LECTURE 24-25: QUANTUM ERGODICITY

1. Classical dynamics on cotangent bundle T^*M

¶Hamiltonian flow on cotangent bundle.

Recall from Lecture 2 that associated to any Hamiltonian function $H(x, \xi) \in C^\infty(\mathbb{R}^{2n})$, there is a Hamiltonian vector field

$$\Xi_H = \sum_k \left(\frac{\partial H}{\partial \xi_k} \frac{\partial}{\partial x_k} - \frac{\partial H}{\partial x_k} \frac{\partial}{\partial \xi_k} \right)$$

which "dominates" the classical behavior of the system. More precisely, if we denote by $\gamma(t) = \gamma_{x_0,\xi_0}(t)$ the integral curve of Ξ_H starting at the point $\gamma_{x_0,\xi_0}(0) = (x_0, \xi_0)$, then the time-evolution of the system is given by the Hamiltonian flow

$$\rho_t = e^{t\Xi_H} : \mathbb{R}^{2n} \to \mathbb{R}^{2n}, \quad (x, \xi) \mapsto \rho_t(x, \xi) := \gamma_{x,\xi}(t).$$

In particular, we have the conservation of energy

$$\rho_t^* H = H$$

(which implies that each energy surface $H^{-1}(E)$ is preserved under the flow ρ_t), and the evolution equation of classical observable

$$\frac{d}{dt} \rho_t^* a = \{H, a\},$$

where $\{\cdot, \cdot\}$ is the Poisson bracket so that for any $f, g \in C^\infty(\mathbb{R}^{2n})$,

$$\{f, g\} = \sum_{k=1}^n \left(\frac{\partial f}{\partial \xi_k} \frac{\partial g}{\partial x_k} - \frac{\partial f}{\partial x_k} \frac{\partial g}{\partial \xi_k} \right) \in C^\infty(\mathbb{R}^{2n}).$$

Now suppose M be a smooth manifold, and T^*M its cotangent bundle. For any smooth function $p \in C^\infty(T^*M)$, we can also define its *Hamiltonian vector field* [1] via

$$\Xi_p = \sum_k \left(\frac{\partial p}{\partial \xi_k} \frac{\partial}{\partial x_k} - \frac{\partial p}{\partial x_k} \frac{\partial}{\partial \xi_k} \right). \tag{1}$$

Of course one has to check that Ξ_p is well-defined, namely it is independent of the choice of coordinate charts. Recall from Lecture 18 that under the coordinate change $y = y(x)$ on the base manifold, the cotangent variables are related by $\eta = (\frac{\partial x}{\partial y})^T \xi$. So we have

$$(\frac{\partial}{\partial y}) = (\frac{\partial x}{\partial y})^T (\frac{\partial}{\partial x})$$

[1] We will give a more intrinsic definition of Ξ_p via the symplectic structure later.

and
$$(\frac{\partial}{\partial \xi}) = (\frac{\partial \eta}{\partial \xi})^T(\frac{\partial}{\partial \eta}) = (\frac{\partial x}{\partial y})(\frac{\partial}{\partial \eta}).$$
It follows
$$(\frac{\partial p}{\partial x})^T(\frac{\partial}{\partial \xi}) = (\frac{\partial p}{\partial y})^T(\frac{\partial x}{\partial y})^{-1}(\frac{\partial x}{\partial y})(\frac{\partial}{\partial \eta}) = (\frac{\partial p}{\partial y})^T(\frac{\partial}{\partial \eta})$$
and similarly
$$(\frac{\partial p}{\partial \xi})^T(\frac{\partial}{\partial x}) = (\frac{\partial p}{\partial \eta})^T(\frac{\partial}{\partial y}).$$
So the vector field Ξ_p *is* well-defined. By repeating the theory for $\mathbb{R}^n$, we can define the Hamiltonian flow associated to p on T^*M via
$$\rho_t = e^{t\Xi_p} : T^*M \to T^*M, \quad (x, \xi) \mapsto \rho_t(x, \xi) := \gamma_{x,\xi}(t),$$
where $\gamma_{x,\xi}(t)$ is the integral curve of Ξ_p starting at the point $\gamma_{x,\xi}(0) = (x, \xi)$. Again we have the conservation of energy
$$\rho_t^* p = p$$
(which implies that each energy surface $p^{-1}(E)$ is preserved under the flow ρ_t), and the evolution equation of classical observable
$$\frac{d}{dt}\rho_t^* a = \{p, a\},$$
where $\{\cdot, \cdot\}$ is the Poisson bracket on T^*M so that for any $f, g \in C^\infty(T^*M)$,
$$\{f, g\} = \sum_{k=1}^n \left(\frac{\partial f}{\partial \xi_k}\frac{\partial g}{\partial x_k} - \frac{\partial f}{\partial x_k}\frac{\partial g}{\partial \xi_k} \right) = X_f(g) \in C^\infty(T^*M).$$
Remark. Let (M, g) be a smooth Riemannian manifold, and let
$$p(x, \xi) = \frac{1}{2}\|\xi\|_g^2 = \frac{1}{2}\sum g^{ij}\xi_i\xi_j \tag{2}$$
be half of the Riemannian norm square on the cotangent bundle. Then the integral curves of Ξ_p (i.e. the trajectories of the Hamiltonian flow of p), when projected to M, are geodesics of the Riemannian manifold (M, g). Conversely, every parametrized geodesic arises in this way. (For a proof, c.f. my Riemannian geometry notes.) Note that as a consequence of the conservation law of energy, the cosphere bundle
$$S^*M := \{(x, \xi) \in T^*M \mid \|\xi\|_g = 1\}$$
is invariant under the flow ρ_t. As a consequence, we get an induced flow $\rho_t : S^*M \to S^*M$. This flow is usually called the *geodesic flow* on S^*M.

