Available online at www.sciencedirect.com

ScienceDirect

Journal of Approximation Theory 244 (2019) 1–36

JOURNAL OF Approximation Theory

www.elsevier.com/locate/jat

Full Length Article

Best finite constrained approximations of one-dimensional probabilities

Chuang Xu[a], Arno Berger[b,*]

[a] Department of Mathematical Sciences, University of Copenhagen, Copenhagen, Denmark
[b] Mathematical and Statistical Sciences, University of Alberta, Edmonton, Alberta, Canada

Received 6 April 2018; received in revised form 6 February 2019; accepted 4 March 2019
Available online 18 March 2019

Communicated by A. Kuijlaars

Abstract

This paper studies best finitely supported approximations of one-dimensional probability measures with respect to the L^r-Kantorovich (or transport) distance, where either the locations or the weights of the approximations' atoms are prescribed. Necessary and sufficient optimality conditions are established, and the rate of convergence (as the number of atoms goes to infinity) is discussed. In view of emerging mathematical and statistical applications, special attention is given to the case of best uniform approximations (i.e., all atoms having equal weight). The approach developed in this paper is elementary; it is based on best approximations of (monotone) L^r-functions by step functions, and thus different from, yet naturally complementary to, the classical Voronoi partition approach.

MSC: 28A33; 60B10; 62E17

Keywords: Constrained approximation; Best uniform approximation; Asymptotically best approximation; Kantorovich distance; Balanced function; Quantile function

1. Introduction

Finding best finitely supported approximations of a given (Borel) probability measure μ on $\mathbb{R}$ is an important basic problem that has been studied extensively and from several perspectives. Assuming for instance that $\int_{\mathbb{R}} |x|^r \mathrm{d}\mu(x) < +\infty$ for some $r \geq 1$, a classical question asks to minimize the L^r-Kantorovich (or transport) distance $d_r(\nu, \mu)$ over all discrete probabilities ν

* Corresponding author.
E-mail address: berger@ualberta.ca (A. Berger).

https://doi.org/10.1016/j.jat.2019.03.005
0021-9045/

supported on at most n atoms, where n is a given positive integer. A rich theory of *quantization of probability measures* addresses this question, as well as applications thereof in such diverse fields as information theory, numerical integration, and optimal transport, among others; see, e.g., [5,17,26] and the many references therein. As is well known, a minimal value of $d_r(\nu, \mu)$ always is attained for some discrete probability $\nu = \delta_\bullet^{\bullet,n}$ which may or may not be determined uniquely by this minimality property. Moreover, $d_r(\delta_\bullet^{\bullet,n}, \mu) \to 0$ as $n \to \infty$, and the precise rate of convergence has attracted particular interest. A celebrated theorem (see Proposition 5.27) asserts that, under a mild moment condition, $\big(nd_r(\delta_\bullet^{\bullet,n}, \mu)\big)$ converges to a finite positive limit whenever μ is non-singular (w.r.t. Lebesgue measure). Results in a similar spirit have been established for important classes of singular measures, notably self-similar and -conformal probabilities [18,20,30]. While these classical results crucially employ Voronoi partitions (as developed in some detail, e.g., in [17]), alternative tools and extensions to other metrics have recently been studied also [5,7,10].

A second important perspective on the approximation problem is that of *random empirical quantization* [4,9]. To illustrate it, let $(X_j)_{j\geq 1}$ be an iid. sequence of random variables with common law μ, and consider the (random) empirical measure $\mu_n = \frac{1}{n}\sum_{j=1}^n \delta_{X_j}$; here and throughout, δ_a is a Dirac unit mass at $a \in \mathbb{R}$. Then $d_r(\mu_n, \mu) \to 0$ with probability one as $n \to \infty$, as well as $\mathbb{E}d_r(\mu_n, \mu) \to 0$. A comprehensive analysis of the rate of convergence of $\big(\mathbb{E}d_r(\mu_n, \mu)\big)$ is provided by the recent monograph [3] which, in particular, identifies necessary and sufficient conditions for decay to occur at the "standard rate" ($n^{-1/2}$), that is, for $\big(n^{1/2}\mathbb{E}d_r(\mu_n, \mu)\big)$ to be bounded above and below by finite positive constants. Beyond these one-dimensional results, rates of convergence for random empirical quantization have lately been studied in higher dimensions and other settings also; see, e.g., [4,9,13].

The purpose of the present article is to develop a third perspective on the approximation problem that in a sense lies between the two established perspectives briefly recalled above. Specifically, we present an in-depth study of finitely supported approximations that are *non-random* yet *constrained* in that either the locations or the weights of the approximations' atoms are prescribed. To the best of our knowledge, such approximations have not been studied systematically in the literature, though the recent papers [1] and [5] do consider (uniform) "$\mathcal{U}$-quantization" and discrete approximations of absolutely continuous probabilities μ, respectively. The necessary and sufficient conditions for best constrained approximations presented in this article make no assumptions on μ beyond $\int_{\mathbb{R}} |x|^r \mathrm{d}\mu(x) < +\infty$. They follow rather directly from elementary properties of monotone functions and exploit a certain duality between locations and weights of atoms. By contrast, note that Voronoi partitions are typically much less useful if weights, rather than positions, are prescribed [17].

Arguably the simplest special case where our results apply is that of best *uniform* approximations: Given μ and a positive integer n, for which $\nu = \frac{1}{n}\sum_{j=1}^n \delta_{x_j}$ is $d_r(\nu, \mu)$ minimal, where $x_1, \cdots, x_n \in \mathbb{R}$? Theorem 5.5 characterizes the (often unique) minimizer $\delta_\bullet^{\mathbf{u}_n}$; here usage of the superscript $\mathbf{u}_n$ emphasizes the fact that a *best uniform approximation* $\delta_\bullet^{\mathbf{u}_n}$ typically is quite different from any *best approximation* $\delta_\bullet^{\bullet,n}$. The special case of best uniform approximations is of considerable interest in itself: In statistics, when dealing with empirical data sets, practical considerations may demand that all atoms have equal weights, or at least that they be integer multiples of one fixed unit weight [2]. Also, best uniform approximations are close analogues of *support points* [25], the latter being minimizers relative to a slightly different metric (energy distance). One may thus view $\delta_\bullet^{\mathbf{u}_n}$ as a quasi Monte Carlo (MC) tool that minimizes the integration error bound $|\int f\,\mathrm{d}\mu - \sum_{j=1}^n f(x_j)| \leq \mathrm{Lip}(f)\, d_1(\delta_\bullet^{\mathbf{u}_n}, \mu)$ for a wide class of functions (cf. [11]), and consequently a careful analysis of $d_1(\delta_\bullet^{\mathbf{u}_n}, \mu)$ as

$n \to \infty$ is indispensable. In computational mathematics, best uniform approximations also arise naturally in the form of restricted MC methods, and a basic question is how performance of the latter compares to general (non-restricted) MC methods, that is, how $d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)$ compares to $d_r(\delta_\bullet^{\bullet,n}, \mu)$ as $n \to \infty$; see, e.g., [14,15] and the many references therein. Restricted MC methods have recently found applications in "big-data" problems in Bayesian statistics [25] and the numerical solution of SDE, notably in mathematical finance [15,16,27].

Just as for the best unconstrained and the random empirical approximations mentioned earlier, $d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) \to 0$ as $n \to \infty$, which again makes the rate of convergence a natural object of study. Presented in Section 5.2, our results in this regard are quite similar to those of [3], despite their obviously different context. As a simple illustrative example, consider the standard normal distribution, i.e., let $\frac{\mathrm{d}}{\mathrm{d}x}\mu(\,]-\infty, x]) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$ for all $x \in \mathbb{R}$. From Proposition 5.27, it follows that in the case of best (unconstrained) approximations, for all $r \geq 1$,

$$d_r(\delta_\bullet^{\bullet,n}, \mu) = \mathcal{O}(n^{-1}) \quad \text{as } n \to \infty\,,$$

whereas for random empirical approximations, [3, Sec.6.5] shows that

$$\mathbb{E}d_r(\mu_n, \mu) = \begin{cases} \mathcal{O}(n^{-1/2}) & \text{if } 1 \leq r < 2\,, \\ \mathcal{O}(n^{-1/2}(\log\log n)^{1/2}) & \text{if } r = 2\,, \\ \mathcal{O}(n^{-1/r}(\log n)^{1/2}) & \text{if } r > 2\,. \end{cases}$$

By contrast, for best uniform approximations, with $r = 2$ and along the subsequence $n = 2^k$, the sharp rate of convergence for $\big(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$ is $(n\log n)^{-1/2}$, as proved by [15]. Utilizing the main results of the present article, notably Theorem 5.5, one can show that in fact

$$d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) = \begin{cases} \mathcal{O}(n^{-1}(\log n)^{1/2}) & \text{if } r = 1\,, \\ \mathcal{O}(n^{-1/r}(\log n)^{-1/2}) & \text{if } r > 1\,. \end{cases}$$

Not too surprisingly, the rate of convergence of $\big(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$ is slower than that of $\big(d_r(\delta_\bullet^{\bullet,n}, \mu)\big)$, but faster than that of $\big(\mathbb{E}d_r(\mu_n, \mu)\big)$.

Due to the nature of the underlying approximation problem for monotone functions, our approach is not restricted to d_r, and results in a similar spirit can be established for other important metrics and for discrete approximations with countable support. One-dimensionality, on the other hand, is crucial: In multi-dimensional (Euclidean) spaces, upper bounds for the rate of decay of best uniform approximations have been established only recently [8], via a *uniform decomposition approach*. In addition, we mention [15] which analyzes a best uniform approximation problem (referred to as *random bit quadrature*) in a Hilbert space setting with L^2-Kantorovich metric, motivated also by MC applications.

This article is organized as follows. Section 2 introduces the notations used throughout, and recalls definition and basic properties of the metric d_r for the reader's convenience. Section 3 reviews several elementary facts about monotone functions and their quantile and growth sets, as well as the notion of a balanced function, to be used subsequently in Section 4 to characterize best approximations of (monotone) L^r-functions by step functions. While they may also be of independent interest, these results crucially serve as tools in Section 5, the main part of this work. In that section, necessary and sufficient conditions for best finite approximations with prescribed locations (Section 5.1) or weights (Section 5.2) are established. Much attention is devoted to the special case of best uniform approximations $\delta_\bullet^{\mathbf{u}_n}$, and in particular to the rate of convergence of $\big(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$. Convergence theorems and finite range (upper and lower) bounds for such sequences are provided. All results are illustrated via simple examples of μ which include absolutely continuous (exponential, Beta) as well as singular (Cantor, inverse Cantor) probability measures.

2. Notations

The following, mostly standard notations are used throughout. The natural and real numbers are denoted $\mathbb{N}$ and $\mathbb{R}$, respectively. The extended real numbers are $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, +\infty\}$. For any $a \in \overline{\mathbb{R}}$, $\operatorname{sgn} a = 1$ if $a > 0$, $\operatorname{sgn} 0 = 0$, and $\operatorname{sgn} a = -1$ if $a < 0$. The indicator function of any set $A \subset \overline{\mathbb{R}}$ is denoted $\mathbf{1}_A$, and log symbolizes the natural logarithm. For $x \in \mathbb{R}$, let $|x|$ be the absolute value, $\lfloor x \rfloor$ the floor (i.e., the largest integer $\le x$), and $\langle x \rangle = x - \lfloor x \rfloor$ the fractional part of x, respectively. Lebesgue measure on $\overline{\mathbb{R}}$ is symbolized by λ, and δ_a stands for the Dirac measure concentrated at a, i.e., $\delta_a(A) = \mathbf{1}_A(a)$ for all A.

The usual notations for intervals, e.g., $[a, b[= \{x \in \overline{\mathbb{R}} : a \le x < b\}$ are used. Endowed with the topology $\{[-\infty, a[\cup U \cup]b, +\infty] : a, b \in \overline{\mathbb{R}},\ U \subset \mathbb{R} \text{ open}\}$, the space $\overline{\mathbb{R}}$ is compact and homeomorphic to the unit interval $\mathbb{I} = [0, 1]$. Throughout, $I \subset \overline{\mathbb{R}}$ always denotes a closed (and hence compact) interval that is non-degenerate, i.e., $\lambda(I) > 0$. For $A \subset \overline{\mathbb{R}}$, denote by $\# A$, $\mathring{A}$, and $\overline{A}$ the cardinality (number of elements), interior, and closure of A, respectively. Every non-empty A has an infimum $\inf A$ and a supremum $\sup A$; if A is closed, then $\inf A = \min A$ and $\sup A = \max A$. If $A \subset \overline{\mathbb{R}}$ is an interval and $f : A \to \overline{\mathbb{R}}$ is monotone, then $f(a-) = \lim_{\varepsilon \downarrow 0} f(a-\varepsilon)$ and $f(a+) = \lim_{\varepsilon \downarrow 0} f(a+\varepsilon)$ both exist for every $a \in \mathring{A}$. For any set $A \subset \overline{\mathbb{R}}$ and any function $f : A \to \overline{\mathbb{R}}$, the image of A under f is $f(A) = \{f(a) : a \in A\}$, while the pre-image of $B \subset \overline{\mathbb{R}}$ is $f^{-1}(B) = \{a \in A : f(a) \in B\}$. Also, for every $b \in \overline{\mathbb{R}}$, let $\{f \le b\} = f^{-1}([-\infty, b])$; the sets $\{f \ge b\}$, $\{f < b\}$, $\{f > b\}$, and $\{f = b\}$ are defined analogously. Denote by $\operatorname{essinf}_A f$ and $\operatorname{esssup}_A f$ the essential infimum and supremum of f on A, respectively. For $1 \le r < +\infty$ and any (closed, non-degenerate) interval $I \subset \overline{\mathbb{R}}$, let $L^r(I)$ be the space of all measurable functions $f : I \to \overline{\mathbb{R}}$ that are (absolutely) r-integrable with respect to λ, and $L^\infty(I)$ the space of all functions bounded λ-almost everywhere (a.e.). For $f \in L^r(I)$, let $f^+ = \max\{f, 0\}$ and $f^- = (-f)^+$, hence $f = f^+ - f^-$.

Let $\mathcal{P}$ be the family of all Borel probability measures on $\overline{\mathbb{R}}$ with $\mu(\mathbb{R}) = 1$. For every $\mu \in \mathcal{P}$, $F_\mu : \overline{\mathbb{R}} \to \mathbb{I}$ with $F_\mu(x) = \mu([-\infty, x])$ is the associated *distribution function*, F_μ^{-1} the associated *(upper) quantile function*, i.e.,

$$F_\mu^{-1}(t) = \sup\{F_\mu \le t\}, \quad \forall\, t \in]0, 1[\,, \tag{2.1}$$

and $\operatorname{supp} \mu$ the *support* of μ, that is, the smallest closed set of μ-measure 1. Both F_μ and F_μ^{-1} are non-decreasing and right-continuous. As a consequence, F_μ^{-1} generates a positive Borel measure μ^{-1} on $]0, 1[$ via

$$\mu^{-1}(]t, u]) = F_\mu^{-1}(u) - F_\mu^{-1}(t), \quad \forall 0 < t < u < 1\,.$$

Note that μ^{-1}, referred to as the *inverse measure* of μ, is finite if and only if $\operatorname{supp} \mu$ is bounded, since in fact $\mu^{-1}(]0, 1[) = \max \operatorname{supp} \mu - \min \operatorname{supp} \mu$; see, e.g., [3, App.A] for further basic properties of inverse measures.

For every $r \ge 1$, the set of probability measures with finite rth moment is denoted $\mathcal{P}_r$, i.e., $\mathcal{P}_r = \left\{\mu \in \mathcal{P} : \int_{\mathbb{R}} |x|^r \mathrm{d}\mu(x) < +\infty\right\}$. Thus $\mu \in \mathcal{P}_r$ if and only if $F_\mu^{-1} \in L^r(\mathbb{I})$. On $\mathcal{P}_r$, the *L^r-Kantorovich distance d_r* is

$$d_r(\mu, \nu) = \left(\int_{\mathbb{I}} \left|F_\mu^{-1}(t) - F_\nu^{-1}(t)\right|^r \mathrm{d}t\right)^{1/r} = \left\|F_\mu^{-1} - F_\nu^{-1}\right\|_r, \quad \forall \mu, \nu \in \mathcal{P}_r\,. \tag{2.2}$$

For $r = 1$, by Fubini's theorem,

$$d_1(\mu, \nu) = \int_{\mathbb{R}} \left|F_\mu(x) - F_\nu(x)\right| \mathrm{d}x, \quad \forall\, \mu, \nu \in \mathcal{P}_1\,.$$

When endowed with the metric d_r, the space $\mathcal{P}_r$ is separable and complete, and $d_r(\mu_n, \mu) \to 0$ implies that $\mu_n \to \mu$ weakly. Note that $\mathcal{P}_r \supset \mathcal{P}_s$ and $d_r \le d_s$ whenever $r < s$. On $\mathcal{P}_s$, the metrics d_r and d_s are not equivalent, as the example of $\mu_n = (1 - n^{-s})\delta_0 + n^{-s}\delta_n$ shows, for which $d_s(\mu_n, \delta_0) \equiv 1$, and yet $\lim_{n\to\infty} d_r(\mu_n, \delta_0) = 0$ for all $r < s$, and hence $\mu_n \to \delta_0$ weakly. The reader is referred to [12,32] for details on the mathematical background of the Kantorovich distance, and to [17,32] for a discussion of its usefulness in the study of mass transportation and quantization problems.

3. Monotone and balanced functions and their inverses

Quantization, as informally alluded to in the Introduction, may be understood as the approximation of a given probability measure by finite weighted sums of point masses. Every quantile function is non-decreasing; in particular, the quantile function associated with a finitely supported probability measure is a monotone step function. Therefore, it is natural – not least in view of (2.2) – to formulate the ensuing approximation problem more generally as a problem about the best approximation of monotone L^r-functions by step functions. Towards this goal, we first present some relevant properties of monotone functions. For ease of exposition, the focus is on non-*decreasing* functions, but all subsequent arguments hold analogously for non-increasing functions as well.

Given an interval $I \subset \overline{\mathbb{R}}$ and a non-decreasing function $f : I \to \overline{\mathbb{R}}$, define the *t-quantile set* Q_t^f of f as

$$Q_t^f = [\inf\{f \ge t\}, \sup\{f \le t\}], \quad \forall t \in \overline{\mathbb{R}};$$

here and throughout, $\inf \varnothing := \max I$ and $\sup \varnothing := \min I$. Also remember that I is closed and non-degenerate, by convention. As a generalization of (2.1), the *(upper) inverse* function $f^{-1} : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ associated with f is

$$f^{-1}(t) := \sup\{f \le t\} = \max Q_t^f, \quad \forall t \in \overline{\mathbb{R}}.$$

Note that f^{-1} is non-decreasing, right-continuous and, on $f(I)$, coincides with the ordinary inverse of f whenever f is one-to-one. Moreover, $(f^{-1})^{-1}(x) = f(x+)$ for all $x \in \mathring{I}$; in particular, therefore, $(f^{-1})^{-1}$ equals f a.e. on $\mathring{I}$, and in fact everywhere if f is right-continuous. A few elementary properties of quantile sets are as follows.

Proposition 3.1 ([2, Lem. 2.7]).. *Let $f : I \to \overline{\mathbb{R}}$ be non-decreasing. Then, for every $t \in \overline{\mathbb{R}}$, the set Q_t^f is a non-empty, compact (possibly one-point) subinterval of I, and $f(x) = t$ whenever $\min Q_t^f < x < \max Q_t^f$; in particular, Q_t^f equals $\overline{\{f = t\}}$ whenever the latter set is non-empty. Moreover, the following hold:*

(i) *If $t < u$ then $x \le y$ for every $x \in Q_t^f$ and every $y \in Q_u^f$, and the set $Q_t^f \cap Q_u^f$ contains at most one point.*

(ii) *For every $x \in I$ and $t \in \overline{\mathbb{R}}$, $x \in Q_t^f$ if and only if $t \in Q_x^{f^{-1}}$.*

For any non-decreasing function $f : I \to \overline{\mathbb{R}}$, call $x \in I$ a *growth point* of f if $f(y) < f(x)$ for all $y \in I$ with $y < x$, or $f(y) > f(x)$ for all $y > x$; see also [3, p. 97]. Define the *growth set* of f as

$$G^f = \{x \in I : x \text{ is a growth point of } f\}.$$

Thus for example, $G^{F_\mu} = \operatorname{supp} \mu$ for every $\mu \in \mathcal{P}$, and $\{0, 1\} \subset G^{F_\mu^{-1}} \subset \mathbb{I}$. An elementary relation between growth and quantile sets follows directly from the definitions.

Proposition 3.2. *Let $f : I \to \overline{\mathbb{R}}$ be non-decreasing. Then G^f is a closed subset of I, non-empty unless f is constant, and $G^f \cup \{\min I, \max I\} = \bigcup_{t\in\overline{\mathbb{R}}}\{\min Q_t^f, \max Q_t^f\}$.*

Next, we recall a useful terminology from [6]: Given a bounded interval $I \subset \mathbb{R}$, call a measurable function $f : I \to \overline{\mathbb{R}}$ *balanced* if

$$\left|\lambda(\{f > 0\}) - \lambda(\{f < 0\})\right| \le \lambda(\{f = 0\}),$$

and denote by $B^f := \{t \in \mathbb{R} : f - t \text{ is balanced}\}$ the *set of all balanced values* of f. To establish a few basic properties of B^f (in Lemma 3.6), consider the auxiliary function $\ell_f : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ given by

$$\ell_f(t) = \tfrac{1}{2}\big(\min I + \max I + \lambda(\{f < t\}) - \lambda(\{f > t\})\big).$$

The following properties of ℓ_f are straightforward to verify.

