A HEURISTIC PROOF OF THE DECOUPLING THEOREM IN 2D

JIANHUI LI AND PO-LAM YUNG

Contents

1. Introduction

The goal of this short note is to present an informal proof of the Bourgain-Demeter ℓ^2-decoupling theorem for the parabola in $\mathbb{R}^2$.

Let $Q = [-1, 1]$. Let $P = \{(\xi_1, \xi_2) : \xi_1^2 = \xi_2, \xi_1 \in Q\}$ be the truncated parabola. For each $g : Q \to \mathbb{C}$, we define an extension operator E by

$$Eg(x) = \int_Q g(\xi) e^{ix\cdot(\xi,\xi^2)} d\xi \quad , \quad x \in \mathbb{R}^2$$

For $R \geq 1$, let B_R be a spatial square of side length R, and for $p \geq 2$, we define the decoupling constant $\mathcal{D}_p(R)$ to be the best constant such that

$$\|Eg\|_{L^p_{\text{avg}}(B_R)} \leq \mathcal{D}_p(R) \left\| \|Eg_\kappa\|_{L^p_{\text{avg}}(B_R)} \right\|_{\ell^2(\kappa \in \mathcal{P}_{R^{-1/2}}(Q))},$$

where $\mathcal{P}_{R^{-1/2}}(Q)$ means partition of Q into disjoint intervals of length $R^{-1/2}$ and $g_\kappa = g\chi_\kappa$. (Technically the L^p_{avg} norm on the right hand side should come with a weight, but we will gloss over these, in hope of focusing on the ideas behind the subject.) Using some heuristics about wave packets, we will give an informal proof of the following theorem:

Theorem 1. *If $2 \leq p \leq 6$, then*

$$\mathcal{D}_p(R) \lesssim_\varepsilon R^\varepsilon$$

for every $\varepsilon > 0$ and $R \geq 1$.

Our emphasis will be in the range where $4 < p \le 6$, because when $2 \le p \le 4$ the above theorem follows readily from the stronger, and more classical, square function estimate. But for completeness, we still present a unified treatment for the above two ranges of p. The exposition below is based heavily on the study guide [3]. We would like to thank Ciprian Demeter, Shaoming Guo, Larry Guth, Diogo Olivera e Silva, Lillian Pierce, Hong Wang and Ruixiang Zhang for kindly sharing their insights on the subject.

2. Wave packet heuristics

In this section, we introduce wave packets, which help us understand how a function of compact Fourier support behaves in physical space. Precisely, let f be a function in $\mathbb{R}^2$ whose Fourier transform is supported on $[0,1]^2$. We have the following lemma:

Lemma 2. *f is locally constant in every squares $B = B_1$ of side length 1 in the following sense:*

$$\|f\|_{L^\infty(B)} \lesssim \|f\|_{L^1_{\text{avg}}(B)}$$

In particular, for any $1 \le p, q \le \infty$, we have the following:

$$\|f\|_{L^q_{\text{avg}}(B)} \sim \|f\|_{L^p_{\text{avg}}(B)}$$

(Again, technically, the norms used in the lemma should come with a weight which is of rapid decay. However, we pretend that the norms are on B as heuristics.)

To prove the above lemma, we pick a Schwartz function ψ, whose Fourier transform equals 1 on $[0,1]^2$. Then $\hat{f} = \hat{f}\hat{\psi}$ implies $f = f * \psi$ and hence for each $x \in B$,

$$\begin{aligned}|f(x)| &= |\int_{\mathbb{R}^2} f(y)\psi(x-y)dy| \\ &\le \int_{\mathbb{R}^2} |f(y)||\psi(x-y)|dy \\ &\lesssim \int_{\mathbb{R}^2} \frac{|f(y)|}{(1+|x-y|)^{10}}dy \\ &\le \int_{\mathbb{R}^2} \frac{|f(y)|}{(1+\text{dist}(y,B))^{10}}dy\end{aligned}$$

where we have used the rapid decay property of Schwartz function in the second-to-last line. Technically, we only have the above estimate. However, the power 10 can be replaced by any sufficiently large constants so that we can forget about the rapid decay Schwartz tail outside B. Hence, heuristically

$$\|f\|_{L^\infty(B)} \lesssim \int_B \frac{|f(y)|}{(1+\text{dist}(y,B))^{10}}dy = \|f\|_{L^1_{\text{avg}}(B)}.$$

Now, since B is of compact support (or technically with weight that can selected to have total measure 1), by Hölder's inequality, for each $1 \le p \le \infty$,

$$\|f\|_{L^1_{\text{avg}}(B)} \le \|f\|_{L^p_{\text{avg}}(B)} \le \|f\|_{L^\infty(B)}.$$

Combining with above, we have for each $1 \le p, q \le \infty$,

$$\|f\|_{L^q_{\mathrm{avg}}(B)} \sim \|f\|_{L^p_{\mathrm{avg}}(B)}.$$

We remark that only constant functions have the same L^p_{avg} norms for all p. Hence, roughly speaking, f is locally constant in every square of side length 1.

By rotating the axes, dilating each side of the square, the fact that Fourier transform of $f(r_1x_1, r_2x_2)$ is $r_1r_2\hat{f}(r_1^{-1}\xi_1, r_2^{-1}\xi_2)$, and that modulating a function translates the Fourier support without changing the norm of f in physical space, we have the following generalization of the above lemma about wave packets.

Proposition 3. *Suppose that f has Fourier support on a rectangle $T^* \subset \mathbb{R}^2$. Then f is locally constant in every dual rectangle T to T^*. (Here if T^* has sides parallel to directions $\vec{v}_1$ and $\vec{v}_2$, and if r_i is the length of the sides parallel to $\vec{v}_i$, for $i = 1, 2$, then a dual rectangle T to T^* is a rectangle with sides parallel to $\vec{v}_1$ and $\vec{v}_2$, with r_i^{-1} being the length of the sides parallel to $\vec{v}_i$ for $i = 1, 2$.)*

Let α be an interval in $Q = [-1, 1]$. We decompose $g_\alpha = g\chi_\alpha = \sum_{\kappa \in \mathcal{P}_\delta(\alpha)} g_\kappa$, where $\mathcal{P}_\delta(\alpha)$ means the partition of α into disjoint intervals of length δ and $g_\kappa = g\chi_\kappa$ as before. We will use these notations throughout the note.

