Beyond Spectral Gaps

Back to spanning trees: $k = n-1$

$\mathcal{B} \subset \binom{E}{n-1}$ $\quad P = P_{k \leftrightarrow k-1}$ $\quad \pi = \mathrm{Uni}_{\mathcal{B}}$

We proved ~~$\lambda(\mathcal{B})$~~ $\approx 1-$

$$1 - \lambda_1(P) = \Omega(1/k) \quad \text{optimal spectral gap.}$$

so $d_{\chi^2}(\tau P^s, \pi) \leq (1 - 1/k)^s \, d_{\chi^2}(\tau, \pi)$

$$\simeq \exp(-s/k) \, d_{\chi^2}(\tau, \pi)$$

Suppose $\tau = \delta_{\text{spanning tree}}$

$$d_{\chi^2}(\tau, \pi) = \mathbb{E}_{\pi}\left(\frac{d\tau}{d\pi}\right)^2 - 1 = (\#\text{spanning trees})^2 - 1.$$

$\#\text{spanning trees} \leq n^{n-2}$ (Cayley's formula)

So need $\exp(-s/k)\, n^{n-2}$

$s = \Omega(k^2 \log k)$ to prove mixing

Not linear # of steps!

<u>Fix</u> Need to use KL instead

$$d_{KL}(P, Q) = \mathbb{E}_P\left[\log \frac{dP}{dQ}\right] = \mathbb{E}_Q\left[\frac{dP}{dQ} \log \frac{dP}{dQ}\right].$$

$d_{KL}(\tau, \pi) \leq$ ~~$\log n$~~ $k \log k.$

$s = O(k \log k)$ suffices.

Mixing in KL determined by version of "log-Sobolev inequalities"

Modified log-Sobolev f positive-valued

local

$Ent[f] = \mathbb{E}[f \log f] - \mathbb{E}[f] \log \mathbb{E}[f]$

MLSI constant C

$Ent[f] \le C\, \mathcal{E}(f, \log f)$
↑ Dirichlet form.

Ex: MLSI constant upper bound $\frac{1}{1-\lambda_2}$.

Thm [CGM'19, ASKPV'21]

C-SI at all external fields ("fractional log-concave") $\mu^{\star\lambda}$, $\frac{1}{c}$

⇓

"Entropic independence" ⇒ $O(k^C)$ MLSI constant.

Rmk In fact, $Ent[U_{k\to k-1} f] \le (1 - \frac{1}{\Omega(k^C)}) Ent[f]$; Equiv $\frac{1}{k^C} Ent[f] \le (Ent[f] - Ent[\dots])$

Pf Idea: ① local-to-global w/ Ent instead of Var. $= \mathbb{E}[Ent[f|A_1]]$, A_1 = first element of S

Needs analogue of "Spectral independence..."

Def [AJKPV'21] μ on $\binom{[n]}{k}$

μ is C-Entropically independent if

$Ent[U_{1\to k} f] \le \frac{C}{k} Ent[f]$.

(Exactly analogous to C-SI w/ Var.)

② ...

Thm [AJKPV'21] C-SI at all fields ⇒ C-EI at all fields

Pf idea: Fix $q = U_{1\to k} f$ and compute min Ent f.

"maximum entropy problem"; optimizer f has an exponential form related to generating function (partition fn). ∎

Similarly for log-concave measures on $\mathbb{R}^n$

Thm [Bakry-Émery]

If $p(x) = e^{-H(x)}$, $\nabla^2 H(x) \succeq \alpha I$. ($\alpha$-strongly log-concave)

Then

$Ent[f] \le \frac{1}{\alpha} \mathcal{E}(f, \log f)$.

can be proven for Gaussian and then extended via Caffarelli contraction.

Also important consequence

"Herbst argument"

Thm f L-Lipschitz (wrt Euclidean/Hamming metric)

Then
$$\Pr_\mu(|f - \mathbb{E}[f]| > t) \le 2\exp(-t^2/(C\, c_{MLSI} L^2))$$

Pf idea: Let $X = f(y)$, $y \sim \mu$

① Use MLSI to show
$$\mathrm{Ent}[e^{\lambda X}] \le \frac{\lambda^2 \sigma^2 c_{MLSI}}{2}\, \mathbb{E}[e^{\lambda X}]$$

