THE C_p CONJECTURE OF MUCKENHOUPT

ANDREI K. LERNER AND SHELDY OMBROSI

ABSTRACT. We give an affirmative answer to Muckenhoupt's C_p conjecture: for $1 < p < \infty$, the C_p condition is necessary and sufficient for the weighted $L^p(w)$ norm inequality relating the Hilbert transform and the Hardy–Littlewood maximal operator. We also show that the C_p condition is necessary and sufficient for the weighted Fefferman–Stein inequality.

1. INTRODUCTION

The main results proved in this paper hold in every dimension. For clarity of exposition, we present them in the one-dimensional setting; the extension to $\mathbb{R}^n$ is explained in Remark 8.3.

Let H and M denote the Hilbert transform and the uncentered Hardy–Littlewood maximal operator, respectively. For $1 < p < \infty$, a weight w on $\mathbb{R}$ belongs to C_p if there exist $r > 1$ and $K < \infty$ such that

$$\left(\frac{1}{|I|}\int_I w^r\right)^{1/r} \le \frac{K}{|I|}\int_{\mathbb{R}} (M\mathbf{1}_I)^p w \qquad \text{for every bounded interval } I.$$

Muckenhoupt [9] proved that the C_p condition is necessary for the inequality

$$\int_{\mathbb{R}} |Hf|^p w \le C\int_{\mathbb{R}} (Mf)^p w \tag{1.1}$$

and conjectured that it is also sufficient. This conjecture has remained open for 46 years.

Sawyer [10] established sufficiency under the stronger assumption $w \in C_{p+\varepsilon}$ for some $\varepsilon > 0$. This raised the question whether the C_p condition is self-improving in p: an affirmative answer would have settled the conjecture. However, Kahanpää and Mejlbro [5] constructed a weight in C_p which belongs to no C_q with $q > p$, ruling out this approach.

Further developments on C_p weights include reverse Hölder estimates, extensions to rough and vector-valued operators, and characterizations of weak-type inequalities; see [1–4, 6, 7, 11]. Nevertheless, the original sufficiency problem remained unresolved.

Sparse domination, developed during the 2010s, allows us to replace the Hilbert transform by simpler positive operators built from local averages over sparse families of intervals. However, for more than a decade it remained unclear how to prove (1.1) with these sparse operators in place of H, assuming only $w \in C_p$. The difficulty lies in controlling the overlapping nonlocal tails produced by the C_p condition, even when the underlying family is sparse. We establish this bound and thereby prove the Muckenhoupt conjecture.

A family $\mathscr{S}$ of dyadic intervals is η-sparse, $0 < \eta < 1$, if there are pairwise disjoint measurable sets $E_I \subset I$, $I \in \mathscr{S}$, with $|E_I| \ge \eta|I|$. Its sparse operator is

$$A_{\mathscr{S}} f = \sum_{I\in\mathscr{S}} f_I \mathbf{1}_I, \qquad f_I = \frac{1}{|I|}\int_I f.$$

2010 *Mathematics Subject Classification.* 42B20, 42B25.

Key words and phrases. Sparse operator, maximal operator, C_p condition.

Theorem 1.1. *Let $1<p<\infty$, let w satisfy the C_p condition with parameters r,K, and let $\mathscr{S}$ be an η-sparse family in a dyadic grid. Then, for every locally integrable function f,*

$$\int_{\mathbb{R}} |A_{\mathscr{S}} f|^p w \le C_{p,r,K,\eta} \int_{\mathbb{R}} (Mf)^p w.$$

By the standard sparse domination of the Hilbert transform, see [8], Theorem 1.1 immediately gives the following.

Corollary 1.2. *Muckenhoupt's conjecture is true. More precisely, if $1<p<\infty$ and $w\in C_p$, then (1.1) is valid for every bounded compactly supported function f, with a constant depending only on p,r,K.*

1.1. **Idea of the proof.** We outline the main ideas of the proof, leaving the technical details to the later sections. Write $w(I)=\int_I w$ and $w_I=w(I)/|I|$, and assume that $f\ge 0$.

From the sparse operator to a sum. For any weight w and any $r>1$, Lemma 3.1 gives

$$\int_{\mathbb{R}} (A_{\mathscr{S}} f)^p w \lesssim \sum_{I\in\mathscr{S}} f_I^p \left(\frac{1}{|I|}\int_I w^r\right)^{1/r} |I|.$$

The C_p condition is not used at this stage. This lemma is the main analytic use of the sparseness of the original family.

The natural next step would be to bound the sum on the right by $\int_{\mathbb{R}}(Mf)^p w$. Under the C_p condition alone, however, this is false. Appendix A gives a weight in C_p and a bounded compactly supported f for which the maximal-function integral is finite but this sum is infinite for every $r>1$. We must therefore split the operator before using the lemma. The intervals along which the maximal averages of f grow rapidly are handled directly by a pointwise estimate; all the intervals in the counterexample belong to this part. We apply the lemma to the remaining family.

For this remaining sum, the C_p condition bounds the L^r average of w by the nonlocal quantity $|I|^{-1}\int_{\mathbb{R}}(M\mathbf{1}_I)^p w$. Direct summation of these overlapping tails does not give the desired estimate. Our plan is to transform the sum by telescoping. For this we need two structures: monotonicity in the quantities involving f, and an exact relation between the weight terms at a parent and its children.

Preparing the sum for telescoping. The ordinary averages f_I need not be monotone along nested intervals. We therefore introduce the maximal average

$$\omega_I = \sup_{L\supseteq I} f_L$$

and use

$$f_I^p \le \omega_I^{p-1} f_I.$$

As I shrinks, the supremum is taken over a larger collection, so ω_I can only increase; it is also bounded by Mf on I. Retaining one ordinary average is important: the quantity $f_I|I|=\int_I f$ is additive over the children and increases when I is replaced by a containing interval. When ω_I grows rapidly from parent to child, geometric summation bounds the corresponding part of the sparse operator pointwise by Mf. This gives the preliminary removal described above.

For the weight term, we need a form of the tail that can be followed from one generation to the next. For a dyadic grid $\mathscr{D}$, define

$$J_I^{\mathscr{D}} = \sum_{k\ge 0} q^k w_{I^{(k)}}, \qquad q=2^{-(p-1)},$$

where $I^{(k)}$ is the kth dyadic ancestor of I. This discrete analogue of the tail in the C_p condition satisfies $J^{\mathscr{D}}_{I_\pm} = qJ^{\mathscr{D}}_I + w_{I_\pm}$. This parent–child relation, together with the monotonicity of ω_I and the additivity of $\int_I f$, is what makes the telescoping argument possible.

Faithful intervals and three grids. There is a difficulty with this replacement: the C_p condition controls the L^r average by the full tail, which need not be bounded by the dyadic tail in a given grid. We call an interval *faithful* if its full tail is controlled by its dyadic tail. On faithful intervals, the C_p condition therefore gives the bound by $J^{\mathscr{D}}_I$ needed for the telescoping argument.

The original sparse family need not consist of faithful intervals, and we cannot simply discard those that are not faithful. The three-grid theorem resolves this difficulty. After enlarging any interval by a fixed factor, we can place it in one of three fixed grids so that the enlarged interval is faithful and its dyadic tail also controls the full tail of the original interval. This construction dominates the original sparse operator by three sparse operators on faithful intervals, with a fixed loss in the sparseness constant. It also allows us to transfer terms between grids later in the proof.

Telescoping and weight drops. On faithful intervals where $J^{\mathscr{D}}_I$ is comparable to w_I, the C_p condition gives the local comparison required in the second assertion of Lemma 3.1, and that assertion bounds the corresponding sum by $\int_{\mathbb{R}}(Mf)^p w$. After treating this part and the rapid-growth part separately, we are left with large dyadic tails and slow growth of ω_I.

For this remaining sum, the telescoping argument transfers the contributions to parents with small tails but a child with a large tail, up to terms already controlled by $\int_{\mathbb{R}}(Mf)^p w$. The parent–child relation for $J^{\mathscr{D}}_I$ forces a fixed relative drop in the average of w from each such parent to that child. Thus the transformed sum has additional structure: its intervals are selected by a drop in the average of the weight.

Carleson embedding and absorption. An entropy argument, based on the convexity of $t \log t$, controls how often these drops can occur. Where the ratio of the L^r average of w to its ordinary average is bounded, it yields a weighted Carleson estimate: the sum of the w-masses of the selected intervals contained in any test interval is bounded by a constant times its w-mass. The Carleson embedding lemma then controls this part of the sum by $\int_{\mathbb{R}}(Mf)^p w$.

For intervals where that ratio is large, we use the three-grid construction again to transfer the remaining terms to enlarged faithful intervals. The C_p condition supplies a coefficient proportional to the reciprocal of the ratio, so this coefficient can be made small. Apart from contributions controlled by the preceding estimates, the transferred terms have the same form as the sum we are trying to bound. Summing over the three grids gives a small multiple of that same total sum, which we move to the left-hand side. This absorption completes the argument.

Organization of the paper.

Section 2 collects the elementary facts used in the proof. Section 3 reduces the sparse estimate to a sum involving dyadic tails, and Section 4 transfers this sum to the boundary of the stopping trees. Entropy estimates for these boundary intervals are proved in Section 5. Section 6 constructs the assignment to three fixed grids. Section 7 completes the absorption argument and the proof of Theorem 1.1. Section 8 discusses the Fefferman–Stein inequality, the class SC_p, and the extension to higher dimensions and Calderón–Zygmund operators. Appendix A explains why the full sum in (3.1) cannot be estimated by the maximal-function integral.

2. Preliminary facts

All intervals are bounded and may be taken half-open. By a *dyadic grid* we mean a dyadic lattice in the sense of [8, Definition 2.1]. Thus a collection $\mathscr{D}$ of intervals is a dyadic grid if it satisfies the following three conditions:

(1) both dyadic children of every $I \in \mathscr{D}$ belong to $\mathscr{D}$;
(2) any two members of $\mathscr{D}$ have a common dyadic ancestor in $\mathscr{D}$;
(3) every compact subset of $\mathbb{R}$ is contained in some $I \in \mathscr{D}$.

In particular, every finite family of grid intervals has a common ancestor. After a fixed dilation, we take the interval lengths in the main argument to be 2^k, $k \in \mathbb{Z}$. For $I \in \mathscr{D}$, write I_-, I_+ for its children, $\widehat{I}$ for its parent, and $I^{(m)}$ for its mth ancestor.

