Transfer of χ^2-contraction to D_{KL}-contraction

- Two approaches:

  * Entropic Independence ← "tight" transfer
  * Entropy Factorization ← "loose" transfer which ends up OK in many apps

[A-Jain-Koehler-Pham-Vuong '21]

- Entropic Independence Approach:

  If $\lambda * \mu$ has $\lambda_2(D_{k\to 1} U_{1\to k}) \leq \frac{C}{k}$ for every $\lambda \in \mathbb{R}^n_{>0}$, then

  $$D_{KL}(\nu D_{k\to 1} \| \mu D_{k\to 1}) \leq \frac{C}{k} \quad \leftarrow \text{same } C$$

external field → $\boxed{\lambda * \mu(S) := \mu(S) \prod_{i\in S} \lambda_i}$

[Chen-Liu-Vigoda '21]

- Entropy Factorization Approach:

  If μ comes from product space $\Omega_1 \times \cdots \times \Omega_k$ that we look at as $\binom{\Omega_1 \sqcup \cdots \sqcup \Omega_k}{k}$, and μ and its links are spectrally independent and have marginals $\geq \Omega(1)$:

  $$\underset{x\sim\mu_T}{\mathbb{P}}[x_i = \omega_i] \geq \Omega(1)$$

  then

  $$D_{KL}(\nu D_{k\to\ell} \| \mu D_{k\to\ell}) \leq \Big[1 - \underbrace{\big(\tfrac{k-\ell}{k}\big)}_{\text{a constant for } k-\ell=\Omega(k)}\Big]^{O(1)} D_{KL}(\nu \| \mu)$$

Lemma [A-Jain-Koehler-Pham-Vuong]

We have $D_{KL}(\nu D_{k\to 1} \| \mu D_{k\to 1}) \leq \frac{C}{k} D_{KL}(\nu \| \mu)$

if and only if $\lg g_\mu(z^{\frac{1}{c}})$ is upperbounded by

its tangent at $z=1$: $\forall z \in \mathbb{R}_{>0}^n$

$$\lg g_\mu(z^{\frac{1}{c}}) \leq \lg g_\mu(1) + \nabla \lg g_\mu(1) \cdot (z^{\frac{1}{c}} - 1)$$

Bounded by tangent:

x^2 : Locally around $z=1$

D_{KL}: Globally at all z.

Corollary: If C-lg-concave $\Longrightarrow$

D_{KL} contracts by C/k.

Note: C-lg-concave means spectral independence under all external fields $\lambda * \mu$.

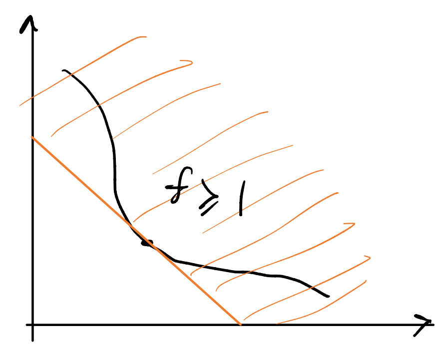
move 1 to any other point.

Lemma: [folklore] If f d-hom over $\mathbb{R}_{>0}^n$ TFAE ($\leftarrow g_\mu(z^{\frac{1}{c}})$)

- $\lg f$ concave
- f quasi-concave $\equiv \{z \mid f(z) \geq 1\}$ convex
- $f^{\frac{1}{d}}$ concave

and TFAE:

- $\lg f$ upperbounded by tangent
- $\{z \mid f(z) \geq 1\}$ lowerbounded by tangent
- $f^{\frac{1}{d}}$ upperbounded by tangent



Proof of Folklore Lemma:

lg-concave $\Rightarrow$ quasi-concave ✓

Now assume quasi-concave:

We want to show

$$f(\lambda x + (1-\lambda)y)^{\frac{1}{d}} \geq \lambda f(x)^{\frac{1}{d}} + (1-\lambda) f(y)^{\frac{1}{d}}$$

Let $x' = \alpha x$, $y' = \beta y$ such that

$f(x') = f(y') = 1$ $\quad \alpha = \frac{1}{f(x)^{\frac{1}{d}}}$ $\quad \beta = \frac{1}{f(y)^{\frac{1}{d}}}$

Then $f(z) \geq 1$ for z any convex combination of x' and y'. In particular

$$z := \Big(\underbrace{\frac{\lambda}{\alpha} x' + \frac{(1-\lambda)}{\beta} y'}_{\lambda x + (1-\lambda) y}\Big) \Big/ \Big(\frac{\lambda}{\alpha} + \frac{1-\lambda}{\beta}\Big)$$

But this means

$$f\Big(\frac{\lambda}{\alpha} x' + \frac{1-\lambda}{\beta} y'\Big) \geq \Big(\frac{\lambda}{\alpha} + \frac{1-\lambda}{\beta}\Big)^d \Rightarrow$$

$$f(\lambda x + (1-\lambda) y)^{\frac{1}{d}} \geq \lambda f(x)^{\frac{1}{d}} + (1-\lambda) f(y)^{\frac{1}{d}} \checkmark$$

$f^{\frac{1}{d}}$ concave $\Longrightarrow$ lg-concave ✓

take lg from above
use concavity of lg

The arguments for tangent versions are very similar ✓

Essentially because of d-hom all info about f is captured in its level set $\{z \mid f(z) \geq 1\}$.

