SHANNON ENTROPY AND LOGARITHMIC SOBOLEV INEQUALITIES

PIOTR NAYAR

Institute for Mathematics and its Applications,
University of Minnesota

Abstract. We review several topics related to the Gross's logarithmic Sobolev inequality. This includes connections to the concentration of measure theory, information theory, combinatorics and the theory of finite Markov chains.

1. Entropy and combinatorics

In the first section we study the Shannon entropy of discrete random variables and use its properties to derive certain results in the field of combinatorics. Let $\Omega =$ be a probability space and let $X : \Omega \to M$ be a discrete random variable, meaning that the range of X is finite. Here M could be any set. Let $p(x) = \mathbb{P}(X = x)$. The Shannon entropy of X is defined via the formula

$$H(X) = -\sum_x p(x) \ln p(x).$$

Here $0 \ln 0$ is interpreted as 0. Since $p(x) \le 1$ we get $H(X) \ge 0$ with equality only when $\mathbb{P}$ is a Dirac delta. Assume that the range of X has cardinality n. Then from Jensen inequality (for concave function $\ln x$) we get

$$H(X) = \sum_x p(x) \ln\left(\frac{1}{p(x)}\right) \le \ln\left(\sum_x \frac{p(x)}{p(x)}\right) = \ln n.$$

Thus we have.

Fact 1. For a discrete random variable X we have $H(X) \le \ln |r(X)|$, where $r(X)$ is the range of X.

For a random variable (X, Y) we define the conditional probability

$$p(x|y) = \frac{p(x,y)}{p(y)}.$$

Note that we have $p(y) = \sum_x p(x,y)$. We define conditional entropy of X given $Y = y$

$$H(X|Y = y) = -\sum_x p(x|y) \ln p(x|y)$$

and the entropy of X given Y

$$H(X|Y) = \mathbb{E}_y H(X|Y = y).$$

E-mail address: nayar@mimuw.edu.pl.

Fact 2. We have $H(X|Y) = H(X,Y) - H(Y)$ and
$$H(X|Y) = \sum_{x,y} p(x,y) \ln\left(\frac{p(y)}{p(x,y)}\right).$$

Proof. We have
$$\begin{aligned} H(X|Y) = \mathbb{E}_y H(X|Y=y) &= \sum_y p(y) H(X|Y=y) = -\sum_y \sum_x p(y) p(x|y) \ln p(x|y) \\ &= -\sum_y \sum_x p(y) \frac{p(x,y)}{p(y)} \ln\left(\frac{p(x,y)}{p(y)}\right) = -\sum_y \sum_x p(x,y) \ln\left(\frac{p(x,y)}{p(y)}\right) \\ &= -\sum_y \sum_x p(x,y) \ln p(x,y) + \sum_y \sum_x p(x,y) \ln p(y) = H(X,Y) - H(Y). \end{aligned}$$
□

The relation
$$H(X,Y) = H(Y) + H(X|Y)$$
is called the *chain rule* for the entropy.

Fact 3. We have $H(X|Y) \leq H(X)$. Moreover, $H(X|Y) = H(X)$ if and only if X and Y are independent.

Proof. Using Jensen inequality we get
$$\begin{aligned} H(X|Y) &= \sum_{x,y} p(x,y) \ln\left(\frac{p(y)}{p(x,y)}\right) = \sum_x p(x) \sum_y \frac{p(x,y)}{p(x)} \ln\left(\frac{p(y)}{p(x,y)}\right) \\ &\leq \sum_x p(x) \ln\left(\sum_y \frac{p(x,y)}{p(x)} \frac{p(y)}{p(x,y)}\right) = \sum_x p(x) \ln\left(\frac{1}{p(x)}\right) = H(X). \end{aligned}$$
The equality in the case of independent random variables follows from the fact that we have equality in Jensen inequality if and only if $p(y)/p(x,y)$ does not depend on y. Thus, $p(y) = h(x)p(x,y)$ fore some h. Summing over y give $h(x) = 1/p(x)$ and thus the condition $p(x,y) = p(x)p(y)$, which means independence. □

Fact 4. We have $H(X|Y,Z) \leq H(X|Y)$. In other words (using chain rule)
$$H(X,Y,Z) + H(Y) \leq H(X,Y) + H(Y,Z).$$

Proof. Again using Jensen inequality one gets
$$\begin{aligned} H(X|Y,Z) &= \sum_{x,y,z} p(x,y,z) \ln\left(\frac{p(y,z)}{p(x,y,z)}\right) = \sum_{x,y} p(x,y) \sum_z \frac{p(x,y,z)}{p(x,y)} \ln\left(\frac{p(y,z)}{p(x,y,z)}\right) \\ &\leq \sum_{x,y} p(x,y) \ln\left(\sum_z \frac{p(y,z)}{p(x,y)}\right) = \sum_{x,y} p(x,y) \ln\left(\frac{p(y)}{p(x,y)}\right). \end{aligned}$$
□

The following fact is the so-called subadditivity of the Shannon entropy.

Fact 5. We have $H(X_1, \ldots, X_n) \leq H(X_1) + \ldots + H(X_n)$. Moreover, there is equality if and only if $X_1, \ldots, X_n$ are independent.

Proof. Using chain rule $n-1$ times (and Fact 3) gives us

$$\begin{aligned} H(X_1,\dots,X_n) &= H(X_1|X_2,\dots,X_n) + H(X_2,\dots,X_n) = \dots \\ &= H(X_1|X_2,\dots,X_n) + H(X_2|X_3,\dots,X_n) + \dots + H(X_{n-1}|X_n) + H(X_n) \\ &\le H(X_1) + H(X_2) + \dots + H(X_n). \end{aligned}$$

□

We are now ready to state the so-called Shearer's lemma.

Proposition 1 (Shearer's lemma)**.** Let $(X_1,\dots,X_n)$ be a random vector and take consider sets $S_1,\dots,S_m \subseteq [n]$. Define $X_S = \{X_i : i \in S\}$. Assume that for any $i \in [n]$ there is at least k sets $S_{i_1},\dots,S_{i_l}$, $l \ge k$ that contain i. Then

$$kH(X_1,\dots,X_n) \le \sum_{i=1}^{m} H(X_{S_i}).$$

Moreover, if S is a random subset of $[n]$ such that for every i we have $\mathbb{P}(i \in S) \ge p$ then $pH(X_1,\dots,X_n) \le \mathbb{E}_S H(X_S)$.

Proof. Using chain rule we have

$$kH(X_1,\dots,X_n) = kH(X_1) + kH(X_2|X_1) + \dots + kH(X_n|X_1,\dots,X_{n-1}).$$

Let us list the elements of S_j in an increasing order, $S_j = \{t_1^{(j)} < \dots < t_{l_j}^{(j)}\}$. Note that

$$\begin{aligned} H(X_{S_j}) &= H(X_{t_1^{(j)}}) + H(X_{t_2^{(j)}}|X_{t_1^{(j)}}) + \dots + H(X_{t_{l_j}^{(j)}}|X_{t_1^{(j)}},\dots X_{t_{l_j-1}^{(j)}}) \\ &\ge H(X_{t_1^{(j)}}|X_{t_1^{(j)}-1}, X_{t_1^{(j)}-2}, X_1) + H(X_{t_2^{(j)}}|X_{t_2^{(j)}-1}, X_{t_2^{(j)}-2}, X_1) + \dots \\ &\quad + H(X_{t_{l_j}^{(j)}}|X_{t_{l_j}^{(j)}-1}, X_{t_{l_j}^{(j)}-2},\dots,X_1). \end{aligned}$$

After using this estimate we are left with terms of the form $H(X_q|X_{q-1},\dots,X_1)$. If we sum those estimates up for $j = 1,\dots,m$ we see that each term of this form will appear at least k times, since each q is contained in at least k sets S_j.

For the probabilistic version, observe that if we set $X_{<i} = (X_{i-1},\dots,X_1)$, then we just observed that $H(X_S) \ge \sum_{i\in S} H(X_i|X_{<i})$. Taking expectation gives

$$\begin{aligned} \mathbb{E}_S H(X_S) &\ge \mathbb{E}_S \sum_{i\in S} H(X_i|X_{<i}) = \mathbb{E}_S \sum_{i\in[n]} \mathbf{1}_{\{i\in S\}} H(X_i|X_{<i}) = \sum_{i\in[n]} \mathbb{P}(i\in S) H(X_i|X_{<i}) \\ &\ge p \sum_{i\in[n]} H(X_i|X_{<i}) = pH(X_1,\dots,X_n). \end{aligned}$$

□

Example 1. If (X_1,X_2,X_3) is our random vector and $S_1 = \{2,3\}$, $S_2 = \{1,3\}$, $S_3 = \{1,2\}$ then we can take $k = 2$ and thus get

$$2H(X_1,X_2,X_3) \le H(X_1,X_2) + H(X_2,X_3) + H(X_3,X_1).$$

This can be generalized to the case of a vector $(X_1,\dots,X_n)$ and $S_j = [n] \setminus \{j\}$, $j = 1,\dots,n$. We then get

$$(n-1)H(X_1,\dots,X_n) \le H(X_1,X_2,\dots,X_{n-1}) + H(X_1,\dots,X_{n-2},X_n) + \dots + H(X_2,X_3,\dots,X_n).$$

Let us derive our first combinatorial statement using the above lemma.

Proposition 2 (Loomis-Whitney inequality)**.** Let P be a finite set of points in $\mathbb{R}^n$. Let P_i be the projection of P onto the hyperplane $\{x_i = 0\}$. Then

$$|P|^{n-1} \leq \prod_{i=1}^{n} |P_i|.$$

Proof. Let $(X_1, \ldots, X_n)$ be the vector uniformly distributed on P. Thus, from Fact 1 we have $H = H(X_1, \ldots, X_n) = \ln |P|$. Note that $H_i = H(X_1, \ldots, X_{i-1}, X_{i+1}, \ldots, X_n)$ has range of cardinality $|P_i|$. Therefore, $H_i \leq \ln |P_i|$. Using Shearer's lemma (actually, the example above) we get

$$(n-1) \ln |P| = (n-1)H \leq \sum_{i=1}^{n} H_i \leq \sum_{i=1}^{n} \ln |P_i| = \ln \prod_{i=1}^{n} |P_i|.$$

□

To state another application let us introduce the so-called fractional cover of graph G.

Definition 1. Let $G = (V, E)$ be a (undirected) graph. A fractional cover of G is a function $\phi : E \to [0, 1]$ such that for every vertex $v \in G$ we have $\sum_{e \in E, e \sim v} \phi(e) \geq 1$. We also take

$$\alpha^\star(G) = \inf \left\{ \sum_{e \in E} \phi(e) \middle| \ \phi \text{ fractional cover of } G \right\}.$$

Definition 2. Let T, G be two graphs. We say that $\psi : V(T) \to V(G)$ is a graph homomorphism if $u \sim v$ implies $\psi(u) \sim \psi(v)$. The sets of all homomorphisms of T into G will be denoted by $\operatorname{Hom}(T, G)$.

We shell prove the following proposition.

Proposition 3. For any two graphs T, G we have $|\operatorname{Hom}(T, G)| \leq (2|E(G)|)^{\alpha^\star(T)}$.

Proof. Let $\sigma : V(T) \to V(G)$ be the random uniform homomorphism. Suppose that we have $V(T) = \{v_1, \ldots, v_n\}$ and let us define the random variables $X_i = \sigma(v_i)$. Take a vector $X = (X_1, \ldots, X_n)$. Note that by uniformity of σ we get $H(X) = |\operatorname{Hom}(T, G)|$. Let $\phi : E(T) \to [0, 1]$ be the optimal fractional cover, i.e. $\sum_{e \in E(T)} \phi(e) = \alpha^\star(T)$. Choose a random edge S (random subset $S \subseteq V(T)$ of cardinality 2 with $\mathbb{P}(e) = \phi(e)/\alpha^\star(T)$. For any i we have $\mathbb{P}(v_i \in S) \geq 1/\alpha^\star(T)$, since $\sum_{e \sim v_i} \phi(e) \geq 1$. Thus,

$$\frac{1}{\alpha^\star(T)} |\operatorname{Hom}(T, G)| = \frac{1}{\alpha^\star(T)} H(X_1, \ldots, X_n) \leq \mathbb{E}_S H(X_S) \leq \ln(2|E(G)|).$$

Example 2. If T is a triangle K_3 then it is easy to see that $\alpha^\star(T) = 3/2$. Thus, we get $|\operatorname{Hom}(K_3, G)| \leq (2|E(G)|)^{3/2}$. Is this the best possible bound (up to a universal constant in front of the right hand side)?

□

2. Isoperimetric inequality on the hypercube

2.1. **Influences.** Let $f : \{-1, 1\}^n \to \{-1, 1\}$. The *influence* of the i-th variable is defined as

$$I_i(f) = \mathbb{P}(f(x) \neq f(\sigma_i(x))) = \frac{1}{2^n} \left| \{ x \in \{-1, 1\}^n : \ f(x) \neq f(\sigma_i(x)) \} \right|.$$

Here $\mathbb{P}$ is the uniform measure on the cube.

There is an one-to-one correspondence between Boolean functions and subsets of the discrete cube. Namely, if $f : \{-1,1\}^n \to \{-1,1\}$ then we can define $A_f = \{x : \ f(x) = 1\}$. If $A \subset \{-1,1\}^n$ then we also have $f_A(x) = 2\mathbf{1}_A(x) - 1$. If we have sets $A, B \subset \{-1,1\}^n$ with then we define

$$E(A,B) = |\{(a,b) : \ a \in A, b \in B, a \sim b\}|.$$

The quantity $E(A, A^c)$ is the so-called the edge boundary of A. We have

$$\frac{|E(A,A^c)|}{2^{n-1}} = \frac{2|E(A,A^c)|}{2^n} = \frac{\sum_{i=1}^n |\{x : \ f_A(x) \neq f_A(\sigma_i(x))\}|}{2^n} = \sum_{i=1}^n I_i(f_A).$$

The influence (total influence) of a Boolean function $f : \{-1,1\}^n \to \{-1,1\}$ is defined as

$$I(f_A) = \sum_{i=1}^n I_i(f_A) = \frac{|E(A,A^c)|}{2^{n-1}}.$$

2.2. **Examples of Boolean functions and their influences.** In this section we analyse some basis examples of Boolean functions.

- Dictator: $\mathrm{Dict}_n(x_1,\ldots,x_n) = x_j$, $1 \le j \le n$,
  Clearly, we have
  $$I_i(\mathrm{Dict}_n) = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}, \qquad I(\mathrm{Dict}_n) = 1, \qquad \mathbb{E}(\mathrm{Dict}_n) = 0.$$
- Junta (k-junta): $f(x_1,\ldots,x_n) = g(x_{i_1},\ldots,x_{i_k})$, where $g : \{-1,1\}^k \to \{-1,1\}$ and $1 \le k < n$.
- Parity: $\mathrm{Par}_n(x_1,\ldots,x_n) = x_1 \cdot \ldots \cdot x_n$. Note that Parity is equal to the Walsh function of highest degree, namely $w_{[n]}$.
  $$I_i(\mathrm{Par}_n) = 1, \qquad I(\mathrm{Par}_n) = n, \qquad \mathbb{E}(\mathrm{Par}_n) = 0.$$
- Majority: $\mathrm{Maj}_n(x_1,\ldots,x_n) = \mathrm{sgn}(x_1 + \ldots + x_n)$, n is odd,
  $$I_i(\mathrm{Maj}_n) = \frac{1}{2^{n-1}}\binom{n-1}{\frac{n-1}{2}} = O\left(\frac{1}{\sqrt{n}}\right), \quad I(\mathrm{Maj}_n) = \frac{n}{2^{n-1}}\binom{n-1}{\frac{n-1}{2}} = O(\sqrt{n}),$$
  $\mathbb{E}(\mathrm{Maj}_n) = 0.$
- AND: $\mathrm{AND}_n(x_1,\ldots,x_n) = \min(x_1,\ldots,x_n)$,
  $$I_i(\mathrm{AND}_n) = \frac{1}{2^{n-1}}, \qquad I(\mathrm{AND}_n) = \frac{n}{2^{n-1}}, \qquad \mathbb{E}(\mathrm{AND}_n) = -1 + \frac{1}{2^{n-1}}.$$
- OR: $\mathrm{OR}_n(x_1,\ldots,x_n) = \max(x_1,\ldots,x_n)$
  $$I_i(\mathrm{OR}_n) = \frac{1}{2^{n-1}}, \qquad I(\mathrm{OR}_n) = \frac{n}{2^{n-1}}, \qquad \mathbb{E}(\mathrm{OR}_n) = 1 - \frac{1}{2^{n-1}}.$$
- Tribes: take $n = mk$ and divide n variables into m groups (tribes), each of cardinality k. The value of our function is 1 if and only if there exists a tribe which says 'yes'. The tribe says 'yes' if all values of spines in this tribe is 1. So the Tribes function is OR of ANDs. We can write
  $$\mathrm{Tribes}_{k,m}(x_1,\ldots,x_n) = OR\left(AND(x_1,...,x_k),...,AND(x_{(m-1)k+1},...,x_{mk})\right).$$

To calculate I_i observe that if x_i wants to decide then others variables in its tribe has to take value 1 and in $m-1$ other tribes there must be at least 1 variable with value 0 in each tribe. Therefore,

$$I_i(\mathrm{Tribes}_{k,m}) = \frac{1}{2^{k-1}}\left(1-\frac{1}{2^k}\right)^{m-1}, \quad I(\mathrm{Tribes}_{k,m}) = \frac{km}{2^{k-1}}\left(1-\frac{1}{2^k}\right)^{m-1},$$
$$\mathbb{E}(\mathrm{Tribes}_{k,m}) = 1-2\left(1-\frac{1}{2^k}\right)^{m}.$$

Now we would like to find the value $k=k(n)$ for which $\mathbb{P}(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}) = p$. Let us take

$$k(n) = \log_2\left(\frac{n}{-\ln(1-p)}\right) - \log_2\log_2 n.$$

Of course $k(n)$ and $n/k(n)$ should be integers, but who cares... Since for a Boolean function f we have $\mathbb{E}f = 2\mathbb{P}(f=1)-1$, therefore

$$\begin{aligned} 1-\mathbb{P}(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}=1) &= \left(1-\frac{1}{2^{k(n)}}\right)^{n/k(n)} \\ &= \left(1+\frac{(\ln(1-p))(\log_2 n)}{n}\right)^{n/k(n)}. \end{aligned}$$

Let

$$a_n = \frac{n}{(\ln(1-p))(\log_2 n)}.$$

Clearly, $\lim_{n\to\infty}|a_n| = +\infty$. Therefore $\lim_{n\to\infty}(1+\frac{1}{a_n})^{a_n} = e$. Moreover,

$$\lim_{n\to\infty}\frac{n}{k(n)a_n} = \lim_{n\to\infty}\frac{(\ln(1-p))(\log_2 n)}{\log_2\left(\frac{n}{-\ln(1-p)}\right) - \log_2\log_2 n} = \ln(1-p).$$

It follows that

$$\lim_{n\to\infty}\mathbb{P}(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}=1) = 1-e^{\ln(1-p)} = p.$$

Let us now calculate the asymptotic behaviour of $I_i(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}})$. We have

$$\begin{aligned} I_i(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}) &= \frac{1}{2^{k(n)-1}}\left(1-\frac{1}{2^k}\right)^{n/k(n)-1} \\ &= \frac{1}{2^{k(n)-1}}\left(1-\frac{1}{2^k}\right)^{-1}\left(1-\mathbb{P}(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}=1)\right) \\ &\approx \frac{1}{2^{k(n)-1}}(1-p) \approx 2(1-p)\ln\left(\frac{1}{1-p}\right)\frac{\log_2 n}{n}. \end{aligned}$$

Therefore,

$$I_i(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}) \approx 2(1-p)\ln\left(\frac{1}{1-p}\right)\frac{\log_2 n}{n}, \quad n\to\infty,$$

$$I_i(\mathrm{Tribes}_{k(n),\frac{n}{k(n)}}) \approx 2(1-p)\ln\left(\frac{1}{1-p}\right)\log_2 n, \quad n\to\infty.$$

If $p \le 1/2$ then we have

$$I_i(\text{Tribes}_{k(n),\frac{n}{k(n)}}) \le Cp\frac{\log_2 n}{n}.$$

2.3. **Isoperimetric inequality on the cube.** We would like to make a connection between Loomis-Whitney inequality and the isoperimetric inequality on the discrete cube. We are going to prove the following proposition.

Proposition 4. Let $A \subseteq \{-1,1\}^n$. Then

$$|E(A,A^c)| \ge 2^n \mu_n(A) \ln\left(\frac{1}{\mu_n(A)}\right).$$

Proof. Fix i and consider 2^{n-1} pairs

$$(x_1,\ldots,x_{i-1},-1,x_{i+1},\ldots,x_n),(x_1,\ldots,x_{i-1},1,x_{i+1},\ldots,x_n).$$

Suppose a is the number of pairs such that both points are not contained in A, b is the number of pair such that both points are contained in A and let c be the number of pairs such that one point is contained in A and the other one is not. We have

$$\mu_n(A) = \frac{b}{2^{n-1}} + \frac{c}{2^n}, \quad I_i = I_i(f_A) = \frac{c}{2^{n-1}}, \quad |P_i(A)| = b + c.$$

Therefore

$$\frac{|P_i(A)|}{2^{n-1}} = \mu_n(A) - \frac{I_i}{2} + I_i = \mu_n(A) + \frac{I_i}{2}, \qquad i = 1,\ldots,n.$$

From the Loomis-Whitney inequality we have

$$\mu_n(A)^{n-1} = \frac{|A|^{n-1}}{2^{n(n-1)}} \le \frac{1}{2^{n(n+1)}}|P_1(A)|\cdot\ldots\cdot|P_n(A)| = \left(\mu_n(A)+\frac{I_1}{2}\right)\ldots\left(\mu_n(A)+\frac{I_n}{2}\right),$$

thus

$$\frac{1}{\mu_n(A)} \le \left(1+\frac{I_1}{2\mu_n(A)}\right)\ldots\left(1+\frac{I_n}{2\mu_n(A)}\right)$$

and therefore

$$\ln\left(\frac{1}{\mu_n(A)}\right) \le \ln\left(1+\frac{I_1}{2\mu_n(A)}\right)+\ldots+\ln\left(1+\frac{I_n}{2\mu_n(A)}\right) \le \frac{I_1+\ldots+I_n}{2\mu_n(A)} = \frac{I(f)}{2\mu_n(A)}.$$

It follows that

$$\frac{|E(A,A^c)|}{2^{n-1}} = I(f) \ge 2\mu_n(A)\ln\left(\frac{1}{\mu_n(A)}\right).$$

□

Recall that on the discrete cube we have a natural graph structure with the set of edges given by $E = \{(x,y) : \ d_H(x,y) = 1\}$, where $d_H(x,y) = |\{i : x_i \ne y_i\}|$. Also, for a set $S \subseteq \{0,1\}^d$ we define its boundary $\partial S = \{(x,y) \in E : x \in S, y \notin S\}$. On $\{0,1\}^d$ we can define the lexicographical order induced by $1 > 0$. Let $L_d[n]$ be the set of first n vertices according to this order.

