Multidimensional variation for quasi-Monte Carlo

Art B. Owen
Stanford University

January 2004

Dedicated to Professor Fang Kai-Tai in honor of his 65th birthday

1 Introduction

This paper collects together some properties of multidimensional definitions of the total variation of a real valued function. The subject has been studied for a long time. Many of the results presented here date back at least to the early 1900s.

The main reason for revisiting this topic is that there has been much recent work in theory and applications of Quasi-Monte Carlo (QMC) sampling. For an account of quasi-Monte Carlo integration see Fang and Wang (1994) and Niederreiter (1992). QMC is especially competitive for multidimensional integrands with bounded variation in the sense of Hardy and Krause (BVHK). For such integrands, over d dimensional domains, one sees QMC errors that are $O(n^{-1}(\log n)^d)$ when using n function evaluations. When $d = 1$, competing methods are usually preferred to QMC. For even modestly large d, Monte Carlo and quasi-Monte Carlo sampling become the methods of choice.

When the integrand is in BVHK, then QMC has superior asymptotic behavior, compared to Monte Carlo sampling. Therefore we may like to know when a specific function is in BVHK. Recent introductory text books on real analysis typically cover the notion of total variation for functions of a single real variable. Few of them say much about multidimensional variation. The not very recent book, Hobson (1927, Chapter 5), does include some discussion of variation beyond the one dimensional case.

Discussions of multidimensional variation usually require ungainly expressions that grow in complexity with the dimension d. For this reason, many authors work out details for $d = 2$ and report that the same results hold for all d. Yet some results that hold for $d = 2$ do not hold for $d > 2$. For example an indicator function in two dimensions must either have positive variation in Vitali's sense, or must have at least one input variable on which it does not truly depend. The same is not true for $d \geq 3$. Similarly, if $f(x)$ and $g(x)$ are linear functions on the d dimensional cube, then $\min(f(x), g(x))$ is BVHK when $d = 2$ but is not necessarily so when $d > 2$.

A result known to hold for all d can be nicely communicated for the case $d = 2$. But to decide if a result holds for all d it is better to use formulas that look the same for all dimensions d. The underlying mathematical operations employed in studying variation require selection and manipulation of subsets of the components of the argument to the integrand. By using these subsets themselves as indices, it is possible to get compact expressions that hold for all $d \geq 1$.

2 One dimensional variation

Let $f(x)$ be a real valued function defined on $[a, b]$ where $-\infty < a \leq b < \infty$. A "ladder" on $[a, b]$ is a set $\mathcal{Y}$ containing a and finitely many, possibly zero, values from (a, b). The ladder $\mathcal{Y}$ does not contain b except when $a = b$. This case is clearly degenerate, but in some settings below it is simpler to include it than to exclude it. Each element $y \in \mathcal{Y}$ has a successor y_+. If $(y, \infty) \cap \mathcal{Y} = \emptyset$ then $y_+ = b$ and otherwise y_+ is the smallest element of $(y, \infty) \cap \mathcal{Y}$. If the elements of $\mathcal{Y}$ are arranged into increasing order, $a = y_0 < y_1 < \cdots < y_m$, then the successor of y_k is y_{k+1} for $k < m$ and it is b for $k = m$. The value y_+ depends on $\mathcal{Y}$ but this dependence will not be made explicit by the notation.

Let $\mathbb{Y}$ denote the set of all ladders on $[a, b]$. Then the total variation of f on $[a, b]$ is

$$V(f; a, b) = \sup_{\mathcal{Y} \in \mathbb{Y}} \sum_{y \in \mathcal{Y}} |f(y_+) - f(y)|. \tag{1}$$

This variation is written $V(f)$ when $[a, b]$ is understood from context. If $V(f) < \infty$ then f is of bounded variation.

The following properties of one dimensional variation are well known:

1. If f is monotone on $[a, b]$, then $V(f; a, b) = |f(b) - f(a)|$.
2. If f has a first derivative f' on $[a, b]$ then $V(f; a, b) = \int_a^b |f'(x)| dx$.
3. For functions f, g on $[a, b]$, $V(f + g; a, b) \leq V(f; a, b) + V(g; a, b)$.
4. For a function f on $[a, b]$ and a scalar α, $V(\alpha f; a, b) = |\alpha| V(f; a, b)$.
5. f is of bounded variation on $[a, b]$ if and only if f can be written as the difference of two bounded monotone functions on $[a, b]$.
6. For $c \in [a, b]$, $V(f; a, b) = V(f; a, c) + V(f; c, b)$.

Item 6 is very useful in QMC settings when extended to $d \geq 1$. Notice that both intervals $[a, c]$ and $[c, b]$ include the point c.

There is no uniquely suitable way to extend the notion of variation to functions of more than one variable. Clarkson and Adams (1933) study six such generalizations, and Adams and Clarkson (1934) mention two more. For quasi-Monte Carlo, the total variation in the sense of Hardy and Krause is the most widely used definition. The early references for that definition are Hardy (1905) and Krause (1903b, Krause (1903a), who were studying double Fourier series. That definition of total variation is constructed using the total variation in Vitali's sense. Only these two definitions are considered in this work.

3 Notation

For $x \in \mathbb{R}^d$, write its j'th component as x^j. Thus $x = (x^1, \ldots, x^d)$. For $a, b \in \mathbb{R}^d$ write $a < b$ or $a \le b$ if these inequalities hold for all d components. For $a, b \in \mathbb{R}^d$ with $a \le b$, the hyperrectangle $[a, b]$ is the set $\{x \in \mathbb{R}^d \mid a \le x \le b\}$. Also $(a, b) = \{x \in \mathbb{R}^d \mid a < x < b\}$ and $[a, b)$ and $(a, b]$ are defined similarly. The d dimensional volume $\prod_{j=1}^d (b^j - a^j)$ of $[a, b]$ is denoted $\mathrm{Vol}([a, b])$.

For arbitrary points $a, b \in \mathbb{R}^d$, let $\mathrm{rect}[a, b]$ denote the hyperrectangle $[\widetilde{a}, \widetilde{b}]$ with $\widetilde{a}^j = \min(a^j, b^j)$ and $\widetilde{b}^j = \max(a^j, b^j)$. We can think of $\mathrm{rect}[a, b]$ as the "rectangular hull" of $\{a, b\}$.

For $u, v \subseteq \{1, \ldots, d\}$ write $|u|$ for the cardinality of u, and $u - v$ for the complement of v with respect to u. For integers $j \le k$, the set $\{j, j+1, \ldots, k\}$ is written $j{:}k$. A unary minus, denotes the complement with respect to $1{:}d$, so that $-u = 1{:}d - u$. In expressions such as $1{:}d - \{j\}$ and $j{:}k \cup u$ the : symbol has highest precedence. In $-u - v$, the unary minus has higher precedence than the binary minus.

For $u \subseteq 1{:}d$, the expression x^u denotes a $|u|$-tuple of real values representing the components x^j for $j \in u$. The domain of x^u is the hyperrectangle $[a^u, b^u]$. Suppose that $u, v \subseteq 1{:}d$ and $x, z \in [a, b]$ with $u \cap v = \emptyset$. Then the symbol $x^u{:}z^v$ represents the point $y \in [a^{u\cup v}, b^{u\cup v}]$ with $y^j = x^j$ for $j \in u$, and $y^j = z^j$ for $j \notin u$. The symbol $x^u{:}z^v$ is well defined for $x^u \in [a^u, b^u]$ and $z^v \in [a^v, b^v]$, when $u \cap v = \emptyset$, even if x^{-u} or z^{-v} is left unspecified. We also use the : symbol to glue together more than two sets of components. For instance $x^u{:}y^v{:}z^w \in [a, b]$ is well defined for $x^u \in [a^u, b^u]$, $y^v \in [a^v, b^v]$, and $z^w \in [a^w, b^w]$, when u, v, w are mutually disjoint sets whose union is $1{:}d$. It will be clear from context whether : pieces together a tuple, as in $x^u{:}x^v$, or denotes a range of integers as in $j{:}k$. The main use of the gluing symbol is to construct the argument to a function by taking components from multiple sources.

Let $f(x)$ be a real valued function defined on the hyperrectangle $[a, b]$. The function f does not depend on x^u if $f(x^u{:}z^{-u}) = f(x^u{:}x^{-u})$ holds for all $x^{-u}, z^{-u} \in [a^{-u}, b^{-u}]$. Similarly, f is a function of x^u alone, if it does not depend on x^{-u}.

For $u \subseteq 1{:}d$ and $x^{-u} \in [a^{-u}, b^{-u}]$ we can define a function g on $[a^u, b^u]$ via $g(x^u) = f(x^u{:}x^{-u})$. We write $f(x^u; x^{-u})$ to denote such a function with the argument x^u on the left of the semi-colon and the parameter x^{-u} on the right.

Many expressions require no special attention for $u = \emptyset$. For instance, when $u = \emptyset$, then the definition of $x^u{:}z^{-u}$ reduces to z. In some other settings, the index set u must be handled specially when it equals $\emptyset$. It is often less trouble to adopt a simplifying convention for $u = \emptyset$ than to explicitly identify it as a special case.

Zero dimensional regions and functions on them are of no direct interest in this work. They do however appear as special cases in some derivations. In the sequel, $x^{\emptyset}$ denotes the "zero-tuple" (), the Cartesian product of zero sets is the set containing the zero-tuple, and the volume of a zero dimensional rectangle is $\prod_{j\in\emptyset}(b^j - a^j) = 1$, just as empty products are conventionally taken to be one. A function f on $[a^{\emptyset}, b^{\emptyset}]$ is necessarily constant, with a value denoted by $f()$.

The hyperrectangles mentioned in the statements of Propositions are assumed to have a dimension d with $1 \le d < \infty$ without repeatedly stating so. Because the components of a and b are real valued we always have $\mathrm{Vol}([a, b]) < \infty$. Proposition statements do include cases with $\mathrm{Vol}([a, b]) = 0$ unless otherwise specified.

4 Multidimensional variation

The d-fold alternating sum of f over $[a, b]$ is

$$\Delta(f; a, b) = \sum_{v\subseteq\{1,\dots,d\}} (-1)^{|v|} f(a^v{:}b^{-v}). \tag{2}$$

Note particularly that in (2), the coefficient of $f(b)$ is one while that of $f(a)$ is $(-1)^d$. Sometimes it will be convenient to write a d-fold alternating sum as $\Delta(f; s)$ where s is a closed hyperrectangle. For $u \subseteq 1{:}d$, define

$$\Delta_u(f; a, b) = \sum_{v\subseteq u} (-1)^{|v|} f(a^v{:}b^{-v}). \tag{3}$$

Notice that $\Delta_u(f; a, b)$ does not depend on a^{-u}.

The alternating sums (2) and (3) are well defined even when $a \le b$ does not hold. In general $\Delta(f; a, b) = \pm\Delta(f; \mathrm{rect}[a, b])$. The sign is negative if $a^j > b^j$ holds for an odd number of $j \in 1{:}d$.