We write $w(E) = \int_E w$ and $w_I = |I|^{-1} w(I)$. The notation $X \lesssim_{a,b} Y$ means that $X \le CY$ with a constant depending only on the displayed parameters. Constants in the main argument may depend on p, r, K, η, but never on the minimum scale or on the cardinality of a sparse family.

It suffices to prove Theorem 1.1 for finite sparse families and nonnegative bounded compactly supported f, since $|A_{\mathscr{S}} f| \le A_{\mathscr{S}} |f|$. Increasing finite subfamilies and the functions $\min(|f|, n)\mathbf{1}_{[-n,n]}$ then give the general result by monotone convergence. We may assume that

$$0 < \int_{\mathbb{R}} (Mf)^p w < \infty.$$

Under this assumption, all the tails in the C_p condition are finite. Indeed, choose an interval L with $f_L > 0$. Then

$$Mf \ge f_L M\mathbf{1}_L.$$

On L this is immediate; outside L, use the shortest interval containing L and the point in question. The maximal functions of the indicators of any two fixed intervals are comparable, with constants depending on those intervals. It follows that $\int_{\mathbb{R}} (M\mathbf{1}_I)^p w < \infty$ for every I. The C_p condition therefore also gives $w^r \in L^1_{\mathrm{loc}}$.

We will divide by averages of w. To avoid zero denominators, we reduce to $w > 0$ almost everywhere. For $\epsilon > 0$, Minkowski's inequality gives

$$\left(\frac{1}{|I|} \int_I (w+\epsilon)^r \right)^{1/r} \le \frac{K}{|I|} \int_{\mathbb{R}} (M\mathbf{1}_I)^p w + \epsilon \le \frac{\max(K,1)}{|I|} \int_{\mathbb{R}} (M\mathbf{1}_I)^p (w+\epsilon).$$

Thus $w + \epsilon$ satisfies the same type of C_p condition, uniformly in ϵ. We also need the maximal-function integral to remain finite after replacing w by $w + \epsilon$. Since $p > 1$ and f is bounded and compactly supported, the unweighted integral $\int_{\mathbb{R}} (Mf)^p$ is finite. Together with our assumption on the original weighted integral, this gives

$$\int_{\mathbb{R}} (Mf)^p (w+\epsilon) = \int_{\mathbb{R}} (Mf)^p w + \epsilon \int_{\mathbb{R}} (Mf)^p < \infty.$$

After proving the estimate for $w + \epsilon$, with a constant independent of ϵ, we let $\epsilon \downarrow 0$.

For every interval I, define the maximal average

$$\omega_I = \sup_{L \supseteq I} f_L.$$

The supremum is over all intervals containing I. This choice makes ω_I independent of any grid, which will allow us to compare intervals from different grids. Directly from the definition,

$$f_I \le \omega_I \le \inf_{x \in I} Mf(x), \qquad \omega_{\widehat{I}} \le \omega_I \le 2\omega_{\widehat{I}}. \tag{2.1}$$

For the last inequality, enlarge an interval containing I to include $\widehat{I}$; its length increases by at most a factor of two. More generally, the same argument shows that

$$I \subset Q, \quad |Q| \le a|I| \quad \Longrightarrow \quad \omega_Q \le \omega_I \le a\omega_Q. \tag{2.2}$$

We also have the useful pointwise inequality

$$\omega_I M\mathbf{1}_I \le Mf. \tag{2.3}$$

Indeed, for every $L \supseteq I$, $Mf \ge f_L M\mathbf{1}_L \ge f_L M\mathbf{1}_I$, and we take the supremum over L. This estimate will control the top intervals of the stopping trees.

Rapid growth of ω_I provides a separate, elementary source of summability. We state it for several generations at once because changing grids can move a jump away from the immediate parent.

Lemma 2.1. *Let $m \ge 1$ be an integer, and let $a > 1$ and $s > 0$. If $\mathscr{F} \subset \mathscr{D}$ satisfies $\omega_I \ge a\omega_{I^{(m)}}$ for every $I \in \mathscr{F}$, then*

$$\sum_{I\in\mathscr{F}} \omega_I^s \mathbf{1}_I \le \frac{m}{1-a^{-s}}(Mf)^s \qquad \text{almost everywhere.}$$

Proof. For every $I \in \mathscr{F}$,

$$(1-a^{-s})\omega_I^s \le \omega_I^s - \omega_{I^{(m)}}^s.$$

All differences on the right are nonnegative by (2.1). At a fixed point outside the grid boundaries, the intervals containing that point form a chain. The sum of these differences over all scales telescopes in m separate chains of generations, and is at most $m(Mf)^s$. Finite partial sums give the same bound, so monotone convergence proves the claim. □

We will also use the Carleson formulation of sparseness. For any family $\mathscr{S} \subset \mathscr{D}$, finite or infinite, the definition in the introduction is equivalent to

$$\sum_{\substack{I\in\mathscr{S}\\ I\subseteq L}} |I| \le \eta^{-1}|L| \qquad (L \in \mathscr{D}). \tag{2.4}$$

See [8, Lemma 6.3]. Summing the measures of the disjoint sets E_I also shows that (2.4) holds for every bounded interval L, even when $L \notin \mathscr{D}$.

The weighted version needed below concerns coefficients rather than lengths. We include the short proof so that all later entropy estimates can be used through the same elementary principle.

Lemma 2.2. *Suppose $a_I \ge 0$, $I \in \mathscr{D}$, and*

$$\sum_{\substack{I\in\mathscr{D}\\ I\subseteq L}} a_I \le \Lambda w(L) \qquad (L \in \mathscr{D}).$$

Then, for every nonnegative measurable function g,

$$\sum_{I\in\mathscr{D}} a_I \operatorname*{ess\,inf}_I g \le \Lambda \int_{\mathbb{R}} g\, w.$$

In particular,

$$\sum_{I\in\mathscr{D}} a_I \omega_I^p \le \Lambda \int_{\mathbb{R}} (Mf)^p w. \tag{2.5}$$

Proof. First restrict to finitely many coefficients. For $t > 0$, the maximal intervals with $\operatorname{ess\,inf}_I g > t$ are disjoint and lie, up to null sets, in $\{g > t\}$. Applying the hypothesis on these maximal intervals gives

$$\sum_{\operatorname{ess\,inf}_I g > t} a_I \le \Lambda w(\{g > t\}).$$

Integration in t proves the first assertion. Increasing finite subfamilies give the general case, and (2.1) gives (2.5). □

3. Sparse estimates and dyadic tails

The first lemma is the only place where the sparseness of the original operator is used analytically. It converts the operator estimate into a sum of local L^r averages of the weight. When these averages are comparable to w_I, the maximal function already controls the sum.

Lemma 3.1. *Let $\mathscr{S} \subset \mathscr{D}$ be finite and η-sparse. Then*

$$\int_{\mathbb{R}} (A_{\mathscr{S}} f)^p w \lesssim_{p,r,\eta} \sum_{I \in \mathscr{S}} f_I^p \left(\frac{1}{|I|}\int_I w^r\right)^{1/r} |I|. \tag{3.1}$$

If, in addition,

$$\left(\frac{1}{|I|}\int_I w^r\right)^{1/r} \le K_0 w_I \qquad (I \in \mathscr{S}),$$

then

$$\sum_{I \in \mathscr{S}} f_I^p \left(\frac{1}{|I|}\int_I w^r\right)^{1/r} |I| \lesssim_{p,r,\eta,K_0} \int_{\mathbb{R}} (Mf)^p w.$$

Proof. We use the standard principle that skipping layers improves the Carleson constant. More precisely, a Λ-Carleson family can be partitioned into m subfamilies with Carleson constant at most $1 + (\Lambda - 1)/m$; see [8, Lemma 6.6]. Taking $\Lambda = \eta^{-1}$ and m sufficiently large, we may therefore reduce to a family satisfying

$$\sum_{\substack{J \in \mathscr{S} \\ J \subsetneq I}} |J| \le \theta |I| \qquad (I \in \mathscr{S}), \tag{3.2}$$

where $\theta > 0$ is as small as needed. Minkowski's inequality and summation over the finitely many subfamilies recover the original family.

Let $\mathscr{S}_k$ be the kth generation of this family, starting with its maximal intervals, and set $U_k = \bigcup_{I \in \mathscr{S}_k} I$. For $I \in \mathscr{S}_k$, iteration of (3.2) gives

$$|I \cap U_{k+\ell}| \le \theta^\ell |I|.$$

Put $E_{I,\ell} = I \cap (U_{k+\ell} \setminus U_{k+\ell+1})$. For each fixed ℓ, these sets are pairwise disjoint as I varies, and $I = \bigsqcup_{\ell \ge 0} E_{I,\ell}$. Writing $r' = r/(r-1)$, Minkowski's and Hölder's inequalities give

$$\begin{aligned} \|A_{\mathscr{S}} f\|_{L^p(w)} &\le \sum_{\ell \ge 0} \left(\sum_{I \in \mathscr{S}} f_I^p w(E_{I,\ell})\right)^{1/p} \\ &\le \sum_{\ell \ge 0} \theta^{\ell/(pr')} \left(\sum_{I \in \mathscr{S}} f_I^p \left(\frac{1}{|I|}\int_I w^r\right)^{1/r} |I|\right)^{1/p}. \end{aligned}$$

The geometric series proves (3.1).

