Topics in concentration of measure: Lecture I

Sourav Chatterjee

Courant Institute, NYU

St. Petersburg Summer School, June 2012

Lecture I: Localization in random structures

Hamiltonian systems

- n particles. Positions at time t: $q_1, \ldots, q_n$. Momenta at time t: $p_1, \ldots, p_n$.
- Hamiltonian: $H(p_1, \ldots, p_n; q_1, \ldots, q_n)$.
- Hamilton's equations:

$$\dot{p}_j := \frac{dp_j}{dt} = -\frac{\partial H}{\partial q_j}, \quad \dot{q}_j := \frac{dq_j}{dt} = \frac{\partial H}{\partial p_j}.$$

- Example:

$$H(p_1, \ldots, p_n; q_1, \ldots, q_n) = \frac{1}{2m}\sum_{j=1}^{n} p_j^2 + V(q_1, \ldots, q_n).$$

  Here m is the 'mass' and V is the 'potential energy'.
- Then we recover Newton's equations of motion:

$$\dot{p}_j = -\frac{\partial V}{\partial q_j}, \quad \dot{q}_j = \frac{p_j}{m}.$$

Conserved quantities

- A Hamiltonian flow has various conserved quantities. These are quantities that do not change over time.
- The Hamiltonian H is always conserved. Noether's theorem relates conserved quantities to symmetries of the system.
- If all the conserved quantities are known, then the system is called integrable. This is rarely the case.

An alternative representation

- Let $\psi_j := q_j + \mathrm{i} p_j$, where $\mathrm{i} = \sqrt{-1}$.
- Then Hamilton's equations become

$$\mathrm{i}\dot{\psi}_j = \frac{\partial H}{\partial \psi_j} := \frac{\partial H}{\partial q_j} + \mathrm{i}\frac{\partial H}{\partial p_j}.$$

The Schrödinger equation

- Consider the Hamiltonian

$$H(\psi_1, \ldots, \psi_n) = \frac{1}{2} \sum_{j=1}^{n} |\psi_{j+1} - \psi_j|^2,$$

where $n + 1 = 1$.

- This gives the equations

$$\mathrm{i}\dot{\psi}_j = -(\psi_{j+1} - 2\psi_j + \psi_{j-1}).$$

- This is the discrete analog of the Schrödinger equation

$$\mathrm{i}\frac{\partial \psi}{\partial t} = -\Delta\psi.$$

Straightforward generalization to higher dimensions.

- It should now be clear why the Schrödinger flow may be called an infinite dimensional Hamiltonian flow with Hamiltonian

$$H(\psi) := \frac{1}{2} \int_{\mathbb{R}^d} |\nabla\psi(t, x)|^2 dx.$$

The nonlinear Schrödinger equation

- The nonlinear Schrödinger equation (NLS) is a famous nonlinear PDE:
$$\mathrm{i}\frac{\partial \psi}{\partial t} = -\Delta\psi + \kappa|\psi|^{p-1}\psi.$$
Here $\kappa = \pm 1$, p is the 'nonlinearity parameter', and $\psi = \psi(t,x)$ is a flow evolving over time, with $x \in \mathbb{R}^d$ and $t \in \mathbb{R}$.
- Crucial to understanding a variety of physical phenomena: Bose-Einstein condensation, Langmuir waves in plasmas, nonlinear optics, envelopes of water waves, etc...
- The case $\kappa = +1$ is called the defocusing equation and the case $\kappa = -1$ is called the focusing equation.
- Just like the Schrödinger equation, the NLS is also an infinite dimensional Hamiltonian system, with Hamiltonian
$$H(\psi) = \frac{1}{2}\int_{\mathbb{R}^d} |\nabla\psi(t,x)|^2 dx + \frac{\kappa}{p+1}\int_{\mathbb{R}^d} |\psi(t,x)|^{p+1} dx.$$

The soliton resolution conjecture

- The long-term behavior of the NLS is shrouded in mystery. The main open problem is the so-called soliton resolution conjecture. (See e.g. Terry Tao's blog entry on this topic for a readable account.)
- The conjecture claims that for 'generic initial data', as $t \to \infty$, the NLS flow decomposes into a union of a finite number of receding 'solitons', plus a term that radiates to zero.
- A soliton is a stationary wave of the form $\psi(t, x) = v(x)e^{\mathrm{i}\omega t}$, where v must satisfy

$$\omega v = \Delta v - \kappa |v|^{p-1} v.$$

- Rigorously verified in a few special cases, but no general result.
- One approach to understanding the long-term behavior is through the study of invariant measures. We will talk about this approach in the next few slides.