¶Dynamical system on $p^{-1}(c)$ generated by the flow ρ_t.

Similar computations also show that the n-form $dx_1 \wedge d\xi_1 \wedge \cdots \wedge dx_n \wedge d\xi_n$ is a well-defined volume form on T^*M, which is usually known as the *Liouville volume form* or the *symplectic volume form*. The induced measure $dxd\xi$ on T^*M is called the *Liouville measure*.

Now suppose $a < b$ and assume that on the set $a \le p(x,\xi) \le b$, $|dp| \ge c_0 > 0$. So in particular for each $c \in [a,b]$, the level set

$$\Sigma_c := p^{-1}(c)$$

is a smooth $(2n-1)$-dimensional hyper-surface in T^*M. Moreover, for each $c \in [a,b]$, there is an *induced Liouville measure* on Σ_c defined via the formula

$$\int_{p^{-1}([a,b])} f dxd\xi = \int_a^b \int_{\Sigma_c} f d\mu_c dc.$$

In other words, $d\mu_c$ is the measure associated to the induced volume form on the orientable hypersurface Σ_c. For example, if $p(x,\xi) = |\xi|$, then each Σ_c is a cosphere bundle of different radius in T^*M, and the induced Liouville measure on $\Sigma_1 = S^*M$ is nothing but $dxd_{S^{n-1}}(\xi)$.

Recall that the Hamiltonian flow of the Hamiltonian function $p(x,\xi)$ preserves any level set Σ_c. We will prove that the Liouville volume form and thus the Liouville measure is invariant under the Hamiltonian flow ρ_t. As a consequence, the induced Liouville measure μ_c on Σ_c is invariant under the flow ρ_t generated by p.

¶Classical ergodicity.

A very important class of dynamical systems, known as *measure preserving flow*, is a triple (X,μ,ρ_t), where (X,μ) is a measure space with $\mu(X) < +\infty$, and $\rho_t : X \to X$ is a measure preserving flow on X, namely

- For any $t \in \mathbb{R}$, $\rho_t : (X,\mu) \to (X,\mu)$ is measure-preserving.
- ρ_t is a flow: $\rho_{t+s} = \rho_t \circ \rho_s$.

Among all classes of dynamical systems, two extremal cases are widely studied: the integrable case and the ergodic case. Roughly speaking, an integrable dynamical system is a system with maximal conserved quantities and thus is very "regular", while ergodic system is very "chaotic". Here is a precise definition of ergodicity:

Definition 1.1. We say a measure-preserving flow $\rho_t : (X,\mu) \to (X,\mu)$ is *ergodic* if any ρ_t-invariant measurable subset of X either has measure 0 or has full measure.

Example. For any compact Riemannian manifold with negative sectional curvature, the geodesic flow is ergodic. (This was first proved by Hopf for $n = 2$, and by Anosov and Sinai for higher dimensions.

Example. The geodesic flow on S^n is NOT ergodic. (It is integrable.)

The following theorem is a classical result in the theory of dynamical systems, which claims that for an ergodic system, the "time-average" of any L^1-function equals to its "space-average":

Theorem 1.2 (Birkhorff). *Suppose ρ_t is an ergodic flow on (X, μ), then for any $f \in L^1(X, \mu)$,*

$$\lim_{T\to\infty} \frac{1}{T}\int_0^T f(\rho_t(x))dt \to \frac{1}{\mu(X)}\int_X f(y)d\mu$$

for a.e. $x \in X$.

In other words, the flows of ergodic systems are equidistributed in the phase space, which is in contrast to the fact that classical completely integrable systems generally have periodic orbits in phase space.

Birkhoff ergodicity theorem is a very strong theorem. What we will need is the following weaker ergodicity theorem:

Theorem 1.3 (L^2-mean ergodic theorem). *Suppose ρ_t is an ergodic flow on (X, μ), then for any $f \in L^2(X, \mu)$,*

$$\lim_{T\to\infty} \int_X \left(\frac{1}{T}\int_0^T f(\rho_t(x))dt - \frac{1}{m(X)}\int_X f(y)d\mu_y\right)^2 d\mu_x = 0.$$

(For a proof, c.f. Zworski, page 367-368.)

Notations for the time-average

$$\langle f\rangle_T(x) := \frac{1}{T}\int_0^T f(\rho_t(x))dt.$$

and the space-average

$$\langle f\rangle_X := \frac{1}{\mu(X)}\int_X f(y)d\mu_y.$$

Then we can rewrite the mean ergodic theorem as

$$\lim_{T\to\infty}\int_X (\langle f\rangle_T(x) - \langle f\rangle_X)^2 d\mu_x = 0$$

while the Birkhoff ergodicity theorem claims that

$$\langle f\rangle_T(x) \to \langle f\rangle_X$$

as $T \to \infty$ for a.e. $x \in X$.