For the second assertion, choose the partition so that $K_0\theta^{1/r'} \le 1/2$. Let F_I be I minus the union of its immediate $\mathscr{S}$-children. The sets F_I are pairwise disjoint and

$$w(I \setminus F_I) \le \theta^{1/r'} \left(\frac{1}{|I|} \int_I w^r \right)^{1/r} |I| \le \tfrac{1}{2} w(I).$$

Consequently,

$$\sum_{I \in \mathscr{S}} f_I^p \left(\frac{1}{|I|} \int_I w^r \right)^{1/r} |I| \le 2K_0 \sum_{I \in \mathscr{S}} f_I^p w(F_I) \le 2K_0 \int_{\mathbb{R}} (Mf)^p w.$$

Summing over the subfamilies proves the second assertion. □

To use the C_p condition in a dyadic argument, we introduce the tail obtained by following the ancestors of an interval:

$$q = 2^{-(p-1)}, \qquad J_I = J_I^{\mathscr{D}} = \sum_{k \ge 0} q^k w_{I^{(k)}}.$$

The superscript will be omitted when the grid is fixed. These tails are finite by the preliminary reductions. If $M_{\mathscr{D}}$ denotes the dyadic maximal operator, then

$$J_I |I| = \frac{1}{1 - 2^{-p}} \int_{\mathbb{R}} (M_{\mathscr{D}} \mathbf{1}_I)^p w. \tag{3.3}$$

Indeed, for $x \in I^{(m)} \setminus I^{(m-1)}$ the dyadic maximal function is 2^{-m}, and on I it is one. Thus

$$\sum_{k \ge 0} 2^{-kp} \mathbf{1}_{I^{(k)}} = \frac{(M_{\mathscr{D}} \mathbf{1}_I)^p}{1 - 2^{-p}}.$$

The ancestors of I exhaust $\mathbb{R}$, by our definition of a grid. Integration proves (3.3). The definition also gives

$$J_{I_\pm} = q J_I + w_{I_\pm}, \qquad J_I \ge w_I. \tag{3.4}$$

The dyadic tail can miss part of the full tail. We call $I \in \mathscr{D}$ *faithful*, with constant F, if

$$\frac{1}{|I|} \int_{\mathbb{R}} (M \mathbf{1}_I)^p w \le F J_I^{\mathscr{D}}. \tag{3.5}$$

This is a property of the interval together with its grid. Section 6 will produce faithful intervals with one constant F depending only on p.

Fix a constant $c > 1$, depending only on p, so large that

$$q + \frac{2}{c} < (2/3)^{p-1}. \tag{3.6}$$

Such a choice is possible because $q = (1/2)^{p-1} < (2/3)^{p-1}$. The significance of this inequality will become clear in the stopping-tree calculation.

There are two easy parts of the sparse estimate. On faithful intervals with $J_I \le c w_I$, the C_p condition gives the local comparison required in Lemma 3.1. Intervals with $\omega_I \ge \frac{3}{2}\omega_{\widehat{I}}$ are controlled by Lemma 2.1. We are therefore left with large tails and slow growth of ω_I.

On the remaining faithful intervals, Lemma 3.1 and the C_p condition lead to terms of the form $f_I^p J_I |I|$. Before defining the sum that we will estimate, we deliberately enlarge these terms using

$$f_I^p \le \omega_I^{p-1} f_I.$$

The maximal average ω_I has the useful monotonicity and growth properties established in Section 2. Retaining one ordinary average f_I, however, preserves the quantity $f_I |I| = \int_I f$.

Its additivity over the two children will be used together with the recursion for J_I in the stopping-tree calculation. It also satisfies

$$f_I|I| = \int_I f \le \int_Q f = f_Q|Q| \qquad (I \subset Q),$$

which will allow us to transfer a contribution to a larger interval in another grid. Replacing f_I^p by ω_I^p would lose this direct comparison of the integrals of f.

With this choice in mind, fix a minimum interval length $h = 2^k$ and define

$$\mathcal{A}_h^{\mathscr{D}} = \sum_{\substack{I\in\mathscr{D},\ |I|\ge h\\ J_I>cw_I,\ \omega_I<\frac{3}{2}\omega_{\widehat{I}}}} \omega_I^{p-1} f_I J_I |I|. \tag{3.7}$$

Only the summation is truncated: each J_I still includes all ancestors.

This sum is finite. To check the upper scales, let I_* contain the support of f. If $f_I > 0$, the convex hull of I and I_* has length at most $|I| + |I_*|$, so

$$\omega_I \ge \frac{\|f\|_1}{|I| + |I_*|}, \qquad \omega_{\widehat{I}} \le \frac{\|f\|_1}{2|I|}.$$

Consequently, $\omega_I < \frac{3}{2}\omega_{\widehat{I}}$ implies $|I| < 3|I_*|$. There are only finitely many scales between h and this upper bound, and only finitely many intervals with $f_I > 0$ at each such scale. This observation will justify the absorption at the end of the proof.

Proposition 3.2. *Let $\mathscr{S} \subset \mathscr{D}$ be a finite η-sparse family of faithful intervals with constant F, and suppose $|I| \ge h$ for $I \in \mathscr{S}$. Then*

$$\int_{\mathbb{R}} (A_{\mathscr{S}} f)^p w \lesssim_{p,r,K,\eta,F} \int_{\mathbb{R}} (Mf)^p w + \mathcal{A}_h^{\mathscr{D}}.$$

Proof. For the subfamily on which $J_I \le cw_I$, the C_p condition and faithfulness (3.5) give

$$\left(\frac{1}{|I|}\int_I w^r\right)^{1/r} \le KFJ_I \le KFcw_I.$$

Both assertions of Lemma 3.1 bound this part of the operator by the required maximal-function integral.

For the subfamily $\mathscr{S}'$ with $J_I > cw_I$ and $\omega_I < \frac{3}{2}\omega_{\widehat{I}}$, the same inequalities and (3.1) give

$$\int_{\mathbb{R}} (A_{\mathscr{S}'} f)^p w \lesssim \sum_{I\in\mathscr{S}'} f_I^p J_I |I| \le \mathcal{A}_h^{\mathscr{D}}.$$

Here we used $f_I \le \omega_I$.

On the remaining subfamily $\mathscr{S}''$, $\omega_I \ge \frac{3}{2}\omega_{\widehat{I}}$. Lemma 2.1, with $m = s = 1$ and $a = 3/2$, gives

$$A_{\mathscr{S}''} f \le \sum_{I\in\mathscr{S}''} \omega_I \mathbf{1}_I \le 3Mf.$$

Combining the three parts proves the proposition. □

The sum $\mathcal{A}_h^{\mathscr{D}}$ runs over all intervals satisfying its two conditions, not just a sparse family. From this point until the final transfer of the original operator, we work with the full dyadic trees.

4. Stopping at small dyadic tails

We now estimate $\mathcal{A}_h^{\mathscr{D}}$ by summing parent–child differences over the intervals with large tails, $J_I > cw_I$. When both an interval and its parent have large tails, the corresponding terms cancel. Apart from a possible top-interval contribution, the positive terms left after cancellation come from intervals I with large tails whose parents P have small tails, $J_P \le cw_P$. We estimate these terms by a sum over their parents P; the other contributions are controlled by $\int_{\mathbb{R}}(Mf)^p w$.

For a fixed grid and minimum interval length h, let

$$\mathscr{P}_h^{\mathscr{D}} = \{P \in \mathscr{D} : \ |P| \ge 2h, \ J_P \le cw_P, \ J_I > cw_I \text{ for some child } I \text{ of } P\}. \tag{4.1}$$

These are the parents with a small tail and at least one child with a large tail. We call them *boundary parents*: each marks a transition from a small tail at P to a large tail at one of its children as we descend through the dyadic grid.

The estimate below retains the product $\omega_P^{p-1} f_P$ introduced in Section 3.

Proposition 4.1. *For every dyadic grid $\mathscr{D}$ and every minimum length $h = 2^k$,*

$$\mathcal{A}_h^{\mathscr{D}} \le C_p \int_{\mathbb{R}} (Mf)^p w + C_p \sum_{P \in \mathscr{P}_h^{\mathscr{D}}} \omega_P^{p-1} f_P w(P). \tag{4.2}$$

Moreover, every $P \in \mathscr{P}_h^{\mathscr{D}}$ satisfies

$$\min(w_{P_-}, w_{P_+}) \le \tau w_P, \qquad \tau = \frac{qc}{c-1} < 1. \tag{4.3}$$

Proof. Choose a single interval $L \in \mathscr{D}$, with $|L| \ge h$, containing every interval that contributes a nonzero summand to $\mathcal{A}_h^{\mathscr{D}}$. This is possible because there are only finitely many such intervals, as shown after (3.7), and a finite family in $\mathscr{D}$ has a common ancestor.

Inside L, consider the finite family

$$\mathscr{B} = \{I \in \mathscr{D} : I \subseteq L, \ |I| \ge h, \ J_I > cw_I\}.$$

The interval L only specifies the region in which we work; it is not automatically included in $\mathscr{B}$. In particular, if $J_L \le cw_L$, then $L \notin \mathscr{B}$, even if $\omega_L < \frac{3}{2}\omega_{\widehat{L}}$. Thus $\mathcal{A}_h^{\mathscr{D}}$ is the sum over those $I \in \mathscr{B}$ for which $\omega_I < \frac{3}{2}\omega_{\widehat{I}}$. The family $\mathscr{B}$ also includes intervals where this inequality fails. We will sum parent–child differences over the whole family $\mathscr{B}$ to obtain cancellation, so both types of intervals must be estimated. For every grid interval I, put

$$m_I = \omega_{\widehat{I}}^{p-1} f_I J_I |I|.$$

We will compare m_I with the sum of its two child terms. The use of $\omega_{\widehat{I}}$ makes their common factor equal to ω_I^{p-1}.

Separate the following three contributions:

$$\begin{aligned} A &= m_L, \\ B &= \sum_{\substack{I \in \mathscr{B}, \ I \ne L \\ \widehat{I} \notin \mathscr{B}}} m_I, \\ C &= 2 \sum_{\substack{I \in \mathscr{B} \\ \omega_I \ge \frac{3}{2}\omega_{\widehat{I}}}} \omega_I^{p-1} f_I w(I). \end{aligned}$$

Here A bounds the possible contribution of the top interval. If $L \notin \mathscr{B}$, this contribution is absent, and the nonnegative term $A = m_L$ merely enlarges the upper bound below. The sum B

collects large-tail intervals whose parents have small tails, and C accounts for intervals where ω grows rapidly. We first show that

$$\kappa \mathcal{A}_h^{\mathscr{D}} \le A + B + C, \qquad \kappa = (2/3)^{p-1} - q - \frac{2}{c} > 0, \tag{4.4}$$

and then estimate A, B, and C separately. Positivity of κ follows from (3.6).

