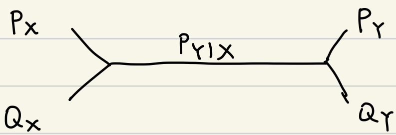
Lec 12: Strong Data Processing Inequalities

Yanjun Han

Recall:



DPI: $D_{KL}(Q_Y \| P_Y) \leq D_{KL}(Q_X \| P_X)$

SDPI: $D_{KL}(Q_Y \| P_Y) \leq \eta(P_{Y|X}) D_{KL}(Q_X \| P_X)$

for some $\eta(P_{Y|X}) < 1$.

Input-independent SDPI.

Defn. Given a channel $P_{Y|X}$, define

$$\boxed{\eta(P_{Y|X}) = \sup_{P_X \neq Q_X} \frac{D_{KL}(Q_Y \| P_Y)}{D_{KL}(Q_X \| P_X)}}$$

Properties: ① $\displaystyle \eta(P_{Y|X}) = \sup_{U-X-Y} \frac{I(U;Y)}{I(U;X)}$

Pf. ($\geq$) $I(U;Y) = \mathbb{E}_U[D_{KL}(P_{Y|U} \| P_Y)]$

$\leq \mathbb{E}_U[\eta(P_{Y|X}) \cdot D_{KL}(P_{X|U} \| P_X)]$ (why?)

$= \eta(P_{Y|X}) \cdot I(U;X)$

($\leq$) Choose $U \sim \text{Bern}(p)$ and $P_{X|U=1} = \tilde{P}_X$. $P_{X|U=0} = \tilde{Q}_X$.

Then $I(U;X) = \mathbb{E}_U[D_{KL}(P_{X|U} \| P_X)]$

$= p \cdot D_{KL}(\tilde{P}_X \| p \cdot \tilde{P}_X + (1-p)\tilde{Q}_X)$

$+ (1-p) D_{KL}(\tilde{Q}_X \| p \cdot \tilde{P}_X + (1-p)\tilde{Q}_X)$

$\Rightarrow \frac{d}{dp} I(U;X)\Big|_{p=0} = D_{KL}(\tilde{P}_X \| \tilde{Q}_X) + \mathbb{E}_{\tilde{Q}_X}\left[\frac{\tilde{P}_X - \tilde{Q}_X}{\tilde{Q}_X}\right]$

$= D_{KL}(\tilde{P}_X \| \tilde{Q}_X)$

$\Rightarrow I(U;X) = p \cdot D_{KL}(\tilde{P}_X \| \tilde{Q}_X) + o(p)$

$I(U;Y) = p \cdot D_{KL}(\tilde{P}_Y \| \tilde{Q}_Y) + o(p)$

$\Rightarrow \frac{I(U;Y)}{I(U;X)} \to \frac{D_{KL}(\tilde{P}_Y \| \tilde{Q}_Y)}{D_{KL}(\tilde{P}_X \| \tilde{Q}_X)}$ as $p \to 0^+$.

② $$\eta(P_{Y|X}) = \sup_{P_X, Q_X \text{ binary}} \frac{D_{KL}(Q_Y \| P_Y)}{D_{KL}(Q_X \| P_X)}.$$

Pf. It suffices to show that, for any $\eta > 0$ and (P_X, Q_X), and

$$f(P_X, Q_X) = D_{KL}(Q_Y \| P_Y) - \eta \cdot D_{KL}(Q_X \| P_X)$$

we can always find binary distributions $(\tilde{P}_X, \tilde{Q}_X)$ s.t. $f(P_X, Q_X) \le f(\tilde{P}_X, \tilde{Q}_X)$.

