Endpoint Sobolev bounds for fractional Hardy–Littlewood maximal operators

Julian Weigt[1]

Abstract

Let $0 < \alpha < d$ and $1 \leq p < d/\alpha$. We present a proof that for all $f \in W^{1,p}(\mathbb{R}^d)$ both the centered and the uncentered Hardy–Littlewood fractional maximal operator $\mathrm{M}_\alpha f$ are weakly differentiable and $\|\nabla \mathrm{M}_\alpha f\|_{p^*} \leq C_{d,\alpha,p}\|\nabla f\|_p$, where $p^* = (p^{-1} - \alpha/d)^{-1}$. In particular it covers the endpoint case $p = 1$ for $0 < \alpha < 1$ where the bound was previously unknown. For $p = 1$ we can replace $W^{1,1}(\mathbb{R}^d)$ by $\mathrm{BV}(\mathbb{R}^d)$. The ingredients used are a pointwise estimate for the gradient of the fractional maximal function, the layer cake formula, a Vitali type argument, a reduction from balls to dyadic cubes, the coarea formula, a relative isoperimetric inequality and an earlier established result for $\alpha = 0$ in the dyadic setting. We use that for $\alpha > 0$ the fractional maximal function does not use certain small balls. For $\alpha = 0$ the proof collapses.

Keywords Fractional maximal function · Variation · Dyadic cubes

Mathematics Subject Classification 42B25 · 26B30

1 Introduction

For $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ and a ball or cube B, we denote

$$f_B = \frac{1}{\mathcal{L}(B)} \int_B |f|.$$

The centered Hardy–Littlewood maximal function is defined by

$$\mathrm{M}^{\mathrm{c}} f(x) = \sup_{r>0} f_{B(x,r)},$$

and the uncentered Hardy–Littlewood maximal function is defined by

$$\widetilde{\mathrm{M}} f(x) = \sup_{B \ni x} f_B$$

✉ Julian Weigt
 julian.weigt@aalto.fi

[1] Department of Mathematics and Systems Analysis, Aalto University, P.O. Box 11100, 00076 Espoo, Finland

where the supremum is taken over all balls that contain x. The regularity of a maximal operator was first studied by Kinnunen in 1997. He proved in [18] that for each $p > 1$ and $f \in W^{1,p}(\mathbb{R}^d)$ the bound

$$\|\nabla \mathrm{M} f\|_p \leq C_{d,p} \|\nabla f\|_p \tag{1.1}$$

holds for $\mathrm{M} = \mathrm{M}^c$. Formula (1.1) also holds for $\mathrm{M} = \widetilde{\mathrm{M}}$. This implies that both Hardy–Littlewood maximal operators are bounded on Sobolev spaces with $p > 1$. His proof does not apply for $p = 1$. Note that unless $f = 0$ also $\|\mathrm{M} f\|_1 \leq C_{d,1}\|f\|_1$ fails since $\mathrm{M} f$ is not in $L^1(\mathbb{R}^d)$. In [16] Hajłasz and Onninen asked whether formula (1.1) also holds for $p = 1$ for the centered Hardy–Littlewood maximal operator. This question has become a well known problem for various maximal operators and there has been lots of research on this topic. So far it has mostly remained unanswered, but there has been some progress. For the uncentered maximal function and $d = 1$ it has been proved in [28] by Tanaka and later in [22] by Kurka for the centered Hardy–Littlewood maximal function. The proof for the centered maximal function turned out to be much more complicated. Aldaz and Pérez Lázaro obtained in [3] the sharp improvement $\|\nabla \widetilde{\mathrm{M}} f\|_{L^1(\mathbb{R})} \leq \|\nabla f\|_{L^1(\mathbb{R})}$ of Tanaka's result. For the uncentered Hardy–Littlewood maximal function Hajłasz's and Onninen's question already also has a positive answer for all dimensions d in several special cases. For radial functions Luiro proved it in [24], for block decreasing functions Aldaz and Pérez Lázaro proved it in [2] and for characteristic functions the author proved it in [30]. As a first step towards weak differentiability, Hajłasz and Malý proved in [15] that for $f \in L^1(\mathbb{R}^d)$ the centered Hardy–Littlewood maximal function is approximately differentiable. In [1] Aldaz et al. proved bounds on the modulus of continuity for all dimensions.

A related question is whether the maximal operator is a continuous operator. Luiro proved in [23] that for $p > 1$ the uncentered maximal operator is continuous on $W^{1,p}(\mathbb{R}^d)$. There is ongoing research for the endpoint case $p = 1$. For example Carneiro et al. proved in [11] that $f \mapsto \nabla \widetilde{\mathrm{M}} f$ is continuous $W^{1,1}(\mathbb{R}) \to L^1(\mathbb{R})$ and in [14] González-Riquelme and Kosz recently improved this to continuity on BV. Carneiro et al. proved in [8] that for radial functions f, the operator $f \mapsto \nabla \widetilde{\mathrm{M}} f$ is continuous as a map $W^{1,1}(\mathbb{R}^d) \to L^1(\mathbb{R}^d)$.

The regularity of maximal operators has also been studied for other maximal operators and on other spaces. We focus on the endpoint $p = 1$. In [12] Carneiro and Svaiter and in [7] Carneiro and González-Riquelme investigated maximal convolution operators M associated to certain partial differential equations. Analogous to the Hardy–Littlewood maximal operator they proved $\|\nabla \mathrm{M} f\|_{L^1(\mathbb{R}^d)} \leq C_d \|\nabla f\|_{L^1(\mathbb{R}^d)}$ for $d = 1$, and for $d > 1$ if f is radial. In [9] Carneiro and Hughes proved $\|\nabla \mathrm{M} f\|_{l^1(\mathbb{Z}^d)} \leq C_d \|f\|_{l^1(\mathbb{Z}^d)}$ for centered and uncentered discrete maximal operators. This bound does not hold on $\mathbb{R}^d$, but because in the discrete setting we have $\|\nabla f\|_{l^1(\mathbb{Z}^d)} \leq C_d \|f\|_{l^1(\mathbb{Z}^d)}$, it is weaker than the still open $\|\nabla \mathrm{M} f\|_{l^1(\mathbb{Z}^d)} \leq C_d \|\nabla f\|_{l^1(\mathbb{Z}^d)}$. In [21] Kinnunen and Tuominen proved the boundedness of a discrete maximal operator in the metric Hajłasz Sobolev space $M^{1,1}$. In [27] Pérez et al. proved the boundedness of certain convolution maximal operators on Hardy-Sobolev spaces $\dot{H}^{1,p}$ for a sharp range of exponents, including $p = 1$. In [29] the author proved $\mathrm{var}\, \mathrm{M}^d f \leq C_d \mathrm{var}\, f$ for the dyadic maximal operator for all dimensions d.

For $0 \leq \alpha \leq d$ the centered fractional Hardy–Littlewood maximal function is defined by

$$\mathrm{M}_\alpha^c f(x) = \sup_{r>0} r^\alpha f_{B(x,r)}.$$

For a ball B we denote the radius of B by $r(B)$. The uncentered fractional Hardy–Littlewood maximal function is defined by

$$\widetilde{M}_\alpha f(x) = \sup_{B \ni x} r(B)^\alpha f_B$$

where the supremum is taken over all balls that contain x. Note that M_α does not make much sense for $\alpha > d$. For $\alpha = 0$ it is the Hardy–Littlewood maximal function. The following is the fractional version of formula (1.1).

Theorem 1.1 *Let $1 \leq p < \infty$ and $0 < \alpha < d/p$ and $M_\alpha \in \{M_\alpha^c, \widetilde{M}_\alpha\}$. Then for all $f \in W^{1,p}(\mathbb{R}^d)$ we have that $M_\alpha f$ is weakly differentiable with*

$$\|\nabla M_\alpha f\|_{(p^{-1}-\alpha/d)^{-1}} \leq C_{d,\alpha,p} \|\nabla f\|_p \tag{1.2}$$

where the constant $C_{d,\alpha,p}$ depends only on d, α and p. In the endpoint $p = 1$ we can replace $f \in W^{1,1}(\mathbb{R}^d)$ by $f \in BV(\mathbb{R}^d)$. The endpoint result for $p = d/\alpha$ holds true as well.

We prove Theorem 1.1 in Sect. 2.1. The study of the regularity of the fractional maximal operator was initiated by Kinnunen and Saksman. They proved in [20,Theorem 2.1] that formula (1.2) holds for $0 \leq \alpha < d/p$ and $1 < p < \infty$. They showed $|\nabla M_\alpha^c f(x)| \leq M_\alpha |\nabla f|(x)$ for almost every $x \in \mathbb{R}^d$, and then concluded formula (1.2) from the $L^{(p^{-1}-\alpha/d)^{-1}}$-boundedness of M_α, which fails for $p = 1$. Another result by Kinnunen and Saksman in [20] is that for all $\alpha \geq 1$ we have $|\nabla M_\alpha^c f(x)| \leq (d-\alpha)M_{\alpha-1} f(x)$ for almost every $x \in \mathbb{R}^d$. In [10] Carneiro and Madrid used this, the $L^{d/(d-\alpha)}$-boundedness of $M_{\alpha-1}$, and Sobolev embedding to concluded formula (1.2). All of this also works for the uncentered fractional maximal function $\widetilde{M}_\alpha$. The strategy fails for $\alpha < 1$.

Our main result is the extension of formula (1.2) to the endpoint $p = 1$ for $0 < \alpha < 1$ which has been an open problem. Our proof of Theorem 1.1 also works for $1 \leq \alpha \leq d$, and further extends to $1 \leq p < \infty$, $0 < \alpha \leq d/p$. We present the proof for this range of parameters here, since it also smoothens out the blowup of the constants for $p \to 1$ which occurs in the previous proof for $p > 1$. Note that interpolation is not immediately available for results on the gradient level. Our approach fails for $\alpha = 0$. The corner point $\alpha = 0$, $p = 1$ is the earlier mentioned question by Hajłasz and Onninen and remains open. Similarly to Carneiro and Madrid, we begin the proof with a pointwise estimate $|\nabla M_\alpha f(x)| \leq (d-\alpha)M_{\alpha,-1} f(x)$ which holds for all $0 < \alpha < d$ for bounded functions. We estimate $M_{\alpha,-1} f$ in Theorem 1.2 and from that conclude Theorem 1.1.

For the centered fractional maximal function define

$$\mathcal{B}_\alpha^c(x) = \{B(x,r)\}$$

where r is the largest radius such that $M_\alpha^c f(x) = r^\alpha f_{B(x,r)}$ and for the uncentered fractional maximal function define

$$\widetilde{\mathcal{B}}_\alpha(x) = \left\{B : x \in \overline{B},\ r(B)^\alpha f_B = \widetilde{M}_\alpha f(x),\ \forall A \supsetneq B\ r(A)^\alpha f_A < \widetilde{M}_\alpha f(x)\right\}.$$

Then for almost every $x \in \mathbb{R}^d$ the sets $\mathcal{B}_\alpha^c(x)$ and $\widetilde{\mathcal{B}}_\alpha(x)$ are nonempty, i.e. the supremum in the definition of the maximal function is attained in a largest ball B with $x \in \overline{B}$, see Lemma 2.2. For $\mathcal{B}_\alpha \in \{\mathcal{B}_\alpha^c, \widetilde{\mathcal{B}}_\alpha\}$ denote $\mathcal{B}_\alpha = \bigcup_{x \in \mathbb{R}^d} \mathcal{B}_\alpha(x)$. For $\beta \in \mathbb{R}$ with $-1 \leq \alpha + \beta < d$ this allows us to define the following maximal functions

$$M_{\alpha,\beta}^c f(x) = \sup_{B \in \mathcal{B}_\alpha^c : x \in \overline{B}} r(B)^{\alpha+\beta} f_B,$$

$$\widetilde{M}_{\alpha,\beta} f(x) = \sup_{B \in \widetilde{\mathcal{B}}_\alpha : x \in \overline{B}} r(B)^{\alpha+\beta} f_B$$

for almost every $x \in \mathbb{R}^d$. Note that also for the centered version the supremum is all balls $B \in \mathcal{B}_\alpha^c$ whose closure contains x, not only over those centered in x.