2. Quantum ergodicity

In this section we will always assume

(1) (M, g) is a compact Riemannian manifold,
(2) $p \in S^m(T^*M)$ is an (almost) elliptic classical symbol, where $m > 0$.
(3) $a < b$, and on the set $a \leq p(x, \xi) \leq b$, $|dp| \geq c_0 > 0$.

Note that by the condition (3), each $c \in [a, b]$ is a regular value and thus each energy level set $\Sigma_c = p^{-1}(c)$ is a smooth compact manifold on which we will always endow the induced Liouville measure μ_c.

¶Quantum ergodicity.

In the classical side, we have the Hamiltonian flow on each energy level set $p^{-1}(c)$ (which is compact since p is proper, where we assume that c is a regular value of p). In the quantum part, we have the eigenvalues/eigenfunctions of P,

$$P\varphi_j = \lambda_j \varphi_j,$$

where $\varphi_j \in C^\infty(M)$ and $\{\varphi_j\}$ form an L^2-orthonormal basis of $L^2(M)$, and

$$\text{Spec}(P): \qquad \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n \le \cdots \to \infty.$$

and people would like to understand the relation between the dynamical behavior of the classical Hamiltonian flow and the quantum eigenvalue/eigenfunction data.

The main theorem in this section is the quantum ergodicity theorem which describes the behavior of eigenfunctions when the corresponding classical system is ergodic.

Theorem 2.1 (Schnirelman-Zelditch-Colin de Verdiere Quantum Ergodicity Theorem, Version 1)**.** *Suppose the Hamiltonian flow of p is ergodic on (Σ_c, μ_c) for each $c \in [a, b]$. . Then there exists a family of subsets $\Lambda(\hbar) \subset \text{Spec}(P) \cap [a, b]$ which has density 1 in the sense*

$$\lim_{\hbar \to 0} \frac{\#\Lambda(\hbar)}{\#(\text{Spec}(P) \cap [a, b])} = 1, \tag{3}$$

such that for any semiclassical pseudodifferential operator $A \in \Psi^0(M)$ whose symbol $\sigma(A)$ satisfy the condition that the quantity

$$\frac{1}{\text{Vol}(\Sigma_c)} \int_{\Sigma_c} \sigma(A) d\mu_c \tag{4}$$

is independent of $c \in [a, b]$, we have, as $\hbar \to 0$,

$$\langle A\varphi_j, \varphi_j \rangle \to \frac{1}{\text{Vol}(p^{-1}([a, b]))} \int_{a \le p \le b} \sigma(A) dx d\xi \tag{5}$$

for $\lambda_j \in \Lambda(\hbar)$.

Before we go to the proof, let's state an important corollary. Let's take $P = \hbar^2 \Delta$, and take $A = M_f$ to be the "multiplication by f" map, where $f \in C^\infty(M)$. Then

- eigenvalues are $\lambda_{j,\hbar} = \hbar^2 \lambda_j$, where λ_j are the standard Laplacian eigenvalues of (M, g)
- the eigenfunctions φ_j are the standard Laplacian eigenfunctions which are independent of $\hbar$
- each Σ_c is a cosphere bundle

- $A \in \Psi^0(M)$ and satisfies the condition (4):

$$\frac{1}{\mathrm{Vol}(\Sigma_c)} \int_{\Sigma_c} \sigma(A) d\mu_c = \frac{1}{\mathrm{Vol}(M)} \int_M f(x)dx.$$

In this case, by taking $[a, b] = [0, b]$ we get

Corollary 2.2 (Quantum ergodicity for Laplacian eigenfunctions)**.** *Suppose* (M, g) *is a compact Riemannian manifold with ergodic geodesic flow. Then there exists a sequence* $j_k \to \infty$ *which has density 1 in the sense*

$$\lim_{N\to\infty} \frac{\#\{j_k \le N\}}{N} = 1,$$

such that for each $f \in C^\infty(M)$*,*

$$\int_M |\varphi_{j_k}|^2 f dx \to \frac{1}{\mathrm{Vol}(M)} \int_M f(x)dx. \tag{6}$$

In particular, if we take $U \subset M$ to be an open subset, and f a sequence of smooth functions that approximates the characteristic function χ_U of U, then we will get

$$\int_U |\varphi_{j_k}|^2 dx \to \frac{\mathrm{Vol}(U)}{\mathrm{Vol}(M)}.$$

In other words, for *most* eigenfunctions, the "mass" will tends to be *equidistributed* on M.

In the statement of quantum ergodicity theorems above, one can't rule out the possibility of the existence of *quantum scar*, namely a density 0 subsequence along which the integral (6) concentrate on a closed geodesic (or the corresponding integral (5) will concentrate on a periodic orbit of the geodesic flow)[2]. For the case of negatively curved manifolds (in which case it is known that the geodesic flow satisfy a stronger ergodicity condition), it was conjectured by Z. Rudnick and P. Sarnak in 1994 that there is no quantum scar. This is known as *Quantum Unique Ergodicity Conjecture*, and is currently one of the most famous open problems in spectral geometry:

Conjecture 2.3 (QUE conjecture)**.** *Suppose* (M, g) *is a compact Riemannian manifold with negative sectional curvature. Then as* $k \to \infty$*,*

$$\int_M |\varphi_k|^2 f dx \to \frac{1}{\mathrm{Vol}(M)} \int_M f(x)dx.$$

[2]The limit can't be too bad. See the theorem at the end of this lecture

¶Proof of Quantum ergodicity theorem (version 2).