Proposition 3.3. *Let I be a bounded interval and $f : I \to \overline{\mathbb{R}}$ a measurable function. Assume that f is finite a.e. Then the following hold:*

(i) *ℓ_f is non-decreasing;*
(ii) *For every $t \in \mathbb{R}$, $\ell_f(t\pm) = \ell_f(t) \pm \frac{1}{2}\lambda(\{f = t\})$, and hence ℓ_f is continuous at t if and only if $\lambda(\{f = t\}) = 0$. Moreover, $\lambda(\{\ell_f^{-1} < t\} \cap I) = \lambda(\{f < t\})$ as well as $\lambda(\{\ell_f^{-1} > t\} \cap I) = \lambda(\{f > t\})$;*
(iii) *$\lim_{t\to-\infty} \ell_f(t) = \ell_f(-\infty) = \min I$ and $\lim_{t\to+\infty} \ell_f(t) = \ell_f(+\infty) = \max I$;*
(iv) *If f is non-decreasing then*

$$\ell_f(t) = \tfrac{1}{2}\left(f^{-1}(t) + f^{-1}(t-)\right), \quad \forall\, t \in \mathbb{R},$$

and also

$$\ell_f^{-1}(x) = (f^{-1})^{-1}(x) = f(x+), \quad \ell_f^{-1}(x-) = f(x-), \quad \forall\, x \in \overset{\circ}{I};$$

(v) *If $f \in L^r(I)$ for some $1 \le r < +\infty$, then $\left\|\ell_f^{-1} - t\right\|_r = \|f - t\|_r$ for every $t \in \mathbb{R}$.*

Example 3.4. Let $I = \mathbb{I}$ and

$$f(x) = \begin{cases} \frac{3}{4}(x+1) & \text{if } 0 \le x < \frac{1}{3}, \\ \frac{1}{2} & \text{if } \frac{1}{3} \le x < \frac{2}{3}, \\ \frac{3}{2}x - 1 & \text{if } \frac{2}{3} \le x \le 1. \end{cases}$$

Here the functions ℓ_f, $\ell_f^{-1} : \overline{\mathbb{R}} \to \overline{\mathbb{R}}$ are given by

$$\ell_f(t) = \begin{cases} \frac{2}{3}\max\{t, 0\} & \text{if } t < \frac{1}{2}, \\ \frac{1}{2} & \text{if } t = \frac{1}{2}, \\ \frac{2}{3} & \text{if } \frac{1}{2} < t < \frac{3}{4}, \\ \frac{4}{3}\min\{t, 1\} - \frac{1}{3} & \text{if } t \ge \frac{3}{4}, \end{cases} \qquad \ell_f^{-1}(x) = \begin{cases} -\infty & \text{if } x < 0, \\ \frac{1}{2}\min\{3x, 1\} & \text{if } 0 \le x < \frac{2}{3}, \\ \frac{1}{4}(3x+1) & \text{if } \frac{2}{3} \le x < 1, \\ +\infty & \text{if } x \ge 1. \end{cases}$$

Fig. 1 illustrates that indeed $\left\|\ell_f^{-1} - t\right\|_r = \|f - t\|_r$ for all t, as asserted by Proposition 3.3(v).

Remark 3.5. By Proposition 3.3(v), minimizing $t \mapsto \|f - t\|_r$ for $f \in L^r(I)$ is equivalent to minimizing $t \mapsto \left\|\ell_f^{-1} - t\right\|_r$ for the *monotone* function ℓ_f^{-1}. Note also that if $f \in L^r(I)$ is non-decreasing then f and ℓ_f^{-1} coincide a.e., by Proposition 3.3(iv).

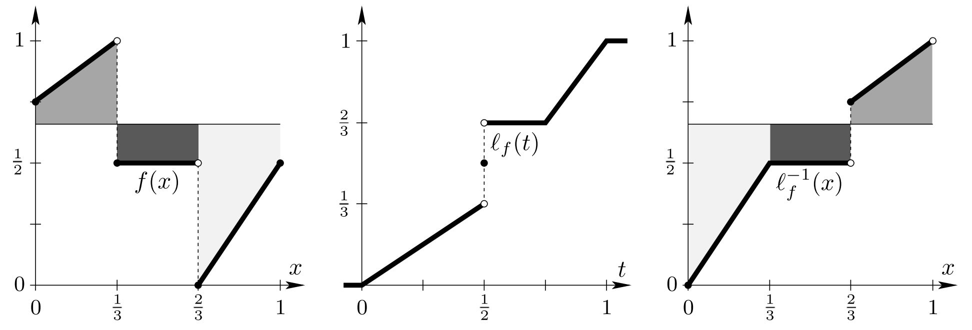

Fig. 1. Graphing f, ℓ_f, and ℓ_f^{-1} of Example 3.4.

Utilizing Propositions 3.1 and 3.3, we now establish a few basic properties of the sets B^f that will be used in the next section.

Lemma 3.6. *Let I be a bounded interval and $f : I \to \overline{\mathbb{R}}$ a measurable function. Assume that f is finite a.e. Then $B^f = Q^{\ell_f}_{\frac{1}{2}(\min I + \max I)}$. Moreover, the following hold:*

(i) *For every $t \in \mathbb{R}$, $\lambda(\{f > t\}) > \lambda(\{f \le t\})$ if $t < \min B^f$, and $\lambda(\{f < t\}) > \lambda(\{f \ge t\})$ if $t > \max B^f$;*
(ii) $\lambda\big(f^{-1}(\overset{\circ}{B^f})\big) = 0$;
(iii) $\lambda\left(\left\{f \le \min B^f\right\}\right) = \lambda\left(\left\{f \ge \max B^f\right\}\right)$.

Proof. For convenience, let $\xi = \frac{1}{2}(\min I + \max I)$, and note that, by definition, $B^f = \{t : |\ell_f(t) - \xi| \le \frac{1}{2}\lambda(\{f = t\})\}$. Define $a = \inf\{\ell_f \ge \xi\}$, $b = \sup\{\ell_f \le \xi\}$, and hence $[a, b] = Q^{\ell_f}_{\xi}$. It is easy to see that a and b are finite, with $a \le b$, and

$$\ell_f(a-) \le \xi \le \ell_f(a+), \quad \ell_f(b-) \le \xi \le \ell_f(b+),$$

which implies that $a, b \in B^f$, by Proposition 3.3(ii). For every $t \in]a, b[$, $\ell_f(t) = \xi$, thus $t \in B^f$, and hence $[a, b] \subset B^f$. For every $t > b$, $\ell_f(t-) > \xi$, so again by Proposition 3.3(ii), $\ell_f(t) - \xi > \frac{1}{2}\lambda(\{f = t\})$, which implies that $t \notin B^f$ and $\lambda(\{f < t\}) > \lambda(\{f \ge t\})$. Similarly, $t \notin B^f$ and $\lambda(\{f > t\}) > \lambda(\{f \le t\})$ for every $t < a$. This proves that $B^f = [a, b] = Q^{\ell_f}_{\xi}$, and also establishes (i).

To prove (ii) and (iii), assume that $\overset{\circ}{B^f} \ne \varnothing$, i.e., $a < b$. For every $t \in \overset{\circ}{B^f}$, $\ell_f(t) = \xi$, i.e., $\lambda(\{f > t\}) = \lambda(\{f < t\})$. Hence for all $a < t_1 < t_2 < b$,

$$\lambda(\{f < t_1\}) \le \lambda(\{f < t_2\}) = \lambda(\{f > t_2\}) \le \lambda(\{f > t_1\}) = \lambda(\{f < t_1\}).$$

Thus $\lambda(\{t_1 \le f < t_2\}) = \lambda(\{t_1 < f \le t_2\}) = 0$. Letting $t_1 \downarrow a$ and $t_2 \uparrow b$, properties (ii) and (iii) immediately follow from the continuity of λ. $\square$

Remark 3.7. If, under the assumptions of Lemma 3.6, the function f is non-decreasing, then $B^f = Q^{f^{-1}}_{\frac{1}{2}(\min I + \max I)}$.

4. Approximating L^r-functions by step functions

This section characterizes the best approximations of a given function by step functions. Two main results (Lemma 4.1 and Theorem 4.3) will be used in Section 5 to identify best finitely supported approximations of a given probability measure $\mu \in \mathcal{P}$; they may also be of independent interest. Throughout this section, we assume that the closed, non-degenerate interval $I \subset \mathbb{R}$ is *bounded*. (For unbounded I, most statements become either trivial or meaningless.)

First, we give a result on the best approximation of a monotone function by a (monotone) step function with a prescribed range and a single jump at a variable location.

Lemma 4.1. *Assume that $f : I \to \overline{\mathbb{R}}$ is non-decreasing, and $f \in L^r(I)$ for some $r \geq 1$. Let $a, b \in \mathbb{R}$ with $a < b$. Then the value of*

$$\left\| f - \left(a\mathbf{1}_{[\min I, \xi[} + b\mathbf{1}_{[\xi, \max I]}\right) \right\|_r, \quad \forall \xi \in I,$$

is minimal if and only if $\xi \in Q^f_{\frac{1}{2}(a+b)}$.

Proof. Given $f \in L^r(I)$ and $a < b$, define $\psi(\xi) = \left\| f - \left(a\mathbf{1}_{[\min I, \xi[} + b\mathbf{1}_{[\xi, \max I]}\right) \right\|_r$ for all $\xi \in I$, and let $c = \frac{1}{2}(a + b)$. Clearly, the function ψ is non-negative and continuous, and so attains a minimal value. If $\xi > f^{-1}(c)$ then there exists $0 < \varepsilon < \xi - f^{-1}(c)$ such that $f(x) > c$ for all $x \in [\xi - \varepsilon, \xi]$. Hence

$$\begin{aligned}
\psi(\xi)^r - \psi\left(f^{-1}(c)\right)^r &= \int_{f^{-1}(c)}^{\xi} (|f(x) - a|^r - |f(x) - b|^r)\,\mathrm{d}x \\
&= \int_{f^{-1}(c)}^{\xi} \left((f(x) - a)^r - (f(x) - b)^r\right) \mathbf{1}_{\{f \geq b\}} \mathrm{d}x \\
&\quad + \int_{f^{-1}(c)}^{\xi} \left((f(x) - a)^r - \left(b - f(x)\right)^r\right) \mathbf{1}_{\{f < b\}} \mathrm{d}x \\
&\geq \int_{\xi-\varepsilon}^{\xi} (b - a)^r \mathbf{1}_{\{f \geq b\}} \mathrm{d}x + \int_{\xi-\varepsilon}^{\xi} (2f(x) - a - b)^r \mathbf{1}_{\{f < b\}} \mathrm{d}x \\
&\geq \varepsilon \min\{b - a, 2(f(\xi - \varepsilon) - c)\}^r > 0,
\end{aligned}$$

i.e., $\psi(\xi) > \psi\left(f^{-1}(c)\right)$. Similarly, $\psi(\xi) > \psi\left(f^{-1}(c)\right)$ whenever $\xi < \inf\{f \geq c\}$. Therefore ψ attains its minimal value on the interval $\left[\inf\{f \geq c\}, f^{-1}(c)\right] = Q^f_c$, and the proof will be complete once it is shown that in fact ψ is constant on Q^f_c. If Q^f_c is a singleton, then, trivially, this is the case. On the other hand, if $\xi, \eta \in \overset{\circ}{Q^f_c}$ with $\xi < \eta$, then $f([\xi, \eta]) = \{c\}$, by Proposition 3.1, and

$$\psi(\eta)^r - \psi(\xi)^r = \int_{\xi}^{\eta} (|f(x) - b|^r - |f(x) - a|^r)\,\mathrm{d}x = \int_{\xi}^{\eta} (|c - b|^r - |c - a|^r)\,\mathrm{d}x = 0.$$

Thus ψ is constant on Q^f_c, as claimed. $\square$

Remark 4.2. The monotonicity of f is essential in Lemma 4.1. To see this, take for instance $I = [0, 5]$ and the (non-monotone) function $f = 16 \cdot \mathbf{1}_{[0,1[} + 8 \cdot \mathbf{1}_{[1,2[} + 18 \cdot \mathbf{1}_{[2,3[} + 9 \cdot \mathbf{1}_{[3,5]}$.

For $a = 0,\ b = 24$, it is straightforward to verify that $\left\| f - 24 \cdot \mathbf{1}_{[\xi,5]} \right\|_r$ is minimal precisely for $\xi \in \{0, 2, 5\}$ if $r = 1$ or $r = 2$, for $\xi = 5$ if $1 < r < 2$, and for $\xi \in \{0, 2\}$ if $r > 2$. In general, therefore, the set of minimizers ξ is not an interval and may depend on r.

The remainder of this section deals with a problem that is dual to the one addressed by Lemma 4.1, namely the best approximation of an L^r-function f by a step function with prescribed locations but variable jumps. By considering intervals of constancy individually, clearly it is enough to consider the approximation of f by a *constant* function. Remember that the closed, non-degenerate interval $I \subset \mathbb{R}$ is assumed to be bounded throughout.

Theorem 4.3. *Assume that $f \in L^{r_0}(I)$ for some $r_0 \geq 1$. Then for every $1 \leq r \leq r_0$, there exists $\tau_r^f \in \mathbb{R}$ such that*

$$\left\| f - \tau_r^f \right\|_r \leq \| f - t \|_r, \quad \forall t \in \mathbb{R}.$$

Moreover, the following hold:

(i) $\tau_r^f \in [\operatorname{essinf}_I f, \operatorname{esssup}_I f]$*;*
(ii) $\| f - t \|_1 = \left\| f - \tau_1^f \right\|_1$ *if and only if* $t \in B^f$*;*
(iii) *For* $1 < r \leq r_0$*, the number* τ_r^f *is unique, and* $r \mapsto \tau_r^f$ *is continuous.*

Proof. Given $f \in L^{r_0}(I)$, recall that $f \in L^r(I)$ for every $1 \leq r \leq r_0$, since I is bounded. Hence the auxiliary function ϕ_r given by

$$\phi_r(t) = \lambda(I)^{-1/r} \| f - t \|_r, \quad \forall t \in \mathbb{R}, \tag{4.1}$$

is well defined and real-valued. Note that $\lim_{|t| \to +\infty} \phi_r(t) = +\infty$. Since ϕ_r is convex, there exists $\tau_r^f \in \mathbb{R}$ such that $\phi_r(\tau_r^f) \leq \phi_r(t)$ for all $t \in \mathbb{R}$.

It remains to prove assertions (i)–(iii). To establish (i), let $b = \operatorname{esssup}_I f$ for convenience, and observe that, for all $t > b$,

$$\begin{aligned} \lambda(I)(\phi_r(t)^r - \phi_r(b)^r) = \int_I \Big(\big(t - f(x)\big)^r - \big(b - f(x)\big)^r \Big) \mathrm{d}x &\geq \int_I (t - b)^r \mathrm{d}x \\ &= \lambda(I)(t - b)^r > 0, \end{aligned}$$

hence $\phi_r(t) > \phi_r(b)$. Similarly, $\phi_r(t) > \phi_r(\operatorname{essinf}_I f)$ whenever $t < \operatorname{essinf}_I f$. This shows that $\tau_r^f \in [\operatorname{essinf}_I f, \operatorname{esssup}_I f]$.

To prove (ii), given $t > \max B^f$, pick any u with $\max B^f < u < t$. Then,

$$\begin{aligned} \lambda(I)\big(\phi_1(t) - \phi_1(u)\big) &= \int_I (|f(x) - t| - |f(x) - u|)\, \mathrm{d}x \\ &\geq \int_{\{f < u\}} \Big(t - f(x) - \big(u - f(x)\big)\Big)\, \mathrm{d}x \\ &\quad + \int_{\{f \geq u\}} \big(f(x) - t - (f(x) - u)\big)\, \mathrm{d}x \\ &= (t - u)\big(\lambda(\{f < u\}) - \lambda(\{f \geq u\})\big) > 0, \end{aligned}$$

by Lemma 3.6(i), and so $\tau_1^f \le \max B^f$. Similarly, $\tau_1^f \ge \min B^f$. On the other hand, if $t, u \in B^f$, then

$$\begin{aligned}\lambda(I)\big(\phi_1(u)-\phi_1(t)\big) &= \int_{\{f\le \min B^f\}} \Big(u - f(x) - \big(t - f(x)\big)\Big)\mathrm{d}x \\ &\quad + \int_{f^{-1}(\overset{\circ}{B^f})} (|f(x)-u| - |f(x)-t|)\mathrm{d}x \\ &\quad + \int_{\{f\ge \max B^f\}} \big(f(x) - u - (f(x)-t)\big)\mathrm{d}x \\ &= (u-t)\big(\lambda(\{f\le \min B^f\}) - \lambda(\{f \ge \max B^f\})\big) = 0\,,\end{aligned}$$

by Lemma 3.6(ii) and (iii). Thus $\phi_1(t)$ is minimal if and only if $t \in B^f$.

Regarding (iii), we claim that the number τ_r^f is unique for $1 < r \le r_0$. Trivially, this is true if f is essentially constant. In any other case, note that ϕ_r^r is differentiable w.r.t. t, and

$$\begin{aligned}\frac{\lambda(I)}{r}\frac{\mathrm{d}\phi_r^r(t)}{\mathrm{d}t} &= \int_I |f(x)-t|^{r-1}\mathrm{sgn}\big(t-f(x)\big)\,\mathrm{d}x \\ &= \int_{\{f<t\}}\big(t-f(x)\big)^{r-1}\mathrm{d}x - \int_{\{f>t\}}(f(x)-t)^{r-1}\mathrm{d}x \\ &= \int_{\{f\le \min B^f\}}\big(t-f(x)\big)^{r-1}\mathrm{d}x - \int_{\{f\ge \max B^f\}}(f(x)-t)^{r-1}\mathrm{d}x\end{aligned} \tag{4.2}$$

is increasing in t. Thus ϕ_r^r is strictly convex, and τ_r^f is unique.

To show that $r \mapsto \tau_r^f$ is continuous on $]1, r_0]$, pick any $1 < r \le r_0$ and any sequence (r_n) in $]1, r_0]$ with $\lim_{n\to\infty} r_n = r$. Given $\varepsilon > 0$, by the strict convexity of ϕ_r, there exists $\delta > 0$ such that $\phi_r(\tau_r^f \pm \varepsilon) > \phi_r(\tau_r^f) + 3\delta$. On the other hand, $\lim_{n\to\infty}\phi_{r_n}(t) = \phi_r(t)$ for every $t \in \mathbb{R}$, by the Dominated Convergence Theorem. Hence for all sufficiently large n,

$$\phi_{r_n}(\tau_r^f \pm \varepsilon) > \phi_r(\tau_r^f) + 2\delta \quad \text{and} \quad \phi_{r_n}(\tau_r^f) < \phi_r(\tau_r^f) + \delta\,,$$

from which it is clear that $|\tau_{r_n}^f - \tau_r^f| < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, $r \mapsto \tau_r^f$ is continuous. $\square$

For monotone functions, Theorem 4.3 takes a particularly simple form.

Corollary 4.4. *Assume that $f : I \to \overline{\mathbb{R}}$ is non-decreasing, and $f \in L^{r_0}(I)$ for some $r_0 \ge 1$. Then for every $1 \le r \le r_0$, there exists $\tau_r^f \in \mathbb{R}$ such that*

$$\left\|f - \tau_r^f\right\|_r \le \|f - t\|_r, \quad \forall\, t \in \mathbb{R}\,.$$

Moreover, the following hold:

(i) $\tau_r^f \in [f(\min I+), f(\max I-)]$.

(ii) $\|f - t\|_1 = \|f - \tau_1^f\|_1$ *if and only if* $t \in Q^{f^{-1}}_{\frac12(\min I + \max I)}$.

(iii) *For* $1 < r \le r_0$, *the number* τ_r^f *is unique and* $r \mapsto \tau_r^f$ *is continuous.*

Remark 4.5. (i) If $f \in L^2(I)$ then simply $\tau_2^f = \frac{1}{\lambda(I)}\int_I f(x)\,\mathrm{d}x$.

(ii) For $r = 1$, Corollary 4.4 immediately yields Lemma 4.1. Indeed, under the assumptions of the latter, $f^{-1}|_{[a,b]} \in L^1([a,b])$, and $\|f - (a\mathbf{1}_{[\min I,\xi[} + b\mathbf{1}_{[\xi,\max I]})\|_1$ is minimal if and only if $\|f^{-1}|_{[a,b]} - \xi\|_1$ is minimal. By Corollary 4.4, this is the case precisely if $\xi \in Q^{g^{-1}}_{\frac12(a+b)}$ with $g = f^{-1}|_{[a,b]}$, which by Proposition 3.1 is equivalent to $\xi \in Q^f_{\frac12(a+b)}$.

Given $r > 1$, the number τ_r^f depends on f in a monotone and continuous way, as the following two simple observations show.

Proposition 4.6. *Assume that $f, \; g \in L^r(I)$ for some $r > 1$, and $f \le g$. Then $\tau_r^f \le \tau_r^g$, and $\tau_r^f = \tau_r^g$ if and only if $f = g$ a.e.*

Lemma 4.7. *Assume that $f, \; f_n \in L^{r_0}(I)$ for some $r_0 > 1$ and all $n \in \mathbb{N}$. If $\lim_{n\to\infty} f_n = f$ in $L^{r_0}(I)$, then $\lim_{n\to\infty} \tau_r^{f_n} = \tau_r^f$ locally uniformly on $]1, r_0]$.*

Proof. Since $f_n \to f$ in $L^{r_0}(I)$ and I is bounded, $\sup_{n\in\mathbb{N}} \|f_n\|_{r_0} < +\infty$ and, for all $r \in]1, r_0]$ and $n \in \mathbb{N}$,

$$\begin{aligned}\left|\tau_r^{f_n}\right| &= \lambda(I)^{-1/r} \left\|\tau_r^{f_n}\right\|_r \le \lambda(I)^{-1/r}\left(\left\|f_n - \tau_r^{f_n}\right\|_r + \|f_n\|_r\right) \\ &\le 2\lambda(I)^{-1/r}\|f_n\|_r \le 2\lambda(I)^{-1/r_0}\|f_n\|_{r_0},\end{aligned}$$

by Hölder's inequality. This shows that $(\tau_r^{f_n})$ is uniformly bounded on $]1, r_0]$.