For each κ of length δ, we cover the part of parabola $\{\xi : \xi_1^2 = \xi_2, \xi_1 \in \kappa\}$ by a rectangle of size $\delta \times \delta^2$, which is the smallest possible rectangle to cover this part of parabola on which $\widehat{Eg_\kappa}$ is supported. By Proposition 3, Eg_κ is locally constant on the dual rectangles T of sizes $\delta^{-1} \times \delta^{-2}$, of which the long sides are parallel to the normal direction of parabola at $(c(\kappa), c(\kappa)^2)$. (We denote the center of κ by $c(\kappa)$.) We tile $B_{\delta^{-2}}$ by such rectangles T and denote the set of rectangles in the tiling by $\mathbb{T}(\kappa)$. Hence morally speaking, we have

$$Eg_\kappa = \sum_{T \in \mathbb{T}(\kappa)} c_T \chi_T$$

for some constants $c_T \sim \|Eg_\kappa\|_{L^p_{\mathrm{avg}}(T)}$.

Now, we explain the local L^2 orthogonality of different wave packets.

Proposition 4. *Let κ_1, κ_2 be two different intervals in α which are of distance at least 2δ. Then Eg_{κ_1} and Eg_{κ_2} are orthogonal on every square B of side length at least δ^{-1}. (Technically, the L^2 norm on each square should be replaced by a rapid decay weight.)*

Let φ be a positive Schwartz function which equals to 1 on B of side length at least B and whose Fourier transform is supported on B_δ centred at 0. For the same reason as before, we forget about the Schwartz tail of φ and assume the integral over B contains some weight outside the ball. Then we have the following:

$$
\begin{aligned}
\int_B Eg_{\kappa_1}\overline{Eg_{\kappa_2}} &\approx \int_{\mathbb{R}^2} (Eg_{\kappa_1}\varphi)(\overline{Eg_{\kappa_2}\varphi}) \\
&= \int_{\mathbb{R}^2} \mathcal{F}(Eg_{\kappa_1}\varphi)\overline{\mathcal{F}(Eg_{\kappa_2}\varphi)} \\
&= \int_{\mathbb{R}^2} (\widehat{Eg_{\kappa_1}} * \hat{\varphi})\overline{(\widehat{Eg_{\kappa_2}} * \hat{\varphi})}
\end{aligned}
$$

where we have used Plancherel theorem in the second-to-last line. Since $\widehat{Eg_{\kappa_1}}$ and $\widehat{Eg_{\kappa_2}}$ are supported on two different arcs on the parabola of distance at least 2δ and $\hat{\varphi}$ is supported on B_δ, $(\widehat{Eg_{\kappa_2}} * \hat{\varphi})$ and $(\widehat{Eg_{\kappa_2}} * \hat{\varphi})$ have disjoint support. Hence, $\int_B Eg_{\kappa_1}\overline{Eg_{\kappa_2}} \approx 0$ and they are orthogonal on B.

3. L^2 DECOUPLING

In this section, we give a proof of Theorem 1 for $p = 2$ based on the above wave packet heuristics.

Let $g = \sum_{\kappa\in\mathcal{P}_{R^{-1/2}}(Q)} g_\kappa$ as before. Take $\delta = R^{-1/2}$ in Proposition 4 and $B = B_R$. We arrange κ according to the the value of its center and name it by κ_i, $1 \le i \lesssim R^{1/2}$. Then Eg_{κ_i} and $Eg_{\kappa_{i+j}}$ are (morally) orthogonal to each other on B_R when $j \ge 3$ and therefore by almost orthogonality principle, we have

$$
\begin{aligned}
\|Eg\|^2_{L^2_{\text{avg}}(B_R)} &= \left\|\sum_\kappa Eg_\kappa\right\|^2_{L^2_{\text{avg}}(B_R)} \\
&= \sum_{\kappa_1}\sum_{\kappa_2}\langle Eg_{\kappa_1}, Eg_{\kappa_2}\rangle \\
&\approx \sum_{\kappa_1}\sum_{j=-3}^{3}\langle Eg_{\kappa_1}, Eg_{\kappa_1+j}\rangle \\
&\le \sum_{j=-3}^{3}\left(\sum_\kappa \|Eg_\kappa\|^2_{L^2_{\text{avg}}(B_R)}\right)^{\frac{1}{2}}\left(\sum_\kappa \|Eg_{\kappa+j}\|^2_{L^2_{\text{avg}}(B_R)}\right)^{\frac{1}{2}} \\
&= 7\left\|\|Eg_\kappa\|_{L^2_{\text{avg}}(B_R)}\right\|^2_{\ell^2(\kappa\in\mathcal{P}_{R^{-1/2}}(Q))}
\end{aligned}
$$

which is Theorem 1 without any ε loss.

4. A bootstrap argument

We now give an outline to an informal proof of Theorem 1 when $2 < p \le 6$. Cauchy-Schwarz gives $\mathcal{D}_p(R) \le R^{1/4}$ for all $R \ge 1$. Theorem 1 then follows easily from the following proposition:

Proposition 5. *Suppose $\eta > 0$ is such that $\mathcal{D}_p(R) \le C_\eta R^\eta$ for all $R \ge 1$.*

(a) If $2 < p \le 4$, then $\mathcal{D}_p(R) \lesssim_\eta R^{\eta/2}$ for all $R \ge 1$.

(b) If $4 < p < 6$, then there exist a (small but positive) constant a, and a (large positive) constant A, both depending only on p, such that $\mathcal{D}_p(R) \lesssim_\eta R^{\eta - a\eta^{-A}}$ for all $R \ge 1$.

(c) If $p = 6$, then there exist a (small but positive) constant a, and a (large positive) constant A, such that $\mathcal{D}_p(R) \lesssim_\eta R^{\eta - ae^{-A/\eta}}$ for all $R \ge 1$.