To prove it, note that

$$\hat{P} \mapsto f\left(\hat{P}, \frac{Q_X}{P_X}\hat{P}\right) = \underbrace{D_{KL}\left(P_{Y|X} \circ \frac{Q_X}{P_X}\hat{P} \,\Big\|\, P_{Y|X} \circ \hat{P}\right) - \eta D_{KL}\left(\frac{Q_X}{P_X}\hat{P} \,\Big\|\, \hat{P}\right)}_{\text{linear in } \hat{P}}$$

is convex on $\left\{\hat{P} : \int \frac{Q_X}{P_X}\hat{P} = 1, \int \hat{P} = 1\right\}$. When $\hat{P} = P_X$, the value is $f(P_X, Q_X)$.

The maximizer $\hat{P}^*$ of this map must belong to the extremal points of this set, i.e. $\hat{P}^*$ must be a binary distribution. $\square$

③ $$\eta(P_{Y|X}) = \sup_{x, x' \in \mathcal{X}} LC_{max}(P_{Y|X=x}, P_{Y|X=x'})$$

$$= \sup_{x, x' \in \mathcal{X}} \sup_{0<\beta<1} \beta(1-\beta) \int \frac{(P_{Y|X=x} - P_{Y|X=x'})^2}{(1-\beta)P_{Y|X=x} + \beta P_{Y|X=x'}}.$$

In particular, $$\frac{1}{2}\text{diam}_{H^2} \le \eta(P_{Y|X}) \le \text{diam}_{H^2} - \frac{\text{diam}_{H^2}^2}{4}$$

where $$\text{diam}_{H^2}(P_{Y|X}) = \sup_{x, x' \in \mathcal{X}} H^2(P_{Y|X=x}, P_{Y|X=x'}).$$

Pf. The first claim follows from ② and computations for binary distributions; see textbook for details. For the second claim:

$(\ge)$ $$LC_{max}(P \| Q) \overset{\beta = 1/2}{\ge} \int \frac{(P-Q)^2}{2(P+Q)} \ge \frac{1}{2}\int (\sqrt{P} - \sqrt{Q})^2 = \frac{H^2(P, Q)}{2}$$

$(\le)$ $$1 - \beta(1-\beta)\int \frac{(P-Q)^2}{(1-\beta)P + \beta Q} \overset{\text{check}}{=} \int \frac{PQ}{(1-\beta)P + \beta Q} \overset{\text{C-S}}{\ge} \left(\int \sqrt{PQ}\right)^2$$

$$= \left(1 - \frac{H^2}{2}(P, Q)\right)^2.$$

$\square$

Example (EC_δ). For EC_δ, $P_{Y|X} = \begin{cases} X & \text{w.p. } 1-\delta \\ ? & \text{w.p. } \delta \end{cases}$, so

$$I(U;Y) = (1-\delta)\, I(U;X) \quad \text{for all } U - X - Y \quad \text{(HW 1)}$$

Therefore, $\eta(EC_\delta) = 1-\delta$.

Example (BSC_δ). For BSC_δ, $X \in \{0,1\}$ and $Y = X \oplus \text{Bern}(\delta)$.

In this case,

$$LC_{max}(P_{Y|X=0}, P_{Y|X=1}) = \sup_{\beta \in (0,1)} \beta(1-\beta)\left(\frac{(1-2\delta)^2}{(1-\beta)(1-\delta)+\beta\delta} + \frac{(1-2\delta)^2}{(1-\beta)\delta + \beta(1-\delta)} \right)$$

$$= (1-2\delta)^2 \sup_{\beta\in(0,1)} \frac{\beta(1-\beta)}{\underbrace{[(1-\beta)(1-\delta)+\beta\delta][(1-\beta)\delta+\beta(1-\delta)]}_{\text{between } \beta \text{ and } 1-\beta \text{ sum into } 1}}$$

$$= (1-2\delta)^2$$

Therefore, $\eta(BSC_\delta) = (1-2\delta)^2$.

Example (tensorization) $\eta(P_{Y|X}^{\otimes n}) \le 1-(1-\eta(P_{Y|X}))^n$.

