Parabolic BMO and the forward-in-time maximal operator

Olli Saari\textsuperscript{1,2}*

Received: 27 March 2017 / Accepted: 15 February 2018 / Published online: 20 February 2018
© Fondazione Annali di Matematica Pura ed Applicata and Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract We study if the parabolic forward-in-time maximal operator is bounded on parabolic BMO. It turns out that for non-negative functions the answer is positive, but the behavior of sign-changing functions is more delicate. The class parabolic BMO and the forward-in-time maximal operator originate from the regularity theory of nonlinear parabolic partial differential equations. In addition to that context, we also study the question in dimension one.

Keywords Parabolic BMO · Forward-in-time · One-sided · Maximal operator · Heat equation · Doubly nonlinear equation · Parabolic equation · p-Laplace

Mathematics Subject Classification Primary: 42B37, 42B25, 42B35; Secondary: 35K92

1 Introduction

The Hardy–Littlewood maximal operator maps functions of bounded mean oscillation back to BMO. This is a classical result of Bennett et al. [3]. In addition to the original approach, which is direct, some alternative proofs relying on the properties of the Muckenhoupt weights are available in [4,5].

The present paper is devoted to studying the counterpart of the boundedness $M : \text{BMO} \to \text{BMO}$ in a context that comes from the regularity theory of parabolic partial differential equations.
The class parabolic BMO is defined through a condition measuring mean oscillation in a special way. As opposed to the ordinary BMO space, a natural time lag appears in connection with parabolic BMO. It causes several challenges, many of which have only been addressed very recently; see [10,14]. Roughly speaking, the positive and negative parts of the deviation from a constant only satisfy bounds in disjoint regions of the space time. For a function $u$ of space and time to be in $\text{PBMO}^+$, it suffices that

$$\sup_{R} \inf_{a \in \mathbb{R}} \left( \frac{1}{2} \int_{R^{-}} (a - u)^+ + \frac{1}{2} \int_{R^{+}} (u - a)^+ \right) < \infty;$$

see Sect. 2.1 for precise definitions. The condition above leads to many properties similar to those of the ordinary BMO, but it also allows for the possibility of arbitrarily fast growth in the negative time direction. Moreover, no decomposition as a sum of a decreasing function and a classical $\text{BMO}$ function is possible so the difference between parabolic BMO and its classical counterpart is remarkable.

Principal examples of partial differential equations with connection to parabolic BMO are the heat equation and its generalizations, most notably the doubly nonlinear equation:

$$\partial \left( |u|^{p-2} u \right) - \text{div} \left( |\nabla u|^{p-2} \nabla u \right) = 0, \quad 1 < p < \infty. \quad (1.1)$$

Logarithms of positive local solutions to (1.1), possibly with measure data, belong to parabolic BMO [1,7,8,12,14,16]. This gives important examples of non-trivial $\text{PBMO}^+$ functions. A similar relation is also known to hold for elliptic partial differential equations and the ordinary BMO. Consequently, the functions in parabolic BMO relate to those in ordinary BMO in a manner analogous to how supersolutions of the heat equation relate to those of the Laplace equation.

In spite of the interesting interaction between parabolic BMO and regularity properties of PDE, the relevant literature is still limited, and the history can be recounted quickly. The seminal papers of Moser [12] and Trudinger [16] established the connection between the partial differential equations and parabolic BMO. In addition, the parabolic John–Nirenberg inequality was proved there. The proof was simplified in [7] and extended to spaces of homogeneous type in [1]. A method to derive global estimates from the local parabolic John–Nirenberg inequality was developed in [14], and it was also used to answer a question about summability of the supersolutions of (1.1). More recent studies are coupled with the new advances in the theory of multidimensional one-sided weights; see [6,9,10,13,15].

In addition to the parabolic John–Nirenberg inequality [1,12], the known results about parabolic BMO include a Coifman–Rochberg type characterization using the parabolic theory of weights [10] and geometric local-to-global properties [14]. The contribution of this paper is to show that the same operator of forward-in-time maximal averages (see Sect. 2 or Fig. 1)

$$M^{\gamma+} f (z) = \sup_{\ell > 0} \int_{R^{+}(z, \ell, \gamma)} |f| \, dx \, dt$$

that generates the parabolic weight theory also maps the class of positive functions in parabolic BMO into itself. The boundedness of the parabolic forward-in-time maximal operator is consistent with the “elliptic” result in [3] when only positive functions are involved. Namely, the results coincide if we restrict our attention to functions with no time-dependency. In the time-dependent case, Theorem 4.3 is more general. On the other hand, if we allow the functions under study to be sign-changing, a full analog of the Bennett–DeVore–Sharpley...
result will be false in the parabolic context. Hence, there are intrinsically parabolic phenomena involved in our result. See Theorem 4.3 and the related discussion for the precise statements.

We conclude the introduction by briefly describing the structure of the present paper. It is divided into three main sections and an additional section discussing the one-dimensional analog of the problem under investigation. Section 2 is used to set up the notation and to introduce the operators and the function classes we study. In Sect. 3, we prove several auxiliary results such as a chain argument, and we also construct a special dyadic grid. These results are needed to solve problems that arise from the time-dependent nature of the main theorems. Once all the preparations have been carried out, we prove the main theorems, namely Lemma 4.1 asserting the boundedness of

\[ M^{γ+} : \text{PBMO}_{\text{positive}}^+ \rightarrow \text{PBMO}_{\text{positive}}^+ \]

and Theorem 4.3 refining the result. At the end of the paper, we show how a simplified argument can be used to prove a slightly stronger result in dimension one. This is in the context of one-sided BMO spaces of Martín-Reyes and de la Torre [11]. To our best knowledge, also the one-dimensional result is new.

## 2 Notation and definitions

We use standard notation. We mostly work in \( \mathbb{R}^{n+1} \) with the last coordinate called time, the first ones space. The results also hold in space time cylinders that are sets of the form \( \Omega \times \mathbb{R} \) with \( \Omega \subset \mathbb{R}^n \) a bounded domain. The notation \( A \lesssim B \) means that there is an uninteresting constant \( C \) such that \( A \leq CB \). We do not keep track of dependencies on dimension \( n \), the growth type of the equation \( p \) or parameters coming from the domain of definition \( \Omega \). It is clear what \( \gtrsim \) and \( \eqsim \) mean.

For a measurable (we always assume it tacitly) set \( E \), we denote by \( |E| \) its \( n+1 \) dimensional Lebesgue measure. For the integral average, we have the standard notation

\[ f_E = \int_E f = \frac{1}{|E|} \int_E f. \]

For a (measurable) function \( f \), we denote

\[ f^+ = f_+ = f 1_{\{f>0\}} \quad \text{and} \quad f^- = f_- = -f 1_{\{f<0\}}. \]

The function \( 1_E \) equals 1 in \( E \) and zero elsewhere.
We continue by introducing the notation for parabolic rectangles. Let $p > 1$ be a number that is fixed throughout the paper. In applications, it would be the $p$ coming from the $p$-Laplace operator. For evolutionary problems, it has an important role in determining how the time variable and the space variables scale.