5. A bilinear reduction

In light of the discussion from the previous section, from now on, we fix $2 < p \le 6$, and fix some $\eta > 0$ that satisfies the hypothesis of Proposition 5. Let K be a large constant to be determined (that depends only on ε). We denote by $\operatorname{geom}_{j=1,2} x_j := \sqrt{x_1 x_2}$ the geometric average of non-negative real numbers x_1 and x_2.

Lemma 6.

$$\|Eg\|_{L^p_{\text{avg}}(B_R)} \le C \max_{\alpha \in \mathcal{P}_{K^{-1}}(Q)} \|Eg_\alpha\|_{L^p_{\text{avg}}(B_R)} + K^2 \max_{\substack{\alpha_1,\alpha_2 \in \mathcal{P}_{K^{-1}}(Q) \\ \textit{transverse}}} \Big\| \operatorname*{geom}_{j=1,2} |Eg_{\alpha_j}| \Big\|_{L^p_{\text{avg}}(B_R)}$$

where α_1, α_2 are called transverse if they are of distance at least $\frac{100}{K}$. In other words, there are at least 100 α between α_1, α_2.

By a standard partition of B_R into smaller squares and the Minkowski inequality, it suffices to prove Lemma 6 where we replace the spatial square B_R with any smaller squares. Also, as in Proposition 3, Eg_α is locally constant in every dual rectangle T^* of size $K \times K^2$. In particular, it is locally constant in every square of side length K. This leads us to prove Lemma 6 with B_R replaced by B_K, any spatial square of side length K.

We denote c_α to be $\|Eg_\alpha\|_{L^p_{\text{avg}}(B_K)}$, which is (morally) absolute value of Eg_α on B_K, in the view of the locally constant property. We then separate our situation into two cases according to the contributions c_α.

If there are at most 100 $\alpha_0 \in \mathcal{P}_{K^{-1}}(Q)$ such that $c_{\alpha_0} > K^{-1} \max_{\alpha \in \mathcal{P}_{K^{-1}}(Q)} c_\alpha$, then

$$\begin{aligned} \|Eg\|_{L^p_{\text{avg}}(B_K)} &\le \sum_{\alpha \in \mathcal{P}_{K^{-1}}(Q)} c_\alpha \\ &\le (K - 100)K^{-1} \max_{\alpha \in \mathcal{P}_{K^{-1}}(Q)} c_\alpha + 100 \max_{\alpha \in \mathcal{P}_{K^{-1}}(Q)} c_\alpha \\ &= 101 \max_{\alpha \in \mathcal{P}_{K^{-1}}(Q)} \|Eg_\alpha\|_{L^p_{\text{avg}}(B_R)} \end{aligned}$$

If there exist $\alpha_1, \alpha_2 \in \mathcal{P}_{K^{-1}}(Q)$ such that $c_{\alpha_1}, c_{\alpha_2} > K^{-1} \max_{\alpha \in \mathcal{P}_{K^{-1}}(Q)} c_\alpha$ and there is at least 100 α between α_1, α_2, then

$$\|Eg\|_{L^p_{\mathrm{avg}}(B_K)} \leq \sum_{\alpha \in \mathcal{P}_{K^{-1}}(Q)} c_\alpha \leq K \max_{\alpha \in \mathcal{P}_{K^{-1}}(Q)} c_\alpha \leq K^2 \operatorname*{geom}_{j=1,2} c_{\alpha_j} = K^2 \left\| \operatorname*{geom}_{j=1,2} |Eg_{\alpha_j}| \right\|_{L^p_{\mathrm{avg}}(B_K)}.$$

Lemma 6 follows immediately after taking maximum over all possible transverse pairs of α_1, α_2.

We note that the decoupling constant is rescaling invariance. Therefore, we can perform a parabolic rescaling from each α to the original parabola, and apply decoupling inequality to partition the rescaled α into $KR^{-1/2}$ pieces. By rescaling back to the original α of length K^{-1}, each $KR^{-1/2}$ pieces represent a κ of length $R^{-1/2}$ in $\mathcal{P}_{R^{-1/2}}(\alpha)$. Hence, we have the following decoupling inequality:

$$\|Eg_\alpha\|_{L^p_{\mathrm{avg}}(B_R)} \leq \mathcal{D}_p(R/K^2) \left\| \|Eg_\kappa\|_{L^p_{\mathrm{avg}}(B_R)} \right\|_{\ell^2(\kappa \in \mathcal{P}_{R^{-1/2}}(\alpha))}.$$

By estimating the maximum by an ℓ^2 norm, we obtain

$$\begin{aligned} \max_{\alpha \in \mathcal{P}_{K^{-1}}(Q)} \|Eg_\alpha\|_{L^p_{\mathrm{avg}}(B_R)} &\leq \mathcal{D}_p(R/K^2) \left\| \left\| \|Eg_\kappa\|_{L^p_{\mathrm{avg}}(B_R)} \right\|_{\ell^2(\kappa \in \mathcal{P}_{R^{-1/2}}(\alpha))} \right\|_{\ell^2(\alpha \in \mathcal{P}_{K^{-1}}(Q))} \\ &= \mathcal{D}_p(R/K^2) \left\| \|Eg_\kappa\|_{L^p_{\mathrm{avg}}(B_R)} \right\|_{\ell^2(\kappa \in \mathcal{P}_{R^{-1/2}}(Q))}. \end{aligned}$$

We define the bilinear decoupling constant $\mathcal{B}_p(R)$ to be the best constant so that

$$\left\| \operatorname*{geom}_{j=1,2} |Eg_{\alpha_j}| \right\|_{L^p_{\mathrm{avg}}(B_R)} \leq \mathcal{B}_p(R) \operatorname*{geom}_{j=1,2} \left\| \|Eg_{\kappa_j}\|_{L^p_{\mathrm{avg}}(B_R)} \right\|_{\ell^2(\kappa_j \in \mathcal{P}_{R^{-1/2}}(\alpha_j))}$$

for all $\alpha_1, \alpha_2 \in \mathcal{P}_{K^{-1}}(Q)$ that are transverse. By Young's inequality, the right hand side is bounded by

$$\mathcal{B}_p(R) (\sum_{j=1,2} \left\| \|Eg_{\kappa_j}\|_{L^p_{\mathrm{avg}}(B_R)} \right\|^2_{\ell^2(\kappa_j \in \mathcal{P}_{R^{-1/2}}(\alpha_j))})^{1/2}.$$