Pf. For $U - X^n - Y^n$,

$$I(U;Y^n) = I(U;Y_2^n) + I(U;Y_1 \mid Y_2^n)$$

$$\le I(U;Y_2^n) + \eta(P_{Y|X})\, I(U;X_1 \mid Y_2^n)$$

$$= (1-\eta(P_{Y|X}))\, I(U;Y_2^n) + \eta(P_{Y|X})\, \underbrace{I(U;X_1,Y_2^n)}_{\le I(U;X^n)}$$

Continuing this process gives

$$\frac{I(U;Y^n)}{I(U;X^n)} \le \eta(P_{Y|X}) \sum_{t=0}^{n-1} (1-\eta(P_{Y|X}))^t = 1-(1-\eta(P_{Y|X}))^n. \quad \boxed{\text{Ⓛ}}$$

(A general result: in a Bayesian network, each vertex v is open w.p. $\eta(P_{v|\text{Parent}(v)})$.
Then $\eta(P_{X_S|X_0}) \le P(\exists \text{ an open path from } 0 \text{ to some vertex in } S)$
= "percolation" probability from 0 to S.)

Input-dependent SDPI.

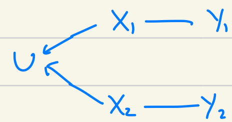
Defn. Given a channel $P_{Y|X}$ and input P_X,

$$\eta(P_X, P_{Y|X}) = \sup_{Q_X} \frac{D_{KL}(Q_Y \| P_Y)}{D_{KL}(Q_X \| P_X)}$$

Properties: ① $\eta(P_X, P_{Y|X}) = \sup_{U-X-Y} \frac{I(U;Y)}{I(U;X)}$

② $\eta(P_X^{\otimes n}, P_{Y|X}^{\otimes n}) = \eta(P_X, P_{Y|X})$.

Pf. By induction, suffice to prove the case $n=2$.



$$I(U; Y_1, Y_2)$$

$$= I(U;Y_1) + I(U;Y_2|Y_1)$$

$U - X_2 - Y_2 \mid Y_1$ and $P_{X_2|Y_1} = P_{X_2}$

$$\le \eta\big(I(U;X_1) + I(U;X_2|Y_1)\big)$$

$$= \eta\big(I(U;X_1) + \underbrace{I(U;X_2|X_1,Y_1)}_{=I(U;X_2|X_1)} + \underbrace{I(X_1;X_2|Y_1)}_{=0} - \underbrace{I(X_1;X_2|Y_1,U)}_{\ge 0}\big)$$

$$\le \eta\, I(U; X_1, X_2)$$ ☒

Unlike $\eta(P_{Y|X})$, the input-dependent SDPI constant $\eta(P_X, P_{Y|X})$ can be much more challenging to characterize. An example is when $P_{Y|X}$ is the transition matrix of a Markov chain, and $P_X = \pi$ is its stationary distribution. Then SDPI says that

$$D_{KL}(\pi_0 P^n \| \pi) = D_{KL}(\pi_0 P^n \| \pi P^n) \le \eta(\pi, P)^n \cdot D_{KL}(\pi_0 \| \pi), \quad \forall \pi_0.$$

This is called the <u>modified log-Sobolev inequality</u>, and leads to upper bounds on the mixing time. Both tasks could be challenging for Markov chains.

Example. $(X,Y) \sim N\left(\begin{pmatrix}0\\0\end{pmatrix}, \begin{pmatrix}1 & \rho\\ \rho & 1\end{pmatrix}\right)$ is jointly Gaussian with correlation ρ.

Claim: $\eta(P_X, P_{Y|X}) = \eta(P_Y, P_{X|Y}) = \rho^2$.

Pf. We'll only prove the upper bound $\eta(P_X, P_{Y|X}) \le \rho^2$; see below for lower bound.

By scaling, WLOG we assume that $Y = X + Z$ with $Z \sim N(0, \frac{1}{\rho^2} - 1)$.

