Math. 212b Problem set 3

Stationary phase.

March 8, 2001

Contents

1 The lattice point problem.

Let D be a domain in the plane with piecewise smooth boundary. The high-school method of computing the area of D is to superimpose a square grid on the plane and count the number of squares "associated" with D. Since some squares may intersect D but not be contained in D, we must make a choice: let us choose to count all squares which intersect $\overline{D}$. Furthermore, in order to avoid unnecessary notation, let us assume that D is taken to include its boundary, i.e. D is closed: $D = \overline{D}$. If we let $\mathbf{Z}^2$ denote the lattice determined by the corners of our grid, then our procedure is to count the number of points in

$$D \cap \mathbf{Z}^2.$$

Of course this is only an approximation to the area of D. To get better and better approximations we would shrink the size of the grid. Our problem is to find an estimate for the error in this procedure.

For notational reasons, it is convenient to keep the lattice fixed, and dilate the domain D. That is, we want to count the number of lattice points in λD where λ is a (large) positive real number. So we set

$$N_D^\sharp(\lambda) := \#(\lambda D \cap \mathbf{Z}^2). \tag{1}$$

Equally well, if χ^D denotes the indicator function (sometimes called the characteristic function) of D:

$$\chi^D(x) = 1 \ \text{ if } x \in D, \qquad \chi^D(x) = 0 \text{ if } x \notin D,$$

then

$$N_D^\sharp(\lambda) = \sum_{\nu \in \mathbf{Z}^2} \chi_\lambda^D(\nu), \tag{2}$$

where

$$\chi_\lambda(x) := \chi\left(\frac{x}{\lambda}\right).$$

(Frequently, in what follows, we will drop the D when D is fixed. Also, we will pass from 2 to n with the obvious minor changes in notation.)

Now it is clear that

$$N_D^\sharp(\lambda) = \lambda^2 \cdot \text{Area}(D) + \text{ error.}$$

Our problem is to estimate the error. Without any further assumptions, it is relatively easy to see that we can certainly say that the error can be estimated by a constant times λ where the constant involves only the length of ∂D. In general, we can not do better, especially if the boundary of D contains straight line segments of rational slope: For the worst possible scenario, consider the case where D is a square centered at the origin. Then every time that λ is such that the vertices of λD lie in $\mathbf{Z}^2$, then the number of boundary points lying in $\mathbf{Z}^2$ will be proportional to λ times the length of the perimeter of D. But

a slightly larger or small value of λ will yield no boundary points in $\mathbf{Z}^2$. We might expect that if the boundary is curved everywhere, we can improve on the estimate of the error.

The main result of this problem set, due to Van der Corput, asserts that if D is convex, with smooth boundary whose curvature is everywhere positive (we will give more precise definitions later) then we can estimate the error terms as being

$$O(\lambda^{\frac{2}{3}}).$$

In fact, Van der Corput shows that this result is sharp if we allow all such strongly convex smooth domains, although we will not establish this result here.

Suppose that we take D to be the unit disk. In this case

$$N_D^\sharp(\lambda) = N(\lambda)$$

where

$$N(\lambda) = \#\{\nu = (m,n) \in \mathbf{Z}^2 | m^2 + n^2 \leq \lambda^2\}. \tag{3}$$

In this case, there will only be lattice points on the boundary of λD if λ^2 is an integer which can be represented as a sum of two squares, and the number of points on the boundary will be the number of ways of representing λ as a sum of two squares. The number of ways of representing an integer n as the sum of two integer squares is closely related to the number of prime factors of n of the form $4k+1$ and the number of prime square factors of the form $4k+3$. In fact, as we shall remind you later on, if $r(n)$ denotes the number of ways of writing n as a sum of two squares then $r(n)$ can be evaluated as follows: Suppose we factorize n into prime powers, collect all the powers of 2, collect all the primes congruent to 1 (mod 4), and collect all the primes which are congruent to 3 (mod 4). In other words, we write

$$n = 2^f n_1 n_2 \tag{4}$$

where

$$n_1 = \prod p^r \quad p \equiv 1 (\text{mod } 4)$$

and

$$n_2 = \prod q^s \quad q \equiv 3 (\text{mod } 4).$$

Then $r(n) = 0$ if any s is odd. If all the s are even, then

$$r(n) = 4d(n_1). \tag{5}$$

So there are relatively few points on the boundary of λD when D is the unit disk, and we might expect special results in this case. Of course our problem is to estimate the number of lattice points close to a given circle, not necessarily exactly on it.

Let us set

$$t := \lambda^2, \tag{6}$$

as the square of λ is the parameter used frequently in the number theoretical literature. Let us define $R(t)$ as the error in terms of t, so

$$\sum_{n\leq t} r(n) = \pi t + R(t). \tag{7}$$

Then the result of Van der Corput cited above asserts that

$$R(t) = O(t^{\frac{1}{3}}). \tag{8}$$

In fact, later work of Van der Corput himself in the twenties and early thirties, involving the theory of "exponent pairs" improves upon this estimate. For example, one consequence of the method of "exponent pairs" is that

$$R(t) = O(t^{\frac{27}{82}}). \tag{9}$$

In fact, the long standing conjecture (going back to Gauss, I believe) has been that

$$R(t) = O(t^{\frac{1}{4}+\epsilon}) \quad \text{for any} \quad \epsilon > 0. \tag{10}$$

Notice the sequence of more and more refined results: trivial arguments, valid for any region with piecewise smooth boundary give an estimate $R(t) = O(t^\rho)$ where $\rho = \frac{1}{2}$. The Van der Corput method valid for all smooth strongly convex domains gives $\rho = \frac{1}{3}$. The method of exponent pairs gives $\rho = (k+\ell)/(2k+2)$ whenever (k,ℓ) is an exponent pair, but although this method improved on $\frac{1}{3}$, it did not yield the desired conjecture - that we may take $\rho = \frac{1}{4} + \epsilon$ for any $\epsilon > 0$.

2 The divisor problem.

Let $d(n)$ denote the number of divisors of the positive integer n. Using elementary arguments, Dirichlet (1849) showed that

$$\sum_{n\leq t} d(n) = t(\log t + 2\gamma - 1) + O(t^{\frac{1}{2}}) \tag{11}$$

where γ is Euler's constant

$$\gamma := \lim_{N\to\infty} \left(\sum_{n\leq N} \frac{1}{n} - \log N \right).$$

Dirichlet's argument is as follows: First of all observe that we can regard the divisor problem as a lattice point counting problem. Indeed, consider the region, T_t, in the (x,y) plane bounded by the hyperbola $xy = t$ and the straight line segments from $(1,1)$ to $(1,t)$ and from $(1,1)$ to $(t,1)$. So T_t is a "triangle" with the hypotenuse replaced by a hyperbola. Then $d(n)$ is the number of lattice points on the "integer hyperbola" $xy = n,\ n \leq t$, and so $\sum_{n\leq t} d(n)$ is the total

number of lattice points in T_t. The area of T_t is $t\log t - t + 1$, which has the same leading term as above. To count the number of lattice points in T_t, observe that T_t is symmetric about the line $y = x$, and there are $[\sqrt{t}]$ lattice points in T_t on this line. For each integer $d \leq [\sqrt{t}]$ the number of lattice points on the horizontal line) $y = d$ in T_t to the right of the diagonal is

$$\left[\frac{t}{d}\right] - d$$

so

$$\sum_{n\leq t} d(n) = 2\sum_{d\leq\sqrt{t}}\left(\left[\frac{t}{d}\right] - d\right) + \left[\sqrt{t}\right].$$

Since $[s] = s + O(1)$ we can write this as

$$2t\sum_{d\leq\sqrt{t}}\frac{1}{d} - 2\cdot\frac{\sqrt{t}(\sqrt{t}+1)}{2} + O(\sqrt{t}).$$

The formula leading to Euler's constant has error term $1/s$:

$$\sum_{n\leq s}\frac{1}{d} = \log s + \gamma + O(\frac{1}{s}) \tag{12}$$

as follows from Euler MacLaurin (see later on). So setting $s = \sqrt{t}$ in the above we get (11).

Once again we may ask if this estimate can be improved: Define

$$\Delta(t) := \sum_{n\leq t} d(n) - t(\log t + 2\gamma - 1) \tag{13}$$

and ask for better σ such that

$$\Delta(t) = O(t^\sigma) \tag{14}$$

It turns out, that the method of exponent pairs yields the same answer as in the circle problem case: If (k, ℓ) is an exponent pair (to be defined, in Chapter **??**) then

$$\sigma = (k + \ell)/(2k + 2)$$

is a suitable exponent in (14). Once again, the conjectured theorem has been that we may take $\sigma = \frac{1}{4} + \epsilon$ for any positive ϵ.