We add the remaining terms to form an ℓ^2 norm over all $\alpha \in \mathcal{P}_{K^{-1}}(Q)$, which gives

$$\begin{aligned} \max_{\substack{\alpha_1, \alpha_2 \in \mathcal{P}_{K^{-1}}(Q) \\ \text{transverse}}} \left\| \operatorname*{geom}_{j=1,2} |Eg_{\alpha_j}| \right\|_{L^p_{\mathrm{avg}}(B_R)} &\leq \mathcal{B}_p(R) \left\| \left\| \|Eg_\kappa\|_{L^p_{\mathrm{avg}}(B_R)} \right\|_{\ell^2(\kappa \in \mathcal{P}_{R^{-1/2}}(\alpha))} \right\|_{\ell^2(\alpha \in \mathcal{P}_{K^{-1}}(Q))} \\ &= \mathcal{B}_p(R) \left\| \|Eg_\kappa\|_{L^p_{\mathrm{avg}}(B_R)} \right\|_{\ell^2(\kappa \in \mathcal{P}_{R^{-1/2}}(Q))} \end{aligned}$$

In summary, we have the following lemma.

Lemma 7.

$$\mathcal{D}_p(R) \leq C\, \mathcal{D}_p(R/K^2) + K^2\, \mathcal{B}_p(R).$$

Hence the goal now is to prove the following bilinear variant of Proposition 5:

Proposition 8. *Suppose $\eta > 0$ is such that $\mathcal{D}_p(R) \leq C_\eta R^\eta$ for all $R \geq 1$.*

(a) If $2 < p \le 4$, then $\mathcal{B}_p(R) \lesssim_\eta R^{\eta/2}$ for all $R \ge 1$.
(b) If $4 < p < 6$, then there exist a (small but positive) constant a, and a (large positive) constant A, both depending only on p, such that $\mathcal{B}_p(R) \lesssim_\eta R^{\eta - a\eta^{-A}}$ for all $R \ge 1$.
(c) If $p = 6$, then there exist a (small but positive) constant a, and a (large positive) constant A, such that $\mathcal{B}_p(R) \lesssim_\eta R^{\eta - ae^{-A/\eta}}$ for all $R \ge 1$.

We claim that Lemma 7 and Proposition 8 gives Proposition 5 by an iteration. We first apply Lemma 7. If the second term dominates the first term, we can conclude from Proposition 8 immediately. If the first term dominates, we can continue our iteration process:

$$\mathcal{D}_p(R) \le C\mathcal{D}_p(R/K^2) + K^2\mathcal{B}_p(R) \le 2C\mathcal{D}_p(R/K^2) \le (2C)(C\mathcal{D}_p(R/K^4) + K^2\mathcal{B}_p(R/K^2)).$$

Similarly, the iteration terminates when the second term dominates. We may assume that the first term always dominates. After m iterations, we get

$$\mathcal{D}_p(R) \le (2C)^m \mathcal{D}_p(R/K^{2m}).$$

Pick the smallest $m \in \mathbb{N}$ such that $R/K^{2m} < 10000$. Using $\mathcal{D}_p(10000) \le 10$, we have

$$\mathcal{D}_p(R) \le 10(2C)^{\frac{\log R}{2\log K}} \lesssim R^{\frac{\log C}{2\log K}} \le R^\varepsilon$$

if we pick K large enough depending on ε.

We remark that in higher dimensions, typically this multilinear reduction is carried out by induction on dimensions, as in the work of Bourgain and Guth [4].

6. Reduction to ball inflation

We now fix $2 < p \le 6$, and fix K as in the previous section. We also fix $\alpha_1, \alpha_2 \in \mathcal{P}_{K^{-1}}(Q)$ that are transverse, and fix some $\eta > 0$ that satisfies the hypothesis of Proposition 8. Let m be a large constant to be determined (that depends only on p and η, but not on R). Then

$$\left\| \operatorname*{geom}_{j=1,2} |Eg_{\alpha_j}| \right\|_{L^p_{\text{avg}}(B_R)} \le R^{\frac{1}{2^m}} \left\| \operatorname*{geom}_{j=1,2} \left\| Eg_{\kappa_j} \right\|_{\ell^2(\kappa_j \in \mathcal{P}_{R^{-\frac{1}{2^m}}}(\alpha_j))} \right\|_{L^p_{\text{avg}}(B_R)}$$

which by wave packet heuristics is approximately

$$\simeq R^{\frac{1}{2^m}} \left\| \operatorname*{geom}_{j=1,2} \left\| \left\| Eg_{\kappa_j} \right\|_{L^q_{\text{avg}}(\Delta)} \right\|_{\ell^2(\kappa_j \in \mathcal{P}_{R^{-\frac{1}{2^m}}}(\alpha_j))} \right\|_{\ell^p_{\text{avg}}(\Delta \in \mathcal{P}_{R^{\frac{1}{2^m}}}(B_R))}$$

for any exponent q. The loss here is only $R^{\frac{1}{2^m}}$, which is small since m will be chosen to be large. We need to estimate what remains of the right hand side, and to this end it will be convenient to introduce some notations.

For $\delta \in (0, 1]$, $r \ge 1$, let

$$M^{p,q}(\delta, r) = \left\| \operatorname*{geom}_{j=1,2} \left\| \left\| Eg_{\kappa_j} \right\|_{L^q_{\text{avg}}(\Delta)} \right\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))} \right\|_{\ell^p_{\text{avg}}(\Delta \in \mathcal{P}_r(B_R))}.$$

Then the above shows that

$$\left\| \operatorname*{geom}_{j=1,2} |Eg_{\alpha_j}| \right\|_{L^p_{\mathrm{avg}}(B_R)} \leq R^{\frac{1}{2^m}} M^{p,q}(R^{-\frac{1}{2^m}}, R^{\frac{1}{2^m}}) \tag{1}$$

for any exponent q. The key now is the following ball inflation lemma:

Lemma 9. *Suppose $4 \leq p \leq 6$, and $q = p/2$. If $\delta \geq R^{-1/2}$, then*

$$M^{p,q}(\delta, \delta^{-1}) \lesssim_\varepsilon R^\varepsilon M^{p,q}(\delta, \delta^{-2}).$$

Indeed, as we will see later, when $p = 4$ we can obtain the conclusion of the lemma without the R^ε loss.

