LECTURE NOTES 8 FOR 247B

TERENCE TAO

1. Oscillatory integrals

A basic problem which comes up whenever performing a computation in harmonic analysis is how to quickly and efficiently compute (or more precisely, to estimate) an explicit integral. Of course, in some cases undergraduate calculus allows one to compute such integrals exactly, after some effort (e.g. looking up tables of special functions), but since in many applications we only need the *order of magnitude* of such integrals, there are often faster, more conceptual, more robust, and less computationally intensive ways to estimate these integrals.

In the case where the integral to evaluate is non-negative, e.g.

$$\int_{\mathbf{R}^d} \langle x-y\rangle^{-\alpha} \langle x-z\rangle^{-\beta}\ dx$$

then the method of *decomposition*, particularly *dyadic decomposition*, works quite well: split the domain of integration into natural regions, such as dyadic annuli on which a key term in the integrand is essentially constant, estimate each sub-integral (which generally reduces to the geometric problem of measuring the volume of some standard geometric set, such as the intersection of two balls), and then sum (generally one ends up with summing a standard series such a geometric series or harmonic series). For non-negative integrands, this approach tends to give answers which only differ above and below from the truth by a constant (possibly depending on things such as the dimension d). Slightly more generally, this type of estimation works well in providing *upper bounds* for integrals which do not oscillate very much. With some more effort, one can often extract *asymptotics* rather than mere upper bounds, by performing some sort of expansion (e.g. Taylor expansion) of the integrand into a main term (which can be integrated exactly, e.g. by methods from undergraduate calculus), plus an error term which can be upper bounded by an expression smaller than the final value of the main term.

However, there are many cases in which one has to deal with integration of highly oscillatory integrands, in which the naive approach of taking absolute values (thus destroying most of the oscillation and cancellation) will give very poor bounds. A typical such *oscillatory integral*[1] takes the form

$$\int_{\mathbf{R}^d} a(x) e^{i\lambda\phi(x)}\ dx, \tag{1}$$

where a is a bump function adapted to some reasonable set B (such as a ball), ϕ is a *real*-valued phase function (usually obeying some smoothness conditions), and $\lambda \in \mathbf{R}$ is a parameter to measure the extent of oscillation. One could consider more general integrals[2] in which the amplitude function a is replaced by something a bit more singular, e.g. a power singularity $|x|^{-\alpha}$, but the aforementioned dyadic decomposition trick can usually decompose such a "singular oscillatory integral" into a dyadic sum of oscillatory integrals of the above type. Also, one can use linear changes of variables to rescale B to be a normalised set, such as the unit ball or unit cube. In one dimension, the definite integral

$$\int_J e^{i\lambda\phi(x)}\ dx \tag{2}$$

is also of interest, where J is now an interval. While one can dyadically decompose around the endpoints of these intervals to reduce this integral to the previous smoother integral (1), in one dimension one can often compute the integrals (2) more directly.

There are two modern tools to estimate (either as upper bounds or as asymptotics) such integrals. One is the *principle of nonstationary phase*, which roughly speaking asserts that (1) is rapidly decreasing in λ whenever ϕ is smooth and non-stationary (thus $\nabla\phi$ does not vanish). This allows one to localise such integrals to the vicinity of the stationary points $\{x : \nabla\phi(x) = 0\}$. If these stationary points are not isolated, then matters can become extremely complicated; however, in many important cases the stationary points are isolated, and then one can apply the *principle of stationary phase*, which roughly speaking asserts that the contribution of each stationary point x_0 to an integral (1) is essentially equal to the amplitude $a(x_0)$ at that point, times the phase $e^{i\lambda\phi(x_0)}$ at that point, times the magnitude $|\{x \approx x_0 : \phi(x) = \phi(x_0) + O(1/\lambda)\}|$ of the region where the phase is close to stationary.

A more classical method is the *method of steepest descent*. This works for certain one-dimensional integrals, using the complex analysis method of contour shifting to shift the integral into a region where the phase acquires a large negative real part, and the integral can then be computed by taking absolute values and using cruder tools such as dyadic decomposition. For instance, one can use this method to show that

$$p.v. \int_{\mathbf{R}} P(x) e^{i\lambda x^2}\ dx = e^{i\pi/4} \int_{\mathbf{R}} P(e^{i\pi/4}x) e^{-\lambda x^2}\ dx \tag{3}$$

[1]This is sometimes also known as an *oscillatory integral of the first kind*, to distinguish it from *oscillatory integral operators* or *oscillatory integrals of the second kind*, which are integral operators whose kernel has significant oscillation.

[2]Another important class of integrals are improper integrals such as $\int_{\mathbf{R}} e^{i\lambda x^2}\ dx$, which are not convergent in an absolute sense but still converge in some weaker sense, e.g. conditional convergence. These can also be largely handled by dyadic decomposition into integrals of the form (1).

for all polynomials P and $\lambda > 0$, where the principal value on the left denotes the limit of the integral $\int_{-R}^{R} P(x)e^{ix^2}\ dx$ as $x \to \infty$. This shows in particular that we expect this integral to be small when λ is large and P vanishes near the origin. However, the method of steepest descent requires analytic extension of all the phases involved (and in particular is incompatible with the use of bump functions), and is difficult to generalise to higher dimensions, and so this method has been largely abandoned as obsolete (though it still is applied for "non-commutative integrals", which are of relevance, among other things, to scattering and inverse scattering problems, and thus to integrable systems. This is unfortunately well beyond the scope of this course).

In the second half of these notes we shall give an application of stationary phase to spherical averages, which in turn will allow us to revisit the Hardy-Littlewood maximal operator in very high dimensions.

2. One dimensional theory

Let us begin with the theory of the one-dimensional definite integrals

$$I(\lambda) = I_{J,\phi}(\lambda) := \int_J e^{i\lambda\phi(x)}\ dx$$

where J is an interval, $\lambda \in \mathbf{R}$, and $\phi : J \to \mathbf{R}$ is a function (which we shall assume to be smooth, in order to avoid technicalities). We observe some simple invariances:

- $I(-\lambda) = \overline{I(\lambda)}$, thus negative λ and positive λ behave similarly;
- Subtracting a constant from ϕ does not affect the magnitude of $I(\lambda)$;
- If $L : \mathbf{R} \to \mathbf{R}$ is any invertible affine-linear transformation, then $I_{L(J),\phi\circ L^{-1}}(\lambda) = |\det(L)| I_{J,\phi}(\lambda)$.
- We have $I_{J,\phi}(\lambda) = I_{J,\lambda\phi}(1)$.

From the triangle inequality we have the trivial bound

$$|I(\lambda)| \leq |J|.$$

This bound is of course sharp if ϕ is constant. But if ϕ is non-constant, we expect $I(\lambda)$ to decay as $\lambda \to \pm\infty$. For instance, we have

Proposition 2.1 (Esseén concentration inequality)**.** *For any $\varepsilon > 0$ and $\phi_0 \in \mathbf{R}$, we have*

$$|\{x \in J : |\phi(x) - \phi_0| \leq \varepsilon\}| \lesssim \varepsilon \int_0^{1/\varepsilon} |I(\lambda)|\ d\lambda.$$

Proof Using the various invariances we can normalise $\phi_0 = 0$ and $\varepsilon = 1$, and reduce to showing that

$$|\{x \in J : |\phi(x)| \leq 1\}| \lesssim \int_{-1}^{1} |I(\lambda)|\ d\lambda.$$

Now let ψ be a bump function adapted to $[-1,1]$. Observe from Fubini's theorem that

$$\int_{-1}^{1} \psi(\lambda) I(\lambda)\ d\lambda = \int_J \check{\psi}(\phi(x)/2\pi)\ dx.$$

One can easily choose ψ so that $\check{\psi}$ is non-negative, and bounded from below by an absolute constant on $[-2\pi, 2\pi]$ (e.g. some variant of the Fejér kernel will work). The claim then easily follows. ■

This simple proposition shows that the *average* decay of $I(\lambda)$ is linked to the non-constancy of ϕ, though it only gives a lower bound on this decay rather than an upper bound.