These "lattice point problems" are closely related to studying the growth of the Riemann zeta function on the critical line, i.e. to obtain power estimates for $\zeta(\frac{1}{2} + it)$. Furthermore, the Riemann hypothesis itself is known to be closely related to soemwhat deeper "approximation" problems. See, for example, the book *Area, Lattice Points, and Exponential Sums* by M.N Huxley, page 15.

3 Stationary phase.

Van der Corput revolutionized the study of the lattice point problem in the 1920's by bringing to bear two classical tools of analysis - the Poisson summation formula and the method of stationary phase. In this section we state a version of the stationary phase formula valid for computing exponential integrals over compact manifolds.

Let M be a smooth compact n-dimensional manifold, and let ψ be a smooth real valued function defined on M. A point $p \in M$ is called a *critical point* of ψ if $d\psi(p) = 0$. This means that $(X\psi)(p) = 0$ for any vector field X on M, and if X itself vanishes at p then $X\phi$ vanishes at p "to second order" in the sense that $YX\psi$ vanishes at p for any vector field Y. Thus $(YX\psi)(p)$ depends only on the value $X(p)$. Furthermore

$$(XY\psi)(p) - (YX\psi)(p) = ([X, Y]\psi)(p) = 0$$

so we get a well defined symmetric bilinear form on the tangent space TM_p called the **Hessian** of ψ at p and denoted by $d_p^2\psi$. For any pair of tangent vectors $v, w \in TM_p$ it is given by

$$d_p^2\psi(p)(v, w) := (XY\psi)(p)$$

where X and Y are any vector fields with

$$X(p) = v, \quad Y(p) = w.$$

A critical point p is called **non-degenerate** if this symmetric bilinear form is non-degenerate. We can then talk of the signature of the quadratic form $d_p^2\psi$ – i.e. the number of +'s minus the number of -'s when we write $d_p^2\psi$ in canonical form as a sum of $\pm(x^i)^2$ where the x^i form an appropriate basis of TM_p^*. We will write this signature as $\operatorname{sgn} d_p^2\psi$ or more simply as $\operatorname{sgn}_p \psi$. The symmetric bilinear form $d_p^2\psi$ determines a symmetric bilinear form on all the exterior powers of TM_p, in particular on the highest exterior power, $\wedge^n TM_p$. This then in turn defines a density at p, assigning to every basis $v_1, \ldots, v_n$ of TM_p the number

$$|d_p^2(\psi)(v_1 \wedge \cdots \wedge v_n, v_1 \wedge \cdots \wedge v_n)|^{\frac{1}{2}}.$$

Replacing $v_1, \ldots, v_n$ by $Av_1, \ldots, Av_n$ has the effect of multiplying the above number by $|\det A|$ which is the defining property of a density. In particular, if we are given some other positive density at p the quotient of these two densities is a number, which we will denote by

$$|\det d_p^2\psi|^{\frac{1}{2}},$$

the second density being understood. The reason for this somewhat perverse notation is as follows: Suppose, as we always can, that we have introduced

coordinates $y^1, \ldots, y^n$ at p such that our second density assigns the number one to the the basis

$$v_1 = \left(\frac{\partial}{\partial y^1}\right)_p, \ \ldots \ v_n = \left(\frac{\partial}{\partial y^n}\right)_p.$$

Then

$$d_p^2(\psi)(v_1 \wedge \cdots \wedge v_n, v_1 \wedge \cdots \wedge v_n) = \det\left(\frac{\partial^2 \psi}{\partial y^i \partial y^j}\right)(p)$$

so

$$|\det d_p^2 \psi|^{\frac{1}{2}} = \left|\det\left(\frac{\partial^2 \psi}{\partial y^i \partial y^j}\right)(p)\right|^{\frac{1}{2}}.$$

With these notation preliminaries we can now state a version of the formula of stationary phase. Suppose we are given a positive density, Ω, on M and that all the critical points of ψ are non-degenerate (so that there are only finitely many of them). Then for any smooth function a on M we have

$$\int_M e^{i\tau\psi} a\Omega = \left(\frac{2\pi}{\tau}\right)^{\frac{n}{2}} \sum_{p | d\psi(p)=0} e^{\frac{1}{4}\pi i \operatorname{sgn}_p \psi} \frac{e^{i\tau\psi(p)} a(p)}{|\det d_p^2 \psi|^{\frac{1}{2}}} + O(\tau^{-\frac{n}{2}-1}) \qquad (15)$$

as $\tau \to \infty$. We will prove this formula after we make use of it to prove Van der Corput's theorem. The proof will yield an actual asymptotic formula, of which the first expression on the right in (15) is leading term.

But in our application, we will not need even this leading term - all we will need to know is that the right hand side is $O(\tau^{-\frac{n}{2}})$. In fact, suppose that ψ depends on some auxiliary parameter, u, and that we can bound the sum uniformly in u. For example, suppose that the number of critical points is bounded uniformly in u and that the denominators occurring in the sum are uniformly bounded from below. Then we may conclude (from the proof of (15)) that

$$\int_M e^{i\tau\psi} a\Omega = O(\tau^{-\frac{n}{2}})$$

this estimate being uniform in u.

Our application will be of the following nature: Recall that a subset of $\mathbf{R}^n$ is convex if it is the intersection of all the half spaces containing it. Suppose that D is a (compact) convex domain with smooth boundary, containing the origin and that u is a unit vector. Then the function $y \mapsto u \cdot y$ achieves a maximum m^+ and a minimum m^- on D and the condition that these be taken on at exactly one point each is what is usually meant by saying that D is "strictly convex". We want to impose the stronger condition: that

the restriction of the function $y \mapsto u \cdot y$ to the boundary is non-degenerate having only two critical points, the maximum and the minimum, for all unit vectors.

This has the following consequence: Let K be a compact subset of $\mathbf{R}^n - \{0\}$ and consider the Fourier transform of the indicator function $\chi = \chi^D$ evaluated

at τx for $x \in K$:

$$\hat{\chi}(\tau x) = \int_D e^{i\tau x\cdot y} dy.$$

1. Show that

$$\hat{\chi}(\tau x) = O(\tau^{-\frac{n+1}{2}}) \tag{16}$$

uniformly for $x \in K$ where K is any compact subset of $\mathbf{R}^n - \{0\}$. [Hint:Show that

$$e^{i\tau x\cdot y} dy = e^{i\tau x\cdot y} dy^1 \wedge \cdots \wedge dy^n = \frac{1}{i\tau |x|^2} d\left(e^{i\tau x\cdot y}\omega\right)$$

where

$$\omega := x^1 dy^2 \wedge \cdots \wedge dy^n - x^2 dy^1 \wedge dy^3 \cdots \wedge dy^n + \cdots \pm x^n dy^1 \wedge \cdots \wedge dy^{n-1}.$$

Apply Stokes, and stationary phase.]

As property (16) is what we will use, we might as well take (16) as the definition of a **strongly convex region**.

4 Poisson summation.

The second theorem from classical analysis that goes into the proof of Van der Corput's theorem is the Poisson summation formula. This says that if f is a smooth function vanishing rapidly with its derivatives at infinity on $\mathbf{R}^n$ then

$$\sum_{\mu\in\mathbf{Z}^n} \hat{f}(2\pi\mu) = \sum_{\nu\in\mathbf{Z}^n} f(\nu). \tag{17}$$

We recall the elementary proof of this fact from last semseter:

Set

$$h(x) := \sum_{\nu\in\mathbf{Z}^n} f(x+\nu)$$

so that h is a smooth periodic function with period the unit lattice, $\mathbf{Z}^n$. By definition

$$h(0) = \sum_{\nu\in\mathbf{Z}^n} f(\nu).$$

Since h is periodic, we may expand it into a Fourier series

$$h(x) = \sum_{\mu\in\mathbf{Z}^n} c_\mu e^{-2\pi i\mu\cdot x}$$

where

$$c_\mu = \int_0^1 \cdots \int_0^1 h(x) e^{2\pi i\mu\cdot x} dx = \int_0^1 \cdots \int_0^1 \sum_{\nu\in\mathbf{Z}^n} f(x+\nu) e^{2\pi i\mu\cdot x} dx.$$

We may interchange the order of summation and integration and make the change of variables $x + \nu \mapsto x$ to obtain

$$c_\mu = \hat{f}(2\pi\mu).$$

Setting $x = 0$ in the Fourier series

$$h(x) = \sum_{\mu \in \mathbf{Z}^n} \hat{f}(2\pi\mu) e^{-2\pi i \mu \cdot x}$$

gives

$$h(0) = \sum_{\mu \in \mathbf{Z}^n} \hat{f}(2\pi\mu).$$

Equating the two expressions for $h(0)$ is (17).