Assuming the lemma for the moment, it is easy to conclude the proof of the Proposition 8 about bilinear decoupling. Indeed, suppose first $2 \leq p \leq 4$. Then by Hölder's inequality and Lemma 9 with $p = 4$, we have

$$M^{p,2}(\delta, \delta^{-1}) \leq M^{4,2}(\delta, \delta^{-1}) \lesssim M^{4,2}(\delta, \delta^{-2}).$$

If $\delta = R^{-1/2}$, then

$$M^{4,2}(\delta, \delta^{-2}) = M^{2,2}(\delta, \delta^{-2}) \lesssim M^{2,2}(\delta^2, \delta^{-2}) = M^{p,2}(\delta^2, \delta^{-2})$$

where we used L^2 decoupling in the above inequality. If $\delta \geq R^{-1/2}$, i.e. if $R \geq \delta^{-2}$, then by partitioning B_R into a disjoint union of squares Δ of side lengths δ^{-2}, applying the above inequality to each Δ, and summing the resulting estimates, one still obtains that

$$M^{p,2}(\delta, \delta^{-1}) \leq C M^{p,2}(\delta^2, \delta^{-2}).$$

This is now good for iteration: continuing from (1), we have

$$\begin{aligned}
\left\| \operatorname*{geom}_{j=1,2} |Eg_{\alpha_j}| \right\|_{L^p_{\mathrm{avg}}(B_R)} &\leq R^{\frac{1}{2^m}} M^{p,2}(R^{-\frac{1}{2^m}}, R^{\frac{1}{2^m}}) \\
&\leq C R^{\frac{1}{2^m}} M^{p,2}(R^{-\frac{1}{2^{m-1}}}, R^{\frac{1}{2^{m-1}}}) \\
&\vdots \\
&\leq C^m R^{\frac{1}{2^m}} M^{p,2}(R^{-\frac{1}{2}}, R^{\frac{1}{2}}) \\
&= C^m R^{\frac{1}{2^m}} \operatorname*{geom}_{j=1,2} \left\| \left\| Eg_{\kappa_j} \right\|_{L^p_{\mathrm{avg}}(B_R)} \right\|_{\ell^2(\kappa_j \in \mathcal{P}_{R^{-1/2}}(\alpha_j))}
\end{aligned}$$

the last line following from Hölder and Minkowski. This shows

$$\mathcal{B}_p(R) \leq C^m R^{\frac{1}{2^m}},$$

which can be made $\lesssim_\eta R^{\eta/2}$ if $m = m(p, \eta)$ is chosen sufficiently large.

Next, suppose $4 < p \leq 6$. As in Lemma 9, let's write $q = p/2$. We define an exponent $\lambda = \lambda(p)$, such that

$$\frac{1}{q} = \frac{1 - \lambda}{2} + \frac{\lambda}{p}.$$

If $\delta \geq R^{-1/2}$, we claim that

$$M^{p,2}(\delta, \delta^{-1}) \lesssim_\varepsilon R^\varepsilon M^{p,2}(\delta^2, \delta^{-2})^{1-\lambda} M^{p,p}(\delta, \delta^{-2})^\lambda. \tag{2}$$

Indeed, $M^{p,2}(\delta, \delta^{-1}) \leq M^{p,q}(\delta, \delta^{-1}) \lesssim_\varepsilon R^\varepsilon M^{p,q}(\delta, \delta^{-2})$ by Hölder and Lemma 9. Also,

$$M^{p,q}(\delta, \delta^{-2}) \leq M^{p,2}(\delta, \delta^{-2})^{1-\lambda} M^{p,p}(\delta, \delta^{-2})^\lambda$$

by Hölder's inequality, and

$$M^{p,2}(\delta, \delta^{-2}) \lesssim M^{p,2}(\delta^2, \delta^2)$$

by L^2 decoupling. Together we have our claim (2).

We are now ready to iterate. By (1) and an $(m-1)$-fold iteration of (2), we have

$$\begin{aligned}
&\Big\| \operatorname*{geom}_{j=1,2} |Eg_{\alpha_j}| \Big\|_{L^p_{\mathrm{avg}}(B_R)} \\
&\lesssim R^{\frac{1}{2^m}} M^{p,2}(R^{-\frac{1}{2^m}}, R^{\frac{1}{2^m}}) \\
&\lesssim_{\varepsilon'} R^{\varepsilon'} R^{\frac{1}{2^m}} M^{p,2}(R^{-\frac{1}{2^{m-1}}}, R^{\frac{1}{2^{m-1}}})^{1-\lambda} M^{p,p}(R^{-\frac{1}{2^m}}, R^{\frac{1}{2^{m-1}}})^\lambda \\
&\lesssim_{\varepsilon'} R^{2\varepsilon'} R^{\frac{1}{2^m}} M^{p,2}(R^{-\frac{1}{2^{m-2}}}, R^{\frac{1}{2^{m-2}}})^{(1-\lambda)^2} M^{p,p}(R^{-\frac{1}{2^m}}, R^{\frac{1}{2^{m-1}}})^\lambda M^{p,p}(R^{-\frac{1}{2^{m-1}}}, R^{\frac{1}{2^{m-2}}})^{\lambda(1-\lambda)} \\
&\lesssim \vdots \\
&\lesssim_{\varepsilon'} R^{(m-1)\varepsilon'} R^{\frac{1}{2^m}} M^{p,2}(R^{-\frac{1}{2}}, R^{\frac{1}{2}})^{(1-\lambda)^{m-1}} \prod_{j=0}^{m-2} M^{p,p}(R^{-\frac{1}{2^{m-j}}}, R^{\frac{1}{2^{m-j-1}}})^{\lambda(1-\lambda)^j}.
\end{aligned} \tag{3}$$