For each $j = 1, \dots, d$ let $\mathcal{Y}^j$ be a ladder on $[a^j, b^j]$. A (multidimensional) ladder on $[a, b]$ has the form $\mathcal{Y} = \prod_{j=1}^d \mathcal{Y}^j$, and we also use $\mathcal{Y}^u = \prod_{j\in u} \mathcal{Y}^j$. For $y \in \mathcal{Y}$, the successor point y_+ is defined by taking y_+^j to be the successor of y^j in $\mathcal{Y}^j$. The variation of f over $\mathcal{Y}$ is

$$V_{\mathcal{Y}}(f) = \sum_{y\in\mathcal{Y}} |\Delta(f; y, y_+)|. \tag{4}$$

A ladder is, with minor differences, what Clarkson and Adams (1933) call a "net". Their nets also include upper boundaries from b. Ladders are sets, which allows some manipulations to be economically written. We avoid the term net here, because in quasi-Monte Carlo, a net is a finite list of points satisfying some equidistribution properties.

For the multidimensional setting, let $\mathbb{Y}^j$ denote the set of all ladders on $[a^j, b^j]$, and put $\mathbb{Y} = \prod_{j=1}^d \mathbb{Y}^j$. Then:

Definition 1 *The variation of f on the hyperrectangle $[a,b]$, in the sense of Vitali, is*

$$V_{[a,b]}(f) = \sup_{\mathcal{Y}\in\mathbb{Y}} V_{\mathcal{Y}}(f).$$

When $[a,b]$ is understood, we simply write $V(f)$. The function f is of bounded variation in Vitali's sense (BV) if $V(f) < \infty$.

As described below, variation in the sense of Vitali is not adequate for the study of quasi-Monte Carlo sampling. Instead, the variation in the sense of Hardy and Krause is used. This notion of variation sums the Vitali variations over $[a,b]$ and its "upper faces".

Definition 2 *The variation of f on the hyperrectangle $[a,b]$, in the sense of Hardy and Krause, is*

$$V_{\mathrm{HK}}(f) = V_{\mathrm{HK}}(f;a,b) = \sum_{u\subsetneq 1:d} V_{[a^{-u},b^{-u}]} f(x^{-u};b^{u}). \tag{5}$$

The function f has bounded variation in the sense of Hardy and Krause (BVHK) if $V_{\mathrm{HK}}(f) < \infty$. The definition of bounded variation in Hardy (1905) requires $V_{[a,b]}(f) < \infty$ and $V_{[a^{-u},b^{-u}]}(f(x^{-u};z^{u})) < \infty$ for all $0 < |u| < d$ and all $z^{u} \in [a^{u},b^{u}]$. Young (1913) shows that definition to be equivalent to the one above.

The premier use of variation in QMC is in Hlawka's inequality (the Koksma-Hlawka theorem) where the quadrature error has an upper bound equal to $V_{\mathrm{HK}}(f)$ times a discrepancy measure of the points $x_1,\dots,x_n$. See Niederreiter (1992).

If one follows the above definitions literally for a zero dimensional hyperrectangle, then $V_{[a^{\emptyset},b^{\emptyset}]}(f) = |f()|$. The variation $V_{\mathrm{HK}}(f)$ is a semi-norm on functions and not a norm, because it vanishes for constant but non-zero functions. The quantity $V_{\mathrm{HK}}(f) + |f(b)|$ is often used in QMC because it is a norm on functions. This norm can be obtained adjoining the case $u = 1{:}d$ to the sum in (5).

5 Splits of hyperrectangles

The properties of variation derive from those of alternating sums. Those in turn are based on properties of splits of hyperrectangles.

Definition 3 *A split of the hyperrectangle $[a,b]$ is a set $\{[a_i,b_i] \mid 1 \le i \le m < \infty\}$ where $\cup_{i=1}^{m}[a_i,b_i] = [a,b]$ and $[a_i,b_i) \cap [a_j,b_j) = \emptyset$ when $i \ne j$.*

Note that $[a_i,b_i] \cap [a_j,b_j]$ is not necessarily empty for $i \ne j$. The most basic split is a coordinate split:

Definition 4 *For $j \in \{1,\dots,d\}$ and $c \in [a^{j},b^{j}]$ the corresponding coordinate split of $[a,b]$ is the set $\{L,R\}$ of left and right pieces*

$$L = L(j,c) = L(j,c;a,b) = \{x \in [a,b] \mid x^{j} \le c\}, \quad \textit{and,}$$
$$R = R(j,c) = R(j,c;a,b) = \{x \in [a,b] \mid x^{j} \ge c\},$$

respectively.

Both L and R are closed hyperrectangles: $L = [a, \widetilde{b}]$ where $\widetilde{b}^k = b^k$ for $k \neq j$ and $\widetilde{b}^j = c$, and $R = [\widetilde{a}, b]$ where $\widetilde{a}^k = a^k$ for $k \neq j$ and $\widetilde{a}^j = c$. Next we show that the alternating sum over $[a, b]$ is the sum of alternating sums over L and R. Propositions 1 through 4 below recapitulate results from Fréchet (1910).

Proposition 1 *Suppose that the hyperrectangle* $[a, b]$ *is split into* $[a, \widetilde{b}]$ *and* $[\widetilde{a}, b]$ *as described above. Then*

$$\Delta(f; a, b) = \Delta(f; a, \widetilde{b}) + \Delta(f; \widetilde{a}, b). \tag{6}$$

Proof: Let $c^{\{j\}}$ denote $c \in [a^j, b^j]$ for use as an argument to f. We write the sum over $v \subseteq 1{:}d$ as a double sum. The outer sum is over $u \subseteq 1{:}d - \{j\}$ and the inner sum is over u and $u \cup \{j\}$. Thus

$$\begin{aligned}\Delta(f; a, \widetilde{b}) &= \sum_{u \subseteq -\{j\}} (-1)^{|u|} \bigl[f(a^u : \widetilde{b}^{-u}) - f(a^{u \cup \{j\}} : \widetilde{b}^{-u-\{j\}}) \bigr] \\ &= \sum_{u \subseteq -\{j\}} (-1)^{|u|} \bigl[f(a^u : c^{\{j\}} : b^{-u-\{j\}}) - f(a^{u \cup \{j\}} : b^{-u-\{j\}}) \bigr], \end{aligned} \tag{7}$$

and similarly,

$$\Delta(f; \widetilde{a}, b) = \sum_{u \subseteq -\{j\}} (-1)^{|u|} \bigl[f(a^u : b^{-u}) - f(a^u : c^{\{j\}} : b^{-u-\{j\}}) \bigr]. \tag{8}$$

Summing (7) and (8),

$$\sum_{u \subseteq 1:d-\{j\}} (-1)^{|u|} \bigl[f(a^u : b^{-u}) - f(a^{u \cup \{j\}} : b^{-u-\{j\}}) \bigr] = \Delta(f; a, b). \quad \square$$

Proposition 2 *Suppose that* $\mathcal{Y} = \prod_{j=1}^d \mathcal{Y}^j$ *is a ladder on* $[a, b]$*. Then* $\{[y, y_+] \mid y \in \mathcal{Y}\}$ *is a split of* $[a, b]$ *and*

$$\Delta(f; a, b) = \sum_{y \in \mathcal{Y}} \Delta(f; y, y_+). \tag{9}$$

Proof: By construction $a \leq y \leq y_+ \leq b$ for $y \in \mathcal{Y}$, so $\cup_{y \in \mathcal{Y}} [y, y_+] \subseteq [a, b]$. Now suppose that $x \in [a, b]$. Consider $y \in \mathcal{Y}$ where $y^j = a^j$ if $x^j = a^j$, and otherwise $y^j = \max\bigl(\mathcal{Y}^j \cap [a^j, x^j)\bigr)$. Then $y^j \leq x^j \leq y_+^j$ so that $x \in [y, y_+]$, and hence $\cup_{y \in \mathcal{Y}} [y, y_+] = [a, b]$. Now suppose that $x \in [y, y_+) \cap [\widetilde{y}, \widetilde{y}_+)$ for $y, \widetilde{y} \in \mathcal{Y}$. Then $y^j \leq x^j < y_+^j$ and $\widetilde{y}^j \leq x^j < \widetilde{y}_+^j$ which implies $y^j = \widetilde{y}^j$, so $y = \widetilde{y}$. Thus $[y, y_+) \cap [\widetilde{y}, \widetilde{y}_+)$ is empty whenever $y \neq \widetilde{y}$, establishing that $\{[y, y_+] \mid y \in \mathcal{Y}\}$ is a split of $[a, b]$. To prove (9), note that the split $\{[y, y_+] \mid y \in \mathcal{Y}\}$ can be obtained by making a sequence of $|\mathcal{Y}| - 1$ coordinate splits of $[a, b]$. $\square$

Proposition 3 *Suppose that* $\{\,[a_i,b_i] \mid 1 \le i \le m < \infty\}$ *is a split of the hyperrectangle* $[a,b]$*. Then*

$$\Delta(f;a,b) = \sum_{i=1}^{m} \Delta(f;a_i,b_i). \tag{10}$$

Proof: Let $\mathcal{Y}^j = \{a_1^j,\dots,a_m^j,b_1^j,\dots,b_m^j\} \cap [a^j,b^j)$, define the ladder $\mathcal{Y} = \prod_{j=1}^d \mathcal{Y}^j$ and define the split $\mathcal{S} = \{[y,y_+] \mid y \in \mathcal{Y}\}$. Next put $\mathcal{S}_i = \{[y,y_+] \mid y \in \mathcal{Y} \cap [a_i,b_i)\}$. Then $\mathcal{S}_i$ is a split of $[a_i,b_i]$ to which Proposition 2 applies. Also $\mathcal{S}_i$ are mutually disjoint with union $\mathcal{S}$. Therefore $\Delta(f;a,b)$ and $\sum_{i=1}^m \Delta(f;a_i,b_i)$ are both equal to $\sum_{i=1}^m \sum_{s\in\mathcal{S}_i} \Delta(f;s)$. $\square$

Proposition 4 *Let* $\mathcal{Y}^j$ *and* $\widetilde{\mathcal{Y}}^j$ *be ladders in* $[a^j,b^j]$ *with* $\mathcal{Y}^j \subseteq \widetilde{\mathcal{Y}}^j$ *for* $j = 1,\dots,d$ *and write* $\mathcal{Y} = \prod_{j=1}^d \mathcal{Y}^j$ *and* $\widetilde{\mathcal{Y}} = \prod_{j=1}^d \widetilde{\mathcal{Y}}^j$*. Then* $V_{\mathcal{Y}}(f;a,b) \le V_{\widetilde{\mathcal{Y}}}(f;a,b)$*.*

