Journal of Universal Computer Science, vol. 14, no. 6 (2008), 938-955
submitted: 31/10/07, accepted: 18/2/08, appeared: 28/3/08 © J.UCS

The Bit-Complexity of Finding Nearly Optimal Quadrature Rules for Weighted Integration

Volker Bosserhoff
(Universität der Bundeswehr, Munich, Germany
volker.bosserhoff@unibw.de)

Abstract: Given a probability measure ν and a positive integer n. How to choose n knots and n weights such that the corresponding quadrature rule has the minimum worst-case error when applied to approximate the ν-integral of Lipschitz functions? This question has been considered by several authors. We study this question whithin the framework of *Turing machine-based real computability and complexity theory* as put forward by [Ko 1991] and others. After having defined the notion of a *polynomial-time computable probability measure* on the unit interval, we will show that there are measures of this type for which there is no computable optimal rule with two knots. We furthermore characterize – in terms of difficult open questions in discrete complexity theory – the complexity of computing rules whose worst-case error is arbitrarily close to optimal.

Key Words: real computational complexity, quadrature rules
Category: F.4.1, F.2, G.1.4

1 Introduction

Let ν be a probability measure on $\mathbb{R}$. Fix an integer $n \geq 1$. In numerical mathematics, the functional $I_\nu := \int d\nu$ is often approximated by means of a quadrature formula

$$I_\nu(f) = \int f \, d\nu \approx \sum_{i=1}^{n} a_i f(x_i)$$

with certain $(\mathbf{x}, \mathbf{a}) \in D_n$, where

$$D_n := \{(\mathbf{x}, \mathbf{a}) \in \mathbb{R}^n \times \mathbb{R}^n \ : \ x_1 \leq \ldots \leq x_n\}.$$

Let us call a pair $(\mathbf{x}, \mathbf{a}) \in D_n$ an *integration rule*.

For every $c \geq 0$, let $\mathfrak{L}(c)$ be the class of all Lipschitz continuous functions with Lipschitz constant $\leq c$. The worst-case error of the integration rule $(\mathbf{x}, \mathbf{a})$ on $\mathfrak{L}(1)$ is defined as

$$e^{\mathrm{wor}}(\nu, n; \mathbf{x}, \mathbf{a}) := \sup_{f \in \mathfrak{L}(1)} \left| I_\nu(f) - \sum_{i=1}^{n} a_i f(x_i) \right|.$$

It is well-known (see e.g. [Rachev 1991, p. 73, eq. (4.3.11)]) that $e^{\mathrm{wor}}(\nu, n; \mathbf{x}, \mathbf{a})$ is equal to the L_1-distance between the distribution function F_ν of ν and the distribution function of the (signed) measure $\sum_{i=1}^{n} a_i \delta_{x_i}$, where δ_{x_i} is the Dirac measure at x_i. This means that if $(\mathbf{x}, \mathbf{c})$ is in D_n,

$$S_{\mathbf{x},\mathbf{c}} := \sum_{i=1}^{n-1} c_i \chi_{[x_i, x_{i+1}[} + c_n \chi_{[x_n, \infty)},$$

$a_1 = c_1$, and $a_i = c_i - c_{i-1}$ for $i = 2, \ldots, n$, then $e^{\mathrm{wor}}(\nu, n; \mathbf{x}, \mathbf{a}) = E(\nu, n; \mathbf{x}, \mathbf{c})$, where

$$E(\nu, n; \mathbf{x}, \mathbf{c}) := \int_{\mathbb{R}} |F_\nu - S_{\mathbf{x},\mathbf{c}}| \, d\lambda. \tag{1}$$

The problem of finding good quadrature rules is hence equivalent to minimizing $E(\nu, n; ., .)$ on D_n.

Using this equivalent formulation, the problem of finding good quadrature rules for the weighted integration of Lipschitz functions has been studied by [Curbera 1998] for ν being Gaussian, and then by [Mathé 1998] for measures that fulfill more general analytic properties.[1] Mathé gives formulas for knots and weights that are asymptotically optimal for $n \to \infty$. [Behrends 1997] considers the "problem to derive conditions under which optimal quadrature rules can explicitly and easily be found". He gives analytic conditions which assure that the optimal rule is unique, as well as an algorithm to approximate the optimal knots and weights in this case.

The present work classifies the problem of finding good quadrature rules in terms of real computational complexity theory as put forward by [Ko 1991]: If the measure ν can be computed in polynomial time, can a minimum point of (1) be computed in polynomial time? We refer to [Ko 1991] for the definition of polynomial-time computability of real numbers and functions. We will call a probability measure "polynomial-time computable" if it is supported on $[0, 1]$ and has a polynomial-time computable distribution function (see [Section 2] for details); the class of all polynomial-time computable probability measures shall be denoted by $\mathbf{P}_{\mathcal{M}}$. If $\nu \in \mathbf{P}_{\mathcal{M}}$ and $n \geq 1$, then $E(\nu, n; ., .)$ always attains its minimum

$$E^{\mathrm{opt}}(\nu, n) := \inf_{(\mathbf{x},\mathbf{c}) \in D_n} E(\nu, n; \mathbf{x}, \mathbf{c}),$$

as we will see below. We will, however, also see:

Theorem 1. *There exists a measure $\nu \in \mathbf{P}_{\mathcal{M}}$ such that $E(\nu, 2; ., .)$ does not attain its minimum at any computable point.*

In view of this result, it does not make sense to ask for the complexity of optimal integration rules in general. But in practical situations, one is in fact not interested in integration rules whose knots and weights are close to optimal knots and weights, but merely in rules whose worst-case error is close to

[1] Mathé also considers more classes of integrands than just Lipschitz functions.

the optimal error. Denote by $\mathbb{D}$ the set of dyadic rationals. Elements $(\mathbf{x}, \mathbf{c})$ of $\bigcup_{n\geq 1}((D_n \cap (\mathbb{D}^n \times \mathbb{D}^n))$ can be encoded by words $\langle \mathbf{x}, \mathbf{c} \rangle \in \Sigma^*$ (where $\Sigma := \{\mathtt{0}, \mathtt{1}\}$) in a canonical way. Let $\mathbf{FP}$ be the class of all word functions $\Sigma^* \to \Sigma^*$ computable by a deterministic Turing machine in polynomial time.[2] For any function complexity class $\mathcal{C}$, let $\mathcal{C}_1$ denote the class $\{f|_{\{\mathtt{0}\}^*} \ : \ f \in \mathcal{C}\}$. We consider the following statements:

Statement 2 *(i) For every $\nu \in \mathbf{P}_\mathcal{M}$ there exists a function $\phi \in \mathbf{FP}_1$ such that for all $n \in \mathbb{N} \setminus \{0\}$ and $k \in \mathbb{N}$ one has that $\phi(\langle \mathtt{0}^n, \mathtt{0}^k \rangle)$ is an (encoded dyadic) element $(\mathbf{x}, \mathbf{c})$ of D_n with*

$$E(\nu, n; \mathbf{x}, \mathbf{c}) - E^{\text{opt}}(\nu, n) \leq 2^{-k}.$$

(ii) For every $\nu \in \mathbf{P}_\mathcal{M}$ there exists a function $\phi \in \mathbf{FP}_1$ such that for all $k \in \mathbb{N}$ one has that $\phi(\mathtt{0}^k)$ is an (encoded dyadic) element $(\mathbf{x}, \mathbf{c})$ of D_2 with

$$E(\nu, 2; \mathbf{x}, \mathbf{c}) - E^{\text{opt}}(\nu, 2) \leq 2^{-k}.$$

We will prove that Statements 2(i), 2(ii) and the following statement in discrete complexity theory are equivalent: For every predicate R on Σ^* and every $v \in \Sigma^*$ consider the expression

$$\text{count}_R(v) := \text{card}\{w \in \Sigma^* \ : \ |w| = |v|, \ R(\langle v, w \rangle)\},$$

and for every $m \in \mathbb{N}$ the expression

$$\text{maxcount}_R(m) := \max\{\text{count}_R(v) \ : \ |v| = m\}.$$

Statement 3 *For every predicate $R \in \mathbf{P}$ there exists a polynomial-time computable function $\phi : \{0\}^* \to \Sigma^*$ such that*

$$(\forall\, m \in \mathbb{N}) \ \text{count}_R(\phi(\mathtt{0}^m)) = \text{maxcount}_R(m).$$

In order to establish this equivalence, we will first consider an auxiliary problem: For every $g \in C[0,1]$, denote by $\overline{g} \in C[0,1]$ the function

$$\overline{g}(s) := \int_0^s g(t)\, dt.$$

Statement 4 *If g is in $\mathbf{P}_{C[0,1]}$, then there exists a function $\gamma \in \mathbf{FP}_1$ such that for all $k \in \mathbb{N}$ one has that $\gamma(\mathtt{0}^k)$ is an (encoded dyadic) element t of $[0,1]$ with*

$$\max(\overline{g}) - \overline{g}(t) \leq 2^{-k}.$$

[2] See [Ko 1991] for the exact definitions of the discrete complexity classes used in this paper.

Our main result is:

Theorem 5. *Statements 2(i), 2(ii), 3 and 4 are equivalent.*

Our proof relies on methods from [Ko 1991] in combination with new constructions. In fact, the problem of finding the maximum *value* of $\overline{f}$ given f is a "concatenation" of problems treated in [Ko 1991], where the following is shown:

$$[g \in \mathbf{P}_{C[0,1]} \Rightarrow \overline{g} \in \mathbf{P}_{C[0,1]}] \iff \mathbf{FP} = \#\mathbf{P},$$
$$\mathbf{P}_1 = \mathbf{P}_1^{\mathbf{NP}} \implies [g \in \mathbf{P}_{C[0,1]} \Rightarrow \max(g) \in \mathbf{P}_{\mathbb{R}}] \implies \mathbf{P}_1 = \mathbf{NP}_1.$$

Here $\#\mathbf{P}$ is the class of all functions that count the number of accepting paths of a nondeterministic polynomial-time Turing machine.

How do the statements shown equivalent in Theorem 5 relate to better-known open questions in discrete complexity theory? Unfortunately, the only result we have in this direction is elementary and leaves a wide gap:

Proposition 6. $\mathbf{FP} = \#\mathbf{P} \implies$ *Statement 3* $\implies \mathbf{P}_1 = \mathbf{NP}_1$.

Proof. First implication: Let R, count_R and $\mathrm{maxcount}_R$ be as in Statement 3. Assume $\mathbf{FP} = \#\mathbf{P}$. Then count_R is in $\mathbf{FP}$. One also has $\mathbf{P} = \mathbf{NP}$, and hence the predicate R_1 with

$$R_1(\langle \mathtt{0}^m, w\rangle) \;:\Leftrightarrow\; (\exists\, v)\ [|v| = m \text{ and } \mathrm{count}_R(v) \geq \mathrm{bin}(w)]$$

is in $\mathbf{P}$ (where $\mathrm{bin}(w) \in \mathbb{N}$ is usual binary interpretation of the word w). A binary search algorithm using R_1 yields $\mathrm{maxcount}_R \in \mathbf{FP}_1$. $\mathbf{P} = \mathbf{NP}$ again yields that the predicate R_2 with

$$R_2(\langle \mathtt{0}^m, v\rangle) \;:\Leftrightarrow\; (\exists\, w)\ [|vw| = m \text{ and } \mathrm{count}_R(vw) = \mathrm{maxcount}_R(m)]$$

is in $\mathbf{P}$. A binary search algorithm using R_2 yields that there is a function ϕ as in Statement 3.

Second implication: Let A be a language in $\mathbf{NP}_1$. There is a predicate R' in $\mathbf{P}$ and a polynomial p (w.l.o.g. $p(n) \geq 1$) such that

$$\mathtt{0}^m \in A \iff (\exists\, w)\ [|w| < p(m) \ \wedge\ R'(\langle \mathtt{0}^m, w\rangle)].$$

Choose the following as the predicate R in Statement 3:

$$R(\langle v, w\rangle) \;:\Longleftrightarrow\; v, w \in \{\mathtt{0}\}^* \text{ or } \left[\begin{array}{l} v \notin \{\mathtt{0}\}^* \text{ and there are } m, n \in \mathbb{N} \text{ and} \\ b, c_0, \ldots, c_{n-1} \in \{\mathtt{0}, \mathtt{1}\}, \\ \text{such that } n < p(m) \text{ and} \\ w = b\mathtt{0}^{p(0)}\mathtt{0}^{p(1)}\mathtt{0}^{p(m-1)}c_0 \cdots c_{n-1}, \\ \text{and } R'(\langle \mathtt{0}^m, c_0 \cdots c_{n-1}\rangle) \end{array}\right].$$

R is designed such that

$$\begin{aligned}\mathtt{0}^m \in A &\iff \bigvee_{n=0}^{p(m)-1} [\mathrm{count}_R(\mathtt{1}\mathtt{1}^{p(0)} \cdots \mathtt{1}^{p(m-1)}\mathtt{1}^n) \geq 2] \\ &\iff \bigvee_{n=0}^{p(m)-1} [\phi(\mathtt{0}\mathtt{0}^{p(0)} \cdots \mathtt{0}^{p(m-1)}\mathtt{0}^n) \notin \{\mathtt{0}\}^*],\end{aligned}$$

where ϕ is as in Statement 3. Hence, if $\phi \in \mathbf{FP}_1$, then $A \in \mathbf{P}_1$. □

2 Polynomial-time computable probability measures

We assume that the reader is familiar with the basic definitions found in [Ko 1991] concerning the computability and complexity of real numbers and functions. In this section, we motivate and define a notion of polynomial-time computability for probability measures on $\mathbb{R}$.

Following a definition by [Weihrauch 1999] (which has been generalization by [Schröder 2007]) one calls a probability measure ν on $\mathbb{R}$ computable if the set

$$\{(r, s, t) \in \mathbb{Q}^3 \ : \ r < s, \, \nu(]r, s[) > t\}$$

is computably enumerable. This definition is motivated by topological considerations on the space of probability measures (see [Weihrauch 1999, Schröder 2007]), but it does not induce a notion of a measure's computational complexity. A measure ν is *diffuse* if $\nu(\{s\}) = 0$ for every $s \in \mathbb{R}$. It is easy to see that a diffuse probability measure is computable in the above sense if, and only if, the function

$$\{(r, s) \in \mathbb{R} \times \mathbb{R} \ : \ r < s\} \to \mathbb{R}, \qquad (r, s) \mapsto \nu(]r, s[)$$

is computable. This is again equivalent to the computability of the measure's distribution function $F_\nu : \mathbb{R} \to [0, 1]$, defined by

$$F_\nu(s) := \nu(] - \infty, s]).$$

In [Ko 1991], polynomial-time computability is only defined for functions with compact domains. We will hence restrict ourselves to diffuse probability measures with $\nu([0, 1]) = 1$. These measures correspond one-to-one to continuous distribution functions F_ν with $F_\nu(0) = 0$ and $F_\nu(1) = 1$.