5 Van der Corput's theorem.

In n-dimensions this says:

Theorem 5.1 *Let D be a strongly convex domain. Then*

$$N_D^\sharp(\lambda) = \lambda^n \operatorname{vol}(D) + O(\lambda^{n-2+\frac{2}{n+1}}) \tag{18}$$

Proof. Let $\chi = \chi^D$ be the indicator function of D so that χ_λ defined by

$$\chi_\lambda(y) := \chi(\frac{y}{\lambda})$$

is the indicator (characteristic) function of λD. Thus

$$N^\sharp(\lambda) = \sum_{\nu \in \mathbf{Z}^n} \chi_\lambda(\nu)$$

where we have written $N^\sharp$ for $N_D^\sharp$. The Fourier transform of χ_λ is given in terms of the Fourier transform of χ by

$$\hat{\chi_\lambda}(x) = \lambda^n \hat{\chi}(\lambda x).$$

Furthermore,

$$\hat{\chi}(0) = \operatorname{vol}(D).$$

If we could apply the Poisson summation formula directly to χ_λ then the contribution from 0 would be $\lambda^n \operatorname{vol}(D)$, and we might hope to control the other terms using (16). (For example, if we could brutally apply (16) to control *all* the remaining terms in the case of the circle, we would be able to estimate the error in the circle problem as $\lambda^{2-3/2} = \lambda^{1/2}$ which is the circle conjecture.) But this will not work directly since χ_λ is not smooth. We must first " regularize" χ_λ and the clever idea will be to choose this regularization to depend the right way on λ.

So let ρ be a non-negative smooth function on $\mathbf{R}^n$ supported in the unit ball with integral one. Let

$$\rho_\epsilon(y) = \frac{1}{\epsilon^n}\rho\left(\frac{y}{\epsilon}\right)$$

so ρ_ϵ is supported in the ball of radius ϵ and has total integral one. Thus

$$\hat{\rho_\epsilon}(x) = \hat{\rho}(\epsilon x)$$

and

$$\hat{\rho}(0) = 1.$$

Define

$$N_\epsilon^\sharp(\lambda) = \sum_{\nu\in\mathbf{Z}^n} (\chi_\lambda \star \rho_\epsilon)(\nu)$$

where $\star$ denotes convolution. If ν lies a distance greater than ϵ from the boundary of λD, then $(\chi_\lambda \star \rho_\epsilon)(\nu) = \chi_\lambda(\nu)$. Thus

$$N_\epsilon^\sharp(\lambda - C\epsilon) \leq N^\sharp(\lambda) \leq N_\epsilon^\sharp(\lambda + C\epsilon)$$

where C is some constant depending only on D. Suppose we could prove that $N_\epsilon^\sharp$ satisfies an estimate of the type (18). Then we could conclude that

$$(\lambda - C\epsilon)^n \operatorname{vol}(D) + O(\lambda^{n-2+\frac{2}{n+1}}) \leq N^\sharp(\lambda) \leq (\lambda + C\epsilon)^n + O(\lambda^{n-2+\frac{2}{n+1}}).$$

Suppose we set

$$\epsilon = \lambda^{-1+\frac{2}{n+1}}. \tag{19}$$

Then

$$(\lambda \pm C\epsilon)^n = \lambda^n + O(\lambda^{n-2+\frac{2}{n+1}})$$

and we obtain the Van der Corput estimate for $N^\sharp(\lambda)$. So it is enough to prove the analogue of (18) with $N_\epsilon^\sharp$ watching out for the dependence on ϵ.

Since $\chi_\lambda \star \rho_\epsilon$ is smooth and of compact support, and since

$$(\chi_\lambda \star \rho_\epsilon)\hat{} = \hat{\chi_\lambda} \cdot \hat{\rho_\epsilon}$$

we may apply the Poisson summation formula to conclude that

$$N_\epsilon^\sharp(\lambda) = \lambda^n \operatorname{vol}(D) + \sum_{\nu\in\mathbf{Z^n}-\{\mathbf{0}\}} \lambda^n \hat{\chi}(2\pi\lambda\nu)\hat{\rho}(2\pi\epsilon\nu)$$

and we must estimate the sum on the right hand side. Now since ρ is of compact support its Fourier transform vanishes faster than any inverse power of $(1+|x|^2)$. So, using (16) we can estimate this sum by

$$\lambda^{n-\frac{n+1}{2}} \sum_{\nu\in\mathbf{Z}^n-\{0\}} |\nu|^{-\frac{n+1}{2}}(1+|\epsilon\nu|^2)^{-K}$$

were K is large, or, what is the same by

$$\lambda^{\frac{n-1}{2}} \int \frac{1}{|x|^{\frac{n+1}{2}}} (1+|\epsilon x|^2)^{-K} dx$$

where K is large. Making the change of variables $x = \epsilon z$ this becomes

$$\lambda^{\frac{n-1}{2}} \epsilon^{-\frac{n-1}{2}} \int \frac{1}{|z|^{\frac{n+1}{2}}} (1+|z|)^{-K} dz.$$

The integral does not depend on anything, and if we substitute (19) for ϵ, the power of λ that we obtain is

$$\frac{n-1}{2} - \frac{n-1}{2}\left(-1+\frac{2}{n+1}\right) = \frac{n-1}{2} + \frac{n-1}{2} - \frac{n+1}{n+1} + \frac{2}{n+1} = n-2+\frac{2}{n+1}$$

proving (18). □

6 Baby Euler-MacLaurin.

We will spend a lot of time later on this semseter studying variants of the Euler-MacLaurin summation formula which estimate a sum over lattice points in terms of an integral. In this section we give an elementary version which involves a minimum of notation. For any real number t we denote the largest integer $\leq t$ by $\lfloor t \rfloor$ and $\lceil t \rceil$ denotes the smallest integer $\geq t$. We will consider integration by parts against the saw tooth function

$$t \mapsto t - \lfloor t \rfloor :$$

Proposition 6.1 *If f has a continuous derivative on the interval $[a,b]$ where $0 < a < b$ then*

$$\sum_{a<k\leq b} f(k) = \int_a^b f(t)dt + \int_a^b (t-\lfloor t \rfloor) f'(t)dt + f(b)(\lfloor b \rfloor - b) - f(a)(\lfloor a \rfloor - a). \quad (20)$$

where the sum on the left ranges over all integers in the interval $(a,b]$.

Proof. For $\lceil a \rceil + 1 \leq k \leq b$, we have

$$\int_{k-1}^{k} \lfloor t \rfloor f'(t)dt = (k-1)\int_{k-1}^{k} f'(t)dt = (k-1)(f(k)-f(k-1))$$

$$= kf(k) - (k-1)f(k-1) - f(k).$$

Summing from over this range of k the first terms on the right telescope and we get

$$\int_{\lceil a \rceil}^{\lfloor b \rfloor} \lfloor t \rfloor f'(t)dt = \lfloor b \rfloor f(\lfloor b \rfloor) - \lceil a \rceil f(\lceil a \rceil) - \sum_{\lceil a \rceil < k \leq b} f(k).$$

We have

$$\begin{aligned}\int_a^b &= \int_{\lceil a\rceil}^{\lfloor b\rfloor} + \int_a^{\lceil a\rceil} + \int_{\lfloor b\rfloor}^b \\ \int_{\lfloor b\rfloor}^b \lfloor t\rfloor f'(t)dt &= \lfloor b\rfloor\,(f(b) - f(\lfloor b\rfloor)) \\ \int_a^{\lceil a\rceil} \lfloor t\rfloor f'(t)dt &= (\lceil a\rceil - 1)(f(\lceil a\rceil) - f(a)) \\ &= \lceil a\rceil f(\lceil a\rceil) - (\lceil a\rceil - 1)f(a) - f(\lceil a\rceil).\end{aligned}$$

If a is not an integer, then $\lceil a\rceil - 1 = \lfloor a\rfloor$ and the term $-f(\lceil a\rceil)$ gets added to the sum over $\lceil a\rceil < k \le b$ to yield a sum over $a < k \le b$. If a is an integer, then $\lceil a\rceil = \lfloor a\rfloor$ so no term is added, but the original sum is over the range $a < k \le b$. In either case we get

$$\sum_{a<k\le b} f(k) = -\int_a^b \lfloor t\rfloor f'(t)dt + \lfloor b\rfloor f(b) - \lfloor a\rfloor f(a).$$

