The Hilbert transform

1 Definition and properties

Recall the distribution $\mathrm{pv}(\frac{1}{x})$, defined by

$$\mathrm{pv}(1/x)(\varphi) := \lim_{\epsilon\to 0}\int_{|x|\geq\epsilon}\frac{\varphi(x)}{x}dx.$$

The Hilbert transform is defined via the convolution with $\mathrm{pv}(1/x)$, namely

$$(Hf)(x) := \frac{1}{\pi}\lim_{\epsilon\to 0}\int_{|t|\geq\epsilon}\frac{f(x-t)}{t}dt.$$

The main theorem we are going to prove in this note is the following.

Theorem 1.1. *For any $p\in(1,+\infty)$, there exists a constant C_p such that*

$$\|Hf\|_p < C_p\|f\|_p \tag{1.1}$$

for all $f\in\mathcal{S}(\mathbb{R})$. Thus, H extends to a bounded linear operator on all of $L^p(\mathbb{R})$.

The first remark we make is that the above theorem is false when $p=1$. To see this, note that for smooth f with compact support, when x is very large, the main contribution to the integral

$$\int\frac{f(x-t)}{t}dt$$

comes from the values t which are not far away from x. This in general gives the decay rate of order $\frac{1}{|x|}$, unless there is a magical cancellation due to the sign changes of f, in which case one can hope for a faster decay. Indeed, it is not hard to show that if f is continuous and has compact support with $\int f(t)dt = a$, then

$$(Hf)(x) = \frac{a}{\pi x} + \mathcal{O}(1/x^2)$$

for all large x. As a consequence, $Hf\in L^1(\mathbb{R})$ if and only if $\int f = 0$. The statement of the above theorem is also not true for $p=+\infty$. The reason is that $\frac{1}{x}$ is not integrable at infinity, so one can not bound on $\|Hf\|_\infty$ merely by using the maximum of f without any other information (e.g., the size of the support).

For $p\in(1,+\infty)$, the above inequality (1.1) does hold. The proof we will give below consists of the following four steps.

1. For $p = 2$, we will show $\|Hf\|_2 = \|f\|_2$. This uses the Fourier transform of $\mathrm{pv}(1/x)$ and then Plancherel's theorem.

2. For $p = 1$, the bound in (1.1) does not hold. However, one has a weak type estimate, namely there exists $C > 0$ independent of f such that

$$\mu\big(\{x : |(Hf)(x)| > \lambda\}\big) < \frac{C}{\lambda}\|f\|_1$$

   for all large λ. This weak type estimate is to replace the invalid L^1 inequality, and its derivation is based on the Calderón-Zygmund decomposition of a function into small and large parts in such a way that some control can be obtained for the large part.

3. Based on the result of the above two steps, we then use interpolation to conclude that the inequality (1.1) holds for all $p \in (1, 2)$. The relevant result we use here is a special case of Marcinkiewicz interpolation theorem. It states that if an operator satisfies weak type $(1, 1)$ and (q, q) estimates, then it satisfies strong (p, p) inequality for all $p \in (1, q)$.

4. Finally, we will prove (1.1) for $p > 2$ by using the duality between L^p and L^q for $\frac{1}{p} + \frac{1}{q} = 1$.

Before we proceed to the proof, we first give a few useful properties of the Hilbert transform. As far as the Fourier transform $\widehat{Hf}$ is concerned, we need the Fourier transform of the tempered distribution $\mathrm{pv}(1/x)$. A formal calculation gives

$$\begin{aligned}
\widehat{\mathrm{pv}(1/x)}(\varphi) : &= \mathrm{pv}(1/x)(\hat{\varphi}) \\
&= \int_{\mathbb{R}} \frac{\hat{\varphi}(\xi)}{\xi} d\xi \\
&= \int_{\mathbb{R}} \frac{1}{\xi}\Big(\int_{\mathbb{R}} \varphi(x) e^{-2\pi i \xi x} dx\Big) d\xi \\
&= \int_{\mathbb{R}} \varphi(x)\Big(\int_{\mathbb{R}} \frac{1}{\xi} \cdot e^{-2\pi i x \xi} d\xi\Big) dx.
\end{aligned}$$

If we let $I = \int_{\mathbb{R}} \frac{1}{\xi} \cdot e^{-2\pi i x \xi} d\xi$, then by symmetry, we have

$$I = -i \int_{\mathbb{R}} \frac{\sin(2\pi x \xi)}{\xi} d\xi = -2i \int_0^{+\infty} \frac{\sin(2\pi x \xi)}{\xi} d\xi.$$

Now, recall the Dirichlet integral $\int_0^{+\infty} \frac{\sin \xi}{\xi} d\xi = \frac{\pi}{2}$; this gives the value of I by

$$I = -\pi i \cdot \mathrm{sgn}(x),$$

where $\mathrm{sgn}(x) = 1$ if $x > 0$ and -1 for $x < 0$. Substituting back into the expression of $\widehat{\mathrm{pv}(1/x)}$, we get

$$\widehat{\mathrm{pv}(1/x)}(\varphi) = -\pi i \int \varphi(x) \cdot \mathrm{sgn}(x) dx.$$

We have thus obtained the following proposition.

