From the holdings of the Department of Special Collections , Princeton University Library, Princeton, NJ

These documents are either in the public domain, or can only be used for educational and research purposes (Fair use) as per U.S. Copyright law (text below). By accessing this file, all users agree that their use falls in the public domain or within fair use as defined by the copyrightlaw. We do not charge any permission or usage fees for the publication of images of material inour collections, including those provided by Princeton University Library, via our website, catalog, or directly from staff.For more information regarding copyright, credit, and citation please review the information onthe Copyright, Credit and Citation Guidelines page of our website.
https://library.princeton.edu/special-collections/policies/copyright-credit-and-citation-guidelines. If you have any questions, including how to order higher-quality images from the collection, contact Special Collections Public Services via our Ask Us! Form. https://library.princeton.edu/special-collections/ask-us

Doctoral Dissertations, 2011-2026 Mathematics

2023

On the sharpness of the Assouad embedding theorem for finitely generated groups of polynomial growth and nilpotent Lie groups

Ryoo, Seung-Yeon

Princeton, NJ : Princeton University

http://arks.princeton.edu/ark:/88435/dsp010p096b180

U.S. Copyright law: The copyright law of the United States (Title 17, §108, United States Code) governs the making of photocopies or other reproductions of copyrighted material. Under certain conditions specified in the law, libraries and archives are authorized to furnish a photocopy or other reproduction. One of these specified conditions is that the photocopy or reproduction is not to be used for any purpose other than private study, scholarship, or research. If a user makes a request for, or later uses, a photocopy or reproduction for purposes in excess of fair use that user may be liable for copyright infringement. This institution reserves the right to refuse to accept a copying order if, in its judgment, fulfillment of the order would involve a violation of copyright law. https://www.copyright.gov/title17/title17.pdf

On the sharpness of the Assouad embedding theorem for finitely generated groups of polynomial growth and nilpotent Lie groups

Seung-Yeon Ryoo

A Dissertation
Presented to the Faculty
of Princeton University
in Candidacy for the Degree
of Doctor of Philosophy

Recommended for Acceptance
by the Department of
Mathematics
Adviser: Assaf Naor

May 2023

© Copyright by Seung-Yeon Ryoo, 2023.

All Rights Reserved

Abstract

Assouad's embedding theorem states that a proper snowflake of a doubling metric space embeds bilipschitzly into a Euclidean space. We investigate the sharpness of the bilipschitz distortion and target dimension of the Assouad embedding theorem for not virtually abelian finitely generated groups of polynomial growth with left-invariant word metric and nonabelian simply connected nilpotent Lie groups with left-invariant sub-Finsler Carnot-Carathéodory metric. We conclude that the distortion, namely $O(\varepsilon^{-1/2})$ for $(1-\varepsilon)$-snowflakes, cannot be improved, whereas the target dimension, namely $O(\varepsilon^{-O(1)})$, can be improved to $O(1)$ while maintaining distortion $O(\varepsilon^{-1/2})$.

We establish the optimal target dimension in Chapter 1 by building on the approach of T. Tao (2020) for the Heisenberg group. Namely, we use a Nash–Moser iteration scheme along with an extension theorem for orthonormal vectorfields, the proof of which combines the Lovász local lemma and the concentration of measure phenomenon on the sphere and is the critical novelty of Chapter 1. We expect these embedding techniques to find use in metric dimension reduction.

We establish the optimality of the distortion in Chapter 2 via 'vertical versus horizontal inequalities,' which quantify the collapse of Lipschitz functions from the above groups into uniformly convex Banach spaces, which was qualitatively described by the Pansu–Semmes nonembeddability argument by Cheeger and Kleiner (2006) and Lee and Naor (2006). We prove the inequalities by using the vector-valued Littlewood–Paley–Stein theory approach of Lafforgue and Naor (2014) and by establishing a version of the classical Dorronsoro theorem inspired by Fässler and Orponen (2020).

As a consequence, we compute the bilipschitz distortion of a ball of radius $n \geq 2$ in the above groups, where the Lie group metric is Riemannian, into Euclidean spaces of sufficiently large dimension or into L^p $(1 < p < \infty)$ spaces to be $\Theta(\sqrt{\log n})$, and into $q(\geq 2)$-uniformly convex Banach spaces to be $\Omega((\log n)^{1/q})$.

Acknowledgements

I thank my advisor, Professor Assaf Naor, for guidance over the past five years. Throughout my doctoral studies, he has provided deep mathematical insight, valuable advice on academic and professional issues, and warm encouragement during difficult times. I also thank him for leading me into the field of metric embeddings, which turned out to be a fascinating trove of phenomena yet to be fully deciphered.

I would like to thank Professors Alexandru Ionescu and Ramon van Handel for participating in my thesis committee and Professor Charles Fefferman for undertaking the tedious responsibility of serving as a thesis reader.

I would like to sincerely thank Jill LeClair for helping me navigate issues in my Ph.D. studies, Ankit Tak for helping with reimbursements, and all the staff of the Mathematics Department for doing their best to make my life in Princeton easier.

I gratefully acknowledge financial support by the Korean Foundation for Advanced Studies. I also thank the Hyperbolic and Kinetic Equations (HYKE) research group at Seoul National University (SNU) for their hospitality during my visit in the fall semester of 2020 and the spring semester of 2021.

My thanks also go to Professors Seung-Yeal Ha, Sang-hyun Kim, Otto van Koert, Seonhee Lim, Terence Tao, and Robert Young for their generous hospitality for when I was visiting their departments. Discussions with them broadened my perspectives on mathematics.

I have tremendously benefited from many other scholars. In particular, I would like to thank my colleagues Alan Chang, Alexandros Eskenazis, Alper Gunes, Otte Heinävaara, Cosmas Kravaris, and Kevin Ren.

I thank Jonathan Zung for his beamer source code and Fabian Gundlach for his sauklaue program which inspired me to take advantage of the aspect ratio of the Fine Hall beam projectors during my Final Public Oral Examination.

I would like to thank my friends that I met in Princeton: Amina Abdurrahman,

Dylan Airey, Ryan Alweiss, Dimitar Chakarov, Eric Y. Chen, Sungjae Chi, Seokyoon Chang, Elena Cho, Kun Woo Cho, Dahyun Choi, Yunseo Choi, Hyungjun Choi, Dor Elboim, Ian Fleschler, Vikram Giri, Linus Hamann, Sokol Hoxha, Jihye Jeon, Sanghak Jeon, Lena Ji, Kexin Jin, Sal Kang, Boeun Kim, Myeonhu Kim, Sabrina Kim, Cooper Kofron, Say-Yeon Kwon, Joonhyun La, Fizay-Noah Lee, Lynn Lee, Sanghoon Lee, Seok Hyeong Lee, Yangyang Li, Maycee Lin, Loki Lin, Andi Liu, Ben Lowe, Vy Luu, Xiao Ma, Thomas Massoni, Eden Michael, Donghoon Nam, Gyujin Oh, Juhyun Park, Sang Woo Park, Adele Peng, Rohil Prasad, Jia Shi, Gawoon Shim, Laura Shou, Ron Shvartsman, Anna Skorobogatova, Seung Wook So, Shruthi Sridhar, Kevin Suh, Seoyeon Yang, Tony Ye, So Hye Yoon, Mengxin Yu, Yoonseo Zoh, and many more not listed here; the HYKE members during my time attending and visiting SNU: Professor Hyunjin Ahn, Junhyeok Byeon, Hangjun Cho, Professor Young-Pil Choi, Gyuyoung Hwang, Professor Jinwook Jung, Myeongju Kang, Professor Doheon Kim, Professor Dohyun Kim, Professor Hwa Kil Kim, Professor Jeongho Kim, Professor Dongnam Ko, Eun Taek Lee, Professor Chan Ho Min, Bora Moon, Hansol Park, Professor Jinyeong Park, So Young Park, Professor Woojoo Shim, Jaeyoung Yoon, and many more not listed; many other friends studying abroad in North America, including but not limited to Taejoo Ahn, Jaedong Hwang, Naehoon Kang, Brianna Kim, Seongsoo Kim, Yunhyae Kim, Dongyeong Ko, Hyunwoo Kwon, Seewoo Lee, Daniel de la Riva Massaad, Wookyeong Song, and Joon-Hyeok Yim; and my finite but large ensemble of friends back in Korea, who are too numerous to list here. Together they have formed tightly-knit communities which have provided strong emotional support.

I dedicate this dissertation to my mother, who has guided me with love throughout my life and who has passed away during my doctoral studies.

사랑하는 어머니께 이 박사학위논문을 바칩니다.

Contents

Abstract iii

Acknowledgements iv

0 Introduction **1**

0.1 Characterization of bilipschitz embeddability into Euclidean spaces: the Lang–Plaut question and the Assouad embedding theorem 2

0.2 Summary of Chapter 1: Sharp Assouad theorem for finitely generated groups of polynomial growth and simply connected nilpotent Lie groups 6

0.3 Summary of Chapter 2: Vertical versus horizontal inequalities 7

1 Embedding finitely generated groups of polynomial growth and nilpotent Lie groups into Euclidean space **13**

1.1 Preliminaries 29

1.1.1 Reduction of Theorem 1.0.1 to Theorem 1.0.2, and reduction of Theorem 1.0.2 to the case of nonabelian simply connected nilpotent Lie groups 29

1.1.2 Hierarchy of constants 33

1.1.3 Basic linear algebra 33

1.1.4 Some metric space geometry 34

1.1.5 Function spaces on nilpotent Lie groups 35

1.1.6 Basic metric geometry of simply connected nilpotent Lie groups 37

1.1.7 Littlewood–Paley theory on nilpotent Lie groups 44
1.2 Nash–Moser Perturbation for a bilinear form 55
1.3 Regular extensions of orthonormal systems 79
1.4 Main Iteration Lemma 95
1.5 Construction of the embedding 120

2 Vertical versus horizontal inequalities on finitely generated groups of polynomial growth and simply connected nilpotent Lie groups 139

2.0.1 Dorronsoro's theorem on Carnot groups 160
2.0.2 L^1-distortion: vertical perimeter versus horizontal perimeter . 166
2.1 Proof of Theorem 2.0.5: differentiation along random paths 170
2.1.1 The case $p = q$ 177
2.1.2 The case $p < q$ 184
2.2 Proof of Theorems 2.0.1, 2.0.2, 2.0.10, 2.0.11, 2.0.14, and 2.0.15 and Corollaries 2.0.3 and 2.0.4: distortion bounds 186
2.3 Proof of Theorem 2.0.6: discretization 192
2.4 Proof of Theorem 2.0.9: sublinear growth of cocycles 204
2.5 Derivation of Theorem 2.0.24 from Dorronsoro's Theorem 2.0.22 . . . 212
2.6 Proof of Dorronsoro's Theorem 2.0.22 for Carnot groups 213
2.6.1 L^1-optimality of the $A^d_{x,r}f$'s 215
2.6.2 Proof of the $\lesssim$ direction, $q = 1$ 217
2.6.3 Proof of the $\gtrsim$ direction 228
2.6.4 L^q β-numbers, $q > 1$ 240

Chapter 0

Introduction

A metric space $(\mathcal{M}, d_{\mathcal{M}})$ is said to embed bilipschitzly into a Banach space $(X, \|\cdot\|_X)$ if there exists a mapping $f : \mathcal{M} \to X$ and a constant $D \geq 1$ such that

$$d_{\mathcal{M}}(p,q) \leq \|f(p) - f(q)\|_X \leq D d_{\mathcal{M}}(p,q), \quad \forall p, q \in \mathcal{M}.$$

In this case, one says that $(\mathcal{M}, d_{\mathcal{M}})$ bilipschitzly embeds with distortion D into X; the infimal possible D is called the X-distortion of $\mathcal{M}$.

Given a metric space $(\mathcal{M}, d_{\mathcal{M}})$ and a Banach space $(X, \|\cdot\|_X)$, how can we detect whether $\mathcal{M}$ embeds bilipschitzly into X? If so, can we compute the X-distortion of $\mathcal{M}$? Of particular interest is the case when $X = \mathbb{R}^d$ for some $d \in \mathbb{N}$. That is, given a metric space, can we know whether it embeds into some Euclidean space $\mathbb{R}^d$? If so, can we compute the corresponding distortion, possibly as a function of d?

Although I do not have answers to the above questions in their full generality, I will provide some answers to them in the special case when the metric space $(\mathcal{M}, d_{\mathcal{M}})$ is either a finitely generated group of polynomial growth, or a simply connected nilpotent Lie group, or a ball thereof. These are important examples since they serve as cases that show the optimality of some embedding theorems, to be discussed below.

I will begin, in Section 0.1, by explaining the central problem of characterizing

bilipschitz embeddability into Euclidean space and the relevance of the Assouad embedding theorem to this problem. Then I will describe the results I have obtained in my dissertation in the following sections.

0.1 Characterization of bilipschitz embeddability into Euclidean spaces: the Lang–Plaut question and the Assouad embedding theorem

One of the long-standing open questions of metric geometry is the problem of intrinsically characterizing metric spaces that embed bilipschitzly into Euclidean space.

One immediate necessary condition is for the space $(\mathcal{M}, d_{\mathcal{M}})$ to be doubling, which means that there exists a constant $K \geq 2$ such that for any $x \in \mathcal{M}$ and $r > 0$ there exist $y_1, \cdots, y_K \in \mathcal{M}$ for which

$$B_r(x) \subset \bigcup_{i=1}^{K} B_{r/2}(y_i).$$

Here, $B_r(x) \coloneqq \{y \in \mathcal{M} : d_{\mathcal{M}}(x, y) \leq r\}$ denotes the ball of radius r centered at x; if one wants to stress the particular K, one says that $(\mathcal{M}, d_{\mathcal{M}})$ is K-doubling.

Another necessary condition is for the space $(\mathcal{M}, d_{\mathcal{M}})$ to embed into Hilbert space. This has been characterized by Linial, London, and Rabinovich [54]: the metric space $(\mathcal{M}, d_{\mathcal{M}})$ embeds $D(\geq 1)$-bilipschitzly into a Hilbert space if and only if for every $n \in \mathbb{N}$ and every positive semidefinite symmetric matrix $A = (a_{ij}) \in M_n(\mathbb{R})$ all of whose rows sum to zero, the following inequality holds:

$$\sum_{i,j=1}^{n} a_{ij} d_{\mathcal{M}}(x_i, x_j)^2 \leq \frac{D^2 - 1}{D^2 + 1} \sum_{i,j=1}^{n} |a_{ij}| d_{\mathcal{M}}(x_i, x_j)^2, \quad \forall x_1, \cdots, x_n \in \mathcal{M}.$$

Lang and Plaut [48] asked whether these two necessary conditions are sufficient.

This can be equivalently asked as follows.

Question 0.1.1 (Lang–Plaut question)**.** *Does every doubling subset of Hilbert space embed bilipschitzly into some $\mathbb{R}^d$?*

Historically, it was first asked whether the doubling condition alone is sufficient,[1] i.e., whether a doubling metric space embeds bilipschitzly into some $\mathbb{R}^d$. In an attempt to answer this in the affirmative, Assouad [3] gave the following characterization.

Theorem 0.1.2 (Assouad embedding theorem, qualitative version)**.** *Given a metric space $(\mathcal{M}, d_{\mathcal{M}})$, the following are equivalent.*

1. *The metric space $(\mathcal{M}, d_{\mathcal{M}})$ is doubling.*
2. *For some $0 < \alpha < 1$, the metric space $(\mathcal{M}, d_{\mathcal{M}}^{\alpha})$ embeds bilipschitzly into some $\mathbb{R}^d$.*
3. *For all $0 < \alpha < 1$, the metric space $(\mathcal{M}, d_{\mathcal{M}}^{\alpha})$ embeds bilipschitzly into some $\mathbb{R}^d$.*

Theorem 0.1.3 (Assouad embedding theorem, quantitative version)**.** *Given $K \geq 2$, there exist constants $C_1, C_2 > 0$ such that the following is true. Let $(\mathcal{M}, d_{\mathcal{M}})$ be a K-doubling metric space. For all $\frac{1}{2} < \alpha < 1$, the metric space $(\mathcal{M}, d_{\mathcal{M}}^{\alpha})$ embeds $\frac{C_1}{\sqrt{1-\alpha}}$-bilipschitzly into $\mathbb{R}^d$, where $d \leq \frac{C_2}{1-\alpha}$.*

In the above, $(\mathcal{M}, d_{\mathcal{M}}^{\alpha})$ denotes the metric space, called the *α-snowflake of* $(\mathcal{M}, d_{\mathcal{M}})$, whose set of points is $\mathcal{M}$ and whose distance between points p and q is defined as $d_{\mathcal{M}}(p, q)^{\alpha}$; one can check that this indeed defines a metric space.

The Assouad embedding theorem falls short of answering the bilipschitz embeddability problem into Euclidean space, because in the statement the exponent α is not allowed to be 1. In fact, Assouad's embedding theorem fails to hold for $\alpha = 1$, i.e., there exists a doubling metric space which fails to embed into Euclidean space;

[1] It is obvious that the embeddability into Hilbert space condition is not sufficient: take $\mathcal{M}$ to be ℓ^2.

the first such example to be found was the Heisenberg group $\mathbb{H}^{2k+1}$, $k \geq 1$, with its (left-invariant) Carnot-Carathéodory metric. This was proven by Semmes [73] using a differentiation argument of Pansu [69]. In fact, the distortion $\frac{1}{\sqrt{1-\alpha}}$ in the quantitative Assouad theorem is sharp up to constant factors, for the same example of the Heisenberg group [62]. Furthermore, the Heisenberg group fails to embed bilipschitzly into Hilbert space [11, 52]. This shows that the above second condition, namely the characterization of embeddability into Hilbert space, is necessary in the Lang–Plaut Question 0.1.1, even when given the doubling condition.

Naor and Neiman [62] suggested the following method of obtaining a negative answer to the Lang–Plaut Question 0.1.1. They argue it is enough to find a K-doubling metric space $(\mathcal{M}, d_{\mathcal{M}})$ with the following property:

1. One cannot take d to depend only on K in the Assouad embedding theorem, i.e., for any fixed integer d and constant $C_1 > 0$ there exists some $\frac{1}{2} < \alpha_0 < 1$ such that $(\mathcal{M}, d_{\mathcal{M}}^{\alpha_0})$ fails to embed $\frac{C_1}{\sqrt{1-\alpha_0}}$-bilipschitzly into $\mathbb{R}^d$.

2. Assouad's embedding theorem for $\mathcal{M}$ holds with doubling image, i.e., there exists some natural number K' and some constant C_1' such that for any $\frac{1}{2} < \alpha < 1$ there exists a $\frac{C_1'}{\sqrt{1-\alpha}}$-bilipschitz mapping $\Phi_\alpha : (\mathcal{M}, d_{\mathcal{M}}^{\alpha}) \to \ell_2$ such that $\Phi_\alpha(\mathcal{M})$ is K'-doubling.

Suppose such a space $\mathcal{M}$ existed and the Lang–Plaut Question 0.1.1 had a positive answer. It is not hard to see that a quantitative Lang–Plaut statement holds, i.e., any subset of Hilbert space of doubling constant $\leq K'$ bilipschitzly embeds, with bilipschitz constant bounded above by a function $D_{K'}$ of K', into some fixed $\mathbb{R}^d$ determined by K'. Thus, by point 2 above, for any $\frac{1}{2} < \alpha < 1$ there exists a $D_{K'}$-bilipschitz mapping $\Psi_\alpha : \Phi_\alpha(\mathcal{M}) \to \mathbb{R}^d$. The composition $\Psi_\alpha \circ \Phi_\alpha : (\mathcal{M}, d_{\mathcal{M}}^{\alpha}) \to \mathbb{R}^d$ is now a $\frac{C_1' D_{K'}}{\sqrt{1-\alpha}}$-bilipschitz mapping, for all $\frac{1}{2} < \alpha < 1$. This contradicts point 1 above.

Naor and Neiman [62] suspected the Heisenberg group to provide the counterex-

ample, as they knew it satisfied property 2 above. They asked if property 1 held as well. Surprisingly, it has been shown that property 1 fails for the Heisenberg group [77].

Theorem 0.1.4 ([77, Theorem 1.1])**.** *There exists a positive integer d and a constant $C > 0$ such that for all $\frac{1}{2} < \alpha < 1$ the metric space $(\mathbb{H}^3, d^{\alpha}_{\mathbb{H}^3})$ embeds $\frac{C}{\sqrt{1-\alpha}}$-bilipschitzly into $\mathbb{R}^d$.*

The guiding goal of this dissertation is to investigate this phenomenon, i.e., I will discuss the sharpness of the Assouad embedding theorem.

Question 0.1.5 (Sharp Assouad question)**.** *Given a $K(\geq 2)$-doubling metric space $(\mathcal{M}, d_{\mathcal{M}})$, do there exist a constant $C > 0$ and a natural number d such that the metric space $(\mathcal{M}, d^{\alpha}_{\mathcal{M}})$ embeds $\frac{C}{\sqrt{1-\alpha}}$-bilipschitzly into $\mathbb{R}^d$, for all $\frac{1}{2} < \alpha < 1$?*

In Chapter 1, I will provide a larger class of examples, namely the category of finitely generated groups of polynomial growth and simply connected nilpotent Lie groups, for which the sharp Assouad Question 0.1.5 has an affirmative answer. In Chapter 2 2, I will prove that the distortion $O\left(\frac{1}{\sqrt{1-\alpha}}\right)$ in the Assouad embedding Theorem 0.1.3 is sharp for not virtually abelian finitely generated groups of polynomial growth and nonabelian simply connected nilpotent Lie groups. The rest of this Chapter is devoted to a summary of Chapters 1 and 2.

0.2 Summary of Chapter 1: Sharp Assouad theorem for finitely generated groups of polynomial growth and simply connected nilpotent Lie groups

In Chapter 1, I establish a sharp Assouad theorem, i.e., an affirmative answer to Question 0.1.5 for finitely generated groups of polynomial growth and simply connected nilpotent Lie groups.

Theorem 0.2.1. *Let G be either a finitely generated group of polynomial growth with left-invariant word metric d_G, or a simply connected nilpotent Lie group with left-invariant sub-Finsler distance d_G. Then there exists a natural number D_G and a constant C_G such that for any $\frac{1}{2} < \alpha < 1$ the metric space (G, d_G^α) admits a bilipschitz embedding into $\mathbb{R}^{D_G}$ with distortion $\frac{C_G}{\sqrt{1-\alpha}}$.*

(A special case of this theorem for Carnot groups and their lattices appeared in [72, Theorem 1.1].)

This can be seen as a generalization of Theorem 0.1.4. The proof of Theorem 0.2.1 builds on the proof of Theorem 0.1.4, which consisted of a new variant of the Nash–Moser iteration scheme combined with a new extension theorem for orthonormal vector fields. Beyond the need to overcome several technical issues that arise in this more general setting, a key point where our proof departs from that of Theorem 0.1.4 is in the proof of the orthonormal vector field extension theorem, stated as follows.

Theorem 0.2.2 ([72, Theorem 4.2])**.** *Let $(\mathcal{M}, d_{\mathcal{M}})$ be a $K(\geq 2)$-doubling metric space, and let D and m be positive integers with $m \leq D - 224K^4 \log K$. If $v_1, \cdots, v_m : \mathcal{M} \to \mathbb{S}^{D-1}$ are 1-Lipschitz functions that form an orthonormal system at each point*

of $\mathcal{M}$, i.e.,

$$v_1(p), \cdots, v_m(p) \text{ form an orthonormal system for each } p \in \mathcal{M},$$

then there exists a $150K^5m(m+1)$-Lipschitz function $v_{m+1} : \mathcal{M} \to \mathbb{S}^{D-1}$ such that $v_1, \cdots, v_m$, and v_{m+1} form an orthonormal system at each point of $\mathcal{M}$.

In [77], Theorem 0.2.2 is established for the Heisenberg group using a quantitative homotopy argument on a CW structure on the Heisenberg group that is left-invariant under the action of a cocompact lattice. This proof does not carry on in the setting of nilpotent Lie groups, as a nilpotent Lie group may not have a cocompact lattice if its structure constants are always irrational. Instead, the proof of Theorem 0.2.2 uses two different ingredients: the Lovász local lemma and the concentration of measure phenomenon on the sphere. This approach has the advantage that the proof applies in the general setting of doubling metric spaces; we anticipate this to be useful in future attempts to answer Question 0.1.5 in the affirmative.

In particular, Theorem 0.2.1 tells us that one cannot find counterexamples to the Lang–Plaut Question 0.1.1 using the aforementioned program of [61] within the category of finitely generated groups of polynomial growth and simply connected nlipotent Lie groups.

0.3 Summary of Chapter 2: Vertical versus horizontal inequalities

In Chapter 2, I show that the distortion that appears in the Assouad embedding Theorem 0.1.3 is sharp for nonabelian simply connected nilpotent Lie groups and not virtually abelian finitely generated groups of polynomial growth. This is proven by establishing "vertical versus horizontal inequalities" on these groups. Previously,

these statements were known for the discrete and continuous Heisenberg groups.

Let G be a nonabelian simply connected nilpotent Lie group endowed with a left-invariant sub-Finsler distance. I will first state the inequality on G. Given left-invariant vector fields $X_1, \cdots, X_k$ that form a basis of the horizontal vector subspace V of the Lie algebra $\mathfrak{g}$ of G, along which the sub-Finsler distance is measured, let us write the horizontal gradient of a function $f : G \to X$ as

$$\nabla f \coloneqq (X_1 f, \cdots, X_k f),$$

if each $X_i f$ exists. As G is nilpotent, one may define the nilpotency step s of G, i.e., the largest integer s such that $\underbrace{[[\mathfrak{g}, \mathfrak{g}], \cdots, \mathfrak{g}]}_{s \text{ times}} \neq 0$; choose any $v \in \underbrace{[[\mathfrak{g}, \mathfrak{g}], \cdots, \mathfrak{g}]}_{s \text{ times}}$ of unit length. Let $\exp : \mathfrak{g} \to G$ be the Lie group exponential map, which is a diffeomorphism, and let μ denote the bi-invariant Haar measure of G, which is the push-forward of the Lebesgue measure on $\mathfrak{g}$ by the exponential map. One has $d_G(\exp(tv), e_G) \asymp_G t^{1/s}$ for all $t \geq 1$.[2]

Theorem 0.3.1 ([71, Theorem 5])**.** *Let G be a nonabelian simply connected nilpotent Lie group. Suppose that $p \in (1, \infty)$ and $q \in [2, \infty)$. Let $(H, \|\cdot\|_H)$ be a Hilbert space, and let $f : G \to H$ be smooth and compactly supported. Then*

$$\begin{aligned} &\left(\int_0^\infty \left(\int_G \left(\frac{\|f(h \exp(tv)) - f(h)\|_H}{t^{1/s}} \right)^p d\mu(h) \right)^{\max\{p,q\}/p} \frac{dt}{t} \right)^{1/\max\{p,q\}} \\ &\lesssim_G (p-1)^{1/q-1} \left(\int_G \|\nabla f(h)\|_{\ell_2^k(H)}^p \, d\mu(h) \right)^{1/p}. \end{aligned} \tag{1}$$

[2]For $P,\ Q > 0$, the notations $P \lesssim Q$, $Q \gtrsim P$, $P = O(Q)$, and $Q = \Omega(P)$ mean that $P \leq KQ$ for a universal constant $K \in (0, \infty)$, and $P \asymp Q$ means $(P \lesssim Q) \wedge (Q \lesssim P)$. Dependence on parameters is indicated by subscripts. For example, in the presence of auxiliary parameters ψ, ξ, the notations $P \lesssim_{\psi,\xi} Q$, $Q \gtrsim_{\psi,\xi} P$, $P = O_{\psi,\xi}(Q)$, $Q = \Omega_{\psi,\xi}(P)$ mean that $P \leq K(\psi, \xi)Q$ where $K(\psi, \xi) \in (0, \infty)$ may depend only on ψ and ξ, and $P \asymp_{\psi,\xi} Q$ means $(P \lesssim_{\psi,\xi} Q) \wedge (Q \lesssim_{\psi,\xi} P)$.

In particular, when $p = q$,

$$\left(\int_0^\infty \int_G \left(\frac{\|f(h\exp(tv)) - f(h)\|_H}{t^{1/s}}\right)^q d\mu(h)\frac{dt}{t}\right)^{1/q} \lesssim_G \left(\int_G \|\nabla f(h)\|_{\ell_2^k(H)}^q \, d\mu(h)\right)^{1/q}.$$

More generally, I can take the target space to be not only Hilbert space but also any uniformly convex space X. In this case, q will range over the values for which X has an equivalent norm which is q-uniformly convex.

I will next state the inequality on groups of polynomial growth. Let Γ be a not virtually abelian finitely generated group of polynomial growth, with left-invariant word metric induced by some finite generating set. By Gromov [30], Γ admits a subgroup Γ' of finite index that is nilpotent. Let T be the (finite) torsion subgroup of Γ' and consider the quotient subgroup $\Gamma'' = \Gamma'/T$, which is a torsion-free nilpotent group, of step $s \geq 2$. Take $v'' \in \underbrace{[[\Gamma'', \Gamma''], \cdots, \Gamma'']}_{s \text{ times}} \setminus \{e_{\Gamma''}\}$ and take v to be any representative of v''. Then $d_W(v^n, e_{\Gamma''}) \asymp_{\Gamma''} n^{1/s}$, $n \geq 2$. Under this setting, the following inequality holds.

Theorem 0.3.2 ([71, Theorem 6])**.** *Let Γ, $v \in \Gamma$, and $s \geq 2$ be as above and let $q \in [2, \infty)$. Suppose that $(H, \|\cdot\|_H)$ is a Hilbert space. Then there exists a natural number $c \geq 2$ such that for every $n \in \mathbb{N}$ and every $f : \Gamma \to H$ one has*

$$\begin{aligned}&\left(\sum_{k=1}^{n^s} \frac{1}{k^{1+\max\{p,q\}/s}} \left(\sum_{x \in B_n(e_\Gamma)} \|f(xv_\Gamma^k) - f(x)\|_H^p\right)^{\max\{p,q\}/p}\right)^{1/\max\{p,q\}} \\ &\lesssim_\Gamma (p-1)^{1/q-1} \left(\sum_{x \in B_{cn}(e_\Gamma)} \sum_{a \in S} \|f(xa) - f(x)\|_H^p\right)^{1/p}.\end{aligned} \tag{2}$$

In particular, when $p = q$,

$$\left(\sum_{k=1}^{n^s} \sum_{x \in B_n(e_\Gamma)} \frac{\|f(xv_\Gamma^k) - f(x)\|_H^q}{k^{1+q/s}}\right)^{1/q} \lesssim_\Gamma \left(\sum_{x \in B_{cn}(e_\Gamma)} \sum_{a \in S} \|f(xa) - f(x)\|_H^q\right)^{1/q}.$$

Again, more generally, one can take the target space to be any uniformly convex space.

These inequalities are proven using the vector-valued Littlewood–Paley–Stein theory approach of Lafforgue and Naor [47]. These inequalities are quantitative nonembeddability statements that show that any Lipschitz mapping from the aforementioned groups into a Hilbert space (or more generally any uniformly convex space) must quantitatively collapse along the center.

As a consequence, any embedding of the α-snowflake, $\frac{1}{2} < \alpha < 1$, of any nonabelian simply connected nilpotent Lie group G and of any not virtually abelian finitely generated group Γ of polynomial growth into Hilbert space must incur distortion at least a constant multiple of $\frac{1}{\sqrt{1-\alpha}}$. With the version of the inequality for uniformly convex spaces, we see that the distortion into a q-uniformly convex space is at least a constant multiple of $\frac{1}{(1-\alpha)^{1/q}}$.

Another consequence is that a ball of radius $n \geq 2$ in the aforementioned groups must incur bilipschitz distortion at least a constant multiple of $\sqrt{\log n}$ into a Hilbert space, or $(\log n)^{1/q}$ into a $q(\geq 2)$-uniformly convex Banach space. This bound is sharp for the L^p $(1 < p < \infty)$ spaces, i.e., the distortion is $\Omega_p(\sqrt{\log n})$ for L^p targets with $1 < p \leq 2$ and is $\Omega((\log n)^{1/p})$ for L^p targets with $p > 2$.

It is an open question what the distortion into L^1 spaces must be.

Question 0.3.3. *What is the smallest possible distortion incurred when embedding a ball of radius n in a nonabelian simply connected nilpotent Lie group or any not virtually abelian finitely generated group Γ of polynomial growth into an $L^1(\sigma)$ space?*

For the Heisenberg groups of dimension 5 or higher it was settled by Naor and Young [64] that the answer is a constant multiple of $\sqrt{\log n}$, and for the Heisenberg group of dimension 3 it was settled by Naor and Young [65] that the answer is a constant multiple of $\sqrt[4]{\log n}$.

A byproduct of the analysis is a version of the classical Dorronsoro theorem for

simply connected nilpotent Lie groups. Dorronsoro's theorem characterizes the L^p $(1 < p < \infty)$ norm of the fractional Laplacian of an element of $L^p(\mathbb{R}^d)$. I show that Dorronsoro's characterization carries on to the setting of simply connected nilpotent Lie groups.

Following Folland [25] and denoting by $\Delta = \sum_{i=1}^{k} X_i^2$ the sub-Laplacian and H_t the corresponding heat kernel, one defines the operator $(-\Delta_p)^\alpha$ for $1 < p < \infty$ and $\operatorname{Re}\alpha > 0$ by

$$(-\Delta_p)^\alpha f = \lim_{\varepsilon \to 0} \frac{1}{\Gamma(\lfloor \operatorname{Re}\alpha \rfloor + 1 - \alpha)} \int_\varepsilon^\infty t^{\lfloor \operatorname{Re}\alpha \rfloor - \alpha} (-\Delta)^{\lfloor \operatorname{Re}\alpha \rfloor + 1} H_t f \, dt$$

for all $f \in L^p(G)$ such that the limit exists in the L^p norm. Then, for $1 < p < \infty$ and $\alpha \geq 0$, the Sobolev space S_α^p is the Banach space $\operatorname{Dom}((-\Delta_p)^{\alpha/2})$ with norm

$$\|\cdot\|_{p,\alpha} \coloneqq \|\cdot\|_{L^p(G)} + \|(-\Delta_p)^{\alpha/2}(\cdot)\|_{L^p(G)}.$$

We now define local and global weightings of polynomials. Recall the horizontal layer $V \leq \mathfrak{g}$ of the sub-Finsler distance of G. Let $\mathcal{A}_d^{\mathrm{loc}}$ denote the class of polynomials of weighted degree at most d, with respect to the

For a locally integrable function $f : G \to \mathbb{R}$, $x \in G$ and $r > 0$, let $A_{x,r}^d f$ denote the unique element of $\mathcal{A}_d$, the class of weighted polynomials of degree at most d on G, such that

$$\int_{B_r(x)} (f(y) - A_{x,r}^d f(y)) A(y) dy = 0, \quad \forall A \in \mathcal{A}_d.$$

One measures how well $A_{x,r}^d f$ approximates f in the ball $B_r(x)$ by the following quantity:

$$\beta_{f,d,q}(B_r(x)) = \left(\fint_{B_r(x)} |f(y) - A_{x,r}^d f(y)|^q dy \right)^{1/q}, \quad 1 \leq q < \infty.$$

Then our version of Dorronsoro's theorem states the following.

Theorem 0.3.4 ([71, Theorem 22])**.** *Let G be a simply connected nilpotent Lie group of global Hausdorff dimension n_h. Let $1 < p < \infty$, $\alpha > 0$, and*

$$1 \leq q < \frac{\min\{p,2\} n_h}{n_h - \min\{p,2\}}.$$

Then for all $f \in L^p(G)$,

$$\left(\int_G \left(\int_0^\infty \left[\frac{1}{r^\alpha} \beta_{f,\lfloor \alpha \rfloor, q}(B_r(x)) \right]^2 \frac{dr}{r} \right)^{p/2} d\mu(x) \right)^{1/p} \asymp_{G,\alpha,p,q} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}.$$

in the sense that $f \in S^p_\alpha(G)$ if and only if the left-hand side of (2.8) *is finite, in which case the above relation holds.*

One consequence of Theorem 0.3.4 is that one obtains Theorem 0.3.1 with $q = 2$ for simply connected nilpotent Lie groups, under the weaker assumption that $v \in [\mathfrak{g}, \mathfrak{g}]$. Thus, when mapping nonabelian simply connected nipotent Lie groups into the L^p $(1 < p < \infty)$ spaces, the quantitative collapse occurs on the commutator subgroup; this is in line with the qualitative Pansu–Semmes nonembeddability argument.

Previously, in the setting of Heisenberg groups, Fässler and Orponen [22] established a one-sided Dorronsoro theorem with a restriction $0 < \alpha < 2$ on the range of exponents α of the Laplacian; this restriction does not appear in the commutative setting and is caused by their use of horizontal polynomials as approximants. The current work identifies the correct class of approximant polynomials, namely polynomials of local or global weighted degree $\leq d$, and proved the two-sided Dorronsoro theorem with the full range $0 < \alpha < \infty$ of exponents in the general setting of simply connected nilpotent Lie groups, thus strengthening and extending the work of Fässler and Orponen.

Chapter 1

Embedding finitely generated groups of polynomial growth and nilpotent Lie groups into Euclidean space

In this chapter, we will prove Theorem 0.2.1, thereby establishing an affirmative answer to the sharp Assouad question 0.1.5 for finitely generated groups of polynomial growth and simply connected nilpotent Lie groups.

Let Γ be a group with finite symmetric generating set S. We say that Γ is of *polynomial growth* if the cardinality of S^n, i.e., the set of group elements expressible as a word of length n, grows at most polynomially in n. Let d_W be the *left-invariant word metric* on Γ induced by S, i.e., for any group element $g \in \Gamma$, the distance $d_W(g, e_\Gamma)$ (where e_Γ is the identity element of Γ) is defined to be the length of the shortest word expressing g using only alphabets from S. For any $g_1, g_2 \in \Gamma$, we define $d_W(g_1, g_2) \coloneqq d_W(g_2^{-1} g_1, e_\Gamma)$. It is clear that this defines a distance on Γ.

One of the main results of this chapter is the following theorem.[1]

Theorem 1.0.1. *Let Γ be a finitely generated group of polynomial growth with left-invariant word metric d_W. Then there exists a natural number D_Γ such that for every $0 < \varepsilon < 1/2$ there exists a bilipschitz embedding of $(\Gamma, d_W^{1-\varepsilon})$ into $\mathbb{R}^{D_\Gamma}$ with distortion $O_\Gamma(\varepsilon^{-1/2})$.*

This theorem is immediate when Γ has a finite index subgroup isomorphic to $\mathbb{Z}^n$; in fact, in this case, there exists a bilipschitz embedding of the original metric space (Γ, d_Γ) into some $\mathbb{R}^{D_\Gamma}$. Thus, the hard part of Theorem 1.0.1 is when Γ does not admit finite index subgroups isomorphic to $\mathbb{Z}^n$, i.e., when Γ is not virtually abelian.

The bilipschitz geometry of finitely generated groups of polynomial growth is intimately related to that of nilpotent Lie groups. In fact, Theorem 1.0.1, which deals with finitely generated groups of polynomial growth, will be a consequence of its counterpart Theorem 1.0.2, which deals with nilpotent Lie groups.

A connected real Lie group G, with real Lie algebra $(\mathfrak{g}, [\cdot,\cdot])$ is said to be *nilpotent* if there exists a positive integer s such that $\underbrace{[\mathfrak{g}, [\mathfrak{g}, [\cdots, \mathfrak{g}]]]}_{s+1 \text{ times}} = 0$. A vector subspace $V \leq \mathfrak{g}$ is said to *generate* $\mathfrak{g}$ *by brackets* if

$$V + [V, V] + \cdots + \underbrace{[V, [V, [\cdots, V]]]}_{s \text{ times}} = \mathfrak{g}.$$

We may define a bundle Δ over G as the extension of V to a left-invariant subbundle:

$$\Delta_p \coloneqq (dL_p)_e V, \quad p \in G.$$

[1]Following widespread convention, $A \lesssim B$, $A = O(B)$, and $B = \Omega(A)$ mean $A \leq CB$ for a universal constant C, and $A \asymp B$ means $(A \lesssim B) \wedge (B \lesssim A)$. If the constant C depends on other parameters, this is denoted using subscripts, e.g., $A \lesssim_{C_0,N_0} B$, $A = O_{C_0,N_0}(B)$, and $B = \Omega_{C_0,N_0}(A)$ mean $A \leq C(C_0, N_0)B$ where $C(C_0, N_0)$ depends only on C_0 and N_0, and $A \asymp_{C_0} B$ means $A \lesssim_{C_0} B$ and $B \lesssim_{C_0} A$.

Let $\|\cdot\|$ be a norm defined on V. We then extend it to Δ as a left-invariant norm:

$$\|(dL_p)_e(v)\| \coloneqq \|v\|, \quad p \in G,\ v \in V.$$

The metric d_G on G is the *left-invariant sub-Finsler Carnot–Carathéodory distance* associated to Δ and $\|\cdot\|$, i.e., for $p, q \in G$,

$$d_G(p,q) \coloneqq \inf\left\{\int_0^1 \|\dot\gamma(t)\|dt : \gamma \in C^\infty_{\mathrm{pw}}([0,1];G), \gamma(0)=p, \gamma(1)=q, \dot\gamma \in \Delta\right\},$$

where $C^\infty_{\mathrm{pw}}([0,1];G)$ consists of the piecewise smooth functions from $[0,1]$ to G.

The other main result of this chapter is the following theorem.

Theorem 1.0.2. *Let G be a connected real nilpotent Lie group with left-invariant sub-Finsler Carnot-Carathéodory distance d_G associated to a bracket generating subspace $V \leq \mathfrak{g}$ and a norm on it. Then there exists a natural number D_G such that for every $0 < \varepsilon < 1/2$ there exists a bilipschitz embedding of $(G, d_G^{1-\varepsilon})$ into $\mathbb{R}^{D_G}$ with distortion $O_G(\varepsilon^{-1/2})$.*

Previously, Theorem 1.0.2 was proven in[2] [77] for the Heisenberg group $\mathbb{H}^3$, and more generally in [72] for *Carnot groups*, which are groups G defined as in Theorem 1.0.2 but with the additional property that the Lie algebra $\mathfrak{g}$ is graded, i.e., it admits the direct sum decomposition

$$\mathfrak{g} = V_1 \oplus V_2 \oplus \cdots \oplus V_s,$$

where $V_{s+1} = 0$ and $V_{r+1} = [V_1, V_r]$ for $r = 1, \cdots, s$, and the distance is induced by the choice of $V = V_1$. These groups are easier to analyze because their group law is cleaner and because they admit dilations: for each $\lambda \in \mathbb{R}^+$, the linear map $\delta_\lambda : \mathfrak{g} \to \mathfrak{g}$

[2]When comparing [77] and this paper, one should keep in mind that the vectorfields and metrics of [77] are right-invariant, whereas this paper uses left-invariant vectorfields and metrics.

defined by

$$\delta_\lambda|_{V_r} = \lambda^i \mathrm{id}_{V_r}, \quad r = 1, \cdots, s$$

is a Lie algebra isomorphism.

We have not attempted to optimize D_Γ or D_G, but from the analysis in this paper it will be clear that D_Γ and D_G are at least polynomial in the doubling constant, to be defined below, of Γ and G.

We will not work in the large epsilon regime $\frac{1}{2} \leq \varepsilon < 1$ because Theorems 1.0.1 and 1.0.2 are simply not true in this case; this is due to obstructions caused by the Assouad dimension, see [3]. Furthermore, we will assume $0 < \varepsilon < \frac{1}{A}$, where A is a very large number. For $\frac{1}{A} \leq \varepsilon < \frac{1}{2}$, a construction of [3] gives such an embedding; we are thus only interested in the small epsilon regime.

Some history behind Theorems 1.0.1 and 1.0.2: the groups under consideration are special cases of doubling metric spaces, where a metric space (X, d_X) is said to be K-doubling for some natural number K if for any $x \in X$ and $R > 0$ there exist points $y_1, \cdots, y_K \in X$ such that

$$B_R(x) \subset \bigcup_{j=1}^{K} B_{R/2}(y_j),$$

where $B_R(x) \coloneqq \{z \in X : d_X(x, z) \leq R\}$ is the ball of radius R centered at x; we say that (X, d_X) is doubling if (X, d_X) is K-doubling for some natural number K. Assouad [3] showed that for $0 < \varepsilon < \frac{1}{2}$, the $(1-\varepsilon)$-snowflake of a K-doubling metric space admits an embedding into $\ell_2^{O_K(\varepsilon^{-O(1)})}$ with distortion $O_K(\varepsilon^{-1/2})$. Here, the distortion $O(\varepsilon^{-1/2})$ is sharp for the Heisenberg group $\mathbb{H}^3$; see [61, Section 4] for a proof of this fact. There is a lot of relevant literature on how to sharpen Assouad's theorem in different ways, with or without various assumptions; see, for example, Abraham, Bartal, and Neiman [1], Bartal, Recht, and Schulman [7], Gottlieb and Krauthgamer [28], and Gupta, Krauthgamer, and Lee [31], but we do not mean this

list to be extensive. The direction that is quite relevant to this paper is that of [61], which showed that one can construct snowflake embeddings of any doubling metric space into bounded dimensional Euclidean spaces at the cost of slightly worsening the distortion: more precisely, one can embed the $(1-\varepsilon)$-snowflake of a K-doubling space into $\ell_2^{O(\log K)}$ with distortion $O(\frac{\log^2 K}{\varepsilon^2})$, or more generally, for any $0 < \delta \leq 1$, into $\ell_2^{O(\frac{\log K}{\delta})}$ with distortion $O((\frac{\log K}{\varepsilon})^{1+\delta})$. One may then further inquire whether we can take the target dimension to be uniformly bounded in $0 < \varepsilon < \frac{1}{2}$ while simultaneously attaining the best possible distortion $O(\varepsilon^{-1/2})$. We state this separately as a question.

Question 1.0.3 (Assouad's theorem with optimal distortion and bounded dimension)**.** *For any natural number $K \geq 2$, does there exist a natural number $D(K)$ such that if (X, d_X) is a K-doubling metric space and $0 < \varepsilon < \frac{1}{2}$, then there exists an embedding of the snowflake $(X, d_X^{1-\varepsilon})$ into $\mathbb{R}^{D(K)}$ with distortion $O_K(\varepsilon^{-1/2})$?*

For completeness, we state a stronger version of the above question, which is motivated by the fact that we must have the lower bound $D(K) \gtrsim \log K$.

Question 1.0.4 (Assouad's theorem with optimal distortion and optimal dimension)**.** *For any natural number $K \geq 2$, does there exist a natural number $D(K) = O(\log K)$ such that if (X, d_X) is a K-doubling metric space and $0 < \varepsilon < \frac{1}{2}$, then there exists an embedding of the snowflake $(X, d_X^{1-\varepsilon})$ into $\mathbb{R}^{D(K)}$ with distortion $O_K(\varepsilon^{-1/2})$?*

These questions are relevant to finding counterexamples to a question raised in [48], which, in our notation, can be stated as follows.

Question 1.0.5 (Lang–Plaut problem [48])**.** *For any natural number $K \geq 2$, does there exist a natural number $D(K)$ such that if X is a subspace of ℓ_2 that is K-doubling under the metric it inherits from ℓ_2, then there exists an embedding of $(X, \|\cdot\|_2)$ into $\mathbb{R}^{D(K)}$ with $O_K(1)$ distortion?*

As observed in [61], if an embedding as in Question 1.0.3 fails to exist for the Heisenberg group $\mathbb{H}^3$, then the Lang–Plaut problem would be answered in the neg-

ative, since it is known that $\mathbb{H}^3$ admits $(1-\varepsilon)$-snowflake embeddings into ℓ_2 with distortion $O(\varepsilon^{-1/2})$ with the additional property that the doubling constant of the image is uniformly bounded. Nevertheless, it was shown in [77] that an embedding as in Question 1.0.3 for $\mathbb{H}^3$ exists, i.e., Theorem 1.0.2 for $G = \mathbb{H}^3$ is true. Thus, the Heisenberg group (or more precisely, the doubling ℓ_2 images of the snowflakes thereof) fails to serve as a counterexample to the Lang–Plaut problem.

The following question was posed in [61] to highlight this connection.

Question 1.0.6. *For any natural number $K \geq 2$, does there exist a natural number $K'(K)$ such that if (X, d_X) is a K-doubling metric space and $0 < \varepsilon < \frac{1}{2}$, then there exists an embedding of the snowflake $(X, d_X^{1-\varepsilon})$ into ℓ_2 with distortion $O_K(\varepsilon^{-1/2})$ with the additional property that image of X is $K'(K)$-doubling?*

Of course, the connection is that if Question 1.0.6 has a positive answer while Question 1.0.3 has a negative answer, then the Lang–Plaut problem (Question 1.0.5) must have a negative answer.

The main purpose of this paper is to expand upon our partial knowledge of Question 1.0.3, by answering it in the affirmative in the setting of finitely generated groups of polynomial growth and nilpotent Lie groups. We first reduce proving Theorem 1.0.1 to proving Theorem 1.0.2, and then establish that it is enough to consider *simply connected* nilpotent Lie groups. We carry the methods of [77] into the setting of simply connected nilpotent Lie groups, and whenever a tool of that paper becomes inadequate in the setting of simply connected nilpotent Lie groups, we replace it with a tool more suitable in the general language of doubling metric spaces. More specifically, the key tools of [77] are a new variant of the Nash–Moser iteration scheme and a certain extension theorem for orthonormal vector fields using a quantitative homotopy argument. When generalizing to the case of nilpotent Lie groups, one potential source of trouble is that an arbitrary nilpotent Lie group might have an arbitrarily large step size s, which could complicate its geometric properties and make the tools of [77]

fail. It will be shown in Section 1.2 that the Nash–Moser iteration scheme, being a first and second order statement, directly generalizes to the setting of nilpotent Lie groups, and we prove an orthogonality statement slightly stronger than that of [77]. However, in Section 1.3, it will be clear that there are some obstructions to the quantitative homotopy argument. Nevertheless, we will prove the orthonormal vector field extension theorem even for general doubling metric spaces, using the concentration of measure phenomenon for the sphere and the Lovász Local Lemma.

Ultimately, we would like to answer Question 1.0.3 for general doubling metric spaces; this paper is an intermediate step in such an investigation, showing that it is at least true for finitely generated groups of polynomial growth and nilpotent Lie groups. We plan to address the case of general doubling metric spaces in future work, by possibly adapting some of the proof methods of this paper. In particular, the orthonormal vector field extension theorem for doubling metric spaces (Theorem 1.3.3) seems to be a good starting point for future work, and we would have to find either a metric analog or a replacement for the Nash–Moser iteration scheme.

We now briefly overview some of the results and strategies of [3, 61, 77] for constructing snowflake embeddings, and describe how these ideas connect to the proof strategy of this paper.

The starting point of constructing snowflake embeddings is the classical fact that if X is a metric space, $A > 1$, $0 < \varepsilon < \frac{1}{2}$, and $\{\phi_m : X \to \ell_2\}_{m\in\mathbb{Z}}$ is a collection of maps such that $|\phi_m| \leq A^m$ and ϕ_m is 1-Lipschitz, then the Weierstrass sum

$$\Phi = \sum_{m\in\mathbb{Z}} A^{-m\varepsilon}\phi_m \tag{1.1}$$

(which is easily seen to be absolutely convergent, say by requiring $\phi_m(x_0) = 0$ for all $m \in \mathbb{Z}$ for some fixed $x_0 \in X$) is $(1-\varepsilon)$-Hölder, with the Hölder norm bounded by $O(\varepsilon^{-1})$. If one assumes in addition that the maps ϕ_m take values in mutually

orthogonal subspaces of ℓ_2, then we can bound the $(1-\varepsilon)$-Hölder norm by $O(\varepsilon^{-1/2})$. In the case where X is a doubling metric space, Assouad [3] constructed functions ϕ_m with the above properties and with the additional property that if $d(p,q) \asymp A^m$ then $|\phi_m(p) - \phi_m(q)| \gtrsim_K A^m$. Then Φ satisfies the Hölder lower bound $|\Phi(p) - \Phi(q)| \gtrsim_K d(p,q)^{1-\varepsilon}$ for any $p, q \in X$, and thus is an embedding of the $(1-\varepsilon)$-snowflake of X into ℓ_2 with distortion $O_K(\varepsilon^{-1/2})$ (we do not stress dependence on A for now, as Assouad chooses A to be a constant).

When improving upon Assouad's theorem to reduce the target dimension (especially when trying to keep the target dimension independent of ε), one usually keeps the idea of using the Weierstrass sum (1.1) to guarantee the upper bound but needs to be more clever to enforce the lower bound. Note that for $A^{m-1} \le d(p,q) \le A^m$, the sum $\sum_{n<m} A^{-n\varepsilon}(\phi_n(p) - \phi_n(q))$ is negligible compared to $d(p,q)^{1-\varepsilon}$,[3] so in order to enforce the lower bound it is enough to have $|\sum_{n\ge m} A^{-n\varepsilon}(\phi_n(p) - \phi_n(q))| \gtrsim d(p,q)^{1-\varepsilon}$. Thus, it is natural to devise an iterative construction: having constructed maps $\{\phi_n : X \to \mathbb{R}^D\}_{n>m}$ such that the partial sum $\sum_{n>m} A^{-n\varepsilon}\phi_n$ is a map that is able to distinguish points which are at least distance A^{m+1} apart, by being a map that "oscillates" at scale A^{m+1} and above, we need to devise a component $A^{-m\varepsilon}\phi_m : X \to \mathbb{R}^D$ which "oscillates" at scale A^m and which, when added to $\psi := \sum_{n>m} A^{-n\varepsilon}\phi_n$, makes the sum $A^{-m\varepsilon}\phi_m + \psi$ able to distinguish points which are at least distance A^m apart. Thus, the challenge is that given the map $\psi : X \to \mathbb{R}^D$, we need to make use of a limited amount of coordinates in constructing a map ϕ_m, so that although ψ and ϕ_m "share coordinates", the sum $A^{-m\varepsilon}\phi_m + \psi$ satisfies the required Hölder lower bound.

In [61], the maps ϕ_m are defined via a probabilistic construction: after they construct random partitions arising from nets (defined in Section 1.1.4) with good proba-

[3] More precisely, we would have to take $A^{m-1+\delta} \le d(p,q) \le A^{m+\delta}$ for a fixed small $\delta > 0$, while taking the length scale A sufficiently large, but for clarity of the introduction we will not address this issue for now. See the proof of Proposition 1.5.5 for a precise argument. We also remark that the length scale A will be chosen last in our "hierarchy of constants" (see subsection 1.1.2), so A will dominate every other parameter used in this chapter.

bilistic padding properties, they define the maps ϕ_m using the distance to the boundaries of these partitions, and by a nested use of the Lovász Local Lemma conditioned on the partial sum $\sum_{n>m} A^{-n\varepsilon}\phi_n$, they show that their maps ϕ_m have the desired property. This construction gives distortion $O(\varepsilon^{-1-\delta})$, where the ε^{-1} factor comes from the Hölder constant of (1.1), and the additional factor $\varepsilon^{-\delta}$ comes from some technicalities of the construction, namely that one needs to introduce a slightly finer length scale than A when constructing the nets to ensure that the Lovász Local Lemma is applicable and thus the Hölder lower bound is achieved.

The result of [61] was surprising at the time since it was the first to achieve a target dimension that is uniformly bounded in the amount of snowflaking. It is also surprising that they have achieved the theoretically best possible target dimension $\log K$, up to constant factors. However, [61] fell short of giving a final answer to the problem of snowflake embeddings, since the distortion they have achieved was $O(\varepsilon^{-1-\delta})$, not $O(\varepsilon^{-1/2})$. They have asked whether it is even possible, citing that if Question 1.0.3 is answered in the negative for $\mathbb{H}^3$, then Question 1.0.5 would be answered in the negative. Surprisingly, Question 1.0.3 was subsequently answered in the affirmative for $\mathbb{H}^3$ in [77], and thus the above method of answering Question 1.0.5 was overturned.

In order to see how one can achieve the optimal distortion $O(\varepsilon^{-1/2})$, let us revisit (1.1). When working with a bounded number of dimensions, one needs to work with a notion of orthogonality weaker than requiring the ϕ_m to have pairwise orthogonal ranges. In fact, one can see that we do not need to require every pair of the ϕ_m to take values in mutually orthogonal spaces; instead, we need only that each ϕ_m be orthogonal in some sense to the partial sum $\sum_{n>m} A^{-n\varepsilon}\phi_n$ (recall that we don't need to consider the partial sum $\sum_{n<m} A^{-n\varepsilon}\phi_n$ when considering points $p, q \in X$ with $d(p,q) \asymp A^m$, as the sum $\sum_{n<m} A^{-n\varepsilon}(\phi_n(p) - \phi_n(q))$ is negligible compared to $A^{m(1-\varepsilon)} \asymp d(p,q)^{1-\varepsilon}$ in this case). Following [77], we will choose the notion of

orthogonality to be that the horizontal derivatives are perpendicular: if ∇ denotes the horizontal gradient in the Carnot group G (see subsection 1.1.6 for the definition of horizontal gradient), then we require the orthogonality $\nabla\phi_m \cdot \nabla \sum_{n>m} A^{-n\varepsilon}\phi_n = 0$ (actually we will derive a slightly stronger orthogonality condition; see Section 1.2). Overall, the ϕ_m are constructed in an iterative manner, and this orthogonality is added to one of the conditions that such an iterative construction must achieve.

Note that this orthogonality easily guarantees the Hölder upper bound of Φ, because an iterated use of the Pythagorean theorem gives

$$\left|\nabla \sum_{n\geq m} A^{-n\varepsilon}\phi_n\right|^2 = \sum_{n\geq m} A^{-2n\varepsilon}|\nabla\phi_n|^2 \lesssim_A \varepsilon^{-1}A^{-2n\varepsilon},$$

(we will have $|\nabla\phi_n| = O(1)$); to get the Hölder lower bound, one needs to guarantee that the partial sum $\sum_{n\geq m} A^{-n\varepsilon}\phi_n$ "oscillates" at scale A^m. In general, proving the Hölder lower bound is more complicated than proving the Hölder upper bound, as the geometry of the particular space X under consideration plays a more significant role.

For simplicity of explanation, assume G is a Carnot group. For such groups, we can use Taylor expansions to exploit the local geometry of G in the following way. Fixing a basis $X_{r,1}, \cdots, X_{r,k_r}$ of each stratum V_r, each point $p \in G$ can be expressed in the coordinates $p = \exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} x_{r,i}X_{r,i}\right)$, while its distance to the identity is roughly $d_G(p,0) \asymp_G \sum_{r=1}^{s}\sum_{j=1}^{k_r}|x_{r,j}|^{1/r}$ (the notation will be explained in full detail in subsection 1.1.6). Under this coordinate system, any $C^{s(s+1)}$-function $\Psi : G \to \mathbb{R}^D$ can be approximated by its Taylor expansion as

$$\Psi(p) = \Psi(0) + \sum_{m=1}^{s}\frac{1}{m!}\left(\sum_{r=1}^{s}\sum_{j=1}^{k_r} x_{r,j}X_{r,j}\right)^m \Psi(0) + (\text{Taylor approx. error})$$

(the smallness of the Taylor approximation error can come from a bound on the

$C^{s(s+1)}$-norm of Ψ and on a smallness assumption on $d_G(p,0)$). It will be seen in the proof of Proposition 1.5.5 that we can rearrange the above into the nicer form

$$\begin{aligned}\Psi(p) =&\Psi(0) + \sum_{r=1}^{s}\sum_{j=1}^{k_r}(x_{r,j} + \text{coordinate errors})X_{r,j}\Psi(0)\\ &+ (\text{cross derivative terms}) + (\text{Taylor approx. error}),\end{aligned}$$

where the "coordinate errors" and the "cross derivative terms" can be made small by a smallness assumption on $d_G(p,0)$ and on control of the second and higher order derivatives of Ψ. Thus, if we also have quantitative linear independence of the derivatives $X_{r,j}\Psi(0)$ (see subsection 1.1.3 for the notion of quantitative linear independence; we will later refer to this as a "freeness property"), then we can deduce

$$\begin{aligned}|\Psi(p) - \Psi(0)| \gtrsim& \max_{1\le r\le s, 1\le j\le k_r} |(x_{r,j} + \text{coordinate errors})X_{r,j}\Psi(0)| - (\text{errors})\\ \gtrsim& d_G(p,0)^s \min_{1\le r\le s, 1\le j\le k_r} |X_{r,j}\Psi(0)| - (\text{errors}).\end{aligned}$$

We remark that this is one point where our proof becomes more complicated due to the more general nilpotent Lie group structure, compared to the setting of $G = \mathbb{H}^3$ in [77]. But this, so far, does not necessitate a fundamental change of techniques.

To deduce the Hölder lower bound for our embedding Φ at scale $A^0 = 1$, we will apply the above approximation to $\Psi = \sum_{n\ge 0} A^{-n\varepsilon}\phi_n$, and consider points at a slightly smaller scale, more precisely $d_G(p,0) \asymp A^{-1/(s+1)}$. (This uses the additional fact that we effectively only need a Hölder lower bound for part of the scales, as long as we can cover the entire set of scales by a finite set of rescalings. See Proposition 1.5.5.) Thus, to enforce the Hölder lower bound, we will have to justify the above approximation and give quantitative bounds on the various errors, namely we will need estimates on the $C^{s(s+1)}$-norm of Ψ and quantitative linear independence of the derivatives $X_{r,j}\Psi(0)$.

Thus, all in all, the iterative step can be stated as follows: upon rescaling so that $m = 0$, and defining $\psi = \sum_{n>0} A^{-n\varepsilon}\phi_n$, the main problem is, given a function $\psi : G \to \mathbb{R}^D$, which "oscillates" at scale A (this will be measured by the rescaled C^m norms which will be introduced in subsection 1.1.5), which is a $C^{s(s+1)}$-function, and whose derivatives are quantitatively linearly independent, to construct a function $\phi : G \to \mathbb{R}^D$, which "oscillates" at scale 1, which is a $C^{s(s+1)}$-function, and with the orthogonality property $\nabla\phi \cdot \nabla\psi = 0$ along with a quantitative linear independence for the derivatives of $\psi + \phi$. The precise statement and proof of the iterative step will be presented in Section 1.4.

Once we have the iterative step as above, the rest of the inductive construction is fairly easy and is given in Section 1.5. This inductive construction mostly follows the line of [77], but we have spelled out the details for completeness. In order to begin the iteration, we need a single function ϕ_m to begin with, which oscillates at a fixed scale A^m, is of class $C^{s(s+1)}$, and whose derivatives are quantitatively linearly independent (we will call this a "freeness" property). This is done in Proposition 1.4.1. It is easy to construct a smooth function with the first two properties, and then we can guarantee the latter freeness property by a Veronese-type embedding often employed in the Nash embedding literature (and also used in [77]). Once we have this single function, we employ the above iterative step a finite number of times to obtain a finite family of mappings $\{\phi_m : G \to \mathbb{R}^D\}_{M_1 \leq m \leq M_2}$ (Proposition 1.5.1), and then use the Arzelà-Ascoli theorem to pass to a full family of mappings $\{\phi_m : G \to \mathbb{R}^D\}_{m\in\mathbb{Z}}$ (Theorem 1.5.3). There is a small issue of losing one degree of regularity when using the Arzelà-Ascoli theorem, but this can easily be fixed by requiring C^{s^2+s+1}-regularity in the first place.

We devote the rest of the introduction to explaining the proof ideas behind the above iterative step.

A naïve approach to solving the equation[4] would be to use the Leibniz rule directly: denoting by $X_1, \cdots, X_k$ a left-invariant vectorfield basis for V_1 so that $\nabla = (X_1, \cdots, X_k)$, the Leibniz rule tells us that $X_i\phi \cdot X_i\psi = X_i(\phi \cdot X_i\psi) - \phi \cdot X_i X_i \psi$, so in order to solve

$$X_i\phi \cdot X_i\psi = 0, \quad i = 1, \cdots, k,$$

we could simply solve

$$\phi \cdot X_i\psi = \phi \cdot X_i X_i \psi = 0, \quad i = 1, \cdots, k$$

(Note that the trivial solution ϕ=0 is not acceptable, as we need a freeness property of $\psi + \phi$ at scale 1.) This latter equation is at least algebraically well-posed in the sense that the vectors $X_i\psi$ and $X_i X_i \psi$, $i = 1, \cdots, k$, are (quantitatively) linearly independent; however it seems inherent that ϕ must be in the same regularity class as $X_i\psi$ and $X_i X_i \psi$; i.e., ϕ must have two fewer degrees of regularity compared to ψ. This "loss of derivatives problem" would make the iteration fail.

The solution proposed by [77] to get around this loss of derivatives problem was to introduce Littlewood–Paley projections $P_{(\leq N_0)}$ on $\mathbb{H}^3$ (see subsection 1.1.7 for Littlewood–Paley projections), and first solve the approximate equation

$$X_i\tilde{\phi} \cdot X_i P_{(\leq N_0)}\psi = 0, \quad i = 1, \cdots, k,$$

by solving

$$\tilde{\phi} \cdot X_i P_{(\leq N_0)}\psi = \tilde{\phi} \cdot X_i X_i P_{(\leq N_0)}\psi = 0, \quad i = 1, \cdots, k.$$

Because lower order derivatives of ψ can control higher order derivatives of $P_{(\leq N_0)}\psi$ (up to losses growing on the order of the Littlewood–Paley frequency N_0; see Theorem 1.1.2 for the precise quantitative statement), and because we are looking at a smaller

[4] In Section 1.2 instead of $\nabla\phi \cdot \nabla\psi = 0$ we will solve the stronger equation $X_i\phi \cdot X_j\psi + X_j\phi \cdot X_i\psi = 0$ for $i, j = 1, \cdots, k$. We are stating the simpler version to keep the Introduction simple.

scale 1 for ϕ compared to the larger scale A for ψ, it is plausible that one can achieve as much control on $\tilde{\phi}$ as ψ, as long as we take A very large compared to N_0. More precisely, because ψ oscillates at scale A, its homogeneous $\dot{C}^m$-norm will behave like A^{-m}, so $\nabla^2 P_{(\leq N_0)}\psi$ will have $\dot{C}^m$-norm roughly A^{-m-2} for $m \leq s^2+s-2$ and $A^{-s^2-s}N_0^{m-s^2-s+2}$ for $m > s^2+s-2$, which are all $O(1)$ if A is chosen sufficiently larger than N_0. This will ensure that any reasonable construction based on $X_i P_{(\leq N_0)}\psi$ and $X_i X_i P_{(\leq N_0)}\psi$ will produce a function that "oscillates" at scale 1 and is of class C^{s^2+s+1}, and thus has the same level of regularity as the function $\tilde{\phi}$ that we want to produce.

We will show in subsection 1.1.7 that simply connected nilpotent Lie groups admit Littlewood–Paley projections that share the same properties as those used in [77].

Thus, the main idea of [77] to solve the equation $\nabla\phi \cdot \nabla\psi = 0$ is to decompose it into two steps. In the first step, we need to solve the low-frequency version of the equation $\nabla\tilde{\phi} \cdot \nabla P_{(\leq N_0)}\psi = 0$ using the Leibniz rule, i.e., we solve the equation

$$\tilde{\phi} \cdot X_i P_{(\leq N_0)}\psi = 0,\ \tilde{\phi} \cdot X_i X_i P_{(\leq N_0)}\psi = 0, \quad i = 1, \cdots, k,$$

while guaranteeing that $\tilde{\phi}$ has the same regularity as $XP_{(\leq N_0)}\psi$ and $XXP_{(\leq N_0)}\psi$. In the second step, we would assume the low-frequency solution $\tilde{\phi}$ as given, and then we would need to correct $\tilde{\phi}$ into a "true" solution ϕ while guaranteeing that ϕ has the same regularity as $\tilde{\phi}$.

The first step, namely solving the approximate equation $\tilde{\phi} \cdot X_i P_{(\leq N_0)}\psi = \tilde{\phi} \cdot X_i X_i P_{(\leq N_0)}\psi = 0$, with $\tilde{\phi}$ having the same regularity as $X_i P_{(\leq N_0)}\psi$ and $X_i X_i P_{(\leq N_0)}\psi$, is the part where our proof in the setting of general simply connected nilpotent Lie groups departs the most from the setting of $\mathbb{H}^3$ in [77], and is the part where the main novelty of this chapter lies in. For $\mathbb{H}^3$, [77] resolves this issue by first observing that $\mathbb{H}^3$ has a cocompact lattice and a CW structure that is compatible with it, and then

provides the extension using quantitative null homotopy on each of the cells of that CW structure (note that, although the null homotopy is used on an infinite number of cells, the final construction is globally controlled, because up to left-translation there is only a finite number of distinct cells). This argument may not work for a general simply connected nilpotent Lie group G, as G might not admit a cocompact lattice (say if the structure constants for every basis of G were irrational). Instead, we give a proof, using the concentration of measure phenomenon on the sphere and the Lovász Local Lemma, that is independent of the topology of the space under consideration, by using only the fact that G is a doubling metric space (see Section 1.3 for details). We essentially do not need the differential structure of G because the uniform continuity of $\tilde{\phi}$ is the main hurdle here; we can automatically gain higher regularity by convolving with a mollifier and using the Gram-Schmidt process.

To put our solution to the approximate equation into context, we will pose, in Section 1.3, a more general question (Question 1.3.1), which asks whether one can extend a given orthonormal system of vectorfields within the same regularity class. It will become clear that our solution provides a partial positive answer to Question 1.3.1, at least when the base space is a doubling metric space, the regularity class is contained in the Lipschitz class, the regularity class is closed under simple algebraic operations, and if smoothening a Lipschitz vectorfield by convolving it with some scalar mollifier provides the resulting vectorfield with the desired regularity (see Theorem 1.3.6 for details).

For the second step, once we have the approximate solution $\tilde{\phi}$, we need to correct it into a true solution ϕ while preserving the regularity. Tao [77] solved this by developing a perturbative theory for the bilinear form $\nabla\phi \cdot \nabla\psi$. More precisely, [77] develops a version of the Nash–Moser iteration scheme to show how one can correct $\tilde{\phi}$ by small amounts into a true solution ϕ to $\nabla\phi \cdot \nabla\psi = 0$, without losing any regularity (technically, the Nash–Moser iteration scheme necessitates that we work in the Hölder

class $C^{m,\alpha}$, $m \geq 3$, $\alpha \in (\frac{1}{2}, 1)$ instead of the usual C^m class, and we will thus have to accommodate for this for the rest of this paper, but this does not affect any of the arguments made so far). This method of solving the orthogonality equation $\nabla\phi \cdot \nabla\psi = 0$ only requires us to look at first and second derivatives. We will show in Section 1.2 that generalizing this Nash–Moser iteration argument of [77] from the case of $\mathbb{H}^3$ to the broader setting of simply connected nilpotent Lie groups does not incur serious difficulties even if the step size of G is greater than 2. We will also show that by modifying the techniques of [77], one can solve the stronger orthogonality equation

$$X_i\phi \cdot X_j\psi + X_j\phi \cdot X_i\psi = 0, \quad i, j = 1, \cdots, k,$$

but we will not be able to obtain the even stronger orthogonality equation

$$X_i\phi \cdot X_j\psi = 0, \quad i, j = 1, \cdots, k.$$

See Proposition 1.2.1 and Corollary 1.2.4 for the precise statement. One can find a detailed description and motivation of this Nash–Moser iteration scheme in the introduction to [77].

In light of the proof method for Theorem 1.0.2 described above, it is natural to ask whether these methods can be generalized further. Many of these methods depend on the fact that nilpotent Lie groups have as tangent spaces and asymptotic cones Carnot groups (recall that Carnot groups arise as tangent spaces of sub-Riemannian manifolds). It would be necessary to revamp many of the ideas here, especially the Nash–Moser iteration scheme, to be applicable to the setting of doubling metric spaces. Even the orthonormal vectorfield extension theorem (Theorem 1.3.3) would require some reformulation since it is unlikely that we will have vectorfields in the setting of doubling metric spaces; at the least, the vectorfields should be replaced by elements of a Grassmannian. One realistic hope is that one could transform Theorem 1.3.3 into

a higher-dimensional block basis variant of the construction of [61] and thus construct an embedding with distortion $O(\varepsilon^{-1/2-\delta})$ using the random net construction of that work. One could also improve upon the results of [61] by using hierarchical nets since these may be easier to control and describe compared to ordinary nets (see [33] for the construction and applications of hierarchical nets).

The rest of this paper is organized as follows. We first introduce some elementary background in Section 1.1, and develop the Nash–Moser perturbation theorem in Section 1.2 and the orthonormal extension theorem in Section 1.3. It is in Section 1.3 that the proof idea differs the most from [77]: whereas that work proves the orthonormal extension theorem in the spirit of quantitative topology, we will prove it using the concentration of measure phenomenon on the sphere and the Lovász local lemma. We then develop the main iteration lemma in Section 1.4 and show how it gives us our desired embedding in Section 1.5.

1.1 Preliminaries

1.1.1 Reduction of Theorem 1.0.1 to Theorem 1.0.2, and reduction of Theorem 1.0.2 to the case of nonabelian simply connected nilpotent Lie groups

It is useful to note the following byproduct of the proof of the Assouad embedding theorem [3].

Theorem 1.1.1. *Let (X, d) be a K-doubling metric space. For $0 < r_1 < r_2$, there exists $d \in \mathbb{N}$ and a 1-Lipschitz mapping $f : X \to \mathbb{R}^d$ such that*

$$\sup_{x \in X} |f(x)| = O_K(r_2),$$

and

$$|f(x_1) - f(x_2)| \gtrsim_K d(x_1, x_2), \quad \text{for all } x_1, x_2 \in X \text{ such that } r_1 \leq d(x_1, x_2) \leq r_2.$$

We now begin by noting that to prove Theorem 1.0.1, it is enough to prove Theorem 1.0.2.

Let Γ be a finitely generated group of polynomial growth. But then Γ possesses a finite index nilpotent subgroup Γ' [30], while if we let T be the torsion subgroup of Γ', it is well-known that T is a finite normal subgroup, and that the quotient subgroup $\Gamma'' = \Gamma'/T$ is a torsion-free nilpotent group. But then Γ'' is a cocompact lattice of some simply connected nilpotent Lie group G [55]. By the Švarc-Milnor lemma [76, 59], Γ is quasi-isometric to Γ' which is quasi-isometric to Γ'', while Γ'' is clearly quasi-isometric to G. Here, a map $f : (\mathcal{M}, d) \to (\mathcal{N}, \delta)$ between metric spaces is called a *quasi-isometry* if there exists $C > 0$ such that

$$\frac{1}{C} d(x, y) - C \leq \delta(f(x), f(y)) \leq C d(x, y) + C \quad \text{for all } x, y \in \mathcal{M}$$

and

$$\text{for all } n \in \mathcal{N} \text{ there exists } m \in \mathcal{M} \text{ such that } \delta(f(m), n) < C.$$

In summary, given Γ as above, there exists a simply connected nilpotent Lie group G and a quasi-isometry $f : \Gamma \to G$. Since G is uniformly discrete, f is Lipschitz, and is bilipschitz at large scales: there exist $C', C'' > 0$ such that

$$\text{for all } g_1, g_2 \in \Gamma, \ d_W(g_1, g_2) > C' \text{ implies } d_G(f(g_1), f(g_2)) > C'' d_W(g_1, g_2).$$

Thus, assuming Theorem 1.0.2, there exists a natural number D_G such that for all $0 < \varepsilon < 1/2$ there exists an $O_G\left(\varepsilon^{-1/2}\right)$-bilipschitz mapping $\Phi_\varepsilon : (G, d_G^{1-\varepsilon}) \to \mathbb{R}^{D_G}$;

by rescaling, say it satisfies, for some constant $C_G > 0$,

$$d_G(h_1, h_2)^{1-\varepsilon} \leq |\Phi_\varepsilon(h_1) - \Phi_\varepsilon(h_2)| \leq \frac{C_G}{\varepsilon^{1/2}} d_G(h_1, h_2)^{1-\varepsilon}, \quad \text{for all } h_1, h_2 \in G.$$

The composition $\Phi_\varepsilon \circ f : (\Gamma, d_W^{1-\varepsilon}) \to \mathbb{R}^{D_G}$ is $O_G(\varepsilon^{-1/2})$-Lipschitz, while it almost satisfies the lower bound:

$$|\Phi_\varepsilon \circ f(g_1) - \Phi_\varepsilon \circ f(g_2)| \geq C''^{1-\varepsilon} d_W(g_1, g_2)^{1-\varepsilon} \quad \text{for all } g_1, g_2 \in \Gamma \text{ with } d_W(g_1, g_2) > C'.$$

But by Theorem 1.1.1, there exists a mapping $\phi : \Gamma \to \mathbb{R}^{O_\Gamma(1)}$ such that $|\phi| = O_\Gamma(1)$ and

$$|\phi(g_1) - \phi(g_2)| \asymp_\Gamma 1 \quad \text{whenever } 1 \leq d_W(g_1, g_2) \leq C'.$$

The map $(\Phi_\varepsilon \circ f) \oplus \phi : \Gamma \to \mathbb{R}^{D_G} \oplus \mathbb{R}^{O_\Gamma(1)}$ gives the desired map.