Theorem 1.2 *Let $1 \le p < \infty$ and $0 < \alpha < d$ and $\beta \in \mathbb{R}$ with $0 \le \alpha + \beta + 1 < d/p$ and $M_{\alpha,\beta} \in \{M_{\alpha,\beta}^c, \widetilde{M}_{\alpha,\beta}\}$. Then for all $f \in W^{1,p}(\mathbb{R}^d)$ we have*

$$\|M_{\alpha,\beta} f\|_{(p^{-1}-(1+\alpha+\beta)/d)^{-1}} \le C_{d,\alpha,\beta,p} \|\nabla f\|_p$$

where the constant $C_{d,\alpha,\beta,p}$ depends only on d, α, β and p. In the endpoint $p = 1$ we can replace $f \in W^{1,1}(\mathbb{R}^d)$ by $f \in BV(\mathbb{R}^d)$. The endpoint result for $p = d/(1 + \alpha + \beta)$ holds true as well.

We prove Theorem 1.2 in Sect. 4. There had also been progress on $0 < \alpha \le 1$ similarly as for the Hardy–Littlewood maximal operator. For the uncentered fractional maximal function Carneiro and Madrid proved Theorem 1.1 for $d = 1$ in [10], and Luiro proved Theorem 1.1 for radial functions in [25]. Beltran and Madrid transferred Luiros result to the centered fractional maximal function in [5]. In [6] Beltran et al. proved Theorem 1.1 for $d \ge 2$ and a centered maximal operator that only uses balls with lacunary radius and for maximal operators with respect to smooth kernels. The next step after boundedness is continuity of the gradient of the fractional maximal operator, as it implies boundedness, but doesn't follow from it. In [4, 26] Beltran and Madrid already proved it for the uncentered fractional maximal operator in the cases where the boundedness is known.

For a dyadic cube Q we denote by $l(Q)$ the sidelength of Q. The fractional dyadic maximal function is defined by

$$M_\alpha^d f(x) = \sup_{Q:Q \ni x} l(Q)^\alpha f_Q,$$

where the supremum is taken over all dyadic cubes that contain x. The dyadic maximal operator has enjoyed a bit less attention than its continuous counterparts, such as the centered and the uncentered Hardy–Littlewood maximal operator. The dyadic maximal operator is different in the sense that formula (1.2) only holds for $\alpha = 0$, $p = 1$ and only in the variation sense, for which formula (1.2) has been proved in [29]. But for any other α and p formula (1.2) fails because $\nabla M_\alpha^d f$ is not a Sobolev function. We can however prove Theorem 1.4, the dyadic analog of Theorem 1.2. For $\alpha \ge 0$ and a function $f \in L^1(\mathbb{R}^d)$ define $\mathcal{Q}_\alpha$ to be the set of all cubes Q such that for all dyadic cubes $P \supsetneq Q$ we have $l(P)^\alpha f_P < l(Q)^\alpha f_Q$.

Remark 1.3 In the uncentered setting one could also define $\mathcal{B}_\alpha$ in a similar way as $\mathcal{Q}_\alpha$.

For $\beta \in \mathbb{R}$ with $-1 \le \alpha + \beta < d$ also define in the dyadic setting

$$M_{\alpha,\beta}^d f(x) = \sup_{Q \in \mathcal{Q}_\alpha : x \in \overline{Q}} l(Q)^{\alpha+\beta} f_Q.$$

Then

Theorem 1.4 *Let $1 \le p < \infty$ and $0 < \alpha < d$ and $\beta \in \mathbb{R}$ with $0 \le \alpha + \beta + 1 < d/p$. Then for all $f \in W^{1,p}(\mathbb{R}^d)$ we have*

$$\|M_{\alpha,\beta}^d f\|_{(p^{-1}-(1+\alpha+\beta)/d)^{-1}} \le C_{d,\alpha,\beta,p} \|\nabla f\|_p$$

where the constant $C_{d,\alpha,\beta,p}$ depends only on d, α, β and p. In the endpoint $p = 1$ we can replace $f \in W^{1,1}(\mathbb{R}^d)$ by $f \in BV(\mathbb{R}^d)$. The endpoint result for $p = d/(1 + \alpha + \beta)$ holds true as well.

Our main result in the dyadic setting is the following.

Theorem 1.5 *Let $1 \leq p < \infty$ and $0 < \alpha < d$. Then for all $f \in W^{1,p}(\mathbb{R}^d)$ we have*

$$\left(\sum_{Q \in \mathcal{Q}_\alpha} (l(Q)^{\frac{d}{p}-1} f_Q)^p \right)^{\frac{1}{p}} \leq C_{d,\alpha,p} \|\nabla f\|_p$$

where the constant $C_{d,\alpha,p}$ depends only on d, α and p. In the endpoint $p = 1$ we can replace $f \in W^{1,1}(\mathbb{R}^d)$ by $f \in \mathrm{BV}(\mathbb{R}^d)$. The endpoint result for $p = \infty$ holds true as well.

Remark 1.6 Note that in Theorem 1.5 we restrict $0 < \alpha < d$ and not $0 < \alpha < d/p$.

In Sect. 2.2 we conclude Theorem 1.4 from Theorem 1.5, and in Sect. 3 we prove Theorem 1.5.

Remark 1.7 Theorem 1.5 fails for $\alpha = 0$. However for $\alpha = 0$ and $p = 1$, a version with f_Q by replaced by $f_Q - \lambda_Q$ holds for certain λ_Q, see [29,Proposition 2.5].

Remark 1.8 For centered, uncentered maximal operator and dyadic maximal operator, Theorems 1.2, 1.4 and 1.5 admit localized versions of the following form. For $D \subset \mathbb{R}^d$ we set $\mathcal{B}_\alpha(D) = \bigcup_{x \in D} \mathcal{B}_\alpha(x)$ and $E = \bigcup \{cB : B \in \mathcal{B}_\alpha(D)\}$ with some large $c > 1$. Then Theorem 1.2 also holds in the form

$$\|\nabla M_{\alpha,-1} f\|_{L^{(p^{-1}-\alpha/d)^{-1}}(D)} \leq C_{d,\alpha,p} \|\nabla f\|_{L^p(E)}.$$

Theorem 1.4 holds with the dyadic version of E and Theorem 1.5 where the sum on the left hand side is over any subset $\mathcal{Q} \subset \mathcal{Q}_\alpha$ and the integral on the right is over $\bigcup \{cQ : Q \in \mathcal{Q}\}$. These localized results directly follow from the same proof as the global results, if one keeps track of the balls and cubes which are being dealt with. The respective localized version of Theorem 1.1 can be proven if one has Lemma 2.4 without the differentiability assumption. Then in the reduction of Theorem 1.1 to Theorem 1.2 one could apply Theorem 1.2 to the same function f and $\mathcal{Q}_\alpha$ for which one is showing Theorem 1.1, bypassing the approximation step and therefore preserving the locality of Theorem 1.2. This is in contrast to the actual local fractional maximal operator, for whom Theorem 1.1 fails by [17,Example 4.2], which works for $\alpha > 0$. However if $\alpha = 0$ and $p > 1$ then the local fractional maximal operator is again bounded due to [19], and by [30] for $\alpha = 0$ and $p = 1$ and characteristic functions.

Dyadic cubes are much easier to deal with than balls, but the dyadic version still serves as a model case for the continuous versions since both versions share many properties. This can be observed in [30], where we proved $\mathrm{var}\, M_0 1_E \leq C_d \,\mathrm{var}\, 1_E$ for the dyadic maximal operator and the uncentered Hardy–Littlewood maximal operator. The proof for the dyadic maximal operator is much shorter, but the same proof idea also works for the uncentered maximal operator. Also in this paper a part of the proof of Theorem 1.4 for the dyadic maximal operator is used also in the proof of Theorem 1.2 for the Hardy–Littlewood maximal operator.

The plan for the proof of Theorem 1.1 is the following. For simplicity we write it down for $p = 1$.

$$\int |\nabla M_\alpha f|^{\frac{d}{d-\alpha}} \leq (d-\alpha)^{\frac{d}{d-\alpha}} \int (M_{\alpha,-1} f)^{\frac{d}{d-\alpha}}$$

$$= d(d-\alpha)^{\frac{\alpha}{d-\alpha}} \int_0^\infty \lambda^{\frac{\alpha}{d-\alpha}} \mathcal{L}(\{M_{\alpha,-1} f > \lambda\}) \, d\lambda$$

$$= d(d-\alpha)^{\frac{\alpha}{d-\alpha}} \int_0^\infty \lambda^{\frac{\alpha}{d-\alpha}} \mathcal{L}\left(\bigcup\{\overline{B} : B \in \mathcal{B}_\alpha, r(B)^{\alpha-1} f_B > \lambda\}\right) d\lambda$$

$$\lesssim_\alpha \int_0^\infty \lambda^{\frac{\alpha}{d-\alpha}} \sum_{B \in \tilde{\mathcal{B}}_\alpha, cr(B)^{\alpha-1} f_B > \lambda} \mathcal{L}(B) \, d\lambda$$

$$= \sum_{B \in \tilde{\mathcal{B}}_\alpha} \int_0^{cr(B)^{\alpha-1} f_B} \lambda^{\frac{\alpha}{d-\alpha}} \, d\lambda$$

$$= \frac{(1-\alpha/d)c^{\frac{d}{d-\alpha}}}{(d\sigma_d)^{\frac{d}{d-\alpha}}} \sum_{B \in \tilde{\mathcal{B}}_\alpha} (f_B \mathcal{H}^{d-1}(\partial B))^{\frac{d}{d-\alpha}}$$

$$\leq \frac{(1-\alpha/d)c^{\frac{d}{d-\alpha}}}{(d\sigma_d)^{\frac{d}{d-\alpha}}} \left(\sum_{B \in \tilde{\mathcal{B}}_\alpha} f_B \mathcal{H}^{d-1}(\partial B)\right)^{\frac{d}{d-\alpha}}$$

$$\lesssim_\alpha \left(\sum_{Q \in \tilde{\mathcal{Q}}_\alpha} f_Q \mathcal{H}^{d-1}(\partial Q)\right)^{\frac{d}{d-\alpha}}$$

$$\leq C_{d,\alpha}(\operatorname{var} f)^{\frac{d}{d-\alpha}},$$

where σ_d is the volume of the d-dimensional unit ball. In the second step we apply the layer cake formula, in the forth step we pass from a union of arbitrary balls to very disjoint balls $\tilde{\mathcal{B}}_\alpha$ with a Vitali covering argument, in the eighth step we pass from those balls to comparable dyadic cubes and as the last step use a result from the dyadic setting.