If $Q \subset \mathbb{R}^n$ is a cube with sides parallel to coordinate axes, we denote its side length by $\ell(Q)$. Take a parameter $\gamma \in (0, 1)$. In previous papers, this has been called the lag, but here the name shape would be better. We specify a parabolic rectangle together with its upper and lower parts by its center $(x, t)$, side length $\ell(Q)$, and shape parameter $\gamma$. We usually drop some or all of the parameters from the notation, but if they are present, they should be understood as follows (see also Fig. 1):

\[
R((x, t), \ell(Q), \gamma) = Q \times (t - \ell(Q)^p, t + \ell(Q)^p)
\]

\[
R^-(\gamma) = Q \times (t - \ell(Q)^p, t - (1 - \gamma)\ell(Q)^p)
\]

\[
R^+(\gamma) = Q \times (t + (1 - \gamma)\ell(Q)^p, t + \ell(Q)^p).
\]

The number $\ell(R) := \ell(Q)$ is called the side length of the rectangle. Addition of a constant to a set in $\mathbb{R}^{n+1}$ is always understood as adding the constant to the time coordinate.

2.1 Classes of BMO type

We say that $u \in \text{PBMO}$ if the following supremum over all parabolic rectangles is finite:

\[
\|u\|_{\text{PBMO}} := \sup_R \inf_{a \in \mathbb{R}} \left( \int_{R^-} (u - a)^+ + \int_{R^+} (\frac{1}{2}) (a - u)^+ \right) < \infty.
\]

This is not a norm in the precise meaning of the word, but we call it a norm. PBMO$^-$ is not a vector space, but we call it a space. Moreover, we say that an operator $T$ is bounded on PBMO$^-$ if $\|Tu\|_{\text{PBMO}} \leq C\|u\|_{\text{PBMO}}$ even if the set up is not the one of normed linear spaces.

The methods developed in the previous works show that given a shape parameter $\gamma \in (0, 1)$ and a lag coefficient $L > (1 - \gamma)$, there are constants $\{b_R\}_R$ such that

\[
\sup_R \left( \int_{R^-} (u - b_R)^+ + \int_{R^+} (\frac{1}{2}) (b_R - u)^+ \right) \approx_{n, p, \gamma, L} \|u\|_{\text{PBMO}}^{-}.
\]

Suitable references for this are [9,14], and the idea of the proof is also contained in the Lemma 3.3 proved in this paper. That lemma is intended to be an easy reference for the numerous applications of the chain argument that we need.

It has been observed already earlier that PBMO$^-$ can be realized as an intersection of two even rougher function classes. We mention [9] as a reference for the multidimensional case. In dimension one, this claim does not make so much sense because the "rough" function classes turn out to coincide and equal to PBMO$^-$. See [11]. However, when it comes to so called one-sided $L^\infty$ functions, similar things also happen in dimension one [2]. Recall the notations of averages, and let

\[
\|u\|_{\text{BMO}}^{+(\gamma, L)} := \sup_R \int_{R^-} (u - u_{R^-} + L\ell(R)^p)^+ < \infty
\]

\[
\|u\|_{\text{BMO}}^{-(\gamma, L)} := \sup_R \int_{R^-} (u_{R^-} - u + L\ell(R)^p)^+ < \infty.
\]
The first inequality is called BMO$^+$ condition and the second one $-$ BMO$^-$ condition. Note that BMO$^-$ would just be BMO$^+$ with $t$-axis of the coordinate space reversed. This notational convention on the direction of time also holds for PBMO$^\pm$ and the maximal functions that we use. The one-sided function classes BMO$^\pm$ are not known to be independent of $\gamma$ or $L$. However, they are useful because of the following:

\[ \text{PBMO}^- = [\text{BMO}^+(\gamma_1, L_1)] \cap [-\text{BMO}^-(\gamma_2, L_2)] \]  

for any choice of the shape and lag parameters. For details about this, see [9] and the discussion preceding the point to which we have advanced.

PBMO$^-$ is closed under addition and multiplication by positive constants. Multiplication by negative constants reverses the direction of time, that is, it maps PBMO$^-$ to PBMO$^+$. For one-sided spaces BMO$^\pm$, proving or disproving that multiplication by negative constants effects as reversing the time axis is an open problem.

We conclude this section by recalling the parabolic John–Nirenberg inequality. This appears in the literature, and it is proved in [1], whereas its formulation in different geometric configurations is studied in [14].

**Lemma 2.1** Let $u \in \text{PBMO}^-$ and $\gamma \in (0, 1)$. Take a parabolic rectangle $R$. Then there are constants $b_R$ and $c_1, c_2 \approx_{n, p, \gamma} 1$ such that

\[
\frac{1}{R^-_-(\gamma)} \exp\left(\frac{c_1}{\|u\|_{\text{PBMO}^-}} (u - b_R)^+\right) \leq c_2
\]

\[
\frac{1}{R^+_-^+(-\gamma)} \exp\left(\frac{c_1}{\|u\|_{\text{PBMO}^-}} (b_R - u)^+\right) \leq c_2.
\]

**2.2 Parabolic maximal function**

The first candidate to be a parabolic backward-in-time maximal operator was introduced in [10], and it reads as

\[
M^\gamma_- f(x) = \sup_{R(x)} \frac{1}{R^-(\gamma)} \int_{R^-(\gamma)} |f|.
\]

The supremum is over parabolic rectangles centered at $x$ and we average the absolute value over the left part. See also Fig. 2. This definition is problematic when dealing PBMO$^-$ functions that are not necessarily positive. Too many of them are mapped to infinity. For the positive functions, however, it coincides with the following maximal function.

We let

\[
M^\gamma_- f(x) := \sup_{\ell > 0} \left( (f^+)^{-}_R(-x, \ell, \gamma) + (f^-)^{+}_R(x, \ell, \gamma) \right)
\]

\[
:= \sup_{R(x)} \left( \int_{R^-(\gamma)} f 1_{\{f > 0\}} - \int_{R^+_-^+(-\gamma)} f 1_{\{f < 0\}} \right)
\]

where the supremum is over all the parabolic rectangles that are centered at $x$. This maximal function measures positivity in the past, negativity in the future. In case the time variable is trivial, that is, we have functions on $\mathbb{R}^n$ as in the study of elliptic PDE or as in the classical Calderón-Zygmund theory, it coincides with the usual Hardy–Littlewood maximal function.

There are, however, delicate issues involved in the behavior of positive and negative parts of the functions, multiplications by negative numbers, and taking absolute values in the parabolic theory. Recall for instance the problem of $- \text{BMO}^-$, and see Remark 3.2. In
connection with $\text{PBMO}^-$, we have that $\text{PBMO}^- = - \text{PBMO}^+$. The maximal function above is designed to have the same property while maintaining the consistency with the classical Hardy–Littlewood maximal function. Namely, $M^\gamma_{\ast}(-f) = M^\gamma_{\ast}f$. Note also that this maximal function has its natural sharp version that can be used to define $\text{PBMO}^-$. 

3 Preliminary results

This section contains the proofs of some lemmas that we need. In many occasions, we have to decompose functions in $\text{PBMO}^-$ into positive and negative parts, and we next prove that this procedure respects the $\text{PBMO}^-$ property. The subsequent section contains a chain lemma, and the remaining two contain a construction of a dyadic grid suitable for our purposes and some remarks on the local integrability of maximal functions of functions in $\text{PBMO}^-$. 

3.1 Truncations

We start with noting that $\text{PBMO}^\pm$ classes are stable with respect to truncation. Even if this property looks very elementary, its proof still differs a bit from the classical analog. Namely, we really need the equivalence with respect to lag and shape in order to get the conclusion.