Now we give some pointwise decay bounds on $I(\lambda)$. As suggested by the above inequality, we will need some non-constancy condition on ϕ. One natural condition might be to impose some lower bound $|\phi'(x)| \geq c$ on the derivative of ϕ. Unfortunately, this by itself is not enough, if ϕ has some significant oscillation at wavelength $1/\lambda$:

Example 2.2. *Consider a phase function ϕ of the form*

$$\phi(x) := 2\pi x + \frac{1}{\lambda} f(\lambda x)$$

where $f : \mathbf{R}/\mathbf{Z} \to \mathbf{R}$ is a smooth 1-periodic function with Lipschitz constant at most $1/2$. Then $1/2 \leq \phi'(x) \leq 3/2$. Now observe that $e^{i\lambda\phi(x)}$ is periodic with period $1/\lambda$. Thus if $|J|$ is a multiple of $1/\lambda$, one quickly computes that

$$I(\lambda) = |J| \int_0^1 e^{2\pi i x} e^{if(x)}\ dx.$$

It is an easy matter to select f so that the integral on the right-hand side is non-zero. Thus this shows that $I(\lambda)$ can be comparable to $|J|$ even when ϕ is non-constant in the sense that $\phi' \sim |J|$.

However, one can get around this example in a number of ways. The first is by assuming control on the second derivative of ϕ:

Lemma 2.3 (Principle of non-stationary phase, toy version)**.** *Let $\phi : \mathbf{R} \to \mathbf{R}$ be a smooth phase such that $|\phi'(x)| \geq c$ and $|\phi''(x)| \leq C$ for some $C, c > 0$ and all $x \in J$. Then for all $\lambda > 0$ we have*

$$|I(\lambda)| \lesssim \frac{1}{\lambda}(\frac{1}{c} + \frac{C}{c^2}|J|).$$

Proof We write

$$\int_J e^{i\lambda\phi(x)}\ dx = \int_J \frac{1}{i\lambda\phi'(x)} \frac{d}{dx} e^{i\lambda\phi(x)}\ dx$$

and integrate by parts to obtain

$$I(\lambda) = \frac{1}{i\lambda\phi'(x)} e^{i\lambda\phi(x)}|_{\partial J} - \int_J (\frac{d}{dx}\frac{1}{i\lambda\phi'(x)}) e^{i\lambda\phi(x)}\ dx. \tag{4}$$

Taking absolute values, we obtain the claim. ■

We remark that one could certainly integrate by parts more times if desired, but one can not improve the decay of $\frac{1}{\lambda}$, as can easily be seen by considering the model case $\phi(x) := x$, although by doing so one does get better asymptotics. However, we shall see that the situation improves markedly if we use a smooth amplitude function.

Another option is to not require control on the second derivative, but merely that the first derivative is monotone:

Lemma 2.4 (Van der Corput lemma, first derivative version)**.** *Let $\phi : \mathbf{R} \to \mathbf{R}$ be a smooth phase such that $|\phi'(x)| \geq c$ for all $x \in J$ and ϕ' is monotone. Then for all $\lambda > 0$ we have*

$$|I(\lambda)| \lesssim \frac{1}{c\lambda}.$$

Proof Again, we start with (4). The first term is $O(1/c\lambda)$ already. As for the second term, we take absolute values to estimate it by

$$\frac{1}{\lambda} \int_J |\frac{d}{dx} \frac{1}{\phi'(x)}| \; dx.$$

But since ϕ' is monotone, so is $\frac{1}{\phi'}$, and so $\frac{d}{dx}\frac{1}{\phi'}$ has a consistent sign. This allows us to *reverse* the triangle inequality and move the absolute values back outside, at which point we can use the fundamental theorem of calculus to conclude. ■

Again, the example $\phi(x) = x$ shows that this lemma is sharp up to constants. One particularly useful feature of this lemma is that it does not depend on the lenght of the interval J. The lemma iterates quite nicely:

Lemma 2.5 (Van der Corput lemma, higher derivative version)**.** *Let $\phi : \mathbf{R} \to \mathbf{R}$ be a smooth phase such that $|\phi^{(k)}(x)| \geq c$ for some $k \geq 2$ and all $x \in J$. Then for all $\lambda > 0$ we have*

$$|I(\lambda)| \lesssim_k \frac{1}{(c\lambda)^{1/k}}.$$

Proof We induct on k. Pick a threshold $\alpha > 0$ to be chosen later. Observe that if $|\phi^{(k)}(x)| \geq c$, then $|\phi^{(k-1)}(x)| \geq \alpha$ will be true outside of an interval of length at most $O(\alpha/c)$. Also, on the remaining portion of the interval $\phi^{(k-1)}$ will be monotone. Applying the inductive hypothesis (or the previous lemma, when $k = 2$) we conclude that

$$|I(\lambda)| \lesssim_k \frac{1}{(\alpha\lambda)^{1/(k-1)}} + \alpha/c.$$

Optimising this in α, we obtain the claim. ■

One can check that the right hand side $\frac{1}{(c\lambda)^{1/k}}$ is consistent with all the symmetries of I mentioned earlier, in particular the dilation symmetry.

Now we consider the smoother integral

$$I_{a,\phi}(\lambda) := \int_{\mathbf{R}} a(x) e^{i\lambda\phi(x)} \; dx$$

in one dimension. The connection of this smoother integral to the previous integrals can be seen by the identity

$$I_{a,\phi}(\lambda) = -\int_{x_0}^{x_1} a'(x) I_{[x_0,x],\phi}(\lambda)\ dx \tag{5}$$

if a is supported on $[x_0, x_1]$, as can easily be seen either by integration by parts or by Fubini's theorem. Thus one can use bounds on the definite integral to obtain bounds on the smoothed out integral. For instance, we now conclude that

$$I_{a,\phi}(\lambda) = O_{a,k}(\lambda^{-1/k}) \tag{6}$$

if $k \geq 2$ and $\phi^{(k)}$ is non-zero on the support of a.

The iterated integration by parts trick works much better in the smooth context (no boundary terms!). Indeed, integration by parts yields the identity

$$I_{a,\phi}(\lambda) = \frac{-1}{i\lambda} I_{\frac{d}{dx}\frac{a}{\phi'},\phi}(\lambda). \tag{7}$$

Iterating this we conclude

Lemma 2.6 (Principle of non-stationary phase, one dimension)**.** *Let $a \in C_0^\infty(\mathbf{R})$, and let $\phi : \mathbf{R} \to \mathbf{R}$ be smooth such that ϕ' is non-zero on the support of a. Then $I_{a,\phi}(\lambda) = O_{N,a,\phi}(\lambda^{-N})$ for all $N \geq 0$.*

Note that this generalises the fact that the Fourier transform of a bump function is rapidly decreasing (this is essentially the special case $\phi(x) := x$). On the other hand, it is very "expensive" in terms of the amount of regularity on a and ϕ needed (one basically requires control on N derivatives of a and $N + 1$ derivatives on ϕ).

Now we consider the question of asymptotics. Our starting point is the basic formula

$$\int_{\mathbf{R}} e^{-\alpha x^2}\ dx = \sqrt{\frac{\pi}{\alpha}}$$

whenever α is a complex number with positive real part, using the standard branch of the square root in this area. In particular we have see that

$$\lim_{\varepsilon \to 0} \int_{\mathbf{R}} e^{-\varepsilon x^2} e^{\lambda i x^2}\ dx = e^{\pi i/4}\sqrt{\frac{\pi}{\lambda}}. \tag{8}$$

for all $\lambda > 0$. (The integral on the left is essentially a Fresnel integral.) On the other hand, from Lemma 2.6 we have

$$\int_{\mathbf{R}} a(x) e^{\lambda i x^2}\ dx = O_{N,a}(\lambda^{-N})$$

for any bump function a which vanishes near the origin and all $N \geq 1$, and thus by scaling

$$\int_{\mathbf{R}} a(x/R) e^{\lambda i x^2}\ dx = O_{N,a}(\lambda^{-N} R^{-N})$$

for any $R \geq 1$. By a dyadic decomposition and (8) we conclude that

$$\int_{\mathbf{R}} a(x) e^{\lambda i x^2}\ dx = e^{\pi i/4}\sqrt{\frac{\pi}{\lambda}} + O_{N,a}(\lambda^{-N})$$

whenever a is a bump function which equals 1 near the origin. More generally we have

Lemma 2.7 (Asymptotic expansion for the Fresnel phase)**.** *Let a be a bump function, and let $\phi(x) := x^2$. If we let $c_0, c_1, \ldots$ be the constants*

$$c_n := e^{\pi i/4}\sqrt{\pi}\frac{i^n a^{(2n)}(0)}{n!}$$

then we have the asymptotic expansion

$$I_{a,\phi}(\lambda) \sim \sum_{n=0}^{\infty} c_n \lambda^{-n-\frac{1}{2}}$$

in the sense that

$$I_{a,\phi}(\lambda) = \sum_{n=0}^{N} c_n \lambda^{-n-\frac{1}{2}} + O_{N,a}(\lambda^{-N-\frac{3}{2}}) \tag{9}$$

for all $N \geq 0$.