To prove the quantum ergodicity theorem, namely, Theorem 2.1, we first prove the following

Theorem 2.4 (Quantum ergodicity theorem, Version 2)**.** *Suppose the Hamiltonian flow of $p(x,\xi)$ is ergodic in $p^{-1}([a,b])$, and assume $A \in \Psi^0(M)$ is a pseudodifferential operator whose symbol $\sigma(A)$ satisfy the condition that the quantity*

$$\alpha := \langle \sigma(A) \rangle_{\Sigma_c} = \frac{1}{\mathrm{Vol}(\Sigma_c)} \int_{\Sigma_c} \sigma(A) d\mu_c$$

is independent of $c \in [a,b]$. Then as $\hbar \to 0$,

$$(2\pi\hbar)^n \sum_{a \le \lambda_j \le b} \left| \langle A\varphi_j, \varphi_j \rangle - \frac{1}{\mathrm{Vol}(p^{-1}([a,b]))} \iint_{a \le p \le b} \sigma(A) dx d\xi \right|^2 \to 0. \tag{7}$$

There are three main ingredients in the proof:

(1) Weak Egorov theorem (PSet 4): Let $U(t) = e^{-itP/\hbar}$. Then for any $a \in S^{-\infty}(T^*M)$, we have
$$\|U(-t)Op(a)U(t) - Op(a \circ \rho_t)\|_{\mathcal{L}(L^2)} = O(\hbar).$$
(2) Generalized Weyl's law (Lecture 23): Suppose $B \in \Psi^0(M)$ and $a < b$. Then as $\hbar \to 0$ we have
$$(2\pi\hbar)^n \sum_{a \le \lambda_j \le b} \langle B\varphi_j, \varphi_j \rangle \to \iint_{a \le p \le b} \sigma(B) dx d\xi.$$
(3) The L^2-mean ergodic theorem (Theorem 1.3 above): Suppose ρ_t is an ergodic flow on (X,μ), then for any $f \in L^2(X,\mu)$,
$$\lim_{T\to\infty} \int_X \left(\langle f \rangle_T - \langle f \rangle_X\right)^2 d\mu_x = 0.$$

Proof of Theorem 2.4. By our assumption on $\sigma(A)$, we have

$$\frac{1}{\mathrm{Vol}(p^{-1}([a,b]))} \int_{p^{-1}([a,b])} \sigma(A) dx d\xi = \langle \sigma(A) \rangle_{\Sigma_c} = \alpha.$$

We choose a cut-off $\chi \in C_0^\infty(\mathbb{R})$ so that $\chi \equiv 1$ on $[a,b]$. Let

$$B = (A - \alpha \mathrm{Id})\chi(P).$$

Then $B \in \Psi^{-\infty}(M)$ (so that we can apply weak Egorov theorem), and in view of the fact $B\varphi_j = A\varphi_j - \alpha\varphi_j$ for $\lambda_j \in [a,b]$, we are reduced to prove

$$(2\pi\hbar)^n \sum_{a \le \lambda_j \le b} |\langle B\varphi_j, \varphi_j \rangle|^2 \to 0. \tag{8}$$

The idea is to consider "the time-average of B",

$$\langle B\rangle_T := \frac{1}{T}\int_0^T U(-t)BU(t)\rangle dt.$$

For any $t \in \mathbb{R}$, we have

$$\langle U(-t)BU(t)\varphi_j, \varphi_j\rangle = \langle B\varphi_j, \varphi_j\rangle.$$

Thus

$$\langle B\varphi_j, \varphi_j\rangle = \langle \langle B\rangle_T\varphi_j, \varphi_j\rangle.$$

It follows from Cauchy-Schwarz inequality that

$$|\langle B\varphi_j, \varphi_j\rangle|^2 \le \|\langle B\rangle_T\varphi_j\|^2 = \big\langle \langle B\rangle_T^*\langle B\rangle_T\varphi_j, \varphi_j\big\rangle.$$

By weak Egorov theorem,

$$\|\langle B\rangle_T - \langle \widetilde{B}\rangle_T\|_{\mathcal{L}(L^2)} = O_T(\hbar),$$

where $O_T(\hbar)$ means $O(\hbar)$ with constants depending on T, and

$$\langle \widetilde{B}\rangle = \frac{1}{T}\int_0^T \widetilde{B}(t)dt,$$

and $\widetilde{B}(t) \in \Psi^0(M)$ with

$$\sigma(\widetilde{B}(t)) = \rho_t^*\sigma(B).$$

It follows

$$\sigma(\langle \widetilde{B}\rangle_T) = \frac{1}{T}\int_0^T \rho_t^*\sigma(B)dt = \langle \sigma(B)\rangle_T.$$

and thus

$$\sigma(\langle \widetilde{B}\rangle_T^*\langle \widetilde{B}\rangle_T) = |\sigma(\langle \widetilde{B}\rangle_T)|^2 = |\langle\sigma(A)\rangle_T - \alpha|^2 = |\langle\sigma(A)\rangle_T - \langle\sigma(A)\rangle_{\Sigma_c}|^2.$$