**Proposition 3.1** Let $u \in \text{PBMO}^-$. Then $u^+, -u^- \in \text{PBMO}^-$. 

**Proof** Take $u \in \text{PBMO}^-$. We start with $u^+$ and the $\text{BMO}^+$ bound. Take any $\gamma \in (0, 1)$. By elementary considerations 

$$
\int_{R^-_{\gamma}} (u^+ - (u^+)^{R^+(\gamma)})^+ \leq \int_{R^-_{\gamma}} (u^+ - (u^{R^+(\gamma)})^+)^+
$$

$$
\leq \int_{R^-_{\gamma}} (u - u^{R^+(\gamma)})^+ \lesssim \|u\|_{\text{PBMO}^-}.
$$

For the $\text{BMO}^-$ bound, take a parabolic rectangle $R$ with side length $\ell(R)$. Let 

$$
Q = R^-_{\gamma}, \quad Q^+ = R^+_{\gamma} \quad \text{and} \quad Q^{++} = R^+_{\gamma} + (0, \ldots, 0, (1 + \gamma)\ell(R)^p).
$$
Then it is easy to see that for every point in $Q^{++}$ we have that
\[
((u^+)_Q - u^+)^+ \leq \left(\left[ (u - u_{Q^+})^+ \right]_Q + (u_{Q^+}^+ - u^+) \right)^+
\]
\[
\leq \int_Q (u - u_{Q^+})^+ + (u_{Q^+}^+ - u^+)
\]
\[
\lesssim \|u\|_{\text{PBMO}^-} + (u_{Q^+}^+ - u^+)^+,
\]
and averaging over the domain $Q^{++}$ completes the proof for $u^+$. See Remark 3.4 for the fact that the increased gap between $Q$ and $Q^{++}$ does not matter.

For the negative part, we note that $-u \in \text{PBMO}^+$ and by the previous step also $(-u)^+ \in \text{PBMO}^+$. Then $-u^- = -(u^+) \in -\text{PBMO}^+ = \text{PBMO}^-$. □

**Remark 3.2** The previous proposition does not say anything about absolute values of $\text{PBMO}^-$ functions. Indeed, $\text{PBMO}^-$ is not closed under subtraction. In general, absolute values may fail to belong to $\text{PBMO}^-$. An easy one-dimensional example is $e^t - e^{-t}$. As an increasing function it trivially satisfies $\text{PBMO}^-$ condition. However, taking an absolute value converts it into a rapidly decreasing function as we go toward $-\infty$.

### 3.2 Chain argument

The following lemma is what we refer to as the chain argument. The formulation and the proof simplify the arguments that were used in [14], and we hope that writing this argument down in a readable way in this simple case is of independent interest. Moreover, as a corollary of this lemma, we can also easily deduce the independence of shape and lag of the definition of $\text{PBMO}^-$ which we often refer to.

**Lemma 3.3** Let $u \in \text{PBMO}^-$. Consider a rectangle $R^-(\theta)$ with spatial side length $\ell$. Let $v \in \mathbb{R}^n$ and $\tau > (1 - \theta)$. Define
\[
P^+(\theta) = R^- (\theta) + (v, \tau \ell^P).
\]
Then
\[
\int_{P^+(\theta)} \int_{R^- (\theta)} (u(x) - u(y))^+ \, dx \, dy \lesssim_{n, p, \theta, |v|/\ell, \tau} \|u\|_{\text{PBMO}^-}
\]

**Proof** This is a very simple instance of the chain argument, but for the reader’s convenience we give a proof. Assume first that $\tau$ is very large. Denote the base of $R^-(\theta)$ by $Q$. It has side length $\ell$. Let
\[
Q_i = Q + \frac{v}{|v|} r_i, \quad r_i > 0.
\]
We choose such a sequence $\{r_i\}_{i=1}^k$ that
\[
\frac{|Q_i \cap Q_{i+1}|}{|Q_0|} = \delta
\]
where $\delta \in (0, 1)$ is a number such that $Q_k = Q_0 + v$. Note that the numbers $\delta$ and $k$ depend only on $|v|$.
The sets $P_{2i-1}$, their translates forwards and backwards in time, and the real parabolic rectangles motivating these boxes

Let $P_1 = R^-(\theta)$. It has the least time coordinate $t_1$. Define

- $P_{2i-1} = Q_i \times (t_1 + (i - 1)(1 + \theta)\ell^p, t_1 + [(i - 1)(1 + \theta) + (1 - \theta)]\ell^p)$
- $P_{2i-1}^+ = P_{2i-1} + (1 + \theta)\ell^p$
- $P_{2i-1}^- = P_{2i-1} - (1 + \theta)\ell^p$
- $S_{2i} = P_{2i-1}^+ \cap P_{2i+1}^-$.

The sets $S_{2i}$ are such that $\frac{|S_{2i}|}{|P_{2i-1}|} = \delta$. Moreover, the pairs of sets $(P_{2i-1}, P_{2i-1}^+)$ can be used in testing the PBMO\(^-\) conditions. See also the Fig. 3.

Since the chain of cubes travels the amount $|v|$ in space, the number of cubes needed is roughly

$$k \approx \frac{|v|}{\ell}.$$

This means that the sequence of space time rectangles $\{P_{2i-1}\}_i$ advances, up to a constant factor, the amount $|v|\ell^p/\ell = |v|\ell^{p-1}$ in time.

By our assumption on $\tau$ being large, we can find $\delta$ such that

$$P_{2(k+l)+1} = P^+(\theta)$$

where $l \lesssim \tau$ is an integer taking care of the possibility that $\tau$ is too large. The exact requirement on $\tau$ is

$$\tau \ell^p \gtrsim |v|\ell^{p-1}. \quad (3.1)$$
Note first that
\[
\int_{P^+(\theta)} \int_{R^-(\theta)} (u(x) - u(y))^+ \, dx \, dy \\
\leq \int_{R^-(\theta)} (u - u_{R^+(\theta)})^+ + (u_{R^+(\theta)} - u_{P^-(\theta)})^+ + \int_{P^+(\theta)} (u_{P^-(\theta)} - u)^+
\]
\[
\lesssim (u_{R^+(\theta)} - u_{P^-(\theta)})^+ + 2\|u\|_{PBM0^-}.
\]
By this reduction, there is no loss in generality in the following estimation:
\[
(u_{R^-(\theta)} - u_{P^+(\theta)})^+ \leq \sum_{i=1}^{k+l} (u_{P_{2i-1}^+} - u_{P_{2i+1}^+})^+ = \sum_{i=1}^{k+l} \int_{S_{2i}} (u_{P_{2i-1}^+} - u_{P_{2i+1}^+})^+
\]
\[
\lesssim \delta \sum_{i=1}^{k+l} \left( \int_{P_{2i}^+} (u_{P_{2i-1}^+} - u)^+ + \int_{P_{2i+1}^+} (u - u_{P_{2i+1}^+})^+ \right)
\]
\[
\lesssim k, \|u\|_{PBM0^-}.
\]
In this estimate, the quantities \(l, k, \delta\) depend only on \(|v|/\ell, \tau, n, p\) and \(\theta\).

To get rid of the restriction of \(\tau\) being a large number, we do the following. We find an auxiliary pair of rectangles \(P^+(\theta)\) and \(R^-(\theta)\) that have side lengths \(\epsilon \ell\) and that form upper and lower halves of a parabolic rectangle \(R_{\epsilon}\) that is centered in the middle of the line connecting the centers of \(P^+(\theta)\) and \(R^-(\theta)\).

We cut the original blocks into pieces with the same side length \(\epsilon \ell\). The pieces of \(R^-(\theta)\) are called \(A_i^-\) and the pieces of \(P^+(\theta)\) are called \(B_i^+\). Then
\[
(u_i^- - u_j^+) \leq (u_i^- - u_{R^-(\theta)})^+ + (u_{P^+(\theta)} - u_{B_j^+})^+.
\]
These two terms are back in the original situation with data \(v_i, v_j \in \mathbb{R}^n, \|v_i\| \approx \|v_j\| \approx |v|,\) and \(\tau_i \approx \tau_j \approx \frac{\tau}{\epsilon |v|}.\) Then
\[
\tau_i (\epsilon \ell)^p \gtrsim \tau \ell^p,
\]
so the lower bound asked by (3.1) is satisfied if
\[
\tau \ell^p \gtrsim |v_i| (\epsilon \ell)^{p-1} \approx \frac{|v|}{\ell} \epsilon^{p-1} \ell^p,
\]
whence we deduce that \(\epsilon \lesssim (\ell \tau/|v|)^{1/(p-1)}\) can be made sufficiently small. This bound only depends on legal quantities, and we may consider \(\epsilon\) a constant in the future.

