

Geometric Characterization of Weyl's Discrepancy Norm in Terms of Its n -Dimensional Unit Balls

Bernhard A. Moser

Received: 29 March 2011 / Revised: 14 August 2012 / Accepted: 22 August 2012 /
Published online: 5 September 2012
© Springer Science+Business Media, LLC 2012

Abstract Weyl's discrepancy measure induces a norm on \mathbb{R}^n which shows a monotonicity and a Lipschitz property when applied to differences of index-shifted sequences. It turns out that its n -dimensional unit ball is a zonotope that results from a multiple sheared projection from the $(n + 1)$ -dimensional hypercube which can be interpreted as a discrete differentiation. This characterization reveals that this norm is the canonical metric between sequences of differences of values from the unit interval in the sense that the n -dimensional unit ball of the discrepancy norm equals the space of such sequences.

Keywords Discrepancy · Lipschitz property · Zonotope · Hypercube

1 Motivation

In the mathematical literature discrepancy theory is devoted to problems related to irregularities of distributions. In this context the term discrepancy refers to a measure that evaluates to which extent a given distribution deviates from total uniformity in measure-theoretic, combinatorial and geometric settings. This theory goes back to Weyl [39] and is still an active field of research, see, e.g., [3, 12, 19]. Applications can be found in the field of numerical integration, especially for Monte Carlo methods in high dimensions, see, e.g., [28, 36, 40], or in computational geometry, see, e.g., [1, 9, 21]. For applications to data storage problems on parallel disks, see [10, 13] and for halftoning images, see [31].

This paper is motivated by [24] which applies Weyl's discrepancy concept in order to derive an ordering-dependent norm for measuring the (dis-)similarity between patterns. In this context the focus lies on evaluating the auto-misalignment that measures

B.A. Moser (✉)
Software Competence Center Hagenberg, Hagenberg im Mühlkreis, Austria
e-mail: bernhard.moser@scch.at

the deviation of some function $f(\cdot)$ with its translated version $f(\cdot - T)$ with respect to the lag T . The function f represents a signal, the intensity profile of a line of an image, an image or volumetric data. The interesting point about this is that based on Weyl's discrepancy concept distance measures can be constructed that guarantee the desirable registration properties: (R1) the measure vanishes if and only if the lag vanishes, (R2) the measure increases monotonically with an increasing lag, and (R3) the measure obeys a Lipschitz condition that guarantees smooth changes also for patterns with high frequencies. As proven in [26], properties (R1)–(R3) are not satisfied simultaneously by commonly used measures in this context like mutual information, Kullback–Leibler distance or the Jensen–Rényi divergence measure which are special variants of f -divergence and f -information measures, see, e.g., [4, 20, 29, 30, 37], nor do the standard measures based on p -norms or the widely used correlation measures due to Pearson or Spearman, see [8, 17, 34].

From the point of view of applications properties (R1)–(R3) are relevant for a variety of problems whenever the misalignment of a pattern with its shifted version has to be evaluated. Such problems encounter as autocorrelation in signal processing, see [5]. In computer vision such problems are particularly encountered in stereo matching as point correspondence problem, see, e.g., [32] and [26], in template matching, e.g., for the purpose of print inspection, see, e.g., [7] and [23], in superpixel matching [16] or in defect detection in textured images, see [6, 25, 35]. In these cases, for high-frequency patterns, the discrepancy norm leads to cost functions with less local extrema and a more distinctive region of convergence in the neighborhood of the global minimum compared to commonly used (dis-)similarity measures.

A further promising field of future application is related to measuring the similarity between event-based signals as encountered in neuroscience due to the all-or-none characteristics of neural signals, see, e.g., [38] and, closely related, in event-based imaging, see, e.g., [18] and [22]. In this context it is interesting to point out that the asynchronicity of neighboring sensor elements can lead to misaligned response sequences of events in time. Figure 1 illustrates a sequence of all-or-none events and its auto-misalignment functions induced by the normalized cross-correlation on the one hand and the discrepancy norm on the other hand. Due to properties (R1)–(R3), the discrepancy norm induces a topology in the space of such signals which is compatible with the asynchronicity effect. This means that slightly shifted versions of a sequence of events are still recognized as similar.

The question addressed in this paper therefore is what makes the discrepancy so special when applied to differences of index shifted sequences. This paper provides a geometric analysis that makes clear that the discrepancy norm is inherently related to measuring the distance between index-shifted sequences.

The paper first recalls Weyl's definition [39] in Sect. 2, formulates it as a norm on \mathbb{R}^n and recalls some of its properties from [23]. As the main result of this paper, in Sect. 3 its unit ball is revealed as special zonotope. Section 4 focusses on geometric properties of this zonotope like the number of k -dimensional faces in Sect. 4.1 and its volume in Sect. 4.2.

2 The Discrepancy Norm

In [39] Weyl introduces a concept of discrepancy in the context of pseudorandomness of sequences of numbers from the unit interval. He proposes the formula

$$D_N = \sup_{0 \leq a < b \leq 1} \left| \frac{N(a, b)}{N} - (b - a) \right| \tag{1}$$

to measure the deviation of a sequence $(y_k)_{k \in \{1, \dots, N\}} \subset (0, 1)$ from a uniformly distributed sequence where $N(a, b) = |\{k \in \{1, \dots, N\} | y_k \in (a, b)\}|$, $a, b \in (0, 1)$, $b > a$. As a generalization, the discrepancy of measures μ and ν is defined as

$$D(\mu, \nu) = \sup_{A \in \mathcal{A} \subset \mathcal{U}} |\mu(A) - \nu(A)|, \tag{2}$$

where \mathcal{A} is a σ -algebra of measurable sets over the domain \mathcal{U} , and μ, ν are signed measures defined on the measure space $(\mathcal{U}, \mathcal{A})$.

For linear combinations of Dirac measures $\delta_{\{i\}}$ on \mathbb{Z} given by $\mu = \sum_{i=1}^n x_i \delta_{\{i\}}$ and $\nu = \sum_{i=1}^n y_i \delta_{\{i\}}$, $x_i, y_i \in \mathbb{R}$, and \tilde{A} a set of index intervals, Definition (2) yields

$$D(\mu, \nu) = \sup_{1 \leq n_1 \leq n_2 \leq n} \left| \sum_{i=n_1}^{n_2} x_i - y_i \right|.$$

Therefore, for a summable sequence of real values $x = (x_i)_{i \in \mathbb{Z}}$, $\sum_{i \in \mathbb{Z}} |x_i| < \infty$, Weyl’s discrepancy concept leads to the definition

$$\|x\|_D = \sup_{n_1, n_2 \in \mathbb{Z}; n_1 \leq n_2} \left| \sum_{i=n_1}^{n_2} x_i \right|, \tag{3}$$

which induces a norm, see [Appendix](#). Applications of the norm (3) can be found in pattern recognition [27] and print inspection in the context of pixel classification [2]. In contrast to p -norms $\|\cdot\|_p$, $\|\mathbf{x}\|_p = (\sum_i |x_i|^p)^{1/p}$, the norm $\|\cdot\|_D$ strongly depends on the sign and also the ordering of the entries, as illustrated by the examples $\|(-1, 1, -1, 1)\|_D = 1$ and $\|(-1, -1, 1, 1)\|_D = 2$.