We use $\alpha > 0$ as follows. Let A be a ball and $B(x, r)$ be a smaller ball that intersects A. Then by $A \subset B(x, 3r(A))$ we have $3^{\alpha-d} r(A)^\alpha f_A \leq (3r(A))^\alpha f_{B(x,3r(A))}$. Thus if $r^\alpha f_{B(x,r)} \leq 3^{\alpha-d} r(A)^\alpha f_A$ then $B(x, r)$ is not used by the fractional maximal operator. Hence it suffices to consider balls B with $3^{d-\alpha}(r(B)/r(A))^\alpha f_B > f_A$. From that we can conclude $f_B > 2f_A$ or $r(B) \gtrsim_\alpha r(A)$. Thus for any two balls B, A used by the fractional maximal operator, one of the following alternatives applies.

(1) The balls B and A are disjoint.
(2) The intervals $(f_B/2, f_B)$ and $(f_A/2, f_A)$ are disjoint.
(3) The radii $r(B)$ and $r(A)$ are comparable.

We use this in the forth step of the proof strategy above. We use a dyadic version of these alternatives in last step. Note that for $\alpha = 0$ optimal balls B of arbitrarily different sizes with similar values f_B can intersect.

Remark 1.9 There is a proof of Theorem 1.1 which has a structure parallel to the one presented above, but three steps are replaced. The estimate $|\nabla M_\alpha f|^{\frac{d}{d-\alpha}} \leq (d-\alpha)^{\frac{d}{d-\alpha}} M_{\alpha,-1} f$ is replaced by $|\nabla M_\alpha f|^{\frac{d}{d-\alpha}} \leq (d-\alpha)^{\frac{\alpha}{d-\alpha}} |\nabla M_\alpha f|(M_{\alpha,-1} f)^{\frac{\alpha}{d-\alpha}}$, the layer cake formula is replaced by the coarea formula [13,Theorem 3.11] and the Vitali covering argument is replaced by [30,Lemma 4.1] which deals with the boundary of balls instead of their volume. Otherwise it is identical to the proof presented in this paper.

$$\int |\nabla M_\alpha f|^{\frac{d}{d-\alpha}} \leq (d-\alpha)^{\frac{\alpha}{d-\alpha}} \int |\nabla M_\alpha f|(M_{\alpha,-1} f)^{\frac{\alpha}{d-\alpha}}$$

$$= (d-\alpha)^{\frac{\alpha}{d-\alpha}} \int_0^\infty \int_{\partial_*\{M_\alpha f > \lambda\}} (M_{\alpha,-1} f)^{\frac{\alpha}{d-\alpha}} \, d\lambda$$

$$= (d-\alpha)^{\frac{\alpha}{d-\alpha}} \int_0^\infty \int_{\partial_* \bigcup \{\overline{B}: B \in \mathcal{B}_\alpha, r(B)^\alpha f_B > \lambda\}} (r(B_x)^{\alpha-1} f_{B_x})^{\frac{\alpha}{d-\alpha}} \, d\mathcal{H}^{d-1}(x) \, d\lambda$$

$$\lesssim_\alpha \int_0^\infty \sum_{B \in \tilde{\mathcal{B}}_\alpha, r(B)^\alpha f_B > \lambda} \mathcal{H}^{d-1}(\partial B)(r(B)^{\alpha-1} f_B)^{\frac{\alpha}{d-\alpha}} \, d\lambda$$

$$\lesssim_\alpha \sum_{B \in \tilde{\mathcal{B}}_\alpha} (f_B \mathcal{H}^{d-1}(\partial B))^{\frac{d}{d-\alpha}}$$

and from there on arrive exactly as before at the bound by $(\text{var } f)^{\frac{d}{d-\alpha}}$. This motivates a similar replacement in the dyadic setting. Instead of proving the boundedness of $\|M_{\alpha,-1} f\|_{d/(d-\alpha)}$, Theorem 1.4, one might bound

$$\int_0^\infty \int_{\partial_* \{M_\alpha f > \lambda\}} (M_{\alpha,-1} f)^{\frac{\alpha}{d-\alpha}} \, d\lambda.$$

Note that formally

$$\int |\nabla M_\alpha f(x)| (M_{\alpha,-1} f(x))^{\frac{\alpha}{d-\alpha}} \, dx$$

is not well defined because $M_{\alpha,-1} f$ jumps where $\nabla M_\alpha f$ is supported.

Remark 1.10 In the proof of Theorems 1.1, 1.2, 1.5 and 1.4 we do not a priori need $f \in L^p(\mathbb{R}^d)$, it suffices to have $f \in L^q(\mathbb{R}^d)$ for some $1 \leq q \leq p$. However from $\|\nabla f\|_p < \infty$ we can then anyways conclude $f \in L^p(\mathbb{R}^d)$ by Sobolev embedding.

2 Reformulation

In order to avoid writing absolute values, we consider only nonnegative functions f for the rest of the paper. We can still conclude Theorems 1.1, 1.2, 1.4 and 1.5 for signed functions because $|f|_B = f_B$ and $\big|\nabla |f|(x)\big| \leq |\nabla f(x)|$. Recall the set of dyadic cubes

$$\bigcup_{n \in \mathbb{Z}} \Big\{ [x_1, x_1 + 2^n) \times \cdots \times [x_d, x_d + 2^n) : \forall i \in \{1, \ldots, n\} \, x_i \in 2^n \mathbb{Z} \Big\}.$$

For a set $\mathcal{B}$ of balls or dyadic cubes we denote

$$\bigcup \mathcal{B} = \bigcup_{B \in \mathcal{B}} B$$

as is commonly used in set theory. By $a \lesssim_{\gamma_1, \ldots, \gamma_n} b$ we mean that there exists a constant $C_{d, \gamma_1, \ldots, \gamma_n}$ that depends only on the values of $\gamma_1, \ldots, \gamma_n$ and the dimension d and such that $a \leq C_{d, \gamma_1, \ldots, \gamma_n} b$.

We work in the setting of functions of bounded variation, as in Evans–Gariepy [13,Section 5]. For an open set $\Omega \subset \mathbb{R}^d$ a function $u \in L^1_{loc}(\Omega)$ is said to have locally bounded variation if for each open and compactly supported $V \subset \Omega$ we have

$$\sup\Big\{ \int_V u \, \text{div} \, \varphi : \varphi \in C^1_c(V; \mathbb{R}^d), \, |\varphi| \leq 1 \Big\} < \infty.$$

Such a function comes with a measure μ and a function $\nu : \Omega \to \mathbb{R}^d$ that has $|\nu| = 1$ μ-a.e. such that for all $\varphi \in C_c^1(\Omega; \mathbb{R}^d)$ we have

$$\int u \operatorname{div} \varphi = \int \varphi \nu \, d\mu.$$

We denote $\nabla u = -\nu \mu$ and define the variation of u by

$$\operatorname{var}_\Omega u = \mu(\Omega) = \|\nabla u\|_{L^1(\Omega)}.$$

If ∇u is a locally integrable function we call u weakly differentiable.

Lemma 2.1 *Let $1 < p \leq \infty$ and $(u_n)_n$ be a sequence of locally integrable functions with*

$$\sup_n \|\nabla u_n\|_p < \infty$$

which converge to u in $L^1_{\text{loc}}(\mathbb{R}^d)$. Then u is weakly differentiable and

$$\|\nabla u\|_p \leq \limsup_n \|\nabla u_n\|_p.$$

Proof By the weak compactness of $L^p(\mathbb{R}^d)$ there is a subsequence, for simplicity also denoted by $(u_n)_n$, and a $v \in L^p(\mathbb{R}^d)^d$ such that $\nabla u_n \to v$ weakly in $L^p(\mathbb{R}^d)$ and $\|v\|_p \leq \limsup_n \|\nabla u_n\|_p$. Let $\varphi \in C_c^\infty(\mathbb{R}^d)$ and $i \in \{1, \ldots, d\}$. Then

$$\int u \partial_i \varphi = \lim_{n \to \infty} \int u_n \partial_i \varphi = -\lim_{n \to \infty} \int \partial_i u_n \varphi = -\int v_i \varphi$$

which means $\nabla u = v$. $\qquad \square$

2.1 Hardy–Littlewood maximal operator

In this section we reduce Theorem 1.1 to Theorem 1.2. Let $1 \leq p < d/\alpha$ and $f \in L^p(\mathbb{R}^d)$. For $x \in \mathbb{R}^d$ consider for the uncentered maximal operator the set of balls B with $x \in \overline{B}$ and $M_\alpha f(x) = r(B)^\alpha f_B$, and for the centered maximal operator such balls B which are centered in x. Recall that we denote by $\mathcal{B}_\alpha(x)$ the subset of those balls that have the largest radius.

Lemma 2.2 *Let $M_\alpha \in \{M_\alpha^c, \widetilde{M}_\alpha\}$ and $1 \leq p < d/\alpha$. Let $f \in L^p(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ be a Lebesgue point of f. Then $\mathcal{B}_\alpha(x)$ is nonempty.*

Proof We formulate one proof that works both for the centered and uncentered fractional maximal operator. Let $(B_n)_n$ a sequence of balls with $x \in B_n$ and

$$M_\alpha f(x) = \lim_{n \to \infty} r(B_n)^\alpha f_{B_n}.$$

Assume there is a subsequence $(n_k)_k$ with $r(B_{n_k}) \to 0$. Then $f_{B_{n_k}} \to f(x)$ and thus

$$\limsup_{k \to \infty} r(B_{n_k})^\alpha f_{B_{n_k}} \leq f(x) \limsup_{n \to \infty} r(B_{n_k})^\alpha = 0,$$

a contradiction. Assume there is a subsequence $(n_k)_k$ with $r(B_{n_k}) \to \infty$. Then

$$\limsup_{k \to \infty} r(B_{n_k})^\alpha f_{B_{n_k}} \leq \limsup_{k \to \infty} r(B_{n_k})^\alpha \mathcal{L}(B_{n_k})^{-1} \mathcal{L}(B_{n_k})^{1 - \frac{1}{p}} \left(\int_{B_{n_k}} f^p \right)^{\frac{1}{p}}$$

$$= \limsup_{k \to \infty} \sigma_d^{-\frac{1}{p}} r(B_{n_k})^{\alpha - \frac{d}{p}} \left(\int_{B_{n_k}} f^p \right)^{\frac{1}{p}}$$

$$\leq \sigma_d^{-\frac{1}{p}} \limsup_{k \to \infty} r(B_{n_k})^{\alpha - \frac{d}{p}} \|f\|_p = 0$$

since $\|f\|_p < \infty$, a contradiction. Hence there is a subsequence $(n_k)_k$ such that $r(B_{n_k})$ converges to some value $r \in (0, \infty)$. We can conclude that there is a ball B with $x \in \overline{B}$ and $r(B) = r$ and $\int_{B_{n_k}} f \to \int_B f$. So we have

$$M_\alpha f(x) = \lim_{k \to \infty} r(B_{n_k})^\alpha f_{B_{n_k}} = r(B)^\alpha f_B.$$

A similar argument shows that there exist a largest ball B for which $\sup_{\overline{B} \ni x} r(B)^\alpha f_B$ is attained. □

Lemma 2.3 *Let* $M_\alpha \in \{M_\alpha^c, \widetilde{M}_\alpha\}$. *and* $f \in L^\infty(\mathbb{R}^d)$ *have bounded variation. Then* $M_\alpha f$ *is locally Lipschitz.*

Proof If $f = 0$ then the statement is obvious, so consider $f \neq 0$. Let B be a ball. Then there is a ball $A \supset B$ with $f_A > 0$. Define

$$r_0 = 2r(A) \left(\frac{f_A}{2^d \|f\|_\infty} \right)^{1/\alpha}$$

and let $x \in B$. Then $A \subset B(x, 2r(A))$ so that for $r < r_0$ we have

$$r^\alpha f_{B(x,r)} < (2r(A))^\alpha \frac{f_A}{2^d \|f\|_\infty} \|f\|_\infty \leq (2r(A))^\alpha f_{B(x,2r(A))}.$$

That means that on B the maximal function $M_\alpha f$ is the supremum over all functions $\sigma_d^{-1} r^{\alpha-d} f * 1_{B(z,r)}$ with $r \geq r_0$ and z such that $0 \in B(z, r)$ for the uncentered operator and $z = 0$ for the centered. Those convolutions are weakly differentiable with

$$\nabla(r^{\alpha-d} f * 1_{B(z,r)}) = r^{\alpha-d} (\nabla f) * 1_{B(z,r)}$$

so that

$$|\nabla(r^{\alpha-d} f * 1_{B(z,r)})| \leq r^{\alpha-d} \operatorname{var} f \leq r_0^{\alpha-d} \operatorname{var} f.$$

Thus on B the maximal function $M_\alpha f$ is a supremum of functions with Lipschitz constant $\sigma_d^{-1} r_0^{\alpha-d} \operatorname{var} f$ and hence itself Lipschitz with the same constant. □

The following has essentially already been observed in [17, 20, 23, 25].

