# SHARP LOGARITHMIC ESTIMATES FOR POSITIVE DYADIC SHIFTS

ADAM OSĘKOWSKI

ABSTRACT. The paper contains the study of sharp logarithmic estimates for positive dyadic shifts $\mathcal{A}$ given on probability spaces $(X,\mu)$ equipped with a tree-like structure. For any $K>0$ we determine the smallest constant $L=L(K)$ such that

$$\int_E |\mathcal{A}f|\mathrm{d}\mu \leq K\int_{\mathbb{R}}\Psi(|f|)\mathrm{d}\mu + L(K)\cdot\mu(E),$$

where $\Psi(t)=(t+1)\log(t+1)-t$, $E$ is an arbitrary measurable subset of $X$ and $f$ is an integrable function on $X$. The proof exploits Bellman function method: we extract the above estimate from the existence of an appropriate special function, enjoying certain size and concavity-type conditions. As a corollary, a dual exponential bound is obtained.

## 1. INTRODUCTION

Let $Q\subset\mathbb{R}^d$ be a given dyadic cube and let $\mathcal{D}(Q)$ stand for the grid of its dyadic subcubes. For a given sequence $\alpha=(\alpha_R)_{R\in\mathcal{D}(Q)}$ of nonnegative numbers, define its Carleson constant by

$$\mathrm{Carl}(\alpha)=\sup_{R\in\mathcal{D}(Q)}\frac{1}{|R|}\sum_{R'\in\mathcal{D}(R)}\alpha_{R'}|R'|,$$

where $|A|$ is the Lebesgue measure of $A$. For any such sequence, we introduce the associated dyadic shift $\mathcal{A}$ which acts on integrable functions $f:Q\to\mathbb{R}$ by the formula

$$\mathcal{A}f=\sum_{R\in\mathcal{D}(Q)}\alpha_R\langle f\rangle_R\chi_R, \tag{1.1}$$

where $\langle f\rangle_R=\frac{1}{|R|}\int_R f\mathrm{d}\mu$ is the average of $f$ over $R$.

The class of positive dyadic shifts arose in the works of A. Lerner during his study of the $A_2$ theorem. Let us discuss this issue in a little more detailed manner. Assume that $T$ is a Calderón-Zygmund operator on $\mathbb{R}^d$ and let $w:\mathbb{R}^d\to(0,\infty)$ be a weight satisfying Muckenhoupt's condition $A_2$. The so-called $A_2$ conjecture asked for the linear dependence of the norm $||T||_{L^2(w)\to L^2(w)}$ on $[w]_{A_2}$, the $A_2$ characterstic of $w$:

$$||Tf||_{L^2(w)}\leq C(T,d)[w]_{A_2}||f||_{L^2(w)}.$$

This question has gained a lot of interest in the recent literature (see e.g. [1, 4, 11, 16, 18, 19, 20, 21, 26]) and was finally answered in the positive by Hÿtonen

2010 *Mathematics Subject Classification.* Primary: 42B25. Secondary: 46E30, 60G42.
*Key words and phrases.* Dyadic shift, Bellman function, best constants.
Research supported by Narodowe Centrum Nauki (Poland), grant DEC-2014/14/E/ST1/00532.

[7], with the use of clever representation of $T$ as an average of good dyadic shifts. Later, Lerner [12] provided a simpler proof of the $A_2$ theorem, which avoided the use of most of the complicated techniques in [7]. The idea was to exploit a general pointwise estimate for $T$ in terms of positive dyadic operators, proven in [10]. This allowed to reduce the $A_2$ problem to a weighted result for the positive dyadic shifts, which had been already shown before in [9] (consult also [4] and [5]). The aforementioned pointwise bound states that for every dyadic cube $Q$,

$$|Tf(x)| \lesssim \sum_{m=0}^{\infty} 2^{-\delta m} \mathcal{A}_{\mathcal{S}}^m |f|(x) \qquad \text{for a.e. } x \in Q, \tag{1.2}$$

where $\delta > 0$ depends on the operator $T$, $\mathcal{S}$ is a collection of dyadic cubes which depends on $f$, $T$ and $m$, and $\mathcal{A}_{\mathcal{S}}^m$ are positive dyadic operators defined by

$$\mathcal{A}_{\mathcal{S}}^m f(x) = \sum_{Q \in \mathcal{S}} \langle f \rangle_{Q^{(m)}} \chi_Q(x),$$

where $Q^{(m)}$ is the $m$-th dyadic parent of $Q$. The collections $\mathcal{S}$ used above are assumed to be *sparse*: for all cubes $Q \in \mathcal{S}$ there exists measurable subsets $E(Q) \subset Q$ with $|E(Q)| \geq |Q|/2$ and $E(Q) \cap E(Q') = \emptyset$ unless $Q = Q'$. Coming back to (1.2), Lerner proves that the operator norm of each $\mathcal{A}_{\mathcal{S}}^m$ is appropriately controlled by the operator norm of $\mathcal{A}_{\mathcal{S}'}^0$, with $\mathcal{S}'$ running over the class of all possible sparse collections. That is, he shows that for any Banach function space $X$,

$$||\mathcal{A}_{\mathcal{S}}^m||_X \lesssim (m+1) \sup_{\mathfrak{D}, \mathcal{S}'} ||\mathcal{A}_{\mathcal{S}'}^0 f||_X,$$

the supremum taken over all dyadic grids $\mathfrak{D}$ and all sparse collections $\mathcal{S}' \subset \mathfrak{D}$.

Hence, any appropriate bound for $\mathcal{A}_{\mathcal{S}}^0$ yields the corresponding statement for the class of Calderón-Zygmund operators. This motivates the question about controlling various norms of $\mathcal{A}_{\mathcal{S}}^0$ efficiently. This class of operators, after an appropriate localization, is contained in the class (1.1) considered at the beginning of the paper. Indeed, suppose that $\mathcal{S} \in \mathfrak{D}$ is a sparse family, let us restrict ourselves to a large cube $Q \in \mathfrak{D}$ and set $\alpha_R = \chi_{\{R \in \mathcal{S}\}}$. Then $(\alpha_R)_{R \in \mathcal{D}(Q)}$ is a Carleson sequence of constant bounded by 2: for any $R \in \mathcal{D}(Q)$, we have

$$\sum_{R' \in \mathcal{D}(R)} \alpha_{R'} |R'| = \sum_{R' \in \mathcal{S},\, R' \subseteq R} |R'| \leq 2 \sum_{R' \in \mathcal{S},\, R' \subseteq R} |E(R')| \leq 2|R|,$$

since the sets $(E(R'))_{R' \in \mathcal{S}}$ are pairwise disjoint.

From now on, we will focus on the dyadic shifts of the form (1.1). It is not difficult to show, using sharp estimates for the dyadic maximal operator and the multisublinear maximal function (cf. [6, 8]), that any $\mathcal{A}$ from this class is bounded on $L^p(Q)$, $1 < p < \infty$. More precisely, we have

$$||\mathcal{A}f||_{L^p(Q)} \leq \frac{p^2}{p-1} \operatorname{Carl}(\alpha) ||f||_{L^p(Q)} \tag{1.3}$$

and the multiplicative constant $p^2/(p-1)$ cannot be improved. For $p = 1$ the $L^p$-boundedness does not hold, but, as proved by Rey and Reznikov [22] (for $d = 1$ only), we have the sharp weak-type bound

$$|\{x \in Q : \mathcal{A}f(x) \geq 1\}| \leq 2 \operatorname{Carl}(\alpha) ||f||_{L^1(Q)}.$$

In this paper, we will establish a related LlogL bound, which can be regarded as another natural substitute for the $L^1$ estimate, and the dual exponential inequality which serves as a version of (1.3) for $p = \infty$. Actually, instead of working with the dyadic lattice in $\mathbb{R}^d$, we will study these estimates in the more general context of probability measures equipped with a tree-like structure. Here is the precise definition.

**Definition 1.1.** Suppose that $(X, \mu)$ is a nonatomic probability space. A set $\mathcal{T}$ of measurable subsets of $X$ will be called a tree if the following conditions are satisfied:

(i) $X \in \mathcal{T}$ and for every $Q \in \mathcal{T}$ we have $\mu(Q) > 0$.

(ii) For every $Q \in \mathcal{T}$ there is a finite subset $C(Q) \subset \mathcal{T}$ containing at least two elements such that
  (a) the elements of $C(Q)$ are pairwise disjoint subsets of $Q$,
  (b) $Q = \bigcup C(Q)$.

(iii) $\mathcal{T} = \bigcup_{m \geq 0} \mathcal{T}^m$, where $\mathcal{T}^0 = \{X\}$ and $T^{m+1} = \bigcup_{Q \in \mathcal{T}^m} C(Q)$.

(iv) We have $\lim_{m \to \infty} \sup_{Q \in \mathcal{T}^m} \mu(Q) = 0$.

The normalization imposed in the above definition (which restricts us to probability spaces) will not affect sharp constants in the estimates under investigation, which can be seen by applying standard dilation and scaling arguments. For any sequence $\alpha = (\alpha_Q)_{Q \in \mathcal{T}}$ of nonnegative numbers, we define its Carleson constant by

$$\mathrm{Carl}(\alpha) = \sup_{Q \in \mathcal{T}} \frac{1}{\mu(Q)} \sum_{R \subseteq Q,\, R \in \mathcal{T}} \alpha_R \mu(R),$$

and the associated shift is given by the formula

$$\mathcal{A}f = \sum_{Q \in \mathcal{T}} \alpha_Q \langle f \rangle_Q \chi_Q.$$

Here, as above, $\langle f \rangle_Q = \frac{1}{\mu(Q)} \int_Q f \mathrm{d}\mu$ is the average of $f$ over $Q$.

The main emphasis in the paper is put on inequalities of the form

$$\int_E |\mathcal{A}f| \mathrm{d}\mu \leq K \int_X \Psi(|f|) \mathrm{d}\mu + L(K) \cdot \mu(E). \tag{1.4}$$

Here $\mathcal{A}$ is the shift associated with some sequence $(\alpha_Q)_{Q \in \mathcal{T}}$ with Carleson constant not exceeding 1, $E$ is a measurable subset of $X$, $f$ is an integrable function on $X$ and $\Psi$ stands for the "LlogL" function $\Psi(t) = (t+1)\log(t+1) - t$, $t \geq 0$. There are two natural questions about (1.4) to be asked:

(i) For which $K$ is there a finite number $L(K)$ such that the inequality holds for all $f$ and $E$?

(ii) For $K$ as in (i), what is the optimal (i.e., the least) value of $L(K)$ allowed?

Both these questions are answered in the theorem below.

**Theorem 1.2.** *For any $K > 0$, the optimal value of the constant $L(K)$ in* (1.4) *is equal to $+\infty$ if $K \leq 1$, and*

$$\frac{K}{K-1} \exp(K^{-1}) - 1 + \int_0^1 (\exp(u/K) - 1) \frac{du}{u}$$

*if $K > 1$.*

Some important observations are in order. First, we may restrict ourselves to nonnegative functions $f$; indeed, the passage from $f$ to $|f|$ does not change the right-hand side of (1.4) and does not decrease the left-hand side. Second, our use of somewhat special LlogL function $\Psi$ is forced by the fact that for the standard functions $t \mapsto t\log t$ or $t \mapsto t\log^+ t$ the above estimate does not hold (for any $K$ one would have to take $L(K)=\infty$). Indeed, otherwise the shifts would be bounded on $L^\infty$: we would have $\int_E \mathcal{A}f \mathrm{d}\mu \leq L(K)\cdot \mu(E)$ for any $f: X\to[0,1]$ and any $E$, which is not the case (see e.g. Theorem 1.3 below).

As a corollary, we obtain the following dual exponential bound. Throughout the paper, $\Phi:[0,\infty)\to[0,\infty)$ is the Young-conjugate to $\Psi$, given by $\Phi(t)=e^t-1-t$.