Generally, $\mathbf{x} = (x_i)_i$ with $x_i \geq 0$ entails $\|\mathbf{x}\|_D = \|\mathbf{x}\|_1$, and $\mathbf{x} = ((-1)^i)_i$ the equality $\|\mathbf{x}\|_D = \|\mathbf{x}\|_\infty$, respectively, indicating that the more there are alternating signs of consecutive entries, the lower is the value of the discrepancy norm. Observe that $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_D \leq \|\mathbf{x}\|_1$; hence, due to Hölder’s inequality, $n^{-1/p} \|\mathbf{x}\|_p \leq \|\mathbf{x}\|_D \leq n^{1-1/p} \|\mathbf{x}\|_p$. For convenience, let us consider a sequence $(x_i)_i$ with $i \in I_n$, $x_i = 0$ for $i \notin I_n$, and denote by $\Delta_{\mathbf{x}}(k) = \|(x_{i+k} - x_i)_i\|_D$ the misalignment function of x with respect to $\|\cdot\|_D$. For the proof of the following properties, see [Appendix](#):

- (P1) $\|(x_i)_{i \in I_n}\|_D$ induces a norm on \mathbb{R}^n .
- (P2) $\Delta_{\mathbf{x}}(0) = 0$ for all summable real sequences \mathbf{x} .
- (P3) $\|(x_i)_{i \in I_n}\|_D = \max\{0, \max_{k \in I_n} \sum_{i=1}^k x_i\} - \min\{0, \min_{k \in I_n} \sum_{i=1}^k x_i\}$.
- (P4) $\Delta_{\mathbf{x}}(k) \leq |k| \cdot L$, where $L = \max_i x_i - \min_i x_i$, $k \in \mathbb{Z}$, and $x_i \geq 0$.
- (P5) $\Delta_{\mathbf{x}}(k) = \Delta_{\mathbf{x}}(-k)$ for $\mathbf{x} = (x_i)_i$ with $x_i \geq 0$ and $k \in \mathbb{Z}$.

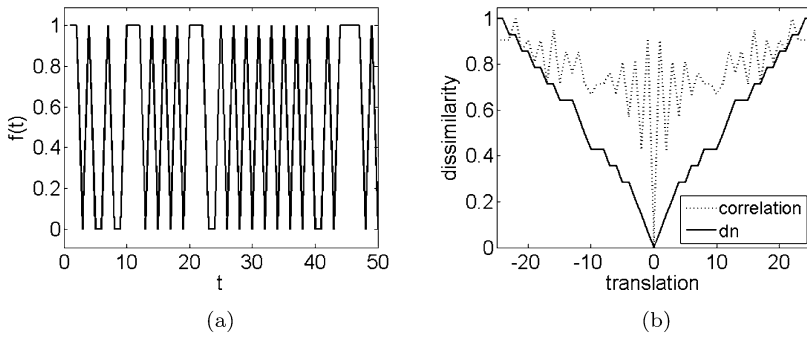


Fig. 1 Figure (a) shows a time series with 0 or 1 values in the depicted interval and 0 outside. Figure (b) shows its misalignment function induced by the normalized cross-correlation (dotted line) and the discrepancy norm (solid line)

(P6) For $\mathbf{x} = (x_i)_i$ with $x_i \geq 0$, the function $\Delta_{\mathbf{x}}(\cdot)$ is monotonically increasing on $\mathbb{N} \cup \{0\}$.

The equation of (P3) allows us to compute the discrepancy of a sequence of length n with $O(n)$ operations instead of $O(n^2)$ number of operations resulting from the original Definition (3). Especially the monotonicity (P6) and the Lipschitz property (P4) are interesting properties for applications in the field of signal analysis. It is instructive to point out that the Lipschitz constant in (P4) does not depend on frequencies or other characteristics of the sequence \mathbf{x} . Properties (P4), (P5), and (P6) are illustrated in Figs. 1(a) and 1(b), which demonstrate the behavior of the misalignment function of a sequence of all-or-none events. While Fig. 1(a) shows typical local minima of the misalignment function with respect to the Euclidean norm, Fig. 1(b) visualizes the symmetry property (P5), the monotonicity property (P6), and the boundedness of its slope due to the Lipschitz property (P4) of the corresponding misalignment function induced by the discrepancy norm.

3 The Unit Ball of the Discrepancy Norm as Convex Polytope

In this section we consider the unit ball of the discrepancy norm in dimension $n \in \mathbb{N}$, $B_D^{(n)} = \{\mathbf{x} \in \mathbb{R}^n \mid \|\mathbf{x}\|_D \leq 1\}$, as geometric object. Definition (3) immediately leads to the representation

$$B_D^{(n)} = \left\{ (x_i)_i \in \mathbb{R}^n \mid -1 \leq \sum_{i=1}^n 1_I(i)x_i \leq 1, I \in \mathcal{I}_n \right\}, \tag{4}$$

where \mathcal{I}_n denotes the set of subintervals from $\{1, \dots, n\}$, and $1_I(\cdot)$ the indicator function given by $1_I(i) = 1$ if and only if $i \in I$. Equation (4) represents the unit ball $B_D^{(n)}$ as bounded intersection of a set of half-spaces, which shows that the unit balls of the discrepancy norm are convex polytopes. Figures 2(a) and 2(b) illustrate the unit balls $B_D^{(n)}$ for $n = 2$ and $n = 3$. Lemma 1 shows a first relationship between $B_D^{(n)}$ and the $(n + 1)$ -hypercube.