For any $\tilde{X}$ and $\tilde{Y} = \tilde{X} + Z$,

$$D_{KL}(P_{\tilde{Y}} \| P_Y) = -h(\tilde{Y}) + \log\frac{\sqrt{2\pi}}{\rho} + \frac{\rho^2}{2}\mathbb{E}[\tilde{Y}^2]$$

$$\overset{EPI}{\le} -\frac{1}{2}\log\left(2\pi e\left(\frac{1}{\rho^2}-1\right) + e^{2h(\tilde{X})}\right) + \log\frac{\sqrt{2\pi}}{\rho} + \frac{\rho^2}{2}\mathbb{E}[\tilde{Y}^2].$$

and $$D_{KL}(P_{\tilde{X}} \| P_X) = -h(\tilde{X}) + \log\sqrt{2\pi} + \frac{1}{2}\mathbb{E}[\tilde{X}^2].$$

Rearranging: $$D_{KL}(P_{\tilde{Y}} \| P_Y) \le -\frac{1}{2}\log\left(1-\rho^2+\rho^2 e^{\mathbb{E}[\tilde{X}^2] - 2D_{KL}(P_{\tilde{X}}\|P_X) - 1}\right) + \frac{\rho^2}{2}\left(\mathbb{E}[\tilde{X}^2]-1\right)$$

concavity of log: $\log(1-p+px) \ge p\log x$

$$\le -\frac{\rho^2}{2}\log\left(e^{\mathbb{E}[\tilde{X}^2] - 2D_{KL}(P_{\tilde{X}}\|P_X) - 1}\right) + \frac{\rho^2}{2}\left(\mathbb{E}[\tilde{X}^2]-1\right)$$

$$= \rho^2 D_{KL}(P_{\tilde{X}} \| P_X) \qquad \boxed{\Sigma}$$

(Another quantity: $$\eta_{\chi^2}(P_X, P_{Y|X}) = \sup_{Q_X} \frac{\chi^2(Q_Y \| P_Y)}{\chi^2(Q_X \| P_X)}.$$

Properties: ① $\eta_{\chi^2} \le \eta_{KL}$

② $\eta_{\chi^2} = \sigma_2(M)^2$, where $\sigma_1(M) \ge \sigma_2(M) \ge \cdots \ge 0$ are singular values of

$$M_{x,y} = \frac{P_{X,Y}(x,y)}{\sqrt{P_X(x)P_Y(y)}}, \quad x \in \mathcal{X}, y \in \mathcal{Y}.$$

③ $\sqrt{\eta_{\chi^2}}$ = maximal correlation between X and Y, defined as

$$\sup_{g_1, g_2} \mathrm{corr}(g_1(X), g_2(Y)) = \sup_{g_1, g_2} \frac{\mathrm{Cov}(g_1(X), g_2(Y))}{\sqrt{\mathrm{Var}(g_1(X))\mathrm{Var}(g_2(Y))}}.$$

④ In Markov chains,

$$\chi^2(\pi_0 P^n \| \pi) \le \eta_{\chi^2}(\pi, P)^n \cdot \chi^2(\pi_0 \| \pi).$$

This is the Poincare's inequality.

By ①+③, for jointly Gaussian (X, Y), $\eta_{KL} \ge \eta_{\chi^2} = (\text{maximal correlation})^2 = \rho^2$.)

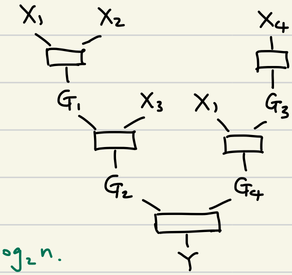
Applications of SDPI.

Example 1 (noisy gates). Suppose a noisy gate is an {AND, OR, NOT} gate with output corrupted by a Bern(δ) noise.

Q: For every $\delta < 1/2$, can we still reliably compute all Boolean functions $\{0,1\}^n \to \{0,1\}$?

Claim: $I(X_i; Y) \le \left(2(1-2\delta)^2\right)^{d_i}$,

where d_i is the minimum distance from X_i to Y.