But by Hölder and Minkowski,

$$M^{p,2}(R^{-\frac{1}{2}}, R^{\frac{1}{2}}) \lesssim \operatorname*{geom}_{j=1,2} \Big\| \big\| Eg_{\kappa_j} \big\|_{L^p_{\mathrm{avg}}(B_R)} \Big\|_{\ell^2(\kappa_j \in \mathcal{P}_{R^{-1/2}}(\alpha_j))},$$

and by parabolic rescaling,

$$M^{p,p}(\delta, \delta^{-2}) \lesssim \mathcal{D}_p(R\delta^2) \operatorname*{geom}_{j=1,2} \Big\| \big\| Eg_{\kappa_j} \big\|_{L^p_{\mathrm{avg}}(B_R)} \Big\|_{\ell^2(\kappa_j \in \mathcal{P}_{R^{-1/2}}(\alpha_j))}$$

whenever $\delta \geq R^{-1/2}$, which in particular implies

$$M^{p,p}(R^{-\frac{1}{2^{m-j}}}, R^{\frac{1}{2^{m-j-1}}}) \lesssim \mathcal{D}_p(R^{1-\frac{1}{2^{m-j-1}}}) \operatorname*{geom}_{j=1,2} \Big\| \big\| Eg_{\kappa_j} \big\|_{L^p_{\mathrm{avg}}(B_R)} \Big\|_{\ell^2(\kappa_j \in \mathcal{P}_{R^{-1/2}}(\alpha_j))}.$$

for $0 \leq j \leq m-2$. It follows that

$$\begin{aligned}
\Big\| \operatorname*{geom}_{j=1,2} |Eg_{\alpha_j}| \Big\|_{L^p_{\mathrm{avg}}(B_R)} &\lesssim_{\varepsilon'} \operatorname*{geom}_{j=1,2} \Big\| \big\| Eg_{\kappa_j} \big\|_{L^p_{\mathrm{avg}}(B_R)} \Big\|_{\ell^2(\kappa_j \in \mathcal{P}_{R^{-1/2}}(\alpha_j))} R^{(m-1)\varepsilon'} R^{\frac{1}{2^m}} \prod_{j=0}^{m-2} \mathcal{D}_p(R^{1-\frac{1}{2^{m-j-1}}})^{\lambda(1-\lambda)^j} \\
&\lesssim_{\varepsilon',\eta} \operatorname*{geom}_{j=1,2} \Big\| \big\| Eg_{\kappa_j} \big\|_{L^p_{\mathrm{avg}}(B_R)} \Big\|_{\ell^2(\kappa_j \in \mathcal{P}_{R^{-1/2}}(\alpha_j))} R^{(m-1)\varepsilon'} R^{\frac{1}{2^m}} \prod_{j=0}^{m-2} (R^{1-\frac{1}{2^{m-j-1}}})^{\eta\lambda(1-\lambda)^j}.
\end{aligned}$$

Hence

$$\mathcal{B}_p(R) \lesssim_{\varepsilon',\eta} R^{(m-1)\varepsilon'} R^{\frac{1}{2^m}} \prod_{j=0}^{m-2} (R^{1-\frac{1}{2^{m-j-1}}})^{\eta\lambda(1-\lambda)^j}.$$

We now consider two cases, namely $4 < p < 6$, and $p = 6$. If $4 < p < 6$, then $\lambda \in (0, 1/2)$, and the above gives

$$\mathcal{B}_p(R) \lesssim_{\varepsilon',\eta} R^{(m-1)\varepsilon'} R^{\frac{1}{2^m}} R^{\eta\left(1-\frac{(1-\lambda)^m}{1-2\lambda}+2^{-m}\frac{2\lambda}{1-2\lambda}\right)} \tag{4}$$

because

$$\sum_{j=0}^{m-2} \lambda(1-\lambda)^j = 1-(1-\lambda)^{m-1}$$

and

$$\sum_{j=0}^{m-2} \frac{\lambda}{2^{m-1}}[2(1-\lambda)]^j = \frac{\lambda}{2^{m-1}} \frac{[2(1-\lambda)]^{m-1}-1}{1-2\lambda} = \frac{\lambda}{1-2\lambda}[(1-\lambda)^{m-1} - 2^{-(m-1)}].$$

We choose $m = m(p, \eta)$ be the smallest positive integer so that

$$2^{-m} + \eta 2^{-m} \frac{2\lambda}{1-2\lambda} \leq \frac{1}{2}\eta \frac{(1-\lambda)^m}{1-2\lambda}.$$

Then $m \lesssim_p -\log \eta$, so

$$R^{2^{-m}} R^{\eta\left(1-\frac{(1-\lambda)^m}{1-2\lambda}+2^{-m}\frac{2\lambda}{1-2\lambda}\right)} \leq R^{\eta\left(1-\frac{1}{2}\frac{(1-\lambda)^m}{1-2\lambda}\right)} \leq R^{\eta-2a\eta^{-A}}$$

for some constants $a > 0$ and $A > 0$, both depending only on p. We also choose $\varepsilon' = \varepsilon'(p, \eta)$ so small so that

$$(m-1)\varepsilon' < a\eta^{-A}.$$

Hence

$$\mathcal{B}_p(R) \lesssim_\eta R^{\eta - a\eta^{-A}}.$$

On the other hand, if $p = 6$, then $\lambda = 1/2$, so

$$\mathcal{B}_p(R) \lesssim_{\varepsilon',\eta} R^{(m-1)\varepsilon'} R^{\frac{1}{2^m}} R^{\eta\left(1-\frac{m+1}{2^m}\right)}$$

We choose $m = m(\eta)$ be the smallest positive integer so that $(m+1)\eta > 2$. Then $m \lesssim 1/\eta$, so

$$R^{\frac{1}{2^m}} R^{\eta\left(1-\frac{m+1}{2^m}\right)} \lesssim R^{\eta - 2ae^{-A/\eta}}$$

for some absolute constants $a > 0$, $A > 0$. We also choose $\varepsilon' = \varepsilon'(\eta)$ so small so that $(m-1)\varepsilon' < ae^{-A/\eta}$. Hence

$$\mathcal{B}_p(R) \lesssim_\eta R^{\eta - ae^{-A/\eta}},$$

as desired. This finishes the proof of Proposition 8.