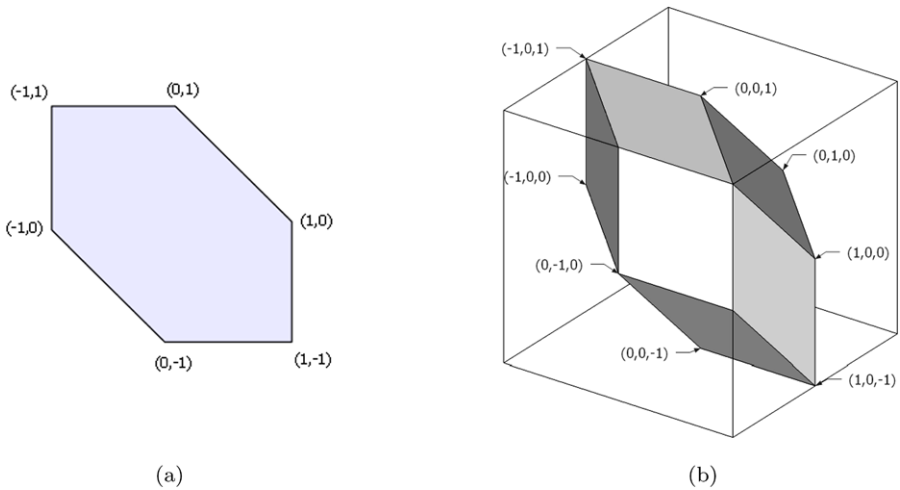


Fig. 2 Illustration of the unit balls of the discrepancy norm in \mathbb{R}^2 and \mathbb{R}^3

Lemma 1 Let $\mathbf{x} = (x_i)_i \in [-1, 1]^n$ with $\|\mathbf{x}\|_D \leq 1$. Then,

$$(c, c + x_1, c + x_1 + x_2, \dots, c + x_1 + \dots + x_n) \in [0, 1]^{n+1} \tag{5}$$

if and only if

$$-\min_{k=1}^n \left\{ 0, \sum_{j=1}^k x_j \right\} \leq c \leq 1 - \max_{k=1}^n \left\{ 0, \sum_{j=1}^k x_j \right\}. \tag{6}$$

The constant c is uniquely determined if and only if $\|\mathbf{x}\|_D = 1$.

Proof Note that if $\min_{k=1}^n \{0, \sum_{j=1}^k x_j\} < 0$, then

$$\min_{i=1}^n \left\{ -\min_{k=1}^n \left\{ 0, \sum_{j=1}^k x_j \right\} + \sum_{j=1}^i x_j \right\} = 0, \tag{7}$$

and that if $\max_{k=1}^n \{0, \sum_{j=1}^k x_j\} > 0$, there holds

$$\max_{i=1}^n \left\{ 1 - \max_{k=1}^n \left\{ 0, \sum_{j=1}^k x_j \right\} + \sum_{j=1}^i x_j \right\} = 1. \tag{8}$$

According to property (P3), the assumption $\|\mathbf{x}\|_D \leq 1$ implies

$$0 \leq -\min_{k=1}^n \left\{ 0, \sum_{j=1}^k x_j \right\} + \sum_{j=1}^i x_j \leq 1 - \max_{k=1}^n \left\{ 0, \sum_{j=1}^k x_j \right\} + \sum_{j=1}^i x_j \leq 1, \tag{9}$$

which shows that condition (6) implies formula (5). Formulas (7) and (8) reveal that the bounds 0 and 1 in inequality (9) are assumed, showing the necessity of condition (6). \square

Given a sequence $\mathbf{x} = (x_1, \dots, x_n)$ with $\|\mathbf{x}\|_D \leq 1$ Lemma 1 reveals that \mathbf{x} can be represented as a sequence of differences $y_{i+1} - y_i, i \in I_n$, with $y_i \in [0, 1]$, and that such a representation is uniquely determined if $\|\mathbf{x}\|_D = 1$. This observation motivates Lemma 2, which points out a fundamental relationship between the discrepancy and the maximum norm.

Lemma 2 *Let $\mathbf{x} = (x_i)_i \in \mathbb{R}^{n+1}, n \in \mathbb{N}$. Then,*

$$\|(x_{i+1} - x_i)_{i \in I_n}\|_D = \|(x_i - \min\{x_i\})_{i \in I_{n+1}}\|_\infty. \tag{10}$$

Proof Consider an \mathbf{x} with $\|\mathbf{x} - \min_i \{x_i\}\|_\infty > 0$ and set $\tilde{x}_i = \frac{x_i - \min_i \{x_i\}}{\|\mathbf{x} - \min_i \{x_i\}\|_\infty} \in [0, 1]$. Then by the Lipschitz property (P4) we get

$$\|(\tilde{x}_{i+1} - \tilde{x}_i)_i\|_D \leq 1. \tag{11}$$

Since $\max_i \{\tilde{x}_i\} = 1$ and $\min_i \{\tilde{x}_i\} = 0$, there are indices i_0 and i_1 such that $\tilde{x}_{i_0} = 1$ and $\tilde{x}_{i_1} = 0$. Without loss of generality, let us assume that $i_0 < i_1$. Then $\|(\tilde{x}_{i+1} - \tilde{x}_i)_i\|_D \geq |\tilde{x}_{i_1} - \tilde{x}_{i_1-1} + \dots + \tilde{x}_{i_0+1} - \tilde{x}_{i_0}| = 1$, which, together with (11), yields $\|(\tilde{x}_{i+1} - \tilde{x}_i)_i\|_D = 1$, and hence $\|(x_{i+1} - x_i)_i\|_D = \|\mathbf{x} - \min_i \{x_i\}\|_\infty$. \square

For convenience, let us define that for a sequence $\mathbf{x} = (x_i)_{i \in I_n} \in \mathbb{R}^n$, the index interval $C \subseteq I_n$ is called a *core discrepancy interval with respect to \mathbf{x}* if and only if for any subset $\tilde{C} \subseteq C$ with $|\sum_{i \in \tilde{C}} x_i| = \|\mathbf{x}\|_D$, it follows that $\tilde{C} = C$. Note that for any \mathbf{x} , due to the definition of the discrepancy norm, there is at least one core discrepancy interval. Further, for convenience, let $\mathbf{0} = (0, \dots, 0)^T$ and $\mathbf{1} = (1, \dots, 1)^T$.

With these prerequisites we come to the central result of this paper that characterizes the vertices $\text{vert}(B_D^{(n)})$ of $B_D^{(n)}$ in terms of vertices of the hypercube of dimension $(n + 1)$.

Lemma 3 *$\mathbf{x} \in \mathbb{R}^n$ is a vertex of the convex polytope $B_D^{(n)}$ if and only if $\|\mathbf{x}\|_D = 1$ and $\mathbf{x} \in \{-1, 0, 1\}^n$.*

Proof First of all we show that $B_D^{(n)}$ equals the convex hull of $\mathcal{D}^{(n)} = \{\mathbf{c} \in \{-1, 0, 1\}^n \mid \|\mathbf{c}\|_D = 1\}$.

Observe that $\text{conv}(\mathcal{D}^{(n)}) \subseteq B_D^{(n)}$ follows immediately from definition (3) and the representation (4) as $\mathbf{x} \in \mathcal{D}^{(n)}$ implies $\sup_{n_1, n_2 \in \mathbb{Z}: n_1 \leq n_2} |\sum_{i=n_1}^{n_2} x_i| \leq 1$.