Definition 7. A probability measure ν on the Borel subsets of $\mathbb{R}$ is polynomial-time computable if it has a distribution function F_ν such that $F_\nu(0) = 0$, $F_\nu(1) = 1$ and $F_\nu|_{[0,1]} \in \mathbf{P}_{C[0,1]}$. Denote by $\mathbf{P}_{\mathcal{M}}$ the class of all polynomial-time computable probability measures.

3 An auxiliary problem

We define the following condition on a function g:

$$g \in \mathbf{P}_{C[0,1]} \cap \mathfrak{L}(1), \qquad g(0) = g(1) = 0, \qquad \int_0^1 g(s)\, ds = 0. \tag{2}$$

The following proposition is merely a variation of Ko's generalization of a well-known example going back to Specker (see [Ko 1991, Corollary 3.3]). It will be needed in the proof of Theorem 1 below.

Proposition 8. *There is a function g such that* (2) *is fulfilled and the maximum of $\overline{g}$ is not attained at any computable point.*

Proof. For all $0 \leq s < t \leq 3$, let $h_{s,t} : [0,3] \to \mathbb{R}$ be the polygon function that is zero outside $[s,t]$, $\wedge$-shaped with height $(t-s)/48$ on $[s,(s+t)/2]$, and $\vee$-shaped with height $-(t-s)/48$ on $[(s+t)/2,t]$. Clearly, all functions $h_{s,t}$ are in $\mathfrak{L}(1/12)$. It is well-known (see e.g. the proof of [Ko 1991, Corollary 3.3]) that there is a computable function $\phi : \mathbb{N} \to \mathbb{D}$ such that $S = \bigcup_{i\in\mathbb{N}}]\phi(2i), \phi(2i+1)[$ is contained in $[0.5, 2.5]$, and $[1,2] \setminus S \neq \emptyset$, and S contains all computable points of $[1,2]$. Define

$$(\forall\, n \in \mathbb{N})\ g_n := h_{\phi(2n),\phi(2n+1)}.$$

Let M be a TM computing ϕ, and let $t(n)$ be the total number of moves for M to run on inputs $0, 1, \ldots, 2n+1$; w.l.o.g. $t(n) \geq n$. Define $f := \sum_{n\in\mathbb{N}} 2^{-t(n)} g_n$. It can be shown (similarly as in the proof of [Ko 1991, Theorem 3.1, "(a)→(c)"]) that $f \in \mathbf{P}_{C[0,3]}$. It is furthermore easy to see that f is in $\mathfrak{L}(1/6)$, $f(0) = f(3) = 0$, the minimum of $\overline{f}$ is 0, and $\overline{f}^{-1}\{0\} \cap [1,2]$ is nonempty and contains only uncomputable points. Let $u : [0,3] \to \mathbb{R}$ be the polygon function that is $\wedge$-shaped with height $1/12$ on $[0,1]$, constantly zero on $[1,2]$, and $\vee$-shaped with height $-1/12$ on $[2,3]$. Consider $v := u - f$. v clearly is in $\mathbf{P}_{C[0,3]} \cap \mathfrak{L}(1/3)$, and we have $v(0) = v(3) = 0$. Note that $\overline{u}$ attains its maximum $1/24$ exactly on $[1,2]$, and hence $\overline{v} = \overline{u} - \overline{f}$ attains this same maximum exactly at the minimum points of $\overline{f}$ that lie in $[1,2]$; recall that these are all uncomputable. Finally, define g by $g(x) := v(3x)$ for all $x \in [0,1]$. Then g has the asserted properties. $\square$

It will be technically convenient to consider the following weaker version of Statement 4:

Statement 9 *If g fulfills* (2)*, then there exists a function $\gamma \in \mathbf{FP}_1$ such that for all $k \in \mathbb{N}$ one has that $\gamma(\mathtt{0}^k)$ is an (encoded dyadic) element t of $[0,1]$ with*

$$\max(\overline{g}) - \overline{g}(t) \leq 2^{-k}. \tag{3}$$

The proof of the next proposition is similar to parts of the proofs of Theorems 3.11 and 5.32 in [Ko 1991].

Proposition 10. *Statement 9 implies Statement 3.*

Proof. For every word $w \in \Sigma^*$, denote by $I_w \subseteq [0,1]$ the closed interval of all real numbers that have a binary expansion of the form $\mathtt{0}.w\ldots$. For every $m \in \mathbb{N}\setminus\{0\}$ define

$$\mathbb{V}_m := \{(v_1,\ldots,v_m) \in (\Sigma^*)^m \ : \ |v_i| = i\},$$
$$\mathbb{W}_m := \{((v_1,\ldots,v_m),w) \in \mathbb{V}_m \times \Sigma^* \ : \ |w| = m\}.$$

With each $\mathbf{v} = (v_1,\ldots,v_m) \in \mathbb{V}_m$ we associate the interval

$$I_{\mathbf{v}} := I_{v_1\tau(v_2)\tau(v_3)\ldots\tau(v_m)},$$

where $\tau : \Sigma^* \setminus \{\varepsilon\} \to \Sigma^*$ is defined by

$$\tau(a_0a_1\ldots a_{n-1}) := \begin{cases} \mathtt{01}a_1\ldots a_{n-1} & \text{if } a_0 = \mathtt{0}, \\ \mathtt{10}a_1\ldots a_{n-1} & \text{else.} \end{cases}$$

For all $(\mathbf{v},w) \in \mathbb{W}_m$, define

$$I^+_{(\mathbf{v},w)} := I_{v_1\tau(v_2)\tau(v_3)\ldots\tau(v_m)\mathtt{00}w} \qquad \text{and} \qquad I^-_{(\mathbf{v},w)} := I_{v_1\tau(v_2)\tau(v_3)\ldots\tau(v_m)\mathtt{11}w}.$$

Note that

$$|I^+_{(\mathbf{v},w)}| = |I^-_{(\mathbf{v},w)}| = 2^{-(2m+m(m+1)/2+1)} =: \delta_m. \tag{4}$$

For $\pm \in \{+,-\}$, let $h^{\pm}_{\mathbf{v},w} : [0,1] \to \mathbb{R}$ be the polygon function that is 0 outside $I^{\pm}_{(\mathbf{v},w)}$, and $\wedge$-shaped with height $\delta_m/2$ on $I^{\pm}_{(\mathbf{v},w)}$. Note that for all $(\mathbf{v},w) \in \mathbb{W}_m$ and $\pm \in \{+,-\}$ we have

$$\int_0^1 h^{\pm}_{\mathbf{v},w}(t)\,dt = \frac{\delta_m^2}{4}. \tag{5}$$

Now let R be a predicate as in Statement 3 and define for every $n \in \mathbb{N}$

$$g_n := \sum_{m=1}^{n} \sum_{\substack{(\mathbf{v},w)\in\mathbb{W}_m,\\ R(v_m,w)}} (h^+_{\mathbf{v},w} - h^-_{\mathbf{v},w}) \qquad \text{and} \qquad g := \lim_{n\to\infty} g_n$$

(see Figure 1).