We remark that, if Γ is not virtually abelian, then the corresponding torsion-free nilpotent group Γ'' is also not virtually abelian [15] and hence its Mal'cev completion G is nonabelian (and simply connected nilpotent). So the "hard case" of Theorem 1.0.1 reduces to the "hard case" of Theorem 1.0.2.

We now reduce Theorem 1.0.2 to the case where the underlying Lie group is simply connected. Deducing Theorem 1.0.2 for G from Theorem 1.0.2 for G/K is similar to our deduction of Theorem 1.0.1 from Theorem 1.0.2; it is enough to construct $O_G(\varepsilon^{-1/2})$-Lipschitz mappings $(G, d_G^{1-\varepsilon}) \to \mathbb{R}^{O_G(1)}$ that satisfy the lower Lipschitz bounds for any points $g_1, g_2 \in G$ in the bounds $d_G(g_1, g_2) = \Omega_G(1)$, $d_G(g_1, g_2) = \Theta_G(1)$, and $d_G(g_1, g_2) = O_G(1)$, respectively. The second component arises from Theorem 1.1.1, so we only need to construct the first and third components.

Let us construct the first component. We are given G, a connected nilpotent Lie group, with Lie algebra $\mathfrak{g}$. Then it has a maximal nontrivial compact subgroup K, with Lie algebra $\mathfrak{k}$, contained in the center of G, while G/K is a simply connected

nilpotent Lie group, with Lie algebra $\mathfrak{g}/\mathfrak{k}$ [80, IV.1.5]. If $V \leq \mathfrak{g}$ is the horizontal subspace along which distances are measured in G, and $\pi : \mathfrak{g} \to \mathfrak{k}$ is the quotient map, let $\pi(V)$ be the horizontal subspace along which distances are measured in G/K. Then the quotient map $\pi : G \to G/K$ is a Lipschitz quasi-isometry. Therefore, composing with π the map resulting from Theorem 1.0.2 applied to G/K gives the bilipschitz mapping that reads scales $\Omega_G(1)$.

To construct the third component, apply Theorem 1.0.2 to the universal cover $\tilde{G}$ of G (e.g., the exponential map $\exp : \mathfrak{g} \to G$ works) to get a $O_{\tilde{G}}(\varepsilon^{-1/2})$-bilipschitz mapping $(\tilde{G}, d_{\tilde{G}}^{1-\varepsilon}) \to \mathbb{R}^{O_{\tilde{G}}(1)}$. Multiply with a smooth bump function that is supported on $B^{\tilde{G}}_{r_{inj}/2}$, where r_{inj} is the injectivity radius of the exponential map $\exp : \tilde{G} \to G$; using the fact that $\tilde{G}$ and G are locally isometric, we obtain an $O_{\tilde{G}}(\varepsilon^{-1/2})$-bilipschitz mapping $\theta : G \to \mathbb{R}^d$, $d = O_{\tilde{G}}(1)$ that is supported on $B^G_{r_{inj}/2}$ and such that

$$\text{for all } g_1, g_2 \in B^G_{r_{inj}/4} \quad |\theta(g_1) - \theta(g_2)| \gtrsim_G d_G(g_1, g_2)^{1-\varepsilon}.$$

Let N be a maximal $r_{inj}/4$-separated set of G. We can partition N into a finite number of sets $N_1, \cdots, N_n$ such that each N_i is r_{inj}-separated. Then

$$\oplus_{i=1}^n \sum_{a \in N_i} \theta(g^{-1}\cdot) : (G, d_G^{1-\varepsilon}) \to \mathbb{R}^{nd}$$

is an $O_G(\varepsilon^{-1/2})$-Lipschitz map that reads scales $O_G(1)$.

This completes the proof that to prove Theorem 1.0.2, it is enough to assume G simply connected.

In case G is an abelian simply connected nilpotent Lie group, i.e., $G = \mathbb{R}^n$ for some n, Assouad [3] constructed $O_n(1)$-bilipschitz mappings $(\mathbb{R}^n, d_{\mathbb{R}^n}^{1-\varepsilon}) \to \mathbb{R}^{d\lceil \frac{1}{1-\varepsilon} \rceil}$. So it is enough to consider the case when G is nonabelian.

For the rest of the chapter, let G be a fixed nonabelian simply connected nilpotent Lie group.

1.1.2 Hierarchy of constants

We will select absolute constants in the following order:

- A Hölder exponent $\alpha \in (\frac{1}{2}, 1)$ and a level of regularity m^* depending on G. For simplicity, one can fix $\alpha = \frac{2}{3}$ and $m^* = s^2 + s + 1$, where s is the step size of G.

- A sufficiently large natural number $C_0 > 1$ depending on G, α and m^*. Specifically, this choice will occur in (1.95), the third inequality of (1.96), the second inequality of (1.97), (1.98), right after (1.113), right before (1.116), (1.117), right after (1.119), and (1.139).

- A sufficiently large dyadic number N_0 depending on G and C_0. This choice will occur in (1.50), in the derivation of (1.62) of Corollary 1.2.4, (1.101), (1.102), (1.104), (1.105), (1.106), and (1.118).

- A sufficiently large dyadic number A depending on G, C_0 and N_0. This choice will occur in (1.51), (1.107), (1.111) and after it, (1.112), right after (1.113), right before (1.116), right after (1.119), right after (1.164), (1.166), (1.167), and (1.168).

1.1.3 Basic linear algebra

Denote by $|\cdot|$ the Euclidean metric and by $\langle \cdot, \cdot \rangle$ the Euclidean inner product on Euclidean spaces $\mathbb{R}^D$.

If $T : \mathbb{R}^{D_1} \to \mathbb{R}^{D_2}$ is a linear map, we also denote by $|T|$ the Frobenius norm of T. Also, for $1 \leq n \leq D$, the exterior power $\bigwedge^n \mathbb{R}^D$ can be identified with $\mathbb{R}^{\binom{D}{n}}$, and so we can also define a Euclidean norm $|\cdot|$ and a Euclidean inner product $\langle \cdot, \cdot \rangle$ on $\bigwedge^n \mathbb{R}^D$. With this norm on $\bigwedge^n \mathbb{R}^D$, the *Cauchy–Binet formula* tells us that for $v_1, \cdots, v_n \in \mathbb{R}^D$,

$$|v_1 \wedge \cdots \wedge v_n|^2 = \det\left(v_i \cdot v_j\right)_{1 \leq i,j \leq n} = \det(TT^*),$$

where $T : \mathbb{R}^D \to \mathbb{R}^n$ is the linear map

$$T(u) \coloneqq (u \cdot v_1, \cdots, u \cdot v_n).$$

More generally, the *polarized Cauchy–Binet formula* tells us that for $u_1, \cdots, u_n$, $v_1, \cdots, v_n \in \mathbb{R}^D$,

$$\left\langle u_1 \wedge \cdots \wedge u_n, v_1 \wedge \cdots \wedge v_n \right\rangle = \det\left(u_i \cdot v_j\right)_{1 \le i,j \le n}.$$

It is not difficult to see that we have a Cauchy–Schwarz-like inequality: for every $v_1, \cdots, v_n \in \mathbb{R}^D$ and $1 \le i < n \le D$, we have

$$|v_1 \wedge \cdots \wedge v_n| \le |v_1 \wedge \cdots \wedge v_i||v_{i+1} \wedge \cdots \wedge v_n|.$$

We will simply refer to this as the Cauchy–Schwarz inequality in the rest of this paper.

1.1.4 Some metric space geometry

Let (X, d) be a metric space. For any $f : X \to \mathbb{R}^D$, we define the norms

$$\|f\|_{C^0} \coloneqq \sup_{x \in X} |f(x)|, \quad \|f\|_{\mathrm{Lip}} \coloneqq \sup_{x,y \in X,\ x \neq y} \frac{|f(x) - f(y)|}{d(x,y)}.$$

These norms satisfy certain algebraic properties. If $f, g : X \to \mathbb{R}^D$ then

$$\|f \cdot g\|_{\mathrm{Lip}} \le \|f\|_{C^0}\|g\|_{\mathrm{Lip}} + \|f\|_{\mathrm{Lip}}\|g\|_{C^0}. \tag{1.2}$$

Also, if $f : X \to \mathbb{R}$ and $f(x) \geq c > 0$ for all $x \in X$ then

$$\|1/f\|_{\text{Lip}} \leq c^{-2}\|f\|_{\text{Lip}}, \tag{1.3}$$

$$\|\sqrt{f}\|_{\text{Lip}} \leq \frac{1}{2\sqrt{c}}\|f\|_{\text{Lip}}. \tag{1.4}$$

One can use these properties to see that for $f : X \to \mathbb{R}^D$ with $|f(x)| \geq c > 0$ for all $x \in X$ we have

$$\left\|\frac{f}{|f|}\right\|_{\text{Lip}} \leq \left\|\frac{1}{|f|}\right\|_{C^0} \|f\|_{\text{Lip}} + \left\|\frac{1}{|f|}\right\|_{\text{Lip}} \|f\|_{C^0} \leq (c^{-1} + c^{-2}\|f\|_{C^0})\|f\|_{\text{Lip}}. \tag{1.5}$$

For $\delta > 0$, a subset $\mathcal{N}_\delta \subset X$ is a *δ-net* if for any distinct $x, y \in \mathcal{N}_\delta$ one has $d(x, y) \geq \delta$. By Zorn's lemma, δ-nets which are maximal with respect to inclusion exist, and if $\mathcal{N}_\delta$ is a maximal δ-net then we have the covering

$$X = \bigcup_{x \in \mathcal{N}_\delta} B_\delta(x).$$

An immediate consequence of the doubling property is that if X is a K-doubling metric space and $m \geq 0$, then for any δ-net $\mathcal{N}_\delta$ we have

$$|\mathcal{N}_\delta \cap B_{2^m\delta}(x)| \leq K^{m+1} \quad \text{for all } x \in X. \tag{1.6}$$

1.1.5 Function spaces on nilpotent Lie groups

We will assume that the norm $\|\cdot\|$ on V is an inner product norm on V (in other words, we may assume G is a sub-Riemannian manifold, as opposed to being a sub-Finsler manifold). This is possible by John's ellipsoid theorem, which allows us to replace $\|\cdot\|$ by an inner product norm while introducing distortion at most $\sqrt{\dim V}$, which is acceptable since this is independent of the amount ε of snowflaking.

We fix a left-invariant orthonormal basis $X_1, \cdots, X_k$ of V with respect to $\|\cdot\|$. If

$\phi : G \to \mathbb{R}^D$ is a differentiable function, we let $\nabla\phi : G \to \mathbb{R}^{kD}$ denote the horizontal gradient

$$\nabla\phi \coloneqq (X_1\phi, \cdots, X_k\phi).$$

By iteration, we have $\nabla^m\phi : G \to \mathbb{R}^{k^m D}$ for any $m \geq 1$, if ϕ is m times differentiable. We recall the C^0 norm

$$\|\phi\|_{C^0} = \sup_{p\in G} |\phi(p)|,$$

and define the higher C^m norms

$$\|\phi\|_{C^m} \coloneqq \sum_{0\leq j\leq m} \|\nabla^j\phi\|_{C^0}.$$

For a fixed spatial scale $R > 0$, we define the C^m_R norm to be the rescaled norm

$$\|\phi\|_{C^m_R} \coloneqq \sum_{0\leq j\leq m} R^j\|\nabla^j\phi\|_{C^0}.$$

Given a Hölder exponent $0 < \alpha < 1$ we may also define the homogeneous Hölder norm

$$\|\phi\|_{\dot{C}^{0,\alpha}} \coloneqq \sup_{p,q\in G, p\neq q} \frac{|\phi(p) - \phi(q)|}{d(p,q)^\alpha}$$

and the higher Hölder norms

$$\|\phi\|_{C^{m,\alpha}} \coloneqq \|\phi\|_{C^m} + \|\nabla^m\phi\|_{\dot{C}^{0,\alpha}}$$

and more generally, the rescaled Hölder norm

$$\|\phi\|_{C^{m,\alpha}_R} \coloneqq \|\phi\|_{C^m_R} + R^{m+\alpha}\|\nabla^m\phi\|_{\dot{C}^{0,\alpha}}.$$

One may easily verify

$$\|\phi\|_{C^{m,\alpha}_R} \lesssim \|\phi\|_{C^{m+1}_R}.$$

By an iterated application of the product rule, one can verify the algebra properties

$$\|\phi\psi\|_{C^m_R} \lesssim_m \|\phi\|_{C^m_R}\|\psi\|_{C^m_R},$$
$$\|\phi\psi\|_{C^{m,\alpha}_R} \lesssim_m \|\phi\|_{C^{m,\alpha}_R}\|\psi\|_{C^{m,\alpha}_R}.$$

These inequalities continue to hold when ϕ and ψ are vector-valued and we take the wedge product or the dot product, where the constants do not depend on the dimension of the codomain of ϕ and ψ:

$$\|\phi\cdot\psi\|_{C^m_R} \lesssim_m \|\phi\|_{C^m_R}\|\psi\|_{C^m_R}, \quad \|\phi\cdot\psi\|_{C^{m,\alpha}_R} \lesssim_m \|\phi\|_{C^{m,\alpha}_R}\|\psi\|_{C^{m,\alpha}_R},$$
$$\|\phi\wedge\psi\|_{C^m_R} \lesssim_m \|\phi\|_{C^m_R}\|\psi\|_{C^m_R}, \quad \|\phi\wedge\psi\|_{C^{m,\alpha}_R} \lesssim_m \|\phi\|_{C^{m,\alpha}_R}\|\psi\|_{C^{m,\alpha}_R}.$$

More generally, one can observe that these algebra properties continue to hold when we replace the above norms with norms of the form

$$\|\phi\|_{C^m_{\{R_j\}_{i=0}^m}} \coloneqq \sum_{0\le j\le m} R_j\|\nabla^j\phi\|_{C^0},$$

where $\{R_j\}_{i=0}^m$ is a "sequence of spatial scales", i.e., a sequence of positive real numbers, that is log-concave: $R_iR_j \ge R_{i+j}$. Examples of such norms include

$$\|\phi\| + R\|\nabla\phi\|_{C^m}, \text{ or } \|\phi\|_{C^0} + \|\nabla\phi\|_{C^m_{1/R}}, \quad R \ge 1. \tag{1.7}$$

1.1.6 Basic metric geometry of simply connected nilpotent Lie groups

As G is a nonabelian nilpotent Lie group, its topological dimension must at least be 3.

As G is nilpotent and simply connected, the exponential map $\exp : \mathfrak{g} \to G$ is a

diffeomorphism.

For $r > 0$, we define the open ball

$$B_r \coloneqq \{h \in G : d_G(h, e_G) < r\},$$

and for $g \in G$ and $r > 0$ we define the open ball

$$B_r(g) \coloneqq \{h \in G : d_G(h, g) < r\} = gB_r,$$

where the last equality follows from left-invariance of the metric.

We will first review Carnot group geometry, as it serves as a model case. Then we will discuss the local and global metric geometry of G.

Carnot group geometry

Recall that in this case we are given the decomposition $\mathfrak{g} = V_1 \oplus V_2 \oplus \cdots \oplus V_s$, and we have set $V = V_1$. In this case, define $\dim G = n$, $\dim V_r = k_r$, $k = k_1$, and the Hausdorff dimension $n_h \coloneqq \sum_{r=1}^{s} rk_r$. Since G is nonabelian, we have $s \geq 2$, since if $s = 1$, then G is just a finite-dimensional Euclidean space. This will give $n_h \geq 4$, as we must have $k_1 \geq 2$ and $k_2 \geq 1$.

For $2 \leq r \leq s$, we fix a basis $X_{r,1}, \cdots, X_{r,k_r}$ of V_r and extend them to left-invariant vectorfields over G. For $r = 1$, we simply write $X_{1,i} = X_i$.

Recall that we can define the scaling maps $\delta_\lambda : \mathfrak{g} \to \mathfrak{g}$ for $\lambda > 0$ by

$$\delta_\lambda|_{V_r} = \lambda^i \mathrm{id}_{V_r}, \quad r = 1, \cdots, s.$$

One may then define the dilation $\delta_\lambda : G \to G$ so that it commutes with exp:

$$\delta_\lambda \circ \exp = \exp \circ \delta_\lambda.$$

One can compute that δ_λ is the unique Lie group automorphism $\delta_\lambda : G \to G$ such that $(\delta_\lambda)_* = \delta_\lambda$. Moreover, δ_λ interacts with the left-invariant vector fields as follows:

$$X_{r,i}(\phi \circ \delta_\lambda) = \lambda^r (X_{r,i}\phi) \circ \delta_\lambda, \quad r = 1, \cdots, s, \ i = 1, \cdots, k_r.$$

The special case $r = 1$ tells us that δ_λ is a scaling in the Carnot-Carathéodory metric:

$$d_G(\delta_\lambda(p), \delta_\lambda(p')) = \lambda d_G(p, p'), \quad p, p' \in G.$$

By iteration, we can also deduce

$$\nabla^m(\phi \circ \delta_\lambda) = \lambda^m (\nabla^m \phi) \circ \delta_\lambda.$$

One can parametrize G by $\mathbb{R}^n$, by first identifying G with $\mathfrak{g}$ via the exponential map exp, and then identifying $\mathfrak{g}$ with $\mathbb{R}^n$ via the basis $\{X_{r,i}\}_{1 \le r \le s, 1 \le i \le k_r}$. Denote the corresponding canonical basis as $\{f_{r,i}\}_{1 \le r \le s, 1 \le i \le k_r}$.

We may define a weighted degree for polynomials in $x_{r,i}$ by assigning degree r to $x_{r,i}$. It is clear that δ_λ acting upon a homogeneous polynomial of degree m is just multiplication by λ^m, so the differential operator $X_{r,i}$ acts on polynomials by reducing the weighted degree by r in each term. One can also see, using the scaling δ_λ, that

$$d_G\left(\exp(\sum_{r=1}^{s}\sum_{i=1}^{k_r} x_{r,i} X_{r,i}), e_G\right) \asymp_G \sum_{r=1}^{s}\sum_{i=1}^{k_r} |x_{r,i}|^{1/r}.$$

The Baker-Campbell-Hausdorff formula gives the group law in this coordinate system:

$$gh = \sum_{m=1}^{s} \frac{(-1)^{m-1}}{m} \sum_{\substack{r_i + s_i > 0 \\ i=1,\ldots,m}} \frac{[g^{r_1} h^{s_1} g^{r_2} h^{s_2} \cdots g^{r_m} h^{s_m}]}{(\sum_{j=1}^{m}(r_j + s_j)) \cdot \prod_{i=1}^{m} r_i! s_i!},$$

where the sum is finite since G is of step s, and we have used the notation

$$[g^{r_1}h^{s_1}\cdots g^{r_m}h^{s_m}] = [\underbrace{g,[g,\cdots[g}_{r_1},[\underbrace{h,[h,\cdots[h}_{s_1},\cdots\ [\underbrace{g,[g,\cdots[g}_{r_m},[\underbrace{h,[h,\cdots h}_{s_m}]]\cdots]]].$$

Thus, we can see that

$$\exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r}x_{r,i}^0X_{r,i}\right)\exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r}x_{r,i}^1X_{r,i}\right)=\exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r}x_{r,i}^2X_{r,i}\right)$$

where

$$x_{r,i}^2 = x_{r,i}^0 + x_{r,i}^1 + (\text{homogeneous polynomial of } \{x_{r',i'}^0\}_{r'<r}, \{x_{r',i'}^1\}_{r'<r} \text{ of degree } r).$$

It follows that the Lebesgue measure on $\mathbb{R}^n$ is a Haar measure on G because the Jacobian of left-multiplication is a unit upper triangular matrix. Also, the left-invariant vectors $\{X_{r,i}\}_{1\le r\le s, 1\le i\le k_r}$ in this coordinate system take the form

$$X_{r,i} = \frac{\partial}{\partial x_{r,i}} + \sum_{r'>r}^{s}\sum_{j=1}^{k_{r'}}(\text{homogeneous polynomial of } \{x_{r'',i'}\}_{r''<r'} \text{ of degree } r'-r)\frac{\partial}{\partial x_{r',j}}. \tag{1.8}$$

For $r, r' = 1,\cdots,s,\ i = 1,\cdots,k_r,\ i' = 1,\cdots,k_{r'}$, we define the lexicographic ordering

$$(r,i)\preceq(r',i') \quad \Leftrightarrow \quad r<r' \text{ or } r=r' \text{ and } i\le i'.$$

A simple volumetric argument gives the following bound for any δ-net $\mathcal{N}_\delta$:

$$|\mathcal{N}_\delta\cap B_R(p)|\le\left(\frac{2R}{\delta}+1\right)^{n_h},\quad p\in G, R>0. \tag{1.9}$$

Local metric geometry

Instead of the grading of $\mathfrak{g}$, we have the following increasing sequence of subspaces

$$V \leq V + [V,V] \leq \cdots \leq V + [V,V] + \underbrace{[V,[V,\cdots,V]]}_{s \text{ times}} = \mathfrak{g}$$

where s denotes the nilpotency step size of G. One may consider the Carnot group with Lie algebra graded in the manner

$$V \oplus (V + [V,V])/V \oplus \cdots \oplus (V + [V,V] + \underbrace{[V,[V,\cdots,V]]}_{s \text{ times}})/(V + [V,V] + \underbrace{[V,[V,\cdots,V]]}_{s-1 \text{ times}})$$

with a natural induced Lie bracket. This is the Carnot group that arises as the tangent space of G, and we can quantitatively approximate the distance, group operation, and left-invariant vectors accordingly.

For $2 \leq r \leq s$, we choose $X^{loc}_{r,1}, \cdots, X^{loc}_{r,k^{loc}_r} \in \underbrace{[V,[V,\cdots,V]]}_{r \text{ times}}$ so that

$$\{X^{loc}_{r',i}\}_{1 \leq r' \leq r,\ 1 \leq i \leq k^{loc}_{r'}} \text{ is a basis of } V + [V,V] + \underbrace{[V,[V,\cdots,V]]}_{r \text{ times}}$$

(it is possible that $k^{loc}_r = 0$ for large r). Extend these vectors to left-invariant vector-fields over G. For $r = 1$, we simply write $X^{loc}_{1,i} = X_i$.

One can parametrize G by $\mathbb{R}^n$, by first identifying G with $\mathfrak{g}$ via the exponential map exp, and then identifying $\mathfrak{g}$ with $\mathbb{R}^n$ via the basis $\{X^{loc}_{r,i}\}_{1 \leq r \leq s, 1 \leq i \leq k^{loc}_r}$. We will denote the corresponding canonical basis as $\{f^{loc}_{r,i}\}_{1 \leq r \leq s, 1 \leq i \leq k^{loc}_r}$; this defines polynomial functions $\{x^{loc}_{r,i}\}_{1 \leq r \leq s, 1 \leq i \leq k^{loc}_r}$.

The following claims can be found in [41, Chapter 2]. We have

$$d_G\left(\exp(\sum_{r=1}^{s}\sum_{i=1}^{k_r} x^{loc}_{r,i} X^{loc}_{r,i}), e_G\right) \asymp_G \sum_{r=1}^{s}\sum_{i=1}^{k_r} |x^{loc}_{r,i}|^{1/r}, \quad \text{if } \forall r,i\ |x^{loc}_{r,i}| \leq 1.$$

We may define the *weighted local degree* for polynomials in $x^{loc}_{r,i}$ by assigning local degree r to $x^{loc}_{r,i}$. (What we call 'local' is called 'nonholonomic' in [41]. We made this choice of terminology to distinguish with the 'global' case.) The group law, in this coordinate system, is given via the Baker-Campbell-Hausdorff formula in the form of

$$\exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} x^{loc}_{r,i}X^{loc}_{r,i}\right)\exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} y^{loc}_{r,i}X^{loc}_{r,i}\right)=\exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} z^{loc}_{r,i}X^{loc}_{r,i}\right)$$

where

$$z^{loc}_{r,i}=x^{loc}_{r,i}+y^{loc}_{r,i}+(\text{polyn in } \{x^{loc}_{r',i'}\},\{y^{loc}_{r',i'}\} \text{ of weighted local degree} \geq r).$$

Thus the left-invariant vectors $\{X^{loc}_{r,i}\}_{1\leq r\leq s, 1\leq i\leq k^{loc}_r}$ in this coordinate system take the form

$$X^{loc}_{r,i}=\frac{\partial}{\partial x^{loc}_{r,i}}+\sum_{r'=1}^{s}\sum_{j=1}^{k^{loc}_{r'}}(\text{polyn in } \{x^{loc}_{r'',i'}\} \text{ of local degree} \geq r'-r)\frac{\partial}{\partial x^{loc}_{r',j}}. \qquad (1.10)$$

So the differential operator $X^{loc}_{r,i}$ acts on polynomials by reducing the weighted degree by r or less in each term.

For $r,r'=1,\cdots,s$, $i=1,\cdots,k^{loc}_r$, $i'=1,\cdots,k^{loc}_{r'}$, we define the lexicographic ordering

$$(r,i)\preceq^{loc}(r',i') \quad \Leftrightarrow \quad r<r' \text{ or } r=r' \text{ and } i\leq i'.$$

Global metric geometry

This time, instead of the grading of $\mathfrak{g}$, we have the following decreasing sequence of subspaces

$$\mathfrak{g}\geq[\mathfrak{g},\mathfrak{g}]\geq\cdots\geq\underbrace{[\mathfrak{g},[\mathfrak{g},\cdots,\mathfrak{g}]]}_{s \text{ times}}$$

where s again denotes the nilpotency step size of G. One may consider the Carnot group with Lie algebra graded in the manner

$$\mathfrak{g}/[\mathfrak{g},\mathfrak{g}] \oplus [\mathfrak{g},\mathfrak{g}]/[\mathfrak{g},[\mathfrak{g},\mathfrak{g}]] \oplus \cdots \oplus \underbrace{[\mathfrak{g},[\mathfrak{g},\cdots,\mathfrak{g}]]}_{s-1 \text{ times}} / \underbrace{[\mathfrak{g},[\mathfrak{g},\cdots,\mathfrak{g}]]}_{s \text{ times}} \oplus \underbrace{[\mathfrak{g},[\mathfrak{g},\cdots,\mathfrak{g}]]}_{s \text{ times}}$$

with a natural induced Lie bracket. This is the Carnot group that arises as the asymptotic cone of G, and we can quantitatively approximate the distance, group operation, and left-invariant vectors accordingly.

For $1 \leq r \leq s$, we choose $X_{r,1}^{glob}, \cdots, X_{r,k_r^{glob}}^{glob} \in \underbrace{[\mathfrak{g},[\mathfrak{g},\cdots,\mathfrak{g}]]}_{r \text{ times}}$ so that

$$\{X_{r',i}^{glob}\}_{r \leq r' \leq s,\ 1 \leq i \leq k_{r'}^{glob}} \text{ is a basis of } \underbrace{[\mathfrak{g},[\mathfrak{g},\cdots,\mathfrak{g}]]}_{r \text{ times}}.$$

(This time, we cannot have $k_r^{glob} = 0$ for any r.) Extend these vectors to left-invariant vectorfields over G. As

$$\underbrace{[\mathfrak{g},[\mathfrak{g},\cdots,\mathfrak{g}]]}_{r \text{ times}} = \underbrace{[V,[V,\cdots,V]]}_{r \text{ times}} + \underbrace{[\mathfrak{g},[\mathfrak{g},\cdots,\mathfrak{g}]]}_{r+1 \text{ times}},$$

we may assume that $X_{r,1}^{glob}, \cdots, X_{r,k_r^{glob}}^{glob} \in \underbrace{[V,[V,\cdots,V]]}_{r \text{ times}}$.

One can parametrize G by $\mathbb{R}^n$, by first identifying G with $\mathfrak{g}$ via the exponential map exp, and then identifying $\mathfrak{g}$ with $\mathbb{R}^n$ via the basis $\{X_{r,i}^{glob}\}_{1 \leq r \leq s, 1 \leq i \leq k_r^{glob}}$. We will denote the corresponding canonical basis as $\{f_{r,i}^{glob}\}_{1 \leq r \leq s, 1 \leq i \leq k_r^{glob}}$; this defines polynomial functions $\{x_{r,i}^{glob}\}_{1 \leq r \leq s, 1 \leq i \leq k_r^{glob}}$.

By [68], we have the distance formula

$$d_G\left(\exp(\sum_{r=1}^{s}\sum_{i=1}^{k_r} x_{r,i}^{glob} X_{r,i}^{glob}), e_G\right) \asymp_G \sum_{r=1}^{s}\sum_{i=1}^{k_r} |x_{r,i}^{glob}|^{1/r}, \quad \text{if } \exists r,i\ |x_{r,i}^{glob}| > 1.$$

We may define the *weighted global degree* for polynomials in $x_{r,i}^{glob}$ by assigning global

degree r to $x_{r,i}^{glob}$. Recalling the Baker-Campbell-Hausdorff formula, in this coordinate system, the group law is given by

$$\exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} x_{r,i}^{glob} X_{r,i}^{glob}\right)\exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} y_{r,i}^{glob} X_{r,i}^{glob}\right) = \exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} z_{r,i}^{glob} X_{r,i}^{glob}\right)$$

where

$$z_{r,i}^{glob} = x_{r,i}^{glob} + y_{r,i}^{glob} + (\text{polyn in } \{x_{r',i'}^{glob}\}, \{y_{r',i'}^{glob}\} \text{ of weighted global degree} \le r).$$

Thus the left-invariant vectors $\{X_{r,i}^{glob}\}_{1\le r\le s, 1\le i\le k_r^{glob}}$ in this coordinate system take the form

$$X_{r,i}^{glob} = \frac{\partial}{\partial x_{r,i}^{glob}} + \sum_{r'>r}^{s}\sum_{j=1}^{k_{r'}^{glob}} (\text{polyn in } \{x_{r'',i'}^{glob}\} \text{ of global degree} \le r'-r)\frac{\partial}{\partial x_{r',j}^{glob}}. \quad (1.11)$$

So the differential operator $X_{r,i}^{glob}$ acts on polynomials by reducing the weighted degree by r or more in each term.

For $r, r' = 1, \cdots, s$, $i = 1, \cdots, k_r^{glob}$, $i' = 1, \cdots, k_{r'}^{glob}$, we define the lexicographic ordering

$$(r,i) \preceq^{glob} (r',i') \quad \Leftrightarrow \quad r < r' \text{ or } r = r' \text{ and } i \le i'.$$

1.1.7 Littlewood–Paley theory on nilpotent Lie groups

A basic tool used in [77] was a Littlewood–Paley theory for functions defined on the Heisenberg group. We present the analogous theorem on nilpotent Lie groups.

One can easily modify the argument in [26] to show the following. For a positive number N and a C^0 function $\phi : G \to \mathbb{R}^D$, one can construct the Littlewood–Paley

projection $P_{(\leq N)}\phi : G \to \mathbb{R}^D$, which is a C^∞ function, and the variants

$$P_{(<N)}\phi \coloneqq P_{(\leq N/2)}\phi, \qquad P_{(N)}\phi \coloneqq P_{(\leq N)}\phi - P_{(<N)}\phi,$$
$$P_{(>N)}\phi \coloneqq \phi - P_{(\leq N)}\phi, \quad P_{(\geq N)}\phi \coloneqq \phi - P_{(<N)}\phi$$

which satisfy the following regularity properties.

Theorem 1.1.2 (Littlewood–Paley Theory)**.** *Let $\phi : G \to \mathbb{R}^D$ be bounded and continuous.*

1. *(Littlewood–Paley decomposition) For any dyadic number N_0, we have*

$$\phi = P_{(\leq N_0)}\phi + \sum_{N > N_0 \text{ dyadic}} P_{(N)}\phi.$$

2. *(regularity) If $N > 0$ is a dyadic number, $M > 0$, $j, l \geq 0$, and ϕ is of class C^l, one has the estimates*[5]

$$\|\nabla^l P_{(\leq N)}\phi\|_{C^j_{1/N}} \lesssim_{G,j,l} \|\nabla^l \phi\|_{C^0}, \tag{1.12}$$

$$\|P_{(N)}\phi\|_{C^j_{1/N}} \lesssim_{G,j,l,\alpha} N^{-l}\|\nabla^l\phi\|_{C^0}, N^{-l-\alpha}\|\nabla^l\phi\|_{\dot{C}^{0,\alpha}}, \tag{1.13}$$

$$j \leq l \quad \Rightarrow \quad \|P_{(>N)}\phi\|_{C^j_{1/N}} \lesssim_{G,j,l,\alpha} N^{-l}\|\nabla^l\phi\|_{C^0}, \tag{1.14}$$

$$j < l \quad \Rightarrow \quad \|P_{(>N)}\phi\|_{C^j_{1/N}} \lesssim_{G,j,l,\alpha} N^{-l-\alpha}\|\nabla^l\phi\|_{\dot{C}^{0,\alpha}}, \tag{1.15}$$

$$\|P_{(\leq N)}\phi\|_{C^l_{1/M}}, \|P_{(N)}\phi\|_{C^l_{1/M}}, \|P_{(>N)}\phi\|_{C^l_{1/M}} \lesssim_{G,l} \|\phi\|_{C^l_{1/M}}. \tag{1.16}$$

Proof. The construction of the Littlewood–Paley projection is as in [26] and [77] is

[5] We note that in [72, Theorem 2.1], the condition $j \leq l$ is erroneously omitted in their counterpart of (1.14).

as follows. As the Laplace-Kohn operator

$$L := -\sum_{i=1}^{k} X_i^2$$

is self-adjoint on $L^2(G)$, with G is given its Haar measure (which is the Lebesgue measure on $\mathbb{R}^n$), for any $m \in L^\infty(\mathbb{R})$ one can define bounded operators $m(L)$ on $L^2(G)$ which commute with each other and with L. Furthermore, if $m \in C_c^\infty(\mathbb{R})$, then this operator is given by convolution with a Schwartz function $K : G \to \mathbb{R}$ [26]:[6]

$$m(L)f = f * K, \quad \text{for all } f \in L^2(G),$$

where a function on G is said to be a Schwarz function if it is a Schwarz function on $\mathbb{R}^n$, and $*$ denotes the convolution operator:

$$f * K(p) = \int_G f(g)K(g^{-1}p)dg.$$

For such m, the operator $m(L)$ can be extended to functions in C^0 using the above convolution formula.

Now, if we choose any smooth function $\varphi : \mathbb{R} \to \mathbb{R}$ supported on $[-4, 4]$ that equals 1 on $[-1, 1]$, we can now define, for $N > 0$, the Littlewood–Paley projection $P_{(\leq N)}$ using the above functional calculus by the formula

$$P_{(\leq N)} := \varphi(L/N^2).$$

Let $K_{(\leq N)}$ be the kernel corresponding to $P_{(\leq N)}$ and $K_{(N)}$ the kernel corresponding to $P_{(N)}$.

Now Theorem 1.1.2 (1) is just [26, Proposition 4.1]. We recall the following fundamental estimates on the kernels:

[6] In the case G is a Carnot group, this follows from [39], using the fact that L is hypoelliptic [37].

Theorem 1.1.3. *[26, Proposition 3.2] Let $\theta \in C_c^\infty(\mathbb{R})$, $\alpha \geq 0$, and $I \in \cup_{\beta\in\mathbb{Z}_{\geq 0}} \{1,\cdots,k\}^\beta$. Let the kernel of $\theta(tL)$ be M_t. Then*

$$\|d_G(\cdot,e_G)^\alpha \nabla^I M_t\|_{L^1(G)} \lesssim_{G,\alpha,I,\theta} t^{(\alpha-|I|)/2}.$$

For example, this tells us that for $\alpha \geq 0$, $I \in \cup_{\beta\in\mathbb{Z}_{\geq 0}}\{1,\cdots,k\}^\beta$

$$\|d_G(\cdot,e_G)^\alpha \nabla^I K_{(\leq N)}\|_{L^1(G)} \lesssim_{G,\alpha,I} N^{|I|-\alpha}. \tag{1.17}$$

The proof of the properties listed in Theorem 1.1.2 (2) is now mostly the same as presented in [77, Theorem 6.1]. When following the proof there, the main part that requires modification in the setting of nilpotent Lie groups is the following. For $i = 1,\cdots,k$, given any Schwarz function $K : G \to \mathbb{R}$, we need to show that there are Schwarz functions $K_j : G \to \mathbb{R}$, $j = 1,\cdots,k$, so that for any C^1-function $\phi : G \to \mathbb{R}^D$,

$$\phi * X_i K = \sum_{j=1}^k X_j \phi * K_j. \tag{1.18}$$

This is a consequence of integration by parts:

$$X\phi * K = -\phi * \tilde{X} K,$$

where X is any left-invariant vector field and $\tilde{X}$ its right-invariant counterpart. First, we have

$$\phi * X_i K = -X_i \phi * K + \phi * (X_i - \tilde{X}_i) K.$$

Now recall (1.10) and (1.11):

$$X_{r,i}^{loc} f = \frac{\partial}{\partial x_{r,i}^{loc}} + \sum_{r'=1}^{s}\sum_{j=1}^{k_{r'}^{loc}} (\text{polyn in } \{x_{r'',i'}^{loc}\} \text{ of local degree} \geq r'-r) \frac{\partial}{\partial x_{r',j}^{loc}}, \tag{1.10}$$

$$X^{glob}_{r,i} = \frac{\partial}{\partial x^{glob}_{r,i}} + \sum_{r'>r}^{s} \sum_{j=1}^{k^{glob}_{r'}} (\text{polyn in } \{x^{glob}_{r'',i'}\} \text{ of global degree} \leq r' - r) \frac{\partial}{\partial x^{glob}_{r',j}}. \quad (1.11)$$

It is not hard to see that we can assume $\{X^{loc}_{r,i}\}$ and $\{X^{glob}_{r,i}\}$ are permutations of each other, with a single left-invariant differential operator possibly having different local and global degrees. Of course, in this case, we may assume the polynomials $\{x^{loc}_{r,i}\}$ and $\{x^{glob}_{r,i}\}$ are permutations of each other.

We have the variants of (1.10) and (1.11) for the right-invariant counterparts:

$$\tilde{X}^{loc}_{r,i} = \frac{\partial}{\partial x^{loc}_{r,i}} + \sum_{r'=1}^{s} \sum_{j=1}^{k^{loc}_{r'}} (\text{polyn in } \{x^{loc}_{r'',i'}\} \text{ of local degree} \geq r' - r) \frac{\partial}{\partial x^{loc}_{r',j}}, \quad (1.19)$$

$$\tilde{X}^{glob}_{r,i} = \frac{\partial}{\partial x^{glob}_{r,i}} + \sum_{r'>r}^{s} \sum_{j=1}^{k^{glob}_{r'}} (\text{polyn in } \{x^{glob}_{r'',i'}\} \text{ of global degree} \leq r' - r) \frac{\partial}{\partial x^{glob}_{r',j}}. \quad (1.20)$$

Local and global degree are independent of left and right, and right-invariant differential operators of local degree r reduce the weighted local degree by r or less. In contrast, their global counterparts reduce the weighted global degree by r or more.

From (1.11) and (1.20), we see that

$$X_i - \tilde{X}_i = \sum_{r=2}^{s} \sum_{j=1}^{k_r} q_{i,r,j} \tilde{X}^{glob}_{r,j}, \quad (1.21)$$

where $q_{i,r,j}$ is a polynomial that commutes with $\tilde{X}^{glob}_{r,j}$, with

$$\text{global degree of } q_{i,r,j} \leq r - 1,$$

while from (1.10) and (1.19), we have that

$$\text{local degree of } q_{i,r,j} \geq \text{local degree of } \tilde{X}^{glob}_{r,j} - 1 \geq r - 1,$$

Similarly, we have

$$\tilde{X}_i - X_i = \sum_{r=2}^{s} \sum_{j=1}^{k_r} \tilde{q}_{i,r,j} X_{r,j}^{glob}, \tag{1.22}$$

where $\tilde{q}_{i,r,j}$ is a polynomial that commutes with $X_{r,j}^{glob}$, with

$$\text{global degree of } \tilde{q}_{i,r,j} \leq r-1,$$

while from (1.10) and (1.19), we have that

$$\text{local degree of } \tilde{q}_{i,r,j} \geq \text{local degree of } X_{r,j}^{glob} - 1 \geq r-1,$$

Thus

$$(X_i - \tilde{X}_i)K = \sum_{r=2}^{s} \sum_{j=1}^{k_r} \tilde{X}_{r,j}^{glob}(q_{i,r,j}K).$$

The operator $\tilde{X}_{r,j}^{glob}$ has global degree r and local degree $\leq r$. Thus, $\tilde{X}_{r,j}^{glob}$ can be written as a linear combination of differential operators of the form $\tilde{X}_{j_1} \cdots \tilde{X}_{j_r}$ (note that $\tilde{X}_j$ has local degree 1 while global degree ≥ 1).

We thus have

$$(X_i - \tilde{X}_i)K = \sum_{j=1}^{k} \tilde{X}_j K_{i,j}$$

for Schwarz functions $K_{i,j}$ of the form

$$K_{i,j} = \sum_{r=1}^{s-1} \sum_{j_1,\cdots,j_r=1}^{k} Q_{i,j,r,\{j_1,\cdots,j_r\}} \tilde{X}_{j_1} \cdots \tilde{X}_{j_r} K$$

where $Q_{i,j,r,\{j_1,\cdots,j_r\}}$ are polynomials such that

$$\text{local degree of } Q_{i,j,r,\{j_1,\cdots,j_r\}} \geq r, \quad \text{global degree of } Q_{i,j,r,\{j_1,\cdots,j_r\}} \leq r,$$

which entails

$$|Q_{i,j,r,\{j_1,\cdots,j_r\}}(g)| \lesssim_{G,i,j,r,\{j_1,\cdots,j_r\}} d_G(g,e_G)^r, \quad g \in G.$$

By (1.22), and a similar expansion of the higher-order derivatives, we have the left-invariant counterpart

$$K_{i,j} = \sum_{r=1}^{s(s-1)} \sum_{j_1,\cdots,j_r=1}^{k} \tilde{Q}_{i,j,r,\{j_1,\cdots,j_r\}} X_{j_1} \cdots X_{j_r} K$$

where $\tilde{Q}_{i,j,r,\{j_1,\cdots,j_r\}}$ are polynomials such that

$$\text{local degree of } \tilde{Q}_{i,j,r,\{j_1,\cdots,j_r\}} \geq r, \quad \text{global degree of } \tilde{Q}_{i,j,r,\{j_1,\cdots,j_r\}} \leq r,$$

which entails

$$|\tilde{Q}_{i,j,r,\{j_1,\cdots,j_r\}}(g)| \lesssim_{G,i,j,r,\{j_1,\cdots,j_r\}} d_G(g,e_G)^r, \quad g \in G.$$

Hence

$$\phi * (X_i - \tilde{X}_i)K = -\sum_{j=1}^{k} X_j \phi * K_{i,j},$$

and we have the form (1.18) (with $K_{i,i}$ replaced by $K_{i,i} + K$, which does not damage the above form).

Upon iteration, we have for $j, l \geq 0$ that

$$\phi * X_{i_1} \cdots X_{i_{j+l}} K = \sum_{j_1,\cdots,j_l=1}^{k} X_{j_1} \cdots X_{j_l} \phi * K_{i_1,\cdots,i_{j+l},j_1,\cdots,j_l},$$

with

$$K_{i_1,\cdots,i_{j+l},j_1,\cdots,j_l} = \sum_{r=j}^{ls(s-1)} \sum_{j'_1,\cdots,j'_r=1}^{k} \tilde{Q}_{i_1,\cdots,i_{j+l},j_1,\cdots,j_l,r,\{j'_1,\cdots,j'_r\}} X_{j'_1} \cdots X_{j'_r} K,$$

where $\tilde{Q}_{i_1,\cdots,i_{j+l},j_1,\cdots,j_l,r,\{j'_1,\cdots,j'_r\}}$ are polynomials with local degree $\geq r-j$, global degree $\leq r-j$, and

$$|\tilde{Q}_{i_1,\cdots,i_{j+l},j_1,\cdots,j_l,r,\{j'_1,\cdots,j'_r\}}(g)| \lesssim_{G,i_1,\cdots,i_{j+l},j_1,\cdots,j_l,r,\{j'_1,\cdots,j'_r\}} d_G(g,e_G)^{r-j}, \quad g \in G.$$

We now have

$$\begin{aligned}
\|\nabla^{j+l} P_{(\leq N)}\phi\|_{C^0} &= \sum_{i_1,\cdots,i_{j+l}=1}^{s} \|\phi * X_{i_1} \cdots X_{i_{j+l}} K_{(\leq N)}\|_{C^0} \\
&\leq \sum_{i_1,\cdots,i_{j+l}=1}^{s} \sum_{j_1,\cdots,j_l=1}^{k} \|X_{j_1} \cdots X_{j_l}\phi * K_{i_1,\cdots,i_{j+l},j_1,\cdots,j_l}\|_{C^0} \\
&\leq \sum_{i_1,\cdots,i_{j+l}=1}^{s} \sum_{j_1,\cdots,j_l=1}^{k} \|X_{j_1} \cdots X_{j_l}\phi\|_{C^0} \cdot \|K_{i_1,\cdots,i_{j+l},j_1,\cdots,j_l}\|_{L^1(G)} \\
&\lesssim_{G,j,l} N^j \|\nabla^l \phi\|_{C^0},
\end{aligned}$$

giving (1.12).

To prove (1.16), the first statement

$$\|P_{(\leq N)}\phi\|_{C^l_{1/M}} \lesssim_{G,l} \|\phi\|_{C^l_{1/M}}$$

follows from (1.12). The other statements follow from

$$P_{(N)}\phi = P_{(\leq N)}\phi - P_{(<N)}\phi, \quad P_{(>N)}\phi = \phi - P_{(\leq N)}\phi.$$

We are left to show (1.13), (1.14) and (1.15). We observe that (1.14) and (1.15)

follow from (1.12) and (1.13). Indeed, if $j < l$, then (1.13) says that for dyadic N'

$$\|\nabla^j P_{(N')}\phi\|_{C^0} \lesssim_{G,j,l,\alpha} (N')^{j-l}\|\nabla^l\phi\|_{C^0}, (N')^{j-l-\alpha}\|\nabla^l\phi\|_{\dot{C}^{0,\alpha}}$$

and summing over all dyadic $N' > N$ gives

$$\|\nabla^j P_{(>N)}\phi\|_{C^0} \lesssim_{G,j,l,\alpha} N^{j-l}\|\nabla^l\phi\|_{C^0}, N^{j-l-\alpha}\|\nabla^l\phi\|_{\dot{C}^{0,\alpha}}.$$

If $j = l$, then similarly (1.13) says that for dyadic N'

$$\|\nabla^l P_{(N')}\phi\|_{C^0} \lesssim_{G,l,\alpha} (N')^{-\alpha}\|\nabla^l\phi\|_{\dot{C}^{0,\alpha}}$$

and summing over all dyaidc $N' > N$ gives

$$\|\nabla^l P_{(>N)}\phi\|_{C^0} \lesssim_{G,l,\alpha} N^{-\alpha}\|\nabla^l\phi\|_{\dot{C}^{0,\alpha}}.$$

On the other hand, by (1.12),

$$\|\nabla^l P_{(>N)}\phi\|_{C^0} \leq \|\nabla^l\phi\|_{C^0} + \|\nabla^l P_{(\leq N)}\phi\|_{C^0} \lesssim_{G,l,\alpha} \|\nabla^l\phi\|_{C^0}.$$

Now it remains to show (1.13). Note that, if $j \geq l$, it follows from (1.12) that

$$\begin{aligned} N^{-j}\|\nabla^j P_{(N)}\phi\|_{C^0} &\leq N^{-l}\|\nabla^l P_{(N)}\phi\|_{C^{j-l}_{1/N}} \\ &\leq N^{-l}\|\nabla^l P_{(\leq N)}\phi\|_{C^{j-l}_{1/N}} + N^{-l}\|\nabla^l P_{(\leq N/2)}\phi\|_{C^{j-l}_{1/N}} \\ &\lesssim_{G,j,l} N^{-l}\|\nabla^l\phi\|_{C^0}. \end{aligned}$$

So it remains to verify

$$N^{-j}\|\nabla^j P_{(N)}\phi\|_{C^0} \lesssim_{G,j,l} N^{-l}\|\nabla^l\phi\|_{C^0} \quad \text{for } j < l$$

and

$$N^{-j}\|\nabla^j P_{(N)}\phi\|_{C^0} \lesssim_{G,j,l} N^{-l-\alpha}\|\nabla^l \phi\|_{C^{0,\alpha}}$$

Note that

$$P_{(N)} = P_{(\leq N)} - P_{(\leq N/2)} = \varphi(L/N^2) - \varphi(4L/N^2) = \varphi_1(L/N^2),$$

where φ_0 is defined as

$$\varphi_0(x) = \varphi(x) - \varphi(4x), \quad x \in \mathbb{R}$$

and is a C_c^∞ function supported on $[-4,-1/4]\cup[1/4,4]$. So, for each $n \in \mathbb{N}$, $\varphi_n(x) = \varphi_0(x)/x^n$ is a well-defined C_c^∞ function supported on $[-4,-1/4]\cup[1/4,4]$, with

$$P_{(N)} = \varphi_0(L/N^2) = N^{-2n}L^n\varphi_n(L/N^2).$$

By Theorem 1.1.3, the kernel $K^n_{(N)}$ of $\varphi_n(L/N^2)$ satisfies

$$\|d_G(\cdot, e_G)^\alpha \nabla^I K^n_{(N)}\|_{L^1(G)} \lesssim_{G,\alpha,I,n} N^{|I|-\alpha}.$$

Hence

$$P_{(N)}\phi = N^{-2n}L^n(\phi * K^n_{(N)}).$$

Choosing $n \in \mathbb{N}$ so that $j + 2n > l$, we have

$$N^{-j}\|\nabla^j P_{(N)}\phi\|_{C^0} \lesssim_{G,j,n} N^{-j-2n}\|\nabla^{j+2n}(\phi * K^n_{(N)})\|_{C^0}.$$

By (1.12) applied to the case where the bounded function that acts on the Laplacian is ϕ_n, we have

$$\|\nabla^{j+2n}(\phi * K^n_{(N)})\|_{C^0} \lesssim_{G,j,n,l} N^{j+2n-l}\|\nabla^l \phi\|_{C^0}.$$

Now it remains to verify for all $j, l \geq 0$

$$N^{-j}\|P_{(N)}\phi\|_{C^0} \lesssim_{G,j,l,\alpha} N^{-l-\alpha}\|\nabla^l \phi\|_{\dot{C}^{0,\alpha}}.$$

Choose $n = l$. Since L commutes with $\varphi_l(L/N^2)$, we have

$$P_{(N)} = N^{-2l}L^l\varphi_l(L/N^2) = N^{-2l}\varphi_l(L/N^2)L^l$$

and so for smooth ϕ,

$$P_{(N)}\phi = N^{-2l}((L^l\phi) * K^l_{(N)})$$

Using integration by parts, we move l derivatives to the right of $*$, to obtain

$$P_{(N)}\phi = N^{-2l}\sum_{j_1,\cdots,j_l=1}^{k}(X_{j_1}\cdots X_{j_l}\phi) * K^l_{(N),j_1,\cdots,j_l}.$$

More precisely, we represent each left-invariant vector X_j in terms of right-invariant vectors using (1.21), and we send them over to the right-side (which is possible since the coefficient polynomial of each right-invariant vector commutes with and is annihilated by the right-invariant vector).

Since ϕ_l vanishes near 0, $\phi_l(L/N^2)$ annihilates the function 1, and the total mass of $K^l_{(N)}$ is 0. Again, using the fact that the coefficient polynomial of each right-invariant vector commutes with and is annihilated by the right-invariant vector, it follows that the total mass of $K^l_{(N),j_1,\cdots,j_l}$ is zero. We have the bound

$$\|d(\cdot, e_G)^\alpha K^l_{(N),j_1,\cdots,j_l}(\cdot)\|_{L^1(G)} \lesssim_{G,l,j_1,\cdots,j_l,\alpha} N^{l-\alpha}.$$

Hence,

$$
\begin{aligned}
&\left|(X_{j_1}\cdots X_{j_l}\phi) * K^l_{(N),j_1,\cdots,j_l}(g)\right| \\
&= \left|\int_G (X_{j_1}\cdots X_{j_l}\phi)(gh^{-1}) K^l_{(N),j_1,\cdots,j_l}(h) d\mu(h)\right| \\
&= \left|\int_G \left((X_{j_1}\cdots X_{j_l}\phi)(gh^{-1}) - (X_{j_1}\cdots X_{j_l}\phi)(g)\right) K^l_{(N),j_1,\cdots,j_l}(h) d\mu(h)\right| \\
&\leq \int_G \|\nabla^l \phi\|_{\dot{C}^{0,\alpha}} d(h,0)^\alpha K^l_{(N),j_1,\cdots,j_l}(h) d\mu(h) \\
&\lesssim_{G,l,\alpha} N^{l-\alpha} \|\nabla^l \phi\|_{\dot{C}^{0,\alpha}}
\end{aligned}
$$

where we have used

$$|\phi(gh) - \phi(g)| \leq \|\phi\|_{\dot{C}^{0,\alpha}} d(h,0)^\alpha.$$

This completes the proof.

□

1.2 Nash–Moser Perturbation for a bilinear form

Fix linearly independent $Y_1, \cdots, Y_\ell \in V$. (We will choose these to be either $X_1, \cdots, X_k$ or $X_1^{\text{glb}} \cdots, X_{k^{\text{glb}}}^{\text{glb}}$.) For two given C^1 functions $\phi, \psi : G \to \mathbb{R}^D$, we define the bilinear form $B(\phi, \psi) : G \to \text{Sym}^2(\mathbb{R}^\ell) \subset \mathbb{R}^\ell \otimes \mathbb{R}^\ell$ as

$$B(\phi, \psi) \coloneqq \text{Sym}\left((Y_i\phi \cdot Y_j\psi)_{i,j=1,\cdots,\ell}\right). \tag{1.23}$$

So, for example, we would define

$$B^{loc}(\phi, \psi) \coloneqq \text{Sym}\left((X_i\phi \cdot X_j\psi)_{i,j=1,\cdots,k}\right), \tag{1.24}$$

and

$$B^{glob}(\phi, \psi) \coloneqq \text{Sym}\left((X_i\phi \cdot X_j\psi)_{i,j=1,\cdots,k_1^{glob}}\right). \tag{1.25}$$

Later, when constructing good embeddings of the simply connected nilpotent Lie group G, we will encounter the following situation. Given $\psi : G \to \mathbb{R}^D$ with certain regularity properties, so that ψ "represents" the geometry of G at scale AR and above, where $R > 0$ is a dyadic number, we will need to find a "nontrivial" solution $\phi : G \to \mathbb{R}^D$ to

$$B(\phi, \psi) = 0, \tag{1.26}$$

so that $Y\phi \cdot Y\psi = 0$ for any $Y \in \text{span}\{Y_1, \cdots, Y_\ell\}$. This way, the Pythagorean theorem will tell us that $|Y(\phi+\psi)|^2 = |Y\phi|^2+|Y\psi|^2$, which, coupled with an Assouad-type summation technique, will give us optimal control on the growth of $|Y\psi|^2$ and hence provide us with the optimal distortion rate $O(\varepsilon^{-1/2})$ (note that Assouad [3] achieved this orthogonality and hence the optimal distortion by allowing the ϕ and ψ to take values in different direct sum components of the target space, but thereby losing control on the dimension of the target space). Here, when we say that ϕ is "nontrivial", we mean that $\psi + \phi$ also has the regularity properties of ψ but at scale R instead of AR. Attempts to solve this system (1.26) directly using the Leibniz rule and linear algebra gives less control on the smoothness on ϕ than that on ψ, which is unsuitable for iteration. The solution proposed by [77] was to first find a nontrivial and approximate solution $\tilde{\phi}$ to (1.26), or more precisely a solution to the low-frequency equation

$$B(\tilde{\phi}, P_{(\leq N_0/R)}\psi) = 0. \tag{1.27}$$

This way, we have control on all levels of smoothness of $P_{(\leq N_0/R)}\psi$ (by Theorem 1.1.2 (3)), and hence also on $\tilde{\phi}$; more precisely, Theorem 1.1.2 (3) tells us that bounds on the C_{AR}-norm of ψ gives bounds on the C_A-norm of $P_{(\leq N_0/R)}\psi$. This $\tilde{\phi}$ will be constructed in later sections. Once we have this approximate solution $\tilde{\phi}$, Tao [77] then proposed to use a variant of the Nash–Moser iteration scheme to find small perturbations of $\tilde{\phi}$, which are small enough to preserve the non-triviality of $\tilde{\phi}$, and which allows us to

solve the original equation (1.26).

Our goal in this section is to show that the Nash–Moser iteration scheme of [77] carries on to the general setting of simply connected nilpotent Lie groups without obstruction, while proving a slightly stronger orthogonality statement (1.26) than that of [77]. The rest of this section follows the argument of Section 7 of [77]; we have reproduced the entire argument here to keep track of certain calculations that arise from higher-dimensional matrix operations, as well as to state and verify various estimates in the setting of simply connected nilpotent Lie groups; in particular, due to the lack of dilation structure, we state theorems and prove them for general scales R.

Because the Nash–Moser iteration process produces error terms, we will need to consider a slightly more general setting. Given $\psi : G \to \mathbb{R}^D$ and $F = (F_{ij})_{i,j=1,\cdots,\ell} : G \to \mathrm{Sym}^2(\mathbb{R}^\ell)$, we consider the problem of finding a solution $\phi : G \to \mathbb{R}^D$ to

$$B(\phi, \psi) = F. \tag{1.28}$$

Let $\{Y_{r,i'} : r = 2, \cdots, s, i' = 1, \cdots, \ell_r\}$ be a linearly independent set in

$$\mathrm{span}\{[Y_i, Y_j] : i, j = 1, \cdots, \ell\},$$

such that $\{Y_i : i = 1, \cdots, \ell\} \cup \{Y_{r,i'} : r = 2, \cdots, s, i' = 1, \cdots, \ell_r\}$ is a basis of

$$\mathrm{span}\{Y_i : i = 1, \cdots, \ell\} + \mathrm{span}\{[Y_i, Y_j] : i, j = 1, \cdots, \ell\},$$

while for all r and i', $Y_{r,i} \in \mathrm{span}\{[Y_{i_1}, [Y_{i_2}, \cdots, Y_{i,r}] : i_1, \cdots, i_r = 1, \cdots, \ell\}$. We may have $\ell_r = 0$ for some r. Later on, in the local case we will just take $\{Y_{2,i'} : i' = 1, \cdots, \ell_2\}$ only, while in the global case we will assign to each $Y_{r,i'}$ its maximal possible r value.

One easy way to solve (1.28) is to solve the system

$$\begin{cases} \phi \cdot Y_i \psi = 0, & i = 1, \cdots, \ell, \\ \phi \cdot Y_i Y_j \psi = -F_{ij}, & 1 \le i, j \le \ell, \\ \phi \cdot Y_{r,i'} \psi = 0, & r = 2, \cdots, s,\ i' = 1, \cdots, \ell_r, \end{cases}$$

for then, since $Y_i Y_j - Y_j Y_i \in \operatorname{span}\{Y_1, \cdots, Y_\ell, Y_{2,1}, \cdots, Y_{s,\ell_s}\}$, $1 \le i, j \le \ell$, we have

$$\begin{cases} \phi \cdot Y_i \psi = 0, \\ \phi \cdot Y_i Y_j \psi = -F_{ij}, \end{cases} \qquad i, j = 1, \cdots, \ell,$$

and so

$$Y_i \phi \cdot Y_j \psi = Y_i(\phi \cdot Y_j \psi) - \phi \cdot Y_i Y_j \psi = F_{ij}, \quad i, j = 1, \cdots, \ell.$$

This system is solvable if $\{Y_i \psi\}_{i=1,\cdots,\ell} \cup \{Y_i Y_j \psi\}_{1 \le i \le j \le \ell} \cup \{Y_{2,i'} \psi\}_{i'=1,\cdots,\ell_2}$ are pointwise independent.

More precisely, for each $p \in G$ define the linear map

$$T_\psi(p) : \mathbb{R}^D \to \mathbb{R}^{\ell + \frac{\ell(\ell+1)}{2} + \ell_2 + \cdots + \ell_s}$$

by setting for each $v \in \mathbb{R}^D$

$$T_\psi(p)(v) \coloneqq \Big((v \cdot Y_i \psi(p))_{1 \le i \le \ell},\ (v \cdot Y_i Y_j \psi(p))_{1 \le i \le j \le \ell},\ (v \cdot Y_{r,i'} \psi(p))_{2 \le r \le s,\ 1 \le i' \le \ell_r} \Big).$$

If $\{Y_i \psi\}_{i=1,\cdots,\ell} \cup \{Y_i Y_j \psi\}_{1 \le i \le j \le \ell} \cup \{Y_{r,i'} \psi\}_{r=2,\cdots,s,\ i'=1,\cdots,\ell_r}$ are pointwise independent, i.e., if $T_\psi(p)$ has full rank, or equivalently (by the Cauchy–Binet formula) if

$$\left| \bigwedge_{i=1}^{\ell} Y_i \psi(p) \wedge \bigwedge_{1 \le i \le j \le \ell} Y_i Y_j \psi(p) \wedge \bigwedge_{r=2}^{s} \bigwedge_{i'=1}^{\ell_r} Y_{r,i'} \psi(p) \right| > 0,$$

then we can define the pseudoinverse $T_\psi(p)^{-1} : \mathbb{R}^{\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_s} \to \mathbb{R}^D$ of $T_\psi(p)$ by the formula

$$T_\psi(p)^{-1} \coloneqq T_\psi(p)^*(T_\psi(p)T_\psi(p)^*)^{-1}.$$

(Note that the linear independence condition necessitates that $D \geq \ell + \frac{\ell(\ell+1)}{2} + \ell_2 + \cdots + \ell_s$, so the pseudoinverse is well-defined.) Then for any continuous functions $a_i : G \to \mathbb{R}$, $i = 1, \cdots, \ell$, $b_{ij} : G \to \mathbb{R}$, $1 \leq i \leq j \leq \ell$, $c_{r,i'} : G \to \mathbb{R}$, $r = 2, \cdots, s$, $i' = 1, \cdots, \ell_r$, we have the pointwise identities

$$\begin{aligned}
&T_\psi(p)^{-1}(a_1(p), \cdots, a_\ell(p), b_{11}(p), \cdots, b_{\ell\ell}(p),\\
&\qquad c_{2,1}(p), \cdots, c_{s,\ell_s}(p)) \cdot Y_i\psi(p) = a_i(p), \qquad i = 1, \cdots, \ell,\\
&T_\psi(p)^{-1}(a_1(p), \cdots, a_\ell(p), b_{11}(p), \cdots, b_{\ell\ell}(p),\\
&\qquad c_{2,1}(p), \cdots, c_{s,\ell_s}(p)) \cdot Y_iY_j\psi(p) = b_{ij}(p), \qquad 1 \leq i \leq j \leq \ell,\\
&T_\psi(p)^{-1}(a_1(p), \cdots, a_\ell(p), b_{11}(p), \cdots, b_{\ell\ell}(p),\\
&\qquad c_{2,1}(p), \cdots, c_{s,\ell_s}(p)) \cdot Y_{r,i'}\psi(p) = c_{r,i'}(p), \qquad 2 \leq r \leq s,\ 1 \leq i' \leq \ell_r.
\end{aligned} \tag{1.29}$$

If we define

$$b_{ji} = b_{ij} + \sum_{h=1}^{\ell} \alpha_{i,j,h} a_h + \sum_{r=2}^{s} \sum_{i'=1}^{\ell_r} \beta_{i,j,r,i'} c_{r,i'}, \quad 1 \leq i < j \leq \ell, \tag{1.30}$$

where $\alpha_{i,j,h}$ and $\beta_{i,j,r,i'}$ are the structure constants so that

$$Y_jY_i - Y_iY_j = \sum_{h=1}^{\ell} \alpha_{i,j,h} Y_h + \sum_{r=2}^{s} \sum_{i'=1}^{\ell_r} \beta_{i,j,r,i'} Y_{r,i'}, \quad 1 \leq i \leq j \leq \ell, \tag{1.31}$$

then we can extend the second equation of (1.29) to

$$T_\psi(p)^{-1}(a(p), b(p), c(p)) \cdot Y_iY_j\psi(p) = b_{ij}(p), \quad i, j = 1, \cdots, \ell. \tag{1.32}$$

So, by using the Leibniz rule as above, we have

$$Y_i\Big(T_\psi^{-1}(a,b,c)\Big)\cdot Y_j\psi = Y_i(a_j) - b_{ij}, \quad i,j = 1,\cdots,\ell. \tag{1.33}$$

As a consequence, one has the explicit solution

$$\phi_{\text{explicit}}(p) \coloneqq T_\psi(p)^{-1}(0,-F(p),0) \tag{1.34}$$

to (1.28) (when we are plugging in $F(p)$ to the above expression, we are using a standard identification $\mathrm{Sym}^2(\mathbb{R}^\ell) \simeq \mathbb{R}^{\frac{\ell(\ell+1)}{2}}$).

The problem with this solution to (1.28) is that the solution ϕ_{explicit} constructed in this manner will have two fewer degrees of regularity than ψ, which will be unsuitable for iteration purposes. We will overcome this issue by applying the above procedure to the Littlewood–Paley components of ψ and F.

Proposition 1.2.1 (Perturbation theorem, analog of [77, Proposition 7.1])**.** *Let R be a dyadic number and let M be a real number with*

$$M \geq C_0^{-1}.$$

Let $m^* \geq s+1$ and $\frac{1}{2} < \alpha < 1$. Suppose we are given a $C^{m^*,\alpha}$-map $\psi : G \to \mathbb{R}^D$ with the following regularity properties:

1. (Hölder regularity at scale AR) We have

$$\|\nabla^2\psi\|_{C_{AR}^{m^*-2,\alpha}} \leq \frac{C_0}{AR}. \tag{1.35}$$

2. (nondegenerate first derivatives) For any $p \in G$, we have

$$C_0^{-1}M \leq |Y_i\psi(p)| \leq C_0 M, \quad i = 1,\cdots,\ell, \tag{1.36}$$

3. (locally free embedding) For any $p \in G$, we have

$$\left| \bigwedge_{i=1}^{\ell} Y_i \psi(p) \wedge \bigwedge_{1 \leq i \leq j \leq \ell} Y_i Y_j \psi(p) \wedge \bigwedge_{r=2}^{s} \bigwedge_{i'=1}^{\ell_r} Y_{r,i'} \psi(p) \right| \gtrsim_{C_0} M^{-\ell} (AR)^{-\frac{\ell(\ell+1)}{2} - \sum_{r=2}^{s} (r-1)\ell_r}. \tag{1.37}$$

Let $F : G \to \mathrm{Sym}^2(\mathbb{R}^\ell)$ be a C^{2m^*-1}-function. Let $\tilde{\phi} : G \to \mathbb{R}^D$ be a C^{2m^*}-solution to the low-frequency equation (1.27)*. Then there exists a $C^{m^*,\alpha}$-solution ϕ to* (1.28) *which is a small perturbation of $\tilde{\phi}$:*

$$\|\phi - \tilde{\phi}\|_{C_R^{m^*,\alpha}} \lesssim_{C_0} AR \|F\|_{C_R^{2m^*-1}} + A^{2-m^*-\alpha} \|\tilde{\phi}\|_{C_R^{m^*,\alpha}}, \tag{1.38}$$

and which makes the following cross terms small:

$$\begin{aligned} &\|Y_i \phi \cdot Y_j \psi - Y_i \tilde{\phi} \cdot Y_j P_{(\leq N_0/R)} \psi\|_{C^0} \\ &\qquad \lesssim_{C_0} \|F\|_{C_R^{2m^*-1}} + N_0^{1-m^*-\alpha} A^{1-m^*-\alpha} R^{-1} \|\tilde{\phi}\|_{C_R^{2m^*}}, \quad i, j = 1, \cdots, \ell. \end{aligned} \tag{1.39}$$

Here, we treat α as a universal constant (we may take $\alpha = \frac{2}{3}$), and we allow the constant C_0 to depend on m^*. This will not contradict our hierarchy of constants, as m^* will be later chosen to depend on G (more precisely, it will equal $s^2 + s + 1$).

Remark 1.2.2. *1. The actual bound needed for m^* is that it is at least the largest number $1 \leq r \leq s$ such that $\ell_r > 0$.*

2. The bilinear form used in [77] was the simpler version

$$\tilde{B}(\phi, \psi) = (Y_1 \phi \cdot Y_1 \psi, \cdots, Y_\ell \phi \cdot Y_\ell \psi),$$

and the corresponding Nash-Moser iteration scheme only established the weaker orthogonality $\tilde{B}(\phi, \psi) = 0$, while still being able to establish the estimate (1.39)*. We may create a version of Proposition 1.2.1 for $\tilde{B}$ by assuming the weaker freeness*

property for (1.37)*:*

$$\left|\bigwedge_{i=1}^{\ell} Y_i\psi(p) \wedge \bigwedge_{i=1}^{\ell} Y_iY_i\psi(p)\right| \gtrsim_{C_0} \prod_{i=1}^{\ell} |Y_i\psi(p)| \prod_{i=1}^{\ell} |Y_iY_i\psi(p)|.$$

The proof of this weaker Proposition would be not so different from the proof of Proposition 1.2.1 given below, where we would modify the pseudoinverse T_ψ^{-1} *and its usage in the obvious way. The statement* (1.39) *becomes slightly harder to prove, but one may directly implement the methods of [77].*

3. *One may imagine strengthening the theorem to obtain the stronger full orthogonality statement*

$$Y_i\phi \cdot Y_j\psi = 0, \quad i,j = 1 \cdots, \ell,$$

but attempts to modify the Nash–Moser iteration scheme to accommodate this difference cause the resulting infinite series to diverge (more specifically, we are then forced to place derivatives on $P_{(N/R)}\psi$ *in* (1.47)*, which we must avoid in order to make the defining series converge). The best one can achieve with the tools outlined in this section is* (1.39)*. Nevertheless, we will be able to choose* $\tilde{\phi}$ *with*

$$Y_i\tilde{\phi} \cdot Y_jP_{(\leq N_0/R)}\psi = 0, \quad i,j = 1 \cdots, \ell,$$

which, along with (1.39)*, establishes that* $Y_i\phi \cdot Y_j\psi$ *is sufficiently small.*

Proof. It will suffice to find a function ϕ with the stated bounds solving the approximate equation

$$\|B(\phi,\psi) - F\|_{C_R^{2m^*-1}} \leq A^{2-m^*}\|F\|_{C_R^{2m^*-1}} + A^{s-2m^*}R^{-1}\|\tilde{\phi}\|_{C_R^{2m^*}} \tag{1.40}$$

rather than the precise equation (1.28), while satisfying (1.38) and (1.39):

$$\|\phi - \tilde{\phi}\|_{C_R^{m^*,\alpha}} \lesssim_{C_0} AR\|F\|_{C_R^{2m^*-1}} + A^{s-m^*-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}}, \tag{1.41}$$

and

$$\begin{aligned}\|Y_i\phi \cdot Y_j\psi - Y_i\tilde{\phi} \cdot Y_j P_{(\leq N_0/R)}\psi\|_{C^0} \\ \lesssim_{C_0} \|F\|_{C_R^{2m^*-1}} + N_0^{1-m^*-\alpha}A^{1-m^*-\alpha}R^{-1}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}}, \quad i,j = 1,\cdots,\ell.\end{aligned} \tag{1.42}$$

One can then iteratively replace $(\tilde{\phi}, F)$ by the error term $(0, F - B(\phi,\psi))$ and sum the resulting solutions to obtain an exact solution to (1.28). This is possible due to the linearity of this equation in ϕ.

We will first construct a low-frequency solution $\phi_{(\leq N_0/R)}$ to the low-frequency equation

$$B(\phi_{(\leq N_0/R)}, P_{(\leq N_0/R)}\psi) = P_{(\leq N_0/R)}F$$

and then, given $\phi_{(\leq N/2R)}$ for dyadic $N > N_0$ by induction, we will add higher frequency components $\phi_{(N/R)}$ to obtain $\phi_{(\leq N/R)} := \phi_{(\leq N/2R)} + \phi_{(N/R)}$, which approximately solves the higher-frequency equation

$$B(\phi_{(\leq N/R)}, P_{(\leq N/R)}\psi) \approx P_{(\leq N/R)}F.$$

More specifically, the construction goes as follows. We first construct the low-frequency component as

$$\phi_{(\leq N_0/R)} := \tilde{\phi} + T^{-1}_{P_{(\leq N_0/R)}\psi}(0, -P_{(\leq N_0/R)}F, 0). \tag{1.43}$$

This form was chosen so that, from (1.33), one has

$$Y_i(\phi_{(\leq N_0/R)} - \tilde{\phi}) \cdot Y_j P_{(\leq N_0/R)}\psi = P_{(\leq N_0/R)}F_{ij}, \quad i,j = 1,\cdots,\ell, \tag{1.44}$$

and from (1.27) and the bilinearity of B, one has

$$B(\phi_{(\leq N_0/R)}, P_{(\leq N_0/R)}\psi) = P_{(\leq N_0/R)}F. \tag{1.45}$$

Next, for every dyadic $N > N_0$ we recursively define the higher-frequency component $\phi_{(N/R)}$ by the formula

$$\phi_{(N/R)} \coloneqq T^{-1}_{P_{(\leq N/R)}\psi}((a^i_N)_{1\leq i\leq \ell}, (b^{ij}_N)_{1\leq i\leq j\leq \ell}, (c^{r,i'}_N)_{2\leq r\leq s,\ 1\leq i'\leq \ell_r}) \tag{1.46}$$

where

$$\begin{cases} a^i_N \coloneqq -(Y_i P_{(\leq N/R)}\phi_{(<N/R)}) \cdot P_{(N/R)}\psi, & i = 1,\cdots,\ell, \\ b^{ij}_N \coloneqq -(Y_i Y_j P_{(\leq N/R)}\phi_{(<N/R)}) \cdot P_{(N/R)}\psi - P_{(N/R)}F_{ij}, & 1 \leq i \leq j \leq \ell, \\ c^{r,i'}_N \coloneqq -(Y_{r,i'} P_{(\leq N/R)}\phi_{(<N/R)}) \cdot P_{(N/R)}\psi, & 2 \leq r \leq s,\ 1 \leq i' \leq \ell_2, \end{cases} \tag{1.47}$$

and

$$\phi_{(<N/R)} \coloneqq \phi_{(\leq N_0/R)} + \sum_{\substack{N_0 < M < N \\ M \text{ dyadic}}} \phi_{(M/R)}.$$

We will also define

$$\phi_{(\leq N/R)} \coloneqq \phi_{(\leq N_0/R)} + \sum_{\substack{N_0 < M \leq N \\ M \text{ dyadic}}} \phi_{(M/R)}.$$

Note that in the definition of $\phi_{(N/R)}$, no derivatives are placed on $P_{(N/R)}\psi$ and there is some mollification of the $\phi_{(<N/R)}$ term, in order to avoid the loss of derivatives

problem. This form of $\phi_{(N/R)}$ was chosen so that

$$B(\phi_{(N/R)}, P_{(\leq N/R)}\psi) + B(P_{(\leq N/R)}\phi_{(<N/R)}, P_{(N/R)}\psi) = P_{(N/R)}F. \tag{1.48}$$

Indeed, by (1.30), (1.31), and (1.32), we have

$$b_N^{ij} = -(Y_iY_jP_{(\leq N/R)}\phi_{(<N/R)}) \cdot P_{(N/R)}\psi - P_{(N/R)}F_{ij}, \quad i,j = 1, \cdots, \ell,$$

so we can compute, for $i,j = 1, \cdots, \ell$,

$$\begin{aligned} &Y_i\phi_{(N/R)} \cdot Y_jP_{(\leq N/R)}\psi + Y_iP_{(\leq N/R)}\phi_{(<N/R)} \cdot Y_jP_{(N/R)}\psi \\ &\overset{(1.33)}{=} \left(Y_ia_N^j - b_N^{ij}\right) + Y_iP_{(\leq N/R)}\phi_{(<N/R)} \cdot Y_jP_{(N/R)}\psi \\ &= P_{(N/R)}F_{ij} - Y_jP_{(\leq N/R)}\phi_{(<N/R)} \cdot Y_iP_{(N/R)}\psi + Y_iP_{(\leq N/R)}\phi_{(<N/R)} \cdot Y_jP_{(N/R)}\psi. \end{aligned} \tag{1.49}$$

Symmetrizing in i and j now gives (1.48).