What remains to be shown is that an arbitrary $\mathbf{x} \in B_D^{(n)}$ can be represented as a convex combination of elements from the set $\mathcal{D}^{(n)}$.

Therefore, suppose that $\mathbf{x} \notin \{-1, 0, 1\}^n$ with $\|\mathbf{x}\|_D = 1$. Let $C = \{n_1, \dots, n_2\} \subseteq I_n$ be a core discrepancy interval with respect to \mathbf{x} . Without loss of generality we may assume that $\sum_{i \in C} x_i > 0$.

Let us consider the cases $n_1 > 1$ or $n_1 = 1$. For the case that $n_1 > 1$, let us set

$$\alpha_i := \sum_{j=i}^{n_1-1} x_j. \tag{12}$$

Observe that $\alpha_{i^*} > 0$ for some index $i^* \in \{1, \dots, n_1 - 1\}$ entails $\sum_{i=i^*}^{n_2} x_i = \alpha_{i^*} + \sum_{i=n_1}^{n_2} x_i > \sum_{i=n_1}^{n_2} x_i = \|\mathbf{x}\|_D$ and, therefore, contradicts the fact that C is a core discrepancy interval. From this it follows that

$$\alpha_i \leq 0 \tag{13}$$

for all indices $i \in \{1, \dots, n_1 - 1\}$.

Now, arrange the partial sums $\alpha_i, i \in \{1, \dots, n_1 - 1\}$, in an increasing order: $0 \leq -\alpha_{r_1} \leq -\alpha_{r_2} \leq \dots \leq -\alpha_{r_{n_1-1}}$, and set $(k \in \{1, \dots, n_1 - 1\})$

$$\begin{aligned} \lambda_1 &:= -\alpha_{r_1}, \\ \lambda_{k+1} &:= \alpha_{r_k} - \alpha_{r_{k+1}}. \end{aligned} \tag{14}$$

Then, we have $\lambda_i \geq 0$ for $i \in \{1, \dots, n_1 - 1\}$, and due to $-\alpha_{r_1} = \lambda_1$ and $-\alpha_{r_k} = \lambda_1 + \dots + \lambda_k$, we obtain $\sum_{i=1}^{n_1-1} \lambda_i \leq \max_{i=1}^{n_1-1} \{-\alpha_i\} \leq 1$. Consequently,

$$\lambda_0 := 1 - \sum_{i=1}^{n_1-1} \lambda_i \in [0, 1]. \tag{15}$$

Finally, we get the representation

$$\alpha_{r_k} = \lambda_1 \cdot v(\alpha_{r_k}, \lambda_1) + \dots + \lambda_{n_1-1} \cdot v(\alpha_{r_k}, \lambda_{n_1-1} + \dots + \lambda_1),$$

where

$$v(\alpha, \lambda) = \begin{cases} -1 & \text{if } \lambda \leq -\alpha \\ 0 & \text{else.} \end{cases}$$

Next let us define the auxiliary vectors $\mathbf{s}^{(0)}, \mathbf{s}^{(1)}, \dots, \mathbf{s}^{(n_1-1)} \in \{-1, 0\}^{(n_1-1)}$ given by

$$\begin{aligned} \mathbf{s}^{(0)} &:= \mathbf{0}, \\ \mathbf{s}^{(j)} &:= \left(v\left(\alpha_1, \sum_{l=1}^j \lambda_l\right), \dots, v\left(\alpha_{n_1-1}, \sum_{l=1}^j \lambda_l\right) \right)^T, \end{aligned} \tag{16}$$

where $j \in \{1, \dots, n_1 - 1\}$. Observe that the vectors (16), the scalars (14) and (15) yield

$$\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_{n_1-1} \end{pmatrix} = \sum_{j=1}^{n_1-1} \lambda_j \mathbf{s}^{(j)} + \lambda_0 \mathbf{s}^{(0)}.$$

Hence,

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n_1-1} \end{pmatrix} = \begin{pmatrix} \alpha_1 - \alpha_2 \\ \vdots \\ \alpha_{n_1-2} - \alpha_{n_1-1} \\ \alpha_{n_1-1} - 0 \end{pmatrix} = \sum_{j=1}^{n_1-1} \lambda_j \tilde{\mathbf{g}}^{(j)} + \lambda_0 \tilde{\mathbf{g}}^{(0)},$$

where $\tilde{\mathbf{g}}^{(0)} = \mathbf{0}$ and

$$\tilde{\mathbf{g}}^{(j)} := \begin{pmatrix} v(\alpha_1, \sum_{l=1}^j \lambda_l) - v(\alpha_2, \sum_{l=1}^j \lambda_l) \\ \vdots \\ v(\alpha_{n_1-2}, \sum_{l=1}^j \lambda_l) - v(\alpha_{n_1-1}, \sum_{l=1}^j \lambda_l) \\ v(\alpha_{n_1-1}, \sum_{l=1}^j \lambda_l) - 0 \end{pmatrix} \in \{-1, 0, 1\}^{n_1-1}.$$

Note that $\tilde{\mathbf{g}}^{(1)}, \dots, \tilde{\mathbf{g}}^{(n_1-1)} \in \mathcal{D}^{(n_1-1)}$ because of

$$\|\tilde{\mathbf{g}}^{(j)}\|_D = \max_k \left\{ 0, v\left(\alpha_k, \sum_{l=1}^j \lambda_l\right) \right\} - \min_k \left\{ 0, v\left(\alpha_k, \sum_{l=1}^j \lambda_l\right) \right\} = 1$$

for $j \in \{1, \dots, n_1 - 1\}$. Note that also $\tilde{\mathbf{g}}^{(0)}$ can be represented as a convex combination of vectors of $\mathcal{D}^{(n_1-1)}$, e.g., $\tilde{\mathbf{g}}^{(0)} = \frac{1}{2}(1, -1, 0, \dots, 0)^T + \frac{1}{2}(-1, 1, 0, \dots, 0)^T$. This proves that (x_1, \dots, x_{n_1-1}) can be represented as a convex combination of elements

$$\mathbf{g}^{(i)} \in \mathcal{D}^{(n_1-1)}. \tag{17}$$

For the other case that $n_1 = 1$, let us set $\beta_i := \sum_{j=n_1}^i x_j$ where $i \in \{n_1, \dots, n_2\}$. If $n_1 = n_2$, the core discrepancy interval property of C entails that $x_{n_1} = 1$. Therefore, let us consider the case $n_2 > n_1$. Then, the assumption $\beta_{i^*} < 0$ for some index $i^* \in \{n_1, \dots, n_2 - 1\}$ leads to $\sum_{i=n_1}^{n_2} x_i = \beta_{i^*} + \sum_{i=i^*}^{n_2} x_i$ implying that $\sum_{i=i^*}^{n_2} x_i > \sum_{i=n_1}^{n_2} x_i$. This contradicts the core discrepancy interval property of C ; hence, in analogy to the case $n_1 > 1$ and formula (13), we get that $\beta_i \geq 0$ for all $i \in \{n_1, \dots, n_2\}$. Now, reasoning steps analogous to the case $n_1 > 1$ can be applied in order to show that $(x_{n_1}, \dots, x_{n_2})$ can be represented as a convex combination of elements