Reduction to the three contributions. Using (3.4) and the identities for child averages, we obtain

$$\begin{aligned} m_{I_-} + m_{I_+} &= \omega_I^{p-1}|I| \left(qJ_I f_I + \frac{f_{I_-} w_{I_-} + f_{I_+} w_{I_+}}{2} \right) \\ &\le \omega_I^{p-1} f_I |I| (qJ_I + 2w_I). \end{aligned}$$

The last step uses nonnegativity together with $f_{I_-} + f_{I_+} = 2f_I$ and $w_{I_-} + w_{I_+} = 2w_I$. For $I \in \mathscr{B}$ with $\omega_I < \frac{3}{2}\omega_{\widehat{I}}$, namely the intervals retained in $\mathcal{A}_h^{\mathscr{D}}$, this gives

$$m_I - m_{I_-} - m_{I_+} \ge \kappa \omega_I^{p-1} f_I J_I |I|. \tag{4.5}$$

We next consider the intervals $I \in \mathscr{B}$ with $\omega_I \ge \frac{3}{2}\omega_{\widehat{I}}$. These do not enter $\mathcal{A}_h^{\mathscr{D}}$, but they occur in the sum over $\mathscr{B}$ used for cancellation. For these intervals, (2.1) gives $\omega_{\widehat{I}}^{p-1} \ge q\omega_I^{p-1}$, and hence

$$m_I - m_{I_-} - m_{I_+} \ge -2\omega_I^{p-1} f_I w(I). \tag{4.6}$$

Summing (4.5) over the slow-growth intervals in $\mathscr{B}$ and (4.6) over the rapid-growth intervals in $\mathscr{B}$ yields

$$\kappa \mathcal{A}_h^{\mathscr{D}} \le \sum_{I \in \mathscr{B}} (m_I - m_{I_-} - m_{I_+}) + C.$$

To see the cancellation, suppose that both I and $\widehat{I}$ belong to $\mathscr{B}$. The summand indexed by I contains $+m_I$, while the summand indexed by $\widehat{I}$ contains $-m_I$, because I is a child of $\widehat{I}$. These two occurrences of the nonnegative number m_I cancel. The only possible uncancelled terms with a plus sign are m_L, present only if $L \in \mathscr{B}$, and the terms defining B. Every other uncancelled term has the form $-m_J \le 0$ and can be discarded in an upper bound. Therefore

$$\sum_{I \in \mathscr{B}} (m_I - m_{I_-} - m_{I_+}) \le m_L + \sum_{\substack{I \in \mathscr{B},\ I \ne L \\ \widehat{I} \notin \mathscr{B}}} m_I = A + B.$$

This proves (4.4). The argument includes the smallest retained intervals: their children have length $h/2$ and occur only among the discarded negative terms.

Estimate of A. By (2.1), (3.3), and (2.3),

$$A = \omega_{\widehat{L}}^{p-1} f_L J_L |L| \le \omega_L^p J_L |L| \le \frac{1}{1 - 2^{-p}} \int_{\mathbb{R}} (Mf)^p w.$$

Estimate of B. If I occurs in B, its parent $P = \widehat{I}$ lies inside L and has length at least $2h$. Since $P \notin \mathscr{B}$, we have $J_P \le c w_P$. Thus $P \in \mathscr{P}_h^{\mathscr{D}}$. Using $w_I \le 2w_P$ and the tail recursion, we obtain

$$J_I = qJ_P + w_I \le (qc + 2) w_P.$$

Since $f_I|I| \le f_P|P|$, it follows that

$$m_I \le (qc + 2)\omega_P^{p-1} f_P w(P). \tag{4.7}$$

Each parent has at most two children. Summing (4.7) therefore gives

$$B \le 2(qc + 2) \sum_{P \in \mathscr{P}_h^{\mathscr{D}}} \omega_P^{p-1} f_P w(P).$$

This is the boundary sum in the statement, with the ordinary average f_P still present.

Estimate of C. Use $f_I \le \omega_I$ and apply Lemma 2.1 with $m = 1$, $a = 3/2$, and $s = p$. We get

$$C \le 2 \sum_{\substack{I \in \mathscr{D} \\ \omega_I \ge \frac{3}{2}\omega_{\widehat{I}}}} \omega_I^p w(I) \lesssim_p \int_{\mathbb{R}} (Mf)^p w.$$

Substituting the three estimates into (4.4) proves (4.2).

The weight drop at a boundary parent. It remains to prove (4.3). If I is a child of $P \in \mathscr{P}_h^{\mathscr{D}}$ with $J_I > cw_I$, then

$$cw_I < J_I = qJ_P + w_I \le qc\, w_P + w_I.$$

Rearranging gives $w_I < \frac{qc}{c-1} w_P = \tau w_P$. Finally, (3.6) implies $q + 2/c < 1$, so $c(1-q) > 1$ and hence $\tau < 1$. □

The boundary intervals in (4.2) need not be sparse with respect to Lebesgue measure. Their useful structure is instead (4.3): each one witnesses a fixed relative change in the average of the weight. The next section turns that change into a weighted counting estimate.

5. Entropy and weight drops

For an interval I, let

$$R_I = \frac{\left(\frac{1}{|I|}\int_I w^r\right)^{1/r}}{w_I}.$$

By Hölder's inequality, $R_I \ge 1$. This ratio measures how much larger the local L^r average is than the ordinary average. The entropy estimate below is useful when this ratio is bounded. When it is large, the C_p condition will instead force a large tail, and we will use that in the absorption argument.

For $0 < \tau < 1$ and a dyadic grid $\mathscr{D}$, define

$$\mathscr{O}_\tau^{\mathscr{D}} = \{I \in \mathscr{D} : \min(w_{I_-}, w_{I_+}) \le \tau w_I\}.$$

Thus $\mathscr{O}_\tau^{\mathscr{D}}$ consists of intervals with a definite drop in the average of w on one child. By (4.3), the boundary family $\mathscr{P}_h^{\mathscr{D}}$ is contained in this family for the fixed value of τ chosen in Section 4.

The containing interval in the next lemma is deliberately allowed to lie outside $\mathscr{D}$. This will be essential when an interval is moved from one grid to another.

Lemma 5.1. *Let $0 < \tau < 1$, let $\mathscr{D}$ be a dyadic grid, and let L be any bounded interval. Then*

$$\sum_{\substack{I \in \mathscr{O}_\tau^{\mathscr{D}} \\ I \subseteq L}} w(I) \le C_\tau \int_L w \log \frac{w}{w_L} \le C_\tau r' (\log R_L)\, w(L), \qquad r' = \frac{r}{r-1}. \tag{5.1}$$

Proof. Put $\phi(t) = t \log t$ for $t > 0$, with $\phi(0) = 0$. For $I \in \mathscr{D}$, the identity $w_I = (w_{I_-} + w_{I_+})/2$ and convexity give

$$\phi(w_I) \le \frac{\phi(w_{I_-}) + \phi(w_{I_+})}{2}.$$

Set

$$d_I = \frac{|I|}{2}\left(\phi(w_{I_-}) + \phi(w_{I_+}) - 2\phi(w_I)\right) \ge 0.$$

Strict convexity shows that $d_I = 0$ exactly when the two child averages are equal. A fixed relative drop on one child gives the following quantitative lower bound. If $I \in \mathscr{O}_\tau^{\mathscr{D}}$, write the two child averages as tw_I and $(2-t)w_I$, with $0 \le t \le \tau$. Then

$$d_I = \frac{w(I)}{2}\big(t\log t + (2-t)\log(2-t)\big) \ge \kappa_\tau w(I),$$

where

$$\kappa_\tau = \tfrac{1}{2}\big(\tau\log\tau + (2-\tau)\log(2-\tau)\big) > 0.$$

The lower bound follows because the expression in parentheses decreases on $[0,1]$.

Suppose first that $L \in \mathscr{D}$, and fix $n \ge 1$. Sum d_I over the dyadic subintervals of L in generations $0,\dots,n-1$. The child terms from one generation cancel the corresponding parent terms from the next, leaving

$$\sum_{\substack{I\in\mathscr{D},\ I\subseteq L\\ |I|>2^{-n}|L|}} d_I = \sum_{\substack{J\in\mathscr{D},\ J\subseteq L\\ |J|=2^{-n}|L|}} |J|\phi(w_J) - |L|\phi(w_L).$$

The intervals J on the right form the bottom generation and partition L. On each such interval, Jensen's inequality for the convex function ϕ gives

$$\phi(w_J) = \phi\left(\frac{1}{|J|}\int_J w\right) \le \frac{1}{|J|}\int_J \phi(w).$$

Multiplying by $|J|$ and summing over these disjoint intervals bounds the sum over J by $\int_L \phi(w)$. Therefore

$$\sum_{\substack{I\in\mathscr{D},\ I\subseteq L\\ |I|>2^{-n}|L|}} d_I \le \int_L \phi(w) - |L|\phi(w_L) = \int_L w\log\frac{w}{w_L}.$$

Since $d_I \ge 0$ for every I and $d_I \ge \kappa_\tau w(I)$ for $I \in \mathscr{O}_\tau^{\mathscr{D}}$, we may discard the other intervals and let $n \to \infty$. This proves the first inequality of (5.1) for dyadic L, with $C_\tau = \kappa_\tau^{-1}$.

For an arbitrary L, first take a finite subfamily of the sum on the left and let $L_1,\dots,L_N$ be its maximal intervals. They are disjoint dyadic intervals contained in L. The preceding argument applies on each L_a. Moreover,

$$\sum_{a=1}^N \int_{L_a} w\log\frac{w}{w_{L_a}} \le \int_L w\log\frac{w}{w_L}. \tag{5.2}$$

To verify this, let $E = L \setminus \bigcup_a L_a$. The difference between the right and left sides is

$$\sum_{a=1}^N |L_a|\phi(w_{L_a}) + \int_E \phi(w) - |L|\phi(w_L).$$

Suppose first that $|E| > 0$, and put $w_E = |E|^{-1}\int_E w$. Jensen's inequality on E gives

$$\int_E \phi(w) \ge |E|\phi(w_E).$$

Next, since $L_1,\dots,L_N,E$ form a partition of L, we have

$$w_L = \sum_{a=1}^N \frac{|L_a|}{|L|}w_{L_a} + \frac{|E|}{|L|}w_E.$$

The coefficients in this expression are nonnegative and sum to one. Applying convexity of ϕ to this weighted average and multiplying by $|L|$, we obtain

$$|L|\phi(w_L) \le \sum_{a=1}^{N} |L_a|\phi(w_{L_a}) + |E|\phi(w_E) \le \sum_{a=1}^{N} |L_a|\phi(w_{L_a}) + \int_E \phi(w).$$

Thus the displayed difference is nonnegative, as required. If $|E| = 0$, the coefficients $|L_a|/|L|$ already sum to one, and the same convexity argument applies with the E term omitted. This proves (5.2). Increasing finite subfamilies proves the first inequality for every L.

