Calderón Commutator Estimate

Patrick Li

August 2025

Let $b \in \mathcal{S}(\mathbb{R})$, and consider the commutator operator $[b, H]$ given by

$$[b, H]f = b \cdot Hf - H(bf),$$

H being the Hilbert transform, $f \in \mathcal{S}(\mathbb{R})$.

We claim that for any $1 < p < \infty$,

$$\|[b, H]f\|_{L^p(\mathbb{R})} \leq C_p \|b\|_{\mathrm{BMO}(\mathbb{R})} \|f\|_{L^p(\mathbb{R})},$$

uniformly in $f \in \mathcal{S}(\mathbb{R})$.

Proof. To do this, we first recall the paraproduct decomposition (valid here due to our technical assumptions on b and f):

$$bf = \sum_N b_N f = \sum_N b_N f_{\ll N} + \sum_N b_N f_{\sim N} + \sum_N b_N f_{\gg N}.$$

We then apply this to the difference, to obtain

$$\begin{aligned}[b, H]f = \sum_N (b_N (Hf)_{\ll N} - H(b_N f_{\ll N})) \\ + \sum_N (b_N (Hf)_{\sim N} - H(b_N f_{\sim N})) \\ + \sum_N (b_N (Hf)_{\gg N} - H(b_N f_{\gg N})),\end{aligned}$$

and we deal with each of these terms in turn.

To estimate the first sum, we consider

$$\sum_N b_N (Hf)_{\ll N} - \sum_N H(b_N f_{\ll N}).$$

We recognize the first term to be

$$\left\| \sum_N b_N (Hf)_{\ll N} \right\|_{L^p(\mathbb{R})} = \|\pi(b, Hf)\|_{L^p(\mathbb{R})},$$

where π is a high-low paraproduct.

It follows from the BMO / Carleson measure theory that we can bound this by

$$\lesssim_p \|b\|_{\mathrm{BMO}(\mathbb{R})}\|Hf\|_{L^p(\mathbb{R})} \lesssim_p \|b\|_{\mathrm{BMO}(\mathbb{R})}\|f\|_{L^p(\mathbb{R})}.$$

For the next term, we sum in N and use linearity of H to recognize

$$\|H(\pi(b,f))\|_{L^p(\mathbb{R})} \lesssim_p \|\pi(b,f)\|_{L^p(\mathbb{R})} \lesssim_p \|b\|_{\mathrm{BMO}(\mathbb{R})}\|f\|_{L^p(\mathbb{R})},$$

using the paraproduct bound with the boundedness of the Hilbert transform.

For the intermediate sum, we can again apply the standard endpoint results for paraproducts, this time bounding

$$\left\|\sum_N b_N (Hf)_{\sim N}\right\|_{L^p(\mathbb{R})} = \|\pi^*(b, Hf)\|_{L^p(\mathbb{R})},$$

where π^* is a high-high paraproduct.

We bound this in the standard way by

$$\lesssim_p \|b\|_{\mathrm{BMO}(\mathbb{R})}\|Hf\|_{L^p(\mathbb{R})} \lesssim_p \|b\|_{\mathrm{BMO}(\mathbb{R})}\|f\|_{L^p(\mathbb{R})}.$$

Next, we control the other term in the form of

$$\|H(\pi^*(b,f))\|_{L^p(\mathbb{R})} \lesssim_p \|\pi^*(b,f)\|_{L^p(\mathbb{R})} \lesssim \|b\|_{\mathrm{BMO}(\mathbb{R})}\|f\|_{L^p(\mathbb{R})}.$$