Theorem 1 (Harper's theorem)**.** We have $|\partial S| \ge |\partial L_d[|S|]|$, i.e., the set of size n minimizing the edge boundary is $L_d[n]$.

Proof. We proceed by induction on d. For $d = 1$ the assertion is trivial. Suppose $d \geq 2$ and the theorem holds for $d - 1$.

Let us first introduce an order on the set of subsets of $\{0,1\}^d$. Each such subset can be identified with a vector in $\{0,1\}^{2^d}$ (since there are 2^d subsets of $\{0,1\}^d$). Here the order of coordinates corresponds to the lexicographical order on $\{0,1\}^d$.

Example 3. For $d = 3$ we have the following order on $\{0,1\}^d$,

$$(000) < (001) < (010) < (011) < (100) < (101) < (110) < (111).$$

Thus, e.g., the vector $(01101001) \in \{0,1\}^{2^3}$ corresponds to the following subset of $\{0,1\}^3$.

$$\{(001), (010), (100), (111)\}.$$

The order $\prec$ on $\{0,1\}^{2^d}$ (and thus the order on subsets of $\{0,1\}^d$) is defined to be the reverse lexicographical order. It is the usual order (where $1 > 0$) but the order of reading the coordinates is reversed.

By the construction we have the following fact.

Fact 6. If $x, y \in \{0,1\}^d$, $y \in T \subseteq \{0,1\}^d$ and $x < y$ then $((T \setminus \{y\}) \cup \{x\}) \prec T$.

We now define the compression of S. Take $T \subseteq \{0,1\}^d$. For every coordinate $i \in [d]$ we can decompose T into two subsets $T_{i=0}, T_{i=1} \subseteq \{0,1\}^{d-1}$ according to the value of ith coordinate. Formally

$$T_{i=\varepsilon} = \{x \in \{0,1\}^{d-1} : (x_1, \dots, x_{i-1}, \varepsilon, x_{i+1}, \dots, x_n) \in T\}, \qquad \varepsilon \in \{0,1\}.$$

Let $C_i(T)$ be the set obtained by replacing $T_{i=0}$ with $L_{d-1}[|T_{i=0}|]$ and $T_{i=1}$ with $L_{d-1}[|T_{i=1}|]$. Of course $|C_i(T)| = |T|$.

Fact 7. We have $|\partial C_i(T)| \leq |\partial T|$.

Proof. Note that

$$\begin{aligned}
|\partial C_i(T)| &= |\partial L_{d-1}[|T_{i=0}|] + |\partial L_{d-1}[|T_{i=1}|] + |L_{d-1}[|T_{i=0}|] \,\Delta\, L_{d-1}[|T_{i=1}|]| \\
&= |\partial L_{d-1}[|T_{i=0}|] + |\partial L_{d-1}[|T_{i=1}|] + ||T_{i=0}| - |T_{i=1}|| \\
&\leq |\partial T_{i=0}| + |\partial T_{i=1}| + |T_{i=0} \,\Delta\, T_{i=1}| = |\partial T|.
\end{aligned}$$

Here the inequalities

$$|\partial L_{d-1}[|T_{i=0}|] \leq |\partial T_{i=0}|, \qquad |\partial L_{d-1}[|T_{i=1}|] \leq |\partial T_{i=1}|$$

follow from the induction assumption and the inequality $||T_{i=0}| - |T_{i=1}|| \leq |T_{i=0} \,\Delta\, T_{i=1}|$ is a general bound $|A \Delta B| \geq ||A| - |B||$ valid for any finite sets A, B. □

We continue the proof of Harper's theorem. From Fact 6 we see that $C_i(T) \prec T$. Let us apply $C_1, \dots, C_n$ in a cyclic fashion,

$$S \to C_1(S) \to C_2C_1(S) \to \dots \to C_dC_{d-1}\dots C_1(S) \to C_1C_dC_{d-1}\dots C_1(S) \to \dots$$

Since in this sequence the (linear) order $\prec$ in non-increasing, we eventually reach a fixed point T of all $C_1, \dots, C_d$.

Let us define a new order $\ll$ on $\{0,1\}^d$ (compressibility order). If all compressed sets containing $y \in \{0,1\}^d$ also contain $x \in \{0,1\}^d$ then we write $x \ll y$. □

Fact 8. We have $x < y$ implies $x \ll y$ unless $x = 01...1$ and $y = 10...0$.

Proof. We first consider the case when $x_i \neq y_i = \varepsilon$ for some $i = 1, \ldots, d$, $\varepsilon \in \{0, 1\}$. Let T be compressed. Suppose $y \in T$ and $x < y$. We are to show that $x \in T$. We have $C_i(T) = T$. Clearly x is in T since $T_{i=\varepsilon} = L_{d-1}[|T_{i=\varepsilon}|]$.

We now consider the case when $x_i \neq y_i$ for all $i = 1, \ldots, d$. Since $x < y$ we get $x_1 = 0$ and $y_1 = 1$. Assume that x, y are not equal to $x = 01...1$ and $y = 10...0$. Thus, there is $i > 1$ such that $x_i = 0$ and $y_i = 1$. Therefore, x, y have the form $x = (0a0b)$ and $y = (1\bar{a}1\bar{b})$, where $\bar{a} = 1 - a$. Take $z = (0a1b)$. We have $x < z$ and $x_1 = z_1$. Thus, from the previous case, $x \ll z$. Moreover, $z < y$ and $z_i = y_i$. Thus, $z \ll y$. We get $x \ll z \ll y$ and therefore $x \ll y$.

Let $L = \{x : x < 01...1\}$ and $R = \{x : x > 10...0\}$. On L and H the orders $<$ and $\ll$ are the same. The only non-comparable points are $x = 01...1$ and $y = 10...0$. To see that they are indeed non-comparable, we take $T = \{(0a) : a \in \{0, 1\}^{d-1}\} \cup (10...0) \setminus (01...1)$. Then T is compressed and contains y but it does not contain x. On the other hand $T = \{0a : a \in \{0, 1\}^{d-1}\}$ is compressed and it contains x but does not contain y. Thus x and y are not comparable in $\ll$.

Take our compressed set T. If $T \cap H \neq \emptyset$ then there is a unique maximal point z in T. Since $z \in T$ we get that $x < z$ implies $x \in T$ for any x. Thus, in this case T is a prefix in $<$.

Let us now assume that $T \cap H = \emptyset$. If $T \cap \{(01...1), (10...0)\} = \emptyset$ then in the same way we get the same conclusion. If $T \cap \{(01...1), (10...0)\} \neq \emptyset$ then we proceed similarly if the cases

$$T \cap \{(01...1), (10...0)\} = \{(01...1), (10...0)\}, \qquad T \cap \{(01...1), (10...0)\} = \{(01...1)\}.$$

The only non-trivial case is $T = L \cup \{(10...0)\}$. In this case we compute the size of edge boundary explicitly,

$$|\partial T| = 2^{d-1} - 2 + 2(d-1) \geq 2^{d-1} = |\partial L_{d-1}[|T|]|.$$

□

3. Harmonic analysis on the hypercube

3.1. **Walsh-Fourier system.** For $S \subset [n]$ consider a function $w_S : \{-1, 1\}^n \to \mathbb{R}$ defined by $w_S(x) = \prod_{i \in S} x_i$. Here we use a convention $w_\emptyset(x) \equiv 1$. Let $\mathbb{E}$ denote the expectation with respect to μ_n. Note that

$$\mathbb{E} w_S = \begin{cases} 0 & S \neq \emptyset \\ 1 & S = \emptyset \end{cases}.$$

Clearly,

$$w_S(x) w_T(x) = \prod_{i \in S} x_i \prod_{j \in T} x_j = \prod_{i \in S \Delta T} x_i \prod_{i \in S \cap T} x_i^2 = \prod_{i \in S \Delta T} x_i = w_{S \Delta T}(x).$$

Since $w_S w_T = w_{S \Delta T}$, we get

$$\mathbb{E} w_S w_T = \begin{cases} 0 & S \neq T \\ 1 & S = T \end{cases}.$$

This means that $(w_S)_{S \subset [n]}$ is an orthonormal system in $L_2(\{-1, 1\}^n, \mu_n)$. Since the dimension of is equal to the number of function w_S (both are equal to 2^n), we get that $(w_S)_{S \subset [n]}$ is an orthonormal basis. It follows that a function $f : \Sigma_n \to \mathbb{R}$ admits an unique expansion

$$f = \sum_{S \subset [n]} \langle f, w_S \rangle \, w_S,$$

where $\langle f, g\rangle = \mathbb{E} fg$. It can be also seen by an elementary argument. Indeed, we have

$$\mathbf{1}_x(y) = \prod_{i=1}^{n} \frac{1 + x_i y_i}{2} = 2^{-n} \sum_{S\subset[n]} w_S(x) w_S(y).$$

Hence,

$$f(x) = \sum_{y\in\Sigma_n} f(y)\mathbf{1}_y(x) = 2^{-n} \sum_{S\subset[n]} \left(\sum_{y\in\Sigma_n} f(y) w_S(y)\right) w_S(x) = \sum_{S\subset[n]} \langle f, w_S\rangle\, w_S(x).$$

The coefficients $a_S = \langle f, w_S\rangle$ are called the **spectrum** of f. Note that we have $\mathbb{E} f = a_\emptyset$ and by orthogonality

$$\mathbb{E} f^2 = \mathbb{E}\left(\sum_S a_S w_S\right)^2 = \sum_{S,T} a_S a_T \mathbb{E} w_S w_T = \sum_S a_S^2.$$

This is the so-called Parseval's identity.

Example 4. Let us prove that $f : \Sigma_n to \mathbb{R}$ satisfies the following Poincaré inequality,

$$\mathrm{Var}_{\mu_n}(f) \le \int_{\Sigma_n} |\nabla f|^2 \mathrm{d}\mu_n.$$

To this end consider the Walsh-Fourier expansion of f, namely $f = \sum_S a_S w_S$. From the Parseval identity we get

$$\mathrm{Var}_{\mu_n}(f) = \mathbb{E} f^2 - (\mathbb{E} f)^2 = \sum_{|S|>0} a_S^2.$$

We now observe that $|\nabla f|^2 = \sum_{i=1}^n |\nabla_i f|^2$. Let us compute the Walsh-Fourier expansion of $\nabla_i f$,

$$(\nabla_i f)(x) = \frac{f(x) - f(\sigma_i(x))}{2} = \sum_{S:i\in S} a_S w_S(x).$$

This is because

$$\nabla_i w_S = \begin{cases} w_S & i \notin S \\ 0 & i \in S \end{cases}.$$

Thus,

$$\int_{\Sigma_n} |\nabla f|^2 \mathrm{d}\mu_n = \sum_{i=1}^n \int_{\Sigma_n} |\nabla_i f|^2 \mathrm{d}\mu_n = \sum_{i=1}^n \sum_{S:i\in S} a_S^2 = \sum_S |S| a_S^2 \ge \sum_{|S|>0} a_S^2 = \mathrm{Var}_{\mu_n}(f).$$

Example 5. It is easy to see that for $f : \{-1,1\}^n \to \{-1,1\}$ the following two conditions are equivalent:

(1) $f(x\cdot y) = f(x)f(y)$, $x, y \in \{-1,1\}^n$,
(2) for some $S \subseteq [n]$ we have $f = w_S$.

Indeed,(2) clearly implies (1). On the other hand, if we assume (1) then we have

$$f(x_1,\ldots,x_n) = \prod_{i=1}^n f(1,\ldots,x_i,\ldots,1).$$

Since $f(1) = f(1\cdot 1) = f(1)^2$ implies $f(1) = 1$ we get that each $f(1,\ldots,x_i,\ldots,1)$ is either identically 1 or is equal to x_i.

Suppose now that we want to consider approximately multiplicative functions. We can define this notion either through point (1) or using (2). The definition (2') reads as follows:

(2') $f : \{-1,1\}^n \to \{-1,1\}$ is ε close to being multiplicative if there is w_S such that $\mathbb{P}_x(f(x) \neq g(x)) \leq \varepsilon$, where x is uniform on $\{-1,1\}^n$.

The definition (1) can be rewritten using the so called Blum-Luby-Rubinfeld test. In BLR test we consider two independent random inputs $x, y \in \Sigma_n$ and accept f if $f(x \cdot y) = f(x)f(y)$. Thus, this test uses only three queries.

(1') We say that f is ε BLR-close to being multiplicative if $\mathbb{P}(f(x \cdot y) = f(x)f(y)) = 1-\varepsilon$, where x, y are independent and uniform in $\{-1,1\}^n$. In other words, BLR test excepts f with probability $1-\varepsilon$.

We show that both definitions are equivalent. First, if f is ε close to certain w_S then BLR test accepts f with probability at least $1-3\varepsilon$,

$$\begin{aligned}\mathbb{P}(f(x\cdot y) \neq f(x)f(y)) &\leq \mathbb{P}(f(x) \neq w_S(x) \text{ or } f(x) \neq w_S(y) \text{ or } f(x\cdot y) \neq w_S(x\cdot y)) \\ &\leq \mathbb{P}(f(x) \neq w_S(x)) + \mathbb{P}(f(y) \neq w_S(y)) + \mathbb{P}(f(x\cdot y) \neq w_S(x\cdot y)) \\ &= 3\mathbb{P}(f(x) \neq w_S(x)) \leq 3\varepsilon.\end{aligned}$$

What is non-trivial is that we have the reverse implication.

Fact 9. If BLR test accepts f with probability $1-\varepsilon$ then f is ε close to certain w_S.

Proof. Take $f : \{-1,1\}^n \to \{-1,1\}$. Let $h(x) = \mathbb{E}_y f(y)f(x\cdot y)$. If $f = \sum_S a_S w_S$ then

$$h(x) = \mathbb{E}_y \left(\sum_S a_S w_S(y)\right)\left(\sum_S a_S w_S(x) w_S(y)\right) =$$
$$\sum_{S,T} a_S a_T w_S(x) \mathbb{E}_y w_S(y) w_T(y) = \sum_S a_S^2 w_S(x).$$

using orthogonality of the Walsh system. We have

$$\frac{1}{2} + \frac{1}{2} f(x)f(y)f(x\cdot y) = \left\{\begin{array}{ll} 1 & f(x)f(y) = f(x\cdot y) \\ 0 & f(x)f(y) \neq f(x\cdot y)\end{array}\right. .$$

Thus,

$$1-\varepsilon = \mathbb{E}\left(\frac{1}{2} + \frac{1}{2} f(x)f(y)f(x\cdot y)\right) = \frac{1}{2} + \frac{1}{2}\mathbb{E}_x f(x) \mathbb{E}_y f(y) f(x\cdot y) = \frac{1}{2} + \frac{1}{2}\mathbb{E}_x f(x)h(x).$$

We get

$$1-2\varepsilon = \mathbb{E}_x f(x)h(x) = \sum_S a_S^3 \leq (\max_S a_S) \sum_S a_S^2 = \max_S a_S.$$

Therefore, there exists w_S such that $1-2\varepsilon \leq \mathbb{E} f w_S = 1 - 2\mathbb{P}_x(f(x) \neq w_S(x))$. Thus, f is ε close to w_S. □

3.2. **Noise semigroup on the cube.** We now compute the action of our semigroup $P_t(f) = e^{tL} f$ on the Walsh functions w_S. We have $L = K - I$ and thus

$$\begin{aligned}(Lw_S)(x) &= (Kw_S)(x) - w_S(x) = \frac{1}{n}\sum_i w_S(\sigma_i(x)) - w_S(x) \\ &= \frac{1}{n}(-|S| w_S(x) + (n-|S|) w_S(x)) - w_S(x) = -2\frac{|S|}{n} w_S(x).\end{aligned}$$

This gives $e^{tL}w_S = e^{-2t\frac{|S|}{n}}w_S$. Thus,

$$P_t\left(\sum_S a_S w_S\right) = \sum_S a_S e^{-2t\frac{|S|}{n}}w_S.$$

To simplify notation in what follows we rescale our operator P_t and define

$$\mathcal{P}_t(f) = P_{nt/2}(f) = \sum_S a_S e^{-t|S|}w_S.$$

The new generator $\mathcal{L}f = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}_t(f)\Big|_{t=0} = \frac{n}{2}Lf$. Therefore the inequality discrete LSI

$$\mathrm{Ent}_{\mu_n}(f^2) \leq 2 \cdot \frac{n}{2}\langle(-Lf), f\rangle$$

now reads

$$\mathrm{Ent}_{\mu_n}(f^2) \leq 2\langle(-\mathcal{L}f), f\rangle = 2\mathcal{E}_{\mathcal{L}}(f, f).$$

3.3. **Arrow's theorem.** Suppose we have three candidates a, b, c and we want to elect one using some voting procedure. Assume we have n voters and each voter has his own ranking of candidates. In other words for each pair (a, b), (b, c), (c, d) a voter gives a number in $\{-1, 1\}$, with 1 meaning that he prefers the first candidate. Thus, each voter V_i delivers a triple $(x_i, y_i, z_i) \in \{-1, 1\}^3$. Note that only six triples are allowed. Indeed, the triples $(1, 1, 1)$ and $(-1, -1, -1)$ are not allowed because a voter can not prefer a than b, b than c and c than a (nor the opposite cycle). So, for each voter we have the following allowed rankings

$$(-1, -1, 1), (-1, 1, -1), (-1, 1, 1), (1, -1, -1), (1, -1, 1), (1, 1, -1).$$

Now suppose we use some function $f : \{-1, 1\}^n \to \{-1, 1\}$ to decide whether the society prefers a than b, etc. by considering $f(x) = f(x_1, \ldots, x_n)$, $f(y) = f(y_1, \ldots, y_n)$ and $f(z) = f(z_1, \ldots, z_n)$. For example $f(x_1, \ldots, x_n) = 1$ means that the society prefers a than b. In other words, w consider all three pairwise elections.

We say that there is a Condorcet winner if there is a candidate who wins all the pairwise elections he participated in. So, there is a Condorcet winner if

$$(f(x), f(y), f(z)) \in \{(-1, -1, 1), (-1, 1, -1), (-1, 1, 1), (1, -1, -1), (1, -1, 1), (1, 1, -1)\}.$$

Here is an example of a voting with Condorcet winner.

	V_1	V_2	V_3	f
$a(+)$ vs. $b(-)$	$+$	$+$	$-$	$+$
$b(+)$ vs. $c(-)$	$-$	$+$	$-$	$-$
$c(+)$ vs. $a(-)$	$+$	$-$	$-$	$-$

Table 1. Voting with $n = 3$ voters using $f(x) = \mathrm{sgn}(x_1 + x_2 + x_3)$. Here we get the ranking $(1, -1, -1)$ which means $c > a > b$ and thus c is the winner.

However, the following voting shows that there may not be a Condorcet winner. This is called the Condorcet paradox.

We show that essentially the only voting scheme free from the Condorcet paradox is dictatorship.

	V_1	V_2	V_3	f
$a(+)$ vs. $b(-)$	$+$	$+$	$-$	$+$
$b(+)$ vs. $c(-)$	$+$	$-$	$+$	$+$
$c(+)$ vs. $a(-)$	$-$	$+$	$+$	$+$

Table 2. Voting with $n = 3$ voters using $f(x) = \mathrm{sgn}(x_1 + x_2 + x_3)$. Here we get the ranking $(1, 1, 1)$ which means $a > b$, $b > c$ and $c > a$ and thus we cannot choose a winner.

Theorem 2 (Arrow's Theorem)**.** Let $f : \{-1, 1\}^n \to \{-1, 1\}$ be unanimous (i.e., $f(1) = 1$ and $f(-1) = -1$) voting rule used in three candidate Condorcet elections. If there is always a Condorcet winner, then $f(x) = x_k$ for some $k \in [n]$.

Proof. Let us do a random election. Each voter chooses one of the 6 possible rankings uniformly at random. We compute the probability of Condorcet winner. For this we need a function $\sigma : \{-1, 1\}^3 \to \{0, 1\}$ which is equal to 1 if and only if the argument (x, y, z) does not belong to the set $\{(-1, -1, -1), (1, 1, 1)\}$. It is easy to see that

$$\sigma(x, y, z) = \frac{3}{4} - \frac{1}{4}(xy + yz + zx).$$

Thus,

$$\begin{aligned}\mathbb{P}(\exists \text{ Condorcet winner}) &= \mathbb{E}\sigma(f(x), f(y), f(z)) \\ &= \frac{3}{4} - \frac{1}{4}\mathbb{E}[f(x)f(y) + f(y)f(z) + f(z)f(x)] = \frac{3}{4} - \frac{3}{4}\mathbb{E}[f(x)f(y)].\end{aligned}$$

Recall that (x_i, y_i), $i = 1, \ldots, n$ are independent. Moreover, the distribution of each (x_i, y_i, z_i) is uniform over all 6 admissible rankings. Therefore, it is easy to see that $\mathbb{E}x_i = \mathbb{E}y_i = 0$ and $\mathbb{E}x_i y_i = -\frac{1}{3}$. Let $f = \sum_S a_S w_S$. We get

$$\begin{aligned}\mathbb{E}[f(x)f(y)] &= \sum_{S,T} a_S a_T \mathbb{E}[w_S(x)w_T(y)] = \sum_S a_S^2 \mathbb{E}[w_S(x)w_S(y)] \\ &= \sum_S a_S^2 (\mathbb{E}[x_1 y_1])^{|S|} = \sum_S a_S^2 \left(-1/3\right)^{|S|}.\end{aligned}$$

We arrive at

$$\mathbb{P}(\exists \text{ Condorcet winner}) = \frac{3}{4} - \frac{3}{4}\sum_S a_S^2 \left(-1/3\right)^{|S|}.$$

Let $W_k[f] = \sum_{|S|=k} a_S^2$. We have

$$\begin{aligned}\frac{3}{4} - \frac{3}{4}\sum_S a_S^2 \left(-1/3\right)^{|S|} &= \frac{3}{4} - \frac{3}{4}\sum_{k=0}^{n} W_k[f]\left(-1/3\right)^k \le \frac{3}{4} - \frac{3}{4}\sum_k W_{2k+1}[f]\left(-1/3\right)^{2k+1} \\ &= \frac{3}{4} + \frac{3}{4}\sum_k W_{2k+1}[f]\left(1/3\right)^{2k+1} \le \frac{3}{4} + \frac{3}{4}\left(\frac{1}{3}W_1[f] + \frac{1}{27}\sum_{k>0} W_{2k+1}[f]\right) \\ &\le \frac{3}{4} + \frac{3}{4}\left(\frac{1}{3}W_1[f] + \frac{1}{27}(1 - W_1[f])\right) = \frac{7}{9} + \frac{2}{9}W_1[f] = \frac{7}{9} + \frac{2}{9}\sum_{k=1}^{n} a_{\{k\}}^2.\end{aligned}$$

Thus,

$$\mathbb{P}(\exists \text{ Condorcet winner}) \leq \frac{7}{9} + \frac{2}{9} \sum_{k=1}^{n} a_{\{k\}}^2.$$

The quantity $\sum_{k=1}^n a_{\{k\}}^2 \leq \sum_S a_S^2 = 1$ can be equal to 1 only if $f(x) = \sum_{k=1}^n a_{\{k\}} x_k$. Taking $x_i = \mathrm{sgn}(a_i)$ we get $\sum_k |a_{\{k\}}| = 1$. Together with $\sum_{k=1}^n a_{\{k\}}^2 = 1$ this gives the existence of l such that $|a_{\{l\}}| = 1$ and $a_{\{k\}} = 0$ for all $k \neq l$. Thus $\mathbb{P}(\exists$ Condorcet winner) implies that f is a dictator.