$$\mathbf{s}^{(j)} \in \mathcal{D}^{(n_2-n_1+1)}. \tag{18}$$

If $n_2 = n$, we are ready, and if $n_2 < n$, then let us consider $\gamma_i := \sum_{j=n_2+1}^i x_j$, which, in analogy to the case $n_1 > 1$ leads to $\gamma_i \leq 0$, for which the same reasoning as in case $n_1 > 1$ can be applied showing that (x_{n_2+1}, \dots, x_n) can be represented as a convex combination of elements

$$\mathbf{l}^{(k)} \in \mathcal{D}^{(n-n_2)}. \tag{19}$$

Putting all together, formulas (17), (18), (19) show that \mathbf{x} can be represented by elements

$$\begin{pmatrix} \mathbf{g}^{(i)} \\ \mathbf{s}^{(j)} \\ \mathbf{l}^{(k)} \end{pmatrix} \in \mathcal{D}^{(n)}$$

showing that $\text{conv}(\mathcal{D}^{(n)}) = B_D^{(n)}$.

Finally we show that all elements of $\mathcal{D}^{(n)}$ are vertices of $\text{conv}(\mathcal{D}^{(n)})$. Suppose that $\mathbf{v}_0 \in \mathcal{D}^{(n)}$ can be represented as convex combination of elements $\mathbf{v}_i \in \mathcal{D}^{(n)} \setminus \{\mathbf{v}_0\}$, $i \in \{1, \dots, |\mathcal{D}^{(n)}| - 1\}$, i.e.,

$$\mathbf{v}_0 = \sum_i \lambda_i \mathbf{v}_i, \tag{20}$$

$\sum_i \lambda_i = 1$, $\lambda_i \geq 0$. Then, due to Lemma 1 and the fact that $\|\mathbf{v}_i\|_D = 1$, $i \in \{0, \dots, |\mathcal{D}^{(n)}| - 1\}$, there are constants c_i such that $\bar{\mathbf{v}}_i = (c_i, c_i + v_1^i, \dots, c_i + v_1^i + \dots + v_n^i) \in [0, 1]^{n+1}$, where $\mathbf{v}_i = (v_1^i, \dots, v_n^i)$. Since $v_j^i \in \{-1, 0, 1\}$, Lemma 1 tells that $c_i = 1 - \max_{k=1}^n \{0, \sum_{j=1}^k v_j^i\} \in \{0, 1\}$. From this it follows that $\bar{\mathbf{v}}_i \in \{0, 1\}^{n+1}$. Further, $\|\mathbf{v}_i\|_D = 1$ implies $\bar{\mathbf{v}}_i \in \{0, 1\}^{n+1} \setminus \{\mathbf{0}, \mathbf{1}\}$. Note that $\mathbf{v}_i \neq \mathbf{v}_j$ implies $\bar{\mathbf{v}}_i \neq \bar{\mathbf{v}}_j$. Now, let us consider the linear mapping

$$\mathcal{D}((x_i)_{i \in I_{n+1}}) = (x_{i+1} - x_i)_{i \in I_n}, \tag{21}$$

where $(x_i)_{i \in I_{n+1}} \in \{0, 1\}^{n+1} \setminus \{\mathbf{0}, \mathbf{1}\}$. Equation (20) expressed in terms of (21) means that $\mathcal{D}(\bar{\mathbf{v}}_0) = \sum_i \lambda_i \mathcal{D}(\bar{\mathbf{v}}_i)$, which entails $\mathcal{D}(\bar{\mathbf{v}}_0) = \mathcal{D}(\sum_i \lambda_i \bar{\mathbf{v}}_i)$. Since $\mathcal{D}(\bar{\mathbf{v}}) = \mathcal{D}(\bar{\mathbf{w}})$ implies $\bar{\mathbf{v}} = \bar{\mathbf{w}} + c \cdot \mathbf{1}$ for some $c \in \mathbb{R}$, we obtain

$$\sum_i \lambda_i \bar{\mathbf{v}}_i = \bar{\mathbf{v}}_0 + c \cdot \mathbf{1}. \tag{22}$$

Since $\{\bar{\mathbf{v}}_0 + c \cdot \mathbf{1} | c \in \mathbb{R}\} \cap [0, 1]^{n+1} = \{\bar{\mathbf{v}}_0\}$, we obtain $c = 0$ in Eq. (22), and hence $\bar{\mathbf{v}} = \bar{\mathbf{w}}$. This proves the injectivity of the mapping (21).

Note that $\bar{\mathbf{v}}_0$ is a vertex of the hypercube and $\sum_i \lambda_i \bar{\mathbf{v}}_i$ is an element of the hypercube $[0, 1]^{n+1}$. But $\bar{\mathbf{v}}_0$ as a vertex of the hypercube cannot be represented as a convex combination of other vertices of the hypercube $[0, 1]^{n+1}$ different from $\bar{\mathbf{v}}_0$, which by means of the injectivity of (21) shows that assumption (20) cannot be true. Consequently, we get $\mathcal{D}^{(n)} = \text{vert}(\text{conv}(\mathcal{D}^{(n)}))$, and together with the first part of the proof, $\text{conv}(\mathcal{D}^{(n)}) = B_D^{(n)}$, we conclude that $\text{vert}(B_D^{(n)}) = \mathcal{D}^{(n)}$, which ends the proof. \square

Next, the main result that characterizes the unit ball of the discrepancy norm by means of the mapping (21) is stated.

Theorem 1 Let $B_D^{(n)}$ denote the n -dimensional unit ball of the discrepancy norm, $n \geq 1$.

- (a) The mapping $\mathcal{D} : \{0, 1\}^{n+1} \setminus \{\mathbf{0}, \mathbf{1}\} \mapsto \text{vert}(B_D^{(n)})$ given by $\mathcal{D}((x_i)_{i \in I_{n+1}}) = (x_{i+1} - x_i)_{i \in I_n}$ is a one-to-one correspondence.