For the final term, we will require a more delicate argument; we proceed by duality. Take $g \in \mathcal{S}(\mathbb{R})$, and consider the term

$$\sum_N \langle b_N (Hf)_{\gg N} - H(b_N f_{\gg N}), g\rangle.$$

We have

$$\mathcal{F}(b_N (Hf)_{\gg N}) = \mathcal{F}(b_N) * \mathcal{F}((Hf)_{\gg N}),$$

as H is a Fourier multiplier operator (and thus commutes with the Littlewood-Paley frequency localizations), so that the Fourier transform of this term is

$$\int_{\mathbb{R}} \widehat{b}(\eta)\varphi_N(\eta)\,\mathrm{sign}(\xi-\eta)\widehat{f}(\xi-\eta)\psi_{\gg N}(\xi-\eta)\,d\eta.$$

Meanwhile, using the fact that the Fourier transform sends products to convolution products, we have that each Fourier transform of an $H(b_N f_{\gg N})$ is given by a convolution integral

$$\mathrm{sign}(\xi)\int_{\mathbb{R}} \widehat{b}(\eta)\varphi_N(\eta)\widehat{f}(\xi-\eta)\psi_{\gg N}(\xi-\eta)\,d\eta.$$

Due to the presence of the $\varphi_N(\eta)$ term, both of these integrals are over bounded regions, where $|\eta| \sim N$ in $\mathbb{R}$. As a consequence, since $\psi_{\gg N}(\eta)$ is supported only in $|\eta| > 4N$, we have that both integrals vanish for any choice of $|\xi| \leq 2N$. This encompasses a neighborhood of the origin, reaching out to a nontrivial proportion of the characteristic length.

Likewise, if $|\xi|$ is sufficiently large — say, $|\xi| > 4N$ — then any integration over η, varying over scales of $\leq 2N$, will never change the sign of $\xi - \eta$; if $|\eta| \ll |\xi|$, then $\xi - \eta$ and ξ will either be both positive or both negative. Consequently,

$$\operatorname{sign}(\xi - \eta) - \operatorname{sign}(\xi) = (\pm 1) - (\pm 1) = 0$$

for all ξ sufficiently large, relative to N.

It follows that if we take and fix a cutoff function in $\mathbb{R} \setminus \{0\}$, identically 1 on $1 \leq |\cdot| \leq 8$, then we can create an operator localizing the difference

$$b_N(Hf)_{\gg N} - H(b_N f_{\gg N}) = \tilde{P}_N(b_N(Hf)_{\gg N} - H(b_N f_{\gg N})).$$

We can thus rewrite our sum using duality as

$$\sum_N \langle b_N(Hf)_{\gg N}, \tilde{P}_N g\rangle - \langle H(b_N f_{\gg N}), \tilde{P}_N g\rangle,$$

and note that for all $M \gg N$ sufficiently large (say, $M - 2N \geq 16N$), $b_N(Hf)_M$ and $H(b_N f_M)$ will have frequency support well beyond the support of $\tilde{P}_N g$. As a consequence, we need only consider the terms

$$\sum_{1<r<5} \sum_N \langle b_N(Hf)_{2^r N}, \tilde{P}_N g\rangle - \sum_{1<r<5} \sum_N \langle H(b_N f_{2^r N}), \tilde{P}_N g\rangle.$$

For these, we use duality to rewrite as

$$\sum_{1<r<5} \sum_N \langle P_{2^r N}(b_N \tilde{P}_N g), Hf\rangle - \langle P_{2^r N}(b_N \tilde{P}_N(Hg)), f\rangle.$$

For each one, we use Plancherel to explicitly compute the L^2 bound of the operator:

$$\left\| \sum_N \tilde{P}_{2^r N}(b_N \tilde{P}_N g) \right\|_{L^2(\mathbb{R})} = \left\| \sum_N \tilde{\varphi}_{2^r N} \mathcal{F}(b_N \tilde{P}_N g) \right\|_{L^2(\mathbb{R})}.$$