□

4. Hypercontractivity

4.1. **Uniform convexity in L_p.** For a given normed space $(V, \|\cdot\|)$ and $\varepsilon > 0$ let us define the quantity

$$\delta_V(\varepsilon) = \inf \left\{ 1 - \left\| \frac{u+v}{2} \right\| : \ \|u\| = \|v\| = 1, \ \|u - v\| \geq 2\varepsilon \right\}.$$

Our goal in to lower bound δ_V for L_p with $1 < p \leq 2$. First, note that the case of L_2 is easy. Indeed for $f, g \in L_2$ we have the parallelogram identity

$$\left\| \frac{f+g}{2} \right\|_2^2 + \left\| \frac{f-g}{2} \right\|_2^2 = \frac{\|f\|_2^2 + \|g\|_2^2}{2}.$$

If $\|f\|_2 = \|g\|_2 = 1$, we get (by using $\sqrt{1-x} \leq 1 - \frac{1}{2}x$, $x \leq -1$)

$$\left\| \frac{f+g}{2} \right\|_2 = \left(1 - \left\| \frac{f-g}{2} \right\|_2^2 \right)^{1/2} \leq 1 - \frac{1}{2} \left\| \frac{f-g}{2} \right\|_2^2.$$

Thus, $\delta_{L^2}(\varepsilon) \geq \frac{1}{2}\varepsilon^2$.

Let us now consider a more general, but still simple, case $p \geq 2$. For numbers $x, y \geq 0$ we have

$$(x^p + y^p)^{1/p} \leq (x^2 + y^2)^{1/2}, \qquad \left(\frac{a^2 + b^2}{2} \right)^{1/2} \leq \left(\frac{a^p + b^p}{2} \right)^{1/p}.$$

Thus, for all a, b we get

$$\left(\left| \frac{a+b}{2} \right|^p + \left| \frac{a-b}{2} \right|^p \right)^{1/p} \leq \left(\left| \frac{a+b}{2} \right|^2 + \left| \frac{a-b}{2} \right|^2 \right)^{1/2} = \left(\frac{a^2+b^2}{2} \right)^{1/2} \leq \left(\frac{|a|^p + |b|^p}{2} \right)^{1/p}.$$

We get

$$\left| \frac{a+b}{2} \right|^p + \left| \frac{a-b}{2} \right|^p \leq \frac{|a|^p + |b|^p}{2}.$$

Taking $a = f(x)$, $b = g(x)$ and integrating yields

$$\left\| \frac{f+g}{2} \right\|_p^p + \left\| \frac{f-g}{2} \right\|_p^p \leq \frac{\|f\|_p^p + \|g\|_p^p}{2}.$$

Again, if $\|f\|_p = \|g\|_p = 1$, we get (by using Bernoulli inequality $(1-x)^{1/p} \leq 1 - x/p$, $x \leq -1$)

$$\left\| \frac{f+g}{2} \right\|_p = \left(1 - \left\| \frac{f-g}{2} \right\|_p^p \right)^{1/p} \leq 1 - \frac{1}{p} \left\| \frac{f-g}{2} \right\|_p^p.$$

This yields $\delta_{L^p}(\varepsilon) \geq \frac{p-1}{2}\varepsilon^2$.

We now prove the following theorem.

Theorem 3. Let $1 < p \leq 2$. Then for every $f, g \in L_p$ we have

$$\left\|\frac{f+g}{2}\right\|_p^2 + (p-1)\left\|\frac{f-g}{2}\right\|_p^2 \leq \frac{\|f\|_p^2 + \|g\|_p^2}{2}.$$

In particular, $\delta_{L^p}(\varepsilon) \geq \frac{p-1}{2}\varepsilon^2$.

Proof. We will prove the complex case. It is enough to consider only step functions of the form

$$f = \sum_j z_j \mathbf{1}_{A_j}, \qquad g = \sum_j w_j \mathbf{1}_{A_j}.$$

Then

$$f + tg = \sum_j (z_j + tw_j)\mathbf{1}_{A_j}.$$

Moreover, we can assume that $z_j + tw_j \neq 0$ for all real t, my imposing the condition $z_j\bar{w}_j \notin \mathbb{R}$. As a consequence $f(x) + tg(x) \neq 0$ and we avoid problems with differentiating in the next step.

Consider the function $Y(t) = \|f + tg\|_p^p$ and let $q = p/2$. We have $\|f + tg\|_p 2 = Y(t)^{2/p} = Y(t)^{1/q}$. Thus,

$$\frac{\mathrm{d}}{\mathrm{d}t^2}\|f + tg\|_p^2 = \frac{1}{q}\left(\frac{1}{q} - 1\right)Y(t)^{\frac{1}{q}-2}(Y')^2 + \frac{1}{q}Y^{\frac{1}{q}-1}Y'' \geq \frac{1}{q}Y^{\frac{1}{q}-1}Y''.$$

Now, our goal is to show that

$$Y''(t) \geq p(p-1)\int |f + tg|^{p-2}|g|^2 \mathrm{d}\mu. \tag{1}$$

It is enough to show that for every complex numbers $a, b \in \mathbb{C}$, such that $a + tb \neq 0$, $t \in \mathbb{R}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t^2}|a + tb|^p \geq p(p-1)|a + tb|^{p-2}|b|^2.$$

Let $a = a_1 + ia_2$, $b - b_1 + ib_2$. Then $|a + tb|^2 = (a_1 + tb_1)^2 + (a_2 + tb_2)^2$. Moreover,

$$\frac{\mathrm{d}}{\mathrm{d}t}|a + tb|^2 = 2\left[(a_1 + tb_1)b_1 + (a_2 + tb_2)b_2\right], \qquad \frac{\mathrm{d}}{\mathrm{d}t^2}|a + tb|^2 = 2|b|^2.$$

We get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t^2}|a + tb|^p &= \frac{\mathrm{d}}{\mathrm{d}t^2}(|a + tb|^2)^{\frac{p}{2}} \\
&= \left(\frac{p}{2} - 1\right)\frac{p}{2}(|a + tb|^2)^{\frac{p}{2}-2} \cdot 4\left[(a_1 + tb_1)b_1 + (a_2 + tb_2)b_2\right]^2 + \frac{p}{2}(|a + tb|^2)^{\frac{p}{2}-1}2|b|^2 \\
&= p(p-2)|a + tb|^{p-4}\left[(a_1 + tb_1)b_1 + (a_2 + tb_2)b_2\right]^2 + p|a + tb|^{p-2}|b|^2.
\end{aligned}$$

Note that by Cauchy-Schwarz

$$\left[(a_1 + tb_1)b_1 + (a_2 + tb_2)b_2\right]^2 \leq |a + tb|^2|b|^2.$$

This, together with the fact that $p - 2 \leq 0$, yields

$$\frac{\mathrm{d}}{\mathrm{d}t^2}|a + tb|^p \geq \left[p(p-2) + p\right]|a + tb|^{p-2}|b|^2 = p(p-1)|a + tb|^{p-2}|b|^2.$$

We arrive at (1). Note that for u, v we have the reverse Hölder inequality,

$$\int |uv|\mathrm{d}\mu \geq \left(\int |u|^r\right)^{1/r}\left(\int |v|^s\right)^{1/s}, \qquad \frac{1}{s}+\frac{1}{r}=1, \quad 0<r\leq 1.$$

We use it with $r=q$, $s=\frac{q}{q-1}=\frac{p}{p-2}$, $u=|g|^2$ and $v=|f+tg|^{2q-2}$,

$$Y''(t)\geq p(p-1)\left(\int |f+tg|^p\mathrm{d}\mu\right)^{1-\frac{1}{q}}\left(\int |g|^p\mathrm{d}\mu\right)^{\frac{2}{p}}=p(p-1)Y(t)^{1-\frac{1}{q}}\left(\int |g|^p\mathrm{d}\mu\right)^{\frac{2}{p}}$$

$$\frac{\mathrm{d}}{\mathrm{d}t^2}\left\|f+tg\right\|_p^2\geq \frac{1}{q}Y^{\frac{1}{q}-1}Y''\geq \frac{1}{q}Y^{\frac{1}{q}-1}\cdot p(p-1)Y(t)^{1-\frac{1}{q}}\left\|g\right\|_p^2=2(p-1)\left\|g\right\|_p^2.$$

Let $\psi(t)=\left\|f+tg\right\|_p^2$ and take $c=(p-1)\left\|g\right\|_p^2$. Then $\psi''(t)\geq 2c$ and thus the function $\varphi(t)=\psi(t)+ct(1-t)$ is convex. This gives $\varphi(1/2)\leq\frac{1}{2}(\varphi(0)+\varphi(1))$, or equivalently

$$\psi(1/2)+\frac{c}{4}\leq\frac{\psi(0)+\psi(1)}{2}.$$

The latter is

$$\left\|f+\frac{g}{2}\right\|_p^2+\frac{p-1}{4}\left\|g\right\|_p^2\leq\frac{\left\|f\right\|_p^2+\left\|f+g\right\|_p^2}{2}.$$

Taking $f=u$ and $g=v-u$ yields

$$\left\|\frac{u+v}{2}\right\|_p^2+(p-1)\left\|\frac{u-v}{2}\right\|_p^2\leq\frac{\left\|u\right\|_p^2+\left\|v\right\|_p^2}{2}.$$

□

4.2. **Hölder and Pinsker inequalities.** Let us show one particular application of Theorem 3 proved in the previous section.

Theorem 4 (Hölder inequality with reminder)**.** Let $1<p\leq 2$ and define q through the relation $\frac{1}{p}+\frac{1}{q}=1$. Assume that $\left\|f\right\|_q=\left\|g\right\|_p=1$. Let θ be such that $e^{i\theta}\int fg\mathrm{d}\mu$ is positive. Then

$$\left|\int fg\mathrm{d}\mu\right|\leq 1-\frac{p-1}{4}\left\|\mathcal{D}_q(f)-e^{i\theta}g\right\|_p^2,$$

where

$$\mathcal{D}_q(f)=\left\|f\right\|_q^{1-q}|f|^{q-2}\overline{f(x)}.$$

Proof. Note that $\int\mathcal{D}_q(f)f\mathrm{d}\mu=\left\|f\right\|_q=1$. Thus

$$1+\left|\int fg\mathrm{d}\mu\right|=1+e^{i\theta}\int fg\mathrm{d}\mu=\int f\left(\mathcal{D}_q(f)+e^{i\theta}g\right)\mathrm{d}\mu\leq\left\|\mathcal{D}_q(f)+e^{i\theta}g\right\|_p.$$

Using the fact that $\left\|\mathcal{D}_q(f)\right\|_p=1$, we get, by strong convexity,

$$\frac{1}{2}+\frac{1}{2}\left|\int fg\mathrm{d}\mu\right|\leq\left\|\frac{\mathcal{D}_q(f)+e^{i\theta}g}{2}\right\|_p\leq 1-\frac{p-1}{2}\left\|\frac{\mathcal{D}_q(f)-e^{i\theta}g}{2}\right\|_p^2$$

Rewriting gives the desired inequality. □

Example 6. Let us consider probability densities ρ, σ. Take $f = \rho^{1/q}$ and $g = \sigma^{1/p}$ with $1/p + 1/q = 1$ and $1 < p \leq 2$. We have $\|f\|_q = \|g\|_p = 1$. Moreover, $\mathcal{D}_q(f) = f^{q-1} = f^{\frac{1}{p-1}} = \rho^{1/p}$. We get

$$\int \rho^{1-\frac{1}{p}} \sigma^{\frac{1}{p}} \mathrm{d}\mu \leq 1 - \frac{p-1}{4} \left\| \rho^{\frac{1}{p}} - \sigma^{\frac{1}{p}} \right\|_p^2 .$$

This is equivalent to

$$\frac{p-1}{4} \left\| \rho^{\frac{1}{p}} - \sigma^{\frac{1}{p}} \right\|_p^2 \leq \int \left(\sigma - \rho^{1-\frac{1}{p}} \sigma^{\frac{1}{p}} \right) \mathrm{d}\mu = \int \sigma \left(1 - (\rho/\sigma)^{1-\frac{1}{p}} \right) \mathrm{d}\mu,$$

which is

$$\frac{p}{4} \left\| \rho^{\frac{1}{p}} - \sigma^{\frac{1}{p}} \right\|_p^2 \leq \frac{1}{1-\frac{1}{p}} \int \sigma \left(1 - (\rho/\sigma)^{1-\frac{1}{p}} \right) \mathrm{d}\mu.$$

Taking $p \to 1^+$ we get

$$\frac{1}{4} \|\rho - \sigma\|_1^2 \leq -\frac{\mathrm{d}}{\mathrm{d}\varepsilon} \int \sigma (\rho/\sigma)^{\varepsilon} \mathrm{d}\mu = -\int \sigma \ln(\rho/\sigma) = \int \sigma \ln(\sigma/\rho) = D(\sigma\|\rho).$$

This is the so-called Pinsker inequality

$$\frac{1}{4} \|\rho - \sigma\|_1^2 \leq D(\sigma\|\rho).$$

In fact the optimal constant is $1/2$, not $1/4$. We leave this improvement as an exercise.

4.3. **Gross's two-point inequality.** If we take $u = f + g$ and $v = f - g$ we get an equivalent form of the inequality from Theorem 3,

$$\|u\|_p^2 + (p-1) \|v\|_p^2 \leq \frac{\|u+v\|_p^2 + \|u-v\|_p^2}{2}.$$

We need the following strengthening of this inequality.

Theorem 5. Let $1 < p \leq 2$. Then for every $f, g \in L_p$ we have

$$\|f\|_p^2 + (p-1) \|g\|_p^2 \leq \left(\frac{\|f+g\|_p^p + \|f-g\|_p^p}{2} \right)^{\frac{2}{p}} .$$

Proof. We use Theorem 3 on $(\Omega \times \{-1, 1\}, \mu \otimes \mu_1)$, where $\mu_1 = \frac{1}{2}\delta_{-1} + \frac{1}{2}\delta_1$ is the symmetric Bernoulli measure. Let $\tilde{f}(x, y) = f(x)$ and $\tilde{g}(x, y) = yg(x)$. We get

$$\begin{aligned} \left\| \tilde{f} \pm \tilde{g} \right\|_p^p &= \int |f(x) \pm yg(x)|^p \mathrm{d}\mu(x) \mathrm{d}\mu_1(y) = \frac{1}{2} \int |f+g|^p \mathrm{d}\mu + \frac{1}{2} \int |f-g|^p \mathrm{d}\mu \\ &= \frac{1}{2} \|f+g\|_p + \frac{1}{2} \|f-g\|_p . \end{aligned}$$

Moreover, $\|\tilde{f}\|_p = \|f\|_p$ and $\|\tilde{g}\|_p = \|g\|_p$. Thus,

$$\begin{aligned} \|f\|_p^2 + (p-1) \|g\|_p^2 = \|\tilde{f}\|_p^2 + (p-1) \|\tilde{g}\|_p^2 &\leq \frac{\|\tilde{f}+\tilde{g}\|_p^2 + \|\tilde{f}-\tilde{g}\|_p^2}{2} = \|\tilde{f}+\tilde{g}\|_p^2 \\ &= \left(\frac{\|f+g\|_p^p + \|f-g\|_p^p}{2} \right)^{2/p} . \end{aligned}$$

$\square$

If we restric the above inequality to two point space $\{-1,1\}$ and take $f(x)=a$, $g(x)=bx$, we get the so-called two-point Gross's inequality

$$(a^2+(p-1)b^2)^{1/2} \le \left(\frac{|a+b|^p+|a-b|^p}{2}\right)^{1/p}. \tag{2}$$

4.4. Gross's hypercontractivity.

Theorem 6. Let $1<p\le 2$. Then

$$e^{-t} \le \sqrt{p-1} \qquad \Longrightarrow \qquad \|\mathcal{P}_t h\|_2 \le \|h\|_p .$$

More generally, if $1<p<q<\infty$ then

$$e^{-t} \le \sqrt{\frac{p-1}{q-1}} \qquad \Longrightarrow \qquad \|\mathcal{P}_t h\|_q \le \|h\|_p .$$

We now prove only the first part.

Proof. For $n=1$ we have $h(x)=a+bx$ Thus, $h=f+g$, where $f(x)=a$ and $g(x)=bx$. We have $\mathcal{P}_t(h)=a+e^{-t}xb$. Clearly, we have

$$\|h\|_p^p = \frac{\|f+g\|_p^p+\|f-g\|_p^p}{2}.$$

Moreover,

$$\left\|\mathcal{P}_{-\ln\sqrt{p-1}}\right\|_2^2 = a^2+(e^{\ln\sqrt{p-1}})b^2 = a^2+(p-1)b^2 = \|f\|_p^2+(p-1)\|g\|_p^2 .$$

Thus, in this case $\|\mathcal{P}_t h\|_2 \le \|h\|_p$ is equivalent to the assertion of Theorem 5.

Let us not provide an induction step. Let us consider $h:\{-1,1\}^n\to\mathbb{R}$. There is a unique decomposition $h=f+x_n g$. Note that $\mathcal{P}_t h = \mathcal{P}_t f + e^{-t}x_n\mathcal{P}_t g$. Let $e^{-t}=p-1$, $\tilde{f}=\mathcal{P}_t f$ and $\tilde{g}=x_n\mathcal{P}_t g$. Then by Theorem 5 we get

$$\begin{aligned}\|\mathcal{P}_t h\|_2^2 &= \|\mathcal{P}_t f\|_2^2+(p-1)\|\mathcal{P}_t g\|_2^2 \le \|f\|_p^2+(p-1)\|g\|_p^2\\ &\le \left(\frac{\|f+g\|_p^p+\|f-g\|_p^p}{2}\right)^{\frac{2}{p}} = \|h\|_p^2 .\end{aligned}$$

Thus, $\|\mathcal{P}_t h\|_2 \le \|h\|_p$. ∎

4.5. Kahn-Kalai-Linial theorem.

We first prove the following theorem due to Talagrand.

Theorem 7. Let $f:\{-1,1\}^n\to\{-1,1\}$ and let $\mu(f)=\mathbb{P}(f=1)$. Then

$$\sum_{i=1}^n \frac{I_i(f)}{\log\left(\frac{1}{I_i(f)}\right)} \ge \frac{4}{15}\mu(f)(1-\mu(f)).$$

We adopt the notation $\frac{0}{\log(1/0)}=0$ and $1/\log(1)=+\infty$. We begin with a lemma.

Lemma 1. Let $g:\{-1,1\}^n\to\mathbb{R}$ with $\|g\|_{3/2}\ne\|g\|_2$, which is equivalent to $|g|$ being not constant. Then

$$\sum_{S\ne\emptyset}\frac{\hat{g}(S)^2}{|S|} \le \frac{5}{2}\frac{\|g\|_2^2}{\log\left(\|g\|_2/\|g\|_{3/2}\right)}.$$

Proof. Using the inequality
$$\|T_\delta g\|_2 \le \|g\|_{1+\delta^2}$$
with $\delta^2 = 1/2$ we obtain
$$\sum_{S:\ |S|=k} \hat{g}(S)^2 \le 2^k \sum_S \frac{1}{2^{|S|}}\hat{g}(S)^2 = 2^k \left\| T_{\sqrt{1/2}}g \right\|_2^2 \le 2^k \|g\|_{3/2}^2 .$$
Now take $m \ge 0$. We have
$$\sum_{S\neq\emptyset} \frac{\hat{g}(S)^2}{|S|} = \sum_{k=1}^m \sum_{S:\ |S|=k} \frac{\hat{g}(S)^2}{k} + \sum_{S:\ |S|>m} \frac{\hat{g}(S)^2}{|S|} \le \sum_{k=1}^m \frac{2^k \|g\|_{3/2}^2}{k} + \sum_{S:\ |S|>m} \frac{\hat{g}(S)^2}{m+1}$$
$$\le \frac{4\cdot 2^m \|g\|_{3/2}^2 + \|g\|_2^2}{m+1},$$
where we have used the inequality
$$\sum_{k=1}^m \frac{2^k}{k} \le \frac{4\cdot 2^m}{m+1},$$
which can be easily proved by induction.

Now we take
$$m = \max\{m \ge 0 \mid 2^m \|g\|_{3/2}^2 \le \|g\|_2^2\}.$$
Then $2^{m+1}\|g\|_{3/2}^2 > \|g\|_2^2$. Hence,
$$m+1 > 2\log\left(\frac{\|g\|_2}{\|g\|_{3/2}}\right).$$
We arrive at
$$\sum_{S\neq\emptyset} \frac{\hat{g}(S)^2}{|S|} \le \frac{5\|g\|_2^2}{m+1} \le \frac{5}{2}\frac{\|g\|_2^2}{\log\left(\|g\|_2 / \|g\|_{3/2}\right)}.$$
$\square$

Proof of Talagrand's theorem. Suppose $I_i(f) \in (0,1)$. Let $g(x) = f(x) - f(x^i)$. It follows that $|g|$ is not constant. We have
$$\frac{\|g\|_2}{\|g\|_{3/2}} = \frac{2I_i(f)^{1/2}}{2I_i(f)^{2/3}} = I_i(f)^{-1/6}.$$
From the lemma we obtain
$$\sum_{S:\ i\in S} \frac{4\hat{f}(S)^2}{|S|} = \sum_S \frac{\hat{g}(S)^2}{|S|} \le \frac{5}{2}\frac{\|g\|_2^2}{\log\left(\|g\|_2 / \|g\|_{3/2}\right)} = \frac{5}{2}\cdot\frac{4I_i(f)}{\log(I_i(f)^{-1/6})} = 60\frac{I_i(f)}{\log(\frac{1}{I_i(f)})}.$$
The inequality
$$\sum_{S:\ i\in S} \frac{4\hat{f}(S)^2}{|S|} \le 60\frac{I_i(f)}{\log(\frac{1}{I_i(f)})}$$
is also true when $I_i(f) \in \{0,1\}$. We obtain
$$16\mu(f)(1-\mu(f)) = 4\operatorname{Var}_\mu(f) = \sum_{Sn\varepsilon\emptyset} 4\hat{f}(S)^2 = \sum_{i=1}^n \sum_{S:\ i\in S} \frac{4\hat{f}(S)^2}{|S|} \le 60\sum_{i=1}^n \frac{I_i(f)}{\log(\frac{1}{I_i(f)})}.$$

The assertion follows. □

We are ready to give state and prove the following celebrated theorem of Kahn, Kalai and Linial.