Lemma 2.4 *Let* $M_\alpha \in \{M_\alpha^c, \widetilde{M}_\alpha\}$ *and let* $M_\alpha f$ *be differentiable in* x. *Then for every* $B \in \mathcal{B}_\alpha(x)$ *we have*

$$|\nabla M_\alpha f(x)| \leq (d - \alpha) r(B)^{\alpha-1} f_B.$$

In the uncentered case if $x \in B$ *we have* $\nabla \widetilde{M}_\alpha f(x) = 0$.

Proof Let $B(z, r) \in \mathcal{B}_\alpha(x)$ and let e be a unit vector. Note that for the centered maximal operator we have $z = x$. Then for all $h > 0$ we have $x + he \in \overline{B(z, r + h)}$. Thus

$$|\nabla M_\alpha f(x)| = \sup_e \lim_{h \to 0} \frac{M_\alpha f(x) - M_\alpha f(x + he)}{h}$$

$$\leq \frac{1}{\sigma_d} \lim_{h \to 0} \frac{1}{h} (r^{\alpha-d} \int_{B(z,r)} f - (r+h)^{\alpha-d} \int_{B(z+eh,r+h)} f)$$

$$\leq \frac{1}{\sigma_d} \lim_{h \to 0} \frac{1}{h} (r^{\alpha-d} \int_{B(z,r)} f - (r+h)^{\alpha-d} \int_{B(z,r)} f)$$

$$= \frac{1}{\sigma_d} \lim_{h \to 0} \frac{1}{h} (r^{\alpha-d} - (r+h)^{\alpha-d}) \int_{B(z,r)} f$$

$$= \frac{1}{\sigma_d} (d-\alpha) r^{\alpha-d-1} \int_{B(z,r)} f.$$

If $x \in B(z,r)$ then since for all $y \in B(z,r)$ we have $M_\alpha f(y) \geq M_\alpha f(x)$ we get $\nabla M_\alpha f(x) = 0$. □

Now we reduce Theorem 1.1 to Theorem 1.2. We prove Theorem 1.2 in Sect. 4.

Proof of Theorem 1.1 For each $n \in \mathbb{N}$ define a cutoff function φ_n by

$$\varphi_n(x) = \begin{cases} 1, & 0 \leq |x| \leq 2^n, \\ 2 - 2^{-n}|x|, & 2^n \leq |x| \leq 2^{n+1}, \\ 0, & 2^{n+1} \leq |x| < \infty. \end{cases}$$

Then $|\nabla \varphi_n(x)| = 2^{-n} 1_{2^n \leq |x| \leq 2^{n+1}}$ and thus

$$\|f \nabla \varphi_n\|_p = 2^{-n} \|f\|_{L^p(B(0,2^{n+1}) \setminus B(0,2^n))} \to 0 \tag{2.1}$$

for $n \to \infty$. Denote $f_n(x) = \min\{f(x), n\} \cdot \varphi_n(x)$. Then by formula (2.1) we have

$$\lim_{n \to \infty} \|\nabla f_n\|_p = \lim_{n \to \infty} \|\nabla f_n - \min\{f, n\} \nabla \varphi_n\|_p = \lim_{n \to \infty} \|\varphi_n \nabla \min\{f, n\}\|_p = \|\nabla f\|_p. \tag{2.2}$$

Since $1 \leq p < d/\alpha$ and $f \in L^p(\mathbb{R}^d)$ we have $M_\alpha f \in L^{(p^{-1}-\alpha/d)^{-1},\infty}(\mathbb{R}^d) \subset L^1_{\text{loc}}(\mathbb{R}^d)$. Then since $M_\alpha f_n \to M_\alpha f$ pointwise from below, $M_\alpha f_n$ converges to $M_\alpha f$ in $L^1_{\text{loc}}(\mathbb{R}^d)$. So from Lemma 2.1 it follows that

$$\|\nabla M_\alpha f\|_{(p^{-1}-\alpha/d)^{-1}} \leq \limsup_{n \to \infty} \|\nabla M_\alpha f_n\|_{(p^{-1}-\alpha/d)^{-1}}.$$

By Lemma 2.3 we have that $M_\alpha f_n$ is weakly differentiable and differentiable almost everywhere, so that by Lemmas 2.2, 2.4 and Theorem 1.2 we have

$$\int |\nabla M_\alpha f_n|^{(p^{-1}-\alpha/d)^{-1}} \leq (d-\alpha) \|M_\alpha f_n / r(B_x)\|_{(p^{-1}-\alpha/d)^{-1}}$$

$$\leq (d-\alpha) \|M_{\alpha,-1} f_n\|_{(p^{-1}-\alpha/d)^{-1}}$$

$$\lesssim_\alpha \|\nabla f_n\|_p,$$

which by formula (2.2) converges to $\|\nabla f\|_p$. for $n \to \infty$. For the endpoint $p = d/\alpha$ the proof works the same. □

2.2 Dyadic maximal operator

In this section we reduce Theorem 1.4 to Theorem 1.5. Let $1 \leq p < d/\alpha$ and $f \in L^p(\mathbb{R}^d)$. Recall that we denote by $\mathcal{Q}_\alpha$ the set of all dyadic cubes Q such that for every dyadic cube ball $P \supsetneq Q$ we have $l(P)^\alpha f_P < l(Q)^\alpha f_Q$. For $x \in \mathbb{R}^d$, we denote by $\mathcal{Q}_\alpha(x)$ the set of dyadic cubes Q with $x \in Q$ and

$$M_\alpha^d f(x) = l(Q)^\alpha f_Q.$$

Lemma 2.5 *Let $1 \leq p < d/\alpha$ and $f \in L^p(\mathbb{R}^d)$ and $x \in \mathbb{R}^d$ be a Lebesgue point of f. Then $\mathcal{Q}_\alpha(x)$ contains a dyadic cube Q_x with*

$$\mathrm{l}(Q_x) = \sup_{Q \in \mathcal{Q}_\alpha(x)} \mathrm{l}(Q)$$

and that cube also belongs to $\mathcal{Q}_\alpha$.

Proof Let $(Q_n)_n$ be a sequence of cubes with $\mathrm{l}(Q_n) \to \infty$. Then

$$\limsup_{n \to \infty} \mathrm{l}(Q_n)^\alpha f_{Q_n} \leq \limsup_{n \to \infty} \mathrm{l}(Q_n)^{\alpha-d} \mathcal{L}(Q_n)^{1-\frac{1}{p}} \left(\int_{Q_n} f^p \right)^{\frac{1}{p}}$$

$$= \limsup_{n \to \infty} \mathrm{l}(Q_n)^{\alpha-d+d-\frac{d}{p}} \left(\int_{Q_n} f^p \right)^{\frac{1}{p}}$$

$$= \limsup_{n \to \infty} \mathrm{l}(Q_n)^{\alpha-\frac{d}{p}} \left(\int_{Q_n} f^p \right)^{\frac{1}{p}}$$

$$\leq \limsup_{n \to \infty} \mathrm{l}(Q_n)^{\alpha-\frac{d}{p}} \|f\|_p = 0.$$

Let $(Q_n)_n$ be a sequence of cubes with $\mathrm{l}(Q_n) \to 0$. Then since $f_{Q_n} \to f(x)$ and $\mathrm{l}(Q_n)^\alpha \to 0$, we have $\mathrm{l}(Q_n)^\alpha f_Q \to 0$. Thus since for each k there are at most 2^d many cubes Q with $\mathrm{l}(Q) = 2^k$ and whose closure contains x, the supremum has to be attained for a finite set of cubes from which we can select the largest. □

Now we reduce Theorem 1.4 to Theorem 1.5. We prove Theorem 1.5 in Sect. 3.

Proof of Theorem 1.4 By Lemma 2.5, $\mathrm{M}_{\alpha,\beta}^{\mathrm{d}} f$ is defined almost everywhere. We have

$$\int (\mathrm{M}_{\alpha,\beta}^{\mathrm{d}} f(x))^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}} \, \mathrm{d}x \leq \int \sum_{Q \in \mathcal{Q}_\alpha} 1_Q(x) (\mathrm{l}(Q)^{\alpha+\beta} f_Q)^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}} \, \mathrm{d}x$$

$$= \sum_{Q \in \mathcal{Q}_\alpha} \mathcal{L}(Q) (\mathrm{l}(Q)^{\alpha+\beta} f_Q)^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}}$$

$$= \sum_{Q \in \mathcal{Q}_\alpha} (\mathrm{l}(Q)^{d/p-1} f_Q)^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}}$$

$$\leq \left(\sum_{Q \in \mathcal{Q}_\alpha} (\mathrm{l}(Q)^{d/p-1} f_Q)^p \right)^{(1-p(1+\alpha+\beta)/d)^{-1}}$$

$$\lesssim_\alpha \|\nabla f\|_p^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}},$$

where the last step follows from Theorem 1.5. In the endpoint case we have by Theorem 1.5

$$\|\mathrm{M}_{\alpha,\beta}^{\mathrm{d}} f\|_\infty = \sup_{Q \in \mathcal{Q}_\alpha} \mathrm{l}(Q)^{\alpha+\beta} f_Q$$

$$= \sup_{Q \in \mathcal{Q}_\alpha} \mathrm{l}(Q)^{\frac{d}{p}-1} f_Q \leq \left(\sum_{Q \in \mathcal{Q}_\alpha} (\mathrm{l}(Q)^{\frac{d}{p}-1} f_Q)^p \right)^{\frac{1}{p}} \lesssim_p \|\nabla f\|_p.$$

□

3 Dyadic maximal operator

In this section we prove Theorem 1.5. For a measurable set $E \subset \mathbb{R}^d$ we define the measure theoretic boundary by

$$\partial_* E = \left\{ x : \limsup_{r \to 0} \frac{\mathcal{L}(B(x,r) \setminus E)}{r^d} > 0, \ \limsup_{r \to 0} \frac{\mathcal{L}(B(x,r) \cap E)}{r^d} > 0 \right\}.$$

We denote the topological boundary by ∂E. As in [29, 30], our approach to the variation is the coarea formula rather then the definition of the variation, see for example [13, Theorem 5.9].

