Multi-parameter Commutators and New Function Spaces of Bounded Mean Oscillation

by

Yumeng Ou

B.Sc., Peking University; Beijing, China, 2011

M.Sc., Brown University; Providence RI, USA, 2013

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Department of Mathematics at Brown University

PROVIDENCE, RHODE ISLAND

May 2016

This dissertation by Yumeng Ou is accepted in its present form
by Department of Mathematics as satisfying the
dissertation requirement for the degree of Doctor of Philosophy.

Date________________ ________________________________

Jill Pipher, Ph.D., Advisor

Recommended to the Graduate Council

Date________________ ________________________________

Benoit Pausader, Ph.D., Reader

Date________________ ________________________________

Sergei Treil, Ph.D., Reader

Approved by the Graduate Council

Date________________ ________________________________

Peter Weber, Dean of the Graduate School

Vitae

Yumeng Ou was born on April 9th, 1989 in the city of Xi'an, China. She earned a B.Sc. degree in mathematics from the School of Mathematical Sciences at Peking University in July 2011, and a M.Sc. degree in mathematics from Brown University in May 2013. During her years at Brown, she taught multiple undergraduate courses including Introductory and Advanced Placement Calculus as both a primary lecturer and a teaching assistant. She has given invited research talks at various institutions including University of Wisconsin-Madison, University of Pennsylvania, Université Paul Sabatier, and University of Helsinki. She has presented posters at conferences such as Joint Mathematics Meetings and AWM Research Symposium. She has also participated in summer schools hosted by Christoph Thiele and Brett Wick. Starting from September 2016, she will be appointed as a C.L.E. Moore Instructor in the Department of Mathematics at Massachusetts Institute of Technology. During Spring 2017, she will go to MSRI as a postdoctoral fellow.

Below is a list of publications and submitted articles for the author.

1. Domination of multilinear singular integrals by positive sparse forms, with Amalia Culiuc and Francesco Di Plinio. Submitted (2016). arXiv:1603.05317v1.

2. Little BMO and Journé commutators, with Stefanie Petermichl. Submitted (2015).

3. A modulation invariant Carleson embedding theorem outside local L^2, with Francesco Di Plinio. Submitted (2015). arXiv:1510.06433v1.

4. Banach-valued multilinear singular integrals, with Francesco Di Plinio. Submitted (2015). arXiv:1506.05827v1.

5. Multi-parameter singular integral operators and representation theorem. Submitted

(2014). arXiv:1410.8055v1.

6. Higher order Journé commutators and characterizations of multi-parameter BMO, with Stefanie Petermichl and Elizabeth Strouse, Adv. Math. **291** (2016), 24-58. DOI: 10.1016/j.aim.2015.12.029.

7. Upper bound for multi-parameter iterated commutators, with Laurent Dalenc, Publ. Mat. **60** (2016), 191-220. DOI: 10.5565/PUBLMAT_60116_07.

8. A $T(b)$ theorem on product spaces, Trans. Amer. Math. Soc. **367** (2015), 6159-6197. DOI: 10.1090/S0002-9947-2015-06246-1. MR3356933.

Acknowledgements

There are many people along my journey through graduate school who have shown me the greatest kindness and given me a tremendous amount of help.

I would like to first thank my Ph.D. advisor, Jill Pipher, for being the most amazing advisor and role model that one can ever dream of having. Her wisdom, dedication, kindness, patience, encouragement, and support have shown me what it is like to be a true researcher and a generous mentor. She gave me the freedom to explore any areas that I'm interested in, worked for hours even on weekends helping me solve problems, encouraged a shy person like me to talk to people and go to conferences, and was always available whenever I was having a hard time in work or life. I am beyond grateful for having the opportunity to learn and work with her, and for having her as a dear friend.

My gratitude also goes to my dissertation committee members, Benoit Pausader and Sergei Treil, for their patience reading through my dissertation and giving me invaluable suggestions, as well as all the help and guidance that I have received from them during my years at Brown.

I wouldn't have understood those subjects and proved those theorems without my collaborators: Amalia Culiuc, who has been growing up together with me in harmonic analysis like a twin sister; Francesco Di Plinio, who has been such a supportive colleague and friend, and has deeply influenced me with his infinite passion for math; Stefanie Petermichl, who is full of brilliant ideas and has always been so encouraging and generous to me; Laurent Dalenc and Elizabeth Strouse, who I have sincerely enjoyed working with.

I am indebted to many harmonic analysts who have guided me: Michael Lacey, Svitlana

Mayboroda, Christoph Thiele, and Brett Wick. There are also many others, whom I have enjoyed hospitality from or inspiring conversations with: Dmitriy Bilyk, Yen Do, Philip Gressman, Steve Hofmann, Tuomas Hytönen, Joshua Isralowitz, Guozhen Lu, Henri Martikainen, Andreas Seeger, Betsy Stovall, and Rodolfo Torres.

I am fortunate to have learned real analysis with Tianquan Chen and harmonic analysis with Heping Liu during college, who made me fall in love with analysis. Thanks to many professors at Brown I was able to further develop my appreciation of this beautiful subject: Brian Cole, Hongjie Dong, Yan Guo, Justin Holmer, Walter Strauss, and John Wermer.

My thanks go to the Mathematics Department at Brown for its superb supporting system. I would like to thank Jeffrey Brock, Daniel Katz, and Yilun Wu for guiding me in teaching, and our amazing staff members Audrey Aguiar, Larry Larrivee, Lori Nascimento, Carol Oliveira, and Doreen Pappas for all the support and warmness that I have received from them. Not to mention the students in analysis: Theresa Anderson, Alexander Barron, Chiara Gallarati, Younghun Hong, Jingguo Lai, and Miles Wheeler, with whom I have had seminars, conversations, and many wonderful memories.

I am also very grateful to have the support from many of my other friends: Paul Carter, Hao Chen, Xiaoshuang Chen, Mengdi Cui, Xue Hao, Semin Kim, Quang Nhat Le, Dawei Lin, Chang Liu, David Lowry, Ningning Ma, Peter McGrath, Wenbo Shen, Zhouli Xu, Bowei Yan, Yingying Yan, Ren Yi, Hong Zhang, and Zihui Zhao.

I could never ask for a husband who is more supportive, understanding, and heart-warming. Kai Hong, the love of my life, has always been by my side through every single moment of joy and sorrow. The best parents in the world, Zhuocheng Ou and Juanjuan Yong, have showered me with the greatest love and trust, and taught me the value of confidence and independence. It is my dad who has kindled my love for logical thinking and taken me to explore the beauty of mathematics since my childhood. It is my mom who had given me the most unconditional love and had the steadfast belief in my potential. In my heart, she will always be with me, forever.

Abstract of “ Multi-parameter Commutators and New Function Spaces of Bounded Mean Oscillation ” by Yumeng Ou, Ph.D., Brown University, May 2016

In this thesis we study the relation between the boundedness of commutators of singular integrals and the function spaces of bounded mean oscillation (BMO) type that their symbols belong to. We discover a new family of BMO spaces in the multi-parameter setting, which are characterized by the boundedness of commutators of singular integrals. More precisely, we show for commutators with a large family of multi-parameter singular integrals $\{T_j\}_{j=1}^t$ and arbitrarily many iterations

$$[\ldots[b,T_1],\ldots,T_t]$$

where $[A,B]:=AB-BA$ and b is understood as the pointwise multiplication, that their operator norms on L^p are equivalent to some BMO norm of the symbols b. Such results have wide applications in the study of Hankel operators, weak factorization of Hardy spaces, and Div-Curl estimates that arise in PDE problems. In one direction of the proof, where the operator norm is shown to be less than the BMO norm, we extend to the multi-parameter setting the representation theorem of Calderón-Zygmund operators as averages of dyadic shifts. This not only enables us to obtain upper estimates for very general multi-parameter Journé commutators, but also yields a $T(1)$ theorem in arbitrarily many parameters, which is of independent interest. In the other direction of the proof, where the BMO norm is shown to be less than the operator norm, we study in particular the case of Hilbert and Riesz commutators, of iterated and mixed type. The argument for Hilbert commutators is of a complex analysis nature, relying heavily on the analytic structure of Hilbert transform and Toeplitz operators. The Riesz commutators case turns out to be very complicated and different from previously known results. We introduce a probabilistic method involving zonal harmonics in our argument, aided by the upper estimates for general Journé commutators we newly obtained. We further study the new family of BMO spaces that we construct and characterize, establishing a partial order among them and demonstrating some weak factorization results for their pre-dual Hardy type spaces. In our proof, some new pararoduct type operators are also studied.

Contents

Chapter One

Introduction

In this thesis, we study the boundedness of commutators of general singular integrals and their connection with a new family of function spaces of bounded mean oscillation, which are called *little product BMO*.

Singular integral theory is a central area of classical harmonic analysis, and has numerous applications in many problems arising in partial differential equations, several complex variables, etc. The class of singular integrals we are concerned with in this work are Calderón-Zygmund operators, in both one-parameter and multi-parameter settings. A classical example of this class is the Hilbert transform

$$Hf(x) := \lim_{\epsilon\to 0}\int_{y\in\mathbb{R},|x-y|>\epsilon}\frac{f(y)}{x-y}\,dy$$

which arises naturally in the study of pointwise convergence of partial sums of Fourier series, as well as in a more familiar setting of conjugate harmonic functions in complex function theory. Note that the integral above fails to be absolutely convergent, and the existence of the limit as a function in x is a consequence of the size and cancellation properties of the kernel $1/x$, which has a singularity at the origin. In higher dimensions, the Hilbert transform has an analog called Riesz transforms

$$R_jf(x) := \lim_{\epsilon\to 0}\int_{y\in\mathbb{R}^d,|y|>\epsilon} f(x-y)\frac{y_j}{|y|^{d+1}}\,dy,\quad 1\le j\le d,$$

whose kernels are also singular at a point. In general, a *Calderón-Zygmund (CZ) operator* is assumed to be L^2 bounded with a distributional kernel being "standard", in the sense that certain size and smoothness conditions are satisfied. The precise definition is given in Section 2.1.

Calderón and Zygmund [CZ52] developed a delicate machinery for studying such operators, and proved that they are all bounded on L^p (the Banach space of functions whose p-th powers are integrable), $1 < p < \infty$, and maps L^1 into $L^{1,\infty}$, the weak L^1 space. At the endpoints, they showed that CZ operators map L^∞ into a *John-Nirenberg space of bounded mean oscillation (BMO)*, and map the Hardy space H^1 into L^1. The Banach space BMO is defined via the norm

$$\|f\|_{\mathrm{BMO}(\mathbb{R}^d)} := \sup_{Q\subset\mathbb{R}^d} \frac{1}{|Q|}\int_Q |f(x) - \langle f\rangle_Q|\,dx,$$

where $\langle f\rangle_Q$ denotes the average value of f over the cube Q. This space plays a very important role in singular integral theory. In fact, under some weak conditions, a singular integral mapping some fixed L^∞ function into BMO guarantees its boundedness on L^2, which are often referred to as $T(1)/T(b)$ type theorems. It is shown by C. Fefferman and Stein [FS72] that BMO is the dual of H^1. In addition to many endpoint results in analysis, BMO arises naturally as well in partial differential equations and probability.

When it comes to the multi-parameter setting, in other words, when one considers singular integrals that are invariant under non-isotropic dilations such as $(x_1,\ldots,x_t) \to (\delta_1 x_1,\ldots,\delta_t x_t)$, $x_i \in \mathbb{R}^{d_i}, \delta_i > 0$, Calderón-Zygmund theory fails to apply. Note that this class is distinct from the higher dimensional Calderón-Zygmund operators, since the latter, for instance the Riesz transforms, are invariant under a one-parameter family of dilations $x \to \delta x$, $x \in \mathbb{R}^d, \delta > 0$. One simple example of an operator in this class is the double Hilbert transform

$$H\otimes Hf(x) := \lim_{\epsilon_1\to 0,\epsilon_2\to 0}\int_{\substack{(y_1,y_2)\in\mathbb{R}\times\mathbb{R}\\ |x_1-y_1|>\epsilon_1,|x_2-y_2|>\epsilon_2}} \frac{f(y)}{(x_1-y_1)(x_2-y_2)}\,dy_1dy_2$$

whose kernel $1/x_1x_2$ has singularity on two lines. Due to its tensor product structure, $H\otimes H$ is bounded on $L^p(\mathbb{R}\times\mathbb{R})$ for $1 < p < \infty$ by iteration. However, most multi-parameter singular integrals are not of tensor product type, and there is usually a vast difference between the machineries that are required in the multi-parameter setting and the one-parameter Calderón-Zygmund world. Such operators arise naturally in many problems including the summation of multiple Fourier series and some boundary value problems in PDE.

The class of multi-parameter singular integrals are first studied by R. Fefferman and Stein [FS82] in the convolutional case, and then are generalized by Journé [Jou85] to the general case. See Section 2.2.1 for the definition of Journé's class. In the same paper, Journé also proved that such operators have similar mapping properties as their one-parameter analog, CZ operators. More precisely, they map L^∞ into *product BMO*, and map product H^1 into L^1. The product BMO, studied by Chang and R. Fefferman in a series of works, is considered as the correct analog of BMO in the multi-parameter setting, in the sense that it is the dual of product H^1. See Subsection 2.2.2 for a definition and detailed discussion of this space. We also mention that there are other spaces of BMO type in the multi-parameter setting that are useful, one example of which that is relevant to us is the *little BMO*, denoted by bmo, containing functions that are uniformly in one-parameter BMO in each variable with all the other variables fixed.

It turns out that BMO spaces play a more profound role in singular integral theory, due to its connection to the boundedness of commutators of singular integrals. A classical result of Nehari [Neh57] says that

$$\|b\|_{\mathrm{BMO}(\mathbb{R})} \approx \|[b,H]\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})},$$

where H denotes the Hilbert transform and $[b,H] := bH - Hb$, b being understood as the pointwise multiplication operator by function b. Note that the cancellation structure of the commutator plays a key role here, since BMO functions, unlike L^∞, are not L^2 multipliers in general. Several remarkable results have then been obtained along this line of research, unveiling the deep connection between the boundedness of commutators and the membership of the symbol functions in BMO. To name a few, Ferguson-Sadosky [FS00] and later Ferguson-Lacey [FL02] showed that

$$\|b\|_{\mathrm{bmo}} \approx \|[b,H_1H_2]\|_{L^2\to L^2}$$

and

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod}}} \approx \|[[b,H_1],H_2]\|_{L^2\to L^2}$$

in the multi-parameter iterated case. Coifman, Rochberg and Weiss [CRW76] extended it to the Riesz transforms case by proving for any $1 < p < \infty$

$$\|b\|_{\mathrm{BMO}(\mathbb{R}^d)} \approx \sup_{1\le j\le d} \|[b, R_j]\|_{L^p(\mathbb{R}^d)\to L^p(\mathbb{R}^d)}.$$

It was obtained by Lacey, Petermichl, Pipher and Wick [LPPW09] that

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod}}} \approx \sup_{\vec{j}} \|[\ldots[[b, R_{1,j_1}], R_{2,j_2}], \ldots, R_{t,j_t}]\|_{L^p(\mathbb{R}^{\vec{d}})\to L^p(\mathbb{R}^{\vec{d}})},$$

where $\vec{j} = (j_1, \ldots, j_t)$ with $1 \le j_s \le d_s$ for $s = 1, \ldots, t$. One can in fact ask a very general question: for what BMO type space and collections of singular integrals, one-parameter or multi-parameter, does one have the following equivalence of norms

$$\|b\|_{\text{“BMO”}} \approx \sup_{\vec{j}} \|[\ldots[b, T_{1,j_1}], \ldots, T_{t,j_t}]\|_{L^p\to L^p}? \tag{1.1}$$

This is the main problem we will explore in this thesis.

In the above type of results, the boundedness of the operator norm by the BMO norm is usually referred to as the “upper bound”, while the other direction is called the “lower bound”. Note that the lower bound estimate usually requires more structure of the operators, as one can easily use multiplication operators to construct a counterexample. We also mention that commutator estimates of this type are in fact closely related with Hankel operators, weak factorization of Hardy spaces, and Div-Curl estimates. More discussions about this will be given in Section 3.1.

Estimates of this type also provide valuable insights for studying BMO spaces in the multi-parameter setting systematically. When entering a setting with several free parameters, a large variety of BMO type spaces are encountered, some of which lose the feature of mean oscillation itself. This makes it a desirable and tangible approach to characterize multi-parameter BMO spaces through the viewpoint of boundedness of commutators.

Our first contribution is a thorough study of the Hilbert and Riesz commutators in the mixed case, i.e. we establish (1.1) in the case that all T_{s,j_s} are tensor products of Hilbert or Riesz transforms with arbitrarily many parameters and iterations. A special case of our result says that

$$\|b\|_{\mathrm{BMO}_{(13)(2)}(\mathbb{R}^3)} \approx \|[[b, H_1 H_3], H_2]\|_{L^2(\mathbb{R}^3)\to L^2(\mathbb{R}^3)}.$$

It turns out that, in order to characterize the boundedness of the mixed type commutators above, the symbol function has to belong to a new BMO type space $\mathrm{BMO}_{(13)(2)}$, defined as the space of functions $b(x_1, x_2, x_3)$ such that $b(x_1, \cdot, \cdot)$ and $b(\cdot, \cdot, x_3)$ are uniformly in product BMO in the remaining two variables, normed by the supremum of the product BMO norms. We call BMO spaces of this type *little product BMO* in general and study them in Chapter 6. A weak factorization result for the pre-dual of little product BMO is also obtained there.

Second, we answer (1.1) in the upper bound direction for very general collections of operators: tensor products of arbitrarily many CZ operators, or "paraproduct free" multi-parameter Journé operators, showing that their commutators are always bounded by the corresponding BMO norm of the symbol. The tensor product case can be effectively reduced to the one-parameter endpoint case:

$$\|[\dots[b, T_1], \dots, T_t]\|_{L^2\to L^2} \lesssim \|b\|_{\mathrm{BMO}_{\mathrm{prod}}},$$

where all T_j's are one-parameter CZ operators. This is the first upper estimate result involving general CZ operators not of convolution type. In the multi-parameter Journé commutator case, the paraproduct free assumption is a technical one, which, for instance in the degenerated one-parameter setting, means that $T(1) = T^*(1) = 0$. This assumption is the only gap that is left for a complete solution to the upper bound direction of (1.1). It is still under our investigation whether it could be removed, and some partial progress has been achieved. Nonetheless, our result is the first of its kind involving Journé commutators, and their upper bound estimate turns out to be very crucial in obtaining the lower bound for the mixed Riesz commutator results mentioned above. Moreover, combining these result together shows that Riesz transforms is a representative class of Journé operators in the sense that the boundedness of Riesz commutators imply that of the Journé commutators. We also remark that, although stated in L^2, all the results above can in fact be extended to hold in L^p for $1 < p < \infty$.

In terms of the proof of the theorems, we would first like to highlight the ingredient of the representation theorem of singular integrals as averages of dyadic shifts. One of the most powerful tools that have emerged in the last fifteen years in singular integral theory is a representation of arbitrary singular integrals as averages of simpler components, which are built from dyadic basis on Hilbert or other L^p spaces in terms of the well known martingale theory, in particular, Haar functions. Take the one-dimensional case as an example. Each dyadic interval I from a fixed dyadic grid, say,

$$\mathcal{D} := \{2^{-k}([0,1) + m) : k \in \mathbb{Z}, m \in \mathbb{Z}\},$$

can be associated with a Haar function $h_I := |I|^{-1/2}(\chi_{I^-} - \chi_{I^+})$, where I^- (I^+) denotes the left (right) dyadic child of I. It is well known that Haar functions form an orthonormal basis of $L^2(\mathbb{R})$, and

$$f = \sum_{I \in \mathcal{D}} \langle f, h_I \rangle h_I, \quad \forall\, f \in L^2(\mathbb{R}).$$

It is derived by Petermichl [Pet00] a representation formula which says that Hilbert transform is equal to the average of the following *Haar shifts* over translations and dilations of some fixed grid:

$$Sf := \sum_{I \in \mathcal{D}} \frac{1}{\sqrt{2}} \langle f, h_I \rangle h_{I^-}.$$

Comparing Haar shift with the identity operator given by Haar expansion, one can observe that Haar shift is an operator that rearranges the Haar coefficients of any L^2 function, which are very easy to approach. Petermichl's result opened up the door of representing continuous singular integrals via simple dyadic operators. Such representation formulas have been extended to hold for arbitrary Calderón-Zygmund operators and even bi-parameter Journé operators, due to Hytönen [Hyt12] and Martikainen [Mar12] respectively. Note that in order to be able to represent more complicated operators than Hilbert transform, more general versions of Haar shifts (*dyadic shifts of higher complexity*) are needed.

A novelty of our work is the application of such representation theorems to the upper estimates of commutators. Briefly speaking, we derive uniform estimates for commutators of dyadic shifts, and pass back to the commutators of singular integrals via appropriate representations. This idea has been used in [LPPW10], where commutators of Haar shift are estimated. However, our result is the first case that commutators of dyadic shifts with arbitrary complexity are studied, which is intrinsically more difficult since uniform bounds that are independent of the complexity become crucial. Moreover, unlike Hilbert transform, which is paraproduct free, general CZ operators have an integrated paraproduct part, which requires very different attention. In fact, estimates for several new paraproduct type operators are obtained along the way of our proof, which might have other applications. In addition, the upper estimate of commutators yields immediately a stability result for characterizing families of BMO. More precisely, following from a trivial triangle inequality argument, it implies that a characterizing family of BMO stays as a characterizing family under small perturbations of singular integrals.

Furthermore, recall that our results are concerned with Journé commutators in arbitrarily many parameters, which means that Martikainen's bi-parameter version of the representation theorem is not sufficient for the execution of the strategy explained above. It is well known to experts in the field that compared with the bi-parameter theory, multi-parameter theory usually involves an additional layer of difficulty. Therefore, in order to tackle this, it is necessary for us to derive a representation theorem for multi-parameter Journé operators, which is not only crucial in our commutator estimates, but also certainly of its own interest. As an example, it yields a multi-parameter $T(1)$ theorem as a direct consequence. We give the formulation and proof of this result in Chapter 7.

We would also like to address an important element in the estimates for Riesz commutators, in particular the lower bound: a probabilistic method involving zonal harmonics. Note that the classical Riesz commutators estimate directly bypassing cone multipliers

([LPPW09]) fails in the general case, where the commutators are of mixed type with tensor products of multiple Riesz transforms being present in some component. In our proof, we construct a special type of Journé commutators, through geometric considerations and an averaging procedure of zonal harmonics on products of spheres, as a bridge that connects cone multiplier commutators to the desired Riesz commutators. The upper estimates for such Journé commutators are crucial in this argument.

The main structure of the thesis is as follows. In Chapter 2, we review some relevant preliminary facts about one-parameter and multi-parameter singular integral theory, together with an introduction of prior representation theorems. In Chapter 3, we discuss our main results regarding commutators of singular integrals, among which the general upper bound proofs are given in Chapter 4, and the arguments for Hilbert and Riesz commutators, in particular the lower bound, are presented in Chapter 5. We then further study our new spaces of bounded mean oscillation (little product BMO) in more details in Chapter 6, where a weak factorization result for their pre-duals are derived as well. Lastly, in Chapter 7, the representation theorem in arbitrarily many parameters is proved. At the end, we record a real variable proof of the Hilbert commutator result in Appendix A. The L^p case of the upper bound estimates for $p \neq 2$ and several more general paraproduct type operators are discussed in Appendix B.

Chapter Two

Preliminary on Multi-parameter Singular Integrals

In this chapter, we review some preliminary facts about singular integral theory, in particular multi-parameter singular integral theory, that will be relevant to us throughout the thesis. Singular integrals are central subjects in this thesis because that: 1. they will appear in the formulation of iterated commutators; 2. their mapping properties are closely related with different types of spaces of *bounded mean oscillation (BMO)*, which will be shown later to be characterizing function spaces for the boundedness of commutators.

In the following, we will first review the one-parameter Calderón-Zygmund theory in Section 2.1, and then move on to the multi-parameter generalization in Section 2.2, where the notion of *Journé operator* will be introduced. In the end, we will discuss a powerful tool that represents singular integrals as averages of simpler components.

2.1 Calderón-Zygmund theory

We have mentioned in the Introduction that a classical example of singular integrals is the *Hilbert transform*

$$Hf(x) := \lim_{\epsilon\to 0}\int_{y\in\mathbb{R},|x-y|>\epsilon}\frac{f(y)}{x-y}\,dy,$$

where the existence of the limit as a function in x is a consequence of the size and cancellation properties of the kernel $1/x$, which has a singularity at the origin. Another example is the so-called *Calderón commutator*

$$T_kf(x) := \lim_{\epsilon\to 0}\int_{y\in\mathbb{R},|x-y|>\epsilon}\left(\frac{A(x)-A(y)}{x-y}\right)^k\frac{f(y)}{x-y}\,dy,$$

where $k\in\mathbb{N}$ and A is a Lipschitz function on $\mathbb{R}$. This is not a convolution operator anymore and it arises in the study of Cauchy integrals along Lipschitz curves.

Many singular integral operators that are important in analysis and PDE (including the two examples above) fall into the category of *Calderón-Zygmund (CZ) operators*, which are L^2 bounded operators whose distributional kernel is "standard", in the sense that it satisfies certain size and smoothness conditions. Here is the definition, where the operator doesn't have to be of convolution type.

Definition 2.1. Let Δ be the diagonal of $\mathbb{R}^d \times \mathbb{R}^d : \Delta = \{(x,x) : x \in \mathbb{R}^d\}$. Let T be a bounded operator on $L^2(\mathbb{R}^d)$, and let K be a function of $\mathbb{R}^d \times \mathbb{R}^d \setminus \Delta$ such that for $f \in L^2(\mathbb{R}^d)$ with compact support,

$$Tf(x) = \int_{\mathbb{R}^d} K(x,y) f(y)\, dy, \quad x \notin \operatorname{spt} f.$$

Further, suppose that K is a *standard kernel*, meaning that there is $\delta > 0$ such that

$$|K(x,y)| \leq \frac{C}{|x-y|^d}, \tag{2.2}$$

$$|K(x,y) - K(x,y')| \leq C\frac{|y-y'|^\delta}{|x-y|^{d+\delta}} \quad \text{if } |x-y| > 2|y-y'|, \tag{2.3}$$

$$|K(x,y) - K(x',y)| \leq C\frac{|x-x'|^\delta}{|x-y|^{d+\delta}} \quad \text{if } |x-y| > 2|x-x'|. \tag{2.4}$$

Then T is called a Calderón-Zygmund (CZ) operator.

Note that all CZ operators form a Banach space, normed by $\|T\|_{\mathrm{CZ}} := \|T\|_{L^2 \to L^2} + |K|$, where $|K|$ denotes the best constant in (2.2), (2.3), (2.4). It can be easily verified that both the Hilbert transform and Calderón commutator are CZ operators.

Calderón and Zygmund [CZ52] studied such operators, and proved that they are all bounded on L^p, $1 < p < \infty$, and maps L^1 into $L^{1,\infty}$. For convolution operators, the L^2 boundedness follows from the bounds of the kernel using Plancherel's theorem. When it comes to non-convolution operators, a more delicate boundedness criterion is needed: the *$T(1)$ theorem* [DJ84] asserts that under a very weak regularity condition, T is bounded on L^2 if and only if $T1$ and T^*1 both lie in a *John-Nirenberg space of bounded mean oscillation (BMO)*. The space BMO plays a very important role in singular integral theory. For example, CZ operators map L^∞ (the Banach space of essentially bounded functions) to BMO. While at the other endpoint, the classical Hardy space (H^1) arises where functions are mapped into L^1 by CZ operators. BMO is the dual space of H^1 ([FS72]).

Now we briefly look at the space BMO, which is normed by

$$\|f\|_{\mathrm{BMO}(\mathbb{R}^d)} := \sup_{Q\subset\mathbb{R}^d} \left(\frac{1}{|Q|}\int_Q |f(x)-\langle f\rangle_Q|^p\,dx\right)^{1/p},$$

where $\langle f\rangle_Q := |Q|^{-1}\int_Q f$ denotes the average value of f on cube Q. And it is a celebrated result of John-Nirenberg that the norm defined above doesn't depend on the choice of $0<p<\infty$ (up to a constant). If one restricts the supremum to be taken over only dyadic cube Q's from a fixed dyadic grid $\mathcal{D}$ on $\mathbb{R}^d$, then the obtained norm characterizes the so-called *dyadic BMO*, denoted by $\mathrm{BMO}_{\mathcal{D}}(\mathbb{R}^d)$. It is shown by Garnett-Jones [GJ82] that a function is in BMO if and only if it is in dyadic BMO for any dyadic grid.

Dyadic BMO is in general more approachable because it has a natural connection to martingale theory, and can be characterized via Haar basis. One of the simplest examples of martingales arises using *Haar functions*. In one-dimension, let $\mathcal{D}$ be a fixed dyadic grid, and denote by I^- (I^+) the left (right) child of I for each dyadic cube $I\in\mathcal{D}$. Then the cancellative Haar function h_I^0 is defined as $h_I^0 := |I|^{-1/2}(\chi_{I^-}-\chi_{I^+})$. It is a fact that the collection $\{h_I^0\}_{I\in\mathcal{D}}$ forms an orthonormal basis of $L^2(\mathbb{R})$, and one has the following expansion

$$f=\sum_{I\in\mathcal{D}}\langle f,h_I^0\rangle h_I^0,\quad \forall f\in L^2, \tag{2.5}$$

where $\langle f,h_I^0\rangle$ is the inner product on L^2. We also define for any dyadic interval I a non-cancellative Haar function $h_I^1 := |I|^{-1/2}\chi_I$. In d dimensions, each cube $I=I_1\times\cdots\times I_d$ is associated with 2^d Haar functions:

$$h_I^\epsilon(x)=h_{I_1\times\cdots\times I_d}^{(\epsilon_1,\dots,\epsilon_d)}(x_1,\dots,x_d)=\prod_{i=1}^d h_{I_i}^{\epsilon_i}(x_i),\ \epsilon\in\{0,1\}^d,$$

where $h_I^{\vec{1}}$ is said to be non-cancellative, while all the other 2^d-1 Haar functions h_I^ϵ for $\epsilon\in\{0,1\}^d\setminus\{1\}$ are cancellative. Note that all the cancellative Haar functions for a fixed grid form an orthonormal basis of $L^2(\mathbb{R}^d)$. In this paper, we usually suppress the parameter

ϵ to abbreviate the notation.

It can be easily verified by letting $p = 2$ in the definition that dyadic BMO has the following equivalent formulation:

$$\|f\|^2_{\mathrm{BMO}_{\mathcal{D}}}(\mathbb{R}^d) = \sup_{Q \in \mathcal{D}} \frac{1}{|Q|} \sum_{I \in \mathcal{D}, I \subset Q} \sum_{\epsilon \neq \vec{1}} |\langle b, h_I^\epsilon \rangle|^2.$$

We also mention that dyadic BMO is the dual of dyadic H^1, which can be characterized via dyadic square function.

$$\|f\|_{H^1_{\mathcal{D}}} := \|S_{\mathcal{D}} f\|_{L^1},$$

where

$$(S_{\mathcal{D}} f) := \left(\sum_{I \in \mathcal{D}} \sum_{\epsilon \neq \vec{1}} |\langle f, h_I^\epsilon \rangle|^2 \frac{\chi_I}{|I|} \right)^{1/2}. \tag{2.6}$$

And similarly as the story of BMO, a function is in H^1 if and only if it is in $H^1_{\mathcal{D}}$ for any dyadic grid. This result together with its multi-parameter version can be found in [Tre09].

Before ending this section, we give another example of CZ operator in dimension $d > 1$: the *Riesz transforms*

$$R_j f(x) := \lim_{\epsilon \to 0} \int_{y \in \mathbb{R}^d, |y| > \epsilon} f(x-y) \frac{y_j}{|y|^{d+1}} \, dy, \quad 1 \leq j \leq d$$

whose kernels are singular at a point and satisfy the size and smoothness conditions so that the limit exists. Such operators are invariant under a one-parameter family of dilations $x \to \delta x$, $x \in \mathbb{R}^d, \delta > 0$. And they have very important applications in PDE.

2.2 Multi-parameter singular integrals

Considering a t-parameter family of dilations $(x_1, \ldots, x_t) \to (\delta_1 x_1, \ldots, \delta_t x_t)$, $x_i \in \mathbb{R}^{d_i}, \delta_i > 0$, we are led to study a class of *multi-parameter singular integrals* outside of the scope of

Calderón-Zygmund theory, which are invariant under such family. A basic example in $\mathbb{R}^2$ is the *double Hilbert transform*

$$H \otimes Hf(x) := \lim_{\epsilon_1 \to 0, \epsilon_2 \to 0} \int_{\substack{(y_1,y_2) \in \mathbb{R}\times\mathbb{R} \\ |x_1-y_1|>\epsilon_1, |x_2-y_2|>\epsilon_2}} \frac{f(y)}{(x_1-y_1)(x_2-y_2)}\, dy_1 dy_2$$

whose kernel $1/x_1x_2$ has singularity on two lines. Due to its tensor product structure, $H \otimes H$ is bounded on $L^p(\mathbb{R}\times\mathbb{R})$ for $1 < p < \infty$ by iteration. However, most multi-parameter singular integrals are not of tensor product type, and there is usually a vast difference between the machineries that are required in the multi-parameter setting and the one-parameter Calderón-Zygmund world. Such operators arise naturally in many problems including the summation of multiple Fourier series, as well as another example being the operator associated with kernel

$$K(x, y) = \frac{x_k}{(|x|^2 + y^2)^{(n+1)/2}} \cdot \frac{1}{|x|^2 + iy}, \quad (x, y) \in \mathbb{R}^n \times \mathbb{R}$$

studied in [FS82] in connection with some boundary value problems and the $\bar{\partial}$-Neumann problem in particular. (See [PS82].)

The study of multi-parameter singular integrals of convolution type was first carried out by R. Fefferman and Stein in [FS82], then Journé established a version of the classical David-Journé $T(1)$ theorem [DJ84] for a class of multi-parameter singular integrals of non-convolution type in [Jou85], which we introduce next.

2.2.1 Journé operators

In [Jou85], Journé developed explicitly a class of bi-parameter singular integral operators and proved a $T(1)$ type theorem for them. It is also pointed out in [Jou85] that, by induction, his approach can be generalized to arbitrarily many parameters. In order to define Journé's operators, we need to first look at the notion of Banach-valued standard kernels. In fact, the key idea of Journé's approach is to view a bi-parameter operator as a vector-valued one-parameter CZ operator.

Definition 2.7. Let B be a Banach space. A continuous function $K : \mathbb{R}^d \times \mathbb{R}^d \setminus \Delta \to B$ is called a *B-δ-standard kernel*, if for some $0 < \delta \leq 1$, $C > 0$ we have

$$\|K(x,y)\|_B \leq \frac{C}{|x-y|^d},$$

$$\|K(x,y) - K(x,y')\|_B \leq C\frac{|y-y'|^\delta}{|x-y|^{d+\delta}}, \quad \text{if } |x-y| > 2|y-y'|,$$

$$\|K(x,y) - K(x',y)\|_B \leq C\frac{|x-x'|^\delta}{|x-y|^{d+\delta}}, \quad \text{if } |x-y| > 2|x-x'|.$$

$|K|$ denotes the best constant in the inequalities above.

Recall that in the previous section, we have mentioned that the spaces of all Calderón-Zygmund operators (for a fixed parameter δ that appears in the standard kernel conditions) form a Banach space (denoted by δCZ). In the following, we will primarily look at the case $B = \delta\text{CZ}$.

Now we define the Journé type bi-parameter singular integral operators (SIO).

Definition 2.8. Let $T : C_0^\infty(\mathbb{R}^{d_1}) \otimes C_0^\infty(\mathbb{R}^{d_2}) \to [C_0^\infty(\mathbb{R}^{d_1}) \otimes C_0^\infty(\mathbb{R}^{d_2})]'$ be a continuous linear mapping. It is a *Journé type bi-parameter δ-SIO* if there exists a pair (K_1, K_2) of δCZ-δ-standard kernels so that, for all $f_1, g_1 \in C_0^\infty(\mathbb{R}^{d_1})$ and $f_2, g_2 \in C_0^\infty(\mathbb{R}^{d_2})$,

$$\langle T(f_1 \otimes f_2), g_1 \otimes g_2 \rangle = \int f_1(y_1) \langle K_1(x_1, y_1) f_2, g_2 \rangle g_1(x_1)\, dx_1 dy_1 \tag{2.9}$$

when $\text{spt} f_1 \cap \text{spt} g_1 = \emptyset$;

$$\langle T(f_1 \otimes f_2), g_1 \otimes g_2 \rangle = \int f_2(y_2) \langle K_2(x_2, y_2) f_1, g_1 \rangle g_2(x_2)\, dx_2 dy_2 \tag{2.10}$$

when $\text{spt} f_2 \cap \text{spt} g_2 = \emptyset$.

We emphasize again that a δCZ-δ-standard kernel is a standard kernel with parameter

δ whose value is in the Banach space δCZ, the space of Calderón-Zygmund operators equipped with the norm $\|T\|_{L^2\to L^2} + \|K\|$.

Let T_1 denote the partial adjoint of T in the first variable,i.e.

$$\langle T_1(f_1 \otimes f_2), g_1 \otimes g_2\rangle := \langle T(g_1 \otimes f_2), f_1 \otimes g_2\rangle,$$

then it is easy to see that T_1 is also a Journé type δ-SIO if T is. And a Journé type δ-SIO T is called a *Journé type bi-parameter δ-CZ operator* if both T and T_1 are bounded on L^2, associated with the norm $\|T\|_{L^2\to L^2} + \|T_1\|_{L^2\to L^2} + \|K_1\|_{\delta\mathrm{CZ}} + \|K_2\|_{\delta\mathrm{CZ}}$.

By induction, one can define Journé type t-parameter singular integrals accordingly for any $t \geq 3$. In the rest of the thesis, without further explanation, we always refer to multi-parameter Journé type CZ operators simply as *Journé operators*.

In the same paper, Journé proved a $T(1)$ theorem for his class of SIO, which is the first result of this type in the multi-parameter setting. He also showed that Journé operators map L^∞ to product BMO (denoted by $\mathrm{BMO}_{\mathrm{prod}}$), which is the "correct" multi-parameter analog of the classical BMO space, in the sense that it is dual to the product Hardy space H^1. (See the pioneer work [Cha79, CF80, CF82] by Chang and R. Fefferman for details.) Product BMO is very important to us, as it is the first BMO type space we have encountered in the multi-parameter setting, and will be seen to have close connection to commutators. We leave its definition and more detailed discussion to the next subsection.

2.2.2 Chang-Fefferman product BMO

As in the one-parameter case, there are several equivalent definitions of the product BMO space. One way to look at it is from the viewpoint of product Hardy space. In one parameter, the Hardy space $H^1(\mathbb{R}^d)$ can be defined with the norm $\sum_{j=0}^d \|R_j f\|_1$ where R_j denotes the j^{th} Riesz transform or the Hilbert transform if the dimension is one. Here and below we adopt the convention that R_0, the 0^{th} Riesz transform, is the identity. This space is invariant under the one-parameter family of isotropic dilations, while the product Hardy space $H^1_{\text{prod}}(\mathbb{R}^{\vec{d}})$ is invariant under dilations of each coordinate separately. That is, it is invariant under a t parameter family of dilations. One way to define a norm on $H^1_{\text{prod}}(\mathbb{R}^{\vec{d}})$ is

$$\|f\|_{H^1_{\text{prod}}} \sim \sum_{0\le j_l\le d_l} \|\bigotimes_{l=1}^{t} \mathrm{R}_{l,j_l} f\|_1.$$

R_{l,j_l} is the Riesz transform in the j_l^{th} direction of the l^{th} variable, and the 0^{th} Riesz transform is the identity operator. Then the t-parameter product BMO space can be defined as the dual of the Hardy space: $H^1_{\text{prod}}(\mathbb{R}^{\vec{d}})^*$.

Another way to characterize product BMO is via a product Carleson measure, a result shown in [CF80, CF82]. Define

$$\|b\|_{\mathrm{BMO}_{\text{prod}}(\mathbb{R}^{\vec{d}})} := \sup_{U\subset\mathbb{R}^{\vec{d}}} \Big(|U|^{-1} \sum_{R\subset U} \sum_{\vec{\varepsilon}\in \mathrm{sig}_{\vec{d}}} |\langle b, w_R^{\vec{\varepsilon}}\rangle|^2 \Big)^{1/2}. \tag{2.11}$$

Here the supremum is taken over all open subsets $U \subset \mathbb{R}^{\vec{d}}$ with finite measure, and we use a wavelet basis $w_R^{\vec{\varepsilon}}$ adapted to rectangles $R = Q_1 \times \cdots \times Q_t$, where each Q_l is a cube. The superscript $\vec{\varepsilon}$ reflects the fact that multiple wavelets are associated to any dyadic cube, like the Haar functions case that appeared in Section 2.1. The fact that the supremum admits all open sets of finite measure cannot be omitted, as Carleson's example shows [Car74]. This fact is responsible for some of the difficulties encountered when working with this space.

In this thesis, rather than the above two approaches, we choose to view product BMO as an average of its dyadic counterparts. Similarly as in the one-parameter case, it is proved by Pipher-Ward in [PW08] and by Treil in [Tre09] from a different approach that in the multi-parameter setting, a function is in product BMO if and only if it is in dyadic product BMO uniformly with respect to any dyadic grids. Dyadic product BMO can be defined in the same way as in (2.11) with the wavelets replaced by Haar functions. Moreover, it is the dual of dyadic H^1_{prod}, which can be characterized via the L^1 norm of the multi-parameter dyadic square functions, i.e. tensor products of one-parameter square functions (2.6).

2.2.3 Martikainen's formulation

Recently, the interest in bi-parameter singular integrals has been rekindled by several authors through their study of new bi-parameter $T(1)/T(b)$ type theorems. Among others, we mention the works of Pott-Villarroya [PV13], Martikainen [Mar12] and Ou [Ou15]. In particular, Martikainen studied a class of bi-parameter singular integrals satisfying certain "mixed" type conditions, adapted from a similar formulation in [PV13]. Roughly speaking, the mixed type conditions are the $T(1)$ type conditions that have to be satisfied by the tensor product of Calderón-Zygmund operators. Surprisingly, it turns out that the class of operators characterized in this way is exactly the same as Journé's bi-parameter SIO, a result recently proved by Grau de la Herran in [GdlH15]. More precisely, in the bi-parameter setting, under the additional assumption that T is bounded on L^2, Grau de la Herran shows that T is a Journé type δ-SIO satisfying certain weak boundedness property if and only if it satisfies Martikainen's mixed type conditions.

We remark that in Chapter 7, we will formulate a class of singular integrals in arbitrarily many parameters that is characterized by a collection of mixed type conditions in the same spirit as Martikainen's. It will also be shown there that this class is equivalent to Journé's SIO in the multi-parameter setting. It is interesting that in the bi-parameter case, our result is in fact an improvement over Grau de la Herran's theorem, in the sense that no L^2 boundedness needs to be assumed for the operators. This comes from an observation that in the proof of [GdlH15], the L^2 boundedness is only used to derive the weak boundedness property, while doesn't come into play in the equivalence between kernel assumptions.

In the following, we give the characterizing conditions of Martikainen's operators, which are now known to also characterize Journé's bi-parameter SIO. In fact, the reader can easily see from the formulation below that these conditions are in spirit simply the "tensor product" of the Calderón-Zygmund operator conditions in one parameter.

Definition 2.12. A linear operator T continuously mapping $C^\infty(\mathbb{R}^n) \times C^\infty(\mathbb{R}^m)$ to its dual is said to be in *Maritkainen's class* if the following *full kernel* and *partial kernel* conditions are satisfied.

Full kernel conditions

For $f = f_1 \otimes f_2, g = g_1 \otimes g_2$ with $f_1, g_1 : \mathbb{R}^n \to \mathbb{C}, f_2, g_2 : \mathbb{R}^m \to \mathbb{C}$, and $\operatorname{spt} f_i \cap \operatorname{spt} g_i = \emptyset$, $i = 1, 2$, we have

$$\langle Tf, g\rangle = \int_{\mathbb{R}^{n+m}} \int_{\mathbb{R}^{n+m}} K(x, y) f(y) g(x)\, dxdy$$

such that

(1) Size condition: $|K(x, y)| \leq C \frac{1}{|x_1-y_1|^n} \frac{1}{|x_2-y_2|^m}$.

(2) Hölder conditions:

1. $|K(x, y) - K(x, (y_1, y_2')) - K(x, (y_1', y_2)) + K(x, y')| \leq C \frac{|y_1-y_1'|^\delta}{|x_1-y_1|^{n+\delta}} \frac{|y_2-y_2'|^\delta}{|x_2-y_2|^{m+\delta}}$ if $|y_1 - y_1'| \leq |x_1 - y_1|/2$ and $|y_2 - y_2'| \leq |x_2 - y_2|/2$,
2. $|K(x, y) - K((x_1, x_2'), y) - K((x_1', x_2), y) + K(x', y)| \leq C \frac{|x_1-x_1'|^\delta}{|x_1-y_1|^{n+\delta}} \frac{|x_2-x_2'|^\delta}{|x_2-y_2|^{m+\delta}}$ if $|x_1 - x_1'| \leq |x_1 - y_1|/2$ and $|x_2 - x_2'| \leq |x_2 - y_2|/2$,
3. $|K(x, y) - K((x_1, x_2'), y) - K(x, (y_1', y_2)) + K((x_1, x_2'), (y_1', y_2))|$ $\leq C \frac{|y_1-y_1'|^\delta}{|x_1-y_1|^{n+\delta}} \frac{|x_2-x_2'|^\delta}{|x_2-y_2|^{m+\delta}}$ if $|y_1 - y_1'| \leq |x_1 - y_1|/2$ and $|x_2 - x_2'| \leq |x_2 - y_2|/2$,
4. $|K(x, y) - K(x, (y_1, y_2')) - K((x_1', x_2), y) + K((x_1', x_2), (y_1, y_2'))|$ $\leq C \frac{|x_1-x_1'|^\delta}{|x_1-y_1|^{n+\delta}} \frac{|y_2-y_2'|^\delta}{|x_2-y_2|^{m+\delta}}$ if $|x_1 - x_1'| \leq |x_1 - y_1|/2$ and $|y_2 - y_2'| \leq |x_2 - y_2|/2$.

(3) Mixed Hölder and size conditions:

1. $|K(x,y)-K((x_1',x_2),y)| \le C\frac{|x_1-x_1'|^\delta}{|x_1-y_1|^{n+\delta}}\frac{1}{|x_2-y_2|^m}$ if $|x_1-x_1'|\le |x_1-y_1|/2$,

2. $|K(x,y)-K(x,(y_1',y_2))| \le C\frac{|y_1-y_1'|^\delta}{|x_1-y_1|^{n+\delta}}\frac{1}{|x_2-y_2|^m}$ if $|y_1-y_1'|\le |x_1-y_1|/2$,

3. $|K(x,y)-K((x_1,x_2'),y)| \le C\frac{1}{|x_1-y_1|^{n}}\frac{|x_2-x_2'|^\delta}{|x_2-y_2|^{m+\delta}}$ if $|x_2-x_2'|\le |x_2-y_2|/2$,

4. $|K(x,y)-K(x,(y_1,y_2'))| \le C\frac{1}{|x_1-y_1|^{n}}\frac{|y_2-y_2'|^\delta}{|x_2-y_2|^{m+\delta}}$ if $|y_2-y_2'|\le |x_2-y_2|/2$.

Partial kernel conditions

If $f=f_1\otimes f_2$ and $g=g_1\otimes g_2$ with $\operatorname{spt} f_1\cap \operatorname{spt} g_1=\emptyset$. Then assume

$$\langle Tf,g\rangle = \int_{\mathbb{R}^n}\int_{\mathbb{R}^n} K_{f_2,g_2}(x_1,y_1)f_1(y_1)g_1(x_1)\,dx_1dy_1.$$

The kernel $K_{f_2,g_2}:(\mathbb{R}^n\times\mathbb{R}^n)\setminus\{x_1=y_1\}$ satisfies:

1. $|K_{f_2,g_2}(x_1,y_1)|\le C(f_2,g_2)\frac{1}{|x_1-y_1|^n}$

2. $|K_{f_2,g_2}(x_1,y_1)-K_{f_2,g_2}(x_1',y_1)|\le C(f_2,g_2)\frac{|x_1-x_1'|^\delta}{|x_1-y_1|^{n+\delta}}$ if $|x_1-x_1'|\le|x_1-y_1|/2$,

3. $|K_{f_2,g_2}(x_1,y_1)-K_{f_2,g_2}(x_1,y_1')|\le C(f_2,g_2)\frac{|y_1-y_1'|^\delta}{|x_1-y_1|^{n+\delta}}$ if $|y_1-y_1'|\le|x_1-y_1|/2$.

Moreover, for any cube $V\subset\mathbb{R}^m$, $C(\chi_V,\chi_V)+C(\chi_V,u_V)+C(u_V,\chi_V)\le C|V|$, whenever u_V is V-adapted with zero-mean (i.e. $\operatorname{spt} u_V\subset V, |u_V|\le 1, \int u_V=0$). Symmetrical representation and properties with a kernel K_{f_1,g_1} in the case $\operatorname{spt} f_2\cap\operatorname{spt} g_2=\emptyset$ hold true as well.

Before ending this section, note that from the formulation given above, it is unclear whether it can be extended to the setting with more parameters to characterize Journé's SIO. In Chapter 7 of the thesis, we will show that such extension is doable by giving a group of mixed type conditions that characterize Journé's SIO in arbitrarily many parameters.

2.3 Representation theorems

Now that we have introduced the class of one-parameter and multi-parameter singular integrals that are relevant to this thesis, as a crucial element in many of our arguments, a representation formula will be discussed in this section. One of the most powerful tools that have emerged in the last fifteen years in singular integral theory is a representation of arbitrary singular integrals as averages of simpler components. Such components are built from dyadic basis on Hilbert or other L^p spaces in terms of the well known martingale theory, in particular, Haar functions. Recall that cancellative Haar functions $\{h_I^\epsilon\}$ form an orthonormal basis of $L^2(\mathbb{R}^d)$. The first instance of such representation type result was derived by Petermichl [Pet00], where she proved that Hilbert transform can be obtained by averaging the translations and dilations of the so-called *Haar shifts*

$$Sf := \sum_{I\in\mathcal{D}} \frac{1}{\sqrt{2}} \langle f, h_I^0\rangle h_{I^-}^0.$$

Compared with the Haar expansion, the Haar shift can be viewed as a dyadic operator that rearranges the Haar coefficients of a given L^2 function, with the rearrangement happening only between two adjacent generations of intervals.

Following this idea, Hytönen introduced the notion of *dyadic shift* of arbitrary complexity (i, j), which generalizes Haar shift (a dyadic shift with complexity $(0, 1)$ in this terminology). By averaging over all complexities, Hytönen [Hyt12] generalized Petermichl's result by deriving a representation formula for general CZ operators as averages of dyadic shifts, which then led to numerous advances in singular integral theory, with an example being the proof of the A_2 theorem. This representation was then generalized to the bi-parameter setting by Martikainen [Mar12]. In the following, we will carefully define dyadic shifts and state Hytönen's and Martikainen's representation theorems. In Chapter 7, we will further generalize the idea of dyadic shifts and prove a representation theorem in arbitrarily many parameters. All the representation results will be very useful to us in the estimates of commutators in Chapters 4 and 5.

2.3.1 Shifted grids and dyadic shifts

Recall that while the standard dyadic grid is defined as

$$\mathcal{D}^0 := \{2^{-k}([0,1)^d + m) : k \in \mathbb{Z}, m \in \mathbb{Z}^d\},$$

for any parameter $\omega = (\omega_j)_{j\in\mathbb{Z}} \in (\{0,1\}^d)^{\mathbb{Z}}$, one can define an associated shifted dyadic grid as

$$\mathcal{D}^\omega := \{I \dot{+} \omega : I \in \mathcal{D}^0\}$$

where

$$I \dot{+} \omega := I + \sum_{j:2^{-j}<\ell(I)} 2^{-j}\omega_j.$$

For a fixed shifted grid $\mathcal{D}^\omega$ and nonnegative integers i, j, a *dyadic shift operator* S_ω^{ij} *of complexity* (i,j) is defined to be the following and is assumed to be bounded on L^2 with operator norm less than 1:

$$S_\omega^{ij} f := \sum_{K\in\mathcal{D}^\omega} \sum_{\substack{I\in\mathcal{D}^\omega, I\subset K \\ \ell(I)=2^{-i}\ell(K)}} \sum_{\substack{J\in\mathcal{D}^\omega, J\subset K \\ \ell(J)=2^{-j}\ell(K)}} a_{IJK}\langle f, h_I\rangle h_J =: \sum_K \sum_{I,J\subset K}^{(i,j)} a_{IJK}\langle f, h_I\rangle h_J,$$

with $|a_{IJK}| \le |I|^{1/2}|J|^{1/2}/|K|$. S_ω^{ij} is called *cancellative* if all the Haar functions in the definition are cancellative, otherwise, it is called *non-cancellative.* Note that in d dimensions, each cube $I = I_1 \times \cdots \times I_d$ is associated with 2^d Haar functions h_I^ϵ. In the definition of S_ω^{ij}, we have suppressed the summation over ϵ. And we will usually do this throughout the thesis to abbreviate the notation. Moreover, the bounds of a_{IJK} in fact already imply that all cancellative dyadic shifts are bounded on L^2 with norm less than 1.