So according to the generalized Weyl law, for each fixed $T > 0$,

$$\begin{aligned}
\limsup_{\hbar\to 0}(2\pi\hbar)^n \sum_{a<\lambda_j<b} |\langle B\varphi_j, \varphi_j\rangle|^2 &\le \limsup_{\hbar\to 0}(2\pi\hbar)^n \sum_{a<\lambda_j<b} \big\langle \langle B\rangle_T^*\langle B\rangle_T\varphi_j, \varphi_j\big\rangle \\
&\le \limsup_{\hbar\to 0}(2\pi\hbar)^n \sum_{a<\lambda_j<b} \big\langle \langle \widetilde{B}\rangle_T^*\langle \widetilde{B}\rangle_T\varphi_j, \varphi_j\big\rangle + O_T(\hbar) \\
&= \iint_{p^{-1}([a,b])} \sigma(\langle \widetilde{B}\rangle_T^*\langle \widetilde{B}\rangle_T)dxd\xi \\
&= \int_a^b \int_{\Sigma_c} |\langle\sigma(A)\rangle_T - \langle\sigma(A)\rangle_{\Sigma_c}|^2 d\mu_c dc.
\end{aligned}$$

Finally we use the L^2-mean ergodic theorem to conclude (8), which completes the proof. □

Remark. Roughly speaking, in the proof we are trying to prove that the operator $\langle B\rangle_T$ is "small", or equivalently, the difference between the "time-average operator" $\langle A\rangle_T$ and the "space-average operator" $\langle A\rangle_X := \langle\sigma(A)\rangle_X \cdot \mathrm{Id}$ is small.

¶Proof of Quantum ergodicity theorem (version 1).

Now we are ready to prove the Schnirelman-Zelditch-Colin de Verdiere Quantum Ergodicity Theorem. To get Theorem 2.1 from Theorem 2.4, it is helpful to notice the following beautiful fact from mathematical analysis:

Lemma 2.5. *Suppose a_i are non-negative real numbers and*

$$\lim_{N\to\infty} \frac{1}{N}\sum_{k=1}^{N} a_k = 0,$$

then there exists a density one subset $\{a_{k_j}\}$ of $\{a_i\}$ such that

$$\lim_{k\to\infty} a_{j_k} = 0.$$

It is easy to see that for the case $P = \hbar^2\Delta_g$ and $[a,b] = [0,\lambda]$, Theorem 2.1 is a direct consequence from Theorem 2.4 together with Lemma 2.5. Unfortunately in the general semiclassical setting, we can't directly apply Lemma 2.5 since the summands in (7) are different for different $\hbar$. So what we need is a variant of Lemma 2.5 adapted to our semiclassical setting.

Proof of Theorem 2.1.

As in the proof of Theorem 2.4, is enough to construct density 1 subsets $\Lambda(\hbar)$ such that for $\lambda_j = \lambda_{j,\hbar} \in \Lambda(\hbar)$ and for all $B = (A - \alpha \mathrm{Id})\chi(P) \in \Psi^{-\infty}(M)$,

$$\langle B\varphi_j, \varphi_j\rangle \to 0$$

as $\hbar \to 0$.

Step 1. Construct $\Lambda(\hbar)$ for a fixed B.

As we have seen in the proof of Theorem 2.4,

$$\varepsilon(\hbar) := (2\pi\hbar)^n \sum_{a\le\lambda_j\le b} |\langle B\varphi_j, \varphi_j\rangle|^2 \to 0.$$

So if we define

$$\Gamma(\hbar) := \{\lambda_j \in [a,b] \ : \ |\langle B\varphi_j,\varphi_j\rangle|^2 \ge \varepsilon(\hbar)^{1/2}\},$$

then we must have

$$(2\pi\hbar)^n \#\Gamma(\hbar) \le \varepsilon(\hbar)^{1/2}.$$

Let's denote

$$\Lambda(\hbar) := (\mathrm{Spec}(P)\cap[a,b]) \setminus \Gamma(\hbar).$$

Then by construction, for $\lambda_j \in \Lambda(\hbar)$ we have

$$|\langle B\varphi_j,\varphi_j\rangle| < \varepsilon(\hbar)^{1/4} \to 0.$$

Moreover, according to Weyl's law, as $\hbar \to 0$,

$$\frac{\#\Lambda(\hbar)}{\#\mathrm{Spec}(P)\cap[a,b]} = 1 - \frac{\#\Gamma(\hbar)}{\#\mathrm{Spec}(P)\cap[a,b]} = 1 - \frac{(2\pi\hbar)^n\#\Gamma(\hbar)}{\mathrm{Vol}(p^{-1}([a,b])) + o(1)} \to 1.$$

Step 2. Construct $\Lambda(\hbar)$ that works for a sequence B_k simultaneously.

This can be done by using the standard diagonal trick. For each k, we have constructed in Step 1 density 1 subsets $\Lambda_k(\hbar) \subset \mathrm{Spec}(P) \cap [a,b]$ so that the theorem holds for B_k and $\Lambda_k(\hbar)$.

