Available online at www.sciencedirect.com

ScienceDirect

Journal of Functional Analysis 257 (2009) 1631–1654

JOURNAL OF Functional Analysis

www.elsevier.com/locate/jfa

Dyadic-like maximal operators on $L \log L$ functions

Antonios D. Melas

Department of Mathematics, University of Athens, Panepistimiopolis 15784, Athens, Greece

Received 15 April 2008; accepted 5 June 2009

Available online 27 June 2009

Communicated by J. Bourgain

Abstract

We study the following well-known property of the dyadic maximal operator M_d on $\mathbb{R}^n$ (see [E.M. Stein, Note on the class $L \log L$, Studia Math. 32 (1969) 305–310] for the case of the Hardy–Littlewood maximal function): If ϕ is integrable and supported in a dyadic cube Q then $M_d\phi$ is integrable over sets of finite measure if and only if $|\phi|\log(1+|\phi|)$ is integrable and the integral of $M_d\phi$ can be estimated both from above and from below in terms of the integral of $|\phi|\log(1+|\phi|)$ over Q. Here we define and explicitly evaluate Bellman functions related to this property and the corresponding estimates (both upper and lower) for the integrals thus producing sharp improved versions of the behavior of M_d on the local $L \log L$ spaces.

Keywords: Bellman; Dyadic; Maximal

1. Introduction

It is well known [13] that the Hardy–Littlewood maximal operator M on $\mathbb{R}^n$ has the following property: If ϕ is supported in a ball B then $M\phi$ is integrable over B if and only if $\int_B |\phi|\log(1+|\phi|) < +\infty$ and that there are constants $C_1, C_2, C_1', C_2' > 0$, depending only on B, such that

$$C_1 \int_B |\phi|\log\bigl(1+|\phi|\bigr) - C_1' \leqslant \int_B M\phi \leqslant C_2 \int_B |\phi|\log\bigl(1+|\phi|\bigr) + C_2' \tag{1.1}$$

holds for all such ϕ. An easy scaling argument shows that C_1', C_2' cannot be removed.

E-mail address: amelas@math.uoa.gr.

0022-1236/$ – see front matter
doi:10.1016/j.jfa.2009.06.005

Clearly the same holds for the case of the dyadic maximal operator on $\mathbb{R}^n$

$$M_d\phi(x) = \sup\left\{\frac{1}{|Q|}\int_Q \left|\phi(u)\right| du\colon\ x \in Q,\ Q \subseteq \mathbb{R}^n \text{ is a dyadic cube}\right\}. \tag{1.2}$$

In this paper we will produce sharp versions of the above property for the dyadic maximal operator. First we study the upper bound in (1.1) and we introduce the Bellman type function:

$$\begin{aligned}\mathcal{B}_{\log}(F, f, k) = \sup\Bigg\{\frac{1}{|Q|}\int_E M_d\phi\colon\ &\phi \text{ nonnegative, measurable,}\\ &\mathrm{Av}_Q\big((\phi+1)\log(\phi+1)\big) \leqslant F,\ \mathrm{Av}_Q(\phi) = f,\\ &E \subseteq Q \text{ measurable with } |E| = k\Bigg\}.\end{aligned} \tag{1.3}$$

The exact determination of this will give further information on the deeper analytic properties of the dyadic maximal operator on functions ϕ supported in a dyadic cube and related to the integral of $M_d\phi$ on sets of finite measure (note that $M_d\phi$ outside the cube Q where ϕ is supported is trivially determined depending only on f). Bellman functions relating different norms of ϕ and $M_d\phi$ have been studied extensively in [5]. However the one defined in (1.3) cannot be studied by the methods there. Here we will use a combination of some of the methods from [5] with those in Section 7 of [3] in order to determine it. For more on Bellman functions and their relation to harmonic analysis we refer to [7–9,17]. For the exact evaluation of Bellman functions in certain cases we refer to [2,1,3,5,6,11,12,14–16,10]. We also note the recent approach initiated in [10], and also used in [16], to certain Bellman functions via PDE methods which has given alternative proofs of the results in [3] plus certain more general ones.

Actually as in [3] we will take the more general approach of defining Bellman functions with respect to the maximal operator on a nonatomic probability space (X, μ) equipped with a tree $\mathcal{T}$ (see Section 2) thus defining

$$\begin{aligned}\mathcal{B}^{\mathcal{T}}_{\log}(F, f, k) = \sup\Bigg\{\int_E M_{\mathcal{T}}\phi\, d\mu\colon\ &\phi \geqslant 0 \text{ is measurable with}\\ &\int_X (\phi+1)\log(\phi+1)\, d\mu \leqslant F,\ \int_X \phi\, d\mu = f,\\ &E \subseteq X \text{ is measurable with } \mu(E) = k\Bigg\}.\end{aligned} \tag{1.4}$$

We denote by $\mathcal{D}^{\mathcal{T}}_{\log}(F, f)$ the function $\mathcal{B}^{\mathcal{T}}_{\log}(F, f, 1)$ corresponding to $E = X$ in (1.4) and we will evaluate this first. To describe the result consider the function $V : [1, +\infty) \to [1, +\infty)$ given by

$$V(z) = \frac{e^{z-1}}{z}. \tag{1.5}$$

Clearly this is strictly increasing and we let $U : [1, +\infty) \to [1, +\infty)$ denote the inverse V^{-1} of V. Moreover differentiating $\frac{\exp(U(z)-1)}{U(z)} = z$ we get

$$U'(z) = \frac{U(z)}{z(U(z)-1)} \tag{1.6}$$

on $z > 1$. Then our first main theorem is the following.

Theorem 1. *For any nonatomic probability space (X, μ), any tree $\mathcal{T}$ on (X, μ) and any F, f with $(f+1)\log(f+1) < F$ the corresponding Bellman function is given by*

$$\mathcal{D}^{\mathcal{T}}_{\log}(F, f) = \begin{cases} (f+1)U(\frac{e^{F/(f+1)}}{f+1}) - 1 & \text{if } F < f(f+1), \\ F + f + f\log\frac{F-f}{f^2} & \text{if } f(f+1) \leqslant F, \end{cases} \tag{1.7}$$

where $U : [1, +\infty) \to [1, +\infty)$ is the inverse V^{-1} of V.

After this theorem and since the right-hand side in (1.7) is strictly increasing in F (for each fixed f) it follows easily that we may replace the $\leqslant F$ in the definition (1.4) of $\mathcal{D}^{\mathcal{T}}_{\log}(F, f) = \mathcal{B}^{\mathcal{T}}_{\log}(F, f, 1)$ by $= F$. But we have initially used $\leqslant F$ instead of the usual $= F$ in (1.4) for technical reasons. Also the double formula in (1.7) for the function $\mathcal{D}^{\mathcal{T}}_{\log}(F, f)$ is a phenomenon not appearing in the Bellman functions studied in [3] but in those studied [5] (mixed norms) where it is also explained.

It is easy to see that $U(\frac{e^{F/(f+1)}}{f+1}) < 1 + \frac{F}{f+1}$ whenever $F < f(f+1)$. Thus we get $\mathcal{D}^{\mathcal{T}}_{\log}(F, f) < F + f$ in this case. Of course due to scaling reasons no estimate of the form $\mathcal{D}^{\mathcal{T}}_{\log}(F, f) < C(F + f)$ can hold for all F, f. However using the rough estimate $\log x < x$ for $x = \sqrt{F/f^2} > 1$ in the formula of the case $f(f+1) \leqslant F$ we have the following estimate

$$\mathcal{D}^{\mathcal{T}}_{\log}(F, f) < F + f + \sqrt{F} \tag{1.8}$$

holding for all F, f. This shows that as $F \to +\infty$ and f is fixed the main term in $\mathcal{D}^{\mathcal{T}}_{\log}(F, f)$ is actually F.

Moreover one can verify directly (although the computation is messy) that the function $\mathcal{D}^{\mathcal{T}}_{\log}$ is concave. However it is much more instructive to show the concavity using the Bellman function dynamics of the problem (see [7]). This has to do with the way the main variables F and f as well as $\mathcal{D}^{\mathcal{T}}_{\log}(F, f)$ are behaved when the space (X, μ) is “split” into probability spaces on the children of X in $\mathcal{T}$. We will describe the situation only in the case where $X = [0, 1)$, $\mu = |.|$ is Lebesgue measure and $\mathcal{T}$ consists of all (left closed, right open) dyadic subintervals of X. This is enough since we infer from Theorem 1 that $\mathcal{D}^{\mathcal{T}}_{\log}(F, f)$ is actually independent of $\mathcal{T}$.

We let $X_- = [0, 1/2)$ and $X_+ = [1/2, 1)$ be the two children of X in $\mathcal{T}$ and given $F, f > 0$ with $(f+1)\log(f+1) < F$ and $\phi \geqslant 0$ measurable with $\int_X (\phi+1)\log(\phi+1) = F$ and $\int_X \phi = f$ we denote by $\phi_\pm$ the restrictions of ϕ to the probability spaces $(X_\pm, 2|.|)$ (which are equivalent to $(X, |.|)$) and we let $F_\pm = 2\int_{X_\pm}(\phi_\pm + 1)\log(\phi_\pm + 1)$ and $f_\pm = 2\int_{X_\pm}\phi_\pm$. Then we clearly have

$$F = \frac{1}{2}(F_- + F_+) \quad \text{and} \quad f = \frac{1}{2}(f_- + f_+) \tag{1.9}$$

these equations constituting the Bellman function dynamics of the problem.

On the other hand given $F_\pm, f_\pm > 0$ with $(f_\pm + 1)\log(f_\pm + 1) < F_\pm$ let F, f be defined from (1.9) and for any $\varepsilon > 0$ we choose $\phi_\pm \geqslant 0$ measurable on $X_\pm$ (respectively) satisfying $F_\pm = 2\int_{X_\pm}(\phi_\pm + 1)\log(\phi_\pm + 1)$, $f_\pm = 2\int_{X_\pm}\phi_\pm$ and $2\int_{X_\pm} M_{\mathcal{T}_\pm}\phi_\pm > \mathcal{D}^{\mathcal{T}_\pm}_{\log}(F_\pm, f_\pm) - \varepsilon$, where $\mathcal{T}_\pm$ are the subtrees of dyadic subintervals of $X_\pm$. But $(X_\pm, 2|.|)$ are equivalent to $(X, |.|)$ so $\mathcal{D}^{\mathcal{T}_\pm}_{\log} = \mathcal{D}^{\mathcal{T}}_{\log}$ and so considering the function ϕ which is equal to ϕ_- on X_- and to ϕ_+ on X_+ we easily get

$$\begin{aligned}
\mathcal{D}^{\mathcal{T}}_{\log}(F, f) &\geqslant \int_X M_{\mathcal{T}}\phi = \int_{X_-} M_{\mathcal{T}_-}\phi_- + \int_{X_+} M_{\mathcal{T}_+}\phi_+ \\
&> \frac{1}{2}\mathcal{D}^{\mathcal{T}_-}_{\log}(F_-, f_-) + \frac{1}{2}\mathcal{D}^{\mathcal{T}_+}_{\log}(F_+, f_+) - \varepsilon \\
&= \frac{1}{2}\big(\mathcal{D}^{\mathcal{T}}_{\log}(F_-, f_-) + \mathcal{D}^{\mathcal{T}}_{\log}(F_+, f_+)\big) - \varepsilon
\end{aligned}$$

which as $\varepsilon \to 0^+$ and in view of (1.9) implies that $\mathcal{D}^{\mathcal{T}}_{\log}(F, f)$ (for this and hence for any tree $\mathcal{T}$) is concave.

