Logarithmic Sobolev inequalities in discrete product spaces: proof by a transportation cost distance

Katalin Marton

Alfréd Rényi Institute of Mathematics
of the Hungarian Academy of Sciences
Relative entropy

Definition

$\mathcal{Z}$: measurable space, $\mu$, $\nu$: probability measures on $\mathcal{Z}$.
$Z$, $U$: random variables, $\mathcal{L}(Z) = \mu$, $\mathcal{L}(U) = \nu$.

Relative entropy:

$$D(\mu||\nu) = D(Z||U) = \int_{\mathcal{Z}} \log \frac{d\mu}{d\nu} d\mu$$  \hspace{1cm} (1)

If $\mathcal{Z}$ is finite:

$$D(\mu||\nu) = D(Z||U) = \sum_{z \in \mathcal{Z}} \mu(z) \log \frac{\mu(z)}{\nu(z)}.$$  \hspace{1cm} (2)
Entropy contraction of Markov kernels

**Definition**

\((\mathcal{Z}, \nu)\) : probability space,
\(\mathcal{P}(\mathcal{Z})\): measures on \(\mathcal{Z}\),
\(\Gamma\): Markov kernel on \(\mathcal{Z}\) with invariant measure \(\nu\).

*Entropy contraction for \((\mathcal{Z}, \nu, \Gamma)\)*

with rate \(1 - c\), \(0 < c \leq 1\):

\[
D(\mu \Gamma || \nu) \leq (1 - c) \cdot D(\mu || \nu). \tag{3}
\]

Equivalently:

\[
c \cdot D(\mu || \nu) \leq (D(\mu || \nu) - D(\mu \Gamma || \nu)), \quad \text{for all} \quad \mu \in \mathcal{P}(\mathcal{Z}) \tag{4}
\]
Gibbs sampler governed by local specifications of $q^n$

**Definition**

$\Gamma_i$: Markov kernel $\mathcal{X}^n \mapsto X^n$

$$\Gamma_i(z^n|y^n) = \delta(\bar{y}_i, \bar{z}_i) \cdot q_i(z_i|\bar{y}_i),$$

$\Gamma$: Markov kernel $\mathcal{X}^n \mapsto X^n$:

$$\Gamma = \frac{1}{n} \cdot \sum_{i=1}^{n} \Gamma_i.$$

**Definition**

$q^n$ has the entropy contraction property if:

its Gibbs sampler $\Gamma$ has.
Entropy contraction in product space

\((\mathcal{X}^n, q^n)\): probability space

**Question**

*Which measures \(q^n\) have the entropy contraction property with a reasonable constant \(c\)?*

\(c\) cannot be larger than \(O(1/n)\).

*Changing notation: WHEN*

\[
\frac{c}{n} \cdot D(p^n||q^n) \leq \left( D(p^n||q^n) - D(p^n\Gamma||q^n) \right) \quad \text{for all} \quad p^n \in \mathcal{P}(\mathcal{X}^n)
\]

\[(5)\]

*Equivalently: WHEN*

\[
D(p^n||q^n) \leq (1 - \frac{c}{n}) \cdot D(p^n\Gamma||q^n)
\]

\[(6)\]
Conditional relative entropy

**Definition**

\( \mathcal{Z} : \) measurable space,

\( \mathcal{V} : \) another measurable space, \( \pi : \) probability measure on \( \mathcal{V} \),

\( V : \) random variable, \( \mathcal{L}(V) = \pi \)

For \( v \in \mathcal{V} \)

probability measures on \( \mathcal{Z} \):

\( \mu(\cdot | v) = \mathcal{L}(Z|V = v) \quad \nu(\cdot | v) = \mathcal{L}(U|V = v) \)

**Conditional relative entropy:**

\[
D(\mu(\cdot | V) || \nu(\cdot | V)) = D(Z|V || U|V)) \\
\triangleq \int_{\mathcal{V}} D(\mu(\cdot | v) || \nu(\cdot | v)) d\pi
\]  