We have the following control on the regularity of $T^{-1}_{P_{(\leq N/R)}\psi}$. Here, for an $m \times n$ matrix A, $1 \leq i \leq j \leq n$ we denote by $[A]^{i,j}$ the submatrix consisting of the ith to jth columns of A.

Lemma 1.2.3 (regularity of the pseudoinverse; analog of [77, Lemma 7.2])**.** *For any $N \geq N_0$ and $l \geq 0$, one has the following estimate for the first ℓ columns:*

$$\left\|[T^{-1}_{P_{(\leq N/R)}\psi}]^{1,\ell}\right\|_{C^{m^*-2}_{AR}} \lesssim_{C_0} M^{-1}, \quad \left\|\nabla^{m^*-2}[T^{-1}_{P_{(\leq N/R)}\psi}]^{1,\ell}\right\|_{C^l_{R/N}} \lesssim_{C_0,l} M^{-1}(AR)^{2-m^*},$$

the following estimate for the next $\frac{\ell(\ell+1)}{2}$ columns:

$$\begin{aligned} \left\|[T^{-1}_{P_{(\leq N/R)}\psi}]^{\ell+1,\ell+\frac{\ell(\ell+1)}{2}}\right\|_{C^{m^*-2}_{AR}} &\lesssim_{C_0} AR, \\ \left\|\nabla^{m^*-2}[T^{-1}_{P_{(\leq N/R)}\psi}]^{\ell+1,\ell+\frac{\ell(\ell+1)}{2}}\right\|_{C^l_{R/N}} &\lesssim_{C_0,l} (AR)^{3-m^*}. \end{aligned}$$

and the following estimate for the ℓ_r columns amongst the rest of the rows:

$$\left\| [T^{-1}_{P_{(\leq N/R)}\psi}]^{\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_{r-1}+1,\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_r} \right\|_{C^{m^*-2}_{AR}} \lesssim_{C_0} (AR)^{r-1},$$

and

$$\left\| \nabla^{m^*-2}[T^{-1}_{P_{(\leq N/R)}\psi}]^{\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_{r-1}+1,\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_r} \right\|_{C^{l}_{R/N}} \lesssim_{C_0,l} (AR)^{r+1-m^*}.$$

Proof. Abbreviating $T = T_{P_{(\leq N/R)}\psi}$, recall the definition of the pseudoinverse

$$T^{-1} = \frac{1}{\det(TT^*)} T^* \mathrm{adj}(TT^*)$$

where $\mathrm{adj}(A)$ denotes the adjugate matrix of A. We then need to show the bounds

$$\left\| \nabla^l \left[\frac{1}{\det(TT^*)} T^* \mathrm{adj}(TT^*) \right]^{1,\ell} \right\|_{C^0} \lesssim_{C_0} M^{-1} B_l,$$

$$\left\| \nabla^l \left[\frac{1}{\det(TT^*)} T^* \mathrm{adj}(TT^*) \right] \ell+1, \ell+\frac{\ell(\ell+1)}{2} \right\|_{C^0} \lesssim_{C_0} AR B_l,$$

$$\left\| \nabla^l \left[\frac{1}{\det(TT^*)} T^* \mathrm{adj}(TT^*) \right]^{\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_{r-1}+1,\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_r} \right\|_{C^0} \lesssim_{C_0} (AR)^{r-2} B_l,$$

for $l \geq 0$, where B_l is the log-convex sequence

$$B_l \coloneqq \begin{cases} (AR)^{-l}, & 0 \leq l \leq m^*-2, \\ (N/R)^{l-m^*+2}(AR)^{2-m^*}, & l > m^*-2. \end{cases}$$

From (1.35), (1.36), and Theorem 1.1.2(3), we have

$$\begin{aligned}
|\nabla^l Y_i P_{(\le N/R)}\psi| &\lesssim_{C_0,l} MB_l, && i=1,\cdots,\ell, \quad l\ge 0,\\
|\nabla^l Y_i Y_j P_{(\le N/R)}\psi| &\lesssim_{C_0,l} (AR)^{-1}B_l, && i,j=1,\cdots,\ell, \quad l\ge 0,\\
|\nabla^l Y_{r,i'} P_{(\le N/R)}\psi| &\lesssim_{C_0,l} (AR)^{-r+1}B_l, && r=2,\cdots,s,\ i'=1,\cdots,\ell_2, \quad l\ge 0.
\end{aligned}$$

Thus, viewing T as a $\left(\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_s\right)\times D$ matrix, for any $W_l\in\nabla^l$ the first ℓ rows of W_lT have norm $O_{C_0}(MB_l)$, the $\frac{\ell(\ell+1)}{2}$ rows below that have norm $O_{C_0}((AR)^{-1}B_l)$, and below that we have $\ell_2,\cdots,\ell_s$ rows, where ℓ_r rows have norm $O_{C_0}((AR)^{-r+1}B_l)$. By the product rule, and noting that $B_lB_{l'}\le B_{l+l'}$ for all $l,l'\ge 0$ we conclude that the $\left(\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_s\right)\times\left(\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_s\right)$ matrix $W_l(TT^*)$ has top left $\ell\times\ell$ block consisting of elements of magnitude $O_{C_0}(M^2B_l)$, the $\ell\times\left(\frac{\ell(\ell+1)}{2}\right)$ block right to that and the $\left(\frac{\ell(\ell+1)}{2}\right)\times\ell$ block below that consisting of elements of magnitude $O_{C_0}(M(AR)^{-1}B_l)$, the$\left(\frac{\ell(\ell+1)}{2}\right)\times\left(\frac{\ell(\ell+1)}{2}\right)$ block, below and right, respectively, of the aforementioned blocks, consisting of elements of magnitude $O_{C_0}((AR)^{-2}B_l)$. The top $\ell\times\ell_r$ block and leftmost $\ell_r\times\ell$ block consist of elements of magnitude $O_{C_0}(M(AR)^{-r+1}B_l)$, the $\left(\frac{\ell(\ell+1)}{2}\right)\times\ell_r$ block and $\ell_r\times\left(\frac{\ell(\ell+1)}{2}\right)$ block consist of elements of magnitude $O_{C_0}((AR)^{-2}B_l)$, and the other $\ell_{r_1}\times\ell_{r_2}$ blocks consist of elements of magnitude $O_{C_0}((AR)^{-r_1-r_2+2}B_l)$.

By the product rule and cofactor expansion, $W_l\mathrm{adj}(TT^*)$ then has top left $\ell\times\ell$ block of norm $O_{C_0}(M^{2\ell-2}(AR)^{-\ell(\ell+1)-\sum_{r=2}^s 2(r-1)\ell_r}B_l)$, the $\ell\times\frac{\ell(\ell+1)}{2}$ block and $\frac{\ell(\ell+1)}{2}\times\ell$ block of norm $O_{C_0}(M^{2\ell-1}(AR)^{-\ell(\ell+1)-\sum_{r=2}^s 2(r-1)\ell_r+1}B_l)$, the $\frac{\ell(\ell+1)}{2}\times\frac{\ell(\ell+1)}{2}$ block of norm $O_{C_0}(M^{2\ell}(AR)^{-\ell(\ell+1)-\sum_{r=2}^s 2(r-1)\ell_r+2}B_l)$, the $\ell\times\ell_{r_1}$ and $\ell_{r_1}\times\ell$ blocks of norm $O_{C_0}(M^{2\ell-1}(AR)^{-\ell(\ell+1)-\sum_{r=2}^s 2(r-1)\ell_r+(r_1-1)}B_l)$, the $\frac{\ell(\ell+1)}{2}\times\ell_{r_1}$ and $\ell_{r_1}\times\frac{\ell(\ell+1)}{2}$ blocks of norm $O_{C_0}(M^{2\ell}(AR)^{-\ell(\ell+1)-\sum_{r=2}^s 2(r-1)\ell_r+r_1}B_l)$, the remaining $\ell_{r_1}\times\ell_{r_2}$ blocks of norm $O_{C_0}(M^{2\ell}(AR)^{-\ell(\ell+1)-\sum_{r=2}^s 2(r-1)\ell_r+r_1+r_2-2}B_l)$.

Again, by the product rule, the $D \times \left(\ell + \frac{\ell(\ell+1)}{2} + \ell_2 + \cdots + \ell_s\right)$ matrix

$$W_l(T^* \mathrm{adj}(TT^*))$$

has first ℓ columns of norm

$$O_{C_0}(M^{2\ell-1}(AR)^{-\ell(\ell+1)-\sum_{r=2}^s 2(r-1)\ell_r} B_l),$$

the next $\frac{\ell(\ell+1)}{2}$ columns of norm

$$O_{C_0}(M^{2\ell}(AR)^{-\ell(\ell+1)-\sum_{r=2}^s 2(r-1)\ell_r+1} B_l),$$

and the remaining ℓ_{r_1} columns each of norm

$$O_{C_0}(M^{2\ell}(AR)^{-\ell(\ell+1)-\sum_{r=2}^s 2(r-1)\ell_r+(r_1-1)} B_l)$$

(due to the presence of R, we cannot say the first ℓ columns are the greatest).

Similarly, $\nabla^l(\det(TT^*))$ has magnitude

$$O_{C_0}(M^{2\ell}(AR)^{-\ell(\ell+1)-\sum_{r=2}^s 2(r-1)\ell_r} B_l).$$

Meanwhile, from (1.36), (1.37), (1.35), and using (1.13) to approximate $P_{(\leq N/R)}\psi$ by ψ up to negligible error (note that $N \geq N_0$ and our hierarchy of constants, namely that N_0 is chosen after C_0), we easily obtain the wedge product lower bound

$$\left| \bigwedge_{i=1}^{\ell} Y_i P_{(\leq N/R)}\psi(p) \wedge \bigwedge_{1\leq i\leq j\leq \ell} Y_i Y_j P_{(\leq N/R)}\psi(p) \wedge \bigwedge_{r=2}^{s} \bigwedge_{i'=1}^{\ell_r} Y_{r,i'} P_{(\leq N/R)}\psi(p) \right| \tag{1.50}$$
$$\gtrsim_{C_0} (AR)^{-\frac{\ell(\ell+1)}{2}-\sum_{r=2}^s (r-1)\ell_r} M^{\ell}.$$

From this and the Cauchy–Binet formula we have the matching lower bound

$$\det(TT^*) \gtrsim_{C_0} M^{2\ell}(AR)^{-\ell(\ell+1)-\sum_{r=2}^{s} 2(r-1)\ell_r}$$

for the determinant. Hence, by the quotient rule, the derivatives $\nabla^l(\det(TT^*)^{-1})$ have magnitude $O_{C_0}(M^{-2\ell}A^{\ell(\ell+1)+\sum_{r=2}^{s} 2(r-1)\ell_r}B_l)$. The claim now follows from the product rule. □

The rest of the proof of Proposition 1.2.1 follows mostly as in [77], except for the proof of (1.39). We have reproduced the argument for completeness.

First, we would need the smoothness of our proposed solution. We begin with the smoothness of the low-frequency component $\phi_{(\leq N_0/R)}$. From the above Lemma we have the estimate for all $l \geq 0$

$$\begin{aligned} \left\| [T^{-1}_{P_{(\leq N/R)}\psi}]^{1,\ell} \right\|_{C^l_R} &\lesssim_{C_0,l} M^{-1}, \\ \left\| [T^{-1}_{P_{(\leq N/R)}\psi}]^{\ell+1,\ell+\frac{\ell(\ell+1)}{2}} \right\|_{C^l_R} &\lesssim_{C_0,l} AR, \\ \left\| [T^{-1}_{P_{(\leq N/R)}\psi}]^{\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_{r-1}+1,\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_r} \right\|_{C^l_R} &\lesssim_{C_0,l} (AR)^{r-1}, \end{aligned} \tag{1.51}$$

(this uses our hierarchy of constants, namely that A is chosen after N_0), while from (1.12) we have

$$\|P_{(\leq N_0/R)}F\|_{C^{2m^*-1}_R} \lesssim_G \|F\|_{C^{2m^*-1}_R}, \quad \|P_{(\leq N_0/R)}F\|_{C^{2m^*}_R} \lesssim_G N_0\|F\|_{C^{2m^*-1}_R}$$

and thus from (1.43),

$$\begin{aligned} \|\phi_{(\leq N_0/R)} - \tilde{\phi}\|_{C^{2m^*-1}_R} &\lesssim \left\| [T^{-1}_{P_{(\leq N/R)}\psi}]^{\ell+1,\ell+\frac{\ell(\ell+1)}{2}} \right\|_{C^{2m^*-1}_R} \|P_{(\leq N_0/R)}F\|_{C^{2m^*-1}_R} \\ &\lesssim_{C_0} AR\|F\|_{C^{2m^*-1}_R}, \end{aligned} \tag{1.52}$$

while, similarly,

$$\|\phi_{(\leq N_0/R)} - \tilde{\phi}\|_{C_R^{2m^*}} \lesssim_{C_0} AN_0R\|F\|_{C_R^{2m^*-1}}. \tag{1.53}$$

Next, we establish the smoothness of the higher-frequency components $\phi_{(N/R)}$. From (1.13) we have, for $N > N_0$ dyadic and $l \geq 0$,

$$R^m\|\nabla^m P_{(\leq N/R)}\phi_{(<N/R)}\|_{C^l_{R/N}} \lesssim_{G,l} R^m\|\nabla^m\phi_{(<N/R)}\|_{C^0} \leq \|\phi_{(<N/R)}\|_{C^m_R}, \quad m \geq 1.$$

This implies in particular that, for $l \geq 0$,

$$R\|Y_i P_{(\leq N/R)}\phi_{(<N/R)}\|_{C^l_{R/N}} \lesssim_{G,l} \|\phi_{(<N/R)}\|_{C^s_R},$$

$$R^2\|Y_iY_j P_{(\leq N/R)}\phi_{(<N/R)}\|_{C^l_{R/N}} \lesssim_{G,l} \|\phi_{(<N/R)}\|_{C^s_R},$$

$$R^r\|Y_{r,i'} P_{(\leq N/R)}\phi_{(<N/R)}\|_{C^l_{R/N}} \lesssim_{G,l} \|\phi_{(<N/R)}\|_{C^s_R}.$$

Again, from (1.13) and (1.35), we also have the estimates, for $l \geq 0$,

$$\begin{aligned}
&\|P_{(N/R)}\psi\|_{C^l_{R/N}} \\
&\lesssim_G (N/R)^{-m^*-\alpha}\|\nabla^{m^*}\psi\|_{\dot{C}^{0,\alpha}} \lesssim (N/R)^{-m^*-\alpha}(AR)^{2-m^*-\alpha}\|\nabla^2\psi\|_{C^{m^*-2,\alpha}_{AR}} \\
&\lesssim_{C_0} RN^{-m^*-\alpha}A^{1-m^*-\alpha}
\end{aligned} \tag{1.54}$$

and

$$\|P_{(N/R)}F\|_{C^l_{R/N}} \lesssim_G (N/R)^{-2m^*+1}\|\nabla^{2m^*-1}F\|_{C^0} \lesssim N^{-2m^*+1}\|F\|_{C_R^{2m^*-1}}.$$

Finally, from Lemma 1.2.3 one has, for $l \geq 0$,

$$\begin{aligned}
\left\| [T^{-1}_{P_{(\leq N/R)}\psi}]^{1,\ell} \right\|_{C^l_{R/N}} &\lesssim_{C_0,l} M^{-1}, \\
\left\| [T^{-1}_{P_{(\leq N/R)}\psi}]^{\ell+1,\ell+\frac{\ell(\ell+1)}{2}} \right\|_{C^l_{R/N}} &\lesssim_{C_0,l} AR, \\
\left\| [T^{-1}_{P_{(\leq N/R)}\psi}]^{\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_{r-1}+1,\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_r} \right\|_{C^l_{R/N}} &\lesssim_{C_0,l} (AR)^{r-1},
\end{aligned}$$

since $A^{1-j}R^{-j} \lesssim A(N/R)^j$ for $0 \leq j \leq m^* - 2$ and $A(AR)^{2-m^*}(N/R)^{j-m^*+2} \lesssim A(N/R)^j$ for $m^* - 2 \leq j \leq l$. Inserting the above estimates into (1.46), we conclude that

$$\begin{aligned}
&\|\phi_{(N/R)}\|_{C^l_{R/N}} \\
&\lesssim \left\| [T^{-1}_{P_{(\leq N/R)}\psi}]^{1,\ell} \right\|_{C^l_{R/N}} \sum_{i=1}^{\ell} \|Y_i P_{(\leq N/R)}\phi_{(<N/R)}\|_{C^l_{R/N}} \|P_{(N/R)}\psi\|_{C^l_{R/N}} \\
&\quad + \left\| [T^{-1}_{P_{(\leq N/R)}\psi}]^{\ell+1,\ell+\frac{\ell(\ell+1)}{2}} \right\|_{C^l_{R/N}} \\
&\qquad \cdot \left(\sum_{i,j=1}^{\ell} \|Y_i Y_i P_{(\leq N/R)}\phi_{(<N/R)}\|_{C^l_{R/N}} \|P_{(N/R)}\psi\|_{C^l_{R/N}} + \|P_{(N/R)}F\|_{C^l_{R/N}} \right) \\
&\quad + \sum_{r=2}^{s} \left\| [T^{-1}_{P_{(\leq N/R)}\psi}]^{\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_{r-1}+1,\ell+\frac{\ell(\ell+1)}{2}+\ell_2+\cdots+\ell_r} \right\|_{C^l_{R/N}} \\
&\qquad\qquad \cdot \sum_{i'=1}^{\ell_r} \|Y_{r,i'} P_{(\leq N/R)}\phi_{(<N/R)}\|_{C^l_{R/N}} \|P_{(N/R)}\psi\|_{C^l_{R/N}} \qquad (1.55) \\
&\lesssim_{C_0,l} M^{-1} \cdot R^{-1} \left\|\phi_{(N/R)}\right\|_{C^s_R} \cdot RN^{-m^*-\alpha}A^{1-m^*-\alpha} \\
&\quad + AR \cdot \left(R^{-2} \left\|\phi_{(N/R)}\right\|_{C^s_R} \cdot RN^{-m^*-\alpha}A^{1-m^*-\alpha} + N^{-2m^*+1}\|F\|_{C^{2m^*-1}_R} \right) \\
&\quad + \sum_{r=2}^{s} (AR)^{r-1} \cdot R^{-r} \left\|\phi_{(N/R)}\right\|_{C^s_R} \cdot RN^{-m^*-\alpha}A^{1-m^*-\alpha} \\
&\lesssim_{C_0} A^{s-m^*-\alpha}N^{-m^*-\alpha}\|\phi_{(<N/R)}\|_{C^s_R} + AN^{-2m^*+1}R\|F\|_{C^{2m^*-1}_R}
\end{aligned}$$

and so

$$\|\phi_{(N/R)}\|_{C_R^{m^*}} \lesssim_{C_0} A^{s-m^*-\alpha}N^{-\alpha}\|\phi_{(<N/R)}\|_{C_R^s} + AN^{-m^*+1}R\|F\|_{C_R^{2m^*-1}}. \tag{1.56}$$

By the triangle inequality we thus have

$$\begin{aligned}\|\phi_{(\leq N/R)} - \tilde{\phi}\|_{C_R^{m^*}} &\leq \|\phi_{(<N/R)} - \tilde{\phi}\|_{C_R^{m^*}} + \|\phi_{(N/R)}\|_{C_R^{m^*}} \\ &\leq (1 + O_{C_0}(A^{s-m^*-\alpha}N^{-\alpha}))\|\phi_{(<N/R)} - \tilde{\phi}\|_{C_R^{m^*}} \\ &\quad + O_{C_0}(A^{s-m^*-\alpha}N^{-\alpha})\|\tilde{\phi}\|_{C_R^{m^*,\alpha}} + O_{C_0}(AN^{-m^*+1}R)\|F\|_{C_R^{2m^*-1}}.\end{aligned}$$

One can easily see by induction, with base case (1.52), that

$$\|\phi_{(\leq N/R)} - \tilde{\phi}\|_{C_R^{m^*}} \lesssim_{C_0} AR\|F\|_{C_R^{2m^*-1}} + A^{s-m^*-\alpha}N_0^{-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}}, \quad N \geq N_0,$$

and so by the triangle inequality, and noting that $m^* \geq s$, we have

$$\|\phi_{(\leq N/R)}\|_{C_R^{m^*}} \lesssim_{C_0} AR\|F\|_{C_R^{2m^*-1}} + \|\tilde{\phi}\|_{C_R^{m^*,\alpha}}, \quad N \geq N_0. \tag{1.57}$$

Inserting this back into (1.55) and (1.56), and noting again that $m^* \geq s \geq 2$, we obtain, for $l \geq 0$,

$$\begin{aligned}\|\phi_{(N/R)}\|_{C^l_{R/N}} &\lesssim_{C_0} A^{s-m^*-\alpha}N^{-m^*-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}} \\ &\quad + (A^{s+1-m^*-\alpha}N^{-m^*-\alpha}R + AN^{-2m^*+1}R)\|F\|_{C_R^{2m^*-1}} \\ &\leq A^{s-m^*-\alpha}N^{-m^*-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}} + AN^{-m^*-\alpha}R\|F\|_{C_R^{2m^*-1}}\end{aligned} \tag{1.58}$$

and

$$\|\phi_{(N/R)}\|_{C_R^{m^*}} \leq N^{m^*}\|\phi_{(N/R)}\|_{C_{R/N}^{m^*+1}} \lesssim_{C_0} A^{s-m^*-\alpha}N^{-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}} + AN^{-\alpha}R\|F\|_{C_R^{2m^*-1}}. \tag{1.59}$$

We thus conclude that the sum

$$\phi := \phi_{(\leq N_0)} + \sum_{N>N_0} \phi_{(N)}$$

converges in the C^{m^*} norm (and consequently also in the C^s norm, as $m^* \geq s$).

We now prove (1.41). From (1.52), it is enough to show

$$\left\| \sum_{N>N_0} \phi_{(N/R)} \right\|_{C_R^{m^*,\alpha}} \lesssim_{C_0} AR\|F\|_{C_R^{2m^*-1}} + A^{s-m^*-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}}.$$

As noted above, from (1.59) we have

$$\begin{aligned}\left\| \sum_{N>N_0} \phi_{(N/R)} \right\|_{C_R^{m^*}} &\leq \sum_{N>N_0} \left\|\phi_{(N/R)}\right\|_{C_R^{m^*}} \\ &\lesssim_{C_0} AN_0^{-\alpha}R\|F\|_{C_R^{2m^*-1}} + A^{s-m^*-\alpha}N_0^{-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}},\end{aligned}$$

so it remains to show Hölder regularity:

$$\begin{aligned}&\left|\nabla^{m^*} \sum_{N>N_0} \phi_{(N/R)}(p) - \nabla^{m^*} \sum_{N>N_0} \phi_{(N/R)}(q)\right| \\ &\lesssim_{C_0} R^{-m^*-\alpha}(AR\|F\|_{C_R^{2m^*-1}} + A^{s-m^*-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}})d(p,q)^\alpha\end{aligned} \tag{1.60}$$

for any $p, q \in G$. By the triangle inequality, it is enough to show

$$\begin{aligned}&\sum_{N>N_0} \left|\nabla^{m^*}\phi_{(N/R)}(p) - \nabla^{m^*}\phi_{(N/R)}(q)\right| \\ &\lesssim_{C_0} R^{-m^*-\alpha}(AR\|F\|_{C_R^{2m^*-1}} + A^{s-m^*-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}})d(p,q)^\alpha\end{aligned}$$

On one hand, we may bound

$$\begin{aligned}&|\nabla^{m^*}\phi_{(N/R)}(p) - \nabla^{m^*}\phi_{(N/R)}(q)| \\ &\lesssim \|\nabla^{m^*}\phi_{(N/R)}\|_{C^0} \lesssim R^{-m^*}\|\phi_{(N/R)}\|_{C_R^{m^*}} \\ &\overset{(1.59)}{\lesssim}_{C_0} R^{-m^*}\left(A^{s-m^*-\alpha}N^{-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}} + AN^{-\alpha}R\|F\|_{C_R^{2m^*-1}}\right) \\ &= R^{-m^*}(A^{s-m^*-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}} + AR\|F\|_{R^{2m^*-1}})N^{-\alpha}.\end{aligned}$$

On the other hand, one has

$$\begin{aligned}&|\nabla^{m^*}\phi_{(N/R)}(p) - \nabla^{m^*}\phi_{(N/R)}(q)| \\ &\lesssim \|\nabla^{m^*+1}\phi_{(N/R)}\|_{C^0} d(p,q) \lesssim (N/R)^{m^*+1}\|\phi_{(N/R)}\|_{C_{R/N}^{m^*+1}} d(p,q) \\ &\overset{(1.58)}{\lesssim}_{C_0} R^{-m^*}(A^{s-m^*-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}} + AR\|F\|_{C_R^{2m^*-1}})N^{-\alpha}(Nd(p,q)/R)\end{aligned}$$

Thus, the left-hand side of (1.60) is bounded by

$$\lesssim_{C_0} R^{-m^*}(A^{s-m^*-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}} + AR\|F\|_{C_R^{2m^*-1}}) \sum_N N^{-\alpha}\min(1, Nd(p,q)/R)$$

and the claim (1.41) follows by summing the double-ended geometric series

$$\sum_N N^{-\alpha}\min(1, Nd(p,q)/R)$$

using the hypothesis $0 < \alpha < 1$.

Now we prove (1.40). As $\phi_{(\leq N/R)}$ converges in C^s to ϕ as $N \to \infty$, and $P_{(\leq N/R)}\psi$ converges in C^s to ψ, we may write $B(\phi, \psi)$ as the uniform limit of

$$B(\phi_{(\leq N/R)}, P_{(\leq N/R)}\psi).$$

Using (1.45) and (1.48), we have the telescoping sum

$$\begin{aligned}
&B(\phi, \psi) \\
&= B(\phi_{(\leq N_0/R)}, P_{(\leq N_0/R)}\psi) + \sum_{N>N_0} (B(\phi_{(\leq N/R)}, P_{(\leq N/R)}\psi) - B(\phi_{(<N/R)}, P_{(<N/R)}\psi)) \\
&= P_{(\leq N_0/R)}F + \sum_{N>N_0} (B(\phi_{(N/R)}, P_{(\leq N/R)}\psi) + B(\phi_{(<N/R)}, P_{(N/R)}\psi)) \\
&= P_{(\leq N_0/R)}F + \sum_{N>N_0} (P_{(N/R)}F + B(P_{(>N/R)}\phi_{(<N/R)}, P_{(N/R)}\psi)) \\
&= F + \sum_{N>N_0} B(P_{(>N/R)}\phi_{(<N/R)}, P_{(N/R)}\psi),
\end{aligned}$$

or

$$B(\phi, \psi) - F = \sum_{N>N_0} B(P_{(>N/R)}\phi_{(<N/R)}, P_{(N/R)}\psi).$$

Each of the terms in the right-hand side, being a "high-high paraproduct" of $\nabla\psi$ and $\nabla\phi$, has much higher regularity (C^{2m^*-1}) than either $\nabla\psi$ or $\nabla\phi$ (C^{m^*-1}). Indeed, by the triangle inequality and product rule, we have

$$\begin{aligned}
\|B(\phi, \psi) - F\|_{C_R^{2m^*-1}} &\leq \sum_{N>N_0} \|B(P_{(\geq N/R)}\phi_{(<N/R)}, P_{(N/R)}\psi)\|_{C_R^{2m^*-1}} \\
&\lesssim \sum_{N>N_0} \sum_{j_1+j_2=2m^*-1} \|\nabla P_{(>N/R)}\phi_{(<N/R)}\|_{C_R^{j_1}} \|\nabla P_{(N/R)}\psi\|_{C_R^{j_2}}.
\end{aligned}$$

For any $0 \le j_1 \le 2m^* - 1$, one has from (1.13), (1.53), and (1.58) that

$$\begin{aligned}
&\|\nabla P_{(>N/R)}\phi_{(<N/R)}\|_{C_R^{j_1}} \\
&\lesssim R^{-1}N^{j_1+1}\|P_{(>N/R)}\phi_{(<N/R)}\|_{C_{R/N}^{j_1+1}} \\
&\lesssim_G R^{2m^*-1}N^{j_1+1-2m^*}\|\nabla^{2m^*}\phi_{(<N/R)}\|_{C^0} \\
&\lesssim_G R^{2m^*-1}N^{j_1+1-2m^*}\left(\|\nabla^{2m^*}\phi_{(\le N_0/R)}\|_{C^0} + \sum_{N_0<N'<N \text{ dyadic}} \|\nabla^{2m^*}\phi_{(N'/R)}\|_{C^0}\right) \\
&\lesssim_{C_0} R^{2m^*-1}N^{j_1+1-2m^*}\cdot R^{-2m^*}\|\tilde{\phi}\|_{C_R^{2m^*}} + R^{2m^*-1}N^{j_1+1-2m^*}\cdot AN_0R^{1-2m^*}\|F\|_{C_R^{2m^*-1}} \\
&\qquad + R^{2m^*-1}N^{j_1+1-2m^*}\cdot R^{-2m^*}\left(A^{s-m^*-\alpha}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}} + AR\|F\|_{C_R^{2m^*-1}}\right) \\
&\qquad\quad \cdot \sum_{N_0<N'<N \text{ dyadic}} (N')^{m^*-\alpha} \\
&\lesssim_{C_0} N^{j_1+1-m^*-\alpha}(A^{s-m^*-\alpha}R^{-1}\|\tilde{\phi}\|_{C_R^{2m^*}} + A\|F\|_{C_R^{2m^*-1}}).
\end{aligned}$$

Also, for any $0 \le j_2 \le 2m^* - 1$, we have from (1.13) and (1.35) that

$$\begin{aligned}
\|\nabla P_{(N/R)}\psi\|_{C_R^{j_2}} &\lesssim R^{-1}N^{j_2+1}\|P_{(N/R)}\psi\|_{C_{R/N}^{j_2+1}} \lesssim_G R^{-1}N^{j_2+1}(N/R)^{-m^*-\alpha}\|\nabla^{m^*}\psi\|_{\dot{C}^{0,\alpha}} \\
&\lesssim_{C_0} N^{j_2+1-m^*-\alpha}A^{1-m^*-\alpha},
\end{aligned}$$

and thus

$$\begin{aligned}
&\|B(\phi,\psi) - F\|_{C_R^{2m^*-1}} \\
&\lesssim_{C_0} \sum_{N>N_0}\sum_{j_1+j_2=2m^*-1} N^{j_1+1-m^*-\alpha}N^{j_2+1-m^*-\alpha}A^{1-m^*-\alpha} \\
&\qquad\qquad \cdot (A^{s-m^*-\alpha}R^{-1}\|\tilde{\phi}\|_{C_R^{2m^*}} + A\|F\|_{C_R^{2m^*-1}}) \\
&\lesssim_{C_0} \sum_{N>N_0} N^{1-2\alpha}A^{1-m^*-\alpha}(A^{s-m^*-\alpha}R^{-1}\|\tilde{\phi}\|_{C_R^{2m^*}} + A\|F\|_{C_R^{2m^*-1}}) \\
&\lesssim_{C_0} N_0^{1-2\alpha}A^{1-m^*-\alpha}(A^{s-m^*-\alpha}R^{-1}\|\tilde{\phi}\|_{C_R^{2m^*}} + A\|F\|_{C_R^{2m^*-1}})
\end{aligned}$$

(we used $\frac{1}{2} < \alpha < 1$) which gives (1.40).

Finally, we prove (1.42). Fix $i, j = 1, \cdots, \ell$. We need to establish

$$\|Y_i\phi \cdot Y_j\psi - Y_i\tilde{\phi} \cdot Y_j P_{(\leq N_0/R)}\psi\|_{C^0} \lesssim_{C_0} \|F\|_{C_R^{2m^*-1}} + N_0^{1-m^*-\alpha} A^{1-m^*-\alpha} \|\tilde{\phi}\|_{C_R^{m^*,\alpha}}.$$

Using telescoping sums,

$$\begin{aligned}
& Y_i\phi \cdot Y_j\psi - Y_i\tilde{\phi} \cdot Y_j P_{(\leq N_0/R)}\psi \\
&= Y_i(\phi_{(\leq N_0/R)} - \tilde{\phi}) \cdot Y_j P_{(\leq N_0/R)}\psi \\
&\quad + \sum_{N>N_0} \left(Y_i\phi_{(\leq N/R)} \cdot Y_j P_{(\leq N/R)}\psi - Y_i\phi_{(<N/R)} \cdot Y_j P_{(<N/R)}\psi \right) \\
&\overset{(1.44)}{=} P_{(\leq N_0/R)}F_{ij} + \sum_{N>N_0} \left(Y_i\phi_{(N/R)} \cdot Y_j P_{(\leq N/R)}\psi + Y_i\phi_{(<N/R)} \cdot Y_j P_{(N/R)}\psi \right) \\
&\overset{(1.49)}{=} P_{(\leq N_0/R)}F_{ij} + \sum_{N>N_0} P_{(N/R)}F_{ij} \\
&\quad + \sum_{N>N_0} \left(-Y_j P_{(\leq N/R)}\phi_{(<N/R)} \cdot Y_i P_{(N/R)}\psi + Y_i\phi_{(<N/R)} \cdot Y_j P_{(N/R)}\psi \right) \\
&= F_{ij} + \sum_{N>N_0} \left(-Y_j P_{(\leq N/R)}\phi_{(<N/R)} \cdot Y_i P_{(N/R)}\psi + Y_i\phi_{(<N/R)} \cdot Y_j P_{(N/R)}\psi \right).
\end{aligned}$$

Thus

$$\begin{aligned}
& \|Y_i\phi \cdot Y_j\psi - Y_i\tilde{\phi} \cdot Y_j P_{(\leq N_0/R)}\psi\|_{C^0} \\
&\leq \|F\|_{C^0} + \sum_{N>N_0} \left(\|\nabla P_{(\leq N/R)}\phi_{(<N/R)}\|_{C^0} + \|\nabla\phi_{(<N/R)}\|_{C^0} \right) \|\nabla P_{(N/R)}\psi\|_{C^0} \\
&\overset{(1.12)}{\lesssim_G} \|F\|_{C^0} + \sum_{N>N_0} \|\nabla\phi_{(<N/R)}\|_{C^0} \|\nabla P_{(N/R)}\psi\|_{C^0} \\
&\overset{(1.57),(1.54)}{\lesssim_{C_0}} \|F\|_{C^0} + \sum_{N>N_0} \left(A\|F\|_{C_R^{2m^*-1}} + R^{-1}\|\tilde{\phi}\|_{C_R^{m^*,\alpha}} \right) N^{1-m^*-\alpha} A^{1-m^*-\alpha} \\
&\lesssim \|F\|_{C_R^{2m^*-1}} + N_0^{1-m^*-\alpha} A^{1-m^*-\alpha} R^{-1} \|\tilde{\phi}\|_{C_R^{m^*,\alpha}},
\end{aligned}$$

as desired. □

We state Proposition 1.2.1 for the case $F = 0$, which is the form that will be used later. (We remark that when converting (1.39) into (1.62) below, we have used our hierarchy of choosing N_0 after C_0.)

Corollary 1.2.4. *Let R be a dyadic number and let M be a real number with*

$$M \geq C_0^{-1}.$$

Let $m^* \geq s+1$ and $\frac{1}{2} < \alpha < 1$. Suppose we are given a $C^{m^*,\alpha}$-map $\psi : G \to \mathbb{R}^D$ with the following regularity properties:

1. (Hölder regularity at scale A) We have

$$\|\nabla^2 \psi\|_{C_{AR}^{m^*-2,\alpha}} \leq \frac{C_0}{AR}.$$

2. (nondegenerate first derivatives) For any $p \in G$, we have

$$C_0^{-1} M \leq |Y_i \psi(p)| \leq C_0 M, \quad i = 1, \cdots, \ell,$$

3. (locally free embedding) For any $p \in G$, we have

$$\left| \bigwedge_{i=1}^{\ell} Y_i \psi(p) \wedge \bigwedge_{1 \leq i \leq j \leq \ell} Y_i Y_j \psi(p) \wedge \bigwedge_{r=2}^{s} \bigwedge_{i'=1}^{\ell_r} Y_{r,i'} \psi(p) \right| \gtrsim_{C_0} M^{-\ell} (AR)^{-\frac{\ell(\ell+1)}{2} - \sum_{r=2}^{s} (r-1)\ell_r}.$$

Let $\tilde{\phi} : G \to \mathbb{R}^D$ be a C^{2m^*}-solution to the low-frequency equation (1.27)*. Then there exists a $C^{m^*,\alpha}$-solution ϕ to* (1.28) *which is a small perturbation of $\tilde{\phi}$:*

$$\|\phi - \tilde{\phi}\|_{C_R^{m^*,\alpha}} \lesssim_{C_0} A^{2-m^*} \|\tilde{\phi}\|_{C_R^{m^*,\alpha}}, \tag{1.61}$$

and which makes the following cross terms small:

$$\|Y_i\phi \cdot Y_j\psi - Y_i\tilde{\phi} \cdot Y_j P_{(\leq N_0/R)}\psi\|_{C^0} \leq A^{1-m^*}R^{-1}\|\tilde{\phi}\|_{C_R^{2m^*}}, \quad i,j = 1,\cdots,\ell. \tag{1.62}$$

Remark 1.2.5. *1. To repeat, we will later on use $\alpha = \frac{2}{3}$ and $m^* = s^2 + s + 1$, where G is of step s, so these are to be considered "geometric constants". Following our hierarchy of constants, we will be choosing C_0, N_0, and A, in that order.*

2. At first glance, Corollary 1.2.4 may seem to be missing a necessary dependence on N_0: the solution $\tilde{\phi}$ to (1.27) *approximates the solution ϕ to* (1.28)*, with the approximation getting better as N_0 gets larger, yet the quantitative estimates* (1.61) *and* (1.62) *lack any explicit dependence on N_0. However, the dependence is implicit:* (1.61) *and* (1.62) *depend on A which in turn depends on N_0, and as N_0 gets larger, the right-hand sides of* (1.61) *and* (1.62) *get smaller. The part when we needed the dependence of A on N_0 in the proof of Proposition 1.2.1 was in* (1.51)*.*

1.3 Regular extensions of orthonormal systems

Another tool that we will need is a certain result on extending orthonormal systems. In order to apply Corollary 1.2.4, we first need to construct a nontrivial solution $\tilde{\phi}$ to the low-frequency equation (1.27), or more generally the stronger equation

$$Y_i\tilde{\phi} \cdot Y_j P_{(\leq N_0/R)}\psi = 0, \quad i,j = 1,\cdots,\ell \tag{1.63}$$

as promised in Remark 1.2.2. One can see using the Leibniz rule that it is enough to solve the system

$$\begin{cases} \tilde{\phi} \cdot Y_i P_{(\leq N_0/R)}\psi = 0, & i = 1,\cdots,\ell, \\ \tilde{\phi} \cdot Y_i Y_j P_{(\leq N_0/R)}\psi = 0, & i,j = 1,\cdots,\ell. \end{cases} \tag{1.64}$$

However, the vectors $\{Y_iY_jP_{(\leq N_0/R)}\psi\}_{i,j=1,\cdots,\ell}$ may not be linearly independent, as $\{Y_iY_j - Y_jY_i : 1 \leq i < j \leq \ell\}$ may be linearly dependent, so the above system (1.64) may be overdetermined. Instead, we will solve the equivalent system

$$\begin{cases} \tilde{\phi} \cdot Y_iP_{(\leq N_0/R)}\psi = 0, & i = 1, \cdots, \ell, \\ \tilde{\phi} \cdot Y_iY_jP_{(\leq N_0/R)}\psi = 0, & i, j = 1, \cdots, \ell,\ i \leq j, \\ \tilde{\phi} \cdot Y_{2,i}P_{(\leq N_0/R)}\psi = 0, & i = 1, \cdots, \ell_2. \end{cases} \tag{1.65}$$

The vectors $\{Y_iP_{(\leq N_0/R)}\psi\}_{i=1,\cdots,\ell} \cup \{Y_iY_jP_{(\leq N_0/R)}\psi\}_{1\leq i\leq j\leq \ell} \cup \{Y_{2,i}P_{(\leq N_0/R)}\psi\}_{i=1,\cdots,\ell_2}$ can be made linearly independent if we require the stronger freeness property

$$\left| \bigwedge_{i=1}^{\ell} Y_i\psi \wedge \bigwedge_{1\leq i\leq j\leq \ell} Y_iY_j\psi \wedge \bigwedge_{i=1}^{\ell_2} Y_{2,i}\psi \right| \gtrsim_{C_0} \prod_{i=1}^{\ell} |Y_i\psi| \cdot \prod_{1\leq i\leq j\leq \ell} |Y_iY_j\psi| \cdot \prod_{i=1}^{\ell_2} |Y_{2,i}\psi| \tag{1.66}$$

along with some regularity, say $\|\nabla^2\psi\|_{C^1_{AR}} \lesssim_{C_0} \frac{1}{AR}$, for then we can apply (1.13) to approximate the derivatives of $P_{(\leq N_0/R)}\psi$ by the corresponding derivatives of ψ.

Now suppose $v_1, \cdots, v_{\ell(\ell+1)/2+\ell_2}$ is the result of applying the Gram-Schmidt process to the vectors

$$\{Y_iP_{(\leq N_0/R)}\psi\}_{i=1,\cdots,\ell} \cup \{Y_iY_jP_{(\leq N_0/R)}\psi\}_{1\leq i\leq j\leq \ell} \cup \{Y_{2,i}P_{(\leq N_0/R)}\psi\}_{i=1,\cdots,\ell_2}.$$

The freeness property (1.66) guarantees some regularity of the v_i, and solving (1.65) is equivalent to solving

$$\tilde{\phi} \cdot v_i = 0, \quad i = 1, \cdots, \frac{\ell(\ell+1)}{2} + \ell_2. \tag{1.67}$$

Thus, the following general question arises:

Question 1.3.1. *Given a space X on which a function space $(\mathcal{F}(X), \|\cdot\|_{\mathcal{F}})$ is defined, and given maps $v_1, \cdots, v_m : X \to \mathbb{S}^{D-1}$, $D \geq m+1$, which form a pointwise orthonor-*

mal system and which have the uniform regularity bound $\|v_i\|_{\mathcal{F}} \lesssim 1$*, when can we extend the system to include a new map* $v_{m+1} : X \to \mathbb{S}^{D-1}$*, such that* $v_1, \cdots, v_m, v_{m+1}$ *forms a pointwise orthonormal system and* $\|v_{m+1}\|_{\mathcal{F}} \lesssim_{X,\mathcal{F}(X),m,D} 1$*?*

We will provide partial positive answers to this question that are applicable to our construction of the embedding. Theorem 1.3.2 is a previous positive answer from [77], while Theorems 1.3.3, 1.3.5, and 1.3.6 and Corollary 1.3.7 are the new positive answers of this paper.

Under conditions that validate the above question, we would be able to simply take $\tilde{\phi}$ to be v_{m+1} to solve (1.67), for then $\tilde{\phi}$ would have regularity similar to that of the v_i; i.e., it would have bounded $C^{m^*,\alpha}$ norm (there is some loss of constant factors when applying (1.12), but this is offset by the fact that there is a change of scale: we have control on the $C^{m^*,\alpha}_{AR}$ norm of ψ, and we need only control the $C^{m^*,\alpha}_{R}$ norm of $\tilde{\phi}$).

Actually, in order to also obtain a freeness property for $\psi + \phi$, we will take $\tilde{\phi}$ to be a linear combination of a larger extension $v_{m+1}, \cdots, v_{m+m'}$ of $v_1, \cdots, v_m$ with variable coefficients; the coefficients will guarantee the freeness property in this case (see (1.99) and (1.110)). Obtaining a larger orthonormal extension will be possible simply by adding new vectors one by one.

In [77, Section 8] Question 1.3.1 has been answered in the affirmative for the case $X = \mathbb{H}^3$ and

$$\|\phi\|_{\mathcal{F}} = \|\phi\|_{C^0} + r\|\nabla\phi\|_{C^j}$$

for any $j \geq 0$ and $r \geq 1$. The proof in [77] used the fact that the Heisenberg group $\mathbb{H}^3$ admits a CW complex structure that is periodic with respect to its standard discrete cocompact lattice. The methods of [77] can be generalized in a straightforward manner to prove the following theorem.

Theorem 1.3.2 ([77, Corollary 8.4])**.** *Let* G *be a simply connected nilpotent Lie group that admits a cocompact lattice* Γ *and a CW structure whose cells can be obtained from*

left Γ-translation from a finite list of cells. Let $1 \leq m \leq D-n-1$ (where n is the topological dimension of G), $j \geq 1$, and let $\{R_i\}_{i=1}^j$ be a log-concave sequence of positive reals, i.e., $R_i R_{i'} \geq R_{i+i'}$ whenever $i+i' \leq j$. Let $v_1, \cdots, v_m : G \to \mathbb{S}^{D-1}$ be functions that form an orthonormal system at each point, with the uniform regularity bound

$$\sum_{k=1}^{j} R_k \|\nabla^k v_i\|_{C^0} \leq 1, \quad i = 1, \cdots, m.$$

Then there exists another function $v_{m+1} : G \to \mathbb{S}^{D-1}$ such that $v_1, \cdots, v_m$ along with v_{m+1} form an orthonormal system at each point, and

$$\sum_{k=1}^{j} R_k \|\nabla^k v_{m+1}\|_{C^0} \lesssim_{G,D,j,\{R_k\}} 1. \tag{1.68}$$

In other words, we are given a bundle B over G, where for each $p \in G$ the fiber of B over p is the collection of unit vectors $v \in \mathbb{S}^{D-1}$ which are perpendicular to $v_1(p), \cdots, v_m(p)$ (so each fibre is homeomorphic to $\mathbb{S}^{D-m-1}$), and we need to show that there is a section of this bundle which has the same level of control on the regularity of the bundle itself. This is achieved in [77, Section 8] for the Heisenberg group $\mathbb{H}^3$ in the spirit of quantitative topology, by imposing a "uniform" CW-structure on $\mathbb{H}^3$ as above and then inductively constructing the section starting from low-dimensional skeleta. In the inductive step in [77], one has to use the fact that the homotopy groups $\pi_i(\mathbb{S}^n)$ vanish for $i < n$, which necessitates the "dimension gap" $m \leq D-3-1$.

The main difficulty in proving Theorem 1.3.2 is to construct a section $\tilde{v}_{m+1}$ which is uniformly continuous, with the modulus of uniform continuity depending only on G, D and R_1. We can then obtain a section v_{m+1} with the stronger regularity property (1.68) simply by mollifying $\tilde{v}_{m+1}$ and then applying the Gram-Schmidt orthogonalization process; the regularity (1.68) will simply be a consequence of the algebra property for norms of the form $\|\cdot\|_{C^0} + \sum_{k=1}^j R_k \|\nabla^k \cdot\|_{C^0}$.

General simply connected nilpotent Lie groups may not admit a cocompact lattice

(as the structure constants for any basis may be irrational), and it is not clear whether G admits a "uniform" CW structure that is amenable to the above proof method. Furthermore, we lack a dilation structure on G, which would have been useful in getting scale-independent bounds. We will avoid the need for a CW structure by only using the fact that G is a doubling metric space. More precisely, we will first prove that the most challenging part of Theorem 1.3.2, i.e., the case $j = 1$, can be done in the setting of doubling metric spaces. We state this result separately in anticipation of future work.

Theorem 1.3.3. *Let (X, d) be a K-doubling metric space ($K \geq 2$), let $m \leq D - 224K^4 \log K$, and let $L > 0$. If $v_1, \cdots, v_m : X \to \mathbb{S}^{D-1}$ are L-Lipschitz functions that form an orthonormal system at each point, then there exists a $150K^5 m(m+1)L$-Lipschitz function $v_{m+1} : X \to \mathbb{S}^{D-1}$ such that $v_1, \cdots, v_m$ along with v_{m+1} form an orthonormal system at each point.*

It will later become clear that Theorem 1.3.3 provides a partial positive answer to Question 1.3.1 for many other function spaces as well; see Theorem 1.3.6.

Proof. By replacing the metric d with $L \cdot d$, we may assume $L = 1$.

Take a maximal δ-net $\mathcal{N}_\delta$ of X, where $\delta = 1/(8Km)$. By (1.6),

$$|\mathcal{N}_\delta \cap B_{2\delta}(p)| \leq K^2 \quad \text{and} \quad |\mathcal{N}_\delta \cap B_{4\delta}(p)| \leq K^3.$$

Let Ω be a probability space on which independent random variables $v'_{m+1}(p) \in \mathbb{S}^{D-1}$ are defined so that for each $\omega \in \Omega$ and $p \in X$, $v'_{m+1}(p)(\omega)$ forms an orthonormal set along with $v_1(p)(\omega)$, $\cdots$, $v_m(p)(\omega)$. Let $\epsilon = 1/(4K^2)$, and for each $p \in \mathcal{N}_\delta$, let us define the event

$$\begin{aligned} A_p = \{&\omega \in \Omega : \text{there exists } q \in \mathcal{N}_\delta \cap (B_{2\delta}(p) \setminus \{p\}) \text{ such that} \\ &|v'_{m+1}(p)(\omega) \cdot v'_{m+1}(q)(\omega)| > \epsilon\}. \end{aligned}$$

Note that for any $p, q \in \mathcal{N}_\delta$ distinct, we may compute

$$\Pr(|v'_{m+1}(p) \cdot v'_{m+1}(q)| > \epsilon) = \mathbb{E}_p \Pr_q(|v'_{m+1}(p) \cdot v'_{m+1}(q)| > \epsilon) \leq \exp\left(-\frac{\epsilon^2}{2}(D-m)\right),$$

where the last inequality follows from a standard computation on the area of caps on the sphere (see [58, Chapter 2] for instance). Therefore, we may estimate the probability of each A_p using a union bound:

$$\begin{aligned}\Pr(A_p) &\leq \sum_{q \in \mathcal{N}_\delta \cap (B_{2\delta}(p) \setminus \{p\})} \Pr(|v'_{m+1}(p) \cdot v'_{m+1}(q)| > \epsilon) \\ &\leq |\mathcal{N}_\delta \cap (B_{2\delta}(p) \setminus \{p\})| \cdot \exp\left(-\frac{\epsilon^2}{2}(D-m)\right) \leq K^2 \cdot \exp\left(-\frac{\epsilon^2}{2}(D-m)\right).\end{aligned}$$

Also note that for each $p \in \mathcal{N}_\delta$, A_p is mutually independent with the collection of events $\{A_q : q \in \mathcal{N}_\delta \setminus B_{4\delta}(p)\}$, which are all the A_q except possibly $|\mathcal{N}_\delta \cap B_{4\delta}(p)| \leq K^3$ of them. By the Lovász local lemma, we see that if

$$e \cdot K^3 \cdot K^2 \cdot \exp\left(-\frac{\epsilon^2}{2}(D-m)\right) < 1, \tag{1.69}$$

then for any finite subcollection $S \subset \mathcal{N}_\delta$ we have $\Pr(\bigcap_{p \in S} A_p^{\mathsf{c}}) > 0$. But by our choice of parameters $D - m \geq 224K^4 \log K$ and $\epsilon = \frac{1}{4K^2}$, the condition (1.69) is indeed satisfied, since the left-hand side is bounded by

$$\begin{aligned}&e \cdot K^5 \cdot \exp\left(-\frac{\epsilon^2}{2}(D-m)\right) \leq e \cdot K^5 \cdot \exp\left(-7 \log K\right) < 1 \\ &\text{(since } 7 \log K > 5 \log K + 1 \text{ as } K \geq 2).\end{aligned}$$

Hence, for any finite subcollection $S \subset \mathcal{N}_\delta$ we have $\Pr(\bigcap_{p \in S} A_p^{\mathsf{c}}) > 0$, and in particular $\bigcap_{p \in S} A_p^{\mathsf{c}} \neq \emptyset$. We can thus find an assignment $\{v'_{m+1,S}(p)\}_{p \in S}$ such that for any distinct $p, q \in S$ with $d(p, q) < 2\delta$ we have $|v'_{m+1,S}(p) \cdot v'_{m+1,S}(q)| \leq \epsilon$. By taking an arbitrary

enumeration of $\mathcal{N}_\delta$, taking a monotone increasing sequence of S's that cover $\mathcal{N}_\delta$ and passing to a limit along a nonprincipal ultrafilter, we conclude the existence of an assignment $\{v'_{m+1}(p)\}_{p\in\mathcal{N}_\delta}$ such that for any distinct $p, q \in \mathcal{N}_\delta$ with $d(p,q) < 2\delta$ we have $|v'_{m+1}(p) \cdot v'_{m+1}(q)| \leq \epsilon$.

We now "interpolate" the discrete vector field $\{v'_{m+1}(p)\}_{p\in\mathcal{N}_\delta}$ to produce a vector field $\{\tilde{v}_{m+1}(p)\}_{p\in X}$ defined on the entirety of X, which nearly has the desired properties. We first construct a "quadratic" partition of unity $\{\phi_q\}_{q\in\mathcal{N}_\delta}$, i.e., functions $\phi_q : X \to [0,1]$ defined for each $q \in \mathcal{N}_\delta$ such that

- supp $\phi_q \subset B_{2\delta}(q)$, $q \in \mathcal{N}_\delta$,
- $\sum_{q\in\mathcal{N}_\delta} \phi_q^2 = 1$ on X,
- $\|\phi_q\|_{\mathrm{Lip}} \leq 2K^2\delta^{-1}$, $q \in \mathcal{N}_\delta$.

Indeed, we start by defining for each $q \in \mathcal{N}_\delta$ the function

$$\tilde{\phi}_q(p) = \begin{cases} 1, & d(p,q) \leq \delta, \\ 2 - \frac{d(p,q)}{\delta}, & \delta < d(p,q) \leq 2\delta, \\ 0, & d(p,q) > 2\delta, \end{cases}$$

and then define

$$\phi_q \coloneqq \frac{\tilde{\phi}_q}{\sqrt{\sum_{r\in\mathcal{N}_\delta} \tilde{\phi}_r^2}}.$$

This obviously satisfies the first two properties, and it remains to compute $\|\phi_q\|_{\mathrm{Lip}}$. We clearly have $\|\tilde{\phi}_q\|_{C^0} \leq 1$, $\|\tilde{\phi}_q\|_{\mathrm{Lip}} \leq \delta^{-1}$.

We first observe that

$$\sum_{r\in\mathcal{N}_\delta} \tilde{\phi}_r^2 \geq 1 \quad \text{and} \quad \left\| \sum_{r\in\mathcal{N}_\delta} \tilde{\phi}_r^2 \right\|_{\mathrm{Lip}} \leq (4K^2 - 2)\delta^{-1}.$$

Indeed, the first follows from the fact that $\mathcal{N}_\delta$ is a maximal δ-net. For the second property, fix any $p, p' \in X$ with $p \neq p'$. If $B_{2\delta}(p) \cap B_{2\delta}(p') \cap \mathcal{N}_\delta = \emptyset$ then we must have $d(p, p') \geq \delta$, because otherwise $B_{2\delta}(p) \cap B_{2\delta}(p') \cap \mathcal{N}_\delta \supseteq B_\delta(p) \cap \mathcal{N}_\delta \neq \emptyset$, by maximality of $\mathcal{N}_\delta$. We have

$$1 \leq \sum_{r \in \mathcal{N}_\delta} \tilde{\phi}_r(p)^2 = \sum_{r \in B_{2\delta}(p) \cap \mathcal{N}_\delta} \tilde{\phi}_r(p)^2 \leq |B_{2\delta}(p) \cap \mathcal{N}_\delta| \leq K^2,$$

and similarly

$$1 \leq \sum_{r \in \mathcal{N}_\delta} \tilde{\phi}_r(p')^2 \leq K^2.$$

Therefore, in this case,

$$\frac{\left| \sum_{r \in \mathcal{N}_\delta} \tilde{\phi}_r(p)^2 - \sum_{r \in \mathcal{N}_\delta} \tilde{\phi}_r(p')^2 \right|}{d(p, p')} \leq (K^2 - 1)\delta^{-1} \leq (4K^2 - 2)\delta^{-1}.$$

On the other hand, if $B_{2\delta}(p) \cap B_{2\delta}(p') \cap \mathcal{N}_\delta \neq \emptyset$ then clearly

$$|B_{2\delta}(p) \cap B_{2\delta}(p') \cap \mathcal{N}_\delta| \leq |B_{2\delta}(p) \cap \mathcal{N}_\delta| + |B_{2\delta}(p') \cap \mathcal{N}_\delta| - 1 \leq 2K^2 - 1.$$

We have

$$\sum_{r \in \mathcal{N}_\delta} \tilde{\phi}_r(p)^2 = \sum_{r \in B_{2\delta}(p) \cap \mathcal{N}_\delta} \tilde{\phi}_r(p)^2 = \sum_{r \in (B_{2\delta}(p) \cup B_{2\delta}(p')) \cap \mathcal{N}_\delta} \tilde{\phi}_r(p)^2$$

and similarly

$$\sum_{r \in \mathcal{N}_\delta} \tilde{\phi}_r(p')^2 = \sum_{r \in (B_{2\delta}(p) \cup B_{2\delta}(p')) \cap \mathcal{N}_\delta} \tilde{\phi}_r(p')^2,$$

so, noting that $|\tilde{\phi}_r(p)^2 - \tilde{\phi}_r(p')^2| \le 2|\tilde{\phi}_r(p)^2 - \tilde{\phi}_r(p')| \le 2\|\tilde{\phi}_r\|_{\mathrm{Lip}} d(p,p') \le 2\delta^{-1} d(p,p')$,

$$\begin{aligned}\left|\sum_{r\in\mathcal{N}_\delta} \tilde{\phi}_r(p)^2 - \sum_{r\in\mathcal{N}_\delta} \tilde{\phi}_r(p')^2\right| &\le \sum_{r\in(B_{2\delta}(p)\cup B_{2\delta}(p'))\cap\mathcal{N}_\delta} |\tilde{\phi}_r(p)^2 - \tilde{\phi}_r(p')^2| \\ &\le |(B_{2\delta}(p)\cup B_{2\delta}(p'))\cap\mathcal{N}_\delta| \cdot 2\delta^{-1} d(p,p') \\ &\le (2K^2-1)\cdot 2\delta^{-1} d(p,p').\end{aligned}$$

This finishes the verification that $\left\|\sum_{r\in\mathcal{N}_\delta} \tilde{\phi}_r^2\right\|_{\mathrm{Lip}} \le (4K^2-2)\delta^{-1}$.

By (1.3) and (1.4), we have

$$\left\|\frac{1}{\sqrt{\sum_{r\in\mathcal{N}_\delta} \tilde{\phi}_r^2}}\right\|_{\mathrm{Lip}} \le \left\|\sqrt{\sum_{r\in\mathcal{N}_\delta} \tilde{\phi}_r^2}\right\|_{\mathrm{Lip}} \le \frac{1}{2}\left\|\sum_{r\in\mathcal{N}_\delta} \tilde{\phi}_r^2\right\|_{\mathrm{Lip}} \le (2K^2-1)\delta^{-1}.$$

Thus, by the definition of ϕ_q and (1.2), we have

$$\begin{aligned}\|\phi_q\|_{\mathrm{Lip}} &\le \left\|\frac{1}{\sqrt{\sum_{r\in\mathcal{N}_\delta} \tilde{\phi}_r^2}}\right\|_{C^0} \left\|\tilde{\phi}_q\right\|_{\mathrm{Lip}} + \left\|\frac{1}{\sqrt{\sum_{r\in\mathcal{N}_\delta} \tilde{\phi}_r^2}}\right\|_{\mathrm{Lip}} \left\|\tilde{\phi}_q\right\|_{C^0} \\ &\le 1\cdot\delta^{-1} + (2K^2-1)\delta^{-1}\cdot 1 \le 2K^2\delta^{-1}.\end{aligned}$$

We now interpolate the vectors $\{v'_{m+1}(q)\}_{q\in\mathcal{N}_\delta}$ using the quadratic partition of unity $\{\phi_q\}_{q\in\mathcal{N}_\delta}$:

$$\tilde{v}_{m+1}(p) \coloneqq \sum_{q\in\mathcal{N}_\delta} \phi_q(p) v'_{m+1}(q), \quad p\in X. \tag{1.70}$$

The idea is that this interpolates nearby v'_{m+1}, which are mutually almost orthogonal, so $\tilde{v}_{m+1}$ should nearly be a unit vector. Moreover, since the 1-Lipschitz functions $v_1,\cdots,v_m$ vary slowly over distance δ, this $\tilde{v}_{m+1}$ should also be nearly orthogonal to $v_1,\cdots,v_m$. This $\tilde{v}_{m+1}$ will oscillate much quicker than $v_1,\cdots,v_m$, but its Lipschitz constant will be controlled by K and δ.

More precisely, we claim that $\tilde{v}_{m+1}$ satisfies the following quantitative estimates:

1. $\frac{3}{4} \leq |\tilde{v}_{m+1}(p)|^2 \leq \frac{5}{4}$ for each $p \in X$,

2. $|\tilde{v}_{m+1}(p) \cdot v_i(p)| \leq \frac{1}{4m}$ for each $p \in X$ and $i = 1, \cdots, m$,

3. $\|\tilde{v}_{m+1}\|_{\text{Lip}} \leq 2K^2(2K^2 - 1)\delta^{-1}$.

Indeed, by the support property of ϕ_q, we see that the summation in (1.70) is locally finite:

$$\tilde{v}_{m+1}(p) = \sum_{q \in \mathcal{N}_\delta \cap B_{2\delta}(p)} \phi_q(p) v'_{m+1}(q).$$

We can verify property (1) using the near-orthogonality and the fact that the sum of squares of the ϕ_q is 1:

$$\begin{aligned}
\left||\tilde{v}_{m+1}(p)|^2 - 1\right| &= \left| \sum_{\substack{q,q' \in \mathcal{N}_\delta \cap B_{2\delta}(p) \\ q \neq q'}} \phi_q(p)\phi_{q'}(p) v'_{m+1}(q) \cdot v'_{m+1}(q') \right| \\
&\leq \sum_{\substack{q,q' \in \mathcal{N}_\delta \cap B_{2\delta}(p) \\ q \neq q'}} \phi_q(p)\phi_{q'}(p) \left| v'_{m+1}(q) \cdot v'_{m+1}(q') \right| \\
&\leq \sum_{\substack{q,q' \in \mathcal{N}_\delta \cap B_{2\delta}(p) \\ q \neq q'}} \frac{\phi_q(p)^2 + \phi_{q'}(p)^2}{2} \epsilon \leq |\mathcal{N}_\delta \cap B_{2\delta}(p)| \epsilon \leq K^2 \epsilon = \frac{1}{4}.
\end{aligned}$$

To verify property (2), we see that for each $p \in X$ and $i = 1, \cdots, m$,

$$\tilde{v}_{m+1}(p) \cdot v_i(p) = \sum_{q \in \mathcal{N}_\delta \cap B_{2\delta}(p)} \phi_q(p) v'_{m+1}(q) \cdot v_i(p).$$

We observe that for each $q \in \mathcal{N}_\delta \cap B_{2\delta}(p)$, we can estimate

$$\begin{aligned}
|v'_{m+1}(q) \cdot v_i(p)| &\leq |v'_{m+1}(q) \cdot v_i(q)| + |v'_{m+1}(q) \cdot (v_i(p) - v_i(q))| \\
&\leq 0 + |v_i(p) - v_i(q)| \leq 2\delta,
\end{aligned}$$

where we have used the orthogonality of $v'_{m+1}(q)$ against $v_i(q)$ and the fact that

$\|v_i\|_{\mathrm{Lip}} \leq 1$. By these facts and Cauchy–Schwarz, we have the bound

$$|\tilde{v}_{m+1}(p) \cdot v_i(p)| \leq \sum_{q \in \mathcal{N}_\delta \cap B_{2\delta}(p)} \phi_q(p) \cdot 2\delta \leq |\mathcal{N}_\delta \cap B_{2\delta}(p)|^{1/2} \cdot 2\delta \leq K \cdot 2\delta = \frac{1}{4m}.$$

To verify (3), we first fix $p, p' \in X$ with $p \neq p'$. If $B_{2\delta}(p) \cap B_{2\delta}(p') \cap \mathcal{N}_\delta = \emptyset$ then we must have $d(p, p') \geq \delta$, because otherwise $B_{2\delta}(p) \cap B_{2\delta}(p') \cap \mathcal{N}_\delta \supseteq B_\delta(p) \cap \mathcal{N}_\delta \neq \emptyset$. By (1), we have $\frac{\sqrt{3}}{2} \leq |\tilde{v}_{m+1}(p)|, |\tilde{v}_{m+1}(p')| \leq \frac{\sqrt{5}}{2}$, so

$$\frac{|\tilde{v}_{m+1}(p) - \tilde{v}_{m+1}(p')|}{d(p, p')} \leq \frac{\sqrt{5} - \sqrt{3}}{2} \delta^{-1} \leq 2K^2(2K^2 - 1)\delta^{-1}.$$

On the other hand, if $B_{2\delta}(p) \cap B_{2\delta}(p') \cap \mathcal{N}_\delta \neq \emptyset$ then clearly

$$|B_{2\delta}(p) \cap B_{2\delta}(p') \cap \mathcal{N}_\delta| \leq |B_{2\delta}(p) \cap \mathcal{N}_\delta| + |B_{2\delta}(p') \cap \mathcal{N}_\delta| - 1 \leq 2K^2 - 1.$$

We have

$$\tilde{v}_{m+1}(p) = \sum_{r \in B_{2\delta}(p) \cap \mathcal{N}_\delta} \phi_r(p) v_{m+1}(r) = \sum_{r \in (B_{2\delta}(p) \cup B_{2\delta}(p')) \cap \mathcal{N}_\delta} \phi_r(p) v_{m+1}(r)$$

and similarly

$$\tilde{v}_{m+1}(p') = \sum_{r \in (B_{2\delta}(p) \cup B_{2\delta}(p')) \cap \mathcal{N}_\delta} \phi_r(p') v_{m+1}(r),$$

so

$$\begin{aligned}
|\tilde{v}_{m+1}(p) - \tilde{v}_{m+1}(p')| &\leq \sum_{r \in (B_{2\delta}(p) \cup B_{2\delta}(p')) \cap \mathcal{N}_\delta} |\phi_r(p) - \phi_r(p')| \cdot |v_{m+1}(r)| \\
&\leq \sum_{r \in (B_{2\delta}(p) \cup B_{2\delta}(p')) \cap \mathcal{N}_\delta} \|\phi_r\|_{\mathrm{Lip}} d(p, p') \\
&\leq |(B_{2\delta}(p) \cup B_{2\delta}(p')) \cap \mathcal{N}_\delta| \cdot 2K^2 \delta^{-1} d(p, p') \\
&\leq 2K^2(2K^2 - 1)\delta^{-1} d(p, p').
\end{aligned}$$

This finishes the verification of properties (1) to (3).

Properties (1)-(3) above finally allow us to use the Gram-Schmidt orthogonalization process to obtain a true v_{m+1} with the desired properties:

$$v_{m+1}(p) := \frac{\tilde{v}_{m+1}(p) - \sum_{i=1}^{m} \tilde{v}_{m+1}(p) \cdot v_i(p)}{|\tilde{v}_{m+1}(p) - \sum_{i=1}^{m} \tilde{v}_{m+1}(p) \cdot v_i(p)|}, \quad p \in X.$$

This is well-defined because

$$|\tilde{v}_{m+1}(p) - \sum_{i=1}^{m} \tilde{v}_{m+1}(p) \cdot v_i(p)| \geq |\tilde{v}_{m+1}(p)| - \sum_{i=1}^{m} |\tilde{v}_{m+1}(p) \cdot v_i(p)| \geq \frac{\sqrt{3}}{2} - \frac{1}{4} > 0,$$

and $v_{m+1}(p)$ clearly forms an orthonormal system along with $v_1(p), \cdots, v_m(p)$. We also note that

$$|\tilde{v}_{m+1}(p) - \sum_{i=1}^{m} \tilde{v}_{m+1}(p) \cdot v_i(p)| \leq |\tilde{v}_{m+1}(p)| + \sum_{i=1}^{m} |\tilde{v}_{m+1}(p) \cdot v_i(p)| \leq \frac{\sqrt{3}}{2} + \frac{1}{4}.$$

To check the Lipschitz regularity of v_{m+1} we first compute (recalling $\delta^{-1} = 8Km$)

$$\begin{aligned}
\left\| \tilde{v}_{m+1} - \sum_{i=1}^{m} \tilde{v}_{m+1} \cdot v_i \right\|_{\text{Lip}} &\leq \|\tilde{v}_{m+1}\|_{\text{Lip}} + \sum_{i=1}^{m} (\|\tilde{v}_{m+1}\|_{C^0} \|v_i\|_{\text{Lip}} + \|\tilde{v}_{m+1}\|_{\text{Lip}} \|v_i\|_{C^0}) \\
&\leq 2K^2(2K^2 - 1)\delta^{-1} + m(\frac{\sqrt{5}}{2} \cdot 1 + 2K^2(2K^2 - 1)\delta^{-1} \cdot 1) \\
&= 2K^2\delta^{-1}\Big(2K^2 - 1 + \frac{\sqrt{5}}{32K^3} + 2K^2 m - m\Big) \\
&\leq 4K^4\delta^{-1}(m+1) = 32K^5 m(m+1).
\end{aligned}$$

By (1.5) we finally have

$$\begin{aligned}\|v_{m+1}\|_{\mathrm{Lip}} &\leq \Big((\frac{\sqrt{3}}{2}-\frac{1}{4})^{-1}+(\frac{\sqrt{3}}{2}-\frac{1}{4})^{-2}(\frac{\sqrt{3}}{2}+\frac{1}{4})\Big)\left\|\tilde{v}_{m+1}-\sum_{i=1}^{m}\tilde{v}_{m+1}\cdot v_i\right\|_{\mathrm{Lip}}\\ &\leq \Big((\frac{\sqrt{3}}{2}-\frac{1}{4})^{-1}+(\frac{\sqrt{3}}{2}-\frac{1}{4})^{-2}(\frac{\sqrt{3}}{2}+\frac{1}{4})\Big)\cdot 32K^5 m(m+1)\\ &\leq 150K^5 m(m+1).\end{aligned}$$

□

Remark 1.3.4. *In Theorem 1.3.2, the required dimension gap $D-m-1$ is precisely the topological dimension of the given domain G. In contrast, in Theorem 1.3.3 the dimension gap is $224K^4\log K$, which is exponential in the "metric dimension" $\log K$ of the given domain X. Note that a dimension gap of $\Omega(\log K)$ is necessary: take for example $X=S^{2n}$, $D=2n+1$, $m=1$, and $v_1 : S^{2n}\hookrightarrow\mathbb{R}^{2n+1}$ the standard inclusion. Then the log of the doubling constant for X is a universal constant multiple of n, and the dimension gap is $D-m-1=2n-1$. However, the high-dimensional hairy ball theorem tells us that no continuous orthonormal extension is possible, let alone a Lipschitz orthonormal extension.*

Perhaps one could reduce the dimension gap in Theorem 1.3.3, say by randomizing the proof using the machinery of random nets and partitions and their padding properties as in [61], though reducing it to somewhere near $\Omega(\log K)$ seems to require more effort (or perhaps it is impossible, but we do not have a counterexample yet).

We now state and prove a version of Theorem 1.3.2 but for general simply connected nilpotent Lie groups, using the idea of proof of Theorem 1.3.3.

Theorem 1.3.5. *Let G be a simply connected nilpotent Lie group with doubling constant K. Let $1\leq m\leq D-224K^4\log K$, $j\geq 1$, $R>0$, and let $\{R_i\}_{i=1}^{j}$ be a log-concave sequence of positive reals. Let $v_1,\cdots,v_m : G\to\mathbb{S}^{D-1}$ form an orthonormal system at*

each point, with the uniform regularity bound

$$\sum_{k=1}^{j} R^k R_k \|\nabla^k v_i\|_{C^0} \leq 1, \quad i = 1, \cdots, m.$$

Then there exists $v_{m+1} : G \to \mathbb{S}^{D-1}$ *such that* $v_1, \cdots, v_m$ *along with* v_{m+1} *form an orthonormal system at each point, and*

$$\sum_{k=1}^{j} R^k R_k \|\nabla^k v_{m+1}\|_{C^0} \lesssim_{G,m,j,\{R_k\}} 1.$$

Proof. Since each v_i is $\frac{1}{RR_1}$-Lipschitz, it follows from Theorem 1.3.3 that there exists a $150K^5 m(m+1)/RR_1$-Lipschitz $\tilde{v}_{m+1} : G \to \mathbb{S}^{D-1}$ such that that forms a pointwise orthonormal system along with $v_1, \cdots, v_m$.

Now, we claim that for each $r > 0$ there exists a C^∞-mollifier $\varphi_r : G \to [0, \infty)$, supported on B_r, with unit mass $\int_G \phi(g) dg = 1$, such that

$$\sum_{k=1}^{j} r^k R_k \|\nabla^k \varphi\|_{C^0} \lesssim_{G,j,\{R_k\}} 1.$$

(The important point is that this final bound is independent of r.) Indeed, it is not hard to construct such families for $r \in (0, 1]$ and $r \in [1, \infty)$ that converge to mollifiers on the Carnot groups that are the tangent space and the asymptotic cones; by compactness, the bound is uniform.

Now take v_{m+1} as the result of applying the Gram-Schmidt orthogonalization process to $v_1, \cdots, v_m$ along with $\tilde{v}_{m+1} * \varphi_{RR_1/600K^5 m^2(m+1)}$. The desired bound follows from the algebra properties of rescaled norms. □

From the proof of Theorem 1.3.5, we can easily see that Theorem 1.3.3 provides a partial positive answer to Question 1.3.1 for many other function spaces as well, described in the following theorem.