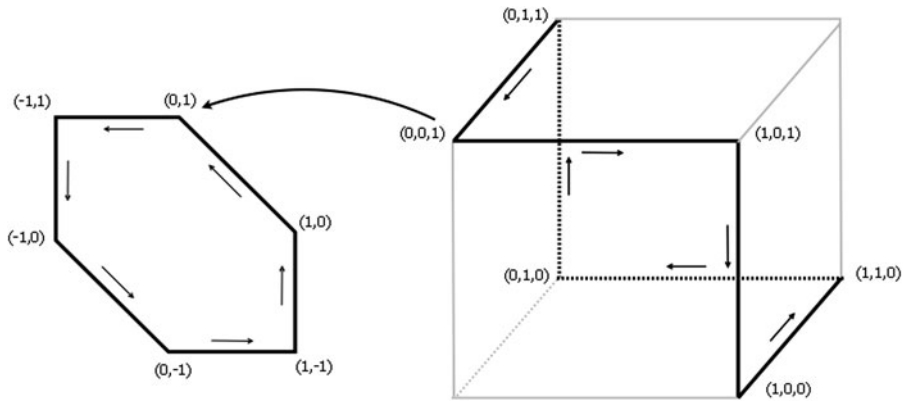


Fig. 3 Illustration of bijection between the set of vertices of $B_D^{(n)}$ and those of $[0, 1]^{n+1}$ without the diagonal elements by means of (21) for $N = 2$

(b) $B_D^{(n)} = \text{conv}(\{(y_{i+1} - y_i)_{i=1}^n \mid y_i \in \{0, 1\}\})$.

Proof The injectivity of the mapping (21) can be shown by induction. In order to prove the surjectivity of the mapping (21), let us consider $\mathbf{x} \in \text{vert}(B_D^{(n)})$, which by Lemma 3 is equivalent to $\|\mathbf{x}\|_D = 1$ and $\mathbf{x} \in \{-1, 0, 1\}^n$. Due to Lemma 1, there is a uniquely determined integration constant $c = 1 - \max_{k \in I_n} \{0, \sum_{j=1}^k x_j\} \in [0, 1]$ such that $(c, c + x_1, \dots, c + x_1 + \dots + x_n) \in [0, 1]^{n+1}$. The assumption $x_i \in \{-1, 0, 1\}$ and $\|\mathbf{x}\|_D = 1$ therefore implies $\mathbf{y} = (c, c + x_1, \dots, c + x_1 + \dots + x_n) \in \{0, 1\}^{n+1} \setminus \{\mathbf{0}, \mathbf{1}\}$; hence there is a sequence $y = (y_i)_{i \in I_{n+1}} \in \{0, 1\}^{n+1} \setminus \{\mathbf{0}, \mathbf{1}\}$ such that $(x_i)_{i \in I_n} = (y_{i+1} - y_i)_{i \in I_n}$. Equation (b) of Theorem 1 directly follows from the bijectivity of the mapping (21). □

Figure 3 illustrates the bijectivity of the mapping (21) for $n = 2$.

4 Geometric Characteristics of the n -Dimensional Unit Ball of the Discrepancy Norm

In this section Theorem 1 is applied in order to determine geometric characteristics of $B_D^{(n)}$ like the number of k -dimensional faces and its volume.

4.1 Number of k -Dimensional Faces

The following corollary relates the number of k -dimensional faces of the n -dimensional unit ball of the discrepancy norm to the number of corresponding faces of the $(n + 1)$ -dimensional hypercube.

Corollary 1 *Let $D_{k,n}$ denote the number of k -faces of $B_D^{(n)}$, $n \in \mathbb{N}$, $0 \leq k < n$. Then*

$$D_{k,n} = 2 \cdot \binom{n+1}{k} (2^{n-k} - 1). \tag{23}$$

Proof First of all let us denote by $H_{k,n}$ the number of k -dimensional faces of the n -dimensional unit hypercube $[0, 1]^n$. As it is well known from the theory of regular polytopes, see, e.g., [11], we have $H_{k,n} = 2^{n-k} \binom{n}{k}$. Observe that k -faces of the $(n + 1)$ -hypercube cannot contain the elements $\mathbf{0}$ and $\mathbf{1}$ together if $0 \leq k < n + 1$, $n \geq 1$. Therefore, for $0 \leq k < n + 1$, there are $Z_{k,n+1} = 2 \binom{n+1}{k}$ k -faces of the $(n + 1)$ -hypercube that contain either $\mathbf{0}$ or $\mathbf{1}$. Further, observe that for $k = n$, all k -faces contain either $\mathbf{0}$ or $\mathbf{1}$, which also can be seen from the identity $Z_{n,n+1} = H_{n,n+1}$.

Now we consider $k < n$ and apply the mapping \mathcal{D} of Theorem 1 on the $(n + 1)$ -hypercube $[0, 1]^{n+1}$. Observe that the elements $\mathbf{0}, \mathbf{1} \in [0, 1]^{n+1}$ are mapped to the inner point $\mathbf{0}$ of $B_D^{(n)}$. Note that a k -face F of the $(n + 1)$ -hypercube can be represented by means of a linear combination of $(n + 1)$ vertices, i.e., $F = \{\mathbf{e}_{i_0} + \sum_{l=1}^k \lambda_l \mathbf{e}_{i_l} \mid \lambda_l \in [0, 1]\}$, where \mathbf{e}_i denotes the i th unit vector. We show that a k -dimensional face of the $(n + 1)$ -hypercube which does not contain either $\mathbf{0}$ or $\mathbf{1}$ is mapped to a k -dimensional face of $B_D^{(n)}$.

For this, let us consider the set of linearly independent vectors $\{\mathbf{e}_{i_0} - \mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_0} - \mathbf{e}_{i_k}\}$. The linear independency of the mapped vectors $\{\mathcal{D}(\mathbf{e}_{i_0} - \mathbf{e}_{i_1}), \dots, \mathcal{D}(\mathbf{e}_{i_0} - \mathbf{e}_{i_k})\}$ follows from the observation that $\sum_{l=1}^k \lambda_l \mathcal{D}(\mathbf{e}_{i_0} - \mathbf{e}_{i_l}) = \mathbf{0}$, i.e., $\mathcal{D}(\sum_{l=1}^k \lambda_l (\mathbf{e}_{i_0} - \mathbf{e}_{i_l})) = \mathbf{0}$, can only be satisfied if there is a real $c \in \mathbb{R}$ such that $\sum_{l=1}^k \lambda_l (\mathbf{e}_{i_0} - \mathbf{e}_{i_l}) = c\mathbf{1}$. Since $k < n$ and $\mathbf{e}_{i_l} \in \{0, 1\}^{n+1}$, there is an index $k^* \in \{1, \dots, n + 1\}$ for which the corresponding coordinate is zero for all vectors \mathbf{e}_{i_l} , $l \in \{0, \dots, k\}$. This implies $c = 0$, hence $\lambda_l = 0$ for $l \in \{1, \dots, k\}$ due to the assumption that the set of vectors $\{\mathbf{e}_{i_0} - \mathbf{e}_{i_1}, \dots, \mathbf{e}_{i_0} - \mathbf{e}_{i_k}\}$ is linearly independent. From this and from Theorem 1 it follows that there is a one-to-one mapping between the set of k -faces of the $(n + 1)$ -hypercube that do not contain $\mathbf{0}$ or $\mathbf{1}$ to the set of k -faces of $B_D^{(n)}$. This implies $D_{k,n} = H_{k,n+1} - Z_{k,n+1}$, which equals (23). \square