We can write this as

$$\begin{aligned} &\int_{\mathbb{R}} \left| \sum_N \tilde{\varphi}_{2^r N}(\xi) \mathcal{F}(b_N \tilde{P}_N g)(\xi) \right|^2 d\xi \\ &\leq \int_{\mathbb{R}} \left(\sum_N \tilde{\varphi}_{2^r N}(\xi)^2 \right) \left(\sum_N |\mathcal{F}(b_N \tilde{P}_N g)(\xi)|^2 \right) d\xi \\ &\leq \int_{\mathbb{R}} 1 \sum_N |\mathcal{F}(b_N \tilde{P}_N g)(\xi)|^2 \, d\xi = \sum_N \int_{\mathbb{R}} |b_N(x)|^2 |(\tilde{P}_N g)(x)|^2 \, dx, \end{aligned}$$

and, by the Carleson measure / square function estimate, we know this is $\lesssim \|b\|_{\mathrm{BMO}(\mathbb{R})}^2 \|g\|_{L^2(\mathbb{R})}^2$. This paraproduct, as a linear operator on g, is (formally) given by a second-generation Calderón-Zygmund kernel of

$$K(x,y) = \sum_N (2^r N)^d \int_{\mathbb{R}^d} (\mathcal{F}^{-1}\varphi)(2^r N z) \\ N^d (\mathcal{F}^{-1}\tilde{\varphi})(N(x-y-z))(Q_N b)(x-z)\, dz,$$

and we can make this rigorous by first considering f with frequency support a compact subset of $\mathbb{R}\setminus\{0\}$, which means only finitely-many of the terms $(Hf)_{2^r N}$ and $H(b_N f_{2^r N})$ will contribute in the first sum, before using duality. Then we may truncate the range of N for corresponding b_N, and derive uniform bounds on the partial sums; by density, this will suffice.

We see that each term in the sum is of the form

$$\lesssim_{d,r,\varphi} \int_{\mathbb{R}^d} \langle z\rangle^{-2d} N^d \langle N(x-y) - z\rangle^{-2d} \|b\|_{\mathrm{BMO}(\mathbb{R}^d)}\, dz,$$

and thus has size

$$\simeq_d N^d \langle N(x-y)\rangle^{-2d} \|b\|_{\mathrm{BMO}(\mathbb{R}^d)},$$

while the derivatives $\nabla_x(K(x,y))$ can be dealt with using the Leibniz rule, and the fact that $|(\nabla_x(Q_N b))(x-z)| \lesssim N^d \|b\|_{\mathrm{BMO}(\mathbb{R}^d)}$ when the derivative lands on the second term; the ∇_y derivative is even easier.

Thus the kernel has a norm of $\lesssim_d \|b\|_{\mathrm{BMO}(\mathbb{R}^d)}$, and so, using the Calderón-Zygmund theory, we have a bound of $\lesssim_p \|b\|_{\mathrm{BMO}(\mathbb{R})}\|g\|_{L^p(\mathbb{R})}$ for all $1 < p < \infty$.

Using this technique, by Hölder's inequality we recover a bound of

$$\lesssim_p \|b\|_{\mathrm{BMO}(\mathbb{R})}\|g\|_{L^{p'}(\mathbb{R})}\|Hf\|_{L^p(\mathbb{R})} + \|b\|_{\mathrm{BMO}(\mathbb{R})}\|Hg\|_{L^{p'}(\mathbb{R})}\|f\|_{L^p(\mathbb{R})}.$$

It follows that we have

$$|\langle [b,H]f, g\rangle| \lesssim_p \|b\|_{\mathrm{BMO}(\mathbb{R})}\|f\|_{L^p(\mathbb{R})}\|g\|_{L^{p'}(\mathbb{R})}$$

for all $g \in \mathcal{S}(\mathbb{R})$, all $f \in \mathcal{S}(\mathbb{R})$ with $\mathrm{supp}(\widehat{f}) \Subset \mathbb{R}\setminus\{0\}$; and this suffices to give the desired conclusion. $\square$

Remark: By the a.e.-convergence of the Hilbert transform, we also have that the action of the commutator operator is given pointwise as

$$([b,H]f)(x) = \lim_{\epsilon\downarrow 0}\left(\int_{|x-y|>\epsilon} \frac{b(x)}{x-y} f(y)\,dy - \int_{|x-y|>\epsilon} \frac{1}{x-y} b(y) f(y)\,dy\right),$$

which is equal to a visibly non-singular (!) integral

$$\int_{\mathbb{R}} \frac{b(x)-b(y)}{x-y} f(y)\,dy.$$

Remark: We now discuss the extension of this result to general $b \in \mathrm{BMO}(\mathbb{R})$.