**Theorem 1.3.** *Let $\mathcal{A}$ be the shift associated with some sequence $(\alpha_Q)_{Q\in\mathcal{T}}$ with Carleson constant not exceeding 1. For any $K>0$ and any $f: X\to[-1,1]$, we have*

$$\int_X \Phi(|\mathcal{A}f|/K)\,d\mu \leq \frac{L(K)}{K}\mu(\{f\neq 0\}). \tag{1.5}$$

*The constant $L(K)/K$ is the best possible for each $K$.*

We should point out here that the constants $L(K)$ and $L(K)/K$ are best possible for *each* probability space $(X,\mu)$ equipped with some tree $\mathcal{T}$. Therefore, in particular, these constants are also optimal in the dyadic context studied at the beginning.

Our approach rests on the so-called Bellman function method. This technique allows to deduce the validity of the LlogL estimate from the existence of a certain special function, which enjoys appropriate size conditions and concavity. This method has its origins in the theory of stochastic optimal control, and, as observed by Burkholder in the eighties during his study on the boundedness properties of the Haar system, it can be used in the investigation of various problems for semimartingales. Following the seminal work [2], Burkholder and others managed to establish a variety of interesting estimates in this probabilistic context (see the monograph [17] for the details). A decisive step towards further applications of the method in harmonic analysis was made by Nazarov, Treil and Volberg [14, 15], who put the approach in a more modern and universal form. Since then, the method has been applied in numerous problems arising in various areas of mathematics (cf. e.g. [19, 24, 25, 27, 28] and consult references therein).

The rest of the paper is organized as follows. In the next section we present an informal reasoning which leads to the discovery of a Bellman function corresponding to our problem. Then, in Section 3, we exploit rigorously the properties of this object to prove the inequalities (1.4) and (1.5). In the final part of the paper we exploit further properties of the Bellman function to show that the constants in the estimates (1.4) and (1.5) cannot be improved.

## 2. A related Bellman function

Throughout this section, we assume that the probability space $(X,\mathcal{T},\mu)$ is the interval $[0,1)$ equipped with Lebesgue's measure and the tree of its dyadic subintervals. We will also use the notation $\mathcal{T}^-=\{Q\in\mathcal{T}: Q\subseteq[0,1/2)\}$ and $\mathcal{T}^+=\{Q\in\mathcal{T}: Q\subseteq[1/2,1)\}$. We start the analysis with introducing the abstract Bellman function $\mathbb{B}:[0,1]\times[0,1]\times[0,\infty)\to\mathbb{R}$ associated with (1.4). This function is given by

$$\mathbb{B}(s,t,x)=\sup\left\{\frac{1}{\mu(R)}\int_E\left(\sum_{Q\subset R,\,Q\in\mathcal{T}}\alpha_Q\langle f\rangle_Q\chi_Q\right)\mathrm{d}\mu-\frac{K}{\mu(R)}\int_R\Psi(f)\mathrm{d}\mu\right\}.$$

Here $R$ is a given dyadic subinterval of $[0,1)$ and the supremum is taken over all measurable sets $E \subseteq X$ satisfying $\langle \chi_E \rangle_R = s$, all sequences $\alpha = (\alpha_Q)_{Q \subseteq R,\, Q \in \mathcal{T}}$ of nonnegative numbers with $\mathrm{Carl}(\alpha) \leq 1$ satisfying

$$\frac{1}{\mu(R)} \sum_{Q \subseteq R} \alpha_Q \mu(Q) = t, \tag{2.1}$$

and all integrable functions $f : R \to [0,\infty)$ satisfying $\langle f \rangle_R = x$. From the formal point of view, the function $\mathbb{B}$ depends also on $R$, however, this is not the case. Indeed, for any two dyadic intervals $R_1$ and $R_2$, an affine mapping of one interval onto another puts the Carleson sequences satisfying (2.1) in one-to-one correspondence, and such a change of the variable preserves the averages. On the other hand, as we will see below, working with different domains $R$ is crucial for the understanding of properties of the Bellman function.

There are two natural questions arising. The first concerns the explicit formula for $\mathbb{B}$, and the second is about the extremizers for a given $(s,t,x)$ (i.e., those $E$, $(\alpha_Q)_Q$ and $f$, for which the supremum defining $\mathbb{B}(s,t,x)$ is attained, or at least almost attained). Clearly, the answer to the first question is all we need: having found $\mathbb{B}$, it remains to take $R = [0,1]$ and determine the least constant $L(K)$ such that $\mathbb{B}(s,t,x) \leq L(K)s$ for all $s$, $t$ and $x$. To identify $\mathbb{B}$, one might try to answer the second question, guessing some structural, fractal-type properties of the extremizers. However, in a sense, we will exploit the implications between the two questions in both directions. Our plan is the following. We will consider a slightly different Bellman function (whose identification is a little easier), then extract the extremizers from its explicit formula, and finally postulate that these extremizers coincide with those corresponding to the initial $\mathbb{B}$. Even though this conjecture is not true (the extremizers coincide only on a part of the domain), plugging them into the formula for $\mathbb{B}$ will return the right Bellman function, which allows a successful treatment of (1.4).

The usual first step in the search for the formula is to exploit some structural, homogeneity-type properties of the Bellman function which follow directly from its abstract definition. Consider a slightly different Bellman function $\overline{\mathbb{B}} : [0,1] \times [0,1] \times [-1,\infty) \to \mathbb{R}$ defined by

$$\overline{\mathbb{B}}(s,t,x) = \sup \left\{ \frac{1}{\mu(R)} \int_E \left( \sum_{Q \in \mathcal{T}} \alpha_Q (\langle f \rangle_Q + 1) \chi_Q \right) \mathrm{d}\mu - \frac{K}{\mu(R)} \int_R \big( \Psi(f) + f \big) \mathrm{d}\mu \right\},$$

where the supremum is taken over the same parameters as previously, with one crucial change. Namely, we slightly enlarge the allowed class of functions, assuming that they take values in $[-1,\infty)$. This new function $\overline{\mathbb{B}}$ has a nice structural property studied in the lemma below; this will allow us to reduce the number of variables to two and then solve the underlying Monge-Ampère equation.

**Lemma 2.1.** *For any $s$, $t \in [0,1]$ and $x \geq -1$ we have*

$$\overline{\mathbb{B}}(s,t,x) = (x+1)\overline{\mathbb{B}}(s,t,0) - K(x+1)\log(x+1).$$

*Proof.* If $x = -1$, then in the definition of $\overline{\mathbb{B}}(s,t,x)$ we must take $f \equiv -1$ and then the expression under the supremum vanishes. Therefore $\overline{\mathbb{B}}(s,t,x) = 0$ and the equality holds. Suppose then that $x > -1$. Take $E$, $\alpha = (\alpha_Q)_{Q \in \mathcal{T}}$ and $f : [0,1] \to [-1,\infty)$ as in the definition of $\overline{\mathbb{B}}(s,t,x)$. Fix a parameter $\lambda > 0$ and consider the function $g:[0,1]\to[-1,\infty)$ given by $g=(f+1)/\lambda-1$. Then we have $\langle g\rangle_Q+1=(x+1)/\lambda$ and

$$\begin{aligned}\int_{[0,1]}(\Psi(f)+f)\mathrm{d}\mu&=\int_{[0,1]}(f+1)\log(f+1)\mathrm{d}\mu\\&=\lambda\int_{[0,1]}(g+1)\log(\lambda(g+1))\mathrm{d}\mu\\&=\lambda\int_{[0,1]}(\Psi(g)+g)\mathrm{d}\mu+\lambda\log\lambda\int_{[0,1]}(g+1)\mathrm{d}\mu,\end{aligned}$$

so

$$\begin{aligned}&\int_E\bigg(\sum_{Q\in\mathcal{T}}\alpha_Q(\langle f\rangle_Q+1)\chi_Q\bigg)\mathrm{d}\mu-K\int_{[0,1]}\big(\Psi(f)+f\big)\mathrm{d}\mu\\&=\lambda\left[\int_E\bigg(\sum_{Q\in\mathcal{T}}\alpha_Q(\langle g\rangle_Q+1)\chi_Q\bigg)\mathrm{d}\mu-K\int_{[0,1]}\big(\Psi(g)+g\big)\mathrm{d}\mu\right]-K(x+1)\log\lambda\\&\leq\lambda\overline{\mathbb{B}}(s,t,(x+1)/\lambda-1)-K(x+1)\log\lambda.\end{aligned}$$

Since $E$, $(\alpha_Q)_{Q\in\mathcal{T}}$ and $f$ were arbitrary, we conclude that

$$\overline{\mathbb{B}}(s,t,x)\leq\lambda\overline{\mathbb{B}}(s,t,(x+1)/\lambda-1)-K(x+1)\log\lambda. \tag{2.2}$$

Now, plug $y=(x+1)/\lambda-1$ and $\gamma=\lambda^{-1}$ to get

$$\overline{\mathbb{B}}(s,t,(y+1)/\gamma-1)\leq\gamma^{-1}\overline{\mathbb{B}}(s,t,y)+K\gamma^{-1}(y+1)\log\gamma^{-1},$$

which, after renaming $x:=y$, $\lambda:=\gamma$, becomes the reverse to (2.2). Consequently, this inequality is actually an equality, and plugging $\lambda=x+1$ yields the assertion. □

Next, we will show that $\overline{\mathbb{B}}$ satisfies the concavity property.

**Lemma 2.2.** *Suppose that $s_\pm$, $t_\pm\in[0,1]$, $x_\pm\in[-1,\infty)$. Then*

$$\overline{\mathbb{B}}\left(\frac{s_-+s_+}{2},\frac{t_-+t_+}{2},\frac{x_-+x_+}{2}\right)\geq\frac{\overline{\mathbb{B}}(s_-,t_-,x_-)+\overline{\mathbb{B}}(s_+,t_+,x_+)}{2}.$$

*Proof.* Take $E^\pm$, $(\alpha_Q^\pm)_Q$ and $f^\pm$ as in the definitions of $\overline{\mathbb{B}}(s_\pm,t_\pm,x_\pm)$. Furthermore, as we have discussed above, for any $(s,t,x)$ we have some freedom in selecting the underlying dyadic interval $R$ "supporting" $\overline{\mathbb{B}}(s,t,x)$. We assume that $\overline{\mathbb{B}}(s_-,t_-,x_-)$ is supported on $[0,1/2)$ and $\overline{\mathbb{B}}(s_+,t_+,x_+)$ is supported on $[1/2,1)$ (so $E^-$ is a subset of $[0,1/2)$, $(\alpha_Q^-)_{Q\in\mathcal{T}^-}$ is a Carleson sequence indexed by dyadic subsets of $[0,1/2)$ and $f$ is supported on $[0,1/2)$; analogous statements, with $[0,1/2)$ replaced by $[1/2,1)$, hold for $E^+$, $(\alpha_Q^+)_{Q\in\mathcal{T}^+}$ and $f^+$). Let us glue these objects as follows: put $E=E^-\cup E^+$, let $(\alpha_Q)_{Q\in\mathcal{T}}$ be given by

$$\alpha_Q=\begin{cases}0 & \text{if } Q=[0,1),\\ \alpha_Q^- & \text{if } Q\in\mathcal{T}^-,\\ \alpha_Q^+ & \text{if } Q\in\mathcal{T}^+,\end{cases}$$

and define $f:[0,1)\to[-1,\infty)$ by $f=f^-\chi_{[0,1/2)}+f^+\chi_{[1/2,1)}$. We easily check that $\mu(E)=(s_++s_-)/2$, $(\alpha_Q)_{Q\in\mathcal{T}}$ has Carleson constant less or equal to 1 and satisfies $\sum_Q \alpha_Q \mu(Q) = (t_- + t_+)/2$, and $\langle f\rangle_{[0,1)} = (x_- + x_+)/2$. Therefore,

$$
\begin{aligned}
&\overline{\mathbb{B}}\left(\frac{s_- + s_+}{2}, \frac{t_- + t_+}{2}, \frac{x_- + x_+}{2}\right)\\
&\geq \int_E \Big(\sum_{Q\in\mathcal{T}} \alpha_Q(\langle f\rangle_Q + 1)\chi_Q\Big)\mathrm{d}\mu - K\int_{[0,1]} \big(\Psi(f) + f\big)\mathrm{d}\mu\\
&= \int_{E^-} \Big(\sum_{Q\in\mathcal{T}^-} \alpha_Q^-(\langle f^-\rangle_Q + 1)\chi_Q\Big)\mathrm{d}\mu - K\int_{[0,1/2)} \big(\Psi(f^-) + f^-\big)\mathrm{d}\mu\\
&\quad + \int_{E^+} \Big(\sum_{Q\in\mathcal{T}^+} \alpha_Q^+(\langle f^+\rangle_Q + 1)\chi_Q\Big)\mathrm{d}\mu - K\int_{[1/2,1)} \big(\Psi(f^+) + f^+\big)\mathrm{d}\mu.
\end{aligned}
$$