But integration by parts gives

$$-\int_a^b f(t)dt = -bf(b) + af(a) + \int_a^b tf'(t)dt.$$

So adding and subtracting $\int_a^b f(t)dt$ to the preceding expression for $\sum f(k)$ gives (20). □

6.1 Application to the Euler constant.

As an application, let us apply (20) to the function

$$F(x,s) = \begin{cases} \frac{x^{1-s}}{1-s} & \text{if } s \ne 1 \\ \\ \log x & \text{if } s = 1 \end{cases} \tag{21}$$

so that

$$F'(x,s) = x^{-s}$$

for all s. Let us first consider the case $s = 1$ and so take $f(t) = 1/t$ in (20) and $(a,b] = (1,x]$. We get

$$\begin{aligned}\sum_{1\le n\le x} \frac{1}{n} &= 1 + \int_0^x \frac{dt}{t} - \int_1^x \frac{(t - \lfloor t\rfloor)}{t^2}dt + \frac{\lfloor x\rfloor - x}{x} \\ &= 1 + \log x - \int_1^x \frac{t - \lfloor t\rfloor}{t^2}dt + O(1/x)\end{aligned}$$

$$
\begin{aligned}
&= 1 + \log x - \int_1^\infty \frac{t - \lfloor t \rfloor}{t^2} dt + \int_x^\infty \frac{t - \lfloor t \rfloor}{t^2} dt + O(1/x) \\
&= 1 + \log x - \int_1^\infty \frac{t - \lfloor t \rfloor}{t^2} dt + O(1/x) \\
&= \log x + C + O(1/x)
\end{aligned}
$$

where

$$C := 1 + \int_1^\infty \frac{\lfloor t \rfloor - t}{t^2} dt$$

the integral on the right being convergent, since the integrand is $O(1/t^2)$. This proves (12) with Euler's constant evaluated as

$$\gamma = 1 + \int_1^\infty \frac{\lfloor t \rfloor - t}{t^2} dt. \tag{22}$$

6.2 Application to the zeta function.

Let us now do the same computation with Re $s > 0$, $s \neq 1$. We get

$$
\begin{aligned}
\sum_{n \leq x} \frac{1}{n^s} &= 1 + \int_1^x \frac{dt}{t^s} - s \int_1^x \frac{t - \lfloor t \rfloor}{t^{s+1}} dt - \frac{x - \lfloor x \rfloor}{x^s} \\
&= \frac{x^{1-s}}{1-s} - \frac{1}{1-s} + 1 - s \int_1^\infty \frac{t - \lfloor t \rfloor}{t^{s+1}} + O(x^{-\sigma})
\end{aligned}
$$

where $\sigma = \text{Re } s$. So

$$\sum_{1 \leq n \leq x} \frac{1}{n^s} = \frac{x^{1-s}}{1-s} + C(s) + O(x^{-\sigma}) \tag{23}$$

where

$$C(s) := 1 - \frac{1}{1-s} - s \int_1^\infty \frac{t - \lfloor t \rfloor}{t^{s+1}} dt.$$

Notice that $C(s)$ is meromorphic in the half plane $\sigma > 0$ with a simple pole of residue 1 at $s = 1$. Also notice that if $\sigma > 1$ the series on the left of (23) converges and so we may let $x \to \infty$ and conclude that $C(s) = \zeta(s)$, the Riemann zeta function. Therefore we see that $C(s)$ is an analytic continuation of the Riemann zeta function and in fact

$$\zeta(s) = \lim_{x \to \infty} \left(\sum_{1 \leq n \leq x} \frac{1}{n^s} - \frac{x^{1-s}}{1-s} \right) \tag{24}$$

for all

$$\text{Re } s > 0, \quad s \neq 1.$$

so we can rewrite (23) as

$$\sum_{1\le n\le x}\frac{1}{n^s}=\frac{x^{1-s}}{1-s}+\zeta(s)+O(x^{-\sigma}) \tag{25}$$

for all

$$\text{Re } s>0,\qquad s\neq 1.$$

6.3 Application to sums of divisor powers.

Observe that if $s=-\alpha$ where $\alpha\ge 0$ then (20) gives

$$\begin{aligned}\sum_{n\le x}n^{\alpha} &= 1+\int_1^x t^{\alpha}dt+\alpha\int_1^x t^{\alpha-1}(t-\lfloor t\rfloor)dt-(x-\lfloor x\rfloor)x^{\alpha}\\ &= \frac{x^{\alpha+1}}{\alpha+1}+O(x^{\alpha})\end{aligned}$$

so

$$\sum_{n\le x}n^{\alpha}=\frac{x^{\alpha+1}}{\alpha+1}+O(x^{\alpha})\quad\text{for }\alpha\ge 0. \tag{26}$$

For any complex number α define the function σ_α on $\mathbf{N}$ by

$$\sigma_\alpha(n):=\sum_{d|n}d^{\alpha}. \tag{27}$$

Thus $\sigma_0(n)=d(n)$, the number of divisors of n and we can ask for analogues of Dirichlet's formula (11), i.e. estimates on

$$\sum_{n\le t}\sigma_\alpha(n).$$

This is again a sum over the lattice points in the triangular region T_t, but this time we are summing the function x^α over the lattice points. For example, if we take $\alpha=1$ and sum successively over the horizontal lines $y=d$ we get

$$\begin{aligned}\sum_{n\le t}\sigma_1(n) &= \sum_{d\le t}\sum_{q\le t/d}q\\ &= \sum_{d\le t}\left[\frac{1}{2}\left(\frac{t}{d}\right)^2+O\left(\frac{t}{d}\right)\right]\quad\text{by (26) with}\quad\alpha=1\\ &= \frac{t^2}{2}\sum_{d\le t}\frac{1}{d^2}+O\left(t\sum_{d\le t}\frac{1}{d}\right)\\ &= \frac{t^2}{2}\left[-\frac{1}{t}+\zeta(2)+O(t^{-2})\right]+O(t\log t)\\ &\qquad\text{by (25) with }\ s=2\ \text{ and (12)}\\ &= \frac{1}{2}\zeta(2)t^2+O(t\log t).\end{aligned}$$

So we have proved

$$\sum_{n\le t}\sigma_1(n)=\frac{1}{2}\zeta(2)t^2+O(t\log t). \tag{28}$$

Let us now do the same computation with $\alpha>0,\ \alpha\neq 1$:

$$\begin{aligned}\sum_{n\le t}\sigma_\alpha(n) &= \sum_{d\le t}\sum_{q\le t/d} q^\alpha\\ &= \sum_{d\le t}\left[\frac{1}{\alpha+1}\left(\frac{t}{d}\right)^{\alpha+1}+O\left(\frac{t^\alpha}{d^\alpha}\right)\right] \text{ by (26)}\\ &= \frac{t^{\alpha+1}}{\alpha+1}\left[\frac{t^{-\alpha}}{-\alpha}+\zeta(\alpha+1)+O(t^{-\alpha-1})\right]\\ &\quad +O\left(t^\alpha\left[\frac{t^{1-\alpha}}{1-\alpha}+\zeta(\alpha)+O(t^{-\alpha})\right]\right)\\ &\quad \text{by (25) with } s=-1-\alpha \text{ and } s=-\alpha\\ &= \frac{1}{\alpha+1}\zeta(\alpha+1)x^{\alpha+1}+O(x)+O(1)+O(x^\alpha),\end{aligned}$$

so

$$\sum_{n\le t}\sigma_\alpha(n)=\frac{1}{\alpha+1}\zeta(\alpha+1)t^{\alpha+1}+O(t^\beta),\quad \text{where}\quad \beta=\max(1,\alpha) \tag{29}$$

when

$$\alpha>0,\ \alpha\neq 1.$$

6.4 Using the function ψ.

For various reasons, including the generalization of the formula (20) to the full fledged Euler MacLaurin formula it is convenient make a slight change in the integrand of the second term on the right in (20) replacing the function $t\mapsto t-\lfloor t\rfloor$ by ψ where

$$\psi(t):=t-\lfloor t\rfloor-\frac{1}{2}. \tag{30}$$

When we recall the definition of the Bernoulli polynomials $B_m(x)$, we will see that

$$B_1(x)=x-\frac{1}{2}.$$

If we let $\{t\}:=t-\lfloor t\rfloor$ denote the fractional part of t then we can restate the preceding definition as

$$\psi(t)=B_1(\{t\}). \tag{31}$$

We can rewrite (20) as

$$\sum_{a<k\le b}f(k)=\int_a^b f(t)dt+\int_a^b \psi(t)f'(t)dt$$

$$+\frac{1}{2}\left(f(b)-f(a)\right)+f(b)(\lfloor b\rfloor - b)-f(a)(\lfloor a\rfloor - a). \tag{32}$$

