POINTWISE AND WEIGHTED HESSIAN ESTIMATES FOR KOLMOGOROV–FOKKER–PLANCK TYPE OPERATORS

ABHISHEK GHOSH AND VIVEK TEWARY

Abstract. In this article, we obtain hessian estimates for Kolmogorov–Fokker–Planck operators in non-divergence form in several Banach function spaces. Our approach relies on a representation formula and newly developed sparse domination techniques in Harmonic Analysis. Our result when restricted to weighted Lebesgue spaces yields sharp quantitative hessian estimates for the Kolmogorov–Fokker–Planck operators.

1. Introduction

In this article, we consider a class of Kolmogorov–Fokker–Planck type operator on \( \mathbb{R}^{N+1} \).

\[
\mathcal{L}u \equiv \sum_{i,j=1}^{s_0} a_{ij}(x,t) \partial_{x_i} \partial_{x_j} u + \sum_{i,j=1}^{N} b_{ij} x_i \partial_{x_j} u - \partial_t u = 0,
\]

where \( z = (x,t) \in \mathbb{R}^{N+1}, 1 \leq s_0 \leq N, \) and \( b_{ij} \) is constant for every \( i,j = 1, \ldots, N. \) The following assumptions on the coefficients of \( \mathcal{L} \):

**\( \text{(H1)} \):** \( a_{ij} = a_{ji} \in L^\infty(\mathbb{R}^{N+1}) \) and there exists a \( \lambda > 0 \) such that

\[
\frac{1}{\lambda} \sum_{i=1}^{s_0} \xi_i^2 \leq \sum_{i,j=1}^{s_0} a_{ij}(x,t) \xi_i \xi_j \leq \lambda \sum_{i=1}^{s_0} \xi_i^2
\]

for every \( (x,t) \in \mathbb{R}^{N+1}, \) and \( \xi \in \mathbb{R}^{s_0}. \)

**\( \text{(H2)} \):** The matrix \( B = (b_{ij})_{N \times N} \) has the form

\[
\begin{pmatrix}
0 & B_1 & 0 & \cdots & 0 \\
0 & 0 & B_2 & \cdots & 0 \\
& & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & B_d \\
0 & 0 & 0 & \cdots & 0
\end{pmatrix}
\]

where \( B_k \) is a matrix \( s_{k-1} \times s_k \) with rank \( s_k \) and \( s_0 \geq s_1 \geq \cdots \geq s_d, s_0 + s_1 + \cdots + s_d = N. \)

It is well known that under the assumptions **(H1)** and **(H2)**, for any fixed \( z_0 \in \mathbb{R}^{N+1} \) the “frozen” operator

\[
\mathcal{L}_0 \equiv \sum_{i,j=1}^{s_0} a_{ij}(z_0) \partial_{x_i} \partial_{x_j} u + \sum_{i,j=1}^{N} b_{ij} x_i \partial_{x_j} u - \partial_t u
\]

satisfies the well-known Hörmander’s hypoellipticity condition. A fundamental property of (1.2) is that the fundamental solution can be explicitly written down since the coefficients \( a_{ij} \) are constant. In fact, it is well known that hypoelliptic operators with smooth coefficients satisfying Hörmander’s hypoellipticity condition are locally solvable [Bra14]. This property has been useful, first in the Euclidean setting, to obtain the Hessian estimates on \( L^p \)-spaces for elliptic and parabolic equations.

2010 Mathematics Subject Classification. 35K65, 35K70, 35B45, 46E30

Key words and phrases. Kolmogorov–Fokker–Planck operators, generalized Orlicz space, weights.

Authors are supported by Centre for Applicable Mathematics, Tata Institute of Fundamental Research.
in non-divergence form by means of Harmonic Analysis techniques, particularly, the analysis of Calderón-Zygmund operators and commutators in the papers [CFL91, BC93] and then in the setting of Kolmogorov–Fokker–Planck operators [BCM96].

This present article addresses hessian estimates for Kolmogorov–Fokker–Planck type operators on several function spaces, for example, weighted $L^p$ spaces, generalized Orlicz spaces and variable Lebesgue spaces. To the best of our knowledge, hessian estimates for (1.1) are not known beyond $L^p$ spaces. The regularity theory of Kolmogorov–Fokker–Planck equations has become an area of extensive research recently, particularly due to its connections with kinetic equations that exhibit a variety of interesting physical and mathematical features [Sil22]. In the ultraparabolic setting, one needs to work with Sobolev spaces adapted to the appropriate vector fields associated to Kolmogorov–Fokker–Planck operators. Indeed, the hypoelliptic structure of the equation determines the directions in which the equation has regularity properties. Our point of departure are the papers [BCM96] and [PR98] where unweighted versions of these results were proved. Other related works are [BCLP13] in the non-divergence setting and [MP98] for the divergence type equations. Two recent works that study weighted versions of $L^p$ estimates for the kinetic Kolmogorov–Fokker–Planck equations are [DY21] and [NZ22].

An important thrust of this work is to employ new techniques in Harmonic Analysis in the context of Kolmogorov–Fokker–Planck operators to obtain pointwise estimates of the hessian. In the Euclidean setting, such techniques have been previously used in [BBD21, AB21]. Our work extends their methods in the following particulars:

- The setting of Kolmogorov–Fokker–Planck operators requires techniques specific to homogeneous spaces. In particular, much of our analysis would rely on the existence of Christ’s dyadic grids in metric spaces with doubling measures.
- Our focus is on obtaining sharp weighted estimates in Lebesgue and variable exponent Lebesgue spaces. See the discussion in Subsection 1.1.
- Our stress is also on the unified nature of the estimates. Hence, our pointwise estimates can give sharp results in more general Banach function spaces. See Subsection 1.2.

Now we present our results systematically. Our first result addresses sharp quantitative weighted estimates on weighted Lebesgue spaces for Kolmogorov–Fokker–Planck operators.

1.1 Sharp regularity estimates. Sharp weighted estimates for the gradient are always of immense interest in the context of partial differential equations, for example we refer the work of Astala–Iwaniec–Saksman [AIS01] where the authors have studied the connections between the quasi regularity of solutions of Beltrami equation with the sharp operator norm estimate of the Ahlfors–Beurling transform. Weighted estimates for the non-divergence elliptic equations were first addressed in [BL15] where they have proved $\|D^2 u\|_{L^p(\omega)} \lesssim \|f\|_{L^p(\omega)}$ for $2 < p < \infty$ and $\omega \in A_{p/2}(\mathbb{R}^n)$. Very recently, the estimate $\|D^2 u\|_{L^p(\omega)} \lesssim \|f\|_{L^p(\omega)}$ is improved for all $1 < p < \infty$ and all $\omega \in A_p(\mathbb{R}^n)$ by the authors in [BBD21]. They have employed the sparse domination technique from Harmonic analysis. Subsequently, the same techniques are employed to the parabolic setting in [AB21] to obtain regularity estimates on generalized Orlicz space. Previously, regularity estimates on generalized Orlicz spaces for elliptic equations of non-divergence type was obtained by Hästö and Ok in [HO19]. In this direction, we obtain the following sharp weighted regularity estimates on weighted $L^p$ spaces.

**Theorem 1.1.** Assume that the matrices $A$ and $B$ satisfy (H1) and (H2). There exists $\delta$ such that if $A$ is $(\delta, \emptyset)$-BMO in the sense of Definition 2.8 then for any $\Omega' \Subset \Omega \subset \mathbb{R}^{N+1}$, $1 < p < \infty$, and $\omega \in A_p(\text{Definition 5.1})$ we have the following estimate

$$
\sum_{i,j=1}^{q} \|u_{x_i x_j}\|_{L^p(\Omega', \omega)} \leq \kappa_{A,C,N,\delta}[\omega]_{A_p}^{\max\left\{\frac{1}{r-1},1\right\}} \left(\|\nabla u\|_{L^p(\Omega, \omega)} + \|u\|_{L^p(\Omega, \omega)}\right).
$$

(1.3)
for any $u \in S_p(\Omega, \mathcal{L})$ (Definition 2.9) such that $u$ is supported in $\Omega \cap \{z = (x, t) \in \mathbb{R}^{N+1} : t > 0\}$.

We would like to highlight certain key features of Theorem 1.1.

- The exponent $\max\left\{ \frac{1}{p-1}, 1 \right\}$ in (1.3) is sharp. Let us consider the following equation

$$\Delta u = f$$

on $\mathbb{R}^n$.

By means of Fourier transform we can write the hessian as $\nabla^2 u = (\mathcal{R} \otimes \mathcal{R})(f)$, where $\mathcal{R} = (R_1, \ldots, R_n)$ is the Riesz transform. Therefore, using estimates from [BB18], we obtain $\|\nabla^2 u\|_{L^p(\omega)} \lesssim \left[ \omega \right]_{\mathcal{A}_p(\mathbb{R}^n)}^\max\left\{ \frac{1}{p-1}, 1 \right\} \|f\|_{L^p(\omega)}$, which validates the sharpness of our result (1.3).

- Our techniques of the proof of Theorem 1.1 are also applicable for elliptic equations of non-divergence form or parabolic equations of non-divergence form. Therefore, the estimate (1.3) is new even for elliptic equations of non-divergence form or parabolic equations of non-divergence form.

Previous results in this directions are motivated by the seminal works [Caf90, CP98]. Most of these techniques heavily rely on the Fefferman-Stein sharp maximal function, small perturbation arguments, good-$\lambda$-type inequalities and reverse Hölder’s inequality(self-improving property) of Muckenhoupt weights. Therefore, it is really not clear how to obtain sharp estimates using these methods. On the other hand, our proof relies on the representation formula (3.13) developed in [BCM96]. We improve the representation formula (3.13) to (4.4), where $u_{x_ix_j}$ is pointwise dominated by appropriate sparse operators (see Definition 4.3). The primary usefulness of the sparse domination lies in the fact that it is easy to get quantitative weighted estimates for sparse operators which can be immediately passed on to the operator at hand. It should be noted that the exponent in (1.3) is same as that of the celebrated $A_2$ conjecture (now a Theorem [Hyt12]).

As mentioned above, our techniques also imply hessian estimates on several Banach function spaces. We state them more precisely.

1.2. Estimates on Banach function spaces. The study of regularity estimates for non-divergence type elliptic and parabolic equations on generalized Orlicz spaces and variable Lebesgue spaces has gained a lot of attention over past decades, due to its physical as well as mathematical importance (see [AM01], [BLO18], [BLO15], [HO19], [BW08], [Zhi86]) we believe our results, that is, Theorem 1.2 and Theorem 1.3 will be of serious interest.

**Theorem 1.2.** Suppose that the matrices $A$ and $B$ satisfy (H1) and (H2). There exists $\delta$ such that if $A$ is $(\delta, \mathcal{R})$-BMO in the sense of Definition 2.8 then for any $\Omega' \subset \Omega \subset \mathbb{R}^{N+1}$, there exists a positive constant $\kappa$ depending on $p, q, \alpha, \beta, \lambda, \Omega, \Omega'$, the matrix $B$ and $\mathcal{R}$ such that for any $u \in S_{p(\cdot)}(\Omega, d, \cdot | \cdot)$ with support in $\Omega \cap \{z = (x, t) \in \mathbb{R}^{N+1} : t > 0\}$, we have

$$\|u\|_{S_{p(\cdot)}(\Omega', d, | \cdot |)} \leq C \left( \|\mathcal{L}u\|_{L^p(\cdot)}(\Omega, d, | \cdot |) + \|u\|_{L^p(\cdot)}(\Omega, d, | \cdot |) \right)$$

where $L^p(\cdot)(\Omega, d, | \cdot |)$ is the generalized Orlicz space satisfying conditions (A0), (A1), (aInc$_p$), and (aDec$_q$). The definition of the Sobolev space $S_{p(\cdot)}(\Omega, d, | \cdot |)$ adapted to the ultraparabolic operator $\mathcal{L}$ may be found in Definition 2.9. The definitions regarding generalized Orlicz spaces may be found in Subsection 5.3.

**Theorem 1.3.** Suppose that the matrices $A$ and $B$ satisfy (H1) and (H2). There exists $\delta$ such that if $A$ is $(\delta, \mathcal{R})$-BMO in the sense of Definition 2.8 then for any $\Omega' \subset \Omega \subset \mathbb{R}^{N+1}$, there exists a positive constant $\kappa$ depending on $p(\cdot), \lambda, \Omega, \Omega'$, the matrix $B$ and $\mathcal{R}$ such that for any $u \in S_{p(\cdot)}(\Omega, \omega)$ with support in $\Omega \cap \{z = (x, t) \in \mathbb{R}^{N+1} : t > 0\}$, we have

$$\|u\|_{S_{p(\cdot)}(\Omega', \omega)} \leq \kappa_{A, B, N, \delta, \gamma_1, \gamma_2, \omega} \left( \|\mathcal{L}u\|_{L^p(\cdot)}(\Omega, \omega) + \|u\|_{L^p(\cdot)}(\Omega, \omega) \right)$$
where $L^{p(\cdot)}(\Omega, \omega)$ is the variable exponent Lebesgue space such that the variable exponent $p(\cdot)$ is globally log-Hölder continuous and $\omega \in A_{p(\cdot)}$. The definition of the Sobolev space $S^{p(\cdot)}(\Omega, \omega)$ adapted to the ultraparabolic operator $\mathcal{L}$ may be found in Definition 2.9. The definitions regarding weighted variable exponent Lebesgue spaces may be found in Subsection 5.2.

We write down the particulars of our approach below:

1. Using the fundamental solution of the frozen operator (1.2), we obtain a pointwise representation of the hessian $(u_{x_i x_j})$ for any test function (3.1) in terms of (i) a singular operator acting on $\mathcal{L}u$, and, (ii) a commutator of the same singular operator with the coefficients $a_{ij}$ acting on the hessian of $u$.

2. In a second step, the variable kernel singular operator is expanded in spherical harmonics ([CZ57], [BCM96]) to obtain a second representation formula (3.13).

3. At this point, our approach departs from that of [BCM96]. We first prove endpoint boundedness of the singular operators (weak-type $L^1$) and more importantly, the appropriate end-point boundedness of the associated grand truncated maximal operator.

4. We derive pointwise estimates for the singular integrals and the commutators in terms of some positive dyadic operators, known as “sparse operators”.

5. Combining the above, we obtain a further representation formula (4.4) for the hessian of $u$ in terms of sparse operators which forms our basis for a unified approach to gradient estimates. As mentioned in the work [AB21], this approach indeed improves all the known estimates and have further implications in the regularity theory of PDEs.

1.3. History of the problem. Kolmogorov–Fokker–Planck equations arise in a variety of physical models in mathematical finance [Bar97, AMP21] among other stochastic models, kinetic theory of gases [Lio94, AV04], clusters in space [Cha43], and image processing [Mum94, AZ03]. A variety of other applications are given in the books [Ris89, Hon94, GvH99].

On the mathematical side of things, Kolmogorov [Kol34] studied a simpler equation as a model for Brownian motion and wrote down its fundamental solution. Generalized models were considered in [Web51, Il’64]. These studies motivated Hörmander to build the notion of hypoellipticity in [H67] where he finds a sufficient condition for an operator $\mathcal{L}$ to be hypoelliptic in terms of a rank condition on the Lie Algebra generated by the distinct first order differential operators in $\mathcal{L}$ now interpreted as vector fields. Subsequently, Stein laid out a program of studying hypoelliptic operators by means of left invariant homogeneous operators on nilpotent graded Lie groups [Ste71] which resulted in such deep work as [RS76], [Fol75], and [NSW85]. Such analysis has found applications in spectral theory of Fokker-Planck type operators with connections to semiclassical analysis [HN05].

1.4. Regularity theory of Kolmogorov–Fokker–Planck equations. The regularity theory for equations in nondivergence form is well-developed in the case when the coefficients are Hölder continuous. In the particular case of Kolmogorov–Fokker–Planck operators, fundamental solution is constructed by Levi’s method of parametrices in [Pol94]. A solution to Dirichlet problem in the Perron-Wiener sense as well as Schauder estimates are obtained in [Man97]. Cauchy problem is studied in [Ter05]. Schauder estimates, Harnack inequality and Gaussian lower bounds are studied in [DFP06]. For less regular coefficients, the regularity theory is much less well developed. For example, the Krylov-Safanov theory for measurable coefficients is missing. Recently, Harnack’s inequality has been proved for less regular coefficients in [AT19].

The Calderón-Zygmund theory has had an important contribution to the $W^{2,p}$-existence theory for equations in nondivergence form. An early work consisting of an existence result involving discontinuous coefficients was the work of Miranda [Mir63] where solutions in $W^{2,2} \cap W_0^{1,2}$ were found for coefficients in $W^{1,1}$. This work can be strengthened by asking for derivatives of coefficients to belong to weak $L^N$. With the advent of viscosity solutions, Caffarelli extended the notion to
$C^{1,\alpha}, C^{2,\alpha}$ and $W^{2,p}$-versions in [Caf89]. However, these works were restricted to $p > N - \varepsilon$ due to the application of the Aleksandrov-Bakel’man-Pucci maximum principle. In the important work of Chiarenza et al [CFL91, CFL93], $W^{2,p}$-existence results were proved for the full range $1 < p < \infty$ through a representation formula. Indeed these methods were generalized to parabolic equations in [BC93] and for ultraparabolic operators in [BCM96], from where we borrow the representation formula. Krylov [Kry07] removed the dependence on the representation formula by exploiting the sharp maximal function of the second derivatives which allows for generalizations to fully nonlinear settings. Other than these approaches, there is also, what is called as, the maximum function free approach of Acerbi and Mingione [AM07] which has been developed for equations in non-divergence form in [Yao12].