Equivalence of Entropy Contraction and Tangent Bound

We want

$$D_{KL}(\nu D_{k\to 1} \| \mu D_{k\to 1}) \leq \frac{C}{k} D_{KL}(\nu \| \mu)$$

Main Idea: For a fixed $q = \nu D_{k\to 1}$, which ν minimizes $D_{KL}(\nu \| \mu)$? It turns out the answer is always some $\lambda * \mu$ for some $\lambda \in \mathbb{R}^n_{>0}$. Moreover

$$\min\{D_{KL}(\nu \| \mu) \mid \nu D_{k\to 1} = q\} = \lg\left(\sup_{z>0} \frac{\prod_i z_i^{kq_i}}{\underbrace{g_\mu(z)}_{\text{enough to upperbound}}}\right)$$

Now if we have C-lg-concavity then

$$\nabla g_\mu(1) = (kp_1, \ldots, kp_n) \text{ where } p = \mu D_{k\to 1}$$

$$\Rightarrow \nabla\left(g_\mu(z^{\frac{1}{C}})^{\frac{C}{k}}\right)\Big|_{z=1} = p$$

So

$$g_\mu(z^{\frac{1}{C}})^{\frac{C}{k}} \leq \cancel{g_\mu(1)} + p\cdot(z - \cancel{1}) = p\cdot z$$

$$\Downarrow$$

$$\underbrace{g_\mu(z) \leq \left(\sum p_i z_i^C\right)^{\frac{k}{C}}}_{\text{the upperbound we want}}$$

This means $\min\{D_{KL}(\nu \| \mu) \mid \nu D_{k\to 1} = q\}$ is at least $\lg\left(\sup_{z>0} \prod_i z_i^{kq_i} / (\sum p_i z_i^C)^{\frac{k}{C}}\right)$. Plug in $z_i = \left(\frac{q_i}{p_i}\right)^{\frac{1}{C}}$ we get

$$D_{KL}(\nu \| \mu) \geq \frac{k}{C} \sum_i q_i \lg \frac{q_i}{p_i} = \frac{k}{C} D_{KL}(\nu D_{k\to 1} \| \mu D_{k\to 1})$$

:)

It remains to show formula for D_{KL}

this is just convex duality

variables

$$\min\{D_{KL}(\nu\|\mu) \mid \nu D_{k\to i} = p\}$$

convex objective linear const.

Note that at optimum we have

$$\nabla_\nu D_{KL}(\nu\|\mu) = \sum_i \alpha_i v_i \quad \text{where}$$

v_i's are the constraint gradients,

namely $v_i \in \mathbb{R}^{\binom{n}{k}}$ and

$$v_i(S) = \frac{1}{k} 1[i \in S].$$

But $\partial_{\nu(S)} D_{KL}(\nu\|\mu) = \lg\frac{\nu(S)}{\mu(S)} - \lg\frac{\sum \nu(T)}{\sum \mu(T)}$.

This means

$$\frac{\nu(S)}{\sum_T \nu(T)} = \frac{\mu(S)}{\sum_T \mu(T)} \cdot \prod_{i\in S} \underbrace{e^{\alpha_i/k}}_{\lambda_i}$$

In other words, $\nu = \lambda * \mu$.

So OPT is found by the external field that achieves the correct marginals q.

We show the "dual" also has the same optimality condition. First we make the dual concave:

$$\text{dual} = \sup_{z>0}\left\{\lg\frac{\prod_i z_i^{kq_i}}{g_\mu(z)}\right\} = \sup_w\left\{\langle y, kq\rangle - \underbrace{\lg g_\mu(e^w)}_{\text{convex in } w}\right\}$$

OPT achieved when

$$\underbrace{\nabla_w \lg g_\mu(e^w)}_{\text{marginals of } e^w * \mu} = kq$$

So the primal and dual have the same conditions of optimality. We show their values are equal: Suppose $\lambda = e^w$ is the external field (we can assume $g_\mu(\lambda) := \sum \lambda^S \mu(S) = 1$, i.e., already normalized)

$$D_{KL}(\lambda * \mu \| \mu) = \sum_S (\lambda * \mu)(S) \lg \prod_{i \in S} \lambda_i =$$

$$\sum_i kq_i \lg \lambda_i \qquad \text{primal}$$

$$\lg \frac{\prod_i \lambda_i^{kq_i}}{g_\mu(\lambda)} = \lg \prod_i \lambda_i^{kq_i} = \sum_i kq_i \lg \lambda_i \qquad \text{dual} \qquad \square$$

Entropy Factorization

Main Ideas:

1. $D_{2\to 1}$ instead of $D_{k\to 1}$
   because quadratic polynomial
   We use spin system + marginal bounds here
2. Alternate local-to-global thm:
   Stitch $D_{2\to 1}$'s together
3. Either use big-step DU or try harder and compare!