Now we look at the bi-parameter dyadic shifts, which are defined as

$$S_{\omega_1,\omega_2}^{i_1 i_2 j_1 j_2} f := \sum_{K_1\in\mathcal{D}^{\omega_1}} \sum_{I_1,J_1\subset K_1}^{(i_1,j_1)} \sum_{K_2\in\mathcal{D}^{\omega_2}} \sum_{I_2,J_2\subset K_2}^{(i_2,j_2)} a_{I_1J_1K_1I_2J_2K_2}\langle f, h_{I_1}\otimes u_{I_2}\rangle h_{J_1}\otimes u_{J_2}$$

where we have used u_I to denote Haar functions in the second variable. We also assume that $|a_{I_1J_1K_1}| \leq |I_1|^{1/2}|J_1|^{1/2}|I_2|^{1/2}|J_2|^{1/2}/|K_1||K_2|$, which implies that all cancellative bi-parameter dyadic shifts are bounded on L^2 with norm less than 1.

2.3.2 One-parameter and bi-parameter representation formulas

We now introduce Hytönen's representation theorem in the one-parameter setting. Interested readers can find its proof and a more detailed discussion in [Hyt12] and [Hyt11]. The operator T mentioned in the following will be a Calderón-Zygmund operator associated with a δ-standard kernel K. Hytönen in [Hyt11] proved the following theorem:

Theorem 2.13. *Let T be a Calderón-Zygmund operator, then it has an expansion, say for $f, g \in C_0^\infty(\mathbb{R}^d)$,*

$$\langle g, Tf \rangle = c \cdot \|T\|_{CZ} \cdot \mathbb{E}_\omega \sum_{i,j=0}^{\infty} 2^{-\max(i,j)\delta/2} \langle g, S_\omega^{ij} f \rangle,$$

where c is a dimensional constant and S_ω^{ij} is a dyadic shift of parameter (i,j) on the dyadic grid $\mathcal{D}^\omega$; all of them except possibly S_ω^{00} are cancellative.

According to the proof of Theorem 2.13, in the representation of any T, only S_ω^{00} may be non-cancellative, and if this is the case, only one of $\{h_I\}, \{h_J\}$ in its definition is non-cancellative, i.e. S_ω^{00} is a paraproduct with some symbol a with the coefficients $a_I := a_{III} = \langle a, h_I \rangle |I|^{-1/2}$. It's also implied from the proof that a will be a BMO function satisfying $\|a\|_{\text{BMO}} \leq 1$.

Recall that a dyadic paraproduct with symbol b is an operator defined as

$$B_0(b, f) := \sum_{I \in \mathcal{D}} \beta_I \langle b, h_I^{\epsilon_1} \rangle \langle f, h_I^{\epsilon_2} \rangle h_I^{\epsilon_3} |I|^{-1/2},$$

where $\epsilon_j \in \{0,1\}^d$ and $\{\beta_I\}_I$ is a sequence satisfying $|\beta_I| \leq 1$ uniformly. If there is at most one of $j = 2, 3$ such that $\epsilon_j = \vec{1}$, then the operator B_0 satisfies

$$\|B_0(b, f)\|_{L^p} \lesssim \|b\|_{\mathrm{BMO}}\|f\|_{L^p}, \quad \forall 1 < p < \infty,$$

a fact that can be seen via Carleson embedding theorem or a square function argument. This implies in particular that non-cancellative dyadic shifts S_ω^{00} appearing in the representation formula are always bounded on L^2 with norm less than 1.

Now we move on to the bi-parameter representation theorem proved in [Mar12], which is Martikainen's original motivation of formulating the mixed type conditions on the bi-parameter singular integrals.

Theorem 2.14. *For a bi-parameter singular integral T in Martikainen's class and satisfies the boundedness and cancellation assumptions in the following, there holds for some bi-parameter shifts $S_{\omega_1,\omega_2}^{i_1 i_2 j_1 j_2}$ that*

$$\langle Tf, g\rangle = C\mathbb{E}_{\omega_1}\mathbb{E}_{\omega_2}\sum_{i_1,i_2,j_1,j_2\geq 0} 2^{-\max(i_1,i_2)\frac{\delta}{2}}2^{-\max(j_1,j_2)\frac{\delta}{2}}\langle S_{\omega_1,\omega_2}^{i_1 i_2 j_1 j_2} f, g\rangle,$$

where non-cancellative shifts may only appear if $(i_1, i_2) = (0, 0)$ or $(j_1, j_2) = (0, 0)$.

The boundedness and cancellation assumptions mentioned in the theorem above are the following:

(1) $T1, T^*1, T_1(1), T_1^*(1) \in \mathrm{BMO}_{\mathrm{prod}}(\mathbb{R}^n \times \mathbb{R}^m)$;

(2) WBP: $|\langle T(\chi_K \otimes \chi_V), \chi_K \otimes \chi_V\rangle| \leq C|K||V|$ for every cube $K \subset \mathbb{R}^n, V \subset \mathbb{R}^m$;

(3) Diagonal BMO conditions: for any cube $K \subset \mathbb{R}^n, V \subset \mathbb{R}^m$, and every zero-mean functions a_K and b_V which are K and V adapted respectively:

1. $|\langle T(a_K \otimes \chi_V), \chi_K \otimes \chi_V\rangle| \leq C|K||V|$,
2. $|\langle T(\chi_K \otimes \chi_V), a_K \otimes \chi_V\rangle| \leq C|K||V|$,

3. $|\langle T(\chi_K \otimes b_V), \chi_K \otimes \chi_V\rangle| \leq C|K||V|$,

4. $|\langle T(\chi_K \otimes \chi_V), \chi_K \otimes b_V\rangle| \leq C|K||V|$.

According to the proof of the theorem, all non-cancellative bi-parameter shifts S^{0000} are bi-parameter dyadic paraproducts for some product BMO symbol function. Recall that a multi-parameter paraproduct associated with function b can be viewed as a bilinear operator which is defined as

$$B_0(b,f) := \sum_{R\in\mathcal{D}_{\vec{d}}} \beta_R \langle b, h_R^{\epsilon_1}\rangle \langle f, h_R^{\epsilon_2}\rangle h_R^{\epsilon_3} |R|^{-1/2}, \tag{2.15}$$

where $\epsilon_j \in \{0,1\}^{\vec{d}}$, $\mathcal{D}_{\vec{d}}$ denotes the tensor product of dyadic grids, and $\{\beta_R\}_R$ is a sequence satisfying $|\beta_R| \leq 1$. Note that $h_R^{\epsilon_j}$ is cancellative if and only if $\epsilon_j \neq \vec{1}$. According to Journé [Jou85] and later on improved by C. Muscalu, J. Pipher, T. Tao and C. Thiele [MPTT04, MPTT06], one has the following boundedness result.

Theorem 2.16. *Let $\vec{d} = (d_1, \ldots, d_t)$ and $\epsilon_j = (\epsilon_{j,1}, \ldots, \epsilon_{j,t})$. If $\epsilon_1 \neq \vec{1}$ and $\forall 1 \leq s \leq t$, there is at most one of $j = 2, 3$ such that $\epsilon_{j,s} = \vec{1}$, then the operator B_0 satisfies*

$$B_0 : BMO_{prod}(\mathbb{R}^{\vec{d}}) \times L^p(\mathbb{R}^{\vec{d}}) \to L^p(\mathbb{R}^{\vec{d}}), \quad \forall 1 < p < \infty.$$

Therefore, all the S^{0000} that might appear in the representation are automatically bounded on L^2 with norm less than 1.

Before the end of this section, we remark that representation theorems yield $T(1)$ theorems as direct corollaries, since dyadic shifts are uniformly bounded on L^2. We will prove a multi-parameter version of the representation theorem in Chapter 7, which also implies a $T(1)$ theorem in arbitrarily many parameters.

Chapter Three

Commutators of Singular Integrals

3.1 Historical remark

A classical result of Nehari [Neh57] shows that a Hankel operator with anti-analytic symbol b mapping analytic functions into the space of anti-analytic functions by $f \mapsto P_{-}bf$ is bounded with respect to an L^2 norm if and only if the symbol belongs to BMO, i.e.

$$\|b\|_{\mathrm{BMO}} := \sup_{Q\subset\mathbb{R}} \frac{1}{|Q|}\int_Q |b(x) - \langle b\rangle_Q|\,dx < \infty,$$

where $\langle b\rangle_Q := |Q|^{-1}\int_Q b(x)\,dx$. This theorem has an equivalent formulation in terms of the boundedness of the commutator of the multiplication operator with symbol function b and the Hilbert transform $[b, H] := bH - Hb$. To be specific, the theorem can be rephrased as that

$$\|b\|_{\mathrm{BMO}(\mathbb{R})} \approx \|[b, H]\|_{L^2(\mathbb{R})\to L^2(\mathbb{R})}. \tag{3.1}$$

The control of the operator norm of the commutator by the BMO norm of the symbol is often referred to as the "upper bound estimate", while the other direction of the boundedness is called the "lower bound estimate".

Ferguson-Sadosky in [FS00] and later Ferguson-Lacey in their groundbreaking paper [FL02] study the symbols of bounded "big" and "little" Hankel operators on the bidisk through commutators of the tensor product or iterated form

$$[b, H_1H_2],\ \text{and}\ [[b, H_1], H_2].$$

Here $b = b(x_1, x_2)$ and H_k are the Hilbert transforms acting in the k^{th} variable. A full characterization of different bi-parameter BMO spaces: Cotlar-Sadosky's little BMO (denoted by bmo) and Chang-Fefferman's product BMO space (see Chapter 2 for definition), are given through these commutators:

$$\|b\|_{\mathrm{bmo}} \approx \|[b, H_1H_2]\|_{L^2\to L^2}; \tag{3.2}$$

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod}}} \approx \|[[b, H_1], H_2]\|_{L^2 \to L^2}. \tag{3.3}$$

And it is shown by Lacey-Terwilleger [LT09] that (3.3) holds true in the multi-parameter setting as well.

Generally, a locally integrable function $b : \mathbb{R}^{\vec{d}} = \mathbb{R}^{d_1} \times \ldots \times \mathbb{R}^{d_s} \to \mathbb{C}$ is in bmo (*little BMO*) if and only if

$$\|b\|_{\mathrm{bmo}} := \sup_{\vec{Q}=Q_1 \times \cdots \times Q_s} |\vec{Q}|^{-1} \int_{\vec{Q}} |b(\vec{x}) - b_{\vec{Q}}| < \infty.$$

Here the Q_k are d_k-dimensional cubes and $b_{\vec{Q}}$ denotes the average of b over $\vec{Q}$. It is easy to see that this space consists of all functions that are uniformly in BMO in each variable separately. Let $\vec{x}_{\hat{v}} = (x_1, \ldots, x_{v-1}, \cdot, x_{v+1}, \ldots, x_s)$. Then $b(\vec{x}_{\hat{v}})$ is a function in x_v only with the other variables fixed. Its BMO norm in x_v is

$$\|b(\vec{x}_{\hat{v}})\|_{\mathrm{BMO}} = \sup_{Q_v} |Q_v|^{-1} \int_{Q_v} |b(\vec{x}) - b(\vec{x}_{\hat{v}})_{Q_v}| dx_v$$

and the little BMO norm becomes

$$\|b\|_{\mathrm{bmo}} = \max_{v} \{ \sup_{\vec{x}_{\hat{v}}} \|b(\vec{x}_{\hat{v}})\|_{\mathrm{BMO}} \}.$$

On the bi-disk, this becomes

$$\|b\|_{\mathrm{bmo}} = \max\{ \sup_{x_1} \|b(x_1, \cdot)\|_{\mathrm{BMO}}, \sup_{x_2} \|b(\cdot, x_2)\|_{\mathrm{BMO}} \},$$

the space that appears in (3.2).

The above results all lean on analytic structure in a crucial way in that they utilize the fact that Hilbert transform is a Fourier projection operator. Through the use of completely different real variable methods, Coifman, Rochberg and Weiss [CRW76] extended Nehari's one-parameter theory to real analysis in the sense that the Hilbert transform was replaced by Riesz transforms. They proved that for any $1 < p < \infty$,

$$\|b\|_{\mathrm{BMO}(\mathbb{R}^d)} \approx \sup_{1 \le j \le d} \|[b, R_j]\|_{L^p(\mathbb{R}^d) \to L^p(\mathbb{R}^d)}.$$

These one-parameter results in [CRW76] were settled in the multi-parameter setting in Lacey-Petermichl-Pipher-Wick [LPPW09]. They demonstrated that for $1 < p < \infty$,

$$\|b\|_{\mathrm{BMO}_{\mathrm{prod}}} \approx \sup_{\vec{j}} \|[\ldots[[b, R_{1,j_1}], R_{2,j_2}], \ldots, R_{t,j_t}]\|_{L^p(\mathbb{R}^{\vec{d}}) \to L^p(\mathbb{R}^{\vec{d}})},$$

where $\vec{j} = (j_1, \ldots, j_t)$ with $1 \le j_s \le d_s$ for $s = 1, \ldots, t$. Both the upper and lower estimate have proofs very different from those in one parameter. In fact, the upper estimate they obtained in the article applies to more general Calderón-Zygmund operators of convolution type with smooth enough kernels. Later, in [LPPW10], they further simplified the upper bound proof for Riesz transforms via the approach of the use of Haar shifts, an idea that motivated the work in Chapter 4.

Outside of the scope of Riesz transforms (or Hilbert transform, in the one-dimensional case), Uchiyama [Uch82] established a very technical and constructive proof of the one-parameter result of [CRW76], where Riesz transforms are replaced by more general classes of kernel operators obeying certain point separation criterion for their Fourier multiplier symbols. Dalenc and Petermichl then extended this direction to several parameters in [DP14], where they characterized product BMO via more general families of singular integrals satisfying similar conditions of Uchiyama type.

Apart from the connection with Hankel operators, commutator estimates of this type establish a natural connection between function spaces of bounded mean oscillation and singular integrals. In particular, it provides a way to construct and study such spaces from the viewpoint of commutators of singular integrals. One simple example is the following observation:

Observation 3.4. If $f_1(x), f_2(y) \in \mathrm{BMO}(\mathbb{R})$, then $f_1 \otimes f_2 \in \mathrm{BMO}_{\mathrm{prod}}(\mathbb{R} \times \mathbb{R})$.

Although this fact is not easy to see directly from the definition of BMO spaces, with the commutator results we stated above, it can be easily proved by splitting the iterated commutator into tensor products:

$$[[f_1 \otimes f_2, H_1], H_2] = [f_1, H_1] \otimes [f_2, H_2].$$

Moreover, it can be seen from the commutators viewpoint that little BMO is contained in product BMO. Indeed, $b(x_1, x_2) \in \mathrm{bmo}(\mathbb{R}^2)$ implies by definition of little BMO and (3.1) that $[b(\cdot, x_2), H_1]$ is uniformly (in x_2) bounded on $L^2(\mathbb{R}^2)$, which then means that the iterated commutator $[[b, H_1], H_2]$ is also bounded. Hence, $b \in \mathrm{BMO}_{\mathrm{prod}}(\mathbb{R}^2)$ according to (3.3). The same result holds true in higher dimensions as well.

Another reason why we are interested in commutator estimates is that they usually imply weak factorization results of Hardy spaces. It is a famous result of C. Fefferman and Stein [FS72] that $\mathrm{BMO}(\mathbb{R}^d)$ is the dual space of the (real) one-parameter Hardy space $H^1(\mathbb{R}^d)$, the class of functions with the norm

$$\sum_{j=0}^{d} \|R_j f\|_{L^1},$$

where R_0 denotes the identity operator. Based on this, as a corollary to the Riesz commutator result, the following theorem is shown in [CRW76]:

Theorem 3.5. *If g_1, g_2 are in $L^2(\mathbb{R}^d)$ then for any $j = 1, \ldots, d$, $f = g_1 R_j(g_2) + R_j(g_1) g_2$ is in $H^1(\mathbb{R}^d)$ and $\|f\|_{H^1} \lesssim \|g_1\|_{L^2} \|g_2\|_{L^2}$. Conversely, every f in $H^1(\mathbb{R}^d)$ can be written as*

$$f = \sum_{i=1}^{\infty} \sum_{j=1}^{d} (g_i^j R_j h_i^j + h_i^j R_j g_i^j)$$

with

$$\sum_{i,j} \|g_i^j\|_{L^2} \|h_i^j\|_{L^2} \lesssim \|f\|_{H^1}.$$

Similar weak factorization results of multi-parameter Hardy spaces are obtained in

[FL02, LPPW09] as well.

Moreover, commutator results are closely related with Div-Curl estimates. Let $E, B \in L^2(\mathbb{R}^n; \mathbb{R}^n)$ be two vector fields on $\mathbb{R}^n$ taking values in $\mathbb{R}^n$. Suppose that

$$\mathrm{div} E(x) = 0, \quad \mathrm{curl} B(x) = 0,$$

then it is shown in [CLMS93] that $E \cdot B \in H^1(\mathbb{R}^n)$ with

$$\|E \cdot B\|_{H^1} \lesssim \|E\|_{L^2} \|B\|_{L^2}.$$

This result can be easily deduced from the aforementioned Riesz commutator theorem in [CRW76] using the duality between BMO and H^1. The extension of this result to the multi-parameter setting is obtained in [LPPW12], where the crucial ingredient of the proof is the iterated Riesz commutator estimate in [LPPW09].

3.2 Main results

In this thesis, we are interested in a more systematic and complete study of iterated commutators with general singular integrals, together with the different BMO type spaces that they characterize.

First of all, we investigate the iterated commutators of "mixed" type. Let's look at the following three dimensional case as an example. Let H_k denote the Hilbert transform acting in the k^{th} variable, $k = 1, 2, 3$. One would like to know what kind of BMO space the mixed commutator $[[b, H_1 H_3], H_2]$ characterizes. Define $\mathrm{BMO}_{(13)(2)}$ as the space of functions $b(x_1, x_2, x_3)$ such that $b(x_1, \cdot, \cdot)$ and $b(\cdot, \cdot, x_3)$ are uniformly in product BMO in the remaining two variables, normed by the supremum of the product BMO norms. Then,

Theorem 3.6. *Let $b \in L^1(\mathbb{T}^3)$. Then the following are equivalent with linear dependence on the respective norms*

(1) $b \in BMO_{(13)(2)}(\mathbb{T}^3)$

(2) The commutators $[[b, H_1], H_2]$ *and* $[[b, H_3], H_2]$ *are bounded on* $L^2(\mathbb{T}^3)$

(3) The commutator $[[b, H_1H_3], H_2]$ *is bounded on* $L^2(\mathbb{T}^3)$.

Corollary 3.7. *We have the following two-sided estimate*

$$\|b\|_{BMO_{(13)(2)}} \lesssim \|[[b, H_1H_3], H_2]\|_{L^2(\mathbb{T}^3)\to L^2(\mathbb{T}^3)} \lesssim \|b\|_{BMO_{(13)(2)}}.$$

Comparing this result to (3.2), (3.3) in Section 3.1, one can easily see that it coincides with the intuition that in the structure of the commutator, iteration relates with product BMO, while tensor product of singular integrals relates with supremum of BMO norms. Indeed, it turns out that Theorem 3.6 is a special case of a grander scheme of commutator estimates of this type, which yields a whole new family of BMO spaces in the multi-parameter setting.

Definition 3.8. Let $b : \mathbb{R}^{\vec{d}} \to \mathbb{C}$ with $\vec{d} = (d_1, \cdots, d_t)$. Take a partition $\mathcal{I} = \{I_s : 1 \le s \le l\}$ of $\{1, 2, \ldots, t\}$ so that $\dot\cup_{1\le s\le l} I_s = \{1, 2, \ldots, t\}$. We say that $b \in \mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ if for any choices $\boldsymbol{v} = (v_s), v_s \in I_s$, b is uniformly in product BMO in the variables indexed by $\{v_s\}$. We call a BMO space of this type a **little product BMO**. If for any $\vec{x} = (x_1, \ldots, x_t) \in \mathbb{R}^{\vec{d}}$, we define $\vec{x}_{\hat{\boldsymbol{v}}}$ by removing those variables indexed by v_s, the little product BMO norm becomes

$$\|b\|_{\mathrm{BMO}_{\mathcal{I}}} = \max_{\boldsymbol{v}}\{\sup_{\vec{x}_{\hat{\boldsymbol{v}}}} \|b(\vec{x}_{\hat{\boldsymbol{v}}})\|_{\mathrm{BMO}}\}$$

where the BMO norm is product BMO in the variables indexed by v_s.

For example, when $\vec{d} = (1, 1, 1) = \vec{1}$, when $t = 3$ and $l = 2$ with $I_1 = (13)$ and $I_2 = (2)$, writing $\mathcal{I} = (13)(2)$ the space $\mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{1}})$ arises, which consists of those functions that are uniformly in product BMO in the variables $(1, 2)$ and $(3, 2)$ respectively, as described above. Moreover, as degenerate cases, it is easy to see that the two endpoint spaces $\mathrm{BMO}_{(12\ldots t)}$ and $\mathrm{BMO}_{(1)(2)\ldots(t)}$ are exactly little BMO and Chang-Fefferman's product BMO respectively. In fact, we will show in Proposition 6.1 in Chapter 6 that the little product BMO spaces can be partially ordered, with little BMO contained in their intersection and product BMO containing their union. Little product BMO spaces on $\mathbb{T}^{\vec{d}}$ can be defined in the same way.

Suppose we are in $\mathbb{R}^{\vec{d}}$ with $\vec{d} = (d_1, \ldots, d_t)$ and a partition $\mathcal{I} = (I_s)_{1\leq s\leq l}$ of $\{1, \ldots, t\}$ is fixed. It is our aim to prove the following characterization theorem of the space $\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$.

Theorem 3.9. *The following are equivalent with linear dependence of the respective norms.*

(1) $b \in BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$

(2) All commutators of the form $[\ldots[b, R_{k_1,j_{k_1}}], \ldots, R_{k_l,j_{k_l}}]$ *are bounded in* $L^2(\mathbb{R}^{\vec{d}})$ *where* $k_s \in I_s$ *and* $R_{k_s,j_{k_s}}$ *is the one-parameter Riesz transform in direction* j_{k_s}*.*

(3) All commutators of the form $[\ldots[b, \vec{R}_{1,\vec{j}^{(1)}}], \ldots, \vec{R}_{l,\vec{j}^{(l)}}]$ *are bounded in* $L^2(\mathbb{R}^{\vec{d}})$ *where* $\vec{j}^{(s)} = (j_k)_{k\in I_s}$*,* $1 \leq j_k \leq d_k$ *and the operators* $\vec{R}_{s,\vec{j}^{(s)}}$ *are a tensor product of Riesz transforms* $\vec{R}_{s,\vec{j}^{(s)}} = \bigotimes_{k\in I_s} R_{k,j_k}$*.*

From the inductive nature of our arguments, it will also be apparent that the characterization holds when we consider intermediate cases, meaning commutators with any fixed number of Riesz transforms in each iterate.

Below we state our most general two-sided estimate through Riesz transforms.

Theorem 3.10. *Let* $1 < p < \infty$*. Under the same assumptions as Corollary 3.11 and for any fixed* $\vec{n} = (n_s)$ *where* $1 \leq n_s \leq |I_s|$*, we have the two-sided estimate*

$$\|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \lesssim \sup_{\vec{j}} \|[\ldots[b, \vec{R}_{1,\vec{j}^{(1)}}], \ldots, \vec{R}_{l,\vec{j}^{(l)}}]\|_{L^p(\mathbb{R}^{\vec{d}})\to L^p(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}$$

where $\vec{j}^{(s)} = (j_k)_{k\in I_s}$*,* $0 \leq j_k \leq d_k$ *and for each* s*, there are* n_s *non-zero choices. A Riesz transform in direction* 0 *is understood as the identity.*

For $p = 2$ and $\vec{n} = \vec{1}$ this is the equivalence *(1)* $\Leftrightarrow$ *(2)* and for $\vec{n} = (|I_1|, \ldots, |I_l|)$ it is the equivalence *(1)* $\Leftrightarrow$ *(3)* from Theorem 3.9.

Our main focus is of course on a two-sided estimate when $\vec{n} = (|I_1|, \ldots, |I_l|)$:

Corollary 3.11. *Let $\vec{j} = (j_1, \ldots, j_t)$ with $1 \le j_k \le d_k$ and let for each $1 \le s \le l$, $\vec{j}^{(s)} = (j_k)_{k \in I_s}$ be associated a tensor product of Riesz transforms $\vec{R}_{s,\vec{j}^{(s)}} = \bigotimes_{k \in I_s} R_{k,j_k}$; here R_{k,j_k} are j_k^{th} Riesz transforms acting on functions defined on the k^{th} variable. We have the two-sided estimate*

$$\|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \lesssim \sup_{\vec{j}} \|[\ldots[b, \vec{R}_{1,\vec{j}^{(1)}}], \ldots, \vec{R}_{t,\vec{j}^{(t)}}]\|_{L^p(\mathbb{R}^{\vec{d}}) \to L^p(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

The statements above also serve as the statement of the general case for products of Hilbert transforms including Theorem 3.7. In fact, when any $d_k = 1$ just replace the Riesz transforms by the Hilbert transform in that variable. In the rest of the thesis, we will focus on proving Theorem 3.10 and Corollary 3.11 in the case of L^2. It will be remarked at the end of Chapter 5 how to obtain the full L^p estimate.

In the upper bound direction, i.e. to bound the operator norm of the commutator by the BMO norm, we are going to prove a much more general result that applies to not only Riesz transforms but also all CZ operators.

Theorem 3.12. *Let $b \in BMO(\mathbb{R}^{\vec{d}})$ and $(T_k)_{1 \le k \le t}$ be a collection of Calderón-Zygmund operators, with each T_k acting on parameter k of $\mathbb{R}^{\vec{d}} = \mathbb{R}^{d_1} \otimes \cdots \otimes \mathbb{R}^{d_t}$. Then for $1 < p < \infty$,*

$$\|[\ldots[[b, T_1], T_2], \ldots, T_t]\|_{L^p(\mathbb{R}^{\vec{d}}) \to L^p(\mathbb{R}^{\vec{d}})} \le C\|b\|_{BMO}$$

where C depends only on $\vec{d}$ and $\prod_{k=1}^{t} \|T_k\|_{CZ}$.

Observing that in general,

$$[b, T_1 T_2] = [b, T_1]T_2 + T_1[b, T_2],$$

by the boundedness of CZ operators one can deduce:

Corollary 3.13. *Let* $\mathcal{I} = (I_s)_{1\le s\le l}$ *be a partition of* $\{1,\dots,t\}$ *and* $(T_k)_{1\le k\le t}$ *be a collection of Calderón-Zygmund operators, with each* T_k *acting on parameter* k *of* $\mathbb{R}^{\vec{d}} = \mathbb{R}^{d_1}\otimes\cdots\otimes\mathbb{R}^{d_t}$*. Then,*

$$\|[\dots[b,\vec{T}_1],\dots,\vec{T}_l]\|_{L^p(\mathbb{R}^{\vec{d}})\to L^p(\mathbb{R}^{\vec{d}})} \le C\|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}, \quad 1<p<\infty,$$

where $\vec{T}_s = \bigotimes_{k\in I_s} T_k$*,* $1\le s\le l$*, and* C *depends only on* $\vec{d}$ *and* $\prod_{k=1}^t \|T_k\|_{CZ}$*.*

Upper bound estimate is important not only in its own right but can also be reflected in the following two aspects: 1. it is a crucial ingredient in the proof of the lower bound estimate; 2. it yields stability results for characterizing families of BMO spaces. More precisely, the second point is formulated as the following corollary, which can be deduced easily from Corollary 3.13 using triangle inequality.

Corollary 3.14. *Let* $1<p<\infty$*. Let* $\mathcal{I} = (I_s)_{1\le s\le l}$ *be a partition of* $\{1,\dots,t\}$ *and for each* $1\le k\le t$*,* $\mathcal{T}_k = \{T_{k,j_k}\}$ *be a family of Calderón-Zygmund operators acting on parameter* k *of* $\mathbb{R}^{\vec{d}} = \mathbb{R}^{d_1}\otimes\cdots\otimes\mathbb{R}^{d_t}$*. Denote* $\vec{T}_{s,\vec{j}_s} = \bigotimes_{k\in I_s} T_{k,j_k}$ *and suppose* $\{\vec{T}_{s,\vec{j}_s}\}$ *characterizes the space* $BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ *via a two-sided commutator estimate*

$$\|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \lesssim \sup_{\vec{j}_1,\dots,\vec{j}_l} \|[\dots[b,\vec{T}_{1,\vec{j}_1}],\dots,\vec{T}_{l,\vec{j}_l}]\|_{L^p(\mathbb{R}^{\vec{d}})\to L^p(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

Then, $\exists\varepsilon>0$ *such that for any families of Calderón-Zygmund operators* $\{\mathcal{T}'_k\}$ *satisfying* $\|T'_{k,j_k}\|_{CZ}\le\epsilon$*,* $\{\vec{T}_{s,\vec{j}_s} + \vec{T}'_{s,\vec{j}_s}\}$ *still characterizes* $BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$*.*

So far, we have considered new BMO spaces that can be characterized by commutators with multi-parameter singular integrals of tensor product type (a special case of Journé operators that we introduced in Chapter 2) in each component. It is natural to ask whether there exist characterizing families that consist of more general Journé operators, i.e. the multi-parameter Journé type CZ operators discussed in Section 2.2 of Chapter 2. As a second main part of our commutator results, we answer this question affirmatively by studying the upper bound of more general Journé commutators.

In the base case, where there is only a bi-parameter Journé operator T involved and no iteration. Suppose that T is *paraproduct free*, meaning that

$$T(1 \otimes \cdot) = T(\cdot \otimes 1) = T^*(1 \otimes \cdot) = T^*(\cdot \otimes 1) = 0.$$

Theorem 3.15. *Let T be as above and b be a little BMO function, there holds*

$$\|[b, T]\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m) \to L^2(\mathbb{R}^n \times \mathbb{R}^m)} \lesssim \|b\|_{bmo(\mathbb{R}^n \times \mathbb{R}^m)},$$

where the underlying constant depends only on the characterizing constants of T.

In the above, the paraproduct free assumption is a technical one, some partial result on how to remove which will be introduced in Chapter 4. Theorem 3.15 can be extended to $\mathbb{R}^{\vec{d}}$, to arbitrarily many parameters and an arbitrary number of iterates in the commutator. To do this, consider multi-parameter Journé operators mentioned in Chapter 2 and will be discussed in details in Chapter 7, which satisfy a weak boundedness property and are paraproduct free, meaning that any partial adjoint of T is zero if acting on some tensor product of functions with one of the components being 1. And consider a little product BMO function $b \in \mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$. One can then prove

Theorem 3.16. *Let us consider $\mathbb{R}^{\vec{d}}$, $\vec{d} = (d_1, \ldots, d_t)$ with a partition $\mathcal{I} = (I_s)_{1 \le s \le l}$ of $\{1, \ldots, t\}$ as discussed before. Let $b \in BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ and let T_s denote a multi-parameter paraproduct free Journé operator acting on functions defined on $\bigotimes_{k \in I_s} \mathbb{R}^{d_k}$. Then we have the estimate below*

$$\|[\ldots[b, T_1], \ldots, T_l]\|_{L^2(\mathbb{R}^{\vec{d}}) \to L^2(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

The assumption that the operators be paraproduct free is sufficient for our lower estimate in Chapter 5. Note that in the case that all the Journé operators are one-parameter CZ operators or are of tensor product type, the above theorem is directly implied by Theorem 3.12 and Theorem 3.13, and the paraproduct free assumption can be fully removed.

Important to our arguments for lower bounds with Riesz transforms is the corollary below, which follows from the control on the norm of the estimate in Theorem 3.16 by an application of triangle inequality. It is a stability result for characterizing families of Journé operators, and implies that for any little product BMO space, there exist infinitely many collections of characterizing families of Journé operators.

Corollary 3.17. *Let for every $1 \le s \le l$ be given a collection $\mathcal{T}_s = \{T_{s,j_s}\}$ of paraproduct free Journé operators on $\bigotimes_{k\in I_s}\mathbb{R}^{d_k}$ that characterize $BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ via a two-sided commutator estimate*

$$\|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \lesssim \sup_{\vec{j}} \|[T_{1,j_1}, \ldots [T_{l,j_l}, b] \ldots]\|_{L^2(\mathbb{R}^{\vec{d}}) \to L^2(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

Then there exists $\varepsilon > 0$ such that for any family of paraproduct free Journé operators $\mathcal{T}'_s = \{T'_{s,j_s}\}$ with characterizing constants $\|T'_{s,j_s}\|_{CZ} \le \varepsilon$, the family $\{T_{s,j_s} + T'_{s,j_s}\}$ still characterizes $BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$.

Summing up the above characterization results in both one-parameter and multi-parameter cases, it is observed that the Riesz transforms are a representative testing class of multi-parameter Journé operators in the sense that, the boundedness of commutators of Riesz transforms implies that of commutators of Journé operators.

In the next two Chapters, we will give detailed proofs of the results mentioned in this section. Specifically, Chapter 4 deals with general upper bound estimates where Theorem 3.12, 3.13, 3.15, 3.16 will be proved; Chapter 5 is devoted to estimates for Hilbert and Riesz commutators where Theorem 3.6, 3.9 will be fully justified. Moreover, in Appendix A, a simple and more transparent proof of the Hilbert commutator estimates in the base case (bi-parameter, no iteration) will be presented. In addition, in Chapter 6, the little product BMO spaces will be further explored, and weak factorization results of their pre-dual Hardy spaces will be obtained.

Chapter Four

Upper Estimates for General Commutators

In this Chapter, we are interested in upper bounds for general commutator norms by means of little product BMO norms of the symbol. In the special case of the Hilbert transform, we will see in the next Chapter that these estimates, even in the iterated case, are straightforward. Other streamlined proofs exist for Hilbert or Riesz transforms when considering dyadic shifts of complexity one, see [Pet00], [PTV02] and [LPPW10].

When considering more general Calderón-Zygmund or multi-parameter Journé operators, the arguments required are more difficult, in each case. The first classical upper bound goes back to [CRW76], where an estimate for one-parameter commutators with convolution type Calderón-Zygmund operators is given. Next, the text [LPPW09] includes a technical estimate for the multi-parameter case for such Calderón-Zygmund operators with a high enough degree of smoothness. In this thesis, we will first remove the requirement of T being a convolution operator and the smoothness assumption by proving Theorem 3.12 from a different approach using the representation formula for Calderón-Zygmund operators by means of infinite complexity dyadic shifts [Hyt11]. See Sections 4.2 and 4.3. We prove only the L^2 case there, where sharp estimates for some of the basic paraproduct-like operators can be obtained. For $p \neq 2$, the proof of the upper bound will be given in Appendix B.

Second, we will derive the first upper bound estimate concerning multi-parameter Journé operators not of tensor product type (Theorem 3.15, 3.16) in Section 4.4. We will give the complete proof of Theorem 3.15, the bi-parameter base case. The part of the proof that targets the multi-parameter Journé operators with more iterates (Theorem 3.16) proceeds exactly the same as the bi-parameter case with the multi-parameter version of the representation theorem proven in Chapter 7. See Subsection 4.4.2 for a more detailed remark. Although Theorem 3.16 only applies to Journé operators that are paraproduct free, we will show that this assumption can be partially removed in Subsection 4.4.3.

The key common idea in all these proofs is a representation of singular integrals as averages of dyadic shift operators of infinite complexity. The one-parameter and bi-parameter cases of this result are proved by Hytönen [Hyt11] and Martikainen [Mar12], respectively, as discussed in Chapter 2. And the multi-parameter case is proved in Chapter 7. Such representation allows us to reduce the problem to estimating commutators with dyadic shifts. The novelty of this approach to the upper bound is twofold. First, the commutators with dyadic shifts which have infinite complexity in our case, are carefully studied and effectively reduced to paraproduct-like operators. In contrast to typical methods dealing with multi-parameter theory, this allows our argument to be iterated. Second, the paraproduct-like operators together with their boundedness estimates are new, and this is where the delicate estimates in product theory are required.

4.1 Preparation: paraproduct-like operators

In this section, we study the boundedness of several dyadic operators in one and multi-parameter that are variants of dyadic paraproducts. Such results will act as building blocks in the upper bound estimates in later sections.

4.1.1 One-parameter case

Let $a, b \in \mathrm{BMO}(\mathbb{R}^d)$ and $\mathcal{D}$ be a fixed dyadic grid. The first operator we study is the bilinear operator B_k, which could be viewed as a generalized dyadic paraproduct *with descendants*:

$$B_k(b, f) := \sum_{I \in \mathcal{D}} \beta_I \langle b, h_{I^{(k)}} \rangle \langle f, h_I^{\epsilon} \rangle h_I^{\epsilon'} |I^{(k)}|^{-1/2},$$

where $\{\beta_I\}_I$ is a sequence satisfying $|\beta_I| \leq 1$, $k \geq 0$ is an arbitrary integer, and $I^{(k)}$ denotes the k-th dyadic ancestor of I. ϵ and ϵ' cannot be $\vec{1}$ simultaneously, i.e. the Haar functions $\{h_I^{\epsilon}\}$ and $\{h_I^{\epsilon'}\}$ cannot both be non-cancellative. Note that when $k = 0$, this is exactly the classical paraproduct that we have introduced in Chapter 2, which is bounded on L^2 when $b \in \mathrm{BMO}$. Lemma 4.1 below shows that such boundedness holds true for any B_k uniformly in $k > 0$, when all Haar functions that appear are cancellative, which is the only case needed in our application. We will discuss the case when $k > 0$ while non-cancellative Haar functions do appear in Appendix B. In fact, the best known dependence of the constant on k so far in this case is $k^{1/2}$.

The second operator is the trilinear operator P defined as

$$P(b, a, f) := \sum_{I \in \mathcal{D}} \beta_I \langle b, h_I \rangle \langle f, h_I \rangle |I|^{-1} \sum_{J : J \subsetneq I} \langle a, h_J \rangle h_J,$$

where $\{\beta_I\}_I$ is again a sequence satisfying $|\beta_I| \leq 1$ and all Haar functions that appear are cancellative. P will be proved to be bounded on $\text{BMO} \times \text{BMO} \times L^p \to L^p$ in Lemma 4.3.

Lemma 4.1. *For any cancellative paraproduct with descendants*

$$B_k(b, f) = \sum_{I \in \mathcal{D}} \beta_I \langle b, h_{I^{(k)}} \rangle \langle f, h_I \rangle h_I |I^{(k)}|^{-1/2},$$

where all Haar functions are cancellative, there holds

$$\|B_k(b, f)\|_{L^2(\mathbb{R}^d)} \lesssim \|b\|_{BMO(\mathbb{R}^d)} \|f\|_{L^2(\mathbb{R}^d)}, \quad \forall f \in L^2.$$

with a constant independent of $k \geq 0$.

Before proceeding with the proof, we remark that for estimates of this type, it suffices to prove the inequality with the BMO norm replaced by the dyadic BMO norm corresponding to the fixed dyadic grid. Indeed, for any dyadic grid $\mathcal{D}$, $\|\cdot\|_{\text{BMO}_{\mathcal{D}}} \leq \|\cdot\|_{\text{BMO}}$. In most cases in the following, when the argument doesn't rely explicitly on the grid $\mathcal{D}$, we will usually denote the dyadic BMO space as BMO.

Proof. $B_k(b, f)$ here is in fact a martingale transform whose uniform boundedness follows directly from the observation $|\langle b, h_{I^{(k)}} \rangle| / |I^{(k)}|^{1/2} \leq \|b\|_{\text{BMO}}$. However, we will present a different proof via square function in the following, which will provide some insight into the estimates of some other operators and the multi-parameter analogs of the result. For any $g \in L^2(\mathbb{R}^d)$,

$$\langle B_k(b,f), g\rangle = \Big\langle b, \sum_I \beta_I \langle f, h_I\rangle \langle g, h_I\rangle h_{I^{(k)}} |I^{(k)}|^{-1/2} \Big\rangle.$$

It thus suffices to show that

$$\Big\| \sum_I \beta_I \langle f, h_I\rangle \langle g, h_I\rangle h_{I^{(k)}} |I^{(k)}|^{-1/2} \Big\|_{H^1} \lesssim \|f\|_{L^2} \|g\|_{L^2},$$

which is equivalent to

$$\left\| S\left(\sum_I \beta_I \langle f, h_I\rangle \langle g, h_I\rangle h_{I^{(k)}} |I^{(k)}|^{-1/2} \right) \right\|_{L^1} \lesssim \|f\|_{L^2} \|g\|_{L^2},$$

where in the above S denotes the dyadic square function.

To see this, write

$$S\left(\sum_I \beta_I \langle f, h_I\rangle \langle g, h_I\rangle h_{I^{(k)}} |I^{(k)}|^{-1/2} \right)^2 = \sum_J \left(\sum_{I:I^{(k)}=J} \beta_I \langle f, h_I\rangle \langle g, h_I\rangle |J|^{-1/2} \right)^2 \frac{\chi_J}{|J|}$$

which together with $\|\cdot\|_{\ell^2} \le \|\cdot\|_{\ell^1}$ and Cauchy-Schwarz inequality implies

$$\begin{aligned}
&S\left(\sum_I \beta_I \langle f, h_I\rangle \langle g, h_I\rangle h_{I^{(k)}} |I^{(k)}|^{-1/2} \right) \\
&\le \sum_J \left(\sum_{I:I^{(k)}=J} |\langle f, h_I\rangle| |\langle g, h_I\rangle| \frac{\chi_J}{|J|} \right) \\
&\le \sum_J \left(\sum_{I:I^{(k)}=J} |\langle f, h_I\rangle|^2 \right)^{1/2} \left(\sum_{I:I^{(k)}=J} |\langle g, h_I\rangle|^2 \right)^{1/2} \frac{\chi_J}{|J|} \\
&\le \left(\sum_J \sum_{I:I^{(k)}=J} |\langle f, h_I\rangle|^2 \frac{\chi_J}{|J|} \right)^{1/2} \left(\sum_J \sum_{I:I^{(k)}=J} |\langle g, h_I\rangle|^2 \frac{\chi_J}{|J|} \right)^{1/2} \\
&=: (S^{(k)} f)(S^{(k)} g).
\end{aligned}$$

where the operator $S^{(k)} f := (\sum_J \sum_{I:I^{(k)}=J} |\langle f, h_I\rangle|^2 |J|^{-1} \chi_J)^{1/2}$. We claim that $S^{(k)}$:

$L^2 \to L^2$ with norm bounded by a dimensional constant, which does not depend on k. This guarantees that our estimate of B_k becomes independent of k. Combining this with another use of Cauchy-Schwarz will complete the proof.

To show the claim, denote $\alpha_J = (\sum_{I:I^{(k)}=J} |\langle f, h_I\rangle|^2)^{1/2}$ for any J and define $F(x) = \sum_J \alpha_J h_J(x)$. Then

$$\begin{aligned}\|S^{(k)} f\|_{L^2}^2 &= \|(\sum_J \alpha_J^2 \frac{\chi_J}{|J|})^{1/2}\|_{L^2}^2 = \|SF\|_{L^2}^2 \\ &\lesssim \|F\|_{L^2}^2 = \sum_J \alpha_J^2 = \sum_J \sum_{I:I^{(k)}=J} |\langle f, h_I\rangle|^2 = \sum_I |\langle f, h_I\rangle|^2 = \|f\|_{L^2}^2,\end{aligned}$$

where the second to last equality holds because that cube I in the previous summation ranges over all the dyadic cubes exactly once. □

Remark 4.2. We mention that the reason why the above argument cannot be used to obtain the L^p boundedness of B_k is because $S^{(k)}$ is not uniformly bounded on L^p when $1 < p < 2$. This can be seen by testing $S^{(k)}$ on Haar functions. However, when $2 \le p < \infty$, we indeed have $\|S^{(k)}\|_{L^p \to L^p} \le 1$, by observing that $\|S^{(k)}\|_{\text{BMO}\to\text{BMO}} \le 1$ and applying interpolation.

Lemma 4.3. *For tri-linear operator*

$$P(b, a, f) = \sum_{I\in\mathcal{D}} \beta_I \langle b, h_I\rangle\langle f, h_I\rangle |I|^{-1} \sum_{J:J\subsetneq I} \langle a, h_J\rangle h_J,$$

there holds

$$\|P(b, a, f)\|_{L^p(\mathbb{R}^d)} \lesssim \|b\|_{BMO(\mathbb{R}^d)} \|a\|_{BMO(\mathbb{R}^d)} \|f\|_{L^p(\mathbb{R}^d)}, \quad \forall f \in L^p,\ 1 < p < \infty.$$

Proof. We present the proof assuming $p = 2$, while the argument proceeds almost identically for other p. The idea is to employ the H^1-BMO duality and the square function characterization of H^1. For any normalized test function $g \in L^2$,

$$\langle P(b, a, f), g\rangle = \Big\langle b, \sum_I \beta_I \langle f, h_I\rangle |I|^{-1} h_I \sum_{J:J\subsetneq I} \langle a, h_J\rangle\langle g, h_J\rangle \Big\rangle.$$

To see where the BMO norm of a comes into play, observe that for any fixed I and some $1 < p < 2$, $\left| \sum_{J:J\subsetneq I} \langle a, h_J\rangle\langle g, h_J\rangle \right| = \left| \left\langle \sum_{J:J\subsetneq I} \langle a, h_J\rangle h_J, g\chi_I \right\rangle \right|$ is bounded by

$$\begin{aligned}
\leq & \Big\| \sum_{J:J\subsetneq I} \langle a, h_J\rangle h_J \Big\|_{L^{p'}} \|g\chi_I\|_{L^p} \lesssim \Big\| \big(\sum_{J:J\subsetneq I} |\langle a, h_J\rangle|^2 \frac{\chi_J}{|J|} \big)^{1/2} \Big\|_{L^{p'}} \|g\chi_I\|_{L^p} \\
\lesssim & \|a\|_{\mathrm{BMO}} |I|^{1/p'} \|g\chi_I\|_{L^p} = \|a\|_{\mathrm{BMO}} |I| \left(\langle |g|^p \rangle_I \right)^{1/p},
\end{aligned}$$

where the last inequality follows from John-Nirenberg inequality.

Therefore,

$$\begin{aligned}
& S\left(\sum_I \beta_I \langle f, h_I\rangle |I|^{-1} h_I \sum_{J:J\subsetneq I} \langle a, h_J\rangle\langle g, h_J\rangle \right) \\
= & \left(\sum_I |\beta_I \langle f, h_I\rangle|^2 |I|^{-2} \Big(\sum_{J:J\subsetneq I} \langle a, h_J\rangle\langle g, h_J\rangle \Big)^2 \frac{\chi_I}{|I|} \right)^{1/2} \\
\leq & \|a\|_{\mathrm{BMO}} \left(\sum_I |\langle f, h_I\rangle|^2 (\langle |g|^p \rangle_I)^{2/p} \frac{\chi_I}{|I|} \right)^{1/2} \\
\leq & \|a\|_{\mathrm{BMO}} \left(\sum_I |\langle f, h_I\rangle|^2 \sup_{I:x\in I} (\langle |g|^p \rangle_I)^{2/p} \frac{\chi_I}{|I|} \right)^{1/2} \\
\leq & \|a\|_{\mathrm{BMO}} M(|g|^p)^{1/p} S(f),
\end{aligned}$$

where M is the Hardy-Littlewood maximal function which is bounded on L^p, $1 < p < \infty$. Hence,

$$\|P(b, a, f)\|_{L^2} \lesssim \|b\|_{\mathrm{BMO}} \|a\|_{\mathrm{BMO}} \|M(|g|^p)^{1/p}\|_{L^2} \|S(f)\|_{L^2} \lesssim \|b\|_{\mathrm{BMO}} \|a\|_{\mathrm{BMO}} \|f\|_{L^2}.$$

□

4.1.2 Multi-parameter case

In the multi-parameter setting, just like the multi-parameter dyadic paraproduct B_0 we have introduced in Chapter 2, which is essentially a tensor product of the one-parameter versions, one expects to encounter "tensor product" of the paraproduct-like basic operators B_k, P that we have discussed in the previous subsection. We will define and study these operators in the bi-parameter setting, and it will be clear that all the arguments iterate well with more parameters. Suppose we are in $\mathbb{R}^n \times \mathbb{R}^m$, and we will use $h_I^\epsilon, u_I^\epsilon$ to denote Haar functions in the first and second variable, respectively.

Lemma 4.4. *Given $b \in BMO_{prod}(\mathbb{R}^n \times \mathbb{R}^m)$ and integers $k, l \geq 0$, define the following operators*

$$BB_{k,l}(b,f) := \sum_{I_1 \in \mathcal{D}_1, I_2 \in \mathcal{D}_2} \beta_{I_1 I_2} \langle b, h_{I_1^{(k)}} \otimes u_{I_2^{(l)}} \rangle \langle f, h_{I_1}^{\epsilon_1} \otimes u_{I_2}^{\epsilon_2} \rangle h_{I_1}^{\epsilon_1'} \otimes u_{I_2}^{\epsilon_2'} |I_1^{(k)}|^{-1/2} |I_2^{(l)}|^{-1/2},$$

where $\beta_{I_1 I_2}$ is a sequence satisfying $|\beta_{I_1 I_2}| \leq 1$. When $k > 0$, all the Haar functions in the first variable are cancellative, while when $k = 0$, there is at most one of $h_{I_1}^{\epsilon_1}, h_{I_1}^{\epsilon_1'}$ being non-cancellative for each I_1. The same assumption goes for the second variable. Then, $\|B_{k,l}(b,f)\|_{L^2} \lesssim \|b\|_{BMO_{prod}} \|f\|_{L^2}$ with a constant independent of k, l.

Without loss of generality, one can assume that all non-cancellative Haar functions appear in the same position for all I_1 (or I_2), since one can always reduce to this case by splitting $BB_{k,l}$ into finitely many parts, each of which satisfies this. We also remark that in the case when $k > 0$ and non-cancellative Haar functions appear in the first variable (or symmetrically, when $l > 0$ and non-cancellative Haar functions appear in the second variable), it is unknown whether one can derive the same uniform bounds as above with respect to k, l. The best known dependence of the boundedness constant on k, l is $k^{1/2}, l^{1/2}$, which we will prove in Appendix B.

Note that when $k = l = 0$, $BB_{k,l}$ becomes the classical bi-parameter paraproduct B_0 we have seen in Chapter 2, which are known to be bounded. Moreover, when all the Haar functions are cancellative, the proof of the lemma proceeds exactly the same as its one-parameter counterpart Lemma 4.1, except that one needs to use bi-parameter dyadic square function as majorization instead. Therefore in the following, we will only prove the lemma assuming that $k = 0, l > 0$, and $h_{I_1}^{\epsilon_1} = h_{I_1}^1$ being the only non-cancellative Haar.

(The $h_{I_1}^{\epsilon_1'} = h_{I_1}^1$ case is the adjoint of this one.)

Proof. We are going to follow the strategy in the proof of Lemma 4.1 and use hybrid maximal-square functions as majorization.

Pairing $BB_{0,l}(b,f)$ with a normalized L^2 function g and applying the product H^1-BMO duality, it suffices to show that

$$\left\| SS\left(\sum_{I_1,I_2} \beta_{I_1 I_2} \langle f, h_{I_1}^1 \otimes u_{I_2}\rangle \langle g, h_{I_1} \otimes u_{I_2}\rangle h_{I_1} \otimes u_{I_2^{(l)}} |I_1|^{-1/2} |I_2^{(l)}|^{-1/2}\right)\right\|_{L^1} \lesssim \|f\|_{L^2},$$

where SS is the dyadic double square function whose L^1 norm characterizes product H^1.