Proof: The ladder $\mathcal{Y}$ can be changed to $\widetilde{\mathcal{Y}}$ by d steps that each refine just one of the ladders $\mathcal{Y}^j$. Therefore it is sufficient to consider the case where $\widetilde{\mathcal{Y}}^j = \mathcal{Y}^j$ for $j \ne k$ for some $k \in \{1,\dots,d\}$. Without loss of generality take $k=1$ and suppose that $\widetilde{\mathcal{Y}}^1 - \mathcal{Y}^1 = \{c\}$ where $y_\ell = \widetilde{y}_\ell < c = \widetilde{y}_{\ell+1} < y_{\ell+1}$ for $0 \le \ell \le m_1$ taking $y_{m_1+1} = b^1$ if $\ell = m_1$. Then $V_{\widetilde{\mathcal{Y}}}(f) - V_{\mathcal{Y}}(f)$ equals

$$\begin{aligned}
&\sum_{\widetilde{y}^{2:d}\in\mathcal{Y}^{2:d}} \sum_{\widetilde{y}^1\in\widetilde{\mathcal{Y}}^1} |\Delta(f;\widetilde{y},\widetilde{y}_+)| - \sum_{y^{2:d}\in y^{2:d}} \sum_{y^1\in\mathcal{Y}^1} |\Delta(f;y,y_+)| \\
&= \sum_{y^{2:d}\in y^{2:d}} \sum_{y^1\in\{y_\ell\}} |\Delta(f;L(1,c;y,y_+))| + |\Delta(f;R(1,c;y,y_+))| \\
&\qquad - |\Delta(f;L(1,c;y,y_+)) + \Delta(f;R(1,c;y,y_+))| \\
&\ge 0. \quad \square
\end{aligned}$$

Proposition 4 allows us to replace the supremum over all ladders in Definition 1 by one over a subset $\widetilde{\mathbb{Y}} \subseteq \mathbb{Y}$ of ladders. If to every $\mathcal{Y} \in \mathbb{Y}$ there is a $\widetilde{\mathcal{Y}} \in \widetilde{\mathbb{Y}}$ with $\mathcal{Y} \subseteq \widetilde{\mathcal{Y}}$, then

$$V(f) = \sup_{\mathcal{Y}\in\widetilde{\mathbb{Y}}} V_{\mathcal{Y}}(f).$$

For instance when $[a,b] = [0,1]^d$ we may suppose that each ladder in $\widetilde{\mathbb{Y}}$ is the d-fold tensor product of some ladder on $[0,1]$. To show this, take $\widetilde{\mathcal{Y}}^k = \cup_{j=1}^d \mathcal{Y}^j$ for $k \in 1{:}d$.

A simple ladder that is sometimes useful is one with an equal number of equispaced points in each direction. Let $m \ge 1$ be an integer and set $\mathcal{Y}^j(m) = \mathcal{Y}^j(m,a,b) = \{a^j, a^j + (b^j-a^j)/m, \dots, \dots, a^j + (b^j-a^j)(m-1)/m\}$ and put

$$\mathcal{Y}(m) = \mathcal{Y}(m,a,b) = \prod_{j=1}^d \mathcal{Y}^j(m,a,b). \tag{11}$$

Simple ladders can be used to show lower bounds on variation, but we cannot replace the supremum in Definition 1 by the supremum over simple ladders, nor even by the supremum over ladders $\mathcal{Y}$ for which $(y^j - a^j)/(b^j - a^j)$ is always a rational number. We can however restrict attention to ladders for which the cells $[y, y_+]$ are nearly congruent and nearly cubic.

Proposition 5 *Let f be a function on the hyperrectangle $[a, b]$ of positive volume. For $\epsilon > 0$, let $\widetilde{\mathbb{Y}} = \widetilde{\mathbb{Y}}(\epsilon)$ be the set of ladders $\widetilde{\mathcal{Y}}$ for which*

$$\max_{\widetilde{y}\in\widetilde{\mathcal{Y}}} \max_{j\in 1:d} (\widetilde{y}^j_+ - \widetilde{y}^j) \le (1+\epsilon) \min_{\widetilde{y}\in\widetilde{\mathcal{Y}}} \min_{j\in 1:d} (\widetilde{y}^j_+ - \widetilde{y}^j). \tag{12}$$

Then $V_{[a,b]}(f) = \sup_{\mathcal{Y}\in\widetilde{\mathbb{Y}}} V_{\mathcal{Y}}(f)$.

Proof: For $\mathcal{Y} \in \mathbb{Y}$, let $\eta = \min_{y\in\widetilde{\mathcal{Y}}} \min_{j\in 1:d}(y^j_+ - y^j)$. Because $\mathrm{Vol}([a, b]) > 0$ we have $\eta > 0$. Set $\widetilde{\mathcal{Y}} = \mathcal{Y}$. Then for each $j \in 1\!:\!d$, while there is $y^j \in \mathcal{Y}^j$ with $y^j_+ - y^j > 2\epsilon$, replace $\widetilde{\mathcal{Y}}^j$ by $\widetilde{\mathcal{Y}}^j \cup \{(y^j + y^j_+)/2\}$. After a finite number of steps $\mathcal{Y} \subseteq \widetilde{\mathcal{Y}}$ and $\widetilde{\mathcal{Y}}$ satisfies (12) for $\epsilon = 1$.

If $\epsilon \ge 1$, we're done. Otherwise, for an integer $m > 1/\epsilon$ let $k(j, \widetilde{y}^j) = \lfloor m(\widetilde{y}^j_+ - \widetilde{y}^j)/\eta \rfloor$, where $\lfloor z \rfloor$ denotes the greatest integer less than or equal to z. We have $m \le k < 2m$ because $\widetilde{y}^j_+ - \widetilde{y}^j \in [\eta, 2\eta)$. Next set $\widehat{\mathcal{Y}} = \prod_{j=1}^d \widehat{\mathcal{Y}}^j$, where

$$\widehat{\mathcal{Y}}^j = \bigcup_{\widetilde{y}^j\in\widetilde{\mathcal{Y}}^j} \bigcup_{r=0}^{k(j,\widetilde{y}^j)-1} \Big\{ \widetilde{y}^j + r(\widetilde{y}^j_+ - \widetilde{y}^j)/k(j, \widetilde{y}^j) \Big\}.$$

The interval $[\widetilde{y}^j, \widetilde{y}^j_+]$ gets split into $k = k(j, \widetilde{y}^j)$ equal width intervals. If an interval in $\widetilde{\mathcal{Y}}$ has been split k ways then its length could have been as small as $k\eta/m$ but not as large as $(k+1)\eta/m$. Thus the shortest interval in the $\hat{\mathcal{Y}}$ ladder has length η/m and the largest has length under $\max_{k\in m:(2m-1)}(k+1)\eta/(km) = (m+1)\eta/m^2$. Now $((m+1)\eta/m^2)/(\eta/m) = (m+1)/m < 1+\epsilon$ because $m > 1/\epsilon$. Thus $\mathcal{Y} \subseteq \widehat{\mathcal{Y}}$ where $\widehat{\mathcal{Y}}$ satisfies (12). $\square$

When computing or bounding $V(f)$ it is often convenient to split the domain of f into hyperrectangular regions and sum the variations from within each of them. The following lemma, stated in Young (1913), justifies such a divisive approach.

Lemma 1 *Let f be defined on the hyperrectangle $[a, b]$. Let $\{[a_i, b_i] \mid 1 \le i \le m < \infty\}$ be a split of $[a, b]$. Then*

$$V_{[a,b]}(f) = \sum_{i=1}^m V_{[a_i,b_i]}(f).$$

Proof: Let $\mathcal{Y}$ be a ladder on $[a, b]$. Let $\widetilde{\mathcal{Y}}^j = (\mathcal{Y}^j \cup \{a^j_1, \dots, a^j_m, b^j_1, \dots, b^j_m\}) \cap [a^j, b^j)$. Then

$$V_{\mathcal{Y}}(f) \le V_{\widetilde{\mathcal{Y}}}(f) = \sum_{i=1}^m \sum_{\widetilde{y}\in\widetilde{\mathcal{Y}}\cap[a_i,b_i)} |\Delta(f; \widetilde{y}, \widetilde{y}_+)| \le \sum_{i=1}^m V_{[a_i,b_i]}(f).$$

Taking the supremum over $\mathcal{Y}$ establishes that $V_{[a,b]}(f) \le \sum_{i=1}^m V_{[a_i,b_i]}(f)$. Now let $\mathcal{Y}_i$ be ladders on $[a_i, b_i]$ for $i = 1, \dots, m$. Let $\widetilde{\mathcal{Y}}$ be the ladder with $\widetilde{\mathcal{Y}}^j = \cup_{i=1}^m \mathcal{Y}_i^j$ and let $\widetilde{\mathcal{Y}}_i = \widetilde{\mathcal{Y}} \cap [a_i, b_i) \supseteq \mathcal{Y}_i$. Then

$$\sum_{i=1}^m \sum_{y \in \mathcal{Y}_i} |\Delta(f; y, y_+)| \le \sum_{i=1}^m \sum_{\widetilde{y} \in \widetilde{\mathcal{Y}}_i} |\Delta(f; \widetilde{y}, \widetilde{y}_+)| = \sum_{\widetilde{y} \in \widetilde{\mathcal{Y}}} |\Delta(f; \widetilde{y}, \widetilde{y}_+)| \le V(f).$$

Taking the supremum over $\mathcal{Y}_1, \dots, \mathcal{Y}_m$ yields $\sum_{i=1}^m V_{[a_i,b_i]}(f) \le V_{[a,b]}(f)$. $\Box$

Suppose that we seek to prove that $V(f) = \infty$. If for $m > 1$ we split $[a, b]$ into m^d congruent hyperrectangles similar in shape to $[a, b]$, then by Lemma 1, at least one of these smaller hyperrectangles has infinite variation. The proof of infinite variation can therefore always be focussed on an arbitrarily small region within $[a, b]$. Of course, matters would be different had we considered unbounded hyperrectangles.

6 Alternating sums

A function f can be easily recovered from its alternating sums, as follows:

Proposition 6 *Let f be a function on the hyperrectangle $[a, b]$. For $x, c \in [a, b]$*

$$f(x) = f(c) + \sum_{\emptyset \ne u \subseteq 1:d} (-1)^{|u|} \Delta_u(f; x, c). \tag{13}$$

Proof: The right hand side of (13) may be written as

$$\begin{aligned}
\sum_{u \subseteq 1:d} (-1)^{|u|} \Delta_u(f; x, c) &= \sum_{u \subseteq 1:d} (-1)^{|u|} \sum_{v \subseteq u} (-1)^{|v|} f(x^v{:}c^{-v}) \\
&= \sum_{v \subseteq 1:d} (-1)^{|v|} f(x^v{:}c^{-v}) \sum_{u \supseteq v} (-1)^{|u|} \\
&= \sum_{v \subseteq 1:d} (-1)^{|v|} f(x^v{:}c^{-v}) 1_{v=1:d} (-1)^{|v|} \\
&= (-1)^{2d} f(x). \quad \Box
\end{aligned}$$

For $a, b \in \mathbb{R}^d$, when f has a (Lebesgue) integral over $\mathrm{rect}[a, b]$ then $\int_{[a,b]} f(x)dx = \pm \int_{\mathrm{rect}[a,b]} f(x)dx$. The sign is negative if and only if there are an odd number of indices $j \in 1:d$ with $a^j > b^j$.