Answer to question: No. Suppose we'd like to compute $XOR(X_1, \dots, X_n) = \sum_{i=1}^{n} X_i \bmod 2$, then $\exists\, i \in [n]$ with $d_i \ge \log_2 n$.

For this i, if $\delta > \frac{1}{2} - \frac{1}{2\sqrt{2}} \approx 0.15$,

$$I(X_i; Y) \le \left(2(1-2\delta)^2\right)^{\log_2 n} \to 0 \quad \text{as } n \to \infty.$$

Since $XOR(X_1, \dots, X_n)$ is sensitive to every X_i, its computation is impossible.

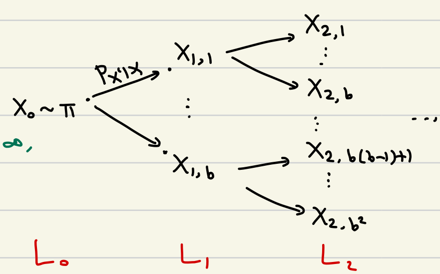
Pf of claim.

$$\begin{aligned} I(X_i; Y) &\le \eta(P_{Y|X_i}) \cdot H(X_i) \\ &\le \eta(P_{Y|X_i}) \\ &\le \text{percolation probability from } X_i \text{ to } Y \\ &= \sum_{\substack{\text{paths from} \\ X_i \text{ to } Y}} (1-2\delta)^{2 \cdot \text{length(path)}}. \end{aligned}$$

When $\text{length(path)} \ge d_i$ and $2(1-2\delta)^2 \le 1$, the sum is $\le \left[2(1-2\delta)^2\right]^{d_i}$ ☒

Example 2 (Broadcast on trees) Let $(\pi, P_{X'|X})$ be a reversible Markov chain.

Consider the broadcasting problem on an infinite b-ary tree.



Q: Given all $(X_{D,i})_{i \in [d^D]}$ on layer $D \to \infty$, can you recover X_0 reliably?

Claim: No if $b \cdot \eta(\pi, P_{X'|X}) < 1$.

Pf. $\quad I(X_0; X_{L_D}) \le \sum_{v \in L_1} I(X_0; X_{L_{D,v}}) \quad (L_{D,v} = \{u \in L_D : v \in \text{ancestor}(u)\})$

$$\le \eta(\pi, P_{X'|X}) \sum_{v \in L_1} I(X_v; X_{L_{D,v}}) \quad (X_{L_{D,v}} \to X_v \to X_0 \quad \uparrow \text{transition is } P_{X'|X} \text{ by reversibility})$$

$$= b\eta(\pi, P_{X'|X}) \cdot I(X_0; X_{L_{D-1}})$$

$$\Rightarrow I(X_0; X_{L_D}) \le (b\eta(\pi, P_{X'|X}))^D \cdot H(X_0) \xrightarrow{D \to \infty} 0 \quad \text{if } b\eta < 1. \qquad \boxed{\text{Ч}}$$

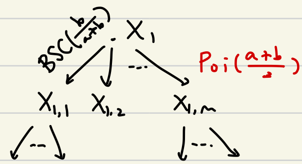
Application: stochastic block model. In 2-SBM$(\frac{a}{n}, \frac{b}{n})$, a vector $X \sim \text{Unif}(\{\pm 1\}^n)$ is drawn, and

$$\mathbb{P}((i,j) \text{ connected} \mid X) = \begin{cases} \frac{a}{n} & \text{if } X_i X_j = 1 \text{ (same community)} \\ \frac{b}{n} & \text{if } X_i X_j = -1 \text{ (different community)} \end{cases}$$

Q: When can we recover $X_1 X_2 \in \{\pm 1\}$ with $\Omega(1)$ probability, as $n \to \infty$?