Theorem 8 (KKL theorem)**.** Let $f : \{-1,1\}^n \to \{-1,1\}$ be a Boolean function. Then

$$\max_i I_i(f) \geq \frac{4}{15}\mu(f)(1-\mu(f))\frac{\log n}{n}.$$

Proof. We show that Talagrand result implies KKL Theorem. Let us first observe that if $a \in (0,1)$ and $\frac{a}{\log(1/a)} \geq c > 0$ then $a \geq \frac{1}{2}c\log(1/c)$. Since $(0,1) \ni a \mapsto \frac{a}{\log(1/a)}$ is increasing, it suffices to assume that $\frac{a}{\log(1/a)} = c$. Then we are to prove

$$a \geq \frac{1}{2}\frac{a}{\log(1/a)}\log\left(\frac{1}{a}\log\left(\frac{1}{a}\right)\right).$$

Taking $x = 1/a \geq 1$ we see that this inequality is equivalent to

$$\log(x) \geq \frac{1}{2}\log(x\log(x)) = \frac{1}{2}\log x + \frac{1}{2}\log\log x.$$

Thus we are to prove $x \geq \log x$. It follows from Bernoulli inequality

$$2^x = (1+1)^x \geq 1 + x \geq x.$$

From Talagrand's inequality we know that there exists i such that

$$\frac{I_i(f)}{\log\left(\frac{1}{I_i(f)}\right)} \geq \frac{1}{n}\cdot\frac{4}{15}\mu(f)(1-\mu(f)).$$

Now take

$$a = I_i(f), \qquad c = \frac{1}{n}\cdot\frac{4}{15}\mu(f)(1-\mu(f)).$$

We have

$$\frac{1}{c} = n\cdot\frac{15}{4}\frac{1}{\mu(f)(1-\mu(f))} \geq 15n.$$

We obtain

$$I_i(f) \geq \frac{1}{2}c\log(1/c) \geq \frac{1}{n}\cdot\frac{4}{15}\mu(f)(1-\mu(f))\log(15n) \geq \frac{4}{15}\mu(f)(1-\mu(f))\frac{\log n}{n}.$$

□

5. Finite space Markov chains

5.1. **Discrete time Markov chains.** Consider a finite set V with $|V| = n$ and a Markov kernel (or transition matrix) $K : V \times V \to \mathbb{R}$, i.e.,

$$K(x,y) \geq 0,\ x,y \in V \qquad \sum_{y\in V} K(x,y) = 1,\ x \in V.$$

The discrete time Markov chain associated with K with an initial distribution ν is a V-valued sequence $(X_n)_{n=0}^{\infty}$ whose law $\mathbb{P}_\nu$ is given by

$$\mathbb{P}_\nu(V_i = v_i,\ 0 \leq i \leq l) = \nu(x_0)K(x_0,x_1)\cdot\ldots\cdot K(x_{l-1},x_l), \qquad l = 0,1,\ldots$$

Consider the Markov chain started at x and set $\mathbb{P}_x = \mathbb{P}_{\delta_x}$. Then the law of X_l is given by $\mathbb{P}_x(X_l = y) = K^l(x,y)$, where K^l is defined recursively via

$$K^l(x,y) = \sum_{z\in V} K^{l-1}(x,z)K(z,y).$$

The kernel K defines an operator

$$(Kf)(x) = \sum_{y\in V} K(x,y)f(y).$$

Clearly, the lth power of this operator has kernel $K^l(x,y)$.

5.2. **Continuous time Markov chains.** In the continuous time Markov chain associated with K (and starting from x) the moves are those of the discrete time Markov chain, however the jumps occur after independent Exp(1) waiting times. Thus, the number of jumps after time t is given by the Poisson process. Therefore, the probability that there have been exactly i jumps until time t is equal to $e^{-t}t^i/i!$. It follows that the probability to be at point y after i jumps is equal to $e^{-t}t^i/i!K^i(x,y)$. Let $P_t(x,y) = P_t^x(y) = \mathbb{P}_x(X_t = y)$ We get

$$P_t(x,y) = e^{-t}\sum_{i=0}^{\infty} K^i(x,y)\frac{t^i}{i!}.$$

This is a kernel of an operator P_t defined by

$$P_t f = e^{-t}\sum_{i=0}^{\infty}\frac{t^i}{i!}K^i f = e^{-t(I-K)}f. \tag{3}$$

Note that

$$P_t(f)(x) = \mathbb{E}f(X_t).$$

The operators $(P_t)_{t\geq 0}$ have the following three properties:

- P_t preserves positivity, i.e. $f \geq 0$ implies $P_t(f) \geq 0$
- $P_t(1) = 1$
- $P_{t+s} = P_t \circ P_s$ (semigroup property)

Thus, $(P_t)_{t\geq 0}$ is a Markov semigroup. The so-called **generator** L of P_t is given by $Lf = \frac{\mathrm{d}}{\mathrm{d}t}P_t f|_{t=0} = (K-I)f$.

Assume that our kernel K is strongly irreducible, i.e., there is i such that $K^i(x,y) > 0$ for every $x, y \in V$. This implies the existence of the unique stationary measure π. This means that

$$\pi(x) = \sum_{y\in V}\pi(y)K(y,x), \qquad \lim_{l\to\infty} K^l(x,y) = \pi(y).$$

Similar convergence holds for P_t,

$$\lim_{l\to\infty} P_t(x,y) = \pi(y).$$

Let us set

$$p_t^x(y) = p_t(x,y) := \frac{P_t^x(y)}{\pi(y)} = \frac{P_t(x,y)}{\pi(y)}.$$

Definition 3. We say that a Markov chain with a transition matrix K and a positive stationary measure π is reversible (or, in other words, satisfies the detailed balance condition) if we have

$$\pi(x)K(x,y) = \pi(y)K(y,x).$$

Let us define the scalar product

$$\langle f,g\rangle=\sum_{x\in V}f(x)g(x)\pi(x),\qquad \mathcal{E}(f,g)=\langle(-L)f,g\rangle\,.$$

We would like to compute the adjoint $K^\star$ of K. We have

$$\langle f,Kg\rangle=\sum_{x,y}f(x)K(x,y)\overline{g(y)}\pi(x)=\sum_y\left(\sum_x\frac{\pi(x)K(x,y)}{\pi(y)}\right)\overline{g(y)}\pi(y).$$

Thus,

$$(K^\star f)(y)=\sum_x\frac{\pi(x)K(x,y)}{\pi(y)}.$$

It follows that the kernel of $K^\star$ is equal to

$$K^\star(x,y)=\frac{\pi(y)K(y,x)}{\pi(x)}.$$

We see that K satisfies the detailed balance condition if and only if $K^\star=K$. We have also $P_t^\star=e^{-t(I-K^\star)}$. The kernel of $P_t^\star$ is equal to

$$P_t^\star(x,y)=\frac{\pi(y)P_t(y,x)}{\pi(x)}.$$

Moreover, $p_t^\star(x,y)=p_t(y,x)$. Let us set

$$\mu(f)=\sum_x f(x)\pi(x).$$

The operator K acts on measures, $\mu\to\mu K$, namely

$$\mu K(x)=\sum_y\mu(y)K(y,x).$$

Thus,

$$(\mu K)(f)=\sum_{x,y}\mu(y)K(y,x)f(x).$$

The operator $P_t\circ P_s$ has kernel $(P_t\circ P_s)(x,y)=\sum_z P_t(x,z)P_s(z,y)$. Thus, since $P_t\circ P_s=P_{t+s}$, we have a chain rule

$$P_{t+s}(x,y)=\sum_z P_t(x,z)P_s(z,y).$$

Equivalently,

$$p_{t+s}(x,y)=\sum_z p_t(x,z)p_s(z,y)\pi(z).$$

5.3. **Dirichlet form and spectral gap.** Define the Dirichlet form,

$$\mathcal{E}(f,g)=\Re\left(\langle(I-K)f,g\rangle\right).$$

Lemma 2. We have

$$\mathcal{E}(f,f)=\left\langle\left(I-\frac{K+K^\star}{2}\right)f,f\right\rangle=\frac{1}{2}\sum_{x,y}|f(x)-f(y)|^2K(x,y)\pi(x).$$

Moreover, if (K, π) is reversible then

$$\mathcal{E}(f,g) = \frac{1}{2}\sum_{x,y}(f(x)-f(y))(g(x)-g(y))K(x,y)\pi(x).$$

Proof. Observe that

$$\left\langle \left(I - \frac{K+K^\star}{2}\right) f, f \right\rangle = \frac{1}{2}\left(\langle (I-K)f,f\rangle + \langle (I-K^\star)f,f\rangle\right).$$

To prove the first inequality it suffices to show that

$$\langle (I-K^\star)f,f\rangle = \overline{\langle (I-K)f,f\rangle}.$$

Indeed, we have

$$\langle (I-K^\star)f,f\rangle = \langle f,f\rangle - \langle K^\star f,f\rangle = \langle f,f\rangle - \langle f,Kf\rangle = \langle f,f\rangle - \overline{\langle Kf,f\rangle} = \overline{\langle (I-K)f,f\rangle}.$$

For the second equality write

$$\begin{aligned}
&\frac{1}{2}\sum_{x,y}|f(x)-f(y)|^2K(x,y)\pi(x)\\
&\quad= \frac{1}{2}\sum_{x,y}(|f(x)|^2+|f(y)|^2-(\overline{f(x)}f(y)+f(x)\overline{f(y)}))K(x,y)\pi(x)\\
&\quad= \frac{1}{2}\sum_{x}|f(x)|^2\pi(x) + \frac{1}{2}\sum_{y}|f(y)|^2\pi(y) - \sum_{x,y}\Re(\overline{f(x)}f(y))K(x,y)\pi(x)\\
&\quad= \langle f,f\rangle - \sum_{x,y}\Re(\overline{f(x)}f(y))K(x,y)\pi(x).
\end{aligned}$$

In the second inequality we have used $\sum_x \pi(x)K(x,y) = \pi(y)$ (stationarity of π) and $\sum_y K(x,y) = 1$. Now it suffices to observe that

$$\mathcal{E}(f,f) = \Re\left(\langle (I-K)f,f\rangle\right) = \langle f,f\rangle - \Re\left(\langle Kf,f\rangle\right)$$

and

$$\langle Kf,f\rangle = \sum_{x,y}\overline{f(x)}f(y)K(x,y)\pi(x).$$

For the second part note that

$$\mathcal{E}(f,g) = \langle (I-K)f,g\rangle = \sum_x f(x)g(x)\pi(x) - \sum_{x,y}K(x,y)f(y)g(x)\pi(x).$$

Moreover,

$$\begin{aligned}
\frac{1}{2}\sum_{x,y}(f(x)-f(y))(g(x)-g(y))K(x,y)\pi(x) &= \frac{1}{2}\sum_{x,y}f(x)g(x)K(x,y)\pi(x) -\\
\frac{1}{2}\sum_{x,y}f(x)g(y)K(x,y)\pi(x) - \frac{1}{2}&\sum_{x,y}f(y)g(x)K(x,y)\pi(x) + \frac{1}{2}\sum_{x,y}f(y)g(y)K(x,y)\pi(x)
\end{aligned}$$

Now it suffices to observe that by stationarity of π we have

$$\sum_{x,y}f(y)g(y)K(x,y)\pi(x) = \sum_y f(y)g(y)\pi(y)$$

and

$$\sum_{x,y} f(x)g(y)K(x,y)\pi(x) = \sum_{x,y} f(x)g(y)K(y,x)\pi(y) = \sum_{x,y} f(y)g(x)K(x,y)\pi(x).$$

□

Remark 1. The Dirichlet forms related to P_t, $P_t^\star$ and $S_t = \exp\left(-t(I - \frac{K+K^\star}{2})\right)$ are the same.

Lemma 3. We have

$$\frac{\partial}{\partial t}\|P_t f\|_2^2 = -2\mathcal{E}(P_t f, P_t f).$$

Proof. We have

$$\begin{aligned}\frac{\partial}{\partial t}\|P_t f\|_2^2 &= \frac{\partial}{\partial t}\langle P_t f, P_t f\rangle = \langle LP_t f, P_t f\rangle + \langle P_t f, LP_t f\rangle = 2\Re\left(\langle LP_t f, P_t f\rangle\right)\\ &= 2\Re\left(\langle (K-I)P_t f, P_t f\rangle\right) = -2\Re\left(\langle (I-K)P_t f, P_t f\rangle\right) = -2\mathcal{E}(P_t f, P_t f).\end{aligned}$$

□

We define the spectral gap $\lambda = \lambda(K)$.

Lemma 4. The following definitions are equivalent.

(a) $\lambda = \min\left\{\frac{\mathcal{E}(f,f)}{\operatorname{Var}_\pi(f)} :\ \operatorname{Var}_\pi(f) \neq 0, f : V \to \mathbb{C}\right\}$,
(a') $\lambda = \min\left\{\frac{\mathcal{E}(f,f)}{\operatorname{Var}_\pi(f)} :\ \operatorname{Var}_\pi(f) \neq 0, f : V \to \mathbb{R}\right\}$,
(b) $\lambda = \{\mathcal{E}(f,f) :\ \|f\|_2 = 1, \pi(f) = 0\}$,
(c) λ is the second smallest eigenvalue of $I - \frac{K+K^\star}{2}$.

The constant λ will be called the spectral gap of K or the Poincaré constant of K.

Proof. The equivalence of (a) and (b) follows from the fact that the quantity $\mathcal{E}(f,f)/\operatorname{Var}_\pi(f)$ is invariant under shifting and rescaling, $f \to af + b$, $a, b \in \mathbb{C}$.

For the equivalence of (a) and (a') let us observe that $\lambda_{\mathbb{R}} \geq \lambda_{\mathbb{C}}$. On the other hand, for $f = u + iv$, where u, v are real, we get

$$\lambda_{\mathbb{R}} \operatorname{Var}_\pi(f) = \lambda_{\mathbb{R}} \operatorname{Var}_\pi(u) + \lambda_{\mathbb{R}} \operatorname{Var}_\pi(v) \leq \mathcal{E}(u,u) + \mathcal{E}(v,v) = \mathcal{E}(f,f).$$

Thus, $\lambda_{\mathbb{R}} \leq \lambda_{\mathbb{C}}$.

We show the equivalence between (a') and (c). Note that $I - \frac{K+K^\star}{2}$ is self adjoint and therefore it has real eigenvalues $\lambda_0 \leq \lambda_1 \leq \lambda_2 \leq \ldots$. Since

$$\mathcal{E}(f,f) = \left\langle \left(I - \frac{K+K^\star}{2}\right) f, f\right\rangle = \frac{1}{2}\sum_{x,y} |f(x) - f(y)|^2 K(x,y)\pi(x),$$

we get that $\lambda_0 \geq 0$. In fact $\lambda_0 = 0$ since for a constant function $f = \mathbf{1}$ we get $\mathcal{E}(f,f) = 0$. Moreover, $\mathcal{E}(f,f) = 0$ if and only if f is constant on every irreducible component of our state space V. Since we assume that our chain is itself irreducible, we get that the only eigenfunction with eigenvalue 0 is a constant function. Thus, in fact $\lambda_1 > 0$ and it is the spectral gap between first two eigenvalues. However, λ_1 can be degenerate (have multiplicity bigger that 1). Let f_k be the (real) eigenfunction with eigenvalue λ_k. We assume that f_k

are orthonormal with respect to $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_\pi$. Take $f : V \to \mathbb{R}$. It has a unique expansion $f = \sum_{k\geq 0} a_k f_k$. We get

$$\pi(f) = \sum_{k\geq 0} a_k \pi(f_k) = \sum_{k\geq 0} a_k \langle f_k, 1\rangle = a_0.$$

Thus,

$$\mathrm{Var}_\pi(f) = \sum_{k\geq 1} a_k^2, \qquad \mathcal{E}(f,f) = \sum_{k\geq 0} \lambda_k a_k^2 = \sum_{k\geq 1} \lambda_k a_k^2.$$

Clearly λ_1 is the best constant λ in the inequality $\lambda \operatorname{Var}_\pi(f) \leq \mathcal{E}(f,f)$. $\square$

Lemma 5. Let λ be the spectral gap of (K,π). Then for any f we have

$$\mathrm{Var}_\pi(P_t f) = \|P_t f - \pi(f)\|_2^2 \leq e^{-2\lambda t} \operatorname{Var}_\pi(f).$$

Moreover, $\|P_t - \pi\|_{2\to 2} \leq e^{-\lambda t}$.

Proof. We have

$$\pi(Kf) = \sum_{x,y} K(x,y) f(y) \pi(x) = \sum_y f(y)\pi(y) = \pi(f).$$

Thus also $\pi(P_t f) = \pi(f)$. Thus, we get the first equality. To show the inequality let us define $u(t) = \mathrm{Var}_\pi(P_t f) = \|P_t(f - \pi(f))\|_2^2$. From the Lemma 3 we get

$$u'(t) = -2\mathcal{E}(P_t(f-\pi(f)), P_t(f-\pi(f))) \leq -2\lambda u(t).$$

Thus, $u(t) \leq e^{-2\lambda} u(0) = e^{-2\lambda} \operatorname{Var}_\pi(f)$.

To prove the second part it suffices to observe that

$$\|P_t f - \pi(f)\|_2^2 \leq e^{-2\lambda t} \operatorname{Var}_\pi(f) \leq e^{-2\lambda t} \|f\|_2^2.$$

$\square$

Proposition 5. Let (K,π) be a Markov chain with spectral gap λ. Then

$$\|p_t^x - 1\|_2 \leq \sqrt{1/\pi(x)} e^{-\lambda t}, \qquad |P_t(x,y) - \pi(y)| \leq \sqrt{\pi(y)/\pi(x)} e^{-\lambda t}.$$

Corollary 1. *Let (K,π) be a Markov chain with spectral gap λ. Then*

$$\|p_t^x - 1\|_2 \leq e^{-C} \qquad \textit{for} \qquad t = \frac{1}{2\lambda}\left(\ln\left(\frac{1}{\pi(x)}\right) + 2C\right)_+ .$$

and

$$|P_t(x,y) - \pi(y)| \leq e^{-C} \qquad \textit{for} \qquad t = \frac{1}{2\lambda}\left(\ln\left(\frac{\pi(y)}{\pi(x)}\right) + 2C\right)_+ .$$

Proof of Proposition 5. Let $P_t^\star$ be the adjoint Markov chain with the spectral gap $\lambda(K^\star) = \lambda(K)$. Define $\delta_x(y) = (1/\pi(x))\mathbf{1}_{y=x}$. We have

$$p_t^x(y) = \frac{P_t(x,y)}{\pi(y)} = \frac{P_t^\star(y,x)}{\pi(x)} = \sum_z P_t^\star(y,z)\delta_x(z) = (P_t^\star \delta_x)(y).$$

We have $\pi(P_t^\star \delta_x) = \pi(\delta_x) = \pi(x)/\pi(x) = 1$. Thus,

$$\|p_t^x - 1\|_2^2 = \|P_t^\star \delta_x - \pi(P_t^\star \delta_x)\|_2^2 = \operatorname{Var}_\pi(P_t^\star \delta_x) \leq e^{-2\lambda t} \operatorname{Var}_\pi(\delta_x) = \left(\frac{1}{\pi(x)} - 1\right) e^{-2\lambda t}$$

We arrive at

$$\|p_t^x - 1\|_2 \leq \sqrt{1/\pi(x) - 1}\, e^{-\lambda t} \leq \sqrt{1/\pi(x)}\, e^{-\lambda t}.$$

For the second part observe that

$$\begin{aligned}&\sum_z (p_{t/2}(x,z)-1)(p_{t/2}(z,y)-1)\pi(z)\\&\quad=\sum_z p_{t/2}(x,z)p_{t/2}(z,y)\pi(z)-\sum_z p_{t/2}(x,z)\pi(z)-\sum_z p_{t/2}(z,y)\pi(z)+\sum_z \pi(z)\\&\quad=p_t(x,y)-\sum_z \frac{P_{t/2}(x,z)}{\pi(z)}\pi(z)-\sum_z \frac{P_{t/2}(z,y)}{\pi(y)}\pi(z)+1=p_t(x,y)-1.\end{aligned}$$

Thus,

$$|p_t(x,y)-1|\le \left\|p^x_{t/2}-1\right\|_2 \|p^{\star y}_{t/2}-1\|_2\le \frac{1}{\sqrt{\pi(x)\pi(y)}}e^{-\lambda t}.$$

Multiplying by $\pi(y)$ give the result. ☐

5.4. **Log-Sobolev inequalities.**

Lemma 6 (Stroock-Varopoulos inequality)**.** If (K,π) is reversible and $f\ge 0$ then for any $p>1$ we have

$$\frac{4(p-1)}{p^2}\mathcal{E}(f^{p/2},f^{p/2})\le \mathcal{E}(f,f^{p-1}).$$

Proof. Take $a>b\ge 0$. By Cauchy-Schwarz we have

$$\left(\frac{a^{p/2}-b^{p/2}}{a-b}\right)^2=\left(\frac{p}{2(a-b)}\int_b^a t^{p/2-1}\mathrm{d}t\right)^2\le \frac{p^2}{4(a-b)}\int_b^a t^{p-2}\mathrm{d}t=\frac{p^2}{4(p-1)}\frac{a^{p-1}-b^{p-1}}{a-b}.$$

We get

$$(a^{p-1}-b^{p-1})(a-b)\ge \frac{4(p-1)}{p^2}(a^{p/2}-b^{p/2})^2.$$

Thus, from Lemma 2 we get

$$\begin{aligned}\frac{4(p-1)}{p^2}\mathcal{E}(f^{p/2},f^{p/2})&=\frac{4(p-1)}{p^2}\cdot\frac{1}{2}\sum_{x,y}|f^{p/2}(x)-f^{p/2}(y)|^2K(x,y)\pi(x)\\&\le \frac{1}{2}\sum_{x,y}(f^{p-1}(x)-f^{p-1}(y))(f(x)-f(y))K(x,y)\pi(x)=\mathcal{E}(f,f^{p-1}).\end{aligned}$$

☐

Lemma 7. Let φ be convex. Then $\varphi(P_tf)\le P_t(\varphi(f))$. Moreover, $\mathbb{E}\varphi(P_tf)\le\mathbb{E}\varphi(f)$. In particular, $\|P_tf\|_p\le\|f\|_p$, $p\ge 1$.

Proof. Any convex function is a supremum of a certain family of convex functions $\varphi(x)=\sup_\alpha(a_\alpha x+b_\alpha)$. We have $a_\alpha f+b_\alpha\le\varphi(f)$. Applying P_t and using the fact that it is linear and preserves positivity, we get $a_\alpha P_tf+b_\alpha\le P_t(\varphi(f))$. Taking supremum over α we get $\varphi(P_tf)\le P_t(\varphi(f))$. To get the second assertion we apply expectation and use the fact that P_t preserves expectation. ☐

Definition 4. For a Markov chain (K, π) the log-Sobolev constant $\alpha = \alpha(K)$ is defined via

$$\alpha = \min\left\{ \frac{\mathcal{E}(f,f)}{\mathrm{Ent}_\pi(|f|^2)} : \ \mathrm{Ent}_\pi(|f|^2) \neq 0 \right\}.$$

Proposition 6. For any Markov chain (K, π) we have $2\alpha \leq \lambda$.