It is not hard to verify that g fulfills (2). (To see that $g \in \mathbf{P}_{C[0,1]}$, note that it is sufficient to construct a polynomial-time computable mapping $\psi : (\mathbb{D} \cap [0,1]) \times \mathbb{N} \to \mathbb{D}$ such that $|\psi(d,k) - g(d)| \le 2^{-k}$ for all d,k; cf. [Ko 1991, Corollary 2.21]. $|g - g_k| \le \delta_{k+1}/2 \le 2^{-k}$ for all $k \ge 1$, so it is sufficient to choose $\psi(d,k) := g_k(d)$, which is easily seen to be computable in polynomial time.)

Under the assumption that Statement 9 holds true, we now construct a function ϕ as in Statement 3. It follows from (5) and the construction of g that

$$\max(\overline{g}) = \frac{1}{4}\sum_{m=1}^{\infty} \mathrm{maxcount}_R(m) \cdot \delta_m^2,$$

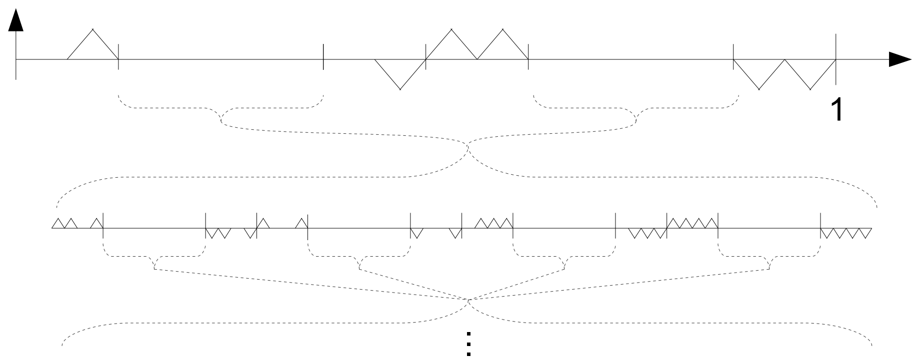

Figure 1: The first two steps in the assembly of the graph of g as constructed in the proof of Proposition 10. In this example R fulfills $\langle 0,1\rangle$, $\langle 1,0\rangle$, $\langle 1,1\rangle \in R$, $\langle 1,0\rangle \notin R$, and $\langle 00,00\rangle$, $\langle 00,01\rangle$, $\langle 00,11\rangle$, $\langle 01,00\rangle$, $\langle 01,11\rangle$, $\langle 10,01\rangle$, $\langle 10,10\rangle$, $\langle 10,11\rangle$, $\langle 11,00\rangle$, $\langle 11,01\rangle$, $\langle 11,10\rangle$, $\langle 11,11\rangle \in R$, $\langle 00,10\rangle$, $\langle 01,01\rangle$, $\langle 01,10\rangle$, $\langle 10,00\rangle \notin R$.

and for every $n \geq 1$, $\mathbf{v} \in \mathbb{V}_n$ one has

$$\sup \overline{g}(I_{\mathbf{v}}) = \frac{1}{4}\Big(\sum_{m=1}^{n} \mathrm{count}_R(v_m) \cdot \delta_m^2 + \sum_{m=n+1}^{\infty} \mathrm{maxcount}_R(m) \cdot \delta_m^2\Big)$$

$$\sup \overline{g}\Big(\bigcup_{|w|=m} I_{\mathbf{v},w}\Big) = \frac{1}{4}\sum_{m=1}^{n} \mathrm{count}_R(v_m)\delta_m^2.$$

If Statement 9 is applied to the function g, this yields that there is a function γ such that (3) holds. It is then clear from the above formulas that for every $m \geq 1$ with $\mathrm{maxcount}_R(m) > 0$ one has that there is a $\mathbf{v} \in \mathbb{V}_m$ such that $\gamma(0^{-\lfloor \log(\delta_m^2/4)\rfloor}) \in I_{\mathbf{v}}$ and $\mathrm{count}_R(v_i) = \mathrm{maxcount}_R(i)$ for $i = 1, \ldots, m$. A suitable ϕ can hence be computed as follows: On input ε, put out ε; on input 0^m, $m \geq 1$, compute $\gamma(0^{-\lfloor \log(\delta_m^2/4)\rfloor}) =: x_m$ and search for a $v \in \Sigma^*$ such that x_m is in an interval of the form $I_{\mathbf{v}} \in \mathbb{V}_m$, $v_m = v$. If such a v is found, put it out; the above considerations show that this output is valid. If no such v exists, then $\mathrm{maxcount}_R(m)$ must be zero, i.e. any output of length m is valid. $\square$

4 Reduction from the auxiliary problem

4.1 Optimal integration rules

This subsection is based on ideas already used in [Curbera 1998, Behrends 1997, Mathé 1998]. We fix a probability measure ν supported on $[0,1]$ whose distribution function F is continuous and strictly increasing on $[0,1]$; we also fix a positive integer n. We can hence omit ν and n in our notation and simply write $E(\mathbf{x}, \mathbf{c})$, E^{opt}. A pair $(\mathbf{x}, \mathbf{c}) \in D_n$ shall be called *optimal* if $E(\mathbf{x}, \mathbf{c}) = E^{\mathrm{opt}}$.

4.1.1 Characterization and existence of optimal rules

For every $(\mathbf{x}, \mathbf{c}) \in D_n$, one can write $E(\mathbf{x}, \mathbf{c})$ as

$$\int_{]-\infty, x_1]} F(t)\, \lambda(dt) + \sum_{i=1}^{n-1} \int_{[x_i, x_{i+1}]} |F(t) - c_i|\, \lambda(dt) + \int_{[x_n, \infty[} |F(t) - c_n|\, \lambda(dt). \quad (6)$$

We first make the following observation:

$$E(\mathbf{x}, \mathbf{c}) < \infty \iff c_n = 1. \quad (7)$$

Let us hence define $D'_n := \{(\mathbf{x}, \mathbf{c}) \in D_n \ : \ c_n = 1\}$.

Lemma 11. *There exists an optimal* $(\mathbf{x}, \mathbf{c})$ *in* $D_n \cap ([0,1]^n \times [0,1]^n)$.

Proof. For any $\mathbf{x} \in \mathbb{R}^n$, put

$$\tau(\mathbf{x}) := (\max(\min(x_1, 1), 0)), \ldots, \max(\min(x_n, 1), 0))).$$

For $(\mathbf{x}, \mathbf{c}) \in D$, one clearly has from (6) and $F(0) = 0$ and $F(1) = 1$

$$E(\mathbf{x}, \mathbf{c}) \geq E(\tau(\mathbf{x}), \tau(\mathbf{c})). \quad (8)$$

The claim now follows from the continuity of E (which is apparent from (6)) and the compactness of $[0,1]^n \times [0,1]^n$.