To see this, one calculates

$$\begin{aligned}
& SS\left(\sum_{I_1,I_2} \beta_{I_1 I_2} \langle f, h_{I_1}^1 \otimes u_{I_2}\rangle \langle g, h_{I_1} \otimes u_{I_2}\rangle h_{I_1} \otimes u_{I_2^{(l)}} |I_1|^{-1/2} |I_2^{(l)}|^{-1/2}\right)^2 \\
&= \sum_{I_1,I_2} \left(\sum_{J_2 : J_2^{(l)} = I_2} \langle f, h_{I_1}^1 \otimes u_{J_2}\rangle \langle g, h_{I_1} \otimes u_{J_2}\rangle |I_1|^{-1/2} |I_2|^{-1/2} \right)^2 \frac{\chi_{I_1} \otimes \chi_{I_2}}{|I_1||I_2|} \\
&\leq \sum_{I_1} \left(\sum_{I_2} \sum_{J_2 : J_2^{(l)} = I_2} \sup_{I_1} (\langle \langle f, u_{J_2}\rangle_2 \rangle_{I_1}) \langle g, h_{I_1} \otimes u_{J_2}\rangle \frac{\chi_{I_2}}{|I_2|} \right)^2 \frac{\chi_{I_1}}{|I_1|},
\end{aligned}$$

where the last inequality follows from $\|\cdot\|_{\ell^2} \leq \|\cdot\|_{\ell^1}$, and $\langle \cdot \rangle_{I_1}$ denotes the average value over I_1. Then the above is controlled by

$$\sum_{I_1} \left(\sum_{I_2} \sum_{J_2 : J_2^{(l)} = I_2} M_1(\langle f, u_{J_2}\rangle_2) |\langle g, h_{I_1} \otimes u_{J_2}\rangle| \frac{\chi_{I_2}}{|I_2|} \right)^2 \frac{\chi_{I_1}}{|I_1|},$$

where M_1 is the Hardy-Littlewood maximal function in the first variable. Next, Cauchy-

Schwarz inequality implies that

$$
\begin{aligned}
&\leq \sum_{I_1}\left(\sum_{I_2}\sum_{J_2:J_2^{(l)}=I_2} M_1(\langle f,u_{J_2}\rangle_2)^2\frac{\chi_{I_2}}{|I_2|}\right)\left(\sum_{I_2}\sum_{J_2:J_2^{(l)}=I_2}|\langle g,h_{I_1}\otimes u_{J_2}\rangle|^2\frac{\chi_{I_2}}{|I_2|}\right)\frac{\chi_{I_1}}{|I_1|}\\
&=\left(\sum_{I_2}\sum_{J_2:J_2^{(l)}=I_2} M_1(\langle f,u_{J_2}\rangle_2)^2\frac{\chi_{I_2}}{|I_2|}\right)\left(\sum_{I_1}\sum_{I_2}\sum_{J_2:J_2^{(l)}=I_2}|\langle g,h_{I_1}\otimes u_{J_2}\rangle|^2\frac{\chi_{I_1}\otimes\chi_{I_2}}{|I_1||I_2|}\right)\\
&=: I\cdot II.
\end{aligned}
$$

II could be written as the square of SS acting on a normalized L^2 function, similarly as the last part of the proof of Lemma 4.1. For I, Fefferman-Stein inequality implies that

$$
\begin{aligned}
\|I^{1/2}\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)} &= \left(\int_{\mathbb{R}^m}\left\|\Big(\sum_{I_2}\sum_{J_2:J_2^{(l)}=I_2} M_1(\langle f,u_{J_2}\rangle_2)^2\frac{\chi_{I_2}}{|I_2|}\Big)^{1/2}\right\|^2_{L^2(\mathbb{R}^n)}dx_2\right)^{1/2}\\
&\lesssim \left(\int_{\mathbb{R}^m}\left\|\Big(\sum_{I_2}\sum_{J_2:J_2^{(l)}=I_2} |\langle f,u_{J_2}\rangle_2|^2\frac{\chi_{I_2}}{|I_2|}\Big)^{1/2}\right\|^2_{L^2(\mathbb{R}^n)}dx_2\right)^{1/2}\\
&\lesssim \left(\int_{\mathbb{R}^m}\|f(\cdot,x_2)\|^2_{L^2(\mathbb{R}^n)}\,dx_2\right)^{1/2} = \|f\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)},
\end{aligned}
$$

where once again the last inequality is due to the same argument in the last part of the proof of Lemma 4.1, thus the proof is complete. □

Lemma 4.5. *Given $b,a\in BMO_{prod}(\mathbb{R}^n\times\mathbb{R}^m)$, define*

$$
\begin{aligned}
&PP(b,a,f):=\\
&\sum_{I_1,I_2}\langle b,h_{I_1}\otimes u_{I_2}\rangle\langle f,h_{I_1}\otimes u_{I_2}\rangle|I_1|^{-1}|I_2|^{-1}\sum_{J_1:J_1\subsetneq I_1}\sum_{J_2:J_2\subsetneq I_2}\langle a,h_{J_1}\otimes u_{J_2}\rangle h_{J_1}\otimes u_{J_2},
\end{aligned}
$$

and let PP_1 be its partial adjoint in the first variable with b,a fixed. Then for $1<p<\infty$,

$$\|PP(b,a,f)\|_{L^p}\lesssim\|b\|_{BMO_{prod}}\|a\|_{BMO_{prod}}\|f\|_{L^p},\tag{4.6}$$

$$\|PP_1(b,a,f)\|_{L^p}\lesssim\|b\|_{BMO_{prod}}\|a\|_{BMO_{prod}}\|f\|_{L^p}.\tag{4.7}$$

Recall that for a bi-parameter singular integral T, its partial adjoint T_1 is defined via

$$\langle T(f_1\otimes f_2), g_1\otimes g_2\rangle = \langle T_1(g_1\otimes f_2), f_1\otimes g_2\rangle.$$

It is known that the L^2 boundedness of T does not imply the L^2 boundedness of T_1 (see [Jou85] or [MO15] for a detailed discussion and counterexamples). Hence, in the following, we need to prove the boundedness of PP and PP_1 separately.

Proof. We assume again that $p=2$, since the almost identical proof works for other p as well. First, note that the proof for PP is essentially the same as the one of Lemma 4.3. In the bi-parameter setting, one needs to use the double square function SS to characterize product H^1 and the strong maximal function M_S as majorization. The key observation is that there holds the following bi-parameter John-Nirenberg inequality (see [CF80]):

$$\left\|\Big(\sum_{R\subset\Omega}|\langle a,h_R\rangle|^2\frac{\chi_R}{|R|}\Big)^{1/2}\right\|_{L^p}\leq \|a\|_{\mathrm{BMO}_{\mathrm{prod}}}|\Omega|^{1/p},\quad 1<p<\infty,$$

where Ω is any open set in $\mathbb{R}^n\times\mathbb{R}^m$ of finite measure, and R denotes dyadic rectangles. It thus easy to verify that a same argument as in Lemma 4.3 implies (4.6).

The estimate (4.7) involves the hybrid maximal-square functions, which we have seen in the proof of Lemma 4.4. Specifically, let $g\in L^2$ be a normalized test function,

$$\begin{aligned}&\langle PP_1(b,a,f),g\rangle\\ =&\Big\langle b,\sum_{I_1,I_2}|I_1|^{-1}|I_2|^{-1}h_{I_1}\otimes u_{I_2}\sum_{J_1:J_1\subsetneq I_1}\sum_{J_2:J_2\subsetneq I_2}\langle a,h_{J_1}\otimes u_{J_2}\rangle\langle f,h_{J_1}\otimes u_{I_2}\rangle\langle g,h_{I_1}\otimes u_{J_2}\rangle\Big\rangle.\end{aligned}$$

Note that by the bi-parameter John-Nirenberg inequality,

$$\begin{aligned}&\Big|\sum_{J_1:J_1\subsetneq I_1}\sum_{J_2:J_2\subsetneq I_2}\langle a,h_{J_1}\otimes u_{J_2}\rangle\langle f,h_{J_1}\otimes u_{I_2}\rangle\langle g,h_{I_1}\otimes u_{J_2}\rangle\Big|\\ =&\Big|\sum_{J_1:J_1\subsetneq I_1}\sum_{J_2:J_2\subsetneq I_2}\langle a,h_{J_1}\otimes u_{J_2}\rangle\langle\langle f,u_{I_2}\rangle_2\otimes\langle g,h_{I_1}\rangle_1,h_{J_1}\otimes u_{J_2}\rangle\Big|\\ \leq&\|a\|_{\mathrm{BMO}_{\mathrm{prod}}}|I_1||I_2|\Big(\big\langle|\langle f,u_{I_2}\rangle_2|^p\big\rangle_{I_1}\Big)^{1/p}\Big(\big\langle|\langle g,h_{I_1}\rangle_1|^p\big\rangle_{I_2}\Big)^{1/p},\end{aligned}$$

for some $1 < p < 2$. Hence,

$$
\begin{aligned}
&SS\left(\sum_{I_1,I_2}|I_1|^{-1}|I_2|^{-1}h_{I_1}\otimes u_{I_2}\sum_{J_1:J_1\subsetneq I_1}\sum_{J_2:J_2\subsetneq I_2}\langle a,h_{J_1}\otimes u_{J_2}\rangle\langle f,h_{J_1}\otimes u_{I_2}\rangle\langle g,h_{I_1}\otimes u_{J_2}\rangle\right)\\
\leq&\|a\|_{\mathrm{BMO_{prod}}}\left(\sum_{I_1,I_2}\left(\left\langle|\langle f,u_{I_2}\rangle_2|^p\right\rangle_{I_1}\right)^{2/p}\left(\left\langle|\langle g,h_{I_1}\rangle_1|^p\right\rangle_{I_2}\right)^{2/p}\frac{\chi_{I_1}\otimes\chi_{I_2}}{|I_1||I_2|}\right)^{1/2}\\
\leq&\|a\|_{\mathrm{BMO_{prod}}}\left(\sum_{I_2}M_1(|\langle f,u_{I_2}\rangle_2|^p)^{2/p}\frac{\chi_{I_2}}{|I_2|}\right)^{1/2}\left(\sum_{I_1}M_2(|\langle g,h_{I_1}\rangle_1|^p)^{2/p}\frac{\chi_{I_1}}{|I_1|}\right)^{1/2}.
\end{aligned}
$$

The two terms on the last line above can be viewed as generalized hybrid maximal-square functions, whose boundedness is easy to obtain. For instance,

$$
\begin{aligned}
&\left\|\Big(\sum_{I_2}M_1(|\langle f,u_{I_2}\rangle_2|^p)^{2/p}\frac{\chi_{I_2}}{|I_2|}\Big)^{1/2}\right\|_{L^2}\\
=&\left(\int_{\mathbb{R}^n}\Big\|\Big(\sum_{I_2}M_1(|\langle f,u_{I_2}\rangle_2|^p)^{2/p}\frac{\chi_{I_2}}{|I_2|}\Big)^{1/2}\Big\|_{L^2(\mathbb{R}^m)}^2dx_1\right)^{1/2}\\
=&\left(\int_{\mathbb{R}^n}\sum_{I_2}M_1(|\langle f,u_{I_2}\rangle_2|^p)^{2/p}\,dx_1\right)^{1/2}\lesssim\left(\sum_{I_2}\int_{\mathbb{R}^n}|\langle f,u_{I_2}\rangle_2|^2\,dx_1\right)^{1/2}=\|f\|_{L^2}.
\end{aligned}
$$

Therefore, $\|PP_1(b,a,f)\|_{L^2}\lesssim\|b\|_{\mathrm{BMO_{prod}}}\|a\|_{\mathrm{BMO_{prod}}}\|f\|_{L^2}$. □

In addition to the above two types of operators, in the bi-parameter setting, a new type of paraproduct-like operators that mix B_k and P together arises naturally in our argument. We show that they have the following uniform BMO estimates.

Lemma 4.8. *Given $b\in BMO_{prod}(\mathbb{R}^n\times\mathbb{R}^m)$. Let $\{a^1_{I_2}\}_{I_2\in\mathcal{D}_2}$ be a sequence of functions in $BMO(\mathbb{R}^n)$ indexed by $I_2\in\mathcal{D}_2$, such that $\sup_{I_2}\|a^1_{I_2}\|_{BMO}=:C_1<\infty$. And let $\{a^2_{I_1}\}_{I_1\in\mathcal{D}_1}$ be a sequence of functions in $BMO(\mathbb{R}^m)$ indexed by $I_1\in\mathcal{D}_1$, such that $\sup_{I_1}\|a^2_{I_1}\|_{BMO}=:$*

$C_2 < \infty$. *For integers* $k, l \geq 0$, *define*

$$BP_k(b, \{a^2_{I_1}\}_{I_1}, f) :=$$
$$\sum_{I_1, I_2} \beta_{I_1} \langle b, h_{I_1^{(k)}} \otimes u_{I_2} \rangle \langle f, h^{\epsilon_1}_{I_1} \otimes u_{I_2} \rangle |I_1^{(k)}|^{-1/2} |I_2|^{-1} h^{\epsilon'_1}_{I_1} \sum_{J_2 : J_2 \subsetneq I_2} \langle a^2_{I_1}, u_{J_2} \rangle_2 u_{J_2},$$

$$PB_l(b, \{a^1_{I_2}\}_{I_2}, f) :=$$
$$\sum_{I_1, I_2} \beta_{I_2} \langle b, h_{I_1} \otimes u_{I_2^{(l)}} \rangle \langle f, h_{I_1} \otimes u^{\epsilon_2}_{I_2} \rangle |I_1|^{-1} |I_2^{(l)}|^{-1/2} h^{\epsilon'_2}_{I_2} \sum_{J_1 : J_1 \subsetneq I_1} \langle a^1_{I_2}, h_{J_1} \rangle_1 h_{J_1},$$

where β_{I_1}, β_{I_2} *are sequences satisfying* $|\beta_{I_1}|, |\beta_{I_2}| \leq 1$. *When* $k > 0$, *all the Haar functions in the first variable are cancellative, while when* $k = 0$, *there is at most one of* $h^{\epsilon_1}_{I_1}, h^{\epsilon'_1}_{I_1}$ *being non-cancellative for each* I_1. *The same assumption goes for the second variable. Then, there holds*

$$\|BP_k(b, \{a^2_{I_1}\}_{I_1}, f)\|_{L^2} \lesssim C_2 \|b\|_{BMO_{prod}} \|f\|_{L^2},$$

$$\|PB_l(b, \{a^1_{I_2}\}_{I_2}, f)\|_{L^2} \lesssim C_1 \|b\|_{BMO_{prod}} \|f\|_{L^2}.$$

We remark that in many applications of the result above, we will be in the case when there is a function $a^2 \in \mathrm{BMO}(\mathbb{R}^m)$ such that $a^2_{I_1} = a^2$ for any I_1. If this is the case, $BP_k(b, \{a^2_{I_1}\}_{I_1}, f)$ will simply be denoted as $BP_k(b, a^2, f)$. And the lemma above says that

$$\|BP_k(b, a^2, f)\|_{L^2} \lesssim \|a^2\|_{\mathrm{BMO}} \|b\|_{\mathrm{BMO_{prod}}} \|f\|_{L^2}.$$

Similarly for PB_l.

Proof. By symmetry, it suffices to estimate PB_l. The strategy is similar as before: a square function argument encoding the product BMO estimate of b, combined with a John-Nirenberg inequality taking advantage of the BMO estimate of $a^1_{I_2}$. Note that the arguments slightly vary depending on whether non-cancellative Haar functions appear.

Taking g such that $\|g\|_{L^2} \leq 1$,

$$\begin{aligned}&\Big\langle PB_l(b, \{a^1_{I_2}\}_{I_2}, f), g\Big\rangle\\ =&\Big\langle b, \sum_{I_1,I_2} \langle f, h_{I_1} \otimes u^{\epsilon_2}_{I_2}\rangle |I_1|^{-1} |I_2^{(l)}|^{-1/2} h_{I_1} \otimes u_{I_2^{(l)}} \sum_{J_1: J_1 \subsetneq I_1} \langle a^1_{I_2}, h_{J_1}\rangle_1 \langle g, h_{J_1} \otimes u^{\epsilon_2'}_{I_2}\rangle\Big\rangle.\end{aligned}$$

A similar application of John-Nirenberg inequality as before implies that

$$\begin{aligned}&SS\left(\sum_{I_1,I_2} \langle f, h_{I_1} \otimes u^{\epsilon_2}_{I_2}\rangle |I_1|^{-1} |I_2^{(l)}|^{-1/2} h_{I_1} \otimes u_{I_2^{(l)}} \sum_{J_1: J_1 \subsetneq I_1} \langle a^1_{I_2}, h_{J_1}\rangle_1 \langle g, h_{J_1} \otimes u^{\epsilon_2'}_{I_2}\rangle\right)\\ \leq& \sup_{I_2} \|a^1_{I_2}\|_{\mathrm{BMO}} \left(\sum_{I_1,J_2}\Big(\sum^{(l)}_{I_2 \subset J_2} \langle f, h_{I_1} \otimes u^{\epsilon_2}_{I_2}\rangle \big(\langle |\langle g, u^{\epsilon_2'}_{I_2}\rangle_2|^p\rangle_{I_1}\big)^{1/p}\Big)^2 \frac{\chi_{I_1} \otimes \chi_{J_2}}{|I_1||J_2|^2}\right)^{1/2}\end{aligned} \tag{4.9}$$

(a) Case $l > 0$.

In this case, all the Haar functions that appear are cancellative, hence by omitting the dependence on ϵ_2, ϵ_2' and applying Cauchy-Schwarz inequality, there holds

$$\begin{aligned}(4.9) &\leq C_1 \left(\sum_{I_1,J_2}\Big(\sum^{(l)}_{I_2\subset J_2} |\langle f, h_{I_1} \otimes u_{I_2}\rangle|^2\Big)\Big(\sum^{(l)}_{I_2\subset J_2} \big(\langle |\langle g, u_{I_2}\rangle_2|^p\rangle_{I_1}\big)^{2/p}\Big) \frac{\chi_{I_1} \otimes \chi_{J_2}}{|I_1||J_2|^2}\right)^{1/2}\\ &\leq C_1 \left(\sum_{J_2}\Big(\sum_{I_1}\sum^{(l)}_{I_2\subset J_2} |\langle f, h_{I_1} \otimes u_{I_2}\rangle|^2 \frac{\chi_{I_1}}{|I_1|}\Big)\Big(\sum^{(l)}_{I_2\subset J_2} M_1(|\langle g, u_{I_2}\rangle_2|^p)^{2/p}\Big)\frac{\chi_{J_2}}{|J_2|^2}\right)^{1/2},\end{aligned}$$

which by $\|\cdot\|_{\ell^2} \leq \|\cdot\|_{\ell^1}$ and another use of Cauchy-Schwarz is bounded by

$$C_1 \left(\sum_{I_1}\sum_{J_2}\sum^{(l)}_{I_2\subset J_2} |\langle f, h_{I_1} \otimes u_{I_2}\rangle|^2 \frac{\chi_{I_1} \otimes \chi_{J_2}}{|I_1||J_2|}\right)^{1/2} \left(\sum_{J_2}\sum^{(l)}_{I_2\subset J_2} M_1(|\langle g, u_{I_2}\rangle_2|^p)^{2/p} \frac{\chi_{J_2}}{|J_2|}\right)^{1/2}.$$

Therefore, a similar double square function and hybrid maximal-square function argument as in Lemma 4.4 and Lemma 4.5 implies that

$$\|(4.9)\|_{L^1} \lesssim C_1 \|f\|_{L^2} \|g\|_{L^2}.$$

(b) Case $l = 0$ and $\epsilon_2 = \vec{1}$.

In this case,

$$
\begin{aligned}
(4.9) &= C_1 \left(\sum_{I_1, I_2} \Big(\langle \langle f, h_{I_1} \rangle_1 \rangle_{I_2} \Big) \Big(\langle | \langle g, u_{I_2} \rangle_2 |^p \rangle_{I_1} \Big)^{2/p} \frac{\chi_{I_1} \otimes \chi_{I_2}}{|I_1||I_2|} \right)^{1/2} \\
&\leq C_1 \left(\sum_{I_1} M_2 (\langle f, h_{I_1} \rangle_1)^2 \frac{\chi_{I_1}}{|I_1|} \right)^{1/2} \left(\sum_{I_2} M_1 (| \langle g, u_{I_2} \rangle_2 |^p)^{2/p} \frac{\chi_{I_2}}{|I_2|} \right)^{1/2},
\end{aligned}
$$

which shows that $\|(4.9)\|_{L^1} \lesssim C_1 \|f\|_{L^2} \|g\|_{L^2}$.

(c) Case $l = 0$ and $\epsilon_2' = \vec{1}$.

This last case can be dealt with similarly by noticing that

$$
\begin{aligned}
(4.9) &= C_1 \left(\sum_{I_1, I_2} |\langle f, h_{I_1} \otimes u_{I_2} \rangle|^2 \big(\langle | \langle g \rangle_{I_2} |^p \rangle_{I_1} \big)^{2/p} \frac{\chi_{I_1} \otimes \chi_{I_2}}{|I_1||I_2|} \right)^{1/2} \\
&\leq C_1 \left(M_1 (|M_2(g)|^p) \right)^{1/p} SS(f).
\end{aligned}
$$

The boundedness of M_1 and M_2 in each variable implies that

$$
\| \left(M_1 (|M_2(g)|^p) \right)^{1/p} \|_{L^2} \lesssim \|g\|_{L^2}.
$$

To conclude, we've demonstrated in each case that

$$
\|PB_l(b, \{a^1_{I_2}\}_{I_2}, f)\|_{L^2} \lesssim C_1 \|b\|_{\mathrm{BMO_{prod}}} \|f\|_{L^2},
$$

which completes the proof. □

Before we end this section, note that in all the paraproduct-like operators discussed above, the BMO symbol function b is always paired with a cancellative Haar function in the summation. When it is not the case, the following results will be of use to us.

Lemma 4.10. *Let b be a little BMO function on $\mathbb{R}^n \times \mathbb{R}^m$. For paraproducts of the following two types:*

$$BB^1_{k,l}(b,f) := \sum_{I_1,I_2} \beta_{I_1,I_2} \langle b, h^1_{I_1^{(k)}} \otimes h_{I_2^{(l)}} \rangle \langle f, h_{I_1} \otimes h^{\epsilon}_{I_2} \rangle h_{I_1} \otimes h^{\epsilon'}_{I_2} |I_1^{(k)}|^{-1/2} |I_2^{(l)}|^{-1/2},$$

$$BB^2_{k,l}(b,f) := \sum_{I_1,I_2} \beta_{I_1,I_2} \langle b, h_{I_1^{(k)}} \otimes h^1_{I_2^{(l)}} \rangle \langle f, h^{\epsilon}_{I_1} \otimes h_{I_2} \rangle h^{\epsilon'}_{I_1} \otimes h_{I_2} |I_1^{(k)}|^{-1/2} |I_2^{(l)}|^{-1/2},$$

where at least one of ϵ, ϵ' is not equal to $\vec{1}$ and $|\beta_{I_1,I_2}| \leq 1$ uniformly, assume that when $l > 0$, all Haar functions that appear in $BB^1_{k,l}$ are cancellative, and when $k > 0$, all Haar functions that appear in $BB^2_{k,l}$ are cancellative. Then there holds

$$\|BB^i_{k,l}(b,f)\|_{L^2} \lesssim \|b\|_{bmo} \|f\|_{L^2}, \quad i = 1, 2$$

Note that operator $BB^i_{k,l}$ is significantly different from the $BB_{k,l}$ that we have studied above, as some non-cancellative Haar functions appear in the pairing with the symbol b. Thanks to the stronger little BMO norm though, this lemma follows from a simple one-parameter argument.

Proof. We only prove the result for $BB^1_{k,l}$ as the other one is completely symmetric. Write

$$\begin{aligned} BB^1_{k,l}(b,f) &= \sum_{I_1} h_{I_1} \otimes \left(\sum_{I_2} \beta_{I_1,I_2} \langle \langle b \rangle_{I_1^{(k)}}, h_{I_2^{(l)}} \rangle_2 \langle \langle f, h_{I_1} \rangle_1, h^{\epsilon}_{I_2} \rangle_2 h^{\epsilon'}_{I_2} |I_2^{(l)}|^{-1/2} \right) \\ &= \sum_{I_1} h_{I_1} \otimes B^2_l (\langle b \rangle_{I_1^{(k)}}, \langle f, h_{I_1} \rangle_1), \end{aligned}$$

where B^2_l is a one-parameter paraproduct with descendants in the second variable. Following from Theorem 4.1,

$$\begin{aligned}\|BB^1_{k,l}(b,f)\|^2_{L^2} &= \sum_{I_1} \|B^2_l(\langle b\rangle_{I_1^{(k)}}, \langle f, h_{I_1}\rangle_1)\|^2_{L^2(\mathbb{R}^m)} \\ &\lesssim \sum_{I_1} \|\langle b\rangle_{I_1^{(k)}}\|^2_{\mathrm{BMO}(\mathbb{R}^m)} \|\langle f, h_{I_1}\rangle_1\|^2_{L^2(\mathbb{R}^m)} \\ &\leq \|b\|^2_{\mathrm{bmo}} \sum_{I_1} \|\langle f, h_{I_1}\rangle_1\|^2_{L^2(\mathbb{R}^m)} = \|b\|^2_{\mathrm{bmo}} \|f\|^2_{L^2}.\end{aligned}$$

□

4.2 Upper bound for commutators of CZ operators without iteration

In this section, we aim to prove Theorem 3.12 in its one-parameter version, i.e. the base case where there is no iterates. The base case will later be used as a stepping stone in the next section where the proof of the general case is presented. In the one-parameter setting, let T be a Calderón-Zygmund operator, and $b \in \mathrm{BMO}(\mathbb{R}^d)$. Our goal is to demonstrate for any L^2 function f that

$$\|[b,T]f\|_{L^2(\mathbb{R}^d)} \lesssim \|b\|_{\mathrm{BMO}(\mathbb{R}^d)} \|f\|_{L^2(\mathbb{R}^d)}, \tag{4.11}$$

with the underlying constant depending only on the dimension d.

Recall that in Chapter 2, we have introduced a representation formula by Hytönen (Theorem 2.13), using which CZ operators can be represented as an average of dyadic shift operators with respect to a probabilistic measure on a collection of dyadic grids. With this formula, (4.11) can be reduced to the upper bound for commutators of dyadic shifts. Specifically, according to Theorem 2.13, one could represent the commutator $[b,T]$ as an average of $[b, S^{ij}_\omega]$, where S^{ij}_ω is a dyadic shift of complexity (i,j) with dyadic grid $\mathcal{D}^\omega$.

Then, in order to prove (4.11), it suffices to prove that for any $f \in C_0^\infty(\mathbb{R}^d) \cap L^2(\mathbb{R}^d)$,

$$\left\| \sum_{i,j=0}^{\infty} 2^{-\max(i,j)\delta/2} [b, S_\omega^{ij}] f \right\|_{L^2} \lesssim \|b\|_{\mathrm{BMO}} \|f\|_{L^2} \tag{4.12}$$

uniformly in ω. In the following we will write S^{ij} for short as the argument doesn't depend on ω explicitly. As crucial ingredients of the proof of (4.12), the paraproduct-like operators B_k and P studied in Subsection 4.1.1 will be very important to us. In fact, the main theorem we will prove in this section is the following:

Theorem 4.13. *For cancellative dyadic shift S^{ij}, $[b, S^{ij}]f$ can be represented as a finite linear combination of the following terms:*

$$S^{ij}(B_k(b,f)), \quad B_k(b, S^{ij} f) \tag{4.14}$$

where the integer k is such that $0 \le k \le \max(i,j)$ and the total number of terms is bounded by $C(1+\max(i,j))$ for some universal dimensional constant C. For non-cancellative dyadic shift S^{00} (dyadic paraproduct) with BMO symbol a, i.e.

$$S^{00} f := \sum_{I \in \mathcal{D}} \langle a, h_I \rangle |I|^{-1/2} \langle f, h_I^\epsilon \rangle h_I^{\epsilon'}$$

with at least one of ϵ, ϵ' being not $\vec{1}$, $[b, S^{00}]f$ can be represented as a finite linear combination of the following terms:

$$S^{00}(B_0(b,f)), \quad B_0(b, S^{00} f), \quad P(b,a,f), \quad P^*(b,a,f), \tag{4.15}$$

where B_0 is the dyadic paraproduct defined in (2.15), P^* is understood as the adjoint of P with b and a fixed, and the total number of terms is bounded by a universal dimensional constant.

Remark 4.16. The representation claimed in Theorem 4.13 is far from unique. In fact, suggested by its proof, the readers can easily come up with representations of $[b, S^{ij}]f$ using other types of paraproducts, by decomposing the Haar sums differently. Moreover, as shown in the proof, the representation can be made such that except when $k = 0$, all the Haar functions appearing in $B_k(b, f)$ are cancellative.

It is easy to see that (4.12) is implied by Theorem 4.13. Indeed, given the boundedness of S^{ij}, Lemma 4.1, Lemma 4.3 together with Theorem 2.16 will guarantee the uniform boundedness of each of the terms in (4.14) and (4.15). Hence,

$$\|[b, S^{ij}]f\|_{L^2} \lesssim (1 + \max(i, j))\|b\|_{\mathrm{BMO}}\|f\|_{L^2}.$$

Note that the uniform boundedness of B_k with respect to k is crucial in the above argument, which was also the main difficulty of the proof of Lemma 4.1. Then, with the decaying factor $2^{-\max(i,j)\delta/2}$ in front, (4.12) follows from a simple geometric series argument.

Now we begin the proof of Theorem 4.13 and the strategy is the following. First, we decompose b and f using Haar bases. Second, we split the sum into several parts and represent each of them as a linear combination of terms as claimed in Theorem 4.13. Note that the reason why function b can be expanded in this way is that BMO functions are locally in L^2. By assuming that f is compactly supported and observing that dyadic shift S^{ij} are local operators, it suffices to assume that b is also compactly supported, therefore the Haar expansion is legitimate. To start with, one decomposes

$$\begin{aligned} [b, S^{ij}]f &= \sum_{I,J} \langle b, h_I\rangle\langle f, h_J\rangle [h_I, S^{ij}]h_J \\ &= \sum_{I,J} \langle b, h_I\rangle\langle f, h_J\rangle \left(h_I S^{ij} h_J - S^{ij}(h_I h_J)\right) =: I + II, \end{aligned}$$

where in the following I and II will be referred to as the *first term* and the *second term*, respectively. In order to further organize the sum and extract the correct paraproduct structure, even in the simplest one-parameter case, one needs to divide up the sum into many different parts, depending on the relative sizes of I, J.

4.2.1 Cancellative dyadic shift S^{ij}

Let's first look at the case when S^{ij} is cancellative, meaning that all the Haar functions appearing are cancellative. Hence,

$$[b, S^{ij}]f = \sum_{I,J} \langle b, h_I\rangle\langle f, h_J\rangle \left(h_I \sum_{J' \subset J^{(i)}}^{(j)} a_{JJ'J^{(i)}} h_{J'} - \sum_K \sum_{I'', J'' \subset K}^{(i,j)} a_{I''J''K} \langle h_I h_J, h_{I''}\rangle h_{J''} \right).$$

First, we claim that it suffices to consider the part $I \subset J^{(i)}$. Indeed, it is obvious that when $I \cap J^{(i)} = \emptyset$, both terms in the parentheses are zero. Furthermore, by the cancellation structure of the commutator, when $I \supsetneq J^{(i)}$, the term $[h_I, S^{ij}]h_J$ is also zero. To see this, as h_I is constant on $J^{(i)}$, fixing an arbitrary $x_0 \in J^{(i)}$ implies

$$h_I S^{ij} h_J - S^{ij}(h_I h_J) = h_I(x_0) S^{ij} h_J - S^{ij}(h_I(x_0) h_J) = 0.$$

Note that for the case $(i, j) \neq (0, 0)$, this is the only part of the proof where one needs the particular cancellation of the commutator structure. Next, we represent the first term and the second term separately.

First term

Based on the discussion above, for any i, j, the first term containing $h_I S^{ij} h_J$ is equal to

$$\sum_J \sum_{I : I \subset J^{(i)}} \langle b, h_I\rangle\langle f, h_J\rangle h_I \sum_{\substack{J' : J' \subset J^{(i)} \\ \ell(J') = 2^{i-j}\ell(J)}} a_{JJ'J^{(i)}} h_{J'}.$$

Introducing index $K = J^{(i)}$ allows us to rewrite this as

$$\sum_K \sum_{J:J\subset K}^{(i)} \sum_{I:I\subset K} \langle b, h_I\rangle\langle f, h_J\rangle h_I \sum_{J':J'\subset K}^{(j)} a_{JJ'K} h_{J'}$$
$$= \sum_I \langle b, h_I\rangle h_I \left(\sum_{K:K\supset I} \sum_{J:J\subset K}^{(i)} \sum_{J':J'\subset K}^{(j)} a_{JJ'K}\langle f, h_J\rangle h_{J'} \right).$$

Comparing the inner parentheses to the definition of S^{ij} suggests that the expression above is equal to

$$\sum_I \langle b, h_I\rangle h_I \sum_{J':J'^{(j)}\supset I} \langle S^{ij} f, h_{J'}\rangle h_{J'}$$
$$= \sum_I \sum_{J':J'\supsetneq I} \langle b, h_I\rangle\langle S^{ij} f, h_{J'}\rangle h_I h_{J'} + \sum_I \sum_{J':J'\subset I\subset J'^{(j)}} \langle b, h_I\rangle\langle S^{ij} f, h_{J'}\rangle h_I h_{J'} =: I + II.$$

Note that there are only part I and II left because of the supports of Haar functions. For part I, one writes

$$I = \sum_I \langle b, h_I\rangle h_I \left(\sum_{J':J'\supsetneq I} \langle S^{ij} f, h_{J'}\rangle h_{J'} \right) = \sum_I \langle b, h_I\rangle h_I \langle S^{ij} f, h_I^1\rangle h_I^1$$
$$= \sum_I \langle b, h_I\rangle\langle S^{ij} f, h_I^1\rangle h_I |I|^{-1/2},$$

which is of type $B_0(b, S^{ij} f)$. In order to deal with part II, observe that it can be decomposed into finitely many pieces depending on the relative sizes of I and J', i.e.

$$II = \sum_{k=0}^{j} \sum_{J'} \langle b, h_{J'^{(k)}}\rangle\langle S^{ij} f, h_{J'}\rangle h_{J'^{(k)}} h_{J'}$$
$$= \sum_{k=0}^{j} \sum_{J'} \beta_{J'} \langle b, h_{J'^{(k)}}\rangle\langle S^{ij} f, h_{J'}\rangle h_{J'} |J'^{(k)}|^{-1/2} = \sum_{k=0}^{j} B_k(b, S^{ij} f),$$

where $\beta_{J'} \in \{1, -1\}$ and $0 \le k \le j$. Note that the sum at the end contains only $1 + j \le 1 + \max(i, j)$ terms. Therefore, the representation of the first term is demonstrated.

Second term

Now we turn to the second term that contains $S^{ij}(h_I h_J)$. Due to the supports of Haar functions, this part is nontrivial only when $I \cap J \neq \emptyset$. Hence, one can split this term into three parts: $I \subsetneq J$, $I = J$, and $J \subsetneq I \subset J^{(i)}$.

For $I \subsetneq J$, note that the second term becomes

$$\begin{aligned} S^{ij}\left(\sum_{I \subsetneq J} \langle b, h_I\rangle \langle f, h_J\rangle h_I h_J\right) &= S^{ij}\left(\sum_{I} \langle b, h_I\rangle h_I \sum_{J: J \supsetneq I} \langle f, h_J\rangle h_J\right) \\ &= S^{ij}\left(\sum_{I} \langle b, h_I\rangle h_I \langle f, h_I^1\rangle h_I^1\right) \\ &= S^{ij}\left(\sum_{I} \langle b, h_I\rangle \langle f, h_I^1\rangle h_I |I|^{-1/2}\right), \end{aligned}$$

which is $S^{ij}(B_0(b, f))$.

As the diagonal part $I = J$ is obviously of the form $S^{ij}(B_0(b, f))$ already, we move on to the last piece $J \subsetneq I \subset J^{(i)}$, which can be written as

$$S^{ij}\left(\sum_{J} \sum_{I: J \subsetneq I \subset J^{(i)}} \langle b, h_I\rangle \langle f, h_J\rangle h_I h_J\right).$$

Observe that what's inside the parentheses is of an almost identical form as part II that appeared at the end of the discussion of the first term except that j is replaced by i and that f takes the place of $S^{ij} f$. Hence, the same reasoning implies that it is a sum of at most $i \leq \max(i, j)$ terms of $S^{ij}(B_k(b, f))$, $1 \leq k \leq i$. This proves the representation of the second term as well as completes the discussion of the case when S^{ij} is cancellative.

4.2.2 Noncancellative dyadic shift S^{00}

It suffices to assume that

$$S^{00}f = \sum_I a_I \langle f, h_I^1 \rangle h_I,$$

where $a_I := \langle a, h_I \rangle |I|^{-1/2}$ with $\|a\|_{\mathrm{BMO}} \le 1$. Because if we switch the positions of cancellative and noncancellative Haar functions, what we obtain is none other than its adjoint. Moreover, for the Haar expansion

$$[b, S^{00}]f = \sum_{I,J} \langle b, h_I \rangle \langle f, h_J \rangle [h_I, S^{00}] h_J,$$

it is not hard to see, according to a discussion similar to the one at the beginning of the case $(i, j) \neq (0, 0)$, that one needs only to consider the part $I \subset J$ thanks to the commutator structure. We then split the sum into two parts: $I \subsetneq J$ and $I = J$.

Part $I \subsetneq J$

To decompose this part, once again we consider the first term containing $h_I S^{00} h_J$ and the second term containing $S^{00}(h_I h_J)$ separately, without need to exploit more of the cancellation of the commutator. The second term can be dealt with exactly the same as how we treated the $I \subsetneq J$ part of the second term in the case $(i, j) \neq (0, 0)$, which we omit. To study the first term, one observes that for any h_J,

$$S^{00} h_J = \sum_{I : I \subsetneq J} a_I \langle h_J, h_I^1 \rangle h_I = \sum_{I : I \subsetneq J} a_I |I|^{1/2} h_I h_J.$$

Hence, the first term becomes

$$\sum_J \sum_{I, I' \subsetneq J} \langle b, h_I \rangle h_I \langle f, h_J \rangle a_{I'} |I'|^{1/2} h_{I'} h_J = \sum_J \sum_{I \subset I' \subsetneq J} + \sum_J \sum_{I' \subsetneq I \subsetneq J} =: I + II.$$

One writes

$$
\begin{aligned}
I &= \sum_I \langle b, h_I\rangle h_I \left(\sum_{I':I\subset I'} \sum_{J:I'\subsetneq J} a_{I'} \langle f, h_J\rangle h_J |I'|^{1/2} h_{I'} \right) \\
&= \sum_I \langle b, h_I\rangle h_I \left(\sum_{I':I\subset I'} a_{I'} |I'|^{1/2} h_{I'} \langle f, h^1_{I'}\rangle h^1_{I'} \right) \\
&= \sum_I \langle b, h_I\rangle h_I \left(\sum_{I':I\subset I'} a_{I'} \langle f, h^1_{I'}\rangle h_{I'} \right) \\
&= \sum_I \langle b, h_I\rangle h_I \left(\sum_{I':I\subset I'} \langle S^{00} f, h_{I'}\rangle h_{I'} \right) \\
&= \sum_I \langle b, h_I\rangle h_I \langle S^{00} f, h_I\rangle h_I + \sum_I \langle b, h_I\rangle h_I \langle S^{00} f, h^1_I\rangle h^1_I \\
&= \sum_I \beta_I \langle b, h_I\rangle \langle S^{00} f, h_I\rangle h^\epsilon_I |I|^{-1/2} + \sum_I \langle b, h_I\rangle \langle S^{00} f, h^1_I\rangle h_I |I|^{-1/2},
\end{aligned}
$$

which is the sum of two $B_0(b, S^{00} f)$ with $\beta_I \in \{1, -1\}$.

To deal with part II, by summing over index J, one obtains

$$
\begin{aligned}
II &= \sum_{I'\subsetneq I} \langle b, h_I\rangle h_I a_{I'} |I'|^{1/2} h_{I'} \langle f, h^1_I\rangle h^1_I \\
&= \sum_{I'} a_{I'} |I'|^{1/2} h_{I'} \left(\sum_{I:I\supsetneq I'} \langle b, h_I\rangle |I|^{-1/2} \langle f, h^1_I\rangle h_I \right).
\end{aligned}
$$

Define operator $S_b f := \sum_I \langle b, h_I\rangle |I|^{-1/2} \langle f, h^1_I\rangle h_I$, which is a classical paraproduct $B_0(b, f)$, the above is equal to

$$
\begin{aligned}
&=: \sum_{I'} a_{I'} |I'|^{1/2} h_{I'} \sum_{I:I\supsetneq I'} \langle S_b f, h_I\rangle h_I \\
&= \sum_{I'} a_{I'} \langle S_b f, h^1_{I'}\rangle h_{I'} = S^{00}(S_b f),
\end{aligned}
$$

which completes the discussion of part $I \subsetneq J$.

Part $I = J$

In this special case, what we try to decompose becomes

$$\sum_{I} \sum_{\epsilon,\epsilon' \in \{0,1\}^d \setminus \{\vec{1}\}} \langle b, h_I^\epsilon \rangle \langle f, h_I^{\epsilon'} \rangle \left(h_I^\epsilon S^{00} h_I^{\epsilon'} - S^{00}(h_I^\epsilon h_I^{\epsilon'}) \right). \tag{4.17}$$

Here, in order to avoid possible confusion, we wrote out the sum over index ϵ, ϵ' explicitly. Recall that for each cube I, there are 2^d different Haar functions associated: $\{h_I^\epsilon\}$, $\epsilon \in \{0,1\}^d$, and the Haar function is non-cancellative if and only if $\epsilon = \vec{1}$. First, it is useful to observe that if $\epsilon \neq \epsilon'$, $[h_I^\epsilon, S^{00}]h_I^{\epsilon'} = 0$. Indeed, for any fixed I and ϵ, ϵ',

$$h_I^\epsilon S^{00} h_I^{\epsilon'} = \sum_{J : J \subsetneq I} a_J |J|^{1/2} h_J (h_I^\epsilon h_I^{\epsilon'}),$$

and

$$S^{00}(h_I^\epsilon h_I^{\epsilon'}) = \sum_{J : J \supset I} a_J |J|^{-1/2} h_J \left(\int_I h_I^\epsilon h_I^{\epsilon'} \right) + \sum_{J : J \subsetneq I} a_J |J|^{1/2} h_J (h_I^\epsilon h_I^{\epsilon'}).$$

As a result of cancellation and the fact that $\int_I h_I^\epsilon h_I^{\epsilon'}$ is nonzero if and only if $\epsilon = \epsilon'$, i.e. $h_I^\epsilon h_I^{\epsilon'} = |I|^{-1} \chi_I$, $[h_I^\epsilon, S^{00}]h_I^{\epsilon'} \neq 0$ only when $\epsilon = \epsilon'$. Therefore, one can safely suppress the dependence on ϵ when studying this part of the sum.

Furthermore, it is easily seen that the second term containing $S^{00}(h_I h_I)$ here can be estimated exactly the same as before, it thus suffices to deal with the first term containing $h_I S^{00} h_I$, which is equal to

$$\sum_I \langle b, h_I \rangle \langle f, h_I \rangle h_I S^{00} h_I = \sum_I \langle b, h_I \rangle \langle f, h_I \rangle |I|^{-1} \sum_{J : J \subsetneq I} \langle a, h_J \rangle h_J = P(b, a, f),$$

hence the proof is complete.

4.3 Upper bound for commutators of CZ operators with iteration

In this section, we are going to prove Theorem 3.12 in the general case by iterating the one-parameter argument (Theorem 4.13) in the previous section. For the sake of brevity, we consider the bi-parameter case as an example, while the strategy can be easily generalized to work for arbitrarily many parameters. Hence, for any $b \in \mathrm{BMO}_{\mathrm{prod}}(\mathbb{R}^n \times \mathbb{R}^m)$ and Calderón-Zygmund operators T_1, T_2, acting in the first and second variable respectively, we are going to prove for any $f \in L^2(\mathbb{R}^n \times \mathbb{R}^m)$ that

$$\|[[b, T_1], T_2]\|_{L^2} \lesssim \|b\|_{\mathrm{BMO}_{\mathrm{prod}}} \|f\|_{L^2}. \tag{4.18}$$

The general L^p case for $p \neq 2$ will be discussed in Appendix B.

The main idea is to show that, after reduced to the dyadic shift case, the commutator can be represented as a finite linear combination of the bi-parameter analogs of terms that appear in Theorem 4.13, i.e. the bi-parameter paraproduct-like operators $BB_{k,l}$, PP, etc.

Applying Theorem 2.13 for both variables implies that

$$\begin{aligned}
&[[b, T_1], T_2]f \\
&= c\|T_1\|_{CZ}\|T_2\|_{CZ}\mathbb{E}_{\omega_1}\mathbb{E}_{\omega_2} \sum_{i_1, j_1=0}^{\infty} \sum_{i_2, j_2=0}^{\infty} 2^{-\max(i_1, j_1)\frac{\delta}{2}} 2^{-\max(i_2, j_2)\frac{\delta}{2}} [[b, S_{\omega_1}^{i_1 j_1}], S_{\omega_2}^{i_2 j_2}]f.
\end{aligned}$$

Since our estimate in the following doesn't depend on the parameters ω_1, ω_2 explicitly, we will omit them in the notation. Our goal is to prove that

$$\begin{aligned}
&\left\|[[b, S_1^{i_1 j_1}], S_2^{i_2 j_2}]f\right\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)} \\
&\lesssim (1 + \max(i_1, j_1))(1 + \max(i_2, j_2))\|b\|_{\mathrm{BMO}_{\mathrm{prod}}(\mathbb{R}^n \times \mathbb{R}^m)}\|f\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)},
\end{aligned} \tag{4.19}$$

which can be achieved by showing that any $[[b, S_1^{i_1 j_1}], S_2^{i_2 j_2}]f$ can be represented as a finite linear combination of the following terms and their adjoints (which is understood as the adjoint operator with b, a^i fixed):

$$BB_{k,l}(b, S_1^{i_1j_1}S_2^{i_2j_2}f), \quad S_1^{i_1j_1}(BB_{k,l}(b, S_2^{i_2j_2}f)),$$

$$BP_k(b, a^2, S_1^{i_1j_1}f), \quad PB_l(b, a^1, S_2^{i_2j_2}f),$$

$$PP(b, a^1 \otimes a^2, f), \quad PP_1(b, a^1 \otimes a^2, f).$$

where $k, l \geq 0$, and a^i is the BMO symbol of the dyadic shift S^{00} if it appears in the i^{th} variable. The total number of terms in the representation is no greater than $C(1 + \max(i_1, j_1))(1 + \max(i_2, j_2))$ for some universal constant C. Note that for $a^1 \in \text{BMO}(\mathbb{R}^n)$ and $a^2 \in \text{BMO}(\mathbb{R}^m)$, there holds $a^1 \otimes a^2 \in \text{BMO}_{\text{prod}}(\mathbb{R}^n \times \mathbb{R}^m)$, as discussed in Observation 3.4. Hence, implied by Theorem 2.16, Lemma 4.4, Lemma 4.5 and Lemma 4.8, the L^2 norm of all of the terms above are uniformly bounded, independent of k, l in particular. Therefore, (4.19) is implied immediately.

In order to derive such representation, we argue by an iteration of Theorem 4.13.

4.3.1 Cancellative dyadic shifts $S_1^{i_1j_1}$ and $S_2^{i_2j_2}$

In the case when both $S_1^{i_1j_1}$ and $S_2^{i_2j_2}$ are cancellative, only operators $BB_{k,l}$ need to be involved. In order to make the notations clear, in the following, we will use B_k^τ to denote the one-parameter paraproducts that appeared in the previous section for the τ-th variable, where $k \geq 0$ and $\tau = 1, 2$. Calculation shows that

$$[[b, S_1^{i_1j_1}], S_2^{i_2j_2}]f = \sum_{I_1, J_1} \sum_{I_2, J_2} \langle b, h_{I_1} \otimes u_{I_2} \rangle \langle f, h_{J_1} \otimes u_{J_2} \rangle [h_{I_1}, S_1^{i_1j_1}] h_{J_1} \otimes [u_{I_2}, S_2^{i_2j_2}] u_{J_2},$$

which by iteration equals

$$\sum_{I_2,J_2}\Big(\sum_{t_1\in\Lambda_1} B^1_{k,t_1}(\langle b,u_{I_2}\rangle_2, S_1^{i_1j_1}(\langle f,u_{J_2}\rangle_2)) + \sum_{t_2\in\Lambda_2} S_1^{i_1j_1}(B^1_{k,t_2}(\langle b,u_{I_2}\rangle_2,\langle f,u_{J_2}\rangle_2))\Big)\otimes\Big([u_{I_2},S_2^{i_2j_2}]u_{J_2}\Big),$$

where B^1_{k,t_i} are paraproducts of type B^1_k in the first variable, and for each t_i, k is an arbitrary nonnegative integer. Note that in the first parentheses we have a finite linear combination of terms that have already been studied in the previous section, and all of the index set Λ_i satisfy $|\Lambda_i|\le C(1+\max(i_1,j_1))$, $i=1,2$. Since the terms inside the first parentheses can be treated similarly, let's study one of the terms B^1_{k,t_1} as an example. We will also omit the subscript t_1 as the choice is arbitrary. Then, the sum corresponding to B^1_k is equal to

$$\begin{aligned}
&\sum_{I_2,J_2} B^1_k\Big(\langle b,u_{I_2}\rangle_2, S_1^{i_1j_1}(\langle f,u_{J_2}\rangle_2)\Big)\otimes\Big([u_{I_2},S_2^{i_2j_2}]u_{J_2}\Big)\\
&=\sum_{I_2,J_2}\sum_{I_1}\beta_{I_1}\langle b, h_{I_1^{(k)}}\otimes u_{I_2}\rangle\langle S_1^{i_1j_1}(f), h^{\epsilon_1}_{I_1}\otimes u_{J_2}\rangle h^{\epsilon_1'}_{I_1}|I_1^k|^{-1/2}\otimes\Big([u_{I_2},S_2^{i_2j_2}]u_{J_2}\Big)\\
&=\sum_{I_1}\beta_{I_1}h^{\epsilon_1'}_{I_1}|I_1^{(k)}|^{-1/2}\otimes\Big([\langle b,h_{I_1^{(k)}}\rangle_1, S_2^{i_2j_2}]\langle S_1^{i_1j_1}f, h^{\epsilon_1}_{I_1}\rangle_1\Big)\\
&=\sum_{I_1}\beta_{I_1}h^{\epsilon_1'}_{I_1}|I_1^{(k)}|^{-1/2}\otimes\Big(\sum_{s_1\in\Gamma_1}B^2_{l,s_1}(\langle b,h_{I_1^{(k)}}\rangle_1, S_2^{i_2j_2}(\langle S_1^{i_1j_1}(f),h^{\epsilon_1}_{I_1}\rangle_1))\\
&\qquad+\sum_{s_2\in\Gamma_2}S_2^{i_2j_2}(B^2_{l,s_2}(\langle b,h_{I_1^{(k)}}\rangle_1,\langle S_1^{i_1j_1}(f),h^{\epsilon_1}_{I_1}\rangle_1))\Big),
\end{aligned}$$

where the last step above follows by applying Theorem 4.13 in the second variable, B^2_{l,s_i} are paraproducts of type B^2_l in the second variable, and all the index sets Γ_i satisfy $|\Gamma_i|\le C(1+\max(i_2,j_2))$, $i=1,2$. Again, since all the terms in the parentheses are similar, we only consider one of B^2_{l,s_2} and omit the subscript s_2. This is a mixed case, and all the other combinations follow similarly. Thus, noticing that

$$\sum_{I_1} \beta_{I_1} h_{I_1}^{\epsilon_1'} |I_1^{(k)}|^{-1/2} \otimes S_2^{i_2 j_2} \left(B_l^2(\langle b, h_{I_1^{(k)}} \rangle_1, \langle S_1^{i_1 j_1}(f), h_{I_1}^{\epsilon_1} \rangle_1) \right)$$
$$= S_2^{i_2 j_2} \left(\sum_{I_1, I_2} \beta_{I_1} \beta_{I_2} \langle b, h_{I_1^{(k)}} \otimes u_{I_2^{(l)}} \rangle \langle S_1^{i_1 j_1} f, h_{I_1}^{\epsilon_1} \otimes u_{I_2}^{\epsilon_2} \rangle h_{I_1}^{\epsilon_1'} \otimes u_{I_2}^{\epsilon_2'} |I_1^{(k)}|^{-\frac{1}{2}} |I_2^{(l)}|^{-\frac{1}{2}} \right)$$

is exactly $S_2^{i_2 j_2}(BB_{k,l}(b, S_1^{i_1 j_1} f))$, where $BB_{k,l}$ is the bi-parameter paraproduct with descendants we've studied in Lemma 4.4, and the only case involving non-cancellative Haar functions is when the corresponding k or l is 0. We therefore obtain the desired representation of this term. All the other terms can be treated similarly, by noticing that paraproducts $BB_{k,l}$ can be obtained by combining B_k^1 and B_l^2 through the same process described above. And it is easily seen that the total number of terms is bounded by $(1 + \max(i_1, j_1))(1 + \max(i_2, j_2))$ up to a dimensional constant.

