), G_2(y_2), \ldots, G_d(y_d).$$

In the definition (46) the dependent nature of $G_1, \ldots, G_d$ is suppressed merely in order to shorten formulas; however, it seems that by suppress-

---

([5]) The dependent nature of the inverse functions is suppressed in Chelson's notation. What is meant is that $z_2 = G_2^{-1}(x_2|z_1), \ldots, z_d = G_d^{-1}(x_d|z_1, \ldots, z_{d-1})$, which explains why the values $z_1, \ldots, z_d$ have to be calculated consecutively.

ing the dependent nature of $G_1, \ldots, G_d$ in the notation, Chelson made the error of simply ignoring this dependence, which of course is not justified ([6]). A correct form of (49) and (50) would require the set $\{\tilde{G}(\mathbf{y}) : \mathbf{y} \in [\mathbf{0}, \mathbf{a}]\}$ instead of $[\mathbf{0}, \tilde{G}(\mathbf{a})]$ on the right-hand side of the equation; however, because of the dependent nature of $G_1, \ldots, G_d$ the set $\{\tilde{G}(\mathbf{y}) : \mathbf{y} \in [\mathbf{0}, \mathbf{a}]\}$ is in general *not* an axis-parallel box, and consequently the discrepancy of $\mathbf{x}_1, \ldots, \mathbf{x}_N$ cannot be used to estimate the number of elements of the original point set which are contained in such a set (see the example below).

In dimension $d = 1$, Chelson's claim is right (there are no conditional distributions in this case). The simplest counterexample can be given for $d = 2$ and $N = 1$. For instance, let the density g be given by

$$g(\mathbf{y}) = \begin{cases} 1/2 & \text{if } y_1 \leq y_2, \\ 3/2 & \text{if } y_1 > y_2, \end{cases} \quad \text{for } \mathbf{y} = (y_1, y_2) \in [0, 1]^2.$$

Thus $g = 1/2$ on the upper left half of the unit square divided by the diagonal linking $\mathbf{0}$ to $\mathbf{1}$, and $g = 3/2$ on the lower right half. Clearly g is a probability density, and g is positive. Note that g cannot be written as the product of two one-dimensional densities. Constructing g_1 and g_2 in the way described above, we get

$$g_1(y_1) = \int_0^1 g(y_1, y_2) \, dy_2 = (3/2)y_1 + (1/2)(1 - y_1) = y_1 + 1/2,$$

$$g_2(y_2|y_1) = \frac{g(y_1, y_2)}{g_1(y_1)} = \begin{cases} \dfrac{1}{1 + 2y_1} & \text{if } y_1 \leq y_2, \\ \dfrac{3}{1 + 2y_1} & \text{if } y_1 > y_2. \end{cases}$$

Accordingly, we get

$$G_1(y_1) = \frac{y_1^2 + y_1}{2}, \quad G_2(y_2|y_1) = \begin{cases} \dfrac{y_2 + 2y_1}{1 + 2y_1} & \text{if } y_1 \leq y_2, \\ \dfrac{3y_2}{1 + 2y_1} & \text{if } y_1 > y_2. \end{cases}$$

Let $\mathbf{x}$ be the point $\mathbf{x} = (x_1, x_2) = (56/81, 20/23)$. The transformed point is $\mathbf{z} = T\mathbf{x} = (7/9, 20/27)$, since

$$G_1\left(\frac{7}{9}\right) = \frac{56}{81} \quad \text{and} \quad G_2\left(\frac{20}{27}\Big|\frac{7}{9}\right) = \frac{20}{23}.$$

Let $\mathbf{a} = (1, 8/10)$. For the function $\tilde{G}$ defined in (48) we have $\tilde{G}(1, 8/10) = (1, 8/10)$. According to (49) we should have

$$\mathbb{1}_{[\mathbf{0}, \mathbf{a}]}(\mathbf{z}) = \mathbb{1}_{[\mathbf{0}, \tilde{G}(\mathbf{a})]}(\mathbf{x}).$$

([6]) To be more precise, the dependent nature of $G_1, \ldots, G_d$ is only irrelevant when G is of product form; in this case the conditional one-dimensional distributions are just the (unconditional) one-dimensional distributions of the product representation themselves, and the transformation method works correctly; see Theorem 4 below.