Taking the supremum over all $E^\pm$, $(\alpha_Q^\pm)_{Q\in\mathcal{T}^\pm}$ and $f^\pm$, we get the desired claim. □

**Remark 2.3.** Let $(s,t,x) \in [0,1]\times[0,1]\times[-1,\infty)$ be a fixed point. Suppose that $E \subseteq [0,1)$, $(\alpha_Q)_{Q\in\mathcal{T}}$ and $f : [0,1) \to [-1,\infty)$ are the extremizers of $\overline{\mathbb{B}}\,(s,t,x)$, so

$$\overline{\mathbb{B}}\,(s,t,x) = \int_E \Big(\sum_{Q\in\mathcal{T}} \alpha_Q(\langle f\rangle_Q + 1)\chi_Q\Big)\mathrm{d}\mu - K\int_{[0,1]} \big(\Psi(f) + f\big)\mathrm{d}\mu.$$

Next, define $s_- = 2\mu(E\cap[0,1/2))$, $s_+ = 2\mu(E\cap[1/2,1))$, $t_- = 2\sum_{Q\in\mathcal{T}^-}\alpha_Q\mu(Q)$, $t_+ = 2\sum_{Q\in\mathcal{T}^+}\alpha_Q\mu(Q)$ and $f^- = f|_{[0,1/2)}$, $f^+ = f|_{[1/2,1)}$. If $\alpha_{[0,1]} = 0$, then we have

$$
\begin{aligned}
&\overline{\mathbb{B}}\,(s,t,x)\\
&= \int_{E\cap[0,1/2)} \Big(\sum_{Q\in\mathcal{T}^-} \alpha_Q(\langle f\rangle_Q + 1)\chi_Q\Big)\mathrm{d}\mu - K\int_{[0,1/2)} \big(\Psi(f) + f\big)\mathrm{d}\mu\\
&\quad + \int_{E\cap[1/2,1)} \Big(\sum_{Q\in\mathcal{T}^+} \alpha_Q(\langle f\rangle_Q + 1)\chi_Q\Big)\mathrm{d}\mu - K\int_{[1/2,1)} \big(\Psi(f) + f\big)\mathrm{d}\mu\\
&\leq \frac{\overline{\mathbb{B}}(s_-,t_-,x_-) + \overline{\mathbb{B}}(s_+,t_+,x_+)}{2}
\end{aligned}
$$

and hence, in the light of the lemma above, we actually have equality here. So, if $(s_-,t_-,x_-) \neq (s,t,x)$ and $\alpha_{[0,1]} = 0$ (which, as we might hope, holds for most points $(s,t,x)$), then there is a line segment passing through $(s,t,x)$ along which $\overline{\mathbb{B}}$ is linear. This observation will be of fundamental importance in our search for the explicit formula for $\overline{\mathbb{B}}$.

Finally, we will need the following condition.

**Lemma 2.4.** *For any $s \in [0,1]$, $0 \leq t < t+\delta \leq 1$ and $x \in [-1,\infty)$ we have*

$$\overline{\mathbb{B}}(s,t+\delta,x) \geq \overline{\mathbb{B}}(s,t,x) + \delta(x+1)s.$$

*Proof.* We proceed as in the previous lemma, with $s_\pm = s$, $t_\pm = t$ and $x_\pm = x$, and take the appropriate parameters $E^\pm$, $(\alpha_Q^\pm)_{Q\in\mathcal{T}^\pm}$ and $f^\pm$. Then $E^\pm$ and $f^\pm$ are glued into one set and one function with the use of the same procedure; however, we construct a slightly different Carleson sequence, setting

$$\alpha_Q = \begin{cases} \delta & \text{if } Q = [0,1), \\ \alpha_Q^- & \text{if } Q \subseteq [0,1/2), \\ \alpha_Q^+ & \text{if } Q \subseteq [1/2,1). \end{cases}$$

It is easy to check that the Carleson constant of $(\alpha_Q)_{Q\in\mathcal{T}}$ is bounded by 1 and $\sum_{Q\in\mathcal{T}} \alpha_Q \mu(Q) = (t_- + t_+)/2 + \delta = t + \delta$. Therefore,

$$\begin{aligned} &\overline{\mathbb{B}}\,(s, t+\delta, x) \\ &\geq \int_E \Big( \sum_{Q\in\mathcal{T}} \alpha_Q (\langle f\rangle_Q + 1)\chi_Q \Big) \mathrm{d}\mu - K \int_{[0,1]} \big(\Psi(f) + f\big)\mathrm{d}\mu \\ &= \alpha_{[0,1]}(\langle f\rangle_{[0,1]} + 1)\mu(E) \\ &\quad + \int_{E^-} \Big( \sum_{Q\in\mathcal{T}^-} \alpha_Q^- (\langle f^-\rangle_Q + 1)\chi_Q \Big) \mathrm{d}\mu - K \int_{[0,1/2)} \big(\Psi(f^-) + f^-\big)\mathrm{d}\mu \\ &\quad + \int_{E^+} \Big( \sum_{Q\in\mathcal{T}^+} \alpha_Q^+ (\langle f^+\rangle_Q + 1)\chi_Q \Big) \mathrm{d}\mu - K \int_{[1/2,1)} \big(\Psi(f^+) + f^+\big)\mathrm{d}\mu. \end{aligned}$$

However, $\alpha_{[0,1]}(\langle f\rangle_{[0,1]} + 1)\mu(E) = \delta(x+1)s$, so taking the supremum over all $E^\pm$, $(\alpha_Q^\pm)_{Q\in\mathcal{T}^\pm}$ and $f^\pm$, we get the claim. □

Equipped with the structural properties above, we are ready to construct a candidate for $\overline{\mathbb{B}}$. We would like to stress here that the reasoning we will present is informal, in particular we will impose some additional regularity conditions on $\overline{\mathbb{B}}$, we will also guess certain formulas at some points. Thus, we will denote the function differently, by $\overline{\mathbb{B}}_0$. One should keep in mind that the primary goal of this section is to *discover* the function $B$ which will be rigorously exploited later.

We start from the observation that $\overline{\mathbb{B}}(s,t,x) = -K(\Psi(x) + x)$ if $s = 0$ or $t = 0$. Indeed, if $st = 0$, then the contribution from $\int_E \Big( \sum_{Q\in\mathcal{T}} \alpha_Q (\langle f\rangle_Q + 1)\chi_Q \Big)\mathrm{d}\mu$ is zero, and the optimal choice for $f$ is $f \equiv x$ (other choices may only make the expression $-K\int_X (\Psi(f) + f)\mathrm{d}\mu$ smaller, since $u \mapsto \Psi(u) + u$ is a convex function). Thus, in our considerations below, we may restrict ourselves to the domain $(0,1] \times (0,1] \times [-1,\infty)$.

Our first assumption is that $\overline{\mathbb{B}}_0$ is of class $C^2$. Then the concavity studied in Lemma 2.2 implies that the Hessian matrix $D^2\overline{\mathbb{B}}_0$ is nonpositive-definite at each point of the domain. By Lemma 2.1, this amounts to saying that the matrix

$$\begin{bmatrix} \varphi_{ss} & \varphi_{st} & \varphi_s/(x+1) \\ \varphi_{ts} & \varphi_{tt} & \varphi_t/(x+1) \\ \varphi_s & \varphi_t & -K/(x+1) \end{bmatrix},$$

where $\varphi(s,t) = \overline{\mathbb{B}}_0(s,t,0)$, is nonpositive-definite. In particular, we see that

$$\varphi_{ss} \leq 0 \qquad \text{and} \qquad \varphi_{tt} \leq 0, \tag{2.3}$$

which will be important to us later. Next, we use Remark 2.3 and assume that for *any* $(s,t,x)$, there is a (short) line segment $I$ containing $(s,t,x)$ such $\overline{\mathbb{B}}_0$ is linear along $I$. This implies the partial differential equation

$$\det\begin{bmatrix} \varphi_{ss} & \varphi_{st} & \varphi_s/(x+1) \\ \varphi_{ts} & \varphi_{tt} & \varphi_t/(x+1) \\ \varphi_s & \varphi_t & -K/(x+1) \end{bmatrix} = 0,$$

the Monge-Ampère equation for $\overline{\mathbb{B}}_0$. Therefore, for any $(s,t,x)$, there are numbers $a = a(s,t,x)$, $b = b(s,t,x)$, $c = c(s,t,x)$ such that

$$a\varphi_{ss} + b\varphi_{st} + c\frac{\varphi_s}{x+1} = 0, \quad a\varphi_{ts} + b\varphi_{tt} + c\frac{\varphi_t}{x+1} = 0$$

and

$$a\varphi_s + b\varphi_t - c\frac{K}{x+1} = 0. \tag{2.4}$$

To gain some intuition about the solution to this system, multiply the last equation by $\varphi_s/K$ and add it to the first equation; similarly, multiply (2.4) by $\varphi_t/K$ and add it to the second equation. As the result, we get the system

$$a(\varphi_{ss}+(\varphi_s)^2/K)+b(\varphi_{st}+\varphi_s\varphi_t/K) = 0, \quad a(\varphi_{ts}+\varphi_s\varphi_t/K)+b(\varphi_{tt}+(\varphi_t)^2/K) = 0,$$

or, if we put $\psi = \exp(\varphi/K)$,

$$a\psi_{ss} + b\psi_{st} = 0, \qquad a\psi_{ts} + b\psi_{tt} = 0. \tag{2.5}$$

In other words, $\psi$ satisfies the Monge-Ampère equation $\det D^2\psi = 0$, and $a$, $b$ indicate the direction in which $\psi$ is linear. Since $\psi$ depends on $s$, $t$ only, we assume that $a$ and $b$ also have this property.

It follows from general theory of Monge-Ampère equations (see a similar discussion in [28]) that $[0,1]\times[0,1]$, the domain of $\psi$, can be foliated, i.e., split into union of pairwise disjoint line segments along which $\psi$ is linear. Suppose now that $s \leq t$. A little thought reveals that there is essentially only one candidate for the foliation, consisting of the line segments $\{(\alpha u, u) : u \in [0,1]\}$, where $\alpha \in [0,1]$ is arbitrary. This corresponds to the choice $a(s,t,x) = s$, $b(s,t,x) = t$. In other words, we conjecture that the function $r \mapsto \psi(s+sr, t+tr)$ is linear. Assuming that this conjecture holds, we get that $\psi(s,t) = (1-t)\psi(0,0) + t\psi(s/t,1)$. However, $\psi(0,0) = \exp(\overline{\mathbb{B}}_0(0,0,0)/K) = 1$, from the very definition of $\overline{\mathbb{B}}$. Putting $\xi(u) = \psi(u,1)$, we obtain

$$\begin{aligned} &\overline{\mathbb{B}}_0(s,t,x) \\ &\quad = (x+1)\varphi(s,t) - K(x+1)\log(x+1) = -K(x+1)\log\left[\frac{x+1}{t\xi(s/t)+1-t}\right]. \end{aligned} \tag{2.6}$$

To find $\xi$, let us exploit the condition (2.4), which implies $\overline{\mathbb{B}}_{0t}(s,t,x) \geq (x+1)s$. We already know from the previous considerations (see (2.3)) that the function $t \mapsto \overline{\mathbb{B}}_0(s,t,x)$ is concave (for any fixed $s$ and $x$), so the above bound for $\overline{\mathbb{B}}_{0t}$ is equivalent to

$$\overline{\mathbb{B}}_{0t}(s,1,x) \geq (x+1)s.$$

We assume that we have equality here, which, by (2.6), yields a differential equation for $\xi$. A little calculation reveals that this equation is

$$\xi'(u) = \frac{\xi(u)}{u} - \frac{\xi(u)}{K} - \frac{1}{u}.$$

Since $\xi(0) = 1$ (this can be extracted from $\overline{\mathbb{B}}_0(0,t,x) = -K(x+1)\log(x+1)$, which follows directly from the definition of $\overline{\mathbb{B}}$) we conclude that $\xi$ is given by

$$\xi(u) = \frac{K}{K-1} u \exp\left(\frac{1-u}{K}\right) + u\exp\left(-\frac{u}{K}\right)\int_1^{1/u} \exp\left(\frac{1}{rK}\right)\mathrm{d}r, \qquad u > 0.$$

Coming back to (2.4), we obtain that

$$c(s,t,x) = \frac{(x+1)t(\xi(s/t)-1)}{t\xi(s/t)+1-t},$$

which completes the analysis of $\overline{\mathbb{B}}_0$ on the set $\{(s,t) \in [0,1]\times[0,1] : s \leq t\}$.