Lemma 3.1 *Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ with locally bounded variation and $U \subset \mathbb{R}^d$. Then*

$$\mathrm{var}_U \ f = \int_{\mathbb{R}} \mathcal{H}^{d-1}(\partial_* \{f > \lambda\} \cap U) \, \mathrm{d}\lambda.$$

Lemma 3.2 *Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ be weakly differentiable and $U \subset \mathbb{R}^d$ and $\lambda_0 < \lambda_1$. Then*

$$\int_{\{x \in U : \lambda_0 < f(x) < \lambda_1\}} |\nabla f| = \int_{\lambda_0}^{\lambda_1} \mathcal{H}^{d-1}(\partial_* \{f > \lambda\} \cap U) \, \mathrm{d}\lambda.$$

Recall also the relative isoperimetric inequality for cubes.

Lemma 3.3 *Let Q be a cube and E be a measurable set. Then*

$$\min\{\mathcal{L}(Q \cap E), \mathcal{L}(Q \setminus E)\}^{d-1} \lesssim \mathcal{H}^{d-1}(\partial_* E \cap Q)^d.$$

We will use a result from the case $\alpha = 0$. For a subset $\mathcal{Q} \subset \mathcal{Q}_0$ and $Q \in \mathcal{Q}_0$, we denote

$$\lambda_Q^{\mathcal{Q}} = \min\left\{ \max\left\{ \inf\{\lambda : \mathcal{L}(\{f > \lambda\} \cap Q) < 2^{-d-2}\mathcal{L}(Q)\}, \ \sup\{f_P : P \in \mathcal{Q}, \ P \supsetneq Q\} \right\}, \ f_Q \right\}.$$

Proposition 3.4 *Let $1 \le p < \infty$ and $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$ and $|\nabla f| \in L^p(\mathbb{R}^d)$. Then for every set $\mathcal{Q} \subset \mathcal{Q}_0$ we have*

$$\sum_{Q \in \mathcal{Q}} (\mathrm{l}(Q)^{\frac{d}{p}-1}(f_Q - \lambda_Q^{\mathcal{Q}}))^p \lesssim_p \|\nabla f\|_p^p.$$

For $p = 1$ it also holds with $\|\nabla f\|_1$ replaced by $\mathrm{var} \ f$.

Remark 3.5 We have that $\alpha < \beta$ implies $\mathcal{Q}_\beta \subset \mathcal{Q}_\alpha$. This is because for $\mathrm{l}(Q) < \mathrm{l}(P)$, $\mathrm{l}(Q)^\alpha f_Q > \mathrm{l}(P)^\alpha f_P$ becomes a stronger estimate the larger α becomes.

By Remark 3.5 we can apply Proposition 3.4 to $\mathcal{Q} = \mathcal{Q}_\alpha$. For $p = 1$ Proposition 3.4 is Proposition 2.5 in [29]. For the proof for all $p \ge 1$ we follow the strategy in [29]. In particular we use the following result. For $Q \in \mathcal{Q}_0$ we denote

$$\bar{\lambda}_Q = \min\left\{ \max\left\{ \inf\{\lambda : \mathcal{L}(\{f > \lambda\} \cap Q) < \mathcal{L}(Q)/2\}, \ \sup\{f_P : P \in \mathcal{Q}_0, \ P \supsetneq Q\} \right\}, \ f_Q \right\}.$$

Lemma 3.6 (Corollary 3.3 in [29]) *Let $f \in L^1_{\mathrm{loc}}(\mathbb{R}^d)$. Then for every $Q \in \mathcal{Q}_0$ we have*

$$\mathcal{L}(Q)(f_Q - \lambda_Q^\emptyset) \le 2^{d+2} \sum_{P \in \mathcal{Q}_0, P \subsetneq Q} \int_{\bar{\lambda}_P}^{f_P} \mathcal{L}(P \cap \{f > \lambda\}) \, \mathrm{d}\lambda.$$

Note that $f_P > \bar{\lambda}_P$ implies $P \in \mathcal{Q}_0$.

Proof of Proposition 3.4 By Lemmas 3.3, 3.2 we have for each $P \in \mathcal{Q}_0$ and $P \subsetneq Q$ that

$$\int_{\bar{\lambda}_P}^{f_P} \mathcal{L}(\{f > \lambda\} \cap P) \, d\lambda \leq 1(P) \int_{\bar{\lambda}_P}^{f_P} \mathcal{L}(\{f > \lambda\} \cap P)^{1-\frac{1}{d}} \, d\lambda$$

$$\lesssim 1(P) \int_{\bar{\lambda}_P}^{f_P} \mathcal{H}^{d-1}(\partial_* \{f > \lambda\} \cap P) \, d\lambda$$

$$= 1(P) \int_{x \in P : \bar{\lambda}_P < f(x) < f_P} |\nabla f|$$

$$= 1(P) \int_Q |\nabla f| 1_{P \times (\bar{\lambda}_P, f_P)}(x, f(x)) \, dx.$$

We note that for any $Q \in \mathcal{Q}$ we have $\lambda_Q^Q \geq \lambda_Q^\emptyset$ and use Lemma 3.6. Then we apply the above calculation, Hölder's inequality and use that $(\bar{\lambda}_P, f_P)$ and $(\bar{\lambda}_Q, f_Q)$ are disjoint for $P \subsetneq Q$,

$$\sum_{Q \in \mathcal{Q}} \left(1(Q)^{\frac{d}{p}-1}(f_Q - \lambda_Q^Q)\right)^p$$

$$\leq 2^{d+2} \sum_{Q \in \mathcal{Q}} \left(1(Q)^{\frac{d}{p}-1-d} \sum_{P \in \mathcal{Q}_0, P \subsetneq Q} \int_{\bar{\lambda}_P}^{f_P} \mathcal{L}(\{f > \lambda\} \cap P) \, d\lambda\right)^p$$

$$\lesssim \sum_{Q \in \mathcal{Q}} \left(1(Q)^{\frac{d}{p}-1-d} \int_Q |\nabla f| \sum_{P \in \mathcal{Q}_0, P \subsetneq Q} 1(P) 1_{P \times (\bar{\lambda}_P, f_P)}(x, f(x)) \, dx\right)^p$$

$$\leq \sum_{Q \in \mathcal{Q}} \left(1(Q)^{\frac{d}{p}-1-d+d(1-\frac{1}{p})} \left[\int_Q |\nabla f|^p \left(\sum_{P \in \mathcal{Q}_0, P \subsetneq Q} 1(P) 1_{P \times (\bar{\lambda}_P, f_P)}(x, f(x))\right)^p dx\right]^{\frac{1}{p}}\right)^p$$

$$= \sum_{Q \in \mathcal{Q}} \left(1(Q)^{-1} \left[\sum_{P \in \mathcal{Q}_0, P \subsetneq Q} 1(P)^p \int_{(x, f(x)) \in P \times (\bar{\lambda}_P, f_P)} |\nabla f|^p\right]^{\frac{1}{p}}\right)^p$$

$$= \sum_{Q \in \mathcal{Q}} 1(Q)^{-p} \sum_{P \in \mathcal{Q}_0, P \subsetneq Q} 1(P)^p \int_{(x, f(x)) \in P \times (\bar{\lambda}_P, f_P)} |\nabla f|^p$$

$$= \sum_{P \in \mathcal{Q}_0} 1(P)^p \int_{x \in P : f(x) \in (\bar{\lambda}_P, f_P)} |\nabla f|^p \sum_{Q \in \mathcal{Q}, Q \supsetneq P} 1(Q)^{-p}$$

$$\leq \frac{1}{2^p - 1} \sum_{P \in \mathcal{Q}_0} \int_{x \in P : f(x) \in (\bar{\lambda}_P, f_P)} |\nabla f|^p$$

$$\leq \frac{1}{2^p - 1} \int |\nabla f|^p.$$

For $p = 1$ with var f instead of $\|\nabla f\|_1$ we do not use Lemma 3.2 or Hölder's inequality, but interchange the order of summation first and then apply Lemma 3.1. □

For a dyadic cube Q denote by $\mathrm{prt}(Q)$ the dyadic parent cube of Q.

Lemma 3.7 *Let $1 \leq p < d/\alpha$ and $f \in L^p(\mathbb{R}^d)$ and let $\varepsilon > 0$. Then there is a subset $\tilde{\mathcal{Q}}_\alpha$ of $\mathcal{Q}_\alpha$ such that for each $Q \in \mathcal{Q}_\alpha$ with $1(Q)^\alpha f_Q > \varepsilon$ there is a $P \in \tilde{\mathcal{Q}}_\alpha$ with $Q \subset \mathrm{prt}(P)$ and $f_Q \leq 2^d f_P$. Furthermore for any two $Q, P \in \tilde{\mathcal{Q}}_\alpha$ one of the following holds.*

(1) $\mathrm{prt}(Q) = \mathrm{prt}(P)$.
(2) $\mathrm{prt}(Q)$ *and* $\mathrm{prt}(P)$ *don't intersect.*
(3) $f_Q/f_P \notin (2^{-d}, 2^d)$.

Proof Set $\tilde{\mathcal{Q}}_\alpha^0$ to be the set of maximal cubes Q with $\mathrm{l}(Q)^\alpha f_Q > \varepsilon$. For any dyadic cube Q with $\mathrm{l}(Q)^\alpha f_Q > \varepsilon$ we have

$$\varepsilon < \mathrm{l}(Q)^{\alpha-d} \int_Q f \le \mathrm{l}(Q)^{\alpha-d+d-\frac{d}{p}} \left(\int_Q f^p \right)^{\frac{1}{p}} \le \mathrm{l}(Q)^{\alpha-\frac{d}{p}} \|f\|_p$$

which implies

$$\mathrm{l}(Q) < (\|f\|_p/\varepsilon)^{(p^{-1}-\alpha/d)^{-1}}. \tag{3.1}$$

Hence

$$\bigcup \tilde{\mathcal{Q}}_\alpha^0 = \bigcup \{Q \in \mathcal{Q}_\alpha : \mathrm{l}(Q)^\alpha f_Q > \varepsilon\}.$$

Assume we have already defined $\tilde{\mathcal{Q}}_\alpha^n$. Then define $\tilde{\mathcal{Q}}_\alpha^{n+1}$ to be the set of maximal cubes $Q \in \mathcal{Q}_\alpha$ with

$$f_Q > 2^d \sup_{P \in \tilde{\mathcal{Q}}_\alpha^n : Q \subset \mathrm{prt}(P)} f_P. \tag{3.2}$$

Set $\tilde{\mathcal{Q}}_\alpha = \tilde{\mathcal{Q}}_\alpha^0 \cup \tilde{\mathcal{Q}}_\alpha^1 \cup \dots$.