However, instead we have

$$\mathbb{1}_{[\mathbf{0},\mathbf{a}]}(\mathbf{z}) = \mathbb{1}_{[\mathbf{0},(1,8/10)]}\left(\left(\frac{7}{9},\frac{20}{27}\right)\right) = 1,$$

but

$$\mathbb{1}_{[\mathbf{0},\tilde{G}(\mathbf{a})]}(\mathbf{x}) = \mathbb{1}_{[\mathbf{0},(1,8/10)]}\left(\left(\frac{56}{81},\frac{20}{23}\right)\right) = 0.$$

Thus (48) fails to hold. In a similar way we can show that the claim (50) is also false. We have

$$\mu_G([\mathbf{0},\mathbf{a}]) = \int\limits_{[\mathbf{0},(1,8/10)]} g(\mathbf{y})\,d\mu_G = \frac{22}{25} = 0.88, \quad \text{but} \quad \lambda([\mathbf{0},\tilde{G}(\mathbf{a})]) = \frac{8}{10}.$$

Now it is not surprising anymore that (47) is also false. After some calculations we get

$$D_N^*(\mathbf{z};\mu_G) = \mu_G\left(\left[\mathbf{0},\left(1,\frac{20}{27}\right)\right]\right) = \frac{610}{729} \approx 0.84,$$

but

$$D_N^*(\mathbf{x}) = \lambda\left(\left[\mathbf{0},\left(1,\frac{20}{23}\right)\right]\right) = \frac{20}{23} \approx 0.87.$$

As already mentioned, the error is caused by ignoring the fact that the function $G_2(y_2) = G_2(y_2 \,|\, y_1)$ depends on y_1 (and not only on y_2), and by the incorrect assumption that the image of an axis-parallel box under $\tilde{G}$ is again an axis-parallel box. Figure 1 shows, for the example above, the set $A = \{\tilde{G}(\mathbf{y}) : \mathbf{y} \in [\mathbf{0},\mathbf{a}]\}$, the set $B = [\mathbf{0},\tilde{G}(\mathbf{a})]$, and the point $\mathbf{x}$. Note that A and B do not coincide, and A is not an axis-parallel box.
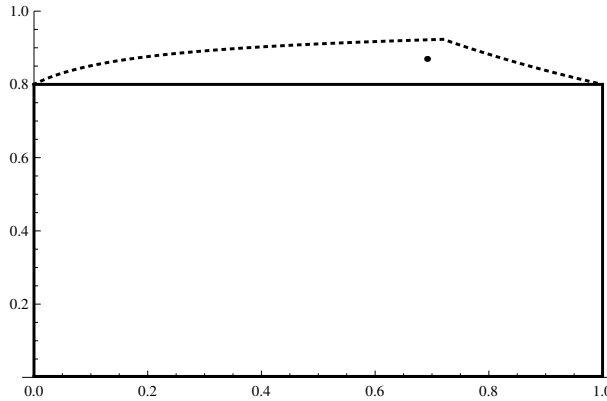

Fig. 1. The sets A (everything below the dotted line) and B (everything below the solid line), and the point $\mathbf{x}$. We have $\mathbf{x} \in A$, so $\mathbf{z} \in [\mathbf{0},\mathbf{a}]$, but $\mathbf{x} \notin B$.

Chelson's generalized Koksma–Hlawka inequality is mentioned several times in the literature; it is explicitly stated, in differing forms, in [16, 17, 35, 39, 40, 41, 46, 47]. In some of these instances it is stated without the transformation procedure and without using the incorrect identity (47) (which means that in those cases it is stated roughly in the same form as our Theorem 1 or Corollary 1). In other cases, the transformation from uniform distribution to another distribution and the (incorrect) identity (47) are explicitly accentuated. For example, after stating Chelson's theorem in its original form, in [47] the authors write

> It is important to note in [the statement of Chelson's theorem] that even though the sampling technique has changed the sequence used to evaluate [the function], the discrepancy appearing in the error bound is still that of the original sequence.