② "Herbst argument" X mean zero
$$\Psi(\lambda) = \log \mathbb{E}[e^{\lambda X}]$$
$$\frac{d}{d\lambda}\frac{\Psi(\lambda)}{\lambda} = \frac{\mathbb{E}[X e^{\lambda X}]}{\lambda \mathbb{E}[e^{\lambda X}]} - \frac{\Psi(\lambda)}{\lambda^2} = \frac{\mathrm{Ent}[e^{\lambda X}]}{\lambda^2 \mathbb{E}[e^{\lambda X}]}$$

so
$$\Psi(\lambda) = \lambda \int_0^\lambda \frac{d}{ds}\frac{\Psi(s)}{s}\, ds \le \frac{\lambda^2 \sigma^2}{2}$$

(Chernoff bound)

Note log-concave measures may not satisfy MLSI with any constant!

$e^{-|x|}$ is not sub-Gaussian....

so analogue of KLS for MLSI is false.

Another useful thm

Thm [Chen–Liu–Vigoda '21]

Suppose μ is

① C-SI at all links

② a graphical model w.r.t. a graph of max degree Δ (e.g. Ising)

③ marginally bounded, ie $\mu(\sigma x)/\mu(y) \in [1/b, b]$ if $x \sim y$

Then P (Glauber) has MLSI $\Omega_{C,\Delta,b}(\frac{1}{n})$. Entropic stability; local-to-global argument.

Idea: do local-to-global up to degree cn, then do "shattering argument"

Open question: Is Δ-dependence necessary?

Some ideas about KLS: Chen: $n^{o(1)}$ bound on $\frac{1}{1-\lambda_2}$

$\mu = \mathcal{O}$ isotropic

analogue of SI: $\|\Sigma_t\|_{op}$ bound.

$\mu_t \leftarrow \frac{1}{t}$-strongly log-concave

① rough analogue of local-to-global: if $\|\Sigma_t\|_{op} = O(1)$ up to time t, get roughly $\frac{1}{\sqrt{t}}$ bound on Poincaré.

② Hard part: show that $\|\Sigma_t\|_{op}$ doesn't blow up.
(or roughly equivalent when a_t stays near 0.)
$da_t = \Sigma_t dW_t$.

Folding Kee-Vempala
... Chen approach

compute $d\,\mathrm{Tr}(\Sigma_t^q)$ for large q and argue it is small

$$\|\Sigma_t\|_{op} \leq \sqrt[q]{\mathrm{Tr}(\Sigma_t^q)}.$$

Klartag-Lehec analysis: $\|\cdot\|^n$ instead, but picture very similar

Read the proof! It's not that complex...

Q: Is there a more general "trickle-up" principle we can formulate?

Trickle-down: control Σ_0 using Σ_t

"trickle-up": control Σ_t using Σ_0?

Last thoughts:

- Besides diffusion, what insights can they bring to practice?
- We mostly skipped connections to geometry, optimal transport...
  - algebraic geometry/topology...
  - complex analysis...
  - representation theory...
  - ~~learning theory~~
  - high-dimensional statistics/estimation...
  - complexity theory...
  - Gibbs measures...
  - variational inference...

Law of total entropy

$$\mathrm{Ent}[f|y] = \mathbb{E}[f \log f|y] - \mathbb{E}[f|y]\log \mathbb{E}[f|y]$$

$$\mathrm{Ent}[f] = \mathrm{Ent}[\mathbb{E}[f|y]] + \mathbb{E}[\mathrm{Ent}[f|y]]$$

$$\mathbb{E}[f \log f] - \mathbb{E}[f]\log \mathbb{E}[f]$$

$$\|$$

$$\mathbb{E}[f \log f] - \mathbb{E}[\mathbb{E}[f|y]\log \mathbb{E}[f|y]] + \mathbb{E}[\mathbb{E}[f|y]\log \mathbb{E}[f|y]] - \mathbb{E}[f]\log \mathbb{E}[f]$$

$$= \mathbb{E}[\mathbb{E}[f \log f|y]]$$

$$= \mathbb{E}[\mathrm{Ent}[f|y]] + \mathrm{Ent}[\mathbb{E}[f|y]]$$

$$KL(p,q) = \mathbb{E}_p\left[\log \frac{dP}{dQ}\right]$$

"given a code optimized for q, how wasteful is it for p?"

$$= \mathbb{E}_p[\log p] - \mathbb{E}_p[\log q]$$

$$= \mathbb{E}_p\left[\log \frac{1}{q}\right] - \mathbb{E}_p\left[\log \frac{1}{p}\right]$$