In particular, $B_D^{(n)}$ has $D_{0,n} = 2^{n+1} - 2$ vertices, $D_{1,n} = (n + 1)(2^n - 2)$ edges, and $D_{n-1,n} = n(n + 1)$ facets $\mathcal{D}(F_{ij})$, $i \neq j$, of dimension $(n - 1)$, where $F_{ij} = \{\mathbf{e}_i + \sum_{k \neq i,j} \lambda_k \mathbf{e}_k \mid 0 \leq \lambda_k \leq 1\}$ and $i, j \in I_{n+1}$. Note that $F_{ij} = F_{ji} + (\mathbf{e}_i - \mathbf{e}_j)$.

4.2 Volume of $B_D^{(n)}$

Using the terminology of the theory of convex polytopes, see, e.g., [14, 41], the volume can be obtained by looking at $B_D^{(n)}$ as zonotope generated by a projection from the hypercube $[0, 1]^{n+1}$ followed by a product of shearing transformations. The $(n + 1)$ unit vectors $\mathbf{e}_1, \dots, \mathbf{e}_{n+1}$ are mapped to the Minkowski sum generators $\mathbf{g}_1, \dots, \mathbf{g}_{n+1}$. Since the unit ball is decomposed by the sheared projection of n faces out of $(n + 1)$ possibilities with each face mapped to an n -dimensional subbody of $B_D^{(n)}$ of volume one, we obtain Corollary 2.

Corollary 2 $V(B_D^{(n)}) = n + 1$, $n \in \mathbb{N}$.

It is interesting that the volume $V(B_D^{(n)}) = n + 1$ of the n -dimensional unit ball of the discrepancy norm increases linearly with its dimension n , while for p -norms with $1 \leq p < \infty$, it can be shown that $r^n V(B_{\|\cdot\|_p}^{(n)}) \xrightarrow{n \rightarrow \infty} 0$ for any $r > 0$, see, e.g., [15, 33].

5 Conclusion

In this paper Weyl’s discrepancy norm was studied from a geometrical point of view by considering its unit ball. Thinking of sequences and differentiation in the sense that two consecutive entries are subtracted, it was shown that the unit ball of Weyl’s discrepancy norm of dimension n results from differentiating the unit hypercube of dimension $(n + 1)$. It was shown how this interpretation helps to derive and prove properties of the discrepancy norm, as, for example, that the volume of the unit ball of dimension n equals $n + 1$. This paper was motivated by considering the discrepancy norm as dissimilarity measure for pattern analysis. In the near future, it is planned to investigate the relevance of the discrepancy norm in various fields of pure and applied mathematics. Particularly, research will be dedicated to the determination of the distribution of the diameter of a random walk, the discrete mathematical foundation of event-based image processing, and the improvement of stereo matching and related algorithms.

Acknowledgements This work is partially supported by the Austrian Science Fund, FWF, grant no. P21496 N23, and the Austrian COMET program.

Thanks to Peter Haslinger, Gernot Stübl, and Jean-Luc Bouchot for fruitful discussions and support for preparing the computer graphics and Johannes Himmelbauer for careful proof reading.

Appendix

Here the proofs of properties (P1)–(P6) of Sect. 2 are outlined.

$\|(x_i)_{i \in I_n}\|_D = 0$ implies $\max_{n,m \in \mathcal{I}} |\sum_{i=m}^n x_i| = 0$; hence $x(i) = 0$ for all $i \in I_n$. Homogeneity and triangle inequality immediately follow from Eq. (3). For (P3), let us consider the sequence $(x_i)_{i \in I_n}$, and let us set $x_i := 0$ for $i \notin I_n$. We are distinguishing two cases. First, let us suppose that there are indices \tilde{n}_1, \tilde{n}_2 with $\sum_{i=-\infty}^{\tilde{n}_1} x_i < 0$ and $\sum_{i=-\infty}^{\tilde{n}_2} x_i > 0$. Then

$$\begin{aligned} \|(x_i)_{i \in I_n}\|_D &= \max_{n_1, n_2 \in I_n} \left| \sum_{i=-\infty}^{n_2} x_i - \sum_{i=-\infty}^{n_1-1} x_i \right| = \max_{n_2 \in \mathbb{Z}} \sum_{i=-\infty}^{n_2} x_i - \min_{n_1 \in \mathbb{Z}} \sum_{i=-\infty}^{n_1} x_i \\ &= \max \left\{ 0, \max_{k \in I_n} \sum_{i=1}^k x_i \right\} - \min \left\{ 0, \min_{k \in I_n} \sum_{i=1}^k x_i \right\}. \end{aligned}$$

Secondly, if either $\sum_{i=-\infty}^{\tilde{k}} x_i \leq 0$ for all $k \in I_k$ or $\sum_{i=-\infty}^{\tilde{k}} x_i \geq 0$ for all $k \in I_k$ we also get $\|(x_i)_{i \in I_n}\|_D = \max\{0, \max_{k \in I_n} \sum_{i=1}^k x_i\} - \min\{0, \min_{k \in I_n} \sum_{i=1}^k x_i\}$. The Lipschitz property (P4), the symmetry property (P5), and the monotonicity (P6) follow from the observation

$$\|(x_i - x_{i-k})_i\|_D = \max \left\{ 0, \sup_{n \in \mathbb{Z}} \sum_{i=n-k+1}^n x_i \right\} - \min \left\{ 0, \underbrace{\inf_{n \in \mathbb{Z}} \sum_{i=n-k+1}^n x_i}_0 \right\}, \tag{24}$$

where $x_i \geq 0$ is assumed.