The results concerning hypoelliptic operators are local in nature, for example, see [BCM96, BZ13]. An analysis at the boundary would require the development of the notion of characteristic points in this context since boundary regularity is expected to fail at such points as noticed by Jerison [Jer81a, Jer81b] for other models.

On the other hand, regularity theory for divergence-form variants of Kolmogorov–Fokker–Planck equations is much well-developed. We point to the survey [AP20] for references to these works. For boundary behaviour of non-negative solutions of Kolmogorov–Fokker–Planck we refer the foundational works [CNP10, CNP12, NP16]. A recent contribution to a nonlinear Kolmogorov–Fokker–Planck model is [GN22].

2. Preliminaries and Notations

2.1. Geometry of the problem. We recall the following translation and dilation structures in the space $\mathbb{R}^{N+1}$ intrinsic to the constant matrix $B$.

**Definition 2.1** (The group of translations). Let $F(\tau) = \exp(-\tau B^T)$. Since the matrix $B$ is a nilpotent matrix, $F(\tau)$ is a polynomial of degree $B$ in $\tau$. For $(x,t), (y,\tau) \in \mathbb{R}^{N+1}$, set

$$(x,t) \circ (y,\tau) = (y + F(\tau)x, t + \tau).$$

Then $(\mathbb{R}^{N+1}, \circ)$ is a group with identity element $(0,0)$; the inverse of an element $(x,t)$ is $(x,t)^{-1} = (-F(-t)x, -t)$. The left translation by $z'$ given by

$$z \mapsto z' \circ z,$$

is a invariant translation to operator $\mathcal{L}_0$.

**Definition 2.2** (Dilations). The compatible dilation to operator $\mathcal{L}_0$ is given by

$$\delta_r = \text{diag}(rI_{s_0}, r^3 I_{s_1}, \ldots, r^{2d+1} I_{s_d}, r^2),$$

where $I_{s_k}$ denotes the $s_k \times s_k$ identity matrix. With respect to this dilation the operator $\mathcal{L}_0$ satisfies the following

$$\mathcal{L}_0(u(\delta_r)) = r^2 \delta_r(\mathcal{L}_0(u)),$$

therefore, $\mathcal{L}_0$ is a homogeneous operator of degree 2. Let

$$Q = s_0 + 3s_1 + \cdots + (2d + 1)s_d.$$

The quantity $Q + 2$ is known as the homogeneous dimension of $\mathbb{R}^{N+1}$ with respect to the dilation $\delta_r$.

**Definition 2.3** (Homogeneous norm). The norm in $\mathbb{R}^{N+1}$, related to the group of translations and dilation to the equation is defined by

$$\||(x,t)|| = \rho$$
Lemma 2.4 is called a space of order \( \beta \) if \( \rho \) is the unique positive solution to the equation

\[
\frac{x_1^2}{\rho^{2\alpha_1}} + \frac{x_2^2}{\rho^{2\alpha_2}} + \cdots + \frac{x_N^2}{\rho^{2\alpha_N}} + \frac{t^2}{\rho^4} = 1,
\]

where \( (x, t) \in \mathbb{R}^{N+1} \setminus \{0\} \) and

\[
a_1 = \cdots = a_{s_0} = 1, \quad a_{s_0+1} = \cdots = a_{s_0+s_1} = 3, \cdots,
\]

\[
a_{s_0+\cdots+s_{d-1}+1} = \cdots = a_N = 2d + 1,
\]

and \( \| (0, 0) \| = 0 \).

The homogeneous norm defined above satisfies the following property.

**Lemma 2.4.** ([BCM96, Prop. 1.3]) There exists \( \beta = \beta(d) \in (0, 1] \), and constants \( \tilde{K}, M \) depending on the matrix \( B \) such that for every \( z, \eta, \zeta \in \mathbb{R}^{N+1} \) satisfying \( \| \eta^{-1} \circ z \| \geq M \| \zeta^{-1} \circ z \| \), it holds that

\[
\| \eta^{-1} \circ z - \eta^{-1} \circ \zeta \| \leq \tilde{K} \| \zeta^{-1} \circ z \| \beta \| \eta^{-1} \circ z \|^{1-\beta},
\]

(2.1)

\[
\| z^{-1} \circ \eta - \zeta^{-1} \circ \eta \| \leq \tilde{K} \| \zeta^{-1} \circ z \| \beta \| \eta^{-1} \circ z \|^{1-\beta},
\]

(2.2)

A homogeneous space satisfying Lemma 2.4 is called a space of order \( \beta \).

**Definition 2.5** (Homogeneous metric). Let \( d \) denotes the metric induced by the norm \( \| \cdot \| \), i.e., \( d(z, \xi) = \| \xi^{-1} \circ z \| \), then \( d \) is a quasimetric. We refer Proposition 1.3 in [BCM96] for more properties of \( d \). Also, we remark that the metric \( d \) is not symmetric but quasi-symmetric i.e. there exist a constant \( \tilde{D} \) depending on the matrix \( B \) such that

\[
\frac{1}{\tilde{D}} d(z, \xi) \leq \tilde{d}(\xi, z) \leq \tilde{D} d(z, \xi).
\]

We define the new metric, \( \tilde{d}(z, \xi) = d(z, \xi) + d(\xi, z) \), which is comparable to the metric \( d \) and is symmetric. Therefore, here onwards we assume the metric \( d \) to be symmetric.

Therefore the space \( (\mathbb{R}^{N+1}, d, |.|) \) is a metric space of homogeneous type in the sense of Coifman and Weiss (see [CW71]).

**Definition 2.6.** A ball centered at \( z = (x, t) \in \mathbb{R}^{N+1} \) is defined as

\[
B_r(z) = \{ w \in \mathbb{R}^{N+1} : d(z, w) < r \}.
\]

**Definition 2.7** (BMO spaces). A function \( b \in L^1_{loc}(\mathbb{R}^{N+1}) \) is said to be in BMO if

\[
\| b \|_{BMO} := \sup \left\{ \frac{1}{|B|} \int_{B} |b(z) - (b)_B| \, dz : B \text{ is a ball in } \mathbb{R}^{N+1} \right\} < \infty,
\]

where \( (b)_B \) denotes the mean of \( b \) over the set \( \mathcal{B} \), that is

\[
(b)_B := \int_{B} b(z) \, dz = \frac{1}{|B|} \int_{B} b(z) \, dz,
\]

and \( L^1_{loc}(\mathbb{R}^{N+1}) \) has the usual meaning of the function space of locally integrable functions with respect to the Lebesgue measure. We also use the notation \( \Omega_B(b) \) to denote the oscillation \( \frac{1}{|B|} \int_{B} |b(z) - (b)_B| \).
Definition 2.8. For given \( \mathcal{R} > 0 \) and \( \delta > 0 \), we say that a matrix \( A \) whose entries are \( L^1_{\text{loc}}(\mathbb{R}^{N+1}) \) functions is \((\delta, \mathcal{R})-\text{BMO}\) if

\[
\sup \left\{ \int_{B_\rho} |A(z) - (A)_\rho| \, dz : B_\rho \text{ is a ball of radius } \rho \text{ in } \mathbb{R}^{N+1}, 0 < \rho < \mathcal{R} \right\} < \delta,
\]

where \((A)_\rho\) denotes the mean of \( A \) over a ball \( B_\rho \) of radius \( \rho \), that is

\[
(A)_\rho := \int_{B_\rho} A(z) \, dz = \frac{1}{|B_\rho|} \int_{B_\rho} A(z) \, dz.
\]

Definition 2.9. For \( \Omega \subset \mathbb{R}^{N+1}, \) an open set, we define a class of Sobolev-type space

\[
S^X(\Omega, \mathcal{L}) = \{ u \in X(\Omega) : u_{x_i}, u_{x_i x_j}, Y u \in X(\Omega), i, j = 1, 2, \ldots, s_0 \},
\]

where \( Y u = \left( \sum_{i=1}^{N} b_{ij} x_i \partial x_j - \partial_t \right) u \) and \( X(\Omega) \) is one of the following Banach function spaces:

\[
X(\Omega) = L^p_w(\Omega) \text{ or } L^p(\Omega),
\]

whose definitions are given in Subsection 5.2 and Subsection 5.3 respectively. On the space \( S^X(\Omega, \mathcal{L}) \), we define the norm

\[
||u||_{S^X(\Omega, \mathcal{L})} = ||u||_{X(\Omega)} + \sum_{i=1}^{s_0} ||u_{x_i}||_{X(\Omega)} + \sum_{i,j=1}^{s_0} ||u_{x_i x_j}||_{X(\Omega)} + ||Y u||^p_{X(\Omega)}.
\]

We may also sometimes use the notation \( S^X(\Omega, d, | \cdot |) \) or \( S^X(\Omega) \). In the specific case of \( X = L^p_w(\Omega) \), we denote \( S^X(\Omega, \mathcal{L}) \) by \( S^p_w(\Omega) \) and when \( X = L^p(\Omega) \), we denote \( S^X(\Omega, \mathcal{L}) \) by \( S^p(\Omega) \).

A complete exposition of Caldéron-Zygmund theory on homogeneous spaces is given in the book of Deng and Han [DH09]. For our purposes, we shall require the Caldéron-Zygmund type decomposition adapted to \((\mathbb{R}^{N+1}, d, | \cdot |)\).

Theorem 2.10. (Caldéron-Zygmund type lemma [DH09, Aim85]) Let \( f \) be a nonnegative integrable function and \( \lambda > 0 \) be any height. Then there exist functions \( h_i \) and balls \( B_i \) such that the sets \( S_i = \{ x \in \mathbb{R}^{N+1} : h_i(x) \neq 0 \} \) are pairwise disjoint and

\[
g = f - \sum_i h_i \in L^1 \cap L^\infty \text{ and } ||g||_{L^\infty} \leq \tilde{D} \lambda,
\]

\[
\int_{\mathbb{R}^{N+1}} h_i(x) \, dx = 0,
\]

\[
S_i \subset B_i, \text{ and } \sum_i |B_i| \leq \tilde{D} \lambda^{-1} ||f||_{L^1},
\]

where the constant \( \tilde{D} \) only depends on the homogeneous space.

3. Representation Formula

3.1. Fundamental solution. The operator \( \mathcal{L}_0 \) is hypoelliptic, for example, see [H67]. Let \( \Gamma_0 \) denotes its fundamental solution with pole at zero. We recall the following expression for \( \Gamma_0 \) as derived in [H67, Kup72, LP94].

Let \( A(z) = (a_{ij}(z))_{s_0 \times s_0} \) denote the \( s_0 \times s_0 \) matrix that features in (1.1). Let \( A_0 \) denote the \( N \times N \) constant matrix defined as

\[
A_0 := \begin{pmatrix} A(z_0) & 0 \\ 0 & 0 \end{pmatrix},
\]
where each $\hat{0}$ represents a block matrix with zeros. Recall the definiton of the matrix polynomial $F(s)$ from Definition 2.1. Let

$$C(\tau, z_0) = \int_0^\tau F(s)A_0F^t(s) \, ds.$$ 

Then, the fundamental solution of $L_0$ with pole at zero is

$$\Gamma_0(z) = \Gamma_0(x, t) = \frac{1}{(4\pi)^{\frac{N}{2}}} \exp \left( -\frac{1}{4}(C^{-1}(t, z_0)x, x) \right).$$

The fundamental solution $\Gamma_0(\cdot, \zeta)$ of $L_0$ with pole in $\zeta \in \mathbb{R}^{N+1}$ is defined as

$$\Gamma_0(z, \zeta) = \Gamma_0(\zeta^{-1} \circ z), \quad z, \zeta \in \mathbb{R}^{N+1}, \quad z \neq \zeta.$$ 

We collect here all the relevant properties of $\Gamma$ as proved in [BCM96].

**Theorem 3.1.** ([BCM96, Theorem 2.2]) Let $z_0 \in \mathbb{R}^{N+1}$ and define $\Gamma(z_0, \cdot) \equiv \Gamma_0(\cdot)$. Then

(i) $\Gamma^0(\cdot) \in C^\infty(\mathbb{R}^{N+1}\setminus\{0\}).$

(ii) $\Gamma^i(\cdot)$ is $\delta_0$-homogeneous of degree $-Q$.

(iii) The function $\Gamma^i = \partial_z \Gamma^i$ is $\delta_0$-homogeneous of degree $-Q - a_i$ for $i = 1, 2, \ldots, N + 1$.

(iv) The function $\Gamma_{ij} = \partial_{x_i x_j} \Gamma^i$ is $\delta_0$-homogeneous of degree $-Q - a_i - a_j$ for $i = 1, 2, \ldots, N + 1$.

(v) The following estimates hold

$$|\Gamma^0_i(z)| \leq \frac{C}{|z|^{Q + a_i}},$$

$$|\Gamma^0_{ij}(z)| \leq \frac{C}{|z|^{Q + a_i + a_j}}, \text{ for } i, j = 1, 2, \ldots, N + 1,$$

where $C = \max \left\{ \sup_{S^N} |\Gamma^0_i(z)|, \sup_{S^N} |\Gamma^0_{ij}(z)|, i, j = 1, 2, \ldots, N + 1 \right\}$.

(vi) For every $m \in \mathbb{N}$, $z \in \mathbb{R}^{N+1}$ and a multiindex $\beta$, it holds that

$$\sup_{||z|| = 1, ||\beta|| = 2m} \left| \partial_\zeta^\beta \Gamma^0_{ij}(z; \zeta) \right| \leq C(m, \mu, B),$$

for all $z \in \mathbb{R}^{N+1}$.

(vii) It holds that

$$\int_{a < ||z|| < b} \Gamma^0_{ij}(z) \, d\zeta = 0 = \int_{||\zeta|| = 1} \Gamma^0_{ij}(z) \, d\sigma(z),$$

for $i, j = 1, 2, \ldots, s_0$ and $0 < a < b$.

The second order derivatives of smooth $u$ have the following pointwise representation.

**Theorem 3.2.** ([BCM96, Theorem 2.4]) Let $u \in C^\infty_0(\mathbb{R}^{N+1})$, $u = 0$ for $t \leq 0$, $z \in \text{spt} u$. Then, it holds that for $i, j = 1, 2, \ldots, s_0$

$$u_{x_i x_j}(z) = -pv \int_{\mathbb{R}^{N+1}} \Gamma_{ij}(z; \zeta^{-1} \circ z) \times \left( \sum_{h,k=1}^{s_0} [a_{hk}(z) - a_{hk}(\zeta)]u_{x_h x_k}(\zeta) + \mathcal{L}u(\zeta) \right) \, d\zeta$$

$$- \mathcal{L}u(z) \cdot \int_{||\zeta|| = 1} \Gamma_{ij}(z; \zeta) \nu_i \, d\sigma(\zeta),$$

(3.1)

where $\nu_i$ is the $i^{th}$ component of the outer normal to the surface $S^N$, the sphere in $\mathbb{R}^{N+1}$.
We may write the expression in (3.1) in a more tractable manner as follows. We define the following:

\[ K_{ij} f(z) = \text{pv} \int_{\mathbb{R}^{N+1}} \Gamma_{ij}(z;\zeta^{-1} \circ z) f(\zeta) \, d\zeta, \]

(3.2)

\[ \alpha_{ij}(z) = \int_{||\zeta||=1} \Gamma_j(z;\zeta) \nu_i \, d\sigma(\zeta), \quad \text{and} \]

for an operator \( K \) and a function \( a \in L^\infty(\mathbb{R}^{N+1}) \), define the commutator as

\[ [a,K](f) = K(af) - a \cdot K(f). \]

(3.3)

Using the notations (3.2) and (3.3) in (3.1), we obtain the following compact expression for the second order derivatives of \( u \).

\[ u_{x_ix_j} = -K_{ij}(\mathcal{L}u) + \sum_{h,k=1}^{s_0} [a_{hk}, K_{ij}](u_{x_hx_k}) + \alpha_{ij} \cdot \mathcal{L}u, \]

(3.4)

for \( i,j = 1,2,\ldots,s_0 \).

In order to study the \( L^p \) boundedness properties of the operators \( K_{ij} \) and \( [a_{hk}, K_{ij}] \), we will make use of spherical harmonics to obtain a Fourier series representation for the singular kernel \( \Gamma_{ij} \). This technique was first used in the classic paper of Calderón and Zygmund [CZ57]. In the case of anisotropic homogeneities, spherical harmonics found use in [FR66].

**Definition 3.3. (Spherical Harmonics)** A harmonic polynomial in \( \mathbb{R}^{N+1} \) which is homogeneous of degree \( m \) is called a solid harmonic of degree \( m \). The restriction of a solid harmonic to a sphere \( S^N \) is called a spherical harmonic of degree \( m \).

The dimension of the space of spherical harmonics of degree \( m \) is

\[ g_m = \binom{m+N}{N} - \binom{m+N-2}{N} \leq C_N \cdot m^{N-1} \]

(3.5)

The orthonormal system of spherical harmonics \( \{Y_{km}\}_{k=1,\ldots,g_m} \) is complete in \( L^2(S^N) \).