Fix any $1 < s < r_0$. To prove that $\lim_{n\to\infty} \tau_r^{f_n} = \tau_r^f$ uniformly on $[s, r_0]$, suppose by way of contradiction that there exists $\varepsilon_0 > 0$, a sequence (r_j) in $[s, r_0]$ and an increasing sequence (n_j) in $\mathbb{N}$ such that

$$\left|\tau_{r_j}^f - \tau_{r_j}^{f_{n_j}}\right| \ge \varepsilon_0, \quad \forall\ j \in \mathbb{N}\,.$$

Assume w.o.l.g. that $r_j \to r^*$ and, by the uniform boundedness of $(\tau_r^{f_n})$, $\tau_{r_j}^{f_{n_j}} \to \tau^* \in \mathbb{R}$. Since $r \mapsto \tau_r^f$ is continuous at r^*, it follows that

$$|\tau_{r^*}^f - \tau^*| \ge \varepsilon_0\,. \tag{4.3}$$

On the other hand,

$$\begin{aligned}\left\|f - \tau_{r_j}^{f_{n_j}}\right\|_{r_j} &\le \left\|f - f_{n_j}\right\|_{r_j} + \left\|f_{n_j} - \tau_{r_j}^{f_{n_j}}\right\|_{r_j} \le \left\|f - f_{n_j}\right\|_{r_j} + \left\|f_{n_j} - \tau_{r_j}^f\right\|_{r_j} \\ &\le 2\left\|f - f_{n_j}\right\|_{r_j} + \left\|f - \tau_{r_j}^f\right\|_{r_j},\end{aligned}$$

and letting $j \to \infty$ yields, $\|f - \tau^*\|_{r^*} \le \left\|f - \tau_{r^*}^f\right\|_{r^*}$ since $(r, t) \mapsto \|f - t\|_r$ is continuous. By Theorem 4.3(iii), $\tau^* = \tau_{r^*}^f$, which clearly contradicts (4.3). $\square$

Remark 4.8. In Lemma 4.7, the convergence $\tau_r^{f_n} \to \tau_r^f$ in general is not uniform on $]1, r_0]$. To see this, take for example $I = [0, 2]$ and $f_n = 2\cdot\mathbf{1}_{[1+2^{-n},2]}$ for all $n \in \mathbb{N}$. With $f = 2\cdot\mathbf{1}_{[1,2]}$, clearly, $f, f_n \in L^\infty(I)$ and $\lim_{n\to\infty} f_n = f$ in $L^r(I)$ for every $r \ge 1$. Still, $\lim_{r\downarrow 1} \tau_r^{f_n} = 0$ for every n, whereas $\tau_r^f = 1$ for all $r > 1$.

Note that if $f : I \to \mathbb{R}$ is affine, i.e., $f(x) = ax + b$ for all $x \in I$ and the appropriate $a, b \in \mathbb{R}$, then $\tau_r^f = f\left(\frac{1}{2}(\min I + \max I)\right)$ for all $r > 1$. In this context, Lemma 4.7 can be given a slightly stronger, quantitative form.

Proposition 4.9. *Assume that $f : I \to \mathbb{R}$ is measurable, and let $\xi = \frac{1}{2}(\min I + \max I)$. If, for some $a, b, c \in \mathbb{R}$,*

$$|f(x) - (ax + b)| \le c|x - \xi|, \quad \forall\ x \in I\,,$$

then $f \in L^\infty(I)$, and $\left|\tau_r^f - f(\xi)\right| \le \frac{1}{2}c\lambda(I)$ for every $r > 1$.

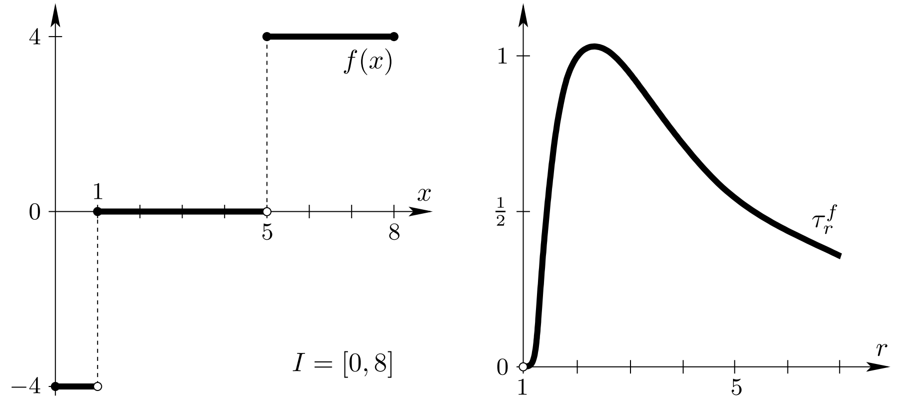

Fig. 2. For the (non-decreasing) function $f = (-4) \cdot \mathbf{1}_{[0,1[} + 4 \cdot \mathbf{1}_{[5,8]}$ the value of τ_r^f depends non-monotonically on r; see Example 4.10(i).

The remainder of this section studies how, given f, the number τ_r^f depends on r. First, this dependence is illustrated by an example, where for simplicity $f \in L^\infty(I)$ is a non-decreasing step function.

Example 4.10. Let $I = [0, 8]$.

(i) Consider the function $f = (-4) \cdot \mathbf{1}_{[0,1[} + 4 \cdot \mathbf{1}_{[5,8]}$, for which $B^f = \{0\}$, and clearly $0 \le \tau_r^f \le 4$ for every $r > 1$. By (4.2),

$$(\tau_r^f + 4)^{r-1} + 4(\tau_r^f)^{r-1} = 3(4 - \tau_r^f)^{r-1}, \tag{4.4}$$

and using (4.4), it is readily deduced that $\tau_{1+}^f := \lim_{r \downarrow 1} \tau_r^f = 0$, but also $\tau_\infty^f := \lim_{r \to +\infty} \tau_r^f = 0$. On the other hand, $\tau_2^f = 1$, and hence $r \mapsto \tau_r^f$ is non-monotone; see Fig. 2. Note that in order for $r \mapsto \tau_r^f$ to be non-monotone, a step function f has to attain at least three different values.

(ii) Consider the function $f = (-a)\mathbf{1}_{[0,1[} + (-1) \cdot \mathbf{1}_{[1,4[} + \mathbf{1}_{[4,5[} + b\mathbf{1}_{[5,8]}$ with real parameters $a, b > 1$. In this case, $B^f = [-1, 1]$, and (4.2) yields, for every $r > 1$,

$$(\tau_r^f + a)^{r-1} + 3(\tau_r^f + 1)^{r-1} = (1 - \tau_r^f)^{r-1} + 3(b - \tau_r^f)^{r-1},$$

from which it is straightforward to deduce that τ_{1+}^f exists and equals the unique real root of

$$g_{a,b}(\tau) := (3b + a + 4)\tau^3 - 3(b^2 + b - a - 1)\tau^2 + (b^3 + 3b^2 + 3a + 1)\tau - b^3 + a = 0. \tag{4.5}$$

Given $\tau \in]-1, 1[$, note that $\lim_{a \to +\infty} g_{a,b}(\tau) = +\infty$ for every $b > 1$, and $\lim_{b \to +\infty} g_{a,b}(\tau) = -\infty$ for every $a > 1$. By the Intermediate Value Theorem, there exists $\overline{a} = \overline{a}(\tau)$, $\overline{b} = \overline{b}(\tau)$ such that $g_{\overline{a},\overline{b}}(\tau) = 0$. Since the real root of (4.5) is unique, $\tau_{1+}^f = \tau$. This shows that with $a, b > 1$ chosen appropriately, τ_{1+}^f can have any value in $]-1, 1[$. Note that, similarly to (i), $\tau_\infty^f = \frac{1}{2}(b - a)$.

As seen in Example 4.10, the number τ_r^f may depend on r in a non-monotone way. In both cases considered, however, the limits $\tau_{1+}^f = \lim_{r \downarrow 1} \tau_r^f$ and $\tau_\infty^f = \lim_{r \to +\infty} \tau_r^f$ exist. Also, by modifying Example 4.10(ii) appropriately, it is clear that, given any compact interval $J \subset \mathbb{R}$

and any $\tau \in J$, one can find $f \in L^\infty(I)$ with $B^f = J$ and $\tau^f_{1+} = \tau$. In fact, one can choose f to be a non-decreasing step function.

This section concludes with a demonstration that, just as in Example 4.10, τ^f_{1+} exists always (Theorem 4.11), whereas, unlike in Example 4.10, τ^f_∞ may not exist (Example 4.15).

Theorem 4.11. *Assume that $f \in L^{r_0}(I)$ for some $r_0 > 1$. Then τ^f_{1+} exists, and $\tau^f_{1+} \in B^f$.*

Proof. We first show that

$$[\liminf_{r\downarrow 1}\tau^f_r, \limsup_{r\downarrow 1}\tau^f_r] \subset B^f\,, \tag{4.6}$$

and then that $\lim_{r\downarrow 1} \tau^f_r$ exists. For any $1 < r \le r_0$, let ϕ_r be defined as in (4.1). Recall that ϕ_r is convex, and $r \mapsto \phi_r(t)$ is continuous and non-decreasing for any $t \in \mathbb{R}$. Assume that $r_n \downarrow 1$ with $\tau^f_{r_n} \to \tau$. Then $\phi_1\left(\tau^f_{r_n}\right) \le \phi_{r_n}\left(\tau^f_{r_n}\right) \le \phi_{r_n}(t)$, and hence $\phi_1(\tau) = \lim_{n\to\infty}\phi_1\left(\tau^f_{r_n}\right) \le \lim_{n\to\infty}\phi_{r_n}(t) = \phi_1(t)$. Since $t \in \mathbb{R}$ was arbitrary, Theorem 4.3(ii) yields $\tau \in B^f$, which in turn establishes (4.6).

It remains to show that $\lim_{r\downarrow 1} \tau^f_r$ exists, which is non-trivial only if B^f is non-degenerate. In this case, define $\Psi : \overset{\circ}{B^f} \to \mathbb{R}$ as

$$\Psi(t) = \int_{\{f\le \min B^f\}} \log\bigl(t - f(x)\bigr)\,\mathrm{d}x - \int_{\{f\ge \max B^f\}} \log(f(x) - t)\,\mathrm{d}x, \quad \forall t \in \overset{\circ}{B^f}\,.$$

Note that Ψ is well-defined and continuous. Moreover, if $t, u \in B^f$ with $t < u$ then, as $B^f \ne I$,

$$\Psi(t) - \Psi(u) = \int_{\{f\le \min B^f\}} \log\frac{t - f(x)}{u - f(x)}\,\mathrm{d}x + \int_{\{f\ge \max B^f\}} \log\frac{f(x) - u}{f(x) - t}\,\mathrm{d}x < 0,$$

showing that Ψ is increasing. By (4.2), $t \mapsto \frac{\lambda(I)}{r}\frac{\mathrm{d}\phi^r_r(t)}{\mathrm{d}t}$ is a real-valued increasing function. To compare the latter to Ψ, notice the elementary inequality

$$|y^{r-1} - 1 - (r-1)\log y| \le (r-1)^2 e^{|\log y|}, \quad \forall y > 0, 1 \le r \le 2. \tag{4.7}$$

With Lemma 3.6 and (4.7), for any fixed $0 < \varepsilon < \min\{1, \frac{1}{2}\lambda(B^f)\}$, there exists $C_\varepsilon > 0$ such that

$$\left|\frac{\lambda(I)}{r}\frac{\mathrm{d}\phi^r_r(t)}{\mathrm{d}t} - (r-1)\Psi(t)\right| \le C_\varepsilon(r-1)^2, \quad \forall 1 < r \le 2,\ t \in [\min B^f + \varepsilon, \max B^f - \varepsilon]. \tag{4.8}$$

Since Ψ is increasing, three cases may be distinguished:

(i) $\Psi(\tau) = 0$ for a unique $\tau \in \overset{\circ}{B^f}$. Pick $\varepsilon > 0$ so that $\min B^f + \varepsilon < \tau < \max B^f - \varepsilon$. Then for every $\delta > 0$, (4.8) implies $\frac{\mathrm{d}\phi_r}{\mathrm{d}t}(\tau + \delta) > 0$ and $\frac{\mathrm{d}\phi_r}{\mathrm{d}t}(\tau - \delta) < 0$ for all $r > 1$ sufficiently small. It follows that $\tau^f_r \in [\tau - \delta, \tau + \delta]$ for all $r > 1$ sufficiently small, and since $\delta > 0$ was arbitrary, $\lim_{r\downarrow 1} \tau^f_r = \tau$.

(ii) $\Psi(\tau) > 0$ for all $\tau \in \overset{\circ}{B^f}$. Similarly to case (i), for every $\delta > 0$, (4.8) yields $\frac{\mathrm{d}\phi_r}{\mathrm{d}t}(\min B^f + \delta) > 0$ for all $r > 1$ sufficiently small. This implies that $\tau^f_r < \min B^f + \delta$ for all $r > 1$ sufficiently small and hence $\limsup_{r\downarrow 1} \tau^f_r \le \min B^f$. By (4.6), $\lim_{r\downarrow 1} \tau^f_r = \min B^f$.

(iii) $\Psi(\tau) < 0$ for all $\tau \in \overset{\circ}{B^f}$. This case is completely analogous to (ii), with $\lim_{r\downarrow 1}\tau^f_r = \max B^f$. $\square$

Corollary 4.12. *Assume that $f : I \to \overline{\mathbb{R}}$ is non-decreasing, and $f \in L^{r_0}(I)$ for some $r_0 > 1$. Then τ_{1+}^f exists, and $\tau_{1+}^f \in Q^{f^{-1}}_{\frac{1}{2}(\min I + \max I)}$.*

Recall that in Example 4.10 the limit τ_∞^f also exists. This is a consequence of the fact that f is bounded, together with the following simple observation.

Theorem 4.13. *Assume that $f \in \bigcap_{r\geq 1} L^r(I)$. If $f^- \in L^\infty(I)$ or $f^+ \in L^\infty(I)$, then* $\lim_{r\to+\infty} \tau_r^f = \frac{1}{2}(\operatorname{essinf}_I f + \operatorname{esssup}_I f)$.

Proof. Let f be non-constant (otherwise, $r \mapsto \tau_r^f$ is constant, too), and assume that $f^- \in L^\infty(I)$, that is, $\operatorname{essinf}_I f > -\infty$. (The case $f^+ \in L^\infty(I)$ is completely analogous.) Let $u = \frac{1}{2}(\operatorname{essinf}_I f + \operatorname{esssup}_I f)$ for convenience, fix any $\operatorname{essinf}_I f < t < u$, and let $\delta = t - \operatorname{essinf}_I f$. For $\tau < t$, note that $\tau - \operatorname{essinf}_I f < \delta$ and $\lambda(\{f \geq \tau + \delta\}) > 0$, and hence, with (4.2),

$$
\begin{aligned}
\frac{\lambda(I)}{r\delta^{r-1}} \frac{\mathrm{d}\phi_r^r(\tau)}{\mathrm{d}\tau} &= \int_{\{f<\tau\}} \left(\frac{\tau - f(x)}{\delta}\right)^{r-1} \mathrm{d}x - \int_{\{f>\tau\}} \left(\frac{f(x) - \tau}{\delta}\right)^{r-1} \mathrm{d}x \\
&\leq \lambda(I) \left(\frac{\tau - \operatorname{essinf}_I f}{\delta}\right)^{r-1} - \int_{\{f \geq \tau+\delta\}} \left(\frac{f(x) - \tau}{\delta}\right)^{r-1} \mathrm{d}x \\
&\quad - \int_{\{\tau \leq f < \tau+\delta\}} \left(\frac{f(x) - \tau}{\delta}\right)^{r-1} \mathrm{d}x \\
&\leq \lambda(I) \left(\frac{\tau - \operatorname{essinf}_I f}{\delta}\right)^{r-1} - \lambda(\{f \geq \tau + \delta\}) < 0,
\end{aligned}
$$

for all sufficiently large r. Thus $\liminf_{r\to+\infty} \tau_r^f \geq \tau$, and since t and $\tau < t$ were arbitrary, $\liminf_{r\to+\infty} \tau_r^f \geq u$. A similar argument shows $\limsup_{r\to+\infty} \tau_r^f \leq u$. $\square$

Corollary 4.14. *If $f \in \bigcap_{r\geq 1} L^r(I)$ is non-decreasing and either $f(\min I+) > -\infty$ or $f(\max I-) < +\infty$, then* $\lim_{r\to+\infty} \tau_r^f = \frac{1}{2}\big(f(\min I+) + f(\max I-)\big)$.

The final example shows that, unlike in Example 4.10, $\lim_{r\to+\infty} \tau_r^f$ may not exist if f is unbounded.

Example 4.15. Consider the function $f : I \to \mathbb{R}$ given by

$$
f = \sum_{n=0}^{\infty} 2n(-1)^{n-1} \mathbf{1}_{I_n} ,
$$

where $I := \overline{\bigcup_{n=0}^{\infty} I_n}$, and $I_0, \ I_1, \ \cdots$ are pairwise disjoint, contiguous half-open intervals, with I_1 to the right of I_0, and generally I_{2n+1} immediately to the right of I_{2n-1}, as well as I_{2n+2} immediately to the left of I_{2n}. (Clearly, f is non-decreasing on $\mathring{I}$.) The lengths $\lambda_n := \lambda(I_n) > 0$ will be determined by induction shortly, subject to the requirement that $\lambda_{n+1} \leq \frac{1}{2}\lambda_n$ for all $n \geq 0$. Thus I is a non-degenerate, closed interval of length $\sum_{n\geq 0} \lambda_n \leq 2\lambda_0$, and $f \in \bigcap_{r\geq 1} L^r(I)$ but clearly $f \notin L^\infty(I)$. For each $N \in \mathbb{N}$, let $f_N = \sum_{n=0}^{N} 2n(-1)^{n-1}\mathbf{1}_{I_n}$ and note that $\lim_{r\to+\infty} \tau_r^{f_N} = (-1)^{N-1}$, by Theorem 4.13. Moreover,

$$
\|f_{N+1} - f_N\|_r = 2(N+1)\lambda_{N+1}^{1/r}, \quad \forall r > 1, \ N \in \mathbb{N}. \tag{4.9}
$$

Let $\lambda_0 = 1$, $r_0 = 1$, and assume that $\lambda_1, \lambda_2, \dots, \lambda_N$ with $\lambda_n \le \frac{1}{2}\lambda_{n-1}$ as well as $r_0 < r_1 < \cdots < r_N$ with $r_n \ge \max\{r_{n-1}, n+1\}$ for $n = 1, \dots, N$ have been chosen in such a way that, for every $1 \le n \le N$,

$$|\tau_{r_j}^{f_n} - (-1)^{j-1}| < 2^{1-j} - 2^{-n}, \quad \forall 1 \le j \le n\,. \tag{4.10}$$

For $N = 1$, clearly such a choice is possible. By Lemma 4.7 and (4.9), choosing $\lambda_{N+1} \le \frac{1}{2}\lambda_N$ sufficiently small guarantees that

$$|\tau_r^{f_{N+1}} - \tau_r^{f_N}| < 2^{-(N+1)}, \quad \forall r \in [r_1, r_N]\,,$$

and consequently

$$\begin{aligned} |\tau_{r_j}^{f_{N+1}} - (-1)^{j-1}| &\le |\tau_{r_j}^{f_{N+1}} - \tau_{r_j}^{f_N}| + |\tau_{r_j}^{f_N} - (-1)^{j-1}| \\ &< 2^{-(N+1)} + 2^{1-j} - 2^{-N} = 2^{1-j} - 2^{-(N+1)}, \quad \forall j = 1, \dots, N. \end{aligned}$$

Also, choose $r_{N+1} \ge \max\{r_N, N+2\}$ such that $|\tau_{r_{N+1}}^{f_{N+1}} - (-1)^N| < 2^{-(N+1)}$. Thus (4.10) holds for $n = N+1$, and in fact for all $n \in \mathbb{N}$, by induction. Furthermore, note that, given any $r > 1$,

$$\|f_N - f\|_r = \left(\sum_{n>N} (2n)^r \lambda_n\right)^{1/r} \le 2\lambda_0^{1/r} \left(\sum_{n>N} n^r 2^{-n}\right)^{1/r} \to 0 \quad \text{as } N \to \infty\,,$$

and so in particular $\lim_{N\to\infty} \|f_N - f\|_{r_j} = 0$ for every $j \in \mathbb{N}$. By Lemma 4.7, $|\tau_{r_j}^{f_N} - \tau_{r_j}^{f}| < 2^{-j}$ for all sufficiently large N, which, together with (4.10), yields $|\tau_{r_j}^{f} - (-1)^{j-1}| < 3 \cdot 2^{-j}$. Since $j \in \mathbb{N}$ was arbitrary and $r_j \uparrow +\infty$, this shows that $\liminf_{r\to+\infty} \tau_r^f \le -1$ and $\limsup_{r\to+\infty} \tau_r^f \ge 1$. On the other hand, using (4.2), it is readily confirmed that $t \frac{\mathrm{d}}{\mathrm{d}t} \|f - t\|_r^r > 0$ for $t = \pm 1$ and all $r > 1$, and consequently $|\tau_r^f| < 1$. Thus $\liminf_{r\to+\infty} \tau_r^f = -1$ and $\limsup_{r\to+\infty} \tau_r^f = 1$.

By modifying Example 4.15 appropriately, it is straightforward to establish

Proposition 4.16. *Given any (bounded) interval $I \subset \mathbb{R}$ and numbers $-\infty \le a \le b \le +\infty$, there exists a non-decreasing function $f \in \bigcap_{r\ge 1} L^r(I)$ such that $\liminf_{r\to+\infty} \tau_r^f = a$ and $\limsup_{r\to+\infty} \tau_r^f = b$.*

5. Best constrained approximations

In this section, we apply results established in previous sections, notably Lemma 4.1 and Theorem 4.3, to investigate best constrained approximations of $\mu \in \mathcal{P}_r$, i.e., approximations of μ by finitely supported probabilities for which either locations (Section 5.1) or weights (Section 5.2) are prescribed. We establish existence of best constrained approximations and study their behaviour as the number of atoms goes to infinity. Finally, in Section 5.3 we relate these results to the classical theory of best (unconstrained) approximations. The main results of this section are Theorems 5.1, 5.5, 5.15, 5.20, 5.21 and 5.33.