① Suppose μ is a spin system with marginals $\geq \Omega(1)$. Then for any other ν we have

$$\frac{\nu D_{k\to 1}(i)}{\mu D_{k\to 1}(i)} = O(1)$$

because marginals of ν are ≤ 1

This makes χ^2 and D_{KL} comparable!

Lemma: With the above assumptions

$$D_{KL}(\nu D_{k\to 1} \| \mu D_{k\to 1}) \leq \left(\tfrac{1}{2} + O\left(\tfrac{1}{k}\right)\right) \times D_{KL}(\nu D_{k\to 2} \| \mu D_{k\to 2})$$

spectral ind + margs

② Alternative local-to-global:

Let $Z_i = D_{KL}(\nu D_{k\to i} \| \mu D_{k\to i})$.

Previous local-to-global would proceed:

$$(Z_1 - Z_0) \leq \bigcirc \cdot (Z_k - Z_0) \Rightarrow (Z_k - Z_1) \geq \square \cdot (Z_k - Z_0)$$

$$(Z_2 - Z_1) \leq \Delta \cdot (Z_k - Z_1) \Rightarrow (Z_k - Z_2) \geq \pentagon \cdot (Z_k - Z_1)$$

$$\vdots$$

we don't have these anymore

$$\Downarrow$$

$$(Z_k - Z_\ell) \geq \square \cdot \pentagon \cdot \cdots \cdot (Z_k - Z_0)$$

$$\Downarrow$$

$$(Z_\ell - Z_0) \leq (1 - \square \cdot \pentagon \cdots)(Z_k - Z_0)$$:)

Instead now we know

$$(z_{i+1} - z_i) \leqslant \left(\tfrac{1}{2} + O\left(\tfrac{1}{k-i}\right)\right)(z_{i+2} - z_i)$$

apply ② to μ_T for $|T|=i$ and average over T.

Note that this can be rewritten as

$$\underbrace{(z_{i+1} - z_i)}_{A_i} \leqslant \underbrace{\left(1 + O\left(\tfrac{1}{k-i}\right)\right)}_{\alpha_i} \underbrace{(z_{i+2} - z_{i+1})}_{A_{i+1}}$$

Question: $A_i \leqslant \alpha_i \cdot A_{i+1} \quad \forall i.$

$$\max \frac{\overbrace{A_0 + A_1 + \cdots + A_{\ell-1}}^{D_{KL}(\nu D_{k\to\ell} \| \mu D_{k\to\ell})}}{\underbrace{A_0 + \cdots + A_{k-1}}_{D_{KL}(\nu \| \mu)}} = ?$$

Answer: When $A_i = \alpha_i A_{i+1}$!

$$\frac{A_i}{A_{k-1}} = \left(1 + \frac{O(1)}{k-i}\right)\left(1 + \frac{O(1)}{k-i-1}\right) \cdots \left(1 + \frac{O(1)}{k-j+1}\right)$$

$$\simeq (k-i)^{O(1)}$$

So we have

$$1 - \frac{A_0 + \cdots + A_{\ell-1}}{A_0 + \cdots + A_{k-1}} \simeq \frac{1 + 2^{O(1)} + \cdots + (k-\ell)^{O(1)}}{1 + 2^{O(1)} + \cdots + k^{O(1)}}$$

$$\simeq \frac{k-\ell}{k} \cdot \frac{(k-\ell)^{O(1)}}{k^{O(1)}} = \left(\frac{k-\ell}{k}\right)^{O(1)}$$

This proves that

$$D_{KL}(\nu D_{k\to\ell} \| \mu D_{k\to\ell}) \leqslant \left(1 - \left(\frac{k-\ell}{k}\right)^{O(1)}\right) D_{KL}(\nu \| \mu)$$

☺

③ For ℓ close to k, the factor

$1-\left(\frac{k-\ell}{k}\right)^{O(1)}$ becomes very bad!

However if $k-\ell=\Omega(k)$, this factor

is always $1-\Omega(1)$:)

Idea 1: Use the $D_{k\to\ell}U_{\ell\to k}$ walk for

$k-\ell=\Omega(k)$! It has $\Omega(1)$ MLSI

and mixes in just $\tilde{O}(1)$ steps.