(7)
If $\mathcal{V}$ is finite:

$$D(\mu(\cdot|V)\|\nu(\cdot|V)) = D(Z|V)\|U|V)) = \sum_{v \in \mathcal{V}} \pi(v) D(\mu(\cdot|v)\|\nu(\cdot|v)).$$  (8)
Sufficient condition for entropy contraction in product space

$$(\mathcal{X}^n, q^n, \Gamma)$$: probability space with Gibbs sampler,

$$q^n = \mathcal{L}(X^n)$$

$$p^n = \mathcal{L}(Y^n)$$: another distribution on $\mathcal{X}^n$

Recall: $$\Gamma = \frac{1}{n} \sum_{i=1}^{n} \Gamma_i$$ \quad \tilde{Y}_i = (Y_1, \ldots, Y_{i-1}, Y_{i+1}, \ldots, Y_n),$$

Proposition

$$c \cdot D(p^n \| q^n) \leq \sum_{i=1}^{n} D(p_i(\cdot | \tilde{Y}_i) \| q_i(\cdot | \tilde{Y}_i)) \quad (*)$$

$$\implies$$

$$D(p^n \Gamma \| q^n) \leq (1 - \frac{c}{n}) D(p^n \| q^n).$$

If $q^n$ is a product measure then: \quad \(c = 1\).
Notation

For $x^n = (x_1, x_2, \ldots, x_n) \in \mathcal{X}^n$:

$$\bar{x}_i \triangleq (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$$

$$\bar{p}_i \triangleq \mathcal{L}(\bar{Y}_i), \quad \bar{q}_i \triangleq \mathcal{L}(\bar{X}_i)$$

Expansion formula

$$D(p^n || q^n) = D(\bar{p}_i || \bar{q}_i) + D(p_i(\cdot|\bar{Y}_i) || q_i(\cdot|\bar{Y}_i))$$

$$\implies$$

$$D(p^n || q^n) = \frac{1}{n} \sum_{i=1}^{n} D(\bar{p}_i || \bar{q}_i) + \frac{1}{n} \sum_{i=1}^{n} D(p_i(\cdot|\bar{Y}_i) || q_i(\cdot|\bar{Y}_i))$$
Proof of Proposition

\[ D(p^n \| q^n) - D(p^n \Gamma \| q^n) \]

\[ \geq \text{(by convexity of entropy)} \]

\[ D(p^n \| q^n) - \frac{1}{n} \cdot \sum_{i=1}^{n} D(\bar{p}_i \| \bar{q}_i) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} D(p_i(\cdot \mid \bar{Y}_i) \| q_i(\cdot \mid \bar{Y}_i)) \]

\[ (by \ assumption) \geq \frac{c}{n} D(p^n \| q^n) \]
\( \mathcal{X} \) finite

\((X^n, q^n, \Gamma)\)

Wanted: inequality

\[
D(p^n || q^n) \leq C \sum_{i=1}^{n} D(p_i(\cdot|\bar{Y}_i) || q_i(\cdot|\bar{Y}_i)) \quad \text{for all} \quad p^n \in \mathcal{P}(X^n) \quad (\ast)
\]

To get (\ast):
use a Wassersteine-like distance.
A reverse Pinsker’s inequality

\( \mu, \nu: \) probability measures on \( \mathcal{X} \quad \mathcal{X} \) finite

**Notation: Variational distance**

\[
|\mu - \nu| = \frac{1}{2} \cdot \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|
\]

**Lemma**

Set

\( \mathcal{X}_+ \triangleq \{ x \in \mathcal{X} : \nu(x) > 0 \}, \quad \alpha \triangleq \min \{ \nu(x) : x \in \mathcal{X}_+ \} \)

Then

\[
D(\mu||\nu) \leq \frac{4}{\alpha} \cdot |\mu - \nu|^2
\]

Follows from the inequality \( D(\mu||\nu) \leq \sum_{x \in \mathcal{X}} \frac{|\mu(x) - \nu(x)|^2}{\nu(x)} \)
\( X^n \): product space

\( \mu^n = \mathcal{L}(Z^n), \quad \nu^n = \mathcal{L}(U^n) \): probability measures on \( X^n \)

