On Gaussian-width gradient complexity and mean-field behavior of interacting particle systems and random graphs - Monday 13th November, 2017

Ronen Eldan

Note Taker: Sylvester Eriksson-Bique

Main references: Arxiv paper "Gaussian-width gradient complexity, reverse log-Sobolev inequalities and non-linear large deviations". `https://arxiv.org/pdf/1612.04346.pdf`

Eldan and Gross (his Ph.D. student): Arxiv "Exponential random graphs behave like mixtures of stochastic block models" `https://arxiv.org/abs/1707.01227`.

For the remaining references consult the first paper above and the discussion below.

1 Introductory

Four motivating examples from combinatorics and statistical mechanics. A priori these may seem disjoint from GFA (geometric functional analysis), but later the methods from GFA become useful. First, define for a graph g the number

$$T(g) = \text{The number of triangles in } g.$$

By $G \sim G(N,p)$ we denote a random graph with distribution of an Erdös-Renyi random graph, with edge probability p, and define the event. When $p = 1/2$ this is the uniform distribution over all graphs. Consider also the event for graphs of N vertices, for $\delta > 0$

$$E_N = \{T(G) \geq (1+\delta)\mathbb{E}(T(G)))\}.$$

Consider now the four motivating questions.

1. What is the asymptotics of $\mathbb{P}(E_N)$ when $G \sim G(N, p)$? Large deviations question.

2. Vague question. Conditioned on E_N, what doeos the graph G look like? Does it look like $G(N, p')$ for larger p', or is there some symmetry breaking involved? Phase transitions?

3. Consider the uniform random graph $G(N, 1/2)$, and define a weighted probability on graphs $p_\beta(g) = Z^{-1}e^{-\frac{\beta T(g)}{N}}$, where $\beta > 0$ and Z is the partition function, which normalizes p to have unit mass. Structure of g with distribution p_β? Can one estimate the partition function $Z(\beta, N)$?

4. $A \subset [n]$ uniformly random subset of integers. Let

   $$f(A) = \text{the number of k} - \text{term arithmetic progressions in } A.$$

   Consider the event

   $$E_n = \{f(A) \geq (1+\delta)\mathbb{E}(f(A))\}.$$

   Ask the same questions as in 1 and 2 for this distribution.

5. Potts/Ising model. For example, let H_n be a sequence of graphs with n-vertices and increasing degree as $n \to \infty$ (more interactions for large n). Consider the Ising model on such a graph, where one places a charge of ± 1 at each vertex and neighboring charges are either positively or negatively correlated. Again, ask asymptotics, large deviations etc.

All of these problems fall in the framework of taking a random point in $C_n = \{-1, 1\}^n$ (for $1, 2$, $n = \binom{N}{2}$). And we have a function $f\colon C_n \to \mathbb{R}$ and we are examining the weighted measure ν with density $d\nu = Z^{-1}e^f \, d\mu$, where μ is the uniform measure on C_n and Z is the partition function. We wish to understand Z.

Remark: The problem 2 and 1 are not too different. For 2 we consider $f(g) = e^{-\beta T(g)/N}$, and for 1 we have $f(g) = 1_{T(g) > (1+\beta)\mathbb{E}(T(G))}$.

2 Methodology and main results: Mean-field method

Existed in physics literature in an imprecise form, and one way to make it precise is presented in this talk. Initially presented by Chaterjee and Dembo (2014). Idea use Gibbs variational principle, which states that the logarithm of the partition function is given by

$$\log(Z) = \log \int e^f \ \mu = \sup_{\nu} \int f \ d\nu - H(\nu||\mu),$$

where

$$H(\nu||\mu) = \int \log \left(\frac{d\nu}{d\mu} \right) \ d\nu,$$

is the relative entropy or Kullback-Leibler divergence.