Proposition 7 *Suppose that f is in $L^1[a, b]$ and that $y, y_+, c \in [a, b]$. Then*

$$\sum_{u \subseteq 1:d} (-1)^{|u|} \int_{[y^{-u}:y_+^u, c]} f(x)dx = \int_{[y, y_+]} f(x)dx. \tag{14}$$

Proof: We proceed by induction on d. For $d=1$ the left hand side of (14) is $\int_{[y,c]} f(x)dx - \int_{[y_+,c]} f(x)dx$ which equals $\int_{[y,y_+]} f(x)dx$. Now suppose that the result holds for dimensions 1 through $d-1$. Then for dimension d the left hand side of (14) is

$$\begin{aligned}
&\sum_{v\subseteq -\{d\}} (-1)^{|v|}\Bigl(\int_{[y^{-v}:y_+^v,c]} f(x)dx - \int_{[y^{-v-\{d\}}:y_+^{v\cup\{d\}},c]} f(x)dx\Bigr)\\
&= \int_{[y^d,c^d]} \sum_{v\subseteq -\{d\}} (-1)^{|v|} \int_{[y^v:y_+^{-v},c]} f(x)dx^{-\{d\}}dx^{\{d\}}\\
&\quad - \int_{[y_+^d,c^d]} \sum_{v\subseteq -\{d\}} (-1)^{|v|} \int_{[y^{v\cup\{d\}}:y_+^{-v-\{d\}},c]} f(x)dx^{-\{d\}}dx^{\{d\}}\\
&= \int_{[y^{\{d\}},c^{\{d\}}]} \int_{[y^{-\{d\}},y_+^{-\{d\}}]} f(x)dx - \int_{[y_+^{\{d\}},c^{\{d\}}]} \int_{[y^{-\{d\}},y_+^{-\{d\}}]} f(x)dx\\
&= \int_{[y,y_+]} f(x)dx. \quad \square
\end{aligned}$$

The result of Proposition 7 is familiar in QMC. There we suppose that $N([a,b])$ denotes the number of points from a list $x_1,\dots,x_n$ that are in $[a,b]$. Then for $a\le x\le y$, $\sum_{u\subseteq 1:d}(-1)^{|u|}N([a,x^u:y^{-u}]) = N([x,y])$.

7 Functions not depending on all variables

The next proposition states a well known deficiency for quasi-Monte Carlo applications, of Vitali's definition of variation:

Proposition 8 *Suppose that $f(x)$ is defined on the hyperrectangle $[a,b]$ and that $f(x)$ does not depend on x^u for non-empty $u\subseteq 1{:}d$. Then $V(f)=0$.*

Proof: Let $j\in u$. Then for $a\le \widetilde{a}\le \widetilde{b}\le b$,

$$\Delta(f;\widetilde{a},\widetilde{b}) = \sum_{v\subseteq -\{j\}} (-1)^{|v|}\bigl(f(\widetilde{a}^v:\widetilde{b}^{-v}) - f(\widetilde{a}^{v\cup\{j\}}:\widetilde{b}^{-v-\{j\}})\bigr) = 0,$$

because f does not depend on whether x^j equals $\widetilde{a}^j$ or $\widetilde{b}^j$. Therefore $V_{\mathcal{Y}}(f)=0$ for all ladders $\mathcal{Y}$ on $[a,b]$, and so $V(f)=0$. $\square$

For the next examples, suppose that $[a,b]$ is a hyperrectangle of positive volume in dimension $d\ge 2$. Let $f_1(x)=0$ for $x^1=a^1$ and $f_1(x)=\sin(1/(x^1-a^1))$ otherwise. Then $V(f_1)=0$ even though f_1 has infinite variation in the one dimensional sense along the line $a^1\le x^1\le b^1$ for any fixed $x^{2:d}\in[a^{2:d},b^{2:d}]$. Similarly $V(f_2)=0$, where $f_2(x)=1$ if x^1 is a rational number and $f_2(x)=0$ otherwise. Finally, suppose that $f_3(x)=0$ if $x^1=a^1<b^1$ and $f_3(x)=1/(x^1-a^1)$ otherwise. Then $V(f_3)=0$, even though f_3 is unbounded. Example f_3 is given in Fréchet (1910).

8 Invariants and closure

Let $f(x)$ be defined on the hyperrectangle $[a,b]$. Let $\widetilde{f}(x)$ be defined on the hyperrectangle $[\widetilde{a},\widetilde{b}]$ by $\widetilde{f}(x) = f(\widetilde{x})$ where $\widetilde{x}^j = \phi_j(x^j)$ with ϕ_j is a strictly monotone (increasing or decreasing) invertible function from $[\widetilde{a}^j,\widetilde{b}^j]$ onto $[a^j,b^j]$.

Proposition 9 *In the notation above* $V_{[\widetilde{a},\widetilde{b}]}(\widetilde{f}) = V_{[a,b]}(f)$.

Proof: Suppose that $\mathcal{Y}$ is a ladder on $[\widetilde{a},\widetilde{b}]$. For $j = 1,\dots,d$, if ϕ_j is increasing, let $\widetilde{\mathcal{Y}}^j = \{\phi_j(y) \mid y \in \mathcal{Y}^j\}$ and otherwise let $\widetilde{\mathcal{Y}}^j = \{\phi_j(a^j)\} \cup \{\phi_j(y) \mid y \in \mathcal{Y}^j - \{a^j\}\}$. Then $V_{\widetilde{\mathcal{Y}}}(\widetilde{f}) = V_{\mathcal{Y}}(f)$, and so $V_{[a,b]}(f) \le V_{[\widetilde{a},\widetilde{b}]}(\widetilde{f})$. A similar argument using the inverses of ϕ_j yields $V_{[\widetilde{a},\widetilde{b}]}(\widetilde{f}) \le V_{[a,b]}(f)$. $\square$

Proposition 10 *In the notation above, if every* ϕ_j *is increasing, then* $V_{\mathrm{HK}}(\widetilde{f};\widetilde{a},\widetilde{b}) = V_{\mathrm{HK}}(f;a,b)$.

Proof: Because all of the ϕ_j are increasing, the function $\widetilde{f}(\widetilde{x}^{-u};\widetilde{b}^u)$ corresponds to $f(x^{-u};b^u)$. Then Proposition 9 applies to each term in (5). $\square$

Proposition 11 *Let* f *and* g *be functions on the hyperrectangle* $[a,b]$. *If* $f,g \in BVHK$, *then* $f+g$, $f-g$, *and* fg *are in BVHK. If* $f \in BVHK$ *with* $|f| > C > 0$ *then* $1/f \in BVHK$. *If* $f,g \in BV$, *then* $f+g$ *and* $f-g$ *are in BV, but* fg *is not necessarily in BV. If for* $u \subset 1:d$ *with* $0 < |u| < d$ *both* $f \in BV[a^u,b^u]$ *and* $g \in BV[a^{-u},b^{-u}]$ *hold, then* $fg \in BV[a,b]$. *If also* $\alpha,\beta \in \mathbb{R}$, *then* $V_{[a,b]}(\alpha+\beta f) = |\beta| V_{[a,b]}(f)$ *and* $V_{\mathrm{HK}}(\alpha+\beta f) = |\beta| V_{\mathrm{HK}}(f)$.

Proof: The closure rules for BVHK are in Hardy (1905). Those for BV are in Fréchet (1910). Let $y \in \mathcal{Y}$ for a ladder $\mathcal{Y}$ on $[a,b]$. Then $|\Delta(\alpha+\beta f;a,b)| = |\beta||\Delta(f;a,b)|$ from which the variation results for $\alpha+\beta f$ follow easily.

The following example proves the nonclosure of BV under multiplication. Suppose that the dimension of $[a,b]$ is $d = 2$ and $\mathrm{Vol}([a,b]) > 0$. Let $f(x) = 1/(x^1 - a^1)$ for $x^1 \in (a^1,b^1]$ and $f(x) = 0$ when $x^1 = a^1$. Also let $g(x) = 1 + x^2 - a^2$. Then $V(f) = V(g) = 0$ by Proposition 8. For $\epsilon > 0$ with $\epsilon \le b^1 - a^1$, let $\widetilde{b}^j(\epsilon)$ equal b^j for $j > 1$ and take $\widetilde{b}^1(\epsilon) = a^1 + \epsilon$. Then

$$\Delta(f;a,\widetilde{b}(\epsilon)) = f(\epsilon,b^2) - f(\epsilon,b^1) - f(a^1,b^2) + f(a^1,a^2) = \frac{b^2-a^2}{\epsilon},$$

and so $V(fg) \ge |b^2 - a^2|/\epsilon$. $\square$

Proposition 12 *The function* f *is in BVHK on* $[a,b]$ *if and only if it can be written* $f = f_1 - f_2$ *where* $\Delta_u(f_i;x,y) \ge 0$ *holds for* $i = 1,2$ *whenever* $x \le y$ *and* $u \subseteq 1{:}d$.

Proof: The "only if" part is due to Hardy (1905) and the "if" part is noted in Adams and Clarkson (1934). $\square$

9 Mixed partial derivatives

Vitali's variation is closely connected with the partial derivative of f, taken once with respect to each variable. We write $\partial^{1:d} f(x)$ for $\partial^d f(x)/\prod_{j=1}^d \partial x^j$. More generally, for $u \subseteq 1{:}d$, the mixed partial derivative of f taken once with respect to every x^j for $j \in u$ is denoted $\partial^u f$ and, by convention $\partial^{\emptyset} f(x) = f(x)$. If $\partial^{1:d} f(x)$ exists for all $x \in [a,b]$, then

$$\int_{[a,b]} \partial^{1:d} f(x) dx = \Delta(f; a, b). \tag{15}$$

Equation (15) is immediate for $d = 1$ and follows by induction for $d \geq 1$. Fréchet (1910) used (15) to get the upper bound (17) below, for $V(f)$ from $\partial^{1:d} f$.