Another slight variation on this formula is as follows: Assume that a and b are integers, so that the last two terms above disappear, and let the sum on the left extend over $a \leq k < b$ so we must add the term $-f(b)+f(a)$ to the right. We get

$$\sum_a^{b-1} f(k) = \int_a^b f(t)dt - \frac{1}{2}(f(b)-f(a)) + \int_a^b B_1(\{t\})f'(t)dt. \tag{33}$$

6.5 The next order term.

We will discuss the full version of the Euler Maclaurin formula later on in class, where I hope to be able to discuss recent delicate higher dimensional generalizations. Basically what is involved is successive integration by parts. To illustrate, let us go over the argument which led to (12) using the function ψ but carrying the integration by parts out one further step so as to get a right hand side with an error which is $O(s^{-2})$. Since $\lfloor u\rfloor = u - \psi(u) + \frac{1}{2}$ and $d\lfloor u\rfloor = \sum \delta_n$ we have

$$\sum_{n\leq s}\frac{1}{n} = \int_{1^-}^s d\lfloor u\rfloor.$$

We will define

$$\Psi(u) := \int_u^1 \psi(r)dr$$

so that Ψ is continuous, periodic with period one, and $\Psi(1)=0$. We integrate by parts

$$\begin{aligned}
\sum_{n\leq s}\frac{1}{n} &= \int_{1^-}^s d\lfloor u\rfloor \\
&= \int_{1^-}^s \frac{1}{u}d(u-\psi(u)) \\
&= \log s + \frac{1}{2} - \int_1^\infty (\psi(u)/u^2)du - \psi(s)/s + \int_s^\infty (\psi(u)/u^2)du \\
&= \log y + \gamma - \psi(s)/s + \int_s^\infty (\psi(u)/u^2)du \\
&= \log y + \gamma - \psi(s)/s - \Psi(s)/s^2 + 2\int_s^\infty (\Psi(u)/u^3)du
\end{aligned}$$

so

$$\sum_{n\leq s}\frac{1}{n} = \log y + \gamma - \frac{\psi(y)}{y} + O\left(\frac{1}{s^2}\right). \tag{34}$$

7 Evaluating $r(n)$.

7.1 $r(p) = 0$ if p is a prime $\equiv 3 \pmod 4$.

A solution of $x^2 + y^2 = p$ implies a non-trivial solution of

$$a^2 + b^2 \equiv 0 \mod p$$

which means (if $a \not\equiv 0 \mod p$ say) that $u \equiv b/a$ is an element of $F_p = \mathbf{Z}/p\mathbf{Z}$ satisfying

$$u^2 = -1.$$

But the group of non-zero elements of F_p is cyclic of order $p-1$, and this equation implies that u is an element of order 4 which is impossible since 4 does not divide $p-1$ by assumption.

This same argument shows that we *can* solve $a^2 + b^2 \equiv 0 \mod p$ if $p \equiv 1 \pmod 4$. In the next subsection we will show that we can find an actual integer solution of $x^2 + y^2 = p$ in this case.

We can strengthen the argument a bit to get

Lemma 7.1 *If $p|n$ and $p \equiv 3 \pmod 4$ then there is no integer solution of*

$$x^2 + y^2 = n, \quad (x, y) = 1.$$

Indeed, if $p|n$ and $p|x$ then $p|y$ and so $(x, y) \neq 1$. Thus p does not divide either x or y. Therefore

$$x^{p-1} \equiv 1 \pmod p$$

so

$$yx^{p-1} \equiv y \pmod p.$$

Setting

$$z := yx^{p-2}$$

we get

$$xz \equiv y \pmod p$$

so

$$x^2(z^2 + 1) \equiv x^2 + y^2 = n \equiv 0 \pmod p$$

or

$$z^2 \equiv -1 \pmod p$$

which is impossible. □

We can conclude that

Proposition 7.1 *If $p \equiv 3 \pmod 4$ is a divisor of n which occurs to an odd power in the factorization of n then $r(n) = 0$.*

Proof. Suppose that $x^2 + y^2 = n$ and $(x, y) = d$ where x and y are integers, and let r be such that $p^r|d$ but $p^{r+1} \not| d$. Write $x = da, y = db$ with $(a, b) = 1$ so $n = d^2(a^2 + b^2)$ and let $m := a^2 + b^2$. Then $p \not| m$ by the lemma. But if s is the highest power of p that divides n, then $p^{s-2r}|m$. So $s - 2r = 0$ or s is even. □

7.2 Minkowski's argument.

Suppose that $p \equiv 1 \pmod 4$, let u be such that $u^2 \equiv -1 \pmod p$, and let $L \subset \mathbf{Z}^2$ consist of all integer lattice points of the form (a, b) where $b \cong ua \pmod p$. So L is a lattice in the plane, in fact a subgroup of $\mathbf{Z}^2$ of index p. Any fundamental domain of the plane relative to this lattice has volume p, and every point $(a, b) \in L$ satisfies $a^2 + b^2 \equiv 0 \pmod p$. Suppose we could show that the disk

$$D_r : \{(x, y) | x^2 + y^2 \leq r^2\}$$

contains a non-zero point, (a, b) of L when

$$r^2 < 2p.$$

Then $a^2 + b^2$ is a non-zero multiple of p and is less than $2p$. so we must have $a^2 + b^2 = p$ showing that $r(p) > 0$. Now we can choose $r^2 < 2p$ such that the area of D_r is $> 4p$, since this area is πr^2 and $\pi > 3$. So our result is a consequence of the following famous theorem of Minkowski:

Theorem 7.1 Minkowski *Let L be a lattice in $\mathbf{R}^n$ whose fundamental domain has volume V. Let C be a bounded convex set which is symmetric with respect to the origin, and such that*

$$vol(C) > 2^n V.$$

Then C contains a non-zero element of L.

Proof. The torus $T^n := \mathbf{R}^n/2L$ has volume $2^n V$. The natural projection of $\mathbf{R}^n :\to T^n$ is locally an isometry (hence locally volume preserving) and so the restriction of this projection to C can not be globally injective. In other words, there exist distinct points $w, z \in C$ such that $z - w \in 2L$. By symmetry, $y := -w \in C$ and by convexity $\frac{1}{2}(z + y) \in C$. But since $z + y \in 2L$ we see that $\frac{1}{2}(z + y) \in L$. $\square$

So we have proved that $r(p) > 0$ if $p \equiv 1 \pmod 4$. Now

$$(x^2 + y^2)(a^2 + b^2) = (xa + yb)^2 + (xb - ya)^2$$

and also

$$(x^2 + y^2)(a^2 + b^2) = (xb + ya)^2 + (xa - yb)^2.$$

Then a little bit of bookkeeping will prove (5).

A convenient way of doing this bookkeeping is to use prime factorization in the Gaussian integers, which we will do in the next subsection. To illustrate the power of Minkowski's method, let us use it to prove another famous 18th century theorem:

Theorem 7.2 Lagrange *Every positive integer is the sum of four integer squares.*

Proof. First observe the identity

$$(a^2+b^2+c^2+d^2)(x^2+y^2+z^2+w^2) = (ax-by-cz-dw)^2+(ay+bx+cw-dz)^2$$

$$+(az - bw + cx + dy)^2 + (aw + bz - cy + dx)^2$$

which shows that if m and n are integers each of which can be written as the sum of four integer squares, so can their product. So it is enough to prove the theorem for primes. Since

$$2 = 1^2 + 1^2 + 0^2 + 0^2$$

it is enough to prove the theorem for odd primes. So let p be an odd prime. The set of values of $u^2 \pmod p$ consists of $(p-1)/2$ non-zero values. Throwing in 0 gives $(p+1)/2$ distinct values. The set of values (mod p) of $-1 - v^2$ must also consist of $(p+1)/2$ distinct values. So these two sets must have at least one value in common, implying that we can always find u, v such that

$$u^2 + v^2 + 1 \equiv 0 (\text{mod } p).$$

Consider the lattice $L \subset \mathbf{Z}^4$ consisting of all (a, b, c, d) with

$$c \equiv ua + vb, \quad d \equiv ub - va \quad (\text{mod } p),$$

so

$$a^2 + b^2 + c^2 + d^2 = (a^2 + b^2)(1 + u^2 + v^2) \equiv 0 \quad (\text{mod } p)$$

for all $(a, b, c, d) \in L$. Clearly L has index p^2 in $\mathbf{Z}^4$. A sphere of radius r in four dimensions has volume $\pi^2 r^4/4$. Now $2\pi^2 > 16$, so we can find an r with $r^2 < 2p$ such that the volume of the ball of radius r is $> 16p^2$. By Minkowski, there is a non-zero element of L inside this ball, and so $a^2 + b^2 + c^2 + d^2$ is a non-zero multiple of p which is less than $2p$ and hence must equal p. $\square$