4.1.2 The partial derivatives of E

For $i = 1, \ldots, n-1$, the partial derivative $(\partial/\partial c_i)E$ exists on D'_n.

$$\frac{\partial}{\partial c_i} E(\mathbf{x}, \mathbf{c}) = \lambda([x_i, x_{i+1}] \cap [F < c_i]) - \lambda([x_i, x_{i+1}] \cap [F > c_i]), \quad (9)$$

and so

$$\frac{\partial}{\partial c_i} E(\mathbf{x}, \mathbf{c}) = 0 \iff \lambda([x_i, x_{i+1}] \cap [F < c_i]) = \lambda([x_i, x_{i+1}] \cap [F > c_i]). \quad (10)$$

Taking into account that F is strictly increasing on $[0,1]$, this can be written as

$$\frac{\partial}{\partial c_i} E(\mathbf{x}, \mathbf{c}) = 0 \iff \left[x_i = x_{i+1} \text{ or } c_i = F\left(\frac{x_{i+1} + x_i}{2} \right) \right] \quad (11)$$

if $0 \leq x_i \leq x_{i+1} \leq 1$.

For $i = 1, \ldots, n$, the partial derivative $(\partial/\partial x_i)E$ exists on the interior of D'_n; taking into account that F is continuous, one has

$$\frac{\partial}{\partial x_i} E(\mathbf{x}, \mathbf{c}) = \begin{cases} F(x_1) - |F(x_1) - c_1| & \text{if } i = 1, \\ |F(x_i) - c_{i-1}| - |F(x_i) - c_i| & \text{if } 2 \le i \le n. \end{cases} \tag{12}$$

This yields

$$\frac{\partial}{\partial x_i} E(\mathbf{x}, \mathbf{c}) = 0 \iff \begin{cases} c_1 = 0 \text{ or } F(x_1) = \frac{c_1}{2} & \text{if } i = 1, \\ c_i = c_{i-1} \text{ or } F(x_i) = \frac{c_i + c_{i-1}}{2} & \text{if } 2 \le i \le n. \end{cases} \tag{13}$$

4.1.3 Properties of optimal and relatively optimal rules

In the following, we will also be interested in relative optima for fixed (maybe subobtimal) x_n-knots. For $t \in \mathbb{R}$, put

$$E^{\text{relopt}}(t) := \inf_{\substack{x_1 \le \ldots \le x_{n-1} \le t \\ c_1, \ldots, c_{n-1}}} E((x_1, \ldots, x_{n-1}, t), (c_1, \ldots, c_{n-1}, 1)).$$

Let us call a pair $(\mathbf{x}, \mathbf{c}) \in D_{n-1}$ *relatively optimal for* t if

$$E((x_1, \ldots, x_{n-1}, t), (c_1, \ldots, c_{n-1}, 1)) = E^{\text{relopt}}(t).$$

For any step function with less than n steps, one can always construct a step function with n steps whose L_1-distance to F is strictly smaller; this is easy to see. Hence, if $(\mathbf{x}, \mathbf{c})$ is an optimal pair, then it cannot be possible to write $S_{\mathbf{x},\mathbf{c}}$ as a step function with less than n steps, i.e. $(\mathbf{x}, \mathbf{c})$ must fulfill

$$x_1 < \ldots < x_n, \tag{14}$$

$$c_1 \neq 0 \text{ and } c_i \neq c_{i-1} \text{ for } 2 \le i \le n, \tag{15}$$

It is also obvious that an optimal pair $(\mathbf{x}, \mathbf{c})$ must fulfill

$$c_1, \ldots, c_n \in [0, 1]. \tag{16}$$

If $t > 0$, then any pair $(\mathbf{x}, \mathbf{c}) \in D_{n-1}$ which is relatively optimal for t must fulfill

$$x_1 < \ldots < x_{n-1} < t, \tag{17}$$

$$c_1 \neq 0 \text{ and } c_i \neq c_{i-1} \text{ for } 2 \le i \le n-1, \tag{18}$$

which can be seen similarly as (14) and (15).

It follows from (14) that any optimal pair $(\mathbf{x}, \mathbf{c})$ must fulfill $(\partial/\partial x_1)E(\mathbf{x}, \mathbf{c}) = (\partial/\partial x_n)E(\mathbf{x}, \mathbf{c}) = 0$. (7), (15) and (13) then yield

$$F(x_1) = \frac{c_1}{2} \qquad \text{and} \qquad F(x_n) = \frac{1 + c_{n-1}}{2}.$$

This in combination with (15) and (16) implies

$$0 < x_1 \qquad \text{and} \qquad x_n < 1. \tag{19}$$

We hence have

$$E^{\text{opt}} = \inf_{0<t<1} E^{\text{relopt}}(t). \tag{20}$$

So fix some $t \in]0,1[$. It follows from (17) that any relatively optimal pair $(\mathbf{x},\mathbf{c})$ must fulfill

$$(\partial/\partial x_1)E(\mathbf{x},\mathbf{c}) = \ldots = (\partial/\partial x_{n-1})E(\mathbf{x},\mathbf{c}) = 0.$$

(18) and (13) then yield

$$F(x_1) = \frac{c_1}{2} \tag{21}$$

and

$$F(x_i) = \frac{c_i + c_{i-1}}{2} \text{ for } i = 2,\ldots,n-1. \tag{22}$$

(21) in combination with (18) implies $x_1 > 0$. Any relatively optimal point must also fulfill

$$(\partial/\partial c_1)E(\mathbf{x},\mathbf{c}) = \ldots = (\partial/\partial c_{n-1})E(\mathbf{x},\mathbf{c}) = 0.$$

Taking $x_1 > 0$ and (17) into account, we have from (11):

$$c_{n-1} = F\left(\frac{t + x_{n-1}}{2}\right) \text{ and } c_i = F\left(\frac{x_{i+1} + x_i}{2}\right) \text{ for } 1 \leq i \leq n-2. \tag{23}$$

4.2 The reduction

We now consider a special $\nu_0 \in \mathbf{P}_{\mathcal{M}}$ which will serve as the substrate of the further construction. Its distribution function F_0 shall be zero on $]-\infty, 0]$, one on $[1,\infty[$, and the polygon with nodes

$$(0,0),\ (1/4,1/2),\ (1/2,3/4),\ (2/3,5/6),\ (13/16,7/8),\ (1,1)$$

on $[0,1]$. ν_0 fulfills all the assumptions made in the previous subsection.