For the second inequality, use the normalized measure

$$d\mu(x) = \frac{w(x)}{w(L)}\,dx \quad \text{on } L, \qquad \mu(L) = 1.$$

Jensen's inequality for the concave function log says that $\int_L \log g\, d\mu \le \log \int_L g\, d\mu$. Apply it to $g = (w/w_L)^{r-1}$. This choice of power makes $gw = w^r/w_L^{r-1}$, so that the resulting integral is controlled by the L^r average of w. Using $\log(w/w_L) = (r-1)^{-1}\log g$, we obtain

$$\begin{aligned}
\frac{1}{w(L)}\int_L w \log\frac{w}{w_L} &= \frac{1}{r-1}\int_L \log g\, d\mu \\
&\le \frac{1}{r-1}\log\left(\frac{1}{w(L)}\int_L w\left(\frac{w}{w_L}\right)^{r-1}\right) \\
&= \frac{1}{r-1}\log\left(\frac{|L|^{-1}\int_L w^r}{w_L^r}\right) \\
&= r'\log R_L.
\end{aligned}$$

This completes the proof. □

The first consequence controls weight drops wherever the concentration ratio is bounded. The intervals involved need not form a sparse family.

Corollary 5.2. *For every $T > 1$,*

$$\sum_{\substack{I\in\mathscr{O}_\tau^{\mathscr{D}} \\ R_I\le T}} \omega_I^p w(I) \le C_{\tau,r}\log T \int_{\mathbb{R}} (Mf)^p w.$$

Proof. Consider a finite subfamily and a dyadic test interval L. Choose the maximal intervals of the subfamily contained in L. On each such interval Q, (5.1) bounds the sum of the weights of all its descendants in the subfamily by $C_\tau r' \log T\, w(Q)$. The maximal intervals are disjoint, so the coefficients $w(I)$ satisfy the hypothesis of Lemma 2.2 with $\Lambda = C_\tau r' \log T$. Applying (2.5) and increasing the finite subfamily proves the claim. □

Corollary 5.2 applies when R_P is bounded on the intervals P where the weight drops. Later we will encounter intervals for which R_P may be large, but for which we can choose a containing interval $Q(P)$ of comparable length with $R_{Q(P)} \le T$. The next lemma shows that this bound on the larger interval still suffices to control the sum over the original intervals P.

The intervals P may belong to several grids $\mathscr{D}_1, \dots, \mathscr{D}_N$. We require all the assigned intervals $Q(P)$ to belong to one grid $\mathscr{D}_*$. In the proof, we sum the w-masses $w(P)$ of all intervals assigned to the same Q and apply the Carleson embedding lemma in $\mathscr{D}_*$. The required Carleson estimate comes from Lemma 5.1: it can be applied with an assigned interval as the containing interval, even when that interval does not belong to the grid of the original intervals P.

Lemma 5.3. *Let $\mathscr{D}_1,\dots,\mathscr{D}_N,\mathscr{D}_*$ be dyadic grids. For each ν, let $\mathscr{F}_\nu\subset\mathscr{O}_\tau^{\mathscr{D}_\nu}$ and assign to every $P\in\mathscr{F}_\nu$ an interval $Q(P)\in\mathscr{D}_*$ such that*

$$P\subset Q(P),\qquad |Q(P)|\le a|P|,\qquad R_{Q(P)}\le T,$$

where $a\ge 1$ and $T>1$. Then

$$\sum_{\nu=1}^N\sum_{P\in\mathscr{F}_\nu}\omega_P^p w(P)\le C_{\tau,r}Na^p\log T\int_{\mathbb{R}}(Mf)^p w.$$

Proof. First assume that each $\mathscr{F}_\nu$ is finite. For every $Q\in\mathscr{D}_*$, sum the w-masses of the original intervals assigned to Q by setting

$$a_Q=\sum_{\nu=1}^N\sum_{\substack{P\in\mathscr{F}_\nu\\ Q(P)=Q}}w(P).$$

An empty sum is understood to be zero. We claim that these coefficients are w-Carleson in $\mathscr{D}_*$. Fix a test interval $L\in\mathscr{D}_*$. Among the assigned intervals $Q(P)\subseteq L$, choose those that are maximal with respect to inclusion, and denote them by Q_0. If $Q(P)\subseteq Q_0$, then $P\subset Q(P)\subseteq Q_0$. We may therefore apply Lemma 5.1 in each grid $\mathscr{D}_\nu$, using Q_0 as the containing interval. This gives

$$\sum_{\substack{Q\in\mathscr{D}_*\\ Q\subseteq Q_0}}a_Q\le C_\tau N\int_{Q_0}w\log\frac{w}{w_{Q_0}}\le C_\tau Nr'\log T\,w(Q_0).$$

Here $R_{Q_0}\le T$ because Q_0 is one of the assigned intervals $Q(P)$. The maximal intervals Q_0 are disjoint, and every $Q\subseteq L$ with $a_Q\neq 0$ is contained in one of them. Hence

$$\sum_{Q\subseteq L}a_Q\le C_\tau Nr'\log T\,w(L).$$

Using (2.2) and Lemma 2.2, we obtain

$$\sum_{\nu=1}^N\sum_{P\in\mathscr{F}_\nu}\omega_P^p w(P)\le a^p\sum_Q\omega_Q^p a_Q\le C_\tau Nr'a^p\log T\int_{\mathbb{R}}(Mf)^p w.$$

Increasing finite subfamilies of each $\mathscr{F}_\nu$ completes the proof. □

6. Three fixed dyadic grids

We now address the difference between a dyadic tail and the full tail in the C_p condition. A single fixed grid cannot always capture that tail. Three adjacent grids suffice after enlarging the interval by a fixed factor. The point is to use the adjacency at two scales: the scale of the enlarged interval and a larger scale reaching a given point of the line.

Apply the Three Lattice Theorem [8, Theorem 3.1] to an auxiliary grid $\mathscr{D}_*$ with interval lengths $2^k/3$. It gives three grids $\mathscr{D}^0,\mathscr{D}^1,\mathscr{D}^2$ such that

$$\mathscr{D}^0\cup\mathscr{D}^1\cup\mathscr{D}^2=\{[a,a+3\ell):[a,a+\ell)\in\mathscr{D}_*\}.$$

Each of these grids satisfies the definition in Section 2, and its interval lengths are 2^k. At a fixed scale 2^k, the left endpoints of the intervals from all three grids, taken together, form an arithmetic progression with spacing $2^k/3$. Within each individual grid, successive left endpoints are 2^k apart, so each grid contributes every third point of this progression. Thus, as we move along the line, the endpoints from the three grids occur in a repeating order, with spacing $2^k/3$

between consecutive points in the combined set. In particular, an interval of length less than $2^k/3$ is contained in a length-2^k interval of at least two grids.

Lemma 6.1. *There are constants C_p, F_p such that every interval P of length 2^ℓ can be assigned an index $j(P) \in \{0,1,2\}$ and an interval $Q(P) \in \mathscr{D}^{j(P)}$ satisfying*

$$P \subset Q(P), \qquad |Q(P)| = 8|P|, \tag{6.1}$$

$$\frac{1}{|P|}\int_{\mathbb{R}} (M\mathbf{1}_P)^p w \le C_p J_{Q(P)}^{\mathscr{D}^{j(P)}}, \tag{6.2}$$

and

$$\frac{1}{|Q(P)|}\int_{\mathbb{R}} (M\mathbf{1}_{Q(P)})^p w \le F_p J_{Q(P)}^{\mathscr{D}^{j(P)}}. \tag{6.3}$$

Thus every assigned interval $Q(P)$ is faithful with the same constant F_p. The assignment depends only on P, w, and the three fixed grids.

Proof. Set $H = 8|P|$. At this scale, each of the three grids partitions $\mathbb{R}$ into intervals of length H. Recall that the endpoints of these three partitions, taken together, are spaced $H/3$ apart. Since $|P| = H/8 < H/3$, the interval P can cross an endpoint of at most one of the three partitions. Thus, in at least two grids, P lies entirely inside one interval of length H. For every index j for which such an interval exists, denote it by Q_j. These are the candidates from which we will choose $Q(P)$.

Fix $x \in \mathbb{R}$ and put $d = |P| + \operatorname{dist}(x, P)$. The convex hull of P and x has length d. To estimate a dyadic maximal function at x, we look for an interval containing both x and one of the candidates Q_j, with length comparable to d. Choose a dyadic length L such that

$$\max(H, 4d) \le L < 2\max(H, 4d) \le 16d.$$

The convex hull has length $d \le L/4 < L/3$. The same argument using the spacing of endpoints at scale L shows that at least two grids have a length-L interval containing both P and x. There are only three grids, so at least one of these also has a candidate Q_j at scale H.

Fix such an index j. The length-L interval and Q_j belong to the same grid and both contain P. Since $L \ge H$, the nesting property implies that the length-L interval contains Q_j. It also contains x, so averaging $\mathbf{1}_{Q_j}$ over this interval gives

$$M_{\mathscr{D}^j}\mathbf{1}_{Q_j}(x) \ge \frac{H}{L} \ge \frac{1}{2}\frac{|P|}{|P| + \operatorname{dist}(x,P)} = \frac{1}{2}M\mathbf{1}_P(x).$$

The index j may depend on x, but the candidates Q_j were fixed before choosing x. Raising to the pth power and summing over all candidates therefore gives a pointwise bound by a fixed sum of at most three dyadic maximal functions to the pth power. Integrating and using (3.3), we obtain

$$\frac{1}{|P|}\int_{\mathbb{R}} (M\mathbf{1}_P)^p w \le 8(2^p - 1)\sum_j J_{Q_j}^{\mathscr{D}^j}.$$

Choose a candidate maximizing $J_{Q_j}^{\mathscr{D}^j}$, resolving ties by the smallest index, and call it $Q(P)$. This proves (6.1) and (6.2).

For nested intervals with $|Q(P)| = 8|P|$, the explicit formula for the maximal function of an indicator gives

$$M\mathbf{1}_{Q(P)} \le 8M\mathbf{1}_P.$$

Consequently,

$$\frac{1}{|Q(P)|}\int_{\mathbb{R}}(M\mathbf{1}_{Q(P)})^p w \le \frac{8^{p-1}}{|P|}\int_{\mathbb{R}}(M\mathbf{1}_P)^p w.$$

Together with (6.2), this proves (6.3). □

We henceforth fix this assignment. Two of its consequences explain how it will be used. The C_p condition and (6.2) imply

$$R_P w_P = \left(\frac{1}{|P|}\int_P w^r\right)^{1/r} \le C_p K J_{Q(P)}^{\mathscr{D}^{j(P)}}. \tag{6.4}$$

Thus a large concentration ratio R_P forces the selected dyadic tail to be large compared with w_P. Also, (6.1) and (2.2) give

$$\omega_P \le 8\omega_{Q(P)}. \tag{6.5}$$

The assignment can be applied to the original sparse family as well as to the boundary intervals. The following proposition ensures that the resulting operators remain sparse.