Proposition 1.2. *The Fourier transform of $pv(1/x)$ is $\widehat{pv(1/x)}(\xi) = -\pi i \cdot sgn(\xi)$.*

The above calculations are only formal in the sense that when changing the order of integration, the hypothesis of Fubini's theorem is not satisfied. One needs the range of integration for ξ to be finite in order for the integrand to be L^1, jointly in (ξ, x). So,

Exercise 1.3. *Justify rigorously the above calculations.*

2 Equality in L^2

The assertion $\|Hf\|_2 = \|f\|_2$ is an immediate consequence of the Fourier transform of $\text{pv}(1/x)$. Since Hf is the convolution of f with $\frac{1}{\pi}\text{pv}(1/x)$, we easily get

$$\widehat{Hf}(\xi) = -i \cdot \text{sgn}(\xi)\hat{f}(\xi),$$

which has the same L^2 norm of $\hat{f}$. Thus, an application of Plancherel's theorem gives

$$\|Hf\|_2 = \|\widehat{Hf}\|_2 = \|\hat{f}\|_2 = \|f\|_2.$$

Furthermore, if we neglect the multiplication of $-i$ for a moment, we see what the operator H does on f is simply changing the sign of its Fourier transform for negative ξ's. If we do it twice, then we just change the sign back as if nothing had happened. Thus, taking into account the multiplication of $-i$, we have the relation

$$H^2 = -\,\text{id}\,.$$

This says H is a unitary operator on L^2 with $H^{-1} = -H$.

3 Weak L^1 estimate

We now establish the weak-type $(1,1)$ estimate for the convolution operator H. For this, we need the Calderón-Zygmund decomposition of an integrable function, which in turn depends on the following lemma.

Lemma 3.1. *Let $f \in L^1$. Then, for any λ, there exists a subset $E_\lambda \subset \mathbb{R}$ such that*

1. *$|f(x)| \le \lambda$ for almost every $x \in E_\lambda^c$;*
2. *$\mu(E_\lambda) < \frac{1}{\lambda}\|f\|_1$;*
3. *E_λ is a union of intervals I_k's whose interiors are disjoint, and that*

$$\lambda \le \frac{1}{|I_k|}\int_{I_k} |f(x)|dx \le 2\lambda$$

for each k.

Proof. Since $f \in L^1$, we can have a family of disjoint intervals with equals size whose union is the whole real line and that for each I in this family, we have

$$\frac{1}{|I|}\int_I |f|dx \leq \lambda.$$

This can be achieved as long as the interval length is large enough. Now, for each such interval I, we decompose it into two equal sub-intervals. If the average of $|f|$ of one sub-interval is bigger than λ, then we label it as I_k and put it as one of the sub-intervals that consist of our set E_λ. We do this process for each of the initial intervals, and stop splitting when the average of $|f|$ of one children is bigger than λ. Thus, we have obtained a sequence of disjoint intervals $\{I_k\}$. If J_k denotes the immediate parent of I_k, then the fact that one splits J_k into two sub-intervals (one of them being I_k) implies

$$\frac{1}{|J_k|}\int_{J_k} |f| < \lambda.$$

Since $|J_k| = 2|I_k|$, we easily have

$$\lambda < \frac{1}{|I_k|}\int_{I_k} |f|dx < 2\lambda.$$

Also, we have

$$|E_\lambda| = \sum_k |I_k| < \sum_k \frac{1}{\lambda}\int_{I_k} |f|dx \leq \frac{1}{\lambda}\|f\|_1.$$

It remains to show f is almost everywhere bounded by λ outside E_λ. First, we note that every $x \in E_\lambda^c$ belongs to a nested sequence of intervals whose lengths shrink to 0, and that each of the interval in this sequence satisfies

$$\frac{1}{|I|}\int_I |f(y)|dy \leq \lambda.$$

By Lebesgue's differentiation theorem, for almost every $x \in E_\lambda^c$, $f(x)$ is the limit of the above average as $|I| \to 0$, which then implies $|f(x)| \leq \lambda$ a.e. in E_λ^c. This completes the proof of the lemma. □