Finally, we may compute
\[
\int_{P^+(\theta)} \int_{R^-(\theta)} (u(x) - u(y))^+ \, dx \, dy \lesssim \sum_{i,j} \int_{A_i^-} \int_{B_j^+} (u(x) - u(y))^+
\]
which reduces the situation to the case already handled. \(\square\)

**Remark 3.4** Given two, possibly rather weird, bounded Borel sets \(A\) and \(B\) from \(\mathbb{R}^{n+1}\) such that
\[
\Delta = \inf \{t : (x, t) \in B\} - \sup \{t : (x, t) \in A\} > 0,
\]
we can conclude by the previous lemma that

\[(u_A - u_B)^+ \leq \int_A \int_B (u(x) - u(y))^+ \, dx \, dy \lesssim \|u\|_{\text{PBMO}}^{-}\]

where the dependency is on the parameters of the Lemma 3.3 associated to a pair of parabolic rectangles \(R_1\) and \(R_2\) that contain the sets \(A\) and \(B\) and satisfy the assumptions of the Lemma. More precisely, there is dependency in terms of a function separately increasing in \(|R_1|/\min(|A|, |B|), \max(d(A), d(B))/\Delta, \) and \(d(A, B)/\Delta\).

**Remark 3.5** The idea of the previous remark also generalizes to the John–Nirenberg inequality of \(\text{PBMO}^{-}\), Lemma 2.1. This is quite a direct consequence of the convexity of the exponential. Hence, we may give an improved formulation of the John–Nirenberg inequality with \(c_1\) and \(c_2\) depending on the same constants as in the above remark:

\[\int_A \int_B \exp\left(\frac{c_1}{\|u\|_{\text{PBMO}}} (u(x) - u(y))^+\right) \, dx \, dy \lesssim c_2.\]

### 3.3 Dyadic grid

In course of the proof of the main theorem, we decompose a function \(u\) into a bounded part and into additional pieces that are roughly of the form

\[b_Q = 1_Q (u - u_{Q^+}).\]

This form is particularly convenient when working with \(u \in \text{PBMO}^{-}\). We want the pieces to have supports with controlled overlap, and obtaining that property in the geometry where the time variable scales as space to power \(p\) requires some work.

For an effective use of the Calderón-Zygmund stopping time argument, it is beneficial to dispose of a dyadic grid. The existence of such a grid is non-trivial in the parabolic geometry if we want to maintain an intuition of how our dyadic parabolic rectangles look like. This intuition is lost if we use the tools from the analysis in metric spaces as black boxes. For example, various general constructions of dyadic sets have been developed since the seminal result by Michael Christ, but their use in our setting does not seem to lead to clearer exposition.

The main problem in the dyadic grid is that if the number \(p\) governing the geometry is an irrational number, then we cannot simply subdivide a box to \(2^n\) cubes in space and \(2^p\) boxes in the space since the latter number is not an integer. This problem is easy to circumvent if \(p\) is rational, but in case of irrational \(p\) one has to resort to an approximative construction.

**Lemma 3.6** Let \(Q_0\) be a half of a parabolic rectangle with side length \(\ell(Q_0)\). Then, there exists boxes with the properties of the dyadic tree that are almost parabolic sub-rectangles of \(Q_0\):

(i) \(\mathcal{D} = \bigcup_{i=0}^{\infty} \mathcal{D}_i\) and \(\mathcal{D}_i = \{Q^{(i)}_j\}_j\). In addition \(\cup_j Q^{(i)}_j = Q_0 \in \mathcal{D}_0\) for all \(i\). The boxes \(Q^{(i)}_j\) with common \(i\) are translates of each other.

(ii) If \(P, Q \in \mathcal{D}\), then \(P \cap Q \in \{\emptyset, P, Q\}\). For each \(Q \in \mathcal{D}_i\), there is a unique \(\hat{Q} \in \mathcal{D}_{i-1}\) with \(\hat{Q} \supseteq Q\). In addition, \(|Q| \approx |\hat{Q}|\).

(iii) Every \(Q \in \mathcal{D}_i\) almost has the dimensions \((\ell, \ell^p)\). Namely, there is a parabolic box \(\tilde{Q}\) obtained from \(Q_0\) by means of parabolic dilation \((x, t) \mapsto (\delta x, \delta^p t)\) and translation so that \(\ell(\tilde{Q}) = 2^{-i} \ell(Q_0), \tilde{Q} \supseteq Q\), and \(|\tilde{Q}| \approx |Q|\).
Proof For simplicity, and without loss of generality, we may assume that $Q_0 = [0, 1]^{n+1}$. For every $p > 1$, it is easy to see that there exists a non-decreasing sequence of integers $k_i$ such that for $i \in \mathbb{Z}_+$,\[
|p - \frac{k_i}{i}| \leq \frac{1}{i}.
\]We denote $q_i = k_i / i$.

At step one, we divide the spatial side length of $Q_0$ by 2 and the temporal one by $2^{k_1}$. These dimensions give us space time boxes in the geometry where time scales as space to power $q_1$. We use them to partition $Q_0$. This gives the collection $D_1$ that serves as a generation in the dyadic grid we are constructing.

At the second step, we keep repeating the process so that boxes in $D_{i-1}$ have their spatial side length halved so that we reach the desired side length $2^{-i}$. The previous side length was $2^{1-i}$. We choose the temporal side length to be $2^{-k_i}$. Note that this is obtained by repeatedly multiplying $2^{-k_{i-1}}$ by $2^{-1}$ (possibly zero times) since the sequence $k_i$ is increasing. With these dimensions, we get space time boxes in the geometry where time scales as the space to power $q_i$. We use these blocks to partition boxes in $D_{i-1}$. Their collection is called $D_i$.

Finally note that if $Q \in D_i$, and $\tilde{Q}$ is a block with the same spatial side length that respects the geometry where time scales as space to power $p$, then\[
\frac{|Q|}{|\tilde{Q}|} = \frac{2^{-k_i}}{2^{-pi}} = 2^{pi-k_i} \in (2^{-1}, 2).
\]Hence, we have good control over the distortion of the geometry at all scales. Consequently, we may always cover the dyadic box with a proper parabolic rectangle so that the error in volume is controlled.

\[\Box\]

3.4 Integrability of maximal functions

The functions in PBMO$^-$ are locally integrable. In spatially unbounded domains, the maximal functions may easily fail this property so some care must be taken. One suitable criterion for local integrability of maximal function of a PBMO$^-$ function is given by its finiteness.

Lemma 3.7 (i) Let $u \in \text{PBMO}^-(\Omega)$ where $\Omega \subset \mathbb{R}^{n+1}$ is a space time cylinder. Then $M^u_{-}\in L^1_{loc}(\Omega)$.