Assume there is a cube Q with $\mathrm{l}(Q)^\alpha f_Q > \varepsilon$ such that for all $P \in \tilde{\mathcal{Q}}_\alpha$ with $Q \subset \mathrm{prt}(P)$ we have $f_Q > 2^d f_P$. Then by formula (3.1) there is a maximal such cube Q. Furthermore there is a smallest $P \in \tilde{\mathcal{Q}}_\alpha$ with $Q \subset \mathrm{prt}(P)$ and an n with $P \in \tilde{\mathcal{Q}}_\alpha^n$. But then Q is a maximal cube that satisfies formula (3.2), which implies $Q \in \tilde{\mathcal{Q}}_\alpha^{n+1}$, a contradiction.

If for $Q, P \in \tilde{\mathcal{Q}}_\alpha$ neither (1) nor (2) holds, then after renaming we have $\mathrm{prt}(Q) \subsetneq \mathrm{prt}(P)$. Then P has been added to $\tilde{\mathcal{Q}}_\alpha$ before Q, and since $Q \subset \mathrm{prt}(P)$ this means $f_Q > 2^d f_P$. $\square$

Lemma 3.8 *Let* $1 \le p < \infty$ *and* $f \in W^{1,p}(\mathbb{R}^d)$ *and let* $\varepsilon > 0$. *Let* $\mathcal{Q} \subset \mathcal{Q}_0$ *be a set of dyadic cubes such that*

(1) *for each* $Q \in \mathcal{Q}$ *there is an ancestor cube* $p(Q) \supsetneq Q$ *with* $\mathrm{l}(p(Q)) \le \mathrm{l}(Q)/\varepsilon$ *and* $f_Q > 2^\varepsilon f_{p(Q)}$,
(2) *and for any two distinct* $Q, P \in \mathcal{Q}$ *such that* $p(Q)$ *and* $p(P)$ *intersect we have* $f_Q/f_P \notin (2^{-\varepsilon}, 2^\varepsilon)$.

Then

$$\left(\sum_{Q \in \mathcal{Q}} (\mathrm{l}(Q)^{\frac{d}{p}-1} f_Q)^p \right)^{\frac{1}{p}} \lesssim_\varepsilon \|\nabla f\|_p.$$

The endpoint $p = \infty$ *holds as well.*

Proof We divide into two types of cubes and deal with them separately. Denote

$$\mathcal{Q}_- = \{Q \in \mathcal{Q} : \mathcal{L}(\{f > 2^{-\varepsilon/3} f_Q\} \cap Q) < 2^{-d-2} \mathcal{L}(Q)\},$$
$$\mathcal{Q}_+ = \{Q \in \mathcal{Q} : \mathcal{L}(\{f > 2^{-\varepsilon/3} f_Q\} \cap Q) \ge 2^{-d-2} \mathcal{L}(Q)\}.$$

Let $Q \in \mathcal{Q}_-$ and recall λ_Q^Q from Proposition 3.4. Then since

$$\sup\{\lambda : \mathcal{L}(\{f > \lambda\} \cap Q) < 2^{-d-2} \mathcal{L}(Q)\} \le 2^{-\varepsilon/3} f_Q,$$

$$\sup\{f_P : P \in \mathcal{Q},\ P \supsetneq Q\} \leq 2^{-\varepsilon} f_Q$$

we have

$$f_Q - \lambda_Q^{\mathcal{Q}} \geq (1 - 2^{-\varepsilon/3}) f_Q.$$

Since $\mathcal{Q} \subset \mathcal{Q}_0$ we conclude from Proposition 3.4

$$\sum_{Q \in \mathcal{Q}_-} \left(l(Q)^{\frac{d}{p}-1} f_Q \right)^p \leq (1 - 2^{-\varepsilon/3})^{-p} \sum_{Q \in \mathcal{Q}_-} \left(l(Q)^{\frac{d}{p}-1} (f_Q - \lambda_Q^{\mathcal{Q}}) \right)^p \lesssim_{\varepsilon,p} \|\nabla f\|_p^p.$$

Let $Q \in \mathcal{Q}_+$ and $\lambda > 2^{-2\varepsilon/3} f_Q$. Since by (1) we have $2^{\varepsilon/3} f_{p(Q)} < 2^{-2\varepsilon/3} f_Q$, we obtain from Chebyshev's inequality

$$\mathcal{L}(p(Q) \cap \{f > \lambda\}) \leq 2^{-\varepsilon/3} \mathcal{L}(p(Q)). \tag{3.3}$$

Since $Q \in \mathcal{Q}_+$, for $\lambda < 2^{-\varepsilon/3} f_Q$ we have

$$2^{-d-2} \varepsilon^d \mathcal{L}(p(Q)) \leq 2^{-d-2} \mathcal{L}(Q) \leq \mathcal{L}(Q \cap \{f > \lambda\}) \leq \mathcal{L}(p(Q) \cap \{f > \lambda\}). \tag{3.4}$$

So for all $2^{-2\varepsilon/3} f_Q \leq \lambda \leq 2^{-\varepsilon/3} f_Q$ we can conclude by the isoperimetric inequality Lemma 3.3 and formulas (3.3) and (3.4) that

$$\begin{aligned}
\mathcal{H}^{d-1}(\partial_* \{f > \lambda\} \cap p(Q))^d &\gtrsim \min\{\mathcal{L}(p(Q) \cap \{f > \lambda\}), \mathcal{L}(p(Q) \backslash \{f > \lambda\})\}^{d-1} \\
&\geq (\mathcal{L}(p(Q)) \min\{\varepsilon^d 2^{-d-2}, 1 - 2^{-\varepsilon/3}\})^{d-1} \\
&\gtrsim_\varepsilon \mathcal{L}(p(Q))^{d-1}.
\end{aligned}$$

Thus for each $Q \in \mathcal{Q}_+$ by Lemma 3.2 and Hölder's inequality we have

$$\begin{aligned}
\int_{2^{-2\varepsilon/3} f_Q}^{2^{-\varepsilon/3} f_Q} l(p(Q))^{d-1} \, d\lambda &\lesssim_\varepsilon \int_{2^{-2\varepsilon/3} f_Q}^{2^{-\varepsilon/3} f_Q} \mathcal{H}^{d-1}(\partial_* \{f > \lambda\} \cap p(Q)) \, d\lambda \\
&= \int_{x \in p(Q): f(x) \in (2^{-2\varepsilon/3}, 2^{-\varepsilon/3}) f_Q} |\nabla f| \\
&\leq l(p(Q))^{d-\frac{d}{p}} \left(\int_{x \in p(Q): f(x) \in (2^{-2\varepsilon/3}, 2^{-\varepsilon/3}) f_Q} |\nabla f|^p \right)^{\frac{1}{p}}.
\end{aligned}$$

Now we use (2) and conclude

$$\begin{aligned}
\sum_{Q \in \mathcal{Q}_+} \left(l(Q)^{\frac{d}{p}-1} f_Q \right)^p &\lesssim_{\varepsilon,p} \sum_{Q \in \mathcal{Q}_+} \left(l(p(Q))^{\frac{d}{p}-1} f_{p(Q)} \right)^p \\
&\lesssim_{\varepsilon,p} \sum_{Q \in \mathcal{Q}_+} \left(l(p(Q))^{\frac{d}{p}-d} \int_{2^{-2\varepsilon/3} f_Q}^{2^{-\varepsilon/3} f_Q} l(p(Q))^{d-1} \, d\lambda \right)^p \\
&\lesssim_{\varepsilon,p} \sum_{Q \in \mathcal{Q}_+} \int_{x \in p(Q): f(x) \in (2^{-2\varepsilon/3}, 2^{-\varepsilon/3}) f_Q} |\nabla f|^p \\
&\leq \int |\nabla f|^p.
\end{aligned}$$

For $p = 1$ with var f instead of $\|\nabla f\|_1$ we use Lemma 3.1 instead of Lemma 3.2 and Hölder's inequality. For $p = \infty$ let $Q \in \mathcal{Q}$. Then by the Sobolev-Poincaré inequality we

have

$$\|\nabla f\|_\infty \geq \|\nabla f\|_{L^\infty(p(Q))} \gtrsim l(p(Q))^{-d-1} \int_{p(Q)} |f - f_{p(Q)}|$$

$$\geq l(Q)^{-d-1} \varepsilon^{d+1} \int_Q |f - f_{p(Q)}|$$

$$\geq l(Q)^{-d-1} \varepsilon^{d+1} \int_Q f - f_{p(Q)}$$

$$= l(Q)^{-1} \varepsilon^{d+1} (f_Q - f_{p(Q)})$$

$$\geq l(Q)^{-1} \varepsilon^{d+1} (1 - 2^{-\varepsilon}) f_Q.$$

$\square$

Proof of Theorem 1.5 Let $\varepsilon > 0$ and $\tilde{\mathcal{Q}}_\alpha$ be the set of cubes from Lemma 3.7. Let $Q \in \mathcal{Q}_\alpha$. Then there is a $P \in \tilde{\mathcal{Q}}_\alpha$ with $Q \subset \mathrm{prt}(P)$ and $f_Q \leq 2^d f_P$. Then $f_Q \leq 4^d f_{\mathrm{prt}(P)}$. Thus since $l(Q)^\alpha f_Q > l(\mathrm{prt}(P))^\alpha f_{\mathrm{prt}(P)}$ we have $l(Q) > 4^{-d/\alpha} l(\mathrm{prt}(P))$. Thus for each P there are at most c_α many $Q \in \mathcal{Q}_\alpha$ with $Q \subset \mathrm{prt}(P)$ and $f_Q \leq 2^d f_P$. We conclude

$$\sum_{Q \in \mathcal{Q}_\alpha, l(Q)^\alpha f_Q > \varepsilon} \left(l(Q)^{\frac{d}{p}-1} f_Q \right)^p \leq \sum_{P \in \tilde{\mathcal{Q}}_\alpha} \sum_{Q \in \mathcal{Q}_\alpha, \ Q \subset \mathrm{prt}(P), \ f_Q \leq 2^d f_P} \left(l(Q)^{\frac{d}{p}-1} f_Q \right)^p$$

$$\lesssim_{\alpha, p} c_\alpha \sum_{P \in \tilde{\mathcal{Q}}_\alpha} \left(l(P)^{\frac{d}{p}-1} f_P \right)^p.$$

For each dyadic cube $P \in \{\mathrm{prt}(Q) : Q \in \tilde{\mathcal{Q}}_\alpha\}$ pick a $Q \in \tilde{\mathcal{Q}}_\alpha$ with $P = \mathrm{prt}(Q)$ such that for all $Q' \in \tilde{\mathcal{Q}}_\alpha$ with $P = \mathrm{prt}(Q')$ we have $f_{Q'} \leq f_Q$. Denote by $\hat{\mathcal{Q}}_\alpha$ the set of all such dyadic cubes Q. Then

$$\sum_{Q \in \tilde{\mathcal{Q}}_\alpha} \left(l(Q)^{\frac{d}{p}-1} f_Q \right)^p \leq \sum_{P \in \{\mathrm{prt}(Q): Q \in \tilde{\mathcal{Q}}_\alpha\}} \sum_{Q \in \tilde{\mathcal{Q}}_\alpha : P = \mathrm{prt}(Q)} \left(l(Q)^{\frac{d}{p}-1} f_Q \right)^p$$

$$\leq \sum_{P \in \{\mathrm{prt}(Q): Q \in \tilde{\mathcal{Q}}_\alpha\}} 2^d \sum_{Q \in \hat{\mathcal{Q}}_\alpha : P = \mathrm{prt}(Q)} \left(l(Q)^{\frac{d}{p}-1} f_Q \right)^p$$

$$= 2^d \sum_{Q \in \hat{\mathcal{Q}}_\alpha} \left(l(Q)^{\frac{d}{p}-1} f_Q \right)^p.$$

We want to show that Lemma 3.8 applies to $\hat{\mathcal{Q}}_\alpha$ with $p(Q) = \mathrm{prt}(Q)$. Since $\hat{\mathcal{Q}}_\alpha \subset \mathcal{Q}_\alpha$ we have $\hat{\mathcal{Q}}_\alpha \subset \mathcal{Q}_0$ by Remark 3.5, and (1) follows from $f_Q > 2^\alpha f_{\mathrm{prt}(Q)}$. For (2) let $Q, P \in \hat{\mathcal{Q}}_\alpha$ be distinct such that $\mathrm{prt}(Q)$ and $\mathrm{prt}(P)$ intersect. Since we have $\mathrm{prt}(Q) \neq \mathrm{prt}(P)$, Lemma 3.7 implies $f_Q/f_P \notin (2^{-d}, 2^d)$. Thus by Lemma 3.8 we have

$$2^d \sum_{Q \in \hat{\mathcal{Q}}_\alpha} \left(l(Q)^{\frac{d}{p}-1} f_Q \right)^p \lesssim_{\alpha, p} \|\nabla f\|_p^p.$$

We have proven for every $\varepsilon > 0$ that

$$\sum_{Q \in \mathcal{Q}_\alpha, l(Q)^\alpha f_Q > \varepsilon} \left(l(Q)^{\frac{d}{p}-1} f_Q \right)^p \lesssim_{\alpha, p} \|\nabla f\|_p^p$$

with constant independent of ε. So we can let ε go to zero and conclude Theorem 1.5.