Proposition 13 *Suppose that f is a function for which $\partial^{1:d} f$ is defined on the hyperrectangle $[a,b]$. Then*

$$V(f; a, b) \leq \int_{[a,b]} |\partial^{1:d} f(x)| dx, \tag{16}$$

$$V(f; a, b) \leq \mathit{Vol}(\,[a,b]\,) \sup_{x \in [a,b]} \left|\partial^{1:d} f(x)\right|, \quad \textit{and,} \tag{17}$$

$$V(f; a, b) \leq \mathit{Vol}(\,[a,b]\,)^{1/2} \left(\int_{x \in [a,b]} \left(\partial^{1:d} f(x)\right)^2 dx \right)^{1/2}. \tag{18}$$

Proof: For any ladder $\mathcal{Y}$, we find that

$$V_{\mathcal{Y}}(f; a, b) \leq \sum_{y \in \mathcal{Y}} \int_{[y, y_+]} \left|\partial^{1:d} f(x)\right| dx = \int_{[a,b]} \left|\partial^{1:d} f(x)\right| dx,$$

so taking suprema over $\mathcal{Y}$ establishes (16). Applying standard inequalities among L^p norms yields (17) and (18). □

Under mild conditions on $\partial^{1:d}$, equality holds in (16). Continuity of $\partial^{1:d}$ is sufficient, though clearly not necessary. Fréchet (1910) states the following result:

Proposition 14 *If $\partial^{1:d} f(x)$ is continuous on the hyperrectangle $[a,b]$ then*

$$V(f) = \int_{[a,b]} \left|\, \partial^{1:d} f(x)\right| dx. \tag{19}$$

Proof: Let $\epsilon > 0$. For an integer $m \geq 1$, define the ladder $\mathcal{Y}(m) = \prod_{j=1}^d \mathcal{Y}^j$ where $\mathcal{Y}^j = \{a + \ell(b-a)/m \mid \ell = 0, \dots, m-1\}$ for $j = 1, \dots, d$. Because $\partial^{1:d} f(x)$ is continuous on the compact set $[a,b]$, it is uniformly continuous there. Thus there is an integer $m \geq 1$ such that

$$\max_{y \in \mathcal{Y}(m)} \left(\max_{x \in [y, y_+]} \partial^{1:d} f(x) - \min_{x \in [y, y_+]} \partial^{1:d} f(x) \right) \leq \epsilon.$$

For each $y \in \mathcal{Y}$,

$$\Bigl|\int_{[y,y_+]} \partial^{1:d} f(x)dx\Bigr| \geq \int_{[y,y_+]} \Bigl(\bigl|\,\partial^{1:d} f(x)\bigr| - \epsilon\Bigr)dx \tag{20}$$

holds. Equation (20) is trivial if $\partial^{1:d} f(x)$ has constant sign on $[y, y_+]$, and otherwise, the left and right sides of (20) are positive and negative respectively. Finally,

$$\begin{aligned}
V_{[a,b]}(f) &\geq V_{\mathcal{Y}(m)}(f) \\
&= \sum_{y\in\mathcal{Y}(m)} \Bigl|\int_{[y,y_+]} \partial^{1:d} f(x)dx\Bigr| \\
&\geq \sum_{y\in\mathcal{Y}(m)} \int_{[y,y_+]} \Bigl(\bigl|\,\partial^{1:d} f(x)\bigr| - \epsilon\Bigr)dx \\
&= \int_{[a,b]} |\partial^{1:d} f(x)|dx - \epsilon \mathrm{Vol}([a,b]). \quad \square
\end{aligned}$$

Proposition 15 *Let f be defined on the hyperrectangle $[a,b]$. Suppose that for some set $u \subseteq 1\!:\!d$ that $\partial^u f$ exists, and satisfies the Lipschitz-like condition*

$$|\Delta_{-u}(\partial^u f; x, y)| \leq A\,Vol(rect[x^{-u}, y^{-u}]), \tag{21}$$

for all $a \leq x \leq y \leq b$. Then $V(f) \leq A\,Vol([a,b])$.

Proof: Let $a \leq x \leq y \leq b$. Then

$$\begin{aligned}
\Delta(f; x, y) &= \sum_{v\subseteq -u} \sum_{w\subseteq u} (-1)^{|v|+|w|} f(x^{v\cup w} : y^{-v-w}) \\
&= \sum_{v\subseteq -u} (-1)^{|v|} \int_{[x^u,y^u]} \partial^u f(z^u : x^{(-u)\cap(v\cup w)} : y^{-u-v-w})dz^u \\
&= \int_{[x^u,y^u]} \Delta_{-u}(\partial^u f; z^u\!:\!x^{-u}, z^u\!:\!y^{-u})dz^u,
\end{aligned}$$

so that $|\Delta(f; x, y)| \leq A|\int_{[x^u,y^u]} \mathrm{Vol}(\mathrm{rect}[x^{-u}, y^{-u}])dx^u| \leq A\mathrm{Vol}(\mathrm{rect}[x, y])$. Therefore for any ladder $\mathcal{Y}$ on $[a,b]$ we find $V_{\mathcal{Y}}(f) \leq A\mathrm{Vol}([a,b])$ and so $V(f) \leq A\mathrm{Vol}([a,b])$. $\square$

When $u = 1\!:\!d$, then the sufficient condition (21) in Proposition 15 reduces to $|\partial^{1:d} f| \leq A$. When $u = \emptyset$ then (21) reduces to $|\Delta(f; x, y)| \leq A\mathrm{Vol}(\mathrm{rect}[x, y])$, a condition in Fréchet (1910). When $u = \{j\}$ then (21) reduces to a Lipschitz condition for $\partial^{-\{j\}} f$, with respect to x^j, holding uniformly in $x^{-\{j\}}$. When condition (21) holds for u then it also holds for $\widetilde{u} \subseteq u$, so that Fréchet's $u = \emptyset$ condition is the most widely applicable version of (21), and the condition on the full partial derivative $\partial^{1:d}$ is the least widely applicable.

We illustrate the use of the propositions above with an example function having a "cusp" along a hyperplane. For integers $d \ge 1$ and $r \ge 0$ let $f_{d,r}$ be a function on $[0,1]^d$ defined by

$$f_{d,r}(x) = \begin{cases} \max(x^1 + \cdots + x^d - 1/2, 0)^r, & r > 0 \\ 1_{x^1+\cdots+x^d>1/2}, & r = 0. \end{cases}$$

For $u \subseteq 1{:}d$ with $|u| < r$,

$$\partial^u f_{d,r}(x) = r(r-1)\cdots(r-|u|+1) f_{d,r-|u|}(x). \tag{22}$$

If $|u| = r$ then (22) holds everywhere except on the set $E = \{x \mid x^1 + \cdots + x^d = 1/2\}$ of d dimensional volume zero. If $|u| > r$ then $\partial^u f_{d,r}(x) = 0$ for $x \notin E$ and is not defined for $x \in E$.

Proposition 16 *$V(f_{d,r}; [0,1]^d)$ is finite for $d \le r$ and infinite for $d \ge r+2$.*

Proof: If $r > d$ then $\partial^{1:d} f$ is bounded. If $r = d$ then $\partial^{2:d} f$ exists and is a Lipschitz continuous function in x^1 uniformly in $x^{2:d}$. Therefore $V(f) < \infty$ by Proposition 15 when $d \le r$.

Now suppose that $d \ge r+2$. Let $\mathcal{Y}^j = \{0, 1/(2m), \ldots, (m-1)/(2m)\}$ be a ladder on $[0, 1/2]$, and put $\mathcal{Y} = \prod_{j=1}^d \mathcal{Y}^j$. Suppose that $y \in \mathcal{Y}$ with $\sum_{j=1}^d y^j = (m-d+1)/(2m)$. Then $\Delta(f_{d,r}; y, y_+) = (2m)^{-r}$. The number of such y is equal to the number of ways to choose d nonnegative integers whose sum is $m-d+1$. Therefore

$$V(f) \ge \frac{1}{(2m)^r}\binom{m}{d-1} \ge \frac{(m-d)^{d-1}}{(d-1)!(2m)^r} \to \infty$$

as $m \to \infty$. Therefore $V(f; [0,1/2]^d) = \infty$ and so $V(f; [0,1]^d) = \infty$ too. $\square$

Taking $r = 0$ in Proposition 16 shows that $V(1_A) = \infty$ for $A = \{x \in [0,1]^d \mid x^1 + \cdots + x^d > 1/2\}$ when $d \ge 2$. Similarly if A is a hyperrectangular region that is not parallel to any of the coordinate axes of $[a,b]$ then 1_A has infinite variation when $d \ge 2$. As d increases, it takes ever greater smoothness along the set E for $f_{d,r}$ to have finite variation.

Proposition 16 has a gap for the case $d = r+1$. Then $\partial^{1:d} f_{d,d-1}$ vanishes for $x \notin E$, but does not exist for $x \in E$. All of the variation of $f_{d,d-1}$ comes from the set E. It is not hard to show that $V(f_{2,1}) = 1$ and that in general $V(f_{d,d-1}) < \infty$. Here is a sketch of the reasoning: By Proposition 5 with $\epsilon = 1$, we need only consider ladders $\mathcal{Y}$ with every $y_+^j - y^j$ in an interval $[\eta, 2\eta)$ where $\eta > 0$. Such a ladder yields fewer than $A\eta^{-d+1}$ hyperrectangles $[y, y_+]$ that intersect the set E, for some $A < \infty$. For each such hyperrectangle, we find that $|\Delta(f_{d,d-1}; y, y_+)| \le B\eta^{d-1}$ for some $B < \infty$. Then $V_{\mathcal{Y}}(f_{d,d-1}) \le AB$.

Proposition 17 considered functions symmetric in their arguments. The infinite variation result is more general:

Proposition 17 *For integer $r \ge 0$, real values $\theta_0, \dots, \theta_d$, and $x \in [a,b]$ let*

$$f(x) = f_{r,\theta}(x) = \begin{cases} \max(\theta_1 x^1 + \cdots + \theta_d x^d - \theta_0, 0)^r, & r > 0 \\ 1_{\theta_1 x^1 + \cdots + \theta_d x^d > \theta}, & r = 0. \end{cases}$$

Let $E = \{x \in [a,b] \mid \theta_1 x^1 + \cdots + \theta_d x^d = \theta_0\}$. If $E \cap [a,b]$ has positive $d-1$ dimensional volume, $d \ge r+2$, and none of $\theta_1, \dots, \theta_d$ is zero, then $V(f) = \infty$.

Proof: The proof follows from two applications of Proposition 9 which reduce the problem to the one handled by Proposition 16.

For $j \in 1:d$, let $\phi_j(x^j) = x^j/\theta_j$. Take $f_1(x) = f_{r,\theta}(\phi(x))$ with ϕ applied componentwise. The domain of f_1 is $s_1 = \mathrm{rect}[\phi^{-1}(a), \phi^{-1}(b)]$ with ϕ^{-1} applied componentwise. By construction $f_1(x) = \max(\sum_{j=1}^d x^j - \theta_0, 0)^r$.

Because E intersects the original $[a,b]$ nontrivially, there is a point $\hat{a} \in s_1$ and a constant $\epsilon > 0$ such that $\sum_{j=1}^d \hat{a}_j = \theta_0 - \epsilon/2$ and $\hat{a} + \epsilon$ (componentwise) is in s_1. For $j \in 1:d$, let $\psi_j(x^j) = \hat{a}^j + \epsilon x^j$. Take $f_2(x) = f_1(\psi(x))$ with ψ applied componentwise. The domain of f_2 is $s_2 = [0,1]^d$. By construction

$$f_2(x) = \max\Bigl(\sum_{j=1}^d \hat{a}^j + \epsilon x^j - \theta_0, 0\Bigr)^r = \epsilon^r \max\Bigl(\sum_{j=1}^d x^j - 1/2, 0\Bigr)^r.$$

Finally $V_{[a,b]}(f_{r,\theta}) = V_{s_1}(f_1) \ge V_{s_2}(f_2) = \epsilon^r V(f_{d,r}) = \infty$. $\square$

10 Functions vanishing except on one face

The next two propositions consider functions that are zero on all of the hyperrectangle $[a,b]$, except for a boundary face. There are two cases, one for a face that is a single corner of $[a,b]$ and one for a face of positive dimension less than d.