Let g be a function that fulfills (2). We define $F_g : \mathbb{R} \to \mathbb{R}$ by "implanting" the graph of $g/7$ onto the segment of the graph of F_0 over $J := [A,B] := [2/3, 13/16]$: Put

$$\widetilde{g}(t) := \begin{cases} |J| \cdot g((t-A)/|J|)/7 & \text{for } t \in J, \\ 0 & \text{else,} \end{cases}$$

and $F_g(t) := F_0(t) - \widetilde{g}(t)$. As $\widetilde{g}$ is in $\mathfrak{L}(1/7)$ and F_0 has slope $2/7$ on J, we have that F_g is still strictly increasing. So F_g is the distribution function of a measure $\nu_g \in \mathbf{P}_{\mathcal{M}}$. Also note that

$$\int_A^t \widetilde{g}(s)\,ds = \frac{|J|^2}{7} \int_0^{(t-A)/|J|} g(s)\,ds,\ \ t \in J, \tag{24}$$

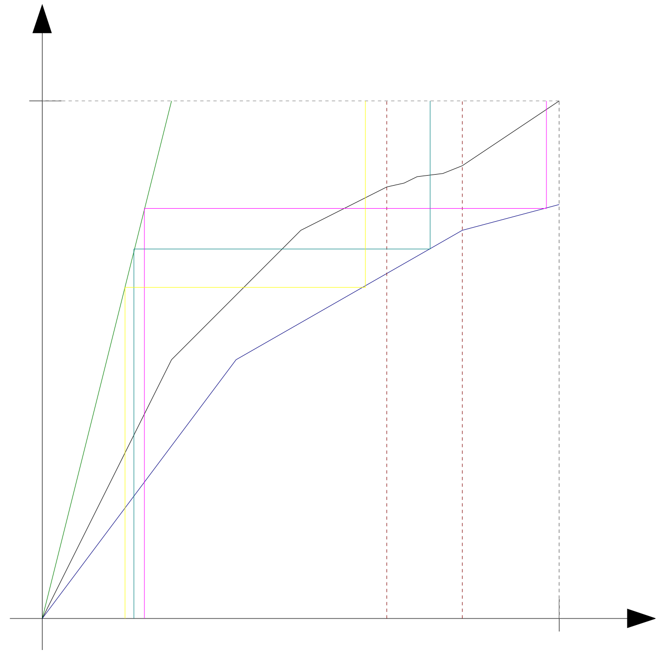

Figure 2: The black polygon is the graph of F_g on $[0,1]$. The green line is comprised of all points (x_1, c_1) with $F_g(x_1) = c_1/2$. The blue polygon is then comprised of all points (x_2, c_1) with $(\exists\, x_1)[F_g(x_1) = c_1/2 \wedge c_1 = F_g((x_2+x_1)/2)]$. The relatively optimal step functions for three different choices of x_2 are depicted in yellow, turquoise and magenta. The dashed red lines bound the stripe $J \times \mathbb{R}$.

so in particular

$$\int_A^B \widetilde{g}(s)\, ds = 0. \tag{25}$$

In the following, we will be interested in optimal and relatively optimal rules for $n = 2$. For abbreviation, we will omit n from our notation. Let us fix some $t \in]0,1[$ and look for relatively optimal x_1, c_1. x_1 and c_1 must fulfill the conditions (21) and (23). In fact, ν_0 is constructed such that for every $t \in]0,1[$, these equations have exactly one solution $x_1 = \xi(t)$, $c_1 = \zeta(t)$ which does furthermore not depend on g; this can be verified elementarily, but it is also apparent in Figure 2. Taking (25) into account, it is not hard to verify that for every g and t

$$E^{\mathrm{relopt}}(\nu_g; t) = \begin{cases} E^{\mathrm{relopt}}(\nu_0; t) - 2\int_A^t \widetilde{g}(s) & \text{for } t \in J, \\ E^{\mathrm{relopt}}(\nu_0; t) & \text{else.} \end{cases} \tag{26}$$

How to choose t such that $E^{\mathrm{relopt}}(\nu_g; t) = E^{\mathrm{opt}}(\nu_g)$? We already know that the optimal t is in $]0,1[$, and that an optimal pair $((x_1, t), (c_1, 1))$ must be relatively optimal for t and additionally fulfill $F_g(t) = (1 + c_1)/2$. These conditions

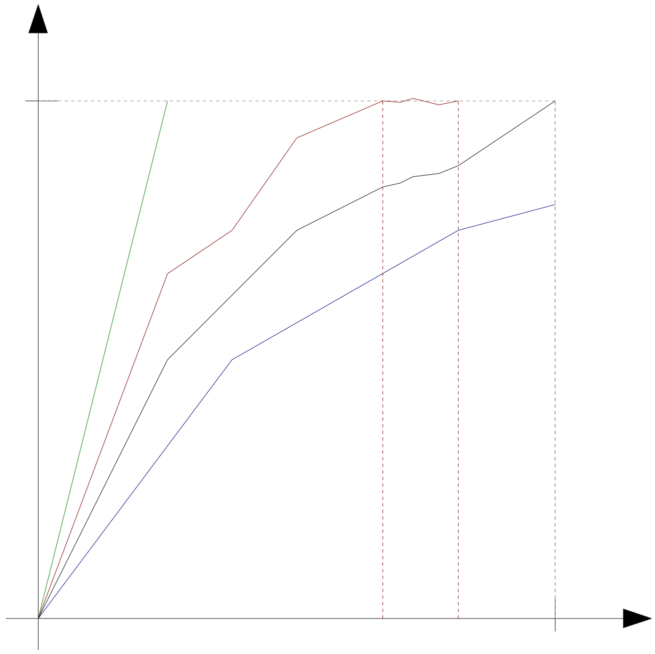

Figure 3: The black, blue, and green polygons are the same as in in Figure 2. The red polygon is comprised of all points (x_2, c_2) such that $(\exists\, x_1)(\exists\, c_1)[F_g(x_1) = c_1/2 \wedge c_1 = F_g((t+x_1)/2 \wedge F_g(x_2) = (c_2+c_1)/2]$. The remaining condition for optimality, i.e. $c_2 = 1$, is fulfilled by a point (x_2, c_2) on the red polygon if, and only if, $x_2 \in J$ and $\widetilde{g}(x_2) = 0$.

are fulfilled if, and only if, (see Figure 3)

$$x_1 = \xi(t),\ c_1 = \zeta(t),\ t \in J,\ \widetilde{g}(t) = 0.$$

Now consider the case $g = 0$. First note that the gradient

$$(\partial/\partial x_1, \partial/\partial x_2, \partial/\partial c_1)\, E(\nu_0; (x_1, x_2), (c_1, 1))$$

is continuous on $\{(x_1, x_2, c_1) \ :\ 0 < x_1 < x_2 < 1\}$. (This is obvious for $\partial/\partial x_1$ and $\partial/\partial x_2$ by (12); for $\partial/\partial c_1$ it follows from (9) taking into account that F_0 is strictly increasing on $[0, 1]$.) We hence have that $E(\nu_0; (.,.), (.,1))$ is totally differentiable. It is elementary to verify (and becomes apparent when looking at Figure 2) that ζ and ξ depend linearly and hence in particular differentiably on $t \in J$. Furthermore, recall that ξ and ζ were chosen such that the above gradient vanishes whenever $x_1 = \xi(t)$, $x_2 = t$, $c_1 = \zeta(t)$, $t \in J$. The chain rule now yields

$$(d/dt)E^{\mathrm{relopt}}(\nu_0; t) = (d/dt)E(\nu_0; (\xi(t), t), (\zeta(t), 1)) = 0$$

for all $t \in J$. So $E^{\mathrm{relopt}}(\nu_0, .)$ is constant on J, which means

$$E(\nu_0; \mathbf{x}, \mathbf{c}) = E^{\mathrm{opt}}(\nu_0) \iff (\exists t \in J)\, (\mathbf{x}, \mathbf{c}) = ((\xi(t), t), (\zeta(t), 1)). \tag{27}$$