4.3.2 Cancellative dyadic shift $S_1^{i_1 j_1}$ and non-cancellative dyadic shift S_2^{00}

We assume that $S_2^{00} f = \sum_{I_2} \langle a^2, u_{I_2} \rangle_2 |I_2|^{-1/2} \langle f, u_{I_2}^1 \rangle_2 u_{I_2}$. Following from Theorem 4.13, in the first variable, the commutator can be represented as a linear combination of paraproducts, i.e.

$$[[b, S_1^{i_1 j_1}], S_2^{00}] f = \sum_{I_1 \subset J_1^{(i_1)}} \sum_{I_2 \subset J_2} \langle b, h_{I_1} \otimes u_{I_2} \rangle \langle f, h_{J_1} \otimes u_{J_2} \rangle [h_{I_1}, S_1^{i_1 j_1}] h_{J_1} \otimes [u_{I_2}, S_2^{00}] u_{J_2}$$
$$= \sum_{I_2 \subset J_2} \Big(\sum_{t_1 \in \Lambda_1} B_{k,t_1}^1(\langle b, u_{I_2} \rangle_2, S_1^{i_1 j_1}(\langle f, u_{J_2} \rangle_2))$$
$$+ \sum_{t_2 \in \Lambda_2} S_1^{i_1 j_1}(B_{k,t_2}^1(\langle b, u_{I_2} \rangle_2, \langle f, u_{J_2} \rangle_2)) \Big) \otimes \left([u_{I_2}, S_2^{00}] u_{J_2} \right).$$

Recall that by Theorem 4.13, in the one-parameter setting, the non-cancellative dyadic shift S^{00} can be represented as a finite linear combination of paraproducts (corresponding to the sum over $I \subsetneq J$ and the second term in the sum over $I = J$) and operator P

(corresponding to the first term in the sum over $I = J$). Hence,

$$
\begin{aligned}
&\sum_{I_2\subset J_2} B^1_{k,t_1}\Big(\langle b, u_{I_2}\rangle_2, S_1^{i_1j_1}(\langle f, u_{J_2}\rangle_2)\Big)\otimes[u_{I_2}, S_2^{00}]u_{J_2}\\
=&\sum_{I_1}\beta_{I_1}h_{I_1}^{\epsilon_1'}|I_1^{(k)}|^{-1/2}\otimes\Big([\langle b, h_{I_1^{(k)}}\rangle_1, S_2^{00}]\langle S_1^{i_1j_1}f, h_{I_1}^{\epsilon_1}\rangle_1\Big)\\
=&\sum_{I_1}\beta_{I_1}h_{I_1}^{\epsilon_1'}|I_1^{(k)}|^{-1/2}\otimes\Big(\sum_{s_1\in\Gamma_1}B^2_{0,s_1}\big(\langle b, h_{I_1^{(k)}}\rangle_1, S_2^{00}(\langle S_1^{i_1j_1}f, h_{I_1}^{\epsilon_1}\rangle_1)\big)\\
&\qquad+\sum_{s_2\in\Gamma_2}S_2^{00}\big(B^2_{0,s_2}(\langle b, h_{I_1^{(k)}}\rangle_1, \langle S_1^{i_1j_1}f, h_{I_1}^{\epsilon_1}\rangle_1)\big)+P\big(\langle b, h_{I_1^{(k)}}\rangle_1, a^2, \langle S_1^{i_1j_1}f, h_{I_1}^{\epsilon_1}\rangle_1\big)\Big)\\
=&\left(\sum_{s_1\in\Gamma_1}BB_{k,0,s_1}(b, S_1^{i_1j_1}S_2^{00}f)\right)+\left(\sum_{s_2\in\Gamma_2}S_2^{00}\Big(BB_{k,0,s_2}(b, S_1^{i_1j_1}f)\Big)\right)+BP_k(b, a^2, S_1^{i_1j_1}f).
\end{aligned}
$$

Similarly, the other term can be treated exactly the same:

$$
\begin{aligned}
&\sum_{I_2\subset J_2} S_1^{i_1j_1}\left(B^1_{k,t_2}\big(\langle b, u_{I_2}\rangle_2, \langle f, u_{J_2}\rangle_2\big)\right)\otimes[u_{I_2}, S_2^{00}]u_{J_2}\\
=&\Big(\sum_{s_1\in\Gamma_1}S_1^{i_1j_1}\big(BB_{k,0,s_1}(b, S_2^{00}f)\big)\Big)+\Big(\sum_{s_2\in\Gamma_2}S_1^{i_1j_1}S_2^{00}\big(BB_{k,0,s_2}(b, f)\big)\Big)\\
&\qquad+S_1^{i_1j_1}\big(BP_k(b, a^2, f)\big).
\end{aligned}
$$

The desired representation is hence obtained. Note that by symmetry and duality, this implies the boundedness of other types of the mixed cases as well.

4.3.3 Non-cancellative dyadic shfits S_1^{00} and S_2^{00}

$$
[[b, S_1^{00}], S_2^{00}]f=\sum_{I_1\subset J_1}\sum_{I_2\subset J_2}\langle b, h_{I_1}\otimes u_{I_2}\rangle\langle f, h_{J_1}\otimes u_{J_2}\rangle[h_{I_1}, S_1^{00}]u_{J_1}\otimes[h_{I_2}, S_2^{00}]u_{J_2}.
$$

First, we deal with the case when both S_1^{00} and S_2^{00} are of the same type, for instance,

$$
S_1^{00}f:=\sum_{I_1}\langle a^1, h_{I_1}\rangle_1|I_1|^{-1/2}\langle f, h_{I_1}^1\rangle h_{I_1},\quad S_2^{00}f:=\sum_{I_2}\langle a^2, u_{I_2}\rangle_2|I_2|^{-1/2}\langle f, u_{I_2}^1\rangle_2 u_{I_2}.
$$

Observe that compared with section 4.3.1 and 4.3.2, after decomposing the commutator in each variable into paraproducts and operator P, the only new case that arises here is the "tensor product" of operator P in both variables, which is equal to

$$\sum_{I_1,I_2}\langle b,h_{I_1}\otimes u_{I_2}\rangle\langle f,h_{I_1}\otimes u_{I_2}\rangle|I_1|^{-1}|I_2|^{-1}\sum_{J_1:J_1\subsetneq I_1}\sum_{J_2:J_2\subsetneq I_2}\langle a^1\otimes a^2,h_{J_1}\otimes u_{J_2}\rangle h_{J_1}\otimes u_{J_2}$$
$$=PP(b,a^1\otimes a^2,f).$$

Second, we discuss the case when S_1^{00} and S_2^{00} are of different types, for instance,

$$S_1^{00}f:=\sum_{I_1}\langle a^1,h_{I_1}\rangle_1|I_1|^{-1/2}\langle f,h_{I_1}\rangle h_{I_1}^1,\quad S_2^{00}f:=\sum_{I_2}\langle a^2,u_{I_2}\rangle_2|I_2|^{-1/2}\langle f,u_{I_2}^1\rangle_2 u_{I_2}.$$

It is implied by Theorem 4.13 that in the first variable, the commutator is a linear combination of paraproducts and operator P^*. Therefore, the only new case that arises here in the representation is P^* in the first variable mixed with P in the second variable, which is

$$\sum_{I_1,I_2}\langle b,h_{I_1}\otimes u_{I_2}\rangle|I_1|^{-1}|I_2|^{-1}\sum_{J_1:J_1\subsetneq I_1}\sum_{J_2:J_2\subsetneq I_2}\langle a^1\otimes a^2,h_{J_1}\otimes u_{J_2}\rangle\langle f,h_{J_1}\otimes u_{I_2}\rangle h_{I_1}\otimes u_{J_2}$$
$$=PP_1(b,a^1\otimes a^2,f),$$

where PP_1 is the partial adjoint operator of PP in the first variable.

Therefore, Theorem 3.12 in the bi-parameter setting is proved. As a final remark, the proof in the multi-parameter setting proceeds exactly the same as this one. Clearly, in the desired representation of commutators with dyadic shifts, one needs to involve a larger number of basic operators which mix together B_k and P in each variable, but the uniform boundedness of such operators can all be obtained similarly as in Lemma 4.4, 4.5 and 4.8.

4.4 Upper bound for commutators of Journé operators

In previous sections, we have proved Theorem 3.12, which gives a complete solution to the upper bound when the commutator is constructed purely with Calderón-Zygmund operators (not necessarily of convolution type) in each component. We would like to discuss in this section what can be concluded if one allows multi-parameter Journé operators to appear in the commutators. The operators we discuss do not need to be of convolution type. If all the Journé operators involved in the commutator are tensor products of one-parameter CZ operators, then they can be estimated by Theorem 3.13, proved in Subsection 4.4.1. In fact, it turns out that the upper bound for commutators of this type of operators follows quite easily from the CZ operators case.

Otherwise, if some of the operators involved are not tensor products, then the currently best upper bound estimate of commutators of this type is given by Theorem 3.16, where one needs to assume that all the non tensor product type T's are paraproduct free. We will give a proof of this result in Subsection 4.4.2. The reason why the paraproduct free assumption is indispensable in our argument lies in the main ingredient of the proof: multi-parameter representation theorems of Journé operators via dyadic shifts (Martikainen's Theorem 2.14 in bi-parameter or Theorem 7.2 in multi-parameter that will be proven in Chapter 7). For example, in the bi-parameter setting, recall that T is called paraproduct free if

$$T(1 \otimes \cdot) = T(\cdot \otimes 1) = T^*(1 \otimes \cdot) = T^*(\cdot \otimes 1) = 0.$$

Applying Theorem 2.14, since T is paraproduct free, one can conclude from observing the proof of Martikainen's theorem that all the dyadic shifts in the representation are cancellative. In other words, there are no dyadic paraproduct terms appearing in the representation formula, which will essentially bring us back to the "tensor product" case, since "locally" in a dyadic sense, cancellative dyadic shifts look as if they were of tensor product type.

In general, suppose T is any bi-parameter Jounré operator, not necessarily paraproduct free. In order to derive upper bound for commutators involving T, according to Theorem 2.14 and similarly as in the argument presented in Sections 4.2 and 4.3, it suffices to bound commutators of bi-parameter dyadic shifts. The case for cancellative dyadic shfits is essentially what one deals with in Subsection 4.4.2, while the case for non-cancellative dyadic shfits can further split into three cases (modulo symmetries and adjoints), depending on the specific type of the non-cancellative shfits as defined in the following:

The standard paraproduct

$$\Pi_1(f) := \sum_{I_1,I_2} a_{I_1 I_2} \langle f, h_{I_1} \otimes u_{I_2} \rangle h^1_{I_1} \otimes u^1_{I_2},$$

the mixed paraproduct

$$\Pi_2(f) := \sum_{I_1,I_2} a_{I_1,I_2} \langle f, h_{I_1} \otimes u^1_{I_2} \rangle h^1_{I_1} \otimes u_{I_2},$$

and the partial paraproduct for any fixed $i, j \geq 0$

$$\Pi_3^{ij}(f) := \sum_{K_1} \sum_{I_1,J_1 \subset K_1}^{(i,j)} \sum_{K_2} a_{I_1 J_1 K_1 K_2} \langle f, h_{I_1} \otimes u_{K_2} \rangle h_{J_1} \otimes u^1_{K_2},$$

where in the first two operators, $a_{I_1 I_2} = \langle a, h_{I_1} \otimes h_{I_2} \rangle |I_1|^{-1/2} |I_2|^{-1/2}$ for some fixed product BMO symbol function a such that $\|a\|_{\text{BMO}_{\text{prod}}} \leq 1$, which also implies that $|a_{I_1 I_2}| \leq 1$. In the last operator, the notation $\sum_{I_1,J_1 \subset K_1}^{(i,j)}$ denotes the sum over all the sub cubes I_1, J_1 contained in K_1 such that $\ell(I_1) = 2^{-i}\ell(K_1), \ell(J_1) = 2^{-j}\ell(K_1)$, and for any fixed I_1, J_1, K_1, there exists a BMO function $a_{I_1 J_1 K_1}$ in the second variable such that $a_{I_1 J_1 K_1 K_2} = \langle a_{I_1 J_1 K_1}, h_{K_2} \rangle_2 |K_2|^{-1/2}$ with $\|a_{I_1 J_1 K_1}\|_{\text{BMO}} \leq \frac{|I_1|^{1/2} |J_1|^{1/2}}{|K_1|}$.

In Subsection 4.4.3, we will show that the upper bound of commutators of standard paraproducts holds true, which brings us one step closer to a complete solution of the upper bound problem of general Journé commutators. However, unfortunately, the cases of mixed and partial paraproducts seem significantly more difficult. Some detailed discussion will be made in the end of the subsection.

Summarizing all of the results above concerning multi-parameter Journé commutators, we have the following corollary, which records the so far best upper bound estimate.

Corollary 4.20. *Consider* $\mathbb{R}^{\vec{d}}$, $\vec{d} = (d_1, \dots, d_t)$ *with a partition* $\mathcal{I} = (I_s)_{1\le s\le l}$ *of* $\{1, \dots, t\}$ *as discussed before. Let* $b \in BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ *and let* T_s *denote a (one-parameter) Calderón-Zygmund or (multi-parameter) Journé operator acting on functions defined on* $\bigotimes_{k\in I_s} \mathbb{R}^{d_k}$ *such that for some partition* $\{I_{s,j}\}_{1\le j\le w_s}$ *of* I_s *with* $1 \le w_s \le |I_s|$, $T_s = \bigotimes_{j=1}^{w_s} T_{s,j}$, *where each* $T_{s,j}$ *is of the following three types: 1. a Calderón-Zygmund operator; 2. a paraproduct free Journé operator; 3. a multi-parameter standard paraproduct or its adjoint. Then we have the estimate below*

$$\|[\dots[b, T_1], \dots, T_l]\|_{L^2(\mathbb{R}^{\vec{d}})\to L^2(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

Before moving on to the proofs, we would like to mention that we will stick to the bi-parameter setting in most cases in the following, as all the arguments iterate well and the multi-parameter proofs are completely the same.

4.4.1 Journé operators of tensor product type

When all the Journé operators in the commutator are of tensor product type (like the cases of Hilbert and Riesz commutators), the upper bound estimate (Theorem 3.13) can be easily deduced from that of the one-parameter Calderón-Zygmund commutators (Theorem 3.12) proved in Section 4.3. Indeed, observing that

$$[b, T_1T_2] = [b, T_1]T_2 + T_1[b, T_2], \tag{4.21}$$

commutator of tensor products splits into finitely many parts, each of which being a composition of CZ operators and commutators of CZ operators. Since CZ operators are all bounded on L^p, each part is bounded according to Theorem 3.12. In the following, to give the readers some flavor of how our argument iterates, we give the proof of Theorem 3.13 in the case where the BMO space involved is $\mathrm{BMO}_{(13)(2)}$. It will be clear through the argument that more parameters or iterates pose no extra difficulties. Let T_k, $k=1,2,3$ be three CZ operators, with T_k acting in the k^{th} variable of $\mathbb{R}^{\vec{d}} = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \times \mathbb{R}^{d_3}$. Let $b \in \mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})$, i.e. $b(\cdot,\cdot,x_3)$ and $b(x_1,\cdot,\cdot)$ are uniformly in product BMO. One wants to show that

$$\|[[b, T_1T_3], T_2]f\|_{L^p} \lesssim \|b\|_{\mathrm{BMO}_{(13)(2)}} \|f\|_{L^p}, \quad \forall f \in L^p,\ 1<p<\infty. \tag{4.22}$$

Proof. Applying (4.21), one has

$$[[b, T_1T_3], T_2]f = [[b, T_1]T_3, T_2]f + [T_1[b, T_3], T_2]f.$$

Our strategy is to show that each of the two terms above is bounded. First,

$$\begin{aligned}\|[[b, T_1]T_3, T_2]f\|^p_{L^p(\mathbb{R}^{\vec{d}})} &= \int_{\mathbb{R}^{d_3}} \|[[b(\cdot,\cdot,x_3), T_1]T_3, T_2]f\|^p_{L^p(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})}\, dx_3 \\ &= \int_{\mathbb{R}^{d_3}} \|[[b(\cdot,\cdot,x_3), T_1], T_2](T_3f(\cdot,\cdot,x_3))\|^p_{L^p(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})}\, dx_3,\end{aligned}$$

then Theorem 3.12 implies that

$$\begin{aligned}&\lesssim \int_{\mathbb{R}^{d_3}} \|b(\cdot,\cdot,x_3)\|^p_{\mathrm{BMO}_{\mathrm{prod}}(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})} \|(T_3f)(\cdot,\cdot,x_3)\|^p_{L^p(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})}\, dx_3 \\ &\le \|b\|^p_{\mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})} \int_{\mathbb{R}^{d_1}\times\mathbb{R}^{d_2}} \|(T_3f)(x_1,x_2,\cdot)\|^p_{L^p(\mathbb{R}^{d_3})}\, dx_1dx_2 \lesssim \|b\|^p_{\mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})} \|f\|^p_{L^p(\mathbb{R}^{\vec{d}})}.\end{aligned}$$

The second term is even simpler,

$$\begin{aligned}&\|[T_1[b, T_3], T_2]f\|^p_{L^p(\mathbb{R}^{\vec{d}})} \lesssim \|[[b, T_3], T_2]f\|^p_{L^p(\mathbb{R}^{\vec{d}})} \\ &= \int_{\mathbb{R}^{d_1}} \|[[b(x_1,\cdot,\cdot), T_3], T_2]f(x_1,\cdot,\cdot)\|^p_{L^p(\mathbb{R}^{d_2}\times\mathbb{R}^{d_3})}\, dx_1 \\ &\lesssim \int_{\mathbb{R}^{d_1}} \|b(x_1,\cdot,\cdot)\|^p_{\mathrm{BMO}_{\mathrm{prod}}(\mathbb{R}^{d_2}\times\mathbb{R}^{d_3})} \|f(x_1,\cdot,\cdot)\|^p_{L^p(\mathbb{R}^{d_2}\times\mathbb{R}^{d_3})}\, dx_1 \le \|b\|^p_{\mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})} \|f\|^p_{L^p(\mathbb{R}^{\vec{d}})}.\end{aligned}$$

□

4.4.2 Paraproduct free Journé operators

First, we prove Theorem 3.15, which is a base case of Theorem 3.16. Let T be a bi-parameter paraproduct free Journé operator in Martikainen's class and $b \in \mathrm{bmo}(\mathbb{R}^n \times \mathbb{R}^m)$. One wants to show that

$$\|[b,T]\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)\to L^2(\mathbb{R}^n\times\mathbb{R}^m)} \lesssim \|b\|_{\mathrm{bmo}(\mathbb{R}^n\times\mathbb{R}^m)}, \tag{4.23}$$

with the underlying constant depends only on the characterizing constants of T.

According to Theorem 2.14, for any sufficiently nice functions f, g, one has the following representation:

$$\langle Tf, g\rangle = C\mathbb{E}_{\omega_1}\mathbb{E}_{\omega_2}\sum_{i_1,j_1=0}^{\infty}\sum_{i_2,j_2=0}^{\infty} 2^{-\max(i_1,j_1)}2^{-\max(i_2,j_2)}\langle S^{i_1j_1i_2j_2}f, g\rangle, \tag{4.24}$$

where the expectation is with respect to a certain parameter of the dyadic grids. Recall that the dyadic shifts $S^{i_1j_1i_2j_2}$ are defined as

$$\begin{aligned} &S^{i_1j_1i_2j_2}f \\ &:= \sum_{K_1\in\mathcal{D}_1}\sum_{\substack{I_1,J_1\subset K_1, I_1,J_1\in\mathcal{D}_1\\ \ell(I_1)=2^{-i_1}\ell(K_1)\\ \ell(J_1)=2^{-j_1}\ell(K_1)}}\sum_{K_2\in\mathcal{D}_2}\sum_{\substack{I_2,J_2\subset K_2, I_2,J_2\in\mathcal{D}_2\\ \ell(I_2)=2^{-i_2}\ell(K_2)\\ \ell(J_2)=2^{-j_2}\ell(K_2)}} a_{I_1J_1K_1I_2J_2K_2}\langle f, h_{I_1}\otimes u_{I_2}\rangle h_{J_1}\otimes u_{J_2} \\ &=: \sum_{K_1}\sum_{I_1,J_1\subset K_1}^{(i_1,j_1)}\sum_{K_2}\sum_{I_2,J_2\subset K_2}^{(i_2,j_2)} a_{I_1J_1K_1I_2J_2K_2}\langle f, h_{I_1}\otimes u_{I_2}\rangle h_{J_1}\otimes u_{J_2}. \end{aligned}$$

The coefficients above satisfy $a_{I_1J_1K_1I_2J_2K_2} \le \frac{\sqrt{|I_1||J_1||I_2||J_2|}}{|K_1||K_2|}$, which also guarantees the normalization $\|S^{i_1j_1i_2j_2}\|_{L^2\to L^2} \le 1$. Moreover, since T is paraproduct free, all the Haar functions appearing above are cancellative. It thus suffices to show that for any dyadic grids $\mathcal{D}_1, \mathcal{D}_2$ and fixed $i_1, j_1, i_2, j_2 \in \mathbb{N}$, one has

$$\|[b, S^{i_1j_1i_2j_2}]f\|_{L^2} \lesssim (1+\max(i_1,j_1))(1+\max(i_2,j_2))\|b\|_{\mathrm{bmo}}\|f\|_{L^2}, \tag{4.25}$$

as the decay factor $2^{-\max(i_1,j_1)}, 2^{-\max(i_2,j_2)}$ in (4.24) will guarantee the convergence of the series.

To see (4.25), one decomposes b and a L^2 test function f using Haar bases:

$$[b, S^{i_1j_1i_2j_2}]f = \sum_{I_1,I_2}\sum_{J_1,J_2} \langle b, h_{I_1}\otimes u_{I_2}\rangle\langle f, h_{J_1}\otimes u_{J_2}\rangle [h_{I_1}\otimes u_{I_2}, S^{i_1j_1i_2j_2}]h_{J_1}\otimes u_{J_2}.$$

A similar argument as that in Section 4.2 implies that $[h_{I_1}\otimes u_{I_2}, S^{i_1j_1i_2j_2}]h_{J_1}\otimes u_{J_2}$ is nonzero only if $I_1 \subset J_1^{(i_1)}$ or $I_2 \subset J_2^{(i_2)}$, where $J_1^{(i_1)}$ denotes the i_1^{th} dyadic ancestor of J_1, similarly for $J_2^{(i_2)}$. Hence, the sum can be decomposed into three parts: $I_1 \subset J_1^{(i_1)}$ and $I_2 \subset J_2^{(i_2)}$ (*regular*), $I_1 \subset J_1^{(i_1)}$ and $I_2 \supsetneq J_2^{(i_2)}$, $I_1 \supsetneq J_1^{(i_1)}$ and $I_2 \subset J_2^{(i_2)}$ (*mixed*).

Regular case

Following Section 4.2 one can decompose the arising sum into sums of classical bi-parameter dyadic paraproducts $B_0(b, f)$ and the paraproducts with descendants

$$BB_{k,l}(b, f) = \sum_{I,J} \beta_{IJ}\langle b, h_{I^{(k)}}\otimes u_{J^{(l)}}\rangle\langle f, h_I^{\varepsilon_1}\otimes u_J^{\varepsilon_2}\rangle h_I^{\varepsilon_1'}\otimes u_J^{\varepsilon_2'}|I^{(k)}|^{-1/2}|J^{(l)}|^{-1/2},$$

where β_{IJ} is a sequence satisfying $|\beta_{IJ}| \le 1$. And it will be observed that when $k > 0$, all Haar functions in the first variable are cancellative, while when $k = 0$, there is at most one of $\{h_I^{\varepsilon_1}\}, \{h_I^{\varepsilon_1'}\}$ being non-cancellative. Same for the second variable. Then, since little BMO functions are contained in product BMO, this part can be well controlled. Specifically, write

$$\begin{aligned}
[b, S^{i_1j_1i_2j_2}]f &= \sum_{I_1,I_2}\sum_{J_1,J_2} \langle b, h_{I_1}\otimes u_{I_2}\rangle\langle f, h_{J_1}\otimes u_{J_2}\rangle h_{I_1}\otimes u_{I_2} S^{i_1j_1i_2j_2}(h_{J_1}\otimes u_{J_2})\\
&\quad - \sum_{I_1,I_2}\sum_{J_1,J_2} \langle b, h_{I_1}\otimes u_{I_2}\rangle\langle f, h_{J_1}\otimes u_{J_2}\rangle S^{i_1j_1i_2j_2}(h_{I_1}h_{J_1}\otimes u_{I_2}u_{J_2})\\
&=: I + II,
\end{aligned}$$

then one can estimate term I and II separately. According to the definition of dyadic shifts, term I is equal to

$$\sum_{J_1,J_2}\sum_{I_1:I_1\subset J_1^{(i_1)}}\sum_{I_2:I_2\subset J_2^{(i_2)}}\langle b,h_{I_1}\otimes u_{I_2}\rangle\langle f,h_{J_1}\otimes u_{J_2}\rangle h_{I_1}\otimes u_{I_2}\cdot$$
$$\Bigg(\sum_{\substack{J_1':J_1'\subset J_1^{(i_1)}\\ \ell(J_1')=2^{i_1-j_1}\ell(J_1)}}\sum_{\substack{J_2':J_2'\subset J_2^{(i_2)}\\ \ell(J_2')=2^{i_2-j_2}\ell(J_2)}} a_{J_1J_1'J_1^{(i_1)}J_2J_2'J_2^{(i_2)}}h_{J_1'}\otimes u_{J_2'}\Bigg)$$
$$=\sum_{K_1,K_2}\sum_{J_1:J_1\subset K_1}^{(i_1)}\sum_{J_2:J_2\subset K_2}^{(i_2)}\sum_{I_1:I_1\subset K_1}\sum_{I_2:I_2\subset K_2}\langle b,h_{I_1}\otimes u_{I_2}\rangle\langle f,h_{J_1}\otimes u_{J_2}\rangle h_{I_1}\otimes u_{I_2}\cdot$$
$$\Bigg(\sum_{J_1':J_1'\subset K_1}^{(j_1)}\sum_{J_2':J_2'\subset K_2}^{(j_2)} a_{J_1J_1'K_1J_2J_2'K_2}h_{J_1'}\otimes u_{J_2'}\Bigg),$$

which by interchanging the order of summations is equal to

$$=\sum_{I_1,I_2}\langle b,h_{I_1}\otimes u_{I_2}\rangle h_{I_1}\otimes u_{I_2}\sum_{\substack{K_1\supset I_1\\ K_2\supset I_2}}\sum_{J_1,J_1'\subset K_1}^{(i_1,j_1)}\sum_{J_2,J_2'\subset K_2}^{(i_2,j_2)} a_{J_1J_1'K_1J_2J_2'K_2}\langle f,h_{J_1}\otimes u_{J_2}\rangle h_{J_1'}\otimes u_{J_2'}$$
$$=\sum_{I_1,I_2}\langle b,h_{I_1}\otimes u_{I_2}\rangle h_{I_1}\otimes u_{I_2}\sum_{J_1':J_1'^{(j_1)}\supset I_1}\sum_{J_2':J_2'^{(j_2)}\supset I_2}\langle S^{i_1j_1i_2j_2}f,h_{J_1'}\otimes u_{J_2'}\rangle h_{J_1'}\otimes u_{J_2'}.$$

Because of the supports of Haar functions, the sum above can be further decomposed into four parts, where

$$I=\sum_{I_1,I_2}\sum_{J_1'\supsetneq I_1}\sum_{J_2'\supsetneq I_2},\quad II=\sum_{I_1,I_2}\sum_{J_1'\supsetneq I_1}\sum_{J_2':J_2'\subset I_2\subset J_2'^{(j_2)}}$$
$$III=\sum_{I_1,I_2}\sum_{J_1':J_1'\subset I_1\subset J_1'^{(j_1)}}\sum_{J_2'\supsetneq I_2},\quad IV=\sum_{I_1,I_2}\sum_{J_1':J_1'\subset I_1\subset J_1'^{(j_1)}}\sum_{J_2':J_2'\subset I_2\subset J_2'^{(j_2)}}.$$

Hence, similarly as in Section 4.2, one has

$$I=\sum_{I_1,I_2}\langle b,h_{I_1}\otimes u_{I_2}\rangle\langle S^{i_1j_1i_2j_2}f,h^1_{J_1'}\otimes u^1_{J_2'}\rangle h_{I_1}\otimes u_{I_2}|I_1|^{-1/2}|I_2|^{-1/2},$$

which is a bi-parameter paraproduct $B_0(b,f)$.

Moreover, one has

$$\begin{aligned} II &= \sum_{I_1, I_2} \langle b, h_{I_1} \otimes u_{I_2} \rangle h_{I_1} \otimes u_{I_2} \sum_{J_2' : J_2' \subset I_2 \subset J_2'^{(j_2)}} \langle S^{i_1 j_1 i_2 j_2} f, h^1_{I_1} \otimes u_{J_2'} \rangle |I_1|^{-1/2} u_{J_2'} \\ &= \sum_{l=0}^{j_2} \sum_{I_1, J_2'} \beta_{J_2'} \langle b, h_{I_1} \otimes h_{J_2'^{(l)}} \rangle \langle S^{i_1 j_1 i_2 j_2} f, h^1_{I_1} \otimes u_{J_2'} \rangle h_{I_1} \otimes u_{J_2'} |I_1|^{-1/2} |J_2'^{(l)}|^{-1/2} \\ &= \sum_{l=0}^{j_2} BB_{0,l}(b, S^{i_1 j_1 i_2 j_2} f), \end{aligned}$$

where constants $\beta_{J_2'} \in \{1, -1\}$, and the $L^2 \to L^2$ operator norm of $BB_{0,l}$ is uniformly bounded by $\|b\|_{\mathrm{BMO_{prod}}}$. Similarly, one can show that

$$III = \sum_{k=0}^{j_1} BB_{k,0}(b, S^{i_1 j_1 i_2 j_2} f), \quad IV = \sum_{k=0}^{j_1} \sum_{l=0}^{j_2} BB_{k,l}(b, S^{i_1 j_1 i_2 j_2} f).$$

Since $\|b\|_{\mathrm{BMO_{prod}}} \lesssim \|b\|_{\mathrm{bmo}}$, all the forms above are L^2 bounded. This completes the discussion of term I.

To estimate of term II, we need to decompose it into finite linear combinations of $S^{i_1 j_1 i_2 j_2}(B_{kl}(b, f))$. By linearity, one can write $S^{i_1 j_1 i_2 j_2}$ on the outside from the beginning, and we will only look at the inside sum. One splits for example the sum regarding the first variable into three parts: $I_1 \subsetneq J_1$, $I_1 = J_1$, $J_1 \subsetneq I_1 \subset J_1^{(i_1)}$. If we split the second variable as well, there are nine mixed parts, and it's not hard to show that each of them can be represented as a finite sum of $BB_{k,l}(b, f)$. We omit the details.

Mixed cases

Let's call the second part $I_1 \subset J_1^{(i_1)}, I_2 \supsetneq J_2^{(i_2)}$, the third part $I_1 \supsetneq J_1^{(i_1)}, I_2 \subset J_2^{(i_2)}$ "mixed" cases, and as the two are symmetric, it suffices to look at the second one. In the first variable, we still have the old case $I_1 \subset J_1^{(i_1)}$ that appeared in Section 4.2, so morally speaking, we only need to nicely play around with the stronger little BMO norm in order to handle the second variable. For any fixed I_1, J_1, I_2, J_2, since $I_2 \supsetneq J_2^{(i_2)}$, the definition of dyadic shifts implies that

$$h_{I_1} \otimes u_{I_2} S^{i_1 j_1 i_2 j_2}(h_{J_1} \otimes u_{J_2}) = h_{I_1} S^{i_1 j_1 i_2 j_2}(h_{J_1} \otimes u_{I_2} u_{J_2})$$

and

$$S^{i_1 j_1 i_2 j_2}(h_{I_1} h_{J_1} \otimes u_{I_2} u_{J_2}) = u_{I_2} S^{i_1 j_1 i_2 j_2}(h_{I_1} h_{J_1} \otimes u_{J_2}).$$

Hence, we still have cancellation in the second variable, which converts the mixed case to

$$\begin{aligned}
&\sum_{I_1 \subset J_1^{(i_1)}} \sum_{I_2 \supsetneq J_2^{(i_2)}} \langle b, h_{I_1} \otimes u_{I_2} \rangle \langle f, h_{J_1} \otimes u_{J_2} \rangle [h_{I_1}, S^{i_1 j_1 i_2 j_2}](h_{J_1} \otimes u_{I_2} u_{J_2}) \\
= &\sum_{I_1 \subset J_1^{(i_1)}} \sum_{J_2} \langle f, h_{J_1} \otimes u_{J_2} \rangle [h_{I_1}, S^{i_1 j_1 i_2 j_2}] \left(h_{J_1} \otimes \sum_{I_2 \supsetneq J_2^{(i_2)}} \langle b, h_{I_1} \otimes u_{I_2} \rangle u_{I_2} u_{J_2} \right) \\
= &\sum_{I_1 \subset J_1^{(i_1)}} \sum_{J_2} \langle f, h_{J_1} \otimes u_{J_2} \rangle [h_{I_1}, S^{i_1 j_1 i_2 j_2}] \left(h_{J_1} \otimes \langle b, h_{I_1} \otimes u^1_{J_2^{(i_2)}} \rangle u^1_{J_2^{(i_2)}} u_{J_2} \right) \\
= &\sum_{I_1 \subset J_1^{(i_1)}} \sum_{J_2} \langle b, h_{I_1} \otimes u^1_{J_2^{(i_2)}} \rangle |J_2^{(i_2)}|^{-1/2} \langle f, h_{J_1} \otimes u_{J_2} \rangle [h_{I_1}, S^{i_1 j_1 i_2 j_2}](h_{J_1} \otimes u_{J_2}) \\
= &\sum_{I_1 \subset J_1^{(i_1)}} \sum_{J_2} \langle \langle b \rangle_{J_2^{(i_2)}}, h_{I_1} \rangle_1 \langle f, h_{J_1} \otimes u_{J_2} \rangle [h_{I_1}, S^{i_1 j_1 i_2 j_2}](h_{J_1} \otimes u_{J_2}),
\end{aligned}$$

where $\langle b \rangle_{J_2^{(i_2)}}$ denotes the average value of b on $J_2^{(i_2)}$, which is a function in the first variable. In the following, we will once again estimate the first term and second term of the commutator separately, and the L^2 norm of each of them will be proved to be bounded by $\|b\|_{\mathrm{bmo}} \|f\|_{L^2}$.

a) First term.

By definition of the dyadic shift, the first term is equal to

$$\sum_{I_1 \subset J_1^{(i_1)}} \sum_{J_2} \langle \langle b \rangle_{J_2^{(i_2)}}, h_{I_1} \rangle_1 h_{I_1} \langle f, h_{J_1} \otimes u_{J_2} \rangle \cdot \\
\left(\sum_{\substack{J_1' \subset J_1^{(i_1)} \\ \ell(J_1') = 2^{i_1 - j_1} \ell(J_1)}} \sum_{\substack{J_2' \subset J_2^{(i_2)} \\ \ell(J_2') = 2^{i_2 - j_2} \ell(J_2)}} a_{J_1 J_1' J_1^{(i_1)} J_2 J_2' J_2^{(i_2)}} h_{J_1'} \otimes u_{J_2'} \right),$$

which by reindexing $K_1 := J_1^{(i_1)}$ is the same as

$$\sum_{I_1,J_2} \langle\langle b\rangle_{J_2^{(i_2)}}, h_{I_1}\rangle_1 h_{I_1} \sum_{K_1:K_1\supset I_1} \sum_{J_1\subset K_1}^{(i_1)} \sum_{J_1'\subset K_1}^{(j_1)} \sum_{J_2'\subset J_2^{(i_2)}}^{(j_2)} a_{J_1J_1'K_1J_2J_2'J_2^{(i_2)}} \langle f, h_{J_1}\otimes u_{J_2}\rangle h_{J_1'}\otimes u_{J_2'}$$
$$= \sum_{I_1,J_2} \langle\langle b\rangle_{J_2^{(i_2)}}, h_{I_1}\rangle_1 h_{I_1} \sum_{J_1':J_1'^{(j_1)}\supset I_1} h_{J_1'}\otimes \left\langle S^{i_1j_1i_2j_2}(\langle f,u_{J_2}\rangle_2\otimes u_{J_2}), h_{J_1'}\right\rangle_1,$$

where the inner sum is the orthogonal projection of the image of $\langle f,u_{J_2}\rangle_2\otimes u_{J_2}$ under $S^{i_1j_1i_2j_2}$ onto the span of $\{h_{J_1'}\}$ such that $J_1'^{(j_1)}\supset I_1$. Taking into account the supports of the Haar functions in the first variable, one can further split the sum into two parts where

$$I := \sum_{J_2}\sum_{I_1\subsetneq J_1'}, \quad II := \sum_{J_2}\sum_{J_1'\subset I_1\subset J_1'^{(j_1)}}.$$

Summing over J_1' first implies that

$$I = \sum_{J_2}\sum_{I_1} \langle\langle b\rangle_{J_2^{(i_2)}}, h_{I_1}\rangle_1 h_{I_1}\left(h_{I_1}^1\otimes\left\langle S^{i_1j_1i_2j_2}(\langle f,u_{J_2}\rangle_2\otimes u_{J_2}), h_{I_1}^1\right\rangle_1\right)$$
$$=: \sum_{J_2} B_0\left(\langle b\rangle_{J_2^{(i_2)}}, S^{i_1j_1i_2j_2}(\langle f,u_{J_2}\rangle_2\otimes u_{J_2})\right)$$

where $B_0(b,f) := \sum_I \langle b,h_I\rangle\langle f,h_I^1\rangle h_I|I|^{-1/2}$ is a classical one-parameter paraproduct in the first variable. Note that its L^2 norm is bounded by $\|b\|_{\mathrm{BMO}}\|f\|_{L^2}$. Moreover, according to the definition of $S^{i_1j_1i_2j_2}$, for any fixed J_2

$$S^{i_1j_1i_2j_2}(\langle f,u_{J_2}\rangle_2\otimes u_{J_2}) = \sum_{J_2':J_2'^{(j_2)}=J_2^{(i_2)}} \left\langle S^{i_1j_1i_2j_2}(\langle f,u_{J_2}\rangle_2\otimes u_{J_2}), u_{J_2'}\right\rangle_2\otimes u_{J_2'}.$$

In other words, $S^{i_1j_1i_2j_2}(\langle f,u_{J_2}\rangle_2\otimes u_{J_2})$ lives only on the span of $\{u_{J_2'} : J_2'^{(j_2)} = J_2^{(i_2)}\}$. Hence, by linearity there holds

$$I = \sum_{J_2}\sum_{J_2':J_2'^{(j_2)}=J_2^{(i_2)}} B_0\left(\langle b\rangle_{J_2^{(i_2)}}, \left\langle S^{i_1j_1i_2j_2}(\langle f,u_{J_2}\rangle_2\otimes u_{J_2}), u_{J_2'}\right\rangle_2\right)\otimes u_{J_2'}$$
$$= \sum_{J_2'} B_0\left(\langle b\rangle_{J_2'^{(j_2)}}, \left\langle S^{i_1j_1i_2j_2}\Big(\sum_{J_2:J_2^{(i_2)}=J_2'^{(j_2)}} \langle f,u_{J_2}\rangle_2\otimes u_{J_2}\Big), u_{J_2'}\right\rangle_2\right)\otimes u_{J_2'}.$$

Thus, orthogonality in the second variable implies that

$$\begin{aligned}
&\|I\|^2_{L^2(\mathbb{R}^n\times\mathbb{R}^m)}\\
&=\sum_{J_2'}\left\|B_0\left(\langle b\rangle_{J_2'^{(j_2)}},\left\langle S^{i_1j_1i_2j_2}\big(\sum_{J_2:J_2^{(i_2)}=J_2'^{(j_2)}}\langle f,u_{J_2}\rangle_2\otimes u_{J_2}\big),u_{J_2'}\right\rangle_2\right)\right\|^2_{L^2(\mathbb{R}^n)}\\
&\lesssim\sum_{J_2'}\|\langle b\rangle_{J_2'^{(j_2)}}\|^2_{\mathrm{BMO}(\mathbb{R}^n)}\left\|\left\langle S^{i_1j_1i_2j_2}\big(\sum_{J_2:J_2^{(i_2)}=J_2'^{(j_2)}}\langle f,u_{J_2}\rangle_2\otimes u_{J_2}\big),u_{J_2'}\right\rangle_2\right\|^2_{L^2(\mathbb{R}^n)}.
\end{aligned}$$

Observing that $\|\langle b\rangle_{J_2'^{(j_2)}}\|_{\mathrm{BMO}(\mathbb{R}^n)}\le\langle\|b\|_{\mathrm{BMO}(\mathbb{R}^n)}\rangle_{J_2'^{(j_2)}}\le\|b\|_{\mathrm{bmo}}$, one has

$$\begin{aligned}
&\le\|b\|^2_{\mathrm{bmo}}\sum_{J_2'}\left\|\left\langle S^{i_1j_1i_2j_2}\big(\sum_{J_2:J_2^{(i_2)}=J_2'^{(j_2)}}\langle f,u_{J_2}\rangle_2\otimes u_{J_2}\big),u_{J_2'}\right\rangle_2\right\|^2_{L^2(\mathbb{R}^n)}\\
&=\|b\|^2_{\mathrm{bmo}}\left\|\sum_{J_2'}\left\langle S^{i_1j_1i_2j_2}\big(\sum_{J_2:J_2^{(i_2)}=J_2'^{(j_2)}}\langle f,u_{J_2}\rangle_2\otimes u_{J_2}\big),u_{J_2'}\right\rangle_2\otimes u_{J_2'}\right\|^2_{L^2(\mathbb{R}^n\times\mathbb{R}^m)}.
\end{aligned}$$

Note that the sum in the L^2 norm

$$\sum_{J_2'}\left\langle S^{i_1j_1i_2j_2}\big(\sum_{J_2:J_2^{(i_2)}=J_2'^{(j_2)}}\langle f,u_{J_2}\rangle_2\otimes u_{J_2}\big),u_{J_2'}\right\rangle_2\otimes u_{J_2'}$$

is in fact very simple, which is the same as

$$\begin{aligned}
&=\sum_{J_2}\sum_{J_2':J_2'^{(j_2)}=J_2^{(i_2)}}\left\langle S^{i_1j_1i_2j_2}\big(\langle f,u_{J_2}\rangle_2\otimes u_{J_2}\big),u_{J_2'}\right\rangle_2\otimes u_{J_2'}\\
&=\sum_{J_2}S^{i_1j_1i_2j_2}\big(\langle f,u_{J_2}\rangle_2\otimes u_{J_2}\big)=S^{i_1j_1i_2j_2}(f).
\end{aligned}$$

Hence, the uniform boundedness of the $L^2\to L^2$ operator norm of dyadic shifts implies that

$$\|I\|^2_{L^2(\mathbb{R}^n\times\mathbb{R}^m)}\lesssim\|b\|^2_{\mathrm{bmo}}\|f\|^2_{L^2(\mathbb{R}^n\times\mathbb{R}^m)}.$$

In order to handle II, we split it into a finite sum depending on how many generations there are between I_1 and J_1', which leads to

$$\begin{aligned} II &= \sum_{k=0}^{j_1} \sum_{J_2} \sum_{J_1'} \langle \langle b \rangle_{J_2^{(i_2)}}, h_{J_1'^{(k)}} \rangle_1 h_{J_1'^{(k)}} h_{J_1'} \otimes \left\langle S^{i_1 j_1 i_2 j_2} \big(\langle f, u_{J_2} \rangle_2 \otimes u_{J_2} \big), h_{J_1'} \right\rangle_1 \\ &= \sum_{k=0}^{j_1} \sum_{J_2} \sum_{J_1'} \beta_{J_1', k} |J_1'^{(k)}|^{-1/2} \langle \langle b \rangle_{J_2^{(i_2)}}, h_{J_1'^{(k)}} \rangle_1 h_{J_1'} \otimes \left\langle S^{i_1 j_1 i_2 j_2} \big(\langle f, u_{J_2} \rangle_2 \otimes u_{J_2} \big), h_{J_1'} \right\rangle_1 \\ &=: \sum_{k=0}^{j_1} \sum_{J_2} B_k \left(\langle b \rangle_{J_2^{(i_2)}}, S^{i_1 j_1 i_2 j_2} \big(\langle f, u_{J_2} \rangle_2 \otimes u_{J_2} \big) \right), \end{aligned}$$

where $B_k(b,f) := \sum_I \beta_{I,k} \langle b, h_{I^{(k)}} \rangle \langle f, h_I \rangle h_I |I^{(k)}|^{-1/2}$ is a one-parameter paraproduct with descendants studied in Lemma 4.1, whose L^2 norm is uniformly bounded by $\|b\|_{\mathrm{BMO}} \|f\|_{L^2}$, independent of k and the coefficients $\beta_{I,k} \in \{1, -1\}$. Then one can proceed as how we dealt with part I to conclude that

$$\|II\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)} \lesssim (1 + j_1) \|b\|_{\mathrm{bmo}} \|f\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)},$$

which together with the estimate for part I implies that

$$\|\text{First term}\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)} \lesssim (1 + j_1) \|b\|_{\mathrm{bmo}} \|f\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)}.$$

b) Second term.

As the second term by linearity is the same as

$$S^{i_1 j_1 i_2 j_2} \left(\sum_{J_2} \sum_{I_1 \subset J_1^{(i_1)}} \langle \langle b \rangle_{J_2^{(i_2)}}, h_{I_1} \rangle_1 \langle f, h_{J_1} \otimes u_{J_2} \rangle h_{I_1} h_{J_1} \otimes u_{J_2} \right),$$

the $L^2 \to L^2$ boundedness of the shift implies that it suffices to estimate the L^2 norm of the term inside the parentheses. Since $I_1 \cap J_1 \neq \emptyset$, one can further split the sum into two parts:

$$I := \sum_{J_2} \sum_{I_1 \subsetneq J_1}, \quad II := \sum_{J_2} \sum_{J_1 \subset I_1 \subset J_1^{(i_1)}}.$$

Summing over J_1 first implies that

$$\begin{aligned} I &= \sum_{J_2} \sum_{I_1} \langle \langle b \rangle_{J_2^{(i_2)}}, h_{I_1} \rangle_1 \langle f, h^1_{I_1} \otimes u_{J_2} \rangle h_{I_1} h^1_{I_1} \otimes u_{J_2} \\ &=: \sum_{J_2} B_0 \left(\langle b \rangle_{J_2^{(i_2)}}, \langle f, u_{J_2} \rangle_2 \right) \otimes u_{J_2}, \end{aligned}$$

where $B_0(b,f) := \sum_I \langle b, h_I \rangle \langle f, h^1_I \rangle h_I |I|^{-1/2}$ is a classical one-parameter paraproduct in the first variable. Hence,

$$\begin{aligned} \|I\|^2_{L^2(\mathbb{R}^n \times \mathbb{R}^m)} &= \sum_{J_2} \left\| B_0 \left(\langle b \rangle_{J_2^{(i_2)}}, \langle f, u_{J_2} \rangle_2 \right) \right\|^2_{L^2(\mathbb{R}^n)} \\ &\lesssim \sum_{J_2} \| \langle b \rangle_{J_2^{(i_2)}} \|^2_{\mathrm{BMO}(\mathbb{R}^n)} \| \langle f, u_{J_2} \rangle_2 \|^2_{L^2(\mathbb{R}^n)} \\ &\leq \|b\|^2_{\mathrm{bmo}} \sum_{J_2} \| \langle f, u_{J_2} \rangle_2 \|^2_{L^2(\mathbb{R}^n)} = \|b\|^2_{\mathrm{bmo}} \|f\|^2_{L^2(\mathbb{R}^n \times \mathbb{R}^m)}. \end{aligned}$$

For part II, note that it can be decomposed as

$$\begin{aligned} II &= \sum_{k=0}^{i_1} \sum_{J_2} \sum_{J_1} \langle \langle b \rangle_{J_2^{(i_2)}}, h_{J_1^{(k)}} \rangle_1 \langle f, h_{J_1} \otimes u_{J_2} \rangle h_{J_1^{(k)}} h_{J_1} \otimes u_{J_2} \\ &= \sum_{k=0}^{i_1} \sum_{J_2} \sum_{J_1} \beta_{J_1,k} |J_1^{(k)}|^{-1/2} \langle \langle b \rangle_{J_2^{(i_2)}}, h_{J_1^{(k)}} \rangle_1 \langle \langle f, u_{J_2} \rangle_2, h_{J_1} \rangle_1 h_{J_1} \otimes u_{J_2} \\ &=: \sum_{k=0}^{i_1} \sum_{J_2} B_k \left(\langle b \rangle_{J_2^{(i_2)}}, \langle f, u_{J_2} \rangle_2 \right) \otimes u_{J_2}, \end{aligned}$$

where coefficients $\beta_{J_1,k} \in \{1, -1\}$ and the L^2 norm of B_k is uniformly bounded as mentioned before. Therefore, same argument as for part I shows that

$$\|II\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)} \lesssim (1 + i_1) \|b\|_{\mathrm{bmo}} \|f\|_{L^2(\mathbb{R}^n \times \mathbb{R}^m)},$$

which completes the discussion of the second term, and thus proves that the mixed case is bounded. Therefore, (4.23) is demonstrated.

Remark 4.26. The upper bound result we just proved can be extended to $\mathbb{R}^{\vec{d}}$, to arbitrarily many parameters and an arbitrary number of iterates in the commutator, as stated in

Theorem 3.16. In the general case, consider a little product BMO function $b \in \mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$, where $\mathcal{I} = (I_s)_{1\leq s\leq l}$ is a partition of $\{1,\ldots,t\}$. Let T_s denote a multi-parameter paraproduct free Journé operator acting on functions defined on $\bigotimes_{k\in I_s} \mathbb{R}^{d_k}$, where T_s is in the class of multi-parameter singular integrals that will be studied in Chapter 7, which satisfies a weak boundedness property and is paraproduct free, meaning that any partial adjoint of T is zero if acting on some tensor product of functions with one of the components being 1. Then we have the following estimate

$$\|[\ldots[b,T_1],\ldots T_l]\|_{L^2(\mathbb{R}^{\vec{d}})\to L^2(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

The part of the proof that targets the Journé operators proceeds exactly the same as the bi-parameter case with the multi-parameter version of the representation theorem that will be established in Chapter 7. Certainly, as the number of parameters increases, more mixed cases will appear. However, if one follows the corresponding argument above for each variable in each case, it is not hard to check that eventually, the boundedness of the arising paraproduct-like operators is implied exactly by the little product BMO norm of the symbol. The difficulty of higher iterates is overcome by observing that the commutator splits into commutators with no iterates, as was done in the proof of Theorem 3.12 in Section 4.3. We omit the details.

4.4.3 Commutators of standard paraproducts

The main theorem we are going to prove in this subsection is the following:

Theorem 4.27. *Let Π_1 be a standard bi-parameter dyadic paraproduct:*

$$\Pi_1(f) := \sum_{I_1,I_2} a_{I_1I_2}\langle f, h_{I_1}\otimes u_{I_2}\rangle h^1_{I_1}\otimes u^1_{I_2},$$

where $a_{I_1I_2} := \langle a, h_{I_1}\otimes u_{I_2}\rangle |I_1|^{-1/2}|I_2|^{-1/2}$ for some product BMO symbol a with $\|a\|_{BMO_{prod}} \leq$

1. *Then for any little BMO function b,*

$$\|[b,\Pi_1]f\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)} \lesssim \|b\|_{bmo}\|f\|_{L^2(\mathbb{R}^n\times\mathbb{R}^m)}, \quad \forall f\in L^2(\mathbb{R}^n\times\mathbb{R}^m).$$

Although this theorem is stated in the base case (bi-parameter without iterates), it will be clear from the proof that similar results hold true when one has an iterated commutator with each component being a multi-parameter standard paraproduct.

We now proceed with the proof. We adopt the by now standard technique: fix a L^2 function f, decompose $[b,\Pi_1]f$ using Haar bases, then represent the sum as finite linear combinations of basic bounded paraproduct-like operators. Decompose

$$[b,\Pi_1]f=\sum_{I_1,I_2}\sum_{J_1,J_2}\langle b,h_{I_1}\otimes u_{I_2}\rangle\langle f,h_{J_1}\otimes u_{J_2}\rangle[h_{I_1}\otimes u_{I_2},\Pi_1]h_{J_1}\otimes u_{J_2}.$$

By the definition of Π_1 and the cancellation of the commutator, it is not hard to observe that $[h_{I_1}\otimes u_{I_2},\Pi_1]h_{J_1}\otimes u_{J_2}$ is nonzero only if $I_1\subset J_1$ or $I_2\subset J_2$. (See Subsection 4.4.2 for more details.) The part $I_1\subset J_1, I_2\subset J_2$ is called the *regular* case, while the part $I_1\subset J_1, I_2\supsetneq J_2$ and the symmetric one are called the *mixed* cases. Moreover, we will deal with the sum corresponding to the first term $h_{I_1}\otimes u_{I_2}\Pi_1(h_{J_1}\otimes u_{J_2})$ and the second term $\Pi_1(h_{I_1}h_{J_1}\otimes u_{I_2}u_{J_2})$ separately.

Second term

The estimate of the sum corresponding to the second term is usually much easier, due to the fact that the operator Π_1 can be pulled outside of the sum by linearity. Since Π_1 is bounded on L^2, it suffices to prove that

$$\left\|\sum_{I_1,I_2}\sum_{J_1,J_2}\langle b,h_{I_1}\otimes u_{I_2}\rangle\langle f,h_{J_1}\otimes u_{J_2}\rangle h_{I_1}h_{J_1}\otimes u_{I_2}u_{J_2}\right\|_{L^2}\lesssim \|b\|_{\mathrm{bmo}}\|f\|_{L^2}.$$

In the regular case when $I_1 \subset J_1, I_2 \subset J_2$, split

$$\sum_{I_1 \subset J_1} \sum_{I_2 \subset J_2} = \sum_{I_1 \subsetneq J_1} \sum_{I_2 \subsetneq J_2} + \sum_{I_1 \subsetneq J_1} \sum_{I_2 = J_2} + \sum_{I_1 = J_1} \sum_{I_2 \subsetneq J_2} + \sum_{I_1 = J_1} \sum_{I_2 = J_2} =: I + II + III + IV,$$

then each of the four parts could be represented as a classical dyadic paraproduct, whose boundedness in terms of the product BMO norm of b (which is less than $\|b\|_{\text{bmo}}$) is well known. For example

$$I = \sum_{I_1, I_2} \langle b, h_{I_1} \otimes u_{I_2} \rangle \langle f, h^1_{I_1} \otimes u^1_{I_2} \rangle h_{I_1} \otimes u_{I_2} |I_1|^{-1/2} |I_2|^{-1/2},$$

$$IV = \sum_{I_1, I_2} \langle b, h_{I_1} \otimes u_{I_2} \rangle \langle f, h_{I_1} \otimes u_{I_2} \rangle h^{\epsilon_1}_{I_1} \otimes u^{\epsilon_2}_{I_2} |I_1|^{-1/2} |I_2|^{-1/2},$$

and similarly for part II and III.