**Definition (P. Massart)**

_The square of the \( W_2 \)-distance of \( \mu^n \) and \( \nu^n \):_

\[
W_2^2(\mu^n, \nu^n) \triangleq \min \sum_{i=1}^{n} Pr^2 \{Z_i \neq U_i\}
\]

_\( \inf \): on couplings of \( \mu^n \) and \( \nu^n \)._
Notation

For $I \subset [1, n]$, $p^n = \mathcal{L}(Y^n)$ and $y^n \in \mathcal{X}^n$:

$$y_I \equiv (y_i : i \in I), \quad \bar{y}_I \equiv (y_i : i \notin I)$$

$$p_I \equiv \mathcal{L}(Y_I), \quad p_I(\cdot|\bar{y}_I) \equiv \mathcal{L}(Y_I|\bar{Y}_I = \bar{y}_I)$$

For Theorem 1 we need the inequality

$$W_2^2(p^n||q^n) \leq C \cdot E \sum_{i=1}^{n} |p_i(\cdot|\bar{Y}_i) - q_i(\cdot|\bar{Y}_i)|^2.$$ 

in a MORE GENERAL FORM:

We require a bound for

$$W_2^2(p_I(\cdot|\bar{y}_I), q_I(\cdot|\bar{y}_I))$$

for ALL subsets $I \subset [1, n]$ (not just $I = [1, n]$), and all $\bar{y}_I$. 
Main Theorem

\((\mathcal{X}^n, q^n), \quad \mathcal{X} \text{ finite!}\)

Theorem 1

Set

\[\alpha = \min \left\{ q_i(x_i|\bar{x}_i) : \quad q^n(x^n) > 0, \quad 1 \leq i \leq n \right\} \quad (10)\]

Fix a \(p^n = \mathcal{L}(Y^n) \in \mathcal{P}(\mathcal{X}^n); \text{ assume}\)

\[q^n(x^n) = 0 \implies p^n(x^n) = 0. \quad (11)\]

Main assumption:

\[W_2^2 \left( p_I (\cdot | \bar{y}_I), q_I (\cdot | \bar{y}_I) \right) \leq C \cdot \mathbb{E} \left\{ \sum_{i \in I} |p_i(\cdot | \bar{Y}_i) - q_i(\cdot | \bar{Y}_i)|^2 \left| \bar{Y}_I = \bar{y}_I \right\} \right\}, \quad (12)\]

for all \( I \subset [1, n], \quad \bar{y}_I \in \mathcal{X}^{[1,n]\setminus I}. \)
Main Theorem continued

Assume all the inequalities

\[
W_2^2(p_I(\cdot|\bar{y}_I), q_I(\cdot|\bar{y}_I)) \\
\leq C \cdot \mathbb{E} \left\{ \sum_{i \in I} |p_i(\cdot|\bar{Y}_i) - q_i(\cdot|\bar{Y}_i)|^2 \middle| \bar{Y}_I = \bar{y}_I \right\},
\]

(13)

where \( I \subset [1, n] \) and \( \bar{y}_I \in \mathcal{X}^{[1,n]\setminus I} \) is fixed.

Then

\[
D(p^n||q^n) \leq \\
\frac{4C}{\alpha} \cdot \sum_{i=1}^{n} \mathbb{E} |p_i(\cdot|\bar{Y}_i) - q_i(\cdot|\bar{Y}_i)|^2 \\
\leq \frac{2C}{\alpha} \cdot \sum_{i=1}^{n} D(p_i(\cdot|\bar{Y}_i)||q_i(\cdot|\bar{Y}_i)).
\]

(14)
An analogous result for densities in $\mathbb{R}^n$

Theorem

$$f(x^n) = \exp(-V(x^n)) : \text{density on } \mathbb{R}^n,$$

$$q^n : \text{probability measure with density } f.$$  

Assume: conditional densities $a f(x_i | \bar{x}_i)$ satisfy a logarithmic Sobolev inequality with constant $\rho$, for all $i, \bar{x}_i$.