References

- Alexander, J.R., Beck, J., Chen, W.W.L.: Geometric discrepancy theory and uniform distribution. In: Handbook of Discrete and Computational Geometry, pp. 185–207. CRC Press, Inc., Boca Raton (1997)
- Bauer, P., Bodenhofer, U., Klement, E.P.: A fuzzy algorithm for pixel classification based on the discrepancy norm. In: Proc. 5th IEEE Int. Conf. on Fuzzy Systems, New Orleans, LA, vol. III, pp. 2007–2012 (1996)
- Beck, J., Chen, W.W.L.: Irregularities of Distribution. Cambridge University Press, New York (2009)
- Bercher, J.F.: On some entropy functionals derived from Rényi information divergence. *Inf. Sci.* **178**(12), 2489–2506 (2008)
- Bouchot, J., Himmelbauer, J., Moser, B.: On autocorrelation based on Hermann Weyl's discrepancy norm for time series analysis. In: Proc. 2010 Int. Joint Conf. on Neural Networks (IJCNN'10), pp. 1–7 (2010)
- Bouchot, J., Stübl, G., Moser, B.: A template matching approach based on the discrepancy norm for defect detection on regularly textured surfaces. In: Pinoli, J., Debayle, J., Gavet, Y., Gruy, F., Lambert, C. (eds.) Proc. 10th Intern. Conf. on Quality Control by Artificial Vision (QCAV'11), vol. 8000, p. 80000K. SPIE, Bellingham (2011)
- Brunelli, R.: Template Matching Techniques in Computer Vision: Theory and Practice. Wiley, New York (2009)
- Campbell, J., Lo, A., MacKinlay, C.: The Econometrics of Financial Markets. Princeton University Press, Princeton (1996)
- Chazelle, B.: The Discrepancy Method: Randomness and Complexity. Cambridge University Press, New York (2000)
- Chen, C.M., Cheng, C.T.: From discrepancy to declustering: near-optimal multidimensional declustering strategies for range queries. *J. ACM* **51**(1), 46–73 (2004)
- Coxeter, H.S.M.: Regular Polytopes. Dover, New York (1973)
- Doerr, B.: Integral approximations. Habilitation thesis, University of Kiel (2005)
- Doerr, B., Hebbinghaus, N., Werth, S.: Improved bounds and schemes for the declustering problem. *Theor. Comput. Sci.* **359**(1), 123–132 (2006)
- Grünbaum, B.: Convex Polytopes. Graduate Texts in Mathematics, vol. 221. Springer, New York (2003)
- Huang, Z.Y., He, B.W.: Volume of unit ball in an n -dimensional normed space and its asymptotic properties. *J. Shanghai Univ.* **12**, 107–109 (2008)
- Jancosek, M., Pajdla, T.: A criterial function for superpixel matching (2010)
- Kendall, M., Stuart, A.: The Advanced Theory of Statistics, vol. 2. MacMillan, New York (1979)
- Kogler, J., Sulzbachner, C., Kubinger, W.: Bio-inspired stereo vision system with Silicon Retina imagers. In: Fritz, M., Schiele, B., Piater, J. (eds.) Computer Vision Systems. Lecture Notes in Computer Science, vol. 5815, pp. 174–183. Springer, Berlin (2009)
- Kuipers, L., Niederreiter, H.: Uniform Distribution of Sequences. Dover Publications, New York (2005)
- Kullback, S.: Information Theory and Statistics. Wiley, New York (1959)
- Matousek, J.: Geometric Discrepancy: An Illustrated Guide (Algorithms and Combinatorics), 1st edn. Springer, Berlin (1999)
- Moser, B.: Discrete geometric foundation of event based imaging. Tech. Rep. SCCH-TR-2011-44, Software Competence Center Hagenberg (2011)
- Moser, B.: A similarity measure for image and volumetric data based on Hermann Weyl's discrepancy. *IEEE Trans. Pattern Anal. Mach. Intell.* **33**(11), 2321–2329 (2011)
- Moser, B., Hoch, T.: Misalignment measure based on Hermann Weyl's discrepancy. In: Kuijper, A., Heise, B., Muresan, L. (eds.) Challenges in the Biosciences: Image Analysis and Pattern Recognition Aspects, Linz, Austria, vol. 232, pp. 187–198 (2008)
- Moser, B., Haslinger, P., Kazmar, T.: On the potential of Hermann Weyl's discrepancy norm for texture analysis. In: Masoud, M. (ed.) Proc. 2008 Int. Conf. on Computational Intelligence for Modelling, Control and Automation; Intelligent Agents, Web Technologies and Internet Commerce; and Innovation in Software Engineering (CIMCA/IAWTIC/ISE), pp. 187–191. IEEE Computer Society, Washington (2008)
- Moser, B., Stübl, G., Bouchot, J.L.: On a non-monotonicity effect of similarity measures. In: Proc. 1st Int. Conf. on Similarity-Based Pattern Recognition, SIMBAD'11, pp. 46–60. Springer, Berlin (2011)

27. Neunzert, H., Wetton, B.: Pattern recognition using measure space metrics. Tech. Rep. 28, University of Kaiserslautern, Department of Mathematics (1987)
28. Niederreiter, H.: Random Number Generation and Quasi-Monte Carlo Methods. Society for Industrial and Applied Mathematics, Philadelphia (1992)
29. Pluim, J., Maintz, J., Viergever, M.: f-Information measures in medical image registration. *IEEE Trans. Med. Imaging* **23**(12), 1506–1518 (2004)
30. Renyi, A.: On measures of entropy and information. In: Proc. 4th Berkeley Symposium on Mathematical Statistics and Probability, vol. 1, pp. 547–561 (1961)
31. Sadakane, K., Chebihi, N., Tokuyama, T.: Discrepancy-based digital halftoning: automatic evaluation and optimization. *Interdiscip. Inf. Sci.* **8**(2), 219–234 (2002)
32. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *Int. J. Comput. Vis.* **47**, 7–42 (2001)
33. Smith, D.J., Vamanamurthy, M.: How small is a unit ball? *Math. Mag.* **62**, 101–107 (1989)
34. Spearman, C.: The proof and measurement of association between two things. *Am. J. Psychol.* **15**, 72–101 (1904)
35. Stübl, G., Bouchot, J.L., Haslinger, P., Moser, B.: Discrepancy norm as fitness function for defect detection on regularly textured surfaces. Tech. Rep. SCCH-TR-, Software Competence Center Hagenberg, Austria (2012)
36. Takhtamyshev, G., Vandewoestyne, B., Cools, R.: Quasi-random integration in high dimensions. *Math. Comput. Simul.* **73**(5), 309–319 (2007)
37. Vajda, I.: Theory of Statistical Inference and Information. Kluwer Academic, Dordrecht (1989)
38. Victor, J.D.: Spike train metrics. *Curr. Opin. Neurobiol.* **15**(5), 585–592 (2005)
39. Weyl, H.: Über die Gleichverteilung von Zahlen mod. Eins. *Math. Ann.* **77**, 313–352 (1916)
40. Zaremba, S.: The mathematical basis of Monte Carlo and quasi-Monte Carlo methods. *SIAM Rev.* **10**(3), 303–314 (1968)
41. Ziegler, G.M.: Lectures on Polytopes. Graduate Texts in Mathematics, vol. 152. Springer, New York (1995)