Proposition 6.2. *Let $\mathscr{S}$ be a finite η-sparse family in any dyadic grid. For $j = 0, 1, 2$, let*

$$\mathscr{S}_j = \{Q(I) : I \in \mathscr{S},\ j(I) = j\},$$

with repeated intervals $Q(I)$ counted only once. Then $\mathscr{S}_j \subset \mathscr{D}^j$ is $\eta/8$-sparse, every interval in $\mathscr{S}_j$ is faithful with constant F_p, and

$$A_{\mathscr{S}} f \le 64\sum_{j=0}^{2} A_{\mathscr{S}_j} f. \tag{6.6}$$

Proof. Faithfulness follows from (6.3). For each $I \in \mathscr{S}$, the size and containment relations give

$$f_I \mathbf{1}_I \le 8 f_{Q(I)} \mathbf{1}_{Q(I)}.$$

For a fixed Q, at most eight intervals I in the original grid satisfy $Q(I) = Q$: they all lie in Q, have length $|Q|/8$, and are disjoint. This proves (6.6).

For $L \in \mathscr{D}^j$ and each $Q \in \mathscr{S}_j$ contained in L, choose one original interval $I(Q) \in \mathscr{S}$ assigned to Q. These intervals $I(Q)$ are distinct and contained in L. If E_I are the disjoint sets witnessing the original sparseness, then

$$\sum_{\substack{Q\in\mathscr{S}_j\\ Q\subseteq L}} |Q| = 8\sum_{Q\subseteq L}|I(Q)| \le 8\eta^{-1}\sum_{Q\subseteq L}|E_{I(Q)}| \le 8\eta^{-1}|L|.$$

The equivalence (2.4) proves the asserted sparseness. □

One more geometric observation is needed. Rapid growth of ω from the parent of $Q(P)$ to $Q(P)$ need not imply rapid growth from the parent of P to P. It does, however, imply a definite increase from the sixth ancestor of P to P in the original grid, which is enough for Lemma 2.1.

Lemma 6.3. *Let $\mathscr{D}$ be a dyadic grid. Suppose $\mathscr{F} \subset \mathscr{D}$ consists of intervals P whose assigned intervals $Q(P)$ satisfy*

$$\omega_{Q(P)} \ge \tfrac{3}{2}\omega_{\widehat{Q(P)}},$$

where the parent of $Q(P)$ is taken in $\mathscr{D}^{j(P)}$. Then

$$\omega_{P^{(6)}} \le \tfrac{5}{6}\omega_P \qquad (P \in \mathscr{F}),$$

and

$$\sum_{P\in\mathscr{F}} \omega_P^p w(P) \lesssim_p \int_{\mathbb{R}} (Mf)^p w.$$

Proof. The intervals $P^{(6)}$ and $\widehat{Q(P)}$ both contain P, and have lengths $64|P|$ and $16|P|$, respectively. Fix any interval $L \supseteq P^{(6)}$, and let $\widetilde{L}$ be the smallest interval containing both L and $\widehat{Q(P)}$. These two intervals intersect, since both contain P. Hence there is no gap between them, and

$$|\widetilde{L}| \le |L| + |\widehat{Q(P)}| = |L| + 16|P| \le \tfrac{5}{4}|L|,$$

where the last inequality uses $|L| \ge 64|P|$.

Since $f \ge 0$ and $L \subseteq \widetilde{L}$, comparison of the integrals gives

$$f_L \le \frac{|\widetilde{L}|}{|L|} f_{\widetilde{L}} \le \tfrac{5}{4}\omega_{\widehat{Q(P)}}.$$

Here $f_{\widetilde{L}} \le \omega_{\widehat{Q(P)}}$ because $\widetilde{L}$ contains $\widehat{Q(P)}$, so its average is included in the supremum defining this maximal average. Taking the supremum over all $L \supseteq P^{(6)}$ gives the first inequality below. The next two follow from the hypothesis of the lemma and the inclusion $P \subset Q(P)$:

$$\omega_{P^{(6)}} \le \tfrac{5}{4}\omega_{\widehat{Q(P)}} \le \tfrac{5}{6}\omega_{Q(P)} \le \tfrac{5}{6}\omega_P,$$

which proves the first assertion. Apply Lemma 2.1 with $m = 6$, $a = 6/5$, and $s = p$, and integrate against w to obtain the second assertion. □

7. Absorption and completion of the proof

We apply the stopping-tree estimate simultaneously in the three grids $\mathscr{D}^0, \mathscr{D}^1, \mathscr{D}^2$. An interval $P \in \mathscr{P}_h^{\mathscr{D}^j}$ may be assigned to an interval $Q(P)$ in another grid. Summing the three main quantities allows such transfers to be absorbed in one inequality.

The families $\mathscr{P}_h^{\mathscr{D}^j}$ below are those defined in (4.1). The threshold T will be chosen only after the coefficient of the absorbable term is known.

Proposition 7.1. *Let w satisfy the C_p condition with exponent $r > 1$ and constant K. For every $T \ge 2$ and every minimum length $h = 2^k$,*

$$\sum_{j=0}^{2} \sum_{P\in\mathscr{P}_h^{\mathscr{D}^j}} \omega_P^{p-1} f_P w(P) \le C_{p,r,K,T} \int_{\mathbb{R}} (Mf)^p w + \frac{C_p K}{T} \sum_{j=0}^{2} \mathcal{A}_h^{\mathscr{D}^j}.$$

In particular, the constant C_p in the second term is independent of T, r, and K.

Proof. Recall that $\mathscr{P}_h^{\mathscr{D}^j}$ consists of intervals $P \in \mathscr{D}^j$ with $|P| \ge 2h$ such that $J_P \le c w_P$, while at least one of the two dyadic children I of P satisfies $J_I > c w_I$. For each such P, (4.3) gives

$$\min(w_{P_-}, w_{P_+}) \le \tau w_P.$$

Thus P belongs to $\mathscr{O}_\tau^{\mathscr{D}^j}$, the family of intervals with a fixed relative drop in the average of w on one child.

We first estimate the contribution of intervals with $R_P \le T$. Using $f_P \le \omega_P$ and Corollary 5.2, we obtain

$$\sum_{j=0}^{2} \sum_{\substack{P\in\mathscr{P}_h^{\mathscr{D}^j}\\ R_P\le T}} \omega_P^{p-1} f_P w(P) \le C_{p,r} \log T \int_{\mathbb{R}} (Mf)^p w.$$

It remains to estimate the intervals $P \in \mathscr{P}_h^{\mathscr{D}^j}$ with $R_P > T$. For each such P, let $Q = Q(P)$ be the assigned interval, and interpret J_Q and $\widehat{Q}$ in the grid $\mathscr{D}^{j(P)}$ containing Q. There are three possibilities.

Small tail at Q: $J_Q \le cw_Q$. The interval Q is faithful with constant F_p, so the C_p condition gives

$$R_Q w_Q \le \frac{K}{|Q|}\int_{\mathbb{R}} (M\mathbf{1}_Q)^p w \le KF_p J_Q \le KF_p c w_Q.$$

Put $T_0 = \max(2, KF_pc)$. For each $k \in \{0,1,2\}$, apply Lemma 5.3 to the intervals P under consideration for which $j(P) = k$, with $\mathscr{D}_* = \mathscr{D}^k$, $N = 3$, $a = 8$, and $T = T_0$. Its hypotheses hold because $P \subset Q$, $|Q| = 8|P|$, and $R_Q \le T_0$. Summing over $k = 0,1,2$ and using $f_P \le \omega_P$, we obtain a total contribution bounded by

$$C_{p,r,K}\int_{\mathbb{R}} (Mf)^p w.$$

The ratio R_P may be arbitrarily large here: the entropy estimate is applied on $Q(P)$, where the ratio is bounded.

Large tail and rapid growth at Q: $J_Q > cw_Q$ and $\omega_Q \ge \frac{3}{2}\omega_{\widehat{Q}}$. For each $j \in \{0,1,2\}$, Lemma 6.3 controls the sum of $\omega_P^p w(P)$ over these intervals $P \in \mathscr{P}_h^{\mathscr{D}^j}$. Again $f_P \le \omega_P$, so the total contribution of this case is at most

$$C_p \int_{\mathbb{R}} (Mf)^p w.$$

Large tail and slow growth at Q: $J_Q > cw_Q$ and $\omega_Q < \frac{3}{2}\omega_{\widehat{Q}}$. The estimate (6.4) gives $w_P \le C_p K J_Q/R_P$. Since $P \subset Q$ and $f \ge 0$, we also have $\int_P f \le \int_Q f$. Using these two estimates and (6.5), we obtain

$$\begin{aligned}
\omega_P^{p-1} f_P w(P) &= \omega_P^{p-1}\left(\int_P f\right) w_P \\
&\le \frac{8^{p-1}C_pK}{R_P}\omega_Q^{p-1}\left(\int_Q f\right) J_Q \\
&= \frac{8^{p-1}C_pK}{R_P}\omega_Q^{p-1} f_Q J_Q |Q| \\
&\le \frac{8^{p-1}C_pK}{T}\omega_Q^{p-1} f_Q J_Q |Q|.
\end{aligned} \tag{7.1}$$

The term indexed by Q belongs to $\mathcal{A}_h^{\mathscr{D}^{j(P)}}$. Indeed, the tail and growth conditions are precisely those in (3.7), and $|Q| = 8|P| \ge 16h$ preserves the lower cutoff.

For each fixed Q and each grid $\mathscr{D}^j$, at most eight intervals $P \in \mathscr{D}^j$ can satisfy $Q(P) = Q$, because they all have length $|Q|/8$ and lie inside Q. The total multiplicity of Q is thus at most 24. Summing (7.1) and absorbing the factor $24 \cdot 8^{p-1}$ into C_p bounds this case by

$$\frac{C_pK}{T}\sum_{j=0}^{2} \mathcal{A}_h^{\mathscr{D}^j}.$$

Combining the low-ratio estimate with the three cases proves the proposition. All counting estimates can first be applied to finite subfamilies of the original intervals and then extended by monotone convergence. □

The factor f_P in estimate (4.7) is used in (7.1) through the inequality $f_P|P| \le f_Q|Q|$. Replacing the boundary cost by $\omega_P^p w(P)$ would lose this comparison with the summand in $\mathcal{A}_h$. This is why we retained the ordinary average throughout the stopping-tree argument.