Next, using Theorem 1, we evaluate the function $\mathcal{B}^{\mathcal{T}}_{\log}(F, f, k)$. Given $0 < k \leqslant 1$ we define

$$\tau_k(x) = \log\frac{1}{x+1} + \frac{x(x+k)}{k(x+1)} \tag{1.10}$$

on $x \geqslant 0$. Since $\tau_k'(x) = \frac{x(x+1-k)}{k(x+1)^2} > 0$ on $x > 0$ and $\tau_k(0) = 0$ we let $T_k : [0, +\infty) \to [0, +\infty)$ denote the inverse function of τ_k. If $F, f > 0$ are such that $F > (f + 1)\log(f + 1)$ it is clear that $T_k(\frac{F}{f+1} - \log(f + 1)) < f$ if and only if $\tau_k(f) > \frac{F}{f+1} - \log(f + 1)$ which is equivalent to $f(\frac{f}{k} + 1) > F$. If this happens we write $\xi_k(F, f) = T_k(\frac{F}{f+1} - \log(f + 1)) \in (0, f)$. Then we can state the following.

Theorem 2. *Given $F, f, k > 0$ with $k \leqslant 1$ and $(f + 1)\log(f + 1) < F$ we have*:

$$\mathcal{B}^{\mathcal{T}}_{\log}(F, f, k) = \begin{cases} \frac{(f+1)(\xi_k(F,f)+k)}{\xi_k(F,f)+1} U\big(\frac{ke^{\xi_k(F,f)/k}}{\xi_k(F,f)+k}\big) - k & \text{if } F < f(\frac{f}{k} + 1), \\ F + f + f\log\frac{k(F-f)}{f^2} & \text{if } f(\frac{f}{k} + 1) \leqslant F. \end{cases} \tag{1.11}$$

It is easy to see, noting that the equation satisfied by $\xi = \xi_1(F, f)$ is $\xi - \log(\xi + 1) = \frac{F}{f+1} - \log(f + 1)$ that by taking $k = 1$ in (1.11) one obtains (1.7). However we have stated Theorem 1 first since it constitutes an essential step in proving the more general Theorem 2.

Now we turn to sharp forms of the lower estimate in (1.1). We will see that no nontrivial lower estimate exists for all trees $\mathcal{T}$. Because of that we will restrict our attention to N-homogeneous trees that resemble the dyadic tree in $\mathbb{R}^n$ which is 2^n-homogeneous (the definition is given in the next section). Then assuming that $\mathcal{T}$ is N-homogeneous on the probability space (X,μ) we define the following function

$$\mathcal{L}^{\mathcal{T}}(F,f)=\inf\left\{\int_X M_{\mathcal{T}}\phi\,d\mu\colon\ \phi\geqslant 0 \text{ is measurable}, \int_X \phi\,d\mu=f \text{ and } \int_X \phi\log^+\frac{\phi}{f}\,d\mu=F\right\}. \tag{1.12}$$

Then we have the following.

Theorem 3. *If the tree $\mathcal{T}$ is N-homogeneous and $F,f>0$ then*

$$\mathcal{L}^{\mathcal{T}}(F,f)=\frac{N-1}{N\log N}F+f. \tag{1.13}$$

Thus in particular for the dyadic maximal operator in $\mathbb{R}^n$ we get for any $\phi\geqslant 0$ measurable and supported in the cube $Q_0=[0,1]^n$ that the following sharp estimate holds

$$\int_{Q_0} M_d\phi\geqslant\frac{2^n-1}{2^n n\log 2}\int_{Q_0}\left(\phi\log^+\frac{\phi}{\|\phi\|_1}\right)+\|\phi\|_1. \tag{1.14}$$

Also by taking $N\to\infty$ we conclude that there is no uniform lower estimate, other than the trivial one $\int_X M_{\mathcal{T}}\phi\,d\mu\geqslant\int_X\phi\,d\mu$, holding for all trees $\mathcal{T}$.

Theorem 3 provides an example where an infimum Bellman type function is evaluated. For another interesting example related to the dyadic Carleson embedding theorem we refer to [16]. Note however that $\mathcal{L}^{\mathcal{T}}(F,f)$ does not satisfy the usual Bellman type dynamics (that is (1.9) in our case) since the definition of F involves the variable f (see (1.12)).

We have chosen the growth functions in the above theorems in order to get results that are readable (this would not be the case if we had chosen $x\log(1+x)$ instead). Of course these combined with the L^1 norm are equivalent size conditions on ϕ.

In Section 2 we give a general procedure that can be used to evaluate Bellman functions involving the integral of $M_{\mathcal{T}}\phi$. We are not aiming at a general theory as in [5] but rather on a more direct computation scheme that can be used for specific growth functions, and especially for the ones like $x\log x$ where the theory in [5] does not apply. As applications other than the proof of Theorem 1, which is given in Section 3, we will compute here the corresponding to (1.12) supremum Bellman function as well as one related to the L^∞ norm of ϕ (the last one has been also found in [5]). Other applications of this will be given elsewhere. In Section 3 we will also prove Theorem 2 by a detailed study of the function in (1.7) combined with certain methods from [3]. Then in Section 4 we will prove Theorem 3.

2. Trees and maximal operators

As in [3] we let (X,μ) be a nonatomic probability space (i.e. $\mu(X)=1$). Then we give the following.

Definition 1. (a) 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 $I \in \mathcal{T}$ we have $\mu(I) > 0$.
(ii) For every $I \in \mathcal{T}$ there corresponds a finite subset $\mathcal{C}(I) \subseteq \mathcal{T}$ containing at least two elements such that the elements of $\mathcal{C}(I)$ are pairwise disjoint subsets of I and $I = \bigcup \mathcal{C}(I)$.
(iii) $\mathcal{T} = \bigcup_{m \geqslant 0} \mathcal{T}_{(m)}$ where $\mathcal{T}_{(0)} = \{X\}$ and $\mathcal{T}_{(m+1)} = \bigcup_{I \in \mathcal{T}_{(m)}} \mathcal{C}(I)$.
(iv) $\lim_{m\to\infty} \sup_{I \in \mathcal{T}_{(m)}} \mu(I) = 0$.

(b) A tree $\mathcal{T}$ on (X, μ) will be called N-homogeneous (where $N > 1$ is an integer) if it satisfies the following additional conditions:

(i) For every $I \in \mathcal{T}$ the set $\mathcal{C}(I)$ consists of exactly N elements of $\mathcal{T}$ each having measure equal to $N^{-1}\mu(I)$.
(ii) The family $\mathcal{T}$ differentiates $L^1(X, \mu)$.

We remark that the above definition can be given under the assumption that the elements of each $\mathcal{C}(I)$ are only pairwise almost disjoint, that is if $A, B \in \mathcal{C}(I)$ and $A \neq B$ then $\mu(A \cap B) = 0$. However by considering $X \setminus E(\mathcal{T})$, where $E(\mathcal{T}) = \bigcup_{I\in\mathcal{T}} \bigcup_{J_1, J_2 \in \mathcal{C}(I),\, J_1 \neq J_2} (J_1 \cap J_2)$ clearly has measure 0, the above makes no difference.

Examples. 1. If Q_0 is the unit cube $\mathbb{R}^n$ we let E be the union of all the boundaries of all dyadic cubes in Q_0 then let $X = Q_0 \setminus E$ and $\mathcal{T}$ be the set of all open dyadic cubes $Q \subseteq Q_0$. Here $N = 2^n$ and each $\mathcal{C}(Q)$ is the set of the 2^n subcubes of Q obtained by bisecting its sides. More generally for any integer $m > 1$ we may consider all m-adic cubes $Q \subseteq Q_0$ with $\mathcal{C}(Q)$ being the set of the m^n open subcubes of Q obtained by dividing each side of it into m equal parts.

2. Given the integers $d_1, \ldots, d_n \geqslant 1$ and $m > 1$ we can define $\mathcal{T}$ on X equal to Q_0 minus a certain set of measure 0 by setting for each open parallelepiped R the family $\mathcal{C}(R)$ to consist of the open parallelepipeds formed by dividing the dimensions of R into $m^{d_1}, \ldots, m^{d_n}$ equal parts respectively. For example if $n = 2$, $m = 2$, $d_1 = 1$ and $d_2 = 2$ we get the set of dyadic parabolic rectangles contained in $[0, 1]^2$.

3. The family of rectangles $\{[0, 1) \times I\colon I$ is a dyadic subinterval of $[0, 1)\}$ on the probability space $[0, 1)^2$ equiped with the Lebesgue measure is a tree that satisfies condition (i) of Definition 1(b) with $N = 2$ but is not 2-homogeneous since it does not satisfy condition (ii) of the same definition.

An easy induction shows that each family $\mathcal{T}_{(m)}$ consists of pairwise disjoint sets whose union is X. Moreover if $x \in X \setminus E(\mathcal{T})$ then for each m there exists exactly one $I_m(x)$ in $\mathcal{T}_{(m)}$ containing x. For every $m > 0$ there is a $J \in \mathcal{T}_{(m-1)}$ such that $I_m(x) \in \mathcal{C}(J)$. Since then $x \in J$ we must have $J = I_{m-1}(x)$. Hence the set $\mathcal{A}(x) = \{I \in \mathcal{T}\colon x \in I\}$ forms a chain $I_0(x) = X \supsetneq I_1(x) \supsetneq \cdots$ with $I_m(x) \in \mathcal{C}(I_{m-1}(x))$ for every $m > 0$. From this remark it easily follows that if $I, J \in \mathcal{T}$ and $I \cap J$ is nonempty then $I \subseteq J$ or $J \subseteq I$. In particular for any $I, J \in \mathcal{T}$ we have either $I \cap J$ is empty or one of them is contained in the other. The condition (ii) in Definition 1(b) can now be described as follows:

"For any $\psi \in L^1(X, \mu)$ we have $\lim_{m\to\infty} \frac{1}{\mu(I_m(x))} \int_{I_m(x)} \psi \, d\mu = \psi(x)$ for μ-almost every x in X."