(ii) Let $u \in \text{PBMO}^-(\mathbb{R}^{n+1})$. If there are $x_1, x_2 \in \mathbb{R}^n$ and numbers $t_2 > t_1$ such that\[
M^u_{-}\in (x_2, t_2) < \infty \quad \text{and} \\
M^u_{+}\in (x_1, t_1) < \infty,
\]then\[
M^u_{-}\in L^1_{loc}((y, \tau) \in \mathbb{R}^{n+1} : t_1 < \tau < t_2).
\]

Proof (i) For the case of the space time cylinder, choose any rectangle $R \subset \Omega$. We may apply the John–Nirenberg from [14] to get exponential integrability for $u\pm$ in the union of the rectangles that the maximal function sees from there. This is enough to show that the maximal function has to be locally integrable.

(ii) In the case where $u$ is defined on $\mathbb{R}^{n+1}$, some more analysis is needed. Assume first that $u \geq 0$. We may restrict our attention to a maximal operator that only sees large rectangles since the case of small rectangles is the same as the case of a space time
cylinder. If we write $U = \max\{U_1, U_2\}$ where $U$ is the maximal function, $U_1$ is the small rectangle supremum, and $U_2$ is the supremum over large rectangles, we see that for all parabolic rectangles $R$ located in the half-space $\{(y, \tau) \in \mathbb{R}^{n+1} : \tau < t_2\}$ we have that
\[
\mathcal{F}_{R^-(\gamma)} U_2 \lesssim_{R, n, p, \|u\|_{\text{PBMO}^-}} U(x, t) < \infty
\]
by the argument in Sect. 4.1. Hence, we get that $M_{\gamma}^- u$ is locally integrable in the half-space. In case we have a general $u = u^+ - u^-$, we may do the previous estimation for the positive and negative parts separately. \square

4 Parabolic BMO and maximal functions

Next we study the boundedness of the parabolic maximal operator on parabolic BMO space. This is inspired by a result in Bennett et al. [3]. Indeed, under the positivity assumption, the following lemma generalizes their result. The case of more general functions is discussed after the proof of this lemma.

We do not distinguish the cases where the domain of definition of $\text{PBMO}^-$ functions is a space time cylinder or a full space. The only difference between these cases is in integrability issues of the maximal function. In what follows, we assume the local integrability of $M_{\gamma}^- u$, and criteria for this condition in the two different cases can be found in Lemma 3.7.

**Lemma 4.1** Let $u \in \text{PBMO}^-$ be such that $u \geq 0$ almost everywhere. Then $M_{\gamma}^- u = M_{\gamma}^- u \in \text{PBMO}^-$ and
\[
\|M_{\gamma}^- u\|_{\text{PBMO}^-} \lesssim_{\gamma} \|u\|_{\text{PBMO}^-}
\]
provided that $M_{\gamma}^- u$ is locally integrable.

**Proof** Let $u \geq 0$ be in $\text{PBMO}^-$. We start by proving that $M_{\gamma}^- u \in \text{BMO}^+$. Let $R_0$ be an arbitrary parabolic rectangle. We will use the following notation:
\[
U(x) = M_{\gamma}^- u(x)
\]
\[
U_1(x) = \sup\{u_{R^-(\gamma)} : x \text{ is the center of } R, \ell(R) \leq 100^{-1} \ell(R_0)\}
\]
\[
U_2(x) = \sup\{u_{R^-(\gamma)} : x \text{ is the center of } R, \ell(R) \leq 100^{-1} \ell(R_0)\}
\]
Note that $U(x) = \max\{U_1(x), U_2(x)\}$. We may choose the configuration in which we attempt to bound the $\text{PBMO}^-$ norm type quantities. This is due to Lemma 3.3. We let $R_0^- = R_0^-(0)$ be a standard half of a parabolic rectangle whereas $R_0^+ = R_0^+ + 100\ell^p(R_0^-)$.

We will obtain estimates for
\[
\mathcal{F}_{R_0^-} (U_1(x) - U_{R_0^+})^+ \quad \text{and} \quad \mathcal{F}_{R_0^-} (U_2(x) - U_{R_0^+})^+.
\]
We start the proof with the case of $U_2$.

4.1 Large rectangles

Take a point $x \in R_0^-$ and any rectangle $R$ that is admissible in the definition of $U_2$, that is, $x$ is the center of $R$ and $\ell(R) \geq 100^{-1} \ell(R_0)$. We denote by $A$ the number satisfying $\ell(R) = A \ell(R_0)$.
Take a point \( z \) from \( R^+ \) and a rectangle \( P \) centered at \( z \) with
\[
 u_{P^-(\gamma)} \leq M^\gamma_{R^+} u(z)
\]
such that it has large enough side length \( B\ell(R_0) \). The constant \( B \) will be fixed later. It has to have the correct relations to the constant \( A \).

Next we look at how close the rectangles \( R^-(\gamma) \) and \( P^-(\gamma) \) are to each other. Let
\[
 \tau^- = \sup \{ t : (x, t) \in R^-(\gamma) \} \\
 \tau^+ = \inf \{ t : (x, t) \in P^-(\gamma) \}.
\]
Note that since
\[
 \tau^+ - \tau^- \geq \ell(R_0)^p (A^p \gamma - B^p),
\]
we may choose the constant \( B \) to satisfy
\[
 B^p = \frac{\gamma}{100} A^p
\]
so that \( P \) and \( R \) are of comparable size and well separated. Independently of \( R \), the separation of the rectangles in both time and space variables is governed by the parameters \( n, p, \gamma \).

Hence, we may apply the Lemma 3.3 in form of Remark 3.4:
\[
 (u_{R^-(\gamma)} - U(z))^+ \leq (u_{R^-(\gamma)} - u_{P^-(\gamma)})^+ \lesssim \|u\|_{\text{PBMO}}.
\]
Taking a supremum over \( R \), we get a pointwise bound for \( U_2 - U(z) \) for every \( z \in R_0^+ \). This also proves the claim about the mean value:
\[
 \int_{R_0^+} \int_{R_0^-} (U_2(x) - U(y))^+ \, dx \, dy \lesssim \|u\|_{\text{PBMO}}.
\]
This estimate controls the large rectangle part trivially.

**Remark 4.2** Note that in this part it was only essential that the positive term was average of a large rectangle, and the size of the rectangle in the negative term could be chosen. Hence, we got
\[
 \sup_{r > 100^{-1}\ell(R_0)} \inf_{p > 0} (u_{R^-(x, r, \gamma)} - u_{R^-(y, p, \gamma)})^+ \lesssim \|u\|_{\text{PBMO}}.
\]

### 4.2 Small rectangles I: Calderén–Zygmund

In this part, we use the dyadic grid constructed in Lemma 3.6. The cases with \( p \in \mathbb{Q} \) and \( p \notin \mathbb{Q} \) are not very different since the fact that the dyadic grid only approximates the real one only comes into the picture in few occasions. In order to keep the notation more simple, we work on the case with rational \( p \) and comment the corrections that should be done with approximate dyadic grid. This saves us from some additional indices. Before attacking our target rectangle \( R_0 \), we make some preparations.

Take a parabolic rectangle with lower part \( Q^-_0 \). Let \( \Omega_1 = \{ x \in Q^-_0 : u_{Q^+_i} < U_1(x) \} \).

We run a Calderén–Zygmund stopping time argument on \( Q^-_0 \) with a stopping rule \( u_{Q^+_i} > \lambda \).
where $\lambda > u_{Q_0^+}$. Denote by $\hat{Q}$ the dyadic parent of $Q$. We decompose $u = b + g$ where the components are

$$b_i = 1_{Q_i}(u - u_{\hat{Q}_i^+}),$$

$$b = \sum_i b_i,$$

$$g = \sum_i 1_{Q_i}u_{\hat{Q}_i^+} + u1_{Q_0^- \setminus \cup_i Q_i}.$$  

In case we use the approximate dyadic grid, the stopping rule applies to $(\hat{\hat{Q}})$ rectangles, while the indicator functions in the decomposition of $u$ are left as they are. Recall the details of the dyadic grid from Lemma 3.6.