In view of (3), we expect an analogy between the theory of the Fresnel phase $e^{\lambda i x^2}$ and the theory of the Gaussian weight $e^{-\lambda x^2}$. It is instructive to obtain analogues of the above lemma for the non-oscillatory integral $\int_{\mathbf{R}} a(x) e^{-\lambda x^2}\, dx$.

Proof If a is odd, then the claim is true by symmetry, so we may assume a is even. We have just shown that the lemma is true when a equals 1 near the origin. If instead a equals x^{2n} near the origin, the claim follows by an induction on n using (7). By linearity, the claim then follows if a is a polynomial in x near the origin. Using Taylor expansion, it then suffices, for each fixed N, to prove the claim (9) when a vanishes near the origin to high order, say $N + 10$. But this follows by a repeated application of (7) (followed at last by a trivial estimation of $I(\lambda)$ using absolute values). ∎

Once one handles the phase x^2, one can use change of variables to deal with other stationary phases, as long as the phase is quadratic at the stationary point:

Lemma 2.8 (Asymptotic expansion for non-degenerate phases)**.** *Let a be a bump function, and let $\phi : \mathbf{R} \to \mathbf{R}$ be smooth and have a stationary point at x_0 with $\phi''(x_0) \neq 0$. If ϕ has no other stationary points on the support of a, then there exist constants $c_0, c_1, \ldots,$ with each c_n depending (in some explicit fashion) only on finitely many derivatives of a, ϕ at x_0, such that we have the asymptotic formula*

$$I_{a,\phi}(\lambda) = \sum_{n=0}^{N} c_n \lambda^{-n-\frac{1}{2}} e^{i\lambda\phi(x_0)} + O_{N,a,\phi}(\lambda^{-N-\frac{3}{2}}) \tag{10}$$

for all $N \geq 0$. Furthermore,

$$c_0 = e^{\pi i \operatorname{sgn}(\phi''(x_0))/4}\sqrt{\frac{2\pi}{|\phi''(x_0)|}}a(x_0).$$

Proof We may translate $x_0 = 0$, and then conjugate and normalise so that $\phi(0) = 0$ and $\phi''(0) = 2$, thus $\phi(x) = x^2 + O(x^3)$. If a vanishes near x_0, the claim follows from the principle of non-stationary phase, so we may assume that a is supported on a very small neighbourhood of 0, so that $\phi(x)$ is comparable to x^2 (and ϕ' is comparable to $2x$). In such a case one can perform a smooth change of variables to deform ϕ to be *exactly* x^2, which changes a in the usual manner; the claim will now follow from the preceding lemma. ■

The coefficients c_n are in principle computable explicitly for any given n, but in practice only the explicit form of c_0 is needed for most applications. The above lemma can also be viewed as a more precise version of the ($k = 2$ case of) (6). The quantity $\sqrt{\frac{2\pi}{|\phi''(x_0)|}}$ present in c_0 measures the size of the interval in which ϕ stays close to $\phi(x_0)$.

There is a similar claim for higher order stationary points:

Lemma 2.9 (Asymptotic expansion for finite order non-degenerate phases)**.** *Let a be a bump function, and let $\phi : \mathbf{R} \to \mathbf{R}$ be smooth and have a stationary point at x_0 with $\phi'(x_0) = \ldots = \phi^{(k-1)}(x_0) = 0$ and $\phi^{(k)}(x_0) \neq 0$ for some $k \geq 2$. If ϕ has no other stationary points on the support of a, then there exist constants $c_0, c_1, \ldots,$ with each c_n depending (in some explicit fashion) only on finitely many derivatives of a, ϕ at x_0, such that we have the asymptotic formula*

$$I_{a,\phi}(\lambda) = \sum_{n=0}^{N} c_n \lambda^{-n/k} e^{i\lambda\phi(x_0)} + O_{N,a,\phi,k}(\lambda^{-(N+1)/k})$$

for all $N \geq 0$. The quantity c_0 obeys the size estimate

$$|c_0| \sim_k |\phi^{(k)}(x_0)|^{-1/k} |a(x_0)|.$$

The claim is proven similarly to the previous claim (reducing to the model phase x^k, and using Taylor expansion to strip out the leading coefficients of a), and is left as an exercise. Again, this can be viewed as a more precise version of (6).

If ϕ has multiple stationary points on the support of a, then one can simply decompose a and obtain a sum over stationary points. Note that as long as all stationary points are of finite order, they cannot accumulate and so one has only finitely many stationary points on the support of a. (The situation unfortunately gets much more complex than this in higher dimensions.) When there is a stationary point of infinite order, Esseén's concentration lemma (adapted to smooth cutoffs) indicates that we do not expect any significant decay in $I(\lambda)$ at all, though as long as the set where ϕ is stationary has zero measure[3], one can show (using the principle of non-stationary phase) that $I(\lambda) \to 0$ as $\lambda \to \infty$.

[3]This can for instance happen if ϕ is constant. Note that Sard's theorem does show that the *image* of the stationary points under ϕ has measure zero, but this is not directly useful for us.

It is worth noting that the asymptotic formulae such as (9) are differentiable in λ once one strips out the phase $e^{i\lambda\phi(x_0)}$, or more specifically that

$$\frac{d^k}{d\lambda^k}[e^{-i\lambda\phi(x_0)}I_{a,\phi}(\lambda)] = \frac{d^k}{d\lambda^k}[\sum_{n=0}^{N} c_n\lambda^{-n-\frac{1}{2}}] + O_{N,a,k}(\lambda^{-N-\frac{3}{2}-k}).$$

This can be explained as follows. First we may normalise $\phi(x_0) = 0$. Then by differentiating under the integral sign we see that

$$\frac{d^k}{d\lambda^k}I_{a,\phi}(\lambda) = I_{(i\phi)^k a,\phi}(\lambda).$$

Thus we see that

$$\frac{d^k}{d\lambda^k}I_{a,\phi}(\lambda) = \sum_{n=0}^{N+k} d_n\lambda^{-n-\frac{1}{2}} + O_{N,a,k}(\lambda^{-N-\frac{3}{2}-k})$$

for some quantities d_n independent of λ. Integrating this k times, we see that this is only compatible with (9) if the series $\sum_{n=0}^{N+k} d_n\lambda^{-n-\frac{1}{2}}$ is the k^{th} derivative of $\sum_{n=0}^{N} c_n\lambda^{-n-\frac{1}{2}}$, and the claim follows. From this we see in particular that

$$I_{a,\phi}(\lambda) = b(\lambda)e^{i\lambda\phi(x_0)}$$

where b is an (inhomogeneous) symbol of order $-1/2$, with implied constants depending of course on a and ϕ.

3. Higher dimensional theory

The higher dimensional theory is less precise than the one-dimensional theory, mainly because the structure of stationary points can be significantly more complicated. Nevertheless, we can still say quite a bit about the higher dimensional oscillatory integrals

$$I_{a,\phi}(\lambda) := \int_{\mathbf{R}^d} a(x)e^{i\lambda\phi(x)}\ dx$$

in many cases. The van der Corput lemma becomes significantly weaker, and will not be discussed here; however, we still have the principle of non-stationary phase.

Lemma 3.1 (Principle of non-stationary phase)**.** *Let $a \in C_0^\infty(\mathbf{R}^d)$, and let $\phi : \mathbf{R}^d \to \mathbf{R}$ be smooth such that $\nabla\phi$ is non-zero on the support of a. Then $I_{a,\phi}(\lambda) = O_{N,a,\phi,d}(\lambda^{-N})$ for all $N \geq 0$.*

Proof Let x_0 lie in the support of a, then by rotation if necessary we may assume that $\partial_{x_1}\phi(x_0) \neq 0$. By smoothness the same is true for a small neighbourhood of x_0. If a is supported on this small neighbourhood then the claim then follows by applying the one-dimensional principle of non-stationary phase in the e_1 direction, followed by Fubini's theorem (here we have to use the fact that the bounds in the above principle depend on only finite many derivatives of a, ϕ, so that one has uniformity in the $e_2, \ldots, e_d$ directions). The general case then follows by a standard partition of unity argument exploiting the compactness of the support of a. ■

Now we look at quadratic phases. We again begin with a model case, in which the Fresnel phase x^2 is now replaced by a more general non-degenerate quadratic form.