Under some conditions on

$$\frac{1}{\rho} \cdot Hess\ V(x^n)$$

(expressing that $V$ is not too far from being uniformly convex):

$$D(p^n || q^n) \leq C \cdot \sum_{i=1}^{n} D(p_i(\cdot | \bar{Y}_i) || q_i(\cdot | \bar{Y}_i)) \quad \text{for all } p^n$$

($C = C(q^n)$)
Proof of Theorem 1

By induction on $n$. Assume Theorem 1 for $n - 1$

Notation

\[ \bar{p}_i(\cdot | y_i) \triangleq \mathcal{L}(\bar{Y}_i | Y_i = y_i) \]

For every $i \in [1, n]$

\[
D(p^n || q^n) = D(Y^n || X^n) = D(Y_i || X_i) + D(\bar{p}_i(\cdot | Y_i) || \bar{q}_i(\cdot | Y_i))
\]

\[
\implies D(p^n || q^n) = \frac{1}{n} \sum_{i=1}^{n} D(Y_i || X_i) + \frac{1}{n} \sum_{i=1}^{n} D(\bar{p}_i(\cdot | Y_i) || \bar{q}_i(\cdot | Y_i))
\]

(15)

By the induction hypothesis the second term is

\[
\leq (1 - \frac{1}{n}) \cdot \frac{4C}{\alpha} \sum_{i=1}^{n} \mathbb{E}\left| p_i(\cdot | \bar{Y}_i) - q_i(\cdot | \bar{Y}_i) \right|^2
\]
Proof of Theorem 1 Cont’d

Second term:

\[
\leq \left(1 - \frac{1}{n}\right) \cdot \frac{4C}{\alpha} \sum_{i=1}^{n} \mathbb{E} \left| p_i(\cdot | \bar{Y}_i) - q_i(\cdot | \bar{Y}_i) \right|^2
\]

First term:

\[
\frac{1}{n} \sum_{i=1}^{n} D(Y_i || X_i) \leq \quad \text{(by the Lemma)} \quad \frac{1}{n} \cdot \frac{4}{\alpha} \cdot \left| \mathcal{L}(Y_i) - \mathcal{L}(X_i) \right|^2
\]

\[
\leq \sum_{i=1}^{n} \Pr^2 \{ Y_i \neq X_i \} \quad \text{in any coupling of} \quad p^n, q^n
\]

\[
= \frac{1}{n} \cdot \frac{4}{\alpha} \cdot W_2^2(p^n, q^n) \quad \text{for the best coupling}
\]

\[
\leq \quad \text{(by the assumption of Theorem 1)}
\]

\[
\frac{1}{n} \cdot \frac{4C}{\alpha} \cdot \sum_{i=1}^{n} \left| p_i(\cdot | \bar{Y}_i) - q_i(\cdot | \bar{Y}_i) \right|^2
\]
Entropy contraction

$(\mathcal{X}^n, q^n, \Gamma), \quad \mathcal{X}$ finite
$\Gamma$: Gibbs sampler

**Corollary 1**

*If $q^n$ satisfies the conditions of Theorem 1 then*

$$D(p^n\Gamma\|q^n) \leq \left(1 - \frac{\alpha}{2nC}\right) \cdot D(p^n\|q^n).$$  \hspace{1cm} (17)
**Notation**

$\mathcal{E} :$ Dirichlet form associated with $\Gamma$ is the quadratic form

$$\mathcal{E}(f, g) = \langle (\text{Id} - \Gamma)f, g \rangle_{q^n}$$

**Definition**

$q^n$ satisfies a logarithmic Sobolev inequality with constant $c > 0$ if:

$$c \cdot D(p^n || q^n) \leq \mathcal{E} \left( \sqrt{\frac{p^n}{q^n}}, \sqrt{\frac{p^n}{q^n}} \right) \quad \text{for all} \quad p^n \in \mathcal{P}(\mathcal{X}^n)$$
Under conditions of Theorem 1, the logarithmic Sobolev inequality holds true:

$$\frac{1}{n} \cdot D(p^n||q^n) \leq \frac{4C}{\alpha} \cdot \mathcal{E}_\Gamma \left( \sqrt{\frac{p^n}{q^n}}, \sqrt{\frac{p^n}{q^n}} \right)$$

$$= \frac{4C}{\alpha n} \cdot \sum_{i=1}^{n} \mathbb{E} \left( 1 - \left( \sum_{y_i \in \mathcal{X}} \sqrt{p_i(y_i|\bar{Y}_i)} \cdot q_i(y_i|\bar{Y}_i) \right)^2 \right).$$

$$\implies$$ hypercontractivity
Application: Gibbs measures with Dobrushin’s uniqueness condition