What happens if $s > t$? We have showed above that $\overline{\mathbb{B}}_{0ss} \leq 0$ and it is clear from the very definition that for any $x$ and $t$, the function $s \mapsto \overline{\mathbb{B}}_0(s,t,x)$ is nondecreasing. However, one easily checks, exploiting the above formula for $\overline{\mathbb{B}}_0$ on the set $s \leq t$, that

$$\lim_{s\uparrow t} \overline{\mathbb{B}}_{0s}(s,t,x) = 0 \qquad \text{for any } t \in [0,1] \text{ and } x \geq -1.$$

These observations imply that $\overline{\mathbb{B}}_0(s,t,x) = \overline{\mathbb{B}}_0(\min(s,t),t,x)$. This also shows that the family of line segments $\{(1-\alpha+\alpha t, t) : t \in [0,1]\}$, $\alpha \in [0,1]$, forms the foliation of $\{(s,t) : s \geq t\}$. Furthermore, $a(s,t,x) = t(1-s)/(1-t)$, $b(s,t,x) = t$ and

$$c(s,t,x) = c(t,t,x) = \frac{(x+1)t}{t+K-1}.$$

One can prove that the function $\overline{\mathbb{B}}_0$ we have constructed majorizes the abstract function $\overline{\mathbb{B}}$ (similar arguments will appear in the next section). The examples we are going to study in Section 4 will show the reverse inequality and hence the two functions coincide.

## 3. A formal proof

Throughout this section, $K > 1$ is a fixed number. Consider the function $\xi : [0,1] \to (0,\infty)$, given by $\xi(0) = 1$ and

$$\xi(u) = \frac{K}{K-1} u \exp\left(\frac{1-u}{K}\right) + u\exp\left(-\frac{u}{K}\right)\int_1^{1/u} \exp\left(\frac{1}{rK}\right)\mathrm{d}r, \qquad u > 0.$$

One easily verifies that $\xi$ is continuous and $\xi(1) = K/(K-1)$.

**Lemma 3.1.** *For any $s \in (0,1)$ we have $\xi(s) > 1$, $\xi'(s) > 0$ and*

$$\frac{K\xi(s) - K}{s} - 1 + \log s \leq L(K). \tag{3.1}$$

*Proof.* Observe that $\xi$ satisfies the differential equation

$$\xi'(u) = \frac{\xi(u)}{u} - \frac{\xi(u)}{K} - \frac{1}{u}. \tag{3.2}$$

Suppose that there is $u_0 \in (0,1)$ such that $\xi(u_0) \leq 1$. Then $\xi'(u_0) \leq -\xi(u_0)/K < 0$ and hence $\xi(u) < 1$ for all $u \in (u_0, 1]$, a contradiction (we have $u(1) > 1$). This establishes the first part of the assertion. To show the second part, note that (3.2) implies $\lim_{u\to 1-} \xi'(u) = 0$. Therefore, we will be done if we show that $\xi$ is concave. To this end, we differentiate (3.2) and obtain

$$-u\xi''(u) = \frac{\xi(u) + u\xi'(u)}{K} = \frac{2\xi(u) - u\xi(u)/K - 1}{K} \geq \frac{\xi(u)-1}{K} > 0. \tag{3.3}$$

Finally, we turn our attention to (3.1). The equality (3.2) and the estimate $\xi \geq 1$ we have just established imply that

$$\left(\frac{K\xi(s)-K}{s} - 1 + \log s\right)' = \frac{K}{s}\left(\xi'(s) - \frac{\xi(s)}{s} + \frac{1}{s}\right) + \frac{1}{s} = \frac{1-\xi(s)}{s} \leq 0,$$

so it suffices to show (3.1) as $s \to 0$. Directly from the definition of $\xi$ and integration by parts, we obtain

$$\begin{aligned}
&\frac{K\xi(s)-K}{s} - 1 + \log s \\
&= \frac{K^2}{K-1}\exp\left(\frac{1-s}{K}\right) + K\exp\left(-\frac{s}{K}\right)\int_1^{1/s}\exp(1/(rK))\mathrm{d}r - \frac{K}{s} - 1 + \log s \\
&= \frac{K}{K-1}\exp\left(\frac{1-s}{K}\right) + \exp\left(-\frac{s}{K}\right)\int_1^{1/s}\exp(1/(rK))\frac{\mathrm{d}r}{r} - 1 + \log s.
\end{aligned}$$

We have $\lim_{s\to 0}(\exp(-s/K)-1)\log s = 0$, so adding $(\exp(-s/K)-1)\log s$ to both sides above yields

$$\begin{aligned}
\lim_{s\to 0}\left(\frac{K\xi(s)-K}{s} - 1 + \log s\right) &= \frac{K}{K-1}\exp(K^{-1}) - 1 + \int_1^{\infty}(\exp(1/(rK))-1)\frac{\mathrm{d}r}{r} \\
&= L(K)
\end{aligned}$$

and the claim is proved. □

Let $\mathbb{D} = \{(s,t,x) \in [0,1]\times[0,1]\times[0,\infty) : s \leq t\}$ and consider $B : \mathbb{D} \to \mathbb{R}$ defined by the formula

$$B(s,t,x) = Kx - s - s\log(t/s) + K(x+1)\log\left[\frac{t\xi(s/t)+1-t}{x+1}\right],$$

where we use the convention $0/0 = 1$ and $0\log(t/0) = 0$. Extend $B$ to the whole set $[0,1]\times[0,1]\times[0,\infty)$, putting $B(s,t,x) = B(\min(s,t),t,x)$. The function $B$ is "almost" the function $\overline{\mathbb{B}}$ constructed in the preceding section: the key differences will be discussed in the last part of the paper.

The following properties of $B$ will be needed later.

**Lemma 3.2.** *We have $B_s(s,t,x) \geq 0$ and $B_t(s,t,x) \geq xs/t$ in the interior of $\mathbb{D}$.*

*Proof.* We compute that

$$B_s(s,t,x) = \log(s/t) + \frac{K(x+1)}{t\xi(s/t)+1-t}\xi'(s/t) \geq \log(s/t) + \frac{K\xi'(s/t)}{\xi(s/t)}, \tag{3.4}$$

where the inequality follows from the assertion of the previous lemma and the estimates $K > 1$, $x+1 \geq 1$, $\xi \geq 1$. If we let $s/t \to 1$, then the latter expression converges to 0. Therefore, the inequality $B_s \geq 0$ will follow once we have shown that the function $u \mapsto \log u + K\xi'(u)/\xi(u)$ is nonincreasing. If we compute its derivative, we see that we must prove that

$$\frac{K\xi''(u)}{\xi(u)} - \frac{K(\xi'(u))^2}{(\xi(u))^2} + \frac{1}{u} \leq 0.$$

By (3.3), this can be rewritten in the form

$$-\frac{u\xi'(u)}{\xi(u)} - \frac{K(\xi'(u))^2}{(\xi(u))^2} \leq 0,$$

which is evident ($\xi$ is increasing, see the previous lemma). To show the estimate for $B_t$, we reformulate it as

$$-u+\frac{K(x+1)}{t\xi(u)+1-t}\left(\xi\left(u\right)-u\xi'\left(u\right)-1\right)\geq xu$$

(here, as above, we use the notation $u=s/t$) or, by (3.2),

$$-u+\frac{x+1}{t\xi(u)+1-t}\cdot u\xi\left(u\right)\geq xu.$$

This can be further transformed into $\xi(u)\geq t\xi(u)+1-t$, which holds true due to the estimate $\xi\geq 1$ established above. □

**Lemma 3.3.** *We have $B(s,0,x)=B(0,0,x)=-K\Psi(x)$ for any $s\in[0,1]$ and any $x\geq 0$. Furthermore, if $(s,t,x)\in\mathbb{D}$, then*

$$B(s,t,x)\leq L(K)\cdot s. \tag{3.5}$$

*Proof.* The first part is trivial. To show the majorization (3.5), note that $B_t\geq 0$ by the previous lemma, and hence

$$B(s,t,x)\leq B(s,1,x)=Kx-s+s\log s+K(x+1)\log\left[\frac{\xi(s)}{x+1}\right].$$

Furthermore, we derive that $B_x(s,1,x)=K\log\left(\xi(s)/(x+1)\right)$, which is positive if and only if $x+1<\xi(s)$. Consequently,

$$B(s,1,x)\leq B(s,1,\xi(s)-1)=s\left[\frac{K\xi(s)-K}{s}-1+\log s\right]\leq L(K)\cdot s,$$

where in the last passage we have exploited (3.1). □

**Lemma 3.4.** *The Hessian matrix of $B$ is nonpositive-definite in the interior of $\mathbb{D}$.*

*Proof.* It suffices to check that $B_{xx}\leq 0$, $\det\begin{bmatrix} B_{xx} & B_{xs}\\ B_{xs} & B_{ss}\end{bmatrix}\geq 0$ and $\det D^2B\leq 0$, in the light of Sylvester's criterion. We have $B_{xx}(s,t,x)=-K/(x+1)$, so the first inequality is obvious. To check the second one, we differentiate (3.4) to obtain

$$\begin{aligned}
B_{sx}(s,t,x)&=\frac{K\xi'(s/t)}{t\xi(s/t)+1-s/t},\\
B_{ss}(s,t,x)&=\frac{1}{s}-\frac{K(x+1)(\xi'(s/t))^2}{(t\xi(s/t)+1-t)^2}+\frac{K(x+1)\xi''(s/t)}{t(t\xi(s/t)+1-t)}
\end{aligned}$$

and hence

$$\det\begin{bmatrix} B_{xx} & B_{xs}\\ B_{xs} & B_{ss}\end{bmatrix}=-\frac{K}{(x+1)s}-\frac{K^2\xi''(s/t)}{t(t\xi(s/t)+1-t)}.$$

Since $x+1\geq 1$, the desired positivity of the determinant will follow if we prove the estimate

$$-1-\frac{Ku\xi''(u)}{t\xi(u)+1-t}\geq 0,$$

where $u=s/t$. This, by (3.3), is equivalent to

$$-1+\frac{\xi(u)+u\xi'(u)}{t\xi(u)+1-t}\geq 0.$$

However, for a fixed $u$, the left-hand side above is the smallest for $t=1$; but then its value is equal to $u\xi'(u)/\xi(u)$ which is nonnegative due to Lemma 3.1.

It remains to show that $\det D^2B \leq 0$; actually, we will show that the determinant vanishes. To this end, we compute that

$$\begin{aligned}&\left[\frac{s}{t}B_{xs}+B_{xt},\frac{s}{t}B_{ts}+B_{tt},\frac{s}{t}B_{ss}+B_{st}\right]\\&=-\frac{(x+1)(\xi(u)-1)}{t\xi(u)+1-t}\left[-\frac{K}{x+1},\frac{u\xi(u)}{t\xi(u)+1-t},\frac{K\xi'(u)}{t\xi(u)+1-t}\right]\\&=-\frac{(x+1)(\xi(u)-1)}{t\xi(u)+1-t}\left[B_{xx},B_{xt},B_{xs}\right].\end{aligned}$$

In other words, the rows of the Hessian are linearly dependent. This completes the proof. □

As a consequence of the lemmas above, we get the following concavity-type property of $B$.