For the endpoint $p = \infty$ let $Q \in \mathcal{Q}_\alpha$. Then we use $f_{\mathrm{prt}(Q)} \le 2^{-\alpha} f_Q$ and copy the proof of the endpoint in Lemma 3.8 with $p(Q) = \mathrm{prt}(Q)$ and $\varepsilon = 1/2$. □

4 Hardy–Littlewood maximal operator

In this section we prove Theorem 1.2.

4.1 Making the balls disjoint

Lemma 4.1 *Let* $\mathrm{M}_\alpha \in \{\mathrm{M}_\alpha^c, \widetilde{\mathrm{M}}_\alpha\}$ *and* $1 \le p < d/(1 + \alpha + \beta)$ *and* $f \in L^p(\mathbb{R}^d)$ *and let* $\varepsilon > 0$. *Then for any* $c_1 \ge 2, c_2 \ge 1$ *there is a set of balls* $\widetilde{\mathcal{B}} \subset \mathcal{B}_\alpha$ *such that for two balls* $B, A \in \widetilde{\mathcal{B}}$ *we have* $c_1 B \cap c_1 A = \emptyset$ *or* $f_A/f_B \notin (c_2^{-1}, c_2)$, *and furthermore*

$$\int_\varepsilon^\infty \lambda^{(p^{-1} - (1+\alpha+\beta)/d)^{-1} - 1} \mathcal{L}\Big(\bigcup\{B \in \mathcal{B}_\alpha : r(B)^{\alpha+\beta} f_B > \lambda\}\Big) \, d\lambda$$

$$\lesssim_{\alpha,\beta,p,c_1,c_2} \Big(\sum_{B \in \widetilde{\mathcal{B}}} \big(r(B)^{\frac{d}{p}-1} f_B\big)^p\Big)^{(1 - p(1+\alpha+\beta)/d)^{-1}}.$$

Proof Let $B \in \mathcal{B}_\alpha$ with $r(B)^{\alpha+\beta} f_B > \varepsilon$. Then

$$\varepsilon < r(B)^{\alpha+\beta} f_B \le r(B)^{\alpha+\beta} \mathcal{L}(B)^{-1} \mathcal{L}(B)^{1-1/p} \Big(\int_B f^p\Big)^{1/p} \le \sigma_d^{-1/p} r(B)^{\alpha+\beta-d/p} \|f\|_p,$$

which means that $r(B)$ is bounded by

$$K = (\sigma_d^{-1/p} \|f\|_p / \varepsilon)^{1/(d/p - \alpha - \beta)}.$$

Define $\mathcal{B}^0 = \{B \in \mathcal{B}_\alpha : r(B) \in [1/2, 1]K\}$. Then for all $B \in \mathcal{B}^0$ we have that $r(B)^\alpha f_B$ is uniformly bounded. Inductively define a sequence of balls as follows. For $B_0, \ldots, B_{k-1}$ already defined choose a ball $B_k \in \mathcal{B}^0$ such that $c_1 B_k$ is disjoint from $c_1 B_0, \ldots, c_1 B_{k-1}$ and which attains at least half of

$$\sup\{f_B : B \in \mathcal{B}^0, c_1 B \cap (c_1 B_0 \cup \ldots \cup c_1 B_{k-1}) = \emptyset\}$$

if one exists. Set $\widetilde{\mathcal{B}^0} = \{B_0, B_1, \ldots\}$. Then for all $B \in \mathcal{B}^0$ we have that $c_1 B$ intersects $\bigcup\{c_1 B : B \in \widetilde{\mathcal{B}^0}\}$. Define

$$\overline{\mathcal{B}^0} = \{B(x,r) \in \mathcal{B}_\alpha : \exists A \in \widetilde{\mathcal{B}^0} \; A \subset B(x, 5c_1 r(A)), \; f_{B(x,r)} \le c_2 f_A\}.$$

Then $\mathcal{B}^0 \subset \overline{\mathcal{B}^0}$. We proceed by induction. For each $n \in \mathbb{N}$ define

$$\mathcal{B}^n = \big\{B \in \mathcal{B}_\alpha \backslash (\overline{\mathcal{B}^0} \cup \ldots \cup \overline{\mathcal{B}^{n-1}}) : r(B) \in [1/2, 1]2^{-n} K\big\},$$

as above greedily select a sequence $\widetilde{\mathcal{B}^n}$ of balls $B \in \mathcal{B}^n$ with almost maximal f_B such that for every already selected $A \in \widetilde{\mathcal{B}^n}$ we have $c_1 B \cap c_1 A = \emptyset$, and define

$$\overline{\mathcal{B}^n} = \big\{B(x,r) \in \mathcal{B}_\alpha : \exists A \in \widetilde{\mathcal{B}^n} \; A \subset B(x, 5c_1 r(A)), \; f_{B(x,r)} \le c_2 f_A\big\}.$$

Note that we have $\mathcal{B}^n \subset \overline{\mathcal{B}^n}$. Finally set $\widetilde{\mathcal{B}} = \widetilde{\mathcal{B}^0} \cup \widetilde{\mathcal{B}^1} \cup \ldots$. For $A \in \widetilde{\mathcal{B}}$, we denote

$$U_{A,\lambda} = \big\{B(x,r) \in \mathcal{B}_\alpha : A \subset B(x, 5c_1 r(A)), \; f_{B(x,r)} \le c_2 f_A, \; r^{\alpha+\beta} f_{B(x,r)} > \lambda\big\}.$$

Let $\lambda > \varepsilon$ and $B \in \mathcal{B}_\alpha$ with $r(B)^{\alpha+\beta} f_B > \lambda$. Then there is an n with $B \in \overline{\mathcal{B}^n}$, and hence a $A \in \widetilde{\mathcal{B}^n}$ with $B \in U_{A,\lambda}$. Let $A \in \widetilde{\mathcal{B}}$ and $B(x,r) \in U_{A,\lambda}$. Then $A \subset B(x, 5c_1 r(A))$. Since $r \in R_\alpha f(x)$ we have

$$r^\alpha f_{B(x,r)} \geq (5c_1 r(A))^\alpha f_{B(x,5c_1 r(A))} \geq (5c_1 r(A))^\alpha (5c_1)^{-d} f_A$$

which implies

$$r \geq (5c_1)^{1-d/\alpha} r(A)(f_A/f_{B(x,r)})^{1/\alpha} \geq (5c_1)^{1-d/\alpha} c_2^{1/\alpha} r(A).$$

Since $r \leq 5c_1 r(A)$ it follows that

$$r^\beta \leq r(A)^\beta \begin{cases} (5c_1)^\beta, & \beta \geq 0, \\ (5c_1)^{\beta - d\beta/\alpha} c_2^{\beta/\alpha}, & \beta < 0. \end{cases}$$

Together with

$$r^\alpha f_{B(x,r)} \leq (5c_1 r(A))^\alpha c_2 f_A$$

we obtain

$$r^{\alpha+\beta} f_{B(x,r)} \leq c_3 r(A)^{\alpha+\beta} f_A,$$

where

$$c_3 = \begin{cases} (5c_1)^{\alpha+\beta} c_2, & \beta \geq 0, \\ (5c_1)^{\alpha+\beta - d\beta/\alpha} c_2^{1+\beta/\alpha}, & \beta < 0. \end{cases}$$

Thus $U_{A,\lambda}$ is only nonempty if

$$\lambda < c_3 r(A)^{\alpha+\beta} f_A.$$

We can conclude

$$\int_\varepsilon^\infty \lambda^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}-1} \mathcal{L}\left(\bigcup \{ B \in \mathcal{B}_\alpha : r(B)^{\alpha+\beta} f_B > \lambda \} \right) d\lambda$$

$$= \int_\varepsilon^\infty \lambda^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}-1} \mathcal{L}\left(\bigcup_{A \in \widetilde{\mathcal{B}}} \bigcup U_{A,\lambda} \right) d\lambda$$

$$\leq \sum_{A \in \widetilde{\mathcal{B}}} \int_\varepsilon^\infty \lambda^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}-1} \mathcal{L}\left(\bigcup U_{A,\lambda} \right) d\lambda$$

$$= \sum_{A \in \widetilde{\mathcal{B}}} \int_\varepsilon^{c_3 r(A)^{\alpha+\beta} f_A} \lambda^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}-1} \mathcal{L}\left(\bigcup U_{A,\lambda} \right) d\lambda$$

$$\leq \sum_{A \in \widetilde{\mathcal{B}}} (5c_1)^d \mathcal{L}(A) \int_\varepsilon^{c_3 r(A)^{\alpha+\beta} f_A} \lambda^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}-1} d\lambda$$

$$\leq (1/p - (1+\alpha+\beta)/d) \sum_{A \in \widetilde{\mathcal{B}}} (5c_1)^d \mathcal{L}(A) \left(c_3 r(A)^{\alpha+\beta} f_A \right)^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}}$$

$$= (1/p - (1+\alpha+\beta)/d)(5c_1)^d c_3^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}} \sigma_d \sum_{A \in \widetilde{\mathcal{B}}} \left(r(A)^{\frac{d}{p}-1} f_A \right)^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}}$$

$$\leq (1/p - (1+\alpha+\beta)/d)(5c_1)^d c_3^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}} \sigma_d \left(\sum_{A \in \widetilde{\mathcal{B}}} \left(r(A)^{\frac{d}{p}-1} f_A \right)^p \right)^{(1-p(1+\alpha+\beta)/d)^{-1}}.$$

$\square$

4.2 Transfer to dyadic cubes

In this subsection we pass from disjoint balls to dyadic cubes and then conclude Theorem 1.2 using a result from the dyadic setting.