Initially this idea may seem too simplistic, because the supremum on the right is actually attained by $d\nu = Z^{-1}e^f d\mu$, which is the quantity we wish to understand. However, the revised idea is that we expect in certain situations that the sup is approximated by a product measure, or a mixture thereof. Further, the supremum over such ν can in some cases be explicitly computed, because product measures on C_n can be described by their center of mass in $[-1, 1]^n$. This has been done for example for problem 1 and 2 in (Lubetzky and Zhao 2014, and Bhattacharya et al. 2017).

Chatterjee and Dembo realized this strategy under two assumptions on f, one consisting of a complexity bound for the discrete gradient ∇f (amounting to a low covering number in terms of balls of definite size), and an additional technical condition on the second derivative corresponding to certain smoothness. This is Theorem 1.1 in Chaterjee, Dembo 2014.

In this talk this result is generalized to larger classes of functions, where convexity is measured using Gaussian width, and no assumption on the second derivative is necessary (for some of the results).

Remark: Recall, that the discrete gradient is given by

$$\partial_i f(x_1, \ldots, x_i, \ldots, x_n) = f(x_1, \ldots, 1, \ldots, x_n) - f(x_1, \ldots, -1, \ldots x_n),$$

where the co-orindate changing is the i'th one.

To realize this strategy, we would need to find a partition of C_n to sets A such that $f|_A$ would be close to linear. Since the optimizers for linear functions are product measures, this would give an approximate maximizer in terms of a mixture of product measures.

Define the Gaussian width of a compact $K \subset \mathbb{R}^n$

$$GW(K) = \mathbb{E}_{R \sim N(0,Id)} \sup_{x \in K} \langle x, R \rangle,$$

where $N(0, Id)$ is the distribution of a standard Gaussian vector in $\mathbb{R}^n$. Also, recall that the Wasserstein distance

$$W_H(\nu, \mu) = \inf_{(X,Y)} \mathbb{E}(d_H(X, Y)),$$

where the infimum is taken over all product measures with margials ν, μ, and the distance is measured in the Hamming metric.

Also, to state the main theorem we need to define tilts of measures. The tilt $\tau_\theta \nu$ is defined as

$$d\tau_\theta \nu = \exp(\langle x, \theta \rangle)\ d\nu Z^{-1}.$$

Theorem 1 (Theorem 1)**.** *If $d\nu = e^f\ d\mu$, and* Lip $(f) = O(1)$ *(w.r.t. the Hamming metric), and $GW(\{\nabla f\}) = o(n)$, then there exists a probability measure m on $[-\epsilon, \epsilon]^n$, where $\epsilon = o(1)$, s.t.*

$$\nu = \int \tau_\theta \nu\ dm(\theta),$$

and such that $\exists \Theta \subset \mathbb{R}^n$ with $m(\Theta) = 1 - o(1)$ and $\forall \theta \in \Theta$ there exists a product measure ξ_θ such that

$$W_H(\tau_\theta \nu, \xi_\theta) = o(n).$$

These error bounds turn out to be sufficient for the desired Large deviation estimates. In another paper by the speaker and his student Gross, they have strengthened the conclusion of the previous theorem in the following way.

Theorem 2. *In addition to the above, if one assumes a technical smoothness assumption for ∇f, we can also say that $\forall \theta \in \Theta$ if*

$$x_\theta = \int x \; d\tau_\theta \nu,$$

then

$$x_\theta \sim \tanh(\nabla f(x_\theta)),$$

which is the mean field equation.

The speaker gave a very high level description of the proof as below.

$$\text{Low Complexity} \Rightarrow \text{Reverse log} - \text{Sobolev inequality of } \nu \text{ with respect to } \mu,$$

which then can be used with classical techniques from (Fathi Indrei Ledoux, "Quantitative logarithmic Sobolev inequalities and stability estimates" 2014), and (Nourdin, Picatti, Ledoux, "Stein's method, logarithmic Sobolev and transport inequalities", '14), to conclud that the measure is close to a product measure or a mixture of product measures.

Other conclusions/discussion: As a result of this and other work, one observes that the structure of the conditioning on E_N has a phase transition in δ. For small δ, the measure looks like a $G(N, p')$ for larger p', but for large δ there is some symmetry breaking.