We list below some properties of the spherical harmonics which may be found in [CZ57] and [FR66]:

1. We have

\[ \left| \left( \frac{\partial}{\partial x} \right)^\beta Y_{km}(x) \right| \leq C_N \cdot m^{\frac{N-1}{2}+|\beta|}, \quad \text{for } x \in S^N, k = 1,2,\ldots,g_m. \]

2. Let \( f \in C^\infty(S^N) \). Let \( f(x) \sim \sum_{k,m} b_{km} Y_{km} \) be its expansion in Fourier series, so that \( b_{km} = \int_{S^N} f(x) Y_{km}(x) \, d\sigma \), then for every \( l \in \mathbb{N} \), there is a constant \( c \) depending on \( l \) such that

\[ |b_{km}| \leq c \cdot m^{-2l} \sup_{|\beta|=2l} \left| \left( \frac{\partial}{\partial x} \right)^\beta f(x) \right|. \]

3. For a fixed \( z \in \mathbb{R}^{N+1} \) and \( \zeta \in S^N \), we obtain the Fourier series expansion

\[ \Gamma_{ij}(z;\zeta) = \sum_{m=0}^{\infty} \sum_{k=1}^{g_m} \kappa_{ij}^{km}(z) Y_{km}(\zeta), \]
and then for a general $\zeta \in \mathbb{R}^{N+1}$, using the dilation $\zeta' = D(||\zeta||^{-1})\zeta$ and exploiting the homogeneity property of $\Gamma_{ij}$ we receive

$\Gamma_{ij}(z; \zeta) = \sum_{m=0}^{\infty} \sum_{k=1}^{g_m} c_{ij}^{km}(z) \frac{Y_{km}(\zeta')}{||\zeta||^{Q+2}}.$

(4) $c_{ij}^{km} = 0$ for $m = 0$.

(5) For $m > 0$, we have the bound

$$|c_{ij}^{km}(z)| \leq C \cdot m^{-2l}$$

for any $l > 1, z \in \mathbb{R}^{N+1}$.

Now, define

$$K_{km}(z) = \frac{Y_{km}(z')}{||z||^{Q+2}}.$$

Theorem 3.4. ([BCM96]) The kernels $K_{km}$ satisfy the following properties:

(1) $K_{km} \in C^\infty(\mathbb{R}^{N+1} \setminus \{0\})$.

(2) $K_{km}$ is $\delta_r$-homogeneous of degree $-(Q + 2)$.

(3) $K_{km}$ satisfies the following growth condition, for any $z, \zeta \in \mathbb{R}^{N+1} \setminus \{0\}$,

$$|K_{km}(z)| \leq \frac{c_{km}}{||z||^{Q+2}}$$

with $c_{km} \leq C_N \cdot m^{\frac{N-1}{2}}$.

(4) $K_{km}$ satisfies the following vanishing property

$$\int_{||\zeta||=1} K_{km}(\zeta) d\sigma(\zeta) = 0.$$

We shall also require the following property of $K_{km}$.

Theorem 3.5. (Hörmander’s condition [BCM96, Proposition 3.4]) There exists $\beta = \beta(r) \in (0, 1]$, $M = M(B) > 1$, $c_{km} = c(N) \cdot m^{\frac{N+1}{2}}$ such that

$$|K_{km}(\eta^{-1} \circ \zeta) - K_{km}(\eta^{-1} \circ z)| \leq c_{km} \frac{||\zeta^{-1} \circ z||^{\beta}}{||\eta^{-1} \circ z||^{Q+2+\beta}},$$

and

$$|K_{km}(\zeta^{-1} \circ \eta) - K_{km}(\zeta^{-1} \circ \eta)| \leq c_{km} \frac{||\zeta^{-1} \circ z||^{\beta}}{||\eta^{-1} \circ z||^{Q+2+\beta}},$$

for every $z, \zeta, \eta \in \mathbb{R}^{N+1}$ satisfying $||\eta^{-1} \circ z|| \geq M ||\zeta^{-1} \circ z||$.

Remark 3.6. By an abuse of notation, we use the same factor $c_{km}$ in (3.8) and Theorem 3.5 since in our estimates, we will always majorize by the larger number.

Now, we define a further singular operator

$$T_{km} f(z) = \text{p.v.} \int_{\mathbb{R}^{N+1}} K_{km}(\zeta^{-1} \circ z) f(\zeta) d\zeta,$$

so that (3.4) unfolds to

$$u_{x_i x_j} = -\sum_{m=1}^{\infty} \sum_{k=1}^{g_m} c_{ij}^{km} T_{km}(Lu) + \sum_{h,k=1}^{s_0} \sum_{m=1}^{\infty} \sum_{k=1}^{g_m} c_{ij}^{km} [a_{hk}, T_{km}] (u_{x_h x_k}) + \alpha_{ij} \cdot Lu,$$

for $i, j = 1, 2, \ldots, s_0$. 
4. Pointwise estimates

4.1. Sparse Domination Preliminaries. We begin with the uncentered Hardy-Littlewood maximal function is defined as follows

\[ \mathcal{M}f(\zeta) = \sup_{B \ni \zeta} \int_B |f(z)| \, dz, \]

where the supremum is over all balls \( B \) containing the points \( \zeta \). Another useful variant of the Hardy-Littlewood maximal function is the dyadic Hardy-Littlewood maximal function and to define that we need the following notion of dyadic grids for homogeneous spaces, see [HK12].

Let \( 0 < c_1 \leq C_1 < \infty \) and \( \delta \in (0, 1) \). By a general dyadic grid \( \mathcal{U} = \bigcup_{k \in \mathbb{Z}} \mathcal{U}_k \) on \( \mathbb{R}^{N+1} \), we mean a countable collection of sets \( \mathcal{U}_k^\alpha \) for \( k \in \mathbb{Z} \), each associated with a point \( z_k^\alpha \), \( \alpha \) coming from a countable index, with the following properties:

- \( \mathbb{R}^{N+1} = \bigcup \mathcal{U}_k^\alpha \) for every \( k \in \mathbb{Z} \).
- If \( l \geq k \), then either \( \mathcal{U}_l^\beta \subset \mathcal{U}_k^\alpha \) or \( \mathcal{U}_l^\beta \cap \mathcal{U}_k^\alpha = \emptyset \).
- There exist \( c_1, C_1 > 0 \) such that \( B(z_k^\alpha, c_1 \delta^k) \subset B(z_k^\alpha, C_1 \delta^k) := B(z_k^\alpha) \).
- If \( l \geq k \) and \( \mathcal{U}_l^\beta \subset \mathcal{U}_k^\alpha \), then \( B(z_l^\beta) \subset B(z_k^\alpha) \).
- For all \( k \in \mathbb{Z} \), there exists \( \alpha \) such that \( x_0 = z_k^\alpha \).

Hytönen and Kairema [HK12, Theorem 4.1] proved the existence of a finite collection of dyadic grids \( \mathcal{U}_t \), \( t = 1, 2, \ldots, L \) such that for every ball \( B(z, r) \subset \mathbb{R}^{N+1} \) with \( \delta^{k+2} \leq r < \delta^{k+1} \), there exists some \( t \in \{1, 2, \ldots, L\} \) and \( \mathcal{U}_k^\alpha \subset \mathcal{U}_t \) such that \( B(z, r) \subset \mathcal{U}_k^\alpha \) and \( \text{diam} \mathcal{U}_k^\alpha \leq C r \), where \( C \) depends on \( \delta \).

Corresponding to a dyadic grid \( \mathcal{U} \), the dyadic Hardy-Littlewood maximal function is defined as follows:

\[ \mathcal{M}_\mathcal{U}f(\zeta) = \sup_{P \ni \zeta; P \in \mathcal{U}} \int_P |f(z)| \, dz. \]

It is obvious that \( \mathcal{M}_\mathcal{U} \) is pointwise controlled by the Hardy-Littlewood maximal function \( \mathcal{M} \), but more interestingly, as proved in [HK12, Prop. 7.9], we have

\[ (4.1) \quad \mathcal{M}f \lesssim \sum_{t=1}^L \mathcal{M}_{\mathcal{U}_t}f. \]

In the rest of the paper, an element of a dyadic grid as above will be called as an ultraparabolic dyadic cube \( P \in \mathcal{U} \). We reserve the notation \( \mathcal{U}(P) \) for the collection of all dyadic cubes \( T \in \mathcal{U} \) such that \( T \subset P \) and \( \mathcal{M}_P f(\zeta) := \sup_{T \ni \zeta; T \in \mathcal{U}(P)} \int_T |f| \) is the restricted Hardy-Littlewood maximal function.

Definition 4.1 (Sparse family). Let \( \mathcal{G} \) be a family of cubes belonging to some general dyadic grid and \( \eta \in (0, 1) \). We call \( \mathcal{G} \) to be a \( \eta \)-sparse family if for every \( P \in \mathcal{G} \) there exists a set \( F_P \) such that \( |F_P| \geq \eta |P| \) and \( \{F_P\}_{P \in \mathcal{G}} \) are mutually disjoint.

Remark 4.2. In application we shall work with collection of ultraparabolic dyadic cubes \( \mathcal{G} \subset \mathcal{U} \) such that for any \( T \in \mathcal{G} \) we have

\[ (4.2) \quad \sum_{T' \in \mathcal{U}, T' \in \mathcal{U}(T)} |T'| \leq \frac{1}{2} |T|. \]

The above condition ensures that the collection \( \mathcal{G} \) is sparse since for each \( T \in \mathcal{G} \) if we define

\[ F_T = T \setminus \bigcup_{T' \in \mathcal{U}(T)} T', \]
then by the dyadic structure of \( \mathcal{U} \), the collection \( \{ F_T \}_{T} \) is pairwise disjoint and (4.2) implies
\[
|F_T| \geq |T| - \sum_{T' \in \mathcal{U}, T' \supseteq T} |T'| \geq \frac{1}{2}|T|.
\]

**Definition 4.3** (Sparse operator). Given a sparse family \( \mathcal{G} \), the associated sparse operator \( \mathcal{A}_G \) is defined as follows
\[
\mathcal{A}_G f(x) := \sum_{P \in \mathcal{G}} \left( \frac{1}{|P|} \int_P |f| \right) \chi_P(x).
\]

We also introduce the following sparse operators in order to estimate the commutator terms appearing in (3.13). Corresponding to a sparse family \( \mathcal{G} \) and a function \( b \), the operator \( \mathcal{A}^b_{\mathcal{G}} \) is defined as follows
\[
\mathcal{A}^b_{\mathcal{G}} f(x) := \sum_{P \in \mathcal{G}} |b(x) - b_P| \left( \frac{1}{|P|} \int_P |f| \right) \chi_P(x).
\]

The adjoint of \( \mathcal{A}^b_{\mathcal{G}} \) is given by
\[
\mathcal{A}^{b,x}_{\mathcal{G}} f(x) := \sum_{P \in \mathcal{G}} \left( \frac{1}{|P|} \int_P |b - b_P||f| \right) \chi_P(x).
\]

We let \( \mathcal{M} \) stand for the uncentered Hardy–Littlewood maximal function on \((\mathbb{R}^{N+1}, d, |\cdot|)\) and for \( 1 < r < \infty \), \( \mathcal{M}_r f := (\mathcal{M}(|f|^r))^{1/r} \). Our proofs on sparse domination depend on a general sparse domination principle by Lerner–Ombrosi, see Theorem 1.1 in [LO20]. For any linear operator \( T \), let us consider the following version of the grand maximal truncated operator: For \( s > 0 \)
\[
\mathcal{M}^T_{r,s} f(x) = \sup_{B: B \ni x} \text{ess sup}_{y,z \in B} \left| T(f \chi_{\mathbb{R}^{N+1}\setminus sB})(y) - T(f \chi_{\mathbb{R}^{N+1}\setminus sB})(z) \right|,
\]
where the supremum is taken over all balls \( B \) containing the point \( x \).

Let us state the following pointwise result which is even new in the context of elliptic and parabolic equations.

**Theorem 4.4.** Suppose that the matrices \( A \) and \( C \) satisfy (H1) and (H2). Fix \( i, j \in \{1, 2, \ldots, s_0\} \). Then for any ultraparabolic cube \( P \subset \Omega \), there exist sparse families \( \{ \mathcal{G}_{km,l} : k = 1, \ldots, g_m, 1 \leq l \leq C(N), m \in \mathbb{N} \} \) such that for any \( u \in \mathcal{S}^X(\Omega, \mathcal{L}) \), we have
\[
|u_{x_i x_j}(z)| \leq \kappa_{A,N,C} \sum_{m=1}^{s_0} \sum_{k=1}^{g_m} c_{km} |c_{ij}^{km}| \sum_{l=1}^{C(N)} \mathcal{A}_{\mathcal{G}_{km,l}}(\mathcal{L} u)(z)
\]
\[
+ \sum_{h,k=1}^{s_0} \sum_{m=1}^{g_m} c_{km} |c_{ij}^{km}| \sum_{l=1}^{C(N)} \mathcal{A}^{a_{hk}}_{\mathcal{G}_{km,l}}(u_{x_h x_k})(z)
\]
\[
+ \sum_{h,k=1}^{s_0} \sum_{m=1}^{g_m} c_{km} |c_{ij}^{km}| \sum_{l=1}^{C(N)} \mathcal{A}^{a_{hk},x}_{\mathcal{G}_{km,l}}(u_{x_h x_k})(z) + |a_{ij}| \cdot |\mathcal{L} u(z)|,
\]
for a.e. \( z \in P \).

To conclude the proof of the above theorem we need to individually handle each term appearing in (3.13). For the rest of this section fix some \( k, m \) such that \( m \in \mathbb{N} \) and \( k = 1, \cdots, g_m \).
4.2. **End-point boundedness.** In order to kickstart the sparse domination method, we will require a weak-$L^1$ type estimate for the operator $T_{km}$. This, in turn, requires the $L^2$ boundedness of the operator $T_{km}$. The proof of $L^2$ boundedness can be obtained from the proof given in Aimar [Aim85].

**Theorem 4.5.** [BCM96, Theorem 3.6] *(L$^2$ boundedness of $T_{km}$)* Let us define

$$T_{km}f(z) = \text{p.v.} \int_{\mathbb{R}^{N+1}} K_{km}(\zeta^{-1} \circ z) f(\zeta) \, d\zeta.$$ 

Then

$$||T_{km}f||_{L^2} \leq c_{km} ||f||_{L^2},$$

where $c_{km}$ is the constant that appears in Theorem 3.5.

**Theorem 4.6.** *(Weak-$L^1$ type estimates for $T_{km}$)* Let us define

$$T_{km}f(z) = \text{p.v.} \int_{\mathbb{R}^{N+1}} K_{km}(\zeta^{-1} \circ z) f(\zeta) \, d\zeta.$$ 

Then $T_{km}$ maps $L^1(\mathbb{R}^{N+1}, d, | \cdot |)$ to $L^{1,\infty}(\mathbb{R}^{N+1}, d, | \cdot |)$ boundedly, that is, for all $\lambda > 0$

$$|\{z \in \mathbb{R}^{N+1} : |T_{km}f(z)| > \lambda\}| \leq \frac{c_{km}}{\lambda} \int_{\mathbb{R}^{N+1}} |f(z)| \, dz,$$

where $c_{km}$ is the constant that appears in Theorem 3.5.

**Proof.** Let $f \in L^1(\mathbb{R}^{N+1}, d, | \cdot |)$ and $\lambda > 0$ be fixed. We apply the Calderon-Zygmund type decomposition adapted for homogeneous spaces (as in Theorem 2.10) to the function $f$ and level $\lambda$. In consequence, we obtain

$$f = g + \sum h_i$$

where $||g||_{L^\infty} \leq \lambda,$

the functions $h_i$ are supported in balls $B_i$ whose total measure is controlled by $\sum |B_i| \leq \tilde{D} \lambda^{-1} ||f||_{L^1}.$

The set

$$\{z \in \mathbb{R}^{N+1} : T_{km}f(z) > \lambda\}$$

is contained in the union of the sets

$$\left\{ z \in \mathbb{R}^{N+1} : T_{km}g(z) > \frac{\lambda}{2} \right\}_{G} \text{ and } \left\{ z \in \mathbb{R}^{N+1} : T_{km}\left(\sum h_i\right)(z) > \frac{\lambda}{2} \right\}_{B},$$

hence, it suffices to estimate their measures. For the first of these sets $G$, we use Theorem 4.5. The estimate for the good set is as follows:

$$(a) \leq \frac{C}{\lambda^2} ||T_{km}g||^2_{L^2}$$

$$(b) \leq \frac{c_{km}}{\lambda^2} ||g||^2_{L^2}$$

$$= \frac{c_{km}}{\lambda^2} \int_{\mathbb{R}^{N+1}} |g(z)|^2 \, dz$$

$$(c) \leq \frac{c_{km}}{\lambda} \int_{\mathbb{R}^{N+1}} |g(z)| \, dz$$
Theorem 3.5 \((4.5)\), \((4.8)\)\), Theorem 4.5, \((c)\) is a consequence of \((4.5)\).

To estimate the measure of the remaining set \(B\), we begin by excising the supports of \(h_i\) from \(B\) since they are already controlled in the desired way. Indeed, in the ensuing calculation, this removal is crucial to create a gap between points in \(B_i\) and their complements which allows the application of Theorem 3.5.