First, we fix a few notations specific to this section. Given $n \in \mathbb{N}$, let $\Xi_n = \{\mathbf{x} \in \mathbb{R}^n : x_1 \le \cdots \le x_n\}$ and $\Pi_n = \{\mathbf{p} \in \mathbb{R}^n : p_i \ge 0,\ \sum_{i=1}^n p_i = 1\}$. For any $\mathbf{x} \in \Xi_n$, the conventions $x_0 = -\infty$ and $x_{n+1} = +\infty$ are adopted, and for any $\mathbf{p} \in \Pi_n$, let $P_i = \sum_{j=1}^i p_j$, $i = 0, 1, \dots, n$; note that $P_0 = 0$ and $P_n = 1$. Given $\mathbf{x} \in \Xi_n$ and $\mathbf{p} \in \Pi_n$, let $\delta_{\mathbf{x}}^{\mathbf{p}} = \sum_{i=1}^n p_i \delta_{x_i}$. Throughout, usage of the symbol $\delta_{\mathbf{x}}^{\mathbf{p}}$ tacitly assumes that $\mathbf{x} \in \Xi_n$, $\mathbf{p} \in \Pi_n$, with $n \in \mathbb{N}$ either specified explicitly or else clear from the context.

5.1. Best approximations with prescribed locations

Let $\mu \in \mathcal{P}_r$ for some $r \geq 1$, and $n \in \mathbb{N}$. Given $\mathbf{x} \in \Xi_n$, call $\delta_{\mathbf{x}}^{\mathbf{p}}$ with $\mathbf{p} \in \Pi_n$ a *best r-approximation of μ, given* $\mathbf{x}$ if

$$d_r(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu) \leq d_r(\delta_{\mathbf{x}}^{\mathbf{q}}, \mu), \quad \forall \mathbf{q} \in \Pi_n .$$

Denote by $\delta_{\mathbf{x}}^{\bullet}$ any (possibly not unique) best r-approximation of μ, given $\mathbf{x}$. (Note that $\delta_{\mathbf{x}}^{\bullet}$ also depends on r. In the interest of readability, this dependence is made explicit by a subscript only when necessary to avoid ambiguities.)

The existence of best r-approximations with prescribed locations can be established using the results of Sections 3 and 4.

Theorem 5.1. *Assume that $\mu \in \mathcal{P}_r$ for some $r \geq 1$, and $n \in \mathbb{N}$. For every $\mathbf{x} \in \Xi_n$, there exists a best r-approximation of μ, given $\mathbf{x}$. Moreover, $d_r\left(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu\right) = d_r\left(\delta_{\mathbf{x}}^{\bullet}, \mu\right)$ with $\mathbf{p} \in \Pi_n$ if and only if, for every $i = 1, \ldots, n$,*

$$x_i < x_{i+1} \text{ implies } P_i \in Q_{\frac{1}{2}(x_i + x_{i+1})}^{F_\mu^{-1}} . \tag{5.1}$$

Proof. For convenience, let $A_i = Q_{\frac{1}{2}(x_i + x_{i+1})}^{F_\mu^{-1}}$ for $0 \leq i \leq n$; note that $A_0 = [-\infty, 0]$, $A_n = [1, +\infty]$, and every A_i is a compact (possibly one-point) interval, by Proposition 3.1. Since the theorem trivially is correct for $n = 1$, henceforth assume $n \geq 2$. We first establish (5.1), as the asserted existence of best r-approximations will follow directly from it.

Labelling $\mathbf{x} \in \Xi_n$ as

$$x_{i_0+1} = \cdots = x_{i_1} < x_{i_1+1} = \cdots = x_{i_2} < x_{i_2+1} = \cdots < \cdots < x_{i_{m-1}+1} = \cdots = x_{i_m} \tag{5.2}$$

with integers $j \leq i_j \leq n$ for $1 \leq j \leq m \leq n$, and $i_0 = 0$, $i_m = n$, note first that $d_r(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu) = d_r(\delta_{\overline{\mathbf{x}}}^{\overline{\mathbf{p}}}, \mu)$, where $\overline{\mathbf{x}} \in \Xi_m$ and $\overline{\mathbf{p}} \in \Pi_m$, with $\overline{x}_j = x_{i_j}$, and $\overline{P}_j = P_{i_j}$ for $1 \leq j \leq m$. Moreover, (5.1) reduces to $\overline{P}_j \in Q_{\frac{1}{2}(\overline{x}_j + \overline{x}_{j+1})}^{F_\mu^{-1}}$ for all $1 \leq j \leq m-1$. To establish (5.1), therefore, it can be assumed w.o.l.g. that $x_i < x_{i+1}$ for all i.

To prove that (5.1) is necessary, let $\delta_{\mathbf{x}}^{\mathbf{p}}$ be a best r-approximation of μ, given x. Given any $1 \leq i \leq n-1$, let $\widetilde{\mathbf{p}} \in \Pi_n$ satisfy $\widetilde{p}_j = p_j$ for all $j \neq i, i+1$, and $0 \leq \widetilde{p}_i \leq p_i + p_{i+1}$. Note that $P_{i-1} \leq \widetilde{P}_i \leq P_{i+1}$.

If $P_{i-1} < P_{i+1}$, then $d_r\left(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu\right) \leq d_r\left(\delta_{\mathbf{x}}^{\widetilde{\mathbf{p}}}, \mu\right)$ implies

$$\| f_i - (x_i \mathbf{1}_{[P_{i-1}, P_i[} + x_{i+1} \mathbf{1}_{[P_i, P_{i+1}]}) \|_r \leq \| f_i - (x_i \mathbf{1}_{[P_{i-1}, \widetilde{P}_i[} + x_{i+1} \mathbf{1}_{[\widetilde{P}_i, P_{i+1}]}) \|_r ,$$

with $f_i = F_\mu^{-1}|_{[P_{i-1}, P_{i+1}]}$. Since $\widetilde{P}_i \in [P_{i-1}, P_{i+1}]$ was arbitrary, Lemma 4.1 and Proposition 3.1 yield $P_i \in Q_{\frac{1}{2}(x_i + x_{i+1})}^{f_i} = A_i$.

If $P_{i-1} = P_{i+1}$, let i^- and i^+ be the minimum and maximum, respectively, of the (non-empty) set $\{0 \leq j \leq n : \ P_j = P_i\}$. Clearly, $0 \leq i^- \leq i-1$, $i+1 \leq i^+ \leq n$, and $i^+ - i^- \geq 2$. Assume first that $i^- = 0$, in which case $i^+ \leq n-1$ and $P_i = P_{i^+} = 0$. Lemma 4.1, applied to f_{i^+} yields $0 \in A_{i^+}$. Recall that $A_i \subset \mathbb{I}$ and $\max A_i \leq \min A_{i^+}$, by Proposition 3.1. Thus $0 \leq \min A_i \leq \min A_{i^+} \leq 0$, and hence $0 = P_i \in A_i$. By a completely analogous argument, the case of $i^+ = n$, where $i^- \geq 1$ and $P_i = P_{i^-} = 1$, leads to $1 = P_i \in A_i$. Finally, assume that $1 \leq i^- < i^+ \leq n-1$. In this case, Lemma 4.1, applied to f_{i^-} and f_{i^+} yields $P_{i^-} \in A_{i^-}$ and $P_{i^+} \in A_{i^+}$, respectively. Thus $P_i = P_{i^-} = P_{i^+} \in A_{i^-} \cap A_{i^+}$. Since $j \mapsto \frac{1}{2}(x_j + x_{j+1})$ is increasing, Proposition 3.1 implies that $A_i = \{P_i\}$, and hence trivially $P_i \in A_i$. This completes the proof that (5.1) holds whenever $d_r(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu)$ is minimal, i.e., (5.1) is necessary.

To see that (5.1) also is sufficient, let $\mathbf{p} \in \Pi_n$ satisfy (5.1) and consider $\widetilde{\mathbf{p}} \in \Pi_n$ with $\widetilde{P}_i = \max A_i$ for all i. As $d_r(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu)$ then has the same value for every $\mathbf{p}$ satisfying (5.1), clearly, it is enough to show that $d_r(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu) = d_r(\delta_{\mathbf{x}}^{\widetilde{\mathbf{p}}}, \mu)$. To see the latter, note that by Proposition 3.1(i), $P_i \le \widetilde{P}_i \le P_{i+1}$, for all $1 \le i \le n-1$, and $|x_i - F_\mu^{-1}(t)| = |x_{i+1} - F_\mu^{-1}(t)|$ for all $P_i < t < \widetilde{P}_i$. Consequently,

$$
\begin{aligned}
d_r\left(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu\right)^r &= \sum_{i=1}^n \int_{P_{i-1}}^{P_i} \left|x_i - F_\mu^{-1}(t)\right|^r \mathrm{d}t \\
&= \sum_{i=1}^n \left(\int_{P_{i-1}}^{\widetilde{P}_{i-1}} \left|x_i - F_\mu^{-1}(t)\right|^r \mathrm{d}t + \int_{\widetilde{P}_{i-1}}^{P_i} \left|x_i - F_\mu^{-1}(t)\right|^r \mathrm{d}t \right) \\
&= \sum_{i=1}^n \left(\int_{P_{i-1}}^{\widetilde{P}_{i-1}} \left|x_{i-1} - F_\mu^{-1}(t)\right|^r \mathrm{d}t + \int_{\widetilde{P}_{i-1}}^{P_i} \left|x_i - F_\mu^{-1}(t)\right|^r \mathrm{d}t \right) \\
&= \sum_{i=1}^n \left(\int_{P_i}^{\widetilde{P}_i} \left|x_i - F_\mu^{-1}(t)\right|^r \mathrm{d}t + \int_{\widetilde{P}_{i-1}}^{P_i} \left|x_i - F_\mu^{-1}(t)\right|^r \mathrm{d}t \right) \\
&= \sum_{i=1}^n \int_{\widetilde{P}_{i-1}}^{\widetilde{P}_i} \left|x_i - F_\mu^{-1}(t)\right|^r \mathrm{d}t = d_r(\delta_{\mathbf{x}}^{\widetilde{\mathbf{p}}}, \mu)^r .
\end{aligned}
$$

As indicated earlier, the asserted existence of a best r-approximation of μ, given x, is a direct consequence of (5.1). Indeed, when $x \in \Xi_n$ is written as in (5.2), Proposition 3.1(i) guarantees that the m intervals $A_{i_1-1}, A_{i_2-1}, \dots, A_{i_m-1} \subset \mathbb{I}$ are arranged in such a way that $t \le u$ for all $t \in A_{i_j-1}$ and $u \in A_{i_{j+1}-1}$. It is possible, therefore, to choose $\mathbf{p} \in \Pi_n$ satisfying (5.1). $\square$

Given $\mu \in \mathcal{P}_r$ and $\mathbf{x}_n \in \Xi_n$ for all n, it is natural to ask whether $d_r(\delta_{\mathbf{x}_n}^{\bullet}, \mu) \to 0$ as $n \to \infty$. The following example illustrates that this may or may not be the case.

Example 5.2. Let μ be the standard *exponential distribution* with $F_\mu(x) = 1 - e^{-x}$ for all $x \ge 0$. Note that $\mu \in \bigcap_{r \ge 1} \mathcal{P}_r$. Given $\mathbf{x}_n = (1, 2, \dots, n)/\sqrt{n} \in \Xi_n$, Theorem 5.1 yields a unique best r-approximation of μ, namely, $\delta_{\mathbf{x}_n}^{\mathbf{p}_n}$ with $P_{n,i} = F_\mu(\frac{2i+1}{2\sqrt{n}}) = 1 - e^{-(2i+1)/(2\sqrt{n})}$ for $1 \le i \le n-1$. It is readily confirmed that $\lim_{n\to\infty} \sqrt{n} d_r(\delta_{\mathbf{x}_n}^{\bullet}, \mu) = \frac{1}{2}(r+1)^{-1/r}$ for every $r \ge 1$; in particular, therefore, $\lim_{n\to\infty} d_r(\delta_{\mathbf{x}_n}^{\mathbf{p}_n}, \mu) = 0$. By contrast, consider $\mathbf{y}_n = (0, 2, \dots, 2n-2) \in \Xi_n$, for which $\lim_{n\to\infty} d_r(\delta_{\mathbf{y}_n}^{\bullet}, \mu) = d_r(\nu, \mu) > 0$ with $\nu = (1 - e^{-1})\delta_0 + 2\sinh 1 \sum_{i=1}^\infty e^{-2i}\delta_{2i}$. Note that while every point in supp $\mu = [0, +\infty]$ is the limit of an appropriate sequence (x_{n,i_n}), this clearly is not the case for $(\mathbf{y}_n)$.

As Example 5.2 suggests, a condition has to be imposed on $(\mathbf{x}_n)$, with $\mathbf{x}_n \in \Xi_n$ for all n, in order to guarantee that $\lim_{n\to\infty} d_r(\delta_{\mathbf{x}_n}^{\bullet}, \mu) = 0$.

Theorem 5.3. *Assume that $\mu \in \mathcal{P}_r$ for some $r \ge 1$, and $\mathbf{x}_n \in \Xi_n$ for every $n \in \mathbb{N}$. Then $\lim_{n\to\infty} d_r(\delta_{\mathbf{x}_n}^{\bullet}, \mu) = 0$ if and only if*

$$
\lim_{n\to\infty} \min_{1\le i\le n} |x - x_{n,i}| = 0, \quad \forall x \in \mathbb{R} \cap \operatorname{supp} \mu . \tag{5.3}
$$

In particular, (5.3) holds whenever

$$
\lim_{n\to\infty} \left(F_\mu(x_{n,1}) + \max_{1\le i\le n-1} (x_{n,i+1} - x_{n,i}) + 1 - F_\mu(x_{n,n}) \right) = 0 .
$$

Proof. For convenience, let $P_{n,i} = F_\mu\big(\frac{1}{2}(x_{n,i} + x_{n,i+1})\big)$ for all $n \in \mathbb{N}$ and $0 \le i \le n$, as well as $A = \mathbb{I} \setminus \{P_{n,i} : n \in \mathbb{N},\ 0 \le i \le n\}$ and $f_n = F_{\delta_{\mathbf{x}_n}^{\mathbf{p}_n}}$. Note that $|F_\mu^{-1}(t) - x_{n,i}| = \min_{1\le j\le n} |F_\mu^{-1}(t) - x_{n,j}|$ whenever $P_{n,i-1} < t < P_{n,i}$, and hence

$$|F_\mu^{-1}(t) - f_n^{-1}(t)| = \min_{1\le j\le n} |F_\mu^{-1}(t) - x_{n,j}|, \quad \forall t \in A. \tag{5.4}$$

We first show that (5.3) is necessary. To see this, assume that (5.3) fails. Then, with the appropriate $\varepsilon > 0$, $x \in \operatorname{supp} \mu$ and sequence (n_k),

$$\min_{1\le i\le n_k} |x - x_{n_k,i}| \ge 2\varepsilon, \quad \forall k \in \mathbb{N}.$$

Since f_n is constant on $[x - \varepsilon, x + \varepsilon]$ whereas F_μ is not,

$$d_1(\delta^\bullet_{\mathbf{x}_{n_k}}, \mu) = d_1(\delta^{\mathbf{p}_{n_k}}_{\mathbf{x}_{n_k}}, \mu) \ge \min_{c\in\mathbb{R}} \int_{x-\varepsilon}^{x+\varepsilon} |F_\mu(y) - c|\,\mathrm{d}y > 0, \quad \forall k \in \mathbb{N},$$

and so $\limsup_{n\to\infty} d_r(\delta^\bullet_{\mathbf{x}_n}, \mu) > 0$ as well.

To see that (5.3) also is sufficient, note first that if F_μ^{-1} is continuous at $t \in A$, then $F_\mu^{-1}(t) \in \operatorname{supp} \mu$, and hence $f_n^{-1}(t) \to F_\mu^{-1}(t)$, by (5.4). Since F_μ^{-1} is monotone, $f_n^{-1} \to F_\mu^{-1}$ a.e. on $\mathbb{I}$. If supp μ is bounded then $f_n^{-1} \to F_\mu^{-1}$ in $L^r(\mathbb{I})$, by the Dominated Convergence Theorem, i.e., $\lim_{n\to\infty} d_r\big(\delta^{\mathbf{p}_n}_{\mathbf{x}_n}, \mu\big) = 0$, and thus $\lim_{n\to\infty} d_r\big(\delta^\bullet_{\mathbf{x}_n}, \mu\big) = 0$. If, on the other hand, supp μ is unbounded, then, given any $\varepsilon > 0$, choose $\nu \in \mathcal{P}$ with bounded support and $d_r(\mu, \nu) < \varepsilon$. Then $d_r(\delta^\bullet_{\mathbf{x}_n}, \mu) \le d_r(\widetilde{\delta}^\bullet_{\mathbf{x}_n}, \mu) \le d_r(\widetilde{\delta}^\bullet_{\mathbf{x}_n}, \nu) + d_r(\nu, \mu)$, where $\widetilde{\delta}^\bullet_{\mathbf{x}_n}$ denotes a best r-approximation of ν, given $\mathbf{x}_n$. By the above, $\limsup_{n\to\infty} d_r(\delta^\bullet_{\mathbf{x}_n}, \mu) \le \varepsilon$, and since $\varepsilon > 0$ was arbitrary, $\lim_{n\to\infty} d_r(\delta^\bullet_{\mathbf{x}_n}, \mu) = 0$. $\square$

Example 5.4. Let μ be the Beta(2, 1) distribution, i.e., $F_\mu(x) = x^2$ for all $x \in \mathbb{I}$, and consider $\mathbf{x}_n = (1, \sqrt{2}, \dots, \sqrt{n})/\sqrt{n} \in \Xi_n$. By Theorem 5.3, $\lim_{n\to\infty} d_r(\delta^\bullet_{\mathbf{x}_n}, \mu) = 0$ for every $r \ge 1$. Unlike in Example 5.2, however, the rate of convergence depends on r: With $\alpha_r = \frac{1}{2} + \frac{1}{\max\{2,r\}}$ and the appropriate $0 < C_r < +\infty$,

$$\lim_{n\to\infty} n^{\alpha_r} d_r(\delta^\bullet_{\mathbf{x}_n}, \mu) = C_r$$

whenever $r \ne 2$, whereas

$$\lim_{n\to\infty} \frac{n}{\sqrt{\log n}} d_2(\delta^\bullet_{\mathbf{x}_n}, \mu) = \frac{1}{4\sqrt{3}}.$$

Thus $\big(d_r(\delta^\bullet_{\mathbf{x}_n}, \mu)\big)$ decays like $(n^{-\alpha_r})$ and $(n^{-1}\sqrt{\log n})$ if $r \ne 2$ and $r = 2$, respectively.

5.2. *Best approximations with prescribed weights*

Let $\mu \in \mathcal{P}_r$ for some $r \ge 1$, and $n \in \mathbb{N}$. Given $\mathbf{p} \in \Pi_n$, call $\delta^{\mathbf{p}}_{\mathbf{x}}$ with $\mathbf{x} \in \Xi_n$ a *best r-approximation of μ, given* $\mathbf{p}$ if

$$d_r(\delta^{\mathbf{p}}_{\mathbf{x}}, \mu) \le d_r(\delta^{\mathbf{p}}_{\mathbf{y}}, \mu), \quad \forall \mathbf{y} \in \Xi_n.$$

Denote by $\delta^{\mathbf{p}}_\bullet$ any best r-approximation of μ, given $\mathbf{p}$. (Again in the interest of readability, the r-dependence of $\delta^{\mathbf{p}}_\bullet$ is made explicit by a subscript only when necessary to avoid ambiguity.) An important special case of $\mathbf{p} \in \Pi_n$ is the uniform probability vector $\mathbf{u}_n = (1, \dots, 1)/n$. Best r-approximations of μ, given $\mathbf{u}_n$, will be referred to as best *uniform* r-approximations, and denoted $\delta^{\mathbf{u}_n}_\bullet$. As in the case of prescribed locations studied in Section 5.1, the existence of best

r-approximations with prescribed weights follows from results in Sections 3 and 4. Due to the nature of (2.2), the proof of the following theorem even is simpler than that of its counterpart, Theorem 5.1.

Theorem 5.5. *Assume that $\mu \in \mathcal{P}_r$ for some $r \geq 1$, and $n \in \mathbb{N}$. For every $\mathbf{p} \in \Pi_n$, there exists a best r-approximation of μ, given $\mathbf{p}$. Moreover, $d_1\left(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu\right) = d_1\left(\delta_{\bullet}^{\mathbf{p}}, \mu\right)$ if and only if, for every $i = 1, \dots, n$,*

$$P_{i-1} < P_i \text{ implies } x_i \in Q^{F_\mu}_{\frac{1}{2}(P_{i-1}+P_i)}, \tag{5.5}$$

and for $r > 1$, $d_r\left(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu\right) = d_r\left(\delta_{\bullet}^{\mathbf{p}}, \mu\right)$ if and only if, for every $i = 1, \dots, n$,

$$P_{i-1} < P_i \text{ implies } x_i = \tau_r^{f_i}, \text{ where } f_i = F_\mu^{-1}\big|_{[P_{i-1}, P_i]}. \tag{5.6}$$

Proof. As in the proof of Theorem 5.1, existence follows immediately, once (5.5) and (5.6) are established. Labelling $\mathbf{P}$ as

$$P_{i_0} = \dots = P_{i_1 - 1} < P_{i_1} = \dots = P_{i_2 - 1} < P_{i_2} = \dots < \dots < P_{i_{m-1}} = \dots = P_{i_m - 1}$$

with integers $j \leq i_j \leq n+1$ for $1 \leq j \leq m \leq n$, and $i_0 = 0$, $i_m = n+1$, note that $d_r\left(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu\right) = d_r\left(\delta_{\overline{\mathbf{x}}}^{\overline{\mathbf{p}}}, \mu\right)$, where $\overline{\mathbf{x}} \in \Xi_m$ and $\overline{\mathbf{p}} \in \Pi_m$, with $\overline{x}_j = x_{i_j}$, and $\overline{P}_j = P_{i_j}$ for $1 \leq j \leq m$. Moreover, (5.5) reduces to $\overline{x}_j \in Q^{F_\mu}_{\frac{1}{2}(\overline{P}_{j-1}+\overline{P}_j)}$ for all $1 \leq j \leq m$, whereas (5.6) reduces to $\overline{x}_j = \tau_r^{f_j}$ with $f_j = F_\mu^{-1}\big|_{[\overline{P}_{j-1}, \overline{P}_j]}$. Thus, to establish (5.5) and (5.6), it can be assumed w.o.l.g. that $P_{i-1} < P_i$ for all i.