7.3 Factorization of Gaussian integers.

We recall that the ring of Gaussian integers, $\mathbf{Z}[i]$ is a Euclidean domain, hence a unique factorization domain whose units consist of the four elements $\pm 1, \pm i$ and whose primes are: $1 + i, 1 - i$ (the prime divisors of 2), the elements $a \pm ib$ where $a^2 + b^2 = p$ and p is a prime of the form $4k + 1$, and the rational primes q of the form $4k + 3$. Accordingly the factorization (4) of a positive rational integer n into primes is of the form

$$n = (1+i)^f (1-i)^f \left(\prod_{p=a^2+b^2=4k+1} (a+ib)^r (a-ib)^r \right) \prod_{q=4k+3} q^z \qquad (35)$$

where each of the z's must be even if n is a sum of two squares, i.e. $z = 2s$. Any decomposition $n = u^2 + v^2$ of n into a sum of squares can be written as a factorization

$$n = (u + iv)(u - iv)$$

in $\mathbf{Z}[i]$ and unique factorization tells us that

$$\begin{aligned} u+iv &= i^t(1+i)^{f_1}(1-i)^{f_2}\left(\prod_{p=a^2+b^2=4k+1}(a+ib)^{r_1}(a-ib)^{r_2}\right)\prod_{q=4k+3}q^{s_1} \\ u-iv &= \overline{u+iv} \quad \text{hence} \\ f_1+f_2 &= f \\ r_1+r_2 &= r \\ s_1 &= s. \end{aligned}$$

Since t can range over $0,1,2,3$ and f_1 can range over $0,1,2,\dots,f$ while r_1 can range over $0,1,\dots,r$ we see that we have $4(f+1)\prod(r+1)$ choices. But not all lead to different factorizations since $(1+i)/(1-i)=i$ and so the choice of f can be compensated by a choice of t. Hence there are $4\prod_p(r+1)$ different ways of writing n as a sum of two squares. Since $\prod_p(r+1)=d(n_1)$, the number of divisors of n_1 in the factorization (4) we conclude (5): that

$$r(n)=4d(n_1)$$

if all prime factors q of n of the form $4k+3$ occur to an even power and $r(n)=0$ otherwise.

8 Stationary phase

In this section we prove a version of the stationary phase formula. We begin by proving the version that we used in above to prove van der Corput's theorem, which relates to oscillatory integrals of smooth functions over manifolds without boundary.

8.1 Proof of the stationary phase formula.

We will prove the stationary phase formula by a series of reductions. Given any finite cover of M by coordinate neighborhoods, we may apply a partition of unity to reduce our integral to a finite sum of similar integrals, each with the function a supported in one of these neighborhoods. Our first step is to prove Morse's lemma which asserts that we can choose coordinates about each non-degenerate critical point of ψ so that the ψ is actually quadratic in terms of these coordinates:

8.1.1 Morse's lemma.

Suppose that f is a smooth function defined in a neighborhood of the origin in a vector space V with $f(0)=0, df_0=0$ and such that

$$Q:=\frac{1}{2}d_0^2 f$$

is a non-degenerate quadratic form. Morse's lemma asserts that *there is a neighborhood U of the origin, and a diffeomorphism ϕ of U onto a neighborhood of the origin with $\phi(0) = 0$ and*

$$f \circ \phi = Q. \tag{36}$$

The proof we present, due to Palais, is based on a technique of Moser. Set

$$f^t := Q + t(f - Q).$$

Thus

$$f^1 = f, \quad f^0 = Q, \quad \text{and } \dot{f}^t := \frac{df^t}{dt} = f - Q$$

while

$$f^t(0) \equiv 0, \qquad df^t(0) \equiv 0, \qquad d_0^2 f \equiv Q$$

identically in t. Instead of our original problem (36), we will seek to solve the harder problem of finding a one parameter family of diffeomorphisms, ϕ^t such that.

$$f^t \circ \phi^t = f^0. \tag{37}$$

Of course, if we solved this harder problem, setting $\phi = \phi^1$ gives us a solution of our original problem.

Suppose for the moment that we have solved this harder problem, and let X^t denote the vector field tangent to ϕ^t so

$$X^t(\phi^t(x)) := \frac{d\phi^t}{dt}(x).$$

Differentiate (37) with respect to t to obtain

$$\left(\dot{f}^t + X^t f^t\right) \circ \phi^t \equiv 0. \tag{38}$$

If we could find a time dependent vector field X^t such that

$$\dot{f}^t + X^t f^t \equiv 0, \quad \text{and } X^t(0) \equiv 0, \tag{39}$$

then the existence theorem for ordinary differential equations says that we can integrate X^t to find a one parameter family of diffeomorphisms all defined in some neighborhood with $\phi^t(0) = 0$ satisfying (38) and hence (37). Since $\dot{f}^t \equiv f^1 - f^0$ our problem is reduced to solving the equation

$$df^t(X^t) \equiv f^0 - f^1, \quad X^t(0) \equiv 0 \tag{40}$$

for X^t.

For any $x \in V$ near 0 and for any $v \in V$ we have

$$df_x^t(v) = \int_0^1 \frac{d}{ds} df_{sx}^t(v) ds$$

since $df_0^t = 0$. In the above, and in the following, we are using the subscript to denote the point at which we are computing the differential, and the v occurring in the parenthesis denotes a vector in V under the identification of the tangent space T_xV with V. So if g is any differentiable function, the chain rule says that

$$\frac{d}{ds}g(sx) = dg_{sx}(x).$$

As we are in a vector space, the notion of second derivative makes sense at all points, not just critical points, and we can write the integrand above as

$$\frac{d}{ds}df_{sx}^t(v) = d^2f_{sx}(x, v).$$

So if we define the symmetric form B_x^t by

$$B_x^t(u, v) = \int_0^1 d^2f_{sx}(u, v)ds$$

we have

$$df_x^t(v) = B_x^t(x, v).$$

From the definition of f^t we see that

$$B_x^t = B_x^0 + t(B_x^1 - B_x^0)$$

where

$$B_x^0 = 2Q$$

does not depend on x and $B_x^1 = B_x^0$ at $x = 0$. Hence B_x^t is non-singular for all x in some neighborhood of 0 for all $0 \leq t \leq 1$. We can rewrite (40) as

$$B_x^t(x, X^t(x)) = f^0(x) - f^1(x). \tag{41}$$

Let $g := f^0 - f^1$ so that $g(0) = 0$ and $dg_0 = 0$ and hence

$$\begin{aligned} g &= \int_0^1 \frac{d}{ds}g(sx)ds \\ &= \int_0^1 dg_{sx}(x)ds \\ &= \int_0^1 \int_0^1 d^2g_{rsx}(sx, x)drds \\ &= C_x(x, x) \end{aligned}$$

where C_x is the symmetric bilinear form defined by

$$C_x(u, v) := \int_0^1 \int_0^1 d^2g_{rsx}(su, v)drds.$$

We now choose X^t to be the unique solution to

$$B_x^t(u, X^t(x)) = C_x(u, x) \quad \forall u \in V$$

which is possible since B_x is non-singular. The X^t so obtained is smooth and vanishes at the origin. This completes the proof of Morse's lemma.

By partition of unity, our proof of the stationary phase formula thus reduces to estimating integrals over Euclidean space of the form

$$\int e^{i\tau\psi(y)} a(y) dy$$

where a is a smooth function of compact support and where either

1. $d\psi \neq 0$ on supp a so that

$$|d\psi|^2 := \left(\frac{\partial\psi}{\partial y^1}\right)^2 + \cdots + \left(\frac{\partial\psi}{\partial y^n}\right)^2 > \epsilon > 0$$

   on supp a, or

2. ψ is a non-degenerate quadratic form, which, by Sylvester's theorem in linear algebra, we may take to be of the form

$$\psi(y) = \frac{1}{2}\left((y^1)^2 + \cdots (y^k)^2 - (y^{k+1})^2 - \cdots - (y^n)^2\right)$$

   (with, of course, the possibility that $k = 0$ in which case all the signs are negative and $k = n$ in which case all the signs are positive). The number $2k - n$ is the signature of the quadratic form ψ and is what we have denoted by $\mathrm{sgn}(d_0^2\psi)$ in the stationary phase formula.