**Corollary 3.5.** *Let $n \geq 2$ be an integer and suppose that $(s_i, t_i, x_i) \in [0,1]\times[0,1]\times[0,\infty)$, $i = 1, 2, \ldots, n$, be arbitrary points. Assume further that $\alpha_1, \alpha_2, \ldots, \alpha_n$ are nonnegative numbers summing up to 1 and set $s = \sum_{i=1}^n \alpha_i s_i$, $x = \sum_{i=1}^n \alpha_i x_i$. If $t \in [0,1]$ satisfies $t \geq \sum_{i=1}^n \alpha_i t_i$, then*

$$B(s,t,x) \geq \sum_{i=1}^n \alpha_i B(s_i,t_i,x_i) + \left(t - \sum_{i=1}^n \alpha_i t_i\right)\frac{x\cdot\min(s,t)}{t}. \tag{3.6}$$

*Proof.* Set $\tilde{s} = \sum_{i=1}^n \alpha_i \min(s_i,t_i)$, $\tilde{t} = \sum_{i=1}^n \alpha_i t_i$ and observe that

$$\tilde{s} \leq \sum_{i=1}^n \alpha_i s_i = s, \qquad \tilde{s} \leq \sum_{i=1}^n \alpha_i t_i \leq \tilde{t}.$$

By the previous lemma, the function $B$ is concave on $\mathbb{D}$, so

$$\sum_{i=1}^n \alpha_i B(s_i,t_i,x_i) = \sum_{i=1}^n \alpha_i B(\min(s_i,t_i),t_i,x_i) \leq B(\tilde{s},\tilde{t},x).$$

Now consider two cases. If $s \leq \tilde{t}$, then by Lemma 3.2,

$$\begin{aligned}B(\tilde{s},\tilde{t},x) \leq B(s,\tilde{t},x) &= B(s,t,x) - \int_{\tilde{t}}^t B_t(s,u,x)\mathrm{d}u\\&\leq B(s,t,x) - \int_{\tilde{t}}^t \frac{xs}{u}\mathrm{d}u\\&\leq B(s,t,x) - \int_{\tilde{t}}^t \frac{xs}{t}\mathrm{d}u\\&= B(s,t,x) - (t-\tilde{t})\frac{x\min(s,t)}{t},\end{aligned}$$

which combined with the previous estimate yields (3.6). On the other hand, if $\tilde{t} < s$, then we apply Lemma 3.2 to get $B(\tilde{s},\tilde{t},x) \leq B(\tilde{t},\tilde{t},x)$. If we consider the function

$$\eta(u) = B\big(\tilde{t} + u(\min(s,t)-\tilde{t}), \tilde{t} + u(t-\tilde{t}), x\big),$$

then we see that

$$\begin{aligned} B(s,t,x) - B(\tilde{t},\tilde{t},x) &= \eta(1) - \eta(0) \\ &= \int_0^1 \eta'(u)\mathrm{d}u \\ &= \int_0^1 \Big( B_s(\dots)(\min(s,t) - \tilde{t}) + B_t(\dots)(t - \tilde{t}) \Big)\mathrm{d}u \\ &\geq \int_0^1 B_t(\dots)(t-\tilde{t})\mathrm{d}u \\ &\geq x(t-\tilde{t}) \int_0^1 \frac{\tilde{t} + u(\min(s,t) - \tilde{t})}{\tilde{t} + u(t-\tilde{t})}\mathrm{d}u, \end{aligned}$$

where in the last line we have exploited Lemma 3.2 again. It suffices to note that for any $u \in [0,1]$ we have

$$\frac{\tilde{t} + u(\min(s,t) - \tilde{t})}{\tilde{t} + u(t-\tilde{t})} \geq \frac{\min(s,t)}{t},$$

since the left-hand side is concave in $u$ and the inequality holds for $u \in \{0,1\}$. $\square$

We are ready to establish the logarithmic estimate.

*Proof of* (1.4). Fix a sequence $(\alpha_Q)_{Q\in\mathcal{T}}$ with a Carleson constant less or equal to 1, a measurable set $E$ and a nonnegative function $f$. By a simple limiting argument, we may assume that all the terms $\alpha_Q$ corresponding to sufficiently small $Q$ vanish (i.e., there is $M$ such that $\alpha_Q = 0$ if $Q \in \bigcup_{n>M} \mathcal{T}^n$). We will use the following notation: for $Q \in \mathcal{T}$ we will write

$$s_Q = \frac{\mu(E\cap Q)}{\mu(Q)}, \qquad t_Q = \frac{1}{\mu(Q)} \sum_{Q'\in\mathcal{T}(Q)} \alpha_{Q'}\mu(Q'), \qquad x_Q = \langle f\rangle_Q.$$

Observe that if $Q_1$, $Q_2$, …, $Q_n$ are direct children of $Q$, then

$$s_Q = \sum_{i=1}^n \frac{\mu(Q_i)}{\mu(Q)} s_{Q_i}, \qquad t_Q = \alpha_Q + \sum_{i=1}^n \frac{\mu(Q_i)}{\mu(Q)} t_{Q_i} \quad \text{and} \quad x_Q = \sum_{i=1}^n \frac{\mu(Q_i)}{\mu(Q)} x_{Q_i}.$$

Therefore, if we apply the inequality (3.6) with $(s,t,x) = (s_Q,t_Q,x_Q)$, $(s_i,t_i,x_i) = (s_{Q_i},t_{Q_i},x_{Q_i})$ and $\alpha_i = \mu(Q_i)/\mu(Q)$, then we get

$$\begin{aligned} &B(s_Q,t_Q,x_Q) \\ &\geq \sum_{i=1}^n \frac{\mu(Q_i)}{\mu(Q)} B\big(s_{Q_i},t_{Q_i},x_{Q_i}\big) + \alpha_Q \langle f\rangle_Q \cdot \frac{\min(s_Q,t_Q)}{t_Q} \\ &\geq \sum_{i=1}^n \frac{\mu(Q_i)}{\mu(Q)} B\big(s_{Q_i},t_{Q_i},x_{Q_i}\big) + \alpha_Q \langle f\rangle_Q s_Q. \end{aligned} \tag{3.7}$$

Multiply throughout by $\mu(Q)$ and sum the obtained inequalities over all $Q \in \mathcal{T}^N$ (for some fixed $N \geq 0$) to get

$$\sum_{Q\in\mathcal{T}^N} B(s_Q,t_Q,x_Q)\mu(Q) \geq \sum_{Q\in\mathcal{T}^{N+1}} B(s_Q,t_Q,x_Q)\mu(Q) + \sum_{Q\in\mathcal{T}^N} \alpha_Q\langle f\rangle_Q \mu(E\cap Q).$$

Writing this estimate for $N = 0, 1, 2, \ldots, M$ and summing, we obtain

$$B(s_X,t_X,x_X) \geq \sum_{Q\in\mathcal{T}^{M+1}} B(s_Q,t_Q,x_Q)\mu(Q) + \sum_{N=0}^{M}\sum_{Q\in\mathcal{T}^N} \alpha_Q\langle f\rangle_Q \mu(E\cap Q).$$

By (3.5), we have $B(s_X,t_X,x_X) \leq L(K)\cdot \min(s_X,t_X) \leq L(K)\cdot\mu(E)$. Furthermore, we have assumed that $\alpha_Q = 0$ for $Q \in \bigcup_{n>M}\mathcal{T}^n$, which implies that the second sum above is equal to

$$\sum_{Q\in\mathcal{T}} \alpha_Q\langle f\rangle_Q \mu(E\cap Q) = \int_E \mathcal{A}f\mathrm{d}\mu.$$

Furthermore, we have $t_Q = 0$ for $Q \in \mathcal{T}^{M+1}$, so Lemma 3.3 and Jensen's inequality give

$$\sum_{Q\in\mathcal{T}^{M+1}} B(s_Q,t_Q,x_Q)\mu(Q) = -K\sum_{Q\in\mathcal{T}^{M+1}} \Psi(\langle f\rangle_Q)\mu(Q) \geq -K\int_X \Psi(f)\mathrm{d}\mu.$$

Putting all the above facts together, we obtain the desired estimate (1.4). □

Finally, we turn our attention to Theorem 1.3.

*Proof.* It suffices to show the claim for nonnegative functions $f$. Pick $K > 1$, $M > 0$, $f : X \to [0,1]$ and let

$$g = \exp\Big(\min\left(\mathcal{A}f/K, M\right)\Big) - 1.$$

Since $\mathcal{A}$ is self-adjoint, we may write

$$\begin{aligned}\int_X \frac{\mathcal{A}f}{K} g\mathrm{d}\mu &= \frac{1}{K}\int_X f\mathcal{A}g\mathrm{d}\mu \\ &= \frac{1}{K}\int_{\{f\neq 0\}} f\mathcal{A}g\mathrm{d}\mu \leq \int_X \Psi(g)\mathrm{d}\mu + \frac{L(K)}{K}\mu(\{f\neq 0\}),\end{aligned}$$

by virtue of (1.4). After simple manipulations, this can be shown to be equivalent to the estimate

$$\begin{aligned}&\int_X \left(\exp\Big(\min\left(\mathcal{A}f/K, M\right)\Big) - 1 - \frac{\mathcal{A}f}{K}\right)\mathrm{d}\mu \\ &\quad + \int_{\{\mathcal{A}f > KM\}} \left(\frac{\mathcal{A}f}{K} - M\right)\exp\Big(\min\left(\mathcal{A}f/K, M\right)\Big)\mathrm{d}\mu \leq \frac{L(K)}{K}\mu(\{f\neq 0\}).\end{aligned}$$

Letting $M \to \infty$ and applying Fatou's lemma, we get (1.5). Let us now show how to deduce the optimality of $L(K)/K$ from the fact that (1.4) is sharp (which will be proved later below). Suppose that (1.5) holds with some constant $L'(K)/K$ on the right. Then for any nonnegative $f$ and any measurable $E$ we have, by Young's inequality,

$$\begin{aligned}\int_E \mathcal{A}f\mathrm{d}\mu &= \int_X f\mathcal{A}\chi_E\mathrm{d}\mu \\ &\leq K\int_X \Psi(f)\mathrm{d}\mu + \int_X \Phi(\mathcal{A}\chi_E/K)\mathrm{d}\mu \\ &\leq K\int_X \Psi(f)\mathrm{d}\mu + \frac{L'(K)}{K}\mu(E).\end{aligned}$$

Therefore $L'(K)\geq L(K)$ for each $K$ and hence the sharpness is proved. It remains to show that for $K\leq 1$ the exponential bound does not hold with any finite constant; this follows immediately from the facts that the left-hand side of (1.5) decreases as $K$ increases, and $L(K)\to\infty$ as $K\downarrow 1$. □

## 4. Sharpness and the formula for $\mathbb{B}$

Throughout this section we return to the context of general probability space $(X,\mu)$ equipped with some tree of its subsets. We will use the following fact which can be found in [13].