Theorem 1.3.6. *Let (X, d) be a K-doubling metric space ($K \geq 2$), and let*

$$\mathcal{F}_1 \subset \mathrm{Lip}(X, \mathbb{R}) \coloneqq \{f : X \to \mathbb{R} | f \text{ is Lipschitz}\}$$

and

$$\mathcal{F} \subset \mathrm{Lip}(X, \mathbb{R}^D) \coloneqq \{f : X \to \mathbb{R}^D | f \text{ is Lipschitz}\}$$

be spaces of functions on X such that

1. *The function spaces $(\mathcal{F}_1, \|\cdot\|_{\mathcal{F}_1})$, $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ are normed linear spaces.*

2. *$\|f\|_{\mathrm{Lip}(X,\mathbb{R})} \lesssim \|f\|_{\mathcal{F}_1}$ for all $f \in \mathcal{F}$ and $\|v\|_{\mathrm{Lip}(X,\mathbb{R}^D)} \lesssim \|v\|_{\mathcal{F}}$ for all $v \in \mathcal{F}$.*

3. *(closure under algebraic operations) For $f \in \mathcal{F}_1, v, w \in \mathcal{F}$ we have $fv \in \mathcal{F}$ and $v \cdot w \in \mathcal{F}_1$ with*

$$\|fv\|_{\mathcal{F}} \lesssim \|f\|_{\mathcal{F}_1}\|v\|_{\mathcal{F}}, \quad \|v \cdot w\|_{\mathcal{F}_1} \lesssim \|v\|_{\mathcal{F}}\|w\|_{\mathcal{F}}.$$

   Also, if $f \in \mathcal{F}_1$, $f(x) \geq c > 0$ for all $x \in X$, then $1/f \in \mathcal{F}_1$ with

$$\|1/f\|_{\mathcal{F}_1} \lesssim_c \|f\|_{\mathcal{F}_1}.$$

   If $v \in \mathcal{F}$, $|v(x)| \geq c > 0$ for all $x \in X$, then $|v| \in \mathcal{F}_1$ and

$$\||v|\|_{\mathcal{F}_1} \lesssim_c \|v\|_{\mathcal{F}}.$$

4. *(density in the space of Lipschitz functions) For any $\delta > 0$ and $w \in \mathrm{Lip}(X, \mathbb{R}^D)$ with $\|w\|_{\mathrm{Lip}} \leq 1$ there exists $v \in \mathcal{F}(X)$ such that*

$$\|v - w\|_{L^\infty} < \delta, \quad \|v\|_{\mathcal{F}} \lesssim_\delta \|w\|_{L^\infty} + \|w\|_{\mathrm{Lip}}.$$

Let $1 \leq m \leq D - 224K^4 \log K$. If $v_i : X \to \mathbb{S}^{D-1}$, $\|v_i\|_{\mathcal{F}} \leq 1$, $i = 1, \cdots, m$, form an orthonormal system at each point $p \in X$, then there exists $v_{m+1} : X \to \mathbb{S}^{D-1}$ with $\|v_{m+1}\|_{\mathcal{F}} \lesssim_{K,m,\mathcal{F}} 1$ such that $v_1, \cdots, v_m$ along with v_{m+1} form an orthonormal system at each point $p \in X$.

In words, Theorem 1.3.6 says that as long as the function space $\mathcal{F}$ behaves "better than" the space of Lipschitz functions in the sense of (2), is closed under the operations used in the Gram-Schmidt process in the sense of (3), and can approximate Lipschitz functions arbitrarily well (4) (for simply connected nilpotent Lie groups we accomplished this by convolving with a smooth mollifier), then we have a positive answer to Question 1.3.1.

By applying Theorem 1.3.5 several times, we obtain the following.

Corollary 1.3.7. *Let G be a simply connected nilpotent Lie group with doubling constant K. Let $1 \leq m \leq D - 224K^4 \log K$, $j \geq 1$, $R > 0$, and let $\{R_i\}_{i=1}^{j}$ be a log-concave sequence of positive reals. Let $v_1, \cdots, v_m : G \to \mathbb{S}^{D-1}$ form an orthonormal system at each point, with the uniform regularity bound*

$$\sum_{k=1}^{j} R^k R_k \|\nabla^k v_i\|_{C^0} \leq 1, \quad i = 1, \cdots, m.$$

Then there exist $v_{m+1}, \cdots, v_{m+m'} : G \to \mathbb{S}^{D-1}$ such that $v_1, \cdots, v_m$ along with v_{m+1}, $\cdots$, $v_{m+m'}$ form an orthonormal system at each point, and

$$\sum_{k=1}^{j} R^k R_k \|\nabla^k v_i\|_{C^0} \lesssim_{G,m,m',j,\{R_k\}} 1, \quad i = m+1, \cdots, m+m'.$$

Remark 1.3.8. *The need for a normal field in embedding problems dates back to [66], where the existence of a section was demonstrated using a homotopy argument from [74] based on the fact that the base space is contractible. However, such an argument in this situation will fail to control the regularity of the section at points of the base*

far from the contraction point. In [77, Section 8] this was achieved for the Heisenberg group $\mathbb{H}^3$ by imposing a "uniform" CW-structure on $\mathbb{H}^3$ and then inductively defining the section starting from low-dimensional skeleta. In the inductive step in [77], one has to use the fact that the homotopy groups $\pi_i(\mathbb{S}^n)$ vanish for $i < n$, which necessitates the "dimension gap" $D - m - 1 \geq 3$. Here we have chosen the section over a 0-skeleton such that the section is locally roughly an orthonormal set and obtained a section by directly interpolating. This allows us to avoid the need for a CW structure by only using the doubling property of the base space, but we have thus increased the dimension gap exponentially.

1.4 Main Iteration Lemma

The starting point of the iterative construction is a function that oscillates at a fixed scale while satisfying a suitable freeness property. (This will also be an ingredient in the inductive step of the iterative construction, when we pass from a larger scale A^{m+1} down to a smaller scale A^m.) We begin by constructing this single oscillating function on the simply connected nilpotent Lie group G. (See also Proposition 5.2 of [77] for an analogous statement.)

Recall that K is the doubling constant of G, and recall the definitions of $X_{r,i}^{loc}$, $f_{r,i}^{loc}$, $x_{r,i}^{loc}$, $X_{r,i}^{glob}$, $f_{r,i}^{glob}$, $x_{r,i}^{glob}$, from subsection 1.1.6.

Proposition 1.4.1. *There exists a dimension $d = O_G(1)$ with the following property. For each dyadic $R > 0$, there exists a smooth map $\phi^R : G \to \mathbb{R}^d$ with the following properties.*

1. (smoothness) For all $j \geq 1$,

$$\|\phi^R\|_{C_R^j} \lesssim_{G,j} R.$$

2. (locally free embedding) If $R \leq 1$ (local case), then for all $p \in G$, we have

$$\left| \bigwedge_{\substack{W = X^{loc}_{r_1,j_1} \cdots X^{loc}_{r_m,j_m} \\ 1 \leq m \leq s,\ 1 \leq j_i \leq k^{loc}_{r_i}, \\ (r_1,j_1) \preceq^{loc} \cdots \preceq^{loc} (r_m,j_m)}} W\phi^R(p) \right| \geq \prod_{\substack{1 \leq m \leq s,\ 1 \leq j_i \leq k^{loc}_{r_i}, \\ (r_1,j_1) \preceq^{loc} \cdots \preceq^{loc} (r_m,j_m)}} R^{1-\sum_{i=1}^m r_i}. \tag{1.71}$$

If $R \geq 1$ (global case), then for all $p \in G$, we have

$$\left| \bigwedge_{\substack{W = X^{glob}_{r_1,j_1} \cdots X^{glob}_{r_m,j_m} \\ 1 \leq m \leq s,\ 1 \leq j_i \leq k^{glob}_{r_i}, \\ (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)}} W\phi^R(p) \right| \geq \prod_{\substack{1 \leq m \leq s,\ 1 \leq j_i \leq k^{glob}_{r_i}, \\ (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)}} R^{1-\sum_{i=1}^m r_i}. \tag{1.72}$$

Proof. Inspired by the Veronese-type embedding used in [77], in the global case $R \geq 1$ we first begin with the function

$$\varphi^R : G \to \bigoplus_{m=1}^{s} \otimes^m \mathbb{R}^n = \mathbb{R}^{\sum_{m=1}^s \binom{n}{m}},$$

defined by

$$\varphi^R \left(\exp \left(\sum_{r=1}^{s} \sum_{i=1}^{k_r} x^{glob}_{r,i} X^{glob}_{r,i} \right) \right) = \bigoplus_{m=1}^{s} \frac{R}{m!} \otimes^m \left(\sum_{r=1}^{s} \sum_{i=1}^{k_r} \frac{x^{glob}_{r,i}}{R^r} f^{glob}_{r,i} \right) \quad x^{glob}_{r,i} \in \mathbb{R}.$$

Recall that

$$X^{glob}_{r,i}$$
$$= \frac{\partial}{\partial x_{r,i}} + \sum_{r'>r}^{s} \sum_{j=1}^{k_{r'}} (\text{polyn in } \{x^{glob}_{r'',i'}\}_{r''<r'} \text{ of weighted global degree} \leq r - r') \frac{\partial}{\partial x^{glob}_{r',j}},$$

and recall that $X^{glob}_{r,i}$ acts on polynomials by reducing the weighted global degree

by r or more. Consequently, for each $(r_1, j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m, j_m)$, the differential operator $X^{glob}_{r_1,j_1} \cdots X^{glob}_{r_m,j_m}$ reduces weighted degrees by $\sum_{i=1}^m r_i$, and so

$$\begin{aligned}
&X^{glob}_{r_1,j_1} \cdots X^{glob}_{r_m,j_m} \varphi^R \\
&= R^{1-\sum_{i=1}^m r_i} f^{glob}_{r_1,j_1} \otimes \cdots \otimes f^{glob}_{r_m,j_m} \\
&\quad + \sum_{\substack{(r'_1,j'_1) \preceq^{glob} \cdots \preceq^{glob} (r'_{m'},j'_{m'}) \\ r'_1+\cdots+r'_{m'} > r_1+\cdots+r_m \\ m' \le s}} (\text{polyn of global degree } \le \Sigma r'_i - \Sigma r_i) f^{glob}_{r'_1,j'_1} \otimes \cdots \otimes f^{glob}_{r'_{m'},j'_{m'}}.
\end{aligned}$$

Thus, when taking the wedge product of all these differentials, we can rearrange the differentials in order of their degree and cancel the higher degree terms. This leads to

$$\begin{aligned}
&\left| \bigwedge_{\substack{W = X^{glob}_{r_1,j_1} \cdots X^{glob}_{r_m,j_m} \\ 1 \le m \le s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)}} W\varphi^R(p) \right| \\
&= \left| \bigwedge_{1 \le m \le s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)} R^{1-\sum_{i=1}^m r_i} f^{glob}_{r_1,j_1} \otimes \cdots \otimes f^{glob}_{r_m,j_m} \right| \\
&= \prod_{1 \le m \le s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)} R^{1-\sum_{i=1}^m r_i}.
\end{aligned}$$

In the local case $R \le 1$ we proceed similarly but we need to work with global degrees instead of local degrees, as the global degree structure gives the desired upper-triangular structure. Again, we begin with the function

$$\varphi^R : G \to \bigoplus_{m=1}^s \otimes^m \mathbb{R}^n = \mathbb{R}^{\sum_{m=1}^s \binom{n}{m}},$$

defined by

$$\varphi^R \left(\exp \left(\sum_{r=1}^s \sum_{i=1}^{k_r} x^{loc}_{r,i} X^{loc}_{r,i} \right) \right) = \bigoplus_{m=1}^s \frac{R}{m!} \otimes^m \left(\sum_{r=1}^s \sum_{i=1}^{k_r} \frac{x^{loc}_{r,i}}{R^r} f^{loc}_{r,i} \right) \quad x^{loc}_{r,i} \in \mathbb{R}.$$

Recall that $X^{loc}_{r,i}$, $f^{loc}_{r,i}$, $x^{loc}_{r,i}$ can be chosen to be permutations of the objects $X^{glob}_{r,i}$, $f^{glob}_{r,i}$, $x^{glob}_{r,i}$. Thus,

$$\begin{aligned}
&X^{loc}_{r_1,j_1} \cdots X^{loc}_{r_m,j_m} \varphi^0 \\
&= R^{1-\sum_{i=1}^m r_i} f^{loc}_{r_1,j_1} \otimes \cdots \otimes f^{loc}_{r_m,j_m} \\
&\quad + \sum_{\substack{(r'_1,j'_1) \preceq^{glob} \cdots \preceq^{glob} (r'_{m'},j'_{m'}) \\ r'_1+\cdots+r'_{m'} > r_1+\cdots+r_m \\ m' \leq s}} (\text{polyn of global degree} \leq \Sigma r'_i - \Sigma r_i) f^{loc}_{r'_1,j'_1} \otimes \cdots \otimes f^{loc}_{r'_{m'},j'_{m'}}.
\end{aligned}$$

Using this upper-triangular structure, we again obtain

$$\begin{aligned}
&\left| \bigwedge_{\substack{W = X^{loc}_{r_1,j_1} \cdots X^{loc}_{r_m,j_m} \\ 1 \leq m \leq s, (r_1,j_1) \preceq^{loc} \cdots \preceq^{loc} (r_m,j_m)}} W\varphi^R(p) \right| \\
&= \left| \bigwedge_{1 \leq m \leq s, (r_1,j_1) \preceq^{loc} \cdots \preceq^{loc} (r_m,j_m)} R^{1-\sum_{i=1}^m r_i} f^{loc}_{r_1,j_1} \otimes \cdots \otimes f^{loc}_{r_m,j_m} \right| \\
&= \prod_{1 \leq m \leq s, (r_1,j_1) \preceq^{loc} \cdots \preceq^{loc} (r_m,j_m)} R^{1-\sum_{i=1}^m r_i}.
\end{aligned}$$

Hence, in both the local and global cases, we can create a mapping with the required freeness property. It remains to modify this construction so that we have bounded C^j_R norms as well. This can be done by viewing φ^0 as a mapping that works locally around the origin and then pasting together mollifications of φ^0, using the doubling property of G to ensure that the pasting process only increases the C^j norm by a bounded factor. (This construction is inspired by [3].)

Take a smooth function $\eta_R : G \to [0,1]$ which is identically 1 on the unit ball B_R, which vanishes on $B^c_{1.5R}$, and such that

$$\|\eta_R\|_{C^j_R} \lesssim_{G,R} 1.$$

Then the function $\eta_R \varphi^R : G \to \mathbb{R}^{\sum_{r=1}^{s} \binom{n}{r}}$ has bounded C_R^j norm for all $j \geq 1$, vanishes outside $B_{1.5R}$, and, if $R \geq 1$, satisfies

$$\left| \bigwedge_{W = X^{glob}_{r_1,j_1}, \cdots, X^{glob}_{r_m,j_m}} R^{\sum_i r_i - 1} W \varphi^1 \right| = 1 \quad \text{on } B_R,$$

and if $R \leq 1$, satisfies

$$\left| \bigwedge_{W = X^{loc}_{r_1,j_1}, \cdots, X^{loc}_{r_m,j_m}} R^{\sum_i r_i - 1} W \varphi^1 \right| = 1 \quad \text{on } B_R.$$

Now take a maximal R-net $\mathcal{N}_R$ of G. We claim that we can take the decomposition

$$\mathcal{N}_R = \bigsqcup_{a=1}^{O_G(1)} \mathcal{N}_{3R}^a,$$

where each $\mathcal{N}_{3R}^a$ is a $3R$-net of G. (This is a standard coloring argument.) Indeed, each point $g \in \mathcal{N}_R$ has at most $O_G(1)$ points in $\mathcal{N}_R$ in its $3R$-neighborhood by (1.9); having inductively assigned a finite number of points of $\mathcal{N}_R$ in one of the $O_G(1)$ sets $\{\mathcal{N}_{3R}^a\}_{a=1}^{O_G(1)}$, any other point of $\mathcal{N}_R$ can be assigned to one of them consistently.

We now define $\phi^R : G \to \mathbb{R}^{O_G(1) \sum_{r=1}^{s} \binom{n}{r}}$ as

$$\phi^R(p) := \bigoplus_{a=1}^{O_G(1)} \sum_{g \in \mathcal{N}_{3R}^a} (\eta_R \varphi^R)(g^{-1}p), \quad p \in G.$$

Then this satisfies the given properties because for each $a = 1, \cdots, O_G(1)$, the function

$$\sum_{g \in \mathcal{N}_{3R}^a} (\eta_R \varphi^R)(g^{-1}p)$$

is a sum of smooth compactly supported functions whose supports $gB_{1.5R}$ are disjoint,

thus has bounded C^j_R norm, and satisfies the wedge product bound

$$\left| \bigwedge_{\substack{W = X^{glob}_{r_1,j_1} \cdots X^{glob}_{r_m,j_m} \\ 1 \leq m \leq s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)}} R^{\sum_i r_i - 1} W \sum_{g \in \mathcal{N}^a_{3R}} (\eta_R \varphi^R)(g^{-1}p) \right| \geq 1, \quad p \in \mathcal{N}^a_{3R} B_R$$

if $R \geq 1$ and

$$\left| \bigwedge_{\substack{W = X^{loc}_{r_1,j_1} \cdots X^{loc}_{r_m,j_m} \\ 1 \leq m \leq s, (r_1,j_1) \preceq^{loc} \cdots \preceq^{loc} (r_m,j_m)}} R^{\sum_i r_i - 1} W \sum_{g \in \mathcal{N}^a_{3R}} (\eta_R \varphi^R)(g^{-1}p) \right| \geq 1, \quad p \in \mathcal{N}^a_{3R} B_R$$

if $R \leq 1$. As $\{\mathcal{N}^a_3 B_1\}_{a=1}^{O_G(1)}$ covers G, we see that φ^0 satisfies (1.71). □

Remark 1.4.2. *The method of proof of Proposition 1.4.1 in [77] for the case $G = \mathbb{H}^3$, or more generally for the case when G admits a cocompact lattice Γ, is the following. Since the nilmanifold G/Γ is a smooth compact n-dimensional manifold, by the strong Whitney immersion theorem [81] there exists a smooth immersion $G/\Gamma \to \mathbb{R}^{2n-1}$. By precomposing with the projection $G \to G/\Gamma$, one obtains a map $\varphi^1 : G \to \mathbb{R}^{2n-1}$ that satisfies the weaker freeness property*

$$\left| \bigwedge_{r=1}^{s} \bigwedge_{i=1}^{k_r} X_{r,i} \varphi^1 \right| \gtrsim 1$$

while having bounded C^j norms due to the compactness of G/Γ. We can then obtain the stronger freeness property (1.71) *by composing φ^1 with a Veronese-type embedding, say*

$$\phi^1 : G \to \bigoplus_{r=1}^{s} \otimes^r \mathbb{R}^{2n-1} = \mathbb{R}^{\sum_{r=1}^{s} \binom{2n-1}{r}}, \quad \phi^1 := \bigoplus_{r=1}^{s} \frac{1}{r!} (\varphi^1)^{\otimes r}.$$

(One can indeed prove the stronger freeness property by using a simple change of coordinates argument.) There is an exponential saving in the target dimension: the

target dimension, in this case, is polynomial in the topological dimension of G, whereas the target dimension d in Proposition 1.4.1 is exponential in the Hausdorff dimension of G. Perhaps one could improve the target dimension in Proposition 1.4.1 to be polynomial in the Hausdorff dimension of G, say by using random nets and partitions as in [61] and being more careful about how we paste the different embeddings.

Now we state the inductive step. Having constructed a map $\psi : G \to \mathbb{R}^D$ which "represents" the geometry of G at scale AR and above, where $R > 0$ is a dyadic number, we need to construct a correction $\phi : G \to \mathbb{R}^D$ which oscillates at scale R such that $\psi + \phi$ represents the geometry of G at scale R. When we say that a mapping represents the geometry of G at scale R, we mean that its $C_R^{m^*,2/3}$ norm is controlled and that it has good freeness properties. To make a viable induction argument, we need to show that the quantitative controls on the $C_R^{m^*,2/3}$ norm and the freeness properties are preserved when we pass from ψ to $\psi + \phi$. The precise statement for the inductive step is as follows.

Proposition 1.4.3 (main iteration step – global version)**.** *For all sufficiently large dimensions $D = \Omega_G(1)$ the following holds. Let M be a real number with*

$$M \geq C_0^{-1},$$

let $R \geq 1$ be a dyadic number, and let $m^* \geq s^2$. Suppose a map $\psi : G \to \mathbb{R}^D$ obeys the following estimates:

1. (nondegenerate first derivatives) For any $p \in G$, we have

$$C_0^{-1} M \leq |X_i^{glob} \psi(p)| \leq C_0 M, \quad i = 1, \cdots, k_1^{glob}, \tag{1.73}$$

and

$$\left| \bigwedge_{m=1}^{i} X_{j_m}^{glob} \psi(p) \right| \geq C_0^{-i^2-i+2} \prod_{m=1}^{i} \left| X_{j_m}^{glob} \psi(p) \right| \tag{1.74}$$

for $2 \leq i \leq k_1^{glob}$ *and* $1 \leq j_1 < \cdots < j_i \leq k_1^{glob}$.

2. *(locally free embedding) For any* $p \in G$, *we have*

$$\begin{aligned} &\left| \bigwedge_{\substack{W = X^{glob}_{r_1,j_1} \cdots X^{glob}_{r_m,j_m} \\ 1 \leq m \leq s, (r_1,j_1) \preceq^{glob} \ldots \preceq^{glob} (r_m,j_m)}} W\psi(p) \right| \\ &\geq C_0^{-2(k_1^{glob})^2 - 7k_1^{glob} + 2} \prod_{2 \leq m \leq s, (r_1,j_1) \preceq^{glob} \ldots \preceq^{glob} (r_m,j_m)} (AR)^{1 - \sum_{i=1}^m r_i} \\ &\quad \cdot \prod_{j=1}^{k_1^{glob}} \left| X_j^{glob} \psi(p) \right|. \end{aligned} \tag{1.75}$$

3. *(Hölder regularity at scale A) We have*

$$\|\nabla^2 \psi\|_{C^{m^*-2,2/3}_{AR}} \leq \frac{C_0}{AR}. \tag{1.76}$$

Then there exists a map $\phi : G \to \mathbb{R}^D$ *such that* ϕ *obeys the following estimates:*

1. *(regularity at scale 1) We have*

$$\|\phi\|_{C^{m^*,2/3}_R} \lesssim_G R. \tag{1.77}$$

2. *(orthogonality) We have*

$$B^{glob}(\phi, \psi) = 0. \tag{1.78}$$

3. *(nondegenerate first derivatives) We have*

$$|X_i^{glob} \phi| \gtrsim_G 1, \quad i = 1, \cdots, k_1^{glob}. \tag{1.79}$$

and the sum $\psi + \phi$ *obeys the following regularity estimates:*

1. *(nondegenerate first derivatives) For any $p \in G$, we have*

$$C_0^{-1}\sqrt{M^2+1} \leq |X_i^{glob}(\psi+\phi)(p)| \leq C_0\sqrt{M^2+1}, \quad i = 1, \cdots, k_1^{glob}, \tag{1.80}$$

and

$$\left| \bigwedge_{m=1}^{i} X_{j_m}^{glob}(\psi+\phi)(p) \right| \geq C_0^{-i^2-i+2} \prod_{m=1}^{i} \left| X_{j_m}^{glob}(\psi+\phi)(p) \right| \tag{1.81}$$

for $2 \leq i \leq k_1^{glob}$ and $1 \leq j_1 < \cdots < j_i \leq k_1^{glob}$.

2. *(locally free embedding) For any $p \in G$, we have*

$$\begin{aligned} &\left| \bigwedge_{\substack{W = X_{r_1,j_1}^{glob} \cdots X_{r_m,j_m}^{glob} \\ 1 \leq m \leq s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)}} W(\psi+\phi)(p) \right| \\ &\geq C_0^{-2(k_1^{glob})^2 - 7k_1^{glob}+3} \prod_{2 \leq m \leq s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)} (AR)^{1-\sum_{i=1}^m r_i} \\ &\quad \cdot \prod_{j=1}^{k_1^{glob}} \left| X_j^{glob}(\psi+\phi)(p) \right|. \end{aligned} \tag{1.82}$$

3. *(Hölder regularity at scale 1) We have*

$$\|\nabla^2(\psi+\phi)\|_{C_R^{m^*-2,2/3}} \leq \frac{C_0}{R}. \tag{1.83}$$

Remark 1.4.4. *The largeness of the dimension D will be necessitated in Proposition 1.4.1 along with Corollary 1.3.7; see Lemma 1.4.7 below; D will be polynomial in the doubling constant of G.*

Remark 1.4.5. *Proposition 1.4.3 shows that if ψ is a map with the regularity and freeness properties* (1.73)-(1.76) *at scale AR, then we can find a correction ϕ so that $\psi + \phi$ is a map with the same regularity and freeness properties* (1.80)-(1.83) *but at scale R. In particular, the constants in the freeness properties are the same. We*

will make this possible by creating a "hierarchy" of freeness properties. Namely, the freeness property (1.81) *for i-fold wedge products of horizontal derivatives of $\psi+\phi$ will be based on the freeness property* (1.74) *for $(i-1)$-fold wedge products of horizontal derivatives of ψ. Also, the freeness property* (1.82) *for the wedge product of up to s-order derivatives of $\psi+\phi$ will be based on the freeness property* (1.74) *for k_1^{glob}-fold wedge products of horizontal derivatives of ψ. Thus, we do not lose constants when passing from ψ to $\psi+\phi$, which will allow us to close the iteration.*

Proposition 1.4.3 will be a consequence of the following lemma, which is a generalization of Proposition 5.1 of [77] for the case $G = \mathbb{H}^3$.

Lemma 1.4.6 (main iteration lemma – global version)**.** *For all sufficiently large dimensions $D = \Omega_G(1)$ the following holds. Let M be a real number with*

$$M \geq C_0^{-1},$$

let $R \geq 1$ be a dyadic number, and let $m^* \geq s^2$. Suppose a map $\psi : G \to \mathbb{R}^D$ obeys the following estimates:

1. (nondegenerate first derivatives) For any $p \in G$, we have

$$C_0^{-1} M \leq |X_i^{glob}\psi(p)| \leq C_0 M, \quad i = 1, \cdots, k_1^{glob}, \tag{1.84}$$

and

$$\left| \bigwedge_{m=1}^{i} X_{j_m}^{glob}\psi(p) \right| \geq C_0^{-i^2-2i+2} M^i, \tag{1.85}$$

for $2 \leq i \leq k_1^{glob}$ and $1 \leq j_1 < \cdots < j_i \leq k_1^{glob}$.

2. *(locally free embedding) For any $p \in G$, we have*

$$\left| \bigwedge_{\substack{W = X^{glob}_{r_1,j_1} \cdots X^{glob}_{r_m,j_m} \\ 1 \le m \le s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)}} (AR)^{\sum_{i=1}^m r_i - 1} W\psi(p) \right| \ge C_0^{-2(k_1^{glob})^2 - 8k_1^{glob} + 2} M^k. \quad (1.86)$$

3. *(Hölder regularity at scale A) We have*

$$\|\nabla^2 \psi\|_{C^{m^*-2,\alpha}_{AR}} \le \frac{C_0}{AR}. \quad (1.87)$$

Then there exists a map $\phi : G \to \mathbb{R}^D$ obeying the following estimates.

1. *(nondegenerate first derivatives) For any $p \in G$, we have*

$$|X_i^{glob} \phi(p)| \gtrsim_G 1 \quad (1.88)$$

and

$$\left| \bigwedge_{m=1}^i X^{glob}_{j_m} (\psi + \phi)(p) \right|^2 - \left| \bigwedge_{m=1}^i X^{glob}_{j_m} \psi(p) \right|^2 \ge C_0^{-2i^2+5} M^{2(2i-1)}, \quad (1.89)$$

for $2 \le i \le k_1^{glob}$ and $1 \le j_1 < \cdots < j_i \le k_1^{glob}$.

2. *(locally free embedding) For any $p \in G$, we have*

$$\left| \bigwedge_{\substack{W = X^{glob}_{r_1,j_1} \cdots X^{glob}_{r_m,j_m} \\ 1 \le m \le s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)}} R^{\sum_{i=1}^m r_i - 1} W(\psi + \phi)(p) \right| \ge C_0^{-2(k_1^{glob})^2 - 5k_1^{glob} + 3} M^{k_1^{glob}}. \quad (1.90)$$

3. (Hölder regularity at scale 1) We have

$$\|\phi\|_{C_R^{m^*,\alpha}} \lesssim_G R. \tag{1.91}$$

4. (orthogonality) We have

$$B^{glob}(\phi, \psi) = 0. \tag{1.92}$$

We now show why Proposition 1.4.3 follows from Lemma 1.4.6.

proof of Proposition 1.4.3 assuming Lemma 1.4.6. Let ψ be as in Proposition 1.4.3. One can easily verify the hypotheses of Lemma 1.4.6 for ψ, as (1.85) and (1.86) each follow from (1.74) and (1.75) combined with (1.73). Thus, by Lemma 1.4.6, there exists a function $\phi : G \to \mathbb{R}^D$ that satisfies (1.77), (1.78), and (1.79) (as these are exactly (1.91), (1.92), and (1.88), respectively) and the following (which are just restatements of (1.90) and (1.91)):

1. (nondegenerate first derivatives) For $p \in G$, $2 \leq i \leq k_1^{glob}$ and $1 \leq j_1 < \cdots < j_i \leq k_1^{glob}$,
$$\left| \bigwedge_{q=1}^{i} X_{j_q}^{glob}(\psi + \phi)(p) \right|^2 - \left| \bigwedge_{q=1}^{i} X_{j_q}^{glob} \psi(p) \right|^2 \geq C_0^{-2i^2+5} M^{2(i-1)}. \tag{1.93}$$

2. (locally free embedding) For any $p \in G$, we have
$$\left| \bigwedge_{\substack{W = X_{r_1,j_1}^{glob} \cdots X_{r_m,j_m}^{glob} \\ 1 \leq m \leq s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)}} R^{\sum_{i=1}^m -1} W(\psi + \phi)(p) \right| \geq C_0^{-2(k_1^{glob})^2 - 5k_1^{glob} + 3} M^{k_1^{glob}}. \tag{1.94}$$

We now verify that the map $\psi + \phi$ satisfies the properties (1.80) to (1.83).

Verification of (1.80): By (1.78), we have $|X_i^{glob}(\psi + \phi)|^2 = |X_i^{glob}\psi|^2 + |X_i^{glob}\phi|^2$. But

by (1.79) and (1.77), we have $|X_i^{glob}\phi| \sim_G 1$, so we have

$$C_0^{-1} \leq |X_i^{glob}\phi| \leq C_0. \tag{1.95}$$

(This uses our hierarchy of constants in subsection 1.1.2, by choosing C_0 depending on G.) Combining these facts with (1.73) we obtain (1.80).

Verification of (1.81): This follows from (1.93):

$$\begin{aligned}
\left|\bigwedge_{q=1}^{i} X_{j_q}^{glob}(\psi+\phi)\right|^2 &\geq \left|\bigwedge_{q=1}^{i} X_{j_q}^{glob}\psi\right|^2 + C_0^{-2i^2+5}M^{2(i-1)} \\
&\geq C_0^{-2i^2-2i+4}\prod_{q=1}^{i}\left|X_{j_q}^{glob}\psi\right|^2 + C_0^{-2i^2+5}M^{2(i-1)} \\
&\geq C_0^{-2i^2-2i+4}\prod_{q=1}^{i}\left(\left|X_{j_q}^{glob}\psi\right|^2 + |X_{j_q}^{glob}\phi|^2\right) \\
&= C_0^{-2i^2-2i+4}\prod_{q=1}^{i}|X_{j_q}^{glob}(\psi+\phi)|^2
\end{aligned} \tag{1.96}$$

where in the third inequality we used $|X_{j_q}^{glob}\psi| \leq C_0 M$ and $|X_{j_q}^{glob}\phi| \lesssim_G 1$.

Verification of (1.82): This follows from (1.94):

$$\begin{aligned}
&\left|\bigwedge_{\substack{W=X_{r_1,j_1}^{glob}\cdots X_{r_m,j_m}^{glob} \\ 1\leq m\leq s,(r_1,j_1)\preceq^{glob}\cdots\preceq^{glob}(r_m,j_m)}} R^{\sum_{i=1}^m r_i-1}W(\psi+\phi)(p)\right| \\
&\geq C_0^{-2(k_1^{glob})^2-5k_1^{glob}+3}M^{k_1^{glob}} \\
&\geq C_0^{-2(k_1^{glob})^2-7k_1^{glob}+3}\prod_{i=1}^{k_1^{glob}}|X_i^{glob}(\psi+\phi)(p)|,
\end{aligned} \tag{1.97}$$

where in the last inequality we used that $|X_i^{glob}(\psi+\phi)|^2 = |X_i^{glob}\psi|^2 + |X_i^{glob}\phi|^2 \leq C_0^2M^2 + O_G(1) \leq C_0^4M^2$ (recall that $M \geq C_0^{-1}$), and our hierarchy of constants where

we choose C_0 depending on G.

Verification of (1.83): This follows from (1.76) and (1.77):

$$\|\nabla^2(\psi+\phi)\|_{C_R^{m^*-2,2/3}} \leq \|\nabla^2\psi\|_{C_{AR}^{m^*-2,2/3}} + \|\phi\|_{C_R^{m^*,2/3}} \leq \frac{C_0}{R}, \tag{1.98}$$

and our hierarchy of constants where we choose C_0 depending on G. □

For the rest of the section, we will prove Lemma 1.4.6.

Suppose that ψ is as in Lemma 1.4.6. We will first construct a solution $\tilde{\phi}$ to the low-frequency equation (1.27) as

$$\tilde{\phi}(p) = U(p)\left(\phi^R(p)\right), \tag{1.99}$$

where $\phi^0 : G \to \mathbb{R}^d$ is as in Proposition 1.4.1, and $U(p) : \mathbb{R}^d \to \mathbb{R}^D$ is a linear isometry with the following properties, which is constructed using Corollary 1.3.7. (See also Lemma 9.1 of [77] for an analogous statement for $\mathbb{H}^3$.)

Lemma 1.4.7. *Let $d = O_G(1)$. Then for $D = d + \Omega(1)$, there exists $U : G \to \hom(\mathbb{R}^d, \mathbb{R}^D)$ such that*

1. *For each $p \in G$, $U(p) \in \hom(\mathbb{R}^d, \mathbb{R}^D)$ is an isometry.*

2. *For each $p \in G$, $s \in \mathbb{R}^d$, we have, for $1 \leq i, j \leq k_1^{glob}$,*

$$(U(p)(s))\cdot X_i^{glob} P_{(\leq N_0/R)}\psi(p) = 0,\ (U(p)(s))\cdot X_i^{glob} X_j^{glob} P_{(\leq N_0/R)}\psi(p) = 0. \tag{1.100}$$

3. *We have the smoothness*

$$\|\nabla U\|_{C_R^{2m^*}} \lesssim_{N_0} \frac{1}{AR}.$$

Proof. Let $\{Y_{r,i}\}_{1\leq i\leq \ell_r}$ be a subset of $\{X_{r,i}^{glob}\}_{1\leq i\leq k_r^{glob}}$, for $2 \leq r \leq s$, such that

$\{X_i^{glob}\}_{i=1}^{k_1^{glob}} \cup \{Y_{r,i}\}_{2\le r\le s,\ 1\le i\le \ell_r}$ forms a basis of

$$\mathrm{span}\{X_i^{glob}\}_{i=1}^{k_1^{glob}} + \{[X_i^{glob}, X_j^{glob}]\}_{i,j=1}^{k_1^{glob}}.$$

Let $W_1, \cdots, W_{k_1^{glob}+\frac{k_1^{glob}(k_1^{glob}+3)}{2}+\ell_2+\cdots+\ell_s}$ denote the rescaled differential operators

$$((M^{-1}X_i^{glob})_{i=1}^{k_1^{glob}}, (ARX_i^{glob}X_j^{glob})_{1\le i\le j\le k_1^{glob}}, ((AR)^{r-1}Y_{r,i})_{2\le r\le s,\ 1\le i\le \ell_r}).$$

Let $w_i(p) \coloneqq W_i P_{(\le N_0/R)}\psi(p)$ for $1 \le i \le k_1^{glob} + \frac{k_1^{glob}(k_1^{glob}+1)}{2} + \ell_2 + \cdots + \ell_s$ for all $p \in G$. Then, by (1.84), (1.87), and Theorem 1.1.2 (3) (more specifically, (1.13) with $j = 1$, $l = 2$, and recalling $m^* \ge s^2$),

$$w_i = W_i\psi + O_G\left(\frac{C_0}{AN_0M}\right) = O(C_0), \quad i = 1, \cdots, k_1^{glob}, \tag{1.101}$$

and by (1.87) and Theorem 1.1.2 (3) (more specifically, (1.13) with $j \ge 2$, $l = j+1$),

$$w_i = W_i\psi + O_G\left(\frac{C_0}{AN_0}\right) = O(C_0) \tag{1.102}$$

for $i = k_1^{glob} + 1, \cdots, k_1^{glob} + \frac{k_1^{glob}(k_1^{glob}+1)}{2} + \ell_2 + \cdots + \ell_s$. (Note that we have used our hierarchy of constants in (1.101) and (1.102), choosing N_0 depending on G and C_0, while we have only used $A \ge 1$ so far.)

On the other hand, (1.86) tells us that

$$\left|\bigwedge_{i=1,\cdots,k_1^{glob}+\frac{k_1^{glob}(k_1^{glob}+1)}{2}+\ell_2+\cdots+\ell_s} W_i\psi(p)\right| \gtrsim_{C_0} 1,$$

so by applying the triangle inequality and Cauchy–Schwarz inequality, we conclude

that

$$\left| \bigwedge_{i=1,\cdots,k_1^{glob}+\frac{k_1^{glob}(k_1^{glob}+1)}{2}+\ell_2+\cdots+\ell_s} w_i(p) \right| \gtrsim_{C_0} 1.$$

By applying Cauchy–Schwarz again, we conclude that

$$\left| \bigwedge_{i=1,\cdots,j} w_i(p) \right| \sim_{C_0} 1, \quad j = 1, \cdots, k_1^{glob} + \frac{k_1^{glob}(k_1^{glob}+1)}{2} + \ell_2 + \cdots + \ell_s. \tag{1.103}$$

Again, using (1.87) along with Theorem 1.1.2 (3), we can see that

$$\|w_i\|_{C_0} + AR\|\nabla w_i\|_{C_R^{2m^*-1}} \lesssim_{N_0} 1, \quad j = 1, \cdots, k_1^{glob} + \frac{k_1^{glob}(k_1^{glob}+1)}{2} + \ell_2 + \cdots + \ell_s, \tag{1.104}$$

because for $i = 1, \cdots, k_1^{glob}$, we use (1.87), and (1.12) with $l = 2$, $j = m^*$, to see

$$\begin{aligned} \|w_i\|_{C_0} + AR\|\nabla w_i\|_{C_R^{2m^*-1}} &\overset{(1.101)}{\lesssim} C_0 + \frac{AR}{M}\|\nabla^2 P_{(\leq N_0/R)}\psi\|_{C_R^{2m^*-1}} \\ &\leq C_0 + \frac{ARN_0^{2m^*-1}}{M}\|\nabla^2 P_{(\leq N_0/R)}\psi\|_{C_{R/N_0}^{2m^*-1}} \\ &\overset{(1.12)}{\lesssim_{G,m^*}} C_0 + \frac{ARN_0^{2m^*-1}}{M}\|\nabla^2\psi\|_{C^0} \\ &\overset{(1.87)}{\leq} C_0 + \frac{ARN_0^{m^*}}{M}\cdot\frac{C_0}{AR} \lesssim_{N_0} 1, \end{aligned} \tag{1.105}$$

and for $i = k_1^{glob} + 1, \cdots, k_1^{glob} + \frac{k_1^{glob}(k_1^{glob}+1)}{2} + \ell_2 + \cdots + \ell_s$, we use (1.87), and (1.12) with $l = 3$, $j = m^*$, with w_i corresponding to $2 \leq r \leq s$, to see

$$\begin{aligned} \|w_i\|_{C_0} + AR\|\nabla w_i\|_{C_R^{2m^*-1}} &\overset{(1.102)}{\lesssim} C_0 + (AR)^r\|\nabla^{r+1} P_{(\leq N_0/R)}\psi\|_{C_R^{2m^*-1}} \\ &\leq C_0 + (AR)^r N_0^{2m^*-1}\|\nabla^{r+1} P_{(\leq N_0/R)}\psi\|_{C_{R/N_0}^{2m^*-1}} \\ &\overset{(1.12)}{\lesssim_{G,m^*}} C_0 + (AR)^r N_0^{2m^*-1}\|\nabla^{r+1}\psi\|_{C^0} \\ &\overset{(1.87)}{\leq} C_0 + (AR)^r N_0^{2m^*-1}\cdot\frac{C_0}{(AR)^r} \lesssim_{N_0} 1. \end{aligned} \tag{1.106}$$

(We have used our hierarchy of constants in (1.104), (1.105), and (1.106), by choosing

N_0 after G and C_0.) Now, because the norm of (1.104) is of the form (1.7), we can apply the product rule to this norm: for instance,

$$\left\| \bigwedge_{i=1,\cdots,j} w_i \right\|_{C_0} + AR \left\| \nabla \bigwedge_{i=1,\cdots,j} w_i \right\|_{C_R^{2m^*-1}} \lesssim_{N_0} 1$$

for $j = 1, \cdots, k_1^{glob} + \frac{k_1^{glob}(k_1^{glob}+1)}{2} + \ell_2 + \cdots + \ell_s$. Now let us consider the orthonormal system $v_1, \cdots, v_{k_1^{glob}+\frac{k_1^{glob}(k_1^{glob}+1)}{2}+\ell_2+\cdots+\ell_s}$ formed by applying the Gram-Schmidt process to the vectors w_i, i.e., we inductively define

$$v_i := \frac{|\bigwedge_{j<i} w_j|}{|\bigwedge_{j\leq i} w_j|} \left(w_i - \sum_{j<i} (w_i \cdot v_j) v_j \right).$$

By (1.103) this is well-defined, and by a repeated application of the aforementioned product rule, one can deduce the smoothness

$$\|v_i\|_{C_0} + AR\|\nabla v_i\|_{C_R^{2m^*-1}} \lesssim_{N_0} 1, \quad i = 1, \cdots, k_1^{glob} + \frac{k_1^{glob}(k_1^{glob}+1)}{2} + \ell_2 + \cdots + \ell_s.$$

We now apply Corollary 1.3.7 with $m = k_1^{glob} + \frac{k_1^{glob}(k_1^{glob}+1)}{2} + \ell_2 + \cdots + \ell_s$, $m' = d$, $j = 2m^*$, $R_i = A$ to the above v_i. (We use our liberty to take $D = d + \Omega_G(1)$ in this step.) Thus, we have maps $v_{k_1^{glob}+\frac{k_1^{glob}(k_1^{glob}+1)}{2}+\ell_2+\cdots+\ell_s+1}, \cdots, v_{k_1^{glob}+\frac{k_1^{glob}(k_1^{glob}+1)}{2}+\ell_2+\cdots+\ell_s+d} : G \to \mathbb{R}^D$ such that

$$\|v_i\|_{C_0} + AR\|\nabla v_i\|_{C_R^{m^*}} \lesssim_{N_0} 1, \quad i = 1, \cdots, k_1^{glob} + \frac{k_1^{glob}(k_1^{glob}+1)}{2} + \ell_2 + \cdots + \ell_s + d$$

and such that $v_1(p), \cdots, v_{k_1^{glob}+\frac{k_1^{glob}(k_1^{glob}+1)}{2}+\ell_2+\cdots+\ell_s+d}(p)$ are orthonormal for all $p \in G$.

Now define $U(p) : \mathbb{R}^d \to \mathbb{R}^D$, $p \in G$, to be the map

$$U(p)(s) = \sum_{i=1}^{d} s_i v_{k_1^{glob}+\frac{k_1^{glob}(k_1^{glob}+1)}{2}+\ell_2+\cdots+\ell_s+i}(p).$$

This clearly has the properties (1) and (3) asserted above, and we can also deduce property (2) once we note that

$$X_i^{glob} X_j^{glob} - X_j^{glob} X_i^{glob} \in \text{span}\left(\{X_i^{glob}\}_{i=1}^{k_1^{glob}} \cup \{Y_{r,i}\}_{2 \leq r \leq s,\ 1 \leq i \leq \ell_r} \right).$$

□

Continuing with the proof of Lemma 1.4.6, let $\tilde{\phi}$ be as in (1.99). By Lemma 1.4.7 and Proposition 1.4.1(1), and using our hierarchy of choosing A after C_0 and N_0, we have

$$\|\tilde{\phi}\|_{C^{2m^*}} \lesssim_G R. \tag{1.107}$$

Also, by (1.100) and the Leibniz rule, it is clear that $\tilde{\phi}$ satisfies (1.63) and a fortiori solves the low-frequency equation (1.27). It is also clear that ψ satisfies the hypothesis of Corollary 1.2.4. By applying Corollary 1.2.4, there exists a $C^{m^*,2/3}$-function $\phi : G \to \mathbb{R}^D$ such that

$$\begin{aligned} B(\phi, \psi) &= 0, \\ \|\phi - \tilde{\phi}\|_{C_R^{m^*,2/3}} &\lesssim_{C_0} A^{2-m^*} R, \end{aligned} \tag{1.108}$$

and, as $\tilde{\phi}$ satisfies (1.63), we have for $i, j = 1, \cdots, k_1^{glob}$,

$$\|X_i^{glob}\phi \cdot X_j^{glob}\psi\|_{C^0} = \|X_i^{glob}\phi \cdot X_j^{glob}\psi - X_i^{glob}\tilde{\phi} \cdot X_j^{glob} P_{(\leq N_0/R)}\psi\|_{C^0} \lesssim_{C_0} A^{1-m^*}. \tag{1.109}$$

It remains to verify conditions (1.88)-(1.92). Conditions (1.91) and (1.92) are imme-

diate from the construction. For later use, we note that

$$\begin{aligned} X_i^{glob}\phi &= X_i^{glob}\tilde{\phi} + X_i^{glob}(\phi - \tilde{\phi}) = U(X_i^{glob}\phi^R) + (X_i^{glob}U)\left(\phi^R\right) + O_{C_0}\left(A^{2-m^*}\right) \\ &= U(X_i^{glob}\phi^0) + O_{N_0}\left(A^{-1}\right) \end{aligned} \tag{1.110}$$

for $i = 1, \cdots, k$. From this and Proposition 1.4.1, we immediately have (again using our hierarchy of constants that A is chosen after N_0)

$$|X_i^{glob}\phi| \sim_G 1, \quad \left|\bigwedge_{i=1}^{k_1^{glob}} X_i^{glob}\phi\right| \sim_G 1, \tag{1.111}$$

so in particular (1.88) immediately follows. It remains to verify (1.89) and (1.90).

Verification of (1.89): Recall from (1.109) that

$$|X_i^{glob}\phi \cdot X_j^{glob}\psi| \leq \frac{1}{A} \text{ for } i, j = 1, \cdots, k_1^{glob}.$$

(We have used the fact that A is chosen after C_0.) Now we observe that for $1 \leq j_1 < \cdots < j_i \leq k_1^{glob}$ we have the expansion

$$\bigwedge_{m=1}^{i} X_{j_m}^{glob}(\psi + \phi) = \bigwedge_{m=1}^{i} X_{j_m}^{glob}\psi + \sum_{n=1}^{2^i-1} \bigwedge_{m=1}^{i} X_{j_m}^{glob} f_m^n,$$

for some sequence $\{f_m^n\}_{n=1,\cdots,2^i-1,\ m=1,\cdots,i}$ of functions, each being either ϕ or ψ. Note that for each n there must exist some m such that $f_m^n = \phi$. This expansion implies

$$\begin{aligned} &\left|\bigwedge_{m=1}^{i} X_{j_m}^{glob}(\psi + \phi)\right|^2 - \left|\bigwedge_{m=1}^{i} X_{j_m}^{glob}\psi\right|^2 \\ &= \left|\sum_{n=1}^{2^i-1} \bigwedge_{m=1}^{i} X_{j_m}^{glob} f_m^n\right|^2 + 2\sum_{n=1}^{2^i-1} \left\langle \bigwedge_{m=1}^{i} X_{j_m}^{glob}\psi, \bigwedge_{m=1}^{i} X_{j_m}^{glob} f_m^n \right\rangle. \end{aligned}$$

For each fixed n, the polarized Cauchy–Binet formula shows that

$$\left\langle \bigwedge_{m=1}^{i} X_{j_m}^{glob}\psi, \bigwedge_{m=1}^{i} X_{j_m}^{glob} f_m^n \right\rangle$$

can be represented as the determinant of an $i \times i$ matrix, whose entries are each of magnitude $O_{C_0}(M^2)$ and one of whose columns (the m-th column, where m is such that $f_m^n = \phi$) consists of entries of magnitude $O(A^{-1})$ (because of (1.109)). Thus, the determinant of this $i \times i$ matrix is of magnitude $O_{C_0}(A^{-1}M^{2(i-1)})$, or equivalently $\left\langle \bigwedge_{m=1}^{i} X_{j_m}^{glob}\psi, \bigwedge_{m=1}^{i} X_{j_m}^{glob} f_m^n \right\rangle = O_{C_0}(A^{-1}M^{2(i-1)})$. Summing over all n, we obtain

$$2\sum_{n=1}^{2^i-1} \left\langle \bigwedge_{m=1}^{i} X_{j_m}^{glob}\psi, \bigwedge_{m=1}^{i} X_{j_m}^{glob} f_m^n \right\rangle = O_{C_0}(A^{-1}M^{2(i-1)}),$$

and so (again using our hierarchy of choosing A after C_0) it is enough to show that

$$\left| \sum_{n=1}^{2^i-1} \bigwedge_{m=1}^{i} X_{j_m}^{glob} f_m^n \right|^2 \geq C_0^{-2i^2+5.5} M^{2(i-1)}. \tag{1.112}$$

But by Cauchy–Schwarz,

$$\begin{aligned}
&\left| \sum_{n=1}^{2^i-1} \bigwedge_{m=1}^{i} X_{j_m}^{glob} f_m^n \right|^2 \\
&\asymp \left| \sum_{n=1}^{2^i-1} \bigwedge_{m=1}^{i} X_{j_m}^{glob} f_m^n \right|^2 \left| \bigwedge_{m=2}^{i} X_{j_m}^{glob}\phi \right|^2 \geq \left| \left(\sum_{n=1}^{2^i-1} \bigwedge_{m=1}^{i} X_{j_m}^{glob} f_m^n \right) \wedge \bigwedge_{m=2}^{i} X_{j_m}^{glob}\phi \right|^2 \\
&= \left| \left(X_{j_1}^{glob}\phi \wedge \bigwedge_{m=2}^{i} X_{j_m}^{glob}\psi \right) \wedge \bigwedge_{m=2}^{i} X_{j_m}^{glob}\phi \right|^2 = \left| \bigwedge_{m=2}^{i} X_{j_m}^{glob}\psi \wedge \bigwedge_{m=1}^{i} X_{j_m}^{glob}\phi \right|^2 .
\end{aligned}$$

By the Cauchy–Binet formula, this is the determinant of a certain $(2i-1) \times (2i-1)$ matrix, which, by (1.109), is close to being block-diagonal: the upper-left $(i-1)\times(i-1)$ block consists of entries of size $O_{C_0}(M^2)$, the lower-right $i \times i$ block consists of entries

of size $O_{C_0}(1)$, while the off-block-diagonal entries are of size $O(A^{-1})$ (by (1.109)). Therefore, we may estimate the total determinant with the determinant of the block diagonal approximation, with error $O_{C_0}(A^{-2}M^{2(i-2)})$:

$$\left|\bigwedge_{m=2}^{i} X_{j_m}^{glob}\psi \wedge \bigwedge_{m=1}^{i} X_{j_m}^{glob}\phi\right|^2 = \left|\bigwedge_{m=2}^{i} X_{j_m}^{glob}\psi\right|^2 \left|\bigwedge_{m=1}^{i} X_{j_m}^{glob}\phi\right|^2 + O_{C_0}(A^{-2}M^{2(i-2)}).$$

But by (1.85), we have

$$\left|\bigwedge_{m=2}^{i} X_{j_m}^{glob}\psi\right|^2 \geq C_0^{-2i^2+6}M^{2(i-1)}. \tag{1.113}$$

This completes the verification of (1.89) (again using the hierarchy that C_0 is chosen after G and A is chosen after C_0).

Verification of (1.90): We first observe that for differential operators W of global degree at least 2, $W\phi$ dominates $W\psi$, and so we may approximate $\bigwedge_W W(\psi+\phi)$ by $\bigwedge_W W\phi$, where W ranges over such operators.

More precisely, for $W = X_{r_1,j_1}^{glob}\cdots X_{r_m,j_m}^{glob}$, where $r := \sum_{i=1}^m r_i \geq 2$, $m \leq s$, we have

$$\begin{aligned} W(\psi+\phi) &= W\tilde{\phi} + W(\phi-\tilde{\phi}) + W\psi \\ &= W\Big(U(\phi^0)\Big) + O_{C_0}(A^{2-m^*}R^{1-r}) + O_{C_0}((AR)^{-r+1}) \\ &= W\Big(U(\phi^0)\Big) + O_{C_0}(A^{-1}R^{1-r}). \end{aligned}$$

(The second equation holds because $r \leq s^2$ and our choice of $m^* \geq s^2$ allows us to use our bounds (1.108) and (1.87) on $\|\phi-\tilde{\phi}\|_{C^{m^*}}$ and $\|\nabla^2\psi\|_{C_{AR}^{m^*-2}}$). Many applications of the Leibniz rule tell us that $W\Big(U(\phi^0)\Big) - U(W\phi^0)$ is a linear combination of derivatives of U times ϕ^0 or derivatives of ϕ^0, so from $\|\nabla U\|_{C_R^{m^*}} \lesssim_{N_0} \frac{1}{AR}$, and counting derivatives, we have

$$W\Big(U(\phi^0)\Big) - U(W\phi^0) = O_{N_0}(A^{-1}R^{1-r})$$

and consequently

$$W(\psi+\phi) = U(W\phi^0) + O_{N_0}(A^{-1}R^{1-r}).$$

Therefore, we have

$$\bigwedge_{\substack{W=X^{glob}_{r_1,j_1}\cdots,X^{glob}_{r_m,j_m} \\ m=1 \text{ and } r_1\geq 2, \text{ or } 2\leq m\leq s}} R^{\sum_{i=1}^m r_i-1} W(\psi+\phi) = \omega + O_{N_0}(A^{-1}), \tag{1.114}$$

where

$$\omega \coloneqq \bigwedge_{\substack{W=X^{glob}_{r_1,j_1}\cdots,X^{glob}_{r_m,j_m} \\ m=1 \text{ and } r_1\geq 2, \text{ or } 2\leq m\leq s}} R^{\sum_{i=1}^m r_i-1} U(W\phi^0).$$

Since U is an isometry, and ϕ^0 has the freeness property (1.71), we have

$$|\omega| = \left| \bigwedge_{\substack{W=X^{glob}_{r_1,j_1}\cdots,X^{glob}_{r_m,j_m} \\ m=1 \text{ and } r_1\geq 2, \text{ or } 2\leq m\leq s}} R^{\sum_{i=1}^m r_i-1} W\phi^0 \right| \asymp_G 1. \tag{1.115}$$

Therefore, by (1.84), (1.95), (1.114) and our hierarchy that C_0 is chosen after G and that A is chosen after C_0 and N_0, to verify (1.90) it is enough to verify

$$\left| \bigwedge_{i=1}^{k_1^{glob}} X_i^{glob}(\psi+\phi) \wedge \omega \right| \gtrsim C_0^{-2(k_1^{glob})^2-5k_1^{glob}+3.3} M^{k_1^{glob}}. \tag{1.116}$$

By Cauchy–Schwarz, and

$$\left| \bigwedge_{i=1}^{k_1^{glob}} X_i^{glob} P_{(\leq N_0/R)}\psi \right| \leq \prod_{i=1}^{k_1^{glob}} \left| X_i^{glob} P_{(\leq N_0/R)}\psi \right| \overset{(1.12)}{\lesssim_G} \|\nabla\psi\|_{C^0}^{k_1^{glob}} \leq C_0^{k_1^{glob}} M^{k_1^{glob}},$$

(we used (1.12) with $l=1$), we see (using our hierarchy that C_0 is chosen after G)

that it is enough to verify

$$\left\langle \bigwedge_{i=1}^{k_1^{glob}} X_i^{glob} P_{(\leq N_0/R)}\psi \wedge \omega, \bigwedge_{i=1}^{k_1^{glob}} X_i^{glob}(\psi+\phi)\wedge\omega \right\rangle \gtrsim C_0^{-2(k_1^{glob})^2-4k_1^{glob}+3.7} M^{2k_1^{glob}}. \quad (1.117)$$

Since all the components of ω are orthogonal to the vectors $X_i^{glob} P_{(\leq N_0/R)}\psi$, we can use Cauchy–Binet twice to see that the left-hand side is equal to

$$\left\langle \bigwedge_{i=1}^{k_1^{glob}} X_i^{glob} P_{(\leq N_0/R)}\psi, \bigwedge_{i=1}^{k_1^{glob}} X_i^{glob}(\psi+\phi) \right\rangle |\omega|^2.$$

Using that for $1 \leq i,j \leq k_1^{glob}$ we have

$$\begin{aligned}
&X_i^{glob} P_{(\leq N_0/R)}\psi \cdot X_j^{glob}(\psi+\phi) \\
&= X_i^{glob}\psi \cdot X_j^{glob}\psi - X_i^{glob} P_{(>N_0/R)}\psi \cdot X_j^{glob}(\psi+\phi) + X_i^{glob}\psi \cdot X_j^{glob}\phi \\
&\overset{(1.109)}{=} X_i^{glob}\psi \cdot X_j^{glob}\psi - X_i^{glob} P_{(>N_0/R)}\psi \cdot X_j^{glob}(\psi+\phi) + O(A^{-1}), \\
&\overset{(1.84),(1.95),(1.13)}{=} X_i^{glob}\psi \cdot X_j^{glob}\psi + O_G\left(\frac{C_0}{AN_0}\right)\cdot(O(C_0M)+O(C_0)) + O(A^{-1}), \\
&= X_i^{glob}\psi \cdot X_j^{glob}\psi + O(C_0A^{-1}M),
\end{aligned} \quad (1.118)$$

(where in the penultimate equation we used (1.13) with $j=1$, $l=2$, and in the last equation we used our hierarchy that N_0 is chosen after C_0), we see using Cauchy–Binet twice that

$$\left\langle \bigwedge_{i=1}^{k_1^{glob}} X_i^{glob} P_{(\leq N_0)}\psi, \bigwedge_{i=1}^{k_1^{glob}} X_i^{glob}(\psi+\phi) \right\rangle = \left| \bigwedge_{i=1}^{k_1^{glob}} X_i^{glob}\psi \right|^2 + O_{C_0}(A^{-1}M^{2k_1^{glob}-1}).$$

But, from (1.85), we have

$$\left| \bigwedge_{i=1}^{k_1^{glob}} X_i^{glob}\psi \right|^2 \geq C_0^{-2(k_1^{glob})^2-4k_1^{glob}+4} M^{2k_1^{glob}}. \quad (1.119)$$

The claim (1.117) follows from the above, (1.115), and our hierarchy of choosing C_0 after G and A after C_0.

This concludes the proof of Lemma 1.4.6.

The local version of Proposition 1.4.6 is stated, similarly to the global version, as follows. We omit its proof, since it is very similar to that of the global version.

Proposition 1.4.8 (main iteration step – local version)**.** *For all sufficiently large dimensions $D = \Omega_G(1)$ the following holds. Let M be a real number with*

$$M \geq C_0^{-1},$$

let $R \leq 1$ be a dyadic number, and let $m^* \geq s^2$. Suppose a map $\psi : G \to \mathbb{R}^D$ obeys the following estimates:

1. *(nondegenerate first derivatives) For any $p \in G$, we have*

$$C_0^{-1} M \leq |X_i \psi(p)| \leq C_0 M, \quad i = 1, \cdots, k, \tag{1.120}$$

and

$$\left| \bigwedge_{m=1}^{i} X_{j_m} \psi(p) \right| \geq C_0^{-i^2-i+2} \prod_{m=1}^{i} |X_{j_m} \psi(p)| \tag{1.121}$$

for $2 \leq i \leq k$ and $1 \leq j_1 < \cdots < j_i \leq k$.

2. *(locally free embedding) For any $p \in G$, we have*

$$\begin{aligned} &\left| \bigwedge_{\substack{W = X^{loc}_{r_1,j_1} \cdots X^{loc}_{r_m,j_m} \\ 1 \leq m \leq s, (r_1,j_1) \preceq^{loc} \cdots \preceq^{loc} (r_m,j_m)}} W\psi(p) \right| \\ &\geq C_0^{-2k^2-7k+2} \prod_{2 \leq m \leq s, (r_1,j_1) \preceq^{loc} \cdots \preceq^{loc} (r_m,j_m)} (AR)^{1-\sum_{i=1}^m r_i} \\ &\quad \cdot \prod_{j=1}^{k} |X_j \psi(p)| . \end{aligned} \tag{1.122}$$

3. (Hölder regularity at scale A) We have

$$\|\nabla^2\psi\|_{C_{AR}^{m^*-2,2/3}} \leq \frac{C_0}{AR}. \tag{1.123}$$

Then there exists a map $\phi : G \to \mathbb{R}^D$ such that ϕ obeys the following estimates:

1. (regularity at scale 1) We have

$$\|\phi\|_{C_R^{m^*,2/3}} \lesssim_G R. \tag{1.124}$$

2. (orthogonality) We have

$$B^{loc}(\phi, \psi) = 0. \tag{1.125}$$

3. (nondegenerate first derivatives) We have

$$|X_i\phi| \gtrsim_G 1, \quad i = 1, \cdots, k. \tag{1.126}$$

and the sum $\psi + \phi$ obeys the following regularity estimates:

1. (nondegenerate first derivatives) For any $p \in G$, we have

$$C_0^{-1}\sqrt{M^2+1} \leq |X_i(\psi+\phi)(p)| \leq C_0\sqrt{M^2+1}, \quad i = 1, \cdots, k,$$

and

$$\left|\bigwedge_{m=1}^{i} X_{j_m}(\psi+\phi)(p)\right| \geq C_0^{-i^2-i+2} \prod_{m=1}^{i} |X_{j_m}(\psi+\phi)(p)|$$

for $2 \leq i \leq k$ and $1 \leq j_1 < \cdots < j_i \leq k$.

2. (locally free embedding) For any $p \in G$, we have

$$\left| \bigwedge_{\substack{W = X^{loc}_{r_1,j_1} \cdots X^{loc}_{r_m,j_m} \\ 1 \le m \le s, (r_1,j_1) \preceq^{loc} \cdots \preceq^{loc} (r_m,j_m)}} W(\psi+\phi)(p) \right|$$
$$\geq C_0^{-2k^2-7k+3} \prod_{2 \le m \le s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)} (AR)^{1-\sum_{i=1}^m r_i}$$
$$\cdot \prod_{j=1}^{k} |X_j(\psi+\phi)(p)| .$$

3. (Hölder regularity at scale 1) We have

$$\|\nabla^2(\psi+\phi)\|_{C_R^{m^*-2,2/3}} \leq \frac{C_0}{R}.$$

1.5 Construction of the embedding

One obtains the following proposition by repeating Proposition 1.4.3 a finite number of times (see also Claim 5.4 of [77] for an analogous statement for $\mathbb{H}^3$).

Proposition 1.5.1 (finite iteration – global version)**.** *Let $0 < \varepsilon \leq 1/A$, let $R_0 \geq 1$ be a dyadic number, and let $0 \leq M_1 \leq M_2$ be nonnegative integers. One can find maps $\phi_m : G \to \mathbb{R}^D$ for $M_1 \leq m \leq M_2$ obeying the following bounds, where $\phi_{(\geq m)} : G \to \mathbb{R}^D$ is defined by*

$$\phi_{(\geq m)} := \sum_{m \leq m' \leq M_2} A^{-\varepsilon(m'-m)} \phi_{m'}.$$

1. (smoothness at scale $R_0 A^m$) For all $M_1 \leq m \leq M_2$, we have

$$\|\phi_m\|_{C^{s^2+s+1}_{R_0 A^m}} \leq C_0 R_0 A^m \tag{1.127}$$

$$\|\nabla^2 \phi_{(\geq m)}\|_{C^{s^2+s-1,2/3}_{R_0 A^m}} \leq C_0 R_0^{-1} A^{-m}. \tag{1.128}$$

2. *(orthogonality) One has, for all* $M_1 \le m \le M_2$,

$$\sum_{m'>m} A^{-\varepsilon(m'-m)} B^{glob}(\phi_m, \phi_{m'}) = 0. \tag{1.129}$$

3. *(nondegeneracy) For all* $p \in G$ *and* $M_1 \le m \le M_2$, *we have the following estimates:*

$$|X_i^{glob}\phi_m(p)| \ge C_0^{-1}, \quad i = 1, \cdots, k_1^{glob}, \tag{1.130}$$

$$\left| \bigwedge_{q=1}^{i} X_{j_q}^{glob}\phi_{(\ge m)}(p) \right| \ge C_0^{-i^2-i+2} \prod_{q=1}^{i} |X_{j_q}^{glob}\phi_{(\ge m)}(p)|, \tag{1.131}$$

for $2 \le i \le k_1^{glob}$ *and* $1 \le j_1 < \cdots < j_i \le k_1^{glob}$, *and*

$$\begin{aligned} &\left| \bigwedge_{\substack{W = X_{r_1,j_1}^{glob} \cdots X_{r_l,j_l}^{glob} \\ 1 \le l \le s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_l,j_l)}} (R_0 A^m)^{\sum_{i=1}^{l} r_i - 1} W\phi_{(\ge m)}(p) \right| \\ &\ge C_0^{-2(k_1^{glob})^2 - 7k_1^{glob} + 3} \prod_{i=1}^{k_1^{glob}} |X_i^{glob}\phi_{(\ge m)}(p)|. \end{aligned} \tag{1.132}$$

Proof. We prove this by induction on $M_2 - M_1$. The case $M_2 - M_1 = 0$ follows directly from Proposition 1.4.1, so we may assume $M_2 - M_1 > 0$. By the inductive hypothesis applied to $M_1' = M_1 + 1$ and $M_2' = M_2$, we may find functions ϕ_m, $M_1 + 1 \le m \le M_2$, so that

1. (smoothness at scale $R_0 A^n$) For all $M_1 + 1 \le m \le M_2$, we have

$$\|\phi_m\|_{C^{s^2+s+1}_{R_0 A^m}} \le C_0 R_0 A^m \tag{1.133}$$

$$\|\nabla^2 \phi_{(\ge m)}\|_{C^{s^2+s-1,2/3}_{R_0 A^m}} \le C_0 R_0^{-1} A^{-m} \tag{1.134}$$

2. (orthogonality) One has, for all $M_1 + 1 \le m \le M_2$,

$$\sum_{m'>m} A^{-\varepsilon(m'-m)} B^{glob}(\phi_m, \phi_{m'}) = 0. \tag{1.135}$$

3. (nondegeneracy) For all $p \in G$ and $M_1 + 1 \le m \le M_2$, we have the following estimates:

$$|X_i^{glob}\phi_m(p)| \ge C_0^{-1}, \quad i = 1, \cdots, k_1^{glob}, \tag{1.136}$$

$$\left| \bigwedge_{q=1}^{i} X_{j_q}^{glob} \phi_{(\ge m)}(p) \right| \ge C_0^{-i^2-i+2} \prod_{q=1}^{i} |X_{j_q}^{glob}\phi_{(\ge m)}(p)|, \tag{1.137}$$

for $2 \le i \le k_1^{glob}$ and $1 \le j_1 < \cdots < j_i \le k_1^{glob}$, and

$$\begin{aligned} &\left| \bigwedge_{\substack{W = X_{r_1,j_1}^{glob} \cdots X_{r_l,j_l}^{glob} \\ 1 \le l \le s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_l,j_l)}} (R_0 A^m)^{\sum_{i=1}^{l} r_i - 1} W\phi_{(\ge m)}(p) \right| \\ &\ge C_0^{-2(k_1^{glob})^2 - 7k_1^{glob} + 3} \prod_{i=1}^{k_1^{glob}} |X_i^{glob}\phi_{(\ge m)}(p)|. \end{aligned} \tag{1.138}$$

We now verify the hypotheses of Proposition 1.4.3 for

$$\psi := A^{-\varepsilon}\phi_{(\ge M_1+1)} = \sum_{M_1+1 \le m \le M_2} A^{-\varepsilon(m-M_1)}\phi_m, \quad \text{with } M = \left(\sum_{1 \le m \le M_2 - M_1} A^{-2\varepsilon m} \right)^{1/2},$$

and

$$m^* = s^2 + s + 1, \quad \alpha = \frac{2}{3}, \quad R = R_0 A^{M_1} \ge 1.$$

We have $M \ge A^{-\varepsilon} \ge A^{-1/A} \ge \frac{1}{2}$, so we have $M \ge C_0^{-1}$.

Verification of (1.73): From (1.135) and (1.136) we have, for $i = 1, \cdots, k_1^{glob}$,

$$|X_i^{glob}\psi|^2 = \sum_{M_1+1 \le m \le M_2} A^{-2\varepsilon(m-M_1)} |X_i^{glob}\phi_m|^2 \ge \sum_{1 \le m \le M_2 - M_1} A^{-2\varepsilon m} C_0^{-2} = C_0^{-2} M^2,$$

and from (1.133) we have, for $i = 1, \cdots, k_1^{glob}$,

$$|X_i^{glob}\psi|^2 = \sum_{M_1+1\le m\le M_2} A^{-2\varepsilon(m-M_1)}|X_i^{glob}\phi_m|^2 \le \sum_{1\le m\le M_2-M_1} A^{-2\varepsilon m}C_0^2 = C_0^2M^2.$$

Verification of (1.74): From (1.137) we have, for $2 \le i \le k_1^{glob}$ and $1 \le j_1 < \cdots <$ $j_i \le k_1^{glob}$,

$$\begin{aligned}\left|\bigwedge_{q=1}^{i} X_{j_q}^{glob}\psi\right| &= A^{-i\varepsilon}\left|\bigwedge_{q=1}^{i} X_{j_q}^{glob}\phi_{(\ge M_1+1)}\right| \ge A^{-i\varepsilon}C_0^{-i^2-i+2}\prod_{q=1}^{i}\left|X_{j_q}^{glob}\phi_{(\ge M_1+1)}\right| \\ &= C_0^{-i^2-i+2}\prod_{q=1}^{i}\left|X_{j_q}^{glob}\psi\right|.\end{aligned}$$

Verification of (1.75): From (1.138) we have

$$\begin{aligned}&\left|\bigwedge_{\substack{W=X_{r_1,j_1}^{glob}\cdots X_{r_m,j_m}^{glob}\\ 1\le m\le s,(r_1,j_1)\preceq^{glob}\cdots\preceq^{glob}(r_m,j_m)}} (R_0A^m)^{\sum_{i=1}^m r_i-1}W\psi(p)\right| \\ &\ge C_0^{-1}\left|\bigwedge_{\substack{W=X_{r_1,j_1}\cdots X_{r_m,j_m}\\ 1\le m\le s,(r_1,j_1)\preceq\cdots\preceq(r_m,j_m)}} W\phi_{(\ge M_1+1)}\right| \\ &\ge C_0^{-1}C_0^{-2k^2-7k+3}\prod_{i=1}^{k}|X_i\phi_{(\ge M_1+1)}(p)| \\ &\ge C_0^{-2k^2-7k+2}\prod_{i=1}^{k}|X_i\psi(p)|.\end{aligned} \tag{1.139}$$

Verification of (1.76): From (1.134) we have

$$\|\nabla^2\psi\|_{C_{AR}^{s^2+s-1,2/3}} = A^{-\varepsilon}\|\nabla^2\phi_{(\ge M_1+1)}\|_{C_{R_0A^{M_1+1}}^{s^2+s-1,2/3}} \le A^{-\varepsilon}C_0R_0^{-1}A^{-1-M_1} \le \frac{C_0}{AR}.$$

Hence, ψ and M satisfy the assumptions of Proposition 1.4.3, and so there exists a

function $\phi_{M_1} : G \to \mathbb{R}^{128 \cdot 23^{n_h}}$ that satisfies the following:

1. (regularity at scale 1) We have

$$\|\phi_{M_1}\|_{C^{s^2+s+1,2/3}_{AR_0^{M_1}}} \lesssim_G AR_0^{M_1}. \tag{1.140}$$

2. (orthogonality) We have

$$B^{glob}(\phi_{M_1}, \psi) = 0. \tag{1.141}$$

3. (nondegenerate first derivatives) For any $p \in G$, we have

$$|X_i^{glob}\phi_{M_1}(p)| \gtrsim_G 1, \quad i = 1, \cdots, k_1^{glob}, \tag{1.142}$$

$$C_0^{-1}\sqrt{M^2+1} \leq |X_i^{glob}(\psi + \phi_{M_1})(p)| \leq C_0\sqrt{M^2+1}, \quad i = 1, \cdots, k_1^{glob}, \tag{1.143}$$

$$\left| \bigwedge_{m=1}^{i} X_{j_m}^{glob}(\psi + \phi_{M_1})(p) \right| \geq C_0^{-i^2-i+2} \prod_{m=1}^{i} \left| X_{j_m}^{glob}(\psi + \phi_{M_1})(p) \right| \tag{1.144}$$

for $2 \leq i \leq k_1^{glob}$ and distinct $1 \leq j_1 < \cdots < j_i \leq k_1^{glob}$.

4. (locally free embedding) For any $p \in G$, we have

$$\begin{aligned} &\left| \bigwedge_{\substack{W = X_{r_1,j_1}^{glob} \cdots X_{r_m,j_m}^{glob} \\ 1 \leq m \leq s, (r_1,j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m,j_m)}} (AR^{M_1})^{\sum_{i=1}^m r_i - 1} W(\psi + \phi_{M_1})(p) \right| \\ &\geq C_0^{-2(k_1^{glob})^2 - 7k_1^{glob} + 3} \prod_{j=1}^{k_1^{glob}} |X_j(\psi + \phi_{M_1})(p)| . \end{aligned} \tag{1.145}$$

5. (Hölder regularity at scale AR^{M_1}) We have

$$\|\nabla^2(\psi + \phi_{M_1})\|_{C^{s^2+s-1,2/3}_{AR^{M_1}}} \leq \frac{C_0}{AR^{M_1}}. \tag{1.146}$$

Note that $\phi_{(\geq M_1)} = \psi + \phi_{M_1}$. To verify that the larger family of maps $\{\phi_m\}_{M_1 \leq m \leq M_2}$ satisfies the properties (1.127) to (1.132), we need only verify these properties for $m = M_1$. But, for $m = M_1$, (1.127) follows directly from (1.140) and (1.130) follows from (1.142), while (1.128), (1.129), (1.131), and (1.132) are precisely conditions (1.146), (1.141), (1.144), and (1.145), respectively. □

The local version is stated similarly as follows. Its proof is similar so we omit it.

Proposition 1.5.2 (finite iteration – local version)**.** *Let $0 < \varepsilon \leq 1/A$, let $R_0 \leq 1$ be a dyadic number, and let $M_1 \leq M_2 \leq 0$ be nonpositive integers. One can find maps $\phi_m : G \to \mathbb{R}^D$ for $M_1 \leq m \leq M_2$ obeying the following bounds, where $\phi_{(\geq m)} : G \to \mathbb{R}^D$ is defined by*

$$\phi_{(\geq m)} := \sum_{m \leq m' \leq M_2} A^{-\varepsilon(m'-m)} \phi_{m'}.$$

1. (smoothness at scale $R_0 A^m$) For all $M_1 \leq m \leq M_2$, we have

$$\|\phi_m\|_{C^{s^2+s+1}_{R_0 A^m}} \leq C_0 R_0 A^m \tag{1.147}$$

$$\|\nabla^2 \phi_{(\geq m)}\|_{C^{s^2+s-1,2/3}_{R_0 A^m}} \leq C_0 R_0^{-1} A^{-m}. \tag{1.148}$$

2. (orthogonality) One has, for all $M_1 \leq m \leq M_2$,

$$\sum_{m' > m} A^{-\varepsilon(m'-m)} B^{loc}(\phi_m, \phi_{m'}) = 0. \tag{1.149}$$

3. (nondegeneracy) For all $p \in G$ and $M_1 \leq m \leq M_2$, we have the estimates

$$|X_i \phi_m(p)| \geq C_0^{-1}, \quad i = 1, \cdots, k, \tag{1.150}$$

$$\left| \bigwedge_{q=1}^{i} X_{j_q} \phi_{(\geq m)}(p) \right| \geq C_0^{-i^2-i+2} \prod_{q=1}^{i} |X_{j_q} \phi_{(\geq m)}(p)|, \quad 2 \leq i \leq k,\ 1 \leq j_1 < \cdots < j_i \leq k, \tag{1.151}$$

$$\left| \bigwedge_{\substack{W = X^{loc}_{r_1,j_1} \cdots X^{loc}_{r_l,j_l} \\ 1 \leq l \leq s, (r_1,j_1) \preceq^{loc} \cdots \preceq^{loc} (r_l,j_l)}} (R_0 A^m)^{\sum_{i=1}^l r_i - 1} W \phi_{(\geq m)}(p) \right| \geq C_0^{-2k^2-7k+3} \prod_{i=1}^k |X_i \phi_{(\geq m)}(p)|. \tag{1.152}$$

We now obtain a full set of lacunary maps in the global case $R \geq 1$ by setting $M_1 = 0$ and taking the limit $M_2 \to \infty$, and in the local case $R \leq 1$ by setting $M_2 = 0$ and taking the limit $M_1 \to -\infty$. (See also Theorem 4.1 of [77] for an analogous statement for $\mathbb{H}^3$.)