Observation: If $\Lambda(\hbar)$ and $\Lambda'(\hbar)$ are density 1 subsets, so is $\Lambda(\hbar) \cap \Lambda'(\hbar)$.

So if we replace $\Lambda_k(\hbar)$ by $\cap_{1\le j\le k}\Lambda_j(\hbar)$, then we get density 1 subsets $\Lambda_k(\hbar) \subset \mathrm{Spec}(P) \cap [a,b]$ with $\Lambda_{k+1}(\hbar) \subset \Lambda_k(\hbar)$, so that the theorem holds for A_k and $\Lambda_k(\hbar)$.

Now for each $k \in \mathbb{N}$, we take $\hbar(k) > 0$ small enough so that

$$\frac{\#\Lambda_k(\hbar)}{\#\mathrm{Spec}(P) \cap [a,b]} \ge 1 - \frac{1}{k}, \qquad \forall 0 < \hbar < \hbar(k).$$

Moreover, we can choose $\hbar(k)$ so that they decrease to 0: $\hbar(k) \searrow 0$ as $k \to \infty$. Now we define

$$\Lambda_\infty(\hbar) := \Lambda_k(\hbar), \quad \hbar(k+1) \le \hbar < \hbar(k).$$

Then we have

$$\frac{\#\Lambda_\infty(\hbar)}{\#\mathrm{Spec}(P) \cap [a,b]} \ge 1 - \frac{1}{k}$$

for all k and all $0 < \hbar < \hbar(k)$, and thus

$$\frac{\#\Lambda_\infty(\hbar)}{\#\mathrm{Spec}(P) \cap [a,b]} \to 1$$

as $\hbar \to 0$. So $\Lambda_\infty(\hbar)$ is a density 1 subset.

Moreover, for each k, since $\Lambda_\infty(\hbar) \subset \Lambda_k(\hbar)$ for $\hbar < \hbar(k)$, for $\lambda_j = \lambda_{j,\hbar} \in \Lambda_\infty(\hbar)$ we still have $|\langle B_k\varphi_j, \varphi_j\rangle| \to 0$ as $\hbar \to 0$. This completes Step 2.

Step 3. Construct $\Lambda(\hbar)$ that works for all B simultaneously.

The idea is to choose a "dense sequence" B_k. Suppose we can find a sequence B_k so that for any $B \in \Psi^{-\infty}(M)$ with $\int_{\Sigma_c} \sigma(B) d\mu_c = 0$ for all $c \in [a,b]$ and for any $\varepsilon > 0$, there exists k and $\hbar_0$ such that for all $0 < \hbar < \hbar_0$, we have

$$\|B_k - B\|_{\mathcal{L}(L^2)} < \varepsilon,$$

then for $\lambda_j = \lambda_{j,\hbar} \in \Lambda_\infty(\hbar)$ (constructed in Step 2),

$$\big|\limsup_{\hbar\to 0}\langle B\varphi_j, \varphi_j\rangle\big| < \varepsilon \quad \text{ and } \quad \big|\liminf_{\hbar\to 0}\langle B\varphi_j, \varphi_j\rangle\big| < \varepsilon$$

and the conclusion follows.

It remains to construct such a dense sequence B_k. According to Corollary 2.2 in Lecture 15 (see also the remark at the end of Lecture 11),

$$\|B_k - B\|_{\mathcal{L}(L^2)} \le \|\sigma(B_k) - \sigma(B)\|_{L^\infty(T^*M)} + O(\hbar^{1/2}).$$

So it is enough to find a sequence $\{b_k\} \subset S^{-\infty}(T^*M)$ with $\int_{\sigma_c} b_k d\mu_c = 0$, such that for any $b \in S^{-\infty}(T^*M)$ with $\int_{\sigma_c} b d\mu_c = 0$ and any $\varepsilon > 0$, there exists k such that $\|b - b_k\|_{L^\infty(T^*M)} < \varepsilon$. This is possible, since

- The space of continuous functions that vanishes at infinity,

$$\mathcal{C}_0(T^*M) = \{f \in \mathcal{C}(T^*M) \mid \forall \varepsilon > 0, \exists \text{ compact } K \text{ s.t. } |f(x)| < \varepsilon \text{ on } K^c\},$$

  is separable (i.e. has a countable dense subset). (This can be proven via the Stone-Weierstrass theorem for locally compact Hausdorff spaces.)
- The space $\mathcal{C}_0(T^*M)$ is a metric space with respect to the L^∞-metric.
- Any subspace of a separable metric space is separable. (In general a subspace of a separable space may be non-separable).
- $S^{-\infty}(T^*M) \subset \mathcal{C}_0(T^*M)$. (We can't use $S^0(T^*M)$ in the proof since the space of bounded continuous functions is not separable.)

This completes the proof. □

3. Semiclassical defect measure

¶Quantum ergodicity from measure point of view.