Proposition 18 *Let $a, b \in \mathbb{R}^d$ with $a \le b$ and let $u \subseteq 1:d$. Suppose that $f(x) = 0$ unless $x^u = a^u$ and $x^{-u} = b^{-u}$. Then*

$$V_{[a,b]}(f) = \begin{cases} |f(a^u : b^{-u})|, & \mathit{Vol}([a,b]) > 0 \\ 0, & \mathit{else}. \end{cases} \tag{23}$$

Proof: If $\mathrm{Vol}([a,b]) = 0$ then $V(f) = 0$ for any real valued f. Assume that $\mathrm{Vol}([a,b]) \ne 0$. Then $V_{\mathcal{Y}}(f; a, b) = |f(a^u : b^{-u})|$ for any ladder $\mathcal{Y}$ on $[a,b]$. $\square$

Proposition 19 *Let $a, b \in \mathbb{R}^d$ with $a \le b$. Let $u, v \subseteq 1:d$ with $u \cap v = \emptyset$ and $|u \cup v| < d$, and set $w = -u - v \ne \emptyset$. Suppose that $f(x)$ is defined on $[a,b]$ with $f(x) = 0$ unless $x^u = a^u$ and $x^v = b^v$. Then*

$$V_{[a,b]}(f) = \begin{cases} V_{[a^w, b^w]}(f(x^w; a^u : b^v)), & \mathit{Vol}([a,b]) > 0 \\ 0, & \mathit{else}. \end{cases} \tag{24}$$

Proof: Suppose that $\mathrm{Vol}([a,b]) > 0$. For any ladder $\mathcal{Y}$ on $[a,b]$ and any $y \in \mathcal{Y}$ we find that $\Delta(f;y,y_+) = 0$ if $y^u \neq a^u$ or $y_+^v \neq b^v$. Then $V_{\mathcal{Y}}(f) = V_{\mathcal{Y}^w}(f(x^w;a^u{:}b^v))$. $\square$

Proposition 18 is the $w = \emptyset$ version of Proposition 19 if we adopt the convention that the variation of f on $[a^\emptyset, b^\emptyset]$ is $|f()|$.

Proposition 20 *Let $a, b, \widetilde{a}, \widetilde{b} \in \mathbb{R}^d$ with $a \le \widetilde{a} \le \widetilde{b} \le b$. Let $f(x)$ be defined on $[a,b]$ with $f(x) = 1$ for $\widetilde{a} \le x \le \widetilde{b}$ and $f(x) = 0$ otherwise. Then*

$$V_{[a,b]}(f) = \prod_{j=1}^{d} \left(1_{a^j < \widetilde{a}^j} + 1_{\widetilde{b}^j < b^j}\right). \tag{25}$$

Proof: Begin by splitting $[a,b]$ into 3^d hyperrectangles of the form $[a^u, \widetilde{a}^u] \times [\widetilde{a}^v, \widetilde{b}^v] \times [\widetilde{a}^w, \widetilde{b}^w]$, where u, v, w are disjoint subsets of $1{:}d$ with $u \cup v \cup w = 1{:}d$. By Lemma 1, $V_{[a,b]}(f)$ is the sum of $V(f)$ taken over these hyperrectangles. Notice that if $v \neq \emptyset$ then f does not depend on x^j over the corresponding hyperrectangle, so $V(f)$ vanishes there. If instead $v = \emptyset$, then f vanishes except at one corner of the hyperrectangle, and so Proposition 18 applies. Therefore

$$\begin{aligned} V_{[a,b]}(f) &= \sum_{u \subseteq 1:d} V_{[a^u, \widetilde{a}^u] \times [\widetilde{b}^u, b^u]}(f) \\ &= \sum_{u \subseteq 1:d} \Bigl(\prod_{j \in u} 1_{a^j < \widetilde{a}^j}\Bigr)\Bigl(\prod_{j \notin u} 1_{\widetilde{b}^j < b^j}\Bigr) \\ &= \prod_{j=1}^{d} \Bigl(1_{a^j < \widetilde{a}^j} + 1_{\widetilde{b}^j < b^j}\Bigr). \quad \square \end{aligned}$$

Proposition 20 includes some interpretable special cases. If $a < \widetilde{a} \le \widetilde{b} < b$, then $V_{[a,b]}(f) = 2^d$, so the variation of the indicator function of a hyperrectangle in general position is 2^d. For the indicator function of a single point in general position, we take $\widetilde{a} = \widetilde{b}$, and find again the variation is 2^d. When the boundary of $[\widetilde{a}, \widetilde{b}]$ intersects that of the containing hyperrectangle $[a,b]$, the variation is smaller. If for any j, there is equality at both boundaries, that is $a^j = \widetilde{a}^j$ and $b^j = \widetilde{b}^j$, then $V(f)$ vanishes, reflecting the fact that f does not depend on x^j. If any $a^j = b^j$ then of necessity there is equality at both boundaries so $V(f)$ vanishes, as it must because $a^j = b^j$ implies $\mathrm{Vol}([a,b]) = 0$.

The next result relates Hardy-Krause variation of f to Vitali variation of f, after extending the domain of f to the "upper-right", and filling in constant values.

Proposition 21 *Suppose that $a, b, \widetilde{b} \in \mathbb{R}^d$ with $a < b < \widetilde{b}$. Let $f(x)$ be a real valued function defined on $[a,b]$. Define $\widetilde{f}(x)$ and $\widehat{f}(x)$ on $[a, \widetilde{b}]$ as follows: For $x \in [a,b]$ let $\widetilde{f}(x) = \widehat{f}(x) = f(x)$. For $x \in [a,\widetilde{b}] - [a,b]$ let $\widetilde{f}(x) = 0$ and $\widehat{f}(x) = f(b)$. Then*

$$V_{[a,\widetilde{b}]}(\widetilde{f}) = V_{\mathrm{HK}}(f) + |f(b)|, \quad \textit{and,} \quad V_{[a,\widetilde{b}]}(\widehat{f}) = V_{\mathrm{HK}}(f).$$

Proof: Let $g(x) = f(x)$ on $[a, b]$ and let $g(x) = c$ on $[a, \widetilde{b}] - [a, b]$. Split $[a, \widetilde{b}]$ into 2^d hyperrectangles of the form $[a^u, b^u] \times [b^{-u}, \widetilde{b}^{-u}]$ for $u \subseteq 1\!:\!d$. For $u = 1 : d$, we find $V_{[a,b]}(g) = V_{[a,b]}(f)$. For $u \neq 1 : d$ note that $g(x) - c$ is a function that vanishes except on one face of $[a^u, b^u] \times [b^{-u}, \widetilde{b}^{-u}]$. If $u = \emptyset$, we find $V_{[b,\widetilde{b}]}(g - c) = |f(b) - c|$ by Proposition 18. For $0 < |u| < d$, we apply Proposition 19 to $g-c$ getting the variation of $f-c$ on a face. By Proposition 11 the variation of $f - c$ equals that of f. Then

$$V_{[a,\widetilde{b}]}(g) = \sum_{u \subseteq 1:d} V_{[a^u,b^u]\times[b^{-u},\widetilde{b}^{-u}]}(g) = \sum_{\emptyset \neq u \subseteq 1:d} V_{[a^u,b^u]}(f(x^u, b^{-u})) + |f(b) - c|.$$

The first result follows for $c = 0$, the second for $c = f(b)$. $\square$

11 Variation and ANOVA

If $f \in L^2[a, b]$ where $\mathrm{Vol}([a, b]) > 0$, then there is an analysis of variance (ANOVA) decomposition of f. Liu and Owen (2003) outline properties and references for ANOVA. The ANOVA takes the form

$$f(x) = \sum_{u \subseteq 1:d} f_u(x)$$

where $f_u(x)$ only depends on x^u. Among such decompositions it is the unique one that satisfies $\int_{a^j}^{b^j} f_u(x)dx^j = 0$ whenever $j \in u$.

By Proposition 8, $V(f_u) = 0$ for $|u| < d$ and so $V(f) = V(f_{1:d})$. Let $E(g) = \mathrm{Vol}([a, b])^{-1} \int_{[a,b]} g(x)dx$ denote the expected value of $g(x)$ for random x uniformly distributed in $[a, b]$. Let $\mathrm{Var}(g) = E((g(x) - E(g(x)))^2)$ denote the variance of $g(x)$. Write $\sigma^2 = \mathrm{Var}(f)$ and $\sigma_u^2 = \mathrm{Var}(f_u)$. The ANOVA decomposition is so named because $\sigma^2 = \sum_{u \subseteq 1:d} \sigma_u^2$.

Proposition 22 *If* $\sigma_{1:d}^2 > 0$ *then* $V(f) > 0$. *The converse does not hold.*

Proof: Liu and Owen (2003) show that $E(\Delta(f; x, \widetilde{x})^2) = \sigma_{1:d}^2$ for independent random x and $\widetilde{x}$, both uniformly distributed on $[a, b]$. Then if $\sigma_{1:d}^2 > 0$ there exist $x, \widetilde{x} \in [a, b]$ with $|\Delta(f; x, \widetilde{x})| \geq \sigma_{1:d}$ and so $V(f) > 0$. As for the converse, let $f(x) = 1$ if $x = b$ and let $f(x) = 0$ otherwise. Then $0 \leq \sigma_{1:d}^2 \leq \sigma^2 = 0$ but $V(f) = 1$ by Proposition 20. $\square$

12 Indicator functions

Let $[a, b]$ be a d dimensional hyperrectangle and let $A \subset [a, b]$. The indicator function of A, also called the characteristic function of A, is given by $1_A(x) = 1$ for $x \in A$ and $1_A(x) = 0$ otherwise. It is clear that $\Delta(1_A; a, b)$ must be an integer and so $V_{\mathcal{Y}}(1_A)$ must also be an integer. Therefore either $V(1_A) = \infty$ or $V(1_A)$

is a nonnegative integer. Also, we easily find that $V_{[a,b]}(1_A) = V_{[a,b]}(1_{[a,b]-A})$ by Proposition 11 because $1_{[a,b]-A} = 1 - 1_A$.

Proposition 20 gives the variation in Vitali's sense for indicator functions of hyperrectangles. Propositions 16 and 17 show how indicator functions can have infinite variation when $d \geq 2$ and A has a planar boundary. The difference between the cases lies in whether the boundary of A is parallel to any of the coordinate axes of $[a, b]$.