In the mixed case when $I_1 \subset J_1, I_2 \supsetneq J_2$, we have to exploit the full strength of the little BMO norm. Summing over I_2 first gives

$$\begin{aligned} &\sum_{I_1 \subset J_1} \sum_{I_2 \supsetneq J_2} \langle b, h_{I_1} \otimes u_{I_2} \rangle \langle f, h_{J_1} \otimes u_{J_2} \rangle h_{I_1} h_{J_1} \otimes u_{I_2} u_{J_2} \\ =& \sum_{I_1 \subset J_1} \sum_{J_2} \langle b, h_{I_1} \otimes u^1_{J_2} \rangle \langle f, h_{J_1} \otimes u_{J_2} \rangle h_{I_1} h_{J_1} \otimes u_{J_2} |J_2|^{-1/2} \\ =& \sum_{J_2} \left(\sum_{I_1 \subsetneq J_1} + \sum_{I_1 = J_1} \right). \end{aligned}$$

In the above, observe that

$$\sum_{J_2} \sum_{I_1 \subsetneq J_1} = \sum_{I_1, J_2} \langle b, h_{I_1} \otimes u^1_{J_2} \rangle \langle f, h^1_{I_1} \otimes u_{J_2} \rangle h_{I_1} \otimes u_{J_2} |I_1|^{-1/2} |J_2|^{-1/2},$$

and

$$\sum_{J_2} \sum_{I_1 = J_1} = \sum_{I_1, J_2} \langle b, h_{I_1} \otimes u^1_{J_2} \rangle \langle f, h_{I_1} \otimes u_{J_2} \rangle h^{\epsilon}_{I_1} \otimes u_{J_2} |I_1|^{-1/2} |I_2|^{-1/2},$$

where ϵ could be equal to $\vec{1}$. Then both of them are paraproducts with little BMO symbol, whose boundedness follows from Lemma 4.10.

First term

In order to deal with the sum corresponding to the first term, other than paraproducts, one needs the aforementioned operators BP_k and PB_l. When $k, l = 0$, we usually omit the subscripts, simply writing BP or PB. By the definition of Π_1, the sum corresponding to the first term is equal to

$$\sum_{I_1,J_1}\sum_{I_2,J_2}\langle b, h_{I_1}\otimes u_{I_2}\rangle\langle f, h_{J_1}\otimes u_{J_2}\rangle h_{I_1}h^1_{J_1}\otimes u_{I_2}u^1_{J_2}a_{J_1J_2}.$$

Different from the second term, here, even in the regular case, the full strength of the little BMO norm of b is needed. Specifically, in the regular case $I_1\subset J_1, I_2\subset J_2$,

$$\sum_{I_1\subset J_1, I_2\subset J_2} = \sum_{I_1=J_1, I_2=J_2} + \sum_{I_1\subsetneq J_1, I_2=J_2} + \sum_{I_1=J_1, I_2\subsetneq J_2} + \sum_{I_1\subsetneq J_1, I_2\subsetneq J_2} =: A+B+C+D.$$

Since $|a_{J_1J_2}|\leq 1$, A is a bounded paraproduct with symbol b. Moreover,

$$\begin{aligned} D &= \sum_{J_1,J_2}\langle a, h_{J_1}\otimes u_{J_2}\rangle\langle f, h_{J_1}\otimes u_{J_2}\rangle|J_1|^{-1}|J_2|^{-1}\sum_{I_1:I_1\subsetneq J_1}\sum_{I_2:I_2\subsetneq J_2}\langle b, h_{I_1}\otimes u_{I_2}\rangle h_{I_1}\otimes u_{I_2} \\ &= PP(a,b,f), \end{aligned}$$

where the operator norm of PP has been shown to be bounded by $\|a\|_{\mathrm{BMO_{prod}}}\|b\|_{\mathrm{BMO_{prod}}}$ in Lemma 4.5. We are thus left with B (as C is symmetric), which can be easily seen to be equal to $PB(a,\{b_{I_2}\}_{I_2}, f)$ with $b_{I_2} := \langle b, u_{I_2}\rangle_2|I_2|^{-1/2}$. Then since

$$\|\langle b, u_{I_2}\rangle_2|I_2|^{-1/2}\|_{\mathrm{BMO}}\leq\|b\|_{\mathrm{bmo}},$$

this term is bounded as well according to Lemma 4.8, which completes the discussion of the regular case.

We are left with the mixed cases of the first term, and by symmetry, it suffices to consider the case $I_1 \subset J_1, I_2 \supsetneq J_2$. Summing over I_2 gives

$$\sum_{I_1 \subset J_1} \sum_{I_2 \supsetneq J_2} \langle b, h_{I_1} \otimes u_{I_2} \rangle \langle f, h_{J_1} \otimes u_{J_2} \rangle h_{I_1} h^1_{J_1} \otimes u_{I_2} u^1_{J_2} a_{J_1 J_1}$$
$$= \sum_{I_1 \subset J_1} \sum_{J_2} \langle b, h_{I_1} \otimes u^1_{J_2} \rangle \langle f, h_{J_1} \otimes u_{J_2} \rangle h_{I_1} h^1_{J_1} \otimes u^1_{J_2} |J_2|^{-1/2} a_{J_1 J_2}$$
$$= \sum_{I_1 = J_1} + \sum_{I_1 \subsetneq J_1} =: E + F.$$

It's easily seen that

$$E = \sum_{J_1, J_2} \beta_{J_1 J_2} \langle a, h_{J_1} \otimes u_{J_2} \rangle \langle f, h_{J_1} \otimes u_{J_2} \rangle h_{J_1} \otimes u^1_{J_2} |J_1|^{-1/2} |J_2|^{-1/2}$$

with $|\beta_{J_1 J_2} := \langle b, h_{J_1} \otimes u^1_{J_2} \rangle |J_1|^{-1/2} |J_2|^{-1/2}| \leq \|b\|_{\mathrm{bmo}}$, which is a bounded classical paraproduct with symbol a. On the other hand, $F = PB(a, \{b_{J_2}\}_{J_2}, f)$ with $b_{J_2} := \langle b, u^1_{J_2} \rangle_2 |J_2|^{-1/2}$, which is also bounded according to Lemma 4.8. Therefore, the proof of Theorem 4.27 is by now complete.

Remark 4.28. The main difficulty blocking us from generalizing Theorem 4.27 to the case of mixed paraproduct Π_2 or partial paraproduct Π_3^{ij} is that, neither the function b nor the symbol a of the mixed paraproduct is necessarily a tensor product. Note that in the argument above, there are many times one essentially interchanges a and b in the process of constructing the correct paraproduct-like operators structure. However, with the added layer of the mixed structure of the paraproduct, in some of the mixed cases as discussed above, one needs to interchange a and b in a particular variable only, which fails to be doable as a, b are not tensor product. Even in the case of partial paraproduct Π_3^{ij}, where the symbol is essentially a one-parameter function, b not being a tensor product still prevents us from performing a similar technique as what was done in the proof of Theorem 4.27.

In fact, the problem we are faced with here might not be an isolated one. The mixed paraproduct is a well known problematic object in multi-parameter theory, as its boundedness is not equivalent to the inclusion of its symbol function in product BMO (see [MO15]), in contrast to the case of standard paraproduct. And this phenomenon is also responsible for the fact that all existing multi-parameter $T(1)/T(b)$ type theorems provide only sufficient but not necessary conditions of the boundedness of the operator. (See Chapter 7 for a more detailed discussion.) We believe that in order to fully tackle the commutator upper bound problem, one needs to first have a more comprehensive understanding of the mixed paraproduct and partial paraproduct themselves. And to do this, a better understanding of the BMO theory in the multi-parameter setting is necessary.

Chapter Five

Estimates for Hilbert and Riesz Commutators

In this chapter, we continue our discussion of commutators by establishing the characterization of little product BMO in terms of the boundedness of corresponding iterated commutators with Hilbert transform or Riesz transforms. More precisely, we are going to prove Theorem 3.6, the Hilbert commutator estimates, in Section 5.1, and Theorem 3.9, 3.10, the Riesz commutator estimates, in Section 5.2. Note that we have developed in Chapter 4 a much more general upper bound theory for commutators, therefore the upper bound estimates for Hilbert and Riesz commutators are simply direct consequences of Theorem 3.13, which leaves our focus on the lower bound direction. However, in the Hilbert commutator case, due to the nature of the lower bound argument, we also obtain an alternative proof of the upper bound automatically.

There is a sharp contrast between the proofs for the Hilbert commutators and Riesz commutators, which is why we present them both even though the former is the one-dimensional case of the latter. Specifically, in the Hilbert commutators case, Toeplitz operators with operator symbol arise naturally. However, using Riesz transforms in $\mathbb{R}^d$ as a replacement, there is absence of analytic structure and tools relying on analytic projection or orthogonal spaces are not readily available. We overcome this difficulty through a first intermediate passage via tensor products of Calderón-Zygmund operators whose Fourier multiplier symbols are adapted to cones. This idea is inspired by [LPPW09]. Such operators are also mentioned in [Uch82]. A class of operators of this type classifies little product BMO through two-sided commutator estimates, but it does not allow the passage to a classification through iterated commutators with tensor products of Riesz transforms. In a second step, we find it necessary to consider upper and lower commutator estimates using a well-chosen family of Journé operators that are not of tensor product type, where the upper estimates for such operators developed in Chapter 4 becomes indispensable. Through geometric considerations and an averaging procedure of zonal harmonics on products of spheres, we construct the multiplier of a special Journé operator that preserves lower commutator estimates and resembles the multiple Hilbert transform: it has large plateaus of constant values and is a polynomial in multiple Riesz transforms. We expect that this construction allows for other applications.

There is an increase in difficulty when the dimension is greater than two, due to the simpler structure of the rotation group on $\mathbb{S}^1$. In higher dimension, there is a rise in difficulty when tensor products involve more than two Riesz transforms.

The actual passage to the Riesz transforms requires for us to take advantage of the stability estimate in commutator norms for certain multi-parameter singular integrals in terms of the mixed BMO class, which is given by Corollary 3.17 introduced in Chapter 3.

In the following, we will first restrict ourselves to the case of the L^2 bounds in Sections 5.1 and 5.2, and then remark on how to obtain the general L^p bounds result in Section 5.3. In Appendix A, we also give an alternative lower bound proof of the Hilbert commutator estimates in its base case (bi-parameter, no iteration), which is a one-dimensional adaptation of the argument in Section 5.2.

5.1 Hilbert commutators

In this section, we prove Theorem 3.6 which characterizes the boundedness of commutators of the form $[[b, H_1H_3], H_2]$ as operators on $L^2(\mathbb{T}^3)$. In the case of the Hilbert transform, this case is representative for the general case and provides a starting point that is easier to read because of the simplicity of the expression of products and sums of projection onto orthogonal subspaces. Its general form can be found in Theorem 3.9, which we prove in the next section. More precisely, to prove Theorem 3.6, we are going to show that the following three statements are equivalent:

(1) $b \in \mathrm{BMO}_{(13)(2)}(\mathbb{T}^3)$;

(2) The commutators $[[b, H_1], H_2]$ and $[[b, H_3], H_2]$ are bounded on $L^2(\mathbb{T}^3)$;

(3) The commutator $[[b, H_1H_3], H_2]$ is bounded on $L^2(\mathbb{T}^3)$.

As preparation, in general, let $b \in L^1(\mathbb{T}^n)$ and let P and Q denote orthogonal projections onto subspaces of $L^2(\mathbb{T}^n)$. We shall first describe relationships between functions in the little product BMOs and several types of projection-multiplication operators. These will be Hilbert transform-type operators of the form $P - P^{\perp}$; and iterated Hankel or Toeplitz type operators of the form $Q^{\perp}bQ$ (Hankel), PbP (Toeplitz), $PQ^{\perp}bQP$ (mixed), where b means the (not a priori bounded) multiplication operator M_b on $L^2(\mathbb{T}^n)$.

We shall use the following simple observation concerning Hilbert transform type operators again and again:

Remark 5.1. If $H = P - P^{\perp}$ and $T : L^2(\mathbb{T}^n) \to L^2(\mathbb{T}^n)$ is a linear operator then

$$[H, T] = 2PTP^{\perp} - 2P^{\perp}TP$$

and $[H, T]$ is bounded if and only if $PTP^{\perp}$ and $P^{\perp}TP$ are.

Proof.

$$\begin{aligned}(P - P^{\perp})T - T(P - P^{\perp}) &= (P - P^{\perp})T(P + P^{\perp}) - (P + P^{\perp})T(P - P^{\perp}) \\ &= 2PTP^{\perp} - 2P^{\perp}TP.\end{aligned}$$

□

Now let's go back to our case. It will be useful to denote by Q_{13} orthogonal projection on the subspace of functions which are either analytic or anti-analytic in the first and third variables; $Q_{13} = P_1P_3 + P_1^{\perp}P_3^{\perp}$. Then the projection $Q_{13}^{\perp}$ onto the orthogonal complement of this subspace is defined by $Q_{13}^{\perp} = P_1^{\perp}P_3 + P_1P_3^{\perp}$. We reformulate properties (2) and (3) in the statement of Theorem 3.6 in terms of Hankel, Toeplitz type operators.

Lemma 5.2. *We have the following algebraic facts on commutators and projection operators.*

(1) The commutators $[[b,H_1],H_2]$ *and* $[[b,H_3],H_2]$ *are bounded on* $L^2(\mathbb{T}^3)$ *if and only if the operators* $P_iP_2bP_i^\perp P_2^\perp, P_i^\perp P_2bP_iP_2^\perp, P_iP_2^\perp bP_i^\perp P_2, P_i^\perp P_2^\perp bP_iP_2$ *with* $i\in\{1,3\}$ *are bounded on* $L^2(\mathbb{T}^3)$.

(2) The commutator $[[b,H_1H_3],H_2]$ *is bounded on* $L^2(\mathbb{T}^3)$ *if and only if all four operators* $P_2Q_{13}bQ_{13}^\perp P_2^\perp, P_2^\perp Q_{13}^\perp bQ_{13}P_2, P_2Q_{13}^\perp bQ_{13}P_2^\perp, P_2^\perp Q_{13}bQ_{13}^\perp P_2$ *are bounded on* $L^2(\mathbb{T}^3)$.

Proof. Using Remark 5.1 it is easy to see that

$$[[b,H_1],H_2]=4\big((P_2P_1bP_1^\perp P_2^\perp-P_2P_1^\perp bP_1P_2^\perp)-(P_2^\perp P_1bP_1^\perp P_2-P_2^\perp P_1^\perp bP_1P_2)\big)$$

and that the corresponding equation for $[[b,H_3],H_2]$ is also true. This, along with the observation that the ranges of all arising summands are mutually orthogonal, gives assertion *(1)*.

To prove *(2)* we just notice that $H_1H_3=Q_{13}-Q_{13}^\perp$ is a Hilbert transform type operator which permits us to repeat the above argument replacing P_1 by Q_{13}. □

The following lemma will allow us to insert an additional Hilbert transform into the commutator without reducing the norm.

Lemma 5.3. $\|P_3P_1^\perp P_2^\perp bP_1P_2P_3\|_{L^2\to L^2}=\|P_1^\perp P_2^\perp bP_1P_2\|_{L^2\to L^2}$.

Proof. The inequality $\le$ is trivial, since P_3 is a projection which commutes with $P_1^\perp$ and $P_2^\perp$. To see $\ge$, notice that $P_3P_1^\perp P_2^\perp bP_1P_2P_3$ is a Toeplitz operator with symbol $P_1^\perp P_2^\perp bP_1P_2$. So $\|P_3P_1^\perp P_2^\perp bP_1P_2P_3\|=\sup_{x_3}\|P_1^\perp P_2^\perp b(\cdot,\cdot,x_3)P_1P_2\|$. The latter is by Wiener's theorem just $\|P_1^\perp P_2^\perp bP_1P_2\|$. For convenience we include a sketch of the facts about Toeplitz operators we use. Let W_3 be the operator of multiplication by z_3, $W_3(f)=z_3f$, acting on $L^2(\mathbb{T}^3)$. If we define $B=P_1^\perp P_2^\perp bP_1P_2$ as well as

$$A_n=W_3^{*n}(P_3P_1^\perp P_2^\perp bP_1P_2P_3)W_3^n \text{ and } C_n=W_3^n(P_3^\perp P_1^\perp P_2^\perp bP_1P_2P_3^\perp)W_3^{*n}$$

as operators acting on $L^2(\mathbb{T}^3)$ then the sequences A_n and C_n converge to B in the strong operator topology: it is easy to see that W_3 , W_3^*; and P_3 commute with $P_1, P_2, P_1^\perp$ and $P_2^\perp$. The multiplier b satisfies the equation $W_3^{*n} b W_3^n = b$ and $W_3^n W_3^{*n} = Id$. So we see that

$$A_n = P_1^\perp P_2^\perp (W_3^{*n} P_3 W_3^n) b P_1 P_2 (W_3^{*n} P_3 W_3^n).$$

But if $f \in L^2(\mathbb{T}^3)$, then, since W_3^n is a unitary operator:

$$\|W_3^{*n} P_3 W_3^n(f) - f\| = \|P_3 W_3^n(f) - W_3^n(f)\| = \|(P_3 - I)(W_3^n)(f)\| \to 0 \qquad (n \to \infty),$$

as tail of a convergent Fourier series. This means that $W_3^{*n} P_3 W_3^n$ converges to the identity in the strong operator topology. Thus, for each $f \in L^2(\mathbb{T}^3)$ we have $\|(A_n - B)(f)\| \to 0$. So

$$\begin{aligned}\|P_1^\perp P_2^\perp b P_1 P_2\| &\leq \sup_{n \in \mathbb{N}} \|W_3^{*n} (P_3 P_1^\perp P_2^\perp b P_1 P_2 P_3) W_3^n\| \\ &\leq \|P_3 P_1^\perp P_2^\perp b P_1 P_2 P_3\|,\end{aligned}$$

□

Now, we are ready to proceed with the proof of the main theorem of this section.

Proof. (of Theorem 3.6) We show (1) ⇔ (2) and (2) ⇔ (3).

(1) ⇔ (2). Consider $f = f(x_1, x_2)$ and $g = g(x_3)$. Observe that $[[b, H_1], H_2](fg) = g \cdot [[b, H_1], H_2](f)$. So

$$\|[[b, H_1], H_2](fg)\|^2_{L^2(\mathbb{T}^3)} = \|Fg\|^2_{L^2(\mathbb{T})}$$

where $F(x_2) = \|[[b, H_1], H_2](f)\|_{L^2(\mathbb{T}^2)}$. The map $g \mapsto Fg$ has $L^2(\mathbb{T})$ operator norm $\|F\|_\infty$. Now change the roles of x_1 and x_3. The Ferguson-Lacey equivalences $\|[[b, H_i], H_2]\| \sim \|b\|_{\mathrm{BMO_{prod}}}$ give the desired result.

(2) $\Rightarrow$ (3). Boundedness of the commutators $[[b, H_1], H_2]$ and $[[b, H_3], H_2]$ implies the boundedness of the mixed commutator $[[b, H_1H_3], H_2]$ by the identity

$$[[b, H_1H_3], H_2] = H_1[[b, H_3], H_2] + [[b, H_1], H_2]H_3.$$

(3) $\Rightarrow$ (2). This part relies on Lemma 5.3. We wish to conclude from the boundedness of $[[b, H_1H_3], H_2]$ the boundedness of $[[b, H_1], H_2]$ and $[[b, H_3], H_2]$. To see boundedness of $[[b, H_1], H_2]$, let us look at one of the Hankels from Lemma 5.2. Lemma 5.3 shows that $P_2^\perp P_1^\perp bP_2P_1$ is bounded if and only if the operator $P_3P_1^\perp P_2^\perp bP_1P_2P_3$ is. And the latter is an operator found in the list from part *(2)* of Lemma 5.2. The analogous reasoning shows that all eight Hankels in Lemma 5.2 are bounded and so (2) is proved. □

5.2 Riesz commutators

In this section, we are again in $\mathbb{R}^{\vec{d}}$ with $\vec{d} = (d_1, \ldots, d_t)$ and a partition $\mathcal{I} = (I_s)_{1\leq s\leq l}$ of $\{1, \ldots, t\}$. It is our aim to prove Theorem 3.9, i.e. the equivalence of the following three statements:

(1) $b \in \mathrm{BMO}_\mathcal{I}(\mathbb{R}^{\vec{d}})$;

(2) All commutators of the form $[\ldots[b, R_{k_1,j_{k_1}}], \ldots, R_{k_l,j_{k_l}}]$ are bounded in $L^2(\mathbb{R}^{\vec{d}})$ where $k_s \in I_s$ and $R_{k_s,j_{k_s}}$ is the one-parameter Riesz transform in direction j_{k_s};

(3) All commutators of the form $[\ldots[b, \vec{R}_{1,\vec{j}^{(1)}}], \ldots, \vec{R}_{l,\vec{j}^{(l)}}]$ are bounded in $L^2(\mathbb{R}^{\vec{d}})$ where $\vec{j}^{(s)} = (j_k)_{k\in I_s}$, $1 \leq j_k \leq d_k$ and the operators $\vec{R}_{s,\vec{j}^{(s)}}$ are a tensor product of Riesz transforms $\vec{R}_{s,\vec{j}^{(s)}} = \bigotimes_{k\in I_s} R_{k,j_k}$.

It will be easy to see from our argument that the L^2 case of the more general Theorem 3.10 holds true as well, i.e. for any fixed $\vec{n} = (n_s)$ with $1 \leq n_s \leq |I_s|$ the two-sided estimate

$$\|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \lesssim \sup_{\vec{j}} \|[\ldots[b, \vec{R}_{1,\vec{j}^{(1)}}], \ldots, \vec{R}_{l,\vec{j}^{(l)}}]\|_{L^2(\mathbb{R}^{\vec{d}}) \to L^2(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \tag{5.4}$$

where $\vec{j}^{(s)} = (j_k)_{k \in I_s}$, $0 \leq j_k \leq d_k$ and for each s, there are n_s non-zero choices. A Riesz transform in direction 0 is understood as the identity. In fact, for $\vec{n} = \vec{1}$ this is the equivalence (1) $\Leftrightarrow$ (2) and for $\vec{n} = (|I_1|, \ldots, |I_l|)$ it is the equivalence (1) $\Leftrightarrow$ (3) from Theorem 3.9. All the other intermediate cases can be similarly deduced as in the proof of Theorem 3.9.

In the following, we only consider the case $d_k \geq 2$ for $1 \leq k \leq t$ and thus iterated commutators with tensor products of Riesz transforms only. The special case when $d_k = 1$ for some k (i.e. when Hilbert transform appears) is easier but requires extra care for notation, which is why we omit it here.

The proof in the Hilbert transform case relied heavily on analytic projections and orthogonal spaces, a feature that we do not have when working with Riesz transforms. We are going to simulate the one-dimensional case by a two-step passage via intermediary Calderón-Zygmund operators whose multiplier symbols are adapted to cones.

5.2.1 Intermediate step: commutators with cone multipliers

In dimension $d \geqslant 2$, a cone $C \subset \mathbb{R}^d$ with cubic base is given by the data (ξ, Q) where $\xi \in \mathbb{S}^{d-1}$ is the direction of the cone and the cube $Q \subset \xi^{\perp}$ centered at the origin is its aperture. The cone consists of all vectors θ that take the form $(\theta_\xi \xi, \theta_\perp)$ where $\theta_\xi = \langle \theta, \xi \rangle$ and $\theta_\perp \in \theta_\xi Q$. By λC we mean the dilated cone with data $(\xi, \lambda Q)$.

A cone D with ball base has data (ξ, r) for $0 < r < \pi/2$ and $\xi \in \mathbb{S}^{d-1}$ and consists of the vectors $\{\eta \in \mathbb{R}^d : d(\xi, \eta/\|\eta\|) \leq r\}$ where d is the geodesic distance (with distance of antipodal points being π.)

Given any cone C or D, we consider its Fourier projection operator defined via $\widehat{P_C f} = \chi_C \hat{f}$. When the apertures are cubes, such operators are combinations of Fourier projections onto half spaces and as such admit uniform L^p bounds. Among others, this fact made cubic cones necessary in the considerations in [LPPW09] and [DP14] that we are going to need. For further technical reasons in the proof these operators are not quite good enough, mainly because they are not of Calderón-Zygmund type. For a given cone C, consider a Calderón-Zygmund operator T_C with a kernel K_C whose Fourier symbol $\widehat{K_C} \in C^\infty$ and satisfies the estimate $\chi_C \leqslant \widehat{K_C} \leqslant \chi_{(1+\tau)C}$. This is accomplished by mollifying the symbol χ_C of the cone projection associated to cone C on $\mathbb{S}^{d-1}$ and then extending radially. We use the same definition for T_D.

Given a collection of cones $\vec{C} = (C_k)$ we denote by $T_{\vec{C}}, P_{\vec{C}}$ the corresponding tensor product operators.

In [LPPW09] it has been proven that Calderón-Zygmund operators adapted to certain cones of cubic aperture classify product BMO via commutators. As part of the argument, it was observed that test functions with opposing Fourier supports made the commutator large. In [DP14] a refinement was proven, that will be helpful to us. We prefer to work with cones with round base. Lower bounds for such commutators can be deduced from the assertion of the main theorem in [DP14], but we need to preserve the information on the Fourier support of the test function in order to succeed with our argument. Information on this test function is instrumental to our argument: it reduces the terms arising in the commutator to those resembling Hankel operators. We have the following lemma, very similar to that in [LPPW09] and [DP14], the only difference being that the cones are based on balls instead of cubes.

Lemma 5.5. *For every parameter $1 \leq k \leq t$ there exist a finite set of directions $\Upsilon_k \in \mathbb{S}^{d_k - 1}$ and an aperture $0 < r_k < \pi/2$ so that for every symbol b belonging to product BMO, there exist cones $D_k = D(\xi_k, r_k)$ with $\xi_k \in \Upsilon_k$ as well as a normalized test function $f = \bigotimes_{k=1}^t f_k$*

whose components have Fourier support in the opposing cones $D(-\xi_k, r_k)$ *so that*

$$\|[\ldots[b, T_{1,D_1}], \ldots, T_{t,D_t}]f\|_2 \gtrsim \|b\|_{BMO_{(1)\ldots(t)}(\mathbb{R}^{\vec{d}})}.$$

The stress is on the fact that the collection is finite, somewhat specific and serves all admissible product BMO functions.

Proof. The lemma in [DP14] supplies us with the sets of directions Υ_k as well as cones of cubic aperture Q_k and a test function f supported in the opposing cones. Now choose the aperture r_k large enough so that $(1+\tau)C(\xi_k, Q_k) \subset D(\xi_k, r_k)$. Then we have the commutator estimate

$$\|[\ldots[b, T_{1,D_1}], \ldots, T_{t,D_t}]f\|_2 \gtrsim \|b\|_{\mathrm{BMO}_{(1)\ldots(t)}(\mathbb{R}^{\vec{d}})}.$$

In fact, both commutators with cones C and D are L^2 bounded and reduce to $\|T_{\vec{D}}(bf)\|_2$ or $\|T_{\vec{C}}(bf)\|_2$ respectively thanks to the opposing Fourier support of f. Observe that $T_{\vec{C}}(bf) = T_{\vec{D}}(T_{\vec{C}}(bf)) = T_{\vec{C}}(T_{\vec{D}}(bf))$. With $\|T_{\vec{C}}\|_{2\to 2} \leq 1$, we see that $\|T_{\vec{D}}(bf)\|_2 \geq \|T_{\vec{C}}(bf)\|_2$. □

Using this a priori lower estimate, we are going to prove the lemma below.

Lemma 5.6. *Let us suppose we are in* $\mathbb{R}^{\vec{d}}$ *with* $\vec{d} = (d_1, \ldots, d_t)$ *and a partition* $\mathcal{I} = (I_s)_{1\leq s\leq l}$. *For every* $1 \leq k \leq t$ *there exists a finite set of directions* $\Upsilon_k \subset \mathbb{S}^{d_k - 1}$ *and an aperture* r_k *so that the following hold for all* $b \in BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$ *:*

(1) For every $1 \leq s \leq l$ *there exists a coordinate* $v_s \in I_s$ *and a direction* $\xi_{v_s} \in \Upsilon_{v_s}$ *and so that with the choice of cone* $D_{v_s} = D(\xi_{v_s}, r_{v_s})$ *and arbitrary* D_k *for coordinates* $k \in I_s \setminus \{v_s\}$ *and their tensor product* $\vec{D}_s$ *we have*

$$\|[\ldots[b, T_{1,\vec{D}_1}], \ldots, T_{l,\vec{D}_l}]\|_{2\to 2} \gtrsim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})},$$

(2) The test function $f = \bigotimes_{k=1}^{t} f_k$ which gives us a large L^2 norm in (1) has Fourier supports of the f_k contained in $D(-\xi_k, r_k)$ when $k = v_s$ and in D_k else.

Before we can begin with the proof of Lemma 5.6, we will need a real variable version of the facts on Toeplitz operators used earlier in Section 5.1.

Lemma 5.7. *Let D_k for $1 \leq k \leq t$ denote any cones with respect to the k^{th} variable. Let T_{D_k} denote the adapted Calderón-Zygmund operators. Let K be any proper subset of $\{k : 1 \leq k \leq t\}$, let $\vec{D}_K = \bigotimes_{k\in K} D_k$ and $T_{\vec{D}_K}$ the associated tensor product of Calderón-Zygmund operators. Let $P^{\sigma}_{\vec{D}_K}$ be a tensor product of projection operators on cones $D(\xi_k, r_k)$ or opposing cones $D(-\xi_k, r_k)$. Let $j \notin K$. Then*

$$\|T_{\vec{D}_K} T_{D_j} b P^{\sigma}_{\vec{D}_K} P_{D_j}\|_{L^2(\mathbb{R}^{\vec{d}})\to L^2(\mathbb{R}^{\vec{d}})} = \|T_{\vec{D}_K} b P^{\sigma}_{\vec{D}_K}\|_{L^2(\mathbb{R}^{\vec{d}})\to L^2(\mathbb{R}^{\vec{d}})}.$$

Proof. We will establish this by composing some unilateral shift operators and studying their Fourier transform in the j variable. Let ξ_j denote the direction of the cone D_j, for any l define the shift operator

$$S_l g(x_j) = \int_{\mathbb{R}^{d_j}} \hat{g}(\eta_j) e^{2\pi i (l\xi_j + \eta_j) x_j} \, d\eta_j.$$

S_l is a translation operator on the Fourier side along the direction ξ_j of the cone D_j. It is not hard to observe that $S_l^* = S_{-l}$. Now define

$$A_l = S_{-l} T_{\vec{D}_K} T_{D_j} b P^{\sigma}_{\vec{D}_K} P_{D_j} S_l, \text{ and } B = T_{\vec{D}_K} b P^{\sigma}_{\vec{D}_K}.$$

We will prove that as $l \to +\infty$, $A_l \to B$ in the strong operator topology. As in the argument in Lemma 5.3, this together with the fact that S_l is an isometry will complete the proof. To see the convergence, let's first remember that S_l only acts on the j variable, and one always has the identities

$$S_l S_{-l} = \mathrm{Id} \quad \text{and} \quad S_{-l} b S_l = b.$$

This implies

$$
\begin{aligned}
A_l &= T_{\vec{D}_K}(S_{-l}T_{D_j}S_l)(S_{-l}bS_l)P^{\sigma}_{\vec{D}_K}(S_{-l}P_{D_j}S_l) \\
&= T_{\vec{D}_K}(S_{-l}T_{D_j}S_l)bP^{\sigma}_{\vec{D}_K}(S_{-l}P_{D_j}S_l).
\end{aligned}
$$

We claim that both $S_{-l}T_{D_j}S_l$ and $S_{-l}P_{D_j}S_l$ converge to the identity operator in the strong operator topology, which then imply that $A_l \to B$ as $l \to \infty$. We will only prove $S_{-l}T_{D_j}S_l \to \mathrm{Id}$ as the second limit is almost identical. Observe that $\|S_{-l}T_{D_j}S_lf - f\| = \|(T_{D_j} - I)S_lf\|$. Given any L^2 function f and any fixed large $l \geq 0$, if f has frequencies supported in $\mathbb{R}^{d_1} \times \ldots \times (D_j - l\xi_j) \times \ldots \times \mathbb{R}^{d_t}$, then S_lf has Fourier support in $\mathbb{R}^{d_1} \times \ldots \times D_j \times \ldots \times \mathbb{R}^{d_t}$ where the symbol of T_{D_j} equals 1. Thus, for such f, we have $S_{-l}T_{D_j}S_lf = f$. Note that the sets $\mathbb{R}^{d_1} \times \ldots \times (D_j - l\xi_j) \times \ldots \times \mathbb{R}^{d_t}$ exhaust the frequency space. With $\|T_{D_j} - I\|_{2\to 2} \leq 1$ the operators $S_{-l}T_{D_j}S_l$ converge to the Identity in the strong operator topology, and the lemma is proved. Observe that the aperture of the cone D_j is not relevant to the proof. □

We proceed with the proof of the lower estimate for cone transforms.

Proof. (of Lemma 5.6) For a given symbol $b \in \mathrm{BMO}_{\mathcal{I}}$, there exist for all $1 \leq s \leq l$ coordinates $\boldsymbol{v} = (v_s), v_s \in I_s$ and a choice of variables not indexed by v_s, $\vec{x}^{\,0}_{\hat{\boldsymbol{v}}}$ so that up to an arbitrarily small error

$$
\|b\|_{\mathrm{BMO}_{\mathcal{I}}} = \|b(\vec{x}^{\,0}_{\hat{\boldsymbol{v}}})\|_{\mathrm{BMO}_{(v_1)\ldots(v_l)}}.
$$

By Lemma 5.5, there exist cones $D_{v_s} = D(\xi_{v_s}, r_{v_s})$ with directions $\xi_{v_s} \in \Upsilon_{v_s}$ and a normalized test function f_H in variables v_s with opposing Fourier support such that we have the lower estimate

$$
\|[\ldots[b(\vec{x}^{\,0}_{\hat{\boldsymbol{v}}}), T_{v_1,D_{v_1}}], \ldots, T_{v_l,D_{v_l}}](f_H)\|_{L^2(\mathbb{R}^{\vec{d}_{\boldsymbol{v}}})} \gtrsim \|b(\vec{x}^{\,0}_{\hat{\boldsymbol{v}}})\|_{\mathrm{BMO}_{(v_1)\ldots(v_l)}}
$$

where $\mathbb{R}^{\vec{d}_{\boldsymbol{v}}} = \mathbb{R}^{d_{v_1}} \times \ldots \times \mathbb{R}^{d_{v_l}}$.

We now consider the commutator with the same cones but with full symbol $b = b(\cdot, \dots, \cdot)$. Due to the lack of action on the variables not indexed by v_s, in the commutator, we have

$$[\dots[b, T_{v_1,D_{v_1}}], \dots, T_{v_l,D_{v_l}}](f_H g) = g \cdot [\dots[b, T_{v_1,D_{v_1}}], \dots, T_{v_l,D_{v_l}}](f_H)$$

for g that only depends upon variables not indexed by v_s. Again using that multiplication operators in L^2 have norms equal to the L^∞ norm of their symbol, for the "worst" L^2-normalized g we have

$$\begin{aligned}
&\|[\dots[b, T_{v_1,D_{v_1}}], \dots, T_{v_l,D_{v_l}}](f_H g)\|_{L^2(\mathbb{R}^{\vec{d}})} \\
=&\sup_{\vec{x}_{\hat{\boldsymbol{v}}}} \|[\dots[b(\vec{x}_{\hat{\boldsymbol{v}}}), T_{v_1,D_{v_1}}], \dots, T_{v_l,D_{v_l}}](f_H)\|_{L^2(\mathbb{R}^{\vec{d}_{\boldsymbol{v}}})} \\
\geqslant&\|[\dots[b(\vec{x}^{\,0}_{\hat{\boldsymbol{v}}}), T_{v_1,D_{v_1}}], \dots, T_{v_l,D_{v_l}}](f_H)\|_{L^2(\mathbb{R}^{\vec{d}_{\boldsymbol{v}}})} \\
\gtrsim&\|b(\vec{x}^{\,0}_{\hat{\boldsymbol{v}}})\|_{\mathrm{BMO}_{(v_1)\dots(v_l)}(\mathbb{R}^{\vec{d}_{\boldsymbol{v}}})} = \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.
\end{aligned}$$

Note that the test function g can be chosen with well distributed Fourier transform. Take any cones in the variables not indexed by v_s and let $\vec{D}$ denote the tensor product of their projections. $f_T = P_{\vec{D}} g$. Notice that

$$\|[\dots[b, T_{v_1,D_{v_1}}], \dots, T_{v_l,D_{v_l}}](f_H f_T)\| \gtrsim \|[\dots[b, T_{v_1,D_{v_1}}], \dots, T_{v_l,D_{v_l}}](f_H g)\|$$

with constants depending upon smallness of the aperture of the chosen cones. Notice that the test function $f := f_H f_T$ has the Fourier support as required in part *(2)* of the statement of Lemma 5.6.

Now build cones $\vec{D}_s$ from the D_{v_s} and the other chosen cones D_k as well as operators $T_{s,\vec{D}_s}$. Notice that the commutators

$$[\dots[b, T_{v_1,D_{v_1}}], \dots, T_{v_l,D_{v_l}}] \text{ and } [\dots[b, T_{1,\vec{D}_1}], \dots, T_{l,\vec{D}_l}]$$

reduce significantly when applied to a test function f with Fourier support like ours.

When the operators $T_{v_s,D_{v_s}}$ or any tensor product $T_{s,\vec{D}_s}$ fall directly on f, the contribution is zero due to opposing Fourier supports of the test function and the symbols of the operators. The only terms left in the commutators $[\ldots[b,T_{1,\vec{D}_1}],\ldots,T_{l,\vec{D}_{v_l}}](f)$ and $[\ldots[b,T_{v_1,D_{v_1}}],\ldots,T_{v_l,D_{v_l}}](f)$ have the form $\bigotimes_s T_{s,\vec{D}_s}(bf)$ and $\bigotimes_s T_{v_s,D_{v_s}}(bf)$ respectively.

By repeated use of Lemma 5.7 we have the operator norm estimates for any symbol b, valid on the subspace of functions with Fourier support as described for f: $\|\bigotimes_s T_{s,\vec{D}_s} b\|_{2\to 2} = \|\bigotimes_s T_{v_s,D_{v_s}} b\|_{2\to 2}$. We conclude the existence of a normalized test function f with Fourier support as described in the statement *(2)* of Lemma 5.6 so that $\|\bigotimes_s T_{s,\vec{D}_s}(bf)\|_2 \gtrsim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}$. In particular, we get the desired estimate in *(1)*. $\square$

5.2.2 Intermediate step: Journé commutators via zonal harmonics

It seems not possible to pass directly to a lower commutator estimate in tensor products of Riesz transforms from that in tensor products of cone operators. Just using tensor products of operators adapted to cones merely gives us *some* lower bound where we are unable to control that a Riesz transform does appear in every variable such as required in (3) of Theorem 3.9. The reason for this will become clear as we advance in the argument. Instead of using operators $T_{s,\vec{D}_s}$ directly, we will build upon them more general multi-parameter Journé type cone operators not of tensor product type that we now describe.

Let us explain the multiplier we need for i copies of $\mathbb{S}^{d-1}$ when all dimensions are the same. We will explain how to pass to the case of i copies of varying dimension d_k below. A picture illustrating a base case, a product of two 1-spheres, can be found at the end of this subsection.

For $0 < b < a < 1$, let $\varphi : [-1,1] \to [-1,1]$ be a smooth function with $\varphi(x) = 1$ when $a \leqslant x \leqslant 1$ and $\varphi(x) = 0$ when $b \geqslant x \geqslant 0$. And let φ be odd, meaning antisymmetric about $t = 0$. The function φ gives rise to a zonal function with pole ξ_1 on the first copy of $\mathbb{S}^{d-1}$, denoted by $C_1(\xi_1;\eta_1)$. This is the multiplier of a one-parameter Calderón-

Zygmund operator adapted to a cone $D(\xi_1, r)$ for $r = \pi/2(1-a)$. For $i > 1$ we define $C_k(\xi_1, \ldots, \xi_k; \eta_1, \ldots, \eta_k)$ for $1 < k \leqslant i$ inductively. In what follows, expectation is taken with respect to traces of surface measure. When $\eta_i = \pm\xi_i$, then conditional expectation is over a one-point set.

$$
\begin{aligned}
&C_k(\xi_1, \ldots, \xi_k; \eta_1, \ldots, \eta_k) \\
&= \mathbb{E}_{a_{k-1}}(C_{k-1}(\xi_1, \ldots, a_{k-1}; \eta_1, \ldots, \eta_{k-1}) \mid d(a_{k-1}, \xi_{k-1}) = d(\eta_k, \xi_k)).
\end{aligned}
$$

If the dimensions are not equal take $d = \max_{d_j}$ and imbed $\mathbb{S}^{d_j-1}$ into $\mathbb{S}^{d-1}$ by the map $\xi = (\xi_1, \ldots, \xi_{d_j}) \mapsto (\xi_1, \ldots, \xi_{d_j}, 0, \ldots, 0)$. Obtain in this manner the function C_i and then restrict to the original number of variables when the dimension is smaller than d.

The multiplier $\vec{J} = C_i(\vec{\xi}; \cdot)$ gives rise to a multi-parameter Calderón-Zygmund operator of convolution type (but not of tensor product type), $\vec{T}_{\vec{J}} = \vec{T}_{C_i(\vec{\xi}; \cdot)}$. In fact, it is defined through principal value convolution against a kernel $\vec{K}_{\vec{J}} = \vec{K}_{C_i(\vec{\xi}; \cdot)}(x_1, \ldots, x_i)$ such that

$$
\forall\, l : \int_{\alpha < |x_l| < \beta} \vec{K}_{\vec{J}}(x_1, \ldots, x_i) dx_l = 0, \forall\, 0 < \alpha < \beta, x_j \in \mathbb{R}^{d_j} \text{ fixed } \forall\, j \neq l,
$$

$$
\left| \frac{\partial^{|\vec{n}|}}{\partial x_1^{n_1} \ldots \partial x_i^{n_i}} \vec{K}_{\vec{J}}(x_1, \ldots, x_i) \right| \leqslant A_{\vec{n}} |x_1|^{-d_1 - n_1} \ldots |x_i|^{-d_i - n_i}, n_j \geqslant 0.
$$

This kind of operator is a special case of the more general, non-convolution type discussed in Chaper 2. It has many other nice features that will facilitate our passage to Riesz transforms. One of them is its very special representation in terms of homogeneous polynomials, the other one a lower commutator estimate in terms of the $\mathrm{BMO}_{\mathcal{I}}$ norm.

Lemma 5.8. *Let C_i be a multiplier in $\bigotimes_{k=1}^{i} \mathbb{R}^{d_k}$ as described above, with any fixed direction and aperture. Let m be an integer of order $d = \max d_k$. For any $\delta > 0$, the function C_i has an approximation by a polynomial C_i^N in the $\prod_{k=1}^{i} d_k$ variables $\{\prod_{k:1 \leqslant k \leqslant i} \eta_{k, j_k} \mid 1 \leqslant j_k \leqslant d_k\}$ so that $\|C_i - C_i^N\|_{\mathcal{C}^m(\mathbb{S}^{d_k - 1})} < \delta$ in each variable separately.*

$\mathcal{C}^m$ indexes the norm of uniform convergence in functions that are m times continuously differentiable. On the space side, C_i^N corresponds to an operator that is a polynomial in Riesz transforms of the variables $\bigotimes_k R_{k,j_k}$.

Lemma 5.9. *We are in $\mathbb{R}^{\vec{d}}$ with partition $\mathcal{I} = (I_s)_{1\leq s\leq l}$. Let $\vec{\Upsilon}$ consist of vectors $\vec{\xi} = (\xi_k)_{k=1}^t$ with $\xi_k \in \Upsilon_k$. Let $\vec{\Upsilon}^{(s)}$ consist of $\vec{\xi}^{(s)} = (\xi_k)_{k\in I_s}$. Let us consider the class of Journé type cone multipliers $\vec{J}_s = C_{i_s}(\vec{\xi}^{(s)};\cdot)$ of aperture r_s with associated multi-parameter Journé operators $\vec{T}_{s,\vec{J}_s}$. Then we have the two-sided estimate*

$$\|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})} \lesssim \sup_{\vec{\xi}\in\vec{\Upsilon}} \|[\ldots[b,\vec{T}_{1,\vec{J}_1}],\ldots,\vec{T}_{l,\vec{J}_l}]\|_{L^2(\mathbb{R}^{\vec{d}})\to L^2(\mathbb{R}^{\vec{d}})} \lesssim \|b\|_{BMO_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}.$$

In order to proceed with the proof of these lemmas, we will use some well known facts about zonal harmonics. Fix a pole $\xi \in \mathbb{S}^{d-1}$. The zonal harmonic with pole ξ of degree n is written as $Z_\xi^{(n)}(\eta)$. With $t = \langle \xi,\eta\rangle \in [-1,1]$, one writes $Z_\xi^{(n)}(\eta) = P_n(t)$ where P_n is the Legendre polynomial of degree n. It is common to suppress the dependence on d in the notation for $Z_\xi^{(n)}$ and P_n. $Z_\xi^{(n)}$ are reproducing for spherical harmonics of degree n, $Y^{(n)}$.

When $Y^{(n)}$ is harmonic and homogeneous of degree n with $Y^{(n)}(\xi) = 1$ and $Y^{(n)}(R\eta) = Y^{(n)}(\eta)$ for any rotation $R \in \mathcal{O}(d)$ with $R\xi = \xi$, then $Y^{(n)} = Z_\xi^{(n)}$.

The lemma below will aid us in understanding the special form of the functions C_i.

Lemma 5.10. *Let $\xi_1,\xi_2 \in \mathbb{S}^{d-1}$. We have*

$$\begin{aligned} Z_{\xi_1}^{(n)}(\eta_1)Z_{\xi_2}^{(n)}(\eta_2) &= \mathbb{E}_{a_1}(Z_{\eta_1}^{(n)}(a_1) \mid d(\xi_1,a_1) = d(\xi_2,\eta_2)) \\ &= \mathbb{E}_{a_2}(Z_{\eta_2}^{(n)}(a_2) \mid d(\xi_2,a_2) = d(\xi_1,\eta_1)). \end{aligned}$$

Proof. The first equality is a change of variable, thanks to symmetry of the zonal harmonic in its variables and invariance with respect to action of the measure preserving elements of the orthogonal group fixing poles ξ_1 or ξ_2, that we now detail. By a rotation in one of the spheres, assume $\xi_1 = \xi_2 = \xi$. Take a small ball

$$B_{\xi,\eta_1}(a_2^0; \varepsilon_2) = \{a_2 : d(a_2, a_2^0) < \varepsilon_2\} \cap \{a_2 : d(a_2, \xi) = d(\eta_1, \xi)\}.$$

Note $\{a_2 : d(a_2, \xi) = d(\eta_1, \xi)\} \sim \mathbb{S}^{d-2}$. Every $a_2 \in B_{\xi,\eta_1}(a_2^0; \varepsilon_2)$ gives rise to a canonical orthogonal map σ_{a_2} along geodesics in a scaled copy of $\mathbb{S}^{d-2}$. Lifted to $\mathbb{S}^{d-1}$, these are orthogonal maps fixing ξ. Let σ^0 fix ξ and map a_2^0 to η_1. Let $a_1^0 = \sigma^0(\eta_2)$. We observe that $\{\sigma^0 \sigma_{a_2}(\eta_2) : a_2 \in B_{\xi,\eta_1}(a_2^0; \varepsilon_2)\} = B_{\xi,\eta_2}(a_1^0; \varepsilon_1)$ with ε_1 so that

$$\mathbb{P}(d(a_2, a_2^0) < \varepsilon_2 \mid d(\xi, a_2) = d(\xi, \eta_1)) = \mathbb{P}(d(a_1, a_1^0) < \varepsilon_1 \mid d(\xi, a_1) = d(\xi, \eta_2)).$$

Together with the symmetry and the rotation property $Z_\eta^{(n)}(a) = Z_a^{(n)}(\eta) = Z_{\sigma(a)}^{(n)}(\sigma(\eta))$, we obtain the first equality.

For fixed a_1, the function $Z_{\eta_1}^{(n)}(a_1) = Z_{a_1}^{(n)}(\eta_1)$ is a function harmonic in $\mathbb{R}^d$, homogeneous of degree n. These properties are preserved when taking expectation in a_1. So the expression $\mathbb{E}(Z_{\eta_1}^{(n)}(a_1) \mid d(\xi_1, a_1) = d(\xi_2, \eta_2))$ remains harmonic (regarded as a function in $\mathbb{R}^d$), n-homogeneous. From the form $\mathbb{E}(Z_{\eta_2}^{(n)}(a_2) \mid d(\xi_2, a_2) = d(\xi_1, \eta_1))$ we learn that its restriction to $\mathbb{S}^{d-1}$ depends only upon $d(\xi_1, \eta_1)$. This implies that it is a constant multiple of the zonal harmonic with pole ξ_1. Exchanging the roles of η_1 and η_2 gives

$$\mathbb{E}(Z_{\eta_1}^{(n)}(a_1) \mid d(\xi_1, a_1) = d(\xi_2, \eta_2)) = c_n Z_{\xi_1}^{(n)}(\eta_1) Z_{\xi_2}^{(n)}(\eta_2).$$

When assuming the normalization $Z_\xi^{(n)}(\xi) = 1$ then $c_n = 1$. This is a generalization of the classical symmetrizing of the cosines sum formula $1/2(\cos(x+y) + \cos(x-y)) = \cos(x)\cos(y)$. $\square$

Proof. (of Lemma 5.8) It is well known that zonal harmonic series have convergence properties when representing smooth zonal functions similar to that of the Fourier transform.

For any given m and sufficiently smooth φ of the type described above, then

$$C_1(\xi_1;\eta_1) = \sum_n \varphi_n Z^{(n)}_{\xi_1}(\eta_1)$$

where the convergence is $\mathcal{C}^m$-uniform. The degree of smoothness required for φ to obtain convergence in the $\mathcal{C}^m$ in the above expression depends upon m and the dimension d. For our purpose, we choose $m \geqslant d$.

Let us denote this function's representation of degree N by a series of zonal harmonics by $C_1^{(N)}(\xi_1;\eta_1)$.

$$C_1^{(N)}(\xi_1;\eta_1) = \sum_{n\leqslant N} \varphi_n Z^{(n)}_{\xi_1}(\eta_1).$$

For every $\delta > 0$ there exists N so that we have the estimate

$$\|C_1^{(N)}(\xi_1;\eta_1) - C_1(\xi_1;\eta_1)\|_{\mathcal{C}^m(\mathbb{S}^{d_1-1})} < \delta.$$

In the case of i copies of spheres, we define $C_i^{(N)}$ inductively in the same manner as C_i. Let us for the moment make all dimensions equal using the argument discussed above. So we set

$$\begin{aligned}&C_k^{(N)}(\xi_1,\dots,\xi_k;\eta_1,\dots,\eta_k)\\ &= \mathbb{E}_{a_{k-1}}(C_{k-1}^{(N)}(\xi_1,\dots,a_{k-1};\eta_1,\dots,\eta_{k-1}) \mid d(a_{k-1},\xi_{k-1}) = d(\eta_k,\xi_k)).\end{aligned}$$

We claim the identity

$$C_i^{(N)}(\vec{\xi};\eta_1,\eta_2,\dots,\eta_i) = \sum_{n\leqslant N} \varphi_n \prod_{k=1}^{i} Z^{(n)}_{\xi_k}(\eta_k). \tag{5.11}$$

This is trivially true for $i = 1$. For $i > 1$ induct on the number of parameters:

$$\begin{aligned}C_i^{(N)}(\vec{\xi};\eta_1,\dots,\eta_i) &= \mathbb{E}_{a_{i-1}}(C_{i-1}^{(N)}(\xi_1,\xi_2,\dots,a_{i-1};\eta_1,\dots,\eta_{i-1}) \mid d(a_{i-1},\xi_{i-1}) = d(\eta_i,\xi_i))\\ &= \mathbb{E}_{a_{i-1}}\left(\sum_{n\leqslant N} \varphi_n \Big(\prod_{k=1}^{i-2} Z^{(n)}_{\xi_k}(\eta_k)\Big) Z^{(n)}_{a_{i-1}}(\eta_{i-1}) \mid d(a_{i-1},\xi_{i-1}) = d(\eta_i,\xi_i)\right),\end{aligned}$$

where the first equality is the definition of $C_i^{(N)}$ and the second one is the induction hypothesis. Then

$$
\begin{aligned}
&= \sum_{n\leqslant N} \varphi_n \prod_{k=1}^{i-2} Z_{\xi_k}^{(n)}(\eta_k)\mathbb{E}_{a_{i-1}}(Z_{a_{i-1}}^{(n)} \mid d(a_{i-1},\xi_{i-1}) = d(\eta_i,\xi_i)) \\
&= \sum_{n\leqslant N} \varphi_n \prod_{k=1}^{i} Z_{\xi_k}^{(n)}(\eta_k).
\end{aligned}
$$

where the last equality is an application of Lemma 5.10.

It follows that neither C_i nor $C_i^{(N)}$ depends on the order chosen in their definition and

$$
C_i(\vec{\xi};\eta_1,\dots,\eta_i) = \sum_n \varphi_n \prod_{k=1}^{i} Z_{\xi_k}^{(n)}(\eta_k)
$$

where the convergence is in $\mathcal{C}^m$ in each variable.