Of course in the conclusion of Corollary 2.2, we may replace $f \in C^\infty(M)$ by $f \in C(M)$. In other words, we can restate the quantum ergodicity as

> If the geodesic flow of (M, g) is ergodic, then there exists a density 1 subset $\{j_k\} \subset \mathbb{N}$ such that the sequence of probability measure $|\varphi_{j_k}|^2 dx$ on M converges weakly to the uniform probability measure:
>
> $$|\varphi_{j_k}(x)|^2 dx \rightharpoonup \frac{1}{\mathrm{Vol}(M)} dx.$$

Similarly, one can explain the conclusion of Theorem 2.1 using the language of measure: For *most* eigenfunctions, if we regard the expected value of the quantum observable A to a system in the state $\varphi_{j,\hbar}$ (c.f. Lecture 2),

$$\langle A\varphi_{j,\hbar}, \varphi_{j,\hbar}\rangle,$$

as an integral of the symbol $\sigma(A)$ with respect to a measure on the classical phase space[3], then the measure converges weakly to the uniform Liouville measure

$$\frac{1}{\mathrm{vol}(p^{-1}([a,b]))} dx d\xi$$

on $p^{-1}([a,b])$ (or the uniform Liouville measure $\frac{1}{\mathrm{vol}(\Sigma_c)} d\mu_c$ on Σ_c). This explains the following description of quantum ergodicity in Wikipedia:

> [Wiki: Quantum ergodicity states, roughly, that in the high-energy limit, the probability distributions associated to energy eigenstates of a quantized ergodic Hamiltonian tend to a uniform distribution in the classical phase space.]

[3]We will make this sentence more clear in the next page.

¶Semiclassical defect measure.

Let $\{u_\hbar\}$ be a family of L^2-normalized functions.

Definition 3.1. A *semiclassical defect measure* (also known as *semiclassical measure* in some literature) associated with $\{u_\hbar\}$ is a nonnegative Radon measure μ such that for some subsequence $h_j \to 0$, we have

$$\lim_{j\to\infty} \langle \widehat{a}^W u_{\hbar_j}, u_{\hbar_j}\rangle \to \int_{\mathbb{R}^{2n}} a(x,\xi) d\mu$$

for any symbol $a \in C_0^\infty(\mathbb{R}^{2n})$.

The existence of a defect measure is guaranteed by the standard diagonal trick:

Proposition 3.2. *For any $\{u_\hbar\}$, there exists at least one semiclassical defect measure $d\mu$.*

Proof. We first choose a countable dense subset $\{a_k\}$ in the set $C_0^\infty(\mathbb{R}^{2n})$. By boundedness of $\widehat{a}^W$ (Corollary 2.2 in Lecture 15), we can pick $\hbar_j^{(1)} \to 0$ such that

$$\langle \widehat{a_1}^W u_\hbar, u_\hbar\rangle_{\hbar=\hbar_j^{(1)}} \to \alpha_1.$$

Similarly we can pick a subsequence $\hbar_j^{(2)}$ of $\hbar_j^{(1)}$ such that

$$\langle \widehat{a_2}^W u_\hbar, u_\hbar\rangle_{\hbar=\hbar_j^{(2)}} \to \alpha_2$$

and so on. Now we define $\hbar_j := \hbar_j^{(j)}$. Then

$$\langle \widehat{a_k}^W u_\hbar, u_\hbar\rangle_{\hbar=\hbar_j} \to \alpha_k$$

for any k.

Now consider a map Φ which maps a_k to α_k. Note that by Corollary 2.2 in Lecture 15, we have $|\alpha_k| \le \|a_k\|_{L^\infty}$. So Φ is bounded on the dense subset α_k, it also keep any possible "linear relation" on the set $\{a_k\}$, and thus can be extended to a bounded linear map on $\Phi : C_0^\infty(\mathbb{R}^{2n}) \to \mathbb{R}$ with

$$|\Phi(a)| \le \|a\|_{L^\infty}.$$

By Riesz representation theorem, there exists a unique (complex-valued) Radon measure μ on $\mathbb{R}^{2n}$ such that

$$\Phi(a) = \int a d\mu.$$

It remains to prove that $d\mu$ is nonnegative, namely $\Phi(a) \ge 0$ for $a \ge 0$. For this purpose we need a more precise formula for $\Phi(a)$. In fact, for any $a \in C_0^\infty(\mathbb{R}^{2n})$, we may approximate a by a sequence a_{k_j}. If $\|a_k - a\|_{L^\infty} < \varepsilon$, then

$$\limsup_{j\to\infty} \left|\langle \widehat{a}^W u_\hbar, u_\hbar\rangle - \langle \widehat{a_k}^W u_\hbar, u_\hbar\rangle\right|_{\hbar=\hbar_j} \le \varepsilon$$

which implies

$$\Phi(a) = \lim_{j\to\infty} \langle \widehat{a}^W u_\hbar, u_\hbar\rangle_{\hbar=\hbar_j}.$$

Thus according to the sharp Gårding inequality, for any $a \geq 0$,

$$\Phi(a) = \lim_{\hbar_j\to 0} \langle \widehat{a}^W u_\hbar, u_\hbar\rangle_{\hbar=\hbar_j} \geq 0.$$