Invariant manifolds

- A property of a finite dimensional Hamiltonian flow is that it preserves Lebesgue measure. (Liouville's theorem)
- Given a Hamiltonian system, Liouville's theorem implies that the uniform probability distribution on a manifold where a subset of all conserved quantities take specified values is an invariant measure for the flow.
- For example, for the NLS, the mass $M(\psi) = \int |\psi(t,x)|^2 dx$ and the energy $H(\psi)$ are conserved quantities. Therefore, the uniform probability distribution on the space of all functions with $M(\psi) = m$ and $H(\psi) = E$, if one can make sense of such a thing, would be an invariant measure for this flow.
- The study of invariant measures is one approach to understanding the long-term behavior of Hamiltonian flows; in particular, the NLS. (Long history: Lebowitz, Rose, Speer, Bourgain, Zhidkov, McKean, Vaninsky, Rider, Tzvetkov, Oh,...)

Uniform distributions on manifolds defined by a single equation

- Classical example: Uniform distribution on the simplex

$$\{(x_1, \ldots, x_n) \in \mathbb{R}_+^n : x_1 + \cdots + x_n = n\}.$$

- In this example, it is known that for n large, the coordinates behave like i.i.d. $Exp(1)$ random variables.
- Another classical example:

$$\{(x_1, \ldots, x_n) \in \mathbb{R}^n : x_1^2 + \cdots + x_n^2 = n\}.$$

- Here, the coordinates behave like i.i.d. $N(0,1)$ random variables.

Manifolds defined by two equations

- Simplest example:

$$\{(x_1, \ldots, x_n) \in \mathbb{R}_+^n : x_1 + \cdots + x_n = n,$$
$$x_1^2 + \cdots + x_n^2 = bn\}.$$

  Here b is a fixed constant.

- The uniform distribution on this set is an invariant measure for a flow with two conserved quantities.
- How does the uniform distribution on this manifold behave when n is large? I tried this out as a toy version of the NLS question.
- Note that we must have $b \geq 1$ for this set to be non-empty.

The subcritical phase

Theorem (C., 2010)

Let $(X_1, \ldots, X_n)$ be a vector chosen uniformly from the manifold defined in the previous slide. Suppose $1 < b \leq 2$. Then there exist unique $r, s \in \mathbb{R}$ such that the probability density proportional to $\exp(-rx^2 - sx)$ on $[0, \infty)$ has first moment 1 and second moment b. Let $Z_1, Z_2, \ldots$ be i.i.d. random variables following this density. Then following hold:

(a) *For any fixed k, the random vector $(X_1, \ldots, X_k)$ converges in law to $(Z_1, \ldots, Z_k)$ as $n \to \infty$.*

(b) *If $b < 2$, there is a constant C, possibly depending on b, such that*

$$\lim_{n\to\infty} \mathbb{P}\Big(\max_{1\leq i\leq n} X_i > C\sqrt{\log n}\Big) = 0.$$

(c) *When $b = 2$, part (c) holds but with $\log n$ instead of $\sqrt{\log n}$.*

Localization in the supercritical phase

Theorem (C., 2010)

Suppose $b > 2$. Let $Z_1, Z_2, \ldots$ be i.i.d. Exp(1) random variables. Then the following hold.

(a) *For any fixed k, the random vector $(X_1, \ldots, X_k)$ converges in law to $(Z_1, \ldots, Z_k)$ as $n \to \infty$.*

(b) *Let $M = \max_{1 \le i \le n} X_i$. Then*

$$\frac{M^2}{(b-2)n} \to 1 \text{ in probability.}$$

Consequently, the sum of squares of all other coordinates is roughly $2n$ with high probability.

(c) *Let M_2 be the value of the second largest coordinate. Then*

$$\frac{M_2^2}{n} \to 0 \text{ in probability.}$$

More on localization

- The proof indicates that the localization that we saw in the supercritical phase is typical in systems with multiple conserved quantities.
- In fact, the soliton resolution conjecture is itself a localization phenomenon: the solitons are localized features. There is no localization for the ordinary Schrödinger equation; the flow tends to zero in sup-norm.
- A preliminary analysis was done in Chatterjee & Kirkpatrick (2010) where localization was proved in the cubic discrete NLS, but a continuum limit could not be taken.
- Very recently, I was finally able to figure out how to take a continuum limit and consequently, prove a 'statistical version' of the soliton resolution conjecture (next slide).