Theorem 1.5.3 (Maps oscillating at lacunary scales – global version)**.** *Let $0 < \varepsilon \leq 1/A$ and let $1 \leq R_0 < A$ be a dyadic number. Then one can find a map $\phi_m : G \to \mathbb{R}^D$ for each nonnegative integer $m \geq 0$ obeying the following bounds:*

- *(smoothness at scale $R_0 A^m$) For all nonnegative integers $m \geq 0$, one has*

$$\|\phi_m\|_{C^{s^2+s}_{R_0 A^m}} \lesssim_{C_0} R_0 A^m.$$

In particular, we have

$$X^{glob}_{r,i} \phi_m(p) = O_{C_0}((R_0 A^m)^{-(r-1)}), \quad \text{for all } r, i. \tag{1.153}$$

- *(orthogonality) For all nonnegative integers $m \geq 0$, one has*

$$\sum_{m' > m} A^{-\varepsilon(m'-m)} B^{glob}(\phi_m, \phi_{m'}) = 0$$

identically on G. (By (1.153) this sum is absolutely convergent.)

- *(nondegeneracy and immersion) For all nonnegative integers $m \leq 0$ and all $p \in G$, one has*

$$|X^{glob}_i \phi_m(p)| \gtrsim_{C_0} 1, \quad i = 1, \cdots, k^{glob}_1,$$

and

$$\left|\bigwedge_{\substack{W=X^{glob}_{r_1,j_1}\cdots X^{glob}_{r_i,j_i}\\ 1\le i\le s,(r_1,j_1)\preceq^{glob}\ldots\preceq^{glob}(r_i,j_i)}} W\phi_{(\geq m)}(p)\right| \gtrsim_{C_0} \prod_{\substack{W=X^{glob}_{r_1,j_1}\cdots X^{glob}_{r_i,j_i}\\ 1\le i\le s,(r_1,j_1)\preceq^{glob}\ldots\preceq^{glob}(r_i,j_i)}} \left|W\phi_{(\geq m)}(p)\right|, \tag{1.154}$$

where

$$\phi_{(\geq m)}(p) = \sum_{m'\geq m} A^{-\varepsilon(m'-m)}\phi_{m'}.$$

Proof. For each $M \in \mathbb{N}$, we can apply Proposition 1.5.1 to find $\phi_m^M : G \to \mathbb{R}^D$ for $0 \leq m \leq M$ obeying the following bounds, where $\phi^M_{(\geq m)} : G \to \mathbb{R}^D$ is defined by

$$\phi^M_{(\geq m)} \coloneqq \sum_{m\leq m'\leq M} A^{-\varepsilon(m'-m)}\phi^M_{m'}.$$

1. (smoothness at scale A^m) For all $0 \leq m \leq M$, we have

$$\|\phi_m^M\|_{C^{s^2+s+1}_{R_0A^m}} \leq C_0R_0A^m. \tag{1.155}$$

2. (orthogonality) One has, for all $0 \leq m \leq M$,

$$\sum_{m'>m} A^{-\varepsilon(m'-m)}B(\phi_m^M, \phi_{m'}^M) = 0.$$

3. (nondegeneracy) For all $p \in G$ and $0 \leq m \leq M$, we have the estimates

$$|X_i^{glob}\phi_m^M(p)| \geq C_0^{-1}, \quad i = 1, \cdots, k_1^{glob},$$

and

$$\left| \bigwedge_{\substack{W=X^{glob}_{r_1,j_1}\cdots X^{glob}_{r_i,j_i} \\ 1\le i\le s,(r_1,j_1)\preceq^{glob}\cdots\preceq^{glob}(r_i,j_i)}} (AR_0)^{\sum_{j=1}^{i} r_j - 1} W\phi^M_{(\ge m)}(p) \right| \tag{1.156}$$
$$\ge C_0^{-2(k_1^{glob})^2-7k_1^{glob}+3} \prod_{i=1}^{k_1^{glob}} |X_i^{glob}\phi^M_{(\ge m)}(p)|.$$

From (1.155) and (1.156) we see that

$$\left| \bigwedge_{\substack{W=X^{glob}_{r_1,j_1}\cdots X^{glob}_{r_i,j_i} \\ 1\le i\le s,(r_1,j_1)\preceq^{glob}\cdots\preceq^{glob}(r_i,j_i)}} W\phi^M_{(\ge m)}(p) \right|$$
$$\ge C_0^{-2(k_1^{glob})^2-7k_1^{glob}+3} \prod_{\substack{W=X^{glob}_{r_1,j_1}\cdots X^{glob}_{r_i,j_i} \\ 1\le i\le s,(r_1,j_1)\preceq^{glob}\cdots\preceq^{glob}(r_i,j_i)}} C_0^{-1}|W\phi^M_{(\ge m)}(p)|.$$

For each $m \in \mathbb{Z}_{\ge 0}$, the sequence $\{\phi^M_m\}_{M\ge m}$ is bounded in the $C^{s^2+s+1}_{R_0A^m}$ norm, so by the Arzelà-Ascoli theorem, one can find a subsequence of $\{M_k\}$ such that $\{\phi^{M_k}_m\}_{M_k\ge m}$ locally converges in the C^{s^2+s} topology, say to ϕ_m, for every $m \in \mathbb{Z}_{\ge 0}$. It now readily follows that these ϕ_m satisfy the above properties. □

The local version of Theorem 1.5.3 is similar and is stated as follows. Its proof is similar and we omit it.

Theorem 1.5.4 (Maps oscillating at lacunary scales – local version)**.** *Let $0 < \varepsilon \le 1/A$ and let $\frac{1}{A} \le R_0 < 1$ be a dyadic number. Then one can find a map $\phi_m : G \to \mathbb{R}^D$ for each nonpositive integer $m \le 0$ obeying the following bounds:*

- *(smoothness at scale R_0A^m) For all nonpositive integers $m \le 0$, one has*

$$\|\phi_m\|_{C^{s^2+s}_{R_0A^m}} \lesssim_{C_0} R_0A^m.$$

In particular, we have

$$X_{r,i}^{loc}\phi_m(p) = O_{C_0}((R_0A^m)^{-(r-1)}), \quad \textit{for all } r,i. \tag{1.157}$$

- *(orthogonality) For all nonpositive integers $m \leq 0$, one has*

$$\sum_{m<m'\leq 0} A^{-\varepsilon(m'-m)} B^{loc}(\phi_m, \phi_{m'}) = 0$$

identically on G.

- *(nondegeneracy and immersion) For all nonnegative integers $m \leq 0$ and all $p \in G$, one has*

$$|X_i\phi_m(p)| \gtrsim_{C_0} 1, \quad i = 1, \cdots, k,$$

and

$$\left| \bigwedge_{\substack{W=X_{r_1,j_1}^{loc}\cdots X_{r_i,j_i}^{loc} \\ 1\leq i\leq s,(r_1,j_1)\preceq^{loc}\cdots\preceq^{loc}(r_i,j_i)}} W\phi_{(\geq m)}(p) \right| \gtrsim_{C_0} \prod_{\substack{W=X_{r_1,j_1}^{loc}\cdots X_{r_i,j_i}^{loc} \\ 1\leq i\leq s,(r_1,j_1)\preceq^{loc}\cdots\preceq^{loc}(r_i,j_i)}} \left|W\phi_{(\geq m)}(p)\right|, \tag{1.158}$$

where

$$\phi_{(\geq m)}(p) = \sum_{m\leq m'\leq 0} A^{-\varepsilon(m'-m)}\phi_{m'}.$$

Once we have the lacunary families guaranteed by Theorems 1.5.3 and 1.5.4, we can construct components of the mappings. For $n = 0, \cdots, \log_2 A - 1$, let

$$\{\phi_m^{n,glob}\}_{m\geq 0}$$

be the lacunary family obtained from Theorem 1.5.3 with $R_0 = 2^n$, and let

$$\{\phi_m^{n-\log_2 A,glob}\}_{m\leq 0}$$

be the lacunary family obtained from Theorem 1.5.3 with $R_0 = 2^n A^{-1}$.

For $n = 0, \cdots, \log_2 A - 1$, we define the global subcomponent $\Phi^{n,glob} : G \to \mathbb{R}^D$ as

$$\Phi^{n,glob}(p) := \sum_{m=0}^{\infty} (2^n A^m)^{-\varepsilon} (\phi_m^{n,glob}(p) - \phi_m^{n,glob}(0)), \quad p \in G,$$

and define the global component $\Phi^{glob} : G \to \mathbb{R}^{D \log_2 A}$ as

$$\Phi^{glob} = \oplus_{n=0}^{\log_2 A - 1} \Phi^{n,glob}.$$

Likewise, for $n = 0, \cdots, \log_2 A - 1$, we define the local subcomponent $\Phi^{n-\log_2 A,glob} : G \to \mathbb{R}^D$ as

$$\Phi^{n-\log_2 A,glob}(p) := \sum_{m=0}^{\infty} (2^n A^{m-1})^{-\varepsilon} (\phi_m^{n-\log_2 A,glob}(p) - \phi_m^{n-\log_2 A,glob}(0)), \quad p \in G,$$

and define the local component $\Phi^{loc} : G \to \mathbb{R}^{D \log_2 A}$ as

$$\Phi^{loc} = \oplus_{n=0}^{\log_2 A - 1} \Phi^{n-\log_2 A,loc}.$$

Then Φ_{glob} and Φ_{loc} are "almost" bi-Lipschitz embeddings of $(G, d_G^{1-\varepsilon})$ into $\mathbb{R}^D$: they are Lipschitz, and satisfy the Lipschitz lower bound at large and at small scales, respectively.

Proposition 1.5.5. *The maps $\Phi^{glob}, \Phi^{loc} : G \to \mathbb{R}^{D \log_2 A}$ satisfy the following estimates.*

1. (Hölder upper bound)

$$|\Phi^{glob}(p) - \Phi^{glob}(p')| \lesssim_A \varepsilon^{-1/2} d_G(p, p')^{1-\varepsilon}, \quad \text{for all } p, p' \in G, \tag{1.159}$$

and

$$|\Phi^{loc}(p) - \Phi^{loc}(p')| \lesssim_A \varepsilon^{-1/2} d_G(p,p')^{1-\varepsilon}, \quad \text{for all } p, p' \in G. \tag{1.160}$$

2. *(partial Hölder lower bound) For $p, p' \in G$ so that $d_G(p,p') \geq A^2 A^{-1/(s+1)}$, we have*

$$|\Phi^{glob}(p) - \Phi^{glob}(p')| \gtrsim_A d_G(p,p')^{1-\varepsilon}. \tag{1.161}$$

For $p, p' \in G$ so that $d_G(p,p') \leq A^{-2} A^{-1/(s+1)}$, we have

$$|\Phi^{loc}(p) - \Phi^{loc}(p')| \gtrsim_A d_G(p,p')^{1-\varepsilon}. \tag{1.162}$$

Proof. 1. By the orthogonality statement of Theorem 1.5.3, we have, for each $n = 0, \cdots, \log_2 A - 1$ and $i = 1, \cdots, k_1^{glob}$,

$$\begin{aligned} |X_i^{glob} \Phi^{n,glob}| &= \left| \sum_{m=0}^{\infty} (2^n A^m)^{-\varepsilon} X_i^{glob} \phi_m^{n,glob} \right| = \left(\sum_{m=0}^{\infty} (2^n A^m)^{-2\varepsilon} |X_i^{glob} \phi_m^{n,glob}|^2 \right)^{1/2} \\ &\asymp_{C_0} M, \end{aligned}$$

where

$$M := \left(\sum_{n=0}^{\infty} A^{-2\varepsilon n} \right)^{1/2} = \left(\frac{1}{1 - A^{-2\varepsilon}} \right)^{1/2} \asymp \frac{1}{\sqrt{\varepsilon \log A}} \lesssim_A \varepsilon^{-1/2},$$

On the other hand, for $r = 2, \cdots, s$, $i = 1, \cdots, k_r^{glob}$, and $n = 0, \cdots, \log_2 A - 1$, by (1.153),

$$|X_{r,i}^{glob} \Phi^{n,glob}| \leq \sum_{m=0}^{\infty} (2^n A^m)^{-\varepsilon} \left| X_{r,i}^{glob} \phi_m^{n,glob} \right| \lesssim_{C_0} \sum_{m=0}^{\infty} (2^n A^m)^{-r-\varepsilon+1} \lesssim 1.$$

Thus,

$$\||\nabla\Phi^{glob}\|_{C^0} \lesssim_A M \lesssim_A \varepsilon^{-1/2},$$

i.e., $\Phi^{glob} : G \to \mathbb{R}^{D\log_2 A}$ is $O_A(\varepsilon^{-1/2})$-Lipschitz. This proves the Hölder upper bound when $d_G(p,p') \leq 1$.

Let $p, p' \in G$ with $d_G(p,p') > 1$, say $A^{m-1} < d_G(p,p') \leq A^m$ for some $m \in \mathbb{Z}_{>0}$. Consider, for $n = 0, \cdots, \log_2 A - 1$, the low-frequency component

$$\Phi^{n,glob}_{\geq m-1}(q) := \sum_{m'=m-1}^{\infty} (2^n A^{m'})^{-\varepsilon}(\phi^{n,glob}_{m'}(q) - \phi^{n,glob}_{m'}(0)), \quad q \in G.$$

Then

$$\Phi^{n,glob}(p) - \Phi^{n,glob}(p') = \Phi^{n,glob}_{\geq m-1}(p) - \Phi^{n,glob}_{\geq m-1}(p') + O_{C_0}(A^{m(1-\varepsilon)}), \tag{1.163}$$

so the Hölder upper bound reduces to

$$|\Phi^{n,glob}_{\geq m-1}(p) - \Phi^{n,glob}_{\geq m-1}(p')| \lesssim_A \varepsilon^{-1/2} A^{m(1-\varepsilon)},$$

and it is enough to show

$$\||\nabla\Phi^{n,glob}_{\geq m-1}|\|_{C^0} \lesssim_A \varepsilon^{-1/2} A^{-m\varepsilon} \tag{1.164}$$

(we just used the hierarchy of choosing A after C_0). But this is the consequence of

$$\begin{aligned} |X_i^{glob}\Phi^{n,glob}_{\geq m-1}| &= \left| \sum_{m'=m-1}^{\infty} (2^n A^{m'})^{-\varepsilon} X_i^{glob}\phi^{n,glob}_{m'} \right| \\ &= \left(\sum_{m'=m-1}^{\infty} (2^n A^{m'})^{-2\varepsilon} |X_i^{glob}\phi^{n,glob}_{m'}|^2 \right)^{1/2} \\ &\asymp_A \varepsilon^{-1/2} A^{-m\varepsilon}, \end{aligned}$$

for $i = 1, \cdots, k_1^{glob}$, and

$$|X_{r,i}^{glob}\Phi_{\geq m-1}^{n,glob}| \leq \sum_{m'=m-1}^{\infty} (2^n A^{m'})^{-\varepsilon} \left| X_{r,i}^{glob}\phi_{m'}^{n,glob} \right| \lesssim_{C_0} \sum_{m'=m-1}^{\infty} (2^n A^{m'})^{-r-\varepsilon+1}$$
$$\lesssim_A A^{-m(1+\varepsilon)}$$

for $r = 2, \cdots, s$ and $i = 1, \cdots, k_r^{glob}$, where we have again used the orthogonality statement of Theorem 1.5.3 and (1.153).

The proof for Φ^{loc} is similar and we omit it.

2. Let $p, p' \in G$ be so that $2^n A^{m_0} A^{-1/(s+1)} \leq d_G(p, p') \leq 2^{n+1} A^{m_0} A^{-1/(s+1)}$ for some integer $m_0 \geq 1$ and $n = 1, \cdots, \log_2 A - 1$. It is enough to show

$$|\Phi^{n,glob}(p) - \Phi^{n,glob}(p')| \gtrsim_A (2^n A^{m_0})^{1-\varepsilon}.$$

Writing $(p')^{-1}p = \exp\left(\sum_{r=1}^{s} \sum_{i=1}^{k_r^{glob}} x_{r,i}^{glob} X_{r,i}^{glob}\right)$, we have

$$\sum_{r=1}^{s} \sum_{j=1}^{k_r^{glob}} |x_{r,j}^{glob}|^{1/r} \asymp_G d_G(p, p') \asymp 2^n A^{m_0} A^{-1/(s+1)}.$$

Equivalently, we have $|x_{r,j}^{glob}| \lesssim_G 2^{nr} A^{m_0 r} A^{-r/(s+1)}$ for all r, j, and there exists some pair (r, j) such that $|x_{r,j}^{glob}| \asymp_G 2^{nr} A^{m_0 r} A^{-r/(s+1)}$. Recall the low-frequency component

$$\Phi_{\geq m_0}^{n,glob}(q) \coloneqq \sum_{m'=m_0}^{\infty} (2^n A^{m'})^{-\varepsilon} (\phi_{m'}^{n,glob}(q) - \phi_{m'}^{n,glob}(0)), \quad q \in G.$$

By Taylor expansion

$$\Phi^{n,glob}_{\geq m_0}(p) = \Phi^{n,glob}_{\geq m_0}(p') + \sum_{m=1}^{s} \frac{1}{m!} \left(\sum_{r=1}^{s} \sum_{j=1}^{k_r^{glob}} x^{glob}_{r,j} X^{glob}_{r,j} \right)^m \Phi^{n,glob}_{\geq m_0}(p') + O_{C_0}(2^{n(1-\varepsilon)} A^{m_0(1-\varepsilon)} A^{-2})$$

(note that this is where we used the $C^{s(s+1)}$-regularity of Ψ), and in light of (1.163), we have

$$\Phi^{n,glob}(p) - \Phi^{n,glob}(p') = \sum_{m=1}^{s} \frac{1}{m!} \left(\sum_{r=1}^{s} \sum_{j=1}^{k_r^{glob}} x^{glob}_{r,j} X^{glob}_{r,j} \right)^m \Phi^{n,glob}_{\geq m_0}(p') + O_{C_0}(2^{n(1-\varepsilon)} A^{m_0(1-\varepsilon)} A^{-1}).$$

This expression contains many terms of the form $X^{glob}_{r_1,j_1} \cdots X^{glob}_{r_m,j_m} \Phi^{n,glob}_{\geq m}$, where the (r_1, j_1), $\cdots$, (r_m, j_m) are arbitrary with $m \leq s$ and are often unordered. In order to use the freeness property (1.154), it would be necessary to modify the above Taylor expansion formula so that the only differential operators acting on Ψ are the ones of the form $X^{glob}_{r_1,j_1} \cdots X^{glob}_{r_m,j_m}$ where $(r_1, j_1) \preceq^{glob} \cdots \preceq^{glob} (r_m, j_m)$.

This modification is possible once we note that for any permutation π of $\{1, \cdots, m\}$ we can express $X^{glob}_{r_1,j_1} \cdots X^{glob}_{r_m,j_m} - X^{glob}_{r_{\pi(1)},j_{\pi(1)}} \cdots X^{glob}_{r_{\pi(m)},j_{\pi(m)}}$ as a linear combination of differential operators $X^{glob}_{r'_1,j'_1} \cdots X^{glob}_{r'_{m'},j'_{m'}}$ where $\sum_{i=1}^{m'} r'_i \geq \sum_{i=1}^{m} r_i$ and $(r'_1, j'_1) \preceq^{glob} \cdots \preceq^{glob} (r'_{m'}, j'_{m'})$. This can be proven by a simple induction argument on m using the fact that, if $r + r' \leq s$, then $[X^{glob}_{r,j}, X^{glob}_{r',j'}] \in \underbrace{[\mathfrak{g}, [\mathfrak{g}, \cdots, \mathfrak{g}]]}_{r+r' \text{ times}}$ is a linear combination of $X^{glob}_{r'',i}$ for $r'' \geq r + r'$ and $i = 1, \cdots, k_{r''}$, and $[X^{glob}_{r,j}, X^{glob}_{r',j'}] = 0$ if $r + r' > s$. Applying this fact to the above Taylor expansion formula and keeping track of the

degrees, we obtain the following modified Taylor expansion formula:

$$\begin{aligned}
&\Phi^{n,glob}(p) - \Phi^{n,glob}(p') \\
&= \sum_{r=1}^{s} \sum_{j=1}^{k_r^{glob}} (x_{r,j}^{glob} + p_{r,j}) X_{r,j}^{glob} \Phi_{\geq m_0}^{n,glob}(p') \\
&\quad + \sum_{m=2}^{s} \sum_{(r_1,j_1) \preceq^{glob} \ldots \preceq^{glob} (r_m,j_m)} p_{(r_1,j_1),\cdots,(r_m,j_m)} X_{r_1,j_1}^{glob} \cdots X_{r_m,j_m}^{glob} \Phi_{\geq m_0}^{n,glob}(p') \\
&\quad + O_{C_0}(2^{n(1-\varepsilon)} A^{m_0(1-\varepsilon)} A^{-1})
\end{aligned} \tag{1.165}$$

where $p_{r,j}$ is a polynomial of global degree $\leq r$ where each monomial contains a term of the form $x_{r',j'}^{glob}$ with $r' < r$ (recall that we define the global degree by assigning weight r to $x_{r,j}^{glob}$), and $p_{(r_1,j_1),\cdots,(r_m,j_m)}$ is a polynomial of global degree $\leq \sum_{i=1}^{m} r_i$.

By the freeness property (1.154), each term in (1.165) (except for the error term) may serve as a lower bound for the entire sum, up to multiplicative constants. Thus, it suffices to show that there exists r, j such that $x_{r,j}^{glob} + p_{r,j}$ has non-negligible size. More precisely, the freeness property (1.154) tells us that

$$\left| \bigwedge_{\substack{W = X_{r_1,j_1}^{glob} \cdots X_{r_m,j_m}^{glob} \\ 1 \leq m \leq s, (r_1,j_1) \preceq^{glob} \ldots \preceq^{glob} (r_m,j_m)}} W\Psi(p) \right| \asymp_{C_0} \prod_{\substack{W = X_{r_1,j_1}^{glob} \cdots X_{r_m,j_m}^{glob} \\ 1 \leq m \leq s, (r_1,j_1) \preceq^{glob} \ldots \preceq^{glob} (r_m,j_m)}} |W\Psi(p)|,$$

which immediately gives us, for each (r_0, j_0), the following control on the main

term of the Taylor expansion:

$$\begin{aligned}&\left|\sum_{r=1}^{s}\sum_{j=1}^{k_r^{glob}}(x_{r,j}^{glob}+p_{r,j})X_{r,j}^{glob}\Phi_{\geq m_0}^{n,glob}(p')\right.\\&\quad\left.+\sum_{m=2}^{s}\sum_{(r_1,j_1)\preceq^{glob}\cdots\preceq^{glob}(r_m,j_m)}p_{(r_1,j_1),\cdots,(r_m,j_m)}X_{r_1,j_1}^{glob}\cdots X_{r_m,j_m}^{glob}\Phi_{\geq m_0}^{n,glob}(p')\right|\\&\gtrsim_{C_0}\left|(x_{r_0,j_0}^{glob}+p_{r_0,j_0})X_{r_0,j_0}^{glob}\Phi_{\geq m_0}^{n,glob}(p')\right|.\end{aligned}$$

So it remains to single out a pair (r_0, j_0) such that the right-hand side is large enough. Indeed, recall that we have $|x_{r,j}^{glob}| \lesssim_G 2^{nr}A^{m_0 r}A^{-r/(s+1)}$ for all r, j, and there exists some pair (r, j) such that $|x_{r,j}^{glob}| \asymp_G 2^{nr}A^{m_0 r}A^{-r/(s+1)}$. Therefore, there exists some (r_0, j_0) such that

$$\begin{aligned}&|x_{r,j}^{glob}| < 2^{nr}A^{m_0 r}A^{-r/(s+1)-1/2(s+1)} \text{ for all } r < r_0, 1 \leq j \leq k_r,\\&|x_{r_0,j_0}^{glob}| \geq 2^{nr_0}A^{m_0 r_0}A^{-r_0/(s+1)-1/2(s+1)}\end{aligned} \tag{1.166}$$

(we have just used the hierarchy that A is chosen after G). Then, as p_{r_0,j_0} is of global degree $\leq r_0$ and is not x_{r_0,j_0}^{glob} itself, we must have

$$|p_{r_0,j_0}| \lesssim_G 2^{nr_0}A^{m_0 r_0}A^{-(r_0+1)/(s+1)}, \quad \text{or} \quad |p_{r_0,j_0}| \leq \frac{1}{2}2^{nr_0}A^{m_0 r_0}A^{-(r_0+1/2)/(s+1)} \tag{1.167}$$

(we have again just used the hierarchy that A is chosen after G). Hence,

$$|x_{r_0,j_0}^{glob}+p_{r_0,j_0}| \geq |x_{r_0,j_0}^{glob}| - |p_{r_0,j_0}| \geq \frac{1}{2}2^{nr_0}A^{m_0 r_0}A^{-(r_0+1/2)/(s+1)},$$

and recalling

$$|X_{r_0,j_0}^{glob}\Phi_{\geq m_0}^{n,glob}(p')| \gtrsim_{C_0} (2^n A^{m_0})^{1-r-\varepsilon}$$

and we see that $\left|(x_{r_0,j_0}^{glob}+p_{r_0,j_0})X_{r_0,j_0}^{glob}\Phi_{\geq m_0}^{n,glob}(p')\right| \gtrsim_{C_0} 2^{n(1-\varepsilon)}A^{m_0(1-\varepsilon)}A^{-(r_0+1/2)/(s+1)}$.

In conclusion, we have $|\Phi^{n,glob}(p) - \Phi^{n,glob}(p')| \gtrsim_{C_0} 2^{n(1-\varepsilon)} A^{m_0(1-\varepsilon)} A^{-(r_0+1/2)/(s+1)}$, or

$$|\Phi^{n,glob}(p) - \Phi^{n,glob}(p')| \gtrsim_A 2^{n(1-\varepsilon)} A^{m_0(1-\varepsilon)} \tag{1.168}$$

(we just used the hierarchy of choosing A after C_0).

The proof of the local case is similar.

□

Now, with Proposition 1.5.5 in hand, we are ready to prove the main theorem.

Proof. By Theorem 1.1.1, there exists an $O_A(\varepsilon^{-1/2})$-bilipschitz mapping

$$\Phi^{meso} : (G, d_G^{1-\varepsilon}) \to \mathbb{R}^{O_A(1)}$$

such that for all $p, p' \in G$ such that $d_G(p, p') = \Theta_A(1)$, we have

$$|\Phi^{meso}(p) - \Phi^{meso}(p')| \gtrsim_A d_G(p, p')^{1-\varepsilon}.$$

Then $\Phi^{glob} \oplus \Phi^{meso} \oplus \Phi^{loc}$ is our desired embedding. □

Remark 1.5.6. *A careful analysis reveals that we have constructed $\Phi : G \to \mathbb{R}^{O_A(1)}$ with the following properties. If $d_G(p, p') \geq 1$ with*

$$(p')^{-1}p = \exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r^{glob}} x_{r,i}^{glob} X_{r,i}^{glob}\right),$$

then

$$|\Phi(p) - \Phi(p')| \asymp_A \varepsilon^{-1/2} \sum_{i=1}^{k_1^{glob}} |x_{1,i}^{glob}| + \sum_{r=2}^{s}\sum_{i=1}^{k_r^{glob}} |x_{r,i}^{glob}|^{1/r},$$

while if $d_G(p,p') \leq 1$ *with*

$$(p')^{-1}p = \exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r^{loc}} x_{r,i}^{loc} X_{r,i}^{loc}\right),$$

then

$$|\Phi(p) - \Phi(p')| \asymp_A \varepsilon^{-1/2}\sum_{i=1}^{k}|x_{1,i}^{loc}| + \sum_{r=2}^{s}\sum_{i=1}^{k_r^{loc}}|x_{r,i}^{loc}|^{1/r}.$$

Chapter 2

Vertical versus horizontal inequalities on finitely generated groups of polynomial growth and simply connected nilpotent Lie groups

It is well known that nonabelian simply connected nilpotent Lie groups and not virtually abelian finitely generated groups of polynomial growth fail to embed bilipschitzly (or quasi-isometrically) into uniformly convex Banach spaces (or, more generally, Banach spaces with the Radon–Nikodým property). This is because nonabelian Carnot groups do not bilipschitzly embed into Banach spaces with the Radon–Nikodým property [11, 52], and nonabelian simply connected nilpotent Lie groups have nonabelian Carnot groups as asymptotic cone and hence do not quasi-isometrically embed into Banach spaces with the Radon-Nikodým property (see [38] for example), while not virtually abelian finitely generated groups of polynomial

growth are quasi-isometric to nonabelian simply connected nilpotent Lie groups [30, 55] and hence fail to embed quasi-isometrically into Banach spaces with the Radon–Nikodým property. The aim of this paper is to provide quantitative counterparts to these qualitative nonembeddability statements.

Let G be a nonabelian simply connected nilpotent Lie group, endowed with a left-invariant Riemannian metric. We write

$$B_r(x) \coloneqq \{y \in G : d_G(x,y) < r\},\ B_r \coloneqq B_r(e_G), \quad x \in G,\ r > 0,$$

where e_G is the identity element of G. Note that $B_r(x) = xB_r$, by left-invariance.

Let Γ be a finitely generated group of polynomial growth. This means that Γ has a finite generating set S, and that if we denote by $d_W(\cdot,\cdot)$ the left-invariant word metric on Γ induced by S and $B_n^\Gamma = \{x \in \Gamma : d_W(x, e_\Gamma) \le n\}$ the corresponding closed ball of radius $n \in \mathbb{N}$, where e_Γ is the identity element of Γ, then the cardinality $|B_n^\Gamma|$ grows at most polynomially in n. We assume in addition that Γ is not virtually abelian, i.e., that it has no finite index subgroup isomorphic to $\mathbb{Z}^n$; by [15, Corollary 1.5] Γ has no finite index subgroups that have abelian groups as quotients by finite normal subgroups.

Let the Banach space $(X, \|\cdot\|_X)$ be a uniformly convex Banach space, that is, for every $\varepsilon \in (0,1)$ there exists $\delta \in (0,1)$ such that every $x, y \in X$ with $\|x\|_X = \|y\|_X = 1$ and $\|x-y\|_X \ge \varepsilon$ satisfy $\|x+y\|_X \le 2(1-\delta)$. By [6, 24, 70], there is an equivalent norm on X such that if we renorm X using this norm, there is some exponent $q \in [2,\infty)$ for which the Banach space $(X, \|\cdot\|_X)$ is q-uniformly convex, which means that the q-uniform convexity constant of X defined by

$$K_q(X) \coloneqq \inf\left\{K > 0 : \forall x, y \in X\ \left(\|x\|_X^q + \frac{1}{K^q}\|y\|_X^q\right)^{1/q} \le \left(\frac{\|x+y\|_X^q + \|x-y\|_X^q}{2}\right)^{1/q}\right\}$$

is finite. We will assume this renorming has happened, because it does not affect our theorems, so that X is q-uniformly convex for some $q \geq 2$. By [32] and [6], the L_p spaces for $1 < p < \infty$ are $\max\{p, 2\}$-uniformly convex with

$$K_2(L_p) \leq \frac{1}{\sqrt{p-1}},\ 1 < p \leq 2, \quad K_p(L_p) \leq 1,\ p \geq 2.$$

A separable metric space (M, d_M) is said to embed *(D-)bilipschitzly* into a normed space $(X, \|\cdot\|_X)$ if there is a finite $D \in [1, \infty)$ and a mapping $f : M \to X$ such that $d_M(x, y) \leq \|f(x) - f(y)\| \leq D d_M(x, y)$ for all $x, y \in M$. The bilipschitz distortion $c_X(M, d_M)$ is defined to be the infimum over those D for which such a mapping exists. We write[1] $c_{L_p(\sigma)}(M, d_M) = c_{L_p([0,1])}(M, d_M) =: c_p(M, d_M)$ and call $c_2(M, d_M)$ the Euclidean distortion of (M, d_M). The space (M, d_M) is said to *quasi-isometrically embed* into $(X, \|\cdot\|_X)$ if there is a mapping $f : M \to X$, $D \in [1, \infty)$ and $C > 0$ such that $d_M(x, y) - C \leq \|f(x) - f(y)\| \leq D d_M(x, y) + C$ for all $x, y \in M$.

Now we have defined all the concepts used in the first sentence of the introduction (except for the Radon–Nikodým property, which will not appear in other parts of this paper), which, in other words, can be written as

$$c_X(G) = c_X(\Gamma) = \infty,$$

for G, Γ, and X given as above. In this paper, we quantify this fact by providing the following growth rates of the bilipschitz constant of balls in such groups.

Theorem 2.0.1. *Let G be a nonabelian simply connected nilpotent Lie group, and*

[1] This equality is due to the fact that any separable subspace of an $L^p(\sigma)$ space is isometric to a subspace of $L^p([0, 1])$, see [67, Fact 1.20].

let X be a q-uniformly convex Banach space, $q \geq 2$. Then[2]

$$c_X(B_r) \gtrsim_G \frac{(\log r)^{1/q}}{K_q(X)}, \quad r \geq 2.$$

Theorem 2.0.2. *Let Γ be a not virtually abelian finitely generated group of polynomial growth, and let X be a q-uniformly convex Banach space, $q \geq 2$. Then*

$$c_X(B_n^\Gamma) \gtrsim_\Gamma \frac{(\log n)^{1/q}}{K_q(X)}, \quad n \geq 2.$$

As a special case, we obtain sharp bounds on the L^p-distortions, the upper bounds following from the Assouad embedding theorem.

Corollary 2.0.3. *Let G be a nonabelian simply connected nilpotent Lie group. Then,*

$$c_p(B_r) \asymp_{G,p} (\log r)^{1/\max\{p,2\}}, \quad 1 < p < \infty,\ r \geq 2.$$

Corollary 2.0.4. *Let Γ be a not virtually abelian finitely generated group of polynomial growth. Then*

$$c_p(B_n^\Gamma) \asymp_{\Gamma,p} (\log n)^{1/\max\{p,2\}}, \quad 1 < p < \infty,\ n \geq 2.$$

We remark that Theorem 2.0.2 and Corollary 2.0.4 were proven for the discrete Heisenberg groups by Lafforgue and Naor [47], where the discrete Heisenberg groups

[2]We will use the following (standard) asymptotic notation. For $P, Q > 0$, the notations $P \lesssim Q$, $Q \gtrsim P$, $P = O(Q)$, and $Q = \Omega(P)$ mean that $P \leq KQ$ for a universal constant $K \in (0, \infty)$, and the notation $P \asymp Q$ means $P \lesssim Q$ and $Q \lesssim P$. If we need to allow for dependence on parameters, we indicate this by subscripts. For example, in the presence of auxiliary parameters ψ, ξ, the notations $P \lesssim_{\psi,\xi} Q$, $Q \gtrsim_{\psi,\xi} P$, $P = O_{\psi,\xi}(Q)$, $Q = \Omega_{\psi,\xi}(P)$ mean that $P \leq K(\psi, \xi)Q$ where $K(\psi, \xi) \in (0, \infty)$ may depend only on ψ and ξ, and $P \asymp_{\psi,\xi} Q$ means that $P \lesssim_{\psi,\xi} Q$ and $Q \lesssim_{\psi,\xi} P$.

$\mathbb{H}_{\mathbb{Z}}^{2k+1}$, $k \in \mathbb{Z}_{>0}$ are defined as the groups with word relations as follows:

$$\mathbb{H}_{\mathbb{Z}}^{2k+1} = \Big\langle a_1, \cdots, a_k, b_1, \cdots, b_k, c \,|\, \forall i \,(c = [a_i, b_i]) \\ \wedge \forall i, j \left(i \neq j \to [a_i, a_j] = [b_i, b_j] = [a_i, b_j] = [a_i, c] = [b_i, c] = e_{\mathbb{H}_{\mathbb{Z}}^{2k+1}} \right) \Big\rangle.$$

A previously weaker result was given by Li [53, Theorem 1.4], who proved that if Γ is a finitely generated nonabelian torsion-free nilpotent group, and X is a uniformly convex Banach space, then there exists $c > 0$ depending on Γ and X such that

$$c_X(B_n^\Gamma) \gtrsim_{G,X} (\log n)^c, \quad n \geq 2.$$

A stronger bound was given for cocompact lattices Γ of Carnot groups by Gartland [27, Corollary 1.6]: if G is a Carnot group of step s (to be defined below) and Γ is a cocompact lattice of G, then

$$c_{\mathbb{R}^d}(B_n^\Gamma) \gtrsim_{\Gamma,d} \frac{(\log n)^{\frac{1}{2}-\frac{1}{2s}}}{(\log\log n)^{\frac{1}{2}+\frac{1}{2s}}}, \quad n \geq 3,$$

under the additional restriction that G contains a copy of the model filiform group $J^{s-1}(\mathbb{R})$ of step s (this requirement is automatically fulfilled if $s \leq 3$), which is the Carnot group whose Lie algebra $\mathfrak{j}^{s-1}$ is spanned by elements $x, y_0, \cdots, y_{s-1}$ with the only nontrivial bracket relations being

$$[x, y_i] = y_{i+1}, \quad i = 0, \cdots, s-2.$$

For example, $J^0(\mathbb{R}) = \mathbb{R}^2$, $J^1(\mathbb{R}) =: \mathbb{H}^3$ is called the real Heisenberg group of dimension 3, and $J^2(\mathbb{R})$ is called the Engel group. Compared to this result, Corollary 2.0.4 removes the exponent gap of $\frac{1}{2s}$ and gets rid of the lower order factor.

We obtain Theorems 2.0.1 and 2.0.2 by proving the following "vertical-versus-

horizontal inequalities" on the groups G and Γ.

We first state the inequality on nilpotent groups. In the following theorem, given left-invariant vector fields $X_1, \cdots, X_k$ that form a basis of the Lie algebra $\mathfrak{g}$ of G at e_G, we write the horizontal gradient of a function $f : G \to X$ as

$$\nabla f \coloneqq (X_1 f, \cdots, X_k f),$$

if each $X_i f$ exists. Also, given $v \in G$ and $t \in \mathbb{R}$, we write $v^t \coloneqq \exp(tw)$, where $w \in \mathfrak{g}$ is such that $v = \exp(w)$, and $\exp : \mathfrak{g} \to G$ is the Lie group exponential map. Let μ denote the bi-invariant Haar measure of G, which is the push-forward of the Lebesgue measure on $\mathfrak{g}$ by the exponential map.

Theorem 2.0.5. *Let G be a nonabelian simply connected nilpotent Lie group. Let $v \in Z(G)$ and $\rho \in \mathbb{N}$ with $\rho \geq 2$ and $d_G(v^t, e_G) \leq t^{1/\rho}$, $t > 0$. Suppose that $p \in (1, \infty)$ and $q \in [2, \infty)$. Let $(X, \|\cdot\|_X)$ be a Banach space with $K_q(X) < \infty$, and let $f : G \to X$ be smooth and compactly supported. Then*

$$\begin{aligned} &\left(\int_0^\infty \left(\int_G \left(\frac{\|f(hv^t) - f(h)\|_X}{t^{1/\rho}} \right)^p d\mu(h) \right)^{\max\{p,q\}/p} \frac{dt}{t} \right)^{1/\max\{p,q\}} \\ &\lesssim \max\left\{ (p-1)^{1/q-1}, K_q(X) \right\} \left(\int_G \|\nabla f(h)\|_{\ell_2^k(X)}^p \, d\mu(h) \right)^{1/p}. \end{aligned} \tag{2.1}$$

In particular, when $p = q$,

$$\left(\int_1^\infty \int_G \left(\frac{\|f(hv^t) - f(h)\|_X}{t^{1/\rho}} \right)^q d\mu(h) \frac{dt}{t} \right)^{1/q} \lesssim K_q(X) \left(\int_G \|\nabla f(h)\|_{\ell_2^k(X)}^q \, d\mu(h) \right)^{1/q}. \tag{2.2}$$

Note that the v described in this theorem exists, for example, if we choose $v \in \underbrace{[[G, G], \cdots, G]}_{s \text{ times}} \setminus \{e_G\}$ where s is the nilpotency step of G, i.e., it is the largest integer s such that $\underbrace{[[\mathfrak{g}, \mathfrak{g}], \cdots, \mathfrak{g}]}_{s \text{ times}} \neq 0$, then we may normalize v so that $d_G(v^t, e_G) \leq t^{1/s}$ for

$t > 0$. We remark that $\exp\left(\underbrace{[[\mathfrak{g},\mathfrak{g}],\cdots,\mathfrak{g}]}_{j \text{ times}}\right) = \underbrace{[[G,G],\cdots,G]}_{j \text{ times}}$ for $j = 1,\cdots,s$, where $[\mathfrak{g},\mathfrak{g}]$ denotes the Lie algebra bracket and $[G,G]$ denotes the commutator subgroup.

We next state the inequality on groups of polynomial growth.

Theorem 2.0.6. *Let Γ be a not virtually abelian finitely generated group of polynomial growth. There exist $v_\Gamma \in \Gamma$, $s \in \mathbb{N}$ with $s \geq 2$, and $c = c(G) \in \mathbb{N}$ such that the following is true. First, $d_W(v_\Gamma^n, e_\Gamma) \asymp_\Gamma n^{1/s}$ for $n \in \mathbb{N}$. Second, let $p \in (1,\infty)$ and $q \in [2,\infty)$. Suppose that $(X, \|\cdot\|_X)$ is a Banach space satisfying $K_q(X) < \infty$. Then for every $n \in \mathbb{N}$ and every $f : \Gamma \to X$ we have*

$$\begin{aligned}&\left(\sum_{k=1}^{n^s} \frac{1}{k^{1+\max\{p,q\}/s}} \left(\sum_{x\in B_n^\Gamma} \|f(xv_\Gamma^k) - f(x)\|_X^p\right)^{\max\{p,q\}/p}\right)^{1/\max\{p,q\}} \\ &\lesssim_{\Gamma,v_\Gamma} \max\left\{(p-1)^{1/q-1}, K_q(X)\right\} \left(\sum_{x\in B_{cn}^\Gamma}\sum_{a\in S} \|f(xa) - f(x)\|_X^p\right)^{1/p}.\end{aligned} \tag{2.3}$$

In particular, when $p = q$,

$$\left(\sum_{k=1}^{n^s}\sum_{x\in B_n^\Gamma} \frac{\|f(xv_\Gamma^k) - f(x)\|_X^q}{k^{1+q/s}}\right)^{1/q} \lesssim_{\Gamma,v_\Gamma} K_q(X) \left(\sum_{x\in B_{cn}^\Gamma}\sum_{a\in S} \|f(xa) - f(x)\|_X^q\right)^{1/q}. \tag{2.4}$$

Our choice of s and v_Γ is as follows. By [30], Γ admits a subgroup Γ' of finite index that is nilpotent. Let T be the torsion subgroup of Γ' and consider the quotient subgroup $\Gamma'' = \Gamma'/T$. We may take any $v'' \in Z(\Gamma'') \setminus \{e_{\Gamma''}\}$ with the property that $d_W((v'')^n, e_{\Gamma''}) \asymp_{\Gamma''} n^{1/s}$, $n \geq 2$, for some integer $s \geq 2$, and take v_Γ to be any representative of v''. For example, we may take s to be the nilpotency step of Γ'' and $v'' \in \underbrace{[[\Gamma'',\Gamma''],\cdots,\Gamma'']}_{s \text{ times}} \setminus \{e_{\Gamma''}\}$. (See Section 2.3 for details.)

One can see that changing to another finite symmetric generating set S of Γ or changing the vertical element v_Γ affects only $c \in \mathbb{N}$ and the constant in the inequalities

(2.3) and (2.4) up to constant factors depending on S and v_Γ.

Theorems 2.0.5 and 2.0.6 are extensions from the real Heisenberg groups $\mathbb{H}^{2k+1}$ to general simply connected nilpotent Lie groups and from the discrete Heisenberg groups $\mathbb{H}_{\mathbb{Z}}^{2k+1}$ to general finitely generated groups of polynomial growth, respectively, of the "vertical versus horizontal inequalities" established by Austin, Naor, and Tessera [4] and Lafforgue and Naor [47]. See also Naor and Young [64, 65] for the endpoint case $p = 1$ for the Heisenberg groups. Here, the real Heisenberg groups $\mathbb{H}^{2k+1}$, $k \in \mathbb{Z}_{>0}$, are defined to be the simply connected nilpotent Lie groups whose Lie algebras $\mathfrak{h}^{2k+1}$ are spanned by the $2k+1$ elements $x_1, \cdots, x_k, y_1, \cdots, y_k$, and z, the only nontrivial bracket relations among which are

$$[x_i, y_i] = z, \quad i = 1, \cdots, k.$$

The discrete Heisenberg groups $\mathbb{H}_{\mathbb{Z}}^{2k+1}$, $k \geq 1$, embed naturally in the real Heisenberg groups $\mathbb{H}^{2k+1}$.

The inequality (2.1) is proven following the argument of Lafforgue and Naor [47] by comparing the left-hand side of (2.1) against convolutions of ∇f with derivatives of the heat kernel on the real line $\mathbb{R}$, and then upper bounding this quantity using the vector-valued Littlewood–Paley–Stein theory of [40, 56, 82]. The inequality (2.3) follows from inequality (2.1) by a discretization argument.

The utility of Theorems 2.0.5 and 2.0.6 is that they prove quantitatively that nonabelian simply connected nilpotent Lie groups and not virtually abelian finitely generated groups of polynomial growth, respectively, do not bilipschitzly embed into uniformly convex spaces. More precisely, they give Theorems 2.0.1 and 2.0.2 as Corollaries; see Section 2.2 for proofs.

We will prove Theorem 2.0.5 in Section 2.1. Actually, Theorem 2.0.5 will be a special case of the more general Theorem 2.1.1 and Corollary 2.1.2, and the resulting

Theorem 2.0.1 will be a special case of the more general Theorem 2.1.3 and Corollary 2.1.4.

It would appear that we may also derive similar constraints for other coarse embeddings (see [67, Definition 1.36] for the definition). For example, we have the following theorem, where for a metric space (M, d_M) and $0 < \alpha < 1$, (M, d_M^α) is a metric space called the *α-snowflake* of (M, d_M).

Theorem 2.0.7. *Let Γ be a not virtually abelian finitely generated group of polynomial growth, and let X be a q-uniformly convex Banach space. Then*

$$c_X(\Gamma, d_W^{1-\varepsilon}) \gtrsim_\Gamma \frac{1}{K_q(X)\varepsilon^{1/q}}.$$

In particular,

$$c_p(\Gamma, d_W^{1-\varepsilon}) \asymp_{p,\Gamma} \varepsilon^{-1/\max\{p,2\}}.$$

Indeed, by (2.4), any D-bilipschitz embedding $(\Gamma, d_W^{1-\varepsilon}) \to X$ must satisfy for all n

$$|B_n^\Gamma|^{1/q} \left(\sum_{k=1}^{n^s} \frac{1}{k^{1+\varepsilon q/s}} \right)^{1/q} \lesssim_\Gamma K_q(X) D |S|^{1/q} |B_{cn}^\Gamma|^{1/q}$$

which gives the stated lower bound. The second assertion of Theorem 2.0.10, namely the estimate for $c_p(N_{r_1,r_2})$, follows from an optimized version [50, Theorem 5.1] of the Assouad embedding theorem [3]:

$$c_p(\Gamma, d_W^{1-\varepsilon}) \lesssim_{G,p} 1/\varepsilon^{1/\max\{p,2\}}, \quad 0 < \varepsilon < 1.$$

Using Theorem 2.0.6, we may also characterize the compression rate of not virtually abelian finitely generated groups of polynomial growth into L^p spaces. Here, the *compression rate* of a Lipschitz function $f : (M, d_M) \to (X, \|\cdot\|_X)$ from a metric space (M, d_M) into a Banach space $(X, \|\cdot\|_X)$ is the largest nondecreasing function $\omega_f : (0, \infty) \to [0, \infty)$ such that for all $x, y \in M$ we have $\|f(x) - f(y)\|_X \geq \omega_f(d_M(x, y))$.

Corollary 2.0.8. *Let Γ be a not virtually abelian finitely generated group of polynomial growth.*

1. *For $n \in \mathbb{N}$, $n \geq 2$, a nondecreasing function $\theta : (0,\infty) \to [0,\infty)$ satisfies $\theta(t) \lesssim \omega_f(t)$ for some 1-Lipschitz function $f : B_n^\Gamma \to L^p$, $p > 1$, if and only if*

$$\int_1^{2n} \left(\frac{\theta(t)}{t}\right)^{\max\{p,2\}} \frac{dt}{t} \lesssim_\Gamma 1.$$

2. *A nondecreasing function $\theta : (0,\infty) \to [0,\infty)$ satisfies $\theta(t) \lesssim \omega_f(t)$ for some 1-Lipschitz function $f : \Gamma \to L^p$, $p > 1$, if and only if*

$$\int_1^{\infty} \left(\frac{\theta(t)}{t}\right)^{\max\{p,2\}} \frac{dt}{t} \lesssim_\Gamma 1.$$

The if direction is due to [78, Theorem 1]. The only if direction is given by Theorem 2.0.6, from which it follows that

$$s\int_1^{2n} \left(\frac{\theta(t)}{t}\right)^{\max\{p,2\}} \frac{dt}{t} = \int_1^{(2n)^s} \frac{\theta(\tau^{1/s})^{\max\{p\}}}{\tau^{1+\max\{p\}/s}} d\tau \lesssim \sum_{k=1}^{(2n)^s} \frac{\theta(k^{1/s})^{\max\{p\}}}{k^{1+\max\{p,2\}/s}} \lesssim_\Gamma 1.$$

Under a slightly more abstract setting, one can obtain the following statement about the compression function.

Theorem 2.0.9. *Let Γ be an amenable group with finite generating set S, with $v \in Z(\Gamma)$ and $\rho > 1$ such that $d_W(v^k, e_\Gamma) \asymp_\Gamma k^{1/\rho}$, $k \in \mathbb{N}$. Let $(X, \|\cdot\|_X)$ be a q-uniformly convex space, and let $f : \Gamma \to X$ be a 1-Lipschitz function. Then for every $t \geq 3$ there exists an integer $t \leq n \leq t^2$ such that*

$$\frac{\omega_f(n)}{n} \lesssim_\Gamma K_q(X) \left(\frac{\log\log n}{\log n}\right)^{1/q}.$$

(Here we consider ρ to be dependent on Γ, so that $\lesssim_\Gamma$ includes dependence on ρ.)

We prove Theorem 2.0.9 in Section 2.4 following the argument of [4]. Theorem 2.0.9 is somewhat weaker than our distortion result of Theorem 2.0.2 and the vertical versus horizontal inequality of Theorem 2.0.6 under the setting of groups of polynomial growth, because the distortion bound one can obtain from Theorem 2.0.9 is of the weaker form $c_X(B_n) \gtrsim_\Gamma \frac{1}{K_q(X)} \left(\frac{\log n}{\log\log n}\right)^{1/q}$ (this deduction follows the argument of [4, Section 6], and requires the assumption that Γ be of polynomial growth).

On a different note, we can investigate the distortion of nets in nonabelian simply connected nilpotent Lie groups. In the following, with $r_1, r_2 > 0$, an *r_1-covering* of a metric space (M, d_M) is a subset $N \subset M$ such that for any $x \in M$ there exists $n_x \in N$ such that $d_M(x, n_x) < r_1$, an *r_2-packing* of (M, d_M) is a subset $N \subset M$ such that for all distinct $n_1, n_2 \in N$ we have $d_M(n_1, n_2) \geq r_2$, and an *(r_1, r_2)-net* of (M, d_M) is a subset which is both an r_1-covering and an r_2-packing.

Theorem 2.0.10. *Let $r_1, r_2 > 1$ with $r_1 \geq 2r_2$. Let G be a nonabelian simply connected nilpotent Lie group, and let N_{r_1,r_2} be an r_2-covering of B_{r_1}. Let X be a q-uniformly convex space, $q \geq 2$. Then*

$$c_X(N_{r_1,r_2}) \gtrsim \left(\frac{C_1 s^{1/q}}{\sqrt{k} K^{2/q} \log K}\right) \frac{(\log(r_1/r_2))^{1/q}}{K_q(X)},$$

where K is the largest of the doubling constant of G and the doubling constant of the measure μ, i.e., it is the smallest constant such that $\mu(B_{2r}) \leq K\mu(B_r)$ for all $r \geq 0$, and that for all $r > 0$ there exist $y_1, \cdots, y_K \in G$ such that $B_{2r} \subset \cup_{i=1}^K B_r(y_i)$, and $C_1 > 0$ denotes a constant such that $C_1 t^{1/s} \leq d(v^t, e_G) \leq t^{1/s}$ for $t \geq 1$.

In particular, if N_{r_1,r_2} is an $(r_2, \Omega(r_2))$-net of B_{r_1}, then

$$c_p(N_{r_1,r_2}) \asymp_{G,p} (\log(r_1/r_2))^{1/\max\{p,2\}}, \quad 1 < p < \infty.$$

We remark that the second assertion of Theorem 2.0.10, namely the estimate for $c_p(N_{r_1,r_2})$, follows from an optimized version [50, Theorem 5.1] of the Assouad

embedding theorem [3]:

$$c_p(G, d_G^{1-\varepsilon}) \lesssim_{G,p} 1/\varepsilon^{1/\max\{p,2\}}, \quad 0 < \varepsilon < 1.$$

When N_{r_1,r_2} is an $\Omega(r_2)$-packing of B_{r_1}, $(N_{r_1,r_2}, d_G^{1-1/2\log(r_1/r_2)})$ is $O(1)$-bilipschitz equivalent to (N_{r_1,r_2}, d_G), so we have the second assertion of Theorem 2.0.10.

Thus, the bound of Theorem 2.0.10 is sharp for $X = L^p$, and therefore is the best result we can attain in terms of q-uniformly convex spaces. These distortion bounds generalizes those established for the Heisenberg group in [47]. The same conclusion holds with L^p replaced by the Schatten class S_p, whose modulus of uniform convexity was computed in [79].

We may also derive the following variant of Theorem 2.0.10 for snowflakes.

Theorem 2.0.11. *Let N'_{c_1,c_2} be a (c_1, c_2)-net of a nonabelian simply connected nilpotent Lie group G, where $c_1, c_2 \geq 1$ with $c_1 \geq c_2/2$. Let X be a q-uniformly convex space. Then*

$$c_X(N'_{c_1,c_2}, d_G^{1-\varepsilon}) \gtrsim \left(\frac{C_1^{1-\varepsilon}(c_2/c_1)^{\varepsilon} s^{(1-\varepsilon)/q}}{k^{(1-\varepsilon)/2} K^{2(1-\varepsilon)/q} \log K} \right) \frac{1}{K_q(X)^{1-\varepsilon} \varepsilon^{1/q}}, \quad 0 < \varepsilon < 1,$$

where K and C_1 are as in Theorem 2.0.10. In particular,

$$c_p(N'_{c_1,c_2}, d_G^{1-\varepsilon}) \asymp_{G,p,c_1,c_2} \varepsilon^{-1/\max\{p,2\}}, \quad 1 < p < \infty.$$

The nonembeddability of nonabelian simply connected nilpotent Lie groups and not virtually abelian finitely generated groups of polynomial growth into uniformly convex Banach spaces is known from the fact that the asymptotic cone of these groups are nonabelian Carnot groups, and that nonabelian Carnot groups do not admit bilipschitz embeddings into uniformly convex Banach spaces. The definition of Carnot groups is as follows.

Definition 2.0.12 ([49]). *A* (sub-Riemannian) Carnot group *is a 5-tuple* $(G, \delta_\lambda, B, |\cdot|, d_G)$*, where:*

- *The simply connected nilpotent Lie group* G *is such that its Lie algebra* $\mathfrak{g}$ *admits an* s-step stratification*,* $s \in \mathbb{Z}_{>0}$*, i.e., a direct sum decomposition*

$$\mathfrak{g} = V_1 \oplus V_2 \oplus \cdots \oplus V_s,$$

  where $V_s \neq 0$*,* $V_{s+1} = 0$*, and* $V_{r+1} = [V_1, V_r]$ *for* $r = 1, \cdots, s$*.*

- *For each* $\lambda \in \mathbb{R}^+$*, the linear map* $\delta_\lambda : \mathfrak{g} \to \mathfrak{g}$ *is defined by*

$$\delta_\lambda|_{V_r} = \lambda^i \mathrm{id}_{V_r}, \quad r = 1, \cdots, s.$$

- *The bundle* B *over* G *is the extension of* V_1 *to a left-invariant subbundle:*

$$B_p \coloneqq (dL_p)_e V_1, \quad p \in G.$$

- *The inner product norm*[3] $|\cdot|$ *is initially defined on* V_1*, and is then extended to* B *as a left-invariant norm:*

$$|(dL_p)_e(v)| \coloneqq |v|, \quad p \in G,\ v \in V_1.$$

- *The metric* d_G *on* G *is the Carnot–Carathéodory distance associated to* B *and* $|\cdot|$*, i.e., for each* $p, q \in G$ *we define*

$$d_G(p,q) \coloneqq \inf\left\{ \int_0^1 |\dot\gamma(t)| dt : \gamma \in C^\infty_{\mathrm{pw}}([0,1]; G), \gamma(0) = p, \gamma(1) = q, \dot\gamma \in B \right\},$$

[3]For sub-Finsler Carnot groups we allow $|\cdot|$ to be more generally a norm. Then, compared to the sub-Riemannian case, the resulting distance d_G is then distorted by a factor of at most $\sqrt{\dim V_1}$ by the John ellipsoid theorem [43], and the results of this paper follow up to multiplicative factors of $\sqrt{\dim V_1}$.

where $C^\infty_{\mathrm{pw}}([0,1];G)$ *consists of the piecewise smooth functions from* $[0,1]$ *to* G.

For simplicity, we will call G *the Carnot group. When we wish to emphasize the step size, we will call* G *an* s*-step Carnot group or a Carnot group of step* s.

It is easy to see that commutative Carnot groups are precisely those of step 1 and are the Euclidean spaces $\mathbb{R}^d$. The nonabelian ones are those of step $s \geq 2$. An example of this are the Heisenberg groups $\mathbb{H}^{2k+1}$ defined above, which are Carnot groups of step 2 whose Lie algebras $\mathfrak{h}^{2k+1}$ admit the 2-step stratification

$$V_1(\mathfrak{h}) = \operatorname{span}\{x_1, \cdots, x_k, y_1, \cdots, y_k\}, \quad V_2(\mathfrak{h}) = \operatorname{span}\{z\}.$$

Another example of Carnot groups is the aforementioned model filiform groups $J^{s-1}(\mathbb{R})$, $s \geq 1$, which are Carnot groups of step s whose Lie algebras $\mathfrak{j}^{s-1}$ admit the s-step grading

$$V_1(\mathfrak{j}^{s-1}) = \operatorname{span}\{x, y_0\},\ V_2(\mathfrak{j}^{s-1}) = \operatorname{span}\{y_1\},\ \cdots,\ V_s(\mathfrak{j}^{s-1}) = \operatorname{span}\{y_{s-1}\}.$$

Let our choice of left-invariant vector fields $X_1, \cdots, X_k$, $k \coloneqq \dim V_1$, be such that $(X_1)_{e_G}, \cdots, (X_k)_{e_G}$ form an orthonormal basis of $(V_1, |\cdot|)$. We have the center $Z(G) = \exp(V_s)$ and commutator subgroup $[G,G] = \exp(V_2 \oplus \cdots \oplus V_s)$. Also, the maps δ_λ act as scalings in the Carnot–Carathéodory distance, in the sense that $d_G(\delta_\lambda(p), \delta_\lambda(q)) = \lambda d_G(p,q)$, for $\lambda > 0$ and $p, q \in G$.

In an algebraic sense, nonabelian Carnot groups are nonabelian simply connected nilpotent Lie groups, but their metrics differ, since we endowed nilpotent Lie groups with *Riemannian* distances while we endowed Carnot groups with *sub-Riemannian* distances. Nevertheless, Theorem 2.0.5 holds for nonabelian Carnot groups G, stated as follows.

Theorem 2.0.13. *Let* G *be a nonabelian Carnot group of step* $s \geq 2$. *Let* $v \in Z(G) \setminus$

$\{e_G\}$ *be normalized so that* $d_G(v, e_G) = 1$. *Suppose that* $p \in (1, \infty)$ *and* $q \in [2, \infty)$. *Let* $(X, \|\cdot\|_X)$ *be a Banach space with* $K_q(X) < \infty$, *and let* $f : G \to X$ *be smooth and compactly supported. Then*

$$\begin{aligned} &\left(\int_0^\infty \left(\int_G \left(\frac{\|f(hv^t) - f(h)\|_X}{t^{1/s}}\right)^p d\mu(h)\right)^{\max\{p,q\}/p} \frac{dt}{t}\right)^{1/\max\{p,q\}} \\ &\lesssim \max\left\{(p-1)^{1/q-1}, K_q(X)\right\} \left(\int_G \|\nabla f(h)\|_{\ell_2^k(X)}^p \, d\mu(h)\right)^{1/p}. \end{aligned} \tag{2.5}$$

In particular, when $p = q$,

$$\left(\int_0^\infty \int_G \left(\frac{\|f(hv^t) - f(h)\|_X}{t^{1/s}}\right)^q d\mu(h)\frac{dt}{t}\right)^{1/q} \lesssim K_q(X) \left(\int_G \|\nabla f(h)\|_{\ell_2^k(X)}^q \, d\mu(h)\right)^{1/q}. \tag{2.6}$$

For nilpotent Lie groups, we have $d_G(v^t, e_G) \asymp_G t$ for $0 < t < 1$ and $d_G(v^t, e_G) \asymp_G t^{1/s}$ for $t \geq 1$ for our choice of v, while for Carnot groups we have $d_G(v^t, e_G) = t^{1/s}$ for all $t > 0$. Thus, Theorem 2.0.5 only gives global nonembeddability of nonabelian simply connected nilpotent Lie groups into uniformly convex spaces, i.e., we only know that the entire group fails to embed into X, while Theorem 2.0.13 gives local nonembeddability of Carnot groups into uniformly convex spaces, i.e., we know also that any nonempty open subset of the group fails to embed into X.

Although Theorems 2.0.5 and 2.0.13 have a formal difference in the sense that Theorem 2.0.5 considers Riemannian distances on nilpotent Lie groups while Theorem 2.0.13 considers sub-Riemannian distances on Carnot groups, it will turn out that they are instances of a more general theorem regarding sub-Riemannian distances on nilpotent Lie groups. This is stated as Corollary 2.1.2 in Section 2.1. See Corollary 2.1.4 for the resulting distortion results.

Theorem 2.0.6 applies to cocompact lattices Γ of a nonabelian Carnot group G, because they are not virtually abelian finitely generated groups of polynomial growth. We can take s to be the step of G and $v_\Gamma \in Z(G) \cap \Gamma \setminus \{e_G\}$; we will not repeat the

statement of Theorem 2.0.6 here.

We may state simplified versions of Theorems 2.0.10 and 2.0.11 in the setting of Carnot groups, thanks to the scale-invariance. In the following, $n_h \coloneqq \sum_{r=1}^{s} rk_r$ is the Hausdorff dimension of G

Theorem 2.0.14. *Let $r_1, r_2 > 0$ with $r_1 \geq 2r_2$. Let G be a nonabelian Carnot group, and let N_{r_1,r_2} be an r_2-covering of B_{r_1}. Let X be a q-uniformly convex space, $q \geq 2$. Then*

$$c_X(N_{r_1,r_2}) \gtrsim \left(\frac{C_1 s^{1/q}}{\sqrt{k} 4^{n_h/q} n_h}\right) \frac{(\log(r_1/r_2))^{1/q}}{K_q(X)}.$$

In particular, if N_{r_1,r_2} is an $(r_2, \Omega(r_2))$-net of B_{r_1}, then

$$c_p(N_{r_1,r_2}) \asymp_{G,p} (\log(r_1/r_2))^{1/\max\{p,2\}}, \quad 1 < p < \infty.$$

Theorem 2.0.15. *Let N'_{c_1,c_2} be a (c_1, c_2)-net of a nonabelian Carnot group G, where $c_1, c_2 > 0$ with $c_1 \geq c_2/2$. Let X be a q-uniformly convex space. Then*

$$c_X(N'_{c_1,c_2}, d_G^{1-\varepsilon}) \gtrsim \left(\frac{C_1^{1-\varepsilon}(c_2/c_1)^{\varepsilon} s^{(1-\varepsilon)/q}}{k^{(1-\varepsilon)/2} 4^{n_h(1-\varepsilon)/q} n_h}\right) \frac{1}{K_q(X)^{1-\varepsilon} \varepsilon^{1/q}}, \quad 0 < \varepsilon < 1.$$

In particular,

$$c_p(N'_{c_1,c_2}, d_G^{1-\varepsilon}) \asymp_{G,p,c_1,c_2} \varepsilon^{-1/\max\{p,2\}}, \quad 1 < p < \infty.$$

That a nonabelian Carnot group G fails to embed bilipschitzly into $\mathbb{R}^n$ was first proven by Semmes [73] using the differentiation theorem of Pansu [69]. Pansu's theorem states that any Lipschitz mapping $f : G \to \mathbb{R}^n$ is differentiable a.e. with the derivative being a Lie group homomorphism that commutes with the dilations δ_λ of G and $\mathbb{R}^n$. In other words, one can 'blow-up' f almost everywhere and the resulting 'blow-up' function is a Carnot group homomorphism. Semmes' observation was that if f were a *bilipschitz* mapping, then the blow-up of f would also be bilipschitz,

resulting in a Carnot group *monomorphism*. This is impossible because a group homomorphism $G \to \mathbb{R}^n$ must send the nontrivial commutator subgroup $[G, G]$ to 0. Later this argument was generalized, independently by Cheeger and Kleiner [11] and Lee and Naor [52], to Banach space targets with the Radon–Nikodým property, which includes uniformly convex spaces.

From the standpoint of the Pansu–Semmes argument which simply outputs the qualitative nonexistence of a bilipschitz mapping $G \to \mathbb{R}^n$, the significance and advantage of Theorems 2.0.5, 2.0.6, and 2.0.13 is that they give quantitative nonembeddability statements, such as Theorems 2.0.1, 2.0.2, 2.0.10, 2.0.11, 2.0.14, and 2.0.15. However, the Pansu–Semmes argument tells us that any Lipschitz function $f : G \to X$ must 'collapse' along directions of the commutator group $[G, G]$, and from this point of view, Theorem 2.0.13 possesses a curious limitation, namely that it requires us to measure the collapse along *central* directions.

This limitation is caused by a technical requirement in the proof of Theorems 2.0.5 and 2.0.13, which is for the horizontal derivative and the convolution along the direction of v to commute (more precisely, equation (2.28) below). Since there is no such limitation in the original Pansu–Semmes proof, we pose the following question.

Question 2.0.16. *Let G be a nonabelian Carnot group. Is Theorem 2.0.13 true for all $v \in [G, G]$ with $d_G(v, e_G) = 1$?*

We answer this question in the affirmative for L^p targets, albeit with some loss of control on the constants.

Theorem 2.0.17. *Let G be a nonabelian Carnot group. For $1 < p \leq 2$, $f \in L^p(G)$ with $\nabla f \in L^p(G; \ell_2^k)$, and $v \in [G, G]$ with $d_G(v, e) = 1$, we have*

$$\left(\int_0^\infty \left[\int_G \left(\frac{|f(h) - f(h\delta_r(v))|}{r} \right)^p d\mu(h) \right]^{2/p} \frac{dr}{r} \right)^{1/2} \lesssim_{G,p} \|\nabla f\|_{L^p(G;\ell_2^k)}. \tag{2.7}$$

Although Theorem 2.0.17 is stated for $\mathbb{R}$-targets, it is easy to tensorize inequality (2.7) using the triangle inequality when $f : G \to L^p(\sigma)$ is smooth and compactly supported.

Remark 2.0.18. *With q defined as in Theorem 2.0.13, Theorem 2.0.17 only considers the exponent $q = 2$. This is because the case $q = \infty$ formally holds, so that the inequality for $q = 2$ implies the inequalities for $q \geq 2$. See Section 2.1, page 175 for a proof of this simple fact. We cannot make this simplification in Theorem 2.0.13 or in Theorem 2.0.5 due to the stated dependence of the constants on q.*

However, we have not managed to prove a strengthening of Theorem 2.0.6 to directions in the commutator group for Carnot groups, formulated as follows.

Question 2.0.19. *Let Γ be a cocompact lattice of a nonabelian Carnot group G. For any $v_\Gamma \in [\Gamma, \Gamma] \setminus \{e_G\}$, does there exist $c \in \mathbb{N}$ such that the following is true? If we let $s' \geq 2$ be the largest integer such that $v_\Gamma \in \underbrace{[[G, G], \cdots, G]}_{s' \text{ times}}$, let $p \in (1, \infty)$ and $q \in [2, \infty)$, and suppose $(X, \|\cdot\|_X)$ is a Banach space satisfying $K_q(X) < \infty$, then for every finitely supported $f : \Gamma \to X$ we have*

$$\left(\sum_{k=1}^{n^{s'}} \frac{1}{k^{1+\max\{p,q\}/s'}} \left(\sum_{x \in B_n^\Gamma} \|f(xv_\Gamma^k) - f(x)\|_X^p \right)^{\max\{p,q\}/p} \right)^{1/\max\{p,q\}}$$

$$\lesssim_{\Gamma, v_\Gamma} \max\left\{(p-1)^{1/q-1}, K_q(X)\right\} \left(\sum_{x \in B_{cn}^\Gamma} \sum_{a \in S} \|f(xa) - f(x)\|_X^p \right)^{1/p}.$$

We will deduce Theorem 2.0.17 from a Dorronsoro theorem for Carnot groups following the argument of Fässler and Orponen [22]. We will also prove the Carnot group Dorronsoro theorem itself, which states that the $L^p(G)$ norm of the L^p fractional Laplacian $\|(-\Delta_p)^\alpha f\|_{L^p(G)}$ ($1 < p < \infty$, $\alpha > 0$) of a function $f \in L^p(G)$, is equivalent up to constant factors to a singular integral that, roughly speaking, measures, at all points of G and at all scales, the deviation of f from being a polynomial of weighted

degree $\leq \lfloor \alpha \rfloor$. This was first proven for $G = \mathbb{R}^n$ for all $\alpha > 0$ by Dorronsoro [19], and for the Heisenberg groups $\mathbb{H}^{2k+1}$ one side of the Dorronsoro statement (namely that the aforementioned singular integral is bounded above by $\|(-\Delta_p)^\alpha f\|_{L^p}$) was proven for the restricted exponent range $0 < \alpha < 2$ in [22]. The reason for this restriction $0 < \alpha < 2$ in [22] is that they considered deviations from the smaller class of "horizontal polynomials", which is inadequate for $\alpha \geq 2$ (see Remark 2.0.23 for a proof).

We will discuss the Dorronsoro theorem for Carnot groups in detail later in subsection 2.0.1. To summarize, the novelty of this paper in this direction is threefold. First, we recognize the correct class of polynomials to approximate by, namely polynomials of weighted degree that depend on the "full set of coordinates" as opposed to just the "horizontal coordinates". Second, we recover the full Dorronsoro theorem, i.e., we prove both directions of the equivalence. Third, we verify that generalizing to higher step Carnot groups introduces no serious problems. The proof method of Theorem 2.0.22 is based on Dorronsoro's original proof in [19] and adds various ingredients used in Fässler and Orponen's proof in [22].

Remark 2.0.20. *So far, we have mentioned two proof methods of the vertical versus horizontal inequality, namely the Littlewood–Paley–Stein theory for Theorems 2.0.5 and 2.0.13 and the Dorronsoro theorem for the variant Theorem 2.0.17. There is a third possible proof method for L^2 targets which follows the representation theoretic proof of [4] for the case $G = \mathbb{H}^3$, $X = L^2$. Specifically, one can reduce proving the vertical versus horizontal inequality to proving a certain inequality regarding 1-cocycles $G \to L^2$ of irreducible unitary representations of G. The irreducible unitary representations of $\mathbb{H}^3$ are given by the Stone–von Neumann theorem, while the irreducible unitary representations have been described by Dixmier [18, 17] and Kirillov [45]. We will not pursue this proof method since the anticipated results do not seem to provide an improvement or advantage over neither Theorem 2.0.5 nor Theorem 2.0.17.*

Focusing on Euclidean distortion for concreteness, Theorem 2.0.14 tells us that

$$c_2(N_R) \asymp_{n_h} \sqrt{\log R},$$

for any Carnot group G and $(1, \Omega(1))$-net N_R of B_R. One may have expected that Carnot groups with higher steps may have a 'hierarchy' of collapsing happening, with distortions being propagated and amplified as one passes from $[G, G]$, $[G, [G, G]]$, all the way up to $\underbrace{[G, [G, \cdots, G]]}_{s \text{ times}}$, with the center ending up "super-collapsed"; certainly if G is high-dimensional and V_1 is low dimensional, the collapsing happens along the subgroup $[G, G]$ of small codimension, and one may have enough room to have many more interesting phenomena such as collapsing happen. In this case, the dependency on s of the asymptotics of $c_2(N_R)$ would appear in the exponent of $\log R$, but actually the asymptotics is exactly that of $\sqrt{\log R}$. Thus, it seems that the collapsing happens "only once" and happens "uniformly across all of $[G, G]$". The algebraic structure of the Carnot group seems to only affect the constants involved while having no effect on the exponent. We then ask what is the correct dependence on the Lie group structure:[4]

Question 2.0.21. *In the case of Euclidean distortion, what are the precise asymptotics? More precisely, we ask the following.*

1. *Let G be a nonabelian simply connected nilpotent Lie group. Does there exist a constant c_G such that we have the following?*

$$c_2(B_r) \asymp c_G \sqrt{\log r}, \quad r \geq 2.$$

Similarly, for $r_1, r_2 > 1$ with $r_1 \geq 2r_2$, let N_{r_1,r_2} be an $(r_2, \Omega(r_2))$-net of B_{r_1}. Does there exist a constant c'_G such that we have the following?

$$c_2(N_{r_1,r_2}) \asymp c'_G \sqrt{\log(r_1/r_2)} \quad \text{as } R \to \infty.$$

[4] I thank Professor Assaf Naor for discussion on this matter and asking this question.

If so, how are the values of c_G and c'_G determined by the algebraic structure of G?

2. *Let Γ be a not virtually abelian finitely generated group of polynomial growth. Does there exist a constant c_Γ such that we have the following?*

$$c_2(B_n^\Gamma) \asymp c_\Gamma \sqrt{\log n} \quad \text{as } n \to \infty.$$

If so, how is the value of c_Γ determined by the algebraic structure of Γ?

We know that, for Carnot groups G, c_G is bounded above and below by functions of n_h, since the constants appearing in the upper bound due to the Assouad embedding theorem and in the lower bound due to Theorem 2.0.14 can all be made depending on n_h.