\((\mathcal{X}^n, q^n), \quad \mathcal{X} \text{ finite}\)

**Definition**

\(q^n\) satisfies (an \(\mathbb{L}_2\)-version of)

Dobrushin’s uniqueness condition with coupling matrix

\[ A = \left( a_{k,i} \right)_{k,i=1}^n, \quad a_{i,i} = 0, \]

if:

(i) \( \max |q_i(\cdot | \bar{z}_i) - q_i(\cdot | \bar{s}_i)| \leq a_{k,i}, \quad k \neq i, \)

\( \max: \text{for all} \quad \bar{z}_i, \bar{s}_i \quad \text{differing only in the} \quad k\text{-th coordinate}, \)

and (ii)

\[ \|A\|_2 < 1. \]
Theorem 2

Assume Dobrushin's uniqueness condition with coupling matrix $A$, $||A||_2 < 1$.

Then conditions of Theorem 1 hold with

$$C = \frac{1}{(1 - ||A||)^2}.$$

Thus

$$D(p^n || q^n) \leq \frac{4}{\alpha} \cdot \frac{1}{(1 - ||A||)^2} \cdot \sum_{i=1}^{n} \mathbb{E} |p_i(\cdot|\bar{Y}_i) - q_i(\cdot|\bar{Y}_i)|^2$$

$$\leq \frac{2}{\alpha} \cdot \frac{1}{(1 - ||A||)^2} \cdot \sum_{i=1}^{n} D(p_i(\cdot|\bar{Y}_i)||q_i(\cdot|\bar{Y}_i)),$$

(20)
Dobrushin’s uniqueness condition implies that $\Gamma$ contracts $W_2$-distance with rate

$$1 - \frac{1}{n} \cdot \left(1 - \|A\|_2\right).$$
Application: Gibbs measures on $\mathbb{Z}^d$

**Notation**

$\mathbb{Z}^d$: $d$-dimensional lattice, $i \in \mathbb{Z}^d$: site

$\rho(k, i) = \max_{1 \leq \nu \leq d} |k_\nu - i_\nu|$: distance on $\mathbb{Z}^d$,

$\Lambda \subset\subset \mathbb{Z}^d$: finite set of sites

$\mathcal{X}$ finite: spin space

$x^{\mathbb{Z}^d} = (x_i : i \in \mathbb{Z}^d) \in \mathcal{X}^{\mathbb{Z}^d}$: spin configuration

$\mathcal{X}^{\mathbb{Z}^d}$: configuration space,

For $x^{\mathbb{Z}^d}$ and $\Lambda \subset \mathbb{Z}^d$

$$x_\Lambda = (x_i : i \in \Lambda), \quad \bar{x}_\Lambda = (x_i : i \notin \Lambda),$$

$\bar{x}_\Lambda$ is called an outside configuration for $\Lambda$. 
Definition

$q_{\Lambda}(\cdot | \bar{x}_{\Lambda})$, $\Lambda \subset \subset \mathbb{Z}^d$ : conditional distributions on $\mathcal{X}^\Lambda$. Assume compatibility conditions.

There exists at least one probability measure $q$ on the space of configurations:

$$q = \mathcal{L}(X) \in \mathcal{P}(\mathcal{X}^{\mathbb{Z}^d})$$

satisfying

$$\mathcal{L}(X_\Lambda | \bar{X}_\Lambda = \bar{x}_\Lambda) = q_{\Lambda}(\cdot | \bar{x}_\Lambda), \quad \text{all} \quad \Lambda \subset \subset \mathbb{Z}^d.$$ 

$q_{\Lambda}(\cdot | \bar{x}_\Lambda)$ : local specifications of $q$. 
Finite range interactions

**Definition**

_The local specifications have finite range of interactions if: there is an \( R > 0 \):_

\[ q_\Lambda (\cdot | x_\Lambda) \text{ only depends on coordinates } k \notin \Lambda \text{ with } \rho(k, \Lambda) \leq R. \]

\( q \) may not be uniquely defined by the local specifications, even for finite range interactions.
Dobrushin-Shlosman’s strong mixing condition

Given local specifications $q_\Lambda(\cdot | \bar{x}_\Lambda)$.