'Statistical' soliton resolution

Theorem (C., 2012; rough statement)

Consider the d-dimensional focusing NLS with nonlinearity p. Suppose that $p < 1 + 4/d$ (subcritical nonlinearity), and that E is a real number bigger than the minimum possible energy at a given mass m. If we attempt to choose a function uniformly at random from all functions satisfying $M(\psi) = m$ and $H(\psi) = E$, by first discretizing the problem and then passing to the infinite volume continuum limit, then the resulting sequence of discrete random functions converges in sup-norm to the soliton of minimum energy at mass m.

Another motivation

- Systems with multiple conserved quantities may arise in other problems too.
- Example: large deviations in random graphs.
- Consider the Erdős-Rényi random graph $G(n, p)$.
- Let T be the number of triangles (or, some other subgraph) in $G(n, p)$. What is the behavior of $\mathbb{P}(T \geq (1+\epsilon)\mathbb{E}(T))$, where $\epsilon > 0$ is fixed, and $n \to \infty$, $p \to 0$? Was a well-known open question in random graphs for a long time.
- In a breakthrough work in 2004, Kim and Vu showed (simultaneously Janson and coauthors) that for any $\epsilon > 0$, there is a constant $C(\epsilon) > 0$ such that whenever $p > C(\epsilon)^{-1} n^{-1} \log n$,

$$e^{-C(\epsilon)^{-1} n^2 p^2 \log(1/p)} \leq \mathbb{P}(T \geq (1+\epsilon)\mathbb{E}(T)) \leq e^{-C(\epsilon) n^2 p^2}.$$

- However, the unmatched log factor between the upper and lower bounds persisted for a while; solved recently.

Upper tail for triangle counts

Theorem (C., 2010)

Let T be the number of triangles in an Erdős-Rényi graph $G(n,p)$. For each $\epsilon > 0$ there is a sufficiently small positive constant $C(\epsilon)$ such that whenever $C(\epsilon)^{-1} n^{-1} \log n \le p \le C(\epsilon)$, we have

$$e^{-C(\epsilon)^{-1} n^2 p^2 \log(1/p)} \le \mathbb{P}(T \ge (1+\epsilon)\mathbb{E}(T)) \le e^{-C(\epsilon) n^2 p^2 \log(1/p)}.$$

- The log factor arises due to the localization of triangles: given that $T \ge (1+\epsilon)\mathbb{E}(T)$, the key step is to show that there is with high probability a small number of vertices supporting a large number of triangles.
- This can be viewed as a problem of understanding the behavior of systems with multiple conserved quantities: here the conserved quantities are the number of edges and the number of triangles. (The number of edges appears in the probability mass function of the Erdős-Rényi random graph.)

Summary

- It is natural to study uniform probability distributions on sets defined by multiple implicit equations.
- Such sets arise as invariant manifolds for Hamiltonian flows with two or more conserved quantities.
- May also arise in various other domains, such as random graphs.
- Often, these probability measures exhibit localization. For example, solitons in nonlinear Schrödinger flows. Or clusters of triangles in random graphs.
- Progress has been made in only a few special problems; many questions remain open.

References

- S. Chatterjee, 2010. A note about the uniform distribution on the intersection of a simplex and a sphere. *arXiv preprint.*
- S. Chatterjee and K. Kirkpatrick, 2010. Probabilistic methods for discrete nonlinear Schrdinger equations. *arXiv preprint. Recently appeared in Comm. Pure Appl. Math.*
- S. Chatterjee, 2010. The missing log in large deviations for triangle counts. *arXiv preprint. To appear in Random Structures and Algorithms.*
- S. Chatterjee, 2012. Invariant measures and the soliton resolution conjecture. *arXiv preprint.*
- Plus: extensive references to the literature from the above manuscripts.

Possible research problems

- Analyze invariant measures for other Hamiltonian systems (hundreds of examples).
- Connect the study of invariant measures with the dynamics.
- Large deviations of sparse random graphs: this is an area with very few theorems. Many open questions in various random graph models, e.g. Erdős-Rényi, exponential random graph models (ERGMs), geometric random graphs, power law graphs, ...

Understanding localization

- We will consider the toy problem of understanding the uniform distribution on the intersection of a simplex and a sphere to grasp the basic philosophy behind localization on manifolds defined by multiple implicit equations.
- Let $X_1, \ldots, X_n$ be i.i.d. non-negative random variables with a continuous distribution function F.
- Let $\hat{F}$ be the empirical distribution function, i.e.

$$\hat{F}(t) := \frac{1}{n}\#\{i : X_i \le t\}.$$

- By the law of large numbers, $\hat{F} \approx F$. In fact, it is well-known that for any $\delta > 0$,

$$\mathbb{P}(\|\hat{F} - F\|_\infty > \delta) \le e^{-C(\delta)n}.$$

- What is the behavior of $\hat{F}$ conditioned on the rare event that $\sum X_i \ge n(1+\epsilon)\mathbb{E}(X_1)$, where ϵ is a fixed positive number?