Proof. It suffices to take $f = 1 + \varepsilon g$ in the above definition (with g real) and observe that $\mathcal{E}(f,f) = \varepsilon^2 \mathcal{E}(g,g)$ and (by easy Taylor expansion) $\mathrm{Ent}_\pi(|f|^2) = 2\varepsilon^2 \mathrm{Var}_\pi(g) + O(\varepsilon^3)$. One gets the result by taking $\varepsilon \to 0$. $\square$

We prove that Log-Sobolev inequality is equivalent to the hypercontractivity property.

Theorem 9. For a reversible chain with a generator L the following statements are equivalent,

(i) (*Log-Sobolev inequality*) for every $f : \Omega \to \mathbb{R}$ satisfying suitable technical assumptions

$$\mathbb{E}(f^2 \ln f^2) - (\mathbb{E} f^2) \ln(\mathbb{E} f^2) \leq C \mathbb{E}(f(-L)f),$$

(ii) (*hypercontractivity*) for every $p > q > 1$ and $f : \Omega \to \mathbb{R}$ we have

$$\|\mathcal{P}_t f\|_p \leq \|f\|_q$$

for $t \geq \frac{C}{4} \ln \frac{p-1}{q-1}$.

Proof. Assume that we have *(i)*. Take $\phi_q : [q, \infty) \to \mathbb{R}$ given by

$$\phi_q(p) = \ln \left\|\mathcal{P}_{t(p)} f\right\|_p = \frac{1}{p} \ln \mathbb{E} \left|\mathcal{P}_{t(p)} f\right|^p,$$

where $t(p) = \frac{C}{4} \ln \frac{p-1}{q-1}$. It suffices to show that $\left\|\mathcal{P}_{t(p)} f\right\|_p \leq \|f\|_q$. Indeed, if $t > t(p)$ then we obtain

$$\|\mathcal{P}_t f\|_p = \left\|\mathcal{P}_{t(p)+t-t(p)} f\right\|_p \leq \left\|\mathcal{P}_{t-t(p)} f\right\|_q \leq \|f\|_q,$$

since $\mathcal{P}_{t-t(p)}$ is a contraction in L^q.

To prove that $\left\|\mathcal{P}_{t(p)} f\right\|_p \leq \|f\|_q$ we can assume that f i nonnegative. Indeed, the inequality $-|f| \leq f \leq |f|$ implies (positivity preserving) that $-\mathcal{P}_t|f| \leq \mathcal{P}_t f \leq \mathcal{P}_t|f|$, hence $|\mathcal{P}_t f| \leq \mathcal{P}_t|f|$. Therefore $\left\|\mathcal{P}_{t(p)} f\right\|_p \leq \left\|\mathcal{P}_{t(p)} |f|\right\|_p$.

Take a nonnegative f. Since $t(q) = 0$, the inequality $\left\|\mathcal{P}_{t(p)} f\right\|_p \leq \|f\|_q$ is equivalent to $\phi_q(p) \leq \phi_q(q)$. Hence, it suffices to show that the function $[q, \infty) \ni p \mapsto \phi_q(p)$ is nonincreasing. Set $\mathcal{P}_{t(p)} f = f_p$. We have

$$\frac{\mathrm{d}}{\mathrm{d}p} \phi_q(p) = \frac{1}{p} \frac{\mathbb{E} \frac{\mathrm{d}}{\mathrm{d}p}(f_p^p)}{\mathbb{E} f_p^p} - \frac{1}{p^2} \ln \mathbb{E} f_p^p$$

and

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}p} f_p^p &= \frac{\mathrm{d}}{\mathrm{d}p} \left(\mathcal{P}_{t(p)} f\right)^p = \frac{\mathrm{d}}{\mathrm{d}p} e^{p \ln(\mathcal{P}_{t(p)} f)} = e^{p \ln(\mathcal{P}_{t(p)} f)} \left(\ln(\mathcal{P}_{t(p)} f) + p \frac{L \mathcal{P}_{t(p)} f}{\mathcal{P}_{t(p)} f} \right) \cdot \frac{\mathrm{d}t(p)}{\mathrm{d}p} \\
&= f_p^p \ln f_p + f_p^{p-1} p (L f_p) \frac{C}{4} \ln \frac{1}{p-1}.
\end{aligned}$$

Thus,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}p}\phi_q(p) &= \frac{1}{p}\cdot\frac{\mathbb{E}f_p^p\ln f_p}{\mathbb{E}f_p^p} + \frac{C}{4}\frac{1}{p-1}\cdot\frac{\mathbb{E}f_p^{p-1}Lf_p}{\mathbb{E}f_p^p} - \frac{1}{p^2}\ln\mathbb{E}f_p^p \\ &= \frac{1}{p^2\mathbb{E}f_p^p}\left(\left(\mathbb{E}f_p^p\ln(f_p^p) - (\mathbb{E}f_p^p)\ln(\mathbb{E}f_p^p)\right) + \frac{Cp}{4(p-1)}\mathbb{E}(f_p^{p-1}Lf_p)\right) \\ &= \frac{1}{p^2\mathbb{E}f_p^p}\left(\mathrm{Ent}(f_p^p) + \frac{Cp}{4(p-1)}\mathbb{E}(f_p^{p-1}Lf_p)\right).\end{aligned}$$

We would like to prove

$$\mathrm{Ent}(f_p^p) \le \frac{Cp^2}{4(p-1)}\mathbb{E}(f_p^{p-1}(-L)f_p).$$

Taking $f = f_p^{p/2}$ in the Log-Sobolev inequality and using Stroock-Varopoulos inequality we obtain

$$\mathrm{Ent}(f_p^p) \le C\mathbb{E}\left(f_p^{p/2}(-L)f_p^{p/2}\right) \le \frac{Cp^2}{4(p-1)}\mathbb{E}(f_p^{p-1}(-L)f_p).$$

To prove that *(ii)* implies *(i)* observe that for a nonnegative function f the inequality $\left\|\mathcal{P}_{t(p)}f\right\|_p \le \|f\|_q$ implies that $\frac{\mathrm{d}}{\mathrm{d}p}\left\|\mathcal{P}_{t(p)}f\right\|_p\Big|_{p=q} \le 0$, which is equivalent to

$$\mathrm{Ent}(f^q) \le \frac{Cq^2}{4(q-1)}\mathbb{E}(f^{q-1}(-L)f).$$

Now it suffices to take $q = 2$ to obtain Log-Sobolev inequality for nonnegative functions. If f is not necessarily nonnegative then we have

$$\mathrm{Ent}(f^2) = \mathrm{Ent}(|f|^2) \le C\mathbb{E}|f|(-L)|f| \le C\mathbb{E}f(-L)f$$

because of the energy stability lemma. □

Since the continuous time random walk on Σ_n satisfy Log-Sobolev inequality with constant 2, we have proved the following theorem.

Theorem 10. *Let $(\mathcal{P}_t)_{t\ge 0}$ be the continuous time random walk on Σ_n. Then for every $p > q > 1$ and $t \ge \frac{1}{2}\ln\frac{p-1}{q-1}$ we have*

$$\|\mathcal{P}_t f\|_p \le \|f\|_q.$$

As an application of the hypercontractivity we prove the following proposition.

Proposition 7 (Khinchin-Kahane inequality)**.** *Let $(F, \|\cdot\|)$ be a normed space and let $v_1, \ldots, v_n \in F$. Then for $p > q > 1$ we have*

$$\left(\mathbb{E}\left\|\sum_{i=1}^n r_i v_i\right\|\right)^{1/p} \le \sqrt{\frac{p-1}{q-1}}\left(\mathbb{E}\left\|\sum_{i=1}^n r_i v_i\right\|\right)^{1/q}.$$

Proof. Let $H(x) = \|\sum_{i=1}^n x_i v_i\|$, $H : \Sigma_n \to [0,\infty)$. We have proved that $(-L)H \le H$. Hence,

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}_t H = L\mathcal{P}_t H = -\mathcal{P}_t LH \ge -\mathcal{P}_t H.$$

Therefore $\mathcal{P}_t H \ge e^{-t}\mathcal{P}_0 H = e^{-t}H$. Take $t = \frac{1}{2}\ln\frac{p-1}{q-1}$. By the hypercontractivity of $\mathcal{P}_t$ we obtain

$$\sqrt{\frac{q-1}{p-1}}\|H\|_p = e^{-t}\|H\|_p \le \|\mathcal{P}_t H\|_p \le \|H\|_q.$$

□

Proposition 8. For all $t, s \geq 0$ we have

$$\|p_{t+s} - 1\|_2 \leq \pi(x)^{-\frac{1}{1+e^{4\alpha s}}} e^{-\lambda t}.$$

Moreover, we have

$$\|p_T^x - 1\|_2 \leq e^{1-C}, \qquad \text{for} \qquad T = \frac{1}{4\alpha} \ln_+ \ln\left(\frac{1}{\pi(x)}\right) + \frac{C}{\lambda}$$

and

$$|p_T(x,y) - 1| \leq e^{2-C}, \qquad \text{for} \qquad T = \frac{1}{4\alpha}\left(\ln_+ \ln\left(\frac{1}{\pi(x)}\right) + \ln_+ \ln\left(\frac{1}{\pi(y)}\right)\right) + \frac{C}{\lambda}.$$

Lemma 8. Let $1 \leq p, r \leq \infty$. Then for any linear operator K we have $\|K\|_{p\to r} = \|K^\star\|_{r'\to p'}$, where r', p' are the Hölder conjugate to r and p.

Proof. We use a well known fact that

$$\|f\|_p = \sup_{\|g\|_{p'} \leq 1} |\langle f, g\rangle|.$$

Thus,

$$\begin{aligned}\|K\|_{p\to r} &= \sup_{\|f\|_p \leq 1} \|Kf\|_r = \sup_{\|f\|_p\leq 1} \sup_{\|g\|_{r'}\leq 1} |\langle Kf, g\rangle| = \sup_{\|g\|_{r'}\leq 1} \sup_{\|f\|_p \leq 1} |\langle K^\star g, f\rangle| \\ &= \sup_{\|g\|_{r'}\leq 1} \|K^\star g\|_{p'} = \|K\|_{r'\to p'}.\end{aligned}$$

□

Proof. Take $q(s) = 1 + e^{4\alpha s}$. By Theorem 9 we have $\|P_s\|_{2\to q(s)} \leq 1$. By Lemma 8 and the fact that $L_2^\star = L_2$ and $L_q^\star = L_p$ with $1/q(s) + 1/p(s) = 1$ we have $\|P_s^\star\|_{p(s)\to 2} \leq 1$. Take $\delta_x(y) = \frac{1}{\pi(x)} \mathbf{1}_{y=x}$. In the proof of Proposition 5 we showed that $p_t(x,y) = (P_t^\star \delta_x)(y)$. Thus

$$p_{t+s}(x,y) - 1 = ((P_{t+s}^\star - \pi)\delta_x)(y) = (P_s^\star(P_t^\star - \pi)\delta_x)(y),$$

since $P_s^\star(P_t^\star - \pi) = P_{t+s}^\star - \pi$. We get

$$\begin{aligned}\left\|p_{t+s}^x - 1\right\|_2 &= \left\|(P_{t+s}^\star - \pi)\delta_x\right\|_2 = \|P_s^\star(P_t^\star - \pi)\delta_x\|_2 \leq \|P_s^\star \delta_x\|_2 \, \|P_t^\star - \pi\|_{2\to 2} \\ &\leq \|\delta_x\|_{p(s)} \, \|P_s^\star\|_{p(s)\to 2} \, \|P_t^\star - \pi\|_{2\to 2}.\end{aligned}$$

First, recall that $\|P_s^\star\|_{p(s)\to 2} \leq 1$. Moreover,

$$\|\delta_x\|_{p(s)} = \left(\left(\frac{1}{\pi(x)}\right)^{\frac{1}{p(s)}} \pi(x)\right)^{1/p(s)} = \pi(x)^{\frac{1}{p(s)} - 1} = \pi(x)^{-\frac{1}{q(s)}}.$$

Finally, by Lemma 5 applied for $P_t^\star$ we have $\|P_t^\star - \pi\|_{2\to 2} \leq 1$.

To prove that the second part take

$$s = \frac{1}{4\alpha} \ln_+ \ln\left(\frac{1}{\pi(x)}\right), \qquad t = \frac{C}{\lambda}.$$

The third part follows from the second and $|p_t(x,y) - 1| \leq \|p_{t/2}^x - 1\|_2 \|p_{t/2}^{\star y} - 1\|_2$ (see the proof of Proposition 5). □

5.5. **Example: continuous time random walk on the cube.** Let us consider a continuous time random walk on the cube $\{-1,1\}^n$. For this walk we have

$$K(x,y)=\begin{cases} \frac{1}{n} & d_H(x,y)=1 \\ 0 & \text{therwise} \end{cases}.$$

Here $d_H(x,y)=|\{1\leq i\leq n:\ x_i\neq y_i\}|$ is the so-called Hamming distance. If $d_H(x,y)=1$ then we will say that x and y are neighbours and we will write $x\sim y$. This relation induces the standard graph structure on the cube. Let us compute the generator $Lf=(K-I)f$. We get

$$(Lf)(x)=\frac{1}{n}\left(\sum_{y\sim x}f(y)\right)-1=\frac{1}{n}\sum_{y\sim x}(f(y)-f(x)).$$

Note that the uniform measure $\pi(x)=2^{-n}$ satisfies the condition

$$\pi(x)=\sum_{y\in\{-1,1\}^n}\pi(y)K(y,x).$$

However, it does not satisfy the condition $\lim_{l\to\infty}K^l(x,y)=\pi(y)$, because, $K^{2l}(x,y)=0$ when $d_H(x,y)$ is odd. However, as we will see later, this problem disappears when we pass to P_t. Thus, $\pi=\mu_n$. The Dirichlet form is equal to,

$$\begin{aligned}\mathcal{E}(f,g)=\langle(-L)f,g\rangle&=\frac{1}{2}\sum_{x,y}(f(x)-f(y))(g(x)-g(y))K(x,y)\pi(x)\\ &=\frac{1}{2^{n+1}n}\sum_{(x,y):\,x\sim y}(f(x)-f(y))(g(x)-g(y)).\end{aligned}$$

Thus,

$$\mathcal{E}(f,f)=\frac{1}{2^{n+1}n}\sum_{(x,y):\,y\sim x}(f(x)-f(y))^2=\frac{1}{2^{n-1}n}\sum_{(x,y):\,y\sim x}\left(\frac{f(x)-f(y)}{2}\right)^2=\frac{2}{n}\int|\nabla f|^2\mathrm{d}\mu_n.$$

We have seen the Poincaré inequality on the cube,

$$\mathrm{Var}_{\mu_n}(f)\leq\int|\nabla f|^2\mathrm{d}\mu_n=\frac{n}{2}\mathcal{E}(f,f).$$

We get that the spectral gap is equal to $\lambda=2/n$.

We have seen that $Lw_S=-2\frac{|S|}{n}w_S$, where w_S is the Walsh-Fourier function.

Recall that the discrete LSI says that

$$\mathrm{Ent}_{\mu_n}(f^2)\leq 2\int|\nabla f|^2\mathrm{d}\mu_n=n\mathcal{E}(f,f).$$

As a consequence, the log-Sobolev constant for the continuous time random walk equals $1/n$. Thus, the eigenvalues of $(-L)=I-K$ are equal to $\lambda_k=2\frac{k}{n}$, each with multiplicity $\binom{n}{k}$. Note that $\lambda_0=0$ and $\lambda=\lambda_1=2/n$.

Let us compute the action of P_t on a function $f=\sum_S a_S w_S$. We get

$$P_tf=\sum_S a_S e^{-2t\frac{|S|}{n}}w_S.$$

Previously we mentioned (and proved for $q = 2$) that the operator $P_{\frac{n}{2}t}$ satisfies the following hypercontractivity property

$$e^{-t} \leq \sqrt{\frac{p-1}{q-1}} \qquad \Longrightarrow \qquad \left\|P_{\frac{n}{2}t}\right\|_q \leq \|f\|_p .$$

From Theorem 9 we get that

$$t \geq \frac{n}{4} \ln\left(\frac{p-1}{q-1}\right) \qquad \Longrightarrow \qquad \|P_t\|_q \leq \|f\|_p .$$

Clearly those two conditions are the same.

Proposition 5 yields

$$\|p_t^x - 1\|_2^2 \leq e^{-C} \qquad \text{for} \qquad t = \frac{n}{4}(n \ln 2 + 2C)_+,$$

which is (say, for $C = 1$) roughly $n^2 \frac{\ln 2}{4}$. As we will see, the log-Sobolev constant give better bound. Indeed, from Proposition 8 we get

$$\|p_t^x - 1\|_2 \leq e^{1-C} \qquad \text{for} \qquad t = \frac{n}{4}\ln(n \ln 2) + \frac{Cn}{2}.$$

For fixed C this is roughly $\frac{n}{4}\ln n$. Let us see that this is in fact the correct order. We have

$$\delta_x(y) = 2^n \mathbf{1}_{y=x} = 2^n \prod_{i=1}^n \frac{1 + x_i y_i}{2} = \sum_S w_S(x) w_S(y).$$

Therefore,

$$P_t \delta_x = \sum_S e^{-2t\frac{|S|}{n}} w_S(x) w_S$$

and

$$\|p_t^x - 1\|_2^2 = \mathrm{Var}_\pi(P_t^x \delta_x) = \sum_{k>0} \binom{n}{k} e^{-4t\frac{k}{n}} = \left(1 + e^{-\frac{4t}{n}}\right)^n - 1.$$

Thus we have $\|p_t^x - 1\|_2^2 = e^{2-2C}$ for $t = -\frac{n}{4}\ln\left((1 + e^{2-2C})^{\frac{1}{n}} - 1\right) \approx \frac{n}{4}\ln n$. To see the last asymptotics it suffices to note that for any $a > 1$ we have $\lim_{n\to\infty}(\ln(a^{\frac{1}{n}} - 1)/\ln n) = -1$.

5.6. **Some spectral graph theory.** Let us recall some properties of symmetric matrices. Suppose M is a symmetric $n \times n$ matrix. Then M has real eigenvalues $\lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n$ with orthonormal eigenvectors $x_1, x_2, \ldots, x_n$, i.e., $Mx_k = \lambda_k x_k$, $k = 1, \ldots, n$. Moreover,

$$\lambda_k = \min_{x \neq 0, x \perp x_1, \ldots, x \perp x_{k-1}} \frac{x^T M x}{x^T x}.$$

Moreover, any minimizer is an eigenvector with eigenvalue λ_k. In particular,

$$\lambda_1 = \min_{x \neq 0} \frac{x^T M x}{x^T x}.$$

Let x_1 be the minimizer in the above expression, thus the eigenvector of M with eigenvalue λ_1. Then

$$\lambda_2 = \min_{x \neq 0, x \perp x_1} \frac{x^T M x}{x^T x}.$$

We also have the following min-max principle,

$$\lambda_k = \min_{V-\text{subspace of } \mathbb{R}^n,\, \dim V = k} \ \max_{x\in V,\, x\neq 0} \frac{x^T M x}{x^T x}.$$

Consider a simple random walk on d regular graph, i.e., let us take

$$K(x,y) = \begin{cases} \frac{1}{d} & x\sim y \\ 0 & x\nsim y \end{cases}.$$

Thus, $(-L) = I - \frac{1}{d}A$, where A is the adjacency matrix of G,

$$A(x,y) = \begin{cases} 1 & x\sim y \\ 0 & x\nsim y \end{cases}.$$

We prove the following proposition.

Proposition 9. Let G be a d regular graph on n vertices. Let $\lambda_1 \leq \ldots \leq \lambda_n$ be eigenvalues of $\mathcal{L} = -L$. Then

(a) $\lambda_1 = 0$,
(b) $\lambda_k = 0$ if and only if G has at least k connected components,
(c) $\lambda_n \leq 2$ and $\lambda_n = 2$ if and only if at least one connected component of G is bipartite.

Proof. (a) From Proposition 2 we get

$$x^T\mathcal{L}x = \frac{1}{d}\sum_{\{u,v\}\in E}(x_u - x_v)^2,$$

where the notation $\{u,v\}\in E$ means that every edge is counted ones. As for general Markov chains we get

$$\lambda_1 = \min_{x\neq 0}\frac{x^T\mathcal{L}x}{x^Tx} = \min_{x\neq 0}\frac{\sum_{\{u,v\}\in E}(x_u-x_v)^2}{d\sum_v x_v^2} \geq 0.$$

Moreover, a constant vector $x = (1,\ldots,1)$ gives $\lambda_1 = 0$ and this vector is an eigenvector of $\mathcal{L}$ with eigenvalue 0.

(b) Assume $\lambda_k = 0$. Since

$$\lambda_k = \min_{V-\text{subspace of } \mathbb{R}^n,\, \dim V = k} \ \max_{x\in V,\, x\neq 0} \frac{x^T M x}{x^T x},$$

we see that there is a k dimensional subspace S such that for every $x\in S$ we have $\sum_{\{u,v\}\in S}(x_u - x_v)^2 = 0$. But this means that x has to be constant on every connected component of G. Thus, the dimension of S is at most the number of connected components of G. Thus, G has at least k connected components.

Conversely, if G has at least k connected components then we can take S to be a subspace of vectors constant on each component of G. We have $\dim(S)\geq k$. For every element of $x\in S$ we have $\sum_{\{u,v\}}(x_u - x_v)^2 = 0$. This gives $\lambda_k = 0$ by the min-max principle.