Given $\mathbf{p} \in \Pi_n$, it is clear from $d_r(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu)^r = \sum_{i=1}^n \|x_i - f_i\|_r^r$ that $d_r(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu)$ is minimal if and only if $\|x_i - f_i\|_r$ is minimal for every i. By Corollary 4.4, the latter is the case precisely if $x_i \in Q^{f_i^{-1}}_{\frac{1}{2}(P_{i-1}+P_i)} = Q^{F_\mu}_{\frac{1}{2}(P_{i-1}+P_i)}$ for $r = 1$, and if $x_i = \tau_r^{f_i}$ for $r > 1$. $\square$

Remark 5.6. (i) For $r = 1$ and $\mathbf{p} = \mathbf{u}_n$, Theorem 5.5 reduces to [2, Theorem 2.8]. In particular, $\frac{1}{n}\sum_{i=1}^n \delta_{F_\mu^{-1}\left(\frac{2i-1}{2n}\right)}$ is a best uniform 1-approximation of $\mu \in \mathcal{P}_1$. For $n = 1$, (5.5) yields the well-known fact that $d_1(\delta_a, \mu)$ is minimal if and only if $a \in \mathbb{R}$ is a *median* of μ.

(ii) For $r = 2$, if $\mu \in \mathcal{P}_2$ and $\mathbf{p} \in \Pi_n$ with $p_i > 0$ for all i, then by Remark 4.5(i), the unique best 2-approximation of μ, given $\mathbf{p}$, is $\delta_{\mathbf{x}}^{\mathbf{p}}$ with $x_i = p_i^{-1}\int_{P_{i-1}}^{P_i} F_\mu^{-1}(t)\,\mathrm{d}t$. In particular, $d_2(\delta_a, \mu)$ is minimal precisely for $a = \int_0^1 F_\mu^{-1}(t)\,\mathrm{d}t$.

Example 5.7. Given $\mu \in \mathcal{P}_r$ and $\mathbf{p} \in \Pi_n$, Theorem 5.5 can also be utilized to minimize $d_r(\sum_{i=1}^n p_i \delta_{x_i}, \mu)$ where $\mathbf{x} \in \mathbb{R}^n$ but not necessarily $\mathbf{x} \in \Xi_n$. For instance, with $\mu = \text{Beta}(2, 1)$ as in Example 5.4 and $\mathbf{p} = (2/3, 1/3)$ as well as $\mathbf{q} = (1/3, 2/3)$, for $r = 1$,

$$\delta_{\bullet}^{\mathbf{p}} = \tfrac{2}{3}\delta_{1/\sqrt{3}} + \tfrac{1}{3}\delta_{\sqrt{5/6}}, \quad \delta_{\bullet}^{\mathbf{q}} = \tfrac{1}{3}\delta_{1/\sqrt{6}} + \tfrac{2}{3}\delta_{2/\sqrt{6}}.$$

Since $d_1(\delta_{\bullet}^{\mathbf{p}}, \mu) \approx 0.12154 > d_1(\delta_{\bullet}^{\mathbf{q}}, \mu) \approx 0.10677$, it follows, that $\min_{\mathbf{x} \in \mathbb{R}^2} d_1(\frac{2}{3}\delta_{x_1} + \frac{1}{3}\delta_{x_2}, \mu) = d_1(\delta_{\bullet}^{\mathbf{q}}, \mu)$. In general, this minimizing problem can be solved by applying Theorem 5.5 to $(p_{\sigma(1)}, \dots, p_{\sigma(n)}) \in \Pi_n$ for all permutations σ of $\{1, \dots, n\}$. The permutations yielding the minimal value may depend on r. Often, not all $n!$ permutations σ have to be considered. For instance, if F_μ^{-1} is concave on $]0, 1[$ as in the above example, then only the (unique) non-decreasing rearrangement of p is relevant.

Given $\mu \in \mathcal{P}_r$ and $\mathbf{p}_n \in \Pi_n$ for all n, it is natural to ask whether $d_r(\delta_\bullet^{\mathbf{p}_n}, \mu) \to 0$ as $n \to \infty$. As in the dual situation of Section 5.1, this may or may not be the case, as illustrated by the following example.

Example 5.8. Consider again the exponential distribution μ of Example 5.2. By (5.5), the unique best uniform 1-approximation of μ is $\delta_{\mathbf{x}_n}^{\mathbf{u}_n}$ with $x_{n,i} = F_\mu^{-1}(\frac{2i-1}{2n}) = \log \frac{2n}{2n-2i+1}$, for every $n \in \mathbb{N}$ and $1 \le i \le n$, and

$$nd_1(\delta_\bullet^{\mathbf{u}_n}, \mu) = -2\sum_{i=1}^{n} i \log \frac{2i-1}{2i} + \log \frac{(2n)!}{2^{2n} n! n^n} = \frac{1}{4} \log n + \mathcal{O}(1) \text{ as } n \to \infty .$$

By Remark 5.6(ii), the best uniform 2-approximation of μ is unique, namely $\delta_{\mathbf{y}_n}^{\mathbf{u}_n}$ with $y_{n,i} = n \int_{(i-1)/n}^{i/n} F_\mu^{-1}(t)\,\mathrm{d}t = \log \dfrac{en(n-i)^{n-i}}{(n-i+1)^{n-i+1}}$, and

$$\sqrt{n} d_2(\delta_\bullet^{\mathbf{u}_n}, \mu) = \sqrt{n - \sum_{i=1}^{n-1} i(i+1)(\log \frac{i}{i+1})^2} = C_2 + \mathcal{O}(n^{-1}) \quad \text{as } n \to \infty ,$$

where $C_2^2 = 1 + \sum_{i=1}^{\infty} \left(1 - i(i+1)\left(\log \frac{i}{i+1}\right)^2\right) \approx 1.0803$. In fact, it can be shown that $\lim_{n\to\infty} n^{1/r} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) = C_r$ whenever $r > 1$, with the appropriate $0 < C_r < +\infty$. Thus $d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) \to 0$ as $n \to \infty$, but the rate of convergence evidently depends on r, and is slower than (n^{-1}). By contrast, consider $\mathbf{p}_n \in \Pi_n$ with $p_{n,i} = \frac{2^{i-1}}{2^n - 1}$ for $1 \le i \le n$. Then $\lim_{n\to\infty} d_r(\delta_\bullet^{\mathbf{p}_n}, \mu) = d_r(\nu, \mu) > 0$ with $\nu = \sum_{i=1}^{\infty} 2^{-i} \delta_{a_i}$, and $a_i = F_\mu^{-1}(3 \cdot 2^{-i-1})$ if $r = 1$ and $a_i = \tau_r^{F_\mu^{-1}|_{[2^{-i}, 2^{-i+1}]}}$ if $r > 1$.

Example 5.8 suggests a simple condition that may be imposed on $(\mathbf{p}_n)$, with $\mathbf{p}_n \in \Pi_n$ for every n, in order to guarantee that $\lim_{n\to\infty} d_r(\delta_\bullet^{\mathbf{p}_n}, \mu) = 0$. The following result is a counterpart of Theorem 5.3. Due to the nature of (2.2), the proof is similar but not identical; recall that $G^{F_\mu^{-1}} \subset \mathbb{I}$ for every $\mu \in \mathcal{P}$.

Theorem 5.9. *Assume that $\mu \in \mathcal{P}_r$ for some $r \ge 1$, and $\mathbf{p}_n \in \Pi_n$ for every $n \in \mathbb{N}$. Then $\lim_{n\to\infty} d_r(\delta_\bullet^{\mathbf{p}_n}, \mu) = 0$ if and only if*

$$\lim_{n\to\infty} \min_{1\le i\le n} |t - P_{n,i}| = 0, \quad \forall t \in G^{F_\mu^{-1}} . \tag{5.7}$$

In particular, (5.7) holds whenever $\lim_{n\to\infty} \max_{1\le i\le n} p_{n,i} = 0$.

Proof. For every $n \in \mathbb{N}$, let $\delta_{\mathbf{x}_n}^{\mathbf{p}_n}$ be a best r-approximation of μ, given $\mathbf{p}_n$, and also $f_n = F_{\delta_{\mathbf{x}_n}^{\mathbf{p}_n}}$, for convenience.

To see that (5.7) is necessary, suppose that

$$\min_{1\le i\le n_k} \left|t - P_{n_k,i}\right| \ge 2\varepsilon, \quad \forall k \in \mathbb{N},$$

for some $0 < t < 1$, $0 < \varepsilon < \min\{t, 1-t\}$, and the appropriate sequence (n_k). (The other cases, $t = 0$ and $t = 1$, are analogous.) Since $f_{n_k}^{-1}$ is constant on $[t-\varepsilon, t+\varepsilon]$ whereas F_μ^{-1} is not,

$$d_r(\delta_\bullet^{\mathbf{p}_{n_k}}, \mu)^r = d_r(\delta_{\mathbf{x}_{n_k}}^{\mathbf{p}_{n_k}}, \mu)^r \ge \min_{c\in\mathbb{R}} \int_{t-\varepsilon}^{t+\varepsilon} \left|F_\mu^{-1}(u) - c\right|^r \mathrm{d}u > 0, \; k \in \mathbb{N},$$

and hence $\limsup_{n\to\infty} d_r(\delta_\bullet^{\mathbf{p}_n}, \mu) > 0$.

To show that (5.7) also is sufficient, assume that t is a continuity point of F_μ^{-1}. If $t \in G^{F_\mu^{-1}}$ then, given $\varepsilon > 0$, there exist $t_1, t_2 \in G^{F_\mu^{-1}}$ with $|F_\mu^{-1}(t_{1,2}) - F_\mu^{-1}(t)| < \varepsilon$ and either $t < t_1 < t_2$ or $t_1 < t_2 < t$. Assume w.o.l.g. that $t < t_1 < t_2$. (The other case is similar.) By (5.7), $t < P_{n,i_n} < P_{n,i_n+1} < t_2$ for all sufficiently large n and the appropriate $1 \le i_n \le n$. Since f_n^{-1} is constant on $[P_{n,i_n}, P_{n,i_n+1}]$ with a value between $F_\mu^{-1}(P_{n,i_n}) \ge F_\mu^{-1}(t)$ and $F_\mu^{-1}(P_{n,i_n+1}) \le F_\mu^{-1}(t) + \varepsilon$, clearly $f_n^{-1}(t) \to F_\mu^{-1}(t)$. If, on the other hand, $t \notin G^{F_\mu^{-1}}$, then let $]a, b[\subset \mathbb{I}$ be the largest interval that contains t but is disjoint from $G^{F_\mu^{-1}}$. Assume w.o.l.g. that $0 < a < b < 1$. (The cases $a = 0$ and $b = 1$ are analogous.) Then $a, b \in G^{F_\mu^{-1}}$. Given $\varepsilon > 0$, since $F_\mu^{-1} - F_\mu^{-1}(t) \in L^r(\mathbb{I})$, there exists $\delta > 0$ such that $\int_A |F_\mu^{-1}(u) - F_\mu^{-1}(t)|^r \mathrm{d}u < \varepsilon$ whenever $\lambda(A) < \delta$. Let $i_n = \min\{1 \le j \le n : \; P_{n,j} > t\}$. Note that $P_{n,i_n-1} \le t < P_{n,i_n}$. If $a \le P_{n,i_n-1} < P_{n,i_n} \le b$, then $f_n^{-1}(t) = F_\mu^{-1}(t)$. If $P_{n,i_n-1} < a$, then $|P_{n,i_n-1} - a| = \min_{1\le i\le n} |a - P_{n,i}|$, $\max\{b, P_{n,i_n}\} - b \le \min_{1\le i\le n} |b - P_{n,i}|$, and $(a - P_{n,i_n-1}) + \max\{b, P_{n,i_n}\} - b < \delta$ for all sufficiently large n, by (5.7). Hence

$$
\begin{aligned}
(t-a)|F_\mu^{-1}(t) - f_n^{-1}(t)|^r &\le \int_{P_{n,i_n-1}}^{P_{n,i_n}} |F_\mu^{-1}(u) - f_n^{-1}(t)|^r \mathrm{d}u \\
&\le \int_{P_{n,i_n-1}}^{P_{n,i_n}} |F_\mu^{-1}(u) - F_\mu^{-1}(t)|^r \mathrm{d}u \\
&= \int_{P_{n,i_n-1}}^{a} |F_\mu^{-1}(u) - F_\mu^{-1}(t)|^r \mathrm{d}u \\
&\quad + \int_{b}^{\max\{b, P_{n,i_n}\}} |F_\mu^{-1}(u) - F_\mu^{-1}(t)|^r \mathrm{d}u < \varepsilon.
\end{aligned}
$$

For $P_{n,i_n} > b$, an analogous argument applies. In summary, $f_n^{-1} \to F_\mu^{-1}$ a.e. on $\mathbb{I}$, and the remaining argument is identical to the one in the proof of Theorem 5.3. $\square$

Since $\delta_\bullet^{\mathbf{p}}$ is a best approximation of $\mu \in \mathcal{P}_r$ w.r.t. the metric d_r, given weights $\mathbf{p}$, it is natural to ask whether $\delta_\bullet^{\mathbf{p}}$ reflects any basic feature of μ. Most basically perhaps, how is supp $\delta_\bullet^{\mathbf{p}}$ related to supp μ ? As the following example shows, it may not be possible to guarantee supp $\delta_\bullet^{\mathbf{p}} \subset$ supp μ.

Example 5.10. Let μ be the Cantor probability measure, i.e., the $\frac{\log 2}{\log 3}$-dimensional Hausdorff measure on the classical Cantor middle third set. Using the fact that $Q_t^{F_\mu}$ is a non-degenerate interval for every dyadic rational $0 < t < 1$, it is readily seen that $\delta_\bullet^{\mathbf{u}_n}$ is not unique for any $n \in \mathbb{N}$ whenever $r = 1$. For instance, $\frac{1}{2}(\delta_{1/5} + \delta_{4/5})$ and $\frac{1}{2}(\delta_{1/9} + \delta_{8/9})$ both are best uniform 1-approximations of μ, and $\{1/5, 4/5\} \cap$ supp $\mu = \varnothing$ whereas $\{1/9, 8/9\} \subset$ supp μ. For $r > 1$, however, $\delta_\bullet^{\mathbf{u}_n}$ always is unique. In fact, $\delta_\bullet^{\mathbf{u}_{2^k}}$ even is independent of $r > 1$, due to symmetry, and supp $\delta_\bullet^{\mathbf{u}_{2^k}} \cap$ supp $\mu = \varnothing$. For example, $\delta_\bullet^{\mathbf{u}_2} = \frac{1}{2}(\delta_{1/6} + \delta_{5/6})$ for all $r > 1$, and $\{1/6, 5/6\} \cap$ supp $\mu = \varnothing$.

To formalize the observations in Example 5.10, note that if $\delta_{\mathbf{x}}^{\mathbf{p}}$ is a best 1-approximation of μ, given $\mathbf{p} \in \Pi_n$, then, by Theorem 5.5, $x_i \in Q_{\frac{1}{2}(P_{i-1}+P_i)}^{F_\mu}$ whenever $P_{i-1} < P_i$. Since the endpoints of all quantile sets $Q_t^{F_\mu}$ belong to supp μ, by Proposition 3.2, it is possible to choose $\mathbf{y} \in \Xi_n$ with $d_1(\delta_{\mathbf{y}}^{\mathbf{p}}, \mu) = d_1(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu)$ and supp $\delta_{\mathbf{y}}^{\mathbf{p}} \subset$ supp μ. Similarly, if $r > 1$, then $x_i = \tau_r^{f_i}$ with $f_i = F_\mu^{-1}\big|_{[P_{i-1}, P_i]}$, and consequently $x_i \in [F_\mu^{-1}(P_{i-1}), F_\mu^{-1}(P_i -)]$. By Corollary 4.4(i), it follows that

$$\min\ \mathrm{supp}\ \mu = F_\mu^{-1}(P_0+) \le x_i \le F_\mu^{-1}(P_n-) = \max\ \mathrm{supp}\ \mu, \quad \forall i = 1, \ldots, n.$$

This establishes

Proposition 5.11. *Assume that $\mu \in \mathcal{P}_r$ for some $r \ge 1$, and $n \in \mathbb{N}$. If $r = 1$ or $\mathrm{supp}\ \mu$ is connected, then there exists a best r-approximation $\delta_\bullet^{\mathbf{p}}$ of μ, given $\mathbf{p} \in \Pi_n$, with $\mathrm{supp}\ \delta_\bullet^{\mathbf{p}} \subset \mathrm{supp}\ \mu$.*

Among the best approximations of μ, given $\mathbf{p} \in \Pi_n$, the case of *uniform* approximations, i.e., $\mathbf{p} = \mathbf{u}_n$, arguably is the most important. In this case, Theorem 5.9 has the following corollary; see also [25, Thm.2].

Corollary 5.12. *Assume that $\mu \in \mathcal{P}_r$ for some $r \ge 1$, and $1 \le s \le r$. For every $n \in \mathbb{N}$, let $\delta_{\bullet,s}^{\mathbf{u}_n}$ be a best uniform s-approximation of μ. Then $\lim_{n\to\infty} d_r(\delta_{\bullet,s}^{\mathbf{u}_n}, \mu) = 0$. In particular, $\lim_{n\to\infty} d_r(\delta_{\mathbf{x}_n}^{\mathbf{u}_n}, \mu) = 0$ for $x_{n,i} = F_\mu^{-1}(\frac{2i-1}{2n})$, i.e., for $\delta_{\mathbf{x}_n}^{\mathbf{u}_n}$ being one best uniform 1-approximation of μ.*

Despite its simplicity, Corollary 5.12 touches upon a recurring theme and motivation of the present article, namely the surprising versatility of best uniform 1-approximations: Not only are they easy to compute (by virtue of Theorem 5.5) and hence preferable for practical computations, but they also provide reasonably good uniform d_r-approximations. In fact, beyond what Corollary 5.12 asserts, they often even capture the precise rate of convergence of $\big(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$; see, e.g., Example 5.18, and the discussion following Proposition 5.27.

Remark 5.13. For $r = s = 2$, Corollary 5.12 yields [1, Thm. 3.6]. In [1], a *convex order* on $\mathcal{P}$ is considered, shown to be preserved by best uniform 2-approximations, and applied to the numerical construction of martingales. We conjecture that best uniform r-approximations preserve this order for all $r > 1$. By contrast, best (unconstrained) 2-approximations, considered in Section 5.3, do not in general preserve the convex order; see [1, Thm.2.1].

The remainder of this subsection is devoted to a study of $d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)$ as $n \to \infty$. Since best uniform r-approximations may be hard to identify explicitly, we will also consider *asymptotically* best uniform r-approximations. Formally, $(\delta_{\mathbf{x}_n}^{\mathbf{u}_n})$ with $\mathbf{x}_n \in \Xi_n$ for all $n \in \mathbb{N}$ is a sequence of *asymptotically best uniform r-approximations* of $\mu \in \mathcal{P}_r \setminus \{\delta_{\mathbf{x}}^{\mathbf{u}_i} : i \in \mathbb{N}, \mathbf{x} \in \Xi_i\}$ if

$$\lim_{n\to\infty} \frac{d_r(\delta_{\mathbf{x}_n}^{\mathbf{u}_n}, \mu)}{d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)} = 1.$$

To illustrate a possible behaviour of $\big(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$, as well as the practical relevance of asymptotically best uniform approximations, we first consider a simple example.

Example 5.14. Let $\mu = \mathrm{Beta}(2, 1)$ as in Example 5.4. Theorem 5.5 yields a unique best uniform r-approximation of μ for every $r \ge 1$. For $r = 1$, a short calculation shows that

$$n d_1(\delta_\bullet^{\mathbf{u}_n}, \mu) = \frac{1}{4} + \mathcal{O}(n^{-1/2}) \quad \text{as } n \to \infty,$$

whereas for $r = 2$,

$$n d_2(\delta_\bullet^{\mathbf{u}_n}, \mu) = \frac{1}{4\sqrt{3}}\sqrt{\log n} + \mathcal{O}(1) \quad \text{as } n \to \infty.$$

For $1 < r < 2$, however, $\delta_\bullet^{\mathbf{u}_n}$ is not easy to calculate explicitly. This not only makes the rate of convergence of $\big(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$ hard to determine, but it also emphasizes the need for simple *asymptotically* best uniform approximations. In fact, Theorem 5.15 shows that, for every $1 \le r < 2$, $\lim_{n\to\infty} nd_r(\delta_\bullet^{\mathbf{u}_n}, \mu) = \left(\frac{2^{1-2r}}{(r+1)(2-r)}\right)^{1/r}$, and $(\delta_{\mathbf{x}_n}^{\mathbf{u}_n})$, with $x_{n,i} = \sqrt{\frac{2i-1}{2n}}$ for $1 \le i \le n$, is a sequence of asymptotically best uniform r-approximations. By contrast, it turns out that $\lim_{n\to\infty} n^{1/2+1/r} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)$ is finite and positive whenever $r > 2$.