Assume $f \in \mathcal{S}(\mathbb{R})$ with $\mathrm{supp}(f) \Subset \mathbb{R} \setminus \{0\}$, and likewise, that $g \in \mathcal{S}(\mathbb{R})$ with $\mathrm{supp}(g) \Subset \mathbb{R} \setminus \{0\}$. We temporarily also assume that $b \in L^\infty(\mathbb{R}) \cap \mathrm{BMO}(\mathbb{R})$.

We consider the effect of taking the limit $N \downarrow 0$ of the paraproduct expansion.

We know that

$$\langle P_{\leq N} b, \phi \rangle = \langle b, P_{\leq N} \phi \rangle \to \langle b, \phi \rangle$$

as $N \to \infty$ when $b \in \mathrm{BMO}(\mathbb{R}^d)$, $\phi \in \mathcal{S}(\mathbb{R}^d)$, due to the convergence of $P_{\leq N}\phi \to \phi$ in $\mathcal{S}(\mathbb{R}^d)$ (and self-adjointness; here we use the L^2 duality bracket, with antilinearity and complex conjugation in the second argument).

As a consequence, we have

$$\langle P_{\leq N} b \cdot Hf, g \rangle = \langle P_{\leq N_0} b, \overline{Hf} \cdot g \rangle + \sum_{N_0 < M \leq N} \langle (P_M b)(Hf), g \rangle.$$

We now claim that the sequence of $(P_{\leq N_0} b)_{N_0 \in 2^{\mathbb{Z}}}$ is qualitatively bounded in $L^\infty(\mathbb{R}^d)$, uniformly in N_0, whenever b is as described. Indeed, from

$$(P_{\leq N_0} b)(x) = N_0^d \int_{\mathbb{R}^d} (\mathcal{F}^{-1}\psi)(N_0(x-y)) b(y)\, dy,$$

we observe that

$$|P_{\leq N_0} b| \lesssim_\psi \|b\|_{L^\infty(\mathbb{R}^d)}.$$

As a consequence, by the separability of $L^1(\mathbb{R}^d)$ and Banach-Alaoglu, there exists a subsequence $(N_0)_k \downarrow 0$ with a limiting function $\ell \in L^\infty(\mathbb{R}^d)$ such that

$$\lim_{k \to \infty} \langle P_{\leq (N_0)_k} b, \phi \rangle \to \int_{\mathbb{R}^d} \ell(x) \overline{\phi(x)}\, dx$$

for all $\phi \in \mathcal{S}(\mathbb{R}^d)$.

In particular, in our situation here, it follows that

$$\langle P_{\leq (N_0)_k} b, \overline{Hf} \cdot g \rangle \to \langle \ell \cdot Hf, g \rangle,$$

$$\langle P_{\leq (N_0)_k} b, \overline{f} \cdot Hg \rangle \to \langle \ell \cdot f, Hg \rangle.$$

Furthermore, we observe the following about the limit function: if $\phi \in \mathcal{S}(\mathbb{R}^d)$ with $\int_{\mathbb{R}^d} \phi(x)\, dx = 0$, then we have

$$\langle P_{\leq (N_0)_k} b, \phi \rangle = \langle b, P_{\leq (N_0)_k} \phi \rangle \to 0,$$

since we know that $(P_{\leq N} f)(x) - N^d (\int_{\mathbb{R}^d} f(y)\, dy)(\mathcal{F}^{-1}\psi)(Nx) \to 0$ in $L^1(\mathbb{R}^d)$ as $N \to 0$, for any $f \in L^1(\mathbb{R}^d)$. Using the mean-zero condition on ϕ and the size condition on b, we identify the limit.