Lemma 3.2 (Asymptotic expansion for quadratic phases)**.** *Let a be a bump function, and let $\phi : \mathbf{R}^d \to \mathbf{R}$ be a non-degenerate quadratic form. Then there exists constants $c_0, c_1, \ldots,$ with each c_k depending on ϕ and on finitely many derivatives of a at zero. Then*

$$I_{a,\phi}(\lambda) \sim \sum_{n=0}^{\infty} c_n \lambda^{-n-\frac{d}{2}}$$

in the sense that

$$I_{a,\phi}(\lambda) = \sum_{n=0}^{N} c_n \lambda^{-n-\frac{d}{2}} + O_{N,a,d}(\lambda^{-N-\frac{d}{2}-1}) \tag{11}$$

for all $N \geq 0$. Furthermore,

$$c_0 = e^{\pi i \mathrm{sgn}(Q)/4} \sqrt{\frac{2\pi}{|\det(Q)|}} a(0)$$

where $\mathrm{sgn}(Q)$ *is the signature of Q (the number of positive eigenvalues minus the number of negative eigenvalues).*

Proof We can diagonalise Q after an affine change of variables into a normal form

$$Q(x) = x_1^2 + \ldots + x_k^2 - x_{k+1}^2 - \ldots - x_d^2$$

for some $0 \leq k \leq d$; one can check that the coefficient c_0 transforms correctly by this procedure.

Suppose first that $a(x)$ factors as a tensor product:

$$a(x) = a_1(x_1) \ldots a_d(x_d).$$

Then the integral $I_{a,\phi}(\lambda)$ factorises into d one-dimensional integrals, and the claim follows from Lemma 2.8. We then obtain the same claim when a is a tensor product times a polynomial,

$$a(x) = a_1(x_1) \ldots a_d(x_d) P(x),$$

since one can split the polynomial into monomials. By Taylor expansion, to prove (11) for a fixed N it thus suffices to verify the case when a vanishes to order $2(N+d+1)$ (say), so that we may factorise $a(x) = |x|^{2(N+d+1)} b$ for some smooth b. But if we write $|x|^2 e^{iQ(x)} = \frac{1}{2i} \langle x, \nabla e^{iQx} \rangle_Q$ and integrate by parts, and repeat this process $N + d$ times, we will obtain a bound of $O(\lambda^{-N-\frac{d}{2}-1})$ as desired. ■

Now we can handle all non-degenerate isolated stationary points.

Lemma 3.3 (Asymptotic expansion for non-degenerate phases)**.** *Let a be a bump function, and let $\phi : \mathbf{R}^d \to \mathbf{R}$ be smooth and have a stationary point at x_0 with* $\det \nabla^2 \phi(x_0) \neq 0$*. If ϕ has no other stationary points on the support of a, then there* *exist constants $c_0, c_1, \ldots$, with each c_n depending (in some explicit fashion) only on finitely many derivatives of a, ϕ at x_0, such that we have the asymptotic formula*

$$I_{a,\phi}(\lambda) = \sum_{n=0}^{N} c_n \lambda^{-n-\frac{d}{2}} e^{i\lambda\phi(x_0)} + O_{N,a,d,\phi}(\lambda^{-N-\frac{d}{2}-1}) \tag{12}$$

for all $N \geq 0$. Furthermore,

$$c_0 = e^{\pi i \operatorname{sgn}(\nabla^2\phi(x_0))/4} \sqrt{\frac{2\pi}{|\det \nabla^2\phi(x_0)|}} a(x_0).$$

Proof We can translate $x_0 = 0$ and $\phi(0) = 0$, so that $\phi(x) = Q(x) + R(x)$ for some non-degenerate quadratic form $Q(x)$ and some $R(x) = O(|x|^3)$. We then Taylor expand[4]

$$e^{i\lambda\phi(x)} = e^{i\lambda Q(x)}[\sum_{j=0}^{2(N+d+1)-1} \frac{i^j}{j!}\lambda^j R(x)^j + \lambda^{2(N+d+1)} R(x)^{2(N+d+1)} \int_0^1 \frac{(1-t)^{2(N+d+1)-1}}{(2(N+d+1)-1)!} e^{it\lambda R(x)}\, dt].$$

The contribution of the finite sum is acceptable as we simply incorporate the $R(x)^j$ factor into the amplitude function a; the loss of λ^j is more than compensated by the order $3j$ decay in $R(x)^j$, as can be seen by integration by parts (and symmetrising to get rid of the odd order terms). The final term can also seen to be $O_{N,a,\phi}(\lambda^{-N-\frac{d}{2}-1})$ by repeated integration by parts. ■

The situation gets significantly more complicated when the $\det \nabla^2\phi$ vanishes; for instance, factors of $\log \lambda$ begin to appear in the asymptotic expansion. When the stationary set no longer consists of isolated points, but contains higher dimensional sets, the asymptotic expansions are not fully understood in general (at a bare minimum, resolution of singularities would be involved).

We make some auxiliary remarks about the above estimates. As stated, the implied constants in the error terms depend in an unspecified manner on the amplitude a and the phase ϕ. However, an inspection of the arguments show in fact that the implied constants depend only on the dimension d, the diameter of the support of a, the L^∞ norm of finitely many derivatives of a and ϕ, the non-degeneracy $|\det \nabla^2\phi(x_0)|$ of ϕ at the stationary point, and a lower bound on $|\nabla\phi|$ outside of a suitable small ball centred at x_0 (the radius of this ball will depend on the previous quantities). In particular, if one has a family of functions a, ϕ in which these quantities are all controlled uniformly, then one has uniform control on the error term. A somewhat related observation is that one also has derivative control on the error term in λ, similar to those mentioned in the one-dimensional case; for instance, with the hypotheses of Lemma 3.3 we have

$$\frac{d^k}{d\lambda^k}[e^{-i\lambda\phi(x_0)} I_{a,\phi}(\lambda)] = \frac{d^k}{d\lambda^k} \sum_{n=0}^{N} c_n \lambda^{-n-\frac{d}{2}} + O_{N,a,\phi,d,k}(\lambda^{-N-\frac{d}{2}-k-1});$$

[4] Another approach is to use Morse theory and apply a diffeomorphism to change ϕ to Q, as in the one-dimensional case.

the proof is as before. Similarly, if a and ϕ depend smoothly on some additional parameter, one can differentiate in that parameter and obtain similar asymptotic expansions; we omit the details.

4. Spherical measure

Let us now apply the above machinery to compute a very specific oscillatory integral, namely the Fourier transform of surface measure μ on the sphere $S^{d-1} \subset \mathbf{R}^d$. We normalise this measure to have total mass one: $\mu(S^{d-1}) = 1$. The Fourier transform $\hat{\mu}$ of this measure is then defined as

$$\hat{\mu}(\xi) := \int_{S^{d-1}} e^{-2\pi i x\cdot\xi}\, d\mu(x).$$

We are interested in the decay and asymptotics of this measure. One can compute this explicitly in terms of Bessel functions (and in the case when d is odd, the formula can even be given exactly in terms of trigonometric functions) but we will present the stationary phase approach as it is more robust (it does not require the measure to have any algebraic structure), and also has a clearer geometric interpretation than a purely algebraic approach. In particular we shall avoid tools such as cylindrical coordinates which are somewhat specific to the sphere.