So $d\mu$ is a positive measure and the proof is finished. □

Recall from Lecture 21 that the wavefront set $\mathrm{WF}(u_\hbar)$ of $\{u_\hbar\}$ (which describes the "concentration" of $\{u_\hbar\}$ in the phase space) is defined via its complement: $(x_0, \xi_0) \notin \mathrm{WF}(u_\hbar)$ if for any $b \in C_0^\infty(\mathbb{R}^{2n})$ with support sufficiently close to (x_0, ξ), we have $\|\widehat{b}^W u_\hbar\|_{L^2} = O(\hbar^2)$. This implies

$$\lim_{j\to\infty} \langle \widehat{b}^W u_\hbar, u_\hbar\rangle_{\hbar=\hbar_j} = 0$$

for any subsequence $\hbar_j \to 0$. As a result, for any semiclassical measure μ associated with $\{u_\hbar\}$, we must have

$$\int_{\mathbb{R}^{2n}} b(x,\xi)d\mu = 0.$$

Since this holds for all b with $b(x_0, \xi_0) = 1$ with support sufficiently close to (x_0, ξ), we must have $\mu(x_0, \xi_0) = 0$. In other words,

Proposition 3.3. *Any semiclassical defect measure μ associated with $\{u_\hbar\}$ is supported in* $\mathrm{WF}(u_\hbar)$.

Remark. The conception of defect measure as well as the propositions just proved can be easily extended (with small modifications) to T^*M, where M is any compact Riemannian manifold.

¶Semiclassical defect measure associated to eigenfunctions.

Now let P be a semiclassical pseudodifferential operator as in §2, and let $u_\hbar$ be L^2-normalized eigenfunctions of P associated with eigenvalue $\lambda_\hbar$:

$$Pu_\hbar = \lambda_\hbar u_\hbar.$$

We have

Theorem 3.4. *Let $\{u_\hbar\}$ be a sequence of eigenfunctions so that $\lambda_\hbar \to E_0$ as $\hbar \to 0$. Then any semiclassical defect measure μ associated with $\{u_\hbar\}$ is a probability measure supported on $p^{-1}(E_0)$ which is invariant under the flow ρ_t generated by p.*

Proof. According to Theorem 2.5 in Lecture 21, we have[4]

$$\mathrm{supp}\mu \subset \mathrm{WF}(u_\hbar) \subset p^{-1}(E_0).$$

[4]Here it is enough to assume $P_\hbar u_\hbar = \lambda_\hbar u_\hbar + o(1)$.

Since $|\int a d\mu| = |\Phi(a)| \le \|a\|_{L^\infty}$, we have $\mu(\mathbb{R}^{2n}) \le 1$. Conversely, according to Theorem 1.3 in Lecture 21, for any $a \in C_0^\infty$ such that $a \le 1$ and $a \equiv 1$ in a neighborhood of $p^{-1}(E_0)$, we have $\widehat{a}^W u_\hbar = u_\hbar + O(\hbar^\infty)$ and thus

$$\mu(\mathbb{R}^{2n}) \ge \int_{\mathbb{R}^{2n}} a d\mu = \lim_{j\to\infty} \langle \widehat{a}^W u_\hbar, u_\hbar\rangle_{\hbar=\hbar_j} \ge 1.$$

To prove the invariance of μ under the Hamiltonian flow ρ_t associated with p, it is enough to prove

$$\Phi(a) = \Phi(\rho_t^* a) \tag{9}$$

for any $a \in C_0^\infty(\mathbb{R}^{2n})$ and any $t \in \mathbb{R}$. In fact, if we denote $A = \widehat{a}^W$, then

$$[P, A] = \frac{\hbar}{i}\widehat{\{p,a\}}^W + O(\hbar^2).$$

But

$$\langle [P,A]u_\hbar, u_\hbar\rangle = \langle Au_\hbar, Pu_\hbar\rangle - \langle Pu_\hbar, Au_\hbar\rangle = \langle Au_\hbar, \lambda_\hbar u_\hbar\rangle - \langle \lambda_\hbar u_\hbar, Au_\hbar\rangle = 0.$$

It follows $\langle \widehat{\{p,a\}}^W u_\hbar, u_\hbar\rangle = O(\hbar)$ and thus [5]

$$\int_{\mathbb{R}^{2n}} \{p,a\} d\mu = \lim_{j\to\infty} \langle \widehat{\{p,a\}}^W u_\hbar, u_\hbar\rangle = 0.$$

Now the equation (9) follows, since

$$\frac{d}{dt}\Phi(\rho_t^* a) = \frac{d}{dt}\int \rho_t^* a d\mu = \int \frac{d}{dt}\rho_t^* a d\mu = \int \{p,a\} d\mu = 0.$$

□

For the geodesic flow on the cosphere bundle of compact Riemannian manifolds, here are some invariant probability measures which are possible candidates of semiclassical defect measures of the Laplacian eigenfunctions:

- The Liouville measure $d\mu_{Liouvlle}$,
- The dirac delta measure on closed geodesics,
- Combinations of the above...

The QUE conjecture claims that for negatively curved manifolds, the only semiclassical defect measure is the Liouville measure. Some recent works:

- Lindenstrauss 2006, Soundararajan 2010: the conjecture is true for modular surface $SL_2(\mathbb{Z}) \setminus H$.
- Anantharaman 2008, Anantharaman-Nonnenmacher 2007: the KS-entropy of the semiclassical measure has a positive lower bound (and in particular rules out the delta measures concentrated on closed geodesics).
- Dyatlov-Jin 2018, Dyatlov-Jin-Nonnenmacher 2019: For negatively curved surfaces, every semiclassical measure has full support.

[5]Here it is enough to assume $P_\hbar u_\hbar = \lambda_\hbar u_\hbar + o(\hbar)$.