Next, we study the terms arising in multipliers of the form $C_i^{(N)}$. When all dimensions are equal, $\prod_{k=1}^i Z_{\xi_k}^{(n)}(\eta_k)$ has the important property that, as a product of n-homogeneous polynomials, has only terms of the form

$$
\prod_{k=1}^{i} \eta_k^{\alpha_k} = \prod_{k=1}^{i} \left(\prod_{j_k=1}^{d} \eta_{k,j_k}^{\alpha_{k,j_k}} \right)
$$

where $\eta_k \in \mathbb{S}^{d-1}$ and $\alpha_k = (\alpha_{k,j_k})$ are multi-indices with $|\alpha_k| = \sum_{j_k} \alpha_{k,j_k} = n$ for all k. This form is inhertited by $C_i^{(N)}$ with varying n. It shows that $C_i^{(N)}$ is indeed a polynomial in the variables $\prod_{k=1}^i \eta_{k,j_k}$. When the dimensions d_k are not equal, observe that by restricting to the actually arising number of variables, we certainly lose harmonicity of the polynomials, but not n-homogeneity or the required form of our polynomials. □

Proof. (of Lemma 5.9) By Lemma 5.6 we know that for each parameter $1 \leqslant s \leqslant l$ there exists a tensor product of cones $\vec{D}_s = \bigotimes_{k\in I_s} D(\xi_k, r_k)$ with $r_s := \sum_{k\in I_s} r_k < \pi/2$ and

$\xi_k \in \Upsilon_k$ and test functions f_s supported as described in Lemma 5.6 part *(2)* so that

$$\|[\dots[b, T_{1,\vec{D}_1}], \dots, T_{l,\vec{D}_l}](f)\|_2 \gtrsim \|b\|_{\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})}$$

where $f = \bigotimes_{s=1}^{l} f_s$. We make a remark about the apertures r_s. Let $d(\cdot,\cdot)$ denote geodesic distance on $\mathbb{S}^{d-1}$, where antipodal points have distance π. Let $\vec{\xi}^{(s)}$ be the set of directions of $\vec{D}_s$. Remember that according to Lemma 5.6, one component had a specific direction $\xi_v^{(s)} \in \Upsilon_v$ and possibly large aperture with $(1+\tau)r_v^{(s)} < \pi/2$. Let us choose the other directions arbitrarily but with apertures $r_{\hat{v}}^{(s)}$ small enough so that $(1+\tau)(r_v^{(s)} + (i-1)r_{\hat{v}}^{(s)}) < \pi/2$. Now choose an aperture $r_s < \pi/2$ so that $(1+\tau)(r_v^{(s)} + (i-1)r_{\hat{v}}^{(s)}) < r_s < \pi/2$.

Writing $i_s = |I_s|$, we find Journé type cone multipliers $\vec{J}_s = C_{i_s}(\vec{\xi}^{(s)}; \cdot)$ according to the construction above with center $\vec{\xi}^{(s)}$ and aperture r_s.

We are going to observe that $\vec{J}_s \equiv 1$ on $\mathrm{spt}(\vec{D}_s)$ and $\vec{J}_s \equiv -1$ on the Fourier support of f_s. Let us drop the dependence on s for the moment. We see in an inductive manner that $C_i(\vec{\xi}; \cdot)$ takes the value 1 in a certain ℓ^1 ball of radius $r < \pi/2$ centered at $\vec{\xi}$. We show that

$$\sum_k d(\xi_k, \eta_k) < r \Rightarrow C_i(\vec{\xi}, \eta_1, \dots, \eta_i) = 1.$$

When $i = 1$, the assertion is obviously true: $d(\xi_1, \eta_1) < r \Rightarrow C_1(\xi_1; \eta_1) = 1$ by the choice of φ, r and definition of C_1. For $i > 1$, we proceed by induction. Assume that $\sum_{k=1}^{i-1} d(\xi_k, n_k) < r$ implies $C_{i-1}(\xi_1, \dots, \xi_{i-1}; \eta_1, \dots, \eta_{i-1}) = 1$. Let us assume that

$$\sum_{k=1}^{i} d(\xi_k, \eta_k) < r.$$

Remembering the definition of $C_i(\vec{\xi}; \cdot)$ we assume $d(a_{i-1}, \xi_{i-1}) = d(\eta_i, \xi_i)$. By the triangle inequality $\sum_{k=1}^{i-2} d(\xi_k, \eta_k) + d(a_{i-1}, \eta_{i-1}) \leqslant \sum_{k=1}^{i-2} d(\xi_k, \eta_k) + d(a_{i-1}, \xi_{i-1}) + d(\xi_{i-1}, \eta_{i-1}) = \sum_{k=1}^{i} d(\xi_k, \eta_k) < r$. So

$$C_{i-1}(\xi_1, \xi_2, \dots, a_{i-1}; \eta_1, \dots, \eta_{i-1}) = 1$$

for all a_{i-1} we condition by. The statement for i follows.

Since $C_i(\vec{\xi};\cdot)$ does not depend upon the order of the variables in its construction, we are also able to see exactly as done above that when $\sigma_k = -1$ for exactly one choice of k, then $\sum_k d(\sigma_k \xi_k, \eta_k) < r \Rightarrow C_i(\vec{\xi}; \eta_1, \dots, \eta_i) = -1$.

Consider associated multi-parameter Journé operators $\vec{T}_{s,\vec{J}_s}$ and $\vec{\mathrm{Id}}_s = \bigotimes_{k\in I_s} \mathrm{Id}_k$ and Id_k the identity on the k^{th} variable. Now

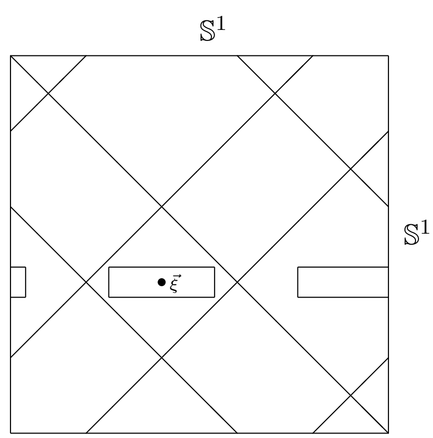
$$\begin{aligned}
[\dots[b, \vec{T}_{1,\vec{J}_1}], \dots, \vec{T}_{l,\vec{J}_l}](f) &= [\dots[b, \vec{T}_{1,\vec{J}_1} + \vec{\mathrm{Id}}_1], \dots, \vec{T}_{l,\vec{J}_l} + \vec{\mathrm{Id}}_l](f) \\
&= \bigotimes_{s=1}^{l} (\vec{T}_{s,\vec{J}_s} + \vec{\mathrm{Id}}_s)(bf).
\end{aligned}$$

With $\| \bigotimes_{s=1}^{l} (\vec{T}_{s,\vec{J}_s} + \vec{\mathrm{Id}}_s)(bf)\|_2 \geqslant \| \bigotimes_{s=1}^{l} T_{s,\vec{D}_s}(bf)\|_2$ and

$$\bigotimes_{s=1}^{l} T_{s,\vec{D}_s}(bf) = [\dots[b, T_{1,\vec{D}_1}], \dots, T_{l,\vec{D}_l}](f)$$

we get the desired lower bound on the Journé commutator as claimed. □

Let us illustrate the base case $(\mathbb{S}^1)^2$ by a picture. The picture is simplified in the sense that the odd function φ above is replaced by an indicator function of an interval.



In the picture, cone functions based on the oblique strips containing $\vec{\xi}$ are averaged. In the two-dimensional case, $\mathbb{S}^1$, expectation is over a one or two point set only. The rectangle around $\vec{\xi}$ with sides parallel to the axes representing $\mathbb{S}^1$ illustrates the support of the tensor product of cone operators with direction $\vec{\xi}$. The longer side is the aperture that arises from the Hankel part. The short sides can be chosen freely as they arise from the Toeplitz part and are chosen small so that the rectangle fits into the oblique square. The other small rectangle corresponds to the Fourier support of the test function f.

5.2.3 Proof of the main theorem

Proof. (of Theorem 3.9) Different from the Hilbert transform case, both lower bounds require separate proofs. This is a notable difference that stems from the loss of orthogonal subspaces in conjunction with the special form of the Hilbert transform only seen in one variable. It does not seem possible to get a lower estimate (3)⇒(2) directly.

(1)⇔(2). The upper bound (1)⇒(2) is an easy consequence of the upper estimates of iterated commutators of single Riesz transforms. The lower bound (2)⇒(1) follows from Wiener's lemma in combination with the main result in [LPPW09], the two-sided estimate for iterated commutators with Riesz transforms, similar to the first arguments used in Lemma 5.6.

(1)⇔(3). The upper bound (1)⇒(3) follows from the tensor product structure and use of the little product BMO norm (see Theorem 3.13 that is proved in Chapter 4). The lower bound (3)⇒(1) uses the considerations leading up to this proof: Suppressing again the dependence on s, we see that the multiplier C_i is an odd, smooth, bounded function in each η_k when the other variables are fixed. Furthermore, since φ, written as a function of $t = \langle \xi, \eta \rangle$ is odd with respect to $t = 0$, then the above series has $\varphi_n \neq 0$ at most when n is odd and thus $Z_\xi^{(n)}$ is odd. So $C_i^{(N)}$ is as a sum of odd functions.

We are now also ready to see that $\vec{T}_{\vec{J}}$, the Journé operator associated to the cone

$\vec{J} = C_i(\vec{\xi};\cdot)$ as well as the operator associated to $C_i^{(N)}(\vec{\xi};\cdot)$ are paraproduct free. In fact, applied to a test function $f = \bigotimes_k f_k$ with f_k acting on the k^{th} variable and where $f_l \equiv 1$ for some l gives $\vec{T}_{\vec{J}}(f) = 0$. To see this, apply the multiplier $C_i^{(N)}(\vec{\xi};\cdot)$ in the l variable (acting on 1) first, leaving the other Fourier variables fixed. The multiplier function

$$\eta_l \mapsto C_i^{(N)}(\vec{\xi};\eta_1,\ldots,\eta_i) = \sum_{n\leqslant N} \varphi_n Z_{\xi_l}^{(n)}(\eta_l) \prod_{k\neq l,k=1}^{i} Z_{\xi_k}^{(n)}(\eta_k)$$

is, as a sum of odd functions, odd on $\mathbb{S}^{d_l-1}$, bounded by 1 and uniformly smooth for all choices of η_k with $k \neq l$. As such it gives rise to a paraproduct free convolution type Journé operator in the l variable whose values are multi-parameter multiplier operators.

Due to the convergence properties proved above, the difference

$$C_i(\vec{\xi};\cdot) - C_i^{(N)}(\vec{\xi};\cdot)$$

gives rise to a paraproduct free Journé operator with norm depending on N. This is seen by an application of an appropriate version of the Marcinkiewicz multiplier theorem.

By our stability result on Journé commutators (Corollary 3.17), discussed in Chaper 3, there exist for all $1 \leq s \leq l$ integers N_s so that $C_s^{(N_s)}(\vec{\xi}_s;\cdot)$ with $\xi_k \in \Upsilon_k$ is a characterizing set of operators via commutators for $\mathrm{BMO}_{\mathcal{I}}(\mathbb{R}^{\vec{d}})$. This is a finite set of possibilities because of the universal choice of the r_s and finiteness of the set $\vec{\Upsilon}$. Using the multi-parameter analog of the observation $[b, AB] = A[b, B] + [b, A]B$ and the special form of the $C_s^{(N_s)}(\vec{\xi};\cdot)$, it leaves us with the desired lower bound: Observe that when $[b, AB]$ has large L^2 norm then either $[b, A]$ or $[b, B]$ has a fair share of the norm. We use this argument finitely many times in a row for operators that are polynomials in tensor products of Riesz transforms $\bigotimes_{k\in I_s} R_{k,j_k}$. This finishes (3)⇒(1). □

We remark that there are two cases of dimension greater than 1, where the proof simplifies. In the case of arbitrarily many copies of $\mathbb{R}^2$, one can work with the multiplicative structure of complex numbers and avoid the symmetrizing procedure to obtain cone functions with the appropriate polynomial approximations. If the dimensions are arbitrary, but only tensor products of two Riesz transforms arise, one can avoid part of the construction above by using the addition formula for zonal harmonics.

5.3 Remarks about our results in L^p

As mentioned before, the two-sided estimates stated in Section 5.2 and in particular Theorem 3.10 hold for all $1 < p < \infty$. The fact that upper estimates hold for the Riesz commutator in L^p in the case where no tensor products are present is proved in [LPPW09] as well as [LPPW10]. It stems from the fact that endpoint estimates for multi-parameter paraproducts hold for all $1 < p < \infty$ [MPTT04], [MPTT06]. This estimate carries over easily to tensor products of Riesz transforms or any other tensor products of operators for which we have L^p estimates on the commutator: one uses $[b, T_1T_2] = T_1[b, T_2] + [b, T_1]T_2$ to handle arising tensor products, followed by a correct use of the little product BMO norm. The argument is very similar to the proof of Theorem 3.13 given in Chapter 4, which we omit.

The lower estimate or the necessity of the BMO condition can be derived from interpolation. In fact, suppose we have uniform boundedness of our commutators with operators running through all choices of Riesz transforms and some symbol b in L^p. Then by duality, we have boundedness in L^q where $1/p + 1/q = 1$. In fact, $[T, b]^* f = -[T^*, \bar{b}]f = -\overline{[\overline{T^*}, b]\bar{f}}$ shows that the boundedness of adjoints is inherited. The same reasoning holds for iterated commutators of tensor products. Thus by interpolation, the boundedness holds in L^2 and the symbol function b necessarily belongs to the required BMO class.

Chapter Six

Little Product BMO and Weak Factorization of Hardy spaces

We have developed so far the characterization of the little product BMO spaces in terms of boundedness of commutators. In this chapter, we want to further investigate this new family of multi-parameter BMO spaces and their relations among each other. First of all, a partial order will be established among the family. In fact, in the proof of the upper bound estimate for general Journé commutators (Theorem 3.16), and in turn the Riesz commutator lower bound, it has been important for us to observe that little product BMOs are contained in product BMO. Second, we will study the pre-dual spaces of little product BMOs and prove weak factorization results for them, similarly as Theorem 3.5 mentioned in Chapter 3.

6.1 Partial orders of little product BMOs

In Section 3.1 of Chapter 3, we have observed that little BMO is contained in product BMO, by means of their commutator characterization. It is not difficult to observe that similar relations are preserved by all little product BMOs. In fact, one can show that there is certain partial order existing among these spaces, with little BMO and product BMO being the two endpoint spaces.

Proposition 6.1. *Consider* $\mathbb{R}^{\vec{d}} = \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_t}$*. Let* $\mathcal{I}_1 = \{I^1_{s_1} : 1 \leq s_1 \leq l_1\}$*,* $\mathcal{I}_2 = \{I^2_{s_2} : 1 \leq s_2 \leq l_2\}$ *be two partitions of* $\{1, 2, \ldots, t\}$ *so that* $\dot\cup_{1\leq s_i \leq l_i} I^i_{s_i} = \{1, 2, \ldots, t\}$*,* $i = 1, 2$*. Suppose* $\mathcal{I}_2$ *is a* ***refinement*** *of* $\mathcal{I}_1$*, meaning that for any* $I^2_{s_2} \in \mathcal{I}_2$*, there exists some* $I^1_{s_1} \in \mathcal{I}_1$ *such that* $I^2_{s_2} \subset I^1_{s_1}$*. Then,* $BMO_{\mathcal{I}_1}(\mathbb{R}^{\vec{d}}) \subset BMO_{\mathcal{I}_2}(\mathbb{R}^{\vec{d}})$*.*

We first observe that iterated commutators are symmetric. More precisely, consider $\mathbb{R}^{\vec{d}}$, $\vec{d} = (d_1, \ldots, d_t)$ with a partition $\mathcal{I} = (I_s)_{1\leq s \leq l}$ of $\{1, \ldots, t\}$. Then

$$[\ldots[b, T_1], \ldots, T_l] = [\ldots[b, T_{\sigma(1)}], \ldots, T_{\sigma(l)}],$$

where T_s denotes a singular integral acting on functions defined on $\bigotimes_{k\in I_s} \mathbb{R}^{d_k}$, $1 \leq s \leq l$, and σ is any permutation of $\{1, \ldots, l\}$. This observation is very natural, as little product

BMOs are all symmetric. And it is straightforward to verify by induction.

Proof. (of Proposition 6.1) Let $b \in \mathrm{BMO}_{\mathcal{I}_1}$. One wants to show that for any choice $\boldsymbol{v} = (v_{s_2}), v_{s_2} \in I^2_{s_2}$, b is uniformly in product BMO in the variables indexed by $\{v_{s_2}\}$. According to the Riesz commutator estimate proved in [LPPW09], this is equivalent to the boundedness of

$$[\ldots[b, R_{v_1, j_{v_1}}], \ldots, R_{v_{l_2}, j_{v_{l_2}}}] \tag{6.2}$$

uniformly in the variables indexed by $\{v_{s_2}\}$ for any choices of Riesz transforms.

To see that it is bounded, choose any subset of $\{v_{s_2}; 1 \leq s_2 \leq l_2\}$, denoted by $\{v'_{t_2}\}$, such that there exists a unique $v'_{t_2} \in I^1_{s_1}$, for any $1 \leq s_1 \leq l_1$. This is possible since $\mathcal{I}_2$ is a refinement of $\mathcal{I}_1$. Then, according to the definition of $\mathrm{BMO}_{\mathcal{I}_1}$, b is uniformly in product BMO in the variables indexed by $\{v'_{t_2}\}$, which is equivalent to the boundedness of $[\ldots[b, R_{v'_1, j_{v'_1}}], \ldots, R_{v'_{l_1}, j_{v'_{l_1}}}]$. The proof is thus complete by observing the symmetry of the commutator (6.2) and the fact that Riesz transforms in the rest of the variables of $\{v_{s_2}\}$ are all bounded. □

Therefore, one has the following relation for any little product BMO space:

$$\mathrm{bmo} \subset \mathrm{BMO}_{\mathcal{I}} \subset \mathrm{BMO}_{\mathrm{prod}}.$$

Moreover, in Proposition 6.1, when $\mathcal{I}_1 \neq \mathcal{I}_2$, the containedness is always proper. To see this, let's take $\mathrm{BMO}_{(123)}$ and $\mathrm{BMO}_{(12)(3)}$ as an example. According to Proposition 6.1, $\mathrm{BMO}_{(123)} \subset \mathrm{BMO}_{(12)(3)}$, and we want to show that it is properly contained. Let $b(x_1, x_2, x_3) = b_1(x_1) \otimes b_2(x_2) \otimes b_3(x_3)$, where $b_1 \in L^\infty(\mathbb{R}^{d_1})$, $b_2 \in L^\infty(\mathbb{R}^{d_2})$, $b_3 \in \mathrm{BMO}(\mathbb{R}^{d_3})$ but not in $L^\infty(\mathbb{R}^{d_3})$. Obviously, $b \in \mathrm{BMO}_{(12)(3)}$, but is not in $\mathrm{BMO}_{(123)}$.

6.2 Pre-duals of little product BMOs

In this section, we study the pre-duals of little product BMO spaces, and prove weak factorization results for them, as a corollary of the two-sided commutator estimates developed in Chapter 5. We will study the pre-dual of $\mathrm{BMO}_{(13)(2)}(\mathbb{T}^{\vec{1}})$ as an example, which is a good model for other cases. We choose the order of variables most convenient for us.

Theorem 6.3. *The pre-dual of the space* $BMO_{(13)(2)}(\mathbb{T}^{\vec{1}})$ *is equal to the space*

$$\begin{aligned} &H^1(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T}) + L^1(\mathbb{T}) \otimes H^1(\mathbb{T}^{(1,1)}) \\ &:= \{f + g : f \in H^1(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T}) \text{ and } g \in L^1(\mathbb{T}) \otimes H^1(\mathbb{T}^{(1,1)})\}. \end{aligned}$$

Proof. The space

$$H^1(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T}) = \{f \in L^1(\mathbb{T}^3) : H_1 f, H_2 f, H_1 H_2 f \in L^1(\mathbb{T}^3)\}$$

equipped with the norm $\|f\| = \|f\|_1 + \|H_1 f\|_1 + \|H_2 f\|_1 + \|H_1 H_2 f\|_1$ is a Banach space. Let $W^1 = L^1(\mathbb{T}^3) \times L^1(\mathbb{T}^3) \times L^1(\mathbb{T}^3) \times L^1(\mathbb{T}^3)$ equipped with the norm

$$\|(f_1, f_2, f_3, f_4)\|_{W_1} = \|f_1\|_1 + \|f_2\|_1 + \|f_3\|_1 + \|f_4\|_1.$$

Then we see that $H^1(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T})$ is isomorphically isometric to the closed subspace

$$V = \{(f, H_1(f), H_2(f), H_1 H_2(f)) : f \in H^1(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T})\}$$

of W^1. Now, the dual of W^1 is equal to $W^\infty = L^\infty(\mathbb{T}^3) \times L^\infty(\mathbb{T}^3) \times L^\infty(\mathbb{T}^3) \times L^\infty(\mathbb{T}^3)$ equipped with the norm $\|(g_1, g_2, g_3, g_4)\|_\infty = \max\{\|g_i\|_\infty : 1 \le i \le 4\}$ so the dual space of V is equal to the quotient of W^∞ by the annihilator U of the subspace V in W^∞. But, using the fact that the Hilbert transforms are self-adjoint up to a sign change, we see that

$$U = \{(g_1, g_2, g_3, g_4) : g_1 + H_1 g_2 + H_2 g_3 + H_1 H_2 g_4 = 0\}$$

and so:

$$V^* \cong W^\infty/U \cong \mathrm{Im}(\theta)$$

where

$$\theta(g_1, g_2, g_3, g_4) = g_1 + H_1 g_2 + H_2 g_3 + H_1 H_2 g_4$$

since $U = \ker(\theta)$. But

$$\mathrm{Im}(\theta) = L^\infty(\mathbb{T}^3) + H_1(L^\infty(\mathbb{T}^3)) + H_2(L^\infty(\mathbb{T}^3)) + H_1(H_2(L^\infty(\mathbb{T}^3)))$$

is equal to the functions that are uniformly in product BMO in variables 1 and 2.

Using the same reasoning we see that the dual of $L^1(\mathbb{T}) \otimes H^1(\mathbb{T}^{(1,1)})$ is equal to $L^\infty(\mathbb{T}^3) + H_2(L^\infty(\mathbb{T}^3)) + H_3(L^\infty(\mathbb{T}^3)) + H_2 H_3(L^\infty(\mathbb{T}^3))$, which is equal to the space of functions that are uniformly in product BMO in variable 2 and 3.

Now, we consider the "L^1 sum" of the spaces $H^1(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T})$ and $L^1(\mathbb{T}) \otimes H^1(\mathbb{T}^{(1,1)})$; that is

$$M_{(13)(2)} = \{(f, g) : f \in H^1(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T}); g \in L^1(\mathbb{T}) \otimes H^1(\mathbb{T}^{(1,1)})\}$$

equipped with the norm

$$\|(f, g)\| = \|f\|_{H^1(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T})} + \|g\|_{L^1(\mathbb{T}) \otimes H^1(\mathbb{T}^{(1,1)})}.$$

We see that, if $\phi : M_{(13)(2)} \to L^1((\mathbb{T}^3)$ is defined by $\phi(f, g) = f + g$, then the image of ϕ is isometrically isomorphic to the quotient of $M_{(13)(2)}$ by the space

$$\begin{aligned} N &= \{(f, g) \in M_{(13)(2)} : f + g = 0\} \\ &= \{(f, -f) : f \in H^1(\mathbb{T}^{(1,1)}) \otimes L^1(\mathbb{T}) \cap L^1(\mathbb{T}) \otimes H^1(\mathbb{T}^{(1,1)})\}. \end{aligned}$$

Now, recall that the dual of the quotient M/N is equal to the annihilator of N. It is easy to see that the annihilator of N is equal to the set of ordered pairs (ϕ, ϕ) with ϕ in the intersection of the duals of the two spaces. Thus the dual of the image of θ is equal to

$\mathrm{BMO}_{(13)(2)}$. The norm of an element in the predual is equal to its norm as an element of the double dual which is easily computed. □

Following this example, the reader may easily find the correct formulation for the pre-dual of other little product BMO spaces as well as those in several variables, replacing the Hilbert transform by all choices of Riesz transforms. For instance, one can prove that the pre-dual of the space $\mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})$ is equal to $H^1(\mathbb{R}^{(d_1,d_2)}) \otimes L^1(\mathbb{R}^{d_3}) + L^1(\mathbb{R}^{d_1}) \otimes H^1(\mathbb{R}^{(d_2,d_3)})$.

It is well known, that theorems of this form have an equivalent formulation in the language of weak factorization of Hardy spaces. We treat the model case $\mathbb{R}^{\vec{d}} = \mathbb{R}^{(d_1,d_2,d_3)}$ and $\mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})$ only for the sake of simplicity. The other statements are an obvious generalization. For the corresponding collections of Riesz transforms $\mathcal{R}_{k,j_k}$ and $b \in \mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})$, $1 \leq s \leq 3$, by unwinding the commutator one can define the operator $\Pi_{\vec{j}}$ such that

$$\langle [[b, R_{1,j_1} R_{3,j_3}], R_{2,j_2}] f, g \rangle_{L^2} = \langle b, \Pi_{\vec{j}}(f,g) \rangle_{L^2}.$$

Consider the Banach space $L^2 * L^2$ of all functions in $L^1(\mathbb{R}^{\vec{d}})$ of the form

$$f = \sum_{\vec{j}} \sum_i \Pi_{\vec{j}}(\phi_i^{\vec{j}}, \psi_i^{\vec{j}})$$

normed by

$$\|f\|_{L^2 * L^2} = \inf\{\sum_{\vec{j}} \sum_i \|\phi_i^{\vec{j}}\|_2 \|\psi_i^{\vec{j}}\|_2\}$$

with the infimum running over all possible decompositions of f. Applying a duality argument and the two-sided estimate in Corollary 3.11 of Chapter 3 we are going to prove the following weak factorization theorem.

Theorem 6.4. $H^1(\mathbb{R}^{(d_1,d_2)}) \otimes L^1(\mathbb{R}^{d_3}) + L^1(\mathbb{R}^{d_1}) \otimes H^1(\mathbb{R}^{(d_2,d_3)})$ *coincides with the space* $L^2 * L^2$. *In other words, for any* $f \in H^1(\mathbb{R}^{(d_1,d_2)}) \otimes L^1(\mathbb{R}^{d_3}) + L^1(\mathbb{R}^{d_1}) \otimes H^1(\mathbb{R}^{(d_2,d_3)})$ *there* *exist sequences* $\phi_i^{\vec{j}}, \psi_i^{\vec{j}} \in L^2$ *such that* $f = \sum_{\vec{j}} \sum_i \Pi_{\vec{j}}(\phi_i^{\vec{j}}, \psi_i^{\vec{j}})$ *and* $\|f\| \sim \sum_{\vec{j}} \sum_i \|\phi_i^{\vec{j}}\|_2 \|\psi_i^{\vec{j}}\|_2$.

Proof. Let's first show that $L^2 * L^2$ is a subspace of $H^1(\mathbb{R}^{(d_1,d_2)}) \otimes L^1(\mathbb{R}^{d_3}) + L^1(\mathbb{R}^{d_1}) \otimes H^1(\mathbb{R}^{(d_2,d_3)})$. Recalling the remark after Theorem 6.3, this is the same as to show $\forall f \in L^2 * L^2$, f is a bounded linear functional on $\mathrm{BMO}_{(13)(2)}(\mathbb{R}^{\vec{d}})$. This follows from the upper bound on the commutators since

$$\Big\langle b, \sum_{\vec{j}} \sum_i \Pi_{\vec{j}}(\phi_i^{\vec{j}}, \psi_i^{\vec{j}}) \Big\rangle = \sum_{\vec{j}} \sum_i \langle [[b, R_{1,j_1} R_{3,j_3}], R_{2,j_2}] \phi_i^{\vec{j}}, \psi_i^{\vec{j}} \rangle.$$

Now we are going to show

$$\sup_{f \in L^2 * L^2} \Big\{ \Big| \int f b \Big| : \|f\|_{L^2 * L^2} \le 1 \Big\} \sim \|b\|_{\mathrm{BMO}_{(13)(2)}}$$

which gives the equivalence of the $H^1(\mathbb{R}^{(d_1,d_2)}) \otimes L^1(\mathbb{R}^{d_3}) + L^1(\mathbb{R}^{d_1}) \otimes H^1(\mathbb{R}^{(d_2,d_3)})$ norm and the $L^2 * L^2$ norm, thus showing that the two spaces are the same.

To see this, note that the direction $\lesssim$ is trivial, and the direction $\gtrsim$ is implied by the lower bound of commutators. For any $b \in \mathrm{BMO}_{(13)(2)}$, there exists $\vec{j}$ such that $\|b\|_{\mathrm{BMO}_{(13)(2)}} \lesssim \|[[b, R_{1,j_1} R_{3,j_3}], R_{2,j_2}]\|$. Hence, there exist $\phi, \psi \in L^2$ with norm 1 such that

$$\|b\|_{\mathrm{BMO}_{(13)(2)}} \lesssim |\langle [[b, R_{1,j_1} R_{3,j_3}], R_{2,j_2}] \phi, \psi \rangle| = |\langle b, \Pi_{\vec{j}}(\phi, \psi) \rangle| \le \mathrm{LHS},$$

which completes the proof. $\square$

Chapter Seven

Multi-parameter Representation Theorem

Steering away from commutators, we would like to prove the multi-parameter version of the representation theorem, which has been shown to be very powerful in the study of upper estimates for Journé commutators and in turn the lower estimates for Riesz commutators. Our work is along the lines of existing representation theorems in the one-parameter ([Hyt11]) and bi-parameter setting ([Mar12]), extending the result to the case of arbitrarily many parameters.

Compared with the bi-parameter theory, the theory of multi-parameter singular integral operators generally involves an additional layer of difficulty. Usually for bi-parameter problems on $\mathbb{R} \times \mathbb{R}$, in the inductive step, by slicing away one dimension one will reduce to the one-parameter setting. This is not the case for t-parameter problems when $t \geq 3$. Furthermore, there are results that are true in the bi-parameter setting but fail to hold in the multi-parameter setting, for example the results regarding rectangle atoms discussed by Fefferman in [Fef81]. (Also see Journé [Jou86].) Naturally, it has been asked by several experts in the field whether one can establish a representation theorem in multi-parameters, which becomes the main motivation and the central problem this chapter will be dealing with.

The first difficulty one encounters is how to generalize Martikainen's class of singular integral operators to more than two parameters, establishing a group of appropriate mixed type conditions that characterizes operators suitable to work with. Recall that in the classical $T(1)$ theorem, the hypotheses involve assumptions on the size and smoothness of the kernel, a weak boundedness property (WBP), and BMO conditions. It is then natural to formulate nine different so-called mixed type conditions (such as kernel/kernel, BMO/WBP and so on) for bi-parameter operators, which is, morally speaking, what Martikainen did in [Mar12]. However, there is no obvious way to generalize to multi-parameters formulations of such mixed type conditions. In fact, although Martikainen has done a brilliant job in [Mar12] to introduce the so-called full kernel and partial kernel assumptions on the operator, his assumptions do not admit easy generalizations to the multi-parameter setting since the following fact has been used in the formulation: once a parameter is taken away, what's left becomes a one-parameter object.

The second difficulty, of course, is the proof of the representation theorem itself. Once the proper assumptions on the operators are formulated, the proof in the multi-parameter setting requires no new techniques. However, verifying that the theorem holds requires a delicate analysis of the symmetries of the operator and the particularly nice formulation of the conditions.

The main contributions of our result are the following. First, mixed type conditions for multi-parameter operators are formulated along the lines of [PV13] and [Mar12], establishing the appropriate class of multi-parameter singular integral operators. Second, we prove a representation theorem in arbitrarily many parameters, which yields a new multi-parameter $T(1)$ theorem. Finally, as an application of our multi-parameter $T(1)$ theorem, we show that our class of multi-parameter singular integrals is equivalent to the class studied by Journé in [Jou85]. This generalizes a recent result of Grau de la Herran [GdlH15] to arbitrarily many parameters. This shows that Journé's class of operators, originally formulated in vector-valued language, can be characterized by conditions that are more intrinsic and easier to verify. In particular, this formulation makes it easier to approach the class of paraproduct free multi-parameter Journé operators, which has appeared in our upper estimates for commutators in Chapter 4.

This chapter is organized as follows. In Sections 7.1 and 7.2, we define a class of multi-parameter singular integral operators characterized by new mixed type conditions. The statement of the multi-parameter representation theorem and its proof are presented in Sections 7.3 and 7.4. We then discuss the equivalence between our class and Journé's class of operators in Section 7.5, followed by a discussion of the necessity of some of the mixed conditions at the end.

7.1 A class of t-parameter singular integrals

In $\mathbb{R}^{\vec{d}} := \mathbb{R}^{d_1} \times \cdots \times \mathbb{R}^{d_t}$, where $t \in \mathbb{N}$ denotes the number of parameters, let T be a linear operator continuously mapping $C_0^\infty(\mathbb{R}^{d_1}) \times \cdots \times C_0^\infty(\mathbb{R}^{d_t})$ to its dual. $\forall S \subset \{1, 2, \ldots, t\}$, define its partial adjoint T_S by exchanging the i^{th} variable, $\forall i \in S$, i.e.

$$\langle T(f_S \otimes f_{S^c}), g_S \otimes g_{S^c} \rangle = \langle T_S(g_S \otimes f_{S^c}), f_S \otimes g_{S^c} \rangle,$$

where f_S, g_S are functions of the i^{th} variables for $i \in S$, and f_{S^c}, g_{S^c} are functions of the i^{th} variables for $i \notin S$.

We say T is in our class of *t-parameter singular integral operators* if for any S, T_S satisfies the following *full kernel* and *partial kernel assumptions*.

7.1.1 Full kernel assumptions

For any $f = \bigotimes_{i=1}^t f_i, g = \bigotimes_{i=1}^t g_i \in C_0^\infty(\mathbb{R}^{d_1}) \times \cdots \times C_0^\infty(\mathbb{R}^{d_t})$ such that $\forall i \in \{1, 2, \ldots, t\}$, $\operatorname{spt} f_i \cap \operatorname{spt} g_i = \emptyset$, there holds

$$\langle T_S f, g \rangle = \int_{\mathbb{R}^{\vec{d}}} \int_{\mathbb{R}^{\vec{d}}} K_S(x, y) f(y) g(x)\, dx dy,$$

where the kernel $K_S(x, y)$ satisfies the following mixed size-Hölder conditions:

For any subset $W \subset \{1, 2, \ldots, t\}$, when $|x_i - x_i'| \le |x_i - y_i|/2, \forall i \in W$, there holds

$$\left| \sum_{\Lambda \subset W} (-1)^{|\Lambda|} K_S^\Lambda(x, x'; y) \right| \lesssim \left(\prod_{i \in W} \frac{|x_i - x_i'|^\delta}{|x_i - y_i|^{d_i + \delta}} \right) \left(\prod_{i \in \{1,2,\ldots,t\} \setminus W} \frac{1}{|x_i - y_i|^{d_i}} \right),$$

where $0 < \delta < 1$ is a fixed constant, and $K_S^\Lambda(x, x'; y)$ is defined as K_S evaluated at x_i for $i \notin \Lambda$, at x_i' for $i \in \Lambda$. Note that when $W = \emptyset$, this is the pure size condition, while when $W = \{1, 2, \ldots, t\}$, this becomes the Hölder conditions we are familiar with in the one-parameter and bi-parameter settings. The reader can compare this condition to the standard kernel conditions of Calderón-Zygmund operators introduced in Chapter 2 to see that when $t=1$ they are exactly the same.

7.1.2 Partial kernel assumptions

Let V be any nonempty proper subset of $\{1,2,\dots,n\}$, and $f=f_V\otimes f_{V^c}, g=g_V\otimes g_{V^c}\in C_0^\infty(\mathbb{R}^{d_1})\times\cdots\times C_0^\infty(\mathbb{R}^{d_t})$, where $f_V=\bigotimes_{i\in V}f_i$ and similarly for others. Suppose for any variable $i\in V$, $\operatorname{spt} f_i\cap\operatorname{spt} g_i=\emptyset$, there holds

$$\langle T_S f,g\rangle=\int_{\otimes_{i\in V}\mathbb{R}^{d_i}}\int_{\otimes_{i\in V}\mathbb{R}^{d_i}}K^V_{S,f_{V^c},g_{V^c}}(x,y)f_V(y)g_V(x)\,dxdy,$$

where the kernel $K^V_{S,f_{V^c},g_{V^c}}$ satisfies the following mixed size-Hölder conditions:

For any subset $W\subset V$, when $|x_i-x_i'|\le|x_i-y_i|/2,\forall i\in W$, there holds

$$\begin{aligned}&\left|\sum_{\Lambda\subset W}(-1)^{|\Lambda|}K^{V,\Lambda}_{S,f_{V^c},g_{V^c}}(x,x';y)\right|\\&\qquad\le C_S^V(f_{V^c},g_{V^c})\left(\prod_{i\in W}\frac{|x_i-x_i'|^\delta}{|x_i-y_i|^{d_i+\delta}}\right)\left(\prod_{i\in V\setminus W}\frac{1}{|x_i-y_i|^{d_i}}\right),\end{aligned}$$

where $K^{V,\Lambda}_{S,f_{V^c},g_{V^c}}(x,x';y)$ is defined as $K^V_{S,f_{V^c},g_{V^c}}$ evaluated at x_i for $i\notin\Lambda$, at x_i' for $i\in\Lambda$. Moreover, we require that constant $C_S^V(f_{V^c},g_{V^c})$ satisfies the following mixed *WBP/BMO* conditions:

For any subset $W\subset V^c$, any cubes $I_i\subset\mathbb{R}^{d_i}$, $i\in W$, there holds

$$\|C_S^V((\otimes_{i\in W}\chi_{I_i})\otimes\left(\otimes_{i\in V^c\setminus W}1\right),(\otimes_{i\in W}\chi_{I_i})\otimes\cdot)\|_{\mathrm{BMO}_{\mathrm{prod}}(\otimes_{i\in V^c\setminus W}\mathbb{R}^{d_i})}\lesssim\prod_{i\in W}|I_i|.$$

Since product BMO is the intersection of all dyadic product BMO, one can express the WBP/BMO condition above via the following Carleson packing condition: For any product dyadic grid $\mathcal{D} = \otimes_{i\in V^c\setminus W}\mathcal{D}_i$,

$$\sup \frac{1}{|\Omega|} \sum_{\substack{R\subset\Omega, R\in\mathcal{D}\\ R=\otimes_{j\in V^c\setminus W} J_j}} |C_S^V((\otimes_{i\in W}\chi_{I_i})\otimes\left(\otimes_{i\in V^c\setminus W}1\right),(\otimes_{i\in W}\chi_{I_i})\otimes\left(\otimes_{j\in V^c\setminus W}h_{J_j}\right))|^2$$
$$\lesssim \prod_{i\in W}|I_i|^2,$$

where the supremum is taken over all the measurable open sets Ω in $\otimes_{i\in V^c\setminus W}\mathbb{R}^{d_i}$ with finite measure.

The expression above is always well defined as the functions involved are all tensor products. In the case when one can naturally extend the definition of operator T to act on more general multivariate functions, one can also rephrase the WBP/BMO condition by duality as the following: For any function $h \in H^1_{\text{prod}}(\otimes_{i\in V^c\setminus W}\mathbb{R}^{d_i})$,

$$|C_S^V((\otimes_{i\in W}\chi_{I_i})\otimes\left(\otimes_{i\in V^c\setminus W}1\right),(\otimes_{i\in W}\chi_{I_i})\otimes h)| \lesssim \left(\prod_{i\in V^c\setminus W}|I_i|\right)\|h\|_{H^1_{\text{prod}}}.$$

This completes our definition of the *t-parameter singular integral operators*. And one can similarly define an *t-parameter Calderón-Zygmund* operator if there are some additional boundedness assumption posed on the operator.

Definition 7.1. T is called an t-parameter Calderón-Zygmund (CZ) operator if it is an t-parameter singular integral operator defined as above and $T_S : L^2 \to L^2$, for any $S \subset \{1, 2, \ldots, t\}$.

In order to derive our main multi-parameter representation theorem for such operators later in this chapter, as a preparation, we will need the definition of the so-called *mixed BMO/WBP assumptions*, which we describe below. Note that these are not characterizing conditions of our class of singular integrals.

7.1.3 BMO/WBP assumptions

We say that an operator T_S satisfies the mixed BMO/WBP conditions if for any subset $W \subset \{1, 2, \dots, t\}$, any cubes $I_i \subset \mathbb{R}^{d_i}$, $i \in W$, there holds

$$\|\langle T_S((\otimes_{i\in W}\chi_{I_i}) \otimes (\otimes_{i\in W^c}1)), (\otimes_{i\in W}\chi_{I_i}) \otimes \cdot\rangle\|_{\mathrm{BMO}_{\mathrm{prod}}(\otimes_{i\in W^c}\mathbb{R}^{d_i})} \lesssim \prod_{i\in W}|I_i|.$$

This is the pure BMO condition when $W = \emptyset$, and the pure dyadic weak boundedness property when $W = \{1, 2, \dots, t\}$. Again, one can interpret the product BMO norm in several different ways, as we described above.

To end this section, we would like to emphasize that the class of singular integral operators defined above is indeed a generalization of the most natural classes of one-parameter and bi-parameter singular integral operators that we have introduced in Chapter 2. When $t = 1$, it coincides with the class of singular integral operators associated with standard kernels. When $t = 2$, it is the same as the class of bi-parameter operators defined by Martikainen in [Mar12] (modulo that some of the conditions in partial kernel assumptions are formulated slightly differently), and is known to be equivalent to the classes of Journé [Jou85] and Pott-Villarroya [PV13], according to Grau de la Herran [GdlH15].

Furthermore, it is not hard to examine that our class of t-parameter singular integrals includes operators of tensor product type as a special case. Let's take a look at the case $t = 3$ as an example. Given Calderón-Zygmund operators T_i defined on $\mathbb{R}^{d_i}$, $i = 1, 2, 3$, it is easy to see that the operator $T_1 \otimes T_2 \otimes T_3$ satisfies the full kernel assumptions. To check one of the partial kernel assumptions, for any test functions with $\mathrm{spt}\, f_1 \cap \mathrm{spt}\, g_1 = \emptyset$, one can define a partial kernel

$$K^{\{1\}}_{f_2\otimes f_3, g_2\otimes g_3}(x_1, y_1) = K_1(x_1, y_1)\langle T_2 \otimes T_3(f_2 \otimes f_3), g_2 \otimes g_3\rangle,$$

where $K_1(x_1, y_1)$ is the kernel of T_1. Observe that $T_2 \otimes T_3$ is a Journé type bi-parameter

CZ operator studied in [Jou85], hence is bounded on L^2 and maps $1\otimes 1$ into product BMO, which thus implies the required WBP/BMO conditions for constants $C^{\{1\}}(f_2\otimes f_3, g_2\otimes g_3)$. We will give a more thorough discussion of the Journé type multi-parameter singular integral operators in Section 7.5.

7.2 Well-definedness of the BMO assumptions

Among the various conditions satisfied by an t-parameter operator T, many of them are establishing certain bounds on pairings involving T acting on function 1 in some of the variables. It is thus necessary to articulate how these objects are defined. For simplicity, let's look at the case $t = 3$.

Recall that in the partial kernel assumptions, if $f = f_1 \otimes f_2 \otimes f_3, g = g_1 \otimes g_2 \otimes g_3$, and $\operatorname{spt} f_1 \cap \operatorname{spt} g_1 = \operatorname{spt} f_2 \cap \operatorname{spt} g_2 = \emptyset$ (i.e. $V = \{1, 2\}$), one wants to show that $C_S^V(1, \cdot) \in \mathrm{BMO}(\mathbb{R}^{d_3})$, which according to [PW08] is the same as showing that for any dyadic system $\mathcal{D}$ of $\mathbb{R}^{d_3}$, it is in dyadic $\mathrm{BMO}_{\mathcal{D}}(\mathbb{R}^{d_3})$.

Hence, it suffices to give a meaning to $C_S^V(1, h_{I_3})$ for any Haar function in the third variable, i.e. to define the pairing $\langle T_S(f_1 \otimes f_2 \otimes 1), g_1 \otimes g_2 \otimes h_{I_3}\rangle$. This can be done by dividing $1 = \chi_{3I_3} + \chi_{(3I_3)^c}$, where the first term makes sense since T is continuous (more precisely, one needs kernel representation, WBP and dominated convergence to justify the well-definedness of the bilinear form of non-smooth functions), while the second term can be defined using the full kernel representation whose convergence is guaranteed by Hölder conditions.

Second, still in the partial kernel assumptions, if one only has $\operatorname{spt} f_3 \cap \operatorname{spt} g_3 = \emptyset$ (i.e. $V = \{3\}$), the well-definedness of constant $C_S^V(\chi_{I_1} \otimes 1, \chi_{I_1} \otimes \cdot)$ is similar as in the case above, so we only look at the meaning of $C_S^V(1 \otimes 1, \cdot)$ as a function in $\mathrm{BMO}_{\mathcal{D}(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})}$. To define $\langle T_S(1 \otimes 1 \otimes f_3), h_{I_1} \otimes h_{I_2} \otimes g_3\rangle$, clearly, one can divide $1 \otimes 1 = \chi_{3I_1} \otimes \chi_{3I_2} + \chi_{3I_1} \otimes$

$\chi_{(3I_2)^c} + \chi_{(3I_1)^c} \otimes \chi_{3I_2} + \chi_{(3I_1)^c} \otimes \chi_{(3I_2)^c}$, where the first and last term are easy to deal with. While for the mixed terms, say, the third one, if $\chi_{(3I_1)^c}$ is replaced by a C_0^∞ function, then the pairing is apparently well defined through the partial kernel representation. Now even though $\chi_{(3I_1)^c}$ is only bounded, we can still define the pairing as

$$\int K^{\{1,3\}}_{S,\chi_{3I_2},h_{I_2}}(x_1, y_1, x_3, y_3)\chi_{(3I_1)^c}(y_1) f_3(y_3) h_{I_1}(x_1) g_3(x_3)\, dx_1 dx_3 dy_1 dy_3,$$

where the integral converges since one can change the kernel to

$$K^{\{1,3\}}_{S,\chi_{3I_2},h_{I_2}}(x_1, y_1, x_3, y_3) - K^{\{1,3\}}_{S,\chi_{3I_2},h_{I_2}}(x_1, y_1, c_{I_3}, y_3)$$

and use the mixed Hölder-size conditions.

Finally, in the BMO/WBP assumptions, to give a meaning to

$$\langle T_S((\otimes_{i\in W}\chi_{I_i}) \otimes (\otimes_{i\in W^c} 1)), (\otimes_{i\in W}\chi_{I_i}) \otimes \cdot\rangle,$$

it is sufficient to define what it means for the function to be paired with tensors of Haar functions. This can be done by dividing $1 \otimes \cdots \otimes 1$ into several parts similarly as above, and use partial kernel representation and Hölder conditions to obtain the convergence of the corresponding integrals.

7.3 Multi-parameter dyadic shifts and the main theorem

In order to formulate the representation theorem in the multi-parameter setting, we first briefly recall the notion of shifted dyadic grids. Denote $\mathcal{D}_i^0 := \{2^{-k}([0,1]^{d_i} + m) : k \in \mathbb{Z}, m \in \mathbb{Z}^{d_i}\}$ as the standard dyadic grid in the i^{th} variable, $1 \le i \le t$. Let $\omega = (\omega_i^j)_{j\in\mathbb{Z}} \in (\{0,1\}^{d_i})^{\mathbb{Z}}$ and $I \dotplus \omega_i := I + \sum_{j:2^{-j}<\ell(I)} 2^{-j}\omega_i^j$, then

$$\mathcal{D}_i^\omega := \{I \dotplus \omega_i : I \in \mathcal{D}_i^0\}$$

is a shifted dyadic grid associated with parameter ω_i. We usually write $\mathcal{D}_i$ for short when the dependence on ω_i is not explicitly needed.

If we assume each ω_i is an independent random variable having an equal probability 2^{-d_i} of taking any of the 2^{d_i} values in $\{0,1\}^{d_i}$, we obtain a random dyadic system $\mathcal{D}_1 \times \cdots \times \mathcal{D}_t$.

A *dyadic shift* with parameter $i_1, j_1, \ldots, i_t, j_t \in \mathbb{N}$ associated with dyadic grids $\mathcal{D}_1, \ldots, \mathcal{D}_t$ is an $L^2 \to L^2$ operator with norm ≤ 1 defined as

$$S^{i_1 j_1, \ldots, i_t j_t}_{\mathcal{D}_1 \ldots \mathcal{D}_t} f := \sum_{K_1 \in \mathcal{D}_1} \cdots \sum_{K_t \in \mathcal{D}_t} \sum_{\substack{I_1, J_1 \in \mathcal{D}_1, I_1, J_1 \subset K_1 \\ \ell(I_1) = 2^{-i_1} \ell(K_1) \\ \ell(J_1) = 2^{-j_1} \ell(K_1)}} \cdots \sum_{\substack{I_t, J_t \in \mathcal{D}_t, I_t, J_t \subset K_t \\ \ell(I_t) = 2^{-i_t} \ell(K_t) \\ \ell(J_t) = 2^{-j_t} \ell(K_t)}} \left(a_{I_1 J_1 K_1 \ldots I_t J_t K_t} \langle f, h_{I_1} \otimes \cdots \otimes h_{I_t} \rangle h_{J_1} \otimes \cdots \otimes h_{J_t} \right),$$

which is written in short as

$$=: \sum_{K_1 \in \mathcal{D}_1} \cdots \sum_{K_t \in \mathcal{D}_t} \sum^{(i_1, j_1)}_{\substack{I_1, J_1 \in \mathcal{D}_1 \\ I_1, J_1 \subset K_1}} \sum^{(i_t, j_t)}_{\substack{I_t, J_t \in \mathcal{D}_t \\ I_t, J_t \subset K_t}} a_{I_1 J_1 K_1 \ldots I_t J_t K_t} \langle f, h_{I_1} \otimes \cdots \otimes h_{I_t} \rangle h_{J_1} \otimes \cdots \otimes h_{J_t},$$

where the coefficients satisfy

$$|a_{I_1 J_1 K_1 \ldots I_t J_t K_t}| \leq \frac{\sqrt{|I_1||J_1| \cdots |I_t||J_t|}}{|K_1| \cdots |K_t|},$$

and h_{I_s} is a Haar function on I_s, similarly for h_{J_s}. Recall that for any dyadic cube $I \subset \mathbb{R}^{d_i}$, there are 2^{d_i} associated Haar functions h_I, with one of them being the non-cancellative function $|I|^{-1/2}\chi_I$ and all the other ones being cancellative. We allow any choices of Haar functions, non-cancellative or cancellative, in the definition of dyadic shifts. Same as the one-parameter and bi-parameter cases, we will call the dyadic shift *cancellative* if all the Haar functions that appear in the sum are cancellative, *non-cancellative* otherwise. It is not hard to show that when the shift is cancellative, the L^2 boundedness requirement in fact follows from the boundedness of the coefficients directly. Furthermore, it is also worth observing that t-parameter dyadic paraproducts with product BMO symbol are particular examples of non-cancellative dyadic shifts.

Now we are ready to state the representation theorem. Recall that T is said to be an t-parameter singular integral operator in our class if it satisfies both the full kernel and partial kernel assumptions defined in Subsections 7.1.1 and 7.1.2.

Theorem 7.2. *For a t-parameter singular integral operator T, which satisfies in addition the BMO/WBP assumptions (see Section 7.1.3) by any of its partial adjoint T_S, there holds for some t-parameter shifts $S^{i_1j_1\dots i_tj_t}_{\mathcal{D}_1\dots\mathcal{D}_t}$ that*

$$\langle Tf,g\rangle = C_T\mathbb{E}_{\omega_1}\mathbb{E}_{\omega_2}\cdots\mathbb{E}_{\omega_t}\sum_{(i_1,j_1)\in\mathbb{N}^2}\cdots\sum_{(i_t,j_t)\in\mathbb{N}^2}\left(\prod_{s=1}^t 2^{-\max(i_s,j_s)\delta/2}\right)\langle S^{i_1j_1\dots i_tj_t}_{\mathcal{D}_1\dots\mathcal{D}_t}f,g\rangle,$$

where non-cancellative shifts may only appear when there is some s such that $(i_s,j_s)=(0,0)$.

f and g above are arbitrary functions taken from some particularly nice dense subset of $L^2(\mathbb{R}^{\vec{d}})$, for example, the finite linear combinations of tensor products of univariate functions in $C_0^\infty(\mathbb{R}^{d_i})$. Hence, according to the uniform boundedness of dyadic shifts, an immediate result implied by the representation theorem is the following.

Corollary 7.3. *A t-parameter singular integral operator T whose partial adjoints all satisfy the BMO/WBP assumptions is bounded on $L^2(\mathbb{R}^{\vec{d}})$.*

Remark 7.4. In the one-parameter and bi-parameter versions of the representation theorem, see [Hyt11], [Mar12], one needs the additional a priori assumption that T is bounded on L^2 in order to justify the convergence of some infinite series in the proof. This makes the $T(1)$ type corollary only a quantitative result. However, very recently, it is suggested by T. Hytönen that one can prove the representation theorem without assuming any a priori bound on T, by first proving a "weak representation" depending on functions f,g, which then implies that T is bounded on L^2. Hence, the corollary obtained above is indeed a $T(1)$ theorem of full strength. To the author's knowledge, the only currently known $T(1)$ type theorem in more than two parameters is proved by Journé in [Jou85] by induction, using a vector-valued argument. The advantage of our $T(1)$ theorem is that the mixed type conditions are expressed in a more transparent way and much easier to verify. In fact, we will see an application of our $T(1)$ theorem later in this chapter, when we establish the relationship between Journé's and our class of multi-parameter singular integral operators.