It follows that $\langle \ell, \phi \rangle = 0$ whenever $\phi \in \mathcal{S}(\mathbb{R}^d)$ has mean-zero. Then ℓ must equal a constant; $\ell(x) = c \in \mathbb{C}$ a.e.

We then carefully write

$$\begin{aligned}\langle (P_{\leq N} b) \cdot Hf, g \rangle &= \langle P_{\leq (N_0)_k} b, \overline{Hf} \cdot g \rangle + \sum_{(N_0)_k < M \leq N} \langle (P_M b)(Hf)_{\ll M}, g \rangle \\ &+ \sum_{(N_0)_k < M \leq N} \langle (P_M b)(Hf)_{\sim M}, g \rangle + \sum_{(N_0)_k < M \leq N} \langle (P_M b)(Hf)_{\gg M}, g \rangle,\end{aligned}$$

and likewise

$$\begin{aligned}\langle (P_{\leq N} b) \cdot f, Hg \rangle &= \langle P_{\leq (N_0)_k} b, \overline{f} \cdot Hg \rangle + \sum_{(N_0)_k < M \leq N} \langle (P_M b) f_{\ll M}, Hg \rangle \\ &+ \sum_{(N_0)_k < M \leq N} \langle (P_M b) f_{\sim M}, Hg \rangle + \sum_{(N_0)_k < M \leq N} \langle (P_M b) f_{\gg M}, Hg \rangle,\end{aligned}$$

while also

$$\langle H(bf), g \rangle = -\langle bf, Hg \rangle = -\lim_{N \to \infty} \langle (P_{\leq N} b) \cdot f, Hg \rangle.$$

Thus we have

$$\langle b \cdot Hf - H(bf), g \rangle = \langle b \cdot Hf, g \rangle - \langle H(bf), g \rangle$$

as being equal to

$$\begin{gathered}\langle P_{\leq (N_0)_k} b, \overline{Hf} \cdot g \rangle + \sum_{(N_0)_k < M \leq N} \langle (P_M b)(Hf)_{\ll M}, g \rangle \\ + \sum_{(N_0)_k < M \leq N} \langle (P_M b)(Hf)_{\sim M}, g \rangle + \sum_{(N_0)_k < M \leq N} \langle (P_M b)(Hf)_{\gg M}, g \rangle \\ + \langle P_{\leq (N_0)_k} b, \overline{f} \cdot Hg \rangle + \sum_{(N_0)_k < M \leq N} \langle (P_M b) f_{\ll M}, Hg \rangle \\ + \sum_{(N_0)_k < M \leq N} \langle (P_M b) f_{\sim M}, Hg \rangle + \sum_{(N_0)_k < M \leq N} \langle (P_M b) f_{\gg M}, Hg \rangle,\end{gathered}$$

and we rearrange so that this is the sum of four terms:

$$\langle P_{\leq (N_0)_k} b, \overline{Hf} \cdot g \rangle + \langle P_{\leq (N_0)_k} b, \overline{f} \cdot Hg \rangle,$$

the ultra-low frequency term;

$$\begin{aligned}&\sum_{(N_0)_k < M \leq N} \langle (P_M b)(Hf)_{\ll M}, g \rangle + \sum_{(N_0)_k < M \leq N} \langle (P_M b) f_{\ll M}, Hg \rangle \\ &= \sum_{(N_0)_k < M \leq N} \langle (P_M b)(Hf)_{\ll M}, g \rangle - \sum_{(N_0)_k < M \leq N} \langle H((P_M b) f_{\ll M}), g \rangle,\end{aligned}$$

the high-low paraproduct term;