Claim: We cannot if $\frac{(a-b)^2}{2(a+b)} < 1$ (Kesten-Stigum threshold)

Pf. Since all edge probabilities are of order $\Theta(\frac{1}{n})$, w.h.p. all vertices with distance $\le d$ from X_1 forms a tree (i.e. no cycles), for some $d = d_n \to \infty$.



In addition, # of children $\sim \text{Poi}(\frac{a+b}{2})$.

label flipping prob $= \frac{b}{a+b}$.

W.h.p. X_2 does not belong to this local neighborhood of X_1, so

$$I(X_1; X_2 \mid G) \le I(X_1; (X_i)_{i \in L_d} \mid G)$$

$$\le \left(\frac{a+b}{2} \cdot \left(1 - 2 \cdot \frac{b}{a+b}\right)^2\right)^d \quad \text{(see HW3 for details)}$$

$$= \left[\frac{(a-b)^2}{2(a+b)}\right]^d \longrightarrow 0 \quad \text{if } \frac{(a-b)^2}{2(a+b)} < 1. \qquad \boxed{\text{Ч}}$$

Example 3 (Spiked Wigner model) $\quad X \sim \text{Unif}(\{\pm 1\}^n)$ unknown

observation: $Y = \sqrt{\frac{\lambda}{n}} XX^T + W \quad (W_{ij} = W_{ji} \overset{\text{i.i.d.}}{\sim} N(0,1))$

Claim: If $\lambda < 1$, then $I(X_1; X_2 \mid Y) = o(1)$

(i.e. weak recovery of X is impossible; the threshold $\lambda \le 1$ is the famous BBP transition)

Pf. The idea is that Y_{ij} is determined by X_iX_j through $Y_{ij} \mid \underbrace{X_iX_j}_{\text{call it } \theta_{ij}\in\{\pm 1\}} \sim N(\sqrt{\frac{\lambda}{n}}X_iX_j, 1)$.

For this Gaussian channel $P = \{N(\sqrt{\frac{\lambda}{n}},1), N(-\sqrt{\frac{\lambda}{n}},1)\}$, can show that

$$\eta := \eta(P) = LC_{1/2}\left(N(\sqrt{\tfrac{\lambda}{n}},1), N(-\sqrt{\tfrac{\lambda}{n}},1)\right) = \frac{\lambda}{n}(1+o(1)).$$

Next we replace Y_{ij} by Z_{ij} with $Z_{ij} \mid \theta_{ij} = \begin{cases} \theta_{ij} & \text{w.p. } \eta \\ ? & \text{w.p. } 1-\eta \end{cases}$ (i.e. $EC(1-\eta)$).

then for any
$$\begin{tikzcd} & & Y_{ij} \\ U \arrow[r] & \theta_{ij} \arrow[ur] \arrow[dr] & \\ & & Z_{ij} \end{tikzcd}$$
, we have

$$I(U;Y_{ij}) \le \eta I(U;\theta_{ij}) = I(U;Z_{ij}).$$

We claim that

$$I(X_1;Y \mid X_2) \le I(X_1; Z \mid X_2) \qquad (*).$$

Indeed, assuming $(*)$,

$$\begin{aligned} I(X_1;X_2 \mid Y) &= I(X_1;X_2,Y) && (I(X_1;Y)=0) \\ &= I(X_1;Y \mid X_2) && (I(X_1;X_2)=0) \\ &\le I(X_1;Z \mid X_2) && (\text{by } (*)) \\ &= I(X_1;X_2 \mid Z) \\ &\le \mathbb{P}(1 \text{ and } 2 \text{ are connected in the graph induced by } Z) \end{aligned}$$

(no edge between (i,j) $\Leftrightarrow$ $Z_{ij} = ?$)

Since this graph is Erdős-Rényi$(\frac{\lambda}{n}(1+o(1)))$, it's known that when $\lambda<1$, the largest connected component has size $O(\log n)$. So $\mathbb{P}(1 \text{ and } 2 \text{ connected}) \to 0$. $\square$