We have the trivial bound

$$\hat{\mu}(\xi) \leq \int_{S^{d-1}} d\mu = 1$$

coming from the triangle inequality, which is attained at (and only at) $\xi = 0$. But we expect some decay as $|\xi| \to \infty$. Writing ξ in polar coordinates, $\xi = r\omega$, we have

$$\hat{\mu}(\xi) = \int_{S^{d-1}} e^{-2\pi i r(x\cdot\omega)}\, d\mu(x);$$

the parameter r thus plays the role of the asymptotic parameter λ in the preceding discussion. There is of course the issue that S^{d-1} is not a Euclidean space, but this can be rectified by an appropriate use of charts and smooth partitions of unity. Suppose for instance that $\omega = e_d$ (we can reduce to this case anyway using the rotational symmetry of μ). We used a smooth partition of unity to split S^{d-1} up into coordinate patches, one near e_d, one near $-e_d$, and a finite number away from both. Consider first the contribution of a patch away from e_d and $-e_d$. After applying a change of variables, this contribution takes the form

$$\int_{\mathbf{R}^{d-1}} a(x) e^{-2\pi i r(\phi(x)\cdot e_d)}\, dx$$

where a is a bump function and ϕ smoothly maps the support of a to the above-mentioned patch. This phase is stationary in x when $\nabla\phi(x)\cdot e_d = 0$; but since we are away from the two points $\pm e_d$ where the sphere is normal to e_d, this cannot happen. Thus the contribution of any such patch is $O_{N,d}(r^{-N})$ for any N.

Now consider the contribution of the patch centred at $+e_d$. Using the standard chart $x \mapsto (x, \sqrt{1-|x|^2})$ in a neighbourhood of 0 in $\mathbf{R}^d$, whose Jacobian can be computed as $\frac{1}{\sqrt{1-|x|^2}}$, the contribution of this patch takes the form

$$\int_{\mathbf{R}^{d-1}} a(x)e^{-2\pi ir\sqrt{1-|x|^2}} \frac{1}{\sqrt{1-|x|^2}}\ dx$$

where a is a bump function which equals 1 near 0 and is supported on a small neighbourhood of the origin; in particular it stays well away from the singularities of $\sqrt{1-|x|^2}$. The phase $\phi(x) := \sqrt{1-|x|^2}$ has a non-degenerate stationary point at zero, with $\nabla^2\phi = -I_{d-1}$; the contribution of this patch thus has an asymptotic expansion $\sum_{k=0}^\infty c_k r^{-(d-1)/2-k}e^{ir}$, where the c_k are explicitly computable (for instance, $c_0 = e^{-\pi i(d-1)/4}\sqrt{2\pi}$). Similarly with the patch near $-e_d$ (but with ir replaced by $-ir$, and the coefficients c_k replaced by their complex conjugates. Putting all this together, we obtain an asymptotic expansion

$$\widehat{d\mu}(\xi) \sim \sum_{k=0}^\infty c_k|\xi|^{-(d-1)/2}e^{i|\xi|} + \sum_{k=0}^\infty \overline{c_k}|\xi|^{-(d-1)/2}e^{-i|\xi|}.$$

Similar estimates hold for derivatives. Indeed it is not hard to use this method to obtain the identity

$$\widehat{d\mu}(\xi) = a(\xi)e^{i|\xi|} + \overline{a(\xi)}e^{-i|\xi|}$$

for $|\xi| \geq 1$ and some symbol $a(\xi)$ of order $-(d-1)/2$. (Informally, we have $\widehat{d\mu}(\xi) \sim e^{\pm i|\xi|}/|\xi|^{(d-1)/2}$ for $|\xi| \geq 1$. For $|\xi| \lesssim 1$, of course, $\widehat{d\mu}(\xi)$ is a smooth function.) In particular we have the useful decay estimate

$$\widehat{d\mu}(\xi) = O_d(\langle\xi\rangle^{-(d-1)/2}).$$

5. Spherical maximal function

Decay estimates for the Fourier transforms of measures have a variety of uses, ranging from restriction theory to dispersive estimates for PDE to geometric measure theory. Here we focus on one particular application, that of spherical averages. We begin with the easy observation that for any continuous function $f : \mathbf{R}^d \to \mathbf{C}$, we have the pointwise limit $\lim_{r\to 0} S_r f(x) = f(x)$, where $S_r f(x)$ is the spherical average

$$S_r f(x) := \int_{S^{d-1}} f(x + r\omega)\ d\mu(\omega).$$

Thus for instance $S_1 f = f * \mu$, and S_r is a rescaling of S_1. A natural question is whether this type of limiting behaviour also holds for, say, L^p functions. As usual, this question will hinge on the behaviour of a maximal operator, in this case the *spherical maximal operator*

$$M_S f(x) := \sup_{r>0} S_r|f|(x).$$

We pause to make a technical remark. If f is merely locally integrable rather than continuous, then Fubini's theorem only guarantees that $S_r|f|$ is defined almost everywhere rather than everywhere. Since there are uncountably many values of r, this may lead to the fact that $M_S f(x)$ is in fact not defined anywhere in the locally integrable case. This turns out to be a problem that can be dealt with later, but for now we avoid the issue by making the *a priori* assumption that f is Schwartz (actually continuous with compact support will suffice).

Note also that the Hardy-Littlewood maximal operator does not immediately appear to control any of the averages $S_r f$, mainly because S_r is an average over sets of measure zero. However, we will be able to improve this with Littlewood-Paley decomposition arguments.

Let f be Schwartz, so $S_r f$ can easily seen to be Schwartz also. From Minkowski's inequality we see that S_r is a contraction on every L^p, $1 \leq p \leq \infty$:

$$\|S_r f\|_{L^p(\mathbf{R}^d)} \leq \|f\|_{L^p(\mathbf{R}^d)}.$$

Now let's see if we can improve this. At first glance we cannot hope to improve the constant, since $S_r 1 = 1$. (And indeed, by truncating 1 at infinity to make it lie in L^p, we see that the L^p operator norm of S_r is indeed 1.) But 1 is a low frequency function - we can do better for high frequencies. Observe that

$$\widehat{S_r f}(\xi) = \widehat{d\sigma}(r\xi)\hat{f}(\xi) \tag{13}$$

and hence by the decay bounds

$$|\widehat{S_r f}(\xi)| \lesssim_d \langle r|\xi|\rangle^{-(d-1)/2}|\hat{f}(\xi)|.$$

If we then apply a Littlewood-Paley projection $\psi_j(D)$, and use Plancherel, we obtain

$$\|\psi_j(D)S_r f\|_{L^2(\mathbf{R}^d)} \lesssim_d \langle 2^j r\rangle^{-(d-1)/2}\|f\|_{L^2(\mathbf{R}^d)}. \tag{14}$$

This is non-trivial for the high frequency case $2^j \gg 1/r$. In order to take suprema in r, we also need to understand some regularity in r. Observe from (13) that

$$\partial_r \widehat{S_r f}(\xi) = \xi \cdot (\nabla\widehat{d\sigma})(r\xi)\hat{f}(\xi)$$

and hence (using the more refined asymptotics available on $\widehat{d\sigma}$)

$$|\partial_r \widehat{S_r f}(\xi)| \lesssim_d |\xi|\langle r|\xi|\rangle^{-(d-1)/2}|\hat{f}(\xi)|$$

and thus

$$\|\partial_r \psi_j(D)S_r f\|_{L^2(\mathbf{R}^d)} \lesssim_d 2^j\langle 2^j r\rangle^{-(d-1)/2}\|f\|_{L^2(\mathbf{R}^d)}. \tag{15}$$

Now we extend the L^2 estimates to L^p estimates.

Lemma 5.1. *Let $1 \leq p \leq 2$, $r > 0$, and $2^j \gtrsim 1/r$. Then for Schwartz f we have*

$$\|\psi_j(D)S_r f\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} (2^j r)^{-(d-1)/p'}\|f\|_{L^p(\mathbf{R}^d)}$$

and

$$\|\partial_r\psi_j(D)S_r f\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} 2^j(2^j r)^{-(d-1)/p'}\|f\|_{L^p(\mathbf{R}^d)}.$$

We remark that it is natural for the bounds for $\partial_r\psi_j(D)S_r f$ to be 2^j larger than those for $\psi_j(D)S_r f$; this reflects the uncertainty principle, that $\psi_j(D)$ introduces a spatial uncertainty of 2^{-j}, and so one should not be able to detect changes in r of less than 2^{-j}.