Another useful observation is that due to the symmetry of the assumptions on the t-parameter singular integral operators, one can conclude that if T is a t-parameter singular integral operator whose partial adjoints all satisfy the BMO/WBP assumptions, then all of its partial adjoints are bounded on L^2. Hence T is a t-parameter CZ operator defined in Section 7.1. In fact, the other direction also holds true, i.e. T being a t-parameter CZ operator implies the BMO/WBP assumptions. We leave the discussion of this point to the end of the chapter.

7.4 Proof of the main theorem

We will prove Theorem 7.2 in the case $t = 3$ as an example, which is sufficient in showing the new difficulties that arise in the multi-parameter setting and in explaining our strategy.

7.4.1 Outline of the proof

Roughly speaking, the proof of representation theorems usually starts with developing some averaging formula, which represents the bilinear form $\langle Tf, g\rangle$ as an expectation of randomized Haar expansion where only "good" dyadic cubes are involved. We will establish a tri-parameter version of this formula in Subsection 7.4.2, where the notions of good and bad cubes will also be recalled.

Next, one targets to decompose the averaging formula that represents $\langle Tf, g\rangle$ into finitely many parts, each of which will be shown to be a convergent sum of bilinear forms

$$\langle S^{i_1 j_1 i_2 j_2 i_3 j_3} f, g\rangle$$

for some dyadic shift $S^{i_1j_1i_2j_2i_3j_3}$. The proof will thus be complete. This step is the key part of the argument. Since we have three free parameters to deal with, there will be a large number of different cases to analyze. More precisely, for each parameter, at some point one splits the summation into four parts: Separated, Inside, Near and Equal, which yields at least 4^3 mixed parts for us to study. Fortunately, many of them can be estimated via kernel assumptions and weak boundedness properties, similarly as the one-parameter and bi-parameter cases treated in [Hyt11] and [Mar12], except for the cases where more complicated tri-parameter paraproducts have to be involved. One typical example of such cases will be referred to as "Inside/Inside/Inside", which we will study in Subsection 7.4.3 with full details. An intrinsic difference between the bi-parameter case and our arbitrarily many parameter case is that, one needs to deal with some *multi-parameter* paraproduct mixed with dyadic shift in our case, which doesn't exist in the bi-parameter setting. This is also one of the reasons why we have to formulate our BMO assumptions on the operators in a global way, in contrast to the local type assumptions in Martikainen's bi-parameter formulation.

7.4.2 Randomizing process and averaging formula

We remind the readers that a cube $I_i \in \mathcal{D}_i$ is called *bad* if there is another $\tilde{I}_i \in \mathcal{D}_i$ such that $\ell(\tilde{I}_i) \geq 2^r \ell(I_i)$ and $d(I_i, \partial \tilde{I}_i) \leq 2\ell(I_i)^{\gamma_i}\ell(\tilde{I}_i)^{1-\gamma_i}$, where r is a fixed large number, $\gamma_i := \delta/(2d_i + 2\delta)$, and δ is the constant that appears in the kernel assumptions of the operator. Naturally, a cube is called *good* if it is not bad. And $\pi^i_{\text{good}} := \mathbb{P}_{\omega_i}(I_i \dot{+} \omega_i \text{ is good})$ is a parameter depending only on δ, d_i and r. One always fixes an r large enough so that $\pi^i_{\text{good}} > 0$ for any $1 \leq i \leq 3$. And in the following, we use $\text{sm}(I_i, J_i)$ to denote the smaller cube of I_i, J_i. Through a similar process of randomization independently in each variable, as described in [Hyt11] and [Mar12], by randomizing over products of shifted dyadic grids, it is not hard to obtain the following tri-parameter version of the key averaging formula:

$$\begin{aligned}&\langle Tf,g\rangle\\=&C\mathbb{E}\sum_{I_1,J_1\in\mathcal{D}_1}\sum_{I_2,J_2\in\mathcal{D}_2}\sum_{I_3,J_3\in\mathcal{D}_3}\chi_{\text{good}}(\text{sm}(I_1,J_1))\chi_{\text{good}}(\text{sm}(I_2,J_2))\chi_{\text{good}}(\text{sm}(I_3,J_3))\\&\langle T(h_{I_1}\otimes h_{I_2}\otimes h_{I_3}),h_{J_1}\otimes h_{J_2}\otimes h_{J_3}\rangle\langle f,h_{I_1}\otimes h_{I_2}\otimes h_{I_3}\rangle\langle g,h_{J_1}\otimes h_{J_2}\otimes h_{J_3}\rangle,\end{aligned}$$

where $\mathbb{E}=\mathbb{E}_{\omega_1}\mathbb{E}_{\omega_2}\mathbb{E}_{\omega_3}$ and $C=1/(\pi^1_{\text{good}}\pi^2_{\text{good}}\pi^3_{\text{good}})$.

In order to demonstrate the desired representation, we will then split the sums on the right hand side of the averaging formula into several pieces depending on the relative sizes of I_i, J_i, $i=1,2,3$, and whether the smaller cubes are far away, strictly inside, exactly equal, or close to the larger cubes (i.e. Separated, Inside, Equal or Near). Specifically, for each variable i, we split the sum

$$\sum_{I_i}\sum_{J_i}=\sum_{\ell(I_i)\le\ell(J_i)}+\sum_{\ell(I_i)>\ell(J_i)}=:I+II.$$

Then decompose

$$\begin{aligned}I=&\sum_{\substack{\ell(I_i)\le\ell(J_i)\\ d(I_i,J_i)>\ell(I_i)^{\gamma_i}\ell(J_i)^{1-\gamma_i}}}+\sum_{I_i\subsetneq J_i}+\sum_{I_i=J_i}+\sum_{\substack{\ell(I_i)\le\ell(J_i)\\ d(I_i,J_i)\le\ell(I_i)^{\gamma_i}\ell(J_i)^{1-\gamma_i}\\ I_i\cap J_i=\emptyset}}\\=:&\ \text{Separated}+\text{Inside}+\text{Equal}+\text{Near}\end{aligned}$$

and similarly for II. The strategy is to prove that each of the terms above can be represented as sums of bilinear forms associated to dyadic shifts.

Many of the cases can be discussed following the same techniques as in [Mar12], while for some mixed cases, new multi-parameter phenomena may appear and require extreme care. The good news is that the new mixed cases won't do us much harm since we have already formulated the proper assumptions on the operators at the beginning to handle them.

As one has already encountered in the bi-parameter setting in [Mar12], different types of mixed paraproducts will appear according to the relative sizes of I_i, J_i. Since the worst situations one would expect are the mixed cases, we will look at the part of the sum corresponding to $|I_1| \leq |J_1|, |I_2| \leq |J_2|, |I_3| > |J_3|$, observing that other cases are symmetric or even simpler. According to the averaging formula, it thus suffices to assume that I_1, I_2, J_3 are all good cubes.

Moreover, recall that in [Hyt11] and [Mar12], the Separated, Near, and Equal parts of the sum can basically be estimated using full kernel assumptions and WBP, while the Inside part, being the most difficult one, involves in addition all the BMO type estimates. Hence, we will study the Inside/Inside/Inside part next, where all the new multi-parameter phenomena will appear. Note that although this is only one of the many cases one needs to discuss in order to obtain a full proof of Theorem 7.2, all the main difficulties in other cases are in fact already embedded in Inside/Inside/Inside, a fact that will become more and more clear throughout the proof. We want to emphasize that the reason why we assumed from the beginning that all the assumptions hold true for any partial adjoint T_S of T is exactly because of the much desired symmetry of the mixed cases.

7.4.3 Inside/Inside/Inside

In this section, we study the case Inside/Inside/Inside, i.e. the summation over $I_1 \subsetneq J_1, I_2 \subsetneq J_2, J_3 \subsetneq I_3$. Recall that I_1, I_2, J_3 are all good cubes. One first decomposes

$$\langle T(h_{I_1} \otimes h_{I_2} \otimes h_{I_3}), h_{J_1} \otimes h_{J_2} \otimes h_{J_3} \rangle = I + II + III + IV + V + VI + VII + VIII,$$

where

$$I := \langle T(h_{I_1} \otimes h_{I_2} \otimes s_{J_3 I_3}), s_{I_1 J_1} \otimes s_{I_2 J_2} \otimes h_{J_3} \rangle,$$

$$II := \langle h_{I_3} \rangle_{J_3} \langle T(h_{I_1} \otimes h_{I_2} \otimes 1), s_{I_1 J_1} \otimes s_{I_2 J_2} \otimes h_{J_3} \rangle,$$

$$III := \langle h_{J_2} \rangle_{I_2} \langle T(h_{I_1} \otimes h_{I_2} \otimes s_{J_3 I_3}), s_{I_1 J_1} \otimes 1 \otimes h_{J_3} \rangle,$$

$$IV := \langle h_{J_2}\rangle_{I_2}\langle h_{I_3}\rangle_{J_3}\langle T(h_{I_1}\otimes h_{I_2}\otimes 1), s_{I_1J_1}\otimes 1\otimes h_{J_3}\rangle,$$

$$V := \langle h_{J_1}\rangle_{I_1}\langle T(h_{I_1}\otimes h_{I_2}\otimes s_{J_3I_3}), 1\otimes s_{I_2J_2}\otimes h_{J_3}\rangle,$$

$$VI := \langle h_{J_1}\rangle_{I_1}\langle h_{I_3}\rangle_{J_3}\langle T(h_{I_1}\otimes h_{I_2}\otimes 1), 1\otimes s_{I_2J_2}\otimes h_{J_3}\rangle,$$

$$VII := \langle h_{J_1}\rangle_{I_1}\langle h_{J_2}\rangle_{I_2}\langle T(h_{I_1}\otimes h_{I_2}\otimes s_{J_3I_3}), 1\otimes 1\otimes h_{J_3}\rangle,$$

$$VIII := \langle h_{J_1}\rangle_{I_1}\langle h_{J_2}\rangle_{I_2}\langle h_{I_3}\rangle_{J_3}\langle T(h_{I_1}\otimes h_{I_2}\otimes 1), 1\otimes 1\otimes h_{J_3}\rangle.$$

In the above, $s_{I_1J_1} := \chi_{Q_1^c}(h_{J_1} - \langle h_{J_1}\rangle_{Q_1})$, $s_{I_2J_2} := \chi_{Q_2^c}(h_{J_2} - \langle h_{J_2}\rangle_{Q_2})$, Q_1, Q_2 being the child of J_1, J_2 containing I_1, I_2, respectively, and $s_{J_3I_3} := \chi_{Q_3^c}(h_{I_3} - \langle h_{I_3}\rangle_{Q_3})$, Q_3 being the child of I_3 containing J_3. The relevant properties are $\operatorname{spt} s_{I_1J_1} \subset Q_1^c$, $\operatorname{spt} s_{I_2J_2} \subset Q_2^c$, $\operatorname{spt} s_{J_3I_3} \subset Q_3^c$, and $|s_{I_1J_1}| \le 2|J_1|^{-1/2}$, $|s_{I_2J_2}| \le 2|J_2|^{-1/2}$, $|s_{J_3I_3}| \le 2|I_3|^{-1/2}$.

Next, we show that the sum corresponding to each of the eight terms above can be realized as a sum of bilinear forms associated to dyadic shifts. The estimate of term I doesn't require any BMO conditions, while all the other terms require delicate BMO norm estimates and boundedness results of paraproducts. Specifically, we will use one-parameter paraproduct to analyze term III, V, II, bi-parameter paraproduct for term IV, VI, VII, and tri-parameter paraproduct for the last term $VIII$. The reader will easily see that when the number of parameters is more than three, analogous argument can be carried out.

Term I

As the functions in the pairing are all disjointly supported, following from the full kernel assumptions, one can argue similarly as in [Mar12] Lemma 7.1 that there holds

$$\begin{aligned}&|\langle T(h_{I_1}\otimes h_{I_2}\otimes s_{J_3I_3}), s_{I_1J_1}\otimes s_{I_2J_2}\otimes h_{J_3}\rangle|\\ &\quad\lesssim \frac{|I_1|^{1/2}}{|J_1|^{1/2}}\left(\frac{\ell(I_1)}{\ell(J_1)}\right)^{\delta/2}\frac{|I_2|^{1/2}}{|J_2|^{1/2}}\left(\frac{\ell(I_2)}{\ell(J_2)}\right)^{\delta/2}\frac{|J_3|^{1/2}}{|I_3|^{1/2}}\left(\frac{\ell(J_3)}{\ell(I_3)}\right)^{\delta/2}.\end{aligned}$$

We omit the details. Hence, term I can be realized in the form

$$C\sum_{i_1=1}^{\infty}\sum_{i_2=1}^{\infty}\sum_{j_3=1}^{\infty}2^{-i_1\delta/2}2^{-i_2\delta/2}2^{-j_3\delta/2}\langle S^{i_10i_200j_3}f,g\rangle.$$

Term III, V, II

Next we deal with term III (symmetric with term V) which can be written in the form

$$\begin{aligned}&\sum_{I_1\subsetneq J_1}\sum_{J_3\subsetneq I_3}\sum_{I_2\subsetneq J_2}\langle h_{J_2}\rangle_{I_2}\langle T(h_{I_1}\otimes h_{I_2}\otimes s_{J_3I_3}),s_{I_1J_1}\otimes 1\otimes h_{J_3}\rangle\cdot\\ &\qquad\langle f,h_{I_1}\otimes h_{I_2}\otimes h_{I_3}\rangle\langle g,h_{J_1}\otimes h_{J_2}\otimes h_{J_3}\rangle\\ &=\sum_{I_1\subsetneq J_1}\sum_{J_3\subsetneq I_3}\sum_{V}\langle\langle g,h_{J_1}\otimes h_{J_3}\rangle_{1,3}\rangle_V\langle T(h_{I_1}\otimes h_V\otimes s_{J_3I_3}),s_{I_1J_1}\otimes 1\otimes h_{J_3}\rangle\cdot\\ &\qquad\langle f,h_{I_1}\otimes h_V\otimes h_{I_3}\rangle.\end{aligned}$$

It is not hard to check the correct normalization of the coefficient

$$|\langle T(h_{I_1}\otimes h_V\otimes s_{J_3I_3}),s_{I_1J_1}\otimes 1\otimes h_{J_3}\rangle|\lesssim\frac{|I_1|^{1/2}}{|J_1|^{1/2}}\left(\frac{\ell(I_1)}{\ell(J_1)}\right)^{\delta/2}\frac{|J_3|^{1/2}}{|I_3|^{1/2}}\left(\frac{\ell(J_3)}{\ell(I_3)}\right)^{\delta/2}|V|^{1/2},$$

which means that term III can be realized in the form

$$C\sum_{i_1=1}^{\infty}\sum_{j_3=1}^{\infty}2^{-i_1\delta/2}2^{-j_3\delta/2}\langle S^{i_10000j_3}f,g\rangle.$$

As $S^{i_10000j_3}$ is a non-cancellative shift, we need to show its boundedness separately, which requires a one-parameter BMO type estimate. Rewrite

$$\begin{aligned}&\sum_V\langle\langle g,h_{J_1}\otimes h_{J_3}\rangle_{1,3}\rangle_V\langle T(h_{I_1}\otimes h_V\otimes s_{J_3I_3}),s_{I_1J_1}\otimes 1\otimes h_{J_3}\rangle\langle f,h_{I_1}\otimes h_V\otimes h_{I_3}\rangle\\ &=\sum_V\langle\langle g,h_{J_1}\otimes h_{J_3}\rangle_{1,3}\rangle_V\langle\langle T^*(s_{I_1J_1}\otimes 1\otimes h_{J_3}),h_{I_1}\otimes s_{J_3I_3}\rangle_{1,3},h_V\rangle_2\cdot\\ &\qquad\qquad\langle\langle f,h_{I_1}\otimes h_{I_3}\rangle_{1,3},h_V\rangle_2\end{aligned}$$

which can be represented as

$$
\begin{aligned}
&=: C2^{-i_1\delta/2}2^{-j_3\delta/2}\langle\langle f, h_{I_1}\otimes h_{I_3}\rangle_{1,3}, \Pi_{b_{I_1J_1J_3I_3}}(\langle g, h_{J_1}\otimes h_{J_3}\rangle_{1,3})\rangle_2\\
&= C2^{-i_1\delta/2}2^{-j_3\delta/2}\langle h_{J_1}\otimes \Pi^*_{b_{I_1J_1J_3I_3}}(\langle f, h_{I_1}\otimes h_{I_3}\rangle_{1,3})\otimes h_{J_3}, g\rangle,
\end{aligned}
$$

where $b_{I_1J_1J_3I_3} = \langle T^*(s_{I_1J_1}\otimes 1\otimes h_{J_3}), h_{I_1}\otimes s_{J_3I_3}\rangle_{1,3}/(C2^{-i_1\delta/2}2^{-j_3\delta/2})$, and Π_b denotes a one-parameter paraproduct in the second variable defined as

$$\Pi_b(f)(x_2) = \sum_V \langle b, h_V\rangle_2\langle f, |V|^{-1/2}\chi_V\rangle_2 h_V(x_2)|V|^{-1/2}.$$

Hence, one has

$$
\begin{aligned}
S^{i_10000j_3}f &= \sum_{J_1}\sum_{\substack{I_1\subset J_1\\ \ell(I_1)=2^{-i_1}\ell(J_1)}}\sum_{I_3}\sum_{\substack{J_3\subset I_3\\ \ell(J_3)=2^{-j_3}\ell(I_3)}} h_{J_1}\otimes \Pi^*_{b_{I_1J_1J_3I_3}}(\langle f, h_{I_1}\otimes h_{I_3}\rangle_{1,3})\otimes h_{J_3}\\
&=: \sum_{J_1}\sum_{I_1\subset J_1}^{(i_1)}\sum_{I_3}\sum_{J_3\subset I_3}^{(j_3)} h_{J_1}\otimes \Pi^*_{b_{I_1J_1J_3I_3}}(\langle f, h_{I_1}\otimes h_{I_3}\rangle_{1,3})\otimes h_{J_3}.
\end{aligned}
$$

One first obverses that there holds the following estimate:

Lemma 7.5. $\|b_{I_1J_1J_3I_3}\|_{BMO(\mathbb{R}^{d_2})} \lesssim \frac{|I_1|^{1/2}}{|J_1|^{1/2}}\frac{|J_3|^{1/2}}{|I_3|^{1/2}}$.

Proof. For any cube V in $\mathbb{R}^{d_2}$, let a be a function on $\mathbb{R}^{d_2}$ with $\mathrm{spt} a \subset V$, $|a|\le 1$ and $\int a = 0$. It suffices to show that

$$|\langle T(h_{I_1}\otimes a\otimes s_{J_3I_3}), s_{I_1J_1}\otimes 1\otimes h_{J_3}\rangle| \lesssim \frac{|I_1|^{1/2}}{|J_1|^{1/2}}\frac{|J_3|^{1/2}}{|I_3|^{1/2}}\left(\frac{\ell(I_1)}{\ell(J_1)}\right)^{\delta/2}\left(\frac{\ell(J_3)}{\ell(I_3)}\right)^{\delta/2}|V|.$$

Since in the pairing, functions of the first and third variables are disjointly supported, one can use partial kernel representation, the standard kernel estimate of $K^{\{1,3\}}_{a,1}$ and boundedness of constant $C^{\{1,3\}}(a,1)$ to derive the desired estimate. We omit the details. □

This then implies that $\Pi^*_{b_{I_1J_1J_3I_3}}$ is bounded on $L^2(\mathbb{R}^{d_2})$ with norm bounded by

$$(|I_1|/|J_1|)^{1/2}(|J_3|/|I_3|)^{1/2}.$$

We now claim that $\|S^{i_1 0000 j_3} f\|_2 \lesssim \|f\|_2$. The idea behind is similar to Proposition 4.5 in [Mar12], but what we face here is more complicated as the relative sizes of cubes in different variables are of mixed type.

Proposition 7.6. *For arbitrary i_1, j_3, there holds*

$$\left\| \sum_{J_1} \sum_{I_1 \subset J_1}^{(i_1)} \sum_{I_3} \sum_{J_3 \subset I_3}^{(j_3)} h_{J_1} \otimes \Pi^*_{b_{I_1J_1J_3I_3}}(\langle f, h_{I_1} \otimes h_{I_3}\rangle_{1,3}) \otimes h_{J_3} \right\|^2_{L^2(\mathbb{R}^{\vec{d}})} \lesssim \|f\|^2_{L^2(\mathbb{R}^{\vec{d}})}.$$

Proof. The orthogonality of Haar systems implies that

$$\begin{aligned}
&\left\| \sum_{J_1} \sum_{I_1 \subset J_1}^{(i_1)} \sum_{I_3} \sum_{J_3 \subset I_3}^{(j_3)} h_{J_1} \otimes \Pi^*_{b_{I_1J_1J_3I_3}}(\langle f, h_{I_1} \otimes h_{I_3}\rangle_{1,3}) \otimes h_{J_3} \right\|^2_{L^2(\mathbb{R}^{\vec{d}})} \\
&= \sum_{J_1} \sum_{J_3} \left\| \sum_{I_1 \subset J_1}^{(i_1)} \Pi^*_{b_{I_1J_1J_3J_3^{(j_3)}}}(\langle f, h_{I_1} \otimes h_{J_3^{(j_3)}}\rangle_{1,3}) \right\|^2_{L^2(\mathbb{R}^{d_2})} \\
&\le \sum_{J_1} \sum_{J_3} \left(\sum_{I_1 \subset J_1}^{(i_1)} \left\| \Pi^*_{b_{I_1J_1J_3J_3^{(j_3)}}}(\langle f, h_{I_1} \otimes h_{J_3^{(j_3)}}\rangle_{1,3}) \right\|_{L^2(\mathbb{R}^{d_2})} \right)^2,
\end{aligned}$$

where $J_3^{(j_3)}$ denotes the j_3^{th} dyadic ancestor of J_3. Now let $P^{i_1}_{J_1}$ denote the orthogonal projection from $L^2(\mathbb{R}^{d_1})$ onto the span of $\{h_{I_1} : I_1 \subset J_1,\ \ell(I_1) = 2^{-i_1}\ell(J_1)\}$, thus,

$$\begin{aligned}
&\left\| \Pi^*_{b_{I_1J_1J_3J_3^{(j_3)}}}(\langle f, h_{I_1} \otimes h_{J_3^{(j_3)}}\rangle_{1,3}) \right\|_{L^2(\mathbb{R}^{d_2})} \lesssim \frac{|I_1|^{1/2}}{|J_1|^{1/2}} \frac{|J_3|^{1/2}}{|J_3^{(j_3)}|^{1/2}} \|\langle f, h_{I_1} \otimes h_{J_3^{(j_3)}}\rangle_{1,3}\|_{L^2(\mathbb{R}^{d_2})} \\
&\le \frac{|I_1|^{1/2}}{|J_1|^{1/2}} \frac{|J_3|^{1/2}}{|J_3^{(j_3)}|^{1/2}} \left(\int_{\mathbb{R}^{d_2}} \int_{I_1} |P^{i_1}_{J_1}(\langle f, h_{J_3^{(j_3)}}\rangle_3)|^2 \, dx_1 dx_2 \right)^{1/2}.
\end{aligned}$$

Therefore, one has by Hölder's inequality that

$$\begin{aligned}
&\left\|\sum_{J_1}\sum_{I_1\subset J_1}^{(i_1)}\sum_{I_3}\sum_{J_3\subset I_3}^{(j_3)} h_{J_1}\otimes \Pi^*_{b_{I_1J_1J_3I_3}}(\langle f, h_{I_1}\otimes h_{I_3}\rangle_{1,3})\otimes h_{J_3}\right\|^2_{L^2(\mathbb{R}^{\vec{d}})}\\
&\lesssim \sum_{J_1}\sum_{J_3}\left(\sum_{I_1\subset J_1}^{(i_1)}\frac{|I_1|^{1/2}}{|J_1|^{1/2}}\frac{|J_3|^{1/2}}{|J_3^{(j_3)}|^{1/2}}\left(\int_{\mathbb{R}^{d_2}}\int_{I_1}|P^{i_1}_{J_1}(\langle f, h_{J_3^{(j_3)}}\rangle_3)|^2\,dx_1dx_2\right)^{1/2}\right)^2,\\
&\lesssim \sum_{J_1}\sum_{J_3}\left(\sum_{I_1\subset J_1}^{(i_1)}\frac{|I_1|}{|J_1|}\frac{|J_3|}{|J_3^{(j_3)}|}\right)\left(\sum_{I_1\subset J_1}^{(i_1)}\int_{\mathbb{R}^{d_2}}\int_{I_1}|P^{i_1}_{J_1}(\langle f, h_{J_3^{(j_3)}}\rangle_3)|^2\,dx_1dx_2\right)\\
&=\sum_{J_3}\frac{|J_3|}{|J_3^{(j_3)}|}\sum_{J_1}\int_{\mathbb{R}^{d_2}}\int_{\mathbb{R}^{d_1}}|P^{i_1}_{J_1}(\langle f, h_{J_3^{(j_3)}}\rangle_3)|^2\,dx_1dx_2,
\end{aligned}$$

which due to the orthogonality of $\{P^{i_1}_{J_1}\}_{J_1}$ is equal to

$$=\sum_{J_3}\frac{|J_3|}{|J_3^{(j_3)}|}\|\langle f, h_{J_3^{(j_3)}}\rangle_3\|^2_{L^2(\mathbb{R}^{d_1+d_2})}.$$

Note that by reindexing $J_3^{(j_3)}$ as I_3, the last expression above can be rewritten as

$$\sum_{I_3}\sum_{J_3\subset I_3}^{(j_3)}\frac{|J_3|}{|I_3|}\|\langle f, h_{I_3}\rangle_3\|^2_{L^2(\mathbb{R}^{d_1+d_2})}=\|f\|^2_{L^2(\mathbb{R}^{\vec{d}})},$$

which completes the proof. □

This finishes the discussion of term III. Though term II is not completely symmetric to III or V, it can be handled similarly by realized in a form of sums of terms involving one-parameter paraproducts and by using the following BMO lemma. The boundedness of the arising dyadic shifts then follows from a similar argument as for Proposition 7.6.

Lemma 7.7. *Define* $b_{I_1J_1I_2J_2}=\langle T(h_{I_1}\otimes h_{I_2}\otimes 1), s_{I_1J_1}\otimes s_{I_2J_2}\rangle_{1,2}/(C2^{-i_1\delta/2}2^{-i_2\delta/2})$, *then*

$$\|b_{I_1J_1I_2J_2}\|_{BMO(\mathbb{R}^{d_3})}\lesssim\frac{|I_1|^{1/2}}{|J_1|^{1/2}}\frac{|I_2|^{1/2}}{|J_2|^{1/2}}.$$

The proof of the lemma is completely the same as Lemma 7.5, which we omit.

Term IV**,** VI**,** VII

Now we turn to term IV (symmetric with term VI), which can be realized in a form involving bi-parameter paraproduct. Write

$$\begin{aligned}&\sum_{I_1\subsetneq J_1}\sum_{I_2\subsetneq J_2}\sum_{J_3\subsetneq I_3}\langle h_{J_2}\rangle_{I_2}\langle h_{I_3}\rangle_{J_3}\langle T(h_{I_1}\otimes h_{I_2}\otimes 1), s_{I_1J_1}\otimes 1\otimes h_{J_3}\rangle\cdot\\ &\qquad\qquad\langle f, h_{I_1}\otimes h_{I_2}\otimes h_{I_3}\rangle\langle g, h_{J_1}\otimes h_{J_2}\otimes h_{J_3}\rangle\\ &=\sum_{I_1\subsetneq J_1}\sum_{V}\sum_{W}\langle\langle g, h_{J_1}\otimes h_W\rangle_{1,3}\rangle_V\langle\langle f, h_{I_1}\otimes h_V\rangle_{1,2}\rangle_W\langle T(h_{I_1}\otimes h_V\otimes 1), s_{I_1J_1}\otimes 1\otimes h_W\rangle,\end{aligned}$$

which is of the form

$$C\sum_{i_1=1}^{\infty}2^{-i_1\delta/2}\langle S^{i_1 00000}f, g\rangle,$$

if one can prove that the following correct normalization holds true:

$$|\langle T(h_{I_1}\otimes h_V\otimes 1), s_{I_1J_1}\otimes 1\otimes h_W\rangle|\lesssim\frac{|I_1|^{1/2}}{|J_1|^{1/2}}\left(\frac{\ell(I_1)}{\ell(J_1)}\right)^{\delta/2}|V|^{1/2}|W|^{1/2}.$$

To see this, recall that by the partial kernel representation,

$$\begin{aligned}&\langle T(h_{I_1}\otimes h_V\otimes 1), s_{I_1J_1}\otimes 1\otimes h_W\rangle=\langle T_2(h_{I_1}\otimes 1\otimes 1), s_{I_1J_1}\otimes h_V\otimes h_W\rangle\\ &=\int_{\mathbb{R}^{d_1}}\int_{\mathbb{R}^{d_1}}K^{\{1\}}_{2,1\otimes 1,h_V\otimes h_W}(x_1,y_1)h_{I_1}(y_1)s_{I_1J_1}(x_1)\,dx_1dy_1,\end{aligned}$$

where the partial kernel $K^{\{1\}}_{2,1\otimes 1,h_V\otimes h_W}$ satisfies standard kernel estimates with constant $C^{\{1\}}(1\otimes 1, h_V\otimes h_W)$, where additionally we have the assumption that $C^{\{1\}}(1\otimes 1,\cdot)$ is a function in $\mathrm{BMO}_{\mathrm{prod}}(\mathbb{R}^{d_2}\times\mathbb{R}^{d_3})$ with norm $\lesssim 1$. Hence, there holds $C^{\{1\}}(1\otimes 1, h_V\otimes h_W)\lesssim|V|^{1/2}|W|^{1/2}$, and the correct normalization of the coefficients then follows from a completely same argument as Lemma 3.10 in [Hyt11]. It is then left to demonstrate the uniform boundedness of the shift $S^{i_1 00000}$. Rewrite

$$\begin{aligned}&\sum_{V}\sum_{W}\langle\langle g, h_{J_1}\otimes h_W\rangle_{1,3}\rangle_V\langle\langle f, h_{I_1}\otimes h_V\rangle_{1,2}\rangle_W\langle T(h_{I_1}\otimes h_V\otimes 1), s_{I_1J_1}\otimes 1\otimes h_W\rangle\\ &=C2^{-i_1\delta/2}\langle h_{J_1}\otimes\Pi_{b_{I_1J_1}}(\langle f, h_{I_1}\rangle_1), g\rangle,\end{aligned}$$

where $b_{I_1J_1} := \langle T_2(h_{I_1}\otimes 1\otimes 1), s_{I_1J_1}\rangle_1/(C2^{-i_1\delta/2})$. The bi-parameter paraproduct

$$\Pi_b(f) := \sum_{V,W}\langle b, h_V\otimes h_W\rangle_{2,3}\langle f, h_V\otimes h_W^1\rangle_{2,3}h_V^1\otimes h_W|V|^{-1/2}|W|^{-1/2},$$

where h_V^1, h_W^1 are non-cancellative Haar functions defined as $|V|^{-1/2}\chi_V, |W|^{-1/2}\chi_W$, respectively. Since the boundedness of $\Pi_{b_{I_1J_1}}$ implies the uniform boundedness of $S^{i_1 00000}$ similarly as in Proposition 7.6, it thus suffices to prove the following result:

Lemma 7.8. $\|b_{I_1J_1}\|_{BMO_{prod}(\mathbb{R}^{d_2}\times\mathbb{R}^{d_3})} \lesssim \frac{|I_1|^{1/2}}{|J_1|^{1/2}}$.

Proof. To see this, one needs to refer to the partial kernel assumption and the WBP/BMO conditions of the constant. Specifically, we will prove that for any dyadic grids $\mathcal{D}_2, \mathcal{D}_3$, and any open set $\Omega\subset\mathbb{R}^{d_2}\times\mathbb{R}^{d_3}$ with finite measure, there holds

$$\frac{1}{|\Omega|}\sum_{\substack{R\subset\Omega, R\in\mathcal{D}_2\times\mathcal{D}_3\\ R=J_2\times J_3}}|\langle T_2(h_{I_1}\otimes 1\otimes 1), s_{I_1J_1}\otimes h_{J_2}\otimes h_{J_3}\rangle|^2/(C^2 2^{-i_1\delta}) \lesssim \frac{|I_1|}{|J_1|}.$$

Due to the disjoint supports of h_{I_1} and $s_{I_1J_1}$, one has

$$\langle T_2(h_{I_1}\otimes 1\otimes 1), s_{I_1J_1}\otimes h_{J_2}\otimes h_{J_3}\rangle = \int_{I_1}\int_{Q_1^c} K^{\{1\}}_{2,1\otimes 1,h_{J_2}\otimes h_{J_3}}(x_1,y_1)h_{I_1}(y_1)s_{I_1J_1}(x_1)\,dx_1dy_1. \tag{7.9}$$

If $\ell(I_1) < 2^{-r}\ell(J_1)$, the goodness of I_1 implies $d(I_1, Q_1^c)\geq \ell(J_1)(\ell(I_1)/\ell(J_1))^{\gamma_1}$. Hence, according to the mean zero property of h_{I_1} and Hölder conditions of the partial kernel, one has

$$\begin{aligned}
&|(7.9)|\\
=&\Big|\int_{I_1}\int_{Q_1^c}[K^{\{1\}}_{2,1\otimes 1,h_{J_2}\otimes h_{J_3}}(x_1,y_1) - K^{\{1\}}_{2,1\otimes 1,h_{J_2}\otimes h_{J_3}}(x_1,c(I_1))]h_{I_1}(y_1)s_{I_1J_1}(x_1)\,dx_1dy_1\Big|\\
\lesssim& C_2^{\{1\}}(1\otimes 1, h_{J_2}\otimes h_{J_3})\|h_{I_1}\|_1\|s_{I_1J_1}\|_\infty\Big|\int_{Q_1^c}\frac{\ell(I_1)^\delta}{d(x_1,I_1)^{d_1+\delta}}\,dx_1\Big|\\
\lesssim& C_2^{\{1\}}(1\otimes 1, h_{J_2}\otimes h_{J_3})\frac{|I_1|^{1/2}}{|J_1|^{1/2}}\left(\frac{\ell(I_1)}{\ell(J_1)}\right)^{\delta/2}.
\end{aligned}$$

If $\ell(I_1) \geq 2^{-r}\ell(J_1)$ instead, we further split (7.9) into two parts. Write

$$
\begin{aligned}
|(7.9)| \leq & \int_{3I_1\setminus I_1} \Big| \int_{I_1} K^{\{1\}}_{2,1\otimes 1,h_{J_2}\otimes h_{J_3}}(x_1,y_1)h_{I_1}(y_1)\,dy_1\Big|\Big|s_{I_1J_1}(x_1)\Big|\,dx_1+\\
& \int_{(3I_1)^c} \Big| \int_{I_1} [K^{\{1\}}_{2,1\otimes 1,h_{J_2}\otimes h_{J_3}}(x_1,y_1) - K^{\{1\}}_{2,1\otimes 1,h_{J_2}\otimes h_{J_3}}(x_1,c(I_1))]h_{I_1}(y_1)\,dy_1\Big|\Big|s_{I_1J_1}(x_1)\Big|\,dx_1\\
&\lesssim C_2^{\{1\}}(1\otimes 1, h_{J_2}\otimes h_{J_3})\|h_{I_1}\|_\infty \|s_{I_1J_1}\|_\infty \int_{3I_1\setminus I_1}\int_{I_1} \frac{1}{|x_1-y_1|^{d_1}}\,dy_1dx_1\\
&\quad + C_2^{\{1\}}(1\otimes 1, h_{J_2}\otimes h_{J_3})\|h_{I_1}\|_1 \|s_{I_1J_1}\|_\infty \int_{(3I_1)^c} \frac{\ell(I_1)^\delta}{d(x_1,I_1)^{d_1+\delta}}\,dx_1,
\end{aligned}
$$

therefore

$$
\lesssim C_2^{\{1\}}(1\otimes 1, h_{J_2}\otimes h_{J_3})\frac{|I_1|^{1/2}}{|J_1|^{1/2}} \lesssim C_2^{\{1\}}(1\otimes 1, h_{J_2}\otimes h_{J_3})\frac{|I_1|^{1/2}}{|J_1|^{1/2}}\left(\frac{\ell(I_1)}{\ell(J_1)}\right)^{\delta/2}.
$$

Combining the two cases together, we obtain

$$
|(7.9)| \lesssim C_2^{\{1\}}(1\otimes 1, h_{J_2}\otimes h_{J_3})\frac{|I_1|^{1/2}}{|J_1|^{1/2}}2^{-i_1\delta/2},
$$

which then implies that

$$
\begin{aligned}
&\frac{1}{|\Omega|}\sum_{\substack{R\subset\Omega, R\in\mathcal{D}_2\times\mathcal{D}_3\\ R=J_2\times J_3}} |\langle T_2(h_{I_1}\otimes 1\otimes 1), s_{I_1J_1}\otimes h_{J_2}\otimes h_{J_3}\rangle|^2/(C^2 2^{-i_1\delta})\\
&\lesssim \frac{1}{|\Omega|}\sum_{\substack{R\subset\Omega, R\in\mathcal{D}_2\times\mathcal{D}_3\\ R=J_2\times J_3}} |C_2^{\{1\}}(1\otimes 1, h_{J_2}\otimes h_{J_3})|^2\frac{|I_1|}{|J_1|} \lesssim \frac{|I_1|}{|J_1|},
\end{aligned}
$$

where the last step follows from the WBP/BMO assumption that $C_2^{\{1\}}(1\otimes 1,\cdot)$ is a product BMO function with norm $\lesssim 1$. □

This finishes the discussion of the term IV. Similarly, term VII can also be expressed as a sum of terms involving bi-parameter paraproducts, where the BMO function and the correct boundedness are given in the following lemma, whose proof is omitted.

Lemma 7.10. *Define* $b_{J_3I_3} = \langle T^*(1 \otimes 1 \otimes h_{J_3}), s_{J_3I_3}\rangle_3/(C2^{-j_3\delta/2})$*, then,*

$$\|b_{J_3I_3}\|_{BMO_{prod}(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2})} \lesssim \frac{|J_3|^{1/2}}{|I_3|^{1/2}}.$$

Term $VIII$

In order to deal with the last term, one needs to realize it into the desired form using tri-parameter paraproducts and apply the assumed mixed BMO/WBP conditions. Specifically, write

$$\begin{aligned}
&\sum_{I_1\subsetneq J_1}\sum_{I_2\subsetneq J_2}\sum_{J_3\subsetneq I_3} \langle h_{J_1}\rangle_{I_1}\langle h_{J_2}\rangle_{I_2}\langle h_{I_3}\rangle_{J_3}\langle T_3^*(1), h_{I_1}\otimes h_{I_2}\otimes h_{J_3}\rangle\cdot \\
&\qquad\qquad \langle f, h_{I_1}\otimes h_{I_2}\otimes h_{I_3}\rangle\langle g, h_{J_1}\otimes h_{J_2}\otimes h_{J_3}\rangle \\
&= \sum_{K,V,W}\langle\langle g, h_W\rangle_3\rangle_{K\times V}\langle\langle f, h_K\otimes h_V\rangle_{1,2}\rangle_W\langle T_3^*(1), h_K\otimes h_V\otimes h_W\rangle \\
&= \sum_{K,V,W}\langle T_3^*(1), h_K\otimes h_V\otimes h_W\rangle\langle f, h_K\otimes h_V\otimes h_W^1\rangle\cdot \\
&\qquad\qquad \langle g, h_K^1\otimes h_V^1\otimes h_W\rangle|K|^{-1/2}|V|^{-1/2}|W|^{-1/2} \\
&=: \langle \Pi_{T_3^*(1)}f, g\rangle,
\end{aligned}$$

where the tri-parameter paraproduct above is defined as

$$\Pi_b(f) := \sum_{K,V,W}\langle b, h_K\otimes h_V\otimes h_W\rangle\langle f, h_K\otimes h_V\otimes h_W^1\rangle h_K^1\otimes h_V^1\otimes h_W|K|^{-1/2}|V|^{-1/2}|W|^{-1/2}.$$

A hybrid square/maximal function argument shows that in the setting of arbitrarily many parameters, the analogue of paraproduct Π_b defined above is always bounded on L^2 for product BMO symbol function b. Since it is one of our mixed BMO/WBP assumptions that $T_3^*(1) \in \mathrm{BMO}_{\mathrm{prod}}$, term $VIII$ can thus be realized as the form $C\langle S^{000000}f, g\rangle$. The proof of the case Inside/Inside/Inside is therefore complete.

Now one can see that for estimate of other cases where not all the pairs of cubes are nested, less multi-parameter paraproduct type estimates are involved. One just needs to carefully apply the suitable standard kernel assumptions to derive the correct normalization, which shouldn't involve any new elements once we have seen what is happening in this more difficult case. It is also not hard to observe that our argument can be easily adapted to handle all the different mixed cases due to the symmetry of our conditions formulated at the beginning of the paper, hence the proof of Theorem 7.2 is complete.

Before ending the section, we emphasize that unlike [Mar12], in the setting of more than two parameters, one has to deal with "partial type" multi-parameter paraproducts (for example for term IV, VI, VII above) in addition to the classical one-parameter ones in the discussion of the above and other cases. This explains why one needs to formulate the full kernel, partial kernel, BMO/WBP assumptions for the operator T in such a particular way as we have done it.

7.5 Comparison to Journé's class

As introduced in Chapter 2, the first general class of bi-parameter singular integral operators containing non-convolution type operators was established by Journé in [Jou85], where he proved a bi-parameter $T(1)$ theorem as well. It is also pointed out in [Jou85] that, by induction, his approach can be generalized to arbitrarily many parameters. We refer the reader to Subsection 2.2.1 of Chapter 2 for the definition of a multi-parameter Journé operator.

The work in this section is inspired from a recent result of Grau de la Herran [GdlH15], where she shows that in the bi-parameter setting, under the additional assumption that T is bounded on L^2, T is a Journé type δ-SIO satisfying certain WBP if and only if it satisfies Martikainen's mixed type conditions in [Mar12]. In the following, we reformulate this theorem without any assumption of the L^2 boundedness and extend it to the multi-parameter setting by comparing multi-parameter Journé operators to our class of singular integrals defined in Section 7.1. Note that in [GdlH15], the L^2 boundedness is used only to compare the two different formulations of WBP. However, in both Journé's and our class of singular integrals, the WBP enter only in the context of the boundedness of the operator.

Our main theorem in this section is formulated as follows:

Theorem 7.11. *T is a t-parameter singular integral operator satisfying both the full kernel and partial kernel assumptions (as stated in Subsections 7.1.1 and 7.1.2) if and only if it is a Journé type t-parameter SIO (see Definition 2.8 in Chapter 2).*

To prove this in the multi-parameter setting, one of the intrinsic new difficulties is the need of some type of multi-parameter $T(1)$ theorem, which we tackle by applying Corollary 7.3.

Proof. We will prove the theorem in the case $t = 3$ as an example, which is enough to show the new multi-parameter phenomena in the problem. And for simplicity of notations, let's assume that the dimensions $d_1 = d_2 = d_3 = 1$. To remind ourselves, T is a Journé type tri-parameter SIO if there exists a triple (K_1, K_2, K_3) of δCZ$(\mathbb{R} \times \mathbb{R})$-$\delta$-standard kernels such that

$$\langle T(f_1 \otimes f_2 \otimes f_3), g_1 \otimes g_2 \otimes g_3\rangle = \int f_1(y_1)\langle K_1(x_1, y_1)f_2 \otimes f_3, g_2 \otimes g_3\rangle g_1(x_1)\, dx_1dy_1 \quad (7.12)$$

when $\operatorname{spt} f_1 \cap \operatorname{spt} g_1 = \emptyset$, and similarly for K_2, K_3.

It is important to keep in mind that for any fixed x_1, y_1, $K_1(x_1, y_1)$ is a Journé type bi-parameter SIO on $\mathbb{R} \times \mathbb{R}$.

To show that any Journé type tri-parameter SIO T satisfies our full and partial kernel assumptions, one can basically follow the strategy in [GdlH15], and note that no L^2 boundedness is needed. Due to the symmetries of the conditions, it suffices to check the kernel assumptions for T while the results for other T_S follow similarly. The full kernel assumptions are straightforward to verify, which we omit. For the partial kernel assumptions, let's look at the most difficult case $V=\{1\}$ as an example, while all the other cases follow symmetrically.

For any $\operatorname{spt} f_1 \cap \operatorname{spt} g_1 = \emptyset$, since T is a Journé type operator, we have

$$\langle T(f_1\otimes f_2\otimes f_3), g_1\otimes g_2\otimes g_3\rangle = \int f_1(y_1)\langle K_1(x_1,y_1)f_2\otimes f_3, g_2\otimes g_3\rangle g_1(x_1)\,dx_1dy_1.$$

Define partial kernel $K^{\{1\}}_{f_2\otimes f_3, g_2\otimes g_3}(x_1,y_1) := \langle K_1(x_1,y_1)f_2\otimes f_3, g_2\otimes g_3\rangle$. Then the mixed size-Hölder conditions are implied by the fact that $K_1(x_1,y_1)$ is a $\delta\mathrm{CZ}(\mathbb{R}\times\mathbb{R})$-$\delta$-standard kernel. Let's first look at the standard kernel estimates and the boundedness of constant $C^{\{1\}}(1\otimes 1,\cdot)$. Since $K_1(x_1,y_1)$ maps $L^\infty(\mathbb{R}\times\mathbb{R})$ boundedly into $\mathrm{BMO}_{\mathrm{prod}}(\mathbb{R}\times\mathbb{R})$ with an operator norm bounded by $\|K_1(x_1,y_1)\|_{\delta\mathrm{CZ}(\mathbb{R}\times\mathbb{R})}$, a result proved by Journé in [Jou85]. $K^{\{1\}}_{1\otimes 1, g_{23}}$ is thus well defined for any function $g_{23}\in H^1(\mathbb{R}\times\mathbb{R})$, not necessarily to be a tensor product.

Then in order to justify the size condition, one writes

$$|K^{\{1\}}_{1\otimes 1, g_{23}}(x_1,y_1)| = |\langle K_1(x_1,y_1)1\otimes 1, g_{23}\rangle| \lesssim \|K_1(x_1,y_1)\|_{\delta\mathrm{CZ}(\mathbb{R}\times\mathbb{R})},$$

where $\|g_{23}\|_{H^1_{\mathrm{prod}}(\mathbb{R}\times\mathbb{R})}\leq 1$. Hence, by the vector-valued standard kernel assumptions of $K_1(x_1,y_1)$,

$$|K^{\{1\}}_{1\otimes 1, g_{23}}(x_1,y_1)| \leq C^{\{1\}}(1\otimes 1, g_{23})\frac{1}{|x_1-y_1|},$$

where $C^{\{1\}}(1\otimes 1, g_{23})$ is some constant that is universally bounded.

For Hölder conditions, one can similarly write

$$\begin{aligned}&|K^{\{1\}}_{1\otimes 1, g_{23}}(x_1,y_1) - K^{\{1\}}_{1\otimes 1, g_{23}}(x_1',y_1)| = |\langle (K_1(x_1,y_1)-K_1(x_1',y_1))1\otimes 1, g_{23}\rangle|\\ &\lesssim \|K_1(x_1,y_1)-K_1(x_1',y_1)\|_{\delta\mathrm{CZ}(\mathbb{R}\times\mathbb{R})} \lesssim C^{\{1\}}(1\otimes 1, g_{23})\frac{|x_1-x_1'|^\delta}{|x_1-y_1|^{1+\delta}},\end{aligned}$$

where the constant $C^{\{1\}}(1\otimes 1, g_{23})$ is the same as before. This completes the proof of the standard kernel estimates and the BMO condition of $C^{\{1\}}(1\otimes 1,\cdot)$ as well.

To prove the bounds for $C^{\{1\}}(\chi_{I_2}\otimes 1,\chi_{I_2}\otimes h)$ (h being an atom of $H^1(\mathbb{R})$ adapted to cube V), for simplicity we only verify the size condition as the Hölder conditions are similar. Split

$$\begin{aligned}&K^{\{1\}}_{\chi_{I_2}\otimes 1,\chi_{I_2}\otimes h}(x_1,y_1)=\langle K_1(x_1,y_1)\chi_{I_2}\otimes 1,\chi_{I_2}\otimes h\rangle\\&=\langle K_1(x_1,y_1)\chi_{I_2}\otimes\chi_{3V},\chi_{I_2}\otimes h\rangle+\langle K_1(x_1,y_1)\chi_{I_2}\otimes\chi_{(3V)^c},\chi_{I_2}\otimes h\rangle=:I+II.\end{aligned}$$

The first term can be estimated using L^2 bounds:

$$|I|\leq\|K_1(x_1,y_1)\|_{\delta\mathrm{CZ}(\mathbb{R}\times\mathbb{R})}\|\chi_{I_2}\otimes\chi_{3V}\|_2\|\chi_{I_2}\otimes h\|_2\lesssim\|K_1(x_1,y_1)\|_{\delta\mathrm{CZ}(\mathbb{R}\times\mathbb{R})}|I_2|.$$

For the second term, noticing that $\chi_{(3V)^c}$ and h are disjointly supported, by the definition of bi-parameter Journé type CZ operator, there exists Calderón-Zygmund operator $K_1^3(x_1,y_1,x_3,y_3)$ such that

$$II=\int\chi_{(3V)^c}(y_3)\langle K_1^3(x_1,y_1,x_3,y_3)\chi_{I_2},\chi_{I_2}\rangle h(x_3)\,dx_3dy_3,$$

which by the vector-valued standard kernel estimate equals

$$\begin{aligned}&=\int\chi_{(3V)^c}(y_3)\left\langle[K_1^3(x_1,y_1,x_3,y_3)-K_1^3(x_1,y_1,x_3,c(V))]\chi_{I_2},\chi_{I_2}\right\rangle h(x_3)\,dx_3dy_3\\&\leq|I_2|\int|\chi_{(3V)^c}(y_3)h(x_3)|\|K_1^3(x_1,y_1,x_3,y_3)-K_1^3(x_1,y_1,x_3,c(V))\|_{\delta\mathrm{CZ}(\mathbb{R})}\,dx_3dy_3\\&\leq|I_2|\|K_1(x_1,y_1)\|_{\delta\mathrm{CZ}(\mathbb{R}\times\mathbb{R})}\int|\chi_{(3V)^c}(y_3)h(x_3)|\frac{\ell(V)^\delta}{d(y_3,V)^{1+\delta}}\,dx_3dy_3\\&\lesssim|I_2|\|K_1(x_1,y_1)\|_{\delta\mathrm{CZ}(\mathbb{R}\times\mathbb{R})}.\end{aligned}$$

One thus has the size condition

$$|K^{\{1\}}_{\chi_{I_2}\otimes 1,\chi_{I_2}\otimes h}(x_1,y_1)|\lesssim C^{\{1\}}(\chi_{I_2}\otimes 1,\chi_{I_2}\otimes h)\frac{1}{|x_1-y_1|},$$

where the constant is taken so that $C^{\{1\}}(\chi_{I_2}\otimes 1,\chi_{I_2}\otimes h)\lesssim|I_2|$, hence satisfies the desired

BMO estimate.

Lastly, the estimate for $C^{\{1\}}(\chi_{I_2}\otimes\chi_{I_3},\chi_{I_2}\otimes\chi_{I_3})$ can be obtained similarly based solely on the L^2 boundedness of $K_1(x_1,y_1)$, which completes the easy direction of the proof of Theorem 7.11.

To prove the other direction, for any given tri-parameter operator T, together with all of its partial adjoints satisfying the full and partial kernel assumptions, we are going to prove that it is a Journé type SIO, i.e. there exist δCZ($\mathbb{R}\times\mathbb{R}$)-$\delta$-standard kernels K_1, K_2 and K_3. By symmetry, it suffices to show the existence of K_1.

For any $\operatorname{spt} f_1\cap\operatorname{spt} g_1=\emptyset$, there holds for some partial kernel $K^{\{1\}}_{f_2\otimes f_3,g_2\otimes g_3}$ that

$$\langle T(f_1\otimes f_2\otimes f_3),g_1\otimes g_2\otimes g_3\rangle=\int K^{\{1\}}_{f_2\otimes f_3,g_2\otimes g_3}(x_1,y_1)f_1(y_1)g_1(x_1)\,dx_1dy_1.$$

This suggests us to define a bi-parameter operator $K_1(x_1,y_1)$ associated with the following bilinear form:

$$\langle K_1(x_1,y_1)f_2\otimes f_3,g_2\otimes g_3\rangle:=K^{\{1\}}_{f_2\otimes f_3,g_2\otimes g_3}(x_1,y_1).$$

It is left to prove that $K_1(x_1,y_1)$ is a Journé type δ-CZ operator on $\mathbb{R}\times\mathbb{R}$ and satisfies the standard kernel estimates. For the sake of brevity, we will focus only on the size condition, i.e. to show that $\|K_1(x_1,y_1)\|_{\delta\mathrm{CZ}(\mathbb{R}\times\mathbb{R})}\lesssim|x_1-y_1|^{-1}$.

For any fixed x_1,y_1, the fact that $K_1(x_1,y_1)$ defined above is indeed a linear continuous mapping follows from the linearity and continuity of T itself, with the aid of Lebesgue differentiation theorem.