By construction, $\|g\|_{L^\infty} \leq \lambda$. By the parabolic John–Nirenberg inequality (Lemma 2.1 and Remark 3.5), we have that

$$-\int_{\hat{Q}_i} e^{2\epsilon b_i^+} \leq -\int_{\hat{Q}_i} e^{2\epsilon (u - u_{\hat{Q}_i^+})} \leq 2$$

if $\epsilon \lesssim \|u\|_{PBMO}^{-1}$ is small enough. By elementary properties of the maximal functions

$$\int_{\hat{Q}_0^-} e^{\gamma M^\gamma - b^+} \leq \left( \int_{\hat{Q}_0^-} (M^\gamma - e^{\epsilon b^+})^2 \right)^{1/2} \lesssim \left( \int_{\hat{Q}_0^-} e^{2\epsilon b^+} \right)^{1/2}.$$ 

By the fact that $Q_i$, the supports of $b_i$, are disjoint, we conclude that

$$\int_{\hat{Q}_0^-} e^{2\epsilon b^+} \leq 1 + \sum_i \frac{|Q_i|}{|Q_0^-|} \int_{\hat{Q}_i} e^{2\epsilon b_i} \lesssim 1 + \sum_i \frac{|Q_i|}{|Q_0^-|} \int_{\hat{Q}_i} e^{2\epsilon b_i} \lesssim 1.$$ 

Moreover,

$$M^\gamma_u(1_{Q_0^-}u) = M^\gamma_u(1_{Q_0^-}(b + g)^+) \leq M^\gamma_u(b^+) + M^\gamma_u(g^+)$$

so using the previous computation, we obtain

$$\int_{\Omega_1} U_1(x) \lesssim |Q_0^-| \|u\|_{PBMO}^{-1} \log \exp \int_{Q_0^-} M^\gamma_u(\epsilon b^+) + |\Omega_1| \|M^\gamma_u(g^+)\|_{L^\infty} \leq C_0 \|u\|_{PBMO}^{-1} |Q_0^-| + |\Omega_1| \|g^+\|_{L^\infty} \leq \|u\|_{PBMO^{-1}} |Q_0^-| + \lambda|\Omega_1|.$$ 

Subtracting $|\Omega_1|\epsilon$ and letting $\epsilon \to u_{Q_0^+}^+$, we get

$$\int_{Q_0^-} (U_1(x) - u_{Q_0^+})^+ \lesssim \|u\|_{PBMO^{-1}}.$$ 

4.3 Small rectangles II: chains

Next consider the function $U_1$ in $R_0^-$. For computing the averages, the maximal function only has the region

$$E = \bigcup_{z \in R_0^-(y), 0<r<100^{-1}l(R_0)} R(z, r, y)$$

Springer
at its disposal. This region is small compared to the gap between $R_0^-$ and $R_0^+$, so we may cover it with one single rectangle $Q_0^-$ such that $Q_0^+$ is still before $R_0^+$.

Now we can localize the maximal function to the region $Q_0^-$. For all $x \in R_0^-$, we have that

$$U_1 \leq M_{*}^{\gamma^-}(1_E u) \leq M_{*}^{\gamma^-}(1_{Q_0^-} u).$$

This together with the previous Calderón–Zygmund consideration gives the estimate

$$\int_{R_0^-} (U_1 - U_{R_0^+})^+ \leq \int_{Q_0^-} (M_{*}^{\gamma^-}(1_{Q_0^-} u) - u_{Q_0^+}^+) + (u_{Q_0^+}^- - u_{R_0^+})^+$$

$$\lesssim |Q_0^-| \|u\|_{\text{PBMO}^-} + (u_{Q_0^+}^- - u_{R_0^+})^+. \quad (4.2)$$

In the first term, recall that $|Q_0^-| \approx |R_0^-|$. For the second term, we may use the Remark 3.4 generalizing Lemma 3.3. This completes the proof of the case of small rectangles.

### 4.4 The case of all sizes

In general, we may estimate the full maximal function

$$U = \max\{U_1, U_2\}$$

by the sum of its two parts:

$$\int_{R_0^-} (U - U_{R_0^+})^+ \leq \int_{R_0^- \cap \{U_1 \geq U_2\}} (U_1 - U_{R_0^+})^+$$

$$+ \int_{R_0^- \cap \{U_2 > U_1\}} (U_2 - U_{R_0^+})^+. \quad (4.3)$$

The first term was bounded in the previous sections, and the second one was the large rectangle case.

Up to now we have proved that if $u \geq 0$ and $u \in \text{PBMO}^-$, then $M_{*}^{\gamma^-} u \in \text{BMO}^+$. In order to get the claimed $M_{*}^{\gamma^-} u \in \text{PBMO}^-$, we still have to show that $U = M_{*}^{\gamma^-} u \in -\text{BMO}^-$, that is,

$$\sup_R \int_{R^+} (U_{R^-} - U)^+ \lesssim \|u\|_{\text{PBMO}^-}. \quad (4.4)$$

Compare to the identity (2.1).

However, this can be reduced to the case just handled. Take three blocks similar to $R_0^\pm$ in the previous considerations:

$$Q = R_0^-, \quad Q^+ = R_0^+, \quad Q^{++} = R_0^++ (0, \ldots, 0, 100\ell(R_0)^p).$$

Now

$$\int_{Q^{++}} (U_Q - U)^+ \leq \int_{Q^{++} \cap \{U_1 > U_2\}} (U_1 - U_2)^+ \, dx \, dy$$

$$= \int_{Q^{++}} \frac{1}{|Q|} \int_{Q \cap \{U_1 > U_2\}} (U_1 - U_2)^+ \, dx \, dy$$

$$+ \int_{Q^{++} \cap \{U_1 \leq U_2\}} (U_2 - U)^+ \, dx \, dy$$

$$= I + II.$$
The second term $I I$ is clear by the same argument that lead to the large rectangle case inequality (4.1). For the first term, we can estimate

$$
I \leq \int_{Q^{++}} \frac{1}{|Q|} \int_{Q \cap \{u_1 > u_2\}} (U_1(x) - u_{Q^+})^+ \, dx \, dy \\
+ \int_{Q^{++}} (u_{Q^+} - U(y))^+ \, dx \, dy
$$

where the second term is bounded by the very definition of PBMO$^{-}$ together with the fact $u \leq U$, and the first term is dealt with the small rectangle inequality (4.2) that we proved previously. This completes the proof that $M_{\gamma}^{-} u \in -\text{BMO}^{-}$, and consequently we have that $M_{\gamma}^{-} u \in \text{PBMO}^{-}$.

4.5 Concluding remarks

The preceding lemma contains essentially all that we wanted to prove. We rephrase the results in the next theorem. Note that for time-independent functions this theorem gives the classical BMO $\rightarrow$ BMO result, whereas the case with non-trivial time-dependency is new and interesting.

**Theorem 4.3** Let $u = u_+ - u_- \in \text{PBMO}^{-}$ have a locally integrable $M_{\gamma}^{-}$ maximal function. Then

(i) $M_{\gamma}^{-} u_+ \in \text{PBMO}^{-}$

(ii) $M_{\gamma}^{-} (-u_-) \in \text{PBMO}^{+}$

(iii) The maximal functions above, $U_+$ and $U_-$, control pointwise $M_{\gamma}^{-} u$ in the following way:

$$
\max\{U_- , U_+\} \leq M_{\gamma}^{-} u \leq U_- + U_+.
$$

**Proof** The first item is Lemma 4.1. The second one follows from the fact that $u_- \in \text{PBMO}^{+}$ by Proposition 3.1, and

$$
M_{\gamma}^{-} (-u_-) = M_{\gamma}^{+} (u_-) \in \text{PBMO}^{+}
$$

by symmetry. The third item is obvious by the previous ones.