A generalization of Chelson's results is apparently contained in the PhD thesis of Maize [28], which we cannot access (but judging from the statement of [47, Theorem 4.2], it probably suffers from the same problem as Chelson's original results). In [35] a generalization of Chelson's results is "proved", repeating Chelson's errors. Actually, in [35] the author proves the analogue of (47) only in the one-dimensional case (in which it is true) and writes that *the generalization to higher dimensions is straightforward* (which is actually not the case). It seems that the first time that a correct version of (47) is stated in the literature is in [40, Theorem 5], where it is assumed that the measure μ has a density which is the product of d one-dimensional densities. Strangely, in [40] this result is attributed to [35], where the additional assumption that μ is of product form does not appear (and the result is stated in an incorrect form, as noted above). However, it is easy to see that the assumption that μ has a density is not necessary in order to ensure that the μ-discrepancy of the transformed point set is dominated by the discrepancy of the original point set when μ is of product form. Thus a general correct version of (47) would look as follows.

THEOREM 4. *Let μ be a normalized Borel measure on $[0,1]^d$, and let $G(\mathbf{x})$ be its distribution function. Assume that there exist d one-dimensional normalized Borel measures $\mu_1, \ldots, \mu_d$ having distribution functions $G_1, \ldots, G_d$ such that*

$$G(x_1, \ldots, x_d) = \prod_{s=1}^{d} G_s(x_s).$$

Let

$$G_s^{-1}(y) = \min\{y \in [0,1] : G_s(x) \geq y\}, \quad y \in [0,1],$$

be the (pseudo-)inverse of G_s for $s = 1, \ldots, d$. Let T be the transformation on $[0,1]^d$ which maps a point $\mathbf{x} \in [0,1]^d$ to

$$T(\mathbf{x}) = T(x_1, \ldots, x_d) = (G_1^{-1}(x_1), \ldots, G_d^{-1}(x_d)).$$

Let $\mathbf{x}_1, \ldots, \mathbf{x}_N$ *be a set of points in* $[0,1]^d$. *Then*

$$(51) \qquad D_N^*(T(\mathbf{x}_1), \ldots, T(\mathbf{x}_N); \mu) \leq D_N^*(\mathbf{x}_1, \ldots, \mathbf{x}_N).$$

If all the functions $G_1, \ldots, G_d$ *are invertible, then in* (51) *we have equality.*

Proof. We define a function $\tilde{T}$ which maps $\mathbf{y} = (y_1, \ldots, y_d)$ to $\tilde{T}(\mathbf{y}) = (G_1(y_1), \ldots, G_d(y_d))$. Let $A \in \mathcal{A}^*$ be given, and let $\tilde{T}(A)$ denote the set $\{\tilde{T}(\mathbf{x}) : \mathbf{x} \in A\}$. Note the important fact that $\tilde{T}(A)$ is again an element of $\mathcal{A}^*$. We have

$$\frac{1}{N} \sum_{n=1}^{N} \underbrace{\mathbb{1}_A(T(\mathbf{x}_n))}_{=\mathbb{1}_{\tilde{T}(A)}(\mathbf{x}_n)} - \underbrace{\mu(A)}_{=\prod_{s=1}^d G_s(a_s)} = \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\tilde{T}(A)}(\mathbf{x}_n) - \lambda(\tilde{T}(A)).$$

Taking absolute values and the supremum over all sets $A \in \mathcal{A}^*$ yields (51).

If all functions $G_1, \ldots, G_d$ are invertible, then the mapping $A \mapsto T(A)$ is a bijection from the class $\mathcal{A}^*$ to itself, and $\tilde{T}$ is the inverse of T. Thus in this case the suprema $\sup_{A \in \mathcal{A}^*}$ and $\sup_{\tilde{T}(A):A \in \mathcal{A}^*}$ coincide, which implies that in (51) there must be equality. ∎

References

[1]   C. R. Adams and J. A. Clarkson, *Properties of functions $f(x,y)$ of bounded variation*, Trans. Amer. Math. Soc. 36 (1934), 711–730.

[2]   C. Aistleitner and J. Dick, *Low-discrepancy point sets for non-uniform measures*, Acta Arith. 163 (2014), 345–369.

[3]   K. Basu and A. B. Owen, *Low discrepancy constructions in the triangle*, arXiv: 1403.2649 (2014).

[4]   J. Beck, *Some upper bounds in the theory of irregularities of distribution*, Acta Arith. 43 (1983), 115–130.

[5]   D. Bilyk, *On Roth's orthogonal function method in discrepancy theory*, Unif. Distrib. Theory 6 (2011), 143–184.