For arbitrary g, we can now deduce from (26) and (27):

$$E^{\text{relopt}}(\nu_g;t)\begin{cases}= E^{\text{opt}}(\nu_0) - 2\int_A^t \widetilde{g}(s) & \text{for } t\in J,\\ > E^{\text{opt}}(\nu_0) & \text{else.}\end{cases} \tag{28}$$

Proof of Theorem 1. Choose g to be the function from Proposition 8. By (28) and (24), any optimal integration rule $(\mathbf{x},\mathbf{a})$ for ν_g fulfills

$$\overline{g}((x_2 - A)/|J|) = \max(\overline{g}).$$

Hence x_2 is not computable. □

Proposition 12. *Statement 2(ii) implies Statement 9*

Proof. Let g be a function that fulfills (2). If $\max(\overline{g}) = 0$, then Statement 9 of course holds true for g. Hence assume that $\max(\overline{g}) > 0$. Assume that Statement 2(ii) holds true for $\nu = \nu_g$, and let ϕ be a function as therein. Then for all $k \in \mathbb{N}$

$$\phi(0^k) = \langle \mathbf{x},\mathbf{c}\rangle \implies E^{\text{relopt}}(\nu_g;x_2) - E^{\text{opt}}(\nu_g) \le 2^{-k}.$$

By (28) and (24), we have $E^{\text{opt}}(\nu_g) < E^{\text{opt}}(\nu_0)$, i.e. there is a k_0 such that $E^{\text{opt}}(\nu_0) - E^{\text{opt}}(\nu_g) \ge 2^{-k_0}$. (28) yields that

$$k \ge k_0 \ \wedge\ E^{\text{relopt}}(\nu_g;x_2) - E^{\text{opt}}(\nu_g) \le 2^{-k} \implies x_2 \in J,$$

and hence again by (28)

$$\begin{aligned} k \ge k_0 \ \wedge\ \phi(0^k) = \langle \mathbf{x},\mathbf{a}\rangle &\implies \max_{t\in J}\int_A^t \widetilde{g}(s) - \int_A^{x_2}\widetilde{g}(s) \le 2^{-(k+1)}\\ &\implies \max(\overline{g}) - \overline{g}((x_2-A)/|J|) \le 2^{-(k+1)}\cdot 7/|J|^2. \end{aligned}$$

It is now obvious that a function γ as in Statement 9 exists. □

5 Proof of Theorem 5

The following implications are obvious:

$$\begin{aligned} \text{Statement 2(i)} &\implies \text{Statement 2(ii)},\\ \text{Statement 4} &\implies \text{Statement 9}, \end{aligned}$$

And we have already shown

$$\begin{aligned} \text{Statement 2(ii)} &\implies \text{Statement 9},\\ \text{Statement 9} &\implies \text{Statement 3}. \end{aligned}$$

In order to prove Theorem 5, it is hence sufficient to prove

$$\text{Statement 3} \implies \text{Statement 2(i)}, \tag{29}$$
$$\text{Statement 3} \implies \text{Statement 4}. \tag{30}$$

The proofs of (29) and (30) use standard techniques from [Ko 1991]. As they are quite similar, we only give the proof of (29). The following lemma (whose prove is an easy exercise) is preparatory:

Lemma 13. *Let* $f : [0,1] \to [0,1]$ *be measurable and non-decreasing. For given* $k \in \mathbb{N}$ *let* $x_1, \dots, x_{2^k} \in [0,1]$ *be numbers with* $|f(i2^{-k}) - x_i| \leq 2^{-k}$ *for* $i = 1, \dots, 2^k$*. Then*

$$\|f - \sum_{i=1}^{2^k} x_i \chi_{[(i-1)2^{-k}, i2^{-k}[}\|_{L_1([0,1])} \leq 2^{-(k-1)}.$$

□

Proof of (29). Let ν be a measure as in Statement 2(i) and let $F \in \mathbf{P}_{C[0,1]}$ be its distribution function. It is clear that for every $n \in \mathbb{N} \setminus \{0\}$ and every $(\mathbf{x}, \mathbf{c})$ in

$$T_n := \{(\mathbf{x}, \mathbf{c}) \in [0,1]^n \times [0,1]^n \ : \ x_1 \leq \ldots \leq x_n, \ c_n = 1\}$$

one has

$$E(\nu, n; \mathbf{x}, \mathbf{c}) = \|F - S_{\mathbf{x},\mathbf{c}}\|_{L_1[0,1]};$$

we furthermore know from Section 4.1 that $E(\nu, n; ., .)$ attains its minimum on T_n. For every $k \in \mathbb{N}$ put

$$\mathbb{D}_k := \{i2^{-k} \ : \ i \in \{1, \dots, 2^k\}\}$$

and

$$\mathbb{T}_{n,k} := T_n \cap (\mathbb{D}_k^n \times \mathbb{D}_k^n).$$

It is not hard to see that for every $n \in \mathbb{N} \setminus \{0\}$ and $(\mathbf{x}, \mathbf{c}) \in T_n$ there is an $(\mathbf{r}, \mathbf{s}) \in \mathbb{T}_{n,k}$ such that

$$\|S_{\mathbf{x},\mathbf{c}} - S_{\mathbf{r},\mathbf{s}}\|_{L_1([0,1])} \leq n2^{-(k-1)}. \tag{31}$$

As $F \in \mathbf{P}_{C[0,1]}$, there is a polynomial-time computable $\psi : (\mathbb{D} \cap [0,1]) \times \mathbb{N} \to \mathbb{D}$ such that

$$(\forall\, d \in \mathbb{D} \cap [0,1])(\forall\, k \in \mathbb{N}) \ [\psi(d,k) \in \mathbb{D}_k \text{ and } |\psi(d,k) - F(d)| \leq 2^{-k}].$$

It follows directly from the previous lemma that for every $k \in \mathbb{N}$

$$\|F - H_k\|_{L_1[0,1]} \leq 2^{-(k-1)}, \tag{32}$$

where $H_k : [0,1] \to [0,1]$ is defined as

$$H_k := \sum_{i=1}^{2^k} \psi(i2^{-k}, k)\chi_{[(i-1)2^{-k}, i2^{-k}[}.$$

So, if $\ell := k + \lceil \log(2+n) \rceil + 1$ and $(\mathbf{x}, \mathbf{c}) \in \mathbb{T}_{n,\ell}$ is chosen such that

$$\|H_\ell - S_{\mathbf{x},\mathbf{c}}\|_{L_1[0,1]} = \min_{(\mathbf{r},\mathbf{s}) \in \mathbb{T}_{n,\ell}} \|H_\ell - S_{\mathbf{r},\mathbf{s}}\|_{L_1[0,1]}, \tag{33}$$

then

$$\begin{aligned}
&E(\nu, n; \mathbf{x}, \mathbf{c}) - E^{\mathrm{opt}}(\nu, n) \\
&= \|F - S_{\mathbf{x},\mathbf{c}}\|_{L_1[0,1]} - \min_{(\mathbf{r},\mathbf{s}) \in T_n} \|F - S_{\mathbf{r},\mathbf{s}}\|_{L_1[0,1]} \\
&\leq \|F - S_{\mathbf{x},\mathbf{c}}\|_{L_1[0,1]} - \min_{(\mathbf{r},\mathbf{s}) \in \mathbb{T}_{n,\ell}} \|F - S_{\mathbf{r},\mathbf{s}}\|_{L_1[0,1]} + n2^{-(\ell-1)} \quad &&\text{[by (31)]} \\
&\leq \|H_\ell - S_{\mathbf{x},\mathbf{c}}\|_{L_1[0,1]} - \min_{(\mathbf{r},\mathbf{s}) \in \mathbb{T}_{n,\ell}} \|H_\ell - S_{\mathbf{r},\mathbf{s}}\|_{L_1[0,1]} \\
&\qquad + 2 \cdot 2^{-(\ell-1)} + n2^{-(\ell-1)} &&\text{[by (32)]} \\
&= 2 \cdot 2^{-(\ell-1)} + n2^{-(\ell-1)} &&\text{[by (33)]} \\
&\leq 2^{-k}.
\end{aligned}$$