For a more general set A we can for integer $m \geq 1$, split $[a, b]$ into m^d congruent hyperrectangles each similar to $[a, b]$. The variation of f is at least as large as the number of those hyperrectangles with nonzero variation. We anticipate that this number grows in proportion to m^{d-1} for typical sets A of interest. Therefore we first consider when an indicator function has non-zero variation. We know that $V(1_A) = 0$ if 1_A does not depend on x^j for some $j \in 1{:}d$. When $d = 2$ there is a converse as follows:

Proposition 23 *Let $[a, b]$ be a rectangle in $\mathbb{R}^2$ with Vol$([a, b]) > 0$. Let $f : [a, b] \to \{0, 1\}$ and suppose that f does depend on x^j for each $j \in \{1, 2\}$. Then $V(f; a, b) \geq 1$.*

Proof: Because f depends on x^2 there is a value $y^1 \in [a^1, b^1]$ such that $f(x^{\{2\}}; y^{\{1\}})$ takes both values 0 and 1. Similarly let y^2 be a point in $[a^2, b^2]$ for which $f(x^{\{1\}}; y^{\{2\}})$ takes both values 0 and 1, and put $y = (y^1, y^2)$. Let $\widetilde{y}^1 \in [a^1, b^1]$ and $\widetilde{y}^2 \in [a^2, b^2]$ satisfy $f(\widetilde{y}^1, y^2) = f(y^1, \widetilde{y}^2) = 1 - f(y)$. Let $[\widetilde{a}, \widetilde{b}] = \mathrm{rect}[y, \widetilde{y}]$. Then

$$|\Delta(f; \widetilde{a}, \widetilde{b})| = |f(\widetilde{y}) - f(\widetilde{y}^1, y^2) - f(y^1, \widetilde{y}^2) + f(y)| = |f(\widetilde{y}) - 2 + 3f(y)| \geq 1$$

for $f(\widetilde{y}), f(y) \in \{0, 1\}$. $\Box$

The natural analogue of Proposition 23 does not hold true for $d \geq 3$. For $d = 3$, consider $[0, 1]^3$. Let $A_1 = A_2 = A_3 = [0, 1/2)$ and define

$$f(x) = \begin{cases} 1, & x^3 \in A_3 \text{ and } x^2 \in A_2, \\ 0, & x^3 \in A_3 \text{ and } x^2 \notin A_2, \\ 1, & x^3 \notin A_3 \text{ and } x^1 \in A_1, \\ 0, & x^3 \notin A_3 \text{ and } x^1 \notin A_1. \end{cases}$$

The function f depends on each of the 3 components of x. This function can be visualized in terms of four blocks of size $1/2 \times 1/2 \times 1$, two opaque blocks for $f(x) = 1$ and two transparent ones for $f(x) = 0$. Taking the x^3 axis to be the vertical direction place an opaque and a transparent block flat on the ground, touching along one of their long sides. On top of these place a second pair of blocks, rotated ninety degrees with respect to the first pair.

Suppose that $[a, b]$ is a hyperrectangular subset of $[0, 1]^3$. If a^3 and b^3 are both in or both not in A_3 then clearly $\Delta(f; a, b) = 0$. Similarly, one can check for $j = 1, 2$ that $\Delta(f; a, b) = 0$ if a^j and b^j are both in or both not in A_j. Finally if $a^j < 1/2 \le b^j$ for $j = 1, 2, 3$, then

$$\begin{aligned}\Delta(f;a,b) &= f(1,1,1) - f(0,1,1) - f(1,0,1) + f(0,0,1)\\ &\quad - f(1,1,0) + f(0,1,0) + f(1,0,0) - f(0,0,0)\\ &= 0 - 1 - 0 + 1 - 0 + 0 + 1 - 1\\ &= 0.\end{aligned}$$

Since $\Delta(f;a,b) = 0$ for any hyperrectangle $[a,b] \subset [0,1]^3$, it follows that $V(f;a,b) = 0$. There is no need for A_j to be the given subintervals. Each of the A_j can be an arbitrary non-empty proper subset of $[0,1]$. Similarly for $d > 3$ one can construct binary functions with variation zero that depend on every one of the d inputs.

Suppose that $A \subset [a,b]$ is a set, open or closed or neither, with a positive d dimensional volume and a smooth boundary. If a portion of that smooth boundary has positive $d-1$ dimensional volume and is not parallel to any of the coordinate axes, then for $d \ge 2$, we expect that $V(1_A) = \infty$. For instance if A is the interior of a sphere of positive radius contained inside $[a,b]$ then $V(1_A) = \infty$. Informally, the argument runs as follows. We can find a small hyperrectangle s inside $[a,b]$ with one face in A, the opposite face not in A, and a nearly linear boundary separating $s \cap A$ from $s \cap (-A)$. Then $V(1_A) \ge V_s(1_A)$ and the latter is infinite. The next proposition fills in details.

Proposition 24 *Let A be a subset of the hyperrectangle $[a,b]$ in dimension $d \ge 2$. Suppose that there exists a subhyperrectangle $[\widetilde{a},\widetilde{b}] \subset [a,b]$ of positive volume, an index $j \in 1:d$, and a function g defined on $[\widetilde{a}^{-\{j\}},\widetilde{b}^{-\{j\}}]$ taking values in $(\widetilde{a}^j,\widetilde{b}^j)$ such that either $\widetilde{x} \in A$ when $\widetilde{x}^j > g(\widetilde{x}^{-\{j\}})$ and $\widetilde{x} \notin A$ when $\widetilde{x}^j < g(\widetilde{x}^{-\{j\}})$ or $\widetilde{x} \in A$ when $\widetilde{x}^j < g(\widetilde{x}^{-\{j\}})$ and $\widetilde{x} \notin A$ when $\widetilde{x}^j > g(\widetilde{x}^{-\{j\}})$. Suppose further that $\partial^{\{k\}} g$ is bounded away from zero for each $k \ne j$. Then $V(1_A) = \infty$.*

Proof: For $m \ge 1$, let S_m be the split of $[\widetilde{a}^{-\{j\}},\widetilde{b}^{-\{j\}}]$ into m^{d-1} congruent hyperrectangles. Let $\widetilde{S}_m = \{s \times [\widetilde{a}^j,\widetilde{b}^j] \mid s \in S_m\}$. Then $\widetilde{S}_m$ is a split of $[\widetilde{a},\widetilde{b}]$ into m^{d-1} long thin hyperrectangles. For each $s \in S_m$ evaluate g at all 2^{d-1} corners and select a value c strictly between the largest and second largest of these values. From a coordinate split of $s \times [\widetilde{a}^j,\widetilde{b}^j]$ along direction j at point c we find that $V_{s\times[\widetilde{a}^j,\widetilde{b}^j]}(1_A) \ge 1$ and so $V(1_A) \ge m^{d-1}$. $\square$

In Proposition 24 the set A was assumed to be of positive d dimensional volume. Thus for example it does not apply to functions like the indicator of a hypershere that nontrivially intersects the $d \ge 2$ dimensional $[a,b]$. For that case we consider a subhyperrectangle $[\widetilde{a},\widetilde{b}]$ for which there is an index j and a function g on $[\widetilde{a}^{-\{j\}},\widetilde{a}^{-\{j\}}]$ with $\widetilde{x} \in A$ if and only if $g(\widetilde{x}^{-\{j\}}) = \widetilde{x}^j$. Once again we can find coordinate splits to show that the variation of 1_A is positive within each of m^{d-1} long thin hyperrectangles constructed as in the proof of Proposition 24.

13 Call and put options

Much work in quasi-Monte Carlo integration has been motivated by some integrands from computational finance. For full details of Monte Carlo applications to computational finance, see Glasserman (2004). Here we present some such integrands, and explain why they are not typically of bounded variation.

For $z \in \mathbb{R}$, let $\varphi(z) = \exp(-z^2/2)/\sqrt{2\pi}$ be the standard normal probability density function, $\Phi(z) = \int_{-\infty}^{z} \varphi(y)dy$ be the corresponding cumulative distribution function, and let Φ^{-1} be the quantile function, mapping $(0,1)$ to $\mathbb{R}$. We also take $\Phi^{-1}(0) = -\infty$ and $\Phi^{-1}(1) = \infty$.

Many call options have a payoff function that can be expressed in the form:

$$C(x) = \max\biggl(0, \sum_{r=1}^{R} \alpha_r \exp\biggl(\beta_{r0} + \sum_{j=1}^{d} \beta_{rj} G^{-1}(x^j)\biggr) - K\biggr), \tag{26}$$

for scalars $\alpha_r > 0$ and β_{rj} and a strike price $K > 0$. It is usual to have $G = \Phi$ but sometimes G is an alternative distribution having fatter tails than does the normal. We will assume that $G^{-1}(0) = -\infty$ and $G^{-1}(1) = \infty$. For simplicity some discount factors have been absorbed into the α_r. The value of the option is $\int_{[0,1]^d} C(x)dx$. In cases of interest there are r and $j \geq 1$ for which $\beta_{rj} \neq 0$ holds. Then C is unbounded on $(0,1)^d$ and hence cannot be BVHK.

For $f_r(x) = \alpha_r \exp(\beta_{r0} + \sum_{j=1}^{d} \beta_{rj} G^{-1}(x^j))$, let

$$P(x) = \max\biggl(0, K - \sum_{r=1}^{R} f_r(x)\biggr). \tag{27}$$

This $P(x)$ is the payoff of a put option whose value $\int_{[0,1]^d} P(x)dx$ is of independent interest. Notice that $C(x) - P(x) = \sum_{r=1}^{R} f_r(x) - K$. When $G = \Phi$, there is a closed form expression for $\int_{[0,1]^d} f_r(x)dx$ and then an estimate of $P(x)$ can then be easily translated into one for $C(x)$. The function $P(x)$ is bounded because all the $\alpha_r > 0$. When $P(x)$ is BVHK, then quasi-Monte Carlo integration yields an estimate of $\int P(x)dx$ and hence also of $\int C(x)dx$ with error rate $O(n^{-1}\log(n)^d)$.

But $P(x)$ is ordinarily not BHVK. It is continuous but has a cusp along the set $E = \{x \mid \sum_{r=1}^{R} f_r(x) = 0\}$. As in the proof of 24 we employ m^{d-1} long thin hyperrectangles that cross E in their long direction. Let j be an index for which $\beta_{rj} \neq 0$ for some $r > 0$. Suppose first that $\beta_{rj} > 0$ so that $f_r(x) \to \infty$ as $x^j \to 1$ for any $x^{-\{j\}}$. The projections of these hyperrectangles in the $-\{j\}$ directions, split a hyperrectangle $[a^{-\{j\}}, b^{-\{j\}}] \subset [0,1]^{d-1}$ such that $P(x) > 0$ at every point of $[a^{-\{j\}}, b^{-\{j\}}] \times c^{\{j\}}$ for some $c^{\{j\}} \in (0,1)$. The hyperrectangles extend from $c^{\{j\}}$ to 1 in the x^j direction. When $d \geq 3$, the variation in each long thin hyperrectangle is larger than a fixed multiple of m^{-1} so that the variation of P is infinite.