Note that the Carnot groups of lowest nonabelian step $s = 2$, such as the Heisenberg group $\mathbb{H}^3$, already give the worst dependence $\sqrt{\log R}$. We however remark that the nonembeddability statements, both qualitative and quantitative, for Carnot groups into uniformly convex spaces do not formally follow from those of the Heisenberg group $\mathbb{H}^3$, because it is not true that any nonabelian Carnot group contains a bilipschitz copy of $\mathbb{H}^3$.[5] For example, the aforementioned model filiform spaces $J^{s-1}(\mathbb{R})$ with $s \geq 3$ do not contain a bilipschitz copy of $\mathbb{H}^3$. Recalling the definition of the model filiform group, $J^{s-1}(\mathbb{R})$ admits the s-step grading $V_1(\mathfrak{j}^{s-1}) = \operatorname{span}\{x, y_0\}$, $V_2(\mathfrak{j}^{s-1}) = \operatorname{span}\{y_1\}$, $\cdots$, $V_s(\mathfrak{j}^{s-1}) = \operatorname{span}\{y_{s-1}\}$. If $J^{s-1}(\mathbb{R})$ were to contain a bilipschitz copy of $\mathbb{H}^3$, then by the Pansu–Semmes differentiation theorem there would be a Carnot group monomorphism $\mathbb{H}^3 \to J^{s-1}(\mathbb{R})$, namely a Lie group monomorphism that commutes with the dilations. This would induce a Lie algebra monomorphism $\mathfrak{h}^3 \to \mathfrak{j}^{s-1}$ that sends $V_i(\mathfrak{h}) \to V_i(\mathfrak{j}^{s-1})$ for all i; by matching dimensions, we actually have vector space isomorphisms $V_1(\mathfrak{h}) \xrightarrow{\approx} V_1(\mathfrak{j}^{s-1})$ and $V_2(\mathfrak{h}) \xrightarrow{\approx} V_2(\mathfrak{j}^{s-1})$, but now we

[5] I thank Professors Assaf Naor and Robert Young for this comment.

have a contradiction since this implies we have a surjection

$$0 = V_3(\mathfrak{h}) = [V_1(\mathfrak{h}), V_2(\mathfrak{h})] \to [V_1(\mathfrak{j}^{s-1}), V_2(\mathfrak{j}^{s-1})] = V_3(\mathfrak{j}^{s-1}) \neq 0.$$

Roadmap. The rest of the introduction is organized as follows. We first discuss in detail Dorronsoro's theorem on Carnot groups (Theorem 2.0.22) and describe the resulting more refined fractional vertical versus horizontal inequalities (Theorem 2.0.24) in subsection 2.0.1. In subsection 2.0.2, we present conjectural quantitative nonembeddability statements of nonabelian simply connected nilpotent Lie groups into L^1 (Conjecture 2.0.30) to suggest a candidate behavior for the L^1 distortion of balls in groups of polynomial growth (Question 2.0.31).

After the introduction, the rest of this paper is organized as follows. We begin by proving our main Theorems 2.0.5 and 2.0.13 in Section 2.1, and we then prove the distortion bounds of Theorems 2.0.1, 2.0.2, 2.0.10, 2.0.11, 2.0.14, and 2.0.15 and Corollaries 2.0.3 and 2.0.4 in Section 2.2. We derive the discretized inequalities of Theorem 2.0.6 in Section 2.3. In Section 2.4, we prove Theorem 2.0.9 by analyzing cocycles. We then, in Section 2.5, derive the fractional vertical versus horizontal inequalities of Theorem 2.0.24, and thereby also prove Theorem 2.0.17, from the Dorronsoro Theorem 2.0.22. Finally, we prove the Dorronsoro Theorem 2.0.22 in Section 2.6.

2.0.1 Dorronsoro's theorem on Carnot groups

As mentioned earlier, a byproduct of this paper is a formulation and proof of a Dorronsoro theorem for Carnot groups, which is used to prove Theorem 2.0.17. In this subsection, we state the Dorronsoro theorem and obtain more refined vertical versus horizontal inequalities.

Let G be a Carnot group in this subsection. Following Folland [25] and denoting

by $\Delta = \sum_{i=1}^{k} X_i^2$ the sub-Laplacian and H_t the corresponding heat kernel, we define the operator $(-\Delta_p)^\alpha$ for $1 < p < \infty$ and $\operatorname{Re}\alpha > 0$ by

$$(-\Delta_p)^\alpha f = \lim_{\varepsilon \to 0} \frac{1}{\Gamma(\lfloor \operatorname{Re}\alpha \rfloor + 1 - \alpha)} \int_\varepsilon^\infty t^{\lfloor \operatorname{Re}\alpha \rfloor - \alpha} (-\Delta)^{\lfloor \operatorname{Re}\alpha \rfloor + 1} H_t f \, dt$$

for all $f \in L^p(G)$ such that the limit exists in the L^p norm. Then, for $1 < p < \infty$ and $\alpha \geq 0$, the Sobolev space S_α^p is the Banach space $\operatorname{Dom}((-\Delta_p)^{\alpha/2})$ with norm

$$\|\cdot\|_{p,\alpha} \coloneqq \|\cdot\|_{L^p(G)} + \|(-\Delta_p)^{\alpha/2}(\cdot)\|_{L^p(G)}.$$

Fix a basis $X_{r,1}, \cdots, X_{r,k_r}$, where $k_r = \dim V_r$, of each stratum V_r. As G is nilpotent and simply connected, the exponential map $\exp : \mathfrak{g} \to G$ is a diffeomorphism. Thus, each point $p \in G$ can be expressed in the coordinates $p = \exp\left(\sum_{r=1}^{s} \sum_{i=1}^{k_r} x_{r,i} X_{r,i}\right)$, $x_{r,i} \in \mathbb{R}$; let this be the single coordinate chart on G. For polynomials of the coordinates $x_{r,i}$, we assign weight r to the variable $x_{r,i}$, $r = 1, \cdots, s$, $i = 1, \cdots, k_r$.

Recall that $n_h = \sum_{r=1}^{s} r k_r$ is the Hausdorff dimension of G. We have $n_h \geq 4$, as we must have $s \geq 2$, $k_1 \geq 2$ and $k_2 \geq 1$, since G is nonabelian.

For $r > 0$ and $g \in G$, recall the notation for open balls

$$B_r = \{h \in G : d_G(h, e_G) < r\}, \; B_r(g) = \{h \in G : d_G(h, g) < r\} = gB_r.$$

Recall that δ_λ is a scaling in the Carnot–Carathéodory metric:

$$d_G(\delta_\lambda(p), \delta_\lambda(p')) = \lambda d_G(p, p'), \quad p, p' \in G.$$

Thus $B_r = \delta_r(B_1)$, $r > 0$. Also, by compactness we have

$$d_G(y, e_G) \asymp_G \sum_{r=1}^{s} \sum_{i=1}^{k_r} |y_{r,i}|^{1/r}.$$

Let $d \in \mathbb{Z}_{\geq 0}$ and let $\mathcal{A}_d$ denote the family of polynomials $G \to \mathbb{R}$ of weighted degree d. (Note that this family is left-invariant. Indeed, one can express the group law in this coordinate system using the Baker–Campbell–Hausdorff formula

$$gh = \sum_{m=1}^{s} \frac{(-1)^{m-1}}{m} \sum_{\substack{r_1+s_1>0 \\ \cdots \\ r_m+s_m>0}} \frac{[g^{r_1}h^{s_1}g^{r_2}h^{s_2}\cdots g^{r_m}h^{s_m}]}{(\sum_{j=1}^{m}(r_j+s_j)) \cdot \prod_{i=1}^{m} r_i! s_i!},$$

where the sum is finite since G is of step s, and we have used the following notation:

$$[g^{r_1}h^{s_1}\cdots g^{r_m}h^{s_m}] = [\underbrace{g,[g,\cdots[g}_{r_1},[\underbrace{h,[h,\cdots[h}_{s_1},\cdots[\underbrace{g,[g,\cdots[g}_{r_m},[\underbrace{h,[h,\cdots h}_{s_m}]]\cdots]].$$

Thus, we can see that

$$\exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} x_{r,i}^0 X_{r,i}\right)\exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} x_{r,i}^1 X_{r,i}\right) = \exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} x_{r,i}^2 X_{r,i}\right)$$

where

$$\begin{aligned} x_{r,i}^2 =& x_{r,i}^0 + x_{r,i}^1 \\ &+ (\text{homogeneous polynomial of } \{x_{r',i'}^0\}_{r'<r}, \{x_{r',i'}^1\}_{r'<r} \text{ of weighted degree } r). \end{aligned}$$

Therefore a polynomial of weighted degree d precomposed with a left translation is still of weighted degree d.) This definition of $\mathcal{A}_d$ is in contrast from [22], where they only considered horizontal polynomials, i.e., polynomials that depend only on the 'horizontal coordinates' $x_{1,1}, \cdots, x_{1,k_1}$; here we are allowing for non-horizontal coordinates, provided they satisfy the weighted degree condition.

For a locally integrable function $f : G \to \mathbb{R}$, $x \in G$ and $r > 0$, let $A_{x,r}^d f$ denote

the unique element of $\mathcal{A}_d$ such that

$$\int_{B_r(x)} (f(y) - A^d_{x,r} f(y)) A(y) dy = 0, \quad \forall A \in \mathcal{A}_d.$$

For example, $A^0_{x,r} f = \langle f \rangle_{B_r(x)}$, the average of f on $B_r(x)$, and a formula for $A^1_{x,r} f$ is given below in (2.54). We measure how well $A^d_{x,r} f$ approximates f in the ball $B_r(x)$ by the following quantity:

$$\beta_{f,d,q}(B_r(x)) = \left(\fint_{B_r(x)} |f(y) - A^d_{x,r} f(y)|^q dy \right)^{1/q}, \quad 1 \le q < \infty.$$

Theorem 2.0.22 (Dorronsoro's theorem for Carnot groups)**.** *Let $1 < p < \infty$, $\alpha > 0$, and*

$$1 \le q < \frac{\min\{p, 2\} n_h}{n_h - \min\{p, 2\}}.$$

Then for all $f \in L^p(G)$,

$$\left(\int_G \left(\int_0^\infty \left[\frac{1}{r^\alpha} \beta_{f,\lfloor \alpha \rfloor, q}(B_r(x)) \right]^2 \frac{dr}{r} \right)^{p/2} d\mu(x) \right)^{1/p} \asymp_{G,\alpha,p,q} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}. \quad (2.8)$$

in the sense that $f \in S^p_\alpha(G)$ if and only if the left-hand side of (2.8) *is finite, in which case the above relation holds.*

Our proof of Theorem 2.0.22 is based on Dorronsoro's original proof [19] and adds on ingredients from Fässler and Orponen's proof [22]. Actually, in the case of $\alpha = 1$ and $G = \mathbb{H}^{2k+1}$ there is a simpler proof of Dorronsoro's theorem involving the Fourier transform [5, Subsection 7.3].[6] It seems likely that there would be a similar simpler proof of Theorem 2.0.22 for $\alpha = 1$, but we have not pursued this direction since we can obtain the full range $\alpha > 0$ with our proof.

Remark 2.0.23. *Fässler and Orponen's version [22] of the Dorronsoro theorem [19]*

[6] I thank Ian Fleschler for pointing out this reference.

for Heisenberg groups states the $\lesssim$ portion of the above inequality (2.8) for $f \in S^p_\alpha(G)$ in the case $G = \mathbb{H}^{2k+1}$ and $0 < \alpha < 2$. (They did not prove the $\gtrsim$ portion because it is not needed when proving the vertical versus horizontal inequalities.) The restriction $0 < \alpha < 2$ was necessary because they were approximating by horizontal polynomials, i.e., polynomials that depend only on the horizontal coordinates; in the range $0 < \alpha < 2$ our formulation of Theorem 2.0.22 agrees with that of [22] because then $\mathcal{A}_1$ is precisely the family of horizontal polynomials. To see why the restriction $0 < \alpha < 2$ is necessary when we consider only horizontal polynomials, let $\alpha \geq 2$, and suppose $\mathcal{A}_d$ were defined as the set of horizontal polynomials of degree at most d. In the case of the Heisenberg group $\mathbb{H}^{2k+1}$ the group structure is given by

$$\begin{aligned}
&\exp\left(\sum_{i=1}^{k}(x_i^0\partial_{x_i} + y_i^0\partial_{y_i}) + z^0\partial_z\right) \cdot \exp\left(\sum_{i=1}^{k}(x_i^1\partial_{x_i} + y_i^1\partial_{y_i}) + z^1\partial_z\right) \\
&= \exp\left(\sum_{i=1}^{k}\left((x_i^0 + x_i^1)\partial_{x_i} + (y_i^0 + y_i^1)\partial_{y_i}\right) + \left(z^0 + z^1 + \frac{1}{2}\sum_{i=1}^{k}(x_i^0 y_i^1 - x_i^1 y_i^0)\right)\partial_z\right)
\end{aligned}$$

and thus the left-invariant horizontal vector fields are given by

$$X_i = \frac{\partial}{\partial x_i} + \frac{y_i}{2}\frac{\partial}{\partial z}, \quad Y_i = \frac{\partial}{\partial y_i} - \frac{x_i}{2}\frac{\partial}{\partial z}, \qquad i = 1, \cdots, k,$$

so that the Laplacian is given by

$$\Delta = \sum_{i=1}^{k}(X_i^2 + Y_i^2) = \sum_{i=1}^{k}\left(\frac{\partial^2}{\partial x_i^2} + \frac{\partial^2}{\partial y_i^2} + y_i\frac{\partial^2}{\partial x_i \partial z} - x_i\frac{\partial^2}{\partial y_i \partial z}\right) + \frac{1}{4}\sum_{i=1}^{k}(x_i^2 + y_i^2)\frac{\partial^2}{\partial z^2}.$$

Let $f : \mathbb{H}^{2k+1} \to \mathbb{R}$ be a function which is smooth, supported on B_2 and agrees with the function z on B_1. Clearly $\|(-\Delta)^{\alpha/2} f\|_{L^p(G)}$ is finite. However, for $r \in (0, \frac{1}{2})$, $A^d_{0,r} f = A^d_{0,r} z = 0$ because B_r is symmetric about reflection with respect to the plane $z = 0$; thus

$$\beta_{f,d,q}(B_r) = \left(\fint_{B_r} |z|^q d\mu\right)^{1/q} \asymp_{G,q} r^2.$$

Since $\mathcal{A}_d$ is invariant under left-translation and left-translation is measure-preserving, at

$$p = \exp\left(\sum_{i=1}^{k}(x_i^0\partial_{x_i} + y_i^0\partial_{y_i}) + z^0\partial_z\right) \in B_{1/2}$$

we have

$$A_{p,r}^d f = z^0 + \frac{1}{2}\sum_{i=1}^{k}(x_i^0 y_i - x_i y_i^0), \quad \beta_{f,d,q}(B_r(x)) = \beta_{f,d,q}(B_r) \asymp_{G,q} r^2.$$

Therefore, the left-hand side of (2.8) *is bounded below by*

$$\begin{aligned}
&\left(\int_{B_{1/2}}\left(\int_0^{1/2}\left[\frac{1}{r^\alpha}\beta_{f,\lfloor\alpha\rfloor,q}(B_r(x))\right]^2\frac{dr}{r}\right)^{p/2} d\mu(x)\right)^{1/p}\\
&\asymp_{G,q} \mu(B_{1/2})^{1/p}\left(\int_0^{1/2}\frac{dr}{r^{1+2(\alpha-2)}}\right)^{1/2}\\
&= \infty.
\end{aligned}$$

Thus (2.8) *fails to hold when $\alpha \geq 2$ when $\mathcal{A}_d$ is restricted to horizontal polynomials.*

Of course, the purpose of the Dorronsoro Theorem 2.0.22 in this paper is to prove the vertical versus horizontal inequality of Theorem 2.0.17. The proof of Theorem 2.0.17 from Theorem 2.0.22 is given in Section 2.5, and uses the special case of $\alpha = 1$ of Theorem 2.0.22 along with the fact that $\|(-\Delta_p)^{1/2} f\|_{L^p(G)} \asymp \|\nabla f\|_{L^p(G;\ell_2^k)}$ [14, (52)] to see that it is enough to upper bound the left-hand side of (2.7) by the left-hand side of (2.8).

However, one may wonder what vertical versus horizontal inequalities emerge when we don't specialize to $\alpha = 1$. In this case, we obtain fractional order generalizations of Theorem 2.0.5.

Theorem 2.0.24. *Let $1 < p \leq 2$, $\alpha > 0$, $n \in \mathbb{N}$, and let $f \in S_\alpha^p(G)$. Let $v \in$*

$\exp(V_{\lfloor \alpha/n \rfloor+1} \oplus \cdots \oplus V_s)$ *with* $d_G(v, e_G) = 1$. *Then*

$$\left(\int_0^\infty \left[\int_G \left(\frac{1}{r^\alpha} \left| \Delta^n_{\delta_r(v)} f(h) \right| \right)^p d\mu(h) \right]^{2/p} \frac{dr}{r} \right)^{1/2} \lesssim_{G,p,\alpha,n} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}, \quad (2.9)$$

where for $g \in G$ *and* $F : G \to \mathbb{R}$, $\Delta_g F(x) := F(xg) - F(x)$ *denotes the finite difference.*

The simple example $n = 1$ is given as follows.

Example 2.0.25. *Let* $1 < p \leq 2$, $\alpha > 0$, *and let* $f \in S^p_\alpha(G)$. *Let* $v \in \exp(V_{\lfloor \alpha \rfloor+1} \oplus \cdots \oplus V_s)$ *with* $d_G(v, e_G) = 1$. *Then*

$$\left(\int_0^\infty \left[\int_G \left(\frac{|f(h) - f(h\delta_r(v))|}{r^\alpha} \right)^p d\mu(h) \right]^{2/p} \frac{dr}{r} \right)^{1/2} \lesssim_{G,p,\alpha} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}.$$

The case $G = \mathbb{H}^{2k+1}$, $n = 1$, $0 < \alpha < 2$ of Example 2.0.25 has been obtained in [23].

In the above, the Dorronsoro Theorem 2.0.22 was used for nonembeddings, but there are also other applications. For example, the case of the Heisenberg groups $\mathbb{H}^{2k+1}$, $k \geq 1$, due to [22], is used in the work [13];[7] it stands to reason that our more general result will have similar applications, but we defer this to future investigations.

2.0.2 L^1-distortion: vertical perimeter versus horizontal perimeter

Given the discussion so far, one may ask whether the results of this paper follows for $L^1(\sigma)$-targets. Since ℓ^1 is not uniformly convex, our results in this paper do not apply; in fact, that general nonabelian simply connected nilpotent Lie groups G do not embed bilipschitzly into L^1 spaces was proven only recently by Eriksson-Bique,

[7] I thank Professor Tuomas Orponen for pointing out this fact.

Gartland, Le Donne, Naples, and Nicolussi-Golo [21]. For the Heisenberg group this was proven previously by Cheeger and Kleiner [12].

The results of this paper give quantitative nonembeddability of nonabelian simply connected nilpotent Lie groups into uniformly convex spaces. In this subsection, which closely follows Section 4 of [47], we will present hypothetical analogues of the results of this paper for quantitative nonembeddability into L^1.

Theorem 2.0.5 states in the case $X = \mathbb{R}$ and $p \in (1, q]$ that for every smooth and compactly supported $f : G \to \mathbb{R}$,

$$\begin{aligned} &\left(\int_0^\infty \left(\int_G |f(hv^t) - f(h)|^p d\mu(h) \right)^{q/p} \frac{dt}{t^{1+q/s}} \right)^{1/q} \\ &\lesssim_G (p-1)^{1/q-1} \left(\int_G \|\nabla f(h)\|_{\ell_2^k}^p \, d\mu(h) \right)^{1/p}. \end{aligned} \tag{2.10}$$

The constant $(p-1)^{1/q-1}$ is unbounded as $p \to 1$; nevertheless, we ask whether the endpoint case $p = 1$ of (2.10) does hold true.

Question 2.0.26. *Let G be a nonabelian simply connected nilpotent Lie group, and let $v \in Z(G)$ be as in Theorem 2.0.5. For which exponents $q \geq 1$ does every smooth and compactly supported $f : G \to \mathbb{R}$ satisfy*

$$\left(\int_0^\infty \left(\int_G |f(hv^t) - f(h)| d\mu(h) \right)^q \frac{dt}{t^{1+q/s}} \right)^{1/q} \lesssim_{G,q} \int_G \|\nabla f(h)\|_{\ell_2^k} \, d\mu(h). \tag{2.11}$$

Similarly, we may ask the following question.

Question 2.0.27. *Let Γ be a not virtually abelian finitely generated group of polynomial growth. Choose v_Γ and $s > 0$ as in Theorem 2.0.6. For which exponents $q \geq 1$ do there exist $c = c(\Gamma) \in \mathbb{N}$ such that for any $n \in \mathbb{N}$ and $f : \Gamma \to L^1$ we have the*

following?

$$\left(\sum_{k=1}^{n^s} \frac{1}{k^{1+q/s}} \sum_{x\in B_n^\Gamma} \|f(xv_\Gamma^k) - f(x)\|_{L^1}^q\right)^{1/q} \lesssim_{\Gamma,q} \sum_{x\in B_{cn}^\Gamma} \sum_{a\in S} \|f(xa) - f(x)\|_{L^1}.$$

By the discretization argument of Section 2.3, an exponent q that answers Question 2.0.26 positively also answers Question 2.0.27 positively.

Because the case $q = \infty$ formally holds for Question 2.0.26 (see Remark 2.0.18), it is clear that the set of q that satisfies Question 2.0.26 is either of the form (q_G, ∞) or $[q_G, \infty)$ for some $q_G \in [1, \infty]$ depending on G. The exponent range is $[4, \infty)$ for $G = \mathbb{H}^3$ [65] and is $[2, \infty)$ for $G = \mathbb{H}^k$, $k \geq 5$ [64]. We wish to know whether there are finite exponents q satisfying (2.11), since then not only would we have a quantitative proof of [21], but we would also have quantitative nonembeddability statements analogous to Theorem 2.0.10, namely that if N_{r_1,r_2} is a r_2-covering of B_{r_1}, $r_1 \geq 2r_2$, $r_1, r_2 > 1$, then

$$c_1(N_{r_1,r_2}) \gtrsim_{G,q} (\log(r_1/r_2))^{1/q}.$$

Also, if Γ is a not virtually abelian finitely generated group of polynomial growth that is quasi-isometric to G, we would have

$$c_1(B_n^\Gamma) \gtrsim_{\Gamma,q} (\log n)^{1/q}, \quad n \geq 2.$$

By the co-area formula it suffices to prove (2.11) when f is an indicator of a measurable set $A \subseteq G$, in which case the right-hand side of (2.11) is interpreted as the horizontal perimeter $\mathrm{PER}(A)$ of A (see [2] and [12, Section 2] for a precise definition).

Definition 2.0.28 (Vertical perimeter at scale t [47, 64])**.** *Let $v \in Z(G)$ be as in Theorem 2.0.5. Let $A \subseteq G$ be measurable and $t \in (0, \infty)$. The vertical perimeter*

$v_t(A)$ of A at scale t is defined as the quantity

$$v_t(A) \coloneqq \mu\left(\left\{h \in A : \ hv^t \notin A \quad \text{or} \quad hv^{-t} \notin A\right\}\right). \tag{2.12}$$

By this definition, we may reformulate Question 2.0.26 into an isoperimetric inequality.

Question 2.0.29. *For which exponents q is it true that for every measurable $A \subseteq G$ one has*

$$\left(\int_0^\infty \frac{v_t(A)^q}{t^{1+q/s}} dt\right)^{1/q} \lesssim_{G,q} \mathrm{PER}(A). \tag{2.13}$$

Of course, the set of q that satisfies Question 2.0.29 is the same as that of Question 2.0.26. In light of the nonembeddability result [21] and the situation in the Heisenberg group, we make the following conjecture.

Conjecture 2.0.30. *There exist finite exponents q that answer Questions 2.0.26 and 2.0.29 positively, and the infimum q_G among such q is attained.*

Conditioned on Conjecture 2.0.30, we would have for a $(r_2, \Omega(r_2))$-net N_{r_1,r_2} of B_{r_1}, where $r_1, r_2 > 1$ with $r_1 \geq 2r_2$, that $c_1(N_{r_1,r_2}) \gtrsim_G (\log(r_1/r_2))^{1/q_G}$, and for a not virtually abelian finitely generated group of polynomial growth Γ we would have $c_1(B_n^\Gamma) \gtrsim_\Gamma (\log n)^{1/q_G}$, $n \geq 2$. However, for the 5 or higher dimensional Heisenberg groups $\mathbb{H}^{2k+1}$, $k \geq 2$, with $q_{\mathbb{H}^{2k+1}} = 2$ we have the matching upper bounds $c_1(N_{r_1,r_2}) \lesssim \sqrt{\log(r_1/r_2)}$ and $c_1(B_n^{\mathbb{H}_{\mathbb{Z}}^{2k+1}}) \lesssim \sqrt{\log n}$ by the Assouad embedding theorem, while for the 3-dimensional Heisenberg group $\mathbb{H}^3$ we have the matching upper bound $c_1(N_{r_1,r_2}) \lesssim (\log(r_1/r_2))^{1/4}$ and $c_1(B_n^{\mathbb{H}_{\mathbb{Z}}^3}) \lesssim (\log n)^{1/4}$ by [65, Theorem 3.1]. We thus pose the following question as well.

Question 2.0.31. *For a $(r_2, \Omega(r_2))$-net N_{r_1,r_2} of B_{r_1}, does*

$$c_1(N_{r_1,r_2}) \asymp_G (\log(r_1/r_2))^{1/q_G},$$

where $r_1, r_2 > 1$ with $r_1 \geq 2r_2$? Furthermore, let Γ be a not virtually abelian finitely generated group of polynomial growth which is quasi-isometric to G. Then does

$$c_1(B_n^\Gamma, d_W) \asymp_G (\log n)^{1/q_G}, \quad n \geq 2$$

hold?

2.1 Proof of Theorem 2.0.5: differentiation along random paths

We will prove a theorem slightly more general than Theorems 2.0.5 and 2.0.13, namely Theorem 2.1.1 below. We first explain the terminology behind the statements.

A Lie group G is said to be unimodular if its left-invariant Haar measure μ is also right-invariant. Given left-invariant vector fields $X_1, \cdots, X_k$, the pointwise span of $X_1, \cdots, X_k$ forms a left-invariant vector subbundle B over G, and on each fibre of the vector bundle B we may define a left-invariant Euclidean norm[8] $|\cdot|$ that has $X_1, \cdots, X_k$ as an orthonormal basis (this generalizes our earlier choice of $X_1, \cdots, X_k$, which we took to be a basis at e_G in the setting of nilpotent Lie groups and a basis of V_1 in the case of Carnot groups). We may define the associated *sub-Riemannian distance* as the Carnot-Carathéodory distance associated to B to $|\cdot|$, i.e., for $p, q \in G$,

$$d_G(p,q) \coloneqq \inf\left\{\int_0^1 |\dot\gamma(t)|dt : \gamma \in C^\infty_{\mathrm{pw}}([0,1]; G), \gamma(0) = p, \gamma(1) = q, \dot\gamma \in B\right\}.$$

If $X_1, \cdots, X_k$ and their brackets generate $\mathfrak{g}$, and if G is connected, then $d_G(\cdot,\cdot)$ is finite everywhere. The center $Z(\mathfrak{g})$ of the Lie algebra $(\mathfrak{g}, [,])$ of G consists of elements

[8] Again, we could also take any general left-invariant norm on B, in which case the distance is sub-Finsler. Compared to the sub-Riemannian case, the resulting distance d_G is then distorted by a factor of at most $\sqrt{\dim V_1}$ by the John ellipsoid theorem [43], and the results of this paper follow up to multiplicative factors of $\sqrt{\dim V_1}$.

$v \in \mathfrak{g}$ such that $[g, h] = 0$ for all $h \in \mathfrak{g}$, and has the property that if $v \in Z(\mathfrak{g})$ then $\exp(v) \in Z(G)$, the group theoretic center of G.

The following is an extension of Theorems 2.0.5 and 2.0.13; it is clear that Theorem 2.1.1 implies Theorems 2.0.5 and 2.0.13.

Theorem 2.1.1. *Let G be a unimodular Lie group with Haar measure μ, left-invariant vector fields $X_1, \cdots, X_k$, and associated sub-Riemannian distance $d_G(\cdot, \cdot)$. Suppose there is an element $v \in Z(\mathfrak{g})$ such that*

$$d_G(\exp(tv), e_G) \leq t^{1/\rho} \quad \forall t > 0$$

for some real number $\rho > 1$ and such that there is a subset S of G with measure ν such that the push-forward of the product measure of ν and the Lebesgue measure of $\mathbb{R}$ under the map $S \times \mathbb{R} \to G$, $(s, t) \mapsto s \exp(tv)$ is the Haar measure μ.

Suppose that $p \in (1, \infty)$ and $q \in [2, \infty)$. Let $(X, \|\cdot\|_X)$ be a Banach space with $K_q(X) < \infty$. If $p \geq q$, then every smooth and compactly supported $f : G \to X$ satisfies

$$\begin{aligned} &\left(\int_0^\infty \int_G \left(\frac{\|f(h \exp(tv)) - f(h)\|_X}{t^{1/\rho}} \right)^p d\mu(h) \frac{dt}{t} \right)^{1/p} \\ &\lesssim \frac{\rho}{\rho - 1} K_q(X) \left(\int_G \|\nabla f(h)\|_{\ell_2^k(X)}^p \, d\mu(h) \right)^{1/p}. \end{aligned} \tag{2.14}$$

If $p < q$, and if for any $t > 0$ we have $\lim_{r \to \infty} \mu(B_{r+t})/\mu(B_r) = 1$, then

$$\begin{aligned} &\left(\int_0^\infty \left(\int_G \left(\frac{\|f(h \exp(tv)) - f(h)\|_X}{t^{1/\rho}} \right)^p d\mu(h) \right)^{q/p} \frac{dt}{t} \right)^{1/q} \\ &\lesssim \frac{\rho}{\rho - 1} \max\left\{ (p-1)^{1/q - 1}, K_q(X) \right\} \left(\int_G \|\nabla f(h)\|_{\ell_2^k(X)}^p \, d\mu(h) \right)^{1/p}. \end{aligned} \tag{2.15}$$

The hypothesis for Theorem 2.1.1 is satisfied when G is a simply connected nilpotent group. The product decomposition of the Haar measure easily follows from that

of the Lebesgue measure on $\mathbb{R}^n$. The distance requirement follows by [10, Proposition 2.13] and [41, Theorem 2.1], which tells us that if $z \in Z(\mathfrak{g}) \setminus \{0\}$ then

$$d_G(\exp(tv), e_G) \asymp_G \begin{cases} t^{1/s}, & t \geq 1, \\ t^{1/r}, & 0 \leq t < 1, \end{cases} \tag{2.16}$$

where $s \geq 2$ is the nilpotency step of G and r is the smallest integer that $v \in \underbrace{[B, [B, \cdots, B]]}_{r \text{ times}}$ (recall B is the left-invariant bundle spanned by $X_1, \cdots, X_k$). By applying Theorem 2.1.1 with $\rho = s$ and $\rho = \max\{r, 1+\varepsilon\}$, we obtain the following.

Corollary 2.1.2. *Let G be a nonabelian simply connected nilpotent Lie group with left-invariant vector fields $X_1, \cdots, X_k$ such that the X_i's and their brackets generate $\mathfrak{g}$. Let $v \in Z(\mathfrak{g}) \setminus \{0\}$. Suppose that $p \in (1, \infty)$ and $q \in [2, \infty)$. Let $(X, \|\cdot\|_X)$ be a Banach space with $K_q(X) < \infty$, and let $f : G \to X$ be smooth and compactly supported.*

1. *Let $v \in Z(\mathfrak{g}) \setminus \{0\}$ be normalized so that $d_G(v^t, e_G) \leq t^{1/s}$ for $t > 0$, where s is the nilpotency step of G. Then*

$$\begin{aligned} &\left(\int_1^\infty \left(\int_G \left(\frac{\|f(hv^t) - f(h)\|_X}{t^{1/s}} \right)^p d\mu(h) \right)^{\max\{p,q\}/p} \frac{dt}{t} \right)^{1/\max\{p,q\}} \\ &\lesssim \max\left\{ (p-1)^{1/q-1}, K_q(X) \right\} \left(\int_G \|\nabla f(h)\|_{\ell_2^k(X)}^p \, d\mu(h) \right)^{1/p}. \end{aligned} \tag{2.17}$$

2. *Let r be the smallest integer that $v \in \underbrace{[B, [B, \cdots, B]]}_{r \text{ times}}$. If $r \geq 2$, and if $v \in Z(\mathfrak{g}) \setminus \{0\}$ is normalized so that $d_G(v^t, e_G) \leq t^{1/r}$ for $t > 0$. Then*

$$\begin{aligned} &\left(\int_0^1 \left(\int_G \left(\frac{\|f(hv^t) - f(h)\|_X}{t^{1/r}} \right)^p d\mu(h) \right)^{\max\{p,q\}/p} \frac{dt}{t} \right)^{1/\max\{p,q\}} \\ &\lesssim \max\left\{ (p-1)^{1/q-1}, K_q(X) \right\} \left(\int_G \|\nabla f(h)\|_{\ell_2^k(X)}^p \, d\mu(h) \right)^{1/p}. \end{aligned} \tag{2.18}$$

If $r = 1$, then for any $\varepsilon > 0$, if $v \in Z(\mathfrak{g}) \setminus \{0\}$ is normalized so that $d_G(v^t, e_G) \leq t^{1/(1+\varepsilon)}$ for $t > 0$, then

$$\begin{aligned} &\left(\int_0^1 \left(\int_G \left(\frac{\|f(hv^t) - f(h)\|_X}{t^{1/(1+\varepsilon)}} \right)^p d\mu(h) \right)^{\max\{p,q\}/p} \frac{dt}{t} \right)^{1/\max\{p,q\}} \\ &\lesssim \frac{1+\varepsilon}{\varepsilon} \max\left\{ (p-1)^{1/q-1}, K_q(X) \right\} \left(\int_G \|\nabla f(h)\|_{\ell_2^k(X)}^p \, d\mu(h) \right)^{1/p}. \end{aligned} \tag{2.19}$$

Based on these theorems, we have the following nonembeddability statements.

Theorem 2.1.3. *Let the Lie group G, left-invariant vector fields $X_1, \cdots, X_k$, and $v \in Z(\mathfrak{g})$ satisfy the hypotheses of Theorem 2.1.1. Suppose $X_1, \cdots, X_k$ and their brackets generate $\mathfrak{g}$. Let X be a q-uniformly convex Banach space, with $q \geq 2$.*

1. *If $d_G(\exp(tv), e_G) \asymp_G t^{1/\rho}$ for $0 < t < 1$ for some $\rho > 1$, then any nonempty open subset U of G fails to embed bilipschitzly into X.*

2. *If $d_G(\exp(tv), e_G) \asymp_G t^{1/\rho}$ for $t \geq 1$ for some $\rho > 1$, and if the Haar measure μ of G is close to being Ahlfors regular at large scales in the sense that for any constant $c' \geq 1$*

$$\mu(B_{c'r})/\mu(B_r) \asymp_G 1, \quad r \geq 1,$$

then $c_X(B_r) \gtrsim_G \frac{\rho-1}{\rho} \cdot \frac{(\rho \log r)^{1/q}}{K_q(X)}$ for $r = \Omega_G(1)$. In particular, G fails to bilipschitzly embed into X. We have $c_p(B_r) \asymp (\log r)^{1/\max\{p,2\}}$ for $1 < p < \infty$.

Indeed, for (1), we may assume by translation that $B_r \subset U$ for some $r > 0$. If $f : U \to X$ were a bilipschitz embedding, $d_G(x,y) \leq \|f(x) - f(y)\|_X \leq D d_G(x,y)$, $x, y \in U$, we may assume by translation that $f(e_G) = 0$. Multiplying f by a smooth cutoff function, we may construct $F : G \to X$ which is Lipschitz, agrees with f on $B_{r/3}$, and is supported on $B_{2r/3}$. By a smooth approximation argument, we may apply (2.14) to F. Let $c_1, c_2 > 0$ be such that $c_1 t^{1/\rho} \leq d_G(\exp(tv), e_G) \leq c_2 t^{1/\rho}$ for $0 < t < 1$. Then the right-hand side of (2.14) is a finite quantity, whereas the left-hand side is

at least

$$\left(\int_0^{r^\rho/c_2^\rho 6^\rho}\int_{B_{r/6}}\left(\frac{\|f(h\exp(tv))-f(h)\|_X}{t^{1/\rho}}\right)^q d\mu(h)\frac{dt}{t}\right)^{1/q}$$
$$=\left(\mu(B_{r/6})\int_0^{r^\rho/c_2^\rho 6^\rho} c_1^q\frac{dt}{t}\right)^{1/q}=\infty,$$

giving a contradiction.

For (2), if $f:B_r\to X$ satisfied $d_G(x,y)\le\|f(x)-f(y)\|_X\le Dd_G(x,y)$ for $x,y\in B_r$ and, then translating and multiplying f by a cutoff function we may construct $F:G\to X$ which is $3D$-Lipschitz, agrees with f on $B_{r/2}$, and is supported on B_r. Again, let $c_1,c_2>0$ be such that $c_1t^{1/\rho}\le d_G(\exp(tv),e_G)\le c_2t^{1/\rho}$ for $t>1$. Applying (2.14), we have

$$c_1\mu(B_{r/4})^{1/q}\log(r^\rho/c_2^\rho 4^\rho)^{1/q}$$
$$\lesssim_G\left(\int_1^{r^\rho/c_2^\rho 4^\rho}\int_{B_{r/4}}\left(\frac{\|f(h\exp(tv))-f(h)\|_X}{t^{1/\rho}}\right)^q d\mu(h)\frac{dt}{t}\right)^{1/q}$$
$$\lesssim\frac{\rho}{\rho-1}K_q(X)\left(\int_{B_r}\|\nabla F\|^q_{\ell_2^k(X)}d\mu\right)^{1/q}\lesssim_G\frac{\rho}{\rho-1}K_q(X)\mu(B_r)^{1/q}D.$$

This gives the stated estimate.

By (2.16), we have the following.

Corollary 2.1.4. *Let G be a nonabelian simply connected nilpotent Lie group with left-invariant vector fields $X_1,\cdots,X_k$ such that the X_i's and their brackets generate $\mathfrak{g}$. Suppose that $q\in[2,\infty)$ and let $(X,\|\cdot\|_X)$ be a Banach space with $K_q(X)<\infty$.*

1. *If $Z(\mathfrak{g})\not\subseteq\operatorname{span}\{X_1,\cdots,X_k\}$, then any nonempty open subset U of G fails to embed bilipschitzly into X.*

2. *If $Z(\mathfrak{g})\subseteq\operatorname{span}\{X_1,\cdots,X_k\}$, we have $c_X(B_r)\gtrsim_G\frac{(\log r)^{1/q}}{K_q(X)}$ for $r\ge 2$. In particular, G fails to bilipschitzly embed into X.*

Before we begin the proof of Theorem 2.1.1, for motivation we begin by proving Remark 2.0.18 which states that the case $q = \infty$ of Theorem 2.0.17 formally holds.

Proof of Remark 2.0.18. It is enough to prove that, for each $r > 0$,

$$\left[\int_G \left(\frac{|f(h) - f(h\delta_r(v))|}{r}\right)^p d\mu(h)\right]^{1/p} \leq \|\nabla f\|_{L^p(G;\ell_2^k)}.$$

Let $\gamma : [0, r] \to G$ be a piecewise smooth horizontal curve of unit speed connecting e_G to $\delta_r(v)$.Then

$$|f(h) - f(h\delta_r(v))| \leq \int_0^r |\nabla f(h\gamma(s)) \cdot dL_h\dot{\gamma}(s)|ds \leq \int_0^r |\nabla f(h\gamma(s))|ds.$$

Thus, by Jensen's inequality and right-invariance of the Haar measure μ,

$$\begin{aligned}\left[\int_G \left(\frac{|f(h) - f(h\delta_r(v))|}{r}\right)^p d\mu(h)\right]^{1/p} &\leq \left[\frac{1}{r}\int_G\int_0^r |\nabla f(h\gamma(s))|^p ds d\mu(h)\right]^{1/p} \\ &\leq \|\nabla f\|_{L^p(G;\ell_2^k)},\end{aligned}$$

and we are done. □

Likewise, it is not hard to obtain a proof of the $q = \infty$ case of Theorem 2.1.1 by the above proof. This proof basically just takes a path from h to $h\delta_r(v)$ and differentiates f along that path; however, the resulting inequality is weak. To make up for this weakness, we will choose over a random distribution of paths, the randomness arising from a heat flow on the line spanned by v. We will then apply the above proof method to each of those random paths (in Lemma 2.1.6). This will result in an upper bound given in terms of a convolution of ∇f with a derivative of the heat kernel, which we treat using Littlewood–Paley–Stein g-function estimates. This proof method is borrowed from and extends that of [47].

We set some notation and terminology.

Recall there is $v \in Z(\mathfrak{g})$ and a subset S of G with measure ν such that the push-forward of the product measure of ν and the Lebesgue measure of $\mathbb{R}$ under the map $S \times \mathbb{R} \to G$, $(s,t) \mapsto s\exp(tv)$ is the Haar measure μ. We denote $v^t \coloneqq \exp(tv) \in G$. For $p \in [1,\infty)$, the Lebesgue–Bochner spaces

$$L^p(\mathbb{R}, \text{Lebesgue}; X) = L^p(\mathbb{R}; X),\ L^p(S, \nu; X) = L^p(S; X),\ L^p(G, \mu; X) = L^p(G; X)$$

are defined. For $f \in L^p(G; X)$ define $f^v : \mathbb{R} \to L^p(S; X)$ by

$$f^v(z)(x) = f(xv^z), \quad x \in S,\ z \in \mathbb{R}. \tag{2.20}$$

Given $\psi \in L^1(\mathbb{R})$, define the convolution $\psi * f \coloneqq \psi * f^v \in L^p(G; X)$,

$$\psi * f(xv^z) \coloneqq \int_{\mathbb{R}} \psi(u) f(xv^{z-u}) du \in X.$$

So $\psi * f$ is the usual group convolution of f with the measure supported on

$$\exp(\text{span}(v))$$

whose density is ψ.

We will take ψ to be the heat kernel or its derivatives. The heat kernel on $\mathbb{R}$ is defined for $t > 0$, $x \in \mathbb{R}$ as

$$h_t(x) \coloneqq \frac{1}{2\sqrt{\pi t}} e^{-\frac{x^2}{4t}}, \quad x \in \mathbb{R},$$

and has derivatives

$$\dot{h}_t(x) = \partial_t h_t(x) = \frac{x^2 - 2t}{8\sqrt{\pi} t^{5/2}} e^{-\frac{x^2}{4t}}, \quad \partial_x h_t(x) = -\frac{x}{4\sqrt{\pi} t^{3/2}} e^{-\frac{x^2}{4t}}.$$

Then for a universal constant C, for all $t > 0$ we have

$$\int_0^\infty x^{1/\rho} h_t(x) dx = \Gamma\left(\frac{1}{2} + \frac{1}{2\rho}\right) \frac{(4t)^{1/\rho}}{2\sqrt{\pi}}, \quad \|\dot{h}_t\|_{L^1(\mathbb{R})} = \frac{C}{t}, \quad \|\partial_x h_t\|_{L^1(\mathbb{R})} = \frac{1}{\sqrt{\pi t}}. \tag{2.21}$$

We now begin the proof of Theorem 2.1.1. First, note that the case $p > q$ of Theorem 2.1.1 follows from that of $p = q$, because $K_q(X) \geq K_p(X)$. Thus, it is enough to consider the cases $p = q$ and $p < q$. The case $p < q$ will also be shown to follow from that of $p = q$, albeit in a more complicated way. We prove the case $p = q$ first.

2.1.1 The case $p = q$

Step 1

The first step is to bound the vertical variations $f(gv^t) - f(g)$ by the quantities $\dot{h}_t * f$. This corresponds to Lemma 2.6 of [47], which is a modification of the corresponding statement in [75].

Lemma 2.1.5 (Analogue of [47, Lemma 2.6])**.** *Fix $p \in [1, \infty)$ and let $(X, \|\cdot\|_X)$ be a Banach space. Every smooth and compactly supported $f : G \to X$ satisfies*

$$\begin{aligned} &\left(\int_0^\infty \int_G \|f(gv^t) - f(g)\|_X^p dg \frac{dt}{t^{1+p/\rho}}\right)^{1/p} \\ &\lesssim \frac{\rho}{\rho - 1} \left(\int_0^\infty t^{p\frac{2\rho-1}{2\rho}-1} \left\|\dot{h}_t * f\right\|_{L^p(G;X)}^p dt\right)^{1/p}. \end{aligned} \tag{2.22}$$

This is the part of the argument that uses that $\rho > 1$. In the nilpotent setting, this means that our group G has nilpotency step $s > 1$ and hence is nonabelian.

Proof. For every $g \in G$ and $t > 0$ we have

$$f(gv^t) - f(g) = [f(gv^t) - h_{t^2} * f(gv^t)] + [h_{t^2} * f(gv^t) - h_{t^2} * f(g)] + [h_{t^2} * f(g) - f(g)]$$

so by the triangle inequality

$$\begin{aligned}
&\left(\int_0^\infty \int_G \|f(gv^t) - f(g)\|_X^p dg \frac{dt}{t^{1+p/\rho}}\right)^{1/p} \\
&\le \left(\int_0^\infty \int_G \|f(gv^t) - h_{t^2} * f(gv^t)\|_X^p dg \frac{dt}{t^{1+p/\rho}}\right)^{1/p} \\
&\quad + \left(\int_0^\infty \int_G \|h_{t^2} * f(gv^t) - h_{t^2} * f(g)\|_X^p dg \frac{dt}{t^{1+p/\rho}}\right)^{1/p} \\
&\quad + \left(\int_0^\infty \int_G \|h_{t^2} * f(g) - f(g)\|_X^p dg \frac{dt}{t^{1+p/\rho}}\right)^{1/p}.
\end{aligned}$$

We bound each term by the right-hand side of (2.22). First,

$$\begin{aligned}
&\left(\int_0^\infty \int_G \|f(gv^t) - h_{t^2} * f(gv^t)\|_X^p dg \frac{dt}{t^{1+p/\rho}}\right)^{1/p} \\
&= \left(\int_0^\infty \int_G \left\| \int_0^{t^2} \dot{h}_\tau * f(gv^t) d\tau \right\|_X^p dg \frac{dt}{t^{1+p/\rho}}\right)^{1/p} \\
&\le \left(\int_0^\infty \left(\int_0^{t^2} \left\|\dot{h}_\tau * f\right\|_{L^p(G;X)} d\tau\right)^p \frac{dt}{t^{1+p/\rho}}\right)^{1/p} \\
&\le 2^{1-1/p}\rho \left(\int_0^\infty t^{p\frac{2\rho-1}{2\rho}-1} \|\dot{h}_t * f\|_{L^p(G;X)}^p dt\right)^{1/p},
\end{aligned}$$

where in the first inequality we used the triangle inequality in $L^p(G;X)$ and the fact that the Haar measure on G is right-invariant, and in the second inequality we used Hardy's inequality [34, Theorem 330]

$$\left(\int_0^\infty \left[x^{-\nu}\int_0^x f(t)dt\right]^p dx\right)^{1/p} \le \frac{1}{\nu - 1/p}\left(\int_0^\infty x^{(1-\nu)p} f(x)^p dx\right)^{1/p}, \quad \nu > \frac{1}{p}, \tag{2.23}$$

with $\nu = \frac{1}{p} + \frac{1}{2\rho}$. The third term of the right-hand side of (2.22) is bounded using the same method.

We now bound the second term of the right-hand side of (2.22). By the semigroup

property we have $h_t = h_{t/2} * h_{t/2}$ for $t > 0$, hence $\dot{h}_t = h_{t/2} * \dot{h}_{t/2}$ and so

$$\partial_t \partial_x h_t = \partial_x \dot{h}_t = \partial_x h_{t/2} * \dot{h}_{t/2}.$$

Since by Young's inequality

$$\|\partial_x h_t * f\|_{L^p(G;X)} \le \|\partial_x h_t\|_{L^1(\mathbb{R})} \|f\|_{L^p(G;X)} = \frac{1}{\sqrt{\pi t}} \|f\|_{L^p(G;X)}, \tag{2.24}$$

we have $\lim_{t\to\infty} \partial_x h_t * f = 0$ in $L^p(G;X)$, and thus

$$\partial_x h_{t^2} * f = -\int_{t^2}^{\infty} \partial_\tau (\partial_x h_\tau * f) d\tau = -\int_{t^2}^{\infty} \partial_x h_{\tau/2} * \dot{h}_{\tau/2} * f d\tau,$$

and we may write

$$h_{t^2} * f(gv^t) - h_{t^2} * f(g) = \int_0^t \partial_x h_{t^2} * f(gv^u) du = -\int_0^t \int_{t^2}^{\infty} \partial_x h_{\tau/2} * \dot{h}_{\tau/2} * f(gv^u) d\tau du.$$

Observe that for every $t \in (0,\infty)$ we have

$$\begin{aligned}
&\left(\int_0^\infty \int_G \|h_{t^2} * f(gv^t) - h_{t^2} * f(g)\|_X^p dg \frac{dt}{t^{1+p/\rho}} \right)^{1/p} \\
&\le \left(\int_0^\infty \int_G \left\| \int_0^t \int_{t^2}^\infty \partial_x h_{\tau/2} * \dot{h}_{\tau/2} * f(gv^u) d\tau du \right\|_X^p dg \frac{dt}{t^{1+p/\rho}} \right)^{1/p} \\
&\le \left(\int_0^\infty \left(\int_0^t \int_{t^2}^\infty \left(\int_G \left\| \partial_x h_{\tau/2} * \dot{h}_{\tau/2} * f(gv^u) \right\|_X^p dg \right)^{1/p} d\tau du \right)^p \frac{dt}{t^{1+p/\rho}} \right)^{1/p} \\
&= \left(\int_0^\infty \left(t \int_{t^2}^\infty \left\| \partial_x h_{\tau/2} * \dot{h}_{\tau/2} * f) \right\|_{L^p(G;X)} d\tau \right)^p \frac{dt}{t^{1+p/\rho}} \right)^{1/p} \\
&\le \left(\int_0^\infty t^{p\frac{\rho-1}{\rho}-1} \left(\int_{t^2}^\infty \left\| \partial_x h_{\tau/2} * \dot{h}_{\tau/2} * f \right\|_{L^p(G;X)} d\tau \right)^p dt \right)^{1/p} \\
&\le \frac{2^{1-1/p}\rho}{\rho-1} \left(\int_0^\infty t^{p\frac{3\rho-1}{2\rho}-1} \left\| \partial_x h_{t/2} * \dot{h}_{t/2} * f \right\|_{L^p(G;X)}^p dt \right)^{1/p},
\end{aligned}$$

where the second inequality uses the triangle inequality in $L^p(G;X)$, the first equality uses the right-invariance of the Haar measure μ, and the fourth inequality uses the second form of Hardy's inequality [34, Theorem 330]

$$\left(\int_0^\infty \left[t^\nu \int_t^\infty f(u)du\right]^p dt\right)^{1/p} \leq \frac{1}{\nu+1/p}\left(\int_0^\infty x^{p(1+\nu)} f(x)^p dx\right)^{1/p}, \quad \nu > -1/p \tag{2.25}$$

with $\nu = \frac{\rho-1}{2\rho} - \frac{1}{p}$. (This is the step where we use $\rho > 1$.) By Young's inequality,

$$\|\partial_x h_{t/2} * \dot{h}_{t/2} * f\|_{L^p(G;X)} \leq \|\partial_x h_{t/2}\|_{L^1(\mathbb{R})} \|\dot{h}_{t/2} * f\|_{L^p(G;X)} = \sqrt{\frac{2}{\pi t}} \|\dot{h}_{t/2} * f\|_{L^p(G;X)}.$$

Therefore

$$\begin{aligned}&\left(\int_0^\infty \int_G \|h_{t^2} * f(gv^t) - h_{t^2} * f(g)\|_X^p dg \frac{dt}{t^{1+p/\rho}}\right)^{1/p} \\ &\leq \frac{2^{3/2-1/p-1/2\rho}\rho}{\sqrt{\pi}(\rho-1)}\left(\int_0^\infty t^{p\frac{2\rho-1}{2\rho}-1} \|\dot{h}_t * f\|_{L^p(G;X)}^p dt\right)^{1/p}.\end{aligned}$$

This completes the proof. □

Step 2

We will next bound $\dot{h}_t * f$ using $\dot{h}_t * \nabla f$. We first prove the following lemma.

Lemma 2.1.6. *Suppose that $p \in [1,\infty)$ and $t \in (0,\infty)$. Then for every Banach space $(X, \|\cdot\|_X)$ and every smooth and compactly supported $f : G \to X$ we have*

$$\left\|\dot{h}_t * f - \dot{h}_{2t} * f\right\|_{L^p(G;X)} \lesssim t^{1/2\rho} \left\|\dot{h}_t * \nabla f\right\|_{L^p(G;\ell_2^k(X))}. \tag{2.26}$$

Proof. Recalling the semigroup property $\dot{h}_{2t} = h_t * \dot{h}_t$, we have

$$\dot{h}_t * f(g) - \dot{h}_{2t} * f(g) = \dot{h}_t * f(g) - h_t * \dot{h}_t * f(g) = \int_{\mathbb{R}} h_t(u)\left(\dot{h}_t * f(g) - \dot{h}_t * f(gv^{-u})\right) du. \tag{2.27}$$

For $u \in [0,\infty)$, let $\gamma_u : [0, d_G(v^u, e_G)] \to G$ be a measurable family of geodesics parametrized by arclength joining e_G to v^u. (Such a measurable family exists by the Aumann measurable selection theorem [9, Theorem 6.9.13]). For every $u \in [0,\infty)$ and $g \in G$,

$$\begin{aligned}\dot{h}_t * f(g) - \dot{h}_t * f(gv^{-u}) &= \dot{h}_t * f(gv^{-u}\gamma_u(d_G(v^u, e_G))) - \dot{h}_t * f(gv^{-u}) \\ &= \int_0^{d_G(v^u,e_G)} \frac{d}{d\theta} \dot{h}_t * f(gv^{-u}\gamma_u(\theta))d\theta.\end{aligned}$$

Because γ_u is horizontal and parametrized by arclength, and because convolution with $\dot{h_t}$ commutes with ∇ (this is because v is in the center of G), we have for every $\theta \in [0, d_G(v^u, e_G)]$

$$\left\| \frac{d}{d\theta} \dot{h}_t * f(gv^{-u}\gamma_u(\theta)) \right\|_X \lesssim \left\| \dot{h}_t * \nabla f(gv^{-u}\gamma_u(\theta)) \right\|_{\ell_2^k(X)}. \tag{2.28}$$

We thus obtain

$$\begin{aligned}&\left(\int_G \left\| \int_0^\infty h_t(u) \left(\dot{h}_t * f(g) - \dot{h}_t * f(gv^{-u}) \right) du \right\|_X^p dg \right)^{1/p} \\ &\lesssim_G \int_0^\infty \int_0^{d_G(v^u,e_G)} h_t(u) \left(\int_G \left\| \dot{h}_t * \nabla f(gv^{-u}\gamma_u(\theta)) \right\|_{\ell_2^k(X)}^p dg \right)^{1/p} d\theta du \\ &\leq \left(\int_0^\infty u^{1/2\rho} h_t(u) du \right) \|\dot{h}_t * \nabla f\|_{L^p(G;\ell_2^k(X))} \\ &\overset{(2.21)-(i)}{\lesssim} t^{1/2\rho} \|\dot{h}_t * \nabla f\|_{L^p(G;\ell_2^k(X))},\end{aligned}$$

where in the first equality we used the right-invariance of the Haar measure on G.

Similarly, for $u \in (-\infty, 0]$,

$$\dot{h}_t * f(g) - \dot{h}_t * f(gv^{-u}) = -\int_0^{d_G(v^u,e_G)} \frac{d}{d\theta} \dot{h}_t * f(g\gamma_{-u}(\theta))d\theta,$$

and by the same reasoning,

$$\left(\int_G \left\| \int_{-\infty}^0 h_t(u) \left(\dot{h}_t * f(g) - \dot{h}_t * f(gv^{-u}) \right) du \right\|_X^p dg \right)^{1/p} \lesssim t^{1/2\rho} \|\dot{h}_t * \nabla f\|_{L^p(G;\ell_2^k(X))}.$$

These two estimates along with (2.27) gives the stated inequality. □

Now we bound $\dot{h}_t * f$ using $\dot{h}_t * \nabla f$.

Lemma 2.1.7. *Fix $p \in [1,\infty)$. For every Banach space $(X, \|\cdot\|_X)$, every smooth and compactly supported $f : G \to X$ satisfies*

$$\left(\int_0^\infty t^{p\frac{2\rho-1}{2\rho}-1} \left\| \dot{h}_t * f \right\|_{L^p(G;X)}^p dt \right)^{1/p} \lesssim_G \left(\int_0^\infty t^{p-1} \left\| \dot{h}_t * \nabla f \right\|_{L^p(G;\ell_2^k(X))}^p dt \right)^{1/p}.$$

Proof. By Young's inequality,

$$\|\dot{h}_t * f\|_{L^p(G;X)} \leq \|\dot{h}_t\|_{L^1(\mathbb{R})} \|f\|_{L^p(G;X)} \overset{(2.21)-(ii)}{\asymp} \frac{1}{t} \|f\|_{L^p(G;X)},$$

so $\lim_{t\to\infty} \dot{h}_t * f = 0$ in $L^p(G;X)$. Therefore

$$\dot{h}_t * f = \sum_{m=1}^\infty \left(\dot{h}_{2^{m-1}t} * f - \dot{h}_{2^m t} * f \right),$$

from which it follows that

$$\begin{aligned}
&\left(\int_0^\infty t^{p\frac{2\rho-1}{2\rho}-1} \left\| \dot{h}_t * f \right\|_{L^p(G;X)}^p dt \right)^{1/p} \\
&\leq \sum_{m=1}^\infty \left(\int_0^\infty t^{p\frac{2\rho-1}{2\rho}-1} \left\| \dot{h}_{2^{m-1}t} * f - \dot{h}_{2^m t} * f \right\|_{L^p(G;X)}^p dt \right)^{1/p} \\
&\lesssim \sum_{m=1}^\infty \left(\int_0^\infty t^{p\frac{2\rho-1}{2\rho}-1} (2^{m-1}t)^{p/\rho} \left\| \dot{h}_{2^{m-1}t} * \nabla f \right\|_{L^p(G;\ell_2^k(X))}^p dt \right)^{1/p} \\
&= \left(\sum_{m=1}^\infty \frac{1}{2^{(m-1)(2\rho-1)/2\rho}} \right) \left(\int_0^\infty t^{p-1} \left\| \dot{h}_t * \nabla f \right\|_{L^p(G;\ell_2^k(X))}^p dt \right)^{1/p},
\end{aligned}$$

where the second inequality uses Lemma 2.1.6. □

Step 3

We are now ready to prove the $p = q$ case of Theorem 2.1.1.

Given a Banach space $(\mathbb{B}, \|\cdot\|_{\mathbb{B}})$, for every function $\phi \in L^p(\mathbb{R};\mathbb{B})$ its generalized Hardy–Littlewood g-function $\mathfrak{G}_p(\phi) : \mathbb{R} \to [0,\infty]$ is defined as follows.

$$\mathfrak{G}_p(\phi)(x) \coloneqq \left(\int_0^\infty t^{p-1} \left\|\dot{h}_t * \phi(x)\right\|_{\mathbb{B}}^p dt\right)^{1/p}. \tag{2.29}$$

By [40, Theorem 17], which is a quantitative version of[9] [56, Theorem 2.1] and [82, Theorem 2]

$$\phi \in L^p(\mathbb{R};\mathbb{B}) \implies \|\mathfrak{G}_p(\phi)\|_{L^p(\mathbb{R};\mathbb{B})} \lesssim K_p(\mathbb{B})\|\phi\|_{L^p(\mathbb{R};\mathbb{B})}. \tag{2.30}$$

We will apply (2.30) to $\mathbb{B} = L^p(S;\ell_2^k(X))$. By [57, Corollary 6.4],

$$K_p(\mathbb{B}) \le \left(\frac{5p^2}{(p-1)^2}\right)^{1-1/p} K_p(\ell_2^k(X)) \le \left(\frac{50p^3}{(p-1)^3}\right)^{1-1/p} K_p(X) \lesssim K_p(X),$$

since $p \ge 2$.

[9] More precisely, [56] proves inequality (2.30) for 'subordinated' semigroups, such as the Poisson semigroup subordinated by the heat semigroup, and [82] proves the inequality for general symmetric diffusion semigroups. The work [40] has the advantage that it obtains, for the heat semigroup, the explicit constant $m_p(\mathbb{B})$, the martingale cotype p constant of $\mathbb{B}$, which in turn is bounded by $K_p(\mathbb{B})$ by [70]. Since [40] and [82] were unavailable when [47] was written, [47] worked with the Poisson semigroup. We could have also started with the Poisson semigroup P_t, which gives the same constant in (2.30) due to semigroup subordination. See [65, Appendix A] for a detailed account of this discussion.

Recalling (2.20), we choose $\phi = (\nabla f)^v : \mathbb{R} \to \mathbb{B}$. Then,

$$\begin{aligned}\left(\int_0^\infty t^{p-1} \left\|\dot{h}_t * \nabla f\right\|^p_{L^p(G;\ell_2^k(X))} dt\right)^{1/p} &= \left(\int_0^\infty \int_{\mathbb{R}} \left(t \left\|\dot{h}_t * \phi(z)\right\|_{\mathbb{B}}\right)^p dz \frac{dt}{t}\right)^{1/p} \\ &\leq \left(\int_{\mathbb{R}} \int_0^\infty \left(t \left\|\dot{h}_t * \phi(z)\right\|_{\mathbb{B}}\right)^p \frac{dt}{t} dz\right)^{1/p} \\ &= \|\mathfrak{G}_p(\phi)\|_{L^p(\mathbb{R};\mathbb{B})} \\ &\lesssim K_p(X)\|\phi\|_{L^p(\mathbb{R};\mathbb{B})} = K_p(X)\|\nabla f\|_{L^p(G;\ell_2^k(X))}.\end{aligned}$$

By this discussion and Lemmas 2.1.6 and 2.1.7, the proof for the case $p = q$ is complete.

2.1.2 The case $p < q$

The following is an imitation of the argument of [65, Appendix A]. Recall that for all $t > 0$, $\lim_{r\to\infty} \mu(B_{r+t})/\mu(B_r) = 1$.

For $M > 1$, let $\beta_M : G \to [0, 1]$ be an $O(1)$-Lipschitz smooth bump function with $\beta_M = 1$ on B_M and $\operatorname{supp} \beta_M \subset B_{M+1}$. For $f : G \to X$, define $F_M : G \to L^p(G; E)$ by

$$F_M(h)(g) = \beta_M(h) f(gh), \quad g, h \in G.$$

Then by F_M is compactly supported and smooth, so by the $p = q$ case,

$$\left(\int_0^\infty \|D_v^t F_M\|^q_{L^q(G;L^p(G;X))} \frac{dt}{t}\right)^{1/q} \lesssim \frac{\rho}{\rho - 1} K_q(L^p(G; X))\|\nabla F_M\|_{L^q(G;\ell_q^k(L^p(G;X)))},$$

where D_v^t is shorthand for $D_v^t f(g) = \frac{f(gv^t) - f(g)}{t^{1/\rho}}$.

For $q \in (1,2]$, the q-uniform smoothness constant of X is defined by

$$S_q(X)$$
$$:= \inf\left\{ K > 0 : \forall x, y \in E \ \ (\|x\|_X^q + K^q \|y\|_X^q)^{1/q} \geq \left(\frac{\|x+y\|_X^q + \|x-y\|_X^q}{2} \right)^{1/q} \right\}.$$

Then we have

$$\begin{aligned} K_q(L^p(G;X)) = S_{q/(q-1)}(L^{p/(p-1)}(G;X^*)) &\leq \max\left\{ \left(\frac{p}{p-1} \right)^{(q-1)/q}, S_{q/(q-1)}(X^*) \right\} \\ &\asymp \max\left\{ (p-1)^{\frac{1}{q}-1}, K_q(X) \right\} \end{aligned}$$

Where the equality and $\asymp$ follow from [6, Lemma 5] and the inequality follows from equation (4.4) of [60]. Therefore we have

$$\begin{aligned} &\left(\int_0^\infty \|D_v^t F_M\|_{L^q(G;L^p(G;X))}^q \frac{dt}{t} \right)^{1/q} \\ &\lesssim \frac{\rho}{\rho-1} \max\{(p-1)^{1/q-1}, K_q(E)\} \|\nabla F_M\|_{L^q(G;\ell_q^k(L^p(G;X)))}. \end{aligned} \tag{2.31}$$

As

$$\nabla F_M(h)(g) = f(gh)\nabla\beta_M(h) + \beta_M(h)\nabla f(gh)$$

from left-invariance of ∇, we have (with the normalization $\mu(B_1) = 1$)

$$\|\nabla F_M(h)\|_{\ell_2^k(L^p(G;X))} \lesssim \|f\|_{L^p(G;X)} \mathbf{1}_{B_{M+1}\setminus B_M}(h) + \|\nabla f\|_{L^p(G;X)} \mathbf{1}_{B_{M+1}}(h), \quad h \in G,$$

and thus (with the normalization $\mu(B_1) = 1$)

$$\|\nabla F_M\|_{L^q(G;\ell_2^k(L^p(G;X)))} \lesssim \mu(B_{M+1} \setminus B_M)^{1/q} \|f\|_{L^p(G;X)} + \mu(B_M)^{1/q} \|\nabla f\|_{L^p(G;\ell_2^k(X))}. \tag{2.32}$$

If $0 < t < M^s$ and $h \in B_{M-t^{1/s}}$, then $hv^t \in B_M$ so $\beta_M(h) = \beta_M(hv^t) = 1$. Thus, for

any $h \in B_{M-t^{1/s}}$, $\|D_v^t F_M(h)\|_{L^p(G;X)} = \|D_v^t f\|_{L^p(G;X)}$. We have

$$\|D_v^t F_M\|_{L^q(G;L^p(G;X))} \gtrsim \mu(B_{M-t^{1/s}})^{1/q} \|D_v^t f\|_{L^p(G;X)},$$

and for $0 < T < M$,

$$\left(\int_0^{T^s} \|D_v^t F_M\|_{L^q(G;L^p(G;X))}^q \frac{dt}{t}\right)^{1/q} \gtrsim \mu(B_{M-T})^{1/q} \left(\int_0^{T^s} \|D_v^t f\|_{L^p(G;X)}^q \frac{dt}{t}\right)^{1/q}. \quad (2.33)$$

Combining (2.31), (2.32), and (2.33), we are left with

$$\begin{aligned}
&\mu(B_{M-T})^{1/q} \left(\int_0^{T^s} \|D_v^t f\|_{L^p(G;X)}^q \frac{dt}{t}\right)^{1/q} \\
&\lesssim \frac{\rho}{\rho-1} \max\{(p-1)^{1/q-1}, K_q(X)\} \\
&\qquad \cdot \left(\mu(B_{M+1} \setminus B_M)^{1/q} \|f\|_{L^p(G;X)} + \mu(B_M)^{1/q} \|\nabla f\|_{L^p(G;\ell_2^k(X))}\right)
\end{aligned}$$

for all $0 < T < M$. Taking $M \to \infty$ and then $T \to \infty$, we obtain the stated inequality. This completes the proof of Theorem 2.1.1.

2.2 Proof of Theorems 2.0.1, 2.0.2, 2.0.10, 2.0.11, 2.0.14, and 2.0.15 and Corollaries 2.0.3 and 2.0.4: distortion bounds

In this section we prove Theorems 2.0.1 and 2.0.2, Corollaries 2.0.3 and 2.0.4, and Theorems 2.0.10, 2.0.11, 2.0.14 and 2.0.15. Theorems 2.0.1, 2.0.10, 2.0.11, 2.0.14 and 2.0.15 will follow from Theorem 2.0.5, Theorem 2.0.2 will follow from Theorem 2.0.6, and Corollaries 2.0.3 and 2.0.4 will follow from Theorems 2.0.1 and 2.0.2, respectively.

First we show the proof of Theorem 2.0.1 from Theorem 2.0.5.

Proof of Theorem 2.0.1 from Theorem 2.0.5. Suppose $f : B_r \to X$ satisfies

$$d_G(x,y) \le \|f(x) - f(y)\|_X \le D d_G(x,y), \quad x, y \in B_r.$$

Translating and multiplying f by a cutoff function we may construct $F : G \to X$ which is $3D$-Lipschitz, agrees with f on $B_{r/2}$, and is supported on B_r. Supposing Theorem 2.0.5 and applying (2.2), we have, using $\|f(hv^t) - f(h)\|_X \ge d_G(v^t, e_G) \gtrsim_G t^{1/\rho}$ for $t \ge 1$,

$$\begin{aligned}\mu(B_{r/4})^{1/q} \log(r^s/4^s)^{1/q} &\lesssim_G \left(\int_1^{r^s/4^s} \int_{B_{r/4}} \left(\frac{\|f(hv^t) - f(h)\|_X}{t^{1/s}} \right)^q d\mu(h) \frac{dt}{t} \right)^{1/q} \\ &\lesssim K_q(X) \left(\int_{B_r} \|\nabla F\|^q_{\ell_2^k(X)} d\mu \right)^{1/q} \lesssim_G K_q(X) \mu(B_r)^{1/q} D.\end{aligned}$$

But by [44, Corollary 4.11], we have $\mu(B_r) \asymp_G r^{n_h}$ for $r = \Omega(1)$ for a fixed positive integer n_h. We therefore obtain Theorem 2.0.1. □

The proof of Theorem 2.0.2 from Theorem 2.0.6 is similar.

Proof of Theorem 2.0.2 from Theorem 2.0.6. Let c be as in Theorem 2.0.6. It suffices to show that $c_X(B_{(c+1)n}) \gtrsim_\Gamma \frac{(\log n)^{1/q}}{K_q(X)}$. Suppose that $f : \Gamma \to X$ satisfies $d_W(x,y) \le \|f(x) - f(y)\|_X \le D d_W(x,y)$ for all $x, y \in B^\Gamma_{(c+1)n}$. Recall that

$$d_W(v^k_\Gamma, e_\Gamma) \asymp_\Gamma k^{1/s}, \quad k \in \mathbb{N},$$

and also that $|B_m| \asymp_\Gamma m^{n_h}$ for every $m \in \mathbb{N}$ (see [10, Theorem 1.1]). Thus, Theorem 2.0.6 applied to f yields the following estimate.

$$n^{n_h/q} (\log n)^{1/q} \lesssim_\Gamma \left(\sum_{k=1}^{n^s} n^{n_h} \frac{k^{q/s}}{k^{1+q/s}} \right)^{1/q} \lesssim_\Gamma K_q(X) n^{n_h/q} D, \tag{2.34}$$

which gives Theorem 2.0.2. □

Proof of Corollary 2.0.4. The lower bound of Corollary 2.0.4 follows from Theorem 2.0.2. The upper bound follows from a version [50, Theorem 5.1] of the Assouad embedding theorem [3]:

$$c_p(\Gamma, d_\Gamma^{1-\varepsilon}) \lesssim_{\Gamma,p} 1/\varepsilon^{1/\max\{p,2\}}, \quad 0 < \varepsilon < 1,$$

Also, $(B_n^\Gamma, d_W^{1-1/2\log n})$ is $O(1)$-bilipschitz equivalent to (B_n^Γ, d_W), so we have the upper bound. □

Proof of Corollary 2.0.3. The lower bound of Corollary 2.0.3 follows from Theorem 2.0.2. To show the upper bound, we note that the proof of the aforementioned version [50, Theorem 5.1] of the Assouad embedding theorem [3], with $\varepsilon = \frac{1}{2\log R}$, produces a mapping $\psi : G \to L^p$ such that ψ is $O_G\left((\log R)^{1/\max\{p,2\}}\right)$-Lipschitz and satisfies

$$\|\psi(g) - \psi(h)\|_{L^p} \gtrsim_G d_G(g,h)^{1-1/2\log R}, \quad g, h \in G \text{ such that } d_G(g,h) = \Omega(1).$$

It is thus enough to construct a mapping $\phi : G \to \mathbb{R}^d$ such that ϕ is $O(1)$-Lipschitz and

$$|\phi(g) - \phi(h)| \gtrsim_G d_G(g,h), \quad g, h \in G \text{ such that } d_G(g,h) \leq 1,$$

for then $\psi \oplus \phi : G \to L^p \oplus \mathbb{R}^d = L^p$ gives the desired mapping.

Since G is Riemannian and $\exp : \mathfrak{g} \to G$ is a diffeomorphism, exp is $O(1)$-bilipschitz on balls of radius $O(1)$. The idea is to 'glue' these mappings together to produce the mapping ϕ.