**Lemma 4.1.** *For every $Q\in\mathcal{T}$ and every $\beta\in(0,1)$ there is a subfamily $F(Q)\subset\mathcal{T}$ consisting of pairwise almost disjoint subsets of $Q$ such that*

$$\mu\left(\bigcup_{R\in F(Q)} R\right) = \sum_{R\in F(Q)}\mu(R) = \beta\mu(Q).$$

The purpose of this section is to complete the analysis of (1.4), by constructing examples which imply the sharpness of this estimate and finding the formula for $B$ (used in Section 3). As we have mentioned above, we expect the examples to be related to the extremizers of the function $\overline{\mathbb{B}}$ discovered above. These extremizers, in turn, are "encoded" in the formula for $\overline{\mathbb{B}}_0$, as we shall explain now. The rough idea can be described as follows. Suppose that $\gamma_1$, $\gamma_2$, …, $\gamma_n$ are nonnegative numbers summing up to 1 and suppose that $(s_i,t_i,x_i)\in[0,1]\times[0,1]\times[-1,\infty)$, $i=1$, 2, …, $n$ satisfy

$$\overline{\mathbb{B}}_0(s,t,x) = \sum_{i=1}^n \gamma_i\overline{\mathbb{B}}_0(s_i,t_i,x_i) + \alpha(x+1)s,$$

where $\alpha\geq 0$ and $s=\sum_{i=1}^n\gamma_i s_i$, $t=\alpha+\sum_{i=1}^n\gamma_i t_i\leq 1$ and $x=\sum_{i=1}^n\gamma_i x_i$. Assume further that for each $i$, any probability space and any tree, we know how to construct the extremizers $E^i$, $(\alpha_Q^i)_Q$ and $f^i$ of $\overline{\mathbb{B}}_0(s_i,t_i,x_i)$. Then these extremizers, after an appropriate splicing, yield the extremizers $E$, $(\alpha_Q)_Q$ and $f$ corresponding to $\overline{\mathbb{B}}_0(s,t,x)$. We have already seen a similar phenomenon in the proofs of Lemmas 2.2 and 2.4, so we will be brief. Apply Lemma 4.1 to obtain a family $A_1$, $A_2$, …, $A_n$ of pairwise disjoint subsets of $X$ such that $\mu(A_i)=\gamma_i$ and such that each $A_i$ can be written as a union $\bigcup_j Q_j^i$ of pairwise disjoint elements of $\mathcal{T}$ (called *the atoms* of $A_i$). Now, on each atom $Q_j^i$ of $A_i$, we construct the extremizers $E^{i,j}$, $(\alpha_Q^{i,j})_Q$ and $f^{i,j}$ of $\overline{\mathbb{B}}(s_i,t_i,x_i)$. Set $E=\bigcup_{i,j}E^{i,j}$, $f=\sum_{i,j}f^{i,j}\chi_{Q_j^i}$ and consider the union of all $(\alpha_Q^{i,j})_Q$, completing the sequence to the full $(\alpha_Q)_{Q\in\mathcal{T}}$ by taking $\alpha_X=\alpha$ and

$\alpha_Q = 0$ for remaining $Q$'s. Then $E$, $(\alpha_Q)_{Q\in\mathcal{T}}$ and $f$ have all the properties required in the definition of $\overline{\mathbb{B}}(s,t,x)$ and

$$\int_E \Big(\sum_{Q\in\mathcal{T}} \alpha_Q(\langle f\rangle_Q + 1)\chi_Q\Big)\mathrm{d}\mu - K\int_X \big(\Psi(f)+f\big)\mathrm{d}\mu = \sum_{i=1}^n \gamma_i \overline{\mathbb{B}}_0(s_i,t_i,x_i) + \alpha x s,$$

so they are indeed the extremizers of $\overline{\mathbb{B}}(s,t,x)$.

It is convenient to split the reasoning into a few intermediate parts.

4.1. **Extremizers corresponding to $\overline{\mathbb{B}}(1,1,x)$.** Let $x \geq -1$ be a given number.

*Step 1. Indication from $\overline{\mathbb{B}}_0$.* Let us look at the definition of $\overline{\mathbb{B}}(1,1,x)$. As the set $E$, we (must) take $E = X$, so that $\mu(E) = s = 1$. Furthermore, since $\overline{\mathbb{B}}_{0t}(1,1,x) = x+1$, we have, for small $\delta > 0$,

$$\overline{\mathbb{B}}_0(1,1,x) = \overline{\mathbb{B}}_0(1,1-\delta,x) + \delta(x+1) + o(\delta). \tag{4.1}$$

Next, we know from the above construction that $\overline{\mathbb{B}}_0$ is linear along the line segment

$$(1,1-\delta,x) + \big(a(1,1-\delta,x), b(1,1-\delta,x), c(1,1-\delta,x)\big)u$$
$$= (1,1-\delta,x) + \left(0, 1-\delta, \frac{(x+1)(1-\delta)}{K-\delta}\right)u, \qquad u \in [-1, \delta/(1-\delta)],$$

so

$$\overline{\mathbb{B}}_0(1,1-\delta,x)$$
$$= \delta\overline{\mathbb{B}}_0\left(1,0,x - \frac{(x+1)(1-\delta)}{K-\delta}\right) + (1-\delta)\overline{\mathbb{B}}_0\left(1,1,x+\frac{(x+1)(1-\delta)}{K-\delta}\cdot\frac{\delta}{1-\delta}\right).$$

Putting the two steps above together, we see that for $\delta$ small enough,

$$\begin{aligned}\overline{\mathbb{B}}_0(1,1,x) &\approx \delta(x+1) + \delta\overline{\mathbb{B}}_0\left(1,0,x-\frac{(x+1)(1-\delta)}{K-\delta}\right)\\ &\quad + (1-\delta)\overline{\mathbb{B}}_0\left(1,1,x+\frac{(x+1)(1-\delta)}{K-\delta}\cdot\frac{\delta}{1-\delta}\right)\\ &= \delta(x+1) + \delta\overline{\mathbb{B}}_0\left(1,0,\frac{(K-1)(x+1)}{K-\delta}-1\right)\\ &\quad + (1-\delta)\overline{\mathbb{B}}_0\left(1,1,\frac{K(x+1)}{K-\delta}-1\right)\end{aligned} \tag{4.2}$$

Now the reader is urged to recall the discussion from the beginning of this section. The above almost equality indicates how to construct the extremizers corresponding to $\overline{\mathbb{B}}(1,1,x)$ if we know how to produce them for the values $\overline{\mathbb{B}}\left(1,0,\frac{(x+1)(K-1)}{K-\delta}-1\right)$ and $\overline{\mathbb{B}}\left(1,1,\frac{K(x+1)}{K-\delta}-1\right)$. The first value is evident: since $s=1$, we must take $E$ to be the entire space (so that $\mu(E) = s = 1$); the condition $t=0$ forces $(\alpha_Q)_Q$ to contain only zeros; finally, we must take $f \equiv \frac{(K-1)(x+1)}{K-\delta} - 1$, since other choices make the expression

$$\int_E \Big(\sum_{Q\in\mathcal{T}} \alpha_Q(\langle f\rangle_Q+1)\chi_Q\Big)\mathrm{d}\mu - K\int_X \big(\Psi(f)+f\big)\mathrm{d}\mu = -K\int_X \big(\Psi(f)+f\big)\mathrm{d}\mu$$

smaller (the function $x \mapsto \Psi(x)+x$ is convex). To handle the second value, we observe that the first two coordinates of the point $\left(1,1,\frac{K(x+1)}{K-\delta}-1\right)$ are equal to

1, so we may repeat the splitting described in (4.2), with $x$ replaced by $\frac{K(x+1)}{K-\delta}-1$. This will imply that the appropriate extremizers can be constructed from those corresponding to $\mathbb{B}\left(1,0,\frac{K(K-1)(x+1)}{(K-\delta)^2}-1\right)$ and $\mathbb{B}\left(1,1,\frac{K^2(x+1)}{(K-\delta)^2}-1\right)$. For the first value the extremizers are clear, while for the second value we use the splitting (4.2) again, and so on. So, the construction has the following inductive algorithm:

1° Start with $n=0$ and put $A_0=X$: this set has only one atom, $A_0$ itself.
2° For any atom $Q$ of $A_n$, set $\alpha_Q=\delta$.
3° Apply Lemma 4.1 to each atom $Q$ of $A_n$, splitting it into two sets $Q^-$ and $Q^+=Q\setminus Q^-$, with $\mu(Q^-)=(1-\delta)\mu(Q)$. Define $A_{n+1}=\bigcup_{Q \text{ atom of } A_n} Q^-$. By Lemma 4.1, each $Q^-$ can be expressed as a union of pairwise disjoint elements of $\mathcal{T}$ and hence $A_{n+1}$ also has this property; call the corresponding elements of $\mathcal{T}$ the atoms of $A_{n+1}$.
4° For all $Q\in\mathcal{T}$ which are proper subsets of some atom of $A_n$ and are not contained in any atom of $A_{n+1}$, set $\alpha_Q=0$.
5° Put $f\equiv\frac{K-1}{K-\delta}\left(\frac{K}{K-\delta}\right)^n(x+1)-1$ on $A_n\setminus A_{n+1}$, increase $n$ by 1 and go to 2°.

It is clear from the above construction that $X=A_0\supset A_1\supset A_2\supset\ldots$. Actually, we see that each atom $Q$ of $A_k$ ($k=0,1,2,\ldots$) is split in the same ratio, so $\mu(Q\cap A_{k+1})=(1-\delta)\mu(Q)$ and in particular, $\mu(A_k)=(1-\delta)^k$.

As the result of the above construction, we get the following explicit candidates for the sequence $(\alpha_Q)_{Q\in\mathcal{T}}$ and the function $f:X\to[-1,\infty)$. Namely, the sequence contains only the terms 0 and $\delta$, and $\alpha_Q=\delta$ if and only if $Q$ is an atom of some $A_n$. The function is given by

$$f=\sum_{n=0}^{\infty}\left(\frac{K-1}{K-\delta}\left(\frac{K}{K-\delta}\right)^n(x+1)-1\right)\chi_{A_n\setminus A_{n+1}}.$$

Let us make a comment which will be used in the study of the sharpness of (1.4).

**Remark 4.2.** If $x\geq 1/(K-1)$, then $f$ is nonnegative.

*Step 2. Formal verification of the properties of* $(\alpha_Q)_{Q\in\mathcal{T}}$ *and* $f$. The above construction was based on the *almost* equality (4.2), which, in addition, was applied infinitely many times. Thus it is not clear whether the objects we obtained satisfy the required conditions, and we need to check them rigorously. We start with the function. By straightforward manipulations on geometric series,

$$\langle f\rangle_X=\sum_{n=0}^{\infty}\left(\frac{K-1}{K-\delta}\left(\frac{K}{K-\delta}\right)^n(x+1)-1\right)(1-\delta)^n\delta=x+1-1=x$$

and, more generally, for any atom $Q$ of $A_k$,

$$\begin{aligned}\langle f\rangle_Q&=\frac{1}{\mu(Q)}\sum_{n=k}^{\infty}\left(\frac{K-1}{K-\delta}\left(\frac{K}{K-\delta}\right)^n(x+1)-1\right)\mu(Q\cap(A_n\setminus A_{n+1}))\\&=\sum_{n=k}^{\infty}\left(\frac{K-1}{K-\delta}\left(\frac{K}{K-\delta}\right)^n(x+1)-1\right)(1-\delta)^{n-k}\delta=\left(\frac{K}{K-\delta}\right)^k(x+1)-1.\end{aligned}$$

Now let us turn our attention to $(\alpha_Q)_{Q\in\mathcal{T}}$. To check that the Carleson constant is bounded by 1, observe first that

$$\sum_{Q\in\mathcal{T}}\alpha_Q\mu(Q)=\sum_{n=0}^\infty\sum_{\substack{Q\text{ atom}\\ \text{of }A_n}}\delta\mu(Q)=\sum_{n=0}^\infty\delta\mu(A_n)=\sum_{n=0}^\infty\delta(1-\delta)^n=1=\mu(X).$$

Next, for a given $Q\in\mathcal{T}$ different from $X$, let $n$ be the largest number such that a certain atom $Q'$ of $A_n$ is a proper superset of $Q$. Using the aforementioned fractal properties of $(A_k)_{k\geq 0}$, we get

$$\begin{aligned}\sum_{R\subseteq Q}\alpha_R\mu(R)&=\sum_{k=n+1}^\infty\sum_{\substack{R\subseteq Q\cap A_k,\\ R\text{ atom of }A_k}}\delta\mu(R)=\sum_{k=n+1}^\infty\delta\mu(Q\cap A_k)\\&=\mu(Q\cap A_{n+1})\sum_{k=n+1}^\infty\delta(1-\delta)^{n+1-k}=\mu(Q\cap A_{n+1})\leq\mu(Q).\end{aligned}$$

Thus, the set $E=X$, $(\alpha_Q)_{Q\in\mathcal{T}}$ and the function $f$ has all the properties required in the definition of $\overline{\mathbb{B}}(1,1,x)$.