This condition will be needed only in Theorem 3 (see Section 4) and all other results in this paper hold without this assumption. In Example 3 above it is easy to see that one can construct functions ϕ with $\int_0^1 \phi(x,y)\,dx = 1$ for all y but with $\int_{[0,1)^2} \phi \log^+ \phi$ arbitrary large. Thus Theorem 3 cannot hold in this case. However Theorems 1 and 2 hold.

The following lemma gives another property of $\mathcal{T}$ that will be useful later. For a proof see [3].

Lemma 1. *For every $I \in \mathcal{T}$ and every α such that $0 < \alpha < 1$ there exists a subfamily $\mathcal{F}(I) \subseteq \mathcal{T}$ consisting of pairwise disjoint subsets of I such that $\mu(\bigcup_{J\in\mathcal{F}(I)} J) = \sum_{J\in\mathcal{F}(I)} \mu(J) = (1-\alpha)\mu(I)$.*

Next let $\mathcal{S}$ be a finite subset of $\mathcal{T}$ such that $X \in \mathcal{S}$. For any $I \in \mathcal{S}$ with $I \neq X$ we let I^* denote the unique minimal ancestor of I in $\mathcal{S}$ (i.e. the minimal element of $\{J \in \mathcal{S}\colon\ I \subsetneq J\}$) and setting

$$A_I = I \setminus \bigcup_{J\in\mathcal{S}:\ J^*=I} J, \qquad a_I = \mu(A_I), \tag{2.1}$$

we easily get

$$I = \bigcup_{\mathcal{S}\ni J\subseteq I} A_J \quad \text{and so} \quad \mu(I) = \sum_{\mathcal{S}\ni J\subseteq I} a_J \tag{2.2}$$

for any $I \in \mathcal{S}$ (if I is a minimal element of $\mathcal{S}$ then clearly $A_I = I$).

Then we will need the following (this is a special case of Theorem 1 in [3] but we include it here since its proof is much simpler).

Lemma 2. *For any finite tree $\mathcal{S}$ and any increasing and convex function $\Psi : [0,+\infty) \to \mathbb{R}$ we have*

$$\sum_{I\in\mathcal{S}} a_I \Psi\left(\sum_{I\subseteq J} \frac{a_J}{\mu(J)}\right) \leqslant \int_0^\infty \Psi(u) e^{-u}\, du. \tag{2.3}$$

Proof. Since

$$\Psi(x) = \Psi(0) + \Psi_r'(0)x + \int_0^\infty (x-\lambda)^+ \, d\Psi_r'(\lambda) \tag{2.4}$$

where $x^+ = \max(x,0)$ and Ψ_r' denotes the right derivative of Ψ, it suffices to prove (2.3) when $\Psi(x) = \Psi_\lambda(x) = (x-\lambda)^+$ where $\lambda \geqslant 0$ is fixed. In this case (2.3) reads

$$\sum_{I\in\mathcal{S}} a_I \left(\sum_{I\subseteq J} \frac{a_J}{\mu(J)} - \lambda\right)^+ \leqslant e^{-\lambda}. \tag{2.5}$$

We will now prove (2.5) for all λ by induction on the size of $\mathcal{S}$.

If $\mathcal{S} = \{X\}$ then $a_X = 1$ and so (2.5) becomes $(1-\lambda)^+ \leqslant e^{-\lambda}$ which holds since $e^{-x} > 1-x$ whenever $0 < x < 1$. Now assuming (2.5) for any $\lambda \geqslant 0$ and any tree (on any $(X,\mu,\mathcal{T})$)

having less elements than $\mathcal{S}$ we let $\{J_1,\dots,J_k\}$ be all the elements J of $\mathcal{S}$ with $J^*=X$. Then when $\lambda \geqslant a_X$ the induction hypothesis applied to the subtrees of $\mathcal{S}$ with tops $J_1,\dots,J_k$ on the probability spaces $(J_i,\frac{1}{\mu(J_i)}\mu)$ and with $\lambda - a_X \geqslant 0$ instead of λ gives

$$B_i = \sum_{\substack{I\in\mathcal{S}\\ I\subseteq J_i}} a_I\left(\sum_{I\subseteq J\subseteq J_i}\frac{a_J}{\mu(J)} - (\lambda - a_X)\right)^+ \leqslant \mu(J_i)e^{-\lambda + a_X} \tag{2.6}$$

for any i and so

$$\sum_{I\in\mathcal{S}} a_I\left(\sum_{I\subseteq J}\frac{a_J}{\mu(J)} - \lambda\right)^+ = a_X(a_X-\lambda)^+ + \sum_{i=1}^k B_i = \sum_{i=1}^k B_i \leqslant \sum_{i=1}^k \mu(J_i)e^{-\lambda+a_X} = (1-a_X)e^{-\lambda+a_X} < e^{-\lambda}. \tag{2.7}$$

On the other hand if $0\leqslant \lambda < a_X < 1$ the left-hand side in (2.5) becomes

$$\sum_{I\in\mathcal{S}} a_I\left(\sum_{I\subseteq J}\frac{a_J}{\mu(J)} - \lambda\right) = \sum_{J\in\mathcal{S}}\frac{a_J}{\mu(J)}\sum_{I\subseteq J} a_I - \lambda\sum_{I\in\mathcal{S}} a_I = \sum_{J\in\mathcal{S}} a_J - \lambda = 1-\lambda < e^{-\lambda} \tag{2.8}$$

and this completes the induction. □

Now given any tree $\mathcal{T}$ we define the maximal operator associated to it as follows

$$M_{\mathcal{T}}\psi(x) = \sup\bigl\{\mathrm{Av}_I\bigl(|\psi|\bigr):\ x\in I\in\mathcal{T}\bigr\} \tag{2.9}$$

for every $\psi\in L^1(X,\mu)$ where for any nonnegative $\phi\in L^1(X,\mu)$ and for any $I\in\mathcal{T}$ we have written $\mathrm{Av}_I(\phi) = \frac{1}{\mu(I)}\int_I \phi\,d\mu$.

Given an integer $m>0$ and $\lambda_P \geqslant 0$ for each $P\in\mathcal{T}_{(m)}$ we consider the function ϕ given by

$$\phi = \sum_{P\in\mathcal{T}_{(m)}} \lambda_P\chi_P \tag{2.10}$$

(where χ_P denotes the characteristic function of P). For every $x\in X$ we let $I_\phi(x)$ denote the unique *largest* element of the set $\{I\in\mathcal{T}:\ x\in I$ and $M_{\mathcal{T}}\phi(x) = \mathrm{Av}_I(\phi)\}$ (which is nonempty since $\mathrm{Av}_J(\phi) = \mathrm{Av}_P(\phi)$ whenever $P\in\mathcal{T}_{(m)}$ and $J\subseteq P$). Next for any $I\in\mathcal{T}$ we define the set

$$A_I = A(\phi,I) = \bigl\{x\in X:\ I_\phi(x) = I\bigr\} \tag{2.11}$$

and we let $\mathcal{S} = \mathcal{S}_\phi$ denote the set of all $I \in \mathcal{T}$ such that A_I is nonempty. It is clear that each such A_I is a union of certain P's from $\mathcal{T}_{(m)}$ and moreover

$$M_{\mathcal{T}}\phi = \sum_{I \in \mathcal{S}} \mathrm{Av}_I(\phi)\chi_{A_I}. \tag{2.12}$$

We also define the correspondence $I \to I^*$ with respect to $\mathcal{S}$ as before. This is defined for every I in $\mathcal{S}$ that is not maximal with respect to $\subseteq$. We also write $y_I = \mathrm{Av}_I(\phi)$ for every $I \in \mathcal{S}$.

The main properties of the above are given in the following (see also [3] and [4]).

Lemma 3.

(i) *For every $I \in \mathcal{S}$ we have $I = \bigcup_{\mathcal{S} \ni J \subseteq I} A_J$.*

(ii) *For every $I \in \mathcal{S}$ we have $A_I = I \setminus \bigcup_{J \in \mathcal{S}\colon J^* = I} J$ and so $\mu(A_I) = \mu(I) - \sum_{J \in \mathcal{S}\colon J^*=I} \mu(J)$ and $\mathrm{Av}_I(\phi) = \frac{1}{\mu(I)} \sum_{J \in \mathcal{S}\colon J \subseteq I} \int_{A_J} \phi\, d\mu$.*

(iii) *For an $I \in \mathcal{T}$ we have $I \in \mathcal{S}$ if and only if $\mathrm{Av}_Q(\phi) < \mathrm{Av}_I(\phi)$ whenever $I \subseteq Q \in \mathcal{T}$, $I \neq Q$. In particular $X \in \mathcal{S}$ and so $I \to I^*$ is defined for all $I \in \mathcal{S}$ such that $I \neq X$.*

(iv) *If $I, J \in \mathcal{S}$ are such that $J^* = I$ then*

$$y_I < y_J \leqslant \frac{\mu(F)}{\mu(J)} y_I \tag{2.13}$$

where F is the unique element of the whole tree $\mathcal{T}$ such that $J \in \mathcal{C}(F)$. In particular if $\mathcal{T}$ is N-homogeneous then $y_I < y_J \leqslant N y_I$.

Proof. (i) Clearly $X = \bigcup_{J \in \mathcal{S}} A_J$. Fix $I \in \mathcal{S}$. Supposing that $x \in A(\phi, J) \cap I$ for some J we have $x \in I \cap J \neq \emptyset$ and so either $I \subseteq J$ or $J \subseteq I$. Suppose now that $I \subsetneq J$. Then also $\mathrm{Av}_J(\phi) = M_{\mathcal{T}}\phi(x) \geqslant \mathrm{Av}_I(\phi)$ and so I cannot be an $I_\phi(z)$ for any $z \in I$. Therefore $A(\phi, I) = \emptyset$ contradicting the assumption $I \in \mathcal{S}$. Hence we must have $J \subseteq I$ and this easily implies that I is the union of all A_J's for $J \subseteq I$.

(ii) Follows easily from (i).

(iii) One direction follows from the definition of the I_ϕ's. For the other assume that $I \in \mathcal{T}_{(s)}$ satisfies the assumption. Since

$$\mathrm{Av}_J(\phi) = \frac{\sum_{F \in \mathcal{C}(J)} \mu(F)\,\mathrm{Av}_F(\phi)}{\sum_{F \in \mathcal{C}(J)} \mu(F)} \tag{2.14}$$

we conclude that for each $J \in \mathcal{T}$ there exists $J' \in \mathcal{C}(J)$ such that $\mathrm{Av}_{J'}(\phi) \leqslant \mathrm{Av}_J(\phi)$. Starting from I and applying the above $m - s$ times we get a chain $I = I_0 \supseteq I_1 \supseteq \cdots \supseteq I_{m-s}$ such that $I_{(r)} \in \mathcal{T}_{(s+r)}$ for each s and moreover $\mathrm{Av}_{I_{m-s}}(\phi) \leqslant \mathrm{Av}_{I_{m-s+1}}(\phi) \leqslant \cdots \leqslant \mathrm{Av}_{I_1}(\phi) \leqslant \mathrm{Av}_{I_0}(\phi) = \mathrm{Av}_I(\phi)$. Now from this and the assumption on I it is clear that $I_\phi(x) = I$ for every $x \in I_{m-s}$ and therefore $I \in \mathcal{S}$.