Observe that due to the doubling property,

\[
\left\lvert \bigcup_i 2B_i \right\rvert \leq C \sum_i |B_i| \leq \frac{C}{\lambda} \|f\|_{L^1}.
\]

Therefore, instead of \(B\) we look at the new set

\[
\left\{ z \in \mathbb{R}^{N+1} \setminus \bigcup_i 2B_i : T_{km} \left( \sum_i h_i \right) (z) > \frac{\lambda}{2} \right\}.
\]

Once more with an application of Chebyshev’s inequality, we have

\[
\left\lvert \left\{ z \in \mathbb{R}^{N+1} \setminus \bigcup_i 2B_i : T_{km} \left( \sum_i h_i \right) (z) > \frac{\lambda}{2} \right\} \right\rvert \leq \frac{C}{\lambda} \int_{\mathbb{R}^{N+1} \setminus \bigcup_i 2B_i} \left\lvert \sum_i T_{km}h_i(z) \right\rvert \, dz
\]
\[
\leq \frac{C}{\lambda} \sum_i \int_{\mathbb{R}^{N+1} \setminus \bigcup_i 2B_i} \left\lvert T_{km}h_i(z) \right\rvert \, dz
\]

(4.8)

Let the ball \(B_i\) be centered at \(z_i \in \mathbb{R}^{N+1}\) and have radius \(r_i\), then

\[
E_i = \int_{\mathbb{R}^{N+1} \setminus \bigcup_i 2B_i} |T_{km}h_i(z)| \, dz
\]
\[
= \int_{\mathbb{R}^{N+1} \setminus \bigcup_i 2B_i} \left| \int_{B_i} K_{km}(\zeta^{-1} \circ z) h_i(\zeta) \, d\zeta \right| \, dz
\]
\[
\overset{(a)}{= \int_{\mathbb{R}^{N+1} \setminus \bigcup_i 2B_i} \left| \int_{B_i} \left( K_{km}(\zeta^{-1} \circ z) - K_{km}(z_i^{-1} \circ z) \right) h_i(\zeta) \, d\zeta \right| \, dz
\]
\[
\leq \int_{\mathbb{R}^{N+1} \setminus \bigcup_i 2B_i} \left| K_{km}(\zeta^{-1} \circ z) - K_{km}(z_i^{-1} \circ z) \right| |h_i(\zeta)| \, d\zeta \, dz
\]
\[
\overset{(b)}{\leq c_{km} \int_{\mathbb{R}^{N+1} \setminus \bigcup_i 2B_i} \int_{B_i} \frac{||\zeta^{-1} \circ z_i||^\beta}{||z^{-1} \circ z_i||^{Q+2+\beta}} |h_i(\zeta)| \, d\zeta \, dz
\]
\[
\overset{(c)}{= c_{km} \int_{B_i} \left( \frac{||\zeta^{-1} \circ z_i||^\beta}{||z^{-1} \circ z_i||^{Q+2+\beta}} \right) |h_i(\zeta)| \, d\zeta
\]
results in the following estimate where (a) is due to the fact that \( \int h_i = 0 \), (b) is due to (3.11), and (c) is due to Fubini’s theorem. Using (4.9) in (4.8) results in the following estimate

\[
\left\{ z \in \mathbb{R}^{N+1} \setminus \bigcup_i 2B_i : T_{km} \left( \sum_i h_i \right)(z) > \frac{\lambda}{2} \right\} \leq \frac{C}{\lambda} \sum_i E_i \\
\leq \frac{ckm}{\lambda} \sum_i \int_{B_i} |h_i(\zeta)| \, d\zeta \\
\leq \frac{ckm}{\lambda} \int_{\mathbb{R}^{N+1}} |f(\zeta)| \, d\zeta.
\]

(4.10)

Now, the weak-\( L^1 \)-type estimate is a consequence of a combination of the three estimates in (4.6), (4.7), and (4.10).

**Theorem 4.7.** (weak \( L^1 \)-type estimate for \( \mathcal{M}_{T_{km},s}^\# \)) Let \( s > 1 \), then the grand truncated maximal function \( \mathcal{M}_{T_{km},s}^\# \) maps \( L^1(\mathbb{R}^{N+1}, d, |\cdot|) \) to \( L^{1,\infty}(\mathbb{R}^{N+1}, d, |\cdot|) \) boundedly.

**Proof.** We recall the definition of the grand truncated maximal function.

\[
\mathcal{M}_{T_{km},s}^\# f(\zeta) = \sup_{B : B \ni \zeta, y, z \in B} \left| T(f\chi_{\mathbb{R}^{N+1}\setminus sB})(y) - T(f\chi_{\mathbb{R}^{N+1}\setminus sB})(z) \right|.
\]

For the proof, we will estimate the grand truncated maximal function by the uncentered Hardy-Littlewood maximal function \( \mathcal{M} \) in a pointwise manner, and this will conclude the proof of the Theorem since \( \mathcal{M} \) is weak type \((1,1)\).

Fix a point \( \zeta \in \mathbb{R}^N \) and choose a ball \( B \ni \zeta \) of radius \( r \). Let \( y, z \in B \) and consider the expression

\[
\left| T_{km}(f\chi_{\mathbb{R}^{N+1}\setminus sB})(y) - T_{km}(f\chi_{\mathbb{R}^{N+1}\setminus sB})(z) \right| = \left| \int_{\mathbb{R}^{N+1}\setminus sB} (K_{km}(\eta^{-1} \circ y) - K_{km}(\eta^{-1} \circ z)) f(\eta) \, d\eta \right|
\leq \int_{\mathbb{R}^{N+1}\setminus sB} \left| K_{km}(\eta^{-1} \circ y) - K_{km}(\eta^{-1} \circ z) \right| |f(\eta)| \, d\eta
\leq c_{km} \int_{\mathbb{R}^{N+1}\setminus sB} \frac{|y^{-1} \circ z|^\beta}{|\eta^{-1} \circ z|^{Q+\beta}} |f(\eta)| \, d\eta
\]
\[
\begin{align*}
&= c_{km} \sum_{k=1}^{\infty} \int_{s^{k+1}B \setminus s^k B} \frac{||y^{-1} \circ z||^\beta}{||\eta^{-1} \circ z||^{Q+2+\beta}} |f(\eta)| \, d\eta \\
&\leq c_{km} \sum_{k=1}^{\infty} \int_{s^{k+1}B \setminus s^k B} \frac{(r)^\beta}{(s^k \eta)^{Q+2+\beta}} |f(\eta)| \, d\eta \\
&\leq c_{km} \sum_{k=1}^{\infty} (s^k \eta)^{Q+2+\beta} \int_{s^k B} |f(\eta)| \, d\eta \\
&\leq \tilde{c}_{km} \int_{s^k B} |f(\eta)| \, d\eta.
\end{align*}
\]

Now, taking supremum over all \( y, z \in B \) followed by supremum over all \( B \ni \zeta \) gives us
\[
\mathcal{M}_{T_{km},s}^b f(\zeta) \leq c_{km} \mathcal{M} f(\zeta).
\]

\[
\square
\]

4.3. **Local pointwise estimates.** Now we are in a position to prove the following local sparse domination result. The following result essentially proves the pointwise domination of commutators by sparse operators. The proof is quite involved and probably is of independent interest due to its applicability for a large class of operators. For any ultraparabolic cube \( \mathcal{P}, \mathcal{P}^* \) denotes the dilated cube \( \delta_{c_N} \mathcal{P} \), where \( c_N \) is a fixed dimensional constant to be chosen later.

**Theorem 4.8** (Pointwise estimates for the commutator). Let \( \mathcal{P} \) be any ultraparabolic cube. Then for any compactly supported bounded function \( f \) there exists a \( \frac{1}{2} \)-sparse family \( \mathcal{G} \subset \mathcal{W}(\mathcal{P}) \) such that

\[
(4.11)
\]

\[
|\langle b, T_{km}(f \chi_{\mathcal{P}^*})(z) \rangle| \lesssim_N c_{km} \left[ \sum_{T \in \mathcal{G}} |b - b_{T^*}| \left( \frac{1}{|T^*|} \int_{T^*} |f| \right) \chi_T(z) + \sum_{T \in \mathcal{G}} \left( \frac{1}{|T^*|} \int_{T^*} |(b - b_{T^*}) f| \right) \chi_T(z) \right]
\]

\[
+ c_{km} ||b||_{BMO} \sum_{T \in \mathcal{G}} \left( \frac{1}{|T^*|} \int_{T^*} |f| \right) \chi_T(z)
\]

holds for a.e. \( z \in \mathcal{P} \).

**Proof.** For simplicity let us denote \( \langle f \rangle_{T^*} = \frac{1}{|T^*|} \int_{T^*} |f| \, dz \). Define the following operator
\[
\mathcal{N}_f(z) := \max \{ |T_{km}(f \chi_{\mathcal{P}^*})(z)|, \mathcal{M}_{T_{km},s}^b (f \chi_{\mathcal{P}^*})(z) \}.
\]

Theorem 4.6 and Theorem 4.7 together imply that \( \mathcal{N} \) is weak-type \( (1,1) \). Let us consider the following sets
\[
F_{\mathcal{P}, \text{bad}}^1 := \{ z \in \mathcal{P} : \mathcal{N}(f \chi_{\mathcal{P}^*})(z) > \alpha_N c_{km}(f \mathcal{P}^*) \},
\]
\[
F_{\mathcal{P}, \text{bad}}^2 := \{ z \in \mathcal{P} : \mathcal{M}(b - b_{\mathcal{P}^*}) f)(z) > \alpha_N c_{km}(b - b_{\mathcal{P}^*} f) \mathcal{P} \}.
\]

Denote \( F_{\mathcal{P}, \text{bad}} = F_{\mathcal{P}, \text{bad}}^1 \cup F_{\mathcal{P}, \text{bad}}^2 \). Since \( \mathcal{M} \) and \( \mathcal{N} \) are weak-type \( (1,1) \), it is possible to choose large enough \( \alpha_N \) such that

\[
(4.12)
\]

\[
|F_{\mathcal{P}, \text{bad}}| \lesssim O \left( \frac{1}{\alpha_N} \right) |\mathcal{P}| \leq \frac{1}{2} |\mathcal{P}|.
\]
Note that we have used the fact $|\mathcal{P}^*| \lesssim |\mathcal{P}|$ in the above estimate. To construct appropriate sparse family, we shall decompose the set $F_{\mathcal{P}, \text{bad}}$ using Calderón-Zygmund decomposition. To that end, let us consider

$$F_{\chi, \mathcal{P}} := \{ z \in \mathcal{P} : \mathcal{M}\mathcal{P}(\chi_{F_{\mathcal{P}, \text{bad}}})(z) > \lambda \}.$$  

Local Calderón-Zygmund decomposition for small enough $\lambda$ produces a collection of pairwise disjoint ultraparabolic cubes $\mathcal{G}_\mathcal{P} \subset \mathcal{U}(\mathcal{P})$ such that $F_{\mathcal{P}, \text{bad}} \subset F_{\chi, \mathcal{P}} = \bigcup_{\mathcal{T} \in \mathcal{G}_\mathcal{P}} \mathcal{T}$ and

$$\lambda < \frac{|\mathcal{T} \cap F_{\mathcal{P}, \text{bad}}|}{|\mathcal{T}|} \leq \frac{1}{2},$$

for all $\mathcal{T} \in \mathcal{G}_\mathcal{P}$. The above implies

$$\sum_{\mathcal{T} \in \mathcal{G}_\mathcal{P}} |\mathcal{T}| \lesssim \sum_{\mathcal{T} \in \mathcal{G}_\mathcal{P}} |\mathcal{T} \cap F_{\mathcal{P}, \text{bad}}| \lesssim \frac{1}{\lambda} |F_{\mathcal{P}, \text{bad}}| \lesssim O\left(\frac{1}{\alpha_N \lambda}\right) |\mathcal{P}| \leq \frac{1}{2} |\mathcal{P}|,$$

provided we choose $\alpha_N$ very large.

Define $\mathcal{G}_0 = \{ \mathcal{P} \}$ and $\mathcal{G}_1 = \mathcal{G}_\mathcal{P}$. For each $\mathcal{T} \in \mathcal{G}_\mathcal{P}$, similar procedure generates the collection $\mathcal{G}_\mathcal{T}$ and we combine them to define $\mathcal{G}_2 = \bigcup_{\mathcal{T} \in \mathcal{G}_\mathcal{P}} \mathcal{G}_\mathcal{T}$. Iteratively we define the collection $\mathcal{G}_k$ for all $k = 0, 1, 2, \ldots$. Therefore, from Remark 4.2 and (4.15) it follows that the collection $\mathcal{G} = \bigcup_{k \geq 0} \mathcal{G}_k$ is a sparse collection.

Rest of the proof is dedicated to conclude that $\mathcal{G}$ is a required collection for which (4.11) holds true. The recursive process implies that $\sum_{\mathcal{T} \in \mathcal{G}_k} |\mathcal{T}| \leq \frac{1}{2^k} |\mathcal{P}|$, consequently, $|\mathcal{G}_k| \to 0$ as $k \to \infty$. Therefore, up to a measure zero set, for each point $z \in \mathcal{P}$ there exists a largest $k = k(z) \in \mathbb{N}$ depending on $z$ such that $z \in \mathcal{G}_j$ for each $0 \leq j \leq k(z)$. Hence, there exist a chain of ultraparabolic dyadic cubes $\{\mathcal{T}_j\}_{j=0}^{k(z)}$ with $\mathcal{T}_j \in \mathcal{G} \mathcal{P}$ such that $z \in \mathcal{T}_{k(z)} \subset \mathcal{T}_{k(z)-1} \subset \cdots \subset \mathcal{T}_0 = \mathcal{P}$. Since $[b, T]f = [b - c, T]f$ for any constant $c \in \mathbb{R}$, note that

$$|\mathcal{B}, T_{km}(f \chi_{\mathcal{P}^*})(z)| \leq \sum_{j=1}^{k(z)} |\mathcal{B}, T_{km}(f \chi_{T_{j-1}^* \setminus T_j^*})(z)| + |\mathcal{B}, T_{km}(f \chi_{T_{k(z)}^*})(z)|
$$

$$= \sum_{j=1}^{k(z)} |\mathcal{B} - b_{T_{j-1}^*}, T_{km}(f \chi_{T_{j-1}^* \setminus T_j^*})(z)| + |\mathcal{B} - b_{T_{k(z)}^*}, T_{km}(f \chi_{T_{k(z)}^*})(z)|
$$

$$\leq \sum_{j=1}^{k(z)} \left( |\mathcal{B} - b_{T_{j-1}^*}, T_{km}(f \chi_{T_{j-1}^* \setminus T_j^*})(z)| + |\mathcal{B} - b_{T_{j-1}^*}, T_{km}(f \chi_{T_{j-1}^* \setminus T_j^*})(z)| \right)
$$

$$+ \left( |\mathcal{B} - b_{T_{k(z)}^*}, T_{km}(f \chi_{T_{k(z)}^*})(z)| + |\mathcal{B} - b_{T_{k(z)}^*}, T_{km}(f \chi_{T_{k(z)}^*})(z)| \right)
$$

$$\lesssim \sum_{j=1}^{k(z)} \left( \left| b - b_{T_{j-1}^*} \right| T_{km}(f \chi_{T_{j-1}^* \setminus T_j^*})(z) + \left| T_{km}(f \chi_{T_{k(z)}^*})(z) \right| \right)
$$

$$+ c_{km} \left( |\mathcal{B} - b_{T_{k(z)}^*}, f\rangle_{T_{k(z)}^*} + |\mathcal{B} - b_{T_{k(z)}^*}, f\rangle_{T_{k(z)}^*} \right),$$

where in the last inequality we have used the fact that $z \notin \mathcal{G}_{k(z)+1}$. Now, we analyze the summands $I_j$ and $II_j$ in the sum in (4.16) for $j \in \{1, 2, \ldots, k(z)\}$. Note that for each $z' \in \mathcal{T}_j$, we can write

$$I_j = \left| b - b_{T_{j-1}^*} \right| T_{km}(f \chi_{T_{j-1}^* \setminus T_j^*})(z).$$
Now to complete the proof we observe that
\[
\left| b - b_{T_{j-1}} \right| T_{km}(f \chi_{T_{j-1}})(z) - T_{km}(f \chi_{T_{j-1}})(z') \right| + \left| b - b_{T_{j-1}} \right| T_{km}(f \chi_{T_{j-1}})(z)\right|, \\
\leq \left| b - b_{T_{j-1}} \right| \inf_{\zeta \in T_j} M_{\cal T_{km,cN}}^z \left( f \chi_{T_{j-1}} \right)(\zeta) + \left| b - b_{T_{j-1}} \right| \left( \inf_{\zeta \in T_j} T_{km}(f \chi_{T_{j-1}})(\zeta) + \inf_{\zeta \in T_j} T_{km}(f \chi_{T_{j-1}})(\zeta) \right),
\]
where the first inequality is obtained by adding and subtracting a suitable term. Similarly, we have

\[
(4.18) \quad II_j = \left| T_{km}(b - b_{T_{j-1}}) f \chi_{T_{j-1}}(z) \right| \\
\leq T_{km}(b - b_{T_{j-1}}) f \chi_{T_{j-1}}(z) - T_{km}(b - b_{T_{j-1}}) f \chi_{T_{j-1}}(z') \right| + \left| T_{km}(b - b_{T_{j-1}}) f \chi_{T_{j-1}}(z) \right|, \\
\leq \inf_{\zeta \in T_j} M_{\cal T_{km,cN}}^z \left( b - b_{T_{j-1}} \right) f \chi_{T_{j-1}}(\zeta) + \left( \inf_{\zeta \in T_j} T_{km}(b - b_{T_{j-1}}) f \chi_{T_{j-1}}(\zeta) + \inf_{\zeta \in T_j} T_{km}(b - b_{T_{j-1}}) f \chi_{T_{j-1}}(\zeta) \right) \\
= \inf_{\zeta \in T_j} M_{\cal T_{km,cN}}^z \left( b - b_{T_{j-1}} \right) f \chi_{T_{j-1}}(\zeta) + \inf_{\zeta \in T_j} T_{km}(b - b_{T_{j-1}}) f \chi_{T_{j-1}}(\zeta) + \inf_{\zeta \in T_j} T_{km}(b - b_{T_{j-1}}) f \chi_{T_{j-1}}(\zeta) \\
\leq \inf_{\zeta \in T_j} M_{\cal T_{km,cN}}^z \left( b - b_{T_{j-1}} \right) f \chi_{T_{j-1}}(\zeta) + \inf_{\zeta \in T_j} T_{km}(b - b_{T_{j-1}}) f \chi_{T_{j-1}}(\zeta) + \inf_{\zeta \in T_j} T_{km}(b - b_{T_{j-1}}) f \chi_{T_{j-1}}(\zeta) + \inf_{\zeta \in T_j} T_{km}(b - b_{T_{j-1}}) f \chi_{T_{j-1}}(\zeta) \\
+ \inf_{\zeta \in T_j} \left( f \chi_{T_{j-1}}(\zeta) + \inf_{\zeta \in T_j} T_{km}(f \chi_{T_{j-1}})(\zeta) \right),
\]

Now to complete the proof we observe that \( |T_j \setminus (F_{T_{j,bad}} \cup F_{T_{j-1,bad}}) | > 0 \), indeed,

\[
|T_j \cap (F_{T_{j,bad}} \cup F_{T_{j-1,bad}}) | \leq |F_{T_{j,bad}} | + |T_j \cap F_{T_{j-1,bad}} | \leq \frac{1}{2} |T_j | + \frac{1}{4} |T_j |.
\]

On account of the above, we obtain the following estimates
Proof. The proof is similar to that of Theorem 4.8 and is inspired by the work [Lor21]. We point out the main steps for the sake of completion.