To see that $K_1(x_1,y_1)$ is a Journé type bi-parameter δ-SIO, according to the definition, we need to show the existence of a pair $(K_1^2(x_1,y_1,x_2,y_2),K_1^3(x_1,y_1,x_3,y_3))$ of δCZ-δ-

standard kernels such that

$$\begin{aligned}\langle K_1(x_1,y_1)f_2\otimes f_3, g_2\otimes g_3\rangle &= K^{\{1\}}_{f_2\otimes f_3,g_2\otimes g_3}(x_1,y_1)\\ &= \int f_2(y_2)\langle K_1^2(x_1,y_1,x_2,y_2)f_3,g_3\rangle g_2(x_2)\,dx_2dy_2\end{aligned} \tag{7.13}$$

when $\operatorname{spt} f_2 \cap \operatorname{spt} g_2 = \emptyset$;

$$\begin{aligned}\langle K_1(x_1,y_1)f_2\otimes f_3, g_2\otimes g_3\rangle &= K^{\{1\}}_{f_2\otimes f_3,g_2\otimes g_3}(x_1,y_1)\\ &= \int f_3(y_3)\langle K_1^3(x_1,y_1,x_3,y_3)f_2,g_2\rangle g_3(x_3)\,dx_3dy_3\end{aligned} \tag{7.14}$$

when $\operatorname{spt} f_3 \cap \operatorname{spt} g_3 = \emptyset$, and the δCZ-valued standard kernel estimates for the operators $K_1^i(x_1,y_1,x_i,y_i)$, $i=2,3$.

The existence of K_1^2, K_1^3 follows from another partial kernel assumption. Let's take K_1^2 as an example. When $\operatorname{spt} f_i \cap \operatorname{spt} g_i = \emptyset$ for $i=1,2$,

$$\begin{aligned}&\langle T(f_1\otimes f_2\otimes f_3), g_1\otimes g_2\otimes g_3\rangle\\ &= \int K^{\{1,2\}}_{f_3,g_3}(x_1,y_1,x_2,y_2)f_1(y_1)f_2(y_2)g_1(x_1)g_2(x_2)\,dx_1dx_2dy_1dy_2\\ &= \int K^{\{1\}}_{f_2\otimes f_3,g_2\otimes g_3}(x_1,y_1)f_1(y_1)g_1(x_1)\,dx_1dy_1.\end{aligned}$$

By Lebesgue differentiation, this implies that

$$\begin{aligned}\langle K_1(x_1,y_1)f_2\otimes f_3, g_2\otimes g_3\rangle &= K^{\{1\}}_{f_2\otimes f_3,g_2\otimes g_3}(x_1,y_1)\\ &= \int K^{\{1,2\}}_{f_3,g_3}(x_1,y_1,x_2,y_2)f_2(y_2)g_2(x_2)\,dx_2dy_2.\end{aligned}$$

It thus natural to define $\langle K_1^2(x_1,y_1,x_2,y_2)f_3,g_3\rangle := K^{\{1,2\}}_{f_3,g_3}(x_1,y_1,x_2,y_2)$.

We next prove $\|K_1^2(x_1,y_1,x_2,y_2)\|_{\delta\mathrm{CZ}} \lesssim |x_1-y_1|^{-1}|x_2-y_2|^{-1}$, which is the size estimate, and the Hölder estimates follow similarly.

First, one can easily check that the operator $K_1^2(x_1,y_1,x_2,y_2)$ is associated with the kernel $K(x_1,y_2,x_2,y_2,\cdot,\cdot)$, which is standard with the correct norm because of the mixed size-Hölder conditions in the full kernel assumption. It thus suffices to prove that

$$\|K_1^2(x_1,y_1,x_2,y_2)\|_{L^2\to L^2}\lesssim |x_1-y_1|^{-1}|x_2-y_2|^{-1},$$

which will follow from Corollary 7.3 in the case $t=1$ provided that $K_1^2(x_1,y_1,x_2,y_2)$ satisfies the BMO/WBP properties. (This is exactly the classical $T(1)$ theorem, rephrased in our language.)

To see this last piece of fact, note that for any normalized H^1 function h, any cube I_3 in the third variable,

$$\begin{aligned}|\langle K_1^2(x_1,y_1,x_2,y_2)1,h\rangle| = |K_{1,h}^{\{1,2\}}(x_1,y_1,x_2,y_2)| &\lesssim C^{\{1,2\}}(1,h)\frac{1}{|x_1-y_1|}\frac{1}{|x_2-y_2|}\\ &\lesssim \frac{1}{|x_1-y_1|}\frac{1}{|x_2-y_2|},\end{aligned}$$

and

$$\begin{aligned}&|\langle K_1^2(x_1,y_1,x_2,y_2)\chi_{I_3},\chi_{I_3}\rangle| = |K_{\chi_{I_3},\chi_{I_3}}^{\{1,2\}}(x_1,y_1,x_2,y_2)|\\ &\lesssim C^{\{1,2\}}(\chi_{I_3},\chi_{I_3})\frac{1}{|x_1-y_1|}\frac{1}{|x_2-y_2|}\lesssim |I_3|\frac{1}{|x_1-y_1|}\frac{1}{|x_2-y_2|},\end{aligned}$$

which are the BMO/WBP assumptions when $t=1$. This demonstrates that $K_1(x_1,y_1)$ is a Journé type bi-parameter δ-SIO on $\mathbb{R}\times\mathbb{R}$.

Now the only gap left in the proof of Theorem 7.11 is to show that as a bi-parameter operator,

$$\|K_1(x_1,y_1)\|_{L^2\to L^2}\lesssim \frac{1}{|x_1-y_1|}, \tag{7.15}$$

together with the same bound for its partial adjoint. We omit the proof of the partial adjoint part as it follows from the same argument by changing T to its corresponding partial adjoint from the beginning.

The proof of (7.15) is exactly where the multi-parameter version of Corollary 7.3 comes into play, as we are in need of a multi-parameter $T(1)$ type theorem of its full strength. It thus suffices to demonstrate that $K_1(x_1, y_1)$ is a bi-parameter singular integral satisfying our full and partial kernel assumptions, as well as the additional BMO/WBP assumptions with the required norm. Note that without loss of generality, we are free to discuss $K_1(x_1, y_1)$ itself only, as the similar results for its partial adjoints will follow from the symmetry of the assumptions on T.

To demonstrate the full kernel assumption, noticing that $K_1(x_1, y_1)$ is associated with kernel $K(x_1, y_1, \cdot, \cdot, \cdot, \cdot)$, then it's not hard to check all the mixed size-Hölder conditions of the kernel.

For the partial kernel assumption, when $\operatorname{spt} f_2 \cap \operatorname{spt} g_2 = \emptyset$, observe that

$$\langle K_1(x_1, y_1) f_2 \otimes f_3, g_2 \otimes g_3 \rangle = \int K^{\{1,2\}}_{f_3, g_3}(x_1, y_1, x_2, y_2) f_2(y_2) g_2(x_2)\, dx_2 dy_2.$$

Then, the partial kernel $K^{\{1,2\}}_{f_3,g_3}$ satisfies the collection of mixed size-Hölder conditions with a constant bounded by $C^{\{1,2\}}(f_3, g_3)|x_1 - y_1|^{-1}$. And for any normalized H^1 function h and any cube I_3,

$$C^{\{1,2\}}(1, h) \lesssim 1, \quad C^{\{1,2\}}(\chi_{I_3}, \chi_{I_3}) \lesssim |I_3|.$$

The Hölder estimates for the partial kernel follow similarly.

It's thus left to check the BMO/WBP assumptions. This will also follow from the partial kernel assumptions of T. First, for any dyadic grids $\mathcal{D}_2, \mathcal{D}_3$ and open set $\Omega \subset \mathbb{R} \times \mathbb{R}$ with finite measure, since

$$|\langle K_1(x_1, y_1) 1 \otimes 1, h_{J_2} \otimes h_{J_3} \rangle| = |K^{\{1\}}_{1 \otimes 1, h_{J_2} \otimes h_{J_3}}(x_1, y_1)| \lesssim C^{\{1\}}(1 \otimes 1, h_{J_2} \otimes h_{J_3}) \frac{1}{|x_1 - y_1|},$$

there holds

$$\begin{aligned}&\frac{1}{|\Omega|}\sum_{\substack{R\subset\Omega,R\in\mathcal{D}_2\times\mathcal{D}_3\\R=J_2\times J_3}}|\langle K_1(x_1,y_1)1\otimes 1,h_{J_2}\otimes h_{J_3}\rangle|^2\\ \lesssim&\frac{1}{|x_1-y_1|}\frac{1}{|\Omega|}\sum_{\substack{R\subset\Omega,R\in\mathcal{D}_2\times\mathcal{D}_3\\R=J_2\times J_3}}|C^{\{1\}}(1\otimes 1,h_{J_2}\otimes h_{J_3})|^2\lesssim\frac{1}{|x_1-y_1|}.\end{aligned}$$

The last inequality above follows from the fact that $C^{\{1\}}(1\otimes 1)$ is a product BMO function with norm $\lesssim 1$.

To verify that other BMO/WBP assumptions hold true, for any normalized $H^1(\mathbb{R})$ function h_3 and cubes I_2, I_3, in the second and third variable respectively, observe that

$$|\langle K_1(x_1,y_1)\chi_{I_2}\otimes\chi_{I_3},\chi_{I_2}\otimes\chi_{I_3}\rangle|\lesssim C^{\{1\}}(\chi_{I_2}\otimes\chi_{I_3},\chi_{I_2}\otimes\chi_{I_3})\frac{1}{|x_1-y_1|}\lesssim|I_2||I_3|\frac{1}{|x_1-y_1|},$$

$$|\langle K_1(x_1,y_1)\chi_{I_2}\otimes 1,\chi_{I_2}\otimes h_3\rangle|\lesssim C^{\{1\}}(\chi_{I_2}\otimes 1,\chi_{I_2}\otimes h_3)\frac{1}{|x_1-y_1|}\lesssim|I_2|\frac{1}{|x_1-y_1|}.$$

Hence, applying Corollary 7.3 in the case $t = 2$ will complete the proof. □

Remark 7.16. Note that when the number of parameters increases, in order to prove Theorem 7.11, one needs to use Corollary 7.3 in the setting of arbitrarily many parameters, where Journé's $T(1)$ theorem fails to be easily applicable due to its many layers of vector-valued formulations. This shows one important aspect of the power of our t-parameter representation theorem for $t \geq 3$.

Once we have Theorem 7.11, the following characterization of Journé type t-parameter δ-CZ operators becomes an immediate consequence.

Corollary 7.17. *T is a Journé type t-parameter δ-CZ operator if and only if it is a t-parameter CZ operator defined in Section 7.1.*

Proof. As we have shown in Theorem 7.11 that Journé's and our classes of t-parameter SIO are equivalent. It is thus left to verify the equivalence between the boundedness of all the partial adjoints of T. This can be shown directly from the inductive definition of Journé type t-parameter CZ operators, observing that in $(t-1)$-parameter, the partial kernels are always CZ operators themselves, satisfying the corresponding L^2 boundedness in $(t-1)$-parameter. □

Up to this point, we have successfully established a set of characterizing conditions for an operator to be a Journé type t-parameter CZ operator. This is very useful in the study of multi-parameter operators since the full kernel, partial kernel, BMO/WBP conditions are usually much easier to verify and use compared with Journé's original vector-valued formulation. From now on, we will refer to our class of multi-parameter singular integrals simply as Journé operators as well.

7.6 Necessity of some of the BMO/WBP conditions

Given a t-parameter singular integral operator T satisfying both full and partial kernel assumptions, one might ask if the mixed BMO/WBP conditions are necessary for T to be bounded on $L^2(\mathbb{R}^{\vec{d}})$. The answer is yes when $t=1$, which is a classical result of Calderón-Zygmund operators introduced in Section 2.1 of Chapter 2, but is no for $t\geq 2$. In fact, a counterexample has been constructed in [Jou85] to show that in the bi-parameter setting, $T_1 1$ and $T_1^* 1\in \mathrm{BMO}_{\mathrm{prod}}$ are not necessary conditions for T to be bounded on L^2.

However, one can indeed prove the necessity of some of the mixed BMO/WBP conditions, more specifically, those that are formulated on T and T^* directly. It is straightforward to verify that pure WBP, i.e.

$$|\langle T(\chi_{I_1}\otimes\cdots\otimes\chi_{I_t}),\chi_{I_1}\otimes\cdots\otimes\chi_{I_t}\rangle|\lesssim\prod_{i=t}^{t}|I_i|$$

is directly implied by the L^2 boundedness of T. For the pure BMO conditions: $T1, T^*1\in$ $\mathrm{BMO}_{\mathrm{prod}}$, the necessity is first pointed out in [GdlH15] for bi-parameters, and is not hard to extend to arbitrarily many parameters using Theorem 7.11. To see this, suppose that there is a L^2 bounded t-parameter singular integral satisfying full and partial kernel assumptions. By Theorem 7.11, T is also a Journé type t-parameter SIO who is bounded on L^2. Hence, Theorem 3 in [Jou85] implies that $T1 \in \mathrm{BMO}_{\mathrm{prod}}$, as well as $T^*1 \in \mathrm{BMO}_{\mathrm{prod}}$ taking into account that T^* is also L^2 bounded.

To prove that for operator T given above, there also hold the mixed BMO/WBP conditions for T, T^*, we take a look at the tri-parameter, $d_1 = d_2 = d_3 = 1$ case as an example. In other words, one wants to show that

$$\|\langle T(\chi_{I_1} \otimes 1 \otimes 1), \chi_{I_1} \otimes \cdot \rangle\|_{\mathrm{BMO}_{\mathrm{prod}}(\mathbb{R}\times\mathbb{R})} \lesssim |I_1|, \tag{7.18}$$

$$\|\langle T(\chi_{I_1} \otimes \chi_{I_2} \otimes 1), \chi_{I_1} \otimes \chi_{I_2} \otimes \cdot \rangle\|_{\mathrm{BMO}(\mathbb{R})} \lesssim |I_1||I_2|, \tag{7.19}$$

and all the other mixed BMO/WBP conditions formulated for T will follow symmetrically, so are the ones for T^*.

In order to prove (7.18), for any cube I_1, one can define an operator $\langle T^1 \chi_{I_1}, \chi_{I_1} \rangle$ mapping $C_0^\infty(\mathbb{R}) \otimes C_0^\infty(\mathbb{R})$ to its dual:

$$\langle \langle T^1 \chi_{I_1}, \chi_{I_1} \rangle f_2 \otimes f_3, g_2 \otimes g_3 \rangle := \langle T(\chi_{I_1} \otimes f_2 \otimes f_3), \chi_{I_1} \otimes g_2 \otimes g_3 \rangle.$$

By taking one parameter away, it is easy to see that $\langle T^1 \chi_{I_1}, \chi_{I_1} \rangle$ is a bi-parameter singular integral, whose full kernel is $K^{\{2,3\}}_{\chi_{I_1}, \chi_{I_1}}(x_2, x_3, y_2, y_3)$ with norm bounded by

$$C^{\{2,3\}}(\chi_{I_1}, \chi_{I_1}) \lesssim |I_1|,$$

while the partial kernel assumptions can be verified similarly. Moreover, following from the definition of $\langle T^1 \chi_{I_1}, \chi_{I_1} \rangle$ and the L^2 boundedness of T, one can conclude that $\langle T^1 \chi_{I_1}, \chi_{I_1} \rangle$ is a L^2 bounded bi-parameter Journé type SIO with norm $\lesssim |I_1|$, thus maps $1 \otimes 1$ boundedly into $\mathrm{BMO}_{\mathrm{prod}}(\mathbb{R} \times \mathbb{R})$, which proves (7.18).

Following the same strategy, one can demonstrate (7.19) by slicing two parameters away and apply the $L^\infty \to \text{BMO}$ estimate for Calderón-Zygmund operators. We omit the details.

This, together with the discussion at the end of Section 7.3, leads us to the following characterizing result for the class of t-parameter CZ operators.

Corollary 7.20. *Given a t-parameter singular integral operator T satisfying both full and partial kernel assumptions, it is then a t-parameter CZ operator if and only if the mixed BMO/WBP assumptions hold true.*

Note that the equivalence above holds true because our CZ operators are defined such that all partial adjoints of T are bounded on L^2, not only T itself.

To end this chapter, we state the following result and sketch the proof, which indicates the generality of our operator class and its inductive nature. Moreover, it also shows that although our class of operators has been proven to be equivalent to Journé's, its mixed type characterizing conditions still provide us with a very helpful tool to study t-parameter operators, especially when t is very large.

Proposition 7.21. *Let $T := T_1 \otimes T_2 \otimes \cdots \otimes T_s$ be an operator on $\mathbb{R}^{\vec{d}} := \mathbb{R}^{\vec{d_1}} \times \cdots \times \mathbb{R}^{\vec{d_s}}$, where for any $1 \leq i \leq s$, T_i is a t_i-parameter CZ operator on $\mathbb{R}^{\vec{d_i}} := \mathbb{R}^{d_i^1} \times \cdots \times \mathbb{R}^{d_i^{t_i}}$. Then T is an t-parameter CZ operator, where $t := t_1 + \cdots + t_s$.*

Proof. Observing that the partial adjoints of T can be expressed as tensor products of some partial adjoints of T_i, it suffices to prove that T itself verifies the full and partial kernel assumptions, as the L^2 boundedness is straightforward.

The full kernel assumption is easy to see, since the tensor product of kernels of T_i is the full kernel and satisfies all the mixed size-Hölder conditions.

To show the partial kernel assumptions, note that in any case, one can always write the partial kernel as a tensor product of some of the full or partial kernels of T_i. And the BMO condition for the constants follow from the fact that the tensor product of partial kernels are always CZ operators with less parameters, hence maps $L^\infty \to \mathrm{BMO}$. To prove the mixed WBP/BMO conditions for the constants, one just needs to take away more parameters and mimic what we did in the proof of (7.18) earlier this section. We omit the details. $\square$

Appendix A

A Real Variable Proof of the Hilbert Commutator Estimate

In this appendix, we give an alternative, direct proof of the lower bound estimate of Ferguson-Sadosky's Hilbert commutator ([FS00]):

$$\|b\|_{\text{bmo}} \lesssim \|[b, H_1H_2]\|_{L^2\to L^2}, \tag{A.1}$$

where $b = b(x, y) \in \text{bmo}(\mathbb{R} \times \mathbb{R})$.

The method presented here completely avoids operator theory, while is an adaptation of the idea of the Riesz commutator lower bound estimate developed in Chapter 5, Section 5.2. As the readers will see in the following, when applied to the Hilbert transform and no iteration case, the argument becomes much simpler and more transparent.

Proof. Let $\Gamma_Q(x, y) = \text{sign}(b(x, y) - \langle b\rangle_Q)\chi_Q(x, y)$. Now

$$\begin{aligned}
&|Q||b(x, y) - \langle b\rangle_Q|\chi_Q(x, y)\\
=&|Q|(b(x, y) - \langle b\rangle_Q)\Gamma_Q(x, y)\\
=&\int_Q \left(\frac{b(x, y) - b(x', y')}{(x - x')(y - y')}\right)(x - x')(y - y')\Gamma_Q(x, y)\, dx'dy'\\
=&[b, H_1H_2]((x - x')(y - y')\chi_Q(x', y'))\Gamma_Q(x, y).
\end{aligned}$$

Integrating a $|Q|^{-2}$ multiple of this equality in $dxdy$ over Q gives us on the left hand side the little BMO expression and on the right hand side

$$\begin{aligned}
&|Q|^{-2}\int_{\mathbb{R}^2}[b, H_1H_2]((x - x')(y - y')\chi_Q(x', y'))\Gamma_Q(x, y)\, dxdy\\
\leq&|Q|^{-2}\|[b, H_1H_2]((x - x')(y - y')\chi_Q(x', y'))\chi_Q(x, y)\|_{L^2}\|\Gamma_Q(x, y)\|_{L^2}\\
\leq&|Q|^{-2}\|[b, H_1H_2]\|_{L^2\to L^2}\|(x - x')(y - y')\chi_Q(x', y')\chi_Q(x, y)\|_{L^2}\|\Gamma_Q(x, y)\|_{L^2}.
\end{aligned}$$

Assuming for a moment that Q is centered about the origin, we obtain due to the supports of Γ_Q and χ_Q the estimate

$$\|(x - x')(y - y')\chi_Q(x', y')\chi_Q(x, y)\|_{L^2}\|\Gamma_Q(x, y)\|_{L^2} \lesssim |Q|^{3/2}|Q|^{1/2},$$

where the first L^2 norm is taken with respect to the variables x', y'. Then a translation argument on b finishes the proof. $\square$

Appendix B

Upper Bound on L^p and Paraproduct Revisited

In this appendix, we complete the proof of Theorem 3.12, whose L^2 case has been proven in Chapter 4. Recall that in Chapter 4, we showed that for Calderón-Zygmund operators T_i, $1 \leq i \leq t$, estimates for the commutator

$$[\dots[b, T_1], \dots, T_t]$$

can be reduced to that for the commutators with dyadic shifts S^{ij}, which can be represented as a finite linear combination of dyadic shifts composed with basic paraproduct-like operators. Given $1 < p < \infty$, note that

$$\|S^{ij}\|_{L^p \to L^p} \lesssim (1 + \max(i, j)),$$

which follows from the fact that

$$\|S^{ij}\|_{L^2(w) \to L^2(w)} \lesssim (1 + \max(i, j))$$

for any A_2 weight w ([Hyt11]) and extrapolation ([RdF84]). The L^p case of Theorem 3.12 can thus be reduced to the boundedness of the paraproduct-like operators (e.g. $BB_{k,l}$, PB_l) on L^p with polynomial dependence on k, l, which will be addressed in Section B.1 for the one-parameter case and Section B.2 for the multi-parameter case.

Moreover, although not required in the upper bound proof, we present some additional results on the boundedness of paraproducts with descendants, in both one-parameter and multi-parameter settings. Take one-parameter as an example. Recall that for $k \geq 1$,

$$B_k(b, f) := \sum_{I \in \mathcal{D}} \beta_I \langle b, h_{I^{(k)}} \rangle \langle f, h_I^1 \rangle h_I |I^{(k)}|^{-1/2},$$

is a non-cancellative paraproduct with descendants, as non-cancellative Haar functions appear in the summation. Although in the cancellative case (Lemma 4.1), we know that the paraproducts with descendants are uniformly bounded in k on L^2, however, in our case here, the best known estimate, even for the L^2 operator norm of such B_k, depends on $k^{1/2}$.

We will prove this result in Section B.1, and extend it to the multi-parameter setting in Section B.2.

Furthermore, it is well known that classical paraproduct estimate (Theorem 2.16) is equivalent to the so-called Carleson embedding theorem:

Theorem B.1. *Suppose that $\{\alpha_I\}_{I\in\mathcal{D}}$ is a sequence of positive numbers indexed by the dyadic cubes in $\mathcal{D}$ such that*

$$\sum_{I:I\subset K} \alpha_I \le C|K|, \quad \forall K \in \mathcal{D}. \tag{B.2}$$

Then,

$$\sum_{J\in\mathcal{D}} \alpha_J \langle f\rangle_J^2 \lesssim C\|f\|_{L^2}^2, \quad \forall f \in L^2.$$

In Section B.3, equivalent to our main paraproduct estimates, generalizations of this result (known us *Carleson embedding theorem with descendants*) will be given.

B.1 One-parameter paraproducts with descendants

We now prove the uniform L^p boundedness of cancellative paraproducts with descendants, where the classical square function is involved.

Lemma B.3. *Let $1 < p < \infty$ and integer $k \ge 1$. For any cancellative paraproduct with descendants*

$$B_k(b,f) = \sum_{I\in\mathcal{D}} \beta_I \langle b, h_{I^{(k)}}\rangle\langle f, h_I\rangle h_I |I^{(k)}|^{-1/2},$$

where $|\beta_I| \le 1$, there holds

$$\|B_k(b,f)\|_{L^p(\mathbb{R}^d)} \lesssim \|b\|_{BMO(\mathbb{R}^d)}\|f\|_{L^p(\mathbb{R}^d)}, \quad \forall f \in L^p.$$

Proof. For any normalized test function $g \in L^{p'}$, applying the H^1-BMO duality, it suffices to show that

$$\left\| \sum_{I\in\mathcal{D}} \beta_I \langle f, h_I\rangle\langle g, h_I\rangle h_{I^{(k)}} |I^{(k)}|^{-1/2} \right\|_{H^1(\mathbb{R}^d)} \lesssim \|f\|_{L^p},$$

which is equivalent to

$$\left\| S\left(\sum_{I\in\mathcal{D}} \beta_I \langle f, h_I\rangle\langle g, h_I\rangle h_{I^{(k)}} |I^{(k)}|^{-1/2} \right) \right\|_{L^1(\mathbb{R}^d)} \lesssim \|f\|_{L^p}. \tag{B.4}$$

The left hand side of (B.4) is equal to

$$\begin{aligned} &\left\| \left(\sum_{J\in\mathcal{D}} \Big(\sum_{I:I^{(k)}=J} \beta_I \langle f, h_I\rangle\langle g, h_I\rangle \Big)^2 \frac{\chi_J}{|J|^2} \right)^{1/2} \right\|_{L^1(\mathbb{R}^d)} \\ \leq &\left\| \sum_{J\in\mathcal{D}} \sum_{I:I^{(k)}=J} |\langle f, h_I\rangle||\langle g, h_I\rangle| \frac{\chi_J}{|J|} \right\|_{L^1(\mathbb{R}^d)}, \end{aligned}$$

which is the same as

$$\left\| \sum_{I\in\mathcal{D}} |\langle f, h_I\rangle||\langle g, h_I\rangle| \frac{\chi_{I^{(k)}}}{|I^{(k)}|} \right\|_{L^1(\mathbb{R}^d)} = \sum_{I\in\mathcal{D}} |\langle f, h_I\rangle||\langle g, h_I\rangle| = \left\| \sum_{I\in\mathcal{D}} |\langle f, h_I\rangle||\langle g, h_I\rangle| \frac{\chi_I}{|I|} \right\|_{L^1(\mathbb{R}^d)}.$$

Applying Cauchy-Schwarz, the above is bounded by

$$\|S(f)S(g)\|_{L^1(\mathbb{R}^d)} \leq \|S(f)\|_{L^p(\mathbb{R}^d)} \|S(g)\|_{L^{p'}(\mathbb{R}^d)} \lesssim \|f\|_{L^p(\mathbb{R}^d)}.$$

□

Next, we show that for non-cancellative paraproduct with descendants, similar bound on L^p still holds true, even though the bound depends on k polynomially.

Lemma B.5. *Let $1 < p < \infty$ and integer $k \geq 1$. For any non-cancellative paraproduct with descendants*

$$B_k(b, f) = \sum_{I\in\mathcal{D}} \beta_I \langle b, h_{I^{(k)}}\rangle\langle f, h_I^1\rangle h_I |I^{(k)}|^{-1/2},$$

where $h_I^1 = |I|^{-1/2}\chi_I$, *there holds*

$$\|B_k(b,f)\|_{L^p(\mathbb{R}^d)} \lesssim k^{1/2}\|b\|_{BMO(\mathbb{R}^d)}\|f\|_{L^p(\mathbb{R}^d)}, \quad \forall f \in L^p.$$

Proof. Define another function $\tilde{b}$ such that

$$\langle \tilde{b}, h_I\rangle = \langle b, h_{I^{(k)}}\rangle 2^{-kd/2}, \quad \forall I \in \mathcal{D},$$

then B_k can be viewed as a classical paraproduct with symbol $\tilde{b}$, i.e.

$$B_k(b,f) = B_0(\tilde{b},f) = \sum_{I\in\mathcal{D}} \beta_I \langle \tilde{b}, h_I\rangle\langle f, h_I^1\rangle h_I |I|^{-1/2}.$$

As mentioned in Chapter 2, $\|B_0(\tilde{b},f)\|_{L^p} \lesssim \|\tilde{b}\|_{\mathrm{BMO}}\|f\|_{L^p}$, it thus suffices to show that

$$\|\tilde{b}\|_{\mathrm{BMO}} \lesssim k^{1/2}\|b\|_{\mathrm{BMO}}. \tag{B.6}$$

To see this, fix any $Q \in \mathcal{D}$. By definition

$$\sum_{I\in\mathcal{D}(Q)} |\langle \tilde{b}, h_I\rangle|^2 = 2^{-kd} \sum_{I\in\mathcal{D}(Q)} |\langle b, h_{I^{(k)}}\rangle|^2.$$

Note that $I \subset Q$ implies $I^{(k)} \subset Q^{(k)}$. Moreover, observe that $\forall 1 \leq l \leq k$,

$$\#\{I \subset Q:\, I^{(k)} = Q^{(l)}\} = 2^{(k-l)d}.$$

Therefore,

$$2^{-kd} \sum_{I\in\mathcal{D}(Q)} \langle b, h_{I^{(k)}}\rangle|^2 = 2^{-kd}\left(\sum_{I:I^{(k)}\subset Q} \langle b, h_{I^{(k)}}\rangle|^2 + \sum_{l=1}^{k} \sum_{\substack{I:I\subset Q\\ I^{(k)}=Q^{(l)}}} \langle b, h_{I^{(k)}}\rangle|^2 \right)$$

which is

$$\leq 2^{-kd}\left(2^{kd}\sum_{J:J\subset Q}\langle b,h_J\rangle|^2+\sum_{l=1}^{k}2^{(k-l)d}\sum_{J:J\subset Q^{(l)}}\langle b,h_J\rangle|^2\right)$$
$$\leq \|b\|_{\mathrm{BMO}}^2|Q|+\sum_{l=1}^{k}2^{-ld}\|b\|_{\mathrm{BMO}}^2|Q^{(l)}|\lesssim k\|b\|_{\mathrm{BMO}}^2|Q|.$$

Hence, (B.6) is justified. □

B.2 Multi-parameter paraproduct-like operators

In this section, we extend the idea of the proof of Lemma B.5 to the multi-parameter setting to obtain L^p bounds for operators $BB_{k,l}$, BP_k, and PB_l. In contrast to the one-parameter setting, even if all Haar functions in the variable with k or l being positive are cancellative, polynomial dependence on k or l will still be present. Therefore, at least in the L^2 case, if these operators are in the forms that are discussed in Lemma 4.4 and 4.8, one can improve the results of this section to obtain uniform dependence on k, l. For other L^p, it is unknown whether these results are sharp.

Lemma B.7. *Let $1<p<\infty$ and integers $k,l\geq 0$. For any dyadic product BMO function b w.r.t. grids $\mathcal{D}_n\times\mathcal{D}_m$ and a paraproduct with descendents defined as*

$$BB_{k,l}(b,f)=\sum_{I_1\in\mathcal{D}_n}\sum_{I_2\in\mathcal{D}_m}\beta_{I_1I_2}\langle b,h_{I_1^{(k)}}\otimes u_{I_2^{(l)}}\rangle\langle f,h_{I_1}^{\epsilon_1}\otimes u_{I_2}^{\epsilon_2}\rangle h_{I_1}^{\epsilon_1'}\otimes u_{I_2}^{\epsilon_2'}|I_1^{(k)}|^{-1/2}|I_2^{(l)}|^{-1/2},$$

where at least one of ϵ_i,ϵ_i' is not $\vec{1}$, for each $i=1,2$, and $|\beta_{I_1I_2}|\leq 1$ uniformly, there holds

$$\|BB_{k,l}(b,f)\|_{L^p}\lesssim \max(k,1)^{1/2}\max(l,1)^{1/2}\|b\|_{BMO_{prod}}\|f\|_{L^p},\quad \forall f\in L^p.$$

Proof. Define function $\tilde{\tilde{b}}$ such that

$$\langle\tilde{\tilde{b}},h_I\otimes u_J\rangle=\langle b,h_{I^{(k)}}\otimes u_{J^{(l)}}\rangle 2^{-kn/2}2^{-lm/2},\quad \forall I\in\mathcal{D}_n,\ J\in\mathcal{D}_m.$$

Then $BB_{k,l}$ can be viewed as a classical bi-parameter paraproduct (2.15) with symbol $\tilde{\tilde{b}}$, which is bounded on L^p according to Theorem 2.16. Therefore, it suffices to prove that

$$\|\tilde{\tilde{b}}\|_{\mathrm{BMO_{prod}}} \lesssim \max(k,1)^{1/2}\max(l,1)^{1/2}\|b\|_{\mathrm{BMO_{prod}}}. \tag{B.8}$$

To see this, define another function $\tilde{b}$ such that

$$\langle \tilde{b}, h_I \otimes u_J\rangle = \langle b, h_I \otimes u_{J^{(l)}}\rangle 2^{-lm/2}, \quad \forall I \in \mathcal{D}_n,\ J \in \mathcal{D}_m.$$

We are going to show that

$$\|\tilde{b}\|_{\mathrm{BMO_{prod}}} \lesssim \max(l,1)^{1/2}\|b\|_{\mathrm{BMO_{prod}}}. \tag{B.9}$$

Then using symmetry and iteration of the argument will yield (B.8).

Assume $l \geq 1$, since otherwise there is nothing to prove. Fix any open set $\Omega \subset \mathbb{R}^n \times \mathbb{R}^m$ with finite measure, note that

$$\sum_{\substack{I\in\mathcal{D}_n, J\in\mathcal{D}_m\\ I\times J\subset\Omega}} |\langle \tilde{b}, h_I\otimes u_J\rangle|^2 = 2^{-lm}\sum_{\substack{I\in\mathcal{D}_n, J\in\mathcal{D}_m\\ I\times J\subset\Omega}} |\langle b, h_I\otimes u_{J^{(l)}}\rangle|^2.$$

For any $0 \leq t \leq l$, define level set

$$\Omega^{(t)} := \{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m : 2^{-tm} \leq M_2(\chi_\Omega) < 2^{(-t+1)m}\},$$

where M_2 is the Hardy-Littlewood maximal function in the second variable. Since $M_2 : L^1(\mathbb{R}^n \times \mathbb{R}^m) \to L^{1,\infty}(\mathbb{R}^n \times \mathbb{R}^m)$, we have $|\Omega^{(t)}| \lesssim 2^{tm}|\Omega|$. Moreover, $\forall x \in \mathbb{R}^n$, denote $\Omega_x := \{y \in \mathbb{R}^m : (x,y) \in \Omega\}$. Then, whenever $I \times J \subset \Omega$, for any $x \in I$, there holds

$$\frac{|J^{(l)} \cap \Omega_x|}{|J^{(l)}|} \geq \frac{|J|}{|J^{(l)}|} = 2^{-lm},$$

hence $M_2(\chi_\Omega)(x,y) \geq 2^{-lm}$ for $(x,y) \in I \times J^{(l)}$. Therefore, decompose

$$2^{-lm} \sum_{I\times J\subset\Omega} |\langle b, h_I \otimes u_{J^{(l)}}\rangle|^2 = 2^{-lm} \sum_{t=0}^{l} \sum_{\substack{I\times J\subset\Omega \\ I\times J^{(l)}\subset\Omega^{(t)}}} |\langle b, h_I \otimes u_{J^{(l)}}\rangle|^2.$$

For $t=0$, one has

$$\begin{aligned}\sum_{\substack{I\times J\subset\Omega \\ I\times J^{(l)}\subset\Omega^{(0)}}} |\langle b, h_I \otimes u_{J^{(l)}}\rangle|^2 &= \sum_{\substack{I\times J: \\ I\times J^{(l)}\subset\Omega}} |\langle b, h_I \otimes u_{J^{(l)}}\rangle|^2 \\ &= 2^{lm} \sum_{I\times J\subset\Omega} |\langle b, h_I \otimes u_J\rangle|^2 \lesssim 2^{lm}\|b\|^2_{\mathrm{BMO_{prod}}}|\Omega|.\end{aligned}$$

For any $1 \leq t \leq l$, rewrite

$$\sum_{\substack{I\times J\subset\Omega \\ I\times J^{(l)}\subset\Omega^{(t)}}} |\langle b, h_I \otimes u_{J^{(l)}}\rangle|^2 = \sum_{I\times Q\subset\Omega^{(t)}} \sum_{\substack{J:J^{(l)}=Q \\ I\times J\subset\Omega}} |\langle b, h_I \otimes u_Q\rangle|^2.$$

Observe that $I \times J \subset \Omega$ and $I \times J^{(l)} \subset \Omega^{(t)}$ yield that $\forall x \in I$,

$$\frac{|J^{(l)} \cap \Omega_x|}{|J^{(l)}|} \leq 2^{(-t+1)m}. \tag{B.10}$$

This, in particular, implies that for any fixed $I \in \mathcal{D}_n$, $Q \in \mathcal{D}_m$,

$$\#\{J \in \mathcal{D}_m : I \times J \subset \Omega, J^{(l)} = Q\} \leq 2^{(l-t+1)m},$$

since $|J|$ times the number on the left hand side is bounded by $2^{(-t+1)m}|J^{(l)}|$ according to (B.10). Hence,

$$\sum_{I\times Q\subset\Omega^{(t)}} \sum_{\substack{J:J^{(l)}=Q \\ I\times J\subset\Omega}} |\langle b, h_I \otimes u_Q\rangle|^2 \leq 2^{(l-t+1)m} \sum_{I\times Q\subset\Omega^{(t)}} |\langle b, h_I \otimes u_Q\rangle|^2 \lesssim 2^{lm}\|b\|^2_{\mathrm{BMO_{prod}}}|\Omega|.$$

Summing over $0 \leq t \leq l$ yields that

$$\sum_{\substack{I\in\mathcal{D}_n, J\in\mathcal{D}_m \\ I\times J\subset\Omega}} |\langle \tilde{b}, h_I \otimes u_J\rangle|^2 \lesssim 2^{-lm} l 2^{lm}\|b\|^2_{\mathrm{BMO_{prod}}}|\Omega|,$$

which completes the proof of (B.9) and thus the lemma. □

Applying the same strategy, one can easily show that there holds the following L^p boundedness of the operators BP_k and PB_l as well.

Lemma B.11. *Given $b \in BMO_{prod}(\mathbb{R}^n \times \mathbb{R}^m)$. Let $\{a^1_{I_2}\}_{I_2\in\mathcal{D}_2}$ be a sequence of functions in $BMO(\mathbb{R}^n)$ indexed by $I_2 \in \mathcal{D}_2$, such that $\sup_{I_2} \|a^1_{I_2}\|_{BMO} =: C_1 < \infty$. And let $\{a^2_{I_1}\}_{I_1\in\mathcal{D}_1}$ be a sequence of functions in $BMO(\mathbb{R}^m)$ indexed by $I_1 \in \mathcal{D}_1$, such that $\sup_{I_1} \|a^2_{I_1}\|_{BMO} =: C_2 < \infty$. For integers $k, l \geq 0$, define*

$$BP_k(b, \{a^2_{I_1}\}_{I_1}, f) :=$$
$$\sum_{I_1,I_2} \beta_{I_1} \langle b, h_{I_1^{(k)}} \otimes u_{I_2}\rangle \langle f, h^{\epsilon_1}_{I_1} \otimes u_{I_2}\rangle |I_1^{(k)}|^{-1/2} |I_2|^{-1} h^{\epsilon_1'}_{I_1} \sum_{J_2:J_2\subsetneq I_2} \langle a^2_{I_1}, u_{J_2}\rangle_2 u_{J_2},$$

$$PB_l(b, \{a^1_{I_2}\}_{I_2}, f) :=$$
$$\sum_{I_1,I_2} \beta_{I_2} \langle b, h_{I_1} \otimes u_{I_2^{(l)}}\rangle \langle f, h_{I_1} \otimes u^{\epsilon_2}_{I_2}\rangle |I_1|^{-1} |I_2^{(l)}|^{-1/2} h^{\epsilon_2'}_{I_2} \sum_{J_1:J_1\subsetneq I_1} \langle a^1_{I_2}, h_{J_1}\rangle_1 h_{J_1},$$

where β_{I_1}, β_{I_2} are sequences satisfying $|\beta_{I_1}|, |\beta_{I_2}| \leq 1$. There is at most one of $h^{\epsilon_1}_{I_1}, h^{\epsilon_1'}_{I_1}$ being non-cancellative for each I_1. The same assumption goes for the second variable. Then, there holds for $1 < p < \infty$,

$$\|BP_k(b, \{a^2_{I_1}\}_{I_1}, f)\|_{L^p} \lesssim \max(k,1)^{1/2} C_2 \|b\|_{BMO_{prod}} \|f\|_{L^p},$$

$$\|PB_l(b, \{a^1_{I_2}\}_{I_2}, f)\|_{L^p} \lesssim \max(l,1)^{1/2} C_1 \|b\|_{BMO_{prod}} \|f\|_{L^p}.$$

Take PB_l as an example. When $l = 0$, the same proof for Lemma 4.8 applies, and the desired L^p bound follows from the L^p bounds of the classical hybrid maximal-square functions. When $l \geq 1$, constructing a function $\tilde{b}$ out of b, similarly as in the proof of Lemma B.7, reduces the problem to the $l = 0$ case. Then the desired estimate is yielded by the BMO norm estimate of $\tilde{b}$, which is identical to the one discussed in Lemma B.7. We omit the details.

B.3 Carleson embedding theorems with descendants

It is proved by Wick in an unpublished note [Wic] recently that there holds the following Carleson embedding theorem for descendants:

Theorem B.12. *Suppose that $\{\alpha_I\}_{I\in\mathcal{D}}$ is a sequence of positive numbers indexed by the dyadic cubes in $\mathcal{D}$ such that*

$$\sum_{I:I\subset K}\alpha_I \leq C|K|, \quad \forall K\in\mathcal{D}.$$

Then for $k\geq 1$,

$$\sum_{J\in\mathcal{D}}\alpha_J \sum_{I:I\subset J}^{(k)} \langle f\rangle_I^2 \lesssim Ck^2 2^{kd}\|f\|_{L^2}^2, \quad \forall f\in L^2.$$

He also shows that the constant 2^{kd} in the above is sharp by constructing a counterexample. In this section, we first establish an upgraded version of Theorem B.12 with k^2 replaced by k, and then extend it to the multi-parameter setting. However, it is still unknown whether this estimate in terms of the constant k is sharp.

Theorem B.13. *Suppose that $\{\alpha_I\}_{I\in\mathcal{D}}$ is a sequence of positive numbers indexed by the dyadic cubes in $\mathcal{D}$ such that*

$$\sum_{I:I\subset K}\alpha_I \leq C|K|, \quad \forall K\in\mathcal{D}. \tag{B.14}$$

Then for any $k\geq 1$,

$$\sum_{J\in\mathcal{D}}\alpha_J \sum_{I:I\subset J}^{(k)} \langle f\rangle_I^2 \lesssim k2^{kd}\|f\|_{L^2}^2, \quad \forall f\in L^2.$$

Proof. We show that this is equivalent to the L^2 case of Lemma B.5. To see that Lemma

B.5 is implied by Theorem B.13, calculate

$$\begin{aligned}
\|B_k(b,f)\|_{L^2}^2 &= \left\| \sum_I \beta_I \langle b, h_{I^{(k)}}\rangle \langle f, h_I^1\rangle h_I |I^{(k)}|^{-1/2} \right\|_{L^2}^2 \\
&= \sum_I \beta_I^2 |\langle b, h_{I^{(k)}}\rangle|^2 |\langle f, h_I^1\rangle|^2 |I^{(k)}|^{-1} \\
&\leq \sum_I |\langle b, h_{I^{(k)}}\rangle|^2 \langle f\rangle_I^2 |I||I^{(k)}|^{-1} \\
&= 2^{-kd} \sum_J |\langle b, h_J\rangle|^2 \sum_{I:I\subset J}^{(k)} \langle f\rangle_I^2.
\end{aligned}$$

Defining sequence $\{\alpha_J := |\langle b, h_J\rangle|^2\}$, one can easily see that condition (B.14) is satisfied with $C = \|b\|_{\mathrm{BMO}}^2$, hence Lemma B.5 is proved.

On the other hand, to derive Theorem B.13 from Lemma B.5, denote B_N as the ball in $\mathbb{R}^d$ centered at the origin with radius N, where N is a large number. For any given sequence $\{\alpha_J\}$ satisfying condition (B.14), define for any fixed N a L^2 function

$$b_N := \sum_{J:J\subset B_N} \alpha_J^{1/2} h_J.$$

Then it is easy to see that

$$\|b_N\|_{\mathrm{BMO}} = \sup_K \left(|K|^{-1} \sum_{J:J\subset K\cap B_N} \alpha_J \right)^{1/2} \leq C^{1/2}.$$

Now consider $B_k(b_N, f)$ with all the coefficients $\beta_I = 1$. From the calculation that

$$\|B_k(b_N,f)\|_{L^2}^2 = 2^{-kd} \sum_{J:J\subset B_N} \alpha_J \sum_{I:I\subset J}^{(k)} \langle f\rangle_I^2,$$

and the boundedness of $B_k(b_N, f)$, one deduces that

$$\sum_{J:J\subset B_N} \alpha_J \sum_{I:I\subset J}^{(k)} \langle f\rangle_I^2 \lesssim k 2^{kd} C \|f\|_{L^2}^2.$$

Theorem B.13 is thus proved by letting $N \to \infty$. $\square$

Theorem B.13 can be extended to the multi-parameter setting. For the exact same reason, the following bi-parameter version of Theorem B.13 is equivalent to the L^2 case of Lemma B.7, similarly for the case with arbitrarily many parameters.

Theorem B.15. *Suppose that $\{\alpha_R\}_{R \in \mathcal{D}_n \times \mathcal{D}_m}$ is a sequence of positive numbers indexed by the dyadic rectangles R in grids $\mathcal{D}_n \times \mathcal{D}_m$ such that*

$$\sum_{R \subset \Omega} \alpha_R \leq C|\Omega|, \quad \forall \text{ open } \Omega \in \mathbb{R}^n \times \mathbb{R}^m \text{ with finite measure.}$$

Then for $k \geq 0, l \geq 0$,

$$\sum_{R = J_1 \times J_2} \alpha_R \sum_{I_1 : I_1 \subset J_1}^{(k)} \sum_{I_2 : I_2 \subset J_2}^{(l)} \langle f \rangle_{I_1 \times I_2}^2 \lesssim C \max(k, 1) \max(l, 1) 2^{kn} 2^{lm} \|f\|_{L^2}^2, \quad \forall f \in L^2.$$

Bibliography

[Car74] Lennart Carleson, *A counterexample for measures bounded on H^p spaces for the bidisk*, Mittag-Leffler Rep. No. 7, Inst. Mittag-Leffler (1974).

[CF80] Sun-Yung Chang and Robert Fefferman, *A continuous version of duality of H^1 with BMO on the bidisc*, Ann. of Math. (2) **112** (1980), no. 1, 179–201. MR 0584078 (82a:32009)

[CF82] ______, *The Calderón-Zygmund decomposition on product domains*, Amer. J. Math. **104** (1982), 445–468. MR 0658542 (84a:42028)

[Cha79] Sun-Yung Chang, *Carleson measure on the bi-disc*, Ann. of Math. (2) **109** (1979), 613–620. MR 0534766 (80j:32009)

[CLMS93] Ronald R. Coifman, Pierre-Louis Lions, Yves Meyer, and Stephen Semmes, *Compensated compactness and Hardy spaces*, J. Math. Pures Appl. (9) **72** (1993), no. 3, 247–286. MR 1225511 (95d:46033)

[CRW76] Ronald R. Coifman, Richard Rochberg, and Guido Weiss, *Factorization theorems for Hardy spaces in several variables*, Ann. of Math. (2) **103** (1976), no. 3, 611–635. MR 0412721 (54 #843)

[CZ52] Alberto P. Calderón and Antoni Zygmund, *On the existence of certain singular integrals*, Acta Math. **88** (1952), 85–139. MR 0052553 (14,637f)

[DJ84] Guy David and Jean-Lin Journé, *A boundedness criterion for generalized Calderón-Zygmund operators*, Ann. of Math. (2) **120** (1984), no. 2, 371–397. MR 763911 (85k:42041)

[DP14] Laurent Dalenc and Stefanie Petermichl, *A lower bound criterion for iterated commutators*, J. Funct. Anal. **266** (2014), no. 8, 5300–5320. MR 3177337

[Fef81] Robert Fefferman, *Singular integrals on product domains*, Bull. Amer. Math. Soc. (N.S.) **4** (1981), no. 2, 195–201. MR 0598687 (83i:42014)

[FL02] Sarah Ferguson and Michael Lacey, *A characterization of product BMO by commutators*, Acta Math. **189** (2002), no. 2, 143–160. MR 1961195 (2004e:42026)

[FS72] Charles Fefferman and Elias M. Stein, *H^p spaces of several variables*, Acta Math. **2** (1972), 137–193. MR 0447953 (56 #6263)

[FS82] Robert Fefferman and Elias M. Stein, *Singular integrals on product spaces*, Adv. Math. **45** (1982), no. 2, 117–143. MR 0664621 (84d:42023)

[FS00] Sarah Ferguson and Cora Sadosky, *Characterizations of bounded mean oscillation on the polydisk in terms of Hankel operators and Carleson measures*, J. Anal. Math. **81** (2000), 239–267. MR 1785283 (2001h:47040)

[GdlH15] Ana Grau de la Herran, *Comparison of $T1$ conditions for multiparameter operators*, Proc. Amer. Math. Soc. (to appear) (2015).

[GJ82] John B. Garnett and Peter W. Jones, *BMO from dyadic BMO*, Pacific J. Math. **99** (1982), no. 2, 351–371. MR 0658065 (85d:42021)

[Hyt11] Tuomas Hytönen, *Representation of singular integrals by dyadic operators, and the A_2 theorem*, preprint arXiv:1108.5119 (2011).

[Hyt12] ______, *The sharp weighted bound for general Calderón-Zygmund operators*, Ann. of Math. **175** (2012), no. 3, 1473–1506. MR 2912709

[Jou85] Jean-Lin Journé, *Calderón-Zygmund operators on product spaces*, Rev. Mat. Iberoamericana **1** (1985), no. 3, 55–91. MR 0836284 (88d:42028)

[Jou86] ______, *A covering lemma for product spaces*, Proc. Amer. Math. Soc. **96** (1986), no. 4, 593–598. MR 0826486 (87g:42028)

[LPPW09] Michael Lacey, Stefanie Petermichl, Jill Pipher, and Brett Wick, *Multiparameter Riesz commutators*, Amer. J. Math. **131** (2009), no. 3, 731–769. MR 2530853 (2010d:42021)

[LPPW10] ______, *Iterated Riesz commutators: a simple proof of boundedness*, Contemp. Math. **505** (2010), 171–178. MR 2664567 (2011g: 42039)

[LPPW12] ______, *Multi-parameter Div-Curl lemmas*, Bull. London Math. Soc. **10** (2012), 1123–1131. MR 3007645

[LT09] Michael Lacey and Erin Terwilleger, *Hankel operators in several complex variables and product BMO*, Houston J. Math. **35** (2009), no. 1, 159–183. MR 2492875 (2010c:47071)

[Mar12] Henri Martikainen, *Representation of bi-parameter singular integrals by dyadic operators*, Adv. Math. **229** (2012), no. 3, 1734–1761. MR 2871155

[MO15] Henri Martikainen and Tuomas Orponen, *Some obstacles in characterising the boundedness of bi-parameter singular integrals*, Math. Z. (2015), 1–11. MR 3448394

[MPTT04] Camil Muscalu, Jill Pipher, Terence Tao, and Christoph Thiele, *Bi-parameter paraproducts*, Acta Math. **193** (2004), no. 2, 269–296. MR 2134868 (2005m:42028)

[MPTT06] ______, *Multi-parameter paraproducts*, Rev. Mat. Iberoam. **22** (2006), no. 3, 963–976. MR 2320408 (2008b:42037)

[Neh57] Zeev Nehari, *On bounded bilinear forms*, Ann. of Math. (2) **65** (1957), 153–162. MR 0082945 (18,633f)

[Ou15] Yumeng Ou, *A $T(b)$ theorem on product spaces*, Trans. Amer. Math. Soc. **367** (2015), 6159–6197. MR 3356933

[Pet00] Stefanie Petermichl, *Dyadic shifts and a logarithmic estimate for Hankel operators with matrix symbol*, Comptes Rendus Acad. Sci. Paris **1** (2000), no. 1, 455–460. MR 1756958 (2000m:42016)

[PS82] Duong H. Phong and Elias M. Stein, *Some further classes of pseudo-differential and singular integral operators arising in boundary value problems. I. Composition of operators*, Amer. J. Math. **104** (1982), no. 1, 141–172. MR 0648484 (84j:35163)

[PTV02] Stefanie Petermichl, Sergei Treil, and Alexander Volberg, *Why the Riesz transforms are averages of the dyadic shifts?*, Publ. Mat. **46** (2002), no. 2, 209–228. MR 1964822 (2003m:42028)

[PV13] Sandra Pott and Paco Villarroya, *A $T(1)$-theorem on product spaces*, preprint arXiv:1105.2516 (2013).

[PW08] Jill Pipher and Lesley A. Ward, *BMO from dyadic BMO on the bidisc*, J. Lond. Math. Soc. (2) **77** (2008), no. 2, 524–544. MR 2400405 (2009f:42024)

[RdF84] José L. Rubio de Francia, *Factorization theory and A_p weights*, Amer. J. Math. **106** (1984), no. 3, 533–547. MR 0745140 (86a:47028a)

[Tre09] Sergei Treil, *H^1 and dyadic H^1*, Linear and complex analysis, Amer. Math. Soc. Transl. Ser. 2, vol. 226, Amer. Math. Soc., Providence, RI (2009), 179–193. MR 2500519 (2010g:42050)

[Uch82] Akihito Uchiyama, *A constructive proof of the Fefferman Stein decomposition of $BMO(\mathbb{R}^n)$*, Acta Math. **148** (1982), 215–241. MR 0666111 (84h:42037)

[Wic] Brett Wick, Personal communication.