In order to prove that there is a function ϕ as in Statement 2(i), it is now clearly sufficient to show that there is a function $\gamma \in \mathbf{FP}$ such that

$$\begin{aligned}
&(\forall\, n \in \mathbb{N} \setminus \{0\})(\forall\, k \in \mathbb{N})\ [\,\gamma(\langle 0^n, 0^k \rangle) = \langle \mathbf{x}, \mathbf{c} \rangle \implies \\
&\qquad\qquad \|H_k - S_{\mathbf{x},\mathbf{c}}\|_{L_1[0,1]} = \min_{(\mathbf{r},\mathbf{s}) \in \mathbb{T}_{n,k}} \|H_k - S_{\mathbf{r},\mathbf{s}}\|_{L_1[0,1]}\,].
\end{aligned}$$

For every $(\mathbf{r}, \mathbf{s}) \in \mathbb{T}_{n,k}$, one has that both H_k and $S_{\mathbf{r},\mathbf{s}}$ are step functions such that the width and the height of each step are positive multiples of 2^{-k}. The L_1-distance of the two functions can hence by computed by partitioning $[0,1] \times [0,1]$ into $2^{-k} \times 2^{-k}$-rectangles and counting how many of these are covered by none or both of the two graphs:

$$\|H_k - S_{\mathbf{r},\mathbf{s}}\|_{L_1[0,1]} = 1 - 2^{-2k} \mathrm{card} A(k, \mathbf{r}, \mathbf{s}),$$

where

$$A(k, \mathbf{r}, \mathbf{s}) := \{(x, y) \in \mathbb{D}_k^2 \ :\ \widetilde{R}(k, \mathbf{r}, \mathbf{s}, x, y)\}$$

and $\widetilde{R}$ is the predicate given by

$$\widetilde{R}(k, \mathbf{r}, \mathbf{s}, x, y) \ :\Longleftrightarrow\ [\min(\psi(x,k), S_{\mathbf{r},\mathbf{s}}(x)) \geq y \text{ or } \max(\psi(x,k), S_{\mathbf{r},\mathbf{s}}(x)) < y]\,.$$

Let $\theta : \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ be a bijective function with $\theta(i,j) \geq 2ij$ and such that θ and θ^{-1} are polynomial-time computable.[3] Let $[.] : \{0,1\}^* \to \mathbb{D}$ be defined by the condition

$$(\forall\, k \in \mathbb{N})(\forall\, p \in \{0,1\}^k)\ [p] - 2^{-k} \text{ is the number represented by } 0.p.$$

Now define the predicate $R \subseteq \{0,1\}^*$ by

$$\begin{aligned} R := \{\langle v, w\rangle\ :\ & \text{if } n,k \in \mathbb{N} \text{ are such that } \theta(n,k) = |v|, \text{ then} \\ & v = p_1 \dots p_n q_1 \dots q_n 0^{|v|-2nk}, \\ & \text{for suitable } p_1, \dots, p_n, q_1, \dots, q_n \in \Sigma^k \\ & \text{with } (([p_1], \dots, [p_n]), ([q_1], \dots, [q_n])) \in \mathbb{T}_{n,k} \\ & \text{and } w = \iota\kappa 0^{|v|-2k} \text{ for suitable } \iota, \kappa \in \{0,1\}^k \text{ and} \\ & \widetilde{R}(k, ([p_1], \dots, [p_n]), ([q_1], \dots, [q_n]), [\iota], [\kappa])\}. \end{aligned}$$

R is designed such that for all $n \in \mathbb{N} \setminus \{0\}$, $k \in \mathbb{N}$, $(\mathbf{r}, \mathbf{s}) \in \mathbb{T}_{n,k}$

$$\mathrm{count}_R([r_1]^{-1} \dots [r_n]^{-1} [s_1]^{-1} \dots [s_n]^{-1} 0^{\theta(n,k)-2nk}) = \mathrm{card}(A(k, \mathbf{r}, \mathbf{s})).$$

Furthermore, for all $n \in \mathbb{N} \setminus \{0\}$, $k \in \mathbb{N}$ and $v \in \Sigma^{\theta(n,k)}$:

$$\begin{aligned} & \mathrm{count}_R(v) = \mathrm{maxcount}_R(\theta(n,k)) \\ \Longleftrightarrow\ & \text{the first } 2nk \text{ bits of } v \text{ encode an } (\mathbf{x}, \mathbf{c}) \in \mathbb{T}_{n,k} \text{ with} \\ & \|H_k - S_{\mathbf{x},\mathbf{c}}\|_{L_1[0,1]} = \min_{(\mathbf{r},\mathbf{s}) \in \mathbb{T}_{n,r}} \|H_k - S_{\mathbf{r},\mathbf{s}}\|_{L_1[0,1]}. \end{aligned}$$

Under the assumption that Statement 3 is true and noting that R is in $\mathbf{P}$, it is now clear that a function γ as above exists. $\square$

Acknowledgements

The author would like to thank Peter Hertling for useful comments on this paper. The work was supported by DFG grant HE 2489/4-1 and by a DFG/NSFC grant.

References

[Behrends 1997] Behrends, E.: Optimal approximative integration of Lipschitz functions. Freie Universität Berlin, Fachbereich Mathematik und Informatik, Preprint A/3/97, 1997,
URL: `ftp://ftp.math.fu-berlin.de/pub/math/publ/pre/1997/Pr-A-97-03.ps`

[Curbera 1998] Curbera, F.: Optimal integration of Lipschitz functions with a Gaussian weight. *J. Complexity*, 14:122 – 149, 1998.

[Ko 1991] Ko, K.-I: *Complexity Theory of Real Functions.* Birkhäuser, Boston, 1991.

[3] One may choose e.g. $\theta(i,j) := j(j+1)/2 + i(j+1) + i(i+1)/2$.

[Mathé 1998] Mathé, P.: Asymptotically optimal weighted numerical integration. *J. Complexity*, 14:34 – 48, 1998.

[Rachev 1991] Rachev, S.T.: *Probability Metrics and the Stability of Stochastic Models.* Wiley, Chichester, 1991.

[Schröder 2007] Schröder, M.: Admissible representations for probability measures. *Math. Logic Quart.*, 53:431 – 445, 2007.

[Weihrauch 1999] Weihrauch, K.: Computability on the probability measures on the Borel sets of the unit interval. *Theoret. Comput. Sci.*, 219(1-2):421–437, 1999.