We treat each of these two cases separately.

8.1.2 The case of no critical points.

2. Show that

$$\int e^{i\tau\psi} a dy = O(\tau^{-k})$$

for any k. [Hint: consider the vector field

$$X := \frac{\partial\psi}{\partial y^1}\frac{\partial}{\partial y^1} + \cdots + \frac{\partial\psi}{\partial y^n}\frac{\partial}{\partial y^n}$$

so that

$$X\left(e^{i\tau\psi}\right) = i\tau|d\psi|^2 e^{i\tau\psi}.$$

Use repreated integration by parts.]

This takes care of the case where there are no critical points. We now turn to the case where ψ is a sum of squares. This reduces to an old friend: Gaussian intergral. so we begin with a reminder.

8.1.3 Gaussian integrals.

Everything stems from the basic computation

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}dx = 1. \tag{42}$$

This is proved by taking the square of the left hand side and then passing to polar coordinates:

$$\begin{aligned}\left[\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}dx\right]^2 &= \frac{1}{2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-(x^2+y^2)/2}dxdy \\ &= \frac{1}{2\pi}\int_0^{2\pi}\int_0^{\infty} e^{-r^2/2}rdrd\theta \\ &= \int_0^{\infty} e^{-r^2/2}rdr \\ &= 1.\end{aligned}$$

Now

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-x^2/2}e^{-\eta x}dx$$

converges for all complex values of η, uniformly in any compact region. Hence it defines an analytic function which may be evaluated by taking η to be real and then using analytic continuation. For real η we complete the square and make a change of variables:

$$\begin{aligned}\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-\frac{x^2}{2}-x\eta)dx &= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp\frac{1}{2}(-(x+\eta)^2+\eta^2)dx \\ &= \exp(\eta^2/2)\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-(x^2+\eta^2)/2)dx \\ &= \exp(\eta^2/2).\end{aligned}$$

As we mentioned, this equation is true for any complex value of η. In particular, setting $\eta = -i\xi$ we get

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-x^2/2+i\xi x)dx = \exp(-\xi^2/2). \tag{43}$$

In short, the Fourier transform of the Gaussian function $x \mapsto \exp(-x^2/2)$ is $\xi \mapsto \sqrt{2\pi}e^{-\xi^2/2}$. If f is any smooth function vanishing rapidly at infinity, and $\hat{f}$ denotes its Fourier transform, then the Fourier transform of $x \mapsto f(cx)$ is $\xi \mapsto \frac{1}{c}\hat{f}(\xi/c)$, a fact that we have already used several times. In particular, if we take $\lambda > 0,\ c = \lambda^{\frac{1}{2}}$ we get

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\exp(-\lambda x^2/2+i\xi x)dx = \left(\frac{1}{\lambda}\right)^{\frac{1}{2}}\exp(-\xi^2/2\lambda). \tag{44}$$

We proved this formula for λ real and positive. But the integral on the left makes sense for all λ with Re $\lambda > 0$, and hence this formula remains true in the entire open right hand plane Re $\lambda > 0$, provided we interpret the square root occurring on the right as arising by analytic continuation from the positive real axis. But we can say more. The integral on the left converges uniformly (but not absolutely) for λ in any region of the form

$$\text{Re } \lambda \geq 0, \quad |\lambda| > \delta > 0.$$

To see this, observe that for any $S > R > 0$ we have

$$e^{-\lambda x^2/2} = -\frac{1}{\lambda x}\frac{d}{dx}\exp(-\lambda x^2/2) \quad \text{for } R \leq x \leq S$$

so we can apply integration by parts to get

$$\int_R^S e^{-\lambda x^2/2} e^{i\xi x} dx =$$

$$\frac{1}{\lambda}\left(\frac{1}{R}e^{-\lambda R^2/2 + i\xi R} - \frac{1}{S}e^{-S^2/2+i\xi S} + \int_R^S e^{-\lambda x^2/2}\frac{d}{dx}\left(\frac{e^{i\xi x}}{x}\right)dx\right)$$

and integrate by parts once more to bound the integral on the right. We conclude that

$$\left|\int_R^S e^{-\lambda x^2/2} e^{i\xi x} dx\right| = O(\frac{1}{|\lambda R|}).$$

This same argument shows that

$$\int e^{-\lambda x^2/2} h(x) dx$$

is convergent for any h with two bounded continuous derivatives. Indeed,

$$\begin{aligned}
\int_R^S e^{-\lambda x^2/2} h(x) dx &= -\frac{1}{\lambda}\int_R^S \frac{h(x)}{x}\frac{d}{dx}e^{-\lambda x^2/2} dx \\
&= -\lambda^{-1} e^{-\lambda x^2/2}(h(x)/x)\Big|_R^S \\
&\quad + \frac{1}{\lambda}\int_R^S e^{-\lambda x^2/2}\frac{d}{dx}(h(x)/x) dx \\
&= -\lambda^{-2} e^{-\lambda x^2/2}[\lambda(h(x)/x) - (1/x)\frac{d}{dx}(h(x)/x)]\Big|_R^S \\
&\quad + \lambda^{-2}\int_R^S e^{-\lambda x^2/2}[(1/x)(h(x)/x)']' dx.
\end{aligned}$$

This last integral is absolutely convergent, and the boundary terms tend to zero as $R \to \infty$. This argument shows that if M is a bound for h and its first

two derivatives, the above expressions can all be estimated purely in terms of M. Thus if h depends on some auxiliary parameters, and is uniformly bounded together with its first two derivatives with respect to these parameters, then the integral $\int_{-\infty}^{\infty} h(x)\exp(-\lambda x^2/2)dx$ converges uniformly with respect to these parameters.

Let us push this argument one step further. Suppose that h has derivatives of all order which are bounded on the entire real axis, and suppose further that $h \equiv 0$ in some neighborhood, $|x| < \epsilon$, of the origin. If we do the integration by parts

$$\int_R^S e^{-\lambda x^2/2}h(x)dx = -\lambda^{-1}e^{-\lambda x^2/2}(h(x)/x)\Big|_R^S + \frac{1}{\lambda}\int_R^S e^{-\lambda x^2/2}\frac{d}{dx}(\frac{h(x)}{x})dx,$$

choose $R < \epsilon$ and let $S \to \infty$. We conclude that

$$\int_\infty^\infty e^{-\lambda^{-1}x^2/2}h(x)dx = \frac{1}{\lambda}\int_\infty^\infty e^{-\lambda x^2/2}\frac{d}{dx}(h(x)/x)dx.$$

The right hand side is a function of the same sort as h. We conclude that

$$\int_{\mathbf{R}} e^{-\lambda x^2/2}h(x)dx = O(\lambda^{-N})$$

for all N if h vanishes in some neighborhood of the origin has derivatives of all order which are each bounded on the entire line. We have already proved this result in the case that h has compact support. We now see that compact support can be replaced by the weaker condition of bounded derivatives of all order.

Getting back to the case $h \equiv 1$, if we take $\lambda = \mp ir, \quad r > 0$ and set $\xi = 0$ in (44) then analytic continuation from the positive real axis gives $\lambda^{\frac{1}{2}} = e^{\mp\pi i/4}$ and we obtain

$$\int_{-\infty}^{\infty} e^{\pm irx^2/2}dx = \left(\frac{2\pi}{r}\right)^{\frac{1}{2}} e^{\pm\pi i/4}. \tag{45}$$

Doing the same computation in n- dimensions gives

$$\int e^{i\tau Q/2}dy = \left(\frac{2\pi}{\tau}\right)^{\frac{n}{2}} e^{i\,\mathrm{sgn}\,Q\pi/4} \tag{46}$$

if

$$Q(y) = \sum \pm (y^i)^2.$$

8.1.4 Completion of the proof.

Now suppose that a is a smooth function of compact support and write

$$a(y) = a(0) + \sum \pm y^i b_i(y)$$

where the b_i are smooth functions which are bounded together with all their derivatives. (Here $b_i(y) = \pm\int_0^1 \frac{\partial a}{\partial y^i}(ty)dy$.)