(iv) The inequality $y_I < y_J$ follows from (iii). For the other first note that clearly $F \subseteq I$. We claim that $\mathrm{Av}_F(\phi) \leqslant y_I$. Indeed $I \in \mathcal{S}$ implies that $\mathrm{Av}_Q(\phi) < y_I$ whenever $I \subseteq Q$, $I \neq Q$ and so if $\mathrm{Av}_F(\phi) > y_I$ there would exist $F' \in \mathcal{T}$ such that $F \subseteq F' \subseteq I$, $F' \neq I$ and $\mathrm{Av}_{F'}(\phi) > \mathrm{Av}_Q(\phi)$

whenever $F' \subseteq Q$, $F' \neq Q$. But this combined with (iii) implies that F' must be in $\mathcal{S}$ contradicting our assumption $J^* = I$. Thus we get since $J \subseteq F$

$$y_J = \frac{1}{\mu(J)} \int_J \phi \, d\mu \leqslant \frac{1}{\mu(J)} \int_F \phi \, d\mu = \frac{\mu(F)}{\mu(J)} \mathrm{Av}_F(\phi) \leqslant \frac{\mu(F)}{\mu(J)} y_I \tag{2.15}$$

which completes the proof. □

The above lemma shows that this linearization $M_{\mathcal{T}}\phi$ may be viewed as a multiscale version of the classical Calderon–Zygmund decomposition.

Now writing $a_I = \mu(A_I)$ and $x_I = a_I^{-1} \int_{A_I} \phi \, d\mu$ for every $I \in \mathcal{S}$ the above lemma and (2.12) imply that:

$$M_{\mathcal{T}} \phi = \sum_{I \in \mathcal{S}} \left(\frac{1}{\mu(I)} \sum_{J \in \mathcal{S},\, J \subseteq I} a_J x_J \right) \chi_{A_I} \quad \text{and} \quad \int_X \phi \, d\mu = \sum_{I \in \mathcal{S}} a_I x_I. \tag{2.16}$$

Next let $G : [0, +\infty) \to [0, +\infty]$ and $\Psi : [0, +\infty) \to [0, +\infty)$ be two convex and increasing functions such that $G(0) = \Psi(0) = 0$ and G is the Legendre transform of Ψ, that is for every $x > 0$ we have

$$G(x) = \sup_{y > 0} \bigl(xy - \Psi(y) \bigr). \tag{2.17}$$

Note that G is allowed to be extended-valued (but Ψ is not). We thus have

$$xy \leqslant G(x) + \Psi(y) \tag{2.18}$$

for all $x, y \geqslant 0$. Moreover we define $\Psi'_+ : \mathbb{R} \to [0, +\infty)$ to be the function which is equal to the right derivative Ψ'_r of Ψ on $[0, +\infty)$ and to 0 on $(-\infty, 0)$. Noting that for any $x, z \geqslant 0$ we have $x\Psi'_r(z) - \Psi(x) \leqslant z\Psi'_r(z) - \Psi(z)$ we conclude that

$$u^+ \Psi'_+(u) = \Psi\bigl(u^+\bigr) + G\bigl(\Psi'_+(u)\bigr) \tag{2.19}$$

for all $u \in \mathbb{R}$.

We define also the following Bellman function related to G

$$\mathcal{D}_G^{\mathcal{T}}(F, f) = \sup \left\{ \int_X M_{\mathcal{T}} \phi \, d\mu \text{: } \phi \geqslant 0 \text{ is measurable}, \right.$$
$$\left. \int_X G \circ \phi \, d\mu \leqslant F, \int_X \phi \, d\mu = f \right\} \tag{2.20}$$

when $0 < f$ and $G(f) < F$.

Using the decomposition of $M_{\mathcal{T}}\phi$ (and ϕ) given in (2.16) we can now prove the following:

Lemma 4. *Given a nonnegative function ϕ of the form* (2.10) *with $\int_X \phi\, d\mu = f$ and $\int_X G \circ \phi\, d\mu \leqslant F$ and given any $c > 0$, $\lambda \in \mathbb{R}$ we have*

$$\int_X M_{\mathcal{T}}\phi\, d\mu \leqslant \frac{F}{c} + \frac{1}{c}\int_0^\infty \Psi\big(c(u+\lambda)^+\big)e^{-u}\, du - \lambda f. \tag{2.21}$$

Proof. Using the above notation we note that by Jensen's inequality

$$\sum_{I\in\mathcal{S}} a_I G(x_I) \leqslant \sum_{I\in\mathcal{S}} \int_{A_I} G(\phi)\, d\mu \leqslant F. \tag{2.22}$$

Now if $\lambda \geqslant 0$ we use (2.18) and Lemma 2 to write

$$\begin{aligned}
c\left(\int_X M_{\mathcal{T}}\phi\, d\mu + \lambda f\right) &= c\sum_{J\in\mathcal{S}} \frac{a_J}{\mu(J)} a_I x_I \sum_{J\in\mathcal{S},\, I\subseteq J} a_I x_I + c\lambda f \\
&= \sum_{I\in\mathcal{S}} c a_I x_I \left(\sum_{J\in\mathcal{S},\, I\subseteq J} \frac{a_J}{\mu(J)} + \lambda\right) \\
&\leqslant \sum_{I\in\mathcal{S}} a_I G(x_I) + \sum_{I\in\mathcal{S}} a_I \Psi\left(c\sum_{J\in\mathcal{S},\, I\subseteq J} \frac{a_J}{\mu(J)} + c\lambda\right) \\
&\leqslant F + \int_0^\infty \Psi\big(c(u+\lambda)\big)e^{-u}\, du.
\end{aligned} \tag{2.23}$$

If $\lambda < 0$ then we write $\mathcal{S}^* = \{I \in \mathcal{S}\colon \sum_{J\in\mathcal{S},\, I\subseteq J} \frac{a_I}{\mu(I)} > -\lambda\}$ and $f^* = \sum_{I\in\mathcal{S}^*} a_I x_I$ to get as in (2.23)

$$\begin{aligned}
&c\left[\sum_{I\in\mathcal{S}^*} a_I x_I \left(\sum_{J\in\mathcal{S},\, I\subseteq J} \frac{a_J}{\mu(J)}\right)\right] + \lambda f^* \\
&\quad\leqslant \sum_{I\in\mathcal{S}^*} a_I G(x_I) + \sum_{I\in\mathcal{S}^*} a_I \Psi\left(c\sum_{J\in\mathcal{S},\, I\subseteq J} \frac{a_J}{\mu(J)} + c\lambda\right) \\
&\quad\leqslant F + \sum_{I\in\mathcal{S}} a_I \Psi\left(c\left(\sum_{J\in\mathcal{S},\, I\subseteq J} \frac{a_J}{\mu(J)} + \lambda\right)^+\right) \\
&\quad\leqslant F + \int_0^\infty \Psi\big(c(u+\lambda)^+\big)e^{-u}\, du
\end{aligned} \tag{2.24}$$

using Lemma 2 for the convex increasing function $x \to \Psi(c(x+\lambda)^+)$, whereas

$$c \sum_{I \in \mathcal{S} \setminus \mathcal{S}^*} a_I x_I \left(\sum_{J \in \mathcal{S},\, I \subseteq J} \frac{a_J}{\mu(J)} \right) \leqslant -c\lambda \sum_{I \in \mathcal{S} \setminus \mathcal{S}^*} a_I x_I = -c\lambda\big(f - f^*\big). \tag{2.25}$$

Adding now (2.24) and (2.25) and using the first equality in (2.23) we get (2.21). □

Using this we have the following.

Proposition 1. *Let $f, F > 0$ be given such that $G(f) < F$. Assume that $c > 0$ and $\lambda \in \mathbb{R}$ satisfy*

$$\int_0^\infty \Psi'_+\big(c(u+\lambda)\big)e^{-u}\,du = f \quad \text{and} \quad \int_0^\infty G \circ \Psi'_+\big(c(u+\lambda)\big)e^{-u}\,du = F. \tag{2.26}$$

Then we have

$$\mathcal{D}_G^{\mathcal{T}}(F, f) = \int_0^\infty u\Psi'_+\big(c(u+\lambda)\big)e^{-u}\,du. \tag{2.27}$$

Proof. Taking any $\theta > 0$ we define $\alpha = \alpha(\theta) = 1 - e^{-\theta} \in (0,1)$ and as in [3], using Lemma 1, we choose for every $I \in \mathcal{T}$ a family $\mathcal{F}(I) \subseteq \mathcal{T}$ of pairwise disjoint subsets of I such that $\sum_{J \in \mathcal{F}(I)} \mu(J) = (1-\alpha)\mu(I)$. Then we define $\mathcal{S} = \mathcal{S}_\alpha$ to be the smallest subset of $\mathcal{T}$ such that $X \in \mathcal{S}$ and for every $I \in \mathcal{S}$, $\mathcal{F}(I) \subseteq \mathcal{S}$. It is clear that defining the correspondence $I \to I^*$ with respect to this $\mathcal{S}$ we have $J^* = I \in \mathcal{S}$ if and only if $J \in \mathcal{F}(I)$ and so writing $A_I = I \setminus \bigcup_{J \in \mathcal{S}:\ J^*=I} J$ we have $a_I = \mu(A_I) = \mu(I) - \sum_{J \in \mathcal{S}:\ J^* = I} \mu(J) = \alpha\mu(I)$ for every $I \in \mathcal{S}$. We define $\mathrm{rank}(I) = r(I)$ of any $I \in \mathcal{S}$ to be the unique integer m such that $I \in \mathcal{S}_{(m)}$ and we define the x_I's by setting

$$x_I = \gamma_m = \frac{1}{\alpha(1-\alpha)^m} \int_{m\theta}^{(m+1)\theta} \Psi'_+\big(c(u+\lambda)\big)e^{-u}\,du \tag{2.28}$$

for every $I \in \mathcal{S}$ where $m = \mathrm{rank}(I)$ and let

$$\phi_\theta = \sum_{I \in \mathcal{S}} x_I \chi_{A_I}. \tag{2.29}$$

For every $I \in \mathcal{S}$ and every $m \geqslant 0$ we have

$$b_m(I) = \sum_{\substack{\mathcal{S} \ni J \subseteq I \\ r(J) = r(I)+m}} \mu(J) = (1-\alpha)^m \mu(I) \tag{2.30}$$

hence

$$\int_X \phi_\theta \, d\mu = \sum_{I \in \mathcal{S}} a_I x_I = \sum_{m \geqslant 0} \sum_{I \in \mathcal{S}_{(m)}} \gamma_m \alpha \mu(J) = \alpha \sum_{m \geqslant 0} \gamma_m b_m(X)$$

$$= \alpha \sum_{m \geqslant 0} \gamma_m (1-\alpha)^m = \int_0^\infty \Psi'_+\big(c(u+\lambda)\big) e^{-u} \, du = f \tag{2.31}$$

and by Jensen's inequality

$$\int_X G(\phi_\theta) \, d\mu = \sum_{I \in \mathcal{S}} a_I G(x_I) = \alpha \sum_{m \geqslant 0} G(\gamma_m)(1-\alpha)^m$$

$$\leqslant \int_0^\infty G\big(\Psi'_+\big(c(u+\lambda)\big)\big) e^{-u} \, du = F. \tag{2.32}$$