Functions of the type $e^t - e^{-t}$ on $\mathbb{R}$ show that the third item is almost the best one may hope for. The method of the proof of Lemma 4.1 does not seem to give a better result than the one above. However, already the maximal function belonging to the sum space PBMO$^{-} + \text{PBMO}^{+}$ would look very much like the classical BMO $\rightarrow$ BMO result for the Hardy–Littlewood maximal function. On the other hand, a function $u \in \text{BMO}(\mathbb{R}^n)$ satisfies $u \in \text{PBMO}^{+} \cap \text{PBMO}^{-}$ and $M_{\gamma}^{+} u = M$ in the time-independent case so the well-known stationary result is covered by our evolutionary theorem.

5 A one-dimensional result

The question about boundedness of the one-sided maximal operators on BMO$^{\pm}$ spaces of Martín-Reyes and de la Torre [11] has also not been studied prior to this work, at least to our best knowledge. In this setting, a statement corresponding to Lemma 4.1 is much easier to prove, but we point out that the correct claim cannot be deduced directly from the multidimensional result.
In dimension one, the relevant maximal function of \( u \geq 0 \) is usually defined as
\[
U(x) = \sup_{h>0} \frac{1}{h} \int_{x-h}^{x} u,
\]
and there is no gap between the evaluation point \( x \) and the domain of integration \((x-h, x)\). It is a general phenomenon that expressions that have a gap in the parabolic context seldom have it in dimension one. It is also usual that the gap is qualitatively inessential in dimension one, whereas it is only quantitatively inessential in the parabolic context. We refer to [9] and [11] for more precise discussion on what is the role of the gap in the theory of one-sided maximal functions, weights, and BMO.

The following exposition is intended to be independent of the rest of the paper. We start by recalling the definitions from [11]. We say that \( u \in \text{BMO}^+(\mathbb{R}) \) if
\[
\|u\|_* := \sup_{I} \frac{1}{|I|} \int_{I} (u - u_{I^+})^+ < \infty.
\]
The supremum is over all intervals, and \( I^+ = I + |I| \). This number is comparable to
\[
\sup_{I} \frac{1}{|I|^2} \int_{I} \int_{I^+} (u(t_1) - u(t_2))^+ dt_2 dt_1
\]
as one can deduce from the theorems 2 and 3 in [11]. Finally, remember the definition of the one-sided maximal function given in the beginning of this section.

**Theorem 5.1** Let \( u \in \text{BMO}^+(\mathbb{R}) \) be positive. If \( U \) is locally integrable, then
\[
\|U\|_* \leq C \|u\|_*
\]

**Proof** Take \( u \geq 0 \) from \( \text{BMO}^+(\mathbb{R}) \). Let \( I \) be an interval. We note that
\[
U(x) = \max\{ U_1(x), U_2(x) \}
\]
where
\[
U_1(x) = \sup\{ u(x-h,x) : h \leq |I| \}
\]
\[
U_2(x) = \sup\{ u(x-h,x) : h \geq |I| \}.
\]

\[ \square \]

**Large \( h \)**

We start by estimating \((U_2(x) - U_1^+)\) where \( I^+ = I + |I| \), and \( x \in I \). Take any \( h \geq |I| \). Choose any \( y \in I^+ \). Suppose that \( x \leq y - h \). Then
\[
(u(x-h,x) - u(y-h,y))^+ \leq \frac{1}{h^2} \int_{x-h}^{x+y-h} \int_{x+y-h}^{y} (u(t_1) - u(t_2))^+ dt_2 dt_1
\]
\[
\leq C \frac{h + y - x}{2} \|u\|_* \leq C \|u\|_*
\]
since \( y - x \leq 2|I| \leq 2h \).

In the complementary situation \( x > y - h \), let \( k \) be the positive integer such that \( x - k(y - x) \in [x - h, y - h] \). The proof of the claim proceeds through an iterative algorithm which either stops after hitting the case \( k = 1 \) or converges asymptotically.
Case $k = 1$

The property $k = 1$ is equivalent to
\[ y - x \geq \frac{1}{2} h. \]

If this holds, we can bisect both $(x - h, x)$ and $(y - h, y)$ into two halves of equal length. Call them $I^+_{x}$ and $I^-_{y}$. Then
\[
(u_{(x-h,x)} - u_{(y-h,y)})^+ = 2(u_{I^-_{x}} + u_{I^+_{x}} - (u_{I^-_{y}} + u_{I^+_{y}})) \\
\leq 2(u_{I^-_{x}} - u_{I^+_{x}}) + 2(u_{I^+_{x}} - u_{I^-_{x}}) \\
\leq C\|u\|_s
\]
by a reduction to the case already handled.

Case $k > 1$

Now the intersection of the intervals is big. Denote $y - x = d$ and $I_j = (x - jd, x - (j - 1)d)$. Then
\[
(u_{(x-h,x)} - u_{(y-h,y)})^+ = \frac{1}{h} \left( \sum_{j=1}^{k} \int I_j u - \sum_{j=1}^{k} \int_{I_{j-1}} u + \int_{x-h}^{x-kd} u - \int_{y-h}^{x-(k-1)d} u \right)^+.
\]

We modify the left-most intervals in the sums by setting
\[
\tilde{I}_k = I_k \cup (x - h, x - kd) \quad \text{and} \quad \tilde{I}_{k-1} = I_{k-1} \cup (y - h, x - (k - 1)d).
\]
The other intervals we keep as they are. With this notation, we can continue to estimate
\[
(u_{(x-h,x)} - u_{(y-h,y)})^+ \leq \frac{|\tilde{I}_k|}{h}(u_{\tilde{I}_k} - u_{\tilde{I}_{k-1}}^+) + \frac{1}{h} \sum_{j=1}^{k-1}|I_j|(u_{I_j} - u_{I_{j-1}}^+) \\
\leq \frac{1}{h} \left( |\tilde{I}_k|(u_{\tilde{I}_k} - u_{\tilde{I}_{k-1}}^+) + (k - 1)d\|u\|_s \right) \\
= \frac{1}{h} \left( |\tilde{I}_k|(u_{\tilde{I}_k} - u_{\tilde{I}_{k-1}}^+) + (h - |\tilde{I}_k|)\|u\|_s \right).
\]

Here, $\tilde{I}_k \cap \tilde{I}_{k-1} \neq \emptyset$ and we can keep iterating the process. Namely, we can repeat the argument for $(u_{\tilde{I}_k} - u_{\tilde{I}_{k-1}}^+)$. If $|\tilde{I}_k \cap \tilde{I}_{k-1}| \leq \frac{1}{2}|\tilde{I}_k|$ the process terminates after an application of Case $k = 1$ on these intervals. Otherwise, we iterate the process as follows.