7. Proof of the ball inflation lemma

It remains to prove the ball inflation Lemma 9. Without loss of generality, assume $\delta = R^{-1/2}$, i.e. $R = \delta^{-2}$. This is because if R is bigger, then we may partition B_R into subcubes of side length δ^{-2}, apply the estimate we are about to prove, and then sum over all such subcubes. Hence assume $\delta = R^{-1/2}$, in which case $M^{p,q}(\delta, \delta^{-2})$ becomes simply

$$\operatorname*{geom}_{j=1,2} \left\| \|Eg_{\kappa_j}\|_{L^q_{\text{avg}}(B_{\delta^{-2}})} \right\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))}.$$

By wave packet heuristics, $M^{p,q}(\delta, \delta^{-1})$ is basically

$$\left\| \operatorname*{geom}_{j=1,2} \|Eg_{\kappa_j}\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))} \right\|_{L^p_{\text{avg}}(B_{\delta^{-2}})}.$$

Hence we are reduced to showing that

$$\left\| \operatorname*{geom}_{j=1,2} \|Eg_{\kappa_j}\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))} \right\|_{L^p_{\text{avg}}(B_{\delta^{-2}})} \lesssim_\varepsilon \delta^{-\varepsilon} \operatorname*{geom}_{j=1,2} \left\| \|Eg_{\kappa_j}\|_{L^q_{\text{avg}}(B_{\delta^{-2}})} \right\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))}. \tag{5}$$

Note that we are taking L^p_{avg} norm on the left hand side, and L^q_{avg} norm on the right hand side. If $p \leq q$, then the above would be simply a consequence of Hölder's inequality and Minkowski. But now $q = p/2 < p$, so we have to proceed differently.

Recall now $4 \leq p \leq 6$. When $p = 4$, it is easy to prove (5), and one can even prove the desired conclusion without the $\delta^{-\varepsilon}$ loss. Indeed, for each $\kappa_j \in \mathcal{P}_\delta(\alpha_j)$, $j = 1, 2$, we write

$$Eg_{\kappa_j} = \sum_{T_j \in \mathbb{T}(\kappa_j)} c_{T_j} \chi_{T_j}$$

on $B_{\delta^{-2}}$ using our wave packet heuristics; here $\mathbb{T}(\kappa_j)$ is the family of dual rectangles of size $\delta^{-1} \times \delta^{-2}$ that tiles $B_{\delta^{-2}}$. Hence writing $\mathbb{T}_j = \bigcup_{\kappa_j \in \mathcal{P}_\delta(\alpha_j)} \mathbb{T}(\kappa_j)$, we have, for any $x \in B_{\delta^{-2}}$, that

$$\operatorname*{geom}_{j=1,2} \|Eg_{\kappa_j}(x)\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))} \lesssim \operatorname*{geom}_{j=1,2} \left(\sum_{T_j \in \mathbb{T}_j} |c_{T_j}|^2 \chi_{T_j}(x) \right)^{1/2} \tag{6}$$

If $p = 4$, then

$$\begin{aligned} \left\| \operatorname*{geom}_{j=1,2} \|Eg_{\kappa_j}\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))} \right\|_{L^p_{\text{avg}}(B_{\delta^{-2}})} &\lesssim \left(\frac{1}{|B_{\delta^{-2}}|} \int_{B_{\delta^{-2}}} \prod_{j=1,2} \sum_{T_j \in \mathbb{T}_j} |c_{T_j}|^2 \chi_{T_j}(x) dx \right)^{1/4} \\ &\leq \left(\sum_{T_1 \in \mathbb{T}_1} \sum_{T_2 \in \mathbb{T}_2} |c_{T_1}|^2 |c_{T_2}|^2 \frac{|T_1 \cap T_2|}{|B_{\delta^{-2}}|} \right)^{1/4}. \end{aligned}$$

But $|T_1 \cap T_2| \lesssim (\delta^{-1})^2 = \frac{|T_1||T_2|}{|B_{\delta^{-2}}|}$, since the tubes from $\mathbb{T}_1$ and $\mathbb{T}_2$ are transverse. Hence the above display equation is bounded by

$$\operatorname*{geom}_{j=1,2} \left\| \left(\sum_{T_j \in \mathbb{T}(\kappa_j)} |c_{T_j}|^2 \frac{|T_j|}{|B_{\delta^{-2}}|} \right)^{1/2} \right\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))} = \operatorname*{geom}_{j=1,2} \left\| \|Eg_{\kappa_j}\|_{L^2_{\text{avg}}(B_{\delta^{-2}})} \right\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))},$$

which is (5) since $q = 2$ when $p = 4$.

We now consider the case $4 < p \leq 6$. First we observe a trivial bound

$$\left\| \operatorname*{geom}_{j=1,2} \|Eg_{\kappa_j}\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))} \right\|_{L^p_{\text{avg}}(B_{\delta^{-2}})} \lesssim \delta^{-2/p} \operatorname*{geom}_{j=1,2} \left\| \|Eg_{\kappa_j}\|_{L^q_{\text{avg}}(B_{\delta^{-2}})} \right\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))}. \tag{7}$$