[6]   D. Bilyk, M. T. Lacey, and A. Vagharshakyan, *On the small ball inequality in all dimensions*, J. Funct. Anal. 254 (2008), 2470–2502.

[7]   N. Bouleau, *On effective computation of expectations in large or infinite dimension*, J. Comput. Appl. Math. 31 (1990), 23–34.

[8]   N. Bouleau and D. Lépingle, *Numerical Methods for Stochastic Processes*, Wiley Ser. Probab. Math. Statist. Appl. Probab. Statist., Wiley, New York, 1994.

[9]   L. Brandolini, L. Colzani, G. Gigante, and G. Travaglini, *A Koksma–Hlawka inequality for simplices*, in: Trends in Harmonic Analysis, Springer INdAM Ser. 3, Springer, Milan, 2013, 33–46.

[10] L. Brandolini, L. Colzani, G. Gigante, and G. Travaglini, *On the Koksma–Hlawka inequality*, J. Complexity 29 (2013), 158–172.

[11] B. Chazelle, *The Discrepancy Method*, Cambridge Univ. Press, Cambridge, 2000.

[12] P. O. Chelson, *Quasi-random techniques for Monte Carlo methods*, Ph.D. thesis, Claremont Graduate Univ., 1976.

[13] J. Dick and F. Pillichshammer, *Digital Nets and Sequences*, Cambridge Univ. Press, Cambridge, 2010.

[14] M. Drmota and R. F. Tichy, *Sequences, Discrepancies and Applications*, Lecture Notes in Math. 1651, Springer, Berlin, 1997.

[15] M. Götz, *Discrepancy and the error in integration*, Monatsh. Math. 136 (2002), 99–121.

[16] J. Hartinger and R. Kainhofer, *Non-uniform low-discrepancy sequence generation and integration of singular integrands*, in: Monte Carlo and Quasi-Monte Carlo Methods 2004, Springer, Berlin, 2006, 163–179.

[17] J. Hartinger, R. F. Kainhofer, and R. F. Tichy, *Quasi-Monte Carlo algorithms for unbounded, weighted integration problems*, J. Complexity 20 (2004), 654–668.

[18] S. Heinrich, E. Novak, G. W. Wasilkowski, and H. Woźniakowski, *The inverse of the star-discrepancy depends linearly on the dimension*, Acta Arith. 96 (2001), 279–302.

[19] E. Hlawka, *Funktionen von beschränkter Variation in der Theorie der Gleichverteilung*, Ann. Mat. Pura Appl. (4) 54 (1961), 325–333.

[20] E. Hlawka and R. Mück, *A transformation of equidistributed sequences*, in: Applications of Number Theory to Numerical Analysis (Proc. Sympos., Univ. Montréal, Montréal, QC, 1971), Academic Press, New York, 1972, 371–388.

[21] E. Hlawka und R. Mück, *Über eine Transformation von gleichverteilten Folgen. II*, Computing (Arch. Elektron. Rechnen) 9 (1972), 127–138.

[22] E. W. Hobson, *The Theory of Functions of a Real Variable and the Theory of Fourier's Series. Vol. I*, 3rd ed., Cambridge, Univ. Press, 1927.

[23] D. Khoshnevisan, *Probability*, Grad. Stud. Math. 80, Amer. Math. Soc., Providence, RI, 2007.

[24] J. F. Koksma, *A general theorem from the theory of uniform distribution modulo 1*, Mathematica Zutphen. B 11 (1942), 7–11 (in Dutch).

[25] L. Kuipers and H. Niederreiter, *Uniform Distribution of Sequences*, Wiley-Interscience, New York, 1974.

[26] C. Lemieux, *Monte Carlo and Quasi-Monte Carlo Sampling*, Springer Ser. Statist., Springer, New York, 2009.

[27] A. S. Leonov, *Remarks on the total variation of functions of several variables and on a multidimensional analogue of Helly's choice principle*, Mat. Zametki 63 (1998), 69–80 (in Russian).

[28] E. H. Maize, *Contributions to the theory of error reduction in Quasi-Monte Carlo methods*, Ph.D. thesis, Claremont Graduate Univ., 1981.

[29] J. Matoušek, *Geometric Discrepancy*, Algorithms Combin. 18, Springer, Berlin, 2010.