Step 1: In a first step, we define an operator

\[ Nf(z) := \max\{|T_{km}(f\chi_{P^*})(z)|, M_{T_{km},s}^+(f\chi_{P^*})(z)\}, \]

which is weak-type \((1, 1)\) by Theorem 4.6 and Theorem 4.7.

Step 2: In a second step, we consider the set

\[ F_{P, \text{bad}} := \{z \in P : \max\{Mf(z), N(f\chi_{P^*})(z)\} > c_Nckm(f)\}, \]

which satisfies

\[ |F_{P, \text{bad}}| \lesssim O\left(\frac{1}{\kappa}\right) |P| \leq \frac{1}{2} |P|, \]

for sufficiently large \(c_N\).

Step 3: Next, we apply Calderón-Zygmund decomposition to the function \(\chi_{F_{P, \text{bad}}}\) at level \(\lambda > 0\) which produces a pairwise disjoint collection of ultraparabolic cubes \(\mathcal{G}_P \subset \mathcal{U}(P)\) such that

\[ F_{P, \text{bad}} \subset \bigcup_{T \in \mathcal{G}_P} T \text{ and } \lambda < \frac{|T \cap F_{P, \text{bad}}|}{|T|} \leq \frac{1}{2}, \]

for all \(T \in \mathcal{G}_P\) for sufficiently small \(\lambda\). We also get

\[ \sum_{T \in \mathcal{G}_P} |T| \leq \frac{1}{\lambda} \sum_{T \in \mathcal{G}_P} |T \cap F_{P, \text{bad}}| \lesssim \frac{1}{\lambda} |F_{P, \text{bad}}| \lesssim O\left(\frac{1}{c_Nckm\lambda}\right) |P| \leq \frac{1}{2} |P|, \]

provided we choose \(c_N\) very large.
**Step 4:** Now, the sparse collection is defined in a recursive manner. Set \( \mathcal{S}_0 = \{ P \} \) and \( \mathcal{S}_1 = \mathcal{S}_P \). For each \( T \in \mathcal{S}_P \), we repeat the steps 1, 2, and 3 to obtain the collection \( \mathcal{S}_T \) and their union \( \mathcal{S}_2 = \bigcup_{T \in \mathcal{S}_P} \mathcal{S}_T \) gives us the second stage of the recursion. Iterating over all sub-cubes indefinitely we receive the collection \( \mathcal{S}_k \) for all \( k = 0, 1, 2, \ldots \) Once again, Remark 4.2 and (4.22) imply that the collection \( \mathcal{S} = \bigcup_{k \geq 0} \mathcal{S}_k \) is a sparse collection.

**Step 5:** It remains to prove that \( \mathcal{S} \) satisfies the property (4.19). The recursive process implies that \( \sum_{T \in \mathcal{S}_k} |T| \leq \frac{1}{2^k} |P| \), consequently, \( |\mathcal{S}_k| \to 0 \) as \( k \to \infty \). Therefore, up to a measure zero set, for each point \( z \in P \) there exists a largest number \( k(z) \) such that \( z \in \mathcal{S}_j \) for each \( 0 \leq j \leq k(z) \). Hence, there exist a chain of ultraparabolic dyadic cubes \( \{ T_j^{(z)} \}_{j=0}^{k(z)} \) with \( T_j^{(z)} \in \mathcal{S}_j \) such that \( z \in T_{k(z)}^{(z)} \subset T_k(z-1) \subset \cdots \subset T_0 = P \). Observe

\[
\left| T_{km}(f \chi_{P^*})(z) \right| \leq \sum_{j=1}^{k(z)} \left| T_{km}(f \chi_{T_{j-1}^*} \setminus T_j^*)(z) \right| + \left| T_{km}(f \chi_{T_{k(z)}^*})(z) \right|
\]

(4.23)

where in the last inequality we have used the fact that \( z \notin \mathcal{S}_{k(z)+1} \).

Now, we shall estimate \( I_j \) in the sum in (4.23). Note that for each \( z' \in T_j \), we can write

\[
|T_{km}(f \chi_{T_{j-1}^*} \setminus T_j^*)(z)| \leq |T_{km}(f \chi_{T_{j-1}^*} \setminus T_j^*)(z) - T_{km}(f \chi_{T_{j-1}^*})(z')| + |T_{km}(f \chi_{T_{j-1}^*})(z')|
\]

\[
\leq \inf_{z \in T_j} |T_{km}(f \chi_{T_{j-1}^*})(z)| + \inf_{z' \in T_j} |T_{km}(f \chi_{T_{j-1}^*})(z')|.
\]

Now to complete the proof we observe that \( |T_j \setminus (F_{T_j, \text{bad}} \cup F_{T_{j-1}, \text{bad}})| > 0 \), indeed,

\[
|T_j \cap (F_{T_j, \text{bad}} \cup F_{T_{j-1}, \text{bad}})| \leq |F_{T_j, \text{bad}}| + |T_j \cap F_{T_{j-1}, \text{bad}}| \leq \frac{1}{2} |T_j| + \frac{1}{4} |T_j|.
\]

Therefore

\[
\inf_{z \in T_j} |T_{km}(f \chi_{T_{j-1}^*})(z)| \leq \kappa(f)_{T_{j-1}^*},
\]

\[
\inf_{z \in T_j} |T_{km}(f \chi_{T_{j-1}^*})(z)| \leq \kappa(f)_{T_{j-1}^*},
\]

\[
\inf_{z \in T_j} |T_{km}(f \chi_{T_{j-1}^*})(z)| \leq \kappa(f)_{T_{j-1}^*}.
\]

Combining the above estimates in (4.23) and (4.24), we obtain

\[
|T_{km}(f \chi_{P^*})(z)| \lesssim_{N} \sum_{T \in \mathcal{S}} \left( \frac{1}{|T^*|} \int_{T^*} |f| \right) \chi_T(z),
\]

for a.e. \( z \in P \).

**Proof of Theorem 4.4.** Recall the representation formula

\[
u_{x(x_j)} = - \sum_{m=1}^{\infty} \sum_{k=1}^{g_m} \alpha_{ij}^{km} T_{km}(L u) + \sum_{m=1}^{\infty} \sum_{k=1}^{g_m} \sum_{h,k=1}^{\infty} \sum_{m=1}^{g_m} \sum_{k=1}^{g_m} \Gamma_{ij}^{km} a_{hk} T_{km}(u_{x(k_hk)}) + \alpha_{ij} \cdot L u.
\]
Now the proof is complete a consequence of the applications of Theorem 4.9 for \( f = Lu \) and Theorem 4.8 to \( f = u_{x_h x_h} \).

\[ \square \]

5. Applications of sparse domination

In this section we highlight the consequences of our sparse domination results.

5.1. Weighted estimates on Lebesgue spaces. This section is devoted to prove appropriate weighted norm inequalities for the sparse operators \( \mathcal{A}_g \) and \( \mathcal{A}_{g,b} \) where \( b \) is in \((\delta, \kappa) - \text{BMO}\) class.

Let us start with the definition of Muckenhoupt \( A_p \) weights in our context.

**Definition 5.1.** For \( 1 < p < \infty \), we say a weight \( \omega \) on \( \mathbb{R}^{N+1} \) belongs to the class \( \mathcal{A}_p(\mathbb{R}^{N+1}, d, |\cdot|) \), if

\[
[w]_{A_p} := \sup_{P} \left( \frac{1}{|P|} \int_P \omega \, dx \right) \left( \frac{1}{|P|} \int_P \omega^{1-p'} \, dx \right)^{p-1} < \infty,
\]

where the supremum is taken over all ultraparabolic cubes \( P \) in the homogeneous space \((\mathbb{R}^{N+1}, d, |\cdot|)\).

\( \sigma \) will denote the dual weight to \( \omega \), i.e., \( \sigma = \omega^{-\frac{1}{p-1}} \). It is well known that \( w \in \mathcal{A}_p(\mathbb{R}^{N+1}, d, |\cdot|) \iff \sigma \in \mathcal{A}_{p'}(\mathbb{R}^{N+1}, d, |\cdot|) \). Here onwards, we simply denote the class \( \mathcal{A}_p(\mathbb{R}^{N+1}, d, |\cdot|) \) as \( \mathcal{A}_p(\mathbb{R}^{N+1}) \).

We shall prove the following estimates in order to conclude weighted gradient estimates for the solution of \((1.1)\).

**Lemma 5.2.** Let \( \mathcal{G} \) be an \( \eta \)-sparse family.

1. For \( 1 < p < \infty \), \( \omega \in \mathcal{A}_p(\mathbb{R}^{N+1}) \) and \( f \in L^p(\omega) \) we have

\[
\|\mathcal{A}_g f\|_{L^p(\omega)} \leq C_{N,p,\eta} [w]_{A_p} \| f \|_{L^p(\omega)}.
\]

2. Let \( 1 < p < \infty \), \( \omega \in \mathcal{A}_p(\mathbb{R}^{N+1}) \), and \( b \in L^\infty \) with small \((\delta, \tilde{K})\) norm where \( \tilde{K} > r(P) \) for all \( P \in \mathcal{G} \). Then

\[
\|\mathcal{A}_{g,b} f\|_{L^p(\omega)} + \|\mathcal{A}_{g,b} f\|_{L^p(\omega)} \leq C_{N,p,\eta} [w]_{A_p} \max \bigg\{ \frac{1}{|P|} \bigg\} O(\delta) \| f \|_{L^p(\omega)},
\]

holds for all \( f \in L^p(\omega) \).

**Proof.** The proof of (5.2) is well known and is available in [Lor21]. We only prove (5.3) by showing that \( \mathcal{A}_{g,b}^* \) is controlled pointwise by composition of two sparse operators. The argument is similar to the proof of Theorem 4.9. Let us divide the proof into several parts.

**Step 1 (Pointwise controlling the oscillation):** We claim that for any \( P \in \mathcal{G} \) there exists a family of pairwise disjoint cubes \( \mathcal{G}_P \subset \mathcal{U}(P) \) such that

\[
|b(z) - (b)_P| \leq \alpha_N \Omega_P(b) + \sum_{T \in \mathcal{G}_P} |b(z) - (b)_T| \chi_T, \text{ a.e. } x \in P,
\]

and \( \sum_{T \in \mathcal{G}_P} |T| \leq \frac{1}{2}|P| \). Let us consider the set \( E := \{ x \in P : M^P(b - (b)_P)(z) > \alpha_N \Omega_P(b) \} \).

Choose \( \alpha_N \) large enough to ensure that \( |E| \leq \frac{1}{2} |P| \). Local Calderón-Zygmund decomposition to the function \( \chi_E \) at level \( \lambda > 0 \) produces a pairwise disjoint collection of ultraparabolic cubes \( \mathcal{G}_P \subset \mathcal{U}(P) \) such that

\[
E \subset \bigcup_{T \in \mathcal{G}_P} T \text{ and } \lambda < \frac{|T \cap E|}{|T|} \leq \frac{1}{2}.
\]
for all $T \in \mathcal{G}_p$ for sufficiently small $\lambda$. Therefore, $|b(z) - (b)_P| \leq \alpha_N \Omega_P(b)$ a.e. $z \notin \bigcup_{T \in \mathcal{G}_p} T$. Also, (5.5) implies that $T \cap E^c \neq \emptyset$, hence for some $\xi \in T \cap E^c$ we have

$$|(b)_T - (b)_P| \leq \frac{1}{|T|} \int_T |b - (b)_P| \leq \mathcal{M}^T (b - (b)_P)(\xi) \leq \alpha_N \Omega_P(b).$$

Combining the above estimates we obtain

$$|b(z) - (b)_P| \leq |b(z) - (b)_P| |\chi_P| \bigg\| \bigcup_{T \in \mathcal{G}_p} T \bigg\| + \sum_{T \in \mathcal{G}_p} |b(z) - (b)_T| |\chi_T| + \sum_{T \in \mathcal{G}_p} |(b)_P - (b)_T| |\chi_T| \leq \alpha_N \Omega_P(b) + \sum_{T \in \mathcal{G}(P)} |b(z) - (b)_T| |\chi_T|.$$  

Also, arguing as in Theorem 4.9, we obtain $\sum_{T \in \mathcal{G}_p} |T| \leq \frac{1}{2} |P|$. This completes the claim. Recursive application of (5.4) produces a $\frac{1}{2}$-sparse family $Z(P) \subset \mathcal{U}(P)$ such that

$$|b(z) - (b)_P| \leq \alpha_N \sum_{T \in \mathcal{G}(P)} \Omega_T(b) |\chi_T| \text{ a.e. } z \in P.$$  

**Step 2 (Augmentation process):** Recursive application to each $P \in \mathcal{G}$ produces a $\frac{1}{2}$-sparse families $Z(P) \subset \mathcal{U}(P)$ such that (5.6) holds. Define the new collection $\tilde{\mathcal{G}} = \mathcal{G} \cup (\bigcup_{P \in \mathcal{G}} Z(P))$. Since $\mathcal{G}$ is a $\eta$-sparse family and each $Z(P)$ is $\frac{1}{2}$-sparse, using Lemma 2.4 from [LORR17], we conclude that $\tilde{\mathcal{G}}$ is $c_\eta := \frac{\eta}{2(1+\eta)}$ sparse. Also, for any arbitrary cube $P \in \tilde{\mathcal{G}}$, denote $S(P) := \{T \in \tilde{\mathcal{G}} : T \subseteq P\}$. Let us decompose $S(P)$ into maximal subfamilies and using the fact that $\tilde{\mathcal{G}}$ is $c_\eta$ sparse, we conclude that

$$|b(z) - (b)_P| \leq \alpha_N \sum_{T \in S(P)} \Omega_T(b) |\chi_T(z)| \leq \alpha_N O(\delta) \sum_{T \in S(P)} |\chi_T(z)|, \text{ a.e. } z \in P,$$

where in the last inequality we have used the fact $\Omega_T(b) \leq O(\delta)$ since $b$ has small $(\delta, \mathfrak{F})$–BMO norm.

**Step 3 (Estimate of $\|\mathcal{A}_{\mathcal{G}}^{b, \ast} f\|_{L^p(\omega)}$):** The sparseness of $\tilde{\mathcal{G}}$ implies there exists $F_T \subset T$ such that $|F_T| \geq c_\eta |T|$ and $\{F_T\}$ is a disjoint family. Now invoking (5.7), for each ultraparabolic cube $P \in \tilde{\mathcal{G}}$, we obtain

$$\int_P |b(z) - (b)_P| |f(z)| \leq \alpha_N O(\delta) \sum_{T \in S(P)} \left( \frac{1}{|T|} \int_T |f| \right) |T| \leq c_\eta \alpha_N O(\delta) \sum_{T \in S(P)} \left( \frac{1}{|T|} \int_T |f| \right) |F_T| \leq c_\eta \alpha_N O(\delta) \int_P |\mathcal{A}_{\mathcal{G}} f(z)| dz.$$  

Therefore,

$$\mathcal{A}_{\mathcal{G}}^{b, \ast} f(z) \leq c_\eta \alpha N O(\delta) |\mathcal{A}_{\mathcal{G}} f(z)|.$$  

From (5.2), we conclude that

$$\|\mathcal{A}_{\mathcal{G}}^{b, \ast} f\|_{L^p(\omega)} \leq c_\eta \alpha N O(\delta) \|\mathcal{A}_{\mathcal{G}} f\|_{L^p(\omega)} \leq c_\eta \alpha N O(\delta) \|f\|_{A_p} \max \left\{ \frac{1}{p-1}, 1 \right\} \|f\|_{L^p(\omega)}.$$
Finally \( \|\mathfrak{g}_f^b\|_{L^p(\omega) \to L^p(\omega)} = \|\mathfrak{g}_f^b \ast f\|_{L^p'(\omega) \to L^p'(\omega)} \leq c_{\eta,N}O(\delta)[\omega]_{A_p}^{2\max\{\frac{1}{p-1},1\}}. \) \( \square \)

Now, in order to prove Theorem 1.1, we begin with the following local estimate.