*Step 3. Checking the value of* $\overline{\mathbb{B}}(1,1,x)$. It remains to note that

$$\begin{aligned}\sum_{Q\in\mathcal{T}}\alpha_Q(\langle f\rangle_Q+1)\mu(E\cap Q)&=\sum_{Q\in\mathcal{T}}\alpha_Q(\langle f\rangle_Q+1)\mu(Q)\\&=\sum_{n=0}^\infty\sum_{\substack{Q\text{ atom}\\ \text{of }A_n}}\delta\left(\frac{K}{K-\delta}\right)^n(x+1)\mu(Q)\\&=\sum_{n=0}^\infty\delta\left(\frac{K}{K-\delta}\right)^n(x+1)\mu(A_n)\\&=\frac{(K-\delta)(x+1)}{K-1}\end{aligned}$$

and

$$\begin{aligned}&\int_X(f+1)\log(f+1)\mathrm{d}\mu\\&=\sum_{n=0}^\infty\frac{K-1}{K-\delta}\left(\frac{K}{K-\delta}\right)^n(x+1)\log\left[\frac{K-1}{K-\delta}\left(\frac{K}{K-\delta}\right)^n(x+1)\right](1-\delta)^n\delta\\&=(x+1)\log\left[\frac{K-1}{K-\delta}(x+1)\right]+\delta\frac{K-1}{K-\delta}(x+1)\log\frac{K}{K-\delta}\sum_{n=0}^\infty n\left(\frac{K(1-\delta)}{K-\delta}\right)^n\\&=(x+1)\log\left[\frac{K-1}{K-\delta}(x+1)\right]+\frac{K(x+1)}{(K-1)\delta}\log\frac{K}{K-\delta}.\end{aligned}$$

Therefore, if we let $\delta\to 0$, then

$$\begin{aligned}&\sum_{Q\in\mathcal{T}}\alpha_Q(\langle f\rangle_Q+1)\mu(E\cap Q)-K\int_X(f+1)\log(f+1)\mathrm{d}\mu\\&\to\frac{K(x+1)}{K-1}-\left\{K(x+1)\log\left[\frac{(K-1)(x+1)}{K}\right]+\frac{K(x+1)}{K-1}\right\}\\&=\overline{\mathbb{B}}(1,1,x).\end{aligned}$$

Thus the function $\overline{\mathbb{B}}_0$ coincides with $\overline{\mathbb{B}}$ at all points of the form $(1,1,x)$, $x\geq -1$.

Our final comment concerns the functions $\mathbb{B}$ and $B$ studied in the previous sections. As we have discussed above, we expect $\mathbb{B}$ and $\overline{\mathbb{B}}$ to have the same extremizers. Let us plug $E$, $(\alpha_Q)_{Q\in\mathcal{T}}$ and $f$ constructed above into the definition of $\mathbb{B}$. Since

$$\sum_{Q\in\mathcal{T}}\alpha_Q\mu(E\cap Q)=\sum_{Q\in\mathcal{T}}\alpha_Q\mu(Q)=1,$$

we obtain in the limit that

$$\sum_{Q\in\mathcal{T}}\alpha_Q\langle f\rangle_Q\mu(E\cap Q)-K\int_X\Psi(f)\mathrm{d}\mu\xrightarrow{\delta\to 0}\overline{\mathbb{B}}(1,1,x)-1+Kx.$$

The latter is precisely the value $B(1,1,x)$.

4.2. **Extremizers corresponding to $\overline{\mathbb{B}}(s,1,x)$.** Let $s\in(0,1)$ and $x\geq -1$ be given numbers. This time the reasoning will be a little different.

*Step 1. Indication from* $\overline{\mathbb{B}}_0$. Let $N$ be a large positive integer and set $\delta=1-s^{1/N}$, so that $s/(1-\delta)^N=1$. Since $\overline{\mathbb{B}}_{0t}(s,1,x)=(x+1)s$, we have

$$\begin{aligned}\overline{\mathbb{B}}_0(s,1,x)&=\delta(x+1)s+\overline{\mathbb{B}}_0(s,1-\delta,x)+o(\delta)\\&=\delta(x+1)s+\delta\overline{\mathbb{B}}_0\left(0,0,x-c(s,1-\delta,x)\right)\\&\quad+(1-\delta)\overline{\mathbb{B}}_0\left(\frac{s}{1-\delta},1,x+c(s,1-\delta,x)\cdot\frac{\delta}{1-\delta}\right)+o(\delta)\\&=\delta(x+1)s+\delta\overline{\mathbb{B}}_0\left(0,0,\frac{x+1}{(1-\delta)\xi(s/(1-\delta))+\delta}-1\right)\\&\quad+(1-\delta)\overline{\mathbb{B}}_0\left(\frac{s}{1-\delta},1,\frac{(x+1)\xi(s/(1-\delta))}{(1-\delta)\xi(s/(1-\delta))+\delta}-1\right)+o(\delta).\end{aligned}\tag{4.3}$$

Here the second passage is a consequence of the linearity of $\overline{\mathbb{B}}_0$ over the appropriate line segment. By induction, after $N$ steps, the above identity yields

$$\begin{aligned}\overline{\mathbb{B}}_0(s,1,x)=\sum_{k=0}^{N-1}\delta(x_k+1)s+\sum_{k=0}^{N-1}\delta(1-\delta)^k\overline{\mathbb{B}}_0\left(0,0,\frac{x_k+1}{(1-\delta)\xi(s_{k+1})+\delta}-1\right)\\+(1-\delta)^N\overline{\mathbb{B}}_0(1,1,x_N)+O(\delta),\end{aligned}\tag{4.4}$$

where $s_k=s/(1-\delta)^k$ and the sequence $(x_k)_{k=0}^N$ is given inductively by $x_0=x$ and

$$x_{k+1}+1=\frac{(x_k+1)\xi(s_{k+1})}{(1-\delta)\xi(s_{k+1})+\delta}.$$

Let us stress that the summand $O(\delta)$ in (4.4) comes from $N\approx\delta^{-1}$ terms $o(\delta)$ appearing in (4.3).

Analogous reasoning to that from the previous case enables to translate the above identities for $\overline{\mathbb{B}}$ into the inductive construction of an appropriate Carleson sequence $(\alpha_Q)_{Q\in\mathcal{T}}$ and a function $f$. Let $A_0\supset A_1\supset A_2\supset\ldots\supset A_N$ be sets given by the same construction as in the previous case. The set $A_N$ has measure $(1-\delta)^N=s$ and the behavior of $E\cap A_N$, $(\alpha_Q)_{Q\in\mathcal{T},Q\subseteq A_N}$ and $f|_{A_N}$ must code the value $\overline{\mathbb{B}}(1,1,x_N)$. So, for any atom $Q$ of $A_N$, we repeat the construction from the previous case (with a different fixed $\delta'$) and "glue" the corresponding sets, sequences and functions (corresponding to different atoms) into one set, one sequence and one function on $A_N$. Clearly, for each $Q$ the set $E$ we take is equal to $Q$; this implies $E \cap A_N = A_N$, and hence we are forced to take $E = A_N$, so that $\mu(E) = s$. To complete the construction, we proceed as follows. If $Q$ is an atom of one of the sets $A_0$, $A_1$, ..., $A_{N-1}$, set $\alpha_Q = \delta$. If $Q \in \mathcal{T}$ does not have this property and is not contained in any atom of $A_N$, put $\alpha_Q = 0$. Finally, the restriction $f$ to $X \setminus A_N$ is given by

$$f|_{X\setminus A_N} = \sum_{k=0}^{N-1} \left( \frac{x_k + 1}{(1-\delta)\xi(s_{k+1}) + \delta} - 1 \right) \chi_{A_k \setminus A_{k+1}}, \tag{4.5}$$

as indicated by the terms appearing under the second sum in (4.4).

The following fact will be useful later.

**Lemma 4.3.** *If $x + 1 = \xi(s)$, then $f$ is nonnegative.*

*Proof.* Fix $s$. Let us first show that for any $k = 0, 1, 2, \ldots, N-1$ we have

$$\xi(s_k) \geq (1-\delta)\xi(s_{k+1}) + \delta. \tag{4.6}$$

To this end, we consider the function

$$F(u) = \xi(u(1-\delta)^{-k}) - \big((1-\delta)\xi(u(1-\delta)^{-k-1}) + \delta\big), \qquad u \in [0, s].$$

We have $F(0) = 0$ and

$$F'(u) = (1-\delta)^{-k}\big(\xi'(u(1-\delta)^{-k}) - \xi'(u(1-\delta)^{-k-1})\big) \geq 0,$$

since $\xi$ is concave, as shown in the proof of Lemma 3.1. This shows $F(s) \geq 0$ and hence (4.6) follows.

Now, to show the nonnegativity of $f$ on $A_N$, we prove inductively the bound

$$x_k + 1 \geq \xi(s_k), \qquad k = 0, 1, 2, \ldots, N. \tag{4.7}$$

When $k = 0$, both sides are equal. Assuming the validity of the above bound for some $k$, we obtain

$$x_{k+1} + 1 = \frac{(x_k+1)\xi(s_{k+1})}{(1-\delta)\xi(s_{k+1}) + \delta} \geq \frac{\xi(s_k)}{(1-\delta)\xi(s_{k+1}) + \delta} \cdot \xi(s_{k+1}) \geq \xi(s_{k+1}),$$

where the last passage is due to (4.6), and hence (4.7) follows. This estimate combined with (4.6) immediately yields $f \geq 0$ on $X \setminus A_N$ (see (4.5)). It also implies the nonnegativity on $A_N$: indeed, plugging $k = N$ we get $x_N + 1 \geq K/(K-1)$ and it suffices to apply Remark 4.2. □

*Step 2. Formal verification of the properties of $E$, $(\alpha_Q)_{Q\in\mathcal{T}}$ and $f$.* We have already checked that $\mu(E) = \mu(A_N) = (1-\delta)^N = s$. The corresponding properties of $(\alpha_Q)_{Q\in\mathcal{T}}$ and $f$ are proved with the same reasoning as previously; we leave the details to the reader.

*Step 3. Checking the value of $\overline{\mathbb{B}}(s, 1, x)$.* Let $N$ be fixed. We know from the analysis of the preceding case that if we let $\delta' \to 0$ (recall that $\delta'$ was the parameter used in the construction of the extremizers on $A_N$), then for each atom $R$ of $A_N$ we have

$$\frac{1}{\mu(R)} \sum_{Q\in\mathcal{T}, Q\subseteq R} \alpha_Q(\langle f\rangle_Q + 1)\mu(E\cap Q) - \frac{K}{\mu(R)} \int_R (\Psi(f) + f)\mathrm{d}\mu \to \overline{\mathbb{B}}(1, 1, x_N),$$

uniformly in $R$ (more precisely, the left-hand side does not depend on $R$, but only on $\delta'$). If we multiply throughout by $\mu(R)$ and sum over all $R$, we get

$$\sum_{\substack{Q\in\mathcal{T},Q\subseteq R,\\ R \text{ atom of } A_N}} \alpha_Q(\langle f\rangle_Q+1)\mu(E\cap Q) - K\int_{A_N}(\Psi(f)+f)\mathrm{d}\mu \to \mu(A_N)\overline{\mathbb{B}}(1,1,x_N)$$
$$= (1-\delta)^N\overline{\mathbb{B}}(1,1,x_N).$$

Furthermore, if $\mathcal{T}' = \{Q\in\mathcal{T} : Q \text{ is not contained in any atom of } A_N\}$, then

$$\begin{aligned}\sum_{Q\in\mathcal{T}'} \alpha_Q(\langle f\rangle_Q+1)\mu(E\cap Q) &= \sum_{k=0}^{N-1}\sum_{\substack{Q \text{ atom}\\ \text{of } A_k}} \delta(x_k+1)\mu(A_N\cap Q)\\ &= \sum_{k=0}^{N-1}\delta(x_k+1)\mu(A_N) = \sum_{k=0}^{N-1}\delta(x_k+1)s\end{aligned}$$

and

$$\begin{aligned}-K\int_{X\setminus A_N}&(f+1)\log(f+1)\mathrm{d}\mu\\ &= \int_{X\setminus A_N}\overline{\mathbb{B}}(0,0,f)\mathrm{d}\mu\\ &= \sum_{k=0}^{N-1}\overline{\mathbb{B}}\left(0,0,\frac{x_k+1}{(1-\delta)\xi(s_{k+1})+\delta}-1\right)\mu(A_k\setminus A_{k+1})\\ &= \sum_{k=0}^{N-1}\delta(1-\delta)^k\overline{\mathbb{B}}\left(0,0,\frac{x_k+1}{(1-\delta)\xi(s_{k+1})+\delta}-1\right).\end{aligned}$$