The observations in Example 5.14 are a special case of a general principle: If the quantile function of $\mu \in \mathcal{P}_r$ is absolutely continuous (and not constant), then $\big(nd_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$ converges to a positive limit. This fact may be seen as an analogue, in the context of best uniform approximations, of a classical result regarding best approximations; cf. Proposition 5.27.

Theorem 5.15. *Assume that $\mu \in \mathcal{P}_r$ for some $r \ge 1$. If μ^{-1} is absolutely continuous (w.r.t. λ) then*

$$\lim_{n\to\infty} nd_r(\delta_\bullet^{\mathbf{u}_n}, \mu) = \frac{1}{2(r+1)^{1/r}} \left(\int_{\mathbb{I}} \left(\frac{\mathrm{d}\mu^{-1}}{\mathrm{d}\lambda} \right)^r \right)^{1/r} . \tag{5.8}$$

Moreover, if $\frac{\mathrm{d}\mu^{-1}}{\mathrm{d}\lambda} \in L^r(\mathbb{I})$ then $(\delta_{\mathbf{x}_n}^{\mathbf{u}_n})$, with $x_{n,i} = F_\mu^{-1}(\frac{2i-1}{2n})$ for $1 \le i \le n$, is a sequence of asymptotically best uniform r-approximations of μ, unless μ is degenerate, i.e., unless $\mu = \delta_a$ for some $a \in \mathbb{R}$.

Proof. For convenience, let $f = F_\mu^{-1}|_{]0,1[}$, as well as $J_{n,i} = [\frac{i-1}{n}, \frac{i}{n}]$ and $x_{n,i} = f(\frac{2i-1}{2n})$ for $n \in \mathbb{N}$ and $1 \le i \le n$. Note that the non-decreasing function f is absolutely continuous, by assumption. For the reader's convenience, the following proof is divided into four steps: First, (5.8) will be established assuming that f has a C^1-extension to $\mathbb{I}$; then (5.8) will be shown to hold in general, regardless of whether both sides are finite (Step 2) or infinite (Step 3); finally, the assertion regarding asymptotically best uniform approximations will be proved (Step 4).

Step 1. Assume f can be extended to a C^1-function on $\mathbb{I}$. Then

$$\begin{aligned} n^r d_r(\delta_{\mathbf{x}_n}^{\mathbf{u}_n}, \mu)^r &= n^r \sum_{i=1}^n \int_{J_{n,i}} |f(t) - x_{n,i}|^r \mathrm{d}t \le n^r \sum_{i=1}^n (\max_{J_{n,i}} f')^r \int_{J_{n,i}} \left| t - \frac{2i-1}{2n} \right|^r \mathrm{d}t \\ &= \frac{1}{2^r(r+1)} \cdot \frac{1}{n} \sum_{i=1}^n (\max_{J_{n,i}} f')^r . \end{aligned}$$

Since $(f')^r$ is Riemann integrable, $\lambda(J_{n,i}) = 1/n$, and similarly

$$n^r d_r(\delta_{\mathbf{x}_n}^{\mathbf{u}_n}, \mu)^r \ge \frac{1}{2^r(r+1)} \cdot \frac{1}{n} \sum_{i=1}^n (\min_{J_{n,i}} f')^r ,$$

it follows that $\lim_{n\to\infty} nd_r(\delta_\bullet^{\mathbf{u}_n}, \mu) = \frac{1}{2(r+1)^{1/r}} (\int_{\mathbb{I}} f'(t)^r \mathrm{d}t)^{1/r} < +\infty$. Moreover, f' is uniformly continuous, hence given $\varepsilon > 0$, there exists $N \in \mathbb{N}$ such that

$$|f'(t) - f'(u)| \le \varepsilon, \quad \forall t, u \in \mathbb{I}, \ |t - u| < \frac{1}{N}.$$

Whenever $n \ge N$, therefore, the Mean Value Theorem yields

$$\left| f(t) - x_{n,i} - f'\left(\frac{2i-1}{2n}\right)\left(t - \frac{2i-1}{2n}\right) \right| \le \varepsilon \left| t - \frac{2i-1}{2n} \right|, \quad \forall t \in J_{n,i},$$

and consequently, with $y_{n,i} = \tau_r^{f|_{J_{n,i}}}$,

$$|y_{n,i} - x_{n,i}| \le \frac{\varepsilon}{n}, \quad \forall 1 \le i \le n,$$

by Proposition 4.9. It follows that

$$n^r d_r(\delta_{\bullet}^{\mathbf{u}_n}, \delta_{\mathbf{x}_n}^{\mathbf{u}_n})^r = n^r d_r(\delta_{\mathbf{y}_n}^{\mathbf{u}_n}, \delta_{\mathbf{x}_n}^{\mathbf{u}_n})^r = n^r \sum_{i=1}^n \int_{J_{n,i}} |y_{n,i} - x_{n,i}|^r \le \varepsilon^r,$$

and since $\varepsilon > 0$ was arbitrary,

$$\limsup_{n\to\infty} |n^r d_r(\delta_{\bullet}^{\mathbf{u}_n}, \mu)^r - n^r d_r(\delta_{\mathbf{x}_n}^{\mathbf{u}_n}, \mu)^r| \le \lim_{n\to\infty} n^r d_r(\delta_{\bullet}^{\mathbf{u}_n}, \delta_{\mathbf{x}_n}^{\mathbf{u}_n})^r = 0,$$

which establishes (5.8), with the same finite value on either side.

Step 2. Let the non-decreasing and absolutely continuous function f be arbitrary, but assume that $f' \in L^r(\mathbb{I})$. Similarly, let $\widetilde{\mu} \in \mathcal{P}_r$ be such that $\widetilde{\mu}^{-1}$ is absolutely continuous, with $\widetilde{f} := F_{\widetilde{\mu}}^{-1}$ and $\widetilde{f}' \in L^r(\mathbb{I})$. For $n \in \mathbb{N}$ and $1 \le i \le n$, pick any $t_{n,i} \in J_{n,i}$ and define $z_{n,i} = f(t_{n,i})$, $\widetilde{z}_{n,i} = \widetilde{f}(t_{n,i})$. Below, it will be shown that, for any $r \ge 1$,

$$|n^r d_r(\delta_{\mathbf{z}_n}^{\mathbf{u}_n}, \mu)^r - n^r d_r(\delta_{\widetilde{\mathbf{z}}_n}^{\mathbf{u}_n}, \widetilde{\mu})^r| \le 2\|f' - \widetilde{f}'\|_r \|f' + \widetilde{f}'\|_r^{r-1}, \quad \forall n \in \mathbb{N}. \tag{5.9}$$

To see that (5.8) follows easily from (5.9), at least under the current assumption that $f' \in L^r(\mathbb{I})$, fix $r \ge 1$ and w.o.l.g. $0 < \varepsilon < \|f'\|_r$. There exists $\widetilde{\mu} \in \mathcal{P}_r$ such that $\widetilde{f}$ has a C^1-extension to $\mathbb{I}$, and $\|f' - \widetilde{f}'\|_r < \varepsilon$. With the appropriate t_n, let $\delta_{\mathbf{z}_n}^{\mathbf{u}_n}$ be a best uniform r-approximation of μ, and $\delta_{\mathbf{x}_n}^{\mathbf{u}_n}$ a best uniform r-approximation of $\widetilde{\mu}$. For all sufficiently large n, Step 1 and (5.9) yield

$$\begin{aligned} n^r d_r(\delta_{\mathbf{z}_n}^{\mathbf{u}_n}, \mu)^r &\le n^r d_r(\delta_{\mathbf{x}_n}^{\mathbf{u}_n}, \mu)^r \le n^r d_r(\delta_{\widetilde{\mathbf{x}}_n}, \mu)^r + 2\varepsilon(\varepsilon + 2\|f'\|_r)^{r-1} \\ &\le \frac{1}{2^r(1+r)}(\|f'\|_r + \varepsilon)^r + \varepsilon + 2\varepsilon(\varepsilon + 2\|f'\|_r)^{r-1}, \end{aligned}$$

but also

$$\begin{aligned} n^r d_r(\delta_{\mathbf{z}_n}^{\mathbf{u}_n}, \mu)^r &\ge n^r d_r(\delta_{\widetilde{\mathbf{z}}_n}^{\mathbf{u}_n}, \widetilde{\mu})^r - 2\varepsilon(\varepsilon + 2\|f'\|_r)^{r-1} \\ &\ge \frac{1}{2^r(1+r)}(\|f'\|_r - \varepsilon)^r - \varepsilon - 2\varepsilon(\varepsilon + 2\|f'\|_r)^{r-1}. \end{aligned}$$

Since $\varepsilon > 0$ was arbitrary, this establishes (5.8).

It remains to verify (5.9), which only requires the elementary inequality, valid for all $r \ge 1$,

$$|a^r - b^r| \le r|a - b|(a^{r-1} + b^{r-1}), \quad \forall a, b \ge 0, \tag{5.10}$$

together with a repeated application of Hölder's inequality, as follows: Note first that

$$d_r(\delta_{\mathbf{z}_n}^{\mathbf{u}_n}, \mu)^r = \sum_{i=1}^n \left(\int_{(i-1)/n}^{t_{n,i}} \left(\int_t^{t_{n,i}} f'(u)\mathrm{d}u \right)^r \mathrm{d}t + \int_{t_{n,i}}^{i/n} \left(\int_{t_{n,i}}^t f'(u)\mathrm{d}u \right)^r \mathrm{d}t \right), \tag{5.11}$$

and consequently

$$\begin{aligned} \left| d_r(\delta_{\mathbf{z}_n}^{\mathbf{u}_n}, \mu)^r - d_r(\delta_{\widetilde{\mathbf{z}}_n}^{\mathbf{u}_n}, \widetilde{\mu})^r \right| \le \sum_{i=1}^n \Bigg(&\int_{(i-1)/n}^{t_{n,i}} \left| \left(\int_t^{t_{n,i}} f'(u)\,\mathrm{d}u \right)^r - \left(\int_t^{t_{n,i}} \widetilde{f}'(u)\,\mathrm{d}u \right)^r \right| \mathrm{d}t \\ &+ \int_{t_{n,i}}^{i/n} \left| \left(\int_{t_{n,i}}^t f'(u)\,\mathrm{d}u \right)^r - \left(\int_{t_{n,i}}^t \widetilde{f}'(u)\,\mathrm{d}u \right)^r \right| \mathrm{d}t \Bigg). \end{aligned}$$

With (5.10), therefore,

$$
\begin{aligned}
&\int_{(i-1)/n}^{t_{n,i}} \left| \left(\int_t^{t_{n,i}} f'(u)\,\mathrm{d}u \right)^r - \left(\int_t^{t_{n,i}} \widetilde{f}'(u)\,\mathrm{d}u \right)^r \right| \mathrm{d}t \\
&\le r \int_{(i-1)/n}^{t_{n,i}} \left| \int_t^{t_{n,i}} \left(f'(u) - \widetilde{f}'(u) \right) \mathrm{d}u \right| \left(\left(\int_t^{t_{n,i}} f'(u)\mathrm{d}u \right)^{r-1} + \left(\int_t^{t_{n,i}} \widetilde{f}'(u)\,\mathrm{d}u \right)^{r-1} \right) \mathrm{d}t \\
&\le 2r \int_{(i-1)/n}^{t_{n,i}} \left| \int_t^{t_{n,i}} \left(f'(u) - \widetilde{f}'(u) \right) \mathrm{d}u \right| \left(\int_t^{t_{n,i}} \left(f'(u) + \widetilde{f}'(u) \right) \mathrm{d}u \right)^{r-1} \mathrm{d}t \\
&\le 2r (a_i^-)^{1/r} (b_i^-)^{(r-1)/r},
\end{aligned}
$$

where, using Hölder's inequality again

$$
\begin{aligned}
a_i^- &= \int_{(i-1)/n}^{t_{n,i}} \left| \int_t^{t_{n,i}} \left(f'(u) - \widetilde{f}'(u) \right) \mathrm{d}u \right|^r \mathrm{d}t \le \frac{1}{rn^r} \int_{(i-1)/n}^{t_{n,i}} \left| f'(t) - \widetilde{f}'(t) \right|^r \mathrm{d}t, \\
b_i^- &= \int_{(i-1)/n}^{t_{n,i}} \left(\int_t^{t_{n,i}} \left(f'(u) + \widetilde{f}'(u) \right) \mathrm{d}u \right)^r \mathrm{d}t \le \frac{1}{rn^r} \int_{(i-1)/n}^{t_{n,i}} \left(f'(t) + \widetilde{f}'(t) \right)^r \mathrm{d}t.
\end{aligned}
$$

By a completely analogous argument,

$$
\int_{t_{n,i}}^{i/n} \left| \left(\int_{t_{n,i}}^t f'(u)\,\mathrm{d}u \right)^r - \left(\int_{t_{n,i}}^t \widetilde{f}'(u)\,\mathrm{d}u \right)^r \right| \mathrm{d}t \le 2r (a_i^+)^{1/r} (b_i^+)^{(r-1)/r}, \quad \forall 1 \le i \le n,
$$

where

$$
\begin{aligned}
a_i^+ &= \int_{t_{n,i}}^{i/n} \left| \int_{t_{n,i}}^t \left(f'(u) - \widetilde{f}'(u) \right) \mathrm{d}u \right|^r \mathrm{d}t \le \frac{1}{rn^r} \int_{t_{n,i}}^{i/n} \left| f'(t) - \widetilde{f}'(t) \right|^r \mathrm{d}t, \\
b_i^+ &= \int_{t_{n,i}}^{i/n} \left(\int_{t_{n,i}}^t \left(f'(u) + \widetilde{f}'(u) \right) \mathrm{d}u \right)^r \mathrm{d}t \le \frac{1}{rn^r} \int_{t_{n,i}}^{i/n} \left(f'(t) + \widetilde{f}'(t) \right)^r \mathrm{d}t.
\end{aligned}
$$

In summary, therefore,

$$
\begin{aligned}
&n^r \left| d_r(\delta_{\mathbf{z}_n}^{\mathbf{u}_n}, \mu)^r - d_r(\delta_{\widetilde{\mathbf{z}}_n}^{\mathbf{u}_n}, \widetilde{\mu})^r \right| \le 2rn^r \sum_{i=1}^n \left((a_i^-)^{1/r} (b_i^-)^{(r-1)/r} + (a_i^+)^{1/r} (b_i^+)^{(r-1)/r} \right) \\
&\le 2rn^r \left(\sum_{i=1}^n (a_i^- + a_i^+) \right)^{1/r} \left(\sum_{i=1}^n (b_i^- + b_i^+) \right)^{(r-1)/r} \\
&\le 2rn^r \left(\frac{1}{rn^r} \int_{\mathbb{I}} |f'(t) - \widetilde{f}'(t)|^r \mathrm{d}t \right)^{1/r} \left(\frac{1}{rn^r} \int_{\mathbb{I}} \left(f'(t) + \widetilde{f}'(t) \right)^r \mathrm{d}t \right)^{(r-1)/r} \\
&= 2 \| f' - \widetilde{f}' \|_r \| f' + \widetilde{f}' \|_r^{r-1},
\end{aligned}
$$

which is just (5.9).

Step 3. To establish (5.8) in case the value on the right is $+\infty$, assume that $f' \notin L^r(\mathbb{I})$. For $N \in \mathbb{N}$, let $g_N = \min\{f', N\}$ and, given $C > 0$, choose N so large that $\|g_N\|_r^r \geq 2^r(1+r)C$. Let μ_N be a probability measure with $(F_{\mu_N}^{-1})' = g_N$. By (5.11),

$$\begin{aligned} d_r(\delta_{\mathbf{z}_n}^{\mathbf{u}_n}, \mu)^r &\geq \sum_{i=1}^{n} \left(\int_{(i-1)/n}^{t_i} \left(\int_t^{t_i} g_N(u)\,\mathrm{d}u \right)^r \mathrm{d}t + \int_{t_i}^{i/n} \left(\int_{t_i}^{t} g_N(u)\,\mathrm{d}u \right)^r \mathrm{d}t \right) \\ &\geq d_r(\delta_\bullet^{\mathbf{u}_n}, \mu_N)^r, \end{aligned}$$

and since Step 2 applies to μ_N,

$$\liminf_{n\to\infty} n^r d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)^r \geq \lim_{n\to\infty} n^r d_r(\delta_\bullet^{\mathbf{u}_n}, \mu_N)^r = \frac{1}{2^r(r+1)} \|g_N\|_r^r \geq C.$$

As $C > 0$ was arbitrary, $n^r d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)^r \to +\infty$ whenever $f' \notin L^r(\mathbb{I})$, i.e., (5.8) is valid in this case also.

Step 4. Finally, to prove the assertion regarding asymptotically best uniform approximations, assume that $f' \in L^r(\mathbb{I})$. Note that $\|f'\|_r > 0$ whenever $\mu \neq \delta_a$ for all $a \in \mathbb{R}$. In this case, given $\varepsilon > 0$, pick $\widetilde{\mu} \in \mathcal{P}_r$ such that $\widetilde{f} = F_{\widetilde{\mu}}^{-1}$ has a C^1-extension to $\mathbb{I}$ and $\|f' - \widetilde{f}'\|_r < \varepsilon$. By Step 1, $\lim_{n\to\infty} n^r d_r(\delta_{\widetilde{\mathbf{x}}_n}^{\mathbf{u}_n}, \widetilde{\mu})^r = \frac{\|\widetilde{f}'\|_r^r}{2^r(r+1)}$, whereas Step 2 guarantees that $\lim_{n\to\infty} n^r d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)^r = \frac{\|f'\|_r^r}{2^r(r+1)}$, and (5.9) yields

$$|n^r d_r(\delta_{\widetilde{\mathbf{x}}_n}^{\mathbf{u}_n}, \widetilde{\mu})^r - n^r d_r(\delta_{\mathbf{x}_n}^{\mathbf{u}_n}, \mu)^r| \leq 2\varepsilon(1 + 2\|f'\|_r)^{r-1}.$$

Combining these three facts leads to

$$\begin{aligned} \limsup_{n\to\infty} \frac{d_r(\delta_{\mathbf{x}_n}^{\mathbf{u}_n}, \mu)^r}{d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)^r} &\leq \limsup_{n\to\infty} \frac{n^r d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)^r + 2\varepsilon(1+2\|f'\|_r)^{r-1}}{2^{-r}\|f'\|_r^r (r+1)^{-1}} \\ &\leq \left(1 + \frac{\varepsilon}{\|f'\|_r}\right)^r + 2^{r+1}(r+1)\varepsilon \frac{(1+2\|f'\|_r)^{r-1}}{\|f'\|_r^r}, \end{aligned}$$

as well as to an analogous lower bound for $\liminf_{n\to\infty} \frac{d_r(\delta_{\mathbf{x}_n}^{\mathbf{u}_n}, \mu)^r}{d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)^r}$. Since $\varepsilon > 0$ was arbitrary, $\lim_{n\to\infty} \frac{d_r(\delta_{\mathbf{x}_n}^{\mathbf{u}_n}, \mu)}{d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)} = 1$, i.e., $(\delta_{\mathbf{x}_n}^{\mathbf{u}_n})$ is a sequence of asymptotically best uniform r-approximations of μ, as claimed. $\square$

The following examples highlight the importance of the absolute continuity and integrability assumptions, respectively, in Theorem 5.15.

Example 5.16. Let μ be the inverse of the Cantor probability measure in Example 5.10. Explicitly, μ is purely atomic, with $\mu(\{j2^{-m}\}) = 3^{-m}$ for every $m \in \mathbb{N}$ and every odd $1 \leq j \leq 2^m$. Note that $F_\mu^{-1}|_{]0,1[}$ simply equals the classical Cantor function, hence is continuous, in fact, $\frac{\log 2}{\log 3}$-Hölder, and $\frac{\mathrm{d}\mu^{-1}}{\mathrm{d}\lambda} = 0$ a.e. While Theorem 5.15, if it did apply, would seem to suggest that $\lim_{n\to\infty} n d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) = 0$, a detailed, elementary analysis shows that this is not the case. In fact, $\left(n^\alpha d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\right)$ may not converge to a finite positive limit for any $r \geq 1$ and $\alpha > 0$. More specifically, let $\alpha_r = \frac{1}{r} + (1 - \frac{1}{r})\frac{\log 2}{\log 3}$ for $r \geq 1$. With this, $3^{\alpha_r} d_r(\delta_\bullet^{\mathbf{u}_{3n}}, \mu) = d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)$ for all n, and hence

$$2^{-2+1/r}3^{-2/r} \leq \liminf_{n\to\infty} n^{\alpha_r} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) = \inf_{n\in\mathbb{N}} n^{\alpha_r} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu),$$

as well as

$$\limsup_{n\to\infty} n^{\alpha_r} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) = \sup_{n\in\mathbb{N}} n^{\alpha_r} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) \le 2^{1/r}.$$

We suspect the bounded sequence $\big(n^{\alpha_r} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$ to be divergent for every $r \ge 1$. This illustrates that the conclusion of Theorem 5.15 may fail if μ^{-1} is not absolutely continuous.