Proof of Theorem 1.1. Sum (4.2) over the three fixed grids and apply Proposition 7.1. For every $T \ge 2$,

$$\sum_{j=0}^{2} \mathcal{A}_h^{\mathscr{D}^j} \le C_{p,r,K,T} \int_{\mathbb{R}} (Mf)^p w + \frac{C_p' K}{T} \sum_{j=0}^{2} \mathcal{A}_h^{\mathscr{D}^j}.$$

The sums on both sides are finite, as verified after (3.7). Choose $T \ge \max(2, 2C_p'K)$ and absorb the last term. We conclude that

$$\sum_{j=0}^{2} \mathcal{A}_h^{\mathscr{D}^j} \lesssim_{p,r,K} \int_{\mathbb{R}} (Mf)^p w, \tag{7.2}$$

uniformly in the minimum length h.

Let $\mathscr{S}$ be the original finite sparse family. Apply Proposition 6.2 to obtain $\eta/8$-sparse families $\mathscr{S}_j \subset \mathscr{D}^j$ of faithful intervals. Choose h no larger than the smallest interval in these families. By (6.6), Proposition 3.2, and (7.2),

$$\begin{aligned}
\int_{\mathbb{R}} (A_{\mathscr{S}} f)^p w &\lesssim_p \sum_{j=0}^{2} \int_{\mathbb{R}} (A_{\mathscr{S}_j} f)^p w \\
&\lesssim_{p,r,K,\eta} \int_{\mathbb{R}} (Mf)^p w + \sum_{j=0}^{2} \mathcal{A}_h^{\mathscr{D}^j} \\
&\lesssim_{p,r,K,\eta} \int_{\mathbb{R}} (Mf)^p w.
\end{aligned}$$

The constants are independent of the finite family. The preliminary regularization of w and monotone convergence in f and in the sparse family now prove Theorem 1.1 in full generality. □

8. Concluding remarks

Remark 8.1 (The Fefferman–Stein inequality). Yabuta [11] proved that C_p is necessary for the weighted Fefferman–Stein inequality and that $C_{p+\varepsilon}$, for some $\varepsilon > 0$, is sufficient. Together with Yabuta's necessity result, Theorem 1.1 shows that the C_p condition is necessary and sufficient for the weighted Fefferman–Stein inequality.

Corollary 8.1. *Let $1 < p < \infty$ and $w \in C_p$. If $f \in L^1_{\mathrm{loc}}(\mathbb{R})$ and $|\{x : |f(x)| > t\}| < \infty$ for every $t > 0$, then*

$$\int_{\mathbb{R}} |f|^p w \le C \int_{\mathbb{R}} (f^{\#})^p w,$$

where $f^{\#}$ is the Fefferman–Stein sharp maximal function.

Proof. For a fixed sufficiently small $\lambda > 0$, the sparse oscillation estimate and the local sharp maximal function $M_\lambda^{\#} f$ satisfy

$$|f| \lesssim A_{\mathscr{S}}(M_\lambda^{\#} f), \qquad M(M_\lambda^{\#} f) \lesssim f^{\#};$$

see [7, Section 2.2 and Theorem 2.3] for the definition of $M_\lambda^{\#} f$ and the above inequalities. Apply Theorem 1.1, first to bounded compactly supported truncations of $M_\lambda^{\#} f$ and then pass to the limit. This gives

$$\int_{\mathbb{R}} |f|^p w \lesssim \int_{\mathbb{R}} [M(M_\lambda^{\#} f)]^p w \lesssim \int_{\mathbb{R}} (f^{\#})^p w. \qquad \square$$

Remark 8.2 (The SC_p condition). There is also a consequence for the class SC_p introduced in [7, Definition 1.1]. In one dimension, a weight w belongs to SC_p if there exist $r > 1$ and $C < \infty$ such that, for every finite family of pairwise disjoint bounded intervals $I_1, \dots, I_N$,

$$\sum_{j=1}^{N} \left(\frac{1}{|I_j|} \int_{I_j} w^r \right)^{1/r} |I_j| \le C \int_{\mathbb{R}} (M\mathbf{1}_{\bigcup_{j=1}^N I_j})^p w.$$

The constants are independent of the family and its cardinality. The formulation for countable families follows by monotone convergence.

Taking a family consisting of a single interval and dividing by its length gives exactly the C_p condition. Thus $SC_p \subset C_p$ follows directly from the definitions. The additional requirement in SC_p is to control the sum of the local L^r averages for an arbitrary disjoint family by one tail, formed from the indicator of its union.

The reverse inclusion follows from the Fefferman–Stein inequality proved above. If $w \in C_p$, then for every $t > 0$ and every f as in Corollary 8.1, Chebyshev's inequality gives

$$t^p w(\{x : |f(x)| > t\}) \le \int_{\mathbb{R}} |f|^p w \le C \int_{\mathbb{R}} (f^{\#})^p w.$$

This is the weighted weak type Fefferman–Stein inequality. By [7, Theorem 1.3], that inequality is equivalent to $w \in SC_p$. Hence $C_p \subset SC_p$, and therefore

$$SC_p = C_p \qquad (1 < p < \infty).$$

In particular, the apparently stronger requirement for disjoint families imposes no additional restriction on the weight.

Remark 8.3 (Higher dimensions). Theorem 1.1, Corollary 8.1, and the identity $SC_p = C_p$ extend to $\mathbb{R}^n$, with intervals replaced by cubes whose sides are parallel to the coordinate axes. We indicate the adjustments to the proof.

A dyadic cube has 2^n children. The appropriate tails are

$$q = 2^{-n(p-1)}, \qquad J_Q^{\mathscr{D}} = \sum_{k \ge 0} q^k w_{Q^{(k)}}.$$

They satisfy the same recursion $J_R = qJ_Q + w_R$ for every child R of Q. In the stopping-tree calculation, the factor 2 coming from the number of children becomes 2^n; we choose c so that $q + 2^n/c < (2/3)^{p-1}$. The resulting positive coefficient and the weight drop are therefore preserved. The entropy argument uses convexity of $t \log t$ for the 2^n child averages: a fixed relative drop on one child still gives an increment bounded below by a positive multiple of $w(Q)$, and the increments telescope as before. The maximal-average comparisons and the embedding argument also extend, with constants depending on n.

For the geometric step, take the 3^n product grids formed from the three one-dimensional grids in Section 6. The two-scale argument in Lemma 6.1 applies coordinate by coordinate. In each coordinate, at least two of the three grids are suitable at each of the two scales, so a common choice exists. The resulting product grid gives an assigned cube $Q(P) \supset P$ with $\ell(Q(P)) = 8\ell(P)$, where ℓ denotes side length, and the same tail comparisons, with constants

depending on n and p. At most 8^n cubes from any one original grid can be assigned to the same cube.

In Lemma 6.3, the sixth ancestor is replaced by the mth ancestor, where $m = m(n)$ is large enough that $(1 + 2^{4-m})^n < 3/2$. Indeed, when $\omega_{Q(P)} \geq \frac{3}{2}\omega_{\widehat{Q(P)}}$, the same enlargement of a cube containing $P^{(m)}$ gives

$$\omega_{P^{(m)}} \leq (1 + 2^{4-m})^n \omega_{\widehat{Q(P)}} \leq \tfrac{2}{3}(1 + 2^{4-m})^n \omega_P.$$

The coefficient of ω_P is strictly less than one, so Lemma 2.1 again applies. These changes preserve the absorption argument and prove the sparse estimate in every dimension. The proofs in Remarks 8.1 and 8.2 then apply to cubes as well.

Finally, standard sparse domination, see [7, Theorem 2.2], shows that C_p is sufficient for the corresponding inequality for Calderón–Zygmund operators. More precisely, if $1 < p < \infty$, $w \in C_p(\mathbb{R}^n)$, and T is an L^2-bounded Calderón–Zygmund operator with a kernel satisfying the standard size and Dini continuity conditions, then

$$\int_{\mathbb{R}^n} |Tf|^p w \leq C \int_{\mathbb{R}^n} (Mf)^p w$$

for every bounded compactly supported f. The constant depends only on n, p, the parameters of the C_p condition, and the operator bounds and kernel constants of T.

Appendix A. Why the sparse sum cannot be estimated directly

The C_p condition does not, in general, allow us to bound the sum on the right-hand side of (3.1) by a constant multiple of $\int_{\mathbb{R}} (Mf)^p w$. The following example shows that this sum may be infinite even when the maximal-function integral is finite. Moreover, choosing the exponent r closer to one does not remove this obstruction.

Proposition A.1. *Let $1 < p < \infty$ and define*

$$w = \sum_{k=2}^{\infty} \frac{2^{k(p-1)}}{k^{1-1/p}(\log k)^{2-1/p}} \mathbf{1}_{[2^k,\, 2^k + 2^k/(k \log k)^{1/p}]}.$$

Then $w \in C_p$, and the defining C_p inequality holds for every $r > 1$. There are a dyadic grid $\mathscr{D}$ and a $3/4$-sparse family $\mathscr{S} \subset \mathscr{D}$ such that, for $f = \mathbf{1}_{[0,1]}$,

$$\int_{\mathbb{R}} (Mf)^p w < \infty, \qquad \sum_{I \in \mathscr{S}} f_I^p \left(\frac{1}{|I|} \int_I w^r \right)^{1/r} |I| = \infty \quad \text{for every } r > 1.$$

In particular, for this fixed weight and any $r > 1$, the right-hand side of (3.1) cannot be bounded by a constant multiple of $\int_{\mathbb{R}} (Mf)^p w$ uniformly over finite sparse families.