(c) Let us recall that

$$\lambda_n = \max_{x\neq 0}\frac{x^T\mathcal{L}x}{x^Tx}.$$

We have

$$x^T\mathcal{L}x = \frac{1}{d}\sum_{\{u,v\}\in E}(x_u-x_v)^2 = |x|^2 - \frac{2}{d}\sum_{\{u,v\}\in E}x_ux_v = 2|x|^2 - \frac{1}{d}\sum_{\{u,v\}\in E}(x_u+x_v)^2.$$

Thus,

$$\lambda_n = \max_{x\neq 0} \frac{x^T \mathcal{L} x}{x^T x} = \max_{x \neq 0} \left(2 - \frac{1}{d} \frac{\sum_{\{u,v\}\in E} (x_u + x_v)^2}{x^T x} \right) \leq 2.$$

Moreover, if $\lambda_n = 2$ then there must be a non-zero vector x such that

$$\sum_{\{u,v\}\in E} (x_u + x_v)^2 = 0.$$

Let v_0 be a vertex with $x_{v_0} = a \neq 0$. Define

$$A = \{v : \ x_v = a\}, \qquad B = \{v : \ x_v = -a\}, \qquad R = \{v : \ |x_v| \neq a\}.$$

We see that $A \cup B$ is disconnected from the rest of the graph R. Otherwise any edge $\{u, v\}$ from R to $A \cup B$ would give $(x_u + x_v)^2 > 0$. Moreover, for the same reason if $v \in A$ and $\{u, v\} \in E$ then $u \in B$. Thus, A and B gives a bipartition of $A \cup B$, which is a sum of connected bipartite components of G. □

5.7. **Maximal Cut.** Let us define the maximal cut for the graph $G = (V, E)$,

$$\mathrm{MaxCut}(G) = \max_{S \subseteq V} \frac{E(S, V \setminus S)}{|E|}.$$

Note that $\mathrm{MaxCut}(G) \leq 1$ and $\mathrm{MaxCut}(G) = 1$ if and only if G is bipartite. Observe that

$$\max_{x\in\{-1,1\}^n} \frac{\sum_{\{u,v\}\in E} (x_u - x_v)^2}{d \sum_v x_v^2} = \max_{S\subseteq V} \frac{4E(S, V \setminus S)}{dn} = 2 \max_{S \subseteq V} \frac{E(S, V \setminus S)}{|E|} = 2\mathrm{MaxCut}(G).$$

We get

$$2\mathrm{MaxCut}(G) \leq \lambda_n.$$

5.8. **Cheeger inequality.** Recall that

$$\lambda_2 = \min_{x\neq 0,\, x \perp \mathbf{1}} \frac{\sum_{\{u,v\}\in E} (x_u - x_v)^2}{d \sum_v x_v^2}.$$

For $x \perp \mathbf{1}$ we have

$$\sum_{u,v\in V} (x_u - x_v)^2 = 2n \sum_v x_v^2 - 2 \sum_{u,v} x_u x_v = 2n \sum_v x_v^2 - 2 \left(\sum_v x_v \right)^2 = 2n \sum_v x_v^2.$$

Thus,

$$\begin{aligned} \lambda_2 &= \min_{x\neq 0,\, x\perp \mathbf{1}} \frac{\sum_{\{u,v\}\in E} (x_u - x_v)^2}{\frac{d}{2n} \sum_{u,v\in V} (x_u - x_v)^2} = \min_{x-\text{non-constant}} \frac{\sum_{\{u,v\}\in E} (x_u - x_v)^2}{\frac{d}{2n} \sum_{u,v\in V} (x_u - x_v)^2} \\ &= \min_{x-\text{non-constant}} \frac{\frac{1}{nd/2} \sum_{\{u,v\}\in E} (x_u - x_v)^2}{\frac{1}{n^2} \sum_{u,v\in V} (x_u - x_v)^2} = \min_{x-\text{non-constant}} \frac{\mathbb{E}_{\{u,v\}\in E} (x_u - x_v)^2}{\mathbb{E}_{u,v\in V} (x_u - x_v)^2}, \end{aligned}$$

where $\mathbb{E}_{\{u,v\}\in E}$ is the expectation with respect to the uniform distribution on E and $\mathbb{E}_{u,v}$ refers to independent uniform choice of u and v. The above minimization problem is a relaxation of uniform sparsest cut problem,

$$\mathrm{USC}(G) = \frac{n}{d} \min_{S\subseteq V} \frac{E(S, V \setminus S)}{|S| \cdot |V \setminus S|} = \min_{\substack{x - \text{non-constant} \\ x \in \{-1,1\}^n}} \frac{\mathbb{E}_{\{u,v\}\in E} (x_u - x_v)^2}{\mathbb{E}_{u,v\in V} (x_u - x_v)^2}.$$

Clearly we have $\mathrm{USC}(G) \geq \lambda_2$.

Definition 5. Let $S \subseteq V$. We define the conductance of S and the conductance of graph G,

$$\phi(S) = \frac{E(S, V \setminus S)}{d|S|}, \qquad \phi(G) = \min_{0<|S|\leq |V|/2} \phi(S).$$

Let us observe that $\mathrm{USC}(G) \leq 2\phi(G)$. Indeed,

$$\begin{aligned}\mathrm{USC}(G) &= \frac{n}{d} \min_{S\subseteq V} \frac{E(S, V\setminus S)}{|S|\cdot|V\setminus S|} \leq \frac{n}{d} \min_{0<|S|\leq|V|/2} \frac{E(S, V\setminus S)}{|S|\cdot|V\setminus S|} \\ &\leq 2 \min_{0<|S|\leq|V|/2} \frac{E(S,V\setminus S)}{d|S|} = 2\phi(G).\end{aligned}$$

Theorem 11. We have $\lambda_2 \leq \mathrm{USC}(G) \leq 2\phi(G) \leq \sqrt{8\lambda_2}$.

Proof. The only non-trivial inequality is $\phi(G) \leq \sqrt{2\lambda_2}$. Given a solution x of the minimization problem for λ_2 we are to find a good Boolean approximation (set S). We do this in several steps.

Step 1. Given a solution x with $x \perp \mathbf{1}$ it is enough to construct a vector $y \in \mathbb{R}^n$ such that $y_v \geq 0$, $|\{v : y_v > 0\}| \leq n/2$, $\max_v y_v = 1$ and

$$\frac{\sum_{\{u,v\}\in E} |y_u - y_v|}{d\sum_v |y_v|} \leq 2\sqrt{\frac{\sum_{\{u,v\}\in E}(x_u - x_v)^2}{d\sum_v x_v^2}} = 2\sqrt{\lambda_2}.$$

Indeed, having such a vector y we construct the set $S \subseteq V$ (in fact we will find $S \subseteq \{v : y_v > 0\}$ and thus we will get $|S| \leq |V|/2$) as follows. Take a random threshold $t \sim \mathrm{Unif}[0, \max_v y_v]$ and define $S = \{v : y_v \geq t\}$. We have

$$\frac{\mathbb{E}E(S, V\setminus S)}{d\mathbb{E}|S|} = \frac{\sum_{\{u,v\}\in E} \mathbb{P}(|\{u,v\}\cap S| = 1)}{d\sum_v \mathbb{P}(v\in S)} = \frac{\sum_{\{u,v\}\in E}|y_u - y_v|}{d\sum_v |y_v|}.$$

Now it suffices to observe that

$$\min_{0<|S|\leq|V|/2} \frac{E(S,V\setminus S)}{d|S|} \leq \frac{\mathbb{E}E(S,V\setminus S)}{d\mathbb{E}|S|}.$$

This is due to the general and easy inequality $\min\left(\frac{X}{Y}\right) \leq \frac{\mathbb{E}X}{\mathbb{E}Y}$ valid for any positive real random variable X, Y. Indeed, the inequality $\frac{X}{Y} > \frac{\mathbb{E}X}{\mathbb{E}Y}$ leads to $X\mathbb{E}Y > Y\mathbb{E}X$ which is, after taking expectation of both sides, a contradiction.

Step 2a. Take $z_v = x - \mathrm{Med}(x)$. Observe that

$$\frac{\sum_{\{u,v\}\in E}(z_u - z_v)^2}{d\sum_v z_v^2} \leq \frac{\sum_{\{u,v\}\in E}(x_u - x_v)^2}{d\sum_v x_v^2}.$$

This follows from the fact that

$$|z|^2 = |x - \mathrm{Med}(x)\mathbf{1}|^2 = |x|^2 - \mathrm{Med}(X)\langle x, \mathbf{1}\rangle + n\,\mathrm{Med}(X)^2 = |x|^2 + n\,\mathrm{Med}(X)^2 \geq |x|^2.$$

Step 2b. Define

$$z_v^+ = \begin{cases} 0 & z_v < 0 \\ z_v & z_v \geq 0 \end{cases}, \qquad z_v^- = \begin{cases} 0 & z_v < 0 \\ -z_v & z_v < 0 \end{cases}.$$

Thus, $z = z^+ - z^-$ and $z^+ \perp z^-$. Note that $|z_u - z_v|^2 \geq |z_u^+ - z_v^+|^2 + |z_u^- - z_v^-|^2$ Therefore,

$$\lambda_2 \geq \frac{\sum_{\{u,v\}\in E}(z_u - z_v)^2}{d\sum_v z_v^2} \geq \frac{\sum_{\{u,v\}\in E}(z_u^+ - z_v^+)^2 + \sum_{\{u,v\}\in E}(z_u^- - z_v^-)^2}{d\sum_v (z_v^+)^2 + d\sum_v (z_v^-)^2}.$$

We get that

$$\lambda_2 \geq \frac{\sum_{\{u,v\}\in E}(z_u^+ - z_v^+)^2}{d\sum_v (z_v^+)^2} \qquad \text{or} \qquad \lambda_2 \geq \frac{\sum_{\{u,v\}\in E}(\sum_{\{u,v\}\in E}(z_u^- - z_v^-)^2}{d\sum_v (z_v^-)^2}.$$

Note that since z has median 0, we have $|\{v : z_v^+ > 0\}| \leq n/2$ and $|\{v : z_v^- > 0\}| \leq n/2$. Moreover $z_v^\pm \geq 0$.

Step 2c. We have constructed a vector w such that $w_v \geq 0$, $|v : w_v > 0| \leq n/2$ and

$$\lambda_2 \geq \frac{\sum_{\{u,v\}\in E}(w_u - w_v)^2}{d\sum_v w_v^2}$$

Take $y_v = w_v^2$. Clearly $y_v \geq 0$, $|v : y_v > 0| \leq n/2$. We have

$$\begin{aligned}\sum_{\{u,v\}\in E} |w_u^2 - w_v^2| &= \sum_{\{u,v\}\in E} |w_u - w_v||w_u + w_v| \\ &\leq \left(\sum_{\{u,v\}\in E} |w_u - w_v|^2\right)^{1/2} \left(\sum_{\{u,v\}\in E} |w_u + w_v|^2\right)^{1/2}.\end{aligned}$$

Moreover,

$$\sum_{\{u,v\}\in E} |w_u + w_v|^2 \leq 2\sum_{\{u,v\}\in E} (w_u^2 + w_v^2) = 2d\sum_v w_v^2.$$

We arrive at

$$\frac{\sum_{\{u,v\}\in E} |y_u - y_v|}{d\sum_v |y_v|} = \frac{\sum_{\{u,v\}\in E} |w_u^2 - w_v^2|}{d\sum_v w_v^2} \leq \sqrt{\frac{\sum_{\{u,v\}\in E} |w_u - w_v|^2}{d\sum_v w_v^2}} \leq \lambda_2.$$

□

6. Gaussian log-Sobolev inequality

6.1. **Tensorization of general LSI.** We say that a probability measure μ on a metric space X satisfies the LSI with constant C if for any Lipschitz $f : \mathbb{R}^n \to \mathbb{R}$ we have

$$\mathrm{Ent}_\mu(f^2) \leq C \int_{\mathbb{R}^n} |\nabla f|^2 \mathrm{d}\mu, \tag{4}$$

where ∇ is some notion of gradient. We have already seen that γ_n satisfies (4) with constant $C = 2$ and with the standard Euclidean gradient. We will provide a certain generalization of this fact. Before that, we prove a tensorization property of LSI.

Lemma 9. Let $(X_i, d_i, \mu_i)_{i=1,\ldots,n}$ be metric probability spaces equipped with some notions of gradient $\nabla_1, \ldots, \nabla_n$. Take $X = X_1 \times \ldots \times X_n$, $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ and assume that X is equipped with a gradient $|\nabla f|^2 = \sum_{i=1}^n |\nabla_i f|^2$. Suppose μ_i satisfies log-Sobolev inequality with constant C_i. Then the measure μ on X satisfies log-Sobolev inequality with constant $C = \max_{1\leq i\leq n} C_i$.

To prove Lemma 9 we need the following sub-additivity property of the entropy.

Lemma 10. Let $\mu_1, \dots, \mu_n$ be probability measures on $X_1, \dots, X_n$. Take the measure $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ on $X = X_1 \times \ldots \times X_n$. Then for $f : X \to (0, \infty)$ we have

$$\mathrm{Ent}_\mu(f) \le \sum_{i=1}^n \int \mathrm{Ent}_{\mu_i}(f) \, \mathrm{d}\mu.$$

Here $\mathrm{Ent}_{\mu_i}(f)$ is the entropy of the function $X_i \ni x_i \mapsto f(x_1, \dots, x_i, \dots, x_n)$, where variables other than x_i are fixed.

Proof. Let $g : X \to \mathbb{R}$ be such that $\int_X g \, \mathrm{d}\mu \le 1$. Take

$$g^i(x_1, \dots, x_n) = \ln \left(\frac{\int e^{g(x_1,\dots,x_n) \mathrm{d}_{\mu_1(x_1)} \dots \mathrm{d}_{\mu_{i-1}(x_{i-1})}}}{\int e^{g(x_1,\dots,x_n) \mathrm{d}_{\mu_1(x_1)} \dots \mathrm{d}_{\mu_i(x_i)}}} \right).$$

We have

$$\sum_{i=1}^n g^i = \ln(e^g) - \ln \left(\int e^g \, \mathrm{d}\mu \right) \ge g.$$

Note that

$$\int e^{g^i} \, \mathrm{d}\mu_i = \int \frac{\int e^g \mathrm{d}_{\mu_1} \dots \mathrm{d}_{\mu_{i-1}}}{\int e^g \, \mathrm{d}\mathrm{d}_{\mu_1} \dots \mathrm{d}_{\mu_i}} \, \mathrm{d}\mu_i = 1.$$

Hence,

$$\int fg \, \mathrm{d}\mu \le \sum_{i=1}^n \int fg^i \, \mathrm{d}\mu = \sum_{i=1}^n \int \int fg^i \, \mathrm{d}\mu_i \, \mathrm{d}\mu \le \sum_{i=1}^n \int \mathrm{Ent}_{\mu_i}(f) \, \mathrm{d}\mu.$$

We finish the proof by taking supremum over all functions g with $\int e^g \, \mathrm{d}\mu \le 1$. ☐

Proof of Lemma 9. We have

$$\mathrm{Ent}_\mu(f^2) \le \sum_{i=1}^n \int \mathrm{Ent}_{\mu_i}(f^2) \, \mathrm{d}\mu \le \sum_{i=1}^n C_i \int \int |\nabla_i f|^2 \, \mathrm{d}\mu_i \, \mathrm{d}\mu \le C \int |\nabla f|^2 \, \mathrm{d}\mu.$$

☐

6.2. **LSI on the discrete cube.** Consider the discrete cube $\{-1, 1\}^n$ equipped with the product measure $\mu_n = \left(\frac{1}{2}\delta_{\{-1\}} + \frac{1}{2}\delta_{\{1\}}\right)^{\otimes n}$. For $x = (x_1, \dots, x_n) \in \{-1, 1\}^n$ take $\sigma_i(x) = (x_1, \dots, x_{i-1}, -x_i, x_{i+1}, \dots, x_n)$. And define the **$i$th gradient** by

$$(\nabla_i f)(x) = \frac{f(x) - f(\sigma_i(x))}{2}.$$

Then the full gradient is defined via $|\nabla f|^2 = \sum_{i=1}^n |\nabla_i f|^2$. We now prove the LSI for the discrete cube $\{-1, 1\}^n$.

Theorem 12. Let $f : \{-1, 1\}^n \to (0, \infty)$. Then

$$\mathrm{Ent}_{\mu_n}(f^2) \le 2 \int |\nabla f|^2 \, \mathrm{d}\mu_n.$$

Proof. Because of the tensorization property of log-Sobolev inequality it suffices to prove the theorem in the case $n = 1$. By homogenity we can assume that $\int f^2 \, \mathrm{d}\mu = (f(1)^2 + f(-1)^2)/2 = 1$. Clearly, there exists $t \in [-1, 1]$ such that $f(1)^2 = 1 + t$, $f(-1)^2 = 1 - t$. We have $||f(1)| - |f(-1)|| \le |f(1) - f(-1)|$, therefore we can assume that $f \ge 0$. Hence

$$|\nabla f|^2 = \frac{1}{4}\left(\sqrt{1+t} - \sqrt{1-t}\right)^2 = \frac{1}{2} - \frac{1}{2}\sqrt{1-t^2}.$$

We also have

$$\mathrm{Ent}_\mu(f^2) = \frac{1+t}{2}\ln(1+t) + \frac{1-t}{2}\ln(1-t).$$

We would like to prove

$$1 - \sqrt{1-t^2} \ge \frac{1+t}{2}\ln(1+t) + \frac{1-t}{2}\ln(1-t).$$

Define

$$\alpha(t) = 1 - \sqrt{1-t^2} - \frac{1+t}{2}\ln(1+t) - \frac{1-t}{2}\ln(1-t).$$

The function α is even, therefore it suffices to prove $\alpha(t) \ge 0$ for $t \ge 0$. Note that $f(0) = 0$. It suffices to prove that

$$\alpha'(t) = \frac{t}{\sqrt{1-t^2}} - \frac{1}{2}\ln(1+t) + \frac{1}{2}\ln(1-t) \ge 0.$$

Again $f'(0) = 0$ and it suffices to observe that

$$\begin{aligned}
\alpha''(t) &= \frac{\sqrt{1-t^2} + \frac{t^2}{\sqrt{1-t^2}}}{1-t^2} - \frac{1}{2}\frac{1}{1+t} - \frac{1}{2}\frac{1}{1-t} \\
&= \frac{1}{1-t^2}\left(\frac{t^2}{\sqrt{1-t^2}} - \sqrt{1-t^2} - 1\right) = \frac{1}{1-t^2}\left(\frac{t^2}{\sqrt{1-t^2}} - \frac{t^2}{1+\sqrt{1-t^2}}\right) \ge 0.
\end{aligned}$$

□

6.3. **From the cube to Gaussian space.** We show that Theorem 12 indeed generalizes the Gaussian LSI. Let γ_1 be the one dimensional standard Gaussian measure and let $f : \mathbb{R} \to \mathbb{R}$ be a bounded function with bounded first and second derivatives. Define $f_n : \{-1, 1\}^n \to \mathbb{R}$ by

$$f_n(x_1, \ldots, x_n) = f\left(\frac{x_1 + \ldots + x_n}{\sqrt{n}}\right).$$

Note that by the Central Limit Theorem we have

$$\lim_{n\to\infty} \int f_n \, \mathrm{d}\mu^n = \int f \, \mathrm{d}\gamma_1.$$

Moreover,

$$\begin{aligned}
|\nabla f_n|^2(x) &= \frac{1}{4}\sum_{i=1}^n \left(f\left(\frac{x_1 + \ldots + x_n}{\sqrt{n}}\right) - f\left(\frac{x_1 + \ldots + x_n}{\sqrt{n}} - \frac{2x_i}{\sqrt{n}}\right)\right)^2 \\
&= \frac{1}{4}\sum_{i=1}^n \left|f'\left(\frac{x_1 + \ldots + x_n}{\sqrt{n}}\right)\right|^2 \frac{4x_i^2}{n} + O(1/n) \\
&= \left|f'\left(\frac{x_1 + \ldots + x_n}{\sqrt{n}}\right)\right|^2 + O(1/n).
\end{aligned}$$

Therefore,

$$\lim_{n\to\infty}\int_{\{-1,1\}^n} |\nabla f_n|^2 \mathrm{d}\mu_n = \int_{\mathbb{R}} |f'|^2 \mathrm{d}\gamma_1.$$

Thus, passing to the limit in $\mathrm{Ent}_{\mu_n}(f^2) \le 2\int |\nabla f|^2 \, \mathrm{d}\mu_n$ we get LSI for γ_1. Tensorization yields LSI for γ_n.

6.4. **Gaussian concentration of measure.**

7. Information theory

7.1. **...** The logarithmic Sobolev inequality (LSI) has been introduced [1] by L. Gross. It states that the standard Gaussian measure γ_n on $\mathbb{R}^n$, i.e. the probability measure with density $\varphi_n(x) = (2\pi)^{-n/2}\exp(-|x|^2/2)$, where $\|\cdot\|$ is the standard Euclidean norm, satisfies the inequality

$$\int_{\mathbb{R}^n} f^2 \ln(f^2)\mathrm{d}\gamma_n - \left(\int_{\mathbb{R}^n} f^2 \mathrm{d}\gamma_n\right)\ln\left(\int_{\mathbb{R}^n} f^2\mathrm{d}\gamma_n\right) \le 2\int_{\mathbb{R}^n} |\nabla f|^2 \mathrm{d}\gamma_n, \tag{5}$$

for every function $f : \mathbb{R}^n$ with $\int_{\mathbb{R}^n} f^2\ln^+(f^2) < \infty$. Here we adopt the standard notation $g^+ = \max\{g,0\}$. One can write (5) using the notion of entropy,

$$\mathrm{Ent}_\mu(f) = \int_{\mathbb{R}^n} f\ln(f)\mathrm{d}\mu - \left(\int_{\mathbb{R}^n} f\mathrm{d}\mu\right)\ln\left(\int_{\mathbb{R}^n} f\mathrm{d}\mu\right). \tag{6}$$

Thus, the log-Sobolev inequality read as

$$\mathrm{Ent}_{\gamma_n}(f^2) \le 2\int_{\mathbb{R}^n} |\nabla f|^2 \mathrm{d}\gamma_n. \tag{7}$$

This inequality has several equivalent formulations. An easy equivalence is a consequence of the homogeneity of both sides under scaling $g \to \lambda g$. Indeed, it is easy to see that for any probability measure μ we have $\mathrm{Ent}_\mu(\lambda g) = \lambda\,\mathrm{Ent}_\mu(g)$. Therefore, in the above inequality we can always assume that $\int f^2 \mathrm{d}\gamma_n = 1$. Then $g = f^2$ is the density of a certain probability measure. We have $|\nabla g|^2 = 4f^2|\nabla f|^2$. As a consequence (7) is implied by

$$\int_{\mathbb{R}^n} g\ln g \; \mathrm{d}\gamma_n \le \frac{1}{2}\int_{\mathbb{R}^n} \frac{|\nabla g|^2}{g}\mathrm{d}\gamma_n, \qquad g \ge 0, \ \int g\mathrm{d}\gamma_n = 1. \tag{8}$$

On the other hand it is easy to show that (7) implies (8). Indeed, it suffices to assume that $g > 0$ and take $f = \sqrt{g}$.

The aim of our next section is to get read of the measure γ_n in the above formulations and thus express the log-Sobolev inequality in terms of the so-called Shannon entropy and Fisher information. These are the main quantities studied in the information theory.