Remark 4.2 There are 3^d dyadic grids $\mathcal{D}_1, \ldots, \mathcal{D}_{3^d}$ such that each ball B is contained in a dyadic cube $Q_B \in \mathcal{D} = \mathcal{D}_1 \cup \cdots \cup \mathcal{D}_{3^d}$ with $l(Q) \lesssim r(B)$.

Lemma 4.3 *Let* $M_\alpha \in \{M_\alpha^c, \widetilde{M}_\alpha\}$ *and* $f \in L^1_{\text{loc}}(\mathbb{R}^d)$. *Then for each* $B \in \mathcal{B}_\alpha$ *we have* $f_{Q_B} \sim f_B$ *and* $l(Q_B) \sim r(B)$.

Proof Let x be the center of B, and Q_B be the cube from Remark 4.2, and $A = B(x, \sqrt{d}\, l(Q))$. Then $r(B) \sim l(Q_B) \sim r(A)$ and $f_B \lesssim f_{Q_B} \lesssim f_A$. Since $B \in \mathcal{B}_\alpha$ we also have $r(A)^\alpha f_A < r(B)^\alpha f_B$ and therefore conclude $f_{Q_B} \lesssim f_A \lesssim f_B$. □

Lemma 4.4 *Let* $M_\alpha \in \{M_\alpha^c, \widetilde{M}_\alpha\}$ *and* $f \in L^1_{\text{loc}}(\mathbb{R}^d)$. *For each* $\alpha > 0$ *and* $B \in \mathcal{B}_\alpha$ *and cube* $P \supset Q_B$ *we have* $l(P)^\alpha f_P \lesssim_\alpha l(Q_B))^\alpha f_{Q_B}$.

Proof For x the center of B define $A = B(x, \sqrt{d}\, l(P))$. Then from $f_P \lesssim f_A$ and $r(A)^\alpha f_A < r(B)^\alpha f_B$ and $f_B \lesssim f_{Q_B}$ we obtain $l(P)^\alpha f_P \lesssim s^\alpha f_{B(x,s)} < r^\alpha f_{B(x,r)} \lesssim_\alpha l(Q_{B(x,r)})^\alpha f_{Q_{B(x,r)}}$. □

Proof of Theorem 1.2 For $B \in \mathcal{B}_\alpha$ denote by P_B the largest cube that attains $\max_{P \supset Q_B} f_P$. Then $P_B \in \mathcal{Q}_0$ and by Lemmas 4.3 and 4.4 we have $l(P_B) \sim_\alpha r(B)$ and $f_{P_B} \sim_\alpha f_B$. By Lemma 4.4 there further exists a cube $p(P_B) \supset P_B$ with $f_{p(P_B)} \leq f_{P_B}/2$ and $l(p(P_B)) \lesssim_\alpha l(P_B)$.

Let $\varepsilon > 0$ and let $\widetilde{\mathcal{B}}$ be the set of balls from Lemma 4.1. By Lemmas 4.3 and 4.4 there are c_1, c_2 such that for any two distinct $B, A \in \widetilde{\mathcal{B}}$ we have that $p(P_B)$ and $p(P_A)$ are disjoint or $f_{P_B}/f_{P_A} \notin (1/2, 2)$. Define $\mathcal{Q} = \{P_B : B \in \widetilde{\mathcal{B}}\}$. By the layer cake formula and Lemmas 4.1, and 4.3 we have

$$\int (M_{\alpha,\beta} f)^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}}$$

$$= (p^{-1} - (1+\alpha+\beta)/d)^{-1} \int_0^\infty \lambda^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}-1} \mathcal{L}(\{M_{\alpha,\beta} f > \lambda\}) \, d\lambda$$

$$= (p^{-1} - (1+\alpha+\beta)/d)^{-1} \lim_{\varepsilon \to 0} \int_\varepsilon^\infty \lambda^{(p^{-1}-(1+\alpha+\beta)/d)^{-1}-1} \mathcal{L}\left(\bigcup \{B \in \mathcal{B}_\alpha : r(B)^{\alpha+\beta} f_B > \lambda\}\right) d\lambda$$

$$\lesssim_{\alpha,\beta,p} \lim_{\varepsilon \to 0} \left(\sum_{B \in \widetilde{\mathcal{B}}} \left(r(B)^{\frac{d}{p}-1} f_B\right)^p\right)^{(1-p(1+\alpha+\beta)/d)^{-1}}$$

$$\sim_{\alpha,\beta,p} \lim_{\varepsilon \to 0} \left(\sum_{Q \in \mathcal{Q}} \left(l(Q)^{\frac{d}{p}-1} f_Q\right)^p\right)^{(1-p(1+\alpha+\beta)/d)^{-1}}.$$

For each $i = 1, \ldots, 3^d$ we apply Lemma 3.8 to $\mathcal{Q} \cap \mathcal{D}_i$ and obtain

$$\sum_{Q \in \mathcal{Q}} \left(l(Q)^{\frac{d}{p}-1} f_Q\right)^p = \sum_{i=1}^{3^d} \sum_{Q \in \mathcal{Q} \cap \mathcal{D}_i} \left(l(Q)^{\frac{d}{p}-1} f_Q\right)^p \lesssim_{\alpha,\beta,p} \|\nabla f\|_p^p.$$

For the endpoint $p = d/(1+\alpha+\beta)$ we use $\|M_{\alpha,\beta} f\|_\infty = \sup_{B \in \mathcal{B}_\alpha} r(B)^{\alpha+\beta} f_B$. Let $B \in \mathcal{B}_\alpha$. Then $f_{2B} \leq 2^{-\alpha} f_B$ and we have by the Sobolev-Poincaré inequality

$$\|\nabla f\|_{d/(1+\alpha+\beta)} \geq \left(\int_{2B} |\nabla f|^{d/(1+\alpha+\beta)}\right)^{(1+\alpha+\beta)/d} \gtrsim r(2B)^{\alpha+\beta-d} \int_{2B} |f - f_{2B}|$$

$$\geq 2^{\alpha+\beta-d} r(B)^{\alpha+\beta-d} \int_B |f - f_{2B}|$$

$$\geq 2^{\alpha+\beta-d} r(B)^{\alpha+\beta-d} \int_B (f - f_{2B})$$

$$= \sigma_d 2^{\alpha+\beta-d} r(B)^{\alpha+\beta} (f_B - f_{2B})$$

$$\geq \sigma_d 2^{\alpha+\beta-d} r(B)^{\alpha+\beta} (1 - 2^{-\alpha}) f_B.$$

$\square$

References

1. Aldaz, J.M., Colzani, L., Pérez Lázaro, J.: Optimal bounds on the modulus of continuity of the uncentered Hardy–Littlewood maximal function. J. Geom. Anal. **22**(1), 132–167 (2012)
2. Aldaz, J.M., Pérez Lázaro, F.J.: Regularity of the Hardy–Littlewood maximal operator on block decreasing functions. Stud. Math. **194**(3), 253–277 (2009)
3. Aldaz, J.M., Pérez Lázaro, J.: Functions of bounded variation, the derivative of the one dimensional maximal function, and applications to inequalities. Trans. Am. Math. Soc. **359**(5), 2443–2461 (2007)
4. Beltran, D., Madrid, J.: Endpoint Sobolev continuity of the fractional maximal function in higher dimensions. Int. Math. Res. Not. **2021**(22), 17316–17342 (2021)
5. Beltran, D., Madrid, J.: Regularity of the centered fractional maximal function on radial functions. J. Funct. Anal. **279**(8), 108686 (2020)
6. Beltran, D., Ramos, J.P., Saari, O.: Regularity of fractional maximal functions through Fourier multipliers. J. Funct. Anal. **276**(6), 1875–1892 (2019)
7. Carneiro, E., González-Riquelme, C.: Gradient bounds for radial maximal functions. Ann. Fenn. Math. **46**(1), 495–521 (2021)
8. Carneiro, E., González-Riquelme, C., Madrid, J.: Sunrise strategy for the continuity of maximal operators. arXiv e-prints (2020)
9. Carneiro, E., Hughes, K.: On the endpoint regularity of discrete maximal operators. Math. Res. Lett. **19**(6), 1245–1262 (2012)
10. Carneiro, E., Madrid, J.: Derivative bounds for fractional maximal functions. Trans. Am. Math. Soc. **369**(6), 4063–4092 (2017)
11. Carneiro, E., Madrid, J., Pierce, L.B.: Endpoint Sobolev and BV continuity for maximal operators. J. Funct. Anal. **273**(10), 3262–3294 (2017)
12. Carneiro, E., Svaiter, B.F.: On the variation of maximal operators of convolution type. J. Funct. Anal. **265**(5), 837–865 (2013)
13. Evans, L.C., Gariepy, R.F.: Measure Theory and Fine Properties of Functions. Textbooks in Mathematics, revised edn. CRC Press, Boca Raton (2015)

14. González-Riquelme, C., Kosz, D.: BV continuity for the uncentered Hardy-Littlewood maximal operator. J. Funct. Anal. **281**(2), 109037 (2021)
15. Hajłasz, P., Malý, J.: On approximate differentiability of the maximal function. Proc. Am. Math. Soc. **138**(1), 165–174 (2010)
16. Hajłasz, P., Onninen, J.: On boundedness of maximal functions in Sobolev spaces. Ann. Acad. Sci. Fenn. Math. **29**(1), 167–176 (2004)
17. Heikkinen, T., Kinnunen, J., Korvenpää, J., Tuominen, H.: Regularity of the local fractional maximal function. Ark. Mat. **53**(1), 127–154 (2015)
18. Kinnunen, J.: The Hardy–Littlewood maximal function of a Sobolev function. Isr. J. Math. **100**, 117–124 (1997)
19. Kinnunen, J., Lindqvist, P.: The derivative of the maximal function. J. Reine Angew. Math. **503**, 161–167 (1998)
20. Kinnunen, J., Saksman, E.: Regularity of the fractional maximal function. Bull. Lond. Math. Soc. **35**(4), 529–535 (2003)
21. Kinnunen, J., Tuominen, H.: Pointwise behaviour of $M^{1,1}$ Sobolev functions. Math. Z. **257**(3), 613–630 (2007)
22. Kurka, O.: On the variation of the Hardy–Littlewood maximal function. Ann. Acad. Sci. Fenn. Math. **40**(1), 109–133 (2015)
23. Luiro, H.: Continuity of the maximal operator in Sobolev spaces. Proc. Am. Math. Soc. **135**(1), 243–251 (2007)
24. Luiro, H.: The variation of the maximal function of a radial function. Ark. Mat. **56**(1), 147–161 (2018)
25. Luiro, H., Madrid, J.: The variation of the fractional maximal function of a radial function. Int. Math. Res. Not. **17**, 5284–5298 (2019)
26. Madrid, J.: Endpoint Sobolev and BV continuity for maximal operators, II. Rev. Mat. Iberoam. **35**(7), 2151–2168 (2019)
27. Pérez, C., Picon, T., Saari, O., Sousa, M.: Regularity of maximal functions on Hardy–Sobolev spaces. Bull. Lond. Math. Soc. **50**(6), 1007–1015 (2018)
28. Tanaka, H.: A remark on the derivative of the one-dimensional Hardy–Littlewood maximal function. Bull. Austral. Math. Soc. **65**(2), 253–258 (2002)
29. Weigt, J.: Variation of the dyadic maximal function. Int. Math. Res. Not. (2020, to appear)
30. Weigt, J.: Variation of the uncentered maximal characteristic function. Rev. Mat. Iberoam. (2020, to appear)