Idea 2: [Chen-Liu-Vigoda] Showed that for

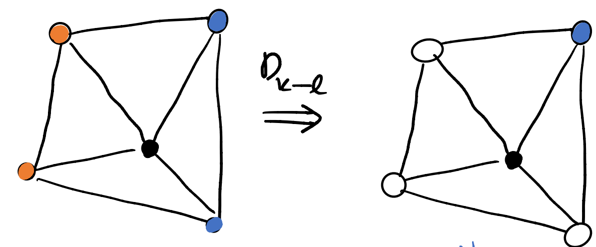
bounded degree spin systems, we can

translate above to $\Omega(\frac{1}{k})$ MLSI for

$D_{k\to k-1}U_{k-1\to k}$, a.k.a. Glauber dynamics.

Let us expand Idea 1. Main problem

is how to implement $U_{\ell\to k}$ step.



⇓

now we need to resample the erased

Lem: If degs bounded by Δ, there is

threshold $\ell_0=k-\Omega_\Delta(k)$ s.t. for $\ell\geq\ell_0$,

the "erased" vertices will be in islands

of size $O_\Delta(\lg n)$ w.h.p.

We can sample each island separately!

Proof: - If we erase $< \frac{1}{\Delta}$ fraction of vertices, then local neighborhood of any v grows at most like a tree with avg. branching factor $< \Delta \cdot \frac{1}{\Delta} = 1$. These trees die off quickly.

- Here is a loose argument:

Let v be some vertex. We know #connected subgraphs containing v of size t is at most $\Delta^{2(t-1)} \leq \Delta^{2t}$

remember Euler tour argument from Patel-Regts?

Suppose $\ell = pk$, $p \in [0,1]$.

Any such conn. neighborhood gets fully erased w.p. $\leq (1-p)^t$

easy exercise

Thus by union bound,

$$\mathbb{P}[\,|v\text{'s island}| \geq t\,] \leq \Delta^{2t}(1-p)^t.$$

So if $1-p \leq \frac{1}{2\Delta^2}$ and $t \geq C \cdot \lg n$, this is at most n^{-C}. Union bounding over all v gives us

$$\mathbb{P}[\text{any island} \geq C \cdot \lg n] \leq n^{1-C}.$$

Note that we could set $p = 1 - \frac{1}{2\Delta^2}$, which means $k - \ell = \Omega(k)$. :)

It remains to prove ①:

$$\boxed{\begin{aligned}\nu^{(i)} &= \nu D_{k\to i}\\ \mu^{(i)} &= \mu D_{k\to i}\end{aligned}}$$

$$D_{KL}(\nu^{(1)} \| \mu^{(1)}) \leqslant \left(\tfrac{1}{2} + O\left(\tfrac{1}{k}\right)\right) D_{KL}(\nu^{(2)} \| \mu^{(2)})$$

Proof: We can use an upperbound for $g_{\mu^{(2)}}$:

$$g_{\mu^{(2)}}(z) \leqslant \lambda\left(\sum p_i z_i^2\right) + (1-\lambda)\left(\sum p_i z_i\right)^2$$

where $\lambda = O\left(\frac{1}{k}\right)$, $p = \mu^{(1)}$.

This follows because $g_{\mu^{(2)}}$ is a quadratic, i.e., $z^t A z$ and A can be bounded by spectral independence as diagonal + rank 1.

Now again if $q = \nu^{(1)}$, then

$$D_{KL}(\nu^{(2)} \| \mu^{(2)}) \geqslant \lg \sup_{z>0} \left\{ \frac{z_1^{2q_1} \cdots z_n^{2q_n}}{g_{\mu^{(2)}}(z)} \right\} \geqslant$$

$$\lg \sup_{z>0} \left\{ \frac{z_1^{2q_1} \cdots z_n^{2q_n}}{\lambda\left(\sum p_i z_i^2\right) + (1-\lambda)\left(\sum p_i z_i\right)^2} \right\}.$$

If we plug in $z_i = \frac{q_i}{p_i}$ we get

$$D_{KL}(\nu^{(2)} \| \mu^{(2)}) \geqslant \lg\left(\frac{e^{2 D_{KL}(q \| p)}}{1 + \lambda \chi^2(q \| p)}\right) \geqslant$$

$$\lg\left(\frac{e^{2 D_{KL}(q \| p)}}{e^{\lambda \chi^2(q \| p)}}\right) = 2 D_{KL}(q \| p) - \lambda \chi^2(q \| p).$$

Lemma: If $q < O(1) \cdot p$ everywhere, and q, p are prob dists, then $\chi^2(q \| p) = O(D_{KL}(q \| p))$.

Exercise: Prove this.