Let N be a $(1,1)$-net of G. Since G is doubling, there is a constant K such that $N = \cup_{i=1}^K N_i$, where N_i are 8-packings of G. For each $i = 1, \cdots, K$ extend the mapping

$$\cup_{g \in N_i} gB_2 \to \mathfrak{g}, \quad h \in gB_2 \mapsto \exp^{-1}(g^{-1}h)$$

to an $O(\log K)$-Lipschitz mapping $\phi_i : G \to \mathfrak{g}$ by [51], and set $\phi = \bigoplus_{i=1}^{K} \phi_i$. Then ϕ has the desired properties. $\square$

Proof of Theorem 2.0.10. Let N_{r_1,r_2} be an r_2-covering of B_{r_1} and let $f : N_{r_1,r_2} \to X$ be a mapping with

$$d_G(x,y) \le \|f(x) - f(y)\| \le D d_G(x,y), \quad x, y \in N_{r_1,r_2}.$$

By [51] there is an extension $F : G \to X$ of f that is $CD \log K$-Lipschitz for a universal constant $C > 0$. Let ξ be a smooth bump function that equals 1 on B_{r_1}, is supported on B_{2r_1} and is $O(1)$-Lipschitz. Then $\phi \coloneqq \xi F$ is $O(D \log K)$-Lipschitz, equals F on B_{r_1} and is supported on B_{2r_1}. To apply Theorem 2.0.5 with $p = q$, we note that

$$\left(\int_G \|\nabla \phi(h)\|_{\ell_2^k(X)}^q \, d\mu(h) \right)^{1/q} \lesssim \sqrt{k} D (\log K) \mu(B_{2r_1})^{1/q}.$$

On the other hand, for $x \in B_{r_1/2}$ and $t \in ((\frac{(4+4CD\log K)r_2}{C_1})^s, (r_1/2)^s]$ (we may assume $(\frac{(4+4CD\log K)r_2}{C_1})^s \le (r_1/2)^s$ since otherwise $D > \frac{C_1(r_1/r_2)}{16C \log K}$ and the proof is complete), we have $\phi(x) = F(x)$ and $\phi(xv^t) = F(xv^t)$. Choosing $n_1, n_2 \in N$ such that $d_G(x, n_1) < r_2$, $d_G(xv^t, n_2) < r_2$, we have

$$\begin{aligned} \|F(xv^t) - F(x)\| &\ge \|f(n_1) - f(n_2)\| - 2CDr_2 \log K \ge d(n_1, n_2) - 2CDr_2 \log K \\ &\ge (C_1 t^{1/s} - 2r_2) - 2CDr_2 \log K > \frac{C_1}{2} t^{1/s}. \end{aligned}$$

Thus

$$\begin{aligned}
&\left(\int_0^\infty \int_G \left(\frac{\|\phi(hv^t)-\phi(h)\|_X}{t^{1/s}}\right)^q d\mu(h)\frac{dt}{t}\right)^{1/q} \\
&\geq \left(\int_{(\frac{(4+4CD\log K)r_2}{C_1})^s}^{(r_1/2)^s} \int_{B_{r_1/2}} \frac{C_1^q}{2^q} d\mu(h)\frac{dt}{t}\right)^{1/q} \\
&\geq \frac{C_1 s^{1/q}}{2}\left(\log(r_1/2)-\log(\frac{(4+4CD\log K)r_2}{C_1})\right)^{1/q}\mu(B_{r_1/2})^{1/q}.
\end{aligned}$$

Thus by Theorem 2.0.5 with $p=q$,

$$\begin{aligned}
&\frac{C_1 s^{1/q}}{2}\left(\log(r_1/2)-\log(\frac{(4+4CD\log K)r_2}{C_1})\right)^{1/q}\mu(B_{r_1/2})^{1/q} \\
&\lesssim K_q(X)\sqrt{k}D(\log K)\mu(B_{2r_1})^{1/q}.
\end{aligned}$$

But we may also assume $\log(r_1/2)-\log(\frac{(4+4CD\log K)r_2}{C_1}) \geq \frac{1}{2}\log(r_1/r_2)$, for otherwise $D > \frac{C_1\sqrt{r_1/r_2}}{16C\log K}$ and the proof is complete. Thus

$$C_1 s^{1/q}(\log(r_1/r_2))^{1/q} \lesssim K_q(X)\sqrt{k}DK^{2/q}\log K,$$

from which it follows that

$$D \gtrsim \left(\frac{C_1 s^{1/q}}{\sqrt{k}K^{2/q}\log K}\right)\frac{(\log(r_1/r_2))^{1/q}}{K_q(X)}.$$

□

Proof of Theorem 2.0.11. Let $N \coloneqq N'_{c_1,c_2}$ be a (c_1,c_2)-net of G and let $f: N \to X$ be a mapping with

$$d_G(x,y)^{1-\varepsilon} \leq \|f(x)-f(y)\| \leq D d_G(x,y)^{1-\varepsilon}, \quad x,y \in N.$$

Then f is $c_2^{-\varepsilon}D$-Lipschitz, so by [51] there is an extension $F: G \to X$ of f that is

$Cc_2^{-\varepsilon} D \log K$-Lipschitz for a universal constant $C > 0$. Fix $R > 0$ sufficiently large, let ξ be a smooth bump function that equals 1 on B_R, is supported on B_{2R} and is $O(1)$-Lipschitz. Then $\phi := \xi F$ is $O(c_2^{-\varepsilon} D \log K)$-Lipschitz, equals F on B_R and is supported on B_{2R}. To apply Theorem 2.0.5 with $p = q$, we note that

$$\left(\int_G \|\nabla \phi(h)\|_{\ell_2^k(X)}^q \, d\mu(h) \right)^{1/q} \lesssim \sqrt{k} c_2^{-\varepsilon} D (\log K) \mu(B_{2R})^{1/q}.$$

On the other hand, for $x \in B_{R/2}$ and $t \in ((8c_1 c_2^{-\varepsilon} D \log K)^{s/(1-\varepsilon)}/C_1^s, (R/2)^s]$ (this interval is nonempty since we took $R > 0$ sufficiently large) $(8c_1 c_2^{-\varepsilon} D \log K)^{s/(1-\varepsilon)}/C_1^s \geq 2^s c_1^s C_1^s$, we have $\phi(x) = F(x)$ and $\phi(xv^t) = F(xv^t)$. Choosing $n_1, n_2 \in N$ such that $d_G(x, n_1) < c_1$, $d_G(xv^t, n_2) < c_1$, we have

$$\begin{aligned}
\|F(xv^t) - F(x)\| \geq & \|f(n_1) - f(n_2)\| - 2c_1 c_2^{-\varepsilon} C D \log K \\
\geq & d(n_1, n_2)^{1-\varepsilon} - 2c_1 c_2^{-\varepsilon} C D \log K \\
\geq & (C_1 t^{1/s} - 2c_1)^{1-\varepsilon} - 2c_1 c_2^{-\varepsilon} C D \log K > \frac{C_1^{1-\varepsilon}}{4} t^{(1-\varepsilon)/s}.
\end{aligned}$$

Thus

$$\begin{aligned}
&\left(\int_0^\infty \int_G \left(\frac{\|\phi(hv^t) - \phi(h)\|_X}{t^{1/s}} \right)^q d\mu(h) \frac{dt}{t} \right)^{1/q} \\
&\geq \left(\int_{(8c_1 c_2^{-\varepsilon} D \log K)^{s/(1-\varepsilon)}/C_1^s}^{(R/2)^s} \int_{B_{R/2}} \frac{C_1^{q(1-\varepsilon)}}{4^q} d\mu(h) \frac{dt}{t^{1+\varepsilon q/s}} \right)^{1/q} \\
&\geq \frac{C_1^{1-\varepsilon}}{4} \left(\frac{s}{\varepsilon q} \left[(8c_1 c_2^{-\varepsilon} D \log K)^{-\varepsilon q/(1-\varepsilon)}/C_1^{-\varepsilon q} - (R/2)^{-\varepsilon q} \right] \right)^{1/q} \mu(B_{R/2})^{1/q}.
\end{aligned}$$

Thus by Theorem 2.0.5 with $p = q$,

$$\begin{aligned}
&\frac{C_1^{1-\varepsilon}}{4} \left(\frac{s}{\varepsilon q} \left[(8c_1 c_2^{-\varepsilon} D \log K)^{-\varepsilon q/(1-\varepsilon)}/C_1^{-\varepsilon q} - (R/2)^{-\varepsilon q} \right] \right)^{1/q} \mu(B_{R/2})^{1/q} \\
&\lesssim K_q(X) \sqrt{k} c_2^{-\varepsilon} D (\log K) \mu(B_{2R})^{1/q}.
\end{aligned}$$

Since $R > 0$ was arbitrarily large,

$$\frac{C_1^{1-\varepsilon} s^{1/q}}{\varepsilon^{1/q}} (8c_1 c_2^{-\varepsilon} D \log K)^{-\varepsilon/(1-\varepsilon)} / C_1^{-\varepsilon} \lesssim K_q(X) \sqrt{k} c_2^{-\varepsilon} D K^{2/q} \log K.$$

It follows that

$$D \gtrsim \left(\frac{C_1^{1-\varepsilon} (c_2/c_1)^{\varepsilon} s^{(1-\varepsilon)/q}}{k^{(1-\varepsilon)/2} K^{2(1-\varepsilon)/q} \log K} \right) \frac{1}{K_q(X)^{1-\varepsilon} \varepsilon^{1/q}}.$$

□

The proof of Theorem 2.0.14 is similar to that of Theorem 2.0.10 and the proof of Theorem 2.0.15 is similar to that of Theorem 2.0.11 and we thus omit them. To see the dependence on the constants, we only need to distinguish when K is used as the metric doubling constant or the measure doubling constant.

2.3 Proof of Theorem 2.0.6: discretization

In this section, we prove Theorem 2.0.6 by discretizing the inequality of Theorem 2.0.5.

Finitely generated groups of polynomial growth have nilpotent subgroups of finite index [30], so let $\Gamma' \leq \Gamma$ be a finite index nilpotent subgroup. We first can see that it is enough to prove Theorem 2.0.6 for Γ', i.e., we may assume Γ is a not virtually abelian finitely generated nilpotent group. Indeed, the distance requirement $d_W(v_\Gamma^n, e_\Gamma) \asymp_G n^{1/s}$ carries over from Γ' to Γ because a finitely generated word metric on a group and its finite index subgroup are known to be quasi-isometric. Also, once we know (2.3) for Γ' and $S' \subset \Gamma'$, it automatically follows that it holds for Γ and $S \supset S'$. (Observe that changing the generating set only affects the right-hand side of (2.3) up to universal constant factors.)

Let $T \coloneqq \{x \in \Gamma' : \exists n \in \mathbb{Z}_{>0} \text{ such that } x^n = e_{\Gamma'}\}$ be the torsion subgroup of Γ'. It is well-known that T is a finite normal subgroup of Γ', so that $\Gamma'' \coloneqq \Gamma'/T$ is a torsion-

free finitely generated nilpotent group. We now see that it is enough to prove Theorem 2.0.6 for Γ'', i.e., we may assume Γ is a torsion-free finitely generated nonabelian nilpotent group. Indeed, again the distance requirement $d_W(v_\Gamma^n, e_\Gamma) \asymp_G n^{1/s}$ carries over from Γ'' to Γ' because a finitely generated word metric on a group and its quotient by a finite normal subgroup are known to be quasi-isometric. Also, suppose we know (2.3) for Γ'' and $S'' \subset \Gamma''$. Let S' be the union of T and a set of representatives for S'' in Γ', and let $v_{\Gamma'}$ be a representative for $v_{\Gamma''}$ in Γ'. Let $\pi : \Gamma' \to \Gamma''$ denote the quotient map, and let $C > 0$ be such that $\pi(B_k^{\Gamma'}(x)) \subset B_{Ck}^{\Gamma''}(\pi(x))$ for $k \in \mathbb{Z}_{>0}$ and $x \in \Gamma'$ and $\pi^{-1}(B_k^{\Gamma''}(y)) \subset B_{Ck}^{\Gamma'}(x)$ for $k \in \mathbb{Z}_{>0}$, $y \in \Gamma''$ and $x \in \pi^{-1}(y)$. Given $f : \Gamma' \to X$, let $[f] : \Gamma'' \to X$ be given as

$$[f](y) \coloneqq \frac{1}{|T|} \sum_{x \in \pi^{-1}(y)} f(x), \quad y \in \Gamma''.$$

Then, denoting $p \vee q \coloneqq \max\{p, q\}$,

$$\begin{aligned}
&\left(\sum_{k=1}^{n^s} \frac{1}{k^{1+(p\vee q)/s}} \left(\sum_{x \in B_n^{\Gamma'}} \|f(xv_{\Gamma'}^k) - f(x)\|_X^p \right)^{p\vee q/p} \right)^{1/p\vee q} \\
&\leq \left(\sum_{k=1}^{n^s} \frac{1}{k^{1+(p\vee q)/s}} \left(\sum_{x \in B_n^{\Gamma'}} \left\| \big(f(xv_{\Gamma'}^k) - [f]([xv_{\Gamma'}^k])\big) - \big(f(x) - [f]([x])\big) \right\|_X^p \right)^{p\vee q/p} \right)^{1/p\vee q} \\
&\quad + \left(\sum_{k=1}^{n^s} \frac{1}{k^{1+(p\vee q)/s}} \left(\sum_{x \in B_n^{\Gamma'}} \left\| [f]([xv_{\Gamma'}^k]) - [f]([x]) \right\|_X^p \right)^{p\vee q/p} \right)^{1/p\vee q} .
\end{aligned} \tag{2.35}$$

The second term is at most

$$\begin{aligned}
&\left(\sum_{k=1}^{n^s} \frac{1}{k^{1+p\vee q/s}} \left(|T| \sum_{y\in B^{\Gamma''}_{Cn}} \|[f](yv^k_{\Gamma''}) - [f](y)\|_X^p\right)^{p\vee q/p}\right)^{1/p\vee q} \\
&\lesssim_\Gamma \max\left\{(p-1)^{1/q-1}, K_q(X)\right\} \left(\sum_{y\in B^{\Gamma'}_{cCn}} \sum_{a\in S''} \|[f](ya) - [f](y)\|_X^p\right)^{1/p} \\
&= \max\left\{(p-1)^{1/q-1}, K_q(X)\right\} \left(\sum_{y\in B^{\Gamma'}_{cCn}} \sum_{a\in S''} \left\|\frac{1}{|T|} \sum_{x\in\pi^{-1}(y)} (f(xa') - f(x))\right\|_X^p\right)^{1/p} \\
&\leq \max\left\{(p-1)^{1/q-1}, K_q(X)\right\} \left(\sum_{y\in B^{\Gamma'}_{cCn}} \sum_{a\in S''} \frac{1}{|T|} \sum_{x\in\pi^{-1}(y)} \|f(xa') - f(x)\|_X^p\right)^{1/p} \\
&\leq \max\left\{(p-1)^{1/q-1}, K_q(X)\right\} \left(\sum_{x\in B^{\Gamma'}_{cC^2n}} \sum_{a\in S''} \frac{1}{|T|} \|f(xa') - f(x)\|_X^p\right)^{1/p},
\end{aligned}$$

and is thus bounded by the right-hand side of (2.3), where for $a \in S''$, $a' \in S'$ denotes an element of $\pi^{-1}(a)$.

It remains to show that the first term of the right-hand side of (2.35) is bounded

by the right-hand side of (2.3). Indeed, it is equal to

$$\left(\sum_{k=1}^{n^s}\frac{1}{k^{1+p\vee q/s}}\left(\sum_{x\in B_n^{\Gamma'}}\left\|\frac{1}{|T|}\sum_{b\in T}\big(f(xv_{\Gamma'}^k)-f(xv_{\Gamma'}^kb)-f(x)+f(xb)\big)\right\|_X^p\right)^{p\vee q/p}\right)^{1/p\vee q}$$
$$\leq\left(\sum_{k=1}^{n^s}\frac{1}{k^{1+p\vee q/s}}\left(\frac{2^{p-1}}{|T|}\sum_{x\in B_n^{\Gamma'}}\sum_{b\in T}\left(\begin{array}{c}\|f(xv_{\Gamma'}^k)-f(xv_{\Gamma'}^kb)\|_X^p\\+\|f(x)-f(xb)\|_X^p\end{array}\right)\right)^{p\vee q/p}\right)^{1/p\vee q}$$
$$\leq\left(\sum_{k=1}^{n^s}\frac{1}{k^{1+p\vee q/s}}\left(\frac{2^p}{|T|}\sum_{x\in B_{c'n}^{\Gamma'}}\sum_{b\in T}\|f(x)-f(xb)\|_X^p\right)^{p\vee q/p}\right)^{1/p\vee q}$$
$$\lesssim_\Gamma\left(\sum_{x\in B_{c'n}^{\Gamma'}}\sum_{b\in T}\|f(x)-f(xb)\|_X^p\right)^{1/p}$$

where $c'>0$ is such that $B_n^{\Gamma'}v_{\Gamma'}^k\subset B_{c'n}^{\Gamma'}$ for $n\in\mathbb{N}$ and $k=1,\cdots,n^s$.

This completes our reduction of Theorem 2.0.6 to proving it for Γ being a torsion-free finitely generated nonabelian nilpotent group. By the Malcev embedding theorem [55], there exists a simply connected nilpotent Lie group G, called the Malcev completion of Γ, such that Γ embeds as a cocompact subgroup of G. Because Γ is nonabelian, G is nonabelian, so Theorem 2.0.5 applies to G.

There exist choices of s and v_Γ as stated. It is enough to take $v\in Z(\Gamma)\setminus\{e_\Gamma\}$ with $d_W(v^n,e_\Gamma)\asymp_\Gamma n^{1/s}$ for some integer $s\geq 2$. For example, let s denote the nilpotency step of G, i.e., the largest integer so that $\underbrace{[[G,G]\cdots,G]}_{s\text{ times}}$ is a nontrivial subgroup. If we choose $v_\Gamma\in\underbrace{[[\Gamma,\Gamma]\cdots,\Gamma]}_{s\text{ times}}\setminus\{e_\Gamma\}\subset Z(\Gamma')\setminus\{e_\Gamma\}$, then

$$d_W(v_\Gamma^k,e_\Gamma)\asymp_{\Gamma,v_\Gamma}k^{1/s},\quad k\in\mathbb{N}$$

because by [10, Theorem 1.3], $|d_W(v_\Gamma^k,e_\Gamma)-d_G(v_\Gamma^k,e_G)|=O_{\Gamma,v_\Gamma}(d_G(v_\Gamma^k,e_G)^{1-\alpha})$ for some fixed $\alpha=\alpha_G$, and $d_G(v^k,e_G)\asymp_G k^{1/s}$ by [10, Proposition 2.13].

Also, $|B_m^\Gamma| \asymp_\Gamma m^{n_h}$ for every $m \in \mathbb{N}$ [10, Theorem 1.1], where n_h is given by the Bass–Guivarc'h formula

$$n_h \coloneqq \sum_{k=1}^{s} k \dim \left(\underbrace{[[G,G]\cdots,G]}_{k \text{ times}} / \underbrace{[[G,G]\cdots,G]}_{k+1 \text{ times}} \right).$$

The rest of this section follows closely the argument of Section 3 of [47].

Before we begin the proof of Theorem 2.0.6, we prove two metric-space valued local Poincaré inequalities on Γ for preparation. Lemma 2.3.1 is an extension of the local Poincaré inequality of Kleiner [46, Thm. 2.2] and was essentially proven in [47, Lemma 3.2] for all finitely generated groups. We repeat the proof for completeness.

Lemma 2.3.1 (Analogue of [47, Lemma 3.2])**.** *Fix $p \in [1,\infty)$ and $n \in \mathbb{N}$. Let (M, d_M) be a metric space. For every $f : \Gamma \to M$,*

$$\sum_{x,y \in B_n^\Gamma} d_M(f(x), f(y))^p \leq (2n)^p |B_{2n}^\Gamma| \sum_{x \in B_{3n}^\Gamma} \sum_{a \in S} d_M(f(xa), f(x))^p.$$

Proof. For every $z \in B_{2n}^\Gamma$ choose $s_1(z), \ldots, s_{2n}(z) \in S \cup \{e_\Gamma\}$ such that

$$z = s_1(z) \cdots s_{2n}(z).$$

For $i \in \{1, \ldots, 2n\}$ write $w_i(z) = s_1(z) \cdots s_i(z)$ and set $w_0(z) = e_\Gamma$. By the triangle inequality and Hölder's inequality, for every $x, y \in B_n^\Gamma$ we have

$$d_M(f(x), f(y))^p \leq (2n)^{p-1} \sum_{i=0}^{2n-1} d_M\left(f\left(xw_i\left(x^{-1}y\right)\right), f\left(xw_i\left(x^{-1}y\right) s_{i+1}\left(x^{-1}y\right)\right)\right)^p.$$

Consequently,

$$\begin{aligned}\sum_{x,y\in B_n^\Gamma} d_M(f(x),f(y))^p &\le (2n)^{p-1}\sum_{z\in B_{2n}^\Gamma}\sum_{i=1}^{2n-1}\sum_{x\in B_n^\Gamma} d_M\left(f\left(xw_i\left(z\right)\right), f\left(xw_i\left(z\right)s_{i+1}\left(z\right)\right)\right)^p \\ &= (2n)^{p-1}\sum_{z\in B_{2n}^\Gamma}\sum_{i=0}^{2n-1}\sum_{g\in B_n^\Gamma w_i(z)} d_M\left(f(g), f(gs_{i+1}(z))\right)^p \\ &\le (2n)^{p-1}\cdot|B_{2n}^\Gamma|\cdot 2n\sum_{x\in B_{3n}^\Gamma}\sum_{a\in S} d_M(f(xa),f(x))^p.\end{aligned}$$

□

The following lemma is an analogue of [47, Lemma 3.4], which was again proven essentially for all finitely generated groups. A mapping $f:\Gamma\to M$ is said to be finitely supported if there exists $m_0\in M$ such that $|f^{-1}(M\setminus\{m_0\})|<\infty$.

Lemma 2.3.2 (Analogue of [47, Lemma 3.4])**.** *Fix $p\in[1,\infty)$ and $n\in\mathbb{N}$. Let (M,d_M) be a metric space. For every finitely supported $f:\Gamma\to M$,*

$$\sum_{x\in\Gamma}\sum_{z\in B_n^\Gamma} d_M(f(xz),f(x))^p \le n^p|B_n^\Gamma|\sum_{x\in\Gamma}\sum_{a\in S} d_M(f(xa),f(x))^p.$$

Proof. Let s_i and w_i be as in the proof of Lemma 2.3.1. By the triangle inequality and Hölder's inequality,

$$\begin{aligned}\sum_{x\in\Gamma}\sum_{z\in B_n^\Gamma} d_M(f(xz),f(x))^p &\le n^{p-1}\sum_{x\in\Gamma}\sum_{z\in B_n^\Gamma}\sum_{i=0}^{n-1} d_M\left(f\left(xw_i\left(z\right)\right), f\left(xw_i\left(z\right)s_{i+1}\left(z\right)\right)\right)^p. \\ &\le n^p|B_n^\Gamma|\sum_{x\in\Gamma}\sum_{a\in S} d_M(f(xa),f(x))^p.\end{aligned}$$

□

With Lemmas 2.3.1 and 2.3.2, we begin by discretizing the inequality of Theorem 2.0.5, as in the following theorem. After this, we will localize the statement into that of Theorem 2.0.6, which states a local version of Theorem 2.3.3 on balls in Γ.

The following theorem corresponds to Theorem 3.3 of [47], whose proof is in turn a variant of the proof of Claim 7.3 in [4]. We may state it in terms of discrete groups of polynomial growth.

Theorem 2.3.3 (Analogue of [47, Theorem 3.3])**.** *Let Γ be a not virtually abelian finitely generated group of polynomial growth. There exist $v_\Gamma \in \Gamma$, $s \in \mathbb{N}$ with $s \geq 2$, and $c = c(\Gamma) \in \mathbb{N}$ such that the following is true. First, $d_W(v_\Gamma^n, e_\Gamma) \asymp_\Gamma n^{1/s}$ for $n \in \mathbb{N}$. Second, let $p \in (1, \infty)$ and $q \in [2, \infty)$. Suppose that $(X, \|\cdot\|_X)$ is a Banach space satisfying $K_q(X) < \infty$. Then for every finitely supported $f : \Gamma \to X$ we have*

$$\left(\sum_{k=1}^{n^s} \frac{1}{k^{1+\max\{p,q\}/s}} \left(\sum_{x \in \Gamma} \|f(xv_\Gamma^k) - f(x)\|_X^p \right)^{\max\{p,q\}/p} \right)^{1/\max\{p,q\}}$$
$$\lesssim_{\Gamma, v_\Gamma} \max\left\{(p-1)^{1/q-1}, K_q(X)\right\} \left(\sum_{x \in \Gamma} \sum_{a \in S} \|f(xa) - f(x)\|_X^p \right)^{1/p}.$$

Proof. With the same argument as in the beginning of this section, we may assume Γ is a torsion-free nonabelian finitely generated nilpotent group, and take v_Γ and s as before.

The idea of the proof is that given a finitely supported function $f : \Gamma \to X$, we extend it to a global function $F : G \to X$ via a partition of unity, and then Theorem 2.0.5 for F will give Theorem 2.3.3 for f.

Let G be the Malcev completion of Γ. Since Γ is a co-compact lattice of G, there exists a compactly supported smooth function $\chi : G \to [0, 1]$ with

$$\forall\, h \in G, \qquad \sum_{x \in \Gamma} \chi_x(h) = 1, \tag{2.36}$$

where for each $x \in \Gamma$, $\chi_x : G \to X$ is given by $\chi_x(h) = \chi(x^{-1}h)$, $h \in G$. Let

$A = \operatorname{supp}\chi$; we may assume $A^{-1} = A$. Note that $\operatorname{supp}\chi_x = xA$ and

$$\bigcup_{x\in\Gamma} xA = G.$$

As A is compact, we may fix $m \in \mathbb{N}$ for which $A^2 \cap \Gamma \subseteq B_m^\Gamma$.

Let $f : \Gamma \to X$ be finitely supported. Define $F : G \to X$ by

$$F(h) \coloneqq \sum_{x\in\Gamma} \chi_x(h) f(x). \tag{2.37}$$

For a fixed $x \in \Gamma$ and $h \in xA$, we have

$$\sum_{y\in\Gamma} \nabla\chi_y(h) = 0,$$

with $\nabla\chi_y(h) \neq 0$ implying $y^{-1}h \in \operatorname{supp}\nabla\chi \subset A$, which implies $y \in hA^{-1} \subset xA^2$, hence $y \in xB_m^\Gamma$. We thus have the bound

$$\begin{aligned}
\|\nabla F(h)\|_{\ell_2^k(X)} &= \left\| \sum_{y\in xB_m^\Gamma} \nabla\chi_y(h)(f(y) - f(x)) \right\|_{\ell_2^k(X)} \\
&\leq \|\nabla\chi\|_{L^\infty(G;\ell_2^k)} \sum_{z\in B_m^\Gamma} \|f(xz) - f(x)\|_X \\
&\lesssim_\Gamma |B_m^\Gamma|^{1-1/p} \left(\sum_{z\in B_m^\Gamma} \|f(yz) - f(y)\|_X^p \right)^{1/p}.
\end{aligned}$$

By integrating over yA and summing over $y \in \Gamma$,

$$\begin{aligned}
&\left(\int_G \|\nabla F(h)\|_{\ell_2^k(X)}^p \, dh\right)^{1/p} \\
&\leq \left(\sum_{y\in\Gamma} \int_{yA} \|\nabla F(h)\|_{\ell_2^k(X)}^p \, dh\right)^{1/p} \lesssim_\Gamma \left(\sum_{y\in\Gamma} \sum_{z\in B_m^\Gamma} \|f(yz) - f(y)\|_X^p\right)^{1/p} \qquad (2.38) \\
&\overset{\text{Lemma 2.3.2}}{\lesssim_\Gamma} \left(\sum_{x\in\Gamma} \sum_{a\in S} \|f(xa) - f(x)\|_X^p\right)^{1/p}.
\end{aligned}$$

Since Γ induces a covering space action on G, we may find a bounded open neighborhood $U \subseteq G$ of e_G such that $U \cap (xU) = \emptyset$ for all $x \in \Gamma \setminus \{e_\Gamma\}$. As $v_\Gamma^{[0,1]} := \{v_\Gamma^\tau : \tau \in [0,1]\}$ and U are bounded in G, we can choose $r \in \mathbb{N}$ such that $(Uv_\Gamma^{[0,1]}A) \cap \Gamma \subseteq B_r^\Gamma$. For $x \in \Gamma$, $h \in xU$, $k \in \mathbb{N}$, and $t \in [k, k+1]$, we have by (2.36) and (2.37)

$$F(hv_\Gamma^t) - f(xv_\Gamma^k) = \sum_{w\in\Gamma} \chi_w(hv_\Gamma^t)(f(w) - f(xv_\Gamma^k)).$$

Every $w \in \Gamma$ that satisfies $\chi_w(hv_\Gamma^t) \neq 0$ should have $hv_\Gamma^t \in wA$, and since $A = A^{-1}$ and $v_\Gamma \in Z(G)$, we have $w \in hv_\Gamma^t A \subseteq xv_\Gamma^k U v_\Gamma^{[0,1]} A$, so that $w \in xv_\Gamma^k B_r^\Gamma$. Therefore

$$\|F(hv_\Gamma^t) - f(xv_\Gamma^k)\|_X \lesssim_\Gamma \left(\sum_{z\in B_r^\Gamma} \|f(xv_\Gamma^k z) - f(xv_\Gamma^k)\|_X^p\right)^{1/p}.$$

The case $k = t = 0$ gives

$$\|F(h) - f(x)\|_X \lesssim_\Gamma \left(\sum_{z\in B_r^\Gamma} \|f(xz) - f(x)\|_X^p\right)^{1/p}.$$

Therefore by the triangle inequality,

$$\|f(xv_\Gamma^k)-f(x)\|_X \lesssim_\Gamma \|F(hv_\Gamma^t)-F(h)\|_X + \left(\sum_{z\in B_r^\Gamma}\left(\|f(xv_\Gamma^k z)-f(xv_\Gamma^k)\|_X^p+\|f(xz)-f(x)\|_X^p\right)\right)^{1/p}.$$

Integration over $h\in xU$ gives

$$\|f(xv_\Gamma^k)-f(x)\|_X \lesssim_\Gamma \left(\int_{xU}\|F(hv_\Gamma^t)-F(h)\|_X^p dh\right)^{1/p} + \left(\sum_{z\in B_r^\Gamma}\left(\|f(xv_\Gamma^k z)-f(xv_\Gamma^k)\|_X^p+\|f(xz)-f(x)\|_X^p\right)\right)^{1/p}.$$

Summing over $x\in\Gamma$ and using that the sets $\{xU\}_{x\in\Gamma}$ are pairwise disjoint,

$$\begin{aligned}&\left(\sum_{x\in\Gamma}\|f(xv_\Gamma^k)-f(x)\|_X^p\right)^{1/p}\\ &\lesssim_\Gamma \left(\int_G\|F(hv_\Gamma^t)-F(h)\|_X^p dh\right)^{1/p}\\ &\quad+\left(\sum_{x\in\Gamma}\sum_{z\in B_r^\Gamma}\left(\|f(xv_\Gamma^k z)-f(xv_\Gamma^k)\|_X^p+\|f(xz)-f(x)\|_X^p\right)\right)^{1/p}\\ &\overset{\text{Lemma 2.3.2}}{\lesssim_\Gamma} \left(\int_G\|F(hv_\Gamma^t)-F(h)\|_X^p dh\right)^{1/p}+\left(\sum_{x\in\Gamma}\sum_{a\in S}\|f(xa)-f(x)\|_X^p\right)^{1/p}.\end{aligned}$$

Integrating over $t\in[k,k+1]$ yields

$$\begin{aligned}&\frac{1}{k^{1+\max\{p,q\}/s}}\left(\sum_{x\in\Gamma}\|f(xv_\Gamma^k)-f(x)\|_X^p\right)^{\max\{p,q\}/p}\\ &\le C^{\max\{p,q\}}\int_k^{k+1}\left(\int_G\|F(hv_\Gamma^t)-F(h)\|_X^p dh\right)^{\max\{p,q\}/p}\frac{dt}{t^{1+\max\{p,q\}/s}}\\ &\quad+\frac{C^{\max\{p,q\}}}{k^{1+\max\{p,q\}/s}}\left(\sum_{x\in\Gamma}\sum_{a\in S}\|f(xa)-f(x)\|_X^p\right)^{\max\{p,q\}/p},\end{aligned}$$

where $C = C_G \in (0, \infty)$ is a constant depending on G. Summing over $k \in \mathbb{N}$,

$$\begin{aligned}
&\left(\sum_{k=1}^{\infty} \frac{1}{k^{1+\max\{p,q\}/s}} \left(\sum_{x \in \Gamma} \|f(xv_\Gamma^k) - f(x)\|_X^p \right)^{\max\{p,q\}/p} \right)^{1/\max\{p,q\}} \\
&\lesssim_\Gamma \left(\int_0^\infty \left(\int_G \|F(hv_\Gamma^t) - F(h)\|_X^p dh \right)^{\max\{p,q\}/p} \frac{dt}{t^{1+\max\{p,q\}/s}} \right)^{1/\max\{p,q\}} \\
&\quad + \left(\sum_{x \in \Gamma} \sum_{a \in S} \|f(xa) - f(x)\|_X^p \right)^{1/p}.
\end{aligned} \tag{2.39}$$

The desired inequality follows from (2.38) and (2.39) along with Theorem 2.0.5 for F (although it may be the case that $d_G(v_\Gamma^t, e_G) \not\leq t^{1/s}$, a simple rescaling argument gives the same inequality up to constant factors depending on v_Γ). □

Now we are ready to prove Theorem 2.0.6. Of course, assume Γ is torsion-free nonabelian finitely generated nilpotent, and choose v_Γ and s as before.

Proof of Theorem 2.0.6. The argument follows the proof of Theorem 3.1 of [47], which in turn follows the proof of Claim 7.2 in [4]. Recall c is such that

$$d_W(e_\Gamma, v_\Gamma^k) \leq ck^{1/s}.$$

Fix $n \in \mathbb{N}$ and an $f : \Gamma \to X$. By translation, we may assume $\sum_{x \in B_{(c+3)n}^\Gamma} f(x) = 0$.

Then

$$\begin{aligned}\left(\sum_{x\in B^\Gamma_{(c+3)n}} \|f(x)\|_X^p\right)^{1/p} &= \left(\sum_{x\in B^\Gamma_{(c+3)n}} \left\|\frac{1}{|B^\Gamma_{(c+3)n}|}\sum_{y\in B^\Gamma_{(c+3)n}} (f(x)-f(y))\right\|_X^p\right)^{1/p}\\ &\le \left(\frac{1}{|B^\Gamma_{(c+3)n}|}\sum_{x,y\in B^\Gamma_{(c+3)n}} \|f(x)-f(y)\|_X^p\right)^{1/p}\\ &\overset{\text{Lemma 2.3.1}}{\lesssim_\Gamma} n\left(\sum_{x\in B^\Gamma_{3(c+3)n}}\sum_{a\in S}\|f(xa)-f(x)\|_X^p\right)^{1/p}.\end{aligned} \tag{2.40}$$

Define the cutoff function $\xi : \Gamma \to [0,1]$ as

$$\xi(x) \coloneqq \begin{cases} 1 & x \in B^\Gamma_{(c+1)n},\\ c+2-\frac{d_W(x,e)}{n} & x \in B^\Gamma_{(c+2)n}\setminus B^\Gamma_{(c+1)n},\\ 0 & x\in \Gamma\setminus B^\Gamma_{(c+2)n}.\end{cases}$$

Then $\phi \coloneqq \xi f$ is supported on $B^\Gamma_{(c+2)n}$ and we may apply Theorem 2.3.3. As ξ is $\frac{1}{n}$-Lipschitz and takes values in $[0,1]$, we have for all $a \in S$ and $x \in \Gamma$

$$\begin{aligned}\|\phi(x)-\phi(xa)\|_X &\le |\xi(x)-\xi(xa)|\cdot\|f(x)\|_X + |\xi(xa)|\cdot\|f(x)-f(xa)\|_X\\ &\le \frac{1}{n}\|f(x)\|_X + \|f(x)-f(xa)\|_X.\end{aligned} \tag{2.41}$$

We have $d_W(e_\Gamma, v_\Gamma^k) \le cn$ for all $k \in \{1,\dots,n^s\}$. Thus, for every $x \in B_n^\Gamma$ we have $xv_\Gamma^k \in B^\Gamma_{(c+1)n}$ and thus $\phi(x) = f(x)$ and $\phi(xv_\Gamma^k) = f(xv_\Gamma^k)$, so

$$\begin{aligned}&\left(\sum_{k=1}^{n^s}\frac{1}{k^{1+q/s}}\left(\sum_{x\in B_n^\Gamma}\|f(xv_\Gamma^k)-f(x)\|_X^p\right)^{q/p}\right)^{1/q}\\ &\le \left(\sum_{k=1}^{\infty}\frac{1}{k^{1+q/s}}\left(\sum_{x\in\Gamma}\|\phi(xv_\Gamma^k)-\phi(x)\|_X^p\right)^{q/p}\right)^{1/q}.\end{aligned} \tag{2.42}$$

On the other hand,

$$\begin{aligned}
&\left(\sum_{x\in\Gamma}\sum_{a\in S}\|\phi(xa)-\phi(x)\|_X^p\right)^{1/p}\\
&=\left(\sum_{x\in B^\Gamma_{(c+3)n}}\sum_{a\in S}\|\phi(xa)-\phi(x)\|_X^p\right)^{1/p}\\
&\overset{(2.41)}{\leq}\frac{|S|^{1/p}}{n}\left(\sum_{x\in B^\Gamma_{(c+3)n}}\|f(x)\|_X^p\right)^{1/p}+\left(\sum_{x\in B^\Gamma_{(c+3)n}}\sum_{a\in S}\|f(xa)-f(x)\|_X^p\right)^{1/p}\\
&\overset{(2.40)}{\lesssim_\Gamma}\left(\sum_{x\in B^\Gamma_{3(c+3)n}}\sum_{a\in S}\|f(xa)-f(x)\|_X^p\right)^{1/p},
\end{aligned}$$

where the first equality holds since ϕ is supported on $B^\Gamma_{(c+2)n}$. The desired inequality now follows from the above two inequalities combined with Theorem 2.3.3 for ϕ. $\square$

2.4 Proof of Theorem 2.0.9: sublinear growth of cocycles

Our goal in this section is to prove Theorem 2.0.9. The proof follows closely that of [4].

Let $(X,\|\cdot\|_X)$ be a q-uniformly convex space. We will see by [63] that is enough to prove Theorem 2.0.9 for 1-Lipschitz 1-cocycles. Here, $f\in Z^1(\pi)$ be a 1-cocycle with respect to an action $\pi:\Gamma\to\mathrm{Aut}(X)$ of Γ on X by linear isometric automorphisms if the function $f:\Gamma\to X$ satisfies $f(xy)=\pi(x)f(y)+f(x)$ for all $x,y\in\Gamma$. The 1-Lipschitz requirement for f is equivalent to $\|f(a)\|_X\leq 1$ for all $a\in S$.

We first recall the following property of q-uniformly convex spaces.

Lemma 2.4.1 ([4, Lemma 3.1])**.** *Let $(X,\|\cdot\|_X)$ be q-uniformly convex. For a fixed*

$z \in X$ and linear operator $T : X \to X$ with $\|T\| \leq 1$, define

$$s_n \coloneqq \frac{1}{2^n} \sum_{j=0}^{2^n-1} T^j z, \quad n \geq 0.$$

Then we have

$$\sum_{i=0}^{\infty} \frac{1}{2^l} \sum_{j=0}^{2^l-1} \left\| s_{(i+1)l} - T^{j2^{il}} s_{il} \right\|_X^q \leq (2K_q(X))^q \|z\|_X^q, \quad l \in \mathbb{N}. \tag{2.43}$$

On the other hand, as X is reflexive, X is ergodic [20, p.662], i.e., for every linear isometry $T : X \to X$ and $x \in X$ the sequence $\left\{ \frac{1}{n} \sum_{j=0}^{n-1} T^j x \right\}_{n=1}^{\infty}$ converges in norm. Thus, the operator $P : X \to X$ defined by

$$Px \coloneqq \lim_{N \to \infty} \frac{1}{N} \sum_{n=0}^{N-1} \pi(v)^n x, \quad x \in X,$$

is well-defined, has norm ≤ 1, is a contraction onto the subspace X_0 of $\pi(v)$-invariant vectors, and is idempotent. As v is a central element of Γ, P commutes with $\pi(g)$ for all $g \in \Gamma$, and so the maps $Pf, (I - P)f : \Gamma \to X$ are both Lipschitz and are 1-cocycles.

Now define linear operators $P_n : X \to X$, $n \in \mathbb{N}$, by

$$P_n \coloneqq \frac{1}{2^n} \sum_{j=0}^{2^n-1} \pi(v)^j.$$

Of course, $\|P_n\| \leq 1$ and P_n commutes with $\pi(g)$ for all $g \in \Gamma$.

Lemma 2.4.2 (Analogue of [4, Lemma 4.1])**.** *Let $(X, \|\cdot\|_X)$ be a q-uniformly convex space. Then for every $l, k, m \in \mathbb{N}$ there exist integers $i \in [k+1, k+m]$ and $j \in [0, 2^l-1]$ such that*

$$\left\| \pi\left(v^{-j2^{il}}\right) P_{(i+1)l} f\left(v^{\lfloor n^\rho \rfloor}\right) - P_{il} f\left(v^{\lfloor n^\rho \rfloor}\right) \right\|_X \lesssim_G \frac{K_q(X) n}{m^{1/p}}. \tag{2.44}$$

Proof. By Lemma 2.4.1 and the fact that $\|f(a)\|_X \leq 1$, we have for each $a \in S$

$$\begin{aligned}
&\sum_{i=k+1}^{k+m} \frac{1}{2^l} \sum_{j=0}^{2^l-1} \left\| \pi\left(v^{-j2^{il}}\right) P_{(i+1)l} f(a) - P_{il} f(a) \right\|_X^q \\
&= \sum_{i=k+1}^{k+m} \frac{1}{2^l} \sum_{j=0}^{2^l-1} \left\| P_{(i+1)l} f(a) - \pi\left(v^{j2^{il}}\right) P_{il} f(a) \right\|_X^q \leq (2K_q(X))^q.
\end{aligned}$$

Thus

$$\begin{aligned}
|S|(2K_q(X))^q &\geq \sum_{i=k+1}^{k+m} \frac{1}{2^l} \sum_{j=0}^{2^l-1} \sum_{a\in S} \left\| \pi\left(v^{-j2^{il}}\right) P_{(i+1)l} f(a) - P_{il} f(a) \right\|_X^q \\
&\geq m \min_{\substack{k+1\leq i\leq k+m \\ 1\leq j\leq 2^l-1}} \sum_{a\in S} \left\| \pi\left(v^{-j2^{il}}\right) P_{(i+1)l} f(a) - P_{il} f(a) \right\|_X^q,
\end{aligned}$$

and so there exist $i \in [k+1, k+m]$, $j \in [0, 2^l - 1]$ such that

$$\max_{a\in S} \left\| \pi\left(v^{-j2^{il}}\right) P_{(i+1)l} f(a) - P_{il} f(a) \right\|_X \lesssim_\Gamma \frac{K_q(X)}{m^{1/q}}.$$

As $v^{\lfloor n^\rho \rfloor} = a_1 \cdots a_b$ for some $a_i \in S$, where $b = O_\Gamma(n)$, we have by the cocycle identity

$$f(v^{\lfloor n^\rho \rfloor}) = \sum_{i=1}^{b} \pi(a_1 \cdots a_{i-1}) f(a_i).$$

Since P_r, $r \geq 0$, and $\pi(g)$, $g \in \Gamma$, commute, and v and $g \in \Gamma$ commute, and $\pi(g)$ is an isometry for $g \in \Gamma$, we have

$$\begin{aligned}
&\left\| \pi\left(v^{-j2^{il}}\right) P_{(i+1)l} f\left(v^{\lfloor n^\rho \rfloor}\right) - P_{il} f\left(v^{\lfloor n^\rho \rfloor}\right) \right\|_X \\
&\leq \sum_{i=1}^{b} \left\| \pi\left(v^{-j2^{il}}\right) P_{(i+1)l} f\left(a_i\right) - P_{il} f\left(a_i\right) \right\|_X \lesssim_\Gamma \frac{K_q(X) n}{m^{1/q}},
\end{aligned}$$

as claimed. □

Lemma 2.4.3 (Analogue of [4, Lemma 4.2])**.** *For every $m, n \in \mathbb{N}$,*

$$\left\|P_m f\left(v^{\lfloor n^\rho \rfloor}\right)\right\|_X \lesssim_\Gamma \frac{n^{(\rho^2+\rho-1)/(2\rho-1)}}{2^{m(\rho-1)/(2\rho-1)}}.$$

Proof. If $\frac{n^\rho}{2^m} \geq \rho - 1$ then we have the obvious bound

$$\left\|P_m f\left(v^{\lfloor n^\rho \rfloor}\right)\right\|_X \leq \left\|f\left(v^{\lfloor n^\rho \rfloor}\right)\right\|_X \leq d_W\left(e_\Gamma, v^{\lfloor n^\rho \rfloor}\right) \lesssim_\Gamma n\left(\frac{n^\rho}{2^m}\right)^{(\rho-1)/(2\rho-1)}.$$

Thus we assume $\frac{n^\rho}{2^m} < \rho - 1$.

Because

$$P_m - \pi(v^k)P_m = \frac{1}{2^m}\sum_{j=0}^{k-1}\pi(v)^j - \frac{1}{2^m}\sum_{j=2^m}^{2^m+k-1}\pi(v)^j,$$

we have

$$\left\|P_m - \pi(v^k)P_m\right\| \leq \frac{2k}{2^m}.$$

By the cocycle identity,

$$f\left(v^{\lfloor k^\rho \rfloor \lfloor n^\rho \rfloor}\right) = \sum_{j=0}^{\lfloor k^\rho \rfloor - 1} \pi\left(v^{j\lfloor n^\rho \rfloor}\right) f\left(v^{\lfloor n^\rho \rfloor}\right)$$

Because f is 1-Lipschitz, $d_W\left(e_\Gamma, v^{\lfloor k^\rho \rfloor \lfloor n^\rho \rfloor}\right) \lesssim_\Gamma kn$, and $\|P_m\| \leq 1$,

$$\begin{aligned}
kn &\gtrsim_\Gamma \left\|P_m f\left(v^{\lfloor k^\rho \rfloor \lfloor n^\rho \rfloor}\right)\right\|_X \\
&= \left\| \lfloor k^\rho \rfloor P_m f\left(v^{\lfloor n^\rho \rfloor}\right) - \sum_{j=0}^{\lfloor k^\rho \rfloor - 1} \left(P_m - \pi\left(v^{j\lfloor n^\rho \rfloor}\right) P_m\right) f\left(v^{\lfloor n^\rho \rfloor}\right) \right\|_X \\
&\geq \lfloor k^\rho \rfloor \left\|P_m f\left(v^{\lfloor n^\rho \rfloor}\right)\right\|_X - \sum_{j=0}^{\lfloor k^\rho \rfloor - 1} \left\|P_m - \pi\left(v^{j\lfloor n^\rho \rfloor}\right) P_m\right\| \cdot \left\|f\left(v^{\lfloor n^\rho \rfloor}\right)\right\|_X \\
&\geq \lfloor k^\rho \rfloor \left\|P_m f\left(v^{\lfloor n^\rho \rfloor}\right)\right\|_X - \sum_{j=0}^{\lfloor k^\rho \rfloor - 1} \frac{2j\lfloor n^\rho \rfloor}{2^m} d_W\left(e_\Gamma, v^{\lfloor n^\rho \rfloor}\right)
\end{aligned}$$

and rearranging terms,

$$\left\|P_m f\left(v^{\lfloor n^\rho \rfloor}\right)\right\|_X \lesssim_\Gamma \frac{n}{k^{\rho-1}} + \frac{n^{\rho+1}k^\rho}{2^m}.$$

Since this is true for all k, choosing $k = \left\lceil \left(\frac{(\rho-1)2^m}{\rho n^\rho}\right)^{1/(2\rho-1)} \right\rceil$ gives the stated bound. Indeed, by $\frac{n^\rho}{2^m} < \rho - 1$ we have $k \asymp \left(\frac{(\rho-1)2^m}{\rho n^\rho}\right)^{1/(2\rho-1)}$ and plugging in gives the desired bound. □

Lemma 2.4.4 (Analogue of [4, Lemma 4.3])**.** *For every $m, n \in \mathbb{N}$,*

$$\left\|f\left(v^{\lfloor n^\rho \rfloor}\right) - P_m f\left(v^{\lfloor n^\rho \rfloor}\right)\right\|_X \lesssim_\Gamma 2^{m/(\rho+1)} n^{1/(\rho+1)}.$$

Proof. If we define $\tilde{f} : \Gamma \to X$ by

$$\tilde{f}(h) \coloneqq f(h) - P_m f(h) = (I - P_m)f(h),$$

then $\tilde{f} \in Z^1(\pi)$. Let $k \geq 1$ be an integer to be determined later. If we set

$$w \coloneqq -\frac{1}{k}\sum_{j=0}^{k-1} \tilde{f}(v^j),$$

then

$$\|w\|_X \lesssim_\Gamma \frac{1}{k}\sum_{j=0}^{k-1} j^{1/\rho} \lesssim k^{1/\rho}.$$

For every $h \in \Gamma$ we have the following identity:

$$\begin{aligned} -\pi(h)w + \tilde{f}(h) &= \frac{1}{k}\sum_{j=0}^{k-1}\left(\pi(h)\tilde{f}(v^j) + \tilde{f}(h)\right) = \frac{1}{k}\sum_{j=0}^{k-1}\tilde{f}(hv^j) = \frac{1}{k}\sum_{j=0}^{k-1}\tilde{f}(v^j h) \\ &= \frac{1}{k}\sum_{j=0}^{k-1}\left(\pi(v^j)\tilde{f}(h) + \tilde{f}(v^j)\right) = \frac{1}{k}\sum_{j=0}^{k-1}\pi(v^j)\tilde{f}(h) - w. \end{aligned} \tag{2.45}$$

But

$$\frac{1}{k}\sum_{j=0}^{k-1}\pi(v^j)\tilde{f}(h) = \frac{1}{k}\sum_{j=0}^{k-1}\left(\pi(v^j)f(h) - \frac{1}{2^m}\sum_{i=0}^{2^m-1}\pi(v^{j+i})f(h)\right)$$
$$= \frac{1}{2^m}\sum_{i=0}^{2^m-1}\left(\frac{1}{k}\sum_{j=0}^{k-1}\pi(v^j) - \frac{1}{k}\sum_{j=i}^{i+k-1}\pi(v^j)\right)f(h),$$

so that

$$\left\|\frac{1}{k}\sum_{j=0}^{k-1}\pi(v^j)\tilde{f}(h)\right\|_X \leq \frac{d_W(h,e_\Gamma)}{2^m}\sum_{i=0}^{2^m-1}\frac{2i}{k} \leq \frac{2^m}{k}d_W(h,e_\Gamma).$$

Because of (2.45), $\tilde{f}$ is close to a coboundary in the following sense:

$$\left\|\tilde{f}(h) - (\pi(h)w - w)\right\| \leq \frac{2^m}{k}d_W(h,e_\Gamma).$$

Writing $v^{\lfloor n^\rho \rfloor} = a_1 \cdots a_b$ for some $a_i \in S$, where $b = O_\Gamma(n)$, we have

$$\begin{aligned}\left\|\tilde{f}\left(v^{\lfloor n^\rho \rfloor}\right)\right\|_X &= \left\|\sum_{i=1}^{b}\pi(a_1\cdots a_{i-1})\tilde{f}(a_i)\right\|_X \\ &\lesssim_\Gamma \left\|\sum_{i=1}^{b}\pi(a_1\cdots a_{i-1})\left(\pi(a_i)w - w\right)\right\|_X + \frac{n2^m}{k} \\ &= \left\|\pi\left(v^{\lfloor n^\rho \rfloor}\right)w - w\right\|_X + \frac{n2^m}{k} \\ &\lesssim_\Gamma k^{1/\rho} + \frac{n2^m}{k}.\end{aligned}$$

With $k = \lceil n^{\frac{\rho}{\rho+1}}2^{\frac{m\rho}{\rho+1}}\rceil$, we have the stated bound. □

We now prove theorem 2.0.9.

Proof of Theorem 2.0.9. By [63, Theorem 9.1] and by amenability of Γ, if X is q-uniformly convex and $f : \Gamma \to X$ is 1-Lipschitz, then there exists a Banach space $(Y, \|\cdot\|_Y)$ that is also q-uniformly convex with $K_q(Y) = K_q(X)$, an action π of Γ on Y by linear isometric automorphisms, and a 1-cocycle $F : \Gamma \to Y$ such that $\omega_F = \omega_f$, where $F : \Gamma \to Y$ is a 1-cocycle if $F(xy) = \pi(x)F(y) + F(x)$ for all $x, y \in \Gamma$. Thus,

we may assume without loss of generality that $f \in Z^1(\pi)$ for some action π of Γ on X by linear isometric automorphisms.

Let $C_1 < C_2$ be constants depending on Γ such that

$$C_1 n \leq d_W\left(v^{\lfloor n^\rho \rfloor}, e_\Gamma\right) \leq C_2 n, \quad n \in \mathbb{N}.$$

We may assume t is sufficiently large so that if m is the largest integer such that

$$\frac{2C_1}{C_2} 2^{\frac{m}{\rho}} m^{\frac{m(\rho^2+2\rho-2)}{p\rho(\rho-1)}} \leq \sqrt{t}, \tag{2.46}$$

then

$$\frac{p(\rho-1)}{\rho^2+2\rho-2} \leq \frac{1}{2}\log_2 m. \tag{2.47}$$

Given m, let k be the smallest integer such that

$$m^{\frac{\rho+1}{\rho p}+\frac{(\rho^2+2\rho-2)(k+1)}{p\rho(\rho-1)}} \geq t, \tag{2.48}$$

and define

$$\ell \coloneqq \left\lceil \frac{\rho^2+2\rho-2}{p(\rho-1)} \log_2 m \right\rceil. \tag{2.49}$$

Using Lemma 2.4.2, we may find integers $i \in [k+1, k+m]$ and $j \in [0, 2^\ell - 1]$ such that for all $n \in \mathbb{N}$,

$$\left\| \pi\left(v^{-j2^{il}}\right) P_{(i+1)l} f\left(v^{\lfloor n^\rho \rfloor}\right) - P_{il} f\left(v^{\lfloor n^\rho \rfloor}\right) \right\|_X \lesssim_\Gamma \frac{K_q(X) n}{m^{1/p}}. \tag{2.50}$$

Finally, define

$$n \coloneqq \left\lceil \frac{1}{C_1} m^{\frac{\rho+1}{\rho p}} 2^{\frac{i\ell}{\rho}} \right\rceil, \tag{2.51}$$

Expanding, we have

$$\begin{aligned} f\left(v^{\lfloor n^\rho \rfloor}\right) =& \pi\left(v^{-j2^{i\ell}}\right) P_{(i+1)\ell} f\left(v^{\lfloor n^\rho \rfloor}\right) + \left(P_{i\ell} f\left(v^{\lfloor n^\rho \rfloor}\right) - \pi\left(v^{-j2^{i\ell}}\right) P_{(i+1)\ell} f\left(v^{\lfloor n^\rho \rfloor}\right)\right) \\ &+ \left(f\left(v^{\lfloor n^\rho \rfloor}\right) - P_{i\ell} f\left(v^{\lfloor n^\rho \rfloor}\right)\right). \end{aligned}$$

Thus by Lemmas 2.4.3 and 2.4.4 and inequality (2.50), we obtain:

$$\begin{aligned} \omega_f\left(d_W\left(v^{\lfloor n^\rho \rfloor}, e_\Gamma\right)\right) \leq \left\| f\left(v^{\lfloor n^\rho \rfloor}\right)\right\|_X &\lesssim_\Gamma \frac{n^{(\rho^2+\rho-1)/(2\rho-1)}}{2^{(i+1)l(\rho-1)/(2\rho-1)}} + \frac{K_q(X)n}{m^{1/p}} + 2^{il/(\rho+1)} n^{1/(\rho+1)} \\ &\overset{(2.49)\wedge(2.51)}{\lesssim_\Gamma} \frac{K_q(X)n}{m^{1/p}}. \end{aligned} \tag{2.52}$$

We compute

$$d_W\left(v^{\lfloor n^\rho \rfloor}, e_\Gamma\right) \geq C_1 n \overset{(2.51)}{\geq} m^{\frac{\rho+1}{\rho p}} 2^{\frac{i\ell}{\rho}} \geq m^{\frac{\rho+1}{\rho p}} 2^{\frac{(k+1)\ell}{\rho}} \overset{(2.49)}{\geq} m^{\frac{\rho+1}{\rho p} + \frac{(\rho^2+2\rho-2)(k+1)}{p\rho(\rho-1)}} \overset{(2.48)}{\geq} t$$

and

$$\begin{aligned} d_W\left(v^{\lfloor n^\rho \rfloor}, e_\Gamma\right) \leq & C_2 n \overset{(2.51)}{\leq} \frac{2C_2}{C_1} m^{\frac{\rho+1}{\rho p}} 2^{\frac{i\ell}{\rho}} \leq \frac{2C_2}{C_1} m^{\frac{\rho+1}{\rho p}} 2^{\frac{(k+m)\ell}{\rho}} \\ &\overset{(2.49)}{\leq} \frac{2C_1}{C_2} 2^{\frac{k+m}{\rho}} m^{\frac{\rho+1}{\rho p} + \frac{k(\rho^2+2\rho-2)}{p\rho(\rho-1)}} m^{\frac{m(\rho^2+2\rho-2)}{p\rho(\rho-1)}} \overset{(2.48)}{<} \frac{2C_1}{C_2} 2^{\frac{k+m}{\rho}} t m^{\frac{m(\rho^2+2\rho-2)}{p\rho(\rho-1)}} \\ &\overset{(2.48)}{\leq} \frac{2C_1}{C_2} 2^{\frac{p(\rho-1)}{\rho^2+2\rho-2}\log_m t} 2^{\frac{m}{\rho}} t m^{\frac{m(\rho^2+2\rho-2)}{p\rho(\rho-1)}} \overset{(2.46)}{\leq} t^{3/2} 2^{\frac{p(\rho-1)}{\rho^2+2\rho-2}\log_m t} \overset{(2.47)}{\leq} t^2. \end{aligned}$$

Therefore $t \leq d_W\left(v^{\lfloor n^\rho \rfloor}, e_\Gamma\right) \leq t^2$. The definition of m implies

$$m \gtrsim_\Gamma \min\left\{\log n, \frac{p \log n}{\log\log n}\right\},$$

and thus by (2.52), we have

$$\frac{\omega_f\left(d_W\left(v^{\lfloor n^\rho \rfloor}, e_\Gamma\right)\right)}{n} \lesssim K_q(X) \left(\frac{\log\log n}{\log n}\right)^{1/p}.$$

The proof of Theorem 2.0.9 is complete. □

2.5 Derivation of Theorem 2.0.24 from Dorronsoro's Theorem 2.0.22

In this section, we prove Theorem 2.0.24, which proves Theorem 2.0.17 since it is a special case. This section follows Section 7 of [22] closely.

Using an argument similar to the Vitali covering lemma, we may find a collection $\mathcal{B}_r$ of balls B of radius r whose union covers G, and such that the concentric balls $\hat{B}$ of radius $(n+1)r$ have bounded overlap. Then

$$\begin{aligned}
&\int_G \left(\frac{1}{r^\alpha} \left| \sum_{j=0}^{n} (-1)^j \binom{n}{j} f(x(\delta_r(v))^j) \right| \right)^p dx \\
&\leq \sum_{B \in \mathcal{B}_r} \int_B \left(\frac{1}{r^\alpha} \left| \sum_{j=0}^{n} (-1)^j \binom{n}{j} f(x(\delta_r(v))^j) \right| \right)^p dx \\
&\lesssim_p \sum_{j=0}^{n} \binom{n}{j}^p \sum_{B \in \mathcal{B}_r} \int_B \left(\frac{1}{r^\alpha} \left| (f(x(\delta_r(v))^j) - A_{\hat{B}}^{\lfloor \alpha \rfloor} f(x(\delta_r(v))^j) \right| \right)^p dx \\
&\qquad + \int_B \left(\frac{1}{r^\alpha} \left| \sum_{j=0}^{n} (-1)^j \binom{n}{j} A_{\hat{B}}^{\lfloor \alpha \rfloor} f(x(\delta_r(v))^j) \right| \right)^p dx \\
&\lesssim_{n,p} \sum_{B \in \mathcal{B}_r} \int_{\hat{B}} \left(\frac{1}{r^\alpha} \left| f(x) - A_{\hat{B}}^{\lfloor \alpha \rfloor} f(x) \right| \right)^p dx + 0 \\
&= \sum_{B \in \mathcal{B}_r} \left[\frac{1}{r^\alpha} \beta_{f, \lfloor \alpha \rfloor, p}(\hat{B}) \right]^p |\hat{B}| \lesssim_{G,\alpha} \int_G \left[\frac{1}{r^\alpha} \cdot \beta_{f, \lfloor \alpha \rfloor, p}(B_{2(n+1)r}(x)) \right]^p dx.
\end{aligned}$$

We have used in the penultimate inequality the fact that since $A_{\hat{B}}^{\lfloor \alpha \rfloor} f$ is a polynomial of weighted degree at most $\lfloor \alpha \rfloor$, and since $\delta_r(v) \in V_{\lfloor \alpha/n \rfloor + 1} \oplus \cdots \oplus V_s$, $A_{\hat{B}}^{\lfloor \alpha \rfloor} f(x(\delta_r(v))^j)$ is a polynomial in j of degree at most $n-1$, and hence

$$\Delta_{\delta_r(v)}^n f(x) = \sum_{j=0}^{n} (-1)^{n-j} \binom{n}{j} A_{\hat{B}}^{\lfloor \alpha \rfloor} f(x(\delta_r(v))^j) = 0.$$

In the last inequality we used Corollary 2.6.2. Now, the inequality (2.9) follows from Minkowski's integral inequality and Theorem 2.0.22 with $q = p$:

$$\left(\int_0^\infty \left[\int_G \left(\frac{1}{r^\alpha}\left|\sum_{j=0}^{n}(-1)^j\binom{n}{j}f(x(\delta_r(v))^j)\right|\right)^p dx\right]^{2/p}\frac{dr}{r}\right)^{1/2}$$
$$\lesssim_{G,\alpha,n,p}\left(\int_0^\infty \left[\int_G \left[\frac{1}{r^\alpha}\beta_{f,\lfloor\alpha\rfloor,p}(B_{2(n+1)r}(x))\right]^p dx\right]^{2/p}\frac{dr}{r}\right)^{1/2}$$
$$\leq \left(\int_G \left[\int_0^\infty \left[\frac{1}{r^\alpha}\beta_{f,\lfloor\alpha\rfloor,p}(B_{2(n+1)r}(x))\right]^2\frac{dr}{r}\right]^{p/2} dx\right)^{1/p} \lesssim_{G,\alpha,n,p} \|(-\Delta_p)^{\alpha/2}f\|_{L^p(G)}.$$

The proof of Theorem 2.0.24 is complete.

2.6 Proof of Dorronsoro's Theorem 2.0.22 for Carnot groups

We need to prove two directions. In one direction, we assume $f \in S^p_\alpha(G)$ and prove the $\lesssim$ direction of (2.8). In the other direction, we assume $f \in L^p(G)$ with the left-hand side of (2.8) finite, and prove $f \in S^p_\alpha(G)$ along with the $\gtrsim$ direction of (2.8). Our proof will be based on Dorronsoro's original proof of the $G = \mathbb{R}^n$ case in [19] and will borrow modifications inspired by the proof of $G = \mathbb{H}^k$ case in [22] to deal with general Carnot groups.

We will first prove Theorem 2.0.22 for $q = 1$, the $\lesssim$ direction in subsection 2.6.2 and the $\gtrsim$ direction in subsection 2.6.3. Since the left-hand side of (2.8) is minimized when $q = 1$, this will finish the proof of the $\gtrsim$ direction; for the $\lesssim$ direction, we will see in subsection 2.6.4 that the $\lesssim$ inequality for $q = 1$ implies the $\lesssim$ inequality for $1 \leq q < \frac{\min\{p,2\}n_h}{n_h - \min\{p,2\}}$.

For the $\lesssim$ direction, we will first see by an approximation argument that it is enough to look at smooth functions f (subsubsection 2.6.2), for which it turns out

that we may as well approximate at all scales simultaneously by Taylor polynomials (subsubsection 2.6.2). Then, for $0 < \alpha < 1$, it will turn out that the desired inequality has already been proven in [14, Theorem 5] (subsubsection 2.6.2), and for $\alpha > 1$ nonintegral we may use an induction argument on α (subsubsection 2.6.2). We finish off the case of $\alpha \geq 1$ integral by an interpolation argument (subsubsection 2.6.2), with the extra terms arising in this process having been taken care of by a homogenization argument (subsubsection 2.6.2).

For the $\gtrsim$ direction, we assume $f \in L^p(G)$ and the finiteness of a certain singular integral and need to derive $f \in S^p_\alpha(G)$. We first see again that we may approximate by a fixed polynomial for all scales (which we suspect to be the Taylor polynomial in the distributive sense but we can only prove this up to first derivatives) in Proposition 2.6.9. For $0 < \alpha \leq 1$ we again use characterizations of the fractional Laplacian given by [14, Theorem 5] for $0 < \alpha < 1$ and [16, Theorem 1.4] for $\alpha = 1$ (subsubsection 2.6.3). We then prove the case $\alpha > 1$ by induction (subsubsection 2.6.3).

For simplicity, we define the L^1-beta numbers

$$\beta_{f,d}(B_r(x)) \coloneqq \beta_{f,d,1}(B_r(x)).$$

If we define the function

$$\mathfrak{G}_\alpha f(x) \coloneqq \left(\int_0^\infty \left(\frac{\beta_{f,\lfloor \alpha \rfloor}(B_r(x))}{r^\alpha} \right)^2 \frac{dr}{r} \right)^{1/2}, \quad x \in G,$$

then Theorem 2.0.22 states that

$$\|\mathfrak{G}_\alpha f\|_{L^p(G)} \asymp_{G,\alpha,p} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}, \quad f \in S^p_\alpha. \tag{2.53}$$

Before we begin the proof, we briefly remark on coordinate and multi-index notation on G. Recall the coordinate system $x = \exp\left(\sum_{r=1}^s \sum_{i=1}^{k_r} x_{r,i} X_{r,i}\right)$. A multi-index

$\gamma = ((\gamma_{r,i})_{i=1}^{k_r})_{r=1}^{s}$ is a multi-index on $\sum_{r=1}^{s} k_r$ entries, and we denote

$$|\gamma| \coloneqq \sum_{r=1}^{s}\sum_{i=1}^{k_r} r|\gamma_{r,i}| \in \mathbb{N}, \quad \gamma! \coloneqq \prod_{r=1}^{s}\prod_{i=1}^{k_r} \gamma_{r,i}!, \quad x^{\gamma} \coloneqq \prod_{r=1}^{s}\prod_{i=1}^{k_r} x_{r,i}^{\gamma_{r,i}} \in \mathbb{R}.$$

The latter should not be confused with the previous notation

$$x^t = \exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} t x_{r,i} X_{r,i}\right)$$

for $t \in \mathbb{R}$.

We now define the weighted degree of linear combinations of 'polynomial differential operators', that is, linear combinations of differential operators of the form

$$x^{\gamma_1}\left(\frac{\partial}{\partial x}\right)^{\gamma_2},$$

by assigning weight $|\gamma_2| - |\gamma_1|$ to the above term. Then, a polynomial differential operator of homogeneous weighted degree d_1 will act on a polynomial of homogeneous weighted degree d_2 to produce a polynomial of homogeneous weighted degree $d_2 - d_1$. We observe that the left-invariant differential operators $X_{r,i}$ is a polynomial differential operator of homogeneous weighted degree r for $r = 1, \cdots, s$ and $i = 1, \cdots, k_r$.

Recall from subsection 2.0.1 that the polynomials of weighted degree $\leq r$ are left-invariant.

2.6.1 L^1-optimality of the $A_{x,r}^d f$'s

We begin the proof by discussing some basic properties of $A_{x,r}^d f$ and $\beta_{f,d}(B_r(x))$.

It is well-known that $A_{x,r}^d f$ is the optimal $L^2(B_r(x))$-approximation of f in $\mathcal{A}_d$. It turns out that it is also an optimal $L^1(B_r(x))$-approximation up to constants. To see this, we first observe that for all $d \geq 0$, there exists an integral formula for the

coefficients of $A^d_{x,r}f$, given by the Gram–Schmidt process. For example in $d = 1$,

$$A^1_{x,r}f(xy) = \langle f \rangle_{B_r(x)} + \sum_{j=1}^{k} \frac{⨍_{B_r(x)} f(z)(z_{1,j} - x_{1,j})dz}{⨍_{B_r(x)}(z_{1,j} - x_{1,j})^2 dz} y_j, \quad y \in G. \tag{2.54}$$

More generally, for each $d \geq 0$, there exists a family $\{p^d_\gamma\}_{|\gamma| \leq d}$ of polynomials such that

$$A^d_{0,1}f(y) = \sum_{|\gamma| \leq d} \left(\int_{B_1} f p^d_\gamma \right) y^\gamma, \quad y \in G,$$

and by rescaling we have for $x \in G$ and $r > 0$ that

$$A^d_{x,r}f(xy) = \sum_{|\gamma| \leq d} r^{-n_h - |\gamma|} \left(\int_{B_r(x)} f(z) p^d_\gamma(\delta_{1/r}(x^{-1}z))dz \right) y^\gamma.$$

Therefore

$$\|A^d_{x,r}f\|_{L^\infty(B_r(x))} \lesssim_{G,d} ⨍_{B_r(x)} |f(y)|dy. \tag{2.55}$$

In other words, $A^d_{x,r}$ is a linear projection of $L^1(B_r(x))$ onto $\mathcal{A}_d|_{B_r(x)}$ which is bounded in the $L^1 \to L^\infty$ norm.

We now prove the L^1-optimality of the $A^d_{x,r}f$'s.

Lemma 2.6.1. *For $x \in G$, $r > 0$, $f \in L^1_{\mathrm{loc}}(B_r(x))$, and $d \in \mathbb{Z}_{\geq 0}$,*

$$⨍_{B_r(x)} |f(y) - A^d_{x,r}f(y)|dy \lesssim_{G,d} ⨍_{B_r(x)} |f(y) - A(y)|dy, \quad A \in \mathcal{A}_d.$$

Proof. From (2.55),

$$⨍_{B_r(x)} |f - A^d_{x,r}f| \leq ⨍_{B_r(x)} |f - A| + ⨍_{B_r(x)} |A^d_{x,r}(f - A)| \lesssim_{G,d} ⨍_{B_r(x)} |f - A|.$$

□

It follows that β possesses a weak monotonicity property.

Corollary 2.6.2. *Let $x, y \in G$, $0 < r < s < \infty$, $d \in \mathbb{Z}_{\geq 0}$ be such that $B_r(x) \subset B_s(y)$. Then*

$$\beta_{f,d}(B_r(x)) \lesssim_{G,d} \left(\frac{s}{r}\right)^{n_h} \beta_{f,d}(B_s(y)).$$

By Corollary 2.6.2, we have that

$$\beta_{f,d}(x,r) \lesssim_{G,d} \beta_{f,d}(x,s) \lesssim_{G,d} \beta_{f,d}(x,2r), \quad r \leq s \leq 2r.$$

Thus $\mathfrak{G}_\alpha f(x)$ is comparable to the series

$$\left(\sum_{i=-\infty}^{\infty} \left[2^{-i\alpha}\beta_{f,d}(x,2^i)\right]^2\right)^{1/2}.$$

One can also see that, if X is a left-invariant differential operator of weighted order n, then because $A^d_{x,r}f$ is a polynomial of weighted degree d,

$$\|XA^d_{x,r}f\|_{L^\infty(B_r(x))} \lesssim_{G,n,d,X} r^{-n}\|A^d_{x,r}f\|_{L^\infty(B_r(x))} \lesssim_{G,d} r^{-n} \fint_{B_r(x)} |f|. \tag{2.56}$$

2.6.2 Proof of the $\lesssim$ direction, $q = 1$

Approximation by smooth compactly supported functions

It is enough to prove the $\lesssim$ statement of Theorem 2.0.22 for f in the space $\mathcal{D}$ of smooth compactly supported functions on G. Indeed, by [25, Theorem 4.5], $\mathcal{D}$ is dense in S^p_α for all $1 < p < \infty$ and $\alpha \geq 0$. Given $f \in S^p_\alpha$, choose a sequence $\{f_j\}_{j=1}^\infty \subset \mathcal{D}$ that converges to f in S^p_α. Then $f_j \to f$ in $L^p(G)$, so by the contraction property (2.55) we have

$$\beta_{f_j,\lfloor\alpha\rfloor}(B_r(x)) \to \beta_{f,\lfloor\alpha\rfloor}(B_r(x)) \text{ as } j \to \infty, \quad x \in G,\ r > 0.$$

Now, by two applications of Fatou's lemma, it follows that

$$\begin{aligned}&\|\mathfrak{G}_\alpha f\|_{L^p(G)}\\&\leq \liminf_{j\to\infty} \|\mathfrak{G}_\alpha f_j\|_{L^p(G)} \lesssim_{G,\alpha,p} \liminf_{j\to\infty} \|(-\Delta_p)^{\alpha/2} f_j\|_{L^p(G)} = \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}.\end{aligned}$$

This completes the proof of our claim.

Taylor polynomials

Recall the coordinate system on G where each $y \in G$ is uniquely expressed as $y = \exp\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} y_{r,i}X_{r,i}\right)$, $y_{r,i} \in \mathbb{R}$. Let $f \in \mathcal{D}$ and $x \in G$. The well-known Taylor formula tells us that for $n \in \mathbb{Z}_{\geq 0}$,

$$f(xy) = f(x) + \sum_{j=1}^{n} \frac{1}{j!} \left.\frac{d^j}{dt^j}\right|_{t=0} f(xy^t) + \frac{1}{n!}\int_0^1 (1-t)^n \frac{d^{n+1}}{dt^{n+1}} f(xy^t)dt.$$

As y^t is the one-parameter subgroup generated by $\sum_{r=1}^{s}\sum_{i=1}^{k_r} y_{r,i}X_{r,i}$, we have

$$\begin{aligned}f(xy) = f(x) &+ \sum_{j=1}^{n} \frac{1}{j!}\left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} y_{r,i}X_{r,i}\right)^j f(x)\\&+ \frac{1}{n!}\int_0^1 (1-t)^n \left(\sum_{r=1}^{s}\sum_{i=1}^{k_r} y_{r,i}X_{r,i}\right)^{n+1} f(xy^t)dt.\end{aligned}$$

For each multi-index γ, if we denote by $\mathrm{Sym}(X^\gamma)$ the 'coefficient operator' of y^γ in $\frac{\gamma!}{j!}\left(\sum_{r=1}^{s}\sum_{j=1}^{k_r} y_{r,j}X_{r,j}\right)^{|\gamma|}$ where $j = \sum_{r=1}^{s}\sum_{i=1}^{k_r}\gamma_{r,i}$, then the above expression can be rewritten as

$$\begin{aligned}f(xy) = &\sum_{\substack{\gamma \text{ multi-index}\\ \sum_{r=1}^{s}\sum_{i=1}^{k_r}\gamma_{r,i}\leq n}} \frac{1}{\gamma!}\left[\mathrm{Sym}(X^\gamma)f(x)\right] y^\gamma\\&+ (n+1)\int_0^1 \sum_{\substack{\gamma \text{ multi-index}\\ \sum_{r=1}^{s}\sum_{i=1}^{k_r}\gamma_{r,i}=n+1}} \frac{1}{\gamma!}\left[\mathrm{Sym}(X^\gamma)f(xy^t)\right] y^\gamma dt.\end{aligned}$$

By isolating the terms with multi-index γ such that $|\gamma| \leq n$, we see that there exist for each multi-index γ with $|\gamma| \geq n+1$ and $\sum_{r=1}^{s} \sum_{i=1}^{k_r} \gamma_{r,i} \leq n+1$ a function $q_{x,\gamma} f \in \mathcal{D}$ such that

$$f(xy) = \underbrace{\sum_{|\gamma| \leq n} \frac{1}{\gamma!} \left[\mathrm{Sym}(X^\gamma) f(x)\right] y^\gamma}_{=: T_x^n f(xy)} + \underbrace{\sum_{\substack{|\gamma| \geq n+1 \\ \sum_{r=1}^{s} \sum_{i=1}^{k_r} \gamma_{r,i} \leq n+1}} q_{x,\gamma} f(y) y^\gamma}_{= o(d_G(y,e_G)^n) \text{ as } d_G(y,e_G) \to 0}, \quad y \in G,$$

where we denote the first sum to be the *Taylor polynomial of f of degree n at $x \in G$*, while we observe the second term to be $o(d_G(y, e_G)^n)$ as $d_G(y, e_G) \to 0$ because $d_G(y, e_G) \asymp_G \sum_{r=1}^{s} \sum_{i=1}^{k_r} |y_{r,i}|^{1/r}$.