On the other hand if $I \in \mathcal{S}$ and $m = \operatorname{rank}(I)$

$$\begin{aligned}
\operatorname{Av}_I(\phi_\theta) &= \frac{1}{\mu(I)} \sum_{J \in \mathcal{S}:\ J \subseteq I} a_J x_J \\
&= \frac{\alpha}{\mu(I)} \sum_{\ell \geqslant 0} \gamma_{\ell + \operatorname{rank}(I)} \sum_{\substack{\mathcal{S} \ni J \subseteq I \\ \operatorname{rank}(J) = \operatorname{rank}(I) + \ell}} \mu(J) \\
&= \alpha \sum_{\ell \geqslant 0} \gamma_{\ell+m} (1-\alpha)^\ell = \frac{1}{(1-\alpha)^m} \int_{m\theta}^\infty \Psi'_+\big(c(u+\lambda)^+\big) e^{-u} \, du
\end{aligned} \tag{2.33}$$

and so

$$\begin{aligned}
\int_X M_{\mathcal{T}} \phi_\theta \, d\mu &\geqslant \sum_{I \in \mathcal{S}} a_I \operatorname{Av}_I(\phi_\theta) \\
&= \sum_{m \geqslant 0} \alpha (1-\alpha)^m \frac{1}{(1-\alpha)^m} \int_{m\theta}^\infty \Psi'_+\big(c(u+\lambda)^+\big) e^{-u} \, du \\
&= \sum_{m \geqslant 0} \big(1 - e^{-\theta}\big) \int_{m\theta}^\infty \Psi'_+\big(c(u+\lambda)^+\big) e^{-u} \, du \\
&= \frac{1 - e^{-\theta}}{\theta} \int_0^\infty \theta \left(\left[\frac{u}{\theta} \right] + 1 \right) \Psi'_+\big(c(u+\lambda)^+\big) e^{-u} \, du
\end{aligned} \tag{2.34}$$

where [.] denotes the integer part. Therefore taking $\theta = \theta_s = 2^{-s} \to 0^+$ (s integer), and using the monotone convergence theorem we get

$$\limsup_{\theta\to 0^+}\int_X M_{\mathcal{T}}\phi_\theta\,d\mu \geqslant \int_0^\infty u\Psi'_+\big(c(u+\lambda)^+\big)e^{-u}\,du \tag{2.35}$$

which proves the lower bound in (2.27).

Next given a nonnegative $\phi \in L^1(X,\mu)$ satisfying $\int_X \phi\,d\mu = f$ and $\int_X G\circ\phi\,d\mu \leqslant F$ we consider the sequence (ϕ_m) where $\phi_m = \sum_{I\in\mathcal{T}_{(m)}} \mathrm{Av}_I(\phi)\chi_I$ and set

$$\Phi_m = \sum_{I\in\mathcal{T}_{(m)}} \max\big\{\mathrm{Av}_J(\phi)\colon\ I\subseteq J\in\mathcal{T}\big\}\chi_I = M_{\mathcal{T}}\phi_m \tag{2.36}$$

since $\mathrm{Av}_J(\phi) = \mathrm{Av}_J(\phi_m)$ whenever $I\subseteq J\in\mathcal{T}$ when $I\in\mathcal{T}_{(m)}$ and note that

$$\int_X \phi_m\,d\mu = \int_X \phi\,d\mu = f, \qquad F_m = \int_X G(\phi_m)\,d\mu \leqslant \int_X G(\phi)\,d\mu \leqslant F \tag{2.37}$$

for all m and that Φ_m converges monotonically almost everywhere to $M_{\mathcal{T}}\phi$. Also since each ϕ_m is of the form (2.10) we can apply (2.21), using the values of c and λ satisfying (2.26), and then combining this with the monotone convergence theorem we get

$$\int_X M_{\mathcal{T}}\phi\,d\mu = \lim_{m\to\infty}\int_X \Phi_m\,d\mu \leqslant \frac{F}{c} + \frac{1}{c}\int_0^\infty \Psi\big(c(u+\lambda)^+\big)e^{-u}\,du - \lambda f. \tag{2.38}$$

But now using (2.26) and (2.19) we have

$$\begin{aligned}
&\frac{F}{c} + \frac{1}{c}\int_0^\infty \Psi\big(c(u+\lambda)^+\big)e^{-u}\,du - \lambda f\\
&\quad = \int_0^\infty \big[(u+\lambda)^+ - \lambda\big]\Psi'_+\big(c(u+\lambda)\big)e^{-u}\,du\\
&\quad = \int_0^\infty u\Psi'_+\big(c(u+\lambda)\big)e^{-u}\,du
\end{aligned} \tag{2.39}$$

the last equality holding since $\Psi'_+(c(u+\lambda)) = 0$ whenever $u+\lambda < 0$. This combined with (2.38) and (2.35) completes the proof of the proposition. □

To illustrate the applicability of the above proposition we will give two examples before turning to the case in Theorem 1.

First let us consider a $\sigma > 0$ and define

$$\Psi(x) = \sigma x \quad \text{which implies that} \quad G(x) = \begin{cases} 0 & \text{if } 0 \leqslant x \leqslant \sigma, \\ +\infty & \text{if } x > \sigma. \end{cases} \tag{2.40}$$

Then we can easily compute the corresponding functions from (2.26) when $\lambda < 0$ to be σe^{λ} and 0 respectively. Thus with f such that $0 < f < \sigma$ and $F = 0$ the system has always a solution with $\lambda < 0$ given by $\lambda = \log(\frac{f}{\sigma})$. Hence we need not examine it for $\lambda \geqslant 0$ and we infer from (2.27) that $\mathcal{D}_G^{\mathcal{T}}(0, f) = \sigma(-\lambda + 1)e^{\lambda} = f + f\log(\frac{\sigma}{f})$. Examining what $\mathcal{D}_G^{\mathcal{T}}(0, f)$ means we have an alternative proof of the following formula proven also in [5]

$$\sup\bigl\{\|M_{\mathcal{T}}\phi\|_{L^1(X)}\colon\ \|\phi\|_{L^1(X)} = f,\ \|\phi\|_{L^\infty(X)} = \sigma\bigr\} = f + f\log\left(\frac{\sigma}{f}\right). \tag{2.41}$$

Next we take

$$\Psi(x) = \begin{cases} x & \text{if } 0 \leqslant x \leqslant 1, \\ e^{x-1} & \text{if } x > 1, \end{cases} \quad \text{which implies that} \quad G(x) = x\log^+ x, \tag{2.42}$$

one computes that again the corresponding functions from (2.26) when $\lambda < 0$ are $e^{\lambda}[1 + \frac{c\exp(-1/c)}{1-c}]$ and $ce^{\lambda}\frac{\exp(-1/c)}{(1-c)^2}$ valid only for $0 < c < 1$. Then with $f = 1$ and $F > 0$ the corresponding system of equations is equivalent to $\lambda = -\log[1 + \frac{c\exp(-1/c)}{1-c}]$ (which is always negative) and to

$$q(z) = \left(1 - \frac{1}{z}\right)\bigl((z-1)e^z + 1\bigr) = \frac{1}{F} \tag{2.43}$$

where $z = \frac{1}{c} > 1$. Observing that q is strictly decreasing, $q(1) = 0$ and $\lim_{z\to+\infty} q(z) = +\infty$ we conclude that the system has always a solution c, λ with $\lambda < 0$, $0 < c < 1$ and then computing the integral in (2.27) to be equal to $\frac{F}{c} + 1 - \lambda$ we have found the value of $\mathcal{U}^{\mathcal{T}}(F, f)$ when $f = 1$ where

$$\mathcal{U}^{\mathcal{T}}(F, f) = \sup\left\{\int_X M_{\mathcal{T}}\phi\, d\mu\colon\ \phi \geqslant 0 \text{ is measurable,} \right.$$
$$\left. \int_X \phi\, d\mu = f \text{ and } \int_X \phi\log^+\frac{\phi}{f}\, d\mu \leqslant F\right\} \tag{2.44}$$

is the corresponding to (1.12) supremum Bellman function. But since it is easy to see that $\mathcal{U}^{\mathcal{T}}(F, f) = f\,\mathcal{U}^{\mathcal{T}}(\frac{F}{f}, 1)$, denoting by W the inverse function of q, straightforward manipulations with Eq. (2.43) give the following.

Corollary 1. *For any tree $\mathcal{T}$ and any $F, f > 0$ we have*

$$\mathcal{U}^{\mathcal{T}}(F, f) = f + FW\left(\frac{f}{F}\right) + f\log\left[1 + \frac{\exp(-1/W(f/F))}{W(f/F) - 1}\right]. \tag{2.45}$$

The above provide examples where the corresponding Bellman type function for any F and for a fixed f is given by a single formula coming from solutions (c,λ) with $\lambda<0$ always. This is the exact opposite of what happens in the Bellman function for the (p,p) inequality in [3] (where the single formula comes from solutions (c,λ) with $\lambda\geqslant 0$ always, see [5]).