**Assumption**

There exists a function $\varphi(\rho)$ of the distance such that:

(i) 
$$\sum_{i \in \mathbb{Z}^d} \varphi(\rho(k, i)) < \infty,$$

and:

(ii) for every

$$\Lambda \subset \subset \mathbb{Z}^d, \quad M \subset \Lambda, \quad k \notin \Lambda$$

and every

$$\bar{y}_\Lambda, \bar{z}_\Lambda, \quad \text{differing only at} \quad k :$$

$$|q_M(\cdot | \bar{y}_\Lambda) - q_M(\cdot | \bar{z}_\Lambda)| \leq \varphi(\rho(k, M)).$$
In case of finite range interactions:

If Dobrushin-Shlosman’s strong mixing condition holds then

$$\varphi(\rho) = C \cdot \exp(-\gamma \cdot \rho)$$

can be taken
\[ \rho(k, M) = \max \text{ length of red segments} \]

\[ |q_M (\cdot | \overline{y}_\Lambda) - q_M (\cdot | \overline{z}_\Lambda)| \leq \varphi(\rho(k, M)) \]

\( \overline{y}_\Lambda, \overline{z}_\Lambda \) differ only at \( k \in \overline{\Lambda} \)
Dobrushin-Shlosman’s strong mixing condition

Cont’d

Meaning: The influence of the spin at \( k \notin \Lambda \) on the spins in \( M \subset \Lambda \) that are far away from \( k \) is small.

Essential:

The spins over \( \bar{\Lambda} \) are fixed in two different ways.

The spins over \( \Lambda \setminus M \) are not fixed.
Logarithmic Sobolev inequality for strongly mixing measures

Earlier results for the case of finite range interactions:
D. Stroock, B. Zegarlinski 1992, F. Ces~i 2001
F. Martinelli, E. Olivieri

**Theorem 3**

\( (X^\Lambda, q^\Lambda(\cdot|\bar{y}^\Lambda)) \) for fixed \( \Lambda \) and \( \bar{y}^\Lambda \)

\[ \alpha = \min \left\{ q_i(x_i|x_i) : q_i(x_i|x_i) > 0 \right\} . \]

*(Finite range is not assumed.)*

\{ Dobrushin-Shlosman’s strong mixing condition + \{ \alpha > 0 \} \} \implies conditions of Theorem 1 for \( q^\Lambda(\cdot|\bar{y}^\Lambda) \), with uniform constant

\[ \implies \]

logarithmic Sobolev inequality for \( q^\Lambda(\cdot|\bar{y}^\Lambda) \) with uniform constant
Logarithmic Sobolev inequality for strongly mixing measures

Cont’d

The proof uses a Gibbs sampler updating cubes of size $m$ depending on the dimension and on the function $\varphi(\rho)$.

We get

$$W_2^2(p_\Lambda, q_\Lambda(\cdot|\bar{y}_\Lambda)) \leq C_m \cdot \sum_{I:m\text{-sided cube}} \mathbb{E}|p_{I \cap \Lambda}(\cdot|\bar{Y}_{I \cap \Lambda}) - q_{I \cap \Lambda}(\cdot|\bar{Y}_{I \cap \Lambda})|^2$$

$$\leq C_{m,\alpha} \cdot \sum_{i \in \Lambda} \mathbb{E}|p_i(\cdot|Y_{\Lambda \setminus i}) - q_i(\cdot|Y_{\Lambda \setminus i}, \bar{y}_\Lambda)|^2$$

for an appropriate $m$ that is good enough for any $\Lambda$ and $\bar{y}_\Lambda$. 