Now, for $f \in L^1$, we let

$$g(x) = \begin{cases} f(x), & x \in E_\lambda^c \\ \frac{1}{|I_k|}\int_{I_k} f(y)dy, & x \in I_k. \end{cases},$$

Let $b_k = f \cdot 1_{\{I_k\}} - g$, then b_k is supported in the interval I_k with $\int_{I_k} b_k dx = 0$, and the function $b = \sum_k b_k$ satisfies

$$f = g + b.$$

This is the Calderón-Zygmund decomposition of f. The tail measure of Hf can be bounded by

$$\mu\big(\{x : |(Hf)(x)| > \lambda\}\big) \leq \mu\big(\{x : |(Hg)(x)| > \frac{\lambda}{2}\}\big) + \mu\big(\{x : |(Hb)(x)| > \frac{\lambda}{2}\}\big),$$

and it suffices to bound each of the two terms on the right hand side. For the first one, we first note that our construction of g satisfies

$$g \in L^2, \qquad \sup_x |g(x)| \leq 2\lambda, \qquad \|g\|_1 = \|f\|_1.$$

Thus, we get

$$\mu\big(\{x : |(Hg)(x)| > \frac{\lambda}{2}\}\big) \leq \frac{C}{\lambda^2}\|Hg\|_2 \leq \frac{C}{\lambda^2}\|g\|_2 \leq \frac{C}{\lambda^2}\|g\|_1\|g\|_\infty < \frac{C}{\lambda}\|g\|_1.$$

We now turn to the functions b_k. Each of them has support in the interval I_k with $\int_{I_k} b_k dx = 0$. We claim that for any function u supported in an interval I with $\int_I u dx = 0$, we have

$$\int_{(I^*)^c} |Hu| dx < C\|u\|_1$$

for some universal constant C, where I^* is the interval with the same center as I but with doubled size. To see this, first note that since $\int_I u dx = 0$, then if $x \in (I^*)^c$, we will have

$$(Hu)(x) = \frac{1}{\pi}\int_I \left(\frac{u(t)}{x-t} - \frac{u(t)}{x-c}\right) dt.$$

In particular, if $x \in (I^*)^c$ and $t \in I$, we will have

$$\left|\frac{1}{x-t} - \frac{1}{x-c}\right| = \left|\frac{t-c}{(x-t)(x-c)}\right| \leq \frac{|I|}{(x-c)^2},$$

and as a consequence, we obtain

$$\int (I^*)^c |(Hu)(x)| dx \leq C|I|\|u\|_1 \int_{(I^*)^c} \frac{1}{(x-c)^2} dx \leq C\|u\|_1.$$

Thus, for our functions b_k's, we have

$$\int_{(I_k^*)^c} |(Hb_k)(x)| dx < C\|b\|_k.$$

We are now ready to bound the tail measure for Hb. We let $E_\lambda = \cup_k I_k^*$, then the construction of the sets I_k's implies

$$|E_\lambda^*| < \frac{2}{\lambda}\|f\|_1.$$

Thus, we have

$$\begin{aligned}
\mu\big(\{x : |(Hb)(x)| > \frac{\lambda}{2}\}\big) &\leq |E_\lambda^*| + \mu\big(\{x \in (E_\lambda^*)^c : |(Hb)(x)| > \frac{\lambda}{2}\}\big) \\
&\leq \frac{2}{\lambda}\|f\|_1 + \frac{C}{\lambda}\int_{(E_\lambda^*)^c} |(Hb)(x)| dx \\
&\leq \frac{2}{\lambda}\|f\|_1 + \frac{C}{\lambda}\sum_k \int_{(I_k^*)^c} |(Hb_k)(x)| dx \\
&\leq \frac{2}{\lambda}\|f\|_1 + \frac{C}{\lambda}\sum_k \|b_k\|_1 \\
&\leq \frac{C}{\lambda}\|f\|_1,
\end{aligned}$$

where in the middle line we have used $(E_\lambda^*)^c \subset (I_k^*)^c$ as well as $|Hb| \leq \sum_k |Hb_k|$. This finishes the proof of the weak $(1,1)$ estimate.

4 Interpolation for $p \in (1,2)$

Given that the transform operator H satisfies both weak tpye $(1,1)$ and $(2,2)$ estimates, the strong (p,p) inequality then becomes a simple consequence of the following interpolation theorem.