3. The upper estimate

To prove Theorem 1 we take in Proposition 1

$$\bar{\Psi}(x)=e^x-x-1 \quad \text{which implies that} \quad \bar{G}(x)=(x+1)\log(x+1)-x \tag{3.1}$$

(this makes certain computations easier) and note that $\bar{\Psi}'_+(x)=(e^x-1)^+$. Also if ϕ satisfies the conditions in (1.4) then $\int_X \bar{G}\circ\phi\,d\mu\leqslant F-f$. Moreover one easily computes that when $0<c<1$ and $\lambda\in\mathbb{R}$ the corresponding functions $a_1(c,\lambda)=\int_0^\infty \bar{\Psi}'_+(c(u+\lambda)^+)e^{-u}\,du$, $a_2(c,\lambda)=\int_0^\infty \bar{G}\circ\bar{\Psi}'_+(c(u+\lambda)^+)e^{-u}\,du$ and $b(c,\lambda)=\int_0^\infty u\bar{\Psi}'_+(c(u+\lambda)^+)e^{-u}\,du$ are given by

$$a_1(c,\lambda)=\begin{cases}\frac{e^{c\lambda}}{1-c}-1 & \text{if } \lambda>0,\\ \frac{ce^{\lambda}}{1-c} & \text{if } \lambda\leqslant 0,\end{cases} \tag{3.2}$$

$$a_2(c,\lambda)=\begin{cases}\frac{ce^{c\lambda}}{1-c}(\frac{1}{1-c}+\lambda)-a_1(c,\lambda) & \text{if } \lambda>0,\\ \frac{ce^{\lambda}}{(1-c)^2}-a_1(c,\lambda) & \text{if } \lambda\leqslant 0,\end{cases} \tag{3.3}$$

and

$$b(c,\lambda)=\begin{cases}\frac{e^{c\lambda}}{(1-c)^2}-1 & \text{if } \lambda>0,\\ \frac{ce^{\lambda}}{1-c}(\frac{1}{1-c}+1-\lambda) & \text{if } \lambda\leqslant 0.\end{cases} \tag{3.4}$$

These functions are infinity when $c\geqslant 1$. The corresponding system of Eqs. (2.26) with $F-f$ replacing F, which thus is equivalent to $a_1(c,\lambda)=f$ and $a_2(c,\lambda)+a_1(c,\lambda)=F$, can be solved as follows.

If this system has a solution with $\lambda\leqslant 0$ then $\frac{ce^{\lambda}}{1-c}=f$, $\frac{ce^{\lambda}}{(1-c)^2}=F$ thus $c=\frac{F-f}{F}$ which is in $(0,1)$ and so $e^{\lambda}=\frac{f^2}{F-f}$. However to have $\lambda\leqslant 0$ we must assume that $F\geqslant f^2+f$. On the other hand when $F\geqslant f^2+f$ the above values furnish a solution to the system for which we have

$$b(c,\lambda)=\frac{ce^{\lambda}}{1-c}\left(\frac{1}{1-c}+1-\lambda\right)=F+f+f\log\frac{F-f}{f^2}. \tag{3.5}$$

Next if the system has a solution with $\lambda>0$ then $\frac{e^{c\lambda}}{1-c}=f+1$, $\frac{ce^{c\lambda}}{1-c}(\frac{1}{1-c}+\lambda)=F$ thus $\frac{c}{1-c}+c\lambda=\frac{F}{f+1}$. Thus setting $z=\frac{1}{1-c}>1$ we get $z\exp(\frac{F}{f+1}-z+1)=f+1$ thus $V(z)=\frac{\exp(F/(f+1))}{f+1}$. This has always a unique solution $z>1$ since $F>(f+1)\log(f+1)$ thus the right-hand side is greater than 1. Thus $z=U(\frac{\exp(F/(f+1))}{f+1})$ and so $c\lambda=\frac{F}{f+1}+1-U(\frac{\exp(F/(f+1))}{f+1})$.

But we must have $\lambda > 0$ for this to work which is equivalent to $V(\frac{F}{f+1}+1) > \frac{\exp(F/(f+1))}{f+1}$ which in turn is equivalent to $F < f^2 + f$. If this happens the solution is unique and we have

$$b(c,\lambda) = \frac{e^{c\lambda}}{(1-c)^2} - 1 = (f+1)z - 1 = (f+1)U\left(\frac{e^{F/(f+1)}}{f+1}\right) - 1. \tag{3.6}$$

The above complete the proof of Theorem 1. Note also that the corresponding system has always a unique solution. This holds in much more general situations as derived in [5]. However we will not need this here.

Next to prove Theorem 2 we consider the convex function $G(x) = (x+1)\log(x+1)$ and argue in a similar as in Section 7 of [3] manner. The basic ingredient here is the fact that, as explained in the Introduction, the function $\mathcal{D}^{\mathcal{T}}_{\log}(x, y)$ given in Theorem 1 is concave (and independent of the tree $\mathcal{T}$). To proceed further we let ϕ, E be as in the definition of $\mathcal{D}^{\mathcal{T}}_{\log}(F, f, k)$ where $0 < k < 1$ and choose $u > 0$ such that

$$\mu\big(\{M_{\mathcal{T}}\phi > u\}\big) \leqslant k \leqslant \mu\big(\{M_{\mathcal{T}}\phi \geqslant u\}\big) \tag{3.7}$$

and then choose a measurable D such that $V_1 = \{M_{\mathcal{T}}\phi > u\} \subseteq D \subseteq \{M_{\mathcal{T}}\phi \geqslant u\} = V_2$ and $\mu(D) = k$. Since $M_{\mathcal{T}}\phi \leqslant u$ on $E \setminus V_1$ it is easy to see that $\int_E M_{\mathcal{T}}\phi\, d\mu \leqslant \int_D M_{\mathcal{T}}\phi\, d\mu$ and defining $s \in [0, 1]$ by $\mu(D) = s\mu(V_1) + (1-s)\mu(V_2)$ we also have (since $M_{\mathcal{T}}\phi = u$ on $V_2 \setminus V_1$)

$$\int\limits_D M_{\mathcal{T}}\phi\, d\mu = s\int\limits_{V_1} M_{\mathcal{T}}\phi\, d\mu + (1-s)\int\limits_{V_2} M_{\mathcal{T}}\phi\, d\mu. \tag{3.8}$$

Now each of the V_1, V_2 is a union of families $\{I_j^{(1)}\}$, $\{I_r^{(2)}\}$ consisting of pairwise disjoint elements maximal under $\mathrm{Av}_I(\phi) > u$ (resp. $\geqslant u$) and we clearly have $M_{\mathcal{T}}\phi = M_{\mathcal{T}(I)}\phi$ (where $\mathcal{T}(I)$ is the subtree of $\mathcal{T}$ with top I on the probability space $(I, \frac{1}{\mu(I)}\mu)$) for each of those I's. Hence, using Theorem 1 for all these trees, arguing as in [3] and using the concavity of the function in Theorem 1 we get

$$\int\limits_E M_{\mathcal{T}}\phi\, d\mu \leqslant k\mathcal{D}^{\mathcal{T}}_{\log}\left(\frac{A}{k}, \frac{B}{k}\right) \tag{3.9}$$

where

$$A = s\int\limits_{V_1} G\circ\phi\, d\mu + (1-s)\int\limits_{V_2} G\circ\phi\, d\mu \leqslant F \tag{3.10}$$

and

$$B = s\int\limits_{V_1} \phi\, d\mu + (1-s)\int\limits_{V_2} \phi\, d\mu \leqslant f. \tag{3.11}$$

Letting $\eta = s\chi_{V_1} + (1-s)\chi_{V_2}$ Jensen's inequality implies that

$$G\left(\frac{B}{k}\right) = G\left(\frac{\int_X \phi\eta\,d\mu}{\int_X \eta\,d\mu}\right) \leqslant \frac{\int_X (G\circ\phi)\eta\,d\mu}{\int_X \eta\,d\mu} = \frac{A}{k} \tag{3.12}$$

and

$$G\left(\frac{f-B}{1-k}\right) = G\left(\frac{\int_X \phi(1-\eta)\,d\mu}{\int_X (1-\eta)\,d\mu}\right) \leqslant \frac{\int_X (G\circ\phi)(1-\eta)\,d\mu}{\int_X (1-\eta)\,d\mu} = \frac{F-A}{1-k}. \tag{3.13}$$

Thus

$$(1-k)G\left(\frac{f-B}{1-k}\right) + kG\left(\frac{B}{k}\right) \leqslant F \tag{3.14}$$

and since $\mathcal{D}_{\log}^{\mathcal{T}}(x,y)$ is strictly increasing in x, (3.9) and (3.13) imply that

$$\int_E M_{\mathcal{T}}\phi\,d\mu \leqslant k\mathcal{D}_{\log}^{\mathcal{T}}\left(\frac{1}{k}\left(F-(1-k)G\left(\frac{f-B}{1-k}\right)\right), \frac{B}{k}\right). \tag{3.15}$$

Conversely supposing B is in $(0,f)$ and satisfies (3.14) we fix $\delta < 1$, choose, using Lemma 1, a family $\{I_1, I_2, \ldots\}$ of pairwise disjoint elements of $\mathcal{T}$ such that $\sum_j \mu(I_j) = k$, we write $E = \bigcup_j I_j$, $A = F - (1-k)G(\frac{f-B}{1-k}) \geqslant kG(\frac{B}{k})$ and using Theorem 1 for each j we choose a nonnegative measurable ϕ_j on I_j such that $\mathrm{Av}_{I_j}(G\circ\phi_j) = \frac{A}{k}$, $\mathrm{Av}_{I_j}(\phi_j) = \frac{B}{k}$ and

$$\int_{I_j} M_{\mathcal{T}(I_j)}(\phi_j)\,d\mu \geqslant \delta\mu(I_j)\mathcal{D}_{\log}^{\mathcal{T}}\left(\frac{A}{k}, \frac{B}{k}\right). \tag{3.16}$$

Next we choose a nonnegative measurable ψ on $X \setminus E$ such that $\int_{X\setminus E} G\circ\psi\,d\mu = F - A > 0$ and $\int_{X\setminus E}\psi\,d\mu = f - B > 0$ which is possible by (3.14) and defining $\phi = \psi\chi_{X\setminus E} + \sum_j \phi_j\chi_{I_j}$ we have $\int_X G\circ\phi\,d\mu = F$, $\int_X \phi\,d\mu = f$ and

$$\int_E M_{\mathcal{T}}\phi\,d\mu \geqslant \delta k\mathcal{D}_{\log}^{\mathcal{T}}\left(\frac{A}{k}, \frac{B}{k}\right). \tag{3.17}$$

Letting now $\delta \to 1^-$ we have proved the following:

Proposition 2. *$\mathcal{B}_{\log}^{\mathcal{T}}(F,f,k)$ is equal to the supremum of the function*

$$R_k(B) = k\mathcal{D}_{\log}^{\mathcal{T}}\left(\frac{1}{k}\left(F-(1-k)G\left(\frac{f-B}{1-k}\right)\right), \frac{B}{k}\right) \tag{3.18}$$

on the set of B in $[0,f]$ *that satisfy the estimate* (3.14).