Denote
\[
(x - h, x) = I^{(1)}_I \quad \text{and} \quad (y - h, y) = I^{(1)}_r
\]
\[
\tilde{I}_k = I^{(2)}_I \quad \text{and} \quad \tilde{I}_k = I^{(2)}_r
\]
and form recursively the intervals $(I^{(i)}_I, I^{(i)}_r)$ for all $i \in \{2, \ldots, K\}$ where $K$ is the value of $i$ at which
\[
|I^{(i)}_I \cap I^{(i)}_r| \geq \frac{1}{2}|I^{(i)}_I|
\]
is violated. At that point, we are done by the case \( k = 1 \). It is also possible that \( K = \infty \). Then, the process does not stop, and we have that
\[
|I_l^{(i)}(u_{l^{(i)}}) - u_{l^{(i)}}|^+ \to 0
\]
as \( i \to \infty \) by local integrability of \( u \) since \( |I_l^{(i)}| < |I_l^{(i-1)}| \) for all \( i \). There are also numbers \( \theta_i = |I_l^{(i)}| - |I_{l^{(i+1)}}| \geq 0 \) such that
\[
\sum_{i=1}^{K} \theta_i \leq h
\]
and
\[
h(u_{(x-h,x)} - u_{(y-h,y)})^+ \leq |I_k|(|I_l^{(k)} - u_{I_{k-1}}|^+ + (h - |I_k|)||u||_*)
\]
\[
= |I_k|(|I_l^{(k)} - u_{I_{k-1}}|^+ + \theta_1||u||_*)
\]
\[
\leq C||u||_* + \sum_{i=1}^{\infty} \theta_i
\]
\[
\leq Ch||u||_*.
\]
Consequently, we always have
\[
(u_{(x-h,x)} - U(y))^+ \leq C||u||_*
\]
for all \( y \in I^+ \). Taking the supremum over \( h \geq |I| \), we see that
\[
\frac{1}{|I|} \int_I (U_2 - U_{I^+}) \leq C||u||_*.
\]

**Small \( h \)**

Then we move to the part dealing with \( U_1 \). Now the maximal function averages over small intervals. To control the quantity
\[
\int_I (U_1 - u_{I^{+} \cup I^{-}})^+
\]
where \( I^2 = I + 2|I| \), we form a “one-sided” Calderón–Zygmund decomposition of \( u \) in \( I \cup I^+ \) where \( I^- = I - |I| \). More precisely, let \( \lambda > u_{I^+ \cup I^+} \). Take the maximal dyadic subintervals \( \{I_i\} \) of \( I \cup I^- \) with \( u_{I_i} > \lambda \). Let \( \hat{I}_i \) be the parent of \( I_i \). We write
\[
b_i = 1_{I_i}(u - u_{\hat{I}_i}^+)
\]
\[
b = \sum_i b_i
\]
\[
g = \sum_i 1_{I_i}u_{\hat{I}_i}^+ + 1_{I\setminus \cup_i I_i}u.
\]
Here, \( g \leq \lambda \). By the definition of \( U_1 \), we have that everywhere in \( I \) it holds that
\[
U_1 \leq M(1_{I^- \cup I} u) \leq M(b^+) + Mg.
\]
Here, $M$ is the standard two-sided Hardy–Littlewood maximal operator. Since $b_i$ have disjoint supports, we can use the John–Nirenberg inequality of [11] to estimate
\[
\int_{\mathbb{R}} (b^+)^2 = \sum_i \int_{I_i} (u - u_{I_i}^+)^2 \leq \sum_i |I_i| \int_{I_i} (u - u_{I_i}^+) \leq C |I| \|u\|^2_*.
\]

Denote $\Omega = I \cap \{ U_1 > u_{I+\cup J^2+} \}$. Then
\[
\int_{\Omega} U_1 \leq |\Omega|^{1/2} \|M(b^+)\|_{L^2} + |\Omega| \lambda
\leq C |I| \|u\|_* + |\Omega| \lambda,
\]
which proves that
\[
\int_I (U_1 - u_{I+\cup J^2+})^+ \leq C \|u\|_* |I|.
\]

Finally, for any interval $J$, we may write $J = I^- \cup I$ where $I^- = I - |I|$. Then
\[
\int_I (U_1 - u_J^+) = \int_{I^-} (U_1 - u_{I+\cup J^2+})^+ + \int_I (U_1 - u_{I+\cup J^2+})^+
= I + II
\]
where $II \leq C \|u\|_* |I|$ by the previous considerations. For the other part, we estimate
\[
\int_{I^-} (U_1 - u_{I+\cup J^2+})^+ \leq \int_{I^-} (U_1 - u_{I+\cup J^2+})^+ + |I|(u_{I+\cup J^2+} - u_{I+\cup J^2+})^+
\leq C \|u\|_* |I| + \frac{|I|}{2} (u_{I+\cup J^2+} - u_{I+\cup J^2+})^+
\leq C \|u\|_* |I| + \frac{|I|}{2} [(u_{I+\cup J^2+} - u_{I+\cup J^2+})^+] + (u_{I+\cup J^2+} - u_{I+\cup J^2+})^+
\leq (C + 1) \|u\|_* |I|.
\]

This proves the estimate for the part of $U_1$.

Putting the pieces together, we see that
\[
\int_I (U - U_I^+) \leq C \|u\|_* |I|
\]
for all intervals $I$ and a numerical $C$. The proof is complete. \qed

**Acknowledgements** The author would like to thank Juha Kinnunen for suggesting this problem. The author would also like to thank Ioannis Parissis for many valuable comments on an earlier version of this manuscript.

**References**

1. Aimar, H.: Elliptic and parabolic BMO and Harnack’s inequality. Trans. Am. Math. Soc. **306**, 265–276 (1988)
2. Aimar, H., Crescimbeni, R.: On one-sided BMO and Lipschitz functions. Ann. Sc. Norm. Super. Pisa **27**, 437–456 (1998)
3. Bennett, C., DeVore, R.A., Sharpley, R.: Weak $L^\infty$ and BMO. Ann. Math. **113**, 601–611 (1981)
4. Chiarenza, F., Frasca, M.: Morrey spaces and Hardy–Littlewood maximal function. Rend. Mat. Appl. **7**, 273–279 (1987)
5. Cruz-Uribe, D., Neugebauer, C.J.: The structure of the reverse Hölder classes. Trans. Am. Math. Soc. **347**, 2941–2960 (1995)
6. Forzani, L., Martín-Reyes, F.J., Ombrosi, S.: Weighted inequalities for the two-dimensional one-sided Hardy–Littlewood maximal function. Trans. Am. Math. Soc. 363, 1699–1719 (2011)
7. Fabes, E.B., Garofalo, N.: Parabolic BMO and Harnack’s inequality. Proc. Am. Math. Soc. 95, 63–69 (1985)
8. Kinnunen, J., Kuusi, T.: Local behaviour of solutions to doubly nonlinear parabolic equations. Math. Ann. 337, 705–728 (2007)
9. Kinnunen, J., Saari, O.: On weights satisfying parabolic Muckenhoupt conditions. Nonlinear Anal. 131, 289–299 (2016)
10. Kinnunen, J., Saari, O.: Parabolic weighted norm inequalities for partial differential equations. Anal. PDE 9, 1711–1736 (2016)
11. Martín-Reyes, F.J., de la Torre, A.: One-sided BMO spaces. J. Lond. Math. Soc. 49, 529–542 (1994)
12. Moser, J.: A Harnack inequality for parabolic differential equations. Commun. Pure Appl. Math. 17, 101–134 (1964)
13. Ombrosi, S.: Weak weighted inequalities for a dyadic one-sided maximal function in $\mathbb{R}^n$. Proc. Am. Math. Soc. 133, 1769–1775 (2005)
14. Saari, O.: Parabolic BMO and global integrability of supersolutions to doubly nonlinear parabolic equations. Rev. Mat. Iberoam. 32, 1001–1018 (2016)
15. Sawyer, E.: Weighted inequalities for the one-sided Hardy–Littlewood maximal functions. Trans. Am. Math. Soc. 297(1), 53–61 (1986)
16. Trudinger, N.S.: Pointwise estimates and quasilinear parabolic equations. Commun. Pure Appl. Math. 21, 205–226 (1968)