If instead $\beta_{rj} < 0$, then take long thin hyperrectangles whose $-\{j\}$ projections split a hyperrectangle $[a^{-\{j\}}, b^{-\{j\}}] \subset [0,1]^{d-1}$ such that $P(x) = 0$ at every point of $[a^{-\{j\}}, b^{-\{j\}}] \times c^{\{j\}}$ for some $c^{\{j\}} \in (0,1)$. Then take the long direction for the hyperrectangles to be from 0 to $c^{\{j\}}$.

14 Low variation extensions

Given a function f defined on a subset K of $[a,b]$ we consider ways of extending it to $\widetilde{f}$ defined on all of $[a,b]$ while keeping some control on the size of $V_{[a,b]}(\widetilde{f})$. One application is in proving results like Theorem 2 of Sobol' (1973). Sobol's proof of that theorem was never published. Professor Sobol' kindly described for me the key ideas underlying the proof. See especially equations (29) and (30) below.

The set K is assumed to have some regularity. First we assume that K is a nonempty closed set. Then we designate some point $c \in K$ as an "anchor" for the extensions. This anchor is commonly taken to be a or b or $(a+b)/2$. Then we suppose that

$$x \in K \implies \mathrm{rect}[x,c] \subseteq K. \tag{28}$$

In case $c = b$ then K has the Pareto property. Given $x \in K$ and $y \notin K$ there is at least one j with $y^j < x^j$. The next result appears in Sobol' (1961).

Proposition 25 *Let f be a function on the hyperrectangle $[a,b]$. Suppose that $\partial^{1:d} f$ exists. Then for $x, c \in [a,b]$*

$$f(x) = f(c) + \sum_{\emptyset \neq u \subseteq 1:d} (-1)^{|u|} \int_{[x^u, c^u]} \partial^u f(y^u; c^{-u}) dy^u. \tag{29}$$

Proof: Similarly to equation (15) we find that $\int_{[x^u,c^u]} \partial^u f(y^u; c^{-u}) dy^u = \Delta_u(f; x, c)$. The rest follows from Proposition 6. $\Box$

The term $f(c)$ in equation (29) corresponds to the case $u = \emptyset$ excluded from the summation there, if we adopt the convention that

$$(-1)^0 \int_{[x^\emptyset, c^\emptyset]} \partial^\emptyset f(c, y^\emptyset) dy^\emptyset = f(c) \int_{[x^\emptyset, c^\emptyset]} dy^\emptyset = f(c).$$

Next we give a representation of f as a sum of functions of varying dimensionalities, using mixed partial derivatives of f taken once with respect to each x^j for j in a set u. When K contains a d dimensional rectangle of nonzero volume, these derivatives are defined as usual. In particular for points x on the boundary of K, only one sided derivatives defined as limits from within K are used. When K is contained inside a zero volume rectangle there are some coordinate directions from which no meaningful limit can be taken. Let $\nu(K) = \{j \in 1:d \mid \sup_{x \in K} x^j > \inf_{x \in K} x^j\}$ be the set of coordinates that truly vary within K. The formulas below will not depend on the value we give to derivatives with respect to coordinates that do not vary. For definiteness, we take

$$\partial_K^u f(x) = \begin{cases} \partial^u f(x), & u \subseteq \nu(K) \\ 0, & \text{else.} \end{cases}$$

Even when $\nu(K) = \emptyset$, which holds when $K = \{c\}$, we still have $\partial_K^{\emptyset} f(c) = f(c)$.

Definition 5 *Let $c \in [a, b]$ and suppose that K is a nonempty closed subset of $[a, b]$ which satisfies* (28)*, and that $\partial_K^u f(x)$ exists for $x \in K$ and $u \subseteq 1:d$. Then the low variation extension of f from K to $[a, b]$ with anchor c is given by*

$$\widetilde{f}(x) = f(c) + \sum_{\emptyset \neq u \subseteq 1:d} (-1)^{|u|} \int_{[x^u, c^u]} 1_{z^u : c^{-u} \in K}\ \partial_K^u f(z^u : c^{-u}) dz^u, \tag{30}$$

for $x \in [a, b]$.

To justify the name "extension" requires that $\widetilde{f}(x) = f(x)$ when $x \in K$. Note that $x \in K$ implies $\mathrm{rect}[x, c] \subseteq K$ so that the expression $1_{z^u:c^{-u}\in K}$ can then be removed from equation (30). Next ∂_K^u only differs from ∂^u in cases where $[x^u, c^u]$ has zero volume, and those terms contribute nothing to the sum. Therefore the subscript K can be removed from the partial derivative symbol. Then $\widetilde{f}(x) = f(x)$ by Proposition 25.

Theorem 1 *For $c \in [a, b]$, let K be a nonempty closed subset of $[a, b]$ which satisfies* (28)*. Let f be a function for which $\partial_K^u f(x)$ exists for $x \in K$, and $u \subseteq 1:d$. Let $\widetilde{f}$ be the low variation extension of f from K to $[a, b]$ with anchor c. Then*

$$V_{[a,b]}(\widetilde{f}) \le \int_K |\partial_K^{1:d} f(x)| dx. \tag{31}$$

If $\partial_K^{1:d} f(x)$ is continuous on K, then

$$V_{[a,b]}(\widetilde{f}) = \int_K |\partial_K^{1:d} f(x)| dx. \tag{32}$$

Proof: If $|u| < d$ then the corresponding term in (30) is a function of x that does not depend on x^{-u}, so it has Vitali variation 0. Therefore the Vitali variation of $\widetilde{f}$ equals that of $\widetilde{f}_{1:d}(x) = (-1)^d \int_{[x,c]\cap K} \partial_K^{1:d} f(z) dz$. Let $\mathcal{Y}$ be a ladder on $[a, b]$. For $y \in \mathcal{Y}$,

$$\begin{aligned}
\Delta(\widetilde{f}_{1:d}; y, y_+) &= \sum_{v \subseteq 1:d} (-1)^{|v|} \widetilde{f}_{1:d}(y^v : y_+^{-v}) \\
&= \sum_{v \subseteq 1:d} (-1)^{|v|} (-1)^d \int_{[y^v, c^v] \times [y_+^{-v}, c^{-v}]} 1_{z \in K}\ \partial_K^{1:d} f(z) dz \\
&= \sum_{v \subseteq 1:d} (-1)^{|-v|} \int_{[y^v, c^v] \times [y_+^{-v}, c^{-v}]} 1_{z \in K}\ \partial_K^{1:d} f(z) dz \\
&= \int_{[y, y_+] \cap K} \partial_K^{1:d} f(z) dz,
\end{aligned}$$

so that $V_{\mathcal{Y}}(\widetilde{f}_{1:d}) \le \int_K |\partial_K^{1:d} f(x)|dx$. Taking the supremum over $\mathcal{Y}$ proves (31). To prove (32), note that K is compact, so $\partial_K^{1:d} f$ is uniformly continuous on K. We may split $[a,b]$ into a regular grid of m^d hyperrectangles, sum $V(f)$ over those hyperrectangles that are contained within K, and let $m \to \infty$. $\Box$

For the Hardy-Krause variation of $\widetilde{f}$ we need to consider which x^{-v} values when glued to b^v give a point in K. Let $K(b^v) = K_{-v}(b^v) = \{x^{-v} \in [a^{-v}, b^{-v}] \mid x^{-v}{:}b^v \in K\}$.

Theorem 2 *Assume the conditions of Theorem 1 and that $c = b$. Then*

$$V_{\mathrm{HK}}(\widetilde{f}) \le \sum_{v \subsetneq 1:d} \int_{K(b^v)} |\partial_K^{-v} f(x^{-v}, b^v)| dx^{-v}. \tag{33}$$

Proof: From the definition, $V_{\mathrm{HK}}(\widetilde{f}) = \sum_{v \subsetneq 1:d} V_{[a^{-v}, b^{-v}]}(\widetilde{f}(x^{-v}; b^v))$. If $-v \subseteq \nu(K)$ then $\widetilde{f}(x^{-v}; b^v)$ is also the low variation extension of $f(x^{-v}; b^v)$ from $K(b^v)$ to $[a^{-v}, b^{-v}]$ with anchor b^{-v}, and so

$$V_{[a^{-v}, b^{-v}]}(\widetilde{f}(x^{-v}; b^v)) \le \int_{K(b^v)} |\partial^{-v} f(x^{-v}; b^v)| dx^{-v}. \tag{34}$$

Now suppose that $j \in -v$ and $j \notin \nu(K)$. Then $\widetilde{f}(x^{-v}; b^v)$ does not depend on x^j, so that $V(\widetilde{f}(x^{-v}; b^v)) = 0$ and again (34) holds. Summing (34) over $v \subsetneq 1{:}d$ establishes (33). $\Box$

Acknowledgments

I thank Ilya M. Sobol' for communicating how he proved Theorem 2 of Sobol' (1973). I thank Linda Yamamoto, Stanford's Mathematics and Computer Science librarian, for help in locating some articles.

References

Adams, C. R. and J. A. Clarkson (1934). Properties of functions $f(x, y)$ of bounded variation. *Transactions of the American Mathematical Society 36*, 711.

Clarkson, J. A. and C. R. Adams (1933). On definitions of bounded variation for functions of two variables. *Transactions of the American Mathematical Society 35*, 824–854.

Fang, K.-T. and Y. Wang (1994). *Number Theoretic Methods in Statistics*. Chapman & Hall.

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

Glasserman, P. (2004). *Monte Carlo methods in financial engineering*. New York: Springer.

Hardy, G. H. (1905). On double Fourier series, and especially those which represent the double zeta-function with real and incommensurable parameters. *Quarterly Journal of Mathematics 37*, 53–79.

Hobson, E. W. (1927). *The theory of functions of a real variable and the theory of Fourier's series* (3rd ed.), Volume 1. Cambridge: Cambridge University Press. Republished by Dover (1957).

Krause, M. (1903a). Über mittelwertsätze im gebiete der doppelsummen and doppelintegrale. *Leipziger Ber. 55*, 239–263.

Krause, M. (1903b). Über Fouriersche reihen mit zwei veränderlichen grössen. *Leipziger Ber. 55*, 164–197.

Liu, R. and A. Owen (2003). Estimating mean dimensionality. Technical report, Stanford University, Department of Statistics.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte Carlo Methods*. Philadelphia, PA: S.I.A.M.

Sobol', I. M. (1961). An exact estimate of the error in multi-dimensional quadrature formulas for function classes $\widetilde{W}_1^0$ and $\widetilde{H}_1^0$. *USSR, Comput Maths and Math Physics*, 228–240.

Sobol', I. M. (1973). Calculation of improper integrals using uniformly distributed sequences. *Soviet Math Dokl 14*(3), 734–738.

Young, W. H. (1913). *Proceedings of the London Mathematical Society (2) 11*, 142.