To proceed further we fix F, f, k and define the following functions on $[0, f]$

$$h(B) = (1-k)G\left(\frac{f-B}{1-k}\right) + kG\left(\frac{B}{k}\right), \qquad A(B) = F - (1-k)G\left(\frac{f-B}{1-k}\right) \tag{3.19}$$

and $y(B) = \frac{f-B}{1-k}$. Since $h'(B) = -G'(\frac{f-B}{1-k}) + G'(\frac{B}{k})$ the convexity of G implies that $h'(B)$ has the same sign as $B - kf$ and since $h(kf) = G(f) < F$ we conclude that the set of all B in $[0, f]$ satisfying (3.14) is a closed interval of the form $[B_1, B_2]$ where $0 \leqslant B_1 < kf < B_2 \leqslant f$. Moreover $B_2 = f$ if $h(f) \leqslant F$ otherwise $B_2 < f$ and $h(B_2) = F$ and similarly $B_1 = 0$ if $h(0) \leqslant F$ otherwise $B_1 > 0$ and $h(B_1) = F$.

Since $\mathcal{D}_{\log}^{\mathcal{T}}(x, y)$ is given by a double formula one must also compare $A(B)$ with $B(\frac{B}{k} + 1)$. Hence we also consider the function

$$\sigma(B) = (1-k)G\left(\frac{f-B}{1-k}\right) + B\left(\frac{B}{k} + 1\right). \tag{3.20}$$

Now using Theorem 1 it is easy to see that on $y^2 + y < x$ we have

$$\frac{\partial \mathcal{D}_{\log}^{\mathcal{T}}}{\partial x}(x, y) = \frac{x}{x-y} > 0 \quad \text{and} \quad \frac{\partial \mathcal{D}_{\log}^{\mathcal{T}}}{\partial y}(x, y) = -\frac{x}{x-y} + \log\frac{x-y}{y^2} > -\frac{x}{x-y} \tag{3.21}$$

so since $A'(B) = G'(\frac{f-B}{1-k}) = 1 + \log(y(B) + 1) > 1$ we get $R_k'(B) > 0$ for every $B \in [B_1, B_2]$ such that $\sigma(B) < F$.

Next on the set where $(y+1)\log(y+1) < x < y^2 + y$ we compute using Theorem 1 and (1.6)

$$\frac{\partial \mathcal{D}_{\log}^{\mathcal{T}}}{\partial x}(x, y) = \frac{U(z)}{U(z)-1} > 0 \quad \text{and} \quad \frac{\partial \mathcal{D}_{\log}^{\mathcal{T}}}{\partial y}(x, y) = \frac{U(z)}{U(z)-1}\left[U(z) - 2 - \frac{x}{y+1}\right] \tag{3.22}$$

where $z = \exp(x/(y+1))$. Comparing (3.21) and (3.22) at $x = y^2 + y$ we conclude that $\mathcal{D}_{\log}^{\mathcal{T}}$ and hence R_k is actually C^1. Also it easily follows from (3.22) that if $B \in [B_1, B_2]$ is such that $\sigma(B) > F$ then $R_k'(B)$ has the same sign as the expression

$$U\left(\frac{k\exp(A(B)/(B+k))}{B+k}\right) - \frac{A(B)}{B+k} - 1 + \log\bigl(y(B) + 1\bigr) \tag{3.23}$$

which since V is strictly decreasing has the same sign as

$$\frac{k\exp(A(B)/(B+k))}{B+k} - \frac{\exp(A(B)/(B+k) - \log(y(B)+1))}{\frac{A(B)}{B+k} + 1 - \log(y(B)+1)} \tag{3.24}$$

if $A(B) > (B+k)\log(y(B)+1)$ and is positive otherwise. But $A(B) > (B+k)\log(y(B)+1)$ holds if and only if

$$F > (f+1)\log\bigl(y(B) + 1\bigr) \tag{3.25}$$

and if this also holds we conclude now that $R'_k(B)$ has the same sign as

$$F - b(B) = F - (f+1)\log\bigl(y(B)+1\bigr) - \frac{(B+k)(B-kf)}{k(f+1-B-k)} \tag{3.26}$$

where $b(B)$ is defined by the above equality. But now comparing (3.25) and (3.26) and also since $R'_k(B)$ is positive if $\sigma(B) < F$ we conclude that $R'_k(B) > 0$ on the whole interval (B_1, kf).

Next if $B \in (kf, f)$ then it is easy to see that $b'(B) = \frac{(B-kf)(2(f+1)-B-k)}{k(f+1-B-k)^2} > 0$ and so since $b(kf) = (f+1)\log(f+1) < F$ we conclude that if $b(B_2) \leqslant F$ then $R'_k(B) \geqslant 0$ for every $B \in [B_1, B_2]$ hence $\mathcal{B}^{\mathcal{T}}_{\log}(F, f, k) = R_k(B_2)$.

Assume now that $b(B_2) > F$. Then there exists a unique $B_0 \in (kf, B_2)$ such that $b(B_0) = F$. By (3.26) this B_0 clearly satisfies (3.25). We will show that B_0 satisfies also the following $F \leqslant \sigma(B_0)$ and therefore $R'_k(B_0) = 0$. Indeed it suffices to prove that $b(B) \leqslant \sigma(B)$ for every $B \in (kf, f)$. But writing $B = kf + (1-k)x$ where $0 < x < f$ straightforward calculations show that $b(B) \leqslant \sigma(B)$ is equivalent to

$$g(x) = \frac{f+1}{f+1-x} + k\log(f+1-x) \leqslant (1-k)x + kf + 1 \tag{3.27}$$

which holds since g is convex and trivially (3.27) holds at the endpoints $x = 0$ and $x = f$. Using the same substitution we also have $b(B) > h(B)$ on (kf, f) since this can be easily computed to be equivalent to the inequality $\log(f+1-x) + \frac{x}{k(f+1-x)} > \log(f+1-x+\frac{x}{k})$ which clearly holds.

Hence if $B_2 < f$ then $\sigma(B_2) \geqslant b(B_2) > h(B_2) = F$ and so B_0 exists and since R_k is C^1 we get that B_0 is its absolute maximum on $[B_1, B_2]$ thus $\mathcal{B}^{\mathcal{T}}_{\log}(F, f, k) = R_k(B_0)$. Considering also the case $B_2 = f$ (that is when $(f+k)\log(\frac{f}{k}+1) \leqslant F$) and since $b(f) = f(\frac{f}{k}+1)$ we get using Proposition 2 the following (noting that if $B_2 < f$ then $f(\frac{f}{k}+1) > (f+k)\log(\frac{f}{k}+1) > F$)

$$\mathcal{B}^{\mathcal{T}}_{\log}(F, f, k) = \begin{cases} R_k(B_0) & \text{if } F < f(\frac{f}{k}+1), \\ R_k(f) & \text{if } f(\frac{f}{k}+1) \leqslant F. \end{cases} \tag{3.28}$$

Obviously $R_k(f) = F + f + f\log\frac{k(F-f)}{f^2}$ and on the other hand if B_0 exists then as we have seen $\mathcal{D}^{\mathcal{T}}_{\log}(\frac{A(B_0)}{k}, \frac{B_0}{k})$ is given by the first part of the formula in (1.7). But now writing

$$\frac{B_0 - kf}{f+1-B_0-k} = \xi_0 \tag{3.29}$$

we observe that ξ_0 satisfies the inequalities $0 < \xi_0 < f$ (since $B_0 \in (kf, f)$) and since $B_0 + k = \frac{(f+1)(\xi_0+k)}{\xi_0+1}$, $y(B_0) + 1 = \frac{f+1}{\xi_0+1}$ the equation $b(B_0) = F$ becomes $\tau_k(\xi_0) = \frac{F}{f+1} - \log(f+1)$ thus $\xi_0 = \xi_k(F, f)$. Then substituting F with $b(B_0)$ in $R_k(B_0) = k\mathcal{D}^{\mathcal{T}}_{\log}(\frac{1}{k}(F - (1-k)G(\frac{f-B_0}{1-k})), \frac{B_0}{k})$ and using ξ_0 it is straightforward to get that $R_k(B_0)$ is equal to the second expression in (1.10). This completes the proof of Theorem 2.

4. The lower estimate

Here we will prove Theorem 3. Assuming that $\mathcal{T}$ is N-homogeneous we let ϕ be a nonnegative function of the form (2.10) such that

$$\int_X \phi\,d\mu = f \quad \text{and} \quad \int_X \phi \log^+\left(\frac{\phi}{f}\right) d\mu = F \tag{4.1}$$

and let $\mathcal{S} = \mathcal{S}_\phi$ be the corresponding subtree of $\mathcal{T}$. Using the notation from Section 2 we make the following two simple observations. First, by Lemma 3(iv), we have $y_{I^*} < y_I \leqslant N y_{I^*}$ for all $I \in \mathcal{S} \setminus \{X\}$ and second, by condition (ii) in Definition 1(b), $\phi(x) \leqslant y_I$ whenever $I \in \mathcal{S}$ and $x \in A_I$.

We consider the function $\tilde{G}(x) = x \log^+(\frac{x}{f})$ which is convex on $x \geqslant 0$. The second remark combined with the convexity of G gives

$$\frac{1}{a_I} \int_{A_I} \tilde{G}\big(\phi(x)\big)\,d\mu(x) \leqslant \frac{1}{a_I} \int_{A_I} \frac{\phi(x)}{y_I} \tilde{G}(y_I)\,d\mu(x) = \frac{x_I}{y_I} \tilde{G}(y_I) \tag{4.2}$$

for all $I \in \mathcal{S}$. Now Lemma 3 implies that

$$\begin{aligned} \int_X M_{\mathcal{T}}\phi\,d\mu &= \sum_{I \in \mathcal{S}} a_I y_I = \sum_{I \in \mathcal{S}} \Big(\mu(I) - \sum_{J \in \mathcal{S}:\ J^* = I} \mu(J)\Big) y_I \\ &= f + \sum_{I \in \mathcal{S},\, I \neq X} \mu(I)(y_I - y_{I^*}) \end{aligned} \tag{4.3}$$

and by using (4.2) we get

$$\begin{aligned} F &= \int_X \phi \log^+\left(\frac{\phi}{f}\right) d\mu \leqslant \sum_{I \in \mathcal{S}} a_I x_I \frac{\tilde{G}(y_I)}{y_I} \\ &= \sum_{I \in \mathcal{S}} \Big(\mu(I) y_I - \sum_{J \in \mathcal{S}:\ J^* = I} \mu(J) y_J\Big) \log^+\left(\frac{y_I}{f}\right) \\ &= \sum_{I \in \mathcal{S}, I \neq X} \mu(I) y_I \left(\log^+\left(\frac{y_I}{f}\right) - \log^+\left(\frac{y_{I^*}}{f}\right)\right) = \sum_{I \in \mathcal{S},\, I \neq X} \mu(I) y_I \log\left(\frac{y_I}{y_{I^*}}\right) \end{aligned} \tag{4.4}$$

since by Lemma 3 $a_I x_I = \mu(I) y_I - \sum_{J \in \mathcal{S}:\ J^* = I} \mu(J) y_J$ and $y_I \geqslant y_X = f$ for all I.