**Proposition 5.3.** For every \( p \in (1, \infty) \), there exists \( C = C(\text{data}) \) and \( r_0 = r_0(\text{data}) \) such that if \( u \in C^\infty_0(\mathbb{R}^N \times (0, \infty)) \), \( \text{spt } u \subset B_r \subset \mathbb{R}^N \times (0, \infty) \) with \( 0 < r < r_0 \), then, for \( i,j = 1,2,...,s_0 \)

\[
\|u_{x_i x_j}\|_{L^p(\omega)} \leq C_{A,B,N,p,\eta}[\omega]_{A_p}^{\max\{\frac{1}{p-1},1\}}\|\mathcal{L}u\|_{L^p(\omega)}.
\]

**Proof.** There is an ultraparabolic cube \( P \) such that \( B_r \subset P \) whose size is determined in Lemma 5.2 and this determines \( r_0 \). Therefore, we take \( L^p(\mathcal{P},\omega) \) norms on both sides of the representation formula (4.4) to get

\[
\|u_{x_i x_j}\|_{L^p(\mathcal{P},\omega)} \leq \kappa_{A,N,C} \sum_{m=1}^{\infty} \sum_{k=1}^{g_m} c_{km} |c_{ijm}| \sum_{l=1}^{C(\mathcal{N})} \|\mathfrak{g}_{h_{km},l}(\mathcal{L}u)\|_{L^p(\mathcal{P},\omega)} + \sum_{h,k=1}^{s_0} \sum_{m=1}^{g_m} c_{km} |c_{ijm}| \sum_{l=1}^{C(\mathcal{N})} \|\mathfrak{g}_{h_{km},l}^a(u_{x_i x_j})\|_{L^p(\mathcal{P},\omega)} + \|\|L^p(\mathcal{P},\omega)\|_{L^p(\mathcal{P},\omega)}.
\]

Applying Lemma 5.2 to the above inequality, we obtain

\[
\|u_{x_i x_j}\|_{L^p(\mathcal{P},\omega)} \leq C_{A,B,N,p,\eta}[\omega]_{A_p}^{\max\{\frac{1}{p-1},1\}}\|\mathcal{L}u\|_{L^p(\mathcal{P},\omega)} \sum_{m=1}^{g_m} \sum_{k=1}^{g_m} c_{km} |c_{ijm}| + C_{A,B,N,p,\eta}[\omega]_{A_p}^{2\max\{\frac{1}{p-1},1\}} O(\delta) \sup_{h,k} \|u_{x_i x_j}\|_{L^p(\mathcal{P},\omega)} \sum_{m=1}^{g_m} \sum_{k=1}^{g_m} c_{km} |c_{ijm}| + \|\|L^p(\mathcal{P},\omega)\|_{L^p(\mathcal{P},\omega)}.
\]

(5.10)

We estimate the sum

\[
\sum_{m=1}^{g_m} \sum_{k=1}^{g_m} c_{km} |c_{ijm}| \leq \sum_{m=1}^{g_m} \sum_{k=1}^{g_m} m^{-2l} m^{\frac{N+1}{m}} \leq \sum_{m=1}^{g_m} \sum_{k=1}^{g_m} m^{-2l} m^{\frac{N-1}{m}} < \infty
\]

provided \( l > \frac{3N-1}{4} \). In the first inequality, we have used (3.6), (3.8) and Theorem 3.5, whereas for the second inequality we used (3.5). Therefore (5.10) if further estimated to

\[
\|u_{x_i x_j}\|_{L^p(\mathcal{P},\omega)} \leq C_{A,B,N,p,\eta}[\omega]_{A_p}^{\max\{\frac{1}{p-1},1\}} \|\mathcal{L}u\|_{L^p(\mathcal{P},\omega)} + C_{A,B,N,p,\eta}[\omega]_{A_p}^{2\max\{\frac{1}{p-1},1\}} O(\delta) \sup_{h,k} \|u_{x_i x_j}\|_{L^p(\mathcal{P},\omega)}.
\]

(5.11)

for all \( i,j = 1,2,...,s_0 \). Therefore, if \( \delta \) is further chosen such that

\[
C_{A,B,N,p,\eta}[\omega]_{A_p}^{2\max\{\frac{1}{p-1},1\}} O(\delta) < 1,
\]

we have proved Theorem 1.1.
we get

\begin{equation}
\sup_{i,j} \|u_{x_ix_j}\|_{L^p(\mathcal{P}, \omega)} \leq C_{A,B,N,p,\eta} \max\left\{ \left[ \omega \right]_{A_p}^{\frac{1}{p-1}}, |\alpha_{ij}| \right\} \|\mathcal{L}u\|_{L^p(\mathcal{P}, \omega)}
\end{equation}

\leq C_{A,B,N,p,\eta} \left[ \omega \right]_{A_p}^{\frac{1}{p-1}} \|\mathcal{L}u\|_{L^p(\mathcal{P}, \omega)},

since $|\alpha_{ij}| \leq 1$. Now, since $u$ is supported in $B_R$, we obtain the desired estimate.

Now, the proof of Theorem 1.1 is immediate.

\textbf{Proof of Theorem 1.1.} In a first step, one obtains the following estimate by a standard argument using cut-off functions and an interpolation lemma, for example, see [GT01, Theorem 9.11].

\begin{equation}
\sup_{i,j} \|u_{x_ix_j}\|_{L^p(B_2, \omega)} \leq C_{A,B,N,p,\eta} \left[ \omega \right]_{A_p}^{\frac{1}{p-1}} \left( \|\mathcal{L}u\|_{L^p(B_2, \omega)} + \frac{\|u\|_{L^p(B_2, \omega)}}{r^2} \right).
\end{equation}

The estimate in Theorem 1.1 follows by covering $\Omega'$ with balls $B$ of radius $R/2$ where where $R$ has a further restriction $R < \text{dist}(\Omega', \Omega)$, other than the restriction from Proposition 5.3.

\hfill \Box

5.2. \textbf{Weighted estimates on Variable Lebesgue spaces.} Let us begin by introducing the variable Lebesgue space and their Sobolev counterparts. We denote by $p(\cdot) = p(z) : \mathbb{R}^{N+1} \to \mathbb{R}$ a measurable function satisfying

\begin{equation}
1 < \gamma_1 \leq p(z) \leq \gamma_2 < \infty, \quad \text{for all } z \in \mathbb{R}^{N+1}
\end{equation}

and we call it a \textit{variable exponent}. The \textit{variable exponent Lebesgue space} $L^{p(\cdot)}(\Omega)$ is defined to be the space of all functions $g : \Omega \to \mathbb{R}$ such that the \textit{modular}

$$
\rho_{p(\cdot)}(g) := \int_{\Omega} |g(z)|^{p(z)} \, dx
$$

is finite.

The \textit{variable exponent Lebesgue space} $L^{p(\cdot)}(\Omega)$ is known to be a reflexive Banach space when equipped with the Luxembourg norm

\begin{equation}
\|g\|_{L^{p(\cdot)}(\Omega)} := \inf \left\{ \lambda > 0 : \rho_{p(\cdot)} \left( \frac{g}{\lambda} \right) \leq 1 \right\},
\end{equation}

and the hypothesis (5.14). For a proof, see [DHHR11]. With the hypothesis (5.14), the dual space $L^{p'(-)}(\Omega)$ is also reflexive where $p'(z) = \frac{p(z)}{p(z)-1}$.

The \textit{variable exponent Sobolev space} $W^{k,p(\cdot)}(\Omega)$ is defined as the space of functions $g \in L^{p(\cdot)}(\Omega)$ such that the distributional derivatives $D^\beta g$ is also in $L^{p(\cdot)}(\Omega)$ for all $|\beta| \leq k$ where $\beta$ is a multi-index. Further, the norm in $W^{k,p(\cdot)}(\Omega)$ is defined to be

\begin{equation}
\|g\|_{W^{k,p(\cdot)}(\Omega)} := \sum_{|\beta| \leq k} \|D^\beta g\|_{L^{p(\cdot)}(\Omega)}.
\end{equation}

The class of variable exponents under consideration is very large to obtain regularity properties of the partial differential equations under consideration. Zhikov [Zhi86], while studying homogenization of highly anisotropic nonlinear media, was among the first people to notice that the variable exponent needs to satisfy additional properties in order to avoid the so called Lavrentiev phenomena, which may be paraphrased as the principle that equations (and their structure/symmetry properties) determine the function spaces in which they should be solved. Zhikov zeroed in on log-Hölder continuity of the variable exponent as an essential requirement in order to obtain regularity results for minimizers of functionals with $p(\cdot)$-growth. Diening [Die04] proved that the log-Hölder...
condition is also required to prove boundedness of the maximal operator on variable exponent Lebesgue spaces. This clearly suggests that the log-Hölder condition is also indispensable when studying gradient $L^p(\cdot)$ estimates.

**Definition 5.4.** We say that a variable exponent $p(\cdot)$ is locally log-Hölder continuous if there exists a constant $C_0$ such that for every $\eta, \zeta \in \mathbb{R}^{N+1}$ such that $d(\eta, \zeta) < 1/2$, it holds that

$$
|p(\eta) - p(\zeta)| \leq \frac{C_0}{\log(d(\eta, \zeta))}. 
$$

(5.17)

On the other hand, $p(\cdot)$ is log-Hölder continuous at infinity with respect to some base point $\zeta_0 \in \mathbb{R}^{N+1}$, if there are constants $C_{\infty}$ and $p_{\infty}$ such that for every $\zeta \in \mathbb{R}^{N+1}$,

$$
|p(\zeta) - p_{\infty}| < \frac{C_{\infty}}{\log(e + d(\zeta, \zeta_0))}. 
$$

(5.18)

If the variable exponent $p(\cdot)$ satisfies (5.17) and (5.18), we say that $p(\cdot)$ is a globally log-Hölder continuous function.

**Remark 5.5.** It turns out that the definition of log-Hölder at infinity is independent of the choice of base point. Hence, there is no loss of generality to choose an arbitrary $\zeta_0$. In our applications, we will only require the boundedness of the maximal function for $L^{p(\cdot)}(\Omega)$ for bounded domains $\Omega$. In this case, the log-Hölder condition at infinity is superfluous.

Let us start with the definition of Muckenhoupt $A_p$ weights in our context.

**Definition 5.6.** For $1 < p < \infty$, we say a weight $\omega$ on $\mathbb{R}^{N+1}$ belongs to the class $A_p(\mathbb{R}^{N+1}, d, |\cdot|)$, if

$$
[\omega]_{A_p(\cdot)} := \sup_{P} |P|^{-1} \|\omega \chi_P |_{L^{p(\cdot)}} \| \omega^{-1} \chi_P |_{L^{p'(\cdot)}} < \infty.
$$

(5.19)

where the supremum is taken over all ultraparabolic cubes $P$ in the homogeneous space $(\mathbb{R}^{N+1}, d, |\cdot|)$. Here onwards, we simply denote the class $A_{p(\cdot)}(\mathbb{R}^{N+1}, d, |\cdot|)$ as $A_{p(\cdot)}(\mathbb{R}^{N+1})$.

We shall require the following result on the boundedness of the maximal function on weighted variable exponent Lebesgue spaces for homogeneous spaces which we state for the ultraparabolic Lebesgue space $L^{p(\cdot)}(\mathbb{R}^{N+1}, d, |\cdot|)$.

**Proposition 5.7.** ([CUC22, Theorem 1.13]) Suppose that the variable exponent $p(\cdot)$ is globally log-Hölder continuous and satisfies (5.14) then

$$
\|(\mathcal{M}f)|_{L^{p(\cdot)}(\omega)} \leq C\|f\|_{L^{p(\cdot)}(\omega)}
$$

(5.20)

if and only if the weight $w \in A_{p(\cdot)}(\mathbb{R}^{N+1})$,

where

$$
\|f\|_{L^{p(\cdot)}(\omega)} := \|f\omega\|_{L^{p(\cdot)}(\mathbb{R}^{N+1})}
$$

**Lemma 5.8.** Let $\mathcal{G}$ be an $\eta$-sparse family.

(1) For $1 < p < \infty$, $\omega \in A_{p(\cdot)}(\mathbb{R}^{N+1})$ and $f \in L^{p(\cdot)}(\omega)$ we have

$$
\|\mathcal{A}_{\mathcal{G}} f\|_{L^{p(\cdot)}(\omega)} \leq C_{N, p, \eta, \omega} \|f\|_{L^{p(\cdot)}(\omega)}.
$$

(5.21)

(2) Let $1 < p < \infty$, $\omega \in A_{p(\cdot)}(\mathbb{R}^{N+1})$, and $b \in L^{\infty}$ with small $(\delta, R)$ norm where $R > r(\mathcal{P})$ for all $\mathcal{P} \in \mathcal{G}$. Then

$$
\|\mathcal{A}_{\mathcal{G}}^b f\|_{L^{p(\cdot)}(\omega)} + \|\mathcal{A}_{\mathcal{G}}^{b, \ast} f\|_{L^{p(\cdot)}(\omega)} \leq C_{N, p, \eta, \omega} O(\delta) \|f\|_{L^{p(\cdot)}(\omega)}
$$

holds for all $f \in L^{p(\cdot)}(\omega)$.

**Proof.** Since the proof is very similar to the proof of Lemma 5.2, we give a brief sketch of the proof.
Proof of (5.21): The proof relies on duality. Let \( g \in L^p(\omega^{-1}) \) be a non-negative function. Then
\[
\int_{\mathbb{R}^{N+1}} \mathcal{A}_g f(z) g(z) \, dz \leq \sum_{T \in \mathcal{G}} \left( \frac{1}{|T|} \int_T |f| \right) \left( \frac{1}{|T|} \int_T g \right) |T| \\
\lesssim_\eta \sum_{T \in \mathcal{G}} \left( \frac{1}{|T|} \int_T |f| \right) \left( \frac{1}{|T|} \int_T g \right) |F_T| \\
\lesssim_\eta \sum_{T \in \mathcal{G}} \inf_{z \in \mathcal{P}} \mathcal{M} f(z) \inf_{z \in \mathcal{P}} \mathcal{M} g(z) |F_T| \\
\lesssim_\eta \sum_{T \in \mathcal{G}} \int_{F_T} \mathcal{M} f(z) \mathcal{M} g(z) \, dz \\
\lesssim_\eta \int_{\mathbb{R}^{N+1}} \mathcal{M} f(z) \mathcal{M} g(z) \, dz \lesssim_\eta \| \mathcal{M} f \|_{L^p(\omega)} \| \mathcal{M} g \|_{L^{p'}(\omega^{-1})} \\
\overset{(a)}{\lesssim_\eta, \omega} \| f \|_{L^p(\omega)} \| g \|_{L^{p'}(\omega^{-1})},
\]
where (a) follows from Proposition 5.7. Now the proof is finished using duality.

Proof of (5.22): In Lemma 5.2, we have proved the following pointwise estimate
\[
\mathcal{A}_g^{h_s} f(z) \leq c_{\eta, N} O(\delta) \mathcal{A}_g (\mathcal{A}_g^s f)(z),
\]
where \( \mathcal{G}^s \) is a \( c_N \)-sparse family containing \( \mathcal{G} \). Now (5.21) implies the following
\[
\| \mathcal{A}_g^{h_s} f \|_{L^p(\omega)} \leq c_{\eta, N} O(\delta) \| \mathcal{A}_g (\mathcal{A}_g^s f) \|_{L^p(\omega)} \\
\leq c_{\eta, N, \omega} O(\delta) \| \mathcal{A}_g (f) \|_{L^p(\omega)} \\
\leq c_{\eta, N, \omega} O(\delta) \| f \|_{L^p(\omega)}.
\]
Thus the proof is complete. \( \square \)

Now, in order to prove Theorem 1.3, we begin with the following local estimate.

Proposition 5.9. There exists \( C = C(\text{data}) \) and \( r_0 = r_0(\text{data}) \) such that if \( u \in C^0(\mathbb{R}^N \times (0, \infty)) \), \( \text{spt} u \subset B_r \subset \mathbb{R}^N \times (0, \infty) \) with \( 0 < r < r_0 \), then, for \( i, j = 1, 2, \ldots, s_0 \)
\[
\| u_{x_i, x_j} \|_{L^p(\omega)} \leq C_{A, B, N, p, \eta, \omega} \| \mathcal{L} u \|_{L^p(\omega)},
\]
where the variable exponent \( p(\cdot) \) satisfies the log-Hölder condition and \( \omega \in \mathcal{A}_{p(\cdot)} \).