Pf of $(*)$. WLOG assume that $Y = (Y_1, Y_2)$, then

$$\begin{aligned} I(X_1;Y \mid X_2) &= I(X_1;Y_1 \mid X_2) + I(X_1;Y_2 \mid X_2,Y_1) \\ &\le I(X_1;Y_1 \mid X_2) + \eta I(X_1;\theta_2 \mid X_2,Y_1) \\ &= I(X_1;Y_1 \mid X_2) + I(X_1;Z_2 \mid X_2,Y_1) \\ &= I(X_1;Y_1,Z_2 \mid X_2) \end{aligned}$$

$$\begin{tikzcd} & & Y_1 \\ & \theta_1 \arrow[ur] \arrow[dr] \arrow[dd, no head] & \\ (X_1,X_2) \arrow[ur] \arrow[dr] & & Z_1 \\ & \theta_2 \arrow[r] \arrow[dr] & Y_2 \\ & & Z_2 \end{tikzcd}$$

Proceeding with same arguments gives $I(X_1;Y \mid X_2) \le I(X_1;Z \mid X_2)$. $\square$

Example 4 (proximal sampling) Suppose we'd like to sample from $\pi(x) \propto e^{-f(x)}$, $f: \mathbb{R}^d \to \mathbb{R}$.

A proximal sampler works as follows: it aims to sample from

$$\pi(x, y) \propto \exp\left(-f(x) - \frac{1}{2\eta}\|x-y\|^2\right)$$

via an iterative procedure. Given initialization $x_0 \sim P_{x_0}$, for each $t = 0, 1, \cdots$:

- Given x_t, sample $y_t \mid x_t \sim N(x_t, \eta I)$;
- Given y_t, sample $x_{t+1} \mid y_t \sim \pi^{X|Y}(\cdot \mid y_t)$. ($\eta$-strongly log-concave for convex f)

Claim: if π satisfies α-LSI, i.e. $D_{KL}(p\|\pi) \le \frac{1}{2\alpha} FI(p\|\pi) := \frac{1}{2\alpha} \mathbb{E}_p[\|\nabla \log \frac{p}{\pi}\|^2]$, $\forall p$,

then
$$D_{KL}(P_{x_t}\|\pi) \le \frac{D_{KL}(P_{x_0}\|\pi)}{(1+\alpha\eta)^{2t}}.$$

Pf. We'll show that

$$D_{KL}(P_{Y_t}\|\pi_Y) \le \frac{D_{KL}(P_{X_t}\|\pi)}{1+\alpha\eta} \quad ①$$

$$D_{KL}(P_{X_{t+1}}\|\pi) \le \frac{D_{KL}(P_{Y_t}\|\pi)}{1+\alpha\eta} \quad ②$$

$(\pi_Y = \pi * N(0, \eta I))$

via SDPIs. This is equivalent to saying that

$$\eta(\pi, N(\cdot, \eta I)) \le \frac{1}{1+\alpha\eta},$$

$$\eta(\pi_Y, \pi^{X|Y}(\cdot \mid y)) \le \frac{1}{1+\alpha\eta}.$$

Forward step. Let $\rho_0 = P_{X_t}$, $\pi_0 = \pi$, and

$$\partial_t \rho_t = \frac{1}{2}\Delta \rho_t$$
$$\partial_t \pi_t = \frac{1}{2}\Delta \pi_t$$

(heat flow)

then $\rho_\eta = P_{Y_t}$, $\pi_\eta = \pi_Y$. Now

$$\begin{aligned}
\partial_t D_{KL}(\rho_t\|\pi_t) &= \partial_t \int \rho_t \log \frac{\rho_t}{\pi_t} \\
&= \frac{1}{2}\int \Delta\rho_t\left(\log\frac{\rho_t}{\pi_t}+1\right) - \frac{1}{2}\int \Delta\pi_t \cdot \frac{\rho_t}{\pi_t} \\
&= -\frac{1}{2}\int \nabla\rho_t \cdot \nabla \log\frac{\rho_t}{\pi_t} + \frac{1}{2}\int \nabla\pi_t \cdot \nabla \frac{\rho_t}{\pi_t} \\
&= -\frac{1}{2}\mathbb{E}_{\rho_t}\left[\nabla \log \rho_t \cdot \nabla \log \frac{\rho_t}{\pi_t}\right] + \frac{1}{2}\mathbb{E}_{\rho_t}\left[\nabla \log \pi_t \cdot \nabla \log \frac{\rho_t}{\pi_t}\right] \\
&= -\frac{1}{2} FI(\rho_t\|\pi_t).
\end{aligned}$$