Next for any $I \in \mathcal{S}$, $I \neq X$ we have $1 < \frac{y_I}{y_{I^*}} \leqslant N$ and so using the easy to verify fact that $\frac{1}{t} \log \frac{1}{1-t}$ is increasing for $t \in (0,1)$ we obtain by taking $t = 1 - \frac{y_{I^*}}{y_I} \in (0, 1 - \frac{1}{N})$ the following

$$y_I \log\left(\frac{y_I}{y_{I^*}}\right) \leqslant \frac{N \log N}{N-1} (y_I - y_{I^*}). \tag{4.5}$$

Using (4.5) in (4.4) and in view of (4.3) we get

$$\int_X M_{\mathcal{T}}\phi\, d\mu \geqslant \frac{N-1}{N\log N}F + f \tag{4.6}$$

for functions ϕ of the form (2.10).

Now for the general case, given $\phi \geqslant 0$ measurable satisfying (4.1) we define ϕ_m, Φ_m as in the proof of Proposition 1 and for each ϕ_m we can apply (4.6) to get

$$\int_X M_{\mathcal{T}}\phi\, d\mu \geqslant \int_X \Phi_m\, d\mu \geqslant \frac{N-1}{N\log N}\int_X \phi_m \log^+\left(\frac{\phi_m}{f}\right) d\mu + f. \tag{4.7}$$

We will now show that the sequence $\psi_m = \phi_m \log^+(\frac{\phi_m}{f})$ is uniformly integrable. Since $\phi_m \to \phi$ almost everywhere, by the second condition in the definition of the homogeneous trees, we get the estimate (4.6) for the general measurable ϕ.

To show the uniform integrability of ψ_m we note that for any $\lambda > e$ and any m the set where $\psi_m > \lambda f$ is contained in the set where $\phi_m > \frac{f\lambda}{\log\lambda}$ and therefore in $E_\lambda = \{M_{\mathcal{T}}\phi > \frac{f\lambda}{\log\lambda}\}$. On the other hand given any $I \in \mathcal{T}_{(m)}$ Jensen's inequality implies that

$$\psi_m = \phi_m \log^+\left(\frac{\phi_m}{f}\right) \leqslant \frac{1}{\mu(I)}\int_I \phi \log^+\left(\frac{\phi}{f}\right) d\mu \tag{4.8}$$

on I thus integrating and summing we get for any m the following

$$\int_{\{\psi_m > \lambda f\}} \psi_m\, d\mu \leqslant \sum_{I\in\mathcal{T}_{(m)},\, I\subseteq E_\lambda} \int_I \phi \log^+\left(\frac{\phi}{f}\right) d\mu \leqslant \int_{E_\lambda} \phi \log^+\left(\frac{\phi}{f}\right) d\mu \tag{4.9}$$

which tends to 0 as $\lambda \to +\infty$ since $\phi \log^+(\frac{\phi}{f})$ is integrable and $\mu(E_\lambda) \leqslant \frac{\log\lambda}{\lambda}$.

These prove that $\mathcal{L}^{\mathcal{T}}(F, f) \geqslant \frac{N-1}{N\log N}F + f$.

To prove the reverse inequality we let $X = I_0 \supseteq I_1 \supseteq \cdots I_s \supseteq I_{s+1} \supseteq \cdots$ be a chain such that $I_s \in \mathcal{T}_{(s)}$ for all $s \geqslant 0$ (and so $\mu(I_s) = N^{-s}$). We write

$$\frac{F}{f\log N} = m_0 + \sum_{k=1}^{\infty} \frac{\ell_k}{N^k} \tag{4.10}$$

for the expansion of $\frac{F}{f\log N}$ in base N, thus $m_0, \ell_1, \ldots, \ell_k, \ldots$ are nonnegative integers such that $\ell_k < N$ for all $k \geqslant 1$, and we define the strictly increasing sequence of integers $m_0 < m_1 < \cdots < m_k < \cdots$ by the rule $m_k - m_{k-1} = \ell_k + 1 > 0$ for all $k \geqslant 1$. Then we define

$$\phi = f\sum_{k=0}^{\infty} N^{m_k - k}\chi_{I_{m_k}\setminus I_{m_k+1}}. \tag{4.11}$$

We have

$$\int_X \phi\, d\mu = f\sum_{k=0}^{\infty} N^{m_k-k}\left(N^{-m_k}-N^{-m_k-1}\right) = f\sum_{k=0}^{\infty} N^{-k}\left(1-\frac{1}{N}\right) = f \tag{4.12}$$

$$\begin{aligned}\int_X \phi \log^+\frac{\phi}{f}\, d\mu &= f\sum_{k=0}^{\infty} N^{m_k-k}\left(N^{-m_k}-N^{-m_k-1}\right)(m_k-k)\log N \\ &= f\log N\left(\sum_{k=0}^{\infty}\frac{m_k-k}{N^k} - \sum_{k=1}^{\infty}\frac{m_{k-1}-k+1}{N^k}\right) \\ &= f\log N\left(m_0+\sum_{k=1}^{\infty}\frac{\ell_k}{N^k}\right) = F \end{aligned} \tag{4.13}$$

and if $m_{k-1} < s \leqslant m_k$ then

$$\mathrm{Av}_{I_s}(\phi) = N^s f\sum_{j=k}^{\infty} N^{m_j-j}\left(N^{-m_j}-N^{-m_j-1}\right) = fN^{s-k} \tag{4.14}$$

and this increases as s increases (if $s=m_k$ then $\mathrm{Av}_{I_s}(\phi) = \mathrm{Av}_{I_{s+1}}(\phi)$). We next claim that $M_{\mathcal{T}}\phi(x) = \mathrm{Av}_{I_s}(\phi)$ whenever $x \in I_s \setminus I_{s+1}$ and $s \geqslant 0$. Indeed suppose that $x \in I_s \setminus I_{s+1}$ and let J be the unique element of $\mathcal{T}_{(s+1)}$ such that $x \in J$ (clearly $J \in \mathcal{C}(I_s)$ and $J \neq I_s$). Then the set of all I's in $\mathcal{T}$ containing x consists of $I_0, \dots, I_s$ and J and certain subintervals of it. But $\mathrm{Av}_{I_s}(\phi) \geqslant \mathrm{Av}_{I_r}(\phi)$ for all $0 \leqslant r < s$ and since ϕ is either 0 on J (if s is not an m_k) or if $s = m_k$ it is equal to $\mathrm{Av}_{I_s}(\phi)$ on J we get that $M_{\mathcal{T}}\phi(x) = \mathrm{Av}_{I_s}(\phi)$. Hence

$$M_{\mathcal{T}}\phi = f\sum_{s=0}^{\infty} N^{s-k(s)}\chi_{I_s\setminus I_{s+1}} \tag{4.15}$$

where $k(s)$ is the smallest integer k with $m_k \geqslant s$ and this implies that

$$\begin{aligned}\int_X M_{\mathcal{T}}\phi\, d\mu &= f\left(1-\frac{1}{N}\right)\sum_{s=0}^{\infty} N^{-k(s)} \\ &= f\left(1-\frac{1}{N}\right)\left(m_0+1+\sum_{k=1}^{\infty}\frac{m_k-m_{k-1}}{N^k}\right) = \frac{f(N-1)}{N}\left(m_0+1+\sum_{k=1}^{\infty}\frac{\ell_k+1}{N^k}\right) \\ &= \frac{N-1}{N\log N}F + f \end{aligned} \tag{4.16}$$

by (4.10). This completes the proof of Theorem 3.

References

[1] D.L. Burkholder, Boundary value problems and sharp inequalities for martingale transforms, Ann. Probab. 12 (1984) 647–702.

[2] D.L. Burkholder, Martingales and Fourier analysis in Banach spaces, in: C.I.M.E. Lectures, Varenna (Como), Italy, 1985, in: Lecture Notes in Math., vol. 1206, 1986, pp. 61–108.

[3] A.D. Melas, The Bellman functions of dyadic-like maximal operators and related inequalities, Adv. in Math. 192 (2005) 310–340.

[4] A.D. Melas, A sharp L^p inequality for dyadic A_1 weights in $\mathbb{R}^n$, Bull. Lond. Math. Soc. 37 (2005) 919–926.

[5] A.D. Melas, Sharp general local estimates for Dyadic-like maximal operators and related Bellman functions, Adv. in Math. 220 (2009) 367–426.

[6] A.D. Melas, E. Nikolidakis, Dyadic-like maximal operators on integrable functions and Bellman functions related to Kolmogorov's inequality, Trans. Amer. Math. Soc., in press.

[7] F. Nazarov, S. Treil, The hunt for a Bellman function: Applications to estimates for singular integral operators and to other classical problems of harmonic analysis, Algebra i Analiz 8 (5) (1996) 32–162.

[8] F. Nazarov, S. Treil, A. Volberg, The Bellman functions and two-weight inequalities for Haar multipliers, J. Amer. Math. Soc. 12 (4) (1999) 909–928.

[9] F. Nazarov, S. Treil, A. Volberg, Bellman function in stochastic optimal control and harmonic analysis (how our Bellman function got its name), Oper. Theory Adv. Appl. 129 (2001) 393–424, Birkhäuser Verlag.

[10] L. Slavin, A. Stokolos, V. Vasyunin, Monge–Ampère equations and Bellman functions: The dyadic maximal operator, C. R. Acad. Paris Sér. I Math., I 346 (2008).

[11] L. Slavin, V. Vasyunin, Sharp results in the integral-form John–Nirenberg inequality, submitted for publication.

[12] L. Slavin, A. Volberg, The explicit BF for a dyadic Chang–Wilson–Wolff theorem. The s-function and the exponential integral, Contemp. Math. 444 (2007).

[13] E.M. Stein, Note on the class $L \log L$, Studia Math. 32 (1969) 305–310.

[14] V. Vasyunin, The sharp constant in the reverse Holder inequality for Muckenhoupt weights, Algebra i Analiz 15 (1) (2003) 73–117.

[15] V. Vasyunin, A. Volberg, The Bellman functions for a certain two weight inequality: The case study, Algebra i Analiz 18 (2) (2006).

[16] V. Vasyunin, A. Volberg, Monge–Ampère equation and Bellman optimization of Carleson embedding theorem, preprint.

[17] A. Volberg, Bellman approach to some problems in harmonic analysis, in: Seminaire des Equations aux derivées partielles, Ecole Polytéchnique, 2002, eposé. XX, pp. 1–14.

Further reading

[18] L. Grafakos, S. Montgomery-Smith, Best constants for uncentered maximal functions, Bull. London Math. Soc. 29 (1) (1997) 60–64.

[19] G. Wang, Sharp maximal inequalities for conditionally symmetric martingales and Brownian motion, Proc. Amer. Math. Soc. 112 (1991) 579–586.