This is because we can apply Minkowski inequality to interchange the L^p_{avg} norm and the ℓ^2 norm on the left hand side, and observe that for $\kappa_j \in \mathcal{P}_\delta(\alpha_j)$,

$$\|Eg_{\kappa_j}\|_{L^p_{\text{avg}}(B_{\delta^{-2}})} = \left\| \|Eg_{\kappa_j}\|_{L^p_{\text{avg}}(\Delta)} \right\|_{\ell^p_{\text{avg}}(\Delta \in \mathcal{P}_{\delta^{-1}}(B_{\delta^{-2}}))} \lesssim \left\| \|Eg_{\kappa_j}\|_{L^q_{\text{avg}}(\Delta)} \right\|_{\ell^q_{\text{avg}}(\Delta \in \mathcal{P}_{\delta^{-1}}(B_{\delta^{-2}}))} \delta^{-2\left(\frac{1}{q} - \frac{1}{p}\right)};$$

in the latter we used the heuristic that Eg_{κ_j} is locally constant on any square Δ of side length δ^{-1}, and Hölder's inequality. Remembering that $q = p/2 > 2$ gives (7). For $j = 1, 2$, let $c_j = \max_{\kappa \in \mathcal{P}_\delta(\alpha_j)} \|Eg_\kappa\|_{L^q_{\text{avg}}(B_{\delta^{-2}})}$. Then on the left hand side of (5), those κ_j's for which $\|Eg_{\kappa_j}\|_{L^q_{\text{avg}}(B_{\delta^{-2}})} < \delta^{2/p+1} c_j$ contributes not more than what is allowed on the right hand side of (5). This shows that we only need to bound the left hand side of (5), where the ℓ^2 norms are only over those κ_j for which $\|Eg_{\kappa_j}\|_{L^q_{\text{avg}}(B_{\delta^{-2}})}$ is between $\delta^{2/p+1} c_j$ and c_j. This is only $\simeq \log \delta^{-1}$ many dyadic ranges of $\|Eg_{\kappa_j}\|_{L^q_{\text{avg}}(B_{\delta^{-2}})}$, so in proving (5), by dyadic pigeonholing, we may assume that for any $j = 1, 2$ and any $\kappa_j \in \mathcal{P}_\delta(\alpha_j)$, either $\|Eg_{\kappa_j}\|_{L^q_{\text{avg}}(B_{\delta^{-2}})} = 0$, or is comparable to a fixed non-zero constant A_j; let N_j be the number of κ_j that satisfy the latter. For $j = 1, 2$, by Hölder's inequality, we have

$$\|Eg_{\kappa_j}(x)\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))} \leq N_j^{\frac{1}{2} - \frac{1}{q}} \|Eg_{\kappa_j}(x)\|_{\ell^q(\kappa_j \in \mathcal{P}_\delta(\alpha_j))} \leq N_j^{\frac{1}{2} - \frac{1}{q}} \left(\sum_{T_j \in \mathbb{T}_j} |c_{T_j}|^q \chi_{T_j}(x) \right)^{1/q},$$

so

$$\operatorname*{geom}_{j=1,2} \|Eg_{\kappa_j}(x)\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))} \lesssim \operatorname*{geom}_{j=1,2} N_j^{\frac{1}{2} - \frac{1}{q}} \left(\sum_{T_j \in \mathbb{T}_j} |c_{T_j}|^q \chi_{T_j}(x) \right)^{1/q}.$$

Since $q = p/2$, we have $\frac{p}{q} = 2$, so we may take $L^p_{\text{avg}}(B_{\delta^{-2}})$ norm of both sides, and obtain, as before, that

$$\left\| \operatorname*{geom}_{j=1,2} \|Eg_{\kappa_j}(x)\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))} \right\|_{L^p_{\text{avg}}(B_{\delta^{-2}})} \lesssim \operatorname*{geom}_{j=1,2} N_j^{\frac{1}{2} - \frac{1}{q}} \left(\sum_{T_j \in \mathbb{T}_j} |c_{T_j}|^q \frac{|T_j|}{|B_{\delta^{-2}}|} \right)^{1/q}.$$

The latter is in turn bounded by

$$\operatorname*{geom}_{j=1,2} N_j^{\frac{1}{2}-\frac{1}{q}} \left(\sum_{\kappa_j \in \mathcal{P}_\delta(\alpha_j)} \|Eg_{\alpha_j}\|^q_{L^q_{\mathrm{avg}}(B_{\delta^{-2}})} \right)^{1/q} \simeq \operatorname*{geom}_{j=1,2} \left\| \|Eg_{\kappa_j}\|_{L^q_{\mathrm{avg}}(B_{\delta^{-2}})} \right\|_{\ell^2(\kappa_j \in \mathcal{P}_\delta(\alpha_j))},$$

as desired.

We remark that in the above argument, the key is the control on the area of the intersection of two $\delta^{-1} \times \delta^{-2}$ rectangles that are transverse to each other. In higher dimensions, such intersections will be more complicated to control, and one useful tool is the multilinear Kakeya inequality of Bennett, Carbery and Tao [1] (see also [5–7]), or more generally the multilinear perturbed Brascamp-Lieb inequality of Zhang [8].

References

[1] Jonathan Bennett, Anthony Carbery, and Terence Tao, *On the multilinear restriction and Kakeya conjectures*, Acta Math. **196** (2006), no. 2, 261–302.

[2] Jean Bourgain and Ciprian Demeter, *The proof of the l^2 decoupling conjecture*, Ann. of Math. (2) **182** (2015), no. 1, 351–389.

[3] ———, *A study guide for the l^2 decoupling theorem*, Chin. Ann. Math. Ser. B **38** (2017), no. 1, 173–200.

[4] Jean Bourgain and Larry Guth, *Bounds on oscillatory integral operators based on multilinear estimates*, Geom. Funct. Anal. **21** (2011), no. 6, 1239–1295.

[5] Anthony Carbery and Stefán Ingi Valdimarsson, *The endpoint multilinear Kakeya theorem via the Borsuk-Ulam theorem*, J. Funct. Anal. **264** (2013), no. 7, 1643–1663.

[6] Larry Guth, *The endpoint case of the Bennett-Carbery-Tao multilinear Kakeya conjecture*, Acta Math. **205** (2010), no. 2, 263–286.

[7] ———, *A short proof of the multilinear Kakeya inequality*, Math. Proc. Cambridge Philos. Soc. **158** (2015), no. 1, 147–153.

[8] Ruixiang Zhang, *The endpoint perturbed Brascamp-Lieb inequalities with examples*, Anal. PDE **11** (2018), no. 3, 555–581.