Since π_0 is α-LSI, can show that $\pi_t = \pi_0 * N(0, tI)$ is $(\frac{1}{\alpha}+t)^{-1}$-LSI.

Therefore,
$$\partial_t D_{KL}(\rho_t \| \pi_t) \le -\frac{1}{\frac{1}{\alpha}+t} D_{KL}(\rho_t \| \pi_t)$$

$$\Rightarrow \quad \frac{D_{KL}(\rho_\eta \| \pi_\eta)}{D_{KL}(\rho_0 \| \pi_0)} \le \exp\left(-\int_0^\eta \frac{1}{\frac{1}{\alpha}+t} dt\right) = \frac{1}{1+\alpha\eta}.$$

<u>Backward step</u>. Let $\rho_0^- = P_{Y_k}$, $\pi_0^- = \pi_Y$, and

$$\partial_t \rho_t^- = -\mathrm{div}(\rho_t^- \nabla \log \pi_t^-) + \frac{1}{2}\Delta \rho_t^- = \mathrm{div}\left(\rho_t^- \nabla \log \frac{\rho_t^-}{\pi_t^-}\right) - \frac{1}{2}\Delta \rho_t^-$$

$$\partial_t \pi_t^- = -\mathrm{div}(\pi_t^- \nabla \log \pi_t^-) + \frac{1}{2}\Delta \pi_t^- = -\frac{1}{2}\Delta \pi_t^-,$$

then $\rho_t^- = P_{X_{k+1}}$, $\pi_t^- = \pi$ (by the reverse process of diffusion model)

Therefore,
$$\partial_t D_{KL}(\rho_t^- \| \pi_t^-) = \partial_t \int \rho_t^- \log \frac{\rho_t^-}{\pi_t^-}$$

$$= \int \left(\mathrm{div}\left(\rho_t^- \nabla \log \frac{\rho_t^-}{\pi_t^-}\right) - \frac{1}{2}\Delta \rho_t^-\right)\left(\log \frac{\rho_t^-}{\pi_t^-} + 1\right)$$

$$- \int \left(-\frac{1}{2}\Delta \pi_t^-\right) \frac{\rho_t^-}{\pi_t^-}$$

$$= -\int \rho_t^- \nabla \log \frac{\rho_t^-}{\pi_t^-} \cdot \nabla \log \frac{\rho_t^-}{\pi_t^-} + \frac{1}{2} FI(\rho_t^- \| \pi_t^-)$$

$$= -\frac{1}{2} FI(\rho_t^- \| \pi_t^-)$$

$$\le -\frac{1}{\frac{1}{\alpha}+(\eta-t)} D_{KL}(\rho_t^- \| \pi_t^-) \quad \left(\begin{array}{l}\pi_t^- = \pi_{\eta-t} \\ \text{is } (\frac{1}{\alpha}+\eta-t)^{-1}\text{-LSI}\end{array}\right)$$

$$\Rightarrow \quad \frac{D_{KL}(\rho_\eta^- \| \pi_\eta^-)}{D_{KL}(\rho_0^- \| \pi_0^-)} \le \exp\left(-\int_0^\eta \frac{dt}{\frac{1}{\alpha}+\eta-t}\right) = \frac{1}{1+\alpha\eta}. \qquad \boxed{S}$$

<u>Special Topic: Guest lecture by Y. Gu on SDPIs</u>