Proof We have already proven these claims for $p = 2$, so by interpolation (either real or complex will do) it suffices to verify them for $p = 1$. The first claim follows since S_r is a contraction, so it suffices to prove the second claim. We rescale $j = 0$ and reduce to showing that

$$\|\partial_r \psi_0(D) S_r f\|_{L^1(\mathbf{R}^d)} \lesssim_d \|f\|_{L^1(\mathbf{R}^d)}$$

when $r \gtrsim 1$. The operator $\partial_r \psi_0(D) S_r$ is an integral operator with kernel

$$K(x, y) := \partial_r \int_{S^{d-1}} \check{\psi}_0(x - y - r\omega)\ d\omega$$

and from the Schwartz nature of $\check{\psi}_0$ one readily verifies that

$$K(x, y) = O_d(r^{-d} \langle |x - y| - r\rangle^{-100d}).$$

The claim then follows from Minkowski's inequality (or Schur's test). ■

For the low frequency case, we have a very satisfactory pointwise estimate:

Lemma 5.2. *If $2^j \lesssim 1/r$, then $|\psi_{\leq j}(D) S_r f(x)| \lesssim_d Mf(x)$.*

Proof We may rescale $j = 0$ and $x = 0$, so $r = O(1)$. From Fubini we observe that

$$\psi_{\leq 0}(D) S_r f(0) = \int_{\mathbf{R}^d} (\int_{S^{d-1}} \check{\psi}_{\leq 0}(-x - r\omega)\ d\omega) f(x)\ dx.$$

Since $r = O(1)$ and $\check{\psi}_{\leq 0}$ is rapidly decreasing, we easily verify that

$$\int_{S^{d-1}} \check{\psi}_{\leq 0}(-x - r\omega)\ d\omega = O(\langle x\rangle^{-100d})$$

(say), and the claim then follows by standard dyadic decomposition. ■

We almost have enough tools to control the full maximal function. Let us first deal with a warm-up case, when the radius is restricted to $1 \leq r \leq 2$.

Proposition 5.3. *Let $d \geq 3$ and $p > \frac{d}{d-1}$. Then for all Schwartz f*

$$\| \sup_{1 \leq r \leq 2} S_r |f| \|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} \|f\|_{L^p(\mathbf{R}^d)}.$$

Proof By interpolation we may take $\frac{d}{d-1} < p < 2$. We may take f non-negative. By Lemma 5.2 we have

$$|\psi_{\leq 0}(D) S_r f(x)| \lesssim_d Mf(x)$$

for all $1 \leq r \leq 2$, hence by the triangle inequality

$$\sup_{1 \leq r \leq 2} S_r f(x) \lesssim_d Mf(x) + \sum_{j=1}^\infty \sup_{1 \leq r \leq 2} |\psi_j(D) S_r f|.$$

Thus by the triangle inequality again, it will suffice to show that

$$\| \sup_{1 \leq r \leq 2} |\psi_j(D) S_r f| \|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} 2^{-\varepsilon j} \|f\|_{L^p(\mathbf{R}^d)}$$

for all $j \geq 1$, and some $\varepsilon > 0$ depending only on p and d.

Of course, we want to use Lemma 5.1. Observe from the fundamental theorem of calculus that for any interval I,

$$\sup_{r\in I}|\psi_j(D)S_rf| \leq \psi_j(D)S_{r_I}f + \int_I |\partial_r\psi_j(D)S_rf|\ dr$$

where r_I is the centre of I. We could apply this directly with $I = [1,2]$ but this gives a bad estimate (the integral over I dominates too much). The optimal size of I (in which both terms on the right-hand side balance) is when $|I| \sim 2^{-j}$. Then from Minkowski's inequality and Lemma 5.1 we see that

$$\|\sup_{r\in I}|\psi_j(D)S_rf|\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} 2^{-j(d-1)/p'}\|f\|_{L^p(\mathbf{R}^d)}.$$

Now if we partition $[1,2]$ into 2^j intervals $I_1,\dots,I_{2^j}$ of length 2^{-j} and use the obvious pointwise bound

$$\sup_{1\leq r\leq 2}|\psi_j(D)S_rf| \leq (\sum_{k=1}^{2^j}\sup_{r\in I_k}|\psi_j(D)S_rf|^p)^{1/p}$$

we conclude that

$$\|\sup_{1\leq r\leq 2}|\psi_j(D)S_rf|\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} 2^{j/p}2^{-j(d-1)/p'}\|f\|_{L^p(\mathbf{R}^d)}.$$

Since $p > d/(d-1)$ by hypothesis, the claim follows. ∎

One may wonder whether the condition $p > d/(d-1)$ is sharp. There are standard counterexamples to establish this (see Q2). The condition $d \geq 3$ can be lowered to $d \geq 2$, but this is somewhat more difficult (and was first achieved by Bourgain, with a significant later simplification by Sogge).

Having tackled the range $1 \leq r \leq 2$, let us now deal with an opposite case, when r is restricted to be a power of two.

Lemma 5.4. *Let $d \geq 2$. Then for any $1 < p \leq \infty$ we have*

$$\|\sup_{n\in\mathbf{Z}} S_{2^n}|f|\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} \|f\|_{L^p(\mathbf{R}^d)}.$$

Proof Again we may take $1 < p \leq 2$ and f non-negative. Using Lemma 5.2, we have the pointwise estimate

$$\sup_{n\in\mathbf{Z}} S_{2^n}f \lesssim_d Mf + \sum_{k=1}^\infty \sup_n |\psi_{-n+k}(D)S_{2^n}f|$$

so it suffices by the triangle inequality to establish a bound of the form

$$\|\sup_n|\psi_{-n+k}(D)S_{2^n}f|\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} 2^{-\varepsilon k}\|f\|_{L^p(\mathbf{R}^d)}$$

for some $\varepsilon > 0$ depending on p, d.

Let's first deal with an L^2 estimate. By estimating a supremum by a square function we have

$$\|\sup_n|\psi_{-n+k}(D)S_{2^n}f|\|_{L^2(\mathbf{R}^d)} \leq (\sum_n \|\psi_{-n+k}(D)S_{2^n}f\|_{L^2(\mathbf{R}^d)}^2)^{1/2}.$$

Next observe that $\psi_{-n+k}(D)S_{2^n}f$ depends only on the Fourier coefficients of f at frequencies $|\xi| \sim 2^{-n+k}$ (note that S_{2^n} and $\psi_{-n+k}(D)$ are both Fourier multipliers and hence commute with each other). Thus we may write $\psi_{-n+k}(D)S_{2^n}f = \psi_{-n+k}(D)S_{2^n}\tilde{\psi}_{-n+k}(D)f$ for some suitable bump function $\tilde{\psi}_{-n+k}$. Applying (14) we conclude

$$\|\sup_n |\psi_{-n+k}(D)S_{2^n}f|\|_{L^2(\mathbf{R}^d)} \lesssim_d 2^{-(d-1)k/2}(\sum_n \|\tilde{\psi}_{-n+k}(D)f\|^2_{L^2(\mathbf{R}^d)})^{1/2}$$

and then by orthogonality we conclude

$$\|\sup_n |\psi_{-n+k}(D)S_{2^n}f|\|_{L^2(\mathbf{R}^d)} \lesssim_d 2^{-(d-1)k/2}\|f\|_{L^2(\mathbf{R}^d)}.$$

Now we obtain a weak $(1,1)$ estimate for the same maximal function, namely

$$|\{\sup_n |\psi_{-n+k}(D)S_{2^n}f| \geq \lambda\}| \lesssim \frac{k}{\lambda}\|f\|_{L^1(\mathbf{R}^d)}; \tag{16}$$

interpolating this with the L^2 bound we obtain the desired L^p bound.