Proof. The proof is exactly similar to the proof of Proposition 5.3. The only difference is we apply Lemma 5.8 to the pointwise representation in Theorem 4.4. \( \square \)

Now, the proof of Theorem 1.3 is immediate.

Proof of Theorem 1.3. In a first step, one obtains the following estimate by a standard argument using cut-off functions and an interpolation lemma, for example, see [GT01, Theorem 9.11].
\[
(5.23) \quad \sup_{i,j} \| u_{x_i, x_j} \|_{L^p(\omega)} \leq C_{A, B, N, p, \eta, \omega} \left( \| \mathcal{L} u \|_{L^p(B_r, \omega)} + \frac{\| u \|_{L^p(B_{2r}, \omega)}}{r^2} \right).
\]

Now, using the fact that
\[
Yu = \mathcal{L} u - \sum_{i,j=1}^{s_0} a_{ij}(z) u_{x_i, x_j}
\]
we readily obtain through (5.23) that
Theorem 1.3. by covering $\Omega'$ with balls $B$ of radius $R/2$ where $R$ has a further restriction $R < \text{dist}(\Omega', \Omega)$, other than the restriction from Proposition 5.3.

\[ (5.24) \quad \|Yu\|_{L^p(\omega)} \leq C_{A,B,N,p,\eta,\omega} \left( \|Lu\|_{L^p(\omega)} + \frac{\|u\|_{L^p(\omega)}}{\tau^2} \right). \]

Now we deduce the estimate in Theorem 1.3 by covering $\Omega'$ with balls $B$ of radius $R/2$ where $R$ has a further restriction $R < \text{dist}(\Omega', \Omega)$, other than the restriction from Proposition 5.3.

Remark 5.10. In the next section, we will study estimates in generalized Orlicz spaces which seemingly subsume the foregoing Lebesgue and variable exponent Lebesgue spaces. However, a full-fledged theory of weights is not yet developed for the generalized Orlicz spaces and in fact, the assumptions required for the boundedness of the maximal functions on generalized Orlicz spaces do not allow for the introduction of sharp weights when restricted to specialized classes such as the variable exponent Lebesgue spaces. See [CUH18] for a related discussion. For this reason, we treat the case of variable exponent Lebesgue spaces separately.

5.3. Estimates on generalized Orlicz spaces. We state the basic facts on generalized Orlicz spaces. For details, we refer to the books [DHHR11] and [HH19]. The concept of Orlicz spaces has found a great synthesis in the work [HH19]. An important success of such spaces is reflected in the great generality in which a number of problems from regularity theory can be studied in a unified manner while often recovering optimal results in specialized cases. An instance of this may be seen in the following works [HO19, BO20, HO21].

A convex, left-continuous function $\varphi : [0, \infty) \to [0, \infty)$ satisfying $\varphi(0) = \lim_{s \to 0^+} \varphi(s) = 0$ and $\lim_{s \to \infty} \varphi(s) = \infty$ is called a $\Phi$-function. We denote the set of all $\Phi$-functions by $\Phi$. Given a subset $E \subset \mathbb{R}^{N+1}$, we define $\Phi(E)$ to be the set of all functions $\varphi : E \times [0, \infty) \to [0, \infty)$ such that $\varphi(z, \cdot) \in \Phi$ for each $z \in E$ and $\varphi(\cdot, s)$ is a measurable function for each $s \geq 0$.

The generalized Orlicz space or the Musielak-Orlicz space $L^{\varphi(\cdot)}(E)$ is defined to be the space of all functions $g : E \to \mathbb{R}$ such that the \emph{modular}

$$\rho_{\varphi(\cdot)}(g) := \int_E \varphi(z, |g(z)|) \, dz$$

is finite.

The generalized Orlicz space $L^{\varphi(\cdot)}(E)$ can be equipped with the Luxembourgh norm

\[ (5.25) \quad \|g\|_{L^{\varphi(\cdot)}(E)} := \inf \left\{ \lambda > 0 : \rho_{\varphi(\cdot)} \left( \frac{g}{\lambda} \right) \leq 1 \right\}. \]

One can impose an equivalence relation on the class $\Phi(E)$ where $\varphi \in \Phi(E)$ is equivalent to $\psi \in \Phi(E)$ if there is $L > 1$ such that $\varphi(z, t/L) \lesssim \psi(z, t) \lesssim \varphi(z, Lt)$ for all $z \in E$ and $t \geq 0$. Equivalent $\Phi$ functions give rise to the same generalized Orlicz spaces with comparable norm. Hence, in many cases, it may be sufficient to consider $\varphi \in \Phi(E)$ with better properties. For example, every $\Phi(E)$ function has a convex (in $t$) equivalent.

A function $\varphi^{-1} : [0, \infty) \to [0, \infty]$ is called the left inverse of $\varphi$ and defined by $\varphi^{-1}(s) = \inf \{ t : \varphi(t) \geq s \}$.

The function $\varphi^* : [0, \infty) \to [0, \infty]$ defined by $\varphi^*(s) = \sup_{t \geq 0} \{ ts - \varphi(t) \}$ is called the conjugate of $\varphi$.

The following Hölder’s inequality holds for generalized Orlicz spaces.

Lemma 5.11. ([HH19, Lemma 3.2.11]) Let $\varphi \in \Phi(E)$, then

$$\int_E fg \, dz \leq 2\|f\|_{L^{\varphi(\cdot)}(E)} \|g\|_{L^{\varphi^*(\cdot)}(E)}$$

for all $f \in L^{\varphi(\cdot)}(E)$ and $g \in L^{\varphi^*(\cdot)}(E)$.\]
Let $\varphi \in \Phi(E)$. For $U \subset E$, we use the notation
\[
\varphi^+_U(s) = \sup_{z \in U} \varphi(z, s), \quad \varphi^-_U(s) = \inf_{z \in U} \varphi(z, s).
\]

The notations $\varphi^+_E$, $\varphi^-_E$ will be shortened to $\varphi^+$, $\varphi^-$ respectively.

A function $g : (0, \infty) \to \mathbb{R}$ is said to be $L$-almost increasing (decreasing) with $L \geq (\leq)1$, if $g(s) \leq (\geq)Lg(t)$ for all $0 < s < t$.

Let $E$ be a domain on $\mathbb{R}^{N+1}$. We restrict the class of $\Phi(E)$ functions to those satisfying the following hypothesis.

(A0): There exists $\alpha \in (0, 1)$ such that $\varphi(z, \alpha) \leq 1 \leq \varphi(z, 1/\alpha)$ for every $z \in \mathbb{R}^{N+1}$.

(A1): There exists $\beta \in (0, 1)$ such that $\beta^{-1}(z, s) \leq \varphi^{-1}(z', s)$

for every $s \in \left[1, \frac{1}{|B|}\right]$ almost everywhere $z, z' \in B \cap E$ and every ball $B$ such that $|B| \leq 1$.

(A2): For every $s > 0$, there exists $\beta \in (0, 1]$ and $h \in L^1(E) \cap L^\infty(E)$ such that $\varphi(\zeta, \beta t) \leq \varphi(\eta, t) + h(\zeta) + h(\eta)$

for almost all $\zeta, \eta \in E$ and $\varphi(y, t) \in [0, s]$.

(aInc): There exists $p > 1$ and $L \geq 1$ such that $\frac{\varphi(z, s)}{s^p}$ is $L$-almost increasing with respect to $s$ for every $z \in E$.

(aDec): There exists $p > 1$ and $L \leq 1$ such that $\frac{\varphi(z, s)}{s^p}$ is $L$-almost decreasing with respect to $s$ for every $z \in E$.

Remark 5.12. If $E$ is bounded, (A2) is superfluous. See [HH19, Lemma 4.2.3]

We shall require the following lemma whose proof may be found in the statements of [HH19, Lemma 3.7.6], [HH19, Lemma 4.1.7], [HH19, Lemma 4.2.4] and [HH19, Proposition 2.4.9].

Lemma 5.13. Let $E$ be an open set in $\mathbb{R}^{N+1}$. Let $\varphi \in \Phi(E)$. If $\varphi$ satisfies (A0), (A1), (A2), (aInc) and (aDec), then $\varphi^*$ satisfies (A0), (A1), (A2), and (aInc)*.

The estimates for sparse operators will require the boundedness of the maximal function on the generalized Orlicz spaces. The proof is deferred to Appendix A.

Lemma 5.14. Let $E$ be an open set in $\mathbb{R}^{N+1}$ and let $\varphi \in \Phi(E)$. If $\varphi$ satisfies (A0), (A1), (A2), (aInc) and (aDec) with $1 < p \leq q < \infty$, then we have
\[
\|\mathcal{M} f\|_{L^{\varphi^*(1)}(E)} \leq C\|f\|_{L^{\varphi^*(1)}(E)}, \quad f \in L^{\varphi^*(1)}(E)
\]
\[
(5.26)
\]
\[
\|\mathcal{M} f\|_{L^{\varphi^*(1)}(E)} \leq C\|f\|_{L^{\varphi^*(1)}(E)}, \quad f \in L^{\varphi^*(1)}(E).
\]
(5.27)

Remark 5.15. As remarked upon earlier, the theory of generalized Orlicz spaces is an important synthesis with regards to the regularity theory of partial differential equations modelled on operators with Orlicz growth as they encompass a great variety of growth conditions such as
\begin{itemize}
  \item $\varphi(z, s) = a(z)s^p(z)$.
  \item $\varphi(z, s) = s^p(z) \log(e + s)$.
  \item $\varphi(z, s) = s^p + a(z)s^q$.
  \item $\varphi(z, s) = s^p + a(z)s^p \log(e + s)$.
\end{itemize}

In particular, in the paper [HO21], essentially optimal regularity results for these classes are recovered under optimal criteria on the variable exponent $p(z)$ and the phase switching factor $a(z)$ that matches the prescribed conditions (A0), (A1), (A2), (aInc) and (aDec).
Remark 5.16. It is worthwhile to note that regularity theory for parabolic equations in the case of Orlicz growth is not very well-developed. For example, the sharp parabolic analogues of [CM15b, CM15a] are not known. Hence, the maximal function estimates in parabolic settings require a different approach that treats the time and space variables separately.

Lemma 5.17. Let $\mathcal{G}$ be an $\eta$-sparse family. Let $\varphi \in \Phi(E)$. If $\varphi$ satisfies (A0), (A1), (A2), (aInc), and (aDec) with $1 < p \leq q < \infty$, then for all $f \in L^{p\ast}(E)$, we have

\begin{equation}
\|\mathcal{A}_g f\|_{L^{p\ast}(E)} \leq C_{N,p,q,\alpha,\beta}\|f\|_{L^{p\ast}(E)}.
\end{equation}

(2) Further, let $b \in L^\infty$ with small $(\delta, \mathcal{R})$ norm where $\mathcal{R} > r(\mathcal{P})$ for all $\mathcal{P} \in \mathcal{G}$. Then, we have

\begin{equation}
\|\mathcal{A}_g b f\|_{L^{p\ast}(E)} + \|\mathcal{A}_g b^* f\|_{L^{p\ast}(E)} \leq C_{N,p,q,\alpha,\beta}O(\delta)\|f\|_{L^{p\ast}(E)}.
\end{equation}

Proof. Since the proof is very similar to the proof of Lemma 5.2, we give a brief sketch of the proof. Proof of (5.28). The proof relies on duality. Let $g \in L^{p\ast}(E)$ be a non-negative function. Then

\begin{equation}
\int_{\mathbb{R}^{N+1}} A_g f(z)g(z)\,dz \leq \sum_{T \in \mathcal{G}} \left( \frac{1}{|T|} \int_T |f| \right) \left( \frac{1}{|T|} \int_T g \right) |T|
\end{equation}

\begin{align*}
&\leq \eta \sum_{T \in \mathcal{G}} \left( \frac{1}{|T|} \int_T |f| \right) \left( \frac{1}{|T|} \int_T g \right) |F_T|
&\leq \eta \sum_{T \in \mathcal{G}} \inf_{z \in \mathcal{P}} Mf(z) \inf_{z \in \mathcal{P}} Mg(z) |F_T|
&\leq \eta \sum_{T \in \mathcal{G}} \int_{F_T} Mf(z) Mg(z) \,dz
\end{align*}

\begin{align*}
&\leq \eta \int_{\mathbb{R}^{N+1}} Mf(z) Mg(z) \,dz \lesssim \|
Mf\|_{L^{p\ast}(\mathbb{R}^{N+1})} \|Mg\|_{L^{p\ast}(\mathbb{R}^{N+1})}
\end{align*}

\begin{align*}
&\lesssim \eta \omega \|f\|_{L^{p\ast}(\mathbb{R}^{N+1})} \|g\|_{L^{p\ast}(\mathbb{R}^{N+1})},
\end{align*}

where (b) follows from Lemma 5.11 and (c) follows from Lemma 5.14. Now the proof is finished using duality ([HI19, Theorem 3.4.6]).

Proof of (5.29). In Lemma 5.2, we have proved the following pointwise estimate

\begin{equation}
\mathcal{A}_g^b h^*(z) \leq c_{n,N}O(\delta)\mathcal{A}_g (\mathcal{A}_g h^*)^*(z),
\end{equation}

where $\mathcal{G}$ is a $\eta$-sparse family containing $\mathcal{G}$. Now (5.28) implies the following

\begin{equation}
\|\mathcal{A}_g b^* f\|_{L^{p\ast}(E)} \leq c_{n,N}O(\delta)\|\mathcal{A}_g (\mathcal{A}_g h^*)^*\|_{L^{p\ast}(E)}
\end{equation}

\begin{equation}
\leq c_{n,N,\omega}O(\delta)\|f\|_{L^{p\ast}(E)}
\end{equation}

Thus the proof is complete. \qed

Proposition 5.18. Let $\varphi$ satisfies (A0), (A1), (A2), (aInc), and (aDec) with $1 < p \leq q < \infty$. There exists $C = C(\text{data})$ and $r_0 = r_0(\text{data})$ such that if $u \in C^\infty_0(\mathbb{R}^N \times (0, \infty))$, $\text{spt} u \subset B_r \subset \mathbb{R}^N \times (0, \infty)$ with $0 < r < r_0$, then, for $i, j = 1, 2, \ldots, s_0$

\begin{equation}
\|u_{x_i x_j}\|_{L^{p\ast}(B_r)} \leq C_{A,B,N,p,\alpha,\beta}\|\mathcal{L}u\|_{L^{p\ast}(B_r)}.
\end{equation}

Proof. The proof is exactly similar to the proof of Proposition 5.3. The only difference is we apply Lemma 5.17 to the pointwise representation in Theorem 4.4. \qed
Now, the proof of Theorem 1.2 is immediate.

**Proof of Theorem 1.2.** In a first step, one obtains the following estimate by a standard argument using cut-off functions and an interpolation lemma, for example, see [GT01, Theorem 9.11].

\[
\sup_{i,j} \| u_{x_i x_j} \|_{L^\varphi(B_r)} \leq C_{A,B,N,p,q,\alpha,\beta} \left( \| Lu \|_{L^\varphi(B_{2r})} + \|u\|_{L^\varphi(B_{2r})} \right). \tag{5.30}
\]

Now, using the fact that \( Yu = Lu - \sum_{i,j=1}^{s_0} a_{ij}(z) \partial_{x_i x_j} u \)
we readily obtain through (5.30) that

\[
\| Yu \|_{L^\varphi(B_r)} \leq C_{A,B,N,p,q,\alpha,\beta} \left( \| Lu \|_{L^\varphi(B_{2r})} + \|u\|_{L^\varphi(B_{2r})} \right). \tag{5.31}
\]

Now we deduce the estimate in Theorem 1.2 by covering \( \Omega' \) with balls \( B \) of radius \( R/2 \) where \( R \) has a further restriction \( R < \text{dist}(\Omega', \Omega) \), other than the restriction from Proposition 5.18.

\[\square\]

**Appendix A. Boundedness of maximal function on Generalized Orlicz spaces**

We sketch below the proof of boundedness of maximal functions in the case of \( \varphi \in \Phi(\mathbb{R}^{N+1}) \).

The case of \( E \subset \mathbb{R}^{N+1} \) is similar. We require the **unit ball property** which asserts that

**Lemma A.1.** Let \( \varphi \in \Phi(\mathbb{R}^{N+1}) \). Then

\[
\| f \|_{L^\varphi(\mathbb{R}^{N+1})} < 1 \implies \rho_{\varphi(\cdot)}(f) \leq 1 \implies \| f \|_{L^\varphi(\mathbb{R}^{N+1})} \leq 1.
\]

**Lemma A.2.** Let \( \varphi \in \Phi(\mathbb{R}^{N+1}) \). If \( \varphi \) satisfies (A0), (A1), (A2), (aInc\_p) and (aDec\_q) with \( 1 < p \leq q < \infty \), then we have

\[
\| \mathcal{M}f \|_{L^\varphi(\mathbb{R}^{N+1})} \leq C \| f \|_{L^\varphi(\mathbb{R}^{N+1})} \quad f \in L^\varphi(\mathbb{R}^{N+1}) \quad \text{(A.1)}
\]

\[
\| \mathcal{M}f \|_{L^{\varphi^*}(\mathbb{R}^{N+1})} \leq C \| f \|_{L^\varphi(\mathbb{R}^{N+1})} \quad f \in L^{\varphi^*}(\mathbb{R}^{N+1}) \quad \text{(A.2)}
\]

**Proof.** **Step 1** (Boundedness of maximal function on ultraparabolic \( L^p \) spaces):

As described in the beginning of Subsection 4.1, there exist finitely many dyadic grids \( \{ \mathcal{U}^t \}_{t=1}^L \) such that

\[
\mathcal{M} f \lesssim \sum_{t=1}^L \mathcal{M}_{\mathcal{U}^t} f,
\]

where \( \mathcal{M}_{\mathcal{U}^t} \) is a dyadic maximal function. Now, the boundedness of maximal function follows from the boundedness of a dyadic maximal function whose proof may be found in [HK12, Eq. 7.8]. Recall that we are strongly using the fact that the ultraparabolic cubes are comparable to the balls defined in (2.3).