Example 5.17. The integrability assumption also is crucial (for the second assertion) in Theorem 5.15. To see this, let μ be the standard exponential distribution, where $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda} \notin L^1(\mathbb{I})$, and (5.8) yields $\lim_{n\to\infty} n d_r(\delta_{\bullet,r}^{\mathbf{u}_n}, \mu) = +\infty$ for all $r \ge 1$, in perfect agreement with the observations in Example 5.8. Deduce from a short calculation that

$$\sqrt{n} d_2(\delta_{\bullet,1}^{\mathbf{u}_n}, \mu) = D_2 + \mathcal{O}(n^{-1}) \quad \text{as } n \to \infty,$$

where $D_2^2 = 1 + 2\sum_{i=1}^\infty \left(1 - i(1 + \log \frac{\sqrt{4i^2-1}}{2i}) \log \frac{2i+1}{2i-1}\right) \approx 1.1749$. Recall from Example 5.8 that $\sqrt{n} d_2(\delta_{\bullet,2}^{\mathbf{u}_n}, \mu) = C_2 + \mathcal{O}(n^{-1})$ with $C_2^2 \approx 1.0803$. Thus while $(\delta_{\bullet,1}^{\mathbf{u}_n})$ identifies the correct rate of decay for $\big(d_2(\delta_{\bullet,2}^{\mathbf{u}_n}, \mu)\big)$, namely $\big(n^{-1/2}\big)$, it is *not* a sequence of asymptotically best uniform 2-approximations of μ, since

$$\lim_{n\to\infty} \frac{d_2(\delta_{\bullet,1}^{\mathbf{u}_n}, \mu)}{d_2\left(\delta_{\bullet,2}^{\mathbf{u}_n}, \mu\right)} = \frac{D_2}{C_2} > 1.$$

Similarly, for any $r > 1$ it can be shown that $\lim_{n\to\infty} n^{1/r} d_r(\delta_{\bullet,1}^{\mathbf{u}_n}, \mu) = D_r$ with the appropriate constant $D_r > C_r$, and C_r as in Example 5.8.

Example 5.18. Let μ be the standard normal distribution. By [15, Thm.1], $\big(d_2(\delta_{\bullet,2}^{\mathbf{u}_n}, \mu)\big)$ decays like $\big(n^{-1/2}(\log n)^{-1/2}\big)$ along the subsequence $n = 2^k$. Moreover, as pointed out in [15, Rem.6], best uniform 1-approximations $\delta_{\bullet,1}^{\mathbf{u}_n}$ converge to μ as fast as $\delta_{\bullet,2}^{\mathbf{u}_n}$ do (w.r.t. d_2). In fact, elementary computations confirm that the sequences

$$\big(n^{1/r}(\log n)^{\alpha_r} d_r(\delta_{\bullet,r}^{\mathbf{u}_n}, \mu)\big), \quad \big(n^{1/r}(\log n)^{\alpha_r} d_r(\delta_{\bullet,1}^{\mathbf{u}_n}, \mu)\big)$$

are bounded above and below by positive constants, where $\alpha_r = -\frac{1}{2}$ if $r = 1$, and $\alpha_r = \frac{1}{2}$ if $r > 1$. Thus, best uniform 1-approximations converge precisely as fast as do best uniform r-approximations. Together with the previous example, this suggests that $(\delta_{\bullet,1}^{\mathbf{u}_n})$, though perhaps not a sequence of asymptotically best uniform r-approximations, may nevertheless often capture the correct rate of convergence of $\big(d_r(\delta_{\bullet,r}^{\mathbf{u}_n}, \mu)\big)$, even when $\frac{\mathrm{d}\mu^{-1}}{\mathrm{d}\lambda} \notin L^r(\mathbb{I})$.

Remark 5.19. For any (non-degenerate) $\mu \in \mathcal{P}$, [3, Prop. A.17] asserts that μ^{-1} is absolutely continuous if and only if supp μ is connected and $\frac{\mathrm{d}\mu_a}{\mathrm{d}\lambda} > 0$ a.e. on supp μ, where μ_a is the absolutely continuous part (w.r.t. λ) of μ; in this case, moreover, $\int_{\mathbb{I}} (\frac{\mathrm{d}\mu^{-1}}{\mathrm{d}\lambda})^r = \int_{\mathrm{supp}\,\mu} (\frac{\mathrm{d}\mu_a}{\mathrm{d}\lambda})^{1-r}$.

If F_μ^{-1} not even is continuous, then the decay of $\big(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$ may be less homogeneous than in Example 5.16. For instance, for the Cantor measure of Example 5.10, it is not hard to see that, for any $r \ge 1$, both numbers

$$\liminf_{n\to\infty} n^{\frac{\log 3}{\log 2}} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) \text{ and } \limsup_{n\to\infty} n^{1/r} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)$$

are finite and positive. Thus, in general it cannot be expected that for some $\alpha_r > 0$, the sequence $\big(n^{\alpha_r} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$ is bounded below and above by positive constants, let alone convergent. Still, it is possible to identify a universal lower bound for $\big(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$ with $\mu \in \mathcal{P}_r$: But for trivial exceptions, this sequence never decays faster than $\big(n^{-1}\big)$.

Theorem 5.20. *Assume that $\mu \in \mathcal{P}_r$ for some $r \geq 1$. Then*

$$\limsup_{n\to\infty} n d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) > 0, \tag{5.12}$$

unless $\mu = \delta_a$ for some $a \in \mathbb{R}$.

Proof. Denote F_μ^{-1} by f for convenience, and for every $n \in \mathbb{N}$, let $a_i = f(\frac{2i-1}{4n})$ and $b_i = f(\frac{2i-1}{4n+2})$ for $1 \leq i \leq 2n$. Then $b_1 \leq a_1 \leq b_2 \leq a_2 \cdots \leq b_{2n} \leq a_{2n} \leq b_{2n+1}$, and

$$\begin{aligned}
&2nd_1(\delta_\bullet^{\mathbf{u}_{2n}}, \delta_\bullet^{\mathbf{u}_{2n+1}})\\
&= 2n \sum_{i=1}^{2n} \left((a_i - b_i)\left(\frac{i}{2n+1} - \frac{i-1}{2n}\right) + (b_{i+1} - a_i)\left(\frac{i}{2n} - \frac{i}{2n+1}\right)\right)\\
&= \frac{1}{2n+1} \sum_{i=1}^{2n} \big((2n+1-i)(a_i - b_i) + i(b_{i+1} - a_i)\big)\\
&\geq \frac{1}{2n+1}\left(\sum_{i=1}^{n} i(b_{i+1} - b_i) + \sum_{i=n+1}^{2n} \big((2n+1-2i)(a_i - b_i) + i(b_{i+1} - b_i)\big)\right)\\
&= \frac{1}{2n+1}\left(\sum_{i=1}^{n} i(b_{i+1} - b_i) + \sum_{i=n+1}^{2n} \big((2n+1-i)(b_{i+1} - b_i) + (2i-2n-1)(b_{i+1} - a_i)\big)\right)\\
&\geq \frac{1}{2n+1}\left(\sum_{i=1}^{n} i(b_{i+1} - b_i) + \sum_{i=n+1}^{2n} (2n+1-i)(b_{i+1} - b_i)\right)\\
&= \sum_{i=n+2}^{2n+1} \frac{b_i}{2n+1} - \sum_{i=1}^{n} \frac{b_i}{2n+1}.
\end{aligned}$$

Since f is locally Riemann integrable on $]0, 1[$, it follows that

$$\limsup_{n\to\infty} 2nd_1(\delta_\bullet^{\mathbf{u}_{2n}}, \delta_\bullet^{u_{2n+1}}) \geq \int_0^{1/2} \big(f(t + \tfrac{1}{2}) - f(t)\big)\, \mathrm{d}t,$$

and consequently

$$\begin{aligned}
\limsup_{n\to\infty} nd_r(\delta_\bullet^{\mathbf{u}_n}, \mu) &\geq \limsup_{n\to\infty} nd_1(\delta_\bullet^{\mathbf{u}_n}, \mu) \geq \tfrac{1}{2} \limsup_{n\to\infty} 2nd_1(\delta_\bullet^{\mathbf{u}_{2n}}, \delta_\bullet^{\mathbf{u}_{2n+1}})\\
&\geq \tfrac{1}{2}\int_0^{1/2} \big(f(t + \tfrac{1}{2}) - f(t)\big)\, \mathrm{d}t > 0
\end{aligned}$$

unless f is constant, i.e., unless $\mu = \delta_a$ for some $a \in \mathbb{R}$. □

It is natural to ask whether Theorem 5.20 has a counterpart in that there also exists a universal *upper* bound on $\big(d_r\big(\delta_\bullet^{\mathbf{u}_n}, \mu\big)\big)$. In general, this is not the case: As an immediate consequence of Theorem 5.33, given $r \geq 1$ and any sequence (a_n) of positive real numbers with $\lim_{n\to\infty} a_n = 0$, there exists $\mu \in \mathcal{P}_r$ such that $d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) \geq a_n$, for all $n \in \mathbb{N}$. Under additional assumptions, however, an upper bound on $\big(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$ can be established.

Theorem 5.21. *Assume that $\mu \in \mathcal{P}_r$ for some $r \geq 1$.*

(i) *If $\mu \in \mathcal{P}_s$ with $s > r$ then* $\lim_{n\to\infty} n^{1/r-1/s} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) = 0$.

(ii) *If* $\operatorname{supp}\mu$ *is bounded then* $\limsup_{n\to\infty} n^{1/r} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) < +\infty$.

Proof. Again, for convenience, let $f = F_\mu^{-1}$, and $x_{n,i} = f(\frac{2i-1}{2n})$ for all $n \in \mathbb{N}$ and $1 \leq i \leq n$. With $t_0 = F_\mu(0)$, assume w.o.l.g. that $0 < t_0 < 1$. (The cases $t_0 = 0$ and $t_0 = 1$ are completely analogous.) Recall that f is non-decreasing and right-continuous, $(t - t_0)f(t) \geq 0$ for all $t \in \mathbb{I}$, and $0 \leq f(t_0),\ -f(t_0-) < +\infty$. For all sufficiently large n, therefore,

$$
\begin{aligned}
d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)^r &\leq d_r(\delta_{\mathbf{x}_n}^{\mathbf{u}_n}, \mu)^r = \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{2i-1}{2n}} \left(\left(x_{n,i} - f(t)\right)^r + \left(f\left(t + \frac{1}{2n}\right) - x_{n,i}\right)^r \right) \mathrm{d}t \\
&\leq \sum_{i=1}^n \int_{\frac{i-1}{n}}^{\frac{2i-1}{2n}} \left(f\left(t + \frac{1}{2n}\right) - f(t)\right)^r \mathrm{d}t \leq \int_{\frac{1}{4n}}^{1-\frac{1}{4n}} \left(f\left(t + \frac{1}{4n}\right) - f\left(t - \frac{1}{4n}\right)\right)^r \mathrm{d}t \\
&= \int_{\frac{1}{4n}}^{t_0-\frac{1}{4n}} \left(\left| f\left(t - \frac{1}{4n}\right)\right| - \left| f\left(t + \frac{1}{4n}\right)\right| \right)^r \mathrm{d}t \\
&\quad + \int_{t_0-\frac{1}{4n}}^{t_0+\frac{1}{4n}} \left(f\left(t + \frac{1}{4n}\right) + \left| f\left(t - \frac{1}{4n}\right)\right| \right)^r \mathrm{d}t \\
&\quad + \int_{t_0+\frac{1}{4n}}^{1-\frac{1}{4n}} \left(f\left(t + \frac{1}{4n}\right) - f\left(t - \frac{1}{4n}\right)\right)^r \mathrm{d}t \\
&\leq \int_0^{t_0-\frac{1}{2n}} |f(t)|^r \mathrm{d}t - \int_{\frac{1}{2n}}^{t_0} |f(t)|^r \mathrm{d}t + 2^{r-1} \int_{t_0-\frac{1}{2n}}^{t_0+\frac{1}{2n}} |f(t)|^r \mathrm{d}t \\
&\quad + \int_{t_0+\frac{1}{2n}}^1 |f(t)|^r \mathrm{d}t - \int_{t_0}^{1-\frac{1}{2n}} |f(t)|^r \mathrm{d}t \\
&= a_n + (2^{r-1} - 1) b_n,
\end{aligned}
$$

where the numbers a_n, b_n are given by

$$
a_n = \int_0^{\frac{1}{2n}} |f(t)|^r \mathrm{d}t + \int_{1-\frac{1}{2n}}^1 |f(t)|^r \mathrm{d}t \quad \text{and} \quad b_n = \int_{t_0-\frac{1}{2n}}^{t_0+\frac{1}{2n}} |f(t)|^r \mathrm{d}t,
$$

respectively. Note that

$$
0 \leq n b_n \leq \max\left\{ f\left(t_0 + \frac{1}{2n}\right), -f\left(t_0 - \frac{1}{2n}\right)\right\}^r \to \max\{f(t_0), -f(t_0-)\}^r \quad \text{as } n \to \infty,
$$

and hence (nb_n) is bounded.

(i) If $\mu \in \mathcal{P}_s$ for some $s > r$ then, by virtue of Hölder's inequality,

$$
0 \leq a_n \leq \left(\left(\int_0^{\frac{1}{2n}} |f(t)|^s \mathrm{d}t \right)^{r/s} + \left(\int_{1-\frac{1}{2n}}^1 |f(t)|^s \mathrm{d}t \right)^{r/s} \right) 2^{r/s} n^{r/s-1},
$$

which shows that $\lim_{n\to\infty} n^{1-r/s} a_n = 0$. It follows that

$$0 \le n^{1-r/s} d_r\left(\delta_\bullet^{\mathbf{u}_n}, \mu\right)^r \le n^{1-r/s} a_n + (2^{r-1}-1) n^{1-r/s} b_n \to 0 \quad \text{as } n \to \infty,$$

and hence $\lim_{n\to\infty} n^{1/r-1/s} d_r\left(\delta_\bullet^{\mathbf{u}_n}, \mu\right) = 0$, as claimed.

(ii) If supp μ is bounded then $\operatorname{esssup}_{\mathbb{I}}|f|$ is finite. In this case, (na_n) is bounded, and so is $\left(n^{1/r} d_r\left(\delta_\bullet^{\mathbf{u}_n}, \mu\right)\right)$. $\square$

Remark 5.22. (i) Boundedness of supp μ is essential in Theorem 5.21(ii), as evidenced, e.g., by Example 5.17 for $r = 1$. Notice, however, that the conclusion of Theorem 5.21(ii) remains valid in this example whenever $r > 1$.

(ii) If supp μ is disconnected, and hence F_μ^{-1} is discontinuous at some $0 < t < 1$, then there exists (n_k) such that $\langle n_k t\rangle \in [1/3, 2/3]$ for all k. For all sufficiently large k, therefore,

$$\begin{aligned} n_k d_r(\delta_\bullet^{\mathbf{u}_{n_k}}, \mu)^r &\ge n_k \min_{c\in\mathbb{R}} \int_{\lfloor n_k t\rfloor/n_k}^{(\lfloor n_k t\rfloor+1)/n_k} \left|F_\mu^{-1}(s) - c\right|^r \mathrm{d}s \\ &\ge \min_{c\in[F_\mu^{-1}(t-), F_\mu^{-1}(t)]} \tfrac{1}{3}\left(\left(F_\mu^{-1}(t) - c\right)^r + \left(c - F_\mu^{-1}(t-)\right)^r\right) \\ &\ge 2^{1-r}3^{-1}\left(F_\mu^{-1}(t) - F_\mu^{-1}(t-)\right)^r . \end{aligned}$$

Hence (5.12) can be strengthened to $\limsup_{n\to\infty} n^{1/r} d_r(\delta_\bullet^{\mathbf{u}_n}, \mu) > 0$ whenever supp μ is disconnected. In fact, by Theorem 5.21(ii), $(n^{-1/r})$ is the sharp upper rate of $\left(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\right)$ in case supp μ is bounded and disconnected, a situation observed for instance for the Cantor measure of Example 5.10.

(iii) By utilizing a *uniform decomposition approach*, a multi-dimensional analogue of Theorem 5.21 is established in [8], with the threshold rates of convergence depending both on r and on the dimension of the ambient (Euclidean) space.

5.3. Best approximations

This final subsection relates the results presented earlier to the classical theory of best (unconstrained) approximations. Let $\mu \in \mathcal{P}_r$ for some $r \ge 1$. Given $n \in \mathbb{N}$, call the probability measure $\delta_{\mathbf{x}}^{\mathbf{p}}$ with $\mathbf{x} \in \Xi_n$ and $\mathbf{p} \in \Pi_n$ a *best* r*-approximation* of μ if

$$d_r(\delta_{\mathbf{x}}^{\mathbf{p}}, \mu) \le d_r(\delta_{\mathbf{y}}^{\mathbf{q}}, \mu), \quad \forall \mathbf{y} \in \Xi_n,\ \mathbf{q} \in \Pi_n.$$

Denote by $\delta_\bullet^{\bullet,n}$ any best r-approximation of μ. (As before, the dependence of $\delta_\bullet^{\bullet,n}$ on r is made explicit by a subscript only where necessary to avoid ambiguities.) It is well known that best r-approximations exist always.

Proposition 5.23 ([17, Sec. 4.1]).. *Assume that $\mu \in \mathcal{P}_r$ for some $r \ge 1$. For every $n \in \mathbb{N}$, there exists a best r-approximation $\delta_\bullet^{\bullet,n}$ of μ. If # supp $\mu \ge n$ then # supp $\delta_\bullet^{\bullet,n} = n$.*

By combining Proposition 5.23 with Theorem 5.1 and 5.5, a description of all best r-approximations is easily established.

Theorem 5.24. *Assume that $\mu \in \mathcal{P}_r$ for some $r \ge 1$, and $n \in \mathbb{N}$. Let $\delta_{\mathbf{x}}^{\mathbf{p}}$ with $\mathbf{x} \in \Xi_n$, $\mathbf{p} \in \Pi_n$ be a best r-approximation of μ. Then, for every $i = 1, \ldots, n$,*

(i) $x_i < x_{i+1}$ *implies* $P_i \in Q^{F_\mu^{-1}}_{\frac{1}{2}(x_i+x_{i+1})}$; *and*

(ii) $P_{i-1} < P_i$ *implies* $x_i \in Q^{F_\mu}_{\frac{1}{2}(P_{i-1}+P_i)}$ *if* $r = 1$, *or* $x_i = \tau_r^{f_i}$ *with* $f_i = F_\mu^{-1}\big|_{[P_{i-1},P_i]}$ *if* $r > 1$.

Moreover, if $\#\,\mathrm{supp}\,\mu \le n$ *then* $\delta_{\mathbf{x}}^{\mathbf{p}} = \mu$, *whereas if* $\#\,\mathrm{supp}\,\mu > n$ *then* $x_i < x_{i+1}$ *and* $P_{i-1} < P_i$ *for all* $i = 1, \ldots, n$.

Proof. Note that $\delta_{\mathbf{x}}^{\mathbf{p}}$ is both a best r-approximation of μ, given $\mathbf{p}$, and a best r-approximation of μ, given $\mathbf{x}$, and thus conclusions (i) and (ii) follow directly from Theorem 5.1 and 5.5. For the non-trivial case where $\#\,\mathrm{supp}\,\mu > n$, Proposition 5.23 implies that $\#\,\mathrm{supp}\,\delta_{\mathbf{x}}^{\mathbf{p}} = n$, or equivalently, $x_i < x_{i+1}$ and $P_{i-1} < P_i$ for all $i = 1, \ldots, n$. □

As an important special case of Theorem 5.24, assume that $\mu \in \mathcal{P}_r$ is continuous. Then $Q_a^{F_\mu^{-1}}$ is a singleton for every $a \in \mathbb{R}$, and Theorem 5.24 asserts that every best 1-approximation $\delta_{\mathbf{x}}^{\mathbf{p}}$ of μ satisfies

$$F_\mu\left(\frac{x_i + x_{i+1}}{2}\right) = P_i, \text{ and } F_\mu(x_i) = \frac{P_{i-1} + P_i}{2}, \quad \forall i = 1, \ldots, n,$$

and hence in particular

$$2F_\mu(x_i) = F_\mu\left(\frac{x_{i-1} + x_i}{2}\right) + F_\mu\left(\frac{x_i + x_{i+1}}{2}\right), \quad \forall i = 1, \ldots, n. \tag{5.13}$$

Similarly, every best 2-approximation of μ satisfies

$$F_\mu\left(\frac{x_i + x_{i+1}}{2}\right) = P_i, \text{ and } (P_i - P_{i-1})x_i = \int_{P_{i-1}}^{P_i} F_\mu^{-1}(t)\,\mathrm{d}t, \quad \forall i = 1, \ldots, n,$$

and consequently

$$x_i F_\mu\left(\frac{x_i + x_{i+1}}{2}\right) - x_i F_\mu\left(\frac{x_{i-1} + x_i}{2}\right) = \int_{\frac{1}{2}(x_{i-1}+x_i)}^{\frac{1}{2}(x_i+x_{i+1})} x\,\mathrm{d}F_\mu(x), \quad \forall i = 1, \ldots, n. \tag{5.14}$$

Note that (5.13) and (5.14) each yield n equations for $x_1, \ldots, x_n$. These equations are exactly the classical optimality conditions, derived, e.g., in [17, Sec. 5.2] by means of Voronoi partitions.

Example 5.25. Let $\mu = \frac{1}{2}\lambda|_{[0,1]} + \frac{1}{2}\delta_1$. While μ is not continuous, and hence not directly amenable to the classical conditions (5.13) and (5.14), Theorem 5.24 applies and yields, for instance, $\delta_{\bullet,r}^{\bullet,2} = \xi(r)\delta_{\xi(r)} + \big(1 - \xi(r)\big)\delta_{3\xi(r)}$ for all $r \ge 1$, where $r \mapsto \xi(r)$ is smooth, decreasing, with $\xi(1) = \frac{1}{3}$, $\xi(2) = \frac{3-\sqrt{3}}{4}$, and $\lim_{r\to+\infty} \xi(r) = \frac{1}{4}$.

If (i) and (ii) in Theorem 5.24 identify only a single probability measure $\delta_{\mathbf{x}}^{\mathbf{p}}$ then the latter clearly is a best r-approximation. In general, however, and unlike in Theorems 5.1 and 5.5, the conditions of Theorem 5.24 are not sufficient, as the following example shows. Moreover, best r-approximations in general are not unique, not even when $r > 1$.