7.2. **From LSI to information theory.** Let us come back to the inequality (7) and take

$$f(x)^2 = (2\pi)^{n/2}e^{|x|^2/2}g(ax), \qquad \text{with } a > 0, \ g \ge 0, \ \int g(x)\mathrm{d}x = 1.$$

Note that

$$f(x)^2\mathrm{d}\gamma_n(x) = g(ax)\mathrm{d}x, \qquad 2f(x)\nabla f(x) = (2\pi)^{n/2}e^{|x|^2/2}(a\nabla g(ax) + xg(ax)).$$

Therefore,

$$|\nabla f(x)|^2 \mathrm{d}\gamma_n(x) = \frac{1}{4} \cdot \frac{(2\pi)^n e^{|x|^2} (a\nabla g(ax) + xg(ax))^2}{(2\pi)^{n/2} e^{|x|^2/2} g(ax)} \cdot \frac{1}{(2\pi)^{n/2}} e^{-|x|^2/2} \mathrm{d}x$$
$$= \frac{1}{4} \frac{(a\nabla g(ax) + xg(ax))^2}{g(ax)} \mathrm{d}x.$$

As a consequence, (7) is equivalent with

$$\int g(ax) \ln\left((2\pi)^{n/2} e^{|x|^2/2} g(ax)\right) \mathrm{d}x - \left(\int g(ax)\mathrm{d}x\right) \ln\left(\int g(ax)\mathrm{d}x\right)$$
$$\leq \frac{1}{2} \int \frac{(a\nabla g(ax) + xg(ax))^2}{g(ax)} \mathrm{d}x.$$

Changing variables ($y = ax$) we get

$$\frac{1}{a^n} \int g(y) \ln\left((2\pi)^{n/2} e^{|y|^2/2a^2} g(y)\right) \mathrm{d}y - \left(\frac{1}{a^n} \int g(y)\mathrm{d}y\right) \ln\left(\frac{1}{a^n} \int g(y)\mathrm{d}y\right)$$
$$\leq \frac{1}{2a^n} \int \frac{(a\nabla g(y) + \frac{y}{a} g(y))^2}{g(y)} \mathrm{d}y.$$

Multiplying both sides by a^n and using $\int g(y)\mathrm{d}y = 1$ gives

$$\ln((2\pi)^{n/2}) + \int g(y) \frac{|y|^2}{2a^2} \mathrm{d}y + \int g(x) \ln g(x) \mathrm{d}x + n \ln a \leq$$
$$\frac{1}{2} \int \left(a^2 \frac{|\nabla g(y)|^2}{g(y)} + y \cdot \nabla g(y) + g(y) \frac{|y|^2}{2a^2}\right) \mathrm{d}y.$$

Let us define the **Shannon entropy**, **Fisher information** and **entropy power** of a probability density g,

$$\mathcal{S}(g) = -\int g(y) \ln g(y) \mathrm{d}y, \qquad \mathcal{I}(g) = \int \frac{|\nabla g(y)|^2}{g(y)} \mathrm{d}y \qquad \mathcal{N}(g) = \frac{1}{2\pi e} \exp\left(\frac{2}{n} \mathcal{S}(g)\right).$$

Integrating by parts we get that

$$\int y \cdot \nabla g(y) \mathrm{d}y = \int \nabla(\frac{1}{2}|y|^2) \cdot \nabla g(y) \mathrm{d}y = -\int \Delta(\frac{1}{2}|y|^2) g(y) = -n.$$

Thus, we can further rewrite the above inequality in the form of

$$\ln((2\pi)^{n/2}) - \mathcal{S}(g) + n \ln a \leq \frac{1}{2} a^2 \mathcal{I}(g) - n.$$

Equivalently,

$$\frac{n}{2} \ln(2\pi) - \mathcal{S}(g) \leq \inf_a \left(\frac{1}{2} a^2 \mathcal{I}(g) - n - n \ln a\right) = -\frac{n}{2} - \frac{n}{2} \ln\left(\frac{n}{\mathcal{I}(g)}\right).$$

After multiplying by $2/n$ and taking the exponent one gets

$$2\pi \exp\left(-\frac{2}{n} \mathcal{S}(g)\right) \leq e^{-1} \frac{\mathcal{I}(g)}{n}.$$

This is

$$\mathcal{N}(g)\mathcal{I}(g) \geq n. \tag{9}$$

Thus, we have written the log-Sobolev inequality in terms of information theoretic quantities.

7.3. **Heat semigroup.** Up to now we did not yet prove the Gross log-Sobolev inequality. Before we do this we need to introduce the notion of heat semigroup of operators $(\mathcal{P}_t)_{t\geq 0}$,

$$(\mathcal{P}_t f)(x) = \int_{\mathbb{R}^n} f\left(x + y\sqrt{t}\right) \mathrm{d}\gamma_n(y).$$

We leave the following easy fact as an exercise for the reader.

Fact 10. The family $(\mathcal{P}_t)_{t\geq 0}$ is a Markov semigroup of operators, namely

- $\mathcal{P}_t(\mathbf{1}) = \mathbf{1}$, $t \geq 0$,
- $f \geq 0$ a.s. $\implies$ $\mathcal{P}_t(f) \geq 0$, a.s.,
- $\mathcal{P}_{t+s} = \mathcal{P}_t \circ \mathcal{P}_s$, $\mathcal{P}_0 = \mathrm{Id}$.

Moreover, $\mathcal{P}_t(f)$ solves the heat equation $\frac{\partial u}{\partial t} = \frac{1}{2}\Delta u$ with an initial condition $u_0 = f$. In other words, we have $\frac{\partial}{\partial t}\mathcal{P}_t(f) = \frac{1}{2}\Delta(P_t(f)) = \frac{1}{2}\mathcal{P}_t(\Delta f)$.

We prove the following lemma.

Lemma 11. Let $(\mathcal{P}_t)_{t\geq 0}$ be the heat semigroup. Then

$$\mathcal{P}_t(f \ln f) - \mathcal{P}_t(f)\ln(\mathcal{P}_t(f)) = \frac{1}{2}\int_0^t \mathcal{P}_s\left(\frac{|\nabla \mathcal{P}_{t-s}(f)|^2}{\mathcal{P}_{t-s}(f)}\right) \mathrm{d}s.$$

Proof. We have

$$\begin{aligned}
\mathcal{P}_t(f \ln f) - \mathcal{P}_t(f)\ln(\mathcal{P}_t(f)) &= \int_0^t \frac{\partial}{\partial s}\left[\mathcal{P}_s\left(\mathcal{P}_{t-s}(f)\ln(\mathcal{P}_{t-s}(f))\right)\right] \mathrm{d}s \\
&= \int_0^t \Big(\frac{\partial}{\partial s_1}\left[\mathcal{P}_{s_1}\left(\mathcal{P}_{t-s_2}(f)\ln(\mathcal{P}_{t-s_2}(f))\right)\right]\Big|_{s_1=s_2=s} \\
&\qquad + \frac{\partial}{\partial s_2}\left[\mathcal{P}_{s_1}\left(\mathcal{P}_{t-s_2}(f)\ln(\mathcal{P}_{t-s_2}(f))\right)\right]\Big|_{s_1=s_2=s}\Big) \mathrm{d}s \\
&= \frac{1}{2}\int_0^t \mathcal{P}_s\left[\Delta\left(\mathcal{P}_{t-s}(f)\ln(\mathcal{P}_{t-s}(f))\right)\right] \mathrm{d}s + \int_0^t \mathcal{P}_s\left[\frac{\partial}{\partial s}\left(\mathcal{P}_{t-s}(f)\ln(\mathcal{P}_{t-s}(f))\right)\right] \mathrm{d}s.
\end{aligned}$$

Note that

$$\begin{aligned}
\Delta(g \ln g) &= \sum_i (g\ln g)_{x_i x_i} = \sum_i (g_{x_i}(1+\ln g))_{x_i} = (\Delta g)(1+\ln g) + \sum_i \frac{g_{x_i}^2}{g} \\
&= (\Delta g)(1+\ln g) + \frac{|\nabla g|^2}{g}.
\end{aligned}$$

Applying this with $g = \mathcal{P}_{t-s}(f)$ we get

$$\begin{aligned}
\mathcal{P}_t(f \ln f) - \mathcal{P}_t(f)\ln(\mathcal{P}_t(f)) &= \frac{1}{2}\int_0^t \mathcal{P}_s\left[\Delta(\mathcal{P}_{t-s}(f))(1+\ln(\mathcal{P}_{t-s}(f))) + \frac{|\nabla \mathcal{P}_{t-s}(f)|^2}{\mathcal{P}_{t-s}(f)}\right] \mathrm{d}s \\
&\qquad - \frac{1}{2}\int_0^t \mathcal{P}_s\left[(1+\ln(\mathcal{P}_{t-s}(f)))\Delta(\mathcal{P}_{t-s}(f))\right] \mathrm{d}s \\
&= \frac{1}{2}\int_0^t \mathcal{P}_s\left(\frac{|\nabla \mathcal{P}_{t-s}(f)|^2}{\mathcal{P}_{t-s}(f)}\right) \mathrm{d}s.
\end{aligned}$$

$\square$

7.4. **First proof of LSI.** Let us first prove that $|\mathcal{P}_s(\nabla f)| \le \mathcal{P}_s(|\nabla f|)$, where we adopt the notation $\mathcal{P}_s(\nabla f) = (\mathcal{P}_s(f_{x_1}), \ldots, \mathcal{P}_s(f_{x_n}))$. Indeed, for any vector $a \in \mathbb{R}^n$ with $|a| = 1$ we have $\langle a, \nabla f\rangle \le |\nabla f|$. Thus, $\langle a, \mathcal{P}_s(\nabla f)\rangle = \mathcal{P}_s(\langle a, \nabla f\rangle) \le \mathcal{P}_s|\nabla f|$. Now it suffices to use the fact that $\sup_{|a|=1}\langle a, \mathcal{P}_s(\nabla f)\rangle = |\mathcal{P}_s(\nabla f)|$.

Note that from the Cauchy-Schwarz inequality we get $(\mathcal{P}_s(fg))^2 \le \mathcal{P}_s(f^2)\mathcal{P}_s(g^2)$. Thus,

$$|\nabla \mathcal{P}_{t-s}(f)|^2 = |\mathcal{P}_{t-s}(\nabla f)|^2 \le \mathcal{P}_{t-s}(|\nabla f|)^2 \le \mathcal{P}_{t-s}(f)\cdot \mathcal{P}_{t-s}\left(\frac{|\nabla f|^2}{f}\right).$$

We arrive at

$$\begin{aligned}\mathcal{P}_t(f\ln f) - \mathcal{P}_t(f)\ln(\mathcal{P}_t(f)) &= \frac{1}{2}\int_0^t \mathcal{P}_s\left(\frac{|\nabla \mathcal{P}_{t-s}(f)|^2}{\mathcal{P}_{t-s}(f)}\right)\mathrm{d}s \\ &\le \frac{1}{2}\int_0^t \mathcal{P}_s\mathcal{P}_{t-s}\left(\frac{|\nabla f|^2}{f}\right)\mathrm{d}s = \frac{t}{2}\mathcal{P}_t\left(\frac{|\nabla f|^2}{f}\right).\end{aligned}$$

This is a poinwise inequality valid for every $x \in \mathbb{R}^n$ and $t \ge 0$. Taking $t = 1$ and $x = 0$ one gets

$$\int_{\mathbb{R}^n} f\ln f\mathrm{d}\gamma_n - \left(\int_{\mathbb{R}^n} f\mathrm{d}\gamma_n\right)\ln\left(\int_{\mathbb{R}^n} f\mathrm{d}\gamma_n\right) \le \frac{1}{2}\int_{\mathbb{R}^n}\frac{|\nabla f|^2}{f}\mathrm{d}\gamma_n,$$

since $\mathcal{P}_1(g)(0) = \int_{\mathbb{R}^n} g\mathrm{d}\gamma_n$. Assuming $\int_{\mathbb{R}^n} f\mathrm{d}\gamma_n = 1$, we get (8).

7.5. **Reverse LSI.** Observe that

$$[\mathcal{P}_t(f_{x_i})]^2 = [\mathcal{P}_s(\mathcal{P}_{t-s}(f_{x_i}))]^2 \le [\mathcal{P}_s(\mathcal{P}_{t-s}(f))]\cdot\left[\mathcal{P}_s\left(\frac{[\mathcal{P}_{t-s}(f_{x_i})]^2}{\mathcal{P}_{t-s}(f)}.\right)\right]$$

Summing over i we get

$$|\mathcal{P}_t(\nabla f)|^2 \le [\mathcal{P}_s(\mathcal{P}_{t-s}(f))]\cdot\left[\mathcal{P}_s\left(\frac{[\mathcal{P}_{t-s}(\nabla f)]^2}{\mathcal{P}_{t-s}(f)}\right)\right].$$

Thus, using Lemma 11, we get

$$\begin{aligned}\mathcal{P}_t(f\ln f) - \mathcal{P}_t(f)\ln(\mathcal{P}_t(f)) &= \frac{1}{2}\int_0^t \mathcal{P}_s\left(\frac{|\nabla \mathcal{P}_{t-s}(f)|^2}{\mathcal{P}_{t-s}(f)}\right)\mathrm{d}s \\ &\ge \frac{1}{2}\int_0^t \frac{|\mathcal{P}_t(\nabla f)|^2}{\mathcal{P}_t(f)}\mathrm{d}s = \frac{t}{2}\frac{|\mathcal{P}_t(\nabla f)|^2}{\mathcal{P}_t(f)}.\end{aligned}$$

Again taking $x = 0$, $t = 1$ and assuming $\int_{\mathbb{R}^n} f\mathrm{d}\gamma_n = 1$, one gets

$$\int_{\mathbb{R}^n} f\ln f\mathrm{d}\gamma_n \ge \frac{1}{2}\frac{\int_{\mathbb{R}^n}|\nabla f|^2\mathrm{d}\gamma_n}{\int_{\mathbb{R}^n} f\mathrm{d}\gamma_n}. \tag{10}$$

This is called the reverse log-Sobolev inequality.

Using the ideas from the Section 7.2 one can show that the reverse LSI is equivalent with the inequality

$$\mathcal{N}(g) \le \frac{\mathrm{Tr}K(g)}{n}, \qquad g \ge 0, \int_{\mathbb{R}^n} g(x)\mathrm{d}x = 1,$$

which is further equivalent with

$$\mathcal{N}(g) \le |K(g)|^{1/n}, \qquad (K(g))_{i,j} = \int x_i x_j g(x)\mathrm{d}x, \tag{11}$$

where $|\cdot|$ denotes the determinant. The matrix $K(g)$ is called the covariance matrix of a random variable X with density g.

Let us give a direct proof of (11). We need the following lemma

Lemma 12. Let K be a symmetric positive definite matrix. Then

$$\varphi_K(x) = \frac{1}{(2\pi)^{n/2}|K|^{1/2}} \exp(-\frac{1}{2}x^T K^{-1}x)$$

is the Gaussian density with covariance matrix K. Moreover,

$$\mathcal{S}(\varphi_K) = \frac{1}{2}\ln\left((2\pi e)^n|K|\right), \qquad \mathcal{N}(\varphi_K) = |K|^{1/n}.$$

Proof. The first part is standard. Let us only compute the entropy,

$$\mathcal{S}(\varphi_K) = -\int \varphi_K \ln \varphi_K = \ln((2\pi)^{n/2}|K|^{1/2}) + \frac{1}{2}\int \varphi_K x^T K^{-1}x.$$

Let $(X_1, \ldots, X_n)$ be the random vector with density φ_K. We have

$$\int \varphi_K x^T K^{-1}x = \mathbb{E}X^T K^{-1}X = \sum_{i,j}\mathbb{E}X_i(K^{-1})_{ij}X_j = \sum_{i,j}K_{ij}(K^{-1})_{ij}$$
$$= \sum_{i,j}K_{ji}(K^{-1})_{ij} = \sum_j (KK^{-1})_{jj} = n.$$

We get

$$\mathcal{S}(\varphi_K) = \ln((2\pi)^{n/2}|K|^{1/2}) + \frac{n}{2} = \frac{n}{2}\ln\left(2\pi e|K|^{1/n}\right) = .$$

Thus,

$$\mathcal{N}(\varphi_K) = \frac{1}{2\pi e}\exp\left(\frac{2}{n}\mathcal{S}(\varphi_K)\right) = |K|^{1/n}.$$

□

To prove the inequality 11 it suffices to establish the following fact.

Fact 11. Let g be a probability density and let φ_g be the Gaussian density with $K(g) = K(\varphi_g)$. Then $\mathcal{S}(g) \le \mathcal{S}(\varphi_g)$.

Proof. Let us define the Kulback-Liebre dirergence (or, in other word, the relative entropy) for the probability densities f, g,

$$D(f\|g) = \int f \ln\left(\frac{f}{g}\right).$$

We first prove that $D(f\|g) \ge 0$. Recall the famous inequality $\ln(1+x) \le x$, $x > -1$. This gives

$$-D(f\|g) = -\int f\ln\left(\frac{f}{g}\right) = \int f\ln\left(\frac{g}{f}\right) \le \int f\left(\frac{g}{f} - 1\right) = \int f - \int g = 0.$$

The inequality $D(g\|\varphi_g) \ge 0$ gives

$$\mathcal{S}(g) = -\int g\ln g \le -\int g\ln\varphi_g = -\int \varphi_g\ln\varphi_g = \mathcal{S}(\varphi_g).$$

□

7.6. **de Bruijn's identity.**

Proposition 10. Let X be a random vector in $\mathbb{R}^n$ and let G be a standard Gaussian in $\mathbb{R}^n$. Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{S}(X+\sqrt{t}G) = \frac{1}{2}\mathcal{I}(X+\sqrt{t}Z).$$

In other words the evolution $\mathcal{P}_t(f)$, where f is the density of X, satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{S}(\mathcal{P}_t(f)) = \frac{1}{2}\mathcal{I}(\mathcal{P}_t(f)).$$

Proof. Note that $\mathcal{P}_t(f)$ satisfies $\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{P}_t(f) = \Delta\mathcal{P}_t(f)$. Thus,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{S}(\mathcal{P}_t(f)) &= -\frac{\mathrm{d}}{\mathrm{d}t}\int \mathcal{P}_t(f)\ln\mathcal{P}_t(f) = -\int \frac{\mathrm{d}\mathcal{P}_t(f)}{\mathrm{d}t}\left(1+\ln\mathcal{P}_t(f)\right) \\
&= -\frac{\mathrm{d}}{\mathrm{d}t}\int \mathcal{P}_t(f) - \frac{1}{2}\int \Delta\mathcal{P}_t(f)\ln\mathcal{P}_t(f) \\
&= -\frac{\mathrm{d}}{\mathrm{d}t}(1) + \frac{1}{2}\int \frac{|\nabla\mathcal{P}_t(f)|^2}{\mathcal{P}_t(f)} = \frac{1}{2}\mathcal{I}(\mathcal{P}_t(f)).
\end{aligned}$$

□

7.7. **Entropy power inequality.** We are now ready to state and prove three equivalent formulation of the famous entropy power inequality.

Proposition 11. Let X, Y be independent random vectors on $\mathbb{R}^n$. The following conditions are equivalent

(a) We have $\mathcal{N}(X+Y) \geq N(G_X+G_Y)$, where G_X, G_Y are independent Gaussian random vectors with **proportional covariance matrices** and $\mathcal{S}(X) = \mathcal{S}(G_X)$, $\mathcal{S}(Y) = \mathcal{S}(G_Y)$,

(b) $\mathcal{N}(X+Y) \geq \mathcal{N}(X) + \mathcal{N}(Y)$,

Proof. We first show that (a) implies (b). Note that $K(G_X+G_Y) = K(G_X)+K(G_Y)$. Since the matrices $K(G_X)$ and $K(G_Y)$ are proportional (say, $K(G_Y) = aK(G_X)$), we have

$$\begin{aligned}
|K(G_X+G_Y)|^{1/n} &= |K(G_X)+K(G_Y)|^{1/n} = |(1+a)K(G_X)|^{1/n} = (1+a)|K(G_X)|^{1/n} \\
&= |K(G_X)|^{1/n} + |aK(G_X)|^{1/n} = |K(G_X)|^{1/n} + |K(G_Y)|^{1/n}.
\end{aligned}$$

Thus, from Lemma 12 we get

$$\begin{aligned}
\mathcal{N}(X+Y) \geq \mathcal{N}(G_X+G_Y) &= |K(G_X+G_Y)|^{1/n} = |K(G_X)|^{1/n} + |K(G_Y)|^{1/n} \\
&= \mathcal{N}(G_X) + \mathcal{N}(G_Y) = \mathcal{N}(X) + \mathcal{N}(Y).
\end{aligned}$$

Similarly, (b) implies (a) since

$$\mathcal{N}(X+Y) \geq \mathcal{N}(X) + \mathcal{N}(Y) = \mathcal{N}(G_X+G_Y).$$

□

To prove the entropy power inequality it suffices to establish the following proposition.

Proposition 12. For any pair of independent random vectors X, Y on $\mathbb{R}^n$ and any $\lambda \in [0,1]$ we have

$$\mathcal{S}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) \geq \lambda\mathcal{S}(X) + (1-\lambda)\mathcal{S}(Y).$$

We first show that Proposition 12 implies inequality (b) from the Proposition 11. Note that

$$\mathcal{S}(X+Y) = \mathcal{S}\left(\sqrt{\lambda}\cdot\frac{X}{\sqrt{\lambda}} + \sqrt{1-\lambda}\cdot\frac{Y}{\sqrt{1-\lambda}}\right) \geq \lambda\mathcal{S}\left(\frac{X}{\sqrt{\lambda}}\right) + (1-\lambda)\mathcal{S}\left(\frac{Y}{\sqrt{1-\lambda}}\right)$$
$$= \lambda\mathcal{S}(X) + (1-\lambda)\mathcal{S}(Y) - \frac{n}{2}\left[\lambda\ln\lambda + (1-\lambda)\ln(1-\lambda)\right].$$

We have used the fact that

$$\mathcal{S}(aX) = \mathcal{S}(X) + n\ln a.$$

The optimal choice of λ is $\lambda = \mathcal{N}(X)/(\mathcal{N}(X)+\mathcal{N}(Y))$. This gives

$$\mathcal{S}(X+Y) \geq \frac{1}{\mathcal{N}(X)+\mathcal{N}(Y)}\Bigg[\mathcal{N}(X)\mathcal{S}(X) + \mathcal{N}(Y)\mathcal{S}(Y)$$
$$- \frac{n}{2}\mathcal{N}(X)\ln\left(\frac{\exp(\frac{2}{n}\mathcal{S}(X))}{\exp(\frac{2}{n}\mathcal{S}(X)) + \exp(\frac{2}{n}\mathcal{S}(Y))}\right) - \frac{n}{2}\mathcal{N}(Y)\ln\left(\frac{\exp(\frac{2}{n}\mathcal{S}(Y))}{\exp(\frac{2}{n}\mathcal{S}(X)) + \exp(\frac{2}{n}\mathcal{S}(Y))}\right)\Bigg]$$
$$= \frac{n}{2}\cdot\frac{1}{\mathcal{N}(X)+\mathcal{N}(Y)}(\mathcal{N}(X)+\mathcal{N}(Y))\ln\left(\exp\left(\frac{2}{n}\mathcal{S}(X)\right) + \exp\left(\frac{2}{n}\mathcal{S}(Y)\right)\right).$$

Equivalently,

$$\frac{2}{n}\mathcal{S}(X+Y) \geq \ln\left(\exp\left(\frac{2}{n}\mathcal{S}(X)\right) + \exp\left(\frac{2}{n}\mathcal{S}(Y)\right)\right).$$

Taking exponent of both sides gives $\mathcal{N}(X+Y) \geq \mathcal{N}(X) + \mathcal{N}(Y)$.

To prove Proposition 12 we need a corresponding fact for Fisher information, called the Blachman-Stam inequality.

Proposition 13. Let X, Y be independent random vectors and let $\lambda \in [0,1]$. Then

$$\mathcal{I}(X+Y) \leq \lambda^2\mathcal{I}(X) + (1-\lambda)^2\mathcal{I}(Y). \tag{12}$$

Moreover,

$$\frac{1}{\mathcal{I}(X+Y)} \geq \frac{1}{\mathcal{I}(X)} + \frac{1}{\mathcal{I}(Y)}. \tag{13}$$

We postpone its proof till the next section and show how it implies Proposition 12.