$$\sum_{(N_0)_k < M \le N} \langle (P_M b)(Hf)_{\sim M}, g\rangle + \sum_{(N_0)_k < M \le N} \langle (P_M b) f_{\sim M}, Hg\rangle$$
$$= \sum_{(N_0)_k < M \le N} \langle (P_M b)(Hf)_{\sim M}, g\rangle - \sum_{(N_0)_k < M \le N} \langle H((P_M b) f_{\sim M}), g\rangle,$$

the high-high paraproduct term;

$$\sum_{(N_0)_k < M \le N} \langle (P_M b)(Hf)_{\gg M}, g\rangle + \sum_{(N_0)_k < M \le N} \langle (P_M b) f_{\gg M}, Hg\rangle$$
$$= \sum_{(N_0)_k < M \le N} \langle (P_M b)(Hf)_{\gg M} - H((P_M b) f_{\gg M}), g\rangle,$$

the low-high frequency difference.

For this last contribution, we saw above that we could write this as

$$\sum_{1<r<5} \sum_{\substack{(N_0)_k \lesssim M \\ M \lesssim N}} \langle P_{2^r M}(b_M \tilde{P}_M g), Hf\rangle - \langle P_{2^r M}(b_M \tilde{P}_M (Hg)), f\rangle,$$

which is a sum of the form $\langle \pi_{a,r}(b,g), Hf\rangle - \langle \pi_{b,r}(b, Hg), f\rangle$ for two high-high paraproducts.

We now observe that any model sum of the form

$$\sum_{A \le M \le B} b_M f_{\ll M}, \quad \sum_{A \le M \le B} b_M f_{\sim M}, \quad \sum_{A \le M \le B} P_M (b_M f_{\sim M})$$

is L^p-bounded, by $\lesssim_{d,p} \|b\|_{\mathrm{BMO}(\mathbb{R}^d)} \|f\|_{L^p(\mathbb{R}^d)}$, with constants independent of the choices of $A, B \in 2^{\mathbb{Z}}$. Indeed, we can use duality, using the localization at scales $\sim M$ for the first sum, and using $\lesssim M$ for the second and rearranging. The L^2 theory only considers the full Carleson measure–type square function (summing over all N), which — by pointwise positivity — dominates everything we obtain from considering partial sums (we apply the square function with f for the first type of sum, and with g for the second). Consequently, we obtain an $L^2 \to L^2$ bound with operator norm $\lesssim \|b\|_{\mathrm{BMO}(\mathbb{R}^d)}$. Finally, the kernel bounds, as computed above, also are controlled by $\lesssim \|b\|_{\mathrm{BMO}(\mathbb{R}^d)}$, uniformly in A and B, since we used nothing but a crude triangle inequality estimate to obtain them; so we get an $L^p \to L^p$ bound of $\lesssim_{d,p} \|b\|_{\mathrm{BMO}(\mathbb{R}^d)}$ by Calderón-Zygmund theory.

Consequently, we have that

$$|\langle (P_{\le N} b) \cdot Hf, g\rangle + \langle (P_{\le N} b) \cdot f, Hg\rangle|$$
$$\lesssim_{p,d} |\langle P_{\le (N_0)_k} b, \overline{Hf} \cdot g\rangle + \langle P_{\le (N_0)_k} b, \overline{f} \cdot Hg\rangle|$$
$$+ \|b\|_{\mathrm{BMO}(\mathbb{R})} \|f\|_{L^p(\mathbb{R})} \|g\|_{L^{p'}(\mathbb{R})},$$

uniformly in N_0 and N. By the compact frequency support of f and g, we see that for sufficiently-large N, the $\langle P_{\leq N} b, \overline{Hf} \cdot g\rangle = \langle b, P_{\leq N}(\overline{Hf} \cdot g)\rangle = \langle b, \overline{Hf} \cdot g\rangle$, and likewise for the second term, so that the term on the left is actually

$$|\langle b \cdot Hf, g\rangle + \langle b \cdot f, Hg\rangle|.$$

For the ultra-low frequency error, we let $k \to \infty$ to obtain

$$\langle \ell, \overline{Hf} \cdot g\rangle + \langle \ell, \overline{f} \cdot Hg\rangle = \langle \ell \cdot Hf, g\rangle + \langle \ell \cdot f, Hg\rangle = \ell(\langle Hf, g\rangle + \langle f, Hg\rangle),$$

due to the constancy of ℓ. This term then vanishes by the skew-symmetry of the Hilbert transform.