Now we prove (16). We can use dilations and homogeneity to rescale $\|f\|_{L^1(\mathbf{R}^d)}, \lambda \sim 1$. We use the Calderón-Zygmund decomposition at level λ to split $f = g + \sum_Q b_Q$, where $\|g\|_2 \lesssim_d 1$, Q are disjoint cubes with $\sum_Q |Q| \lesssim_d 1$, and each b_Q is supported on Q, has mean zero, and $\int_Q |b_Q| \lesssim_d |Q|$. Then, as usual,

$$|\{\sup_n |\psi_{-n+k}(D)S_{2^n}f| \geq \lambda\}| \lesssim |\{\sup_n |\psi_{-n+k}(D)S_{2^n}g| \geq \lambda/2\}| + \sum_Q |Q| + |\{x \notin \bigcup_Q 2Q : \sum_Q \sup_n |\psi_{-n+k}(D)S_{2^n}b_Q|$$

The first term is $O_d(1)$ by the L^2 theory (in fact we even get the much better estimate of $O_d(2^{-(d-1)k})$, but we won't use that here). The second term is also $O_d(1)$. As for the second term, we use Chebyshev's inequality to estimate it by

$$\lesssim \|\sum_Q \sup_n |\psi_{-n+k}(D)S_{2^n}b_Q|\|_{L^1(\mathbf{R}^d \setminus \bigcup_Q 2Q)} = \sum_Q \|\sup_n |\psi_{-n+k}(D)S_{2^n}b_Q|\|_{L^1(\mathbf{R}^d \setminus 2Q)}$$

and so it will suffice to show for each cube Q that

$$\|\sup_n |\psi_{-n+k}(D)S_{2^n}b_Q|\|_{L^1(\mathbf{R}^d \setminus 2Q)} \lesssim_d k|Q|$$

whenever b_Q is supported on Q with mean zero and $\|b_Q\|_{L^1(Q)} \lesssim |Q|$. We may rescale so that Q is the standard unit cube. First consider the high frequency case when $n \leq 0$. Then it is not hard (using the rapid decrease of $\check{\psi}_{-n+k}$, and the fact that we are excluding $2Q$) to obtain the bound

$$\|\psi_{-n+k}(D)S_{2^n}b_Q\|_{L^1(\mathbf{R}^d \setminus 2Q)} \lesssim_d 2^{-100dn}$$

(in fact we even get an arbitrarily large exponential decay in k also, though we do not need this) and so this term sums. Now for the medium frequencies $0 < n \leq k$, each term contributes at most $O(1)$ by Fubini's theorem or Young's inequality, so the net contribution here is $O(k)$ by the triangle inequality. Let's now look at the high frequencies $n > k$. Here we expand out

$$\psi_{-n+k}(D)S_{2^n}b_Q(x) = \int_{\mathbf{R}^d}\int_{S^{d-1}} \check{\psi}_{-n+k}(x - y - 2^n\omega)b_Q(y)\; dyd\mu(\omega).$$

As usual, we use the trick that if b_Q has mean zero, we can subtract a constant from the other factor, to obtain

$$\psi_{-n+k}(D)S_{2^n}b_Q(x) = \int_{\mathbf{R}^d}\int_{S^{d-1}}[\check{\psi}_{-n+k}(x-y-2^n\omega)-\check{\psi}_{-n+k}(x-y_Q-2^n\omega)]b_Q(y)\,dyd\mu(\omega)$$

where y_Q is the centre of Q. We use the fundamental theorem of calculus to write

$$\check{\psi}_{-n+k}(x-y-2^n\omega)-\check{\psi}_{-n+k}(x-y_Q-2^n\omega) = \int_0^1 (y-y_Q)\cdot\nabla\check{\psi}_{-n+k}(x-(1-t)y-ty_Q-2^n\omega)\,dt$$

and then take absolute values everywhere to conclude that

$$\|\psi_{-n+k}(D)S_{2^n}b_Q\|_{L^1(\mathbf{R}^d)} \lesssim_d \|\nabla\check{\psi}_{-n+k}\|_{L^1(\mathbf{R}^d)}\|b_Q\|_{L^1(\mathbf{R}^d)} \lesssim 2^{-n+k}$$

and this sums properly in the region $n \geq k$. This proves (16), and the claim follows. ■

By combining the two arguments together we can now control the full maximal function.

Theorem 5.5. *[Stein's spherical maximal inequality] Let $d \geq 3$ and $p > \frac{d}{d-1}$. Then for all Schwartz f*

$$\|M_S f\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} \|f\|_{L^p(\mathbf{R}^d)}.$$

Proof Once again, we can take $1 < p \leq 2$ and f non-negative. We split $r = 2^n t$, where n is an integer and $1 \leq t < 2$, and use Lemma 5.2 to obtain the pointwise estimate

$$M_S f \lesssim Mf + \sum_{k=1}^\infty \sup_{1\leq t<2}\sup_n |\psi_{n+k}(D)S_{2^n t}f|$$

and so it will suffice to show that

$$\|\sup_{1\leq t<2}\sup_n |\psi_{n+k}(D)S_{2^n t}f|\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} 2^{-\varepsilon k}\|f\|_{L^p(\mathbf{R}^d)}$$

for all $k \geq 1$ and some $\varepsilon > 0$ depending only on p, d. Now from the proof of the previous lemma, we already know for each $1 \leq t < 2$ that

$$\|\sup_n |\psi_{n+k}(D)S_{2^n t}f|\|_{L^2(\mathbf{R}^d)} \lesssim_d 2^{-(d-1)k/2}\|f\|_{L^2(\mathbf{R}^d)}$$

and

$$\|\sup_n |\psi_{n+k}(D)S_{2^n t}f|\|_{L^{1,\infty}(\mathbf{R}^d)} \lesssim_d k\|f\|_{L^1(\mathbf{R}^d)}$$

and thus by Marcinkeiwicz interpolation

$$\|\sup_n |\psi_{n+k}(D)S_{2^n t}f|\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} k2^{-(d-1)k/p'}\|f\|_{L^p(\mathbf{R}^d)}.$$

A similar argument also gives

$$\|\sup_n |\partial_t\psi_{n+k}(D)S_{2^n t}f|\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} k2^k 2^{-(d-1)k/p'}\|f\|_{L^p(\mathbf{R}^d)}$$

and so by the fundamental theorem of calculus as before

$$\|\sup_n\sup_{t\in I} |\psi_{n+k}(D)S_{2^n t}f|\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} k2^{-(d-1)k/p'}\|f\|_{L^p(\mathbf{R}^d)}$$

for any interval $I \subset [1,2]$ of length 2^{-k}. We sum this as before to obtain

$$\| \sup_n \sup_{1 \leq t < 2} |\psi_{n+k}(D) S_{2^n t} f| \|_{L^p(\mathbf{R}^d)} \lesssim_{p,d} k 2^{k/p} 2^{-(d-1)k/p'} \|f\|_{L^p(\mathbf{R}^d)}$$

and the claim follows as before. ∎

Now we obtain a qualitative consequence of the above theorem.

Theorem 5.6 (Stein's spherical maximal theorem, qualitative version)**.** *Let $d \geq 3$, and let $f \in L^p(\mathbf{R}^d)$ for some $p > d/(d-1)$. (For this theorem, it is important that we* do not *identify functions if they agree almost everywhere.) Then for almost every $x \in \mathbf{R}^d$, the averages $S_r f(x)$ are well-defined and finite for all $r > 0$, are continuous in r, and $\lim_{r \to 0} S_r f(x) = f(x)$.*

Proof We may take f non-negative. Let us first deal with a special case when f is zero almost everywhere and bounded by 1. Then for any ε we can cover the support of f by an open set U of measure at most ε. For any Schwartz function $0 \leq g \leq 1$ supported on U, we know from Stein's maximal inequality that

$$\| \sup_r S_r g \|_{L^p(\mathbf{R}^d)} \lesssim \varepsilon$$

and thus by monotone convergence

$$\| \sup_r S_r 1_U \|_{L^p(\mathbf{R}^d)} \lesssim \varepsilon.$$

Since 1_U pointwise dominates f, we then easily conclude that for almost every x, $S_r f = 0$ for all $r > 0$.

The same claim then clearly follows if f is bounded by some other constant than 1, and then by countable additivity and monotone convergence the same is true for unbounded f also. By subadditivity we conclude that we can modify f on sets of measure zero without affecting the conclusion. In particular we may now assume that f is Borel measurable. This implies that the restriction of f to any sphere is also Borel measurable on that sphere, and so $S_r f(x)$ is well-defined but possibly infinite.

Let us now assume temporarily that f is bounded, so that $S_r f$ is also bounded. Now, a standard limiting argument (approximating f pointwise almost everywhere and in L^p by Schwartz functions, using the preceding discussion to neglect the measure zero set where pointwise convergence fails) using Stein's maximal inequality and dominated convergence shows that

$$\| \sup_{0 < r < R} S_r f \|_{L^p(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}$$

for any $R > 0$ (in particular, the maximal function is mesaurable), and thus by monotone convergence

$$\| \sup_{r > 0} S_r f \|_{L^p(\mathbf{R}^d)} \lesssim \|f\|_{L^p(\mathbf{R}^d)}.$$

If we write f as the suitably rapid L^p and pointwise limit of Schwartz functions f_n, we conclude that for almost every x,

$$\sup_{r>0} |S_r f - S_r f_n|(x) \to 0 \text{ as } n \to \infty$$

which in particular implies that for almost every x, $S_r f(x)$ is continuous in r and converges to $f(x)$.