Proof of Proposition 12. Let G_X and G_Y be two independent standard Gaussian random vectors in $\mathbb{R}^n$. Let us define

$$X_t = \sqrt{t}X + \sqrt{1-t}G_X, \qquad Y_t = \sqrt{t}Y + \sqrt{1-t}G_Y.$$

Moreover, let us take

$$V_t = \sqrt{\lambda}X_t + \sqrt{1-\lambda}Y_t.$$

Note that

$$V_t = \sqrt{t}(\sqrt{\lambda}X + \sqrt{1-\lambda}Y) + \sqrt{1-t}(\sqrt{\lambda}G_X + \sqrt{1-\lambda}G_Y) = \sqrt{t}V_1 + \sqrt{1-t}V_0,$$

Take

$$\psi(t) = \mathcal{S}(V_t) - \lambda\mathcal{S}(X_t) - (1-\lambda)\mathcal{S}(Y_t).$$

We have $X_1 = X$, $Y_1 = Y$ and $V_1 = \sqrt{\lambda}X + \sqrt{1-\lambda}Y$. Thus, our goal is to prove that $\psi(1) \geq 0$. Since $X_0 = G_X$, $Y_0 = G_Y$ and $V_0 = \sqrt{\lambda}G_X + \sqrt{1-\lambda}G_Y \sim G_X$, we get $\psi(0) = 0$. As a consequence, we are to prove that $\psi(1) \geq \psi(0)$.

To this end we show that $\psi'(t) \geq 0$ on $[0,1]$. Due to the scaling $\mathcal{S}(aX) = \mathcal{S}(X) + n\ln(|a|)$, we have

$$\psi(t) = \mathcal{S}\left(V_1 + \sqrt{\frac{1-t}{t}}V_0\right) - \lambda\mathcal{S}\left(X_1 + \sqrt{\frac{1-t}{t}}X_0\right) - (1-\lambda)\mathcal{S}\left(Y_1 + \sqrt{\frac{1-t}{t}}Y_0\right).$$

From de Bruijn's identity we get

$$-2t^2\psi'(t) = \mathcal{I}\left(V_1 + \sqrt{\frac{1-t}{t}}V_0\right) - \lambda\mathcal{I}\left(X_1 + \sqrt{\frac{1-t}{t}}X_0\right) - (1-\lambda)\mathcal{I}\left(Y_1 + \sqrt{\frac{1-t}{t}}Y_0\right).$$

Using $\mathcal{I}(aX) = a^{-2}\mathcal{I}(X)$ we get

$$\begin{aligned}
2t\psi'(t) &= -\mathcal{I}(\sqrt{t}V_1 + \sqrt{1-t}V_0) + \lambda\mathcal{I}(\sqrt{t}X_1 + \sqrt{1-t}X_0) + (1-\lambda)\mathcal{I}(\sqrt{t}Y_1 + \sqrt{1-t}Y_0) \\
&= -\mathcal{I}(V_t) + \lambda\mathcal{I}(X_t) + (1-\lambda)\mathcal{I}(Y_t) \\
&= -\mathcal{I}(\sqrt{\lambda}X_t + \sqrt{1-\lambda}Y_t) + \lambda\mathcal{I}(X_t) + (1-\lambda)\mathcal{I}(Y_t).
\end{aligned}$$

Let $\tilde{X}_t = \sqrt{\lambda}X_t$ and $\tilde{Y}_t = \sqrt{1-\lambda}Y_t$. Then

$$2t\psi'(t) = -\mathcal{I}(\tilde{X}_t + \tilde{Y}_t) + \lambda^2\mathcal{I}(\tilde{X}_t) + (1-\lambda)^2\mathcal{I}(\tilde{Y}_t) \geq 0$$

due to Proposition 13. □

7.8. **Blachman-Stam inequality.** For a random vector X with density f let us introduce the notion of **score function**

$$\rho_X(x) = \frac{(\nabla f)(x)}{f(x)} \in \mathbb{R}^n.$$

Note that the Fisher information satisfies

$$\mathcal{I}(X) = \int \frac{|\nabla f|^2}{f} = \mathbb{E}_X|\rho_X|^2,$$

where we set $\mathbb{E}_X g$ to be the expectation of g with respect to X having density f. Note that for any $a \in \mathbb{R}$ and $b \in \mathbb{R}^n$ we have

$$\mathcal{S}(aX+b) = \mathcal{S}(X) + n\ln(|a|), \qquad \mathcal{I}(aX+b) = a^{-2}\mathcal{I}(X), \qquad \mathcal{N}(aX+b) = a^2\mathcal{N}(X). \tag{14}$$

Let us prove one simple lemma.

Lemma 13. Let X, Y be independent random vectors in $\mathbb{R}^n$. Consider $Z = X + Y$ and let ρ_X, ρ_Y, ρ_Z be the corresponding score functions. Then

$$\rho_Z(z) = \mathbb{E}[\rho_X(X)|Z=z] = \mathbb{E}[\rho_Y(Y)|Z=z].$$

Proof. Let f_X, f_Y, f_Z be the densities of X, Y, Z, respectively. Recall that[1]

$$\mathbb{E}[h(X,Y)|Z=z] = \int h(x, z-x)\frac{f_X(x)f_Y(z-x)}{f_Z(z)}\mathrm{d}x.$$

[1]Those who are not familiar with conditional expectation can treat this equality as a definition of the right hand side.

We have

$$\begin{aligned}(\nabla f_Z)(z) &= \nabla_z\left(\int f_X(x)f_Y(z-x)\mathrm{d}x\right) = \int f_X(x)\nabla_z f_Y(z-x)\mathrm{d}x \\ &= -\int f_X(x)\nabla_x f_Y(z-x)\mathrm{d}x = \int \nabla_x f_X(x) f_Y(z-x)\mathrm{d}x.\end{aligned}$$

Thus,

$$\frac{(\nabla f_Z)(z)}{f_Z(z)} = \int \frac{\nabla_x f_X(x)}{f_X(x)}\cdot\frac{f_X(x)f_Y(z-x)}{f_Z(z)}\mathrm{d}x = \mathbb{E}[\rho_X(X)|Z=z].$$

The second equality follows by symmetry. □

We are ready to prove the Blachman-Stam inequality.

Proof of Proposition 13. By Lemma 13 we have

$$\rho_Z(z) = \mathbb{E}\left[\lambda\rho_X(X)+(1-\lambda)\rho_Y(Y)|Z=z\right], \qquad \lambda\in[0,1].$$

Thus,

$$\begin{aligned}\mathcal{I}(X+Y) &= \mathbb{E}_Z[\rho_Z(Z)]^2 = \mathbb{E}_Z\left[\mathbb{E}\left[\lambda\rho_X(X)+(1-\lambda)\rho_Y(Y)|Z=z\right]^2\right] \\ &\le \mathbb{E}_Z\left[\mathbb{E}\left[(\lambda\rho_X(X)+(1-\lambda)\rho_Y(Y))^2|Z=z\right]\right] \\ &= \mathbb{E}\left(\lambda\rho_X(X)+(1-\lambda)\rho_Y(Y)\right)^2 \\ &= \lambda^2\mathcal{I}(X)+(1-\lambda)^2\mathcal{I}(Y)+2\lambda(1-\lambda)\mathbb{E}[\rho_X(X)\cdot\rho_Y(Y)].\end{aligned}$$

Here we have used the inequality

$$\mathbb{E}[h(X,Y)|Z=z]^2 \le \mathbb{E}[h(X,Y)^2|Z=z],$$

which follows from the Cauchy-Schwarz inequality and the very easy equality

$$\mathbb{E}_Z\left[\mathbb{E}[h(X,Y)|Z=z]\right] = \mathbb{E}h(X,Y).$$

Due to independence we have

$$\mathbb{E}[\rho_X(X)\rho_Y(Y)] = \mathbb{E}[\rho_X(X)]\cdot\mathbb{E}[\rho_Y(Y)] = \int\nabla f_X\cdot\int\nabla f_Y = 0\cdot 0 = 0.$$

We thus get

$$\mathcal{I}(X+Y)\le\lambda^2\mathcal{I}(X)+(1-\lambda)^2\mathcal{I}(Y).$$

Optimizing with respect to $\lambda\in[0,1]$ one gets (by taking $\lambda=\frac{\mathcal{I}(Y)}{\mathcal{I}(X)+\mathcal{I}(Y)}$)

$$\mathcal{I}(X+Y)\le\left(\frac{\mathcal{I}(Y)}{\mathcal{I}(X)+\mathcal{I}(Y)}\right)^2\mathcal{I}(X)+\left(\frac{\mathcal{I}(X)}{\mathcal{I}(X)+\mathcal{I}(Y)}\right)^2\mathcal{I}(Y) = \frac{\mathcal{I}(X)\mathcal{I}(Y)}{\mathcal{I}(X)+\mathcal{I}(Y)},$$

which is exactly

$$\frac{1}{\mathcal{I}(X+Y)}\ge\frac{1}{\mathcal{I}(X)}+\frac{1}{\mathcal{I}(Y)}.$$

□

8. Entropic Central Limit Theorem

The simplest version of the Central Limit Theorem (CLT) states that for any sequence of i.i.d. random variables $X_1, \ldots, X_n$ with mean zero and variance 1 the sequence

$$Y_n = \frac{X_1 + \ldots + X_n}{\sqrt{n}}$$

converges in distribution to the standard Gaussian random variable G. Since the random variable Y_n has variance 1, one has $\mathcal{S}(Y_n) \leq \mathcal{S}(G)$, due to Fact 11. From the EPI we deduce

$$e^{2\mathcal{S}(X_1+X_2)} \geq e^{2\mathcal{S}(X_1)} + e^{2\mathcal{S}(X_2)} = 2e^{2\mathcal{S}(X_1)}.$$

Taking the logarithm, we get

$$\mathcal{S}(X_1 + X_2) \geq \ln(\sqrt{2}) + \mathcal{S}(X_1).$$

This gives

$$\mathcal{S}(Y_1) = \mathcal{S}(X_1) \leq \mathcal{S}\left(\frac{X_1 + X_2}{\sqrt{2}}\right) = \mathcal{S}(Y_2).$$

It is therefore natural to conjecture, that the sequence $\mathcal{S}(Y_n)$ is non-decreasing. This is indeed true, due to the celebrated theorem of S. Artstein, K. Ball, F. Barthe and A. Naor.

Theorem 13. Let $X_1, \ldots, X_n$ be a sequence of i.i.d. random variables with mean zero and variance 1. Take $Y_n = (X_1 + \ldots + X_n)/\sqrt{n}$. Then the sequence $\mathcal{S}(Y_n)$ is non-decreasing.

Before we prove this theorem, we need to develop several useful tools.

8.1. **ANOVA decomposition.**
Here we prove the following lemma.

Lemma 14. Let $\mu = \mu_1 \otimes \ldots \otimes \mu_n$ be a product measure on $\mathbb{R}^n$ and let $L^2 = L^2(\mathbb{R}^n, \mu)$. For $S \subset [n]$ let us define linear subspaces

$$\mathcal{H}_S = \left\{ \phi \in L^2 \;\middle|\; \int \phi(x) \mathrm{d}\mu_j(x_j) = \phi(x) \mathbf{1}_{\{j \notin S\}} \ \forall j \in [n] \right\}.$$

Then L^2 is the orthogonal direct sum of $\mathcal{H}_S$. In particular, every $\phi \in L^2$ can be written in the form $\phi = \sum_{S \subset [n]} \phi_S$, where $\phi_S \in \mathcal{H}_S$.

Proof. For $S \subset [n]$ let us define linear operators $\mathbb{E}_S$ by

$$\mathbb{E}_S \phi = \int \phi(x_1, \ldots, x_n) \prod_{j \in S} \mathrm{d}\mu_j(x_j).$$

Moreover, let us set $\mathbb{E}_j = \mathbb{E}_{\{j\}}$. Clearly, $\mathbb{E}_1, \ldots, \mathbb{E}_n$ are commuting projection operators in L^2. We have

$$\phi = \prod_{j=1}^{n} \left[\mathbb{E}_j + (I - \mathbb{E}_j) \right] \phi = \sum_{S \subset [n]} \prod_{j \notin S} \mathbb{E}_j \prod_{j \in S} (I - \mathbb{E}_j) \phi = \sum_{S \subset [n]} \phi_S,$$

where

$$\psi_S = \mathbb{E}_{S^c} \prod_{j \in S} (I - \mathbb{E}_j) \phi = \bar{\mathbb{E}}_S \phi, \qquad \bar{\mathbb{E}}_S := \mathbb{E}_{S^c} \prod_{j \in S} (I - \mathbb{E}_j).$$

We show that $\phi_S \in \mathcal{H}_S$. Indeed, let $j_0 \in S$. Then

$$\mathbb{E}_j \phi_S = \mathbb{E}_{S^c} \prod_{j \in S, j \neq j_0} (I - \mathbb{E}_j) \mathbb{E}_j (I - \mathbb{E}_j) \phi = 0$$

since $\mathbb{E}_j(I-\mathbb{E}_j)=\mathbb{E}_j-\mathbb{E}_j^2=\mathbb{E}_j-\mathbb{E}_j=0$. If $j_0\notin S$, then $\mathbb{E}_{j_0}\mathbb{E}_{S^c}=\mathbb{E}_{S^c}$ and therefore $\mathbb{E}_{j_0}\phi_S=\phi_S$.

Finally, we prove that $\mathcal{H}_S$ are orthogonal. Suppose $S,T\subset[n]$ are such that $S\neq T$ and let $f\in\mathcal{H}_S$, $g\in\mathcal{H}_T$. There is $j\in[n]$ such that $j\in S\Delta T$, for example $j\in S$, $j\notin T$. Thus, $\mathbb{E}_jf=0$ and $\mathbb{E}_jg=g$. We arrive at

$$\mathbb{E}fg=\mathbb{E}\mathbb{E}_j(fg)=\mathbb{E}\mathbb{E}_j(f\mathbb{E}_jg)=\mathbb{E}(\mathbb{E}_jg\mathbb{E}_jf)=0.$$

$\square$

8.2. **Variance drop lemma.** We prove the following lemma.

Lemma 15. Let $\phi:\mathbb{R}^n\to\mathbb{R}$ and let $\mu=\mu_1\otimes\ldots\otimes\mu_n$ be a product measure on $\mathbb{R}^n$. Suppose that for every $j\in[n]$ the function $\phi_j(x)=\phi(x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n)$ has mean 0. Then

$$\mathbb{E}\left(\sum_{j=1}^n\phi_j\right)^2\le(n-1)\sum_{j\in[n]}\mathbb{E}\phi_j^2.$$

Proof. Let $\bar{\mathbb{E}}_S$ be operators defined in the previous section. Then

$$\phi_j=\sum_{S\subset[n]}\bar{\mathbb{E}}_S\phi_j,\qquad j=1,\ldots,n.$$

Moreover, $\bar{\mathbb{E}}_S\phi_j\in\mathcal{H}_S$. If $j\in S$ then we have $\bar{\mathbb{E}}_S\phi_j=\bar{\mathbb{E}}_S\mathbb{E}_j\phi_j=\mathbb{E}_j\bar{\mathbb{E}}_S\phi_j=0$, where the first equality follows from the fact that ϕ_j does not depend on j and the second from the fact that $\bar{\mathbb{E}}_S\in\mathcal{H}_S$. We get

$$\mathbb{E}\left(\sum_{j\in[n]}\phi_j\right)^2=\mathbb{E}\left(\sum_{S\subset[n]}\sum_{j\in[n]}\bar{\mathbb{E}}_S\phi_j\right)^2=\mathbb{E}\left(\sum_{S\subset[n]}\sum_{j\notin S}\bar{\mathbb{E}}_S\phi_j\right)^2=\sum_{S,T\subset[n]}\sum_{j,k\notin S}\mathbb{E}\left(\bar{\mathbb{E}}_S[\phi_j]\bar{\mathbb{E}}_T[\phi_k]\right)$$
$$=\sum_{S\subset[n]}\sum_{j,k\notin S}\mathbb{E}\left(\bar{\mathbb{E}}_S[\phi_j]\bar{\mathbb{E}}_S[\phi_k]\right)=\sum_{S\subset[n]}\mathbb{E}\left(\sum_{j\notin S}\bar{\mathbb{E}}_S\phi_j\right)^2.$$

In the last sum we can ignore $S=\emptyset$, since $\bar{\mathbb{E}}_\emptyset\phi_j=\mathbb{E}\phi_j=0$, due to our assumption. Thus,

$$\mathbb{E}\left(\sum_{j\in[n]}\phi_j\right)^2\le\sum_{S\subset[n],S\neq\emptyset}\mathbb{E}\left(\sum_{j\notin S}\bar{\mathbb{E}}_S\phi_j\right)^2.$$

For $S\neq\emptyset$ the set $\{j:j\notin S\}$ has cardinality at most $n-1$. Thus, by Cauchy-Schwarz inequality we get

$$\left(\sum_{j\notin S}\bar{\mathbb{E}}_S\phi_j\right)^2\le(n-1)\sum_{j\notin S}(\bar{\mathbb{E}}_S\phi_j)^2.$$

We arrive at

$$\mathbb{E}\left(\sum_{j\in[n]}\phi_j\right)^2 \le (n-1)\sum_{S\subset[n],S\neq\emptyset}\mathbb{E}\sum_{j\notin S}(\bar{\mathbb{E}}_S\phi_j)^2 = (n-1)\sum_{S\subset[n]}\mathbb{E}\sum_{j\in[n]}(\bar{\mathbb{E}}_S\phi_j)^2$$
$$= (n-1)\sum_{j\in[n]}\mathbb{E}\left(\sum_{S\subset[n]}\bar{\mathbb{E}}_S\phi_j\right)^2 = (n-1)\sum_{j\in[n]}\mathbb{E}\phi_j^2.$$

□

8.3. **Monotonicity of Fisher information.** Using the techniques developed in the last two chapters, we prove the monotonicity of Fisher information in CLT, i.e. the inequality

$$\mathcal{I}(Y_n) \le \mathcal{I}(Y_{n-1}).$$

This will allow us to deduce (in the next section) the corresponding result for the Shannon entropy.

Let us define

$$V_n = \sum_{i\in[n]} X_i, \qquad V^{(j)} = \sum_{i\neq j} X_i, \qquad Y^{(j)} = \frac{1}{\sqrt{n-1}}\sum_{i\neq j} X_i.$$

Note that $\rho_{aX}(z) = \frac{1}{a}\rho_X(z/a)$. Thus, $\rho_{aX}(aX) = \frac{1}{a}\rho_X(X)$. Using this principle twice we get, for any $j = 1,\dots,n$,

$$\rho_{Y_n}(Y_n) = \sqrt{n}\rho_{V_n}(V_n) = \sqrt{n}\mathbb{E}[\rho_{V^{(j)}}(V^{(j)})|V_n] = \sqrt{\frac{n}{n-1}}\mathbb{E}[\rho_{Y^{(j)}}(Y^{(j)})|V_n]$$
$$= \sqrt{\frac{n}{n-1}}\mathbb{E}[\rho_{Y^{(j)}}(Y^{(j)})|Y_n].$$

Here the second equality follows from Lemma 13 applied to $X = V^{(j)}$, $Y = X_j$. From the linearity of conditional expectation we get

$$\rho_{Y_n}(Y_n) = \frac{1}{\sqrt{n(n-1)}}\sum_{j=1}^n \mathbb{E}[\rho_{Y^{(j)}}(Y^{(j)})|Y_n] = \frac{1}{\sqrt{n(n-1)}}\mathbb{E}\left[\sum_{j=1}^n \rho_{Y^{(j)}}(Y^{(j)})\Big|Y_n\right].$$

Let $\rho_j = \rho_{Y^{(j)}}(Y^{(j)})$. From the Cauchy-Schwarz inequality for the conditional expectation we get

$$\mathcal{I}(Y_n) = \mathbb{E}[\rho_{Y_n}(Y_n)^2] = \frac{1}{n(n-1)}\mathbb{E}\left(\mathbb{E}\left[\sum_{j=1}^n\rho_j\Big|Y_n\right]\right)^2 \le \frac{1}{n(n-1)}\mathbb{E}\mathbb{E}\left[\left(\sum_{j=1}^n\rho_j\right)^2\Big|Y_n\right]$$
$$= \frac{1}{n(n-1)}\mathbb{E}\left(\sum_{j=1}^n\rho_j\right)^2.$$

From the variance drop lemma we get

$$\mathbb{E}\left(\sum_{j=1}^n\rho_j\right)^2 \le (n-1)\sum_{j=1}^n\mathbb{E}[\rho_j^2] = n(n-1)\mathcal{I}(Y_{n-1}).$$

Thus, we get $\mathcal{I}(Y_n) \le \mathcal{I}(Y_{n-1})$.

8.4. **Proof of entropic CLT.** Let G be the standard Gaussian random variable. Define

$$Y_n(t) = \sqrt{t}Y_n + \sqrt{1-t}G, \qquad Y_{n-1}(t) = \sqrt{t}Y_{n-1} + \sqrt{1-t}G, \qquad t \in [0,1].$$

We prove that $\mathcal{S}(Y_n(t)) \geq \mathcal{S}(Y_{n-1}(t))$ for $t \in [0,1]$ and get the desired inequality by taking $t = 1$. For $t = 0$ we clearly have equality. Thus, it suffice to prove that

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{S}(Y_n(t)) \geq \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{S}(Y_{n-1}(t)).$$

Using de Bruijn's identity we get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{S}(Y_n(t)) &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\ln(\sqrt{t}) + \mathcal{S}\left(Y_n + \sqrt{\frac{1-t}{t}}G\right)\right) = \frac{1}{2t} - \frac{1}{2t^2}\mathcal{I}\left(Y_n + \sqrt{\frac{1-t}{t}}G\right) \\ &= \frac{1}{2t} - \frac{1}{2t}\mathcal{I}\left(\sqrt{t}Y_n + \sqrt{1-t}G\right)\end{aligned}$$

Let $G_1, \dots G_n$ be i.i.d. standard Gaussian random variables and take $X_i(t) = \sqrt{t}X_i + \sqrt{1-t}G_i$. Then

$$\sqrt{t}Y_n + \sqrt{1-t}G \sim \frac{(\sqrt{t}X_1 + \sqrt{1-t}G_1) + \ldots + (\sqrt{t}X_n + \sqrt{1-t}G_n)}{\sqrt{n}} = \frac{X_1(t) + \ldots + X_n(t)}{\sqrt{n}}.$$

Thus, from the last section we deduce

$$\mathcal{I}\left(\sqrt{t}Y_n + \sqrt{1-t}G\right) \leq \mathcal{I}\left(\sqrt{t}Y_{n-1} + \sqrt{1-t}G\right)$$

and

$$\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{S}(Y_n(t)) = \frac{1}{2t} - \frac{1}{2t}\mathcal{I}\left(\sqrt{t}Y_n + \sqrt{1-t}G\right) \geq \frac{1}{2t} - \frac{1}{2t}\mathcal{I}\left(\sqrt{t}Y_{n-1} + \sqrt{1-t}G\right) = \frac{\mathrm{d}}{\mathrm{d}t}\mathcal{S}(Y_{n-1}(t)).$$

References

[1] L. Gross, Logarithmic Sobolev Inequalities, American Journal of Mathematics Vol. 97 (4), 1975, 1061–1083.