Proof. Write

$$\ell_k = \frac{2^k}{(k \log k)^{1/p}}, \qquad E_k = [2^k, 2^k + \ell_k], \qquad a_k = \frac{2^{k(p-1)}}{k^{1-1/p}(\log k)^{2-1/p}},$$

so that $w = \sum_{k \geq 2} a_k \mathbf{1}_{E_k}$. We have $\ell_k < 2^k$, so these intervals are disjoint. The two identities

$$a_k = \frac{\ell_k^{p-1}}{\log k}, \qquad \frac{a_k \ell_k}{2^{kp}} = \frac{1}{k(\log k)^2}$$

explain the choice of their lengths and heights. The weight is locally bounded, and the second identity gives

$$\int_4^\infty \frac{w(x)}{x^p}\,dx \le \sum_{k=2}^\infty \frac{1}{k(\log k)^2} < \infty.$$

Consequently, $\int_{\mathbb{R}}(M\mathbf{1}_I)^p w$ is finite for every bounded interval I: on a bounded set this follows from local integrability, and for large positive x one has $M\mathbf{1}_I(x) \lesssim_I x^{-1}$.

Verification of the C_p condition. We prove the stronger estimate

$$\operatorname*{ess\,sup}_I w \lesssim_p \frac{1}{|I|}\int_{\mathbb{R}}(M\mathbf{1}_I)^p w \qquad \text{for every bounded interval } I. \tag{A.1}$$

Fix I, put $L = |I|$, and consider any index $n \ge 2$ for which $|I \cap E_n| > 0$. It is enough to bound a_n by the right-hand side of (A.1), with a constant independent of I and n.

First suppose $L \le \ell_n$. Choose $z \in I \cap E_n$. Since E_n has length at least L, there is a subinterval $F \subset E_n$ of length L such that $|x - z| \le L$ for $x \in F$. An interval containing both I and such an x has length at most $2L$. Therefore $M\mathbf{1}_I(x) \ge 1/2$ on F, and

$$\frac{1}{L}\int_{\mathbb{R}}(M\mathbf{1}_I)^p w \ge \frac{2^{-p}}{L}\int_F w = 2^{-p}a_n.$$

Now suppose $L > \ell_n$. Let $m \ge n$ be the smallest integer such that $L \le 2^m$. Because I meets $E_n \subset [2^n, 2^{n+1}]$, we have

$$I \subset [-2^m, 3\cdot 2^m].$$

If $k \ge m+3$ and $x \in E_k$, an interval containing I and x has length at most $3\cdot 2^k$. Hence

$$M\mathbf{1}_I(x) \ge \frac{L}{3\cdot 2^k},$$

and summing the contributions of these distant intervals gives

$$\begin{aligned}\frac{1}{L}\int_{\mathbb{R}}(M\mathbf{1}_I)^p w &\ge 3^{-p}L^{p-1}\sum_{k=m+3}^\infty \frac{a_k\ell_k}{2^{kp}}\\ &= 3^{-p}L^{p-1}\sum_{k=m+3}^\infty \frac{1}{k(\log k)^2} \gtrsim_p \frac{L^{p-1}}{\log m}.\end{aligned}$$

The last inequality follows by comparison with $\int_{m+3}^\infty dt/(t(\log t)^2) = 1/\log(m+3)$.

If $m = n$, this lower bound is at least a constant multiple of $\ell_n^{p-1}/\log n = a_n$. If $m > n$, the choice of m gives $L > 2^{m-1}$. We use the elementary comparison

$$\frac{2^{n(p-1)}}{\log n} \lesssim_p \frac{2^{m(p-1)}}{\log m} \qquad (m \ge n \ge 2).$$

Indeed, writing $d = m - n$, the ratio of the left-hand side to the right-hand side is at most $(1+d)2^{-d(p-1)}$, which is bounded in terms of p; here $\log m/\log n \le 1 + d$ follows from $\log(m/n) \le d/n$ and $n\log n \ge 1$. It follows that

$$\frac{1}{L}\int_{\mathbb{R}}(M\mathbf{1}_I)^p w \gtrsim_p \frac{2^{m(p-1)}}{\log m} \gtrsim_p \frac{2^{n(p-1)}}{\log n} \ge a_n.$$

Thus in every case the tail controls a_n. Taking the maximum over the finitely many intervals E_n meeting I in positive measure proves (A.1); if there are none, its left-hand side is zero. Finally,

$$\left(\frac{1}{|I|}\int_I w^r\right)^{1/r} \le \operatorname*{ess\,sup}_I w,$$

so (A.1) proves the C_p condition for every $r > 1$, with a constant depending only on p.

Failure of the bound for the sparse sum. Set $f = \mathbf{1}_{[0,1]}$. Since $Mf(x) = 1/x$ for $x > 1$ and w is supported in $[4, \infty)$, the preceding calculation gives

$$\int_{\mathbb{R}} (Mf)^p w = \sum_{k=2}^{\infty} a_k \int_{E_k} x^{-p}\, dx \le \sum_{k=2}^{\infty} \frac{1}{k(\log k)^2} < \infty.$$

For $j \ge 1$, let

$$I_j = [-4^j, 2 \cdot 4^j), \qquad \mathscr{S} = \{I_j : j \ge 1\}.$$

Each I_j is the second of the four dyadic grandchildren of I_{j+1}. The collection of all dyadic subintervals of the I_j therefore forms a dyadic grid. Since $I_j \uparrow \mathbb{R}$, this grid satisfies the definition in Section 2. The sets I_1 and $I_j \setminus I_{j-1}$, $j \ge 2$, witness that $\mathscr{S}$ is 3/4-sparse.

We have $E_{2j} \subset I_j$, $|I_j| = 3 \cdot 4^j$, and $f_{I_j} = |I_j|^{-1}$. Fix $r > 1$ and put

$$\beta = \frac{p - 1 + 1/r}{p} = 1 - \frac{1 - 1/r}{p} < 1.$$

Keeping only the contribution of E_{2j} in the L^r average, we obtain

$$f_{I_j}^p \left(\frac{1}{|I_j|} \int_{I_j} w^r \right)^{1/r} |I_j| \ge |I_j|^{1-p-1/r} a_{2j} \ell_{2j}^{1/r} = \frac{3^{1-p-1/r}}{(2j)^{\beta} (\log(2j))^{1+\beta}}.$$

The series of these lower bounds diverges because $\beta < 1$. Thus the sparse sum is infinite for every $r > 1$. Taking the finite subfamilies $\{I_1, \dots, I_N\}$ and letting $N \to \infty$ proves the asserted failure of a uniform estimate for finite families. □

The role of the rapid-growth subfamily. Every interval in this example belongs to the rapid-growth subfamily. Both I_j and its dyadic parent contain the support of f, so

$$\omega_{I_j} = \frac{1}{|I_j|}, \qquad \omega_{\widehat{I_j}} = \frac{1}{2|I_j|}.$$

In particular, $\omega_{I_j} = 2\omega_{\widehat{I_j}}$, and none of these intervals satisfies the slow-growth restriction in (3.7). Their contribution to the operator itself is harmless. Indeed, if j is the first index for which $x \in I_j$, then

$$A_{\mathscr{S}} f(x) = \sum_{k=j}^{\infty} \frac{1}{3 \cdot 4^k} = \frac{4}{3} \frac{1}{|I_j|} \le \frac{4}{3} Mf(x).$$

Thus the operator is pointwise controlled by Mf, whereas the sum in (3.1) diverges. Proposition 3.2 therefore removes this subfamily before applying Lemma 3.1 to the remaining part.

Acknowledgements and use of AI. AL was supported by ISF grant no. 807/26. SO was partially supported by CONICET, grant KE2.11220250100060CO.

The first version of this paper was written jointly with Fedor Nazarov, who was originally expected to be a coauthor but ultimately chose not to be. We thank Fedor Nazarov for numerous discussions, new ideas, and for his interest in this work. That version was prepared without the use of AI and included a complete proof of the main results. The proof already used Lemma 3.1, after first treating separately the intervals on which the maximal averages of f grow rapidly. The estimate for the remaining sum relied on Bellman-type and probabilistic arguments.

For the present version, ChatGPT was used to develop an alternative treatment of the remaining sum, replacing the Bellman-type and probabilistic arguments with the deterministic arguments presented in Sections 4–7. The authors have checked the resulting proof and take full responsibility for the mathematical content of this paper.

References

[1] S. Buckley, *Harmonic analysis on weighted spaces*, Ph.D. thesis, The University of Chicago, ProQuest LLC, Ann Arbor, MI, 1990.

[2] J. Canto, *Sharp reverse Hölder inequality for C_p weights and applications*, J. Geom. Anal. **31** (2021), no. 4, 4165–4190.

[3] J. Canto, K. Li, L. Roncal, and O. Tapiola, *C_p estimates for rough homogeneous singular integrals and sparse forms*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5) **22** (2021), no. 3, 1131–1168.

[4] M.E. Cejas, K. Li, C. Pérez, and I. P. Rivera-Ríos, *Vector-valued operators, optimal weighted estimates and the C_p condition*, Sci. China Math. **63** (2020), no. 7, 1339–1368.

[5] L. Kahanpää and L. Mejlbro, *Some new results on the Muckenhoupt conjecture concerning weighted norm inequalities connecting the Hilbert transform with the maximal function*, Proceedings of the second Finnish-Polish summer school in complex analysis (Jyväskylä, 1983), 53–72, Bericht, 28, Univ. Jyväskylä, Jyväskylä, 1984.

[6] A.K. Lerner, *Some remarks on the Fefferman–Stein inequality*, J. Anal. Math. **112** (2010), 329–349.

[7] A.K. Lerner, *A characterization of the weighted weak type Coifman–Fefferman and Fefferman–Stein inequalities*, Math. Ann. **378** (2020), no. 1–2, 425–446.

[8] A.K. Lerner and F. Nazarov, *Intuitive dyadic calculus: the basics*, Expo. Math. **37** (2019), no. 3, 225–265.

[9] B. Muckenhoupt, *Norm inequalities relating the Hilbert transform to the Hardy–Littlewood maximal function*, Functional analysis and approximation (Oberwolfach, 1980), 219–231, Internat. Ser. Numer. Math., 60, Birkhäuser, Basel–Boston, Mass., 1981.

[10] E.T. Sawyer, *Norm inequalities relating singular integrals and the maximal function*, Studia Math. **75** (1983), 253–263.

[11] K. Yabuta, *Sharp maximal function and C_p condition*, Arch. Math. **55** (1990), no. 2, 151–155.

Department of Mathematics, Bar-Ilan University, 5290002 Ramat Gan, Israel
Email address: lernera@math.biu.ac.il

Departamento de Matemática, Universidad Nacional del Sur, Bahía Blanca, 8000, Argentina
Email address: sombrosi@uns.edu.ar