Understanding localization, contd.

- If $X_1, \ldots, X_n$ are i.i.d. $Exp(1)$, then conditional on the rare event $\sum X_i \geq n(1+\epsilon)\mathbb{E}(X_1)$, it is easy to show that with high probability, $\hat{F}$ is close to the distribution function of an exponential random variable with mean $1+\epsilon$.
- In particular, conditioning on this rare event has the effect that the empirical distribution of the X_i's changes substantially.

Understanding localization, contd.

- Now suppose that $\mathbb{P}(X_1 > t)$ decays slower than exponential, e.g. like $e^{-\sqrt{t}}$.
- Let A be the event that $\sum X_i \geq n(1+\epsilon)\mathbb{E}(X_1)$, and B be the event that $\|\hat{F} - F\|_\infty > \delta$, where ϵ and δ are fixed constants.
- Then

$$\mathbb{P}(A) \geq \mathbb{P}(X_1 \geq n(1+\epsilon)\mathbb{E}(X_1)) \geq e^{-C(\epsilon)\sqrt{n}}.$$

- On the other hand we know that $\mathbb{P}(B) \leq e^{-C(\delta)n}$.
- Therefore,

$$\mathbb{P}(B|A) \leq \frac{\mathbb{P}(B)}{\mathbb{P}(A)} \leq e^{-C(\epsilon,\delta)n}.$$

- In other words, conditional on A, the empirical distribution is still close to F with high probability.
- This shows that the 'extra' contribution to $\sum X_i$ must come from only a few large values.

Localization in the simplex-sphere problem

- Recall: We wish to study the uniform distribution on the set

$$S := \{(x_1, \ldots, x_n) \in \mathbb{R}_+^n : x_1 + \cdots + x_n = n, \\ x_1^2 + \cdots + x_n^2 = bn\}.$$

- We will use the heuristic outlined in the previous slides to understand why localization occurs when $b > 2$.
- Let $X = (X_1, \ldots, X_n)$ be a point chosen uniformly from the simplex

$$S_1 := \{(x_1, \ldots, x_n) \in \mathbb{R}_+^n : x_1 + \cdots + x_n = n\}.$$

- Conditional on the event $X \in S_2$, where

$$S_2 := \{(x_1, \ldots, x_n) \in \mathbb{R}_+^n : x_1^2 + \cdots + x_n^2 = bn\},$$

the vector X is uniformly distributed on S.

Localization in the simplex-sphere problem, contd.

- But we know that X 'behaves' like a vector of i.i.d. $Exp(1)$ random variables.
- This allows the problem to be reduced to the problem of understanding the conditional behavior of a vector of i.i.d. $Exp(1)$ random variables given that the sum of squares is bn.
- The square of an $Exp(1)$ random variable has an upper tail that decays slower than exponential (indeed, like $e^{-\sqrt{t}}$).
- The expected value of the square of an $Exp(1)$ random variable is 2.
- Therefore if $b > 2$ and n is large, our previous argument via empirical distribution functions shows that if the rare event $\sum X_i^2 \approx bn$ happens, then with high probability there are a few exceptionally large values that contributed to the 'excess' sum of squares.
- It is a harder task to show that exactly one value contributes; we will not discuss this.

Localization in random graphs

- Let A be the event that $T \geq (1+\epsilon)\mathbb{E}(T)$, where T is the number of triangles.
- $\mathbb{P}(A)$ may be easily lower bounded by obtaining all the extra triangles by forming a small clique.
- For each edge e in the graph, consider the number T_e of triangles containing that edge.
- Let N be the number of edges e satisfying $T_e \geq (1+n^{-\delta})\mathbb{E}(T_e)$, where δ is some very small constant.
- Let B be the event that $N > n^{1+3\delta}$.
- It can be shown that $\mathbb{P}(B)$ is negligible compared to the lower bound on $\mathbb{P}(A)$, from which it follows that $\mathbb{P}(B|A)$ is small.
- Thus, if A happens, then with high probability most of the 'extra' triangles come from a small region.
- Actual proof is more complicated. A simpler argument along similar lines was given by DeMarco and Kahn.

Localization in nonlinear Schrödinger flows

- Unfortunately, too many things to summarize in one slide.
- To pique your interest, let me just mention that there is a 99 page article on arXiv with a lot of math, combining techniques from large deviations, PDE and harmonic analysis.

That's all for today!