We may observe that if Y is a left-invariant homogeneous differential operator of degree $d \leq n$, then

$$Yf(xy) = \underbrace{\sum_{|\gamma| \leq n} \frac{1}{\gamma!} \left[\mathrm{Sym}(X^\gamma) f(x)\right] Y y^\gamma}_{= T_x^{n-d} Y f(xy)} + \underbrace{\sum_{\substack{|\gamma| \geq n+1 \\ \sum_{r=1}^{s} \sum_{i=1}^{k_r} \gamma_{r,i} \leq n+1}} Y(q_{x,\gamma} f(y) y^\gamma)}_{= o(d_G(y,e_G)^{n-d}) \text{ as } d_G(y,e_G) \to 0}, \quad y \in G,$$

with the second sum being $o(d_G(y, e_G)^{n-d})$ and the first sum being a polynomial of weighted degree at most $n - d$ and hence is the Taylor polynomial of Yf of degree $n - d$ at $x \in G$. This proves that

$$Y T_x^n f = T_x^{n-d} Y f.$$

In particular, we have for the left-invariant horizontal vector fields that with $d \geq 1$,

$$X_i T_x^d f = T_x^{d-1} X_i f, \quad i = 1, \cdots, k.$$

We may define the β-numbers and $\mathfrak{G}$-function using the Taylor polynomial instead.

$$\tilde{\beta}_{f,d}(B_r(x)) \coloneqq \fint_{B_r(x)} |f - T^{\lfloor\alpha\rfloor}_{x,r} f|,$$

and

$$\tilde{\mathfrak{G}}_\alpha f(x) \coloneqq \left(\int_0^\infty \left(\frac{1}{r^\alpha} \tilde{\beta}_{f,\lfloor\alpha\rfloor}(B_r(x)) \right)^2 \frac{dr}{r} \right)^{1/2}, \quad x \in G.$$

Proposition 2.6.3. *If $f \in \mathcal{D}$ and α is nonintegral, then $\mathfrak{G}_\alpha f(x) \asymp_{G,\alpha} \tilde{\mathfrak{G}}_\alpha f(x)$ for all $x \in G$.*

Proof. That $\mathfrak{G}_\alpha f(x) \lesssim_{G,\alpha} \tilde{\mathfrak{G}}_\alpha f(x)$ follows from Lemma 2.6.1. Denote $d = \lfloor\alpha\rfloor$ and

$$A^d_{x,r} f(xy) = \sum_{|\gamma|\le d} f_\gamma(x,r) y^\gamma,$$

then

$$f_\gamma(x,r) = \frac{1}{\gamma!} \operatorname{Sym}(X^\gamma) A^d_{x,r} f(x).$$

Also denote

$$T^d_x f(xy) = \sum_{|\gamma|\le d} f_\gamma(x) y^\gamma,$$

then, in the case $f \in \mathcal{D}$, $f_\gamma(x,r) \to f_\gamma(x)$ as $r \to 0$ by Taylor expansion. Indeed, recalling that $f(xy) = T^d_x f(xy) + O_f(d_G(y,e_G)^{d+1})$,

$$\begin{aligned} \|A^d_{x,r} f - T^d_x f\|_{L^\infty(B_r(x))} &= \|A^d_{x,r}[f - T^d_x f]\|_{L^\infty(B_r(x))} \overset{(2.55)}{\lesssim}_{G,d} \fint_{B_r(x)} |f - T^d_x f| \\ &= O_{G,d,f}(r^{d+1}) \quad \text{as } r \to 0. \end{aligned}$$

By the above expansions for $A^d_{x,r} f$ and $T^d_x f$, we have

$$\| \sum_{|\gamma|\le d} (f_\gamma(x,r) - f_\gamma(x)) r^{|\gamma|} y^\gamma \|_{L^\infty(B_1)} = \| \sum_{|\gamma|\le d} (f_\gamma(x,r) - f_\gamma(x)) y^\gamma \|_{L^\infty(B_r)} = O_{G,d,f}(r^{d+1}).$$

But because the space $\mathcal{A}^d$ of polynomials of weighted degree $\leq d$ normed by $L^\infty(B_1)$ is finite dimensional, all norms are equivalent, and hence

$$\max_{|\gamma|\leq d} r^{|\gamma|}|f_\gamma(x,r) - f_\gamma(x)| = O_{G,d,f}(r^{d+1}).$$

This completes the proof of the convergence $f_\gamma(x,r) \to f_\gamma(x)$ as $r \to 0$.

With this, we provide a different bound on $|f_\gamma(x,r) - f_\gamma(x)|$ using the β numbers. For each $x \in G$,

$$\begin{aligned}
|f_\gamma(x,r) - f_\gamma(x)| &\leq \sum_{i=-\infty}^{0} |f_\gamma(x,2^i r) - f_\gamma(x,2^{i-1}r)| \\
&= \frac{1}{\gamma!}\sum_{i=-\infty}^{0} \left|\mathrm{Sym}(X^\gamma)(A^d_{x,2^i r}f - A^d_{x,2^{i-1}r}f)(x)\right| \\
&\overset{(2.56)}{\lesssim_{G,|\gamma|}} \sum_{i=-\infty}^{0} 2^{-i|\gamma|}r^{-|\gamma|} \fint_{B_{2^{i-1}r}(x)} |f - A^d_{x,2^i r}f| \qquad (2.57)\\
&= \sum_{i=-\infty}^{0} 2^{-i|\gamma|}r^{-|\gamma|}\beta_{f,d}(B_{2^i r}(x)) \\
&\overset{\text{Corollary 2.6.2}}{\lesssim_{G,d}} \int_0^{2r} \beta_{f,d}(B_u(x))u^{-|\gamma|}\frac{du}{u}.
\end{aligned}$$

(With the Cauchy–Schwartz inequality, we may prove that $|f_\gamma(x,r) - f_\gamma(x)| \lesssim_{G,\alpha} r^{\alpha-|\gamma|}\mathfrak{G}_\alpha f(x)$ for $|\gamma| < \alpha$; see Proposition 2.6.9 for the computation. This is weaker than the $O_{G,d,f}(r^{d+1-|\gamma|})$ bound we get using Taylor approximation. This is because the Taylor approximation argument leverages on the fact that f has $(d+1)$ and higher derivatives, whereas the $\mathfrak{G}_\alpha f$ (in principle) only measures the "α-th derivative" of f at x; however, the latter approach is more natural since we have to work with $\mathfrak{G}_\alpha f$.)

Now we bound

$$\begin{aligned}\tilde{\beta}_{f,d}(B_r(x)) &= \fint_{B_r(x)} |f - T_x^d f| \\ &\leq \fint_{B_r(x)} |f - A_{x,r}^d f| + \fint_{B_r(x)} |T_x^d f - A_{x,r}^d f| \\ &\leq \beta_{f,d}(B_r(x)) + C \sum_{|\gamma| \leq d} |f_\gamma(x,r) - f_\gamma(x)| r^{|\gamma|} \\ &\leq \beta_{f,d}(B_r(x)) + C \sum_{i=0}^{d} r^i \int_0^{2r} \beta_{f,d}(B_u(x)) u^{-i} \frac{du}{u},\end{aligned}$$

where C is a constant depending on G and d. By plugging into the definition of $\tilde{\mathfrak{G}}$ and using the triangle inequality,

$$\tilde{\mathfrak{G}}_\alpha f(x) \leq \mathfrak{G}_\alpha f(x) + C \sum_{i=0}^{d} \left(\int_0^\infty \left[r^{i-\alpha} \int_0^{2r} \beta_{f,d}(B_u(x)) u^{-i} \frac{du}{u} \right]^2 \frac{dr}{r} \right)^{1/2}.$$

But by Hardy's inequality (2.23), we have

$$\begin{aligned}\tilde{\mathfrak{G}}_\alpha f(x) &\leq \mathfrak{G}_\alpha f(x) + C \sum_{i=0}^{d} \frac{1}{\alpha - i} \left(\int_0^\infty \left[\frac{\beta_{f,d}(B_r(x))}{r^\alpha} \right]^2 \frac{dr}{r} \right)^{1/2} \\ &\lesssim_{G,\alpha} \mathfrak{G}_\alpha f(x).\end{aligned}$$

This is where we use the fact that α is nonintegral. □

Inhomogeneous estimates to homogeneous estimates

We provide one more reduction in the proof of Theorem 2.0.22. It is enough to prove the weaker estimate

$$\|\mathfrak{G}_\alpha f\|_{L^p(G)} \lesssim_{G,p,\alpha} \|f\|_{p,\alpha} = \|f\|_{L^p(G)} + \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}, \quad f \in \mathcal{D}.$$

(In fact, the inequality Dorronsoro proves in [19, Theorem 2] is

$$\|f\|_{L^p(G)} + \|\mathfrak{G}_\alpha f\|_{L^p(G)} \asymp_{\alpha,p} \|f\|_{p,\alpha}$$

when $G = \mathbb{R}^n$.) This is because a dimensional analysis reveals that the above inequality is inhomogeneous, so by exploiting scaling properties we may remove the inhomogeneous term $\|f\|_{L^p(G)}$.

Lemma 2.6.4. *For $1 < p < \infty$ and $\alpha > 0$, if we have the estimate,*

$$\|\mathfrak{G}_\alpha f\|_{L^p(G)} \leq C_{G,p,\alpha} \|f\|_{p,\alpha}, \quad f \in \mathcal{D}, \tag{2.58}$$

then with the same constant $C_{G,p,\alpha}$,

$$\|\mathfrak{G}_\alpha f\|_{L^p(G)} \leq C_{G,p,\alpha} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}, \quad f \in \mathcal{D}. \tag{2.59}$$

Proof. Given $f \in \mathcal{D}$, set $f_s \coloneqq f \circ \delta_s \in \mathcal{D}$, for $s > 0$, where δ_s is the Carnot group dilation. It is easy to see that

$$\begin{cases} \|f_s\|_{L^p(G)} = s^{-n_h/p} \|f\|_{L^p(G)}, \\ \|\mathfrak{G}_\alpha f_s\|_{L^p(G)} = s^{\alpha - n_h/p} \|\mathfrak{G}_\alpha f\|_{L^p(G)}, \end{cases} \quad s > 0.$$

Also, it is morally clear that we should have

$$\|(-\Delta_p)^{\alpha/2} f_s\|_{L^p(G)} = s^{\alpha - n_h/p} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}, \quad s > 0,$$

and one may indeed verify this using the definition of $(-\Delta_p)^{\alpha/2}$; see the proof of [22, Lemma 2.6] for details.

We now obtain (2.59) for f from (2.58) for f_s and taking $s \to \infty$. □

The case $0 < \alpha < 1$

We now begin proving Theorem 2.0.22.

Fix $1 < p < \infty$ and $0 < \alpha < 1$. Note that we have

$$\tilde{\mathfrak{G}}_\alpha f(x) = \left(\int_0^\infty \left[\frac{1}{r^\alpha} \fint_{B_r} |f(xy) - f(x)| dy \right]^2 \frac{dr}{r} \right)^{1/2}, \quad f \in \mathcal{D},\ x \in G,$$

then by Proposition 2.6.3

$$\mathfrak{G}_\alpha f(x) \asymp_{G,\alpha} \tilde{\mathfrak{G}}_\alpha f(x), \quad x \in G.$$

But by [14, Theorem 5],

$$\|\tilde{\mathfrak{G}}_\alpha f\|_{L^p(G)} \asymp_{G,\alpha,p} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}, \quad f \in \mathcal{D}.$$

Therefore $\lesssim$ of Theorem 2.0.22 follows in this case. Note that we haven't proven both directions of the inequalities yet due to the restriction $f \in \mathcal{D}$.

The case α **nonintegral,** $\alpha > 1$

We will reduce to the case $0 < \alpha < 1$.

For induction, we need the following result. Although we are currently working with $f \in \mathcal{D}$, we state the proposition below for $f \in S^p_\alpha(G)$, because we will need it later again in the proof of the $\gtrsim$ direction of Dorronsoro's theorem.

Proposition 2.6.5 ([25, Theorem 4.10])**.** *Let $1 < p < \infty$ and $\alpha > 1$, and let $f \in L^p(G)$. Then $f \in S^p_\alpha(G)$ if and only if $f \in S^p_{\alpha-1}(G)$ and the distributional derivatives $X_i f \in S^p_{\alpha-1}(G)$ for $i = 1, \cdots, k$, in which case*

$$\sum_{j=1}^k \|(-\Delta_p)^{(\alpha-1)/2} X_j f\|_{L^p(G)} \asymp_{G,\alpha,p} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}, \quad f \in S^p_\alpha(G). \tag{2.60}$$

Remark 2.6.6. *To be precise, Theorem 4.10 of [25] states the inhomogeneous estimate*

$$\|f\|_{L^p(G)} + \|(-\Delta_p)^{(\alpha-1)/2} f\|_{L^p(G)} + \sum_{j=1}^{k} \|X_j f\|_{L^p(G)} + \|(-\Delta_p)^{(\alpha-1)/2} X_j f\|_{L^p(G)}$$

$$\asymp_{G,\alpha,p} \|f\|_{L^p(G)} + \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}, \quad f \in S^p_\alpha(G).$$

But now (2.60) *easily follows using the homogenization argument of subsubsection 2.6.2.*

The following is an analogue of [19, Theorem 5] and [22, Proposition 4.2].

Proposition 2.6.7 (Analogue of [19, Theorem 5] and [22, Proposition 4.2])**.** *Let* $1 < p < \infty$ *and* $\alpha > 0$*. Then,*

$$\|\mathfrak{G}_{\alpha+1} f\|_{L^p(G)} \lesssim_{G,\alpha} \sum_{j=1}^{k} \|\mathfrak{G}_\alpha(X_j f)\|_{L^p(G)}, \qquad f \in \mathcal{D}.$$

Proof. Let $d = \lfloor \alpha \rfloor$. Define a function $\tilde{A}^{d+1}_{x,r} f \in \mathcal{A}_{d+1}$ for $x \in G$ and $r > 0$ as

$$\tilde{A}^{d+1}_{x,r} f(xy) \coloneqq T^{d+1}_x f(xy) - \langle T^{d+1}_x f \rangle_{B_r(x)} + \langle f \rangle_{B_r(x)}.$$

We choose $C \geq 1$ depending on G such that the weak 1-Poincaré inequality due to Jerison [42] holds:

$$⨍_{B_s(y)} |g - \langle g \rangle_{B_s(y)}| \lesssim_G s ⨍_{B_{Cs}(y)} |\nabla g|, \quad y \in G,\ s > 0,\ g \in \mathcal{D}.$$

As $\langle f - \tilde{A}^{d+1}_{x,r} f\rangle_{B_r(x)} = 0$, we have

$$\begin{aligned}\beta_{f,d+1}(B_r(x)) &\overset{\text{Lemma 2.6.1}}{\lesssim_{G,d}} \fint_{B_r(x)} |f - \tilde{A}^{d+1}_{x,r} f| \\ &\overset{\text{weak 1-Poincaré}}{\lesssim_{G}} r \fint_{B_{Cr}(0)} |\nabla f(xy) - \nabla T^{d+1}_x f(xy)|dy \\ &\asymp_{G} r \sum_{j=1}^{k} \fint_{B_{Cr}(0)} |X_j f(xy) - X_j T^{d+1}_x f(xy)| = r \sum_{j=1}^{k} \tilde{\beta}_{X_j f,d}(B_{Cr}(x)).\end{aligned}$$

because $X_j T^{d+1}_x f = T^d_x X_j f$. Now, by definition,

$$\begin{aligned}\mathfrak{G}_{\alpha+1} f(x) &= \left(\int_0^\infty \left[\frac{1}{r^{\alpha+1}} \beta_{f,d+1}(B_r(x))\right]^2 \frac{dr}{r}\right)^{1/2} \\ &\lesssim_{G,d} \sum_{j=1}^{k} \left(\int_0^\infty \left[\frac{1}{r^{\alpha}} \tilde{\beta}_{X_j f,d}(B_{Cr}(x))\right]^2 \frac{dr}{r}\right)^{1/2} \\ &= C^\alpha \sum_{j=1}^{k} \tilde{\mathfrak{G}}_\alpha (X_j f)(x), \quad x \in G.\end{aligned}$$

By Proposition 2.6.3, the proof is complete. □

If we suppose $\lesssim$ of Theorem 2.0.22 for $a < d$ nonintegral, $d \in \mathbb{N}$, then from the above two Propositions, for $d < \alpha < d+1$ and $f \in \mathcal{D}$,

$$\begin{aligned}\|\mathfrak{G}_\alpha f\|_{L^p(G)} &\lesssim_{G,\alpha} \sum_{j=1}^{k} \|\mathfrak{G}_{\alpha-1}(X_j f)\|_{L^p(G)} \lesssim_{G,\alpha,p} \sum_{j=1}^{k} \|(-\Delta_p)^{(\alpha-1)/2} X_j f\|_{L^p(G)} \\ &\asymp_{G,\alpha,p} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}.\end{aligned}$$

This completes the proof of Theorem 2.0.22 for α nonintegral.

Interpolation and the case α integral

The case $\alpha = d \in \mathbb{Z}$ is proven by complex interpolation. This subsubsection follows Section 5 of [22] closely.

For $1 < p < \infty$ and $0 < \alpha_0 < \alpha_1 < \infty$, the pair $(S^p_{\alpha_0}, S^p_{\alpha_1})$ of Banach spaces is an interpolation pair in the sense of Calderón, as they embed continuously in the space $\mathcal{S}'(G)$ of tempered distributions on G. Thus, we may define the complex interpolation space $[S^p_{\alpha_0}, S^p_{\alpha_1}]_\theta$, $\theta \in (0,1)$.

The following lemma asserts that we have a continuous embedding

$$S^p_{(1-\theta)\alpha_0+\theta\alpha_1} \hookrightarrow [S^p_{\alpha_0}, S^p_{\alpha_1}]_\theta.$$

Lemma 2.6.8 ([22, Lemma 5.1])**.** *Let $1 < p < \infty$, $0 < \alpha_0 < \alpha_1 < \infty$, and $\theta \in (0,1)$. Then*

$$\|f\|_{[S^p_{\alpha_0}, S^p_{\alpha_1}]_\theta} \lesssim_{G,\alpha_0,\alpha_1,p} \|f\|_{p,(1-\theta)\alpha_0+\theta\alpha_1}.$$

The proof is exactly the same as in [22, Lemma 5.1].

Next, we define, for $\alpha > 0$, the Banach space H_α of functions $F : B_1 \times (0,\infty) \to \mathbb{R}$ with norm

$$\|F\|_{H_\alpha} := \left(\int_0^\infty \left[\frac{1}{r^\alpha} \fint_{B_1} |F(y,r)| dy \right]^2 \frac{dr}{r} \right)^{1/2} < \infty.$$

Then, for $1 < p < \infty$, we may define the function space $L^p(G, H_\alpha)$. Now, for $0 < \alpha_0 < \alpha_1 < \infty$,

$$(L^p(G, H_{\alpha_0}), L^p(G, H_{\alpha_1}))$$

is a compatible couple, as both embed continuously into $L^1_{\mathrm{loc}}(G \times B_1 \times (0,\infty))$. As in [19, 22], we apply the results of [8, p. 107, 121] so that for $\theta \in (0,1)$,

$$[L^p(G, H_{\alpha_0}), L^p(G, H_{\alpha_1})]_\theta = L^p(G, [H_{\alpha_0}, H_{\alpha_1}]_\theta) = L^p(G, H_{(1-\theta)\alpha_0+\theta\alpha_1}).$$

Now, for $d - 1 < \alpha_0 < d < \alpha_1 < d + 1$ (recall $d = \alpha$) and $\theta \in (0,1)$, consider the linear map $T : S^p_{\alpha_0} + S^p_{\alpha_1} \to L^p(G, H_{\alpha_0}) + L^p(G, H_{\alpha_1})$ given by

$$f \mapsto Tf(x,r,y) = f(x\delta_r(y)) - A^d_{x,r}(x\delta_r(y)).$$

By Lemma 2.6.1 and Theorem 2.0.22 for $\alpha \in (d-1,d) \cup (d,d+1)$, we have

$$\|Tf\|_{L^p(G,H_\alpha)} \lesssim_{G,d} \|\mathfrak{G}_\alpha f\|_{L^p(G)} \lesssim_{G,\alpha_0,\alpha_1,p} \|f\|_{p,\alpha}, \quad f \in S^p_\alpha,\ \alpha = \alpha_0, \alpha_1,$$

i.e., $T : S^p_{\alpha_0} \to L^p(G, H_{\alpha_0})$ and $T : S^p_{\alpha_1} \to L^p(G, H_{\alpha_1})$ are bounded. By complex interpolation,

$$T : [S^p_{\alpha_0}, S^p_{\alpha_1}]_\theta \to [L^p(G, H_{\alpha_0}), L^p(G, H_{\alpha_1})]_\theta = L^p(G, H_{(1-\theta)\alpha_0+\theta\alpha_1})$$

is bounded.

Now set $\alpha_0 = d - 1/2$, $\alpha_1 = d + 1/2$, and $\theta = 1/2$. Then

$$\|\mathfrak{G}_d f\|_{L^p(G)} = \|Tf\|_{L^p(G,H_d)} \lesssim_{G,d} \|f\|_{[S^p_{d-1/2}, S^p_{d+1/2}]_{1/2}} \overset{\text{Lemma 2.6.8}}{\lesssim_{G,d,p}} \|f\|_{p,d}, \quad f \in \mathcal{D}.$$

This completes the proof of Theorem 2.0.22 for $\alpha = d \in \mathbb{Z}$.

2.6.3 Proof of the $\gtrsim$ direction

Now, given a $f \in L^p(G)$ and $\alpha > 0$ with $\mathfrak{G}_\alpha f \in L^p(G)$, we need to prove that $f \in S^p_\alpha(G)$ and that the $\gtrsim$ direction of (2.8) holds.

We first prove the following preparatory result.

Proposition 2.6.9. *Let $f \in L^p(G)$ and $\alpha > 0$ with $\mathfrak{G}_\alpha f \in L^p(G)$. Denote*

$$A^{\lfloor\alpha\rfloor}_{x,r} f(xy) = \sum_{|\gamma| \le \lfloor\alpha\rfloor} f_\gamma(x,r) y^\gamma,$$

Then

1. *For each multi-index γ with $|\gamma| < \alpha$, $|f_\gamma(x,r) - f_\gamma(x,s)| \lesssim_{G,\alpha} r^{\alpha-|\gamma|} \mathfrak{G}_\alpha f(x)$ for $x \in G$ and $0 < s < r$, and $|f_\gamma(x,r)| \lesssim_{G,\alpha} \mathfrak{G}_\alpha f(x) + Mf(x)$ for $x \in G$*

and $0 < r < 1$*, where* M *denotes the Hardy–Littlewood maximal operator. In particular,* $f_\gamma(\cdot, r) \in L^p(G)$ *with* $\|f_\gamma(\cdot, r)\|_{L^p(G)} \lesssim_{G,\alpha} \|\mathfrak{G}_\alpha f\|_{L^p(G)} + \|f\|_{L^p(G)}$.

2. *For each multi-index* γ *with* $|\gamma| < \alpha$*, there exists* $f_\gamma \in L^p(G)$ *such that* $f_\gamma(\cdot, r) \to f_\gamma(\cdot)$ *as* $r \to 0$*, satisfying*

$$|f_\gamma(x, r) - f_\gamma(x)| \lesssim_{G,|\gamma|} \int_0^{2r} \beta_{f,d}(B_u(x)) u^{-|\gamma|} \frac{du}{u} \lesssim_{G,\alpha} r^{\alpha-|\gamma|} \mathfrak{G}_\alpha f(x)$$

for a.e. $x \in G$ *and* $r > 0$.

3. *If* α *is nonintegral, then for a.e.* $x \in G$*,*

$$\left(\int_0^\infty \left(\frac{1}{r^\alpha} \fint_{B_r} |f(xy) - \sum_{|\gamma| \le \lfloor \alpha \rfloor} f_\gamma(x) y^\gamma | dy \right)^2 \frac{dr}{r} \right)^{1/2} \asymp_{G,\alpha} \mathfrak{G}_\alpha f(x).$$

4. *If* α *is integral and* $|\gamma| = \alpha$*, then* $|f_\gamma(x, r)| \lesssim_{G,\alpha} \log(2/r) \mathfrak{G}_\alpha f(x) + Mf(x)$*,* $0 < r < 1$*. Thus,* $f_\gamma(\cdot, r) \in L^p(G)$ *with* $\|f_\gamma(\cdot, r)\|_{L^p(G)} \lesssim_{G,\alpha} \log(2/r) \|\mathfrak{G}_\alpha f\|_{L^p(G)} + \|f\|_{L^p(G)}$.

5. *If* $0 < \alpha < 1$*, then* $f_0(x) = f(x)$ *for a.e.* $x \in G$.

Proof. 1. By repeating the computation (2.57) made in the proof of Proposition 2.6.3, it follows that

$$|f_\gamma(x, r) - f_\gamma(x, s)| \lesssim_{G,|\gamma|} \int_s^{2r} \beta_{f,d}(B_u(x)) u^{-|\gamma|} \frac{du}{u}, \quad x \in G,\ 0 < s < r.$$

In particular, by Cauchy–Schwartz,

$$
\begin{aligned}
&|f_\gamma(x,r) - f_\gamma(x,s)| \lesssim_{G,|\gamma|} \\
&\left(\int_s^{2r} \left[\frac{\beta_{f,d}(B_u(x))}{u^\alpha}\right]^2 \frac{du}{u}\right)^{1/2} \left(\int_s^{2r} (u^{\alpha-|\gamma|})^2 \frac{du}{u}\right)^{1/2} \\
&\lesssim_\alpha r^{\alpha-|\gamma|} \mathfrak{G}_\alpha f(x), \quad x \in G,\ 0 < s < r.
\end{aligned}
$$

Also, for $0 < r < 1$, $|f_\gamma(x,r)| \le |f_\gamma(x,r) - f_\gamma(x,1)| + |f_\gamma(x,1)| \lesssim_{G,\alpha} \mathfrak{G}_\alpha f(x) + Mf(x)$, where the latter inequality used the fact that $f_\gamma(x,1)$ is given as the integral of f multiplied by a polynomial determined by γ and $\lfloor \alpha \rfloor$. In particular, this implies that $f_\gamma(\cdot, r) \in L^p$.

2. By (1) and the fact that $\mathfrak{G}_\alpha f \in L^p(G)$, $\{f_\gamma(\cdot,r)\}_{r>0}$ is a Cauchy sequence in L^p as $r \to 0$. Thus, there exists $f_\gamma(\cdot) \in L^p(G)$ so that $f_\gamma(\cdot, r) \to f_\gamma(\cdot)$ in $L^p(G)$ as $r \to 0$.

3. The $\gtrsim$ follows from Lemma 2.6.2. For $\lesssim$, we bound

$$
\begin{aligned}
&⨍_{B_r} |f(xy) - \sum_{|\gamma| \le \lfloor \alpha \rfloor} f_\gamma(x) y^\gamma| \\
&\le ⨍_{B_r(x)} |f - A^{\lfloor \alpha \rfloor}_{x,r} f| + ⨍_{B_r} |\sum_{|\gamma| \le \lfloor \alpha \rfloor} f_\gamma(x) y^\gamma - A^{\lfloor \alpha \rfloor}_{x,r} f(xy)| dy \\
&\lesssim_{G,\alpha} \beta_{f,\lfloor \alpha \rfloor}(B_r(x)) + \sum_{|\gamma| \le \lfloor \alpha \rfloor} |f_\gamma(x,r) - f_\gamma(x)| r^{|\gamma|} \\
&\lesssim_{G,\alpha} \beta_{f,\lfloor \alpha \rfloor}(B_r(x)) + \sum_{i=0}^{\lfloor \alpha \rfloor} r^i \int_0^{2r} \beta_{f,\lfloor \alpha \rfloor}(B_u(x)) u^{-i} \frac{du}{u}.
\end{aligned}
$$

By plugging into the definition of $\tilde{\mathfrak{G}}$ and using the triangle inequality,

$$\begin{aligned}
&\left(\int_0^\infty \left(\frac{1}{r^\alpha} \fint_{B_r} |f(xy) - \sum_{|\gamma|\le\lfloor\alpha\rfloor} f_\gamma(x) y^\gamma| dy\right)^2 \frac{dr}{r}\right)^{1/2} \\
&\lesssim_{G,\alpha} \mathfrak{G}_\alpha f(x) + \sum_{i=0}^{\lfloor\alpha\rfloor} \left(\int_0^\infty \left[r^{i-\alpha} \int_0^{2r} \beta_{f,\lfloor\alpha\rfloor}(B_u(x)) u^{-i} \frac{du}{u}\right]^2 \frac{dr}{r}\right)^{1/2} \\
&\lesssim_{G,\alpha} \mathfrak{G}_\alpha f(x) + \sum_{i=0}^{\lfloor\alpha\rfloor} \frac{1}{\alpha - i} \left(\int_0^\infty \left[\frac{\beta_{f,\lfloor\alpha\rfloor}(B_r(x))}{r^\alpha}\right]^2 \frac{dr}{r}\right)^{1/2} \\
&\lesssim_\alpha \mathfrak{G}_\alpha f(x).
\end{aligned}$$

where in the second inequality we used Hardy's inequality (2.23) with $\nu = \alpha - i + \frac{1}{2}$ and $p = 2$ along with the fact that α is nonintegral.

4. By Cauchy–Schwartz,

$$\begin{aligned}
|f_\gamma(x,r) - f_\gamma(x,1)| &\lesssim_{G,|\gamma|} \left(\int_r^2 \left[\frac{\beta_{f,d}(B_u(x))}{u^\alpha}\right]^2 \frac{du}{u}\right)^{1/2} \left(\int_r^2 \frac{du}{u}\right)^{1/2} \\
&\lesssim_\alpha \log(2/r) \mathfrak{G}_\alpha f(x).
\end{aligned}$$

5. If $0 < \alpha < 1$, then $f_0(x,r) = \fint_{B_r(x)} f$. Taking $r \to 0$, the claim follows by the Lebesgue differentiation theorem for doubling metric measure spaces [36, (3.4.10)].

□

Remark 2.6.10. *It seems likely that for $|\gamma| < \alpha$ we have $f_\gamma = \frac{1}{\gamma!} \operatorname{Sym}(X^\gamma) f$, with the latter derivative existing in the distributive sense. This is evident when $0 < \alpha < 1$ and $\gamma = 0$, and we will prove this holds when $\alpha > 1$ and $|\gamma| = 1$ in Lemma 2.6.12. We will not prove this statement in full generality since we do not need it in this paper.*

The case $0 < \alpha \leq 1$

The case $0 < \alpha < 1$ is a repetition and strengthening of subsubsection 2.6.2. If we denote

$$\tilde{\mathfrak{G}}_\alpha f(x) = \left(\int_0^\infty \left[\frac{1}{r^\alpha} \fint_{B_r} |f(xy) - f(x)| dy \right]^2 \frac{dr}{r} \right)^{1/2}, \quad x \in G,$$

then by Proposition 2.6.9 (3) and (5)

$$\mathfrak{G}_\alpha f(x) \asymp_{G,\alpha} \tilde{\mathfrak{G}}_\alpha f(x), \quad \text{for a.e. } x \in G.$$

But by [14, Theorem 5],

$$\|\tilde{\mathfrak{G}}_\alpha f\|_{L^p(G)} \asymp_{G,p} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}.$$

For $\alpha = 1$, it is known from [16, Theorem 1.4] that for $0 < \alpha < 2$, if we define

$$S_\alpha f(x) = \left(\int_0^\infty \left[r^{-\alpha} \fint_{B_r(0)} |\Delta_y^{2,\mathrm{sym}} f(x)| dy \right]^2 \frac{dr}{r} \right)^{1/2},$$

where we denote the symmetric difference

$$\Delta_y^{2,\mathrm{sym}} f(x) = f(xy) + f(xy^{-1}) - 2f(x),$$

then

$$\|f\|_{p,\alpha} \asymp_{G,p,\alpha} \|f\|_{L^p(G)} + \|S_\alpha f\|_{L^p(G)}.$$

By the scaling argument of subsubsection 2.6.2,

$$\|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)} \asymp_{G,p,\alpha} \|S_\alpha f\|_{L^p(G)}.$$

Let f be a locally integrable function, $d \in \mathbb{N}$, $x = z_0 \in G$ and $r > 0$. For

any $y \in B_r(z_0)$ and ball $B_{2^{-k}r}(z_k)$ with $y \in B_{2^{-k}r}(z_k) \subset B_r(z_0)$, let $B_{2^{-k}r}(z_k) \subset B_{2^{-k+1}r}(z_{k-1}) \subset \cdots \subset B_{2^{-1}r}(z_1) \subset B_r(x)$ be a sequence of balls. We have

$$\begin{aligned}
&\fint_{B_{2^{-k}r}(z_k)} |f - A^d_{z_0,r} f| \\
&\leq \fint_{B_{2^{-k}r}(z_k)} |f - A^d_{z_k,2^{-k}r} f| + \sum_{i=0}^{k-1} \fint_{B_{2^{-k}r}(z_k)} |A^d_{z_i,2^{-i}r} f - A^d_{z_{i+1},2^{-i-1}r} f| \\
&\leq \fint_{B_{2^{-k}r}(z_k)} |f - A^d_{z_k,2^{-k}r} f| + \sum_{i=0}^{k-1} \left\| A^d_{z_{i+1},2^{-i-1}r}(f - A^d_{z_i,2^{-i}r} f) \right\|_{L^\infty(B_{2^{-i-1}r}(z_{i+1}))} \\
&\overset{(2.55)}{\lesssim_{G,d}} \fint_{B_{2^{-k}r}(z_k)} |f - A^d_{z_k,2^{-k}r} f| + \sum_{i=0}^{k-1} \fint_{B_{2^{-i-1}r}(z_{i+1})} |f - A^d_{z_i,2^{-i}r} f| \\
&\lesssim_G \sum_{i=0}^{k} \fint_{B_{2^{-i}r}(z_i)} |f - A^d_{z_i,2^{-i}r} f| \overset{\text{Corollary 2.6.2}}{\lesssim_{G,d}} \int_0^{2r} \beta_{f,d}(B_s(y)) \frac{ds}{s}
\end{aligned}$$

so by Lebesgue's differentiation theorem for doubling metric measure spaces, for a.e. $y \in B_r(x)$,

$$|f(y) - A^d_{x,r} f(y)| = \lim_{B_{2^{-k}r}(z_k) \to y} \fint_{B_{2^{-k}r}(z_k)} |f - A^d_{z_0,r} f| \lesssim_{G,d} \int_0^{2r} \beta_{f,d}(B_s(y)) \frac{ds}{s}.$$

As the operator $\Delta_h^{2,\mathrm{sym}}$ annihilates polynomials of weighted degree 1, we obtain for a.e. $x \in G$

$$\begin{aligned}
&\fint_{B_r} |\Delta_h^{2,\mathrm{sym}} f(x)| dh = \fint_{B_r} |\Delta_h^{2,\mathrm{sym}} (f - A^1_{x,2r} f)(x)| dh \\
&= \fint_{B_r} |(f - A^1_{x,2r} f)(xh) + (f - A^1_{x,2r} f)(xh^{-1}) - 2(f - A^1_{x,2r} f)(x)| dh \\
&\lesssim_{G,d} \fint_{B_r} \left(\int_0^{2r} \beta_{f,d}(B_s(xh)) \frac{ds}{s} + \int_0^{2r} \beta_{f,d}(B_s(xh^{-1})) \frac{ds}{s} + 2 \int_0^{2r} \beta_{f,d}(B_s(x)) \frac{ds}{s} \right) dh \\
&\lesssim \int_0^{2r} M \beta_{f,1}(B_s(\cdot))(x) \frac{ds}{s}.
\end{aligned}$$

By Hardy's inequality (2.23) with $\nu = \frac{3}{2}$ and $p = 2$, we obtain

$$\begin{aligned} S_1 f(x) &= \left(\int_0^\infty \left[r^{-1} \fint_{B_r} |\Delta_y^{2,\mathrm{sym}} f(x)| dy \right]^2 \frac{dr}{r} \right)^{1/2} \\ &\lesssim_{G,d} \left(\int_0^\infty \left[M\beta_{f,1}(B_r(\cdot))(x) \right]^2 \frac{dr}{r} \right)^{1/2}. \end{aligned}$$

Now we argue that

$$\begin{aligned} &\left(\int_G \left(\int_0^\infty \left[M\beta_{f,1}(B_r(\cdot))(x) \right]^2 \frac{dr}{r} \right)^{p/2} dx \right)^{1/p} \\ &\lesssim_G \max\{p, \frac{1}{p-1}\} \left(\int_G \left(\int_0^\infty \left[\beta_{f,1}(B_r(\cdot))(x) \right]^2 \frac{dr}{r} \right)^{p/2} dx \right)^{1/p} \end{aligned}$$

This is an instance of the Fefferman–Stein vector-valued maximal function inequality

$$\begin{aligned} &\left(\int_G \left[\sum_{j \in \mathbb{Z}} M(g_j)^u(x) \right]^{v/u} dx \right)^{1/v} \\ &\lesssim_G \max\{1, \frac{1}{u-1}\} \max\{v, \frac{1}{v-1}\} \left(\int_G \left[\sum_{j \in \mathbb{Z}} g_j(x)^u \right]^{v/u} dx \right)^{1/v} \end{aligned} \tag{2.61}$$

on n_h-regular metric measure spaces for measurable functions $(g_j : G \to \mathbb{R})_{j \in \mathbb{Z}}$ and $u, v \in (1, \infty)$, due to [29, Theorem 1.2]. Indeed, we apply (2.61) to the functions

$$g_j(x) = \beta_{f,1}(B_{2^{-j}}(x))$$

and exponents $u = 2 > 1$ and $v = p > 1$, noting by Corollary 2.6.2 that

$$M(\beta_{f,1}(B_r(\cdot))) \lesssim_G M(g_j), \quad r \in [2^{-j-1}, 2^{-j}],$$

and

$$g_j \lesssim_G \beta_{f,1}(B_r(\cdot)), \quad r \in [2^{-j}, 2^{-j+1}],$$

Thus, we have $\|S_1 f\|_{L^p(G)} \lesssim_{G,p} \|\mathfrak{G}_1 f\|_{L^p(G)}$, $1 < p < \infty$, as desired.

The case $\alpha > 1$

We use induction on α.

The following lemma is an analogue of [19, Lemma 1], and the proof proceeds in the same manner.

Lemma 2.6.11 (Analogue of [19, Lemma 1])**.** *Let $f \in L^p(G)$ such that $\mathfrak{G}_\alpha f \in L^p(G)$, and γ be a multi-index with $|\gamma| < \alpha$. With f_γ as in Proposition 2.6.9, we have*

$$\|\mathfrak{G}_{\alpha-|\gamma|} f_\gamma\|_{L^p(G)} \lesssim_{G,\alpha} \|\mathfrak{G}_\alpha f\|_{L^p(G)}.$$

Proof. Fix $x \in G$ and $r > 0$. For $y \in B_r(x)$, by Proposition 2.6.9 and (2.56),

$$\begin{aligned} |f_\gamma(y) - \frac{1}{\gamma!} \operatorname{Sym}(X^\gamma) A^d_{x,2r} f(y)| &\le |f_\gamma(y) - f_\gamma(y,r)| + |\frac{1}{\gamma!} \operatorname{Sym}(X^\gamma)(A^d_{y,r} - A^d_{x,2r} f)(y)| \\ &\lesssim_{G,\alpha} \int_0^r \beta_{f,d}(B_s(y)) s^{-|\gamma|-1} ds + r^{-|\gamma|} \beta_{f,d}(B_{2r}(x)). \end{aligned}$$

Thus, by Lemma 2.6.1,

$$\begin{aligned} \beta_{f_\gamma, d-|\gamma|}(B_r(x)) &\lesssim_{G,\alpha} \fint_{B_r(x)} \left(\int_0^r \beta_{f,d}(B_s(y)) s^{-|\gamma|-1} ds dy + r^{-|\gamma|} \beta_{f,d}(B_{2r}(x)) \right) dy \\ &\le \int_0^r M(\beta_{f,d}(B_s(\cdot)))(x) s^{-|\gamma|-1} ds + r^{-|\gamma|} \beta_{f,d}(B_{2r}(x)). \end{aligned}$$

By Hardy's inequality with $\nu = \alpha - |\gamma| + \frac{1}{2}$ and $p = 2$,

$$G_{\alpha-|\gamma|} f_\gamma(x) \lesssim_{G,\alpha} \mathfrak{G}_\alpha f(x) + \left(\int_0^\infty \left[r^{-\alpha} M \beta_{f,d}(B_r(\cdot))(x) \right]^2 \frac{dr}{r} \right)^{1/2},$$

and the lemma follows by the aforementioned Fefferman–Stein inequality (2.61) on

vector-valued maximal operators with $u = v = 2 > 1$ and

$$g_j(x) = 2^{\alpha j}\beta_{f,d}(B_{2^{-j+1}}(x)).$$

□

The following lemma is an analogue of [19, Lemma 2], and the proof proceeds in the same manner.

Lemma 2.6.12 (Analogue of [19, Lemma 2]). *Let $f \in L^p(G)$ with $\mathfrak{G}_\alpha f \in L^p(G)$, $\alpha > 1$. Then the weak partials $X_i f$ exist and coincide with $f_{(1,i)}$ a.e., where $(1,i)$ stands for the multi-index γ with $\gamma_{r,j} = \delta_{(r,j),(1,i)}$.*

Proof. Denote $e_{r,i} = \exp(x_{r,i})$ for $r = 1, \cdots, s$ and $i = 1, \cdots, k_r$. It is enough to show that $(f(xe_{1,i}^r) - f(x))/r$ tends to $f_{(1,i)}$ in L^p as $r \to 0$. First,

$$\begin{aligned}
&\left|\frac{f(xe_{1,i}^r) - f(x)}{r} - f_{(1,i)}(x)\right| \\
&\leq \frac{|f(xe_{1,i}^r) - f_0(xe_{1,i}^r, 2r)|}{r} + \frac{|f(x) - f_0(x,2r)|}{r} \\
&\qquad + \left|\frac{f_0(xe_{1,i}^r, 2r) - f_0(x,2r)}{r} - f_{(1,i)}(x)\right| \\
&\overset{\text{Proposition 2.6.9(2)}}{\lesssim_{G,\alpha}} r^{\alpha-1}\left(\mathfrak{G}_\alpha f(xe_{1,i}^r) + \mathfrak{G}_\alpha f(x)\right) + \left|\frac{f_0(xe_{1,i}^r, 2r) - f_0(x,2r)}{r} - f_{(1,i)}(x)\right|.
\end{aligned}$$

Next, since $f_0(y,r) = A_{y,r}^{\lfloor\alpha\rfloor} f(y)$,

$$\begin{aligned}
f_0(xe_{1,i}^r, 2r) - f_0(x,2r) &= A_{xe_{1,i}^r,2r}^{\lfloor\alpha\rfloor} f(xe_{1,i}^r) - A_{x,2r}^{\lfloor\alpha\rfloor} f(xe_{1,i}^r) + A_{x,2r}^{\lfloor\alpha\rfloor} f(xe_{1,i}^r) - A_{x,2r}^{\lfloor\alpha\rfloor} f(x) \\
&= A_{xe_{1,i}^r,2r}^{\lfloor\alpha\rfloor} f(xe_{1,i}^r) - A_{x,2r}^{\lfloor\alpha\rfloor} f(xe_{1,i}^r) + \sum_{j=1}^{\lfloor\alpha\rfloor} f_{j(1,i)}(x,2r) r^j/j!.
\end{aligned}$$

Thus

$$\begin{aligned}
&\left|\frac{f_0(xe_{1,i}^r,2r)-f_0(x,2r)}{r}-f_{(1,i)}(x)\right| \\
&\le \left|\frac{f_0(xe_{1,i}^r,2r)-f_0(x,2r)}{r}-f_{(1,i)}(x,2r)\right|+|f_{(1,i)}(x,2r)-f_{(1,i)}(x)| \\
&\overset{\text{Proposition 2.6.9(2)}}{\lesssim_{G,\alpha}} \frac{|A^{\lfloor\alpha\rfloor}_{xe_{1,i}^r,2r}f(xe_{1,i}^r)-A^{\lfloor\alpha\rfloor}_{x,2r}f(xe_{1,i}^r)|}{r}+\sum_{j=2}^{\lfloor\alpha\rfloor}|f_{j(1,i)}(x,2r)|r^{j-1}+r^{\alpha-1}\mathfrak{G}_\alpha f(x) \\
&\lesssim_{G,\alpha} r^{-1}\fint_{B_r(xe_{1,i}^r)}|A^{\lfloor\alpha\rfloor}_{xe_{1,i}^r,2r}f(y)-A^{\lfloor\alpha\rfloor}_{x,2r}f(y)|dy+\sum_{j=2}^{\lfloor\alpha\rfloor}|f_{j(1,i)}(x,2r)|r^{j-1}+r^{\alpha-1}\mathfrak{G}_\alpha f(x) \\
&\le r^{-1}\fint_{B_r(xe_{1,i}^r)}|f-A^{\lfloor\alpha\rfloor}_{xe_{1,i}^r,2r}f(y)|dy+r^{-1}\fint_{B_r(xe_{1,i}^r)}|f-A^{\lfloor\alpha\rfloor}_{x,2r}f(y)|dy \\
&\quad+\sum_{j=2}^{\lfloor\alpha\rfloor}|f_{j(1,i)}(x,2r)|r^{j-1}+r^{\alpha-1}\mathfrak{G}_\alpha f(x) \\
&= r^{-1}\beta_{f,\lfloor\alpha\rfloor}(B_{2r}(xe_{1,i}^r))+r^{-1}\beta_{f,\lfloor\alpha\rfloor}(B_{2r}(x))+\sum_{j=2}^{\lfloor\alpha\rfloor}|f_{j(1,i)}(x,2r)|r^{j-1}+r^{\alpha-1}\mathfrak{G}_\alpha f(x),
\end{aligned}$$

where in the penultimate inequality we used the fact that the $L^1(B_1)$ and $L^\infty(B_1)$ norms on the space of polynomials of weighted degree $\le \lfloor\alpha\rfloor$ are equivalent. Taking L^p norms,

$$\begin{aligned}
\left\|\frac{f(\cdot e_{1,i}^r)-f(\cdot)}{r}-f_{(1,i)}(\cdot)\right\|_{L^p(G)} &\lesssim_{G,\alpha} r^{-1}\|\beta_{f,\lfloor\alpha\rfloor}(B_{2r}(\cdot))\|_{L^p(G)} \\
&\quad+\sum_{j=2}^{\lfloor\alpha\rfloor}r^{j-1}\|f_{j(1,i)}(\cdot,2r)\|_{L^p(G)}+r^{\alpha-1}\|\mathfrak{G}_\alpha f\|_{L^p(G)}.
\end{aligned}$$

By Corollary 2.6.2, $\|\beta_{f,\lfloor\alpha\rfloor}(B_{2r}(\cdot))\|_{L^p(G)} \lesssim_{G,\alpha} r^\alpha\|\mathfrak{G}_\alpha f\|_{L^p(G)}$. By Proposition 2.6.9, if $j<\lfloor\alpha\rfloor$, then

$$\|f_{j(1,i)}(\cdot,2r)\|_{L^p(G)} \lesssim_{G,\alpha} \|\mathfrak{G}_\alpha f\|_{L^p(G)}+\|f\|_{L^p(G)},$$

and if $j=\lfloor\alpha\rfloor=\alpha$, then

$$\|f_{j(1,i)}(\cdot,2r)\|_{L^p(G)} \lesssim_{G,\alpha} (\log r^{-1})(\|\mathfrak{G}_\alpha f\|_{L^p(G)}+\|f\|_{L^p(G)}).$$

Thus $\left\| \frac{f(\cdot e_{1,i}^r)-f(\cdot)}{r} - f_{(1,i)}(\cdot) \right\|_{L^p(G)} \to 0$ as $r \to 0$. $\square$

The following lemma is an analogue of [19, Lemma 3], and the proof proceeds in the same manner.

Lemma 2.6.13 (Analogue of [19, Lemma 3])**.** *Let $f \in L^p(G)$ with $\mathfrak{G}_\alpha f \in L^p(G)$, $\alpha > 1$. Then $\mathfrak{G}_{\alpha-1} f \in L^p(G)$ with*

$$\|\mathfrak{G}_{\alpha-1} f\|_{L^p(G)} \lesssim_{G,\alpha} \|\mathfrak{G}_\alpha f\|_{L^p(G)} + \|f\|_{L^p(G)}.$$

Proof. Write

$$A_{x,r}^{\lfloor\alpha\rfloor} f(xy) = \underbrace{\sum_{|\gamma|<\lfloor\alpha\rfloor} f_\gamma(x,r) y^\gamma}_{=: A_{x,r}^{<\lfloor\alpha\rfloor} f(xy)} + \underbrace{\sum_{|\gamma|=\lfloor\alpha\rfloor} f_\gamma(x,r) y^\gamma}_{=: A_{x,r}^{=\lfloor\alpha\rfloor} f(xy)}.$$

Then by Lemma 2.6.1,

$$\beta_{f,\lfloor\alpha\rfloor-1}(B_r(x)) \lesssim_{G,\alpha} \fint_{B_r(x)} |f - A_{x,r}^{<\lfloor\alpha\rfloor} f| \le \beta_{f,\lfloor\alpha\rfloor}(B_r(x)) + \|A_{x,r}^{=\lfloor\alpha\rfloor} f\|_{L^\infty(B_r(x))}. \quad (2.62)$$

By (2.56), for $i \in \mathbb{Z}_{\ge 0}$ and $|\gamma| = \lfloor\alpha\rfloor$,

$$|f_\gamma(x, 2^i r)| \lesssim_{G,\alpha} (2^i r)^{-\lfloor\alpha\rfloor} \fint_{B_{2^i r}(x)} |f| \lesssim_G (2^i r)^{-\lfloor\alpha\rfloor - n_h/p} \|f\|_{L^p(G)} \to 0 \quad \text{as } i \to \infty$$

and thus

$$\begin{aligned}
|f_\gamma(x,r)| &\le \sum_{i=0}^{\infty} |f_\gamma(x,2^i r) - f_\gamma(x,2^{i+1}r)| \\
&\overset{(2.56)}{\lesssim_{G,\alpha}} \sum_{i=0}^{\infty} (2^i r)^{-\lfloor\alpha\rfloor} \|A^{\lfloor\alpha\rfloor}_{x,2^i r} f - A^{\lfloor\alpha\rfloor}_{x,2^{i+1}r} f\|_{L^\infty(B_{2^i r}(x))} \\
&\overset{(2.55)}{\lesssim_{G,\alpha}} \sum_{i=0}^{\infty} (2^i r)^{-\lfloor\alpha\rfloor} \fint_{B_{2^i r}(x)} |f - A^{\lfloor\alpha\rfloor}_{x,2^{i+1}r} f| \\
&\lesssim_{G,\alpha} \sum_{i=0}^{\infty} (2^i r)^{-\lfloor\alpha\rfloor} \beta_{f,\lfloor\alpha\rfloor}(B_{2^{i+1}r}(x)). \\
&\overset{\text{Corollary 2.6.2}}{\lesssim_{G,\alpha}} \int_r^\infty \beta_{f,\lfloor\alpha\rfloor}(B_\rho(x)) \rho^{-\lfloor\alpha\rfloor-1} d\rho.
\end{aligned}$$

Plugging these results into (2.62), we have

$$\beta_{f,\lfloor\alpha\rfloor-1}(B_r(x)) \lesssim_{G,\alpha} \beta_{f,\lfloor\alpha\rfloor}(B_r(x)) + r^{\lfloor\alpha\rfloor} \int_r^\infty \beta_{f,\lfloor\alpha\rfloor}(B_\rho(x)) \rho^{-\lfloor\alpha\rfloor-1} d\rho,$$

and

$$\begin{aligned}
\mathfrak{G}_{\alpha-1} f(x) &= \left(\int_0^\infty \left[\frac{\beta_{f,\lfloor\alpha\rfloor-1}(B_r(x))}{r^{\alpha-1}} \right]^2 \frac{dr}{r} \right)^{1/2} \\
&\lesssim_{G,\alpha} \left(\int_0^\infty \left[\frac{\beta_{f,\lfloor\alpha\rfloor}(B_r(x))}{r^{\alpha-1}} \right]^2 \frac{dr}{r} \right)^{1/2} \\
&\quad + \left(\int_0^\infty \left[r^{\lfloor\alpha\rfloor-\alpha+1/2} \int_r^\infty \beta_{f,\lfloor\alpha\rfloor}(B_\rho(x)) \rho^{-\lfloor\alpha\rfloor-1} d\rho \right]^2 dr \right)^{1/2} \\
&\lesssim \left(\int_0^\infty \left[\frac{\beta_{f,\lfloor\alpha\rfloor}(B_r(x))}{r^{\alpha-1}} \right]^2 \frac{dr}{r} \right)^{1/2} \\
&\le \left(\int_0^1 \left[\frac{\beta_{f,\lfloor\alpha\rfloor}(B_r(x))}{r^{\alpha-1}} \right]^2 \frac{dr}{r} \right)^{1/2} + \left(\int_1^\infty \left[\frac{\beta_{f,\lfloor\alpha\rfloor}(B_r(x))}{r^{\alpha-1}} \right]^2 \frac{dr}{r} \right)^{1/2} \\
&\overset{(2.55)}{\lesssim_{G,\alpha}} \left(\int_0^1 \left[\frac{\beta_{f,\lfloor\alpha\rfloor}(B_r(x))}{r^{\alpha}} \right]^2 \frac{dr}{r} \right)^{1/2} + \left(\int_1^\infty \left[\frac{Mf(x)}{r^{\alpha-1}} \right]^2 \frac{dr}{r} \right)^{1/2} \\
&\lesssim_\alpha \mathfrak{G}_\alpha f(x) + Mf(x),
\end{aligned}$$

where the second inequality uses the second form of Hardy's inequality (2.25) with $p = 2$ and $\nu = \lfloor \alpha \rfloor - \alpha + 1/2$. This completes the proof. $\square$

To complete the induction, suppose the $\gtrsim$ direction of Theorem 2.8 holds for $\alpha - 1$. If $f \in L^p(G)$ with $\mathfrak{G}_\alpha f \in L^p(G)$, by Lemma 2.6.13, $\mathfrak{G}_{\alpha-1} f \in L^p(G)$, so that $f \in S^p_{\alpha-1}$ by induction. On the other hand, by Proposition 2.6.9 (2) and Lemmas 2.6.11 and 2.6.12, for $i = 1, \cdots, k$, $X_i f \in L^p(G)$ with $\mathfrak{G}_{\alpha-1}(X_i f) \in L^p(G)$, so again by induction $X_i f \in S^p_{\alpha-1}(G)$. Now, by Proposition 2.6.5, $f \in S^p_\alpha(G)$. Tracking the norms involved in this argument, we have that

$$\|\mathfrak{G}_\alpha f\|_{L^p(G)} \overset{\text{Lemma 2.6.11}}{\gtrsim_{G,\alpha}} \sum_{i=1}^k \|\mathfrak{G}_{\alpha-1} f_{(1,i)}\|_{L^p(G)} \overset{\text{Lemma 2.6.12}}{=} \sum_{i=1}^k \|\mathfrak{G}_{\alpha-1}(X_i f)\|_{L^p(G)}$$
$$\overset{\text{induction}}{\gtrsim_{G.\alpha,p}} \sum_{i=1}^k \|(-\Delta_p)^{(\alpha-1)/2} X_i f\|_{L^p(G)} \overset{\text{Proposition 2.6.5}}{\asymp_{G,\alpha,p}} \|(-\Delta_p)^{\alpha/2} f\|_{L^p(G)}.$$

This completes the $\gtrsim$ direction of Theorem 2.8.

2.6.4 L^q β-numbers, $q > 1$

The extension of Theorem 2.0.22 to β numbers defined in the L^q sense are discussed in [19, Section 5] and [22, Section 6]. We will follow their presentation.

By the discussion so far, it is enough to show that the left-hand side of (2.8) is maximized up to constants when $q = 1$.

Denote $d = \lfloor \alpha \rfloor$. We compute that for all $x \in G$, $r > 0$, $y \in B_r(x)$ and $n \in \mathbb{N}$ so

that $B_{2^{-n_r}}(y) \subset B_r(x)$, we have

$$
\begin{aligned}
&\fint_{B_{2^{-n_r}}(y)} |f(z) - A^d_{x,r} f(z)|dz \\
&\le \fint_{B_{2^{-n_r}}(y)} |f(z) - A^d_{y,2^{-n_r}} f(z)|dz + \sum_{k=1}^{n+1} \fint_{B_{2^{-n_r}}(y)} |A^d_{y,2^{k-1-n_r}} f(z) - A^d_{y,2^{k-n_r}} f(z)|dz \\
&\quad + \fint_{B_{2^{-n_r}}(y)} |A^d_{y,2r} f(z) - A^d_{x,r} f(z)|dz \\
&= \beta_{f,d}(B_{2^{-n_r}}(y)) + \sum_{k=1}^{n+1} \fint_{B_{2^{-n_r}}(y)} |A^d_{y,2^{k-1-n_r}}[f - A^d_{y,2^{k-n_r}} f]| \\
&\qquad + \fint_{B_{2^{-n_r}}(y)} |A^d_{x,r}[A^d_{y,2r} f - f]| \\
&\le \beta_{f,d}(B_{2^{-n_r}}(y)) + \sum_{k=1}^{n+1} \left\| A^d_{y,2^{k-1-n_r}}[f - A^d_{y,2^{k-n_r}} f] \right\|_{L^\infty(B_{2^{k-1-n_r}}(y))} \\
&\qquad + \left\| A^d_{x,r}[A^d_{y,2r} f - f] \right\|_{L^\infty(B_r(x))} \\
&\overset{(2.55)}{\lesssim_{G,d}} \beta_{f,d}(B_{2^{-n_r}}(y)) + \sum_{k=1}^{n+1} \fint_{B_{2^{k-1-n_r}}(y)} |f - A^d_{y,2^{k-n_r}} f| + \fint_{B_r(x)} |A^d_{y,2r} f - f| \\
&\lesssim_{G,d} \beta_{f,d}(B_{2^{-n_r}}(y)) + \sum_{k=1}^{n+1} \beta_{f,d}(B_{2^{k-n_r}}(y)) + \beta_{f,d}(B_{2r}(y)) \\
&\overset{\text{Corollary 2.6.2}}{\lesssim_{G,d}} \int_{2^{-n_r}}^{2^{-n+1}r} \beta_{f,d}(B_s(y)) \frac{ds}{s} + \sum_{k=1}^{n+1} \int_{2^{k-n_r}}^{2^{k+1-n_r}} \beta_{f,d}(B_s(y)) \frac{ds}{s} + \int_{2r}^{4r} \beta_{f,d}(B_s(y)) \frac{ds}{s} \\
&\lesssim \int_{2^{-n_r}}^{4r} \beta_{f,d}(B_s(y)) \frac{ds}{s}.
\end{aligned}
$$

Thus by Lebesgue's differentiation theorem for doubling metric measure spaces and Corollary 2.6.2, we infer that for $x \in G$ and $r > 0$, it holds true for a.e. $y \in B_r(x)$ that

$$
|f(y) - A^d_{x,r} f(y)| \lesssim_{G,d} \int_0^{4r} \beta_{f,d}(B_s(y)) \frac{ds}{s} \lesssim_{G,d} \int_0^{4r} \fint_{B_s(y)} \beta_{f,d}(B_{2s}(z))dz \frac{ds}{s}. \quad (2.63)
$$

Now fix

$$
p > 1 \quad \text{and} \quad 1 \le q < \frac{\min\{p, 2\} n_h}{n_h - \min\{p, 2\}}.
$$

Then, choose some $1 < w < \min\{p, 2\}$ and $0 < \eta < \min\{n_h/w, \alpha\}$ such that

$$q = \frac{wn_h}{n_h - \eta w}. \tag{2.64}$$

Given this choice of q, w, and η, [35, Theorem 4.1] tells us that in n_h-regular metric measure spaces, the following fractional Hardy–Littlewood maximal function is a bounded operator $L^w(G) \to L^q(G)$:

$$M_\eta g(y) := \sup_{s>0} \left\{ s^\eta \fint_{B_s(y)} |g(z)|dz \right\}$$

It follows that

$$\begin{aligned}
r^{n_h/q}\beta_{f,d,q}(B_r(x)) &\overset{(2.63)}{\lesssim_{G,\alpha}} \left(\int_{B_r(x)} \left[\int_0^{4r} \fint_{B_s(y)} \beta_{f,d}(B_{2s}(z))dz \frac{ds}{s} \right]^q dy \right)^{1/q} \\
&\leq \int_0^{4r} \left(\int_{B_r(x)} \left[\fint_{B_s(y)} \beta_{f,d}(B_{2s}(z))dz \right]^q dy \right)^{1/q} \frac{ds}{s} \\
&\leq \int_0^{4r} s^{-\eta} \left(\int_{B_r(x)} M_\eta(\beta_{f,d}(B_{2s}(\cdot))\chi_{B_{5r}(x)}(\cdot))(y)^q dy \right)^{1/q} \frac{ds}{s} \\
&\lesssim_{p,q} \int_0^{4r} s^{-\eta} \left(\int_{B_{5r}(x)} \beta_{f,d}(B_{2s}(z))^w dz \right)^{1/w} \frac{ds}{s} \\
&\lesssim_G \int_0^{4r} s^{-\eta} r^{n_h/w} \left[M(\beta_{f,d}(B_{2s}(\cdot))^w)(x) \right]^{1/w} \frac{ds}{s},
\end{aligned}$$

where in the second inequality we used Minkowski's inequality and in the fourth inequality we used the boundedness of $M_\eta : L^w(G) \to L^q(G)$. By (2.64), we have

$$\beta_{f,d,q}(B_r(x)) \lesssim_{G,\alpha,p,q} r^\eta \int_0^{4r} s^{-\eta-1} \left[M(\beta_{f,d}(B_{2s}(\cdot))^w)(x) \right]^{1/w} ds.$$

Next, noting that $\alpha > \eta$ and using Hardy's inequality (2.23) with $\nu = \alpha - \eta + 1/2$,

we obtain

$$\begin{aligned}
&\left(\int_0^\infty \left[\frac{1}{r^\alpha}\beta_{f,d,q}(B_r(x))\right]^2 \frac{dr}{r}\right)^{1/2} \\
&\lesssim_{G,\alpha,p,q} \left(\int_0^\infty r^{2\eta-2\alpha-1}\left[\int_0^{4r} s^{-\eta-1}\left[M(\beta_{f,d}(B_{2s}(\cdot))^w)(x)\right]^{1/w} ds\right]^2 dr\right)^{1/2} \\
&\lesssim_{\alpha,q} \left(\int_0^\infty \left[M(r^{-w\alpha}\beta_{f,d}(B_{2r}(\cdot))^w)(x)\right]^{2/w} \frac{dr}{r}\right)^{1/2}.
\end{aligned}$$

We conclude that

$$\begin{aligned}
&\left(\int_G \left(\int_0^\infty \left[\frac{1}{r^\alpha}\beta_{f,d,q}(B_r(x))\right]^2 \frac{dr}{r}\right)^{p/2} dx\right)^{1/p} \\
&\lesssim_{G,\alpha,p,q} \left(\int_G \left(\int_0^\infty \left[M(r^{-w\alpha}\beta_{f,d}(B_{2r}(\cdot))^w)(x)\right]^{2/w} \frac{dr}{r}\right)^{p/2} dx\right)^{1/p} \\
&\lesssim_{G,\alpha,p,q} \left(\int_G \left(\int_0^\infty \left[\frac{1}{r^\alpha}\beta_{f,d}(B_r(x))\right]^2 \frac{dr}{r}\right)^{p/2} dx\right)^{1/p},
\end{aligned}$$

where in the second inequality we applied the Fefferman–Stein vector-valued maximal function inequality (2.61) with functions

$$g_j(x) = 2^{\alpha w j}\beta_{f,d}(B_{2^{-j+1}}(x))$$

and exponents $u = 2/w > 1$ and $v = p/w > 1$, along with Corollary 2.6.2.

This completes the proof of Theorem 2.0.22.

Bibliography

[1] Abraham, I., Bartal, Y., and Neiman, O. Embedding metric spaces in their intrinsic dimension. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms* (2008), pp. 363–372.

[2] Ambrosio, L. Some fine properties of sets of finite perimeter in Ahlfors regular metric measure spaces. *Advances in Mathematics 159*, 1 (2001), 51–67.

[3] Assouad, P. Plongements lipschitziens dans $\mathbb{R}^n$. *Bulletin de la Société Mathématique de France 111* (1983), 429–448.

[4] Austin, T., Naor, A., and Tessera, R. Sharp quantitative nonembeddability of the Heisenberg group into superreflexive Banach spaces. *Groups, Geometry, and Dynamics 7*, 3 (2013), 497–522.

[5] Azzam, J. Bi-Lipschitz parts of quasisymmetric mappings. *Revista Matemática Iberoamericana 32*, 2 (2016), 589–648.

[6] Ball, K., Carlen, E. A., and Lieb, E. H. Sharp uniform convexity and smoothness inequalities for trace norms. *Inventiones mathematicae 115*, 1 (1994), 463–482.

[7] Bartal, Y., Recht, B., and Schulman, L. J. Dimensionality reduction: beyond the Johnson-Lindenstrauss bound. In *Proceedings of the twenty-second*

annual ACM-SIAM symposium on Discrete Algorithms (2011), SIAM, pp. 868–887.

[8] Bergh, J., and Löfström, J. *Interpolation spaces: an introduction*, vol. 223. Springer Science & Business Media, 2012.

[9] Bogachev, V. I. *Measure theory*, vol. 2. Springer, 2007.

[10] Breuillard, E., and Le Donne, E. Nilpotent groups, asymptotic cones and subfinsler geometry. *Preprint* (2012).

[11] Cheeger, J., and Kleiner, B. On the differentiability of Lipschitz maps from metric measure spaces to Banach spaces. In *Inspired By SS Chern: A Memorial Volume in Honor of A Great Mathematician*. World Scientific, 2006, pp. 129–152.

[12] Cheeger, J., and Kleiner, B. Differentiating maps into L^1, and the geometry of BV functions. *Annals of mathematics* (2010), 1347–1385.

[13] Chousionis, V., Li, S., and Young, R. The strong geometric lemma for intrinsic lipschitz graphs in heisenberg groups. *Journal für die reine und angewandte Mathematik (Crelles Journal) 2022*, 784 (2022), 251–274.

[14] Coulhon, T., Russ, E., and Tardivel-Nachef, V. Sobolev algebras on Lie groups and Riemannian manifolds. *American Journal of Mathematics 123*, 2 (2001), 283–342.

[15] De Cornulier, Y., Tessera, R., and Valette, A. Isometric group actions on Hilbert spaces: growth of cocycles. *GAFA Geometric And Functional Analysis 17*, 3 (2007), 770–792.

[16] De Nápoli, P., and Martín, R. D. Mean value inequalities and characterizations of Sobolev spaces on graded groups. *arXiv preprint arXiv:2109.05064* (2021).

[17] Dixmier, J. Sur les représentations unitaires des groupes de Lie nilpotents. ii. *Bulletin de la Société mathématique de France 85* (1957), 325–388.

[18] Dixmier, J. Sur les représentations unitaires des groupes de Lie nilpotents. i. *American Journal of Mathematics 81*, 1 (1959), 160–170.

[19] Dorronsoro, J. R. A characterization of potential spaces. *Proceedings of the American Mathematical Society 95*, 1 (1985), 21–31.

[20] Dunford, N., and Schwartz, J. T. *Linear operators, part 1: general theory*, vol. 10. John Wiley & Sons, 1988.

[21] Eriksson-Bique, S., Gartland, C., Le Donne, E., Naples, L., and Nicolussi-Golo, S. Nilpotent groups and biLipschitz embeddings into L^1. *arXiv preprint arXiv:2112.11402* (2021).

[22] Fässler, K., and Orponen, T. Dorronsoro's theorem in Heisenberg groups. *Bulletin of the London Mathematical Society 52*, 3 (2020), 472–488.

[23] Fässler, K., and Orponen, T. Vertical versus horizontal Sobolev spaces. *Journal of Functional Analysis 279*, 2 (2020), 108517.

[24] Figiel, T. On the moduli of convexity and smoothness. *Studia Mathematica 56*, 2 (1976), 121–155.

[25] Folland, G. B. Subelliptic estimates and function spaces on nilpotent Lie groups. *Arkiv för matematik 13*, 1 (1975), 161–207.

[26] Furioli, G., Melzi, C., and Veneruso, A. Littlewood–paley decompositions and besov spaces on lie groups of polynomial growth. *Mathematische Nachrichten 279*, 9-10 (2006), 1028–1040.

[27] Gartland, C. Estimates on the markov convexity of carnot groups and quantitative nonembeddability. *Journal of Functional Analysis 279*, 8 (2020), 108697.

[28] Gottlieb, L.-A., and Krauthgamer, R. A nonlinear approach to dimension reduction. *Discrete Comput. Geom. 54*, 2 (2015), 291–315.

[29] Grafakos, L., Liu, L., and Yang, D. Vector-valued singular integrals and maximal functions on spaces of homogeneous type. *Mathematica Scandinavica* (2009), 296–310.

[30] Gromov, M. Groups of polynomial growth and expanding maps (with an appendix by Jacques Tits). *Publications Mathématiques de l'IHÉS 53* (1981), 53–78.

[31] Gupta, A., Krauthgamer, R., and Lee, J. R. Bounded geometries, fractals, and low-distortion embeddings. In *44th Annual IEEE Symposium on Foundations of Computer Science, 2003. Proceedings.* (2003), IEEE, pp. 534–543.

[32] Hanner, O. On the uniform convexity of l^p and ℓ^p. *Arkiv för Matematik 3*, 3 (1956), 239–244.

[33] Har-Peled, S., and Mendel, M. Fast construction of nets in low-dimensional metrics and their applications. *SIAM J. Comput. 35*, 5 (2006), 1148–1184.

[34] Hardy, G. H., Littlewood, J. E., Pólya, G., Pólya, G., et al. *Inequalities*. Cambridge university press, 1952.

[35] Heikkinen, T., Lehrbäck, J., Nuutinen, J., and Tuominen, H. Fractional maximal functions in metric measure spaces. *Analysis and Geometry in Metric Spaces 1*, 2013 (2013), 147–162.

[36] Heinonen, J., Koskela, P., Shanmugalingam, N., and Tyson, J. T. *Sobolev spaces on metric measure spaces*. No. 27. Cambridge University Press, 2015.

[37] Hörmander, L. Hypoelliptic second order differential equations. *Acta Math. 119* (1967), 147–171.

[38] Huang, Y., and Sun, S. Non-embedding theorems of nilpotent Lie groups and sub-Riemannian manifolds. *Frontiers of Mathematics in China 15*, 1 (2020), 91–114.

[39] Hulanicki, A. A functional calculus for Rockland operators on nilpotent Lie groups. *Studia Math. 78* (1984), 253–266.

[40] Hytönen, T., and Naor, A. Heat flow and quantitative differentiation. *Journal of the European Mathematical Society 21*, 11 (2019), 3415–3466.

[41] Jean, F. *Control of nonholonomic systems: from sub-Riemannian geometry to motion planning*. Springer, 2014.

[42] Jerison, D. The Poincaré inequality for vector fields satisfying Hörmander's condition. *Duke Mathematical Journal 53*, 2 (1986), 503–523.

[43] John, F. Extremum problems with inequalities as subsidiary conditions, Studies and Essays Presented to R. Courant on his 60th Birthday, January 8, 1948, 1948.

[44] Karidi, R. Geometry of balls in nilpotent Lie groups. *Duke Mathematical Journal 74*, 2 (1994), 301–317.

[45] Kirillov, A. A. Unitary representations of nilpotent Lie groups. *Russian mathematical surveys 17*, 4 (1962), 53–104.

[46] Kleiner, B. A new proof of Gromov's theorem on groups of polynomial growth. *Journal of the American Mathematical Society 23*, 3 (2010), 815–829.

[47] Lafforgue, V., and Naor, A. Vertical versus horizontal Poincaré inequalities on the Heisenberg group. *Israel Journal of Mathematics 203*, 1 (2014), 309–339.

[48] Lang, U., and Plaut, C. Bilipschitz embeddings of metric spaces into space forms. *Geom. Dedicata 87*, 1-3 (2001), 285–307.

[49] Le Donne, E. A primer on Carnot groups: homogenous groups, Carnot-Carathéodory spaces, and regularity of their isometries. *Analysis and Geometry in Metric Spaces 5*, 1 (2017), 116–137.

[50] Lee, J. R., Mendel, M., and Naor, A. Metric structures in L^1: dimension, snowflakes, and average distortion. *European Journal of Combinatorics 26*, 8 (2005), 1180–1190.

[51] Lee, J. R., and Naor, A. Extending Lipschitz functions via random metric partitions. *Inventiones mathematicae 160*, 1 (2005), 59–95.

[52] Lee, J. R., and Naor, A. L^p metrics on the Heisenberg group and the Goemans-Linial conjecture. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)* (2006), IEEE, pp. 99–108.

[53] Li, S. Coarse differentiation and quantitative nonembeddability for Carnot groups. *Journal of Functional Analysis 266*, 7 (2014), 4616–4704.

[54] Linial, N., London, E., and Rabinovich, Y. The geometry of graphs and some of its algorithmic applications. *Combinatorica 15*, 2 (1995), 215–245.

[55] Malcev, A. On a class of homogeneous spaces, izv. *Akad. Nauk. Armyan. SSSR Ser. Mat 13* (1949), 201–212.

[56] Martínez, T., Torrea, J. L., and Xu, Q. Vector-valued Littlewood–Paley–Stein theory for semigroups. *Advances in Mathematics 203*, 2 (2006), 430–475.

[57] Mendel, M., and Naor, A. Nonlinear spectral calculus and super-expanders. *Publications mathématiques de l'IHÉS 119* (2014), 1–95.

[58] Milman, V. D., and Schechtman, G. *Asymptotic Theory of Finite Dimensional Normed Spaces*. Springer-Verlag, Berlin, Heidelberg, 1986.

[59] Milnor, J. A note on curvature and fundamental group. *Journal of Differential geometry 2*, 1 (1968), 1–7.

[60] Naor, A. Comparison of metric spectral gaps. *Analysis and Geometry in Metric Spaces 2*, 1 (2014).

[61] Naor, A., and Neiman, O. Assouad's theorem with dimension independent of the snowflaking. *Rev. Mat. Iberoam. 28(4)* (2010), 1123–1142.

[62] Naor, A., and Neiman, O. Assouad's theorem with dimension independent of the snowflaking. *Rev. Mat. Iberoam. 28*, 4 (2012), 1123–1142.

[63] Naor, A., and Peres, Y. L^p compression, traveling salesmen, and stable walks. *Duke Mathematical Journal 157*, 1 (2011), 53–108.

[64] Naor, A., and Young, R. Vertical perimeter versus horizontal perimeter. *Annals of Mathematics 188*, 1 (2018), 171–279.

[65] Naor, A., and Young, R. Foliated corona decompositions. *arXiv preprint arXiv:2004.12522* (2020).

[66] Nash, J. C^1 isometric imbeddings. *Ann. of Math.* (1954), 383–396.

[67] Ostrovskii, M. I. Metric embeddings. In *Metric Embeddings*. de Gruyter, 2013.

[68] Pansu, P. Croissance des boules et des géodésiques fermées dans les nilvariétés. *Ergodic Theory and Dynamical Systems 3*, 3 (1983), 415–445.

[69] Pansu, P. Métriques de Carnot-Carathéodory et quasiisométries des espaces symétriques de rang un. *Annals of Mathematics* (1989), 1–60.

[70] Pisier, G. Martingales with values in uniformly convex spaces. *Israel Journal of Mathematics 20*, 3 (1975), 326–350.

[71] Ryoo, S.-Y. Vertical versus horizontal inequalities on simply connected nilpotent Lie groups and groups of polynomial growth. *arXiv preprint arXiv:2207.11305* (2022).

[72] Ryoo, S.-Y. Embedding snowflakes of Carnot groups into bounded dimensional Euclidean spaces with optimal distortion. *Anal. PDE 15*, 8 (2023), 1933–1990.

[73] Semmes, S. On the nonexistence of bilipschitz parameterizations and geometric problems about A_∞-weights. *Revista Matemática Iberoamericana 12*, 2 (1996), 337–410.

[74] Steenrod, N. *The topology of fibre bundles*. Princeton university press, 1999.

[75] Stein, E. M. *Singular Integrals and Differentiability Properties of Functions (PMS-30), Volume 30*. Princeton university press, 2016.

[76] Švarc, A. S. A volume invariant of coverings. In *Dokl. Akad. Nauk SSSR (NS)* (1955), vol. 105, pp. 32–34.

[77] Tao, T. Embedding the Heisenberg group into a bounded-dimensional Euclidean space with optimal distortion. *Rev. Mat. Iberoam. 37*, 1 (2020), 1–44.

[78] Tessera, R. Asymptotic isoperimetry on groups and uniform embeddings into Banach spaces. *Commentarii Mathematici Helvetici 86*, 3 (2011), 499–535.

[79] Tomczak-Jaegermann, N. The moduli of smoothness and convexity and the Rademacher averages of the trace classes S_p $(1 \leq p < \infty)$. *Studia Mathematica 50*, 2 (1974), 163–182.

[80] Varopoulos, N. T., Saloff-Coste, L., and Coulhon, T. Analysis and geometry on groups. *(No Title)* (1993).

[81] Whitney, H. The singularities of a smooth n-manifold in $(2n-1)$-space. *Ann. of Math.* (1944), 247–293.

[82] Xu, Q. Vector-valued Littlewood-Paley-Stein theory for semigroups II. *International Mathematics Research Notices 2020*, 21 (2020), 7769–7791.