Theorem 4.1. *Suppose T is a sub-additive operator that is both weak (p,p) and weak (q,q). Then, T is of strong type (r,r) for all $r \in (p,q)$.*

To prove this theorem, we need the concept of a distribution function. For any function f, its distribution function $\rho : \mathbb{R}^+ \to \mathbb{R}^+$ is defined by

$$\rho_f(\lambda) := \mu\big(\{x : |f(x)| > \lambda\}\big).$$

We will omit the subscript f if there is no confusion of the underlying function. The importance of the distribution function lies in the fact that knowing $\rho(\lambda)$ for all λ would give complete knowledge of $\|f\|_p$. In fact, for any $p > 1$, we have the identity

$$\int_{\mathbb{R}} |f(x)|^p dx = p \int_0^{+\infty} \lambda^{p-1} \rho(\lambda) d\lambda.$$

Now, let $f \in L^r$ and write $f = g + h$ with

$$g = 1_{\{|f|>\lambda\}}, \qquad h = 1_{\{|f|\leq\lambda\}}.$$

In other words, g is the large part of f and h is the small part. By the assumption $r \in (p,q)$, it is easy to see that $g \in L^p$ and $h \in L^q$. Now let T be an operator satisfying the assumptions in the above theorem, then

$$|(Tf)(x)| \leq |(Tg)(x)| + |(Th)(x)|,$$

and as a consequence, we have

$$\big\{x : |(Tf)(x)| > \lambda\big\} \subset \big\{x : |(Tg)(x)| > \frac{\lambda}{2}\big\} \cup \big\{x : |(Th)(x)| > \frac{\lambda}{2}\big\}.$$

If ρ denotes the distribution function of Tf, then the above inclusion together with the assumption T being both weak (p,p) and weak (q,q) gives

$$\rho(\lambda) \leq \frac{C_p}{\lambda^p} \cdot \|g\|_p^p + \frac{C_q}{\lambda^q} \cdot \|h\|_q^q.$$

By the definition of g and h, we get

$$\rho(\lambda) \leq \frac{C}{\lambda^p} \int_{|f|>\lambda} |f(x)| dx + \frac{C}{\lambda^q} \int_{|f|\leq\lambda} |f(x)|^q dx. \tag{4.1}$$

Now, by the relation

$$\int |Tf|^r dx = r \int_0^{+\infty} \lambda^{r-1} \rho(\lambda) d\lambda,$$

we see that to get a bound on $\|Tf\|_r$, it suffices to multiply λ^{r-1} on both sides of (4.1) and integrate λ. For the first term, we have

$$\begin{aligned} \int_0^{+\infty} \lambda^{r-p-1} \Big(\int_{|f|>\lambda} |f| dx \Big) d\lambda &= \int_{\mathbb{R}} |f(x)| \Big(\int_0^{|f(x)|} \lambda^{r-p-1} d\lambda \Big) dx \\ &= \frac{1}{r-p} \int_{\mathbb{R}} |f|^r dx. \end{aligned}$$

For the second one, we similarly have

$$\begin{aligned} \int_0^{+\infty} \lambda^{r-q-1} \Big(\int_{|f|\leq\lambda} |f|^r dx \Big) d\lambda &= \int_{\mathbb{R}} |f(x)|^r \Big(\int_{|f(x)|}^{+\infty} \lambda^{r-1-1} d\lambda \Big) dx \\ &= \frac{1}{q-r} \int_{\mathbb{R}} |f(x)|^r dx. \end{aligned}$$

Since $p < r < q$, we immediately get

$$\|Tf\|_r < C_r \|f\|_r.$$

We can see from the above calculations that $C_r^r \approx \frac{C}{r-p}$ or $\frac{C}{q-r}$ as $r \to p$ or q.

5 $p > 2$ case

Finally, we deal with the case $p > 2$ by duality. Let q be the conjugate of p, so the L^p norm of a function f could be represented by

$$\|f\|_p = \sup_{\|g\|_q=1} \left| \int fg dx \right|.$$

Now, let $f \in L^p$, we then have

$$\begin{aligned} \|Hf\|_p &= \sup_{\|g\|_q=1} \left| \int (Hf)(x)\overline{g(x)} dx \right| \\ &= \sup_{\|g\|_q=1} \left| \int f(x)\overline{(H*g)(x)} dx \right| \\ &\leq \sup_{\|g\|_q=1} \|f\|_p \|Hg\|_q \\ &\leq C\|f\|_p, \end{aligned}$$

which establishes the theorem. Note that here H^* acts on g as a complex conjugate of H, so it does not change the norm anyway.