Multiply both sides of this equation by a smooth function ρ which has compact support and is identically one on the support of a. Then

$$\int e^{i\tau Q(y)^2/2} a(y)dy = \int e^{i\tau Q(y)^2/2} a(0)dy$$

$$+ \int e^{i\tau Q(y)^2/2}(\rho(y)-1)a(0)dy + \int e^{i\tau Q(y)^2/2} \sum \pm y^i \rho(y) b_i(y)dy.$$

3. Complete the proof of the stationary phase formula by evaluating or estimating the three terms on the right.

9 The Fourier inversion formula.

Consider the function $p = p(x,\xi)$ on $\mathbf{R}^n \oplus \mathbf{R}^n$ given by

$$p(x,\xi) = x \cdot (\xi - \eta)$$

where $\eta \in \mathbf{R}^n$. This function has only one critical point, at

$$x = 0, \xi = \eta$$

where its signature is zero. Although we proved the stationary phase formula for functions of compact support, the proof works equally well for functions which vanish rapidly at infinity with all their derivatives. We conclude that for any such function $a = a(x,\xi)$ we have

$$\int\int e^{i\tau x\cdot(\xi-\eta)} a(x,\xi)dxd\xi = \left(\frac{2\pi}{\tau}\right)^n a(0,\eta) + O(\tau^{-(n+1)}).$$

Let us choose $a(x,\xi) = f(x)g(\xi)$ where f and g are smooth functions vanishing rapidly with their derivatives at infinity. We get

$$\left(\frac{1}{\tau^n}\right) f(0)g(\eta) = \frac{1}{(2\pi)^n}\int\int e^{i\tau x\cdot(\xi-\eta)} f(x)g(\xi)dxd\xi + O(\tau^{-(n+1)}).$$

4. Use this formula to prove the Fourier inversion formula.[Hint choose f with $f(0) = 1$ and make the appropriate change of variables.]

10 Group velocity.

In this section we describe one of the most important applications of stationary phase to physics. Let h be a small number (eventually we will take h to be Planck's constant, but for the moment we want to think of h as a parameter which approaches zero, so that $\tau := (1/h) \to \infty$). We want to consider a family of "traveling waves"

$$e^{-(i/h)(E(p)t - p\cdot x)}.$$

For simplicity in exposition we will take p and x to be scalars, but the discussion works as well for x a vector in three (or any) dimensional space and p a vector in the dual space. For each such wave, and for each fixed time t, the wave number of the space variation is h/p. Since we allow E to depend on p, each of these waves will be traveling with a possibly different velocity. Suppose we superimpose a family of such waves, i.e. consider an integral of the form

$$\int a(p)e^{-(i/h)(E(p)t-p\cdot x)}dp. \tag{47}$$

Furthermore, let us assume that the function $a(p)$ has its support in some neighborhood of a fixed value, p_0. Stationary phase says that the only non-negligible contributions to the above integral will come from values of p for which the derivative of the exponent with respect to p vanishes, i.e. for which

$$E'(p)t - x = 0.$$

Since $a(p)$ vanishes unless p is close to p_0, this equation is really a constraint on x and t. It says that the integral is essentially zero except for those values of x and t such that

$$x = E'(p_0)t \tag{48}$$

holds approximately. In other words, the integral looks like a little blip called a *wavepacket* when thought of as a function of x, and this blip moves with velocity $E'(p_0)$ called the *group velocity*.

Let us examine what kind of function E can be of p if we demand invariance under (the two dimensional version of) all Lorentz transformations, which are all linear transformations preserving the quadratic form $c^2t^2 - x^2$. Since (E, p) lies in the dual space to (t, x), the dual Lorentz transformation sends $(E, p) \mapsto (E', p')$ where

$$E^2 - c^2p^2 = (E')^2 - c^2(p')^2$$

and given any (E, p) and (E', p') satisfying this condition, we can find a Lorentz transformation which sends one into the other. Thus the only invariant relation between E and p is of the form

$$E^2 - (pc)^2 = \text{constant}.$$

Let us call this constant m^2c^4 so that $E^2 - (pc)^2 = m^2c^4$ or

$$E(p) = \left((pc)^2 + m^2c^4\right)^{1/2}.$$

Then

$$E'(p) = \frac{pc^2}{E(p)} = \frac{p}{M}$$

where M is defined by

$$E(p) = Mc^2 \quad \text{or} \quad M = \left(m^2 + \left(\frac{p}{c}\right)^2\right)^{1/2}.$$

Notice that if p/c is small in comparison with m then $M \doteq m$. If we think of M as a *mass*, then the relationship between the group velocity $E'(p)$ and p is precisely the relationship between velocity and momentum in classical mechanics. In this way have associated a wave number $k = p/h$ to the momentum p and if we think of E as energy we have associated the frequency $\nu = E/h$ to energy. We have established the three famous formulas

$$\begin{array}{lcccll} E & = & c^2\left(m^2 + \left(\frac{p}{c}\right)^2\right)^{1/2} & \doteq & mc^2 & \text{Einstein's mass energy formula} \\ \lambda & = & \frac{1}{k} & = & \frac{h}{p} & \text{de Broglie's formula} \\ E & = & h\nu & & & \text{Einstein's energy frequency formula.} \end{array}$$

In these formulas we have been thinking of h as a small parameter tending to zero. The great discovery of quantum mechanics is that h should not tend to zero but is a fundamental constant of nature known as *Planck's constant*. In the energy frequency formula it occurs as a conversion factor from inverse time to energy, and hence has units energy $\times$ time. It is given by

$$h = 6.626 \times 10^{-34}\text{J s.}$$

11 Fresnel's version of Huyghen's principle.

Recall that the function

$$\frac{e^{ik(r-t)}}{4\pi r}$$

represents an outgoing spherical wave of frequency k and hence that the function

$$v = \frac{e^{ikr}}{4\pi r}$$

is the fundamental solution of the reduced wave euqation, i.e.

$$\Delta v + k^2 v = \delta_P$$

if r denotes the distance from the point P.

5. From Stokes' theorem (really from Green's fromula)

$$\int\int\int_D (u\Delta v - v\Delta u)dx = \int\int_{\partial D} (u * dv - v * du)$$

deduce Helmholtz formula

$$\frac{1}{4\pi}\int\int_{\partial D}\left[\frac{e^{ikr}}{r} * du - u * d\left(\frac{e^{ikr}}{r}\right)\right] = \begin{cases} u(P) & P \in D \\ 0 & P \notin \overline{D} \end{cases} \tag{49}$$

Apply this to the domain exterior to some surface W and interior to some large sphere of radius R.

6. Show that if u satsifies the "Sommerfeld radiation conditions"

$$\int\int_{S_R} |u| dS = o(1), \qquad \int\int_{S_R} \left| \frac{\partial u}{\partial r} - iku \right| dS = o(R^{-1}) \tag{50}$$

where the integral is over the sphere of radius R, then the value of any solution to the reduced wave equation outside W is given by

$$u(P) = \frac{1}{4\pi} \int\int_W \left[\frac{e^{ikr}}{r} * du - u * d\left(\frac{e^{ikr}}{r} \right) \right]. \tag{51}$$

Assume that near W we have $u = ae^{ik\phi}$ where a and ϕ are smooth and $\|\text{grad } \phi\| = 1$. This would be the case for example if u represented radiation from a single point Q lying interior to W. Assume also that P is sufficiently far from W so that $1/r^2$, and a and da are neglible in comparison with k. Then the top order term in (51) relative to powers of k is

$$\frac{ik}{4\pi} \int\int_W \frac{a}{r} e^{ik(\phi + r)} (*d\phi - *dr).$$

7. Show that the points $y \in W$ of stationary phase for this integral are either where $*d\phi(y) = - * dr(y)$ or where $*d\phi(y) = *dr(y)$ and that at points of the second the top order term in the stationary pahse formule vanishes. (Assume all critical points are non-degenerate.)

Up to lower order terms we may, in the stationary phase formula replace the above integral by

$$\frac{ik}{2\pi} \int\int_W \frac{a}{r} e^{ik(\phi + r} * dr.$$

We can think of this as "secondary radiation" emitted from the surface W. It

- has an amplitiude equal to $1/\lambda$ times the amplitude of the "primary wave" arriving at the surface where $\lambda := 2\pi/k$ is the wave length, and
- its phase is one quarter of a period ahead of the arriving primary wave. (This is one way of thinking about the factor i.)

Fresnel made these two assumptions directly in his formulation of Huyghen's principle of "secondary radiation". This led many people to reject his physical theory as being ad hoc. The above justification for Fresenl's hyopothses was suggested by Helmholtz.