Putting all the above facts together (i.e., summing the three equations above), we get

$$\begin{aligned}\int_E\Big(\sum_{Q\in\mathcal{T}}&\alpha_Q(\langle f\rangle_Q+1)\chi_Q\Big)\mathrm{d}\mu - K\int_X\big(\Psi(f)+f\big)\mathrm{d}\mu\\ &= \sum_{k=0}^{N-1}\delta(x_k+1)s + \sum_{k=0}^{N-1}\delta(1-\delta)^k\overline{\mathbb{B}}\left(0,0,\frac{x_k+1}{(1-\delta)\xi(s_{k+1})+\delta}-1\right)\\ &\quad+(1-\delta)^N\overline{\mathbb{B}}(1,1,x_N)+o(\delta').\end{aligned}$$

This should be compared to (4.4); if we let $\delta'\to 0$ and then $\delta\to 0$, then the expression

$$\int_E\Big(\sum_{Q\in\mathcal{T}}\alpha_Q(\langle f\rangle_Q+1)\chi_Q\Big)\mathrm{d}\mu - K\int_X\big(\Psi(f)+f\big)\mathrm{d}\mu$$

approaches $\overline{\mathbb{B}}(s,1,x)$; hence $\overline{\mathbb{B}}_0=\overline{\mathbb{B}}$ for all points of the form $(s,1,x)$. Furthermore, the extremizers we have constructed give rise to the formula for the function $B$

studied in Section 3. Since $E = A_N$, we have $E \cap A_k = A_{\max\{k,N\}}$ and hence

$$\begin{aligned}\sum_{Q\in\mathcal{T}} \alpha_Q \mu(E\cap Q) &= \sum_{k=0}^{N-1} \delta\mu(A_N) + \sum_{k=N}^{\infty} \delta\mu(A_k) \\ &= N\delta(1-\delta)^N + (1-\delta)^N \xrightarrow{N\to\infty} -s\log s + s.\end{aligned} \tag{4.8}$$

Consequently, if $\delta' \to 0$ and $N \to \infty$, we get

$$\int_E \Big(\sum_{Q\in\mathcal{T}} \alpha_Q \langle f\rangle_Q \chi_Q\Big) \mathrm{d}\mu - K\int_X \Psi(f)\mathrm{d}\mu \to \overline{\mathbb{B}}(s,1,x) + s\log s - s + Kx,$$

and the limit is precisely $B(s,1,x)$.

4.3. **Extremizers corresponding to $\overline{\mathbb{B}}(s,t,x)$.** Suppose that $t < 1$. We will briefly discuss the case $s \le t$ only; if $s > t$, the reasoning is similar. In this case the argument is very simple: we have

$$\overline{\mathbb{B}}(s,t,x) = (1-t)\overline{\mathbb{B}}(0,0,x-c(s,t,x)) + t\overline{\mathbb{B}}\left(\frac{s}{t}, 1, x + c(s,t,x)\cdot\frac{1-t}{t}\right). \tag{4.9}$$

Let us split $X = A \cup (X\setminus A)$, where $\mu(A) = t$. On the set $X \setminus A$ we set $f \equiv x - c(s,t,x)$ and $E = \emptyset$ (formally, we require $E \cap (X\setminus A) = \emptyset$). Next, if $Q$ is not contained in any atom of $A$, we let $\alpha_Q = 0$ (in particular, $\alpha_X = 0$). On the other hand, on each atom $Q$ of $A$ we use the construction of the preceding subsection, corresponding to $\overline{\mathbb{B}}\left(\frac{s}{t}, 1, x + c(s,t,x)\cdot\frac{1-t}{t}\right)$. Now we glue all the sets, sequences and functions (corresponding to different $Q$'s) into one set, sequence and function (on $A$). Together with the preceding requirements, this defines us the desired candidates for the (almost) extremizers. We conclude by deriving the formula for $B$. For $E$, $(\alpha_Q)_{Q\in\mathcal{T}}$ and $f$ as above, we have, by (4.8), that if $R$ is an atom of $A$, then

$$\frac{1}{\mu(R)} \sum_{Q\in\mathcal{T}, Q\subseteq R} \alpha_Q\mu(E\cap Q) \to -(s/t)\log(s/t) + s/t.$$

Therefore

$$\begin{aligned}\sum_{Q\in\mathcal{T}} \alpha_Q\mu(E\cap Q) = \sum_{Q\in\mathcal{T}, Q\subseteq A} \alpha_Q\mu(E\cap Q) &\to \mu(A)\cdot\big(-(s/t)\log(s/t) + s/t\big) \\ &= -s\log(s/t) + s\end{aligned}$$

and hence

$$\int_E \Big(\sum_{Q\in\mathcal{T}} \alpha_Q \langle f\rangle_Q \chi_Q\Big) \mathrm{d}\mu - K\int_X \Psi(f)\mathrm{d}\mu \to \overline{\mathbb{B}}(s,t,x) + s\log(s/t) - s + Kx,$$

and the limit is precisely $B(s,t,x)$.

4.4. **Sharpness of** (1.4)**.** Clearly, it suffices to prove that $L(K)$ is the best for $K > 1$; since $L(K) \to \infty$ as $K \downarrow 1$, this will automatically imply the sharpness for $K \le 1$ as well. Fix small $\varepsilon > 0$, $s > 0$ and suppose that the inequality (1.4) holds with some constant $L'(K)$ in the place of $L(K)$. Consider the extremizers corresponding to $\overline{\mathbb{B}}(s,1,\xi(s)-1)$. As we have shown in Lemma 4.3 above, the (almost) extremal function $f$ is nonnegative and if we take $\delta$, $\delta'$ (the parameters of the construction) sufficiently small, then

$$L'(K)s \geq \int_E \left( \sum_{Q\in\mathcal{T}} \alpha_Q \langle f\rangle_Q \chi_Q \right) \mathrm{d}\mu - K\int_X \Psi(f)\mathrm{d}\mu > B(s,1,\xi(s)-1) - \varepsilon s.$$

Dividing throughout by $s$ and letting $s \to 0$ gives $L'(K) \geq L(K) - \varepsilon$ (see the last display in the proof of Lemma 3.1). Since $\varepsilon$ was arbitrary, the sharpness follows.

## Acknowledgments

The author would like to thank an anonymous referee for the careful reading of the first version of the paper.

## References

[1] S. M. Buckley, *Estimates for operator norms on weighted spaces and reverse Jensen inequalities*, Trans. Amer. Math. Soc. **340** (1993), 253–272.
[2] D. L. Burkholder, *Boundary value problems and sharp inequalities for martingale transforms*, Ann. Probab. **12** (1984), pp. 647–702.
[3] D. Cruz-Uribe, *The minimal operator and the geometric maximal operator in* $\mathbb{R}^n$, Studia Math. **144** (2001), no. 1, 1–37.
[4] D. Cruz-Uribe, J. M. Martell, and C. Pérez, *Sharp weighted estimates for approximating dyadic operators*, Electron. Res. Announc. Math. Sci. **17** (2010), 12–19.
[5] D. Cruz-Uribe, J. M. Martell, and C. Pérez, *Sharp weighted estimates for classical operators*, Adv. Math. **229** (2012), 408–441.
[6] W. Chen and W. Damián, *Weighted estimates for the multisublinear maximal function*, Rend. Circ. Mat. Palermo (2) **62** (2013), no. 3, 379–391.
[7] T. P. Hytönen, *The sharp weighted bound for general Calderón-Zygmund operators*, Ann. of Math. **175** (2012), 1473–1506.
[8] T. P. Hytönen and C. Pérez, *Sharp weighted bounds involving* $A_\infty$, Anal. PDE **6** (2013), no. 4, 777–818.
[9] M. T. Lacey, E. T. Sawyer, and I. Uriarte-Tuero, *Two Weight Inequalities for Discrete Positive Operators*, ArXiv e-prints, Nov. 2009.
[10] A. K. Lerner. A pointwise estimate for the local sharp maximal function with applications to singular integrals. Bull. Lond. Math. Soc. **42** (2010), 843–856.
[11] M. T. Lacey, S. Petermichl, M. C. Reguera, *Sharp* $A_2$ *inequality for Haar shift operators*, Math. Ann. **348** (2010), no. 1, 127–141.
[12] A. K. Lerner, *On an estimate of Calderón-Zygmund operators by dyadic positive operators*, J. Anal. Math. **121** (2013), 141–161.
[13] A. D. Melas, *The Bellman functions of dyadic-like maximal operators and related inequalities*, Adv. Math. 192 (2005), 310–340.
[14] F. L. Nazarov and S. R. Treil, *The hunt for a Bellman function: applications to estimates for singular integral operators and to other classical problems of harmonic analysis,* St. Petersburg Math. J. **8** (1997), pp. 721–824.
[15] F. L. Nazarov, S. R. Treil and A. Volberg, *The Bellman functions and two-weight inequalities for Haar multipliers*, J. Amer. Math. Soc., **12** (1999), pp. 909–928.
[16] F. L. Nazarov, S. Treil, A. Volberg, *The Tb-theorem on non-homogeneous spaces*, Acta Math. **190** (2003), no. 2, 151–239.
[17] A. Osękowski, *Sharp martingale and semimartingale inequalities*, Monografie Matematyczne **72**, Birkhäuser, 2012.
[18] C. Pérez, S. Treil and A. Volberg, *On* $A_2$ *conjecture and corona decomposition of weights*, arXiv:1006.2630.
[19] S. Petermichl, *The sharp bound for the Hilbert transform on weighted Lebesgue spaces in terms of the classical* $A_p$*-characteristic*, Amer. J. Math. **129** (2007), no. 5, 1355Ũ-1375.
[20] S. Petermichl, *The sharp weighted bound for the Riesz transforms*, Proc. Amer. Math. Soc. **136** (2008), 1237-Ũ1249.

[21] S. Petermichl and A. Volberg, *Heating of the Ahlfors-Beurling operator: weakly quasiregular maps on the plane are quasiregular*, Duke Math. J. **112** (2002), 281-Ũ305.

[22] G. Rey and A. Reznikov, *Extremizers and sharp weak-type estimates for positive dyadic shifts*, Adv. Math. **254** (2014), 664–681.

[23] L. Slavin, A. Stokolos, V. Vasyunin, *Monge-Ampère equations and Bellman functions: The dyadic maximal operator*, C. R. Acad. Sci. Paris, Ser. I 346 (2008), 585–588.

[24] L. Slavin and V. Vasyunin, *Sharp results in the integral-form John-Nirenberg inequality*, Trans. Amer. Math. Soc. 363 (2011), 4135?4169.

[25] L. Slavin and A. Volberg, *Bellman function and the $H^1$-BMO duality*, Harmonic analysis, partial differential equations, and related topics, 113?126, Contemp. Math., 428, Amer. Math. Soc., Providence, RI, 2007.

[26] A. Vagharshakyan, *Recovering singular integrals from Haar shifts*, Proc. Amer. Math. Soc. **138** (2010), 4303–4309.

[27] V. Vasyunin, *The exact constant in the inverse Hölder inequality for Muckenhoupt weights (Russian)*, Algebra i Analiz **15** (2003), 73–117; translation in St. Petersburg Math. J. **15** (2004), 49–79.

[28] V. Vasyunin and A. Volberg, *Monge-Ampére equation and Bellman optimization of Carleson embedding theorems*, Linear and complex analysis, pp. 195?238, Amer. Math. Soc. Transl. Ser. 2, 226, Amer. Math. Soc., Providence, RI, 2009.

Department of Mathematics, Informatics and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland

*E-mail address*: ados@mimuw.edu.pl