**Step 2** (Key estimate):

The following lemma from [HH19] goes through as is for the homogeneous space \( (\mathbb{R}^{N+1}, d, | \cdot |) \).
Theorem A.3. ([HH19, Theorem 4.3.2]) Let \( \varphi \in \Phi(\mathbb{R}^{N+1}) \). If \( \varphi \) satisfies (A0), (A1), (A2), (aInc\(p\)) with \( 1 < p < \infty \), then there exists \( \beta > 0 \) and \( h \in L^1(\mathbb{R}^{N+1}) \cap L^\infty(\mathbb{R}^{N+1}) \) such that

\[
(A.4) \quad \varphi \left( x, \beta \int_B |f| \, dy \right)^{\frac{1}{p}} \leq \frac{1}{\beta} \int_B \varphi(y, |f|)^{\frac{1}{p}} \, dy + h(x)^{\frac{1}{p}} \int_B \, h(y)^{\frac{1}{p}} \, dy,
\]

for every ball \( B \subset \mathbb{R}^{N+1} \), \( x \in B \) and \( f \in L^{p(\cdot)}(\mathbb{R}^{N+1}) \) such that \( \rho_{\varphi(\cdot)}(f) \leq 1 \).

Step 3: Taking supremum over all balls in the inequality (A.4), we obtain

\[
(A.5) \quad \varphi \left( x, \beta \mathcal{M}f(x) \right)^{\frac{1}{p}} \leq \mathcal{M}(\varphi(\cdot, f)^{\frac{1}{p}})(x) + \mathcal{M}(h^{\frac{1}{p}})(x),
\]

for every \( x \in \mathbb{R}^{N+1} \) and \( f \in L^{p(\cdot)}(\mathbb{R}^{N+1}) \) such that \( \rho_{\varphi(\cdot)}(f) \leq 1 \). Here, we use the fact that \( h(x)^{\frac{1}{p}} \leq \mathcal{M}(h^{\frac{1}{p}})(x) \).

Step 4: Let \( f \in L^{p(\cdot)}(\mathbb{R}^{N+1}) \) such that \( \|f\|_{L^{p(\cdot)}(\mathbb{R}^{N+1})} > 0 \). Choose \( \varepsilon := \frac{1}{2\|f\|_{L^{p(\cdot)}(\mathbb{R}^{N+1})}} \), then by Lemma A.1, it holds that \( \rho_{\varphi(\cdot)}(\varepsilon f) \leq 1 \) so that we may apply (A.5) to obtain

\[
(A.6) \quad \varphi \left( x, \beta \varepsilon \mathcal{M}f(x) \right)^{\frac{1}{p}} \leq \mathcal{M}(\varphi(\cdot, \varepsilon f)^{\frac{1}{p}})(x) + \mathcal{M}(h^{\frac{1}{p}})(x).
\]

Raising both sides of (A.6) to power \( p \) and integrating over \( \mathbb{R}^{N+1} \), we get

\[
\int_{\mathbb{R}^{N+1}} \varphi \left( x, \beta \varepsilon \mathcal{M}f(x) \right) \, dx \leq \left( \int_{\mathbb{R}^{N+1}} \mathcal{M}(\varphi(\cdot, \varepsilon f)^{\frac{1}{p}})(x) \right)^p \, dx + \left( \int_{\mathbb{R}^{N+1}} \mathcal{M}(h^{\frac{1}{p}})(x) \right)^p \, dx
\]

\[
\leq \int_{\mathbb{R}^{N+1}} \varphi(\cdot, \varepsilon f)(x) \, dx + \int_{\mathbb{R}^{N+1}} h(x) \, dx
\]

\[
= \rho_{\varphi(\cdot)}(f) + \|h\|_{L^1(\mathbb{R}^{N+1})}
\]

\[
\leq 1 + \|h\|_{L^1(\mathbb{R}^{N+1})} = C_1,
\]

where in the second inequality, we invoke boundedness of maximal function on \( L^p(\mathbb{R}^{N+1}) \) from Step 1. The inequality (A.7) is the same as \( \rho_{\varphi(\cdot)}(\beta \varepsilon \mathcal{M}f) \leq C_1 \) so that by aInc\(1\)(which is a consequence of convexity and \( \phi(\cdot, 0) = 0 \), we have \( \rho_{\varphi(\cdot)}(\frac{\beta}{\varepsilon} \varepsilon \mathcal{M}f) \leq 1 \). Once again, applying Lemma A.1, we get

\[
\|\frac{\beta}{\varepsilon} \varepsilon \mathcal{M}f\|_{L^{p(\cdot)}} \leq 1 \text{ so that } \|\mathcal{M}f\|_{L^{p(\cdot)}} \leq \frac{C_1 L}{\|h\|_{L^1(\mathbb{R}^{N+1})}} = \frac{2C_1 L}{\beta} \|f\|_{L^{p(\cdot)}} ,
\]

which is the required result.

Step 5: The result for \( \varphi^*(\cdot) \) follows by the fact that it satisfies (A0), (A1), (A2), and aInc\(q\).

Acknowledgement. We are thankful to Dr. Karthik Adimurthi and Dr. Agnid Banerjee for many fruitful discussions.

References

[AB21] The Anh Bui, Regularity estimates for nondivergence parabolic equations on generalized Orlicz spaces, Int. Math. Res. Not. IMRN (2021), no. 14, 11103–11139.

[Aim85] H. Aimar, Singular integrals and approximate identities on spaces of homogeneous type, Trans. Amer. Math. Soc. 292 (1985), no. 1, 135–153.

[AIS01] Kari Astala, Tadeusz Iwaniec, and Eero Saksman, Beltrami operators in the plane, Duke Math. J. 107 (2001), no. 1, 27–56.

[AM01] Emilio Acerbi and Giuseppe Mingione, Regularity results for a class of functionals with non-standard growth, Arch. Ration. Mech. Anal. 156 (2001), no. 2, 121–140.

[AM07] Emilio Acerbi and Giuseppe Mingione, Gradient estimates for a class of parabolic systems, Duke Math. J. 136 (2007), no. 2, 285–320. MR 2286632
Francesca Anceschi, Silvia Muzzioli, and Sergio Polidoro, Existence of a fundamental solution of partial differential equations associated to Asian options, Nonlinear Anal. Real World Appl. 62 (2021), Paper No. 103373, 29. MR 4273088

Francesca Anceschi and Sergio Polidoro, A survey on the classical theory for Kolmogorov equation, Matematiche (Catania) 75 (2020), no. 1, 221–258. MR 4069607

Farhan Abedin and Giulio Tralli, Harnack inequality for a class of Kolmogorov-Fokker-Planck equations in non-divergence form, Arch. Ration. Mech. Anal. 233 (2019), no. 2, 867–900. MR 3951695

R. Alexandre and C. Villani, On the Landau approximation in plasma physics, Ann. Inst. H. Poincaré C Anal. Non Linéaire 21 (2004), no. 1, 61–95. MR 2037247

J. August and S.W. Zucker, Sketches with curvature: the curve indicator random field and markov processes, IEEE Transactions on Pattern Analysis and Machine Intelligence 25 (2003), no. 4, 387–400.

G. Barles, Convergence of numerical schemes for degenerate parabolic equations arising in finance theory, Numerical methods in finance, Publ. Newton Inst., vol. 13, Cambridge Univ. Press, Cambridge, 1997, pp. 1–21. MR 1470506

Cristina Benea and Frédéric Bernicot, Conservation de certaines propriétés à travers un contrôle épars d’un opérateur et applications au projecteur de Leray-Hopf, Ann. Inst. Fourier (Grenoble) 68 (2018), no. 6, 2329–2379.

The Quan Bui, The Anh Bui, and Xuan Thinh Duong, Global regularity estimates for non-divergence elliptic equations on weighted Lebesgue spaces, Commun. Contemp. Math. 23 (2021).

Marco Bramanti and M. Cristina Cerutti, $W^{1,2}_p$ solvability for the Cauchy-Dirichlet problem for parabolic equations with VMO coefficients, Comm. Partial Differential Equations 18 (1993), no. 9-10, 1735–1763. MR 1239929

M. Bramanti, G. Cupini, E. Lanconelli, and E. Priola, Global $L^p$ estimates for degenerate Ornstein-Uhlenbeck operators with variable coefficients, Math. Nachr. 286 (2013), no. 11-12, 1087–1101. MR 3092273

M. Bramanti, M. C. Cerutti, and M. Manfredini, $L^p$ estimates for some ultraparabolic operators with discontinuous coefficients, J. Math. Anal. Appl. 200 (1996), no. 2, 332–354.

Sun-Sig Byun and Mikyoung Lee, On weighted $W^{2,p}$ estimates for elliptic equations with BMO coefficients in nondivergence form, Internat. J. Math. 26 (2015), no. 1, 1550001, 28.

Sun-Sig Byun, Mikyoung Lee, and Jihoon Ok, $W^{2,p(\cdot)}$-regularity for elliptic equations in nondivergence form with VMO coefficients, Math. Ann. 363 (2015), no. 3-4, 1023–1052.

M. Bramanti, G. Cupini, E. Lanconelli, and E. Priola, Nondivergence parabolic equations in weighted variable exponent spaces, Trans. Amer. Math. Soc. 370 (2018), no. 4, 2263–2298.

Sun-Sig Byun and Jehan Oh, Regularity results for generalized double phase functionals, Anal. PDE 13 (2020), no. 5, 1269–1300. MR 4149062

Marco Bramanti, An invitation to hypoelliptic operators and Hörmander’s vector fields, SpringerBriefs in Mathematics, Springer, Cham, 2014. MR 3154431

Sun-Sig Byun and Lihe Wang, Elliptic equations with BMO nonlinearity in Reifenberg domains, Adv. Math. 219 (2008), no. 6, 1937–1971.

Marco Bramanti and Maochun Zhu, $L^p$ and Schauder estimates for nonvariational operators structured on Hörmander vector fields with drift, Anal. PDE 6 (2013), no. 8, 1793–1855. MR 3198585

Luis A. Caffarelli, Interior a priori estimates for solutions of fully nonlinear equations, Ann. of Math. (2) 130 (1989), no. 1, 189–213. MR 1005611

Luis A. Caffarelli, Interior $W^{2,p}$ estimates for solutions of the Monge-Ampère equation, Ann. of Math. (2) 131 (1990), no. 1, 135–150.

Filippo Chiarenza, Michele Frasca, and Placido Longo, Interior $W^{2,p}$ estimates for nondivergence elliptic equations with discontinuous coefficients, Ricerche Mat. 40 (1991), no. 1, 149–168. MR 1191890

Filippo Chiarenza, Michele Frasca, and Placido Longo, $W^{2,p}$-solvability of the Dirichlet problem for nondivergence elliptic equations with VMO coefficients, Trans. Amer. Math. Soc. 336 (1993), no. 2, 841–853. MR 1088476

S. Chandrasekhar, Stochastic problems in physics and astronomy, Rev. Mod. Phys. 15 (1943), 1–89.

Maria Colombo and Giuseppe Mingione, Bounded minimisers of double phase variational integrals, Arch. Ration. Mech. Anal. 218 (2015), no. 1, 219–273. MR 3360738

Regularity for double phase variational problems, Arch. Ration. Mech. Anal. 215 (2015), no. 2, 443–496. MR 3204408

Chiara Cinti, Kaj Nyström, and Sergio Polidoro, A note on Harnack inequalities and propagation sets for a class of hypoelliptic operators, Potential Anal. 33 (2010), no. 4, 341–354.

Chiara Cinti, Kaj Nyström, and Sergio Polidoro, A boundary estimate for non-negative solutions to Kolmogorov operators in non-divergence form, Ann. Mat. Pura Appl. (4) 191 (2012), no. 1, 1–23.
[Kry07] N. V. Krylov, Parabolic and elliptic equations with VMO coefficients, Comm. Partial Differential Equations 32 (2007), no. 1-3, 453–475. MR 2304157

[Kup72] L. P. Kupcov, The fundamental solutions of a certain class of elliptic-parabolic second order equations, Differential`nye Urvanenija 8 (1972), 1649–1660, 1716. MR 0315290

[Lio94] P.-L. Lions, On Boltzmann and Landau equations, Philos. Trans. Roy. Soc. London Ser. A 346 (1994), no. 1679, 191–204. MR 1275244

[LO20] Andrei K. Lerner and Sheldy Ombrosi, Some remarks on the pointwise sparse domination, J. Geom. Anal. 30 (2020), no. 1, 1011–1027.

[Lor21] Emiel Lorist, On pointwise ℓr-sparse domination in a space of homogeneous type, J. Geom. Anal. 31 (2021), no. 9, 9366–9405. MR 4362224

[LORR17] Andrei K. Lerner, Sheldy Ombrosi, and Israel P. Rivera-Ríos, Some remarks on the pointwise sparse domination, J. Geom. Anal. 30 (2020), no. 1, 1011–1027.

[LP94] E. Lanconelli and S. Polidoro, On a class of hypoelliptic evolution operators, vol. 52, 1994, Partial differential equations, II (Turin, 1993), pp. 29–63. MR 1289901

[Man97] Maria Manfredini, The Dirichlet problem for a class of ultraparabolic equations, Adv. Differential Equations 2 (1997), no. 5, 831–866. MR 1751429

[MP98] Maria Manfredini and Sergio Polidoro, Interior regularity for weak solutions of ultraparabolic equations in divergence form with discontinuous coefficients, Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat. (8) 1 (1998), no. 3, 651–675. MR 1662349

[Mir63] Carlo Miranda, Sulle equazioni ellittiche del secondo ordine di tipo non variazionale, a coefficienti discontinui, Ann. Mat. Pura Appl. (4) 63 (1963), 35–138. MR 170090

[MP98] Maria Manfredini and Sergio Polidoro, Interior regularity for weak solutions of ultraparabolic equations in divergence form with discontinuous coefficients, Boll. Unione Mat. Ital. Sez. B Artic. Ric. Mat. (8) 1 (1998), no. 3, 651–675. MR 1662349

[Mum94] David Mumford, Elastica and computer vision, Algebraic geometry and its applications (West Lafayette, IN, 1990), Springer, New York, 1994, pp. 491–506. MR 1272050

[NSW85] Alexander Nagel, Elias M. Stein, and Stephen Wainger, Balls and metrics defined by vector fields. I. Basic properties, Acta Math. 155 (1985), no. 1-2, 103–147. MR 793239

[NZ22] Lukas Niebel and Rico Zacher, Kinetic maximal Lp-regularity with temporal weights and application to quasilinear kinetic diffusion equations, J. Differential Equations 307 (2022), 29–82. MR 4336467

[Pol94] Sergio Polidoro, On a class of ultraparabolic operators of Kolmogorov-Fokker-Planck type, Matematiche (Catania) 49 (1994), no. 1, 53–105 (1995). MR 1386366

[PR98] Sergio Polidoro and Maria Alessandra Ragusa, Sobolev-Morrey spaces related to an ultraparabolic equation, Manuscripta Math. 96 (1998), no. 3, 371–392. MR 1638177

[RS76] Linda Preiss Rothschild and E. M. Stein, Hypoelliptic differential operators and nilpotent groups, Acta Math. 137 (1976), no. 3-4, 247–320. MR 436223

[Sil22] Luis Silvestre, Regularity estimates and open problems in kinetic equations, arXiv:2204.06401 [math] (2022).

[Ste71] E. M. Stein, Some problems in harmonic analysis suggested by symmetric spaces and semi-simple groups, Actes du Congrès International des Mathématiciens (Nice, 1970), Tome 1, 1971, pp. 173–189. MR 0578903

[Ter05] S. A. Tersenov, On the Cauchy problem for a system of equations of ultraparabolic type, Mat. Zametki 77 (2005), no. 5, 768–774. MR 2178846

[Web51] Maria Weber, The fundamental solution of a degenerate partial differential equation of parabolic type, Trans. Amer. Math. Soc. 71 (1951), 24–37. MR 42035

[Yao12] Fengping Yao, Second-order elliptic equations of nondivergence form with small BMO coefficients in Rn, Potential Anal. 36 (2012), no. 4, 557–568. MR 2904633

[Zhi86] V. V. Zhikov, Averaging of functionals of the calculus of variations and elasticity theory, Izv. Akad. Nauk SSSR Ser. Mat. 50 (1986), no. 4, 675–710, 877. MR 864171

Email address: abhi21@tifrbng.res.in

Email address: vivek2020@tifrbng.res.in

Tata Institute of Fundamental Research, Centre for Applicable Mathematics, Bangalore–560065, Karnataka, India.