Finally, we remove the boundedness hypothesis by a monotone convergence argument and yet another application of the Stein maximal inequality. ■

6. Hardy-Littlewood maximal function in high dimensions

For many weeks now we have taken advantage of boundedness properties of the Hardy-Littlewood maximal operator

$$Mf(x) := \sup_{r>0} A_r|f|(x) = \sup_{r>0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |f(y)|\ dy$$

and in particular the weak $(1,1)$ inequality

$$\|Mf\|_{L^{1,\infty}(\mathbf{R}^d)} \lesssim_d \|f\|_{L^1(\mathbf{R}^d)}$$

and the strong L^p inequality

$$\|Mf\|_{L^p(\mathbf{R}^d)} \lesssim_{d,p} \|f\|_{L^p(\mathbf{R}^d)}.$$

For any fixed dimension d, these estimates have many ramifications for various analytical questions on $\mathbf{R}^d$. However, there is the question of how the implicit constants depend on d as $d \to \infty$. The proof of the weak $(1,1)$ estimate (and hence the strong (p,p)) relies ultimately on the Vitali covering lemma and on the doubling properties of balls in $\mathbf{R}^d$. Since the doubling constant is 2^d, the constants in these arguments will also grow exponentially in d. However, it is possible to do better than this.

Firstly, when $p = \infty$, we of course have the trivial estimate

$$\|Mf\|_{L^\infty(\mathbf{R}^d)} \leq \|f\|_{L^\infty(\mathbf{R}^d)}.$$

For $1 < p < \infty$, we can bound the Hardy-Littlewood function by the spherical maximal function. Indeed, from polar coordinates one sees that

$$A_r f(x) = \int_0^r S_{tr} f(x) dt^{d-1}\ dt$$

and so

$$Mf(x) \leq M_S f(x).$$

Thus to bound the Hardy-Littlewood function independently of dimension, it would suffice to do the same for the spherical maximal function. Of course, our estimates for that operator also rely heavily on the dimension. Nevertheless, by using the deceptively simple *method of rotations*, one can obtain universal estimates:

Proposition 6.1. *Let $d_0 \geq 3$ and $p > d_0/(d_0-1)$. Then for all $d \geq d_0$, we have*

$$\|M_S f\|_{L^p(\mathbf{R}^d)} \lesssim_{p,d_0} \|f\|_{L^p(\mathbf{R}^d)}.$$

The point here is that the implied constant stays bounded even in the limit $d \to \infty$.

Proof We embed $\mathbf{R}^{d_0}$ in $\mathbf{R}^d$ in the usual manner. Now we let $O(\mathbf{R}^d)$ be the orthogonal group on $\mathbf{R}^d$, and let ν be the normalised Haar measure on this compact Lie group (thus $\nu(O(\mathbf{R}^d)) = 1$). We claim the rotation formula

$$\int_{S^{d-1}} f(\omega) d\mu^{(d)}(\omega) = \int_{O(\mathbf{R}^d)} \int_{S^{d_0-1}} f(U\omega) d\mu^{(d_0)}(\omega) d\nu(U)$$

for any continuous function f on S^{d-1}, where we use the superscripts to emphasise the ambient dimension. Indeed, both sides are rotation-invariant bounded linear functionals on $C(S^{d-1})$, and by the uniqueness of Haar measure, they must therefore agree up to a constant. Setting $f \equiv 1$ we obtain the identity.

This rotation formula gives us an expression for the d-dimensional spherical average in terms of d_0-dimensional spherical averages:

$$S_r^{(d)} f(x) = \int_{O(\mathbf{R}^d)} \int_{S^{d_0-1}} f(x + rU\omega y)\ d\mu^{(d_0)}(\omega) d\nu(U)$$

and thus

$$M_S^{(d)} f(x) \leq \int_{O(\mathbf{R}^d)} \sup_{r>0} \int_{S^{d_0-1}} |f(x + rU\omega y)|\ d\mu^{(d_0)}(\omega) d\nu(U).$$

By Minkowski's inequality, it thus suffices to show that

$$\| \sup_{r>0} \int_{S^{d_0-1}} |f(x + rU\omega y)|\ d\mu^{(d_0)}(\omega) \|_{L^p(\mathbf{R}^d)} \lesssim_{p,d_0} \|f\|_{L^p(\mathbf{R}^d)}$$

uniformly in U. But by rotating f by U we may set U to be the identity matrix. Now we split $\mathbf{R}^d = \mathbf{R}^{d_0} \times \mathbf{R}^{d-d_0}$ and $x = (x^{(d_0)}, x')$, and observe that

$$\sup_{r>0} \int_{S^{d_0-1}} |f(x + rU\omega y)|\ d\mu^{(d_0)}(\omega) = M_S^{(d_0)} f_{x'}(x^{(d_0)}),$$

where $f_{x'} : \mathbf{R}^{d_0} \to \mathbf{C}$ is the function $f_{x'}(x^{(d_0)}) := f(x^{(d_0)}, x')$. The claim then follows from the Stein's maximal inequality in $\mathbf{R}^{d_0}$ and Fubini's theorem. ∎

Corollary 6.2 (Stein-Stromberg L^p maximal inequality)**.** *For any* $1 < p < \infty$ *and* $d \geq 1$ *we have*

$$\|Mf\|_{L^p(\mathbf{R}^d)} \lesssim_p \|f\|_{L^p(\mathbf{R}^d)}.$$

Proof Let d_0 be the first integer such that $p > d_0/(d_0 - 1)$. The cases $d \leq d_0$ can be handled by the usual Hardy-Littlewood inequality, since d is bounded by $O_p(1)$. The cases $d > d_0$ follow from the previous proposition. ∎

It is still an open question as to whether M is of weak-type $(1,1)$ uniformly in d. The best bound known is $O(d)$, due to Stein and Stromberg; it is based on comparing M with the maximal operator for the Poisson semigroup and using an abstract maximal inequality for semigroups known as the Dunford-Hopf-Schwartz maximal inequality. A more geometric proof based on covering-type lemmas can give a bound of $O(d \log d)$.

7. Exercises

- Q1. Prove Lemma 2.9.
- Q2. (Stein's counterexample) For any $0 < \delta < 1$, let D_δ denote the disk
$$\{(x', x_d) \in \mathbf{R}^{d-1} \times \mathbf{R} : |x'| \leq \delta; |x_d| \leq \delta^2\}.$$
Use the indicator functions of these disks to show that Proposition 5.3 fails when $p < \frac{d}{d-1}$. Then use a suitable linear combination of these indicator functions to show that Proposition 5.3 also fails for $p = \frac{d}{d-1}$. Manipulate this further (by taking linear combinations of translates of these examples) to show that if $p \leq \frac{d}{d-1}$, then one can find a non-negative $f \in L^p(\mathbf{R}^d)$ such that $\limsup_{r \to 0} S_r f(x) = +\infty$ for almost every $x \in \mathbf{R}^d$, which is about as convincing a counterexample to almost everywhere convergence of spherical means as one can hope for.
- Q3. (Weyl bound for the circle problem) In the plane $\mathbf{R}^2$, show that
$$\hat{1_{B(0,1)}}(\xi) \lesssim \langle \xi \rangle^{-3/2}$$
for all $\xi \in \mathbf{R}^2$. Using this, show that
$$|\mathbf{Z}^2 \cap B(0, R)| = \pi R^2 + O(R^{2/3})$$
for all $R \geq 1$. (Hint: let $0 < r < 1$ be chosen later (the optimal value turns out to be $r = R^{-1/3}$) and use the Poisson summation formula to compute
$$\sum_{n \in \mathbf{Z}^2} 1_{B(0,R)} * \phi_r(n)$$
where ϕ_r is a non-negative approximation to the identity supported on $B(0, r)$. By varying R to $R + r$ or $R - r$ you will then get upper and lower bounds on $|\mathbf{Z}^2 \cap B(0, R)|$.)

  The *circle problem* is to reduce the error term as much as possible, ideally to $O_\varepsilon(R^{1/2+\varepsilon})$ (it is known that $O(R^{1/2})$ is not possible). While some fractional improvement over the 2/3 exponent is known, the full problem remains well out of reach of current technology. (It shares some features in common with the Riemann hypothesis, though the latter is undoubtedly more difficult still.)

Department of Mathematics, UCLA, Los Angeles CA 90095-1555

E-mail address: `tao@math.ucla.edu`