We have thus shown that for $b \in \mathrm{BMO}(\mathbb{R}) \cap L^\infty(\mathbb{R})$,

$$|\langle [b, H]f, g\rangle| \lesssim_p \|b\|_{\mathrm{BMO}(\mathbb{R})} \|f\|_{L^p(\mathbb{R})} \|g\|_{L^{p'}(\mathbb{R})}$$

for all $f, g \in \mathcal{S}(\mathbb{R})$ with compact frequency support in $\mathbb{R} \setminus \{0\}$.

We know that for $b \in \mathrm{BMO}(\mathbb{R})$, $b\phi \in L^p(\mathbb{R})$ for all $1 \leq p < \infty$, whenever $\phi \in \mathcal{S}(\mathbb{R})$ (or, indeed, of any rapid decay).

We now take the sequence $(b_k)_{k \in \mathbb{N}} \subseteq L^\infty(\mathbb{R})$ of truncations of a general $b \in \mathrm{BMO}(\mathbb{R})$ to heights $\leq 2^k$, noting that $\|b_k\|_{\mathrm{BMO}(\mathbb{R})} \lesssim \|b\|_{\mathrm{BMO}(\mathbb{R})}$ uniformly in k. By our remarks immediately-prior, $b_k f \to bf$ in $L^p(\mathbb{R})$, by dominated convergence. As a consequence, $H(b_k f) \to H(bf)$. And by our assumptions on f, $Hf \in \mathcal{S}(\mathbb{R})$, so $\langle b_k \cdot Hf, g\rangle \to \langle b \cdot Hf, g\rangle$. Thus we remove the restriction on b.

Fixing g satisfying our assumptions, if we consider an arbitrary $f \in \mathcal{S}(\mathbb{R})$, and let $(f_n) \subseteq \mathcal{S}(\mathbb{R})$ be a sequence converging to f in $L^p(\mathbb{R})$ $(1 < p < \infty)$ with all terms satisfying the requisite vanishing condition, we have that

$$\langle b \cdot Hf_n, g\rangle = \langle b\overline{g}, \overline{Hf_n}\rangle \to \langle b\overline{g}, \overline{Hf}\rangle = \langle b \cdot Hf, g\rangle$$

by using the convergence in L^p of the Hilbert transforms and the $L^{p'}$-L^p Hölder inequality. Next, we also have

$$\langle H(bf_n), g\rangle = -\langle bf_n, Hg\rangle = -\langle f_n, \overline{b} \cdot Hg\rangle \to -\langle f, \overline{b} \cdot Hg\rangle = -\langle bf, Hg\rangle,$$

and $bf \in L^p(\mathbb{R})$ still, so we may take the adjoint to obtain the bound for all f, and all g supported in frequency away from the origin and infinity.

For Schwartz f, $H(bf)$ is an L^p function. We also see that the pointwise-defined function $b \cdot (Hf)$ lies in $L^q_{\mathrm{loc}}(\mathbb{R})$ for all $1 < q < \infty$. Since the difference defines an $L^{p'}$-bounded form on g, this second term (and the difference) must lie in L^p.

Taking the supremum over g satisfying our restrictions, we have

$$\|[b, H]f\|_{L^p(\mathbb{R})} \lesssim_p \|b\|_{\mathrm{BMO}(\mathbb{R})} \|f\|_{L^p(\mathbb{R})},$$

for all $f \in \mathcal{S}(\mathbb{R})$ and $b \in \mathrm{BMO}(\mathbb{R})$, $1 < p < \infty$.