[30] H. Niederreiter, *Some current issues in quasi-Monte Carlo methods*, J. Complexity 19 (2003), 428–433.

[31] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems. Vol. I: Linear Information*, EMS Tracts Math. 6, Eur. Math. Soc., Zürich, 2008.

[32] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems. Vol. II: Standard Information for Functionals*, EMS Tracts Math. 12, Eur. Math. Soc., Zürich, 2010.

[33] E. Novak and H. Woźniakowski, *Tractability of Multivariate Problems. Vol. III: Standard Information for Operators*, EMS Tracts Math. 18, Eur. Math. Soc., Zürich, 2012.

[34] G. Ökten, *Contributions to the theory of Monte Carlo and quasi-Monte Carlo methods*, Ph.D. thesis, Claremont Graduate Univ., 1997.

[35] G. Ökten, *Error reduction techniques in quasi-Monte Carlo integration*, Math. Comput. Modelling 30, no. 7–8, (1999), 61–69.

[36] A. B. Owen, *Multidimensional variation for quasi-Monte Carlo*, in: Contemporary Multivariate Analysis and Design of Experiments, Ser. Biostat. 2, World Sci., Hackensack, NJ, 2005, 49–74.

[37] T. Pillards and R. Cools, *A theoretical view on transforming low-discrepancy sequences from a cube to a simplex*, Monte Carlo Methods Appl. 10 (2004), 511–529.

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

[39] A. V. Roşca, *A mixed Monte Carlo and quasi-Monte Carlo sequence for multidimensional integral estimation*, Acta Univ. Apulensis Math. Inform. 14 (2007), 141–160.

[40] N. Roşca, *Generation of non-uniform low-discrepancy sequences in the multidimensional case*, Rev. Anal. Numér. Théor. Approx. 35 (2006), 207–219.

[41] N. C. Roşca, *A combined Monte Carlo and quasi-Monte Carlo method with applications to option pricing*, Acta Univ. Apulensis Math. Inform. 19 (2009), 217–235.

[42] M. Rosenblatt, *Remarks on a multivariate transformation*, Ann. Math. Statist. 23 (1952), 470–472.

[43] S. Saks, *Theory of the Integral*, 2nd ed., Dover, New York, 1964.

[44] C. Schretter and H. Niederreiter, *A direct inversion method for non-uniform quasi-random point sequences*, Monte Carlo Methods Appl. 19 (2013), 1–9.

[45] G. R. Shorack and J. A. Wellner, *Empirical processes with applications to statistics*, Wiley Ser. Probab. Math. Statist. Probab. Math. Statist., Wiley, New York, 1986.

[46] J. Spanier and L. Li, *Quasi-Monte Carlo methods for integral equations*, in: Monte Carlo and Quasi-Monte Carlo Methods 1996 (Salzburg), Lecture Notes in Statist. 127, Springer, New York, 1998, 398–414.

[47] J. Spanier and E. H. Maize, *Quasi-random methods for estimating integrals using relatively small samples*, SIAM Rev. 36 (1994), 18–44.

[48] X. Wang and K. S. Tan, *How do path generation methods affect the accuracy of quasi-Monte Carlo methods for problems in finance?*, J. Complexity 28 (2012), 250–277.

[49] X. Wang and K. S. Tan, *Pricing and hedging with discontinuous functions: Quasi-Monte Carlo methods and dimension reduction*, Management Sci. 59 (2013), 376–389.

[50] Y.-J. Xiao, *Variation of product function and numerical solution of some partial differential equations by low-discrepancy sequences*, Monte Carlo Methods Appl. 2 (1996), 321–330.

[51] J. Yeh, *Real Analysis, Theory of Measure and Integration*, 2nd ed., World Sci., Hackensack, NJ, 2006.

[52] S. K. Zaremba, *Some applications of multidimensional integration by parts*, Ann. Polon. Math. 21 (1968), 85–96.

Christoph Aistleitner
Department of Mathematics and Statistics
Graduate School of Science
Kobe University
1-1 Rokkodai, Nada-ku
Kobe 657-8501, Japan
E-mail: aistleitner@math.tugraz.at

Josef Dick
School of Mathematics and Statistics
University of New South Wales
Sydney, NSW 2052, Australia
E-mail: josef.dick@unsw.edu.au