Example 5.26. Consider $\mu = \frac{1}{3}\lambda_{[-1,1]} + \frac{1}{3}\delta_0$ and let $n = 2$. For $r = 1$, Theorem 5.24 identifies exactly three potential best 1-approximations $\delta_{\mathbf{x}_j}^{\mathbf{p}_j}$, $j = 1, 2, 3$, namely

$$\mathbf{x}_1 = (-\tfrac{2}{3}, 0),\ \mathbf{p}_1 = (\tfrac{2}{9}, \tfrac{7}{9}), \quad \mathbf{x}_2 = (-\tfrac{1}{4}, \tfrac{1}{4}),\ \mathbf{p}_2 = (\tfrac{1}{2}, \tfrac{1}{2}), \quad \mathbf{x}_3 = (0, \tfrac{2}{3}),\ \mathbf{p}_3 = (\tfrac{7}{9}, \tfrac{2}{9}).$$

It is clear from $d_1(\delta_{\mathbf{x}_1}^{\mathbf{p}_1}, \mu) = d_1(\delta_{\mathbf{x}_3}^{\mathbf{p}_3}, \mu) = \frac{2}{9} < \frac{7}{24} = d_1(\delta_{\mathbf{x}_2}^{\mathbf{p}_2}, \mu)$ that the two (non-symmetric) probability measure $\delta_{\mathbf{x}_1}^{\mathbf{p}_1}$, $\delta_{\mathbf{x}_3}^{\mathbf{p}_3}$ are best 1-approximations of μ, whereas the (symmetric) $\delta_{\mathbf{x}_2}^{\mathbf{p}_2}$ is not. Similarly, for $r = 2$, Theorem 5.24 yields three candidates of which again only the two non-symmetric ones turn out to be best 2-approximations of μ.

Since $d_r(\delta_\bullet^{\bullet,n}, \mu) \le d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)$ for every $\mu \in \mathcal{P}_r$ and $n \in \mathbb{N}$, clearly $\lim_{n\to\infty} d_r(\delta_\bullet^{\bullet,n}, \mu) = 0$. The rate of convergence of $\big(d_r(\delta_\bullet^{\bullet,n}, \mu)\big)$ has been, and continues to be, studied extensively; see, e.g., [17,19–21,23,28] and the references therein. Arguably the simplest situation occurs if $\mu \in \mathcal{P}_r$ has a non-trivial absolutely continuous part and satisfies a mild moment condition. In this case, $\big(d_r(\delta_\bullet^{\bullet,n}, \mu)\big)$ decays like (n^{-1}) for every r.

Proposition 5.27 ([17, Thm. 6.2])**.** *Assume that $\mu \in \mathcal{P}_r$ for some $r \ge 1$. If $\mu \in \mathcal{P}_s$ with $s > r$ then*

$$\lim_{n\to\infty} n d_r(\delta_\bullet^{\bullet,n}, \mu) = \frac{1}{2(r+1)^{1/r}} \left(\int_{\mathbb{R}} \left(\frac{\mathrm{d}\mu_a}{\mathrm{d}\lambda} \right)^{\frac{1}{r+1}} \right)^{\frac{r+1}{r}},$$

where μ_a is the absolutely continuous part (w.r.t. λ) of μ.

It is instructive to compare Proposition 5.27 to Theorem 5.15. To do so, assume that $\mu \in \mathcal{P}_s$ for some $s > r$ and that μ^{-1} is absolutely continuous. Then $\lim_{n\to\infty} n d_r(\delta_\bullet^{\bullet,n}, \mu)$ and $\lim_{n\to\infty} n d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)$ both are finite and positive, provided that μ is non-singular and $\frac{\mathrm{d}\mu^{-1}}{\mathrm{d}\lambda} \in L^r(\mathbb{I})$. Thus $\big(d_r(\delta_\bullet^{\bullet,n}, \mu)\big)$ and $\big(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$ exhibit the same rate of decay, namely (n^{-1}). Note that while the latter rate is a universal *upper* bound on $\big(d_r(\delta_\bullet^{\bullet,n}, \mu)\big)$, at least under the mild assumption that $\mu \in \mathcal{P}_s$ for some $s > r$, it is a universal *lower* bound on $\big(d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)\big)$, by Theorem 5.20. Even if both sequences decay at the same rate, however, $\lim_{n\to\infty} n d_r(\delta_\bullet^{\bullet,n}, \mu) \le \lim_{n\to\infty} n d_r(\delta_\bullet^{\mathbf{u}_n}, \mu)$, and equality holds only if either $\mu = \frac{1}{\lambda(I)}\lambda|_I$ for some bounded, non-degenerate interval $I \subset \mathbb{R}$ or else $\mu = \delta_a$ for some $a \in \mathbb{R}$. Thus only in the trivial case of a (possibly degenerate) uniform distribution μ does $(\delta_\bullet^{\mathbf{u}_n})$ provide a sequence of asymptotically best r-approximations of μ (as defined below).

Example 5.28. Let μ be the exponential distribution of Example 5.2. For $r = 1$ and every $n \in \mathbb{N}$, (5.13) identifies a unique best 1-approximation $\delta_{\mathbf{x}_n}^{\mathbf{p}_n}$, with

$$x_{n,i} = -2\log\frac{n+1-i}{\sqrt{n(n+1)}}, \quad P_{n,i} = \frac{i(2n+1-i)}{n(n+1)}, \quad \forall i = 1, \ldots, n\,.$$

Here $\delta_\bullet^{\bullet,n}$ is unique, and $n d_1(\delta_\bullet^{\bullet,n}, \mu) = n\log(1+\frac{1}{n}) = 1+\mathcal{O}(n^{-1})$ as $n \to \infty$, in agreement with Proposition 5.27. For comparison, recall from Example 5.8 that $\lim_{n\to\infty} \frac{n}{\log n} d_1(\delta_\bullet^{\mathbf{u}_n}, \mu) = \frac{1}{4}$. For $r > 1$, no explicit expression seems to be known for $\delta_\bullet^{\bullet,n}$, not even when $r = 2$. However, in a sense made precise below, $(\delta_{\mathbf{y}_n}^{\widetilde{\mathbf{P}}_n})$ with

$$y_{n,i} = (r+1)\log\frac{n+1}{n-i+1}, \quad \widetilde{P}_{n,i} = 1 - \left(\frac{(n+1-i)(n-i)}{(n+1)^2}\right)^{\frac{r+1}{2}}, \quad \forall i = 1, \ldots, n$$

yields a sequence of asymptotically best r-approximations of μ for any $r > 1$.

Example 5.28 illustrates that even in very simple situations it may be difficult to compute $\delta_\bullet^{\bullet,n}$ explicitly. Not least from a computational point of view, therefore, it is natural to seek r-approximations that at least are optimal *asymptotically*. Specifically, call $(\delta_{\mathbf{x}_n}^{\mathbf{p}_n})$ with $\mathbf{x}_n \in \Xi_n$,

$\mathbf{p}_n \in \Pi_n$ for all $n \in \mathbb{N}$ a sequence of *asymptotically best r-approximations* of $\mu \in \mathcal{P}_r$ with $\#\,\mathrm{supp}\,\mu = \infty$, if

$$\lim_{n\to\infty} \frac{d_r(\delta_{\mathbf{x}_n}^{\mathbf{p}_n}, \mu)}{d_r(\delta_\bullet^{\bullet,n}, \mu)} = 1.$$

There exists a large literature on asymptotically best approximations. Specifically, mild conditions (such as $\mu \in \mathcal{P}_r$ being absolutely continuous with $\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}$ Hölder continuous and positive on $\mathring{\mathrm{supp}}\,\mu$, among others) have been established which guarantee that $(\delta_{\mathbf{x}_n}^\bullet)$ is a sequence of asymptotically best approximations of μ, where

$$x_{n,i} = F_{\mu_r}^{-1}\left(\frac{i}{n+1}\right), \quad \forall i = 1, \ldots, n, \tag{5.15}$$

with $\frac{\mathrm{d}\mu_r}{\mathrm{d}\lambda} = \frac{\frac{\mathrm{d}\mu}{\mathrm{d}\lambda}^{\frac{1}{r+1}}}{\int_{\mathbb{R}} \frac{\mathrm{d}\mu}{\mathrm{d}\lambda}^{\frac{1}{r+1}}}$; see, e.g., [24,31] and the references therein.

Example 5.29. Let $\mu = \mathrm{Beta}(2, 1)$ as in Example 5.4 and 5.14. While for arbitrary $n \in \mathbb{N}$, the authors do not know of an explicit expression for $\delta_\bullet^{\bullet,n}$ for any $r \geq 1$, (5.15) yields a sequence $(\delta_{\mathbf{x}_n}^{\mathbf{p}_n})$ of asymptotically best r-approximations of μ, with

$$x_{n,i} = \left(\frac{i}{n+1}\right)^{\frac{r+1}{r+2}}, \quad P_{n,i} = \frac{1}{4(n+1)^{\frac{2(r+1)}{r+2}}}\left(i^{\frac{r+1}{r+2}} + (i+1)^{\frac{r+1}{r+2}}\right)^2, \quad \forall i = 1, \ldots, n-1,$$

and $x_{n,n} = (\frac{n}{n+1})^{\frac{r+1}{r+2}}$. From this, a short calculation leads to, for instance,

$$n d_2(\delta_{\mathbf{x}_n}^{\mathbf{p}_n}, \mu) = \frac{3}{8\sqrt{2}} + \mathcal{O}(n^{-1}) \quad \text{as } n \to \infty,$$

which is consistent with Proposition 5.27.

If $\mu \in \mathcal{P}_s$ is singular then Proposition 5.27 only yields $\lim_{n\to\infty} n d_r(\delta_\bullet^{\bullet,n}, \mu) = 0$. The detailed analysis of $\big(d_r(\delta_\bullet^{\bullet,n}, \mu)\big)$ in this case is an active research area, for which already a substantial literature exists, notably for important classes of singular probabilities such as self-similar and -conformal measures; see, e.g., [17,18,20,29,30,32]. A key notion in this context is the so-called *quantization dimension* of $\mu \in \mathcal{P}_r$ of order r, defined as

$$D_r(\mu) = \lim_{n\to\infty} \frac{\log n}{-\log d_r(\delta_\bullet^{\bullet,n}, \mu)},$$

provided that this limit exists. For instance, Proposition 5.27 implies that $D_r(\mu) = 1$ whenever $\mu_a \neq 0$. The relations of $D_r(\mu)$ to various other concepts of dimension have attracted considerable attention [17,20,29,32].

Example 5.30. For the Cantor measure μ of Example 5.10, [22, Cor. 4.7 and Rem. 6.1] show that, for every $r > 1$,

$$0 < \liminf_{n\to\infty} n^{\log 3/\log 2} d_r(\delta_\bullet^{\bullet,n}, \mu) < \limsup_{n\to\infty} n^{\log 3/\log 2} d_r(\delta_\bullet^{\bullet,n}, \mu) < +\infty.$$

From this, it is clear that $D_r(\mu) = \frac{\log 2}{\log 3}$, which is independent of r and coincides with the Hausdorff dimension of $\mathrm{supp}\,\mu$.

Example 5.31. Let μ be the inverse Cantor measure of Example 5.16. Note that μ is not a self-similar, and hence the classical results for self-similar probabilities do not apply. Still, μ is the unique fixed point of a contraction on $\mathcal{P}_1$, namely $\nu \mapsto \frac{1}{3}(\nu \circ T_1^{-1} + \delta_{1/2} + \nu \circ T_2^{-1})$, with the similarities $T_1(x) = \frac{1}{2}x$ and $T_2(x) = \frac{1}{2}(1+x)$. This property enables a fairly complete analysis of $\big(d_r(\delta_\bullet^{\bullet,n}, \mu)\big)$ which the authors intend to present elsewhere. Specifically, with $\beta_r = (1 - \frac{1}{r}) + \frac{1}{r}\frac{\log 3}{\log 2}$, it can be shown that, for every $r \ge 1$, the numbers $\liminf_{n\to\infty} n^{\beta_r} d_r(\delta_\bullet^{\bullet,n}, \mu)$ and $\limsup_{n\to\infty} n^{\beta_r} d_r(\delta_\bullet^{\bullet,n}, \mu)$ both are finite and positive. In particular, $D_r(\mu) = \beta_r^{-1}$. Note that, unlike in the previous example, $D_r(\mu)$ depends on r, and $\frac{\log 2}{\log 3} \le D_r(\mu) < 1$. Thus $D_r(\mu)$ is larger than 0, the Hausdorff dimension of μ, but smaller than 1, the Hausdorff dimension of supp $\mu = \mathbb{I}$.

Proposition 5.27 guarantees that under a mild moment condition, $\big(d_r(\delta_\bullet^{\bullet,n}, \mu)\big)$ decays at least like $\big(n^{-1}\big)$, and in fact may decay faster, as Examples 5.30 and 5.31 illustrate. Even for purely atomic μ, however, the decay of $\big(d_r(\delta_\bullet^{\bullet,n}, \mu)\big)$ can be arbitrarily slow. This final observation, a refinement of [17, Ex.6.4], uses the following simple calculus fact; cf. also [3, Thm.3.3].

Proposition 5.32. *Given any sequence (a_n) of non-negative real numbers with $\lim_{n\to\infty} a_n = 0$, there exists a decreasing sequence (b_n) with $\lim_{n\to\infty} b_n = 0$ such that $(b_n - b_{n+1})$ is decreasing also, and $b_n \ge a_n$ for all n.*

Theorem 5.33. *Given $r \ge 1$ and any sequence (a_n) of non-negative real numbers with $\lim_{n\to\infty} a_n = 0$, there exists $\mu \in \mathcal{P}_r$ such that $d_r(\delta_\bullet^{\bullet,n}, \mu) \ge a_n$ for every $n \in \mathbb{N}$.*

Proof. In view of Proposition 5.32, assume w.o.l.g. that (a_n) and $(a_n^r - a_{n+1}^r)$ both are decreasing. Pick $a_0 > a_1$ such that $a_0^r - a_1^r > a_1^r - a_2^r$, and let $c_r = \sum_{k=1}^\infty 2^{-(k-1)r}\,(a_{k-1}^r - a_k^r)$. Note that c_r is finite and positive. Consider $\mu = \sum_{k=1}^\infty p_k \delta_{x_k}$, where $p_k = c_r^{-1} 2^{-(k-1)r}(a_{k-1}^r - a_k^r)$ and $x_k = 3 \cdot 2^{k-1} c_r^{1/r}$ for all $k \in \mathbb{N}$. Since $\sum_{k=1}^\infty p_k x_k^r = 3^r a_0^r < +\infty$, clearly $\mu \in \mathcal{P}_r$. For every $n \in \mathbb{N}$, define $K_n \subset \mathbb{N}$ as

$$K_n = \{k \in \mathbb{N} :\ \operatorname{supp} \delta_\bullet^{\bullet,n}\ \cap\ [2^k c_r^{1/r}, 2^{k+1} c_r^{1/r}[\, = \varnothing\}\,.$$

Since # supp $\delta_\bullet^{\bullet,n} \le n$ and the intervals $[2^k c_r^{1/r}, 2^{k+1} c_r^{1/r}[$, $k \in \mathbb{N}$, are disjoint, # $(\mathbb{N} \setminus K_n) \le n$. Moreover,

$$\min_{y \in \operatorname{supp} \delta_\bullet^{\bullet,n}} |x_k - y|^r \ge 2^{(k-1)r} c_r \ \text{ for every } k \in K_n\,.$$

Recall from [7, (ii), p.1847] that $d_r(\delta_\bullet^{\bullet,n}, \mu)^r = \int_{\mathbb{R}} \min_{y \in \operatorname{supp}\, \delta_\bullet^{\bullet,n}} |x - y|^r \mathrm{d}\mu(x)$; see also [17, Lemma 3.1]. It follows that, for every $n \in \mathbb{N}$,

$$d_r(\delta_\bullet^{\bullet,n}, \mu)^r = \sum_{k=1}^\infty p_k \min_{y \in \operatorname{supp}\ \delta_\bullet^{\bullet,n}} |x_k - y|^r \ge \sum_{k \in K_n} p_k 2^{(k-1)r} c_r = \sum_{k \in K_n} (a_{k-1}^r - a_k^r).$$

Moreover, recall that $(a_{n-1}^r - a_n^r)$ is decreasing, and # $(\mathbb{N} \setminus K_n) \le n$. Thus

$$d_r(\delta_\bullet^{\bullet,n}, \mu)^r \ge \sum_{k=n+1}^\infty (a_{k-1}^r - a_k^r) = a_n^r,$$

and hence $d_r(\delta_\bullet^{\bullet,n}, \mu) \ge a_n$ for every $n \in \mathbb{N}$. $\square$

Acknowledgments

The first author was supported in part by a Pacific Institute for the Mathematical Sciences (PIMS) Graduate Scholarship and a Josephine Mitchell Graduate Scholarship, both at the University of Alberta. The second author was supported by an NSERC Discovery Grant. Both authors are grateful to F. Dai for helpful conversations. They also like to thank an anonymous referee for bringing to their attention the important reference [25], and for providing several excellent comments that helped improve the exposition.

References

[1] D.M. Baker, Quantizations of probability measures and preservation of the convex order, Statist. Probab. Lett. 107 (2015) 280–285.

[2] A. Berger, T. Hill, K.E. Morrison, Scale-distortion inequalities for mantissas of finite data sets, J. Theor. Probab. 21 (2008) 97–117.

[3] S.G. Bobkov, M. Ledoux, One-dimensional empirical measures, order statistics and kantorovich transport distances, Mem. Amer. Math. Soc. (2016) Preprint http://www-users.math.umn.edu/~bobko001/preprints/2016_BL_Order.statistics_Revised.version.pdf (in press).

[4] E. Boissard, T.L. Gouic, On the mean speed of convergence of empirical and occupation measures in wasserstein distance, Ann. Inst. Henri Poincare Probab. Statist. 50 (2014) 539–563.

[5] G. Bouchitté, C. Jimenez, R. Mahadevan, Asymptotic analysis of a class of optimal location problems, J. Math. Pures Appl. 95 (2011) 382–419.

[6] C.A. Cabrelli, U.M. Molter, The kantorovich metric for probability measures on the circle, J. Comput. Appl. Math. 57 (1995) 345–361.

[7] E. Caglioti, F. Golse, M. Iacobelli, A gradient flow approach to quantization of measures, Math. Models Methods Appl. Sci. 25 (2015) 1845–1885.

[8] J. Chevallier, Uniform decomposition of probability measures: quantization, classification, clustering and rate of convergence, J. Appl. Probab. 55 (2018) 1037–1045.

[9] S. Dereich, M. Scheutzow, R. Schottstedt, Constructive quantization: approximation by empirical measures, Ann. Inst. Henri Poincare Probab. Statist. 49 (2013) 1183–1203.

[10] S. Dereich, C. Vormoor, The high resolution vector quantization problem with orlicz norm distortion, J. Theor. Probab. 24 (2011) 517–544.

[11] J. Dick, F. Pillichshammer, Digital Nets and Sequences, Discrepancy Theory and Quasi-Monte Carlo Integration, Cambridge University Press, 2010.

[12] R. Dudley, Real Analysis and Probability, Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, 2004.

[13] N. Fournier, A. Guillin, On the rate of convergence in wasserstein distance of the empirical measure, Probab. Theory Related Fields 162 (2015) 707–738.

[14] M.B. Giles, M. Hefter, L. Mayer, K. Ritter, Random bit multilevel algorithms for stochastic differential equations, J. Complexity (2019) (in press).

[15] M.B. Giles, M. Hefter, L. Mayer, K. Ritter, Random bit quadrature and approximation of distributions on hilbert spaces, Found. Comput. Math. 19 (2019) 205–238.

[16] P. Glasserman, Monte carlo methods in financial engineering, in: Applications of Mathematics 53, Stochastic Modelling and Applied Probability, Springer, 2004.

[17] S. Graf, H. Luschgy, Foundations of Quantization for Probability Distributions, in: Lecture Notes in Mathematics, vol. 1730, Springer, 2000.

[18] S. Graf, H. Luschgy, The quantization dimension of self-similar probabilities, Math. Nachr. 241 (2002) 103–109.

[19] S. Graf, H. Luschgy, G. Pagès, The local quantization behavior of absolutely continuous probabilities, Ann. Probab. 40 (2012) 1795–1828.

[20] M. Kesseböhmer, S. Zhu, Some recent developments in quantization of fractal measures, in: C. Bandt, K. Falconer, M. Zähle (Eds.), Fractal Geometry and Stochastics V, in: Prog. Probab., 70, 2015, pp. 105–120.

[21] B. Kloeckner, Approximation by finitely supported measures, ESAIM Control Optim. Calc. Var. 18 (2012) 343–359.

[22] W. Kreitmeier, Optimal quantization for dyadic homogeneous cantor distributions, Math. Nachr. 281 (2008) 1307–1327.

[23] W. Kreitmeier, Asymptotic optimality of scalar gersho quantizers, Constr. Approx. 38 (2013) 365–396.

[24] T. Linder, On asymptotically companding quantization, Probl. Control. Inf. 20 (1991) 475–484.

[25] S. Mak, V.R. Joseph, Support points, Ann. Statist. 46 (2018) 2562–2592.

[26] G. Pagès, A space quantization method for numerical integration, J. Comput. Appl. Math. 89 (1997) 1–38.

[27] G. Pagès, H. Pham, J. Printems, Optimal quantization methods and applications to numerical problems in finance, in: Handbook of Computational and Numerical Methods in Finance, Birkhäuser, 2004, pp. 253–297.

[28] G. Pagès, A. Sagna, Asymptotics of the maximal radius of an L^r-optimal sequence of quantizers, Bernoulli 18 (2012) 360–389.

[29] K. Pötzelberger, The quantization dimension of distributions, Math. Proc. Cambridge Philos. Soc. 131 (2001) 507–519.

[30] K. Pötzelberger, The quantization error of self-similar distributions, Math. Proc. Cambridge Philos. Soc. 137 (2004) 725–740.

[31] K. Pötzelberger, K. Felsenstein, An asymptotic result on principal points for univariate distributions, Optimization 28 (1994) 397–406.

[32] S.T. Rachev, L. Rüschendorf, Mass Transportation Problems, 1 and 2, Springer, New York, 1998.