

# A Fuzzy Decomposition Method to Establish Functional Inequalities

Suei-Wen Chen\*

Department of Science, New York University Abu Dhabi, UAE

## Abstract

This paper presents a method to establish functional inequalities via fuzzy decomposition on the state space, which generalizes earlier results dealing with exact partitions of the state space. Given a reversible Markov chain on a finite state space, we define its projection chain and restriction chains from classes of a fuzzy partition on the state space. The Poincaré, log-Sobolev and modified log-Sobolev inequalities associated with the original chain can be estimated from those of its projection chain and restriction chains which tend to have simpler structures and hence easier to work with. An application of this generalization is presented in which the fuzzy decomposition method applies but its exact partitioning counterpart does not.

## Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Reversible Markov Chains and Functional Inequalities . . . . .	1
1.2	Decomposition Methods and Fuzzy Decomposition . . . . .	2
1.3	Main Result . . . . .	3
<b>2</b>	<b>Related Works and Examples</b>	<b>4</b>
<b>3</b>	<b>Proof</b>	<b>5</b>
	<b>References</b>	<b>7</b>

## 1 Introduction

Finite-state Markov chains are used in many fields to describe and predict the dynamics of sequential processes. One powerful technique to study the behavior of Markov chains is via functional inequalities (for instance, see [1, 2, 3]). Applications of functional inequalities includes establishing concentration of measure as well as providing non-asymptotic upper bounds on the mixing time of the underlying Markov chains, which is of importance for applications such as Markov Chain Monte Carlo algorithms (for example, see [4]). Although functional inequalities are powerful in theory, it is often difficult to establish sharp bounds for these inequalities in practice. This paper provides a generalization of a method for establishing functional inequalities via Markov chain decomposition.

### 1.1 Reversible Markov Chains and Functional Inequalities

This paper concerns reversible Markov chains on finite state spaces. Consider a finite probability space  $(\Omega, \mathcal{P}(\Omega), \pi)$  and a Markov generator  $Q$  with respect to  $\pi$ , which is an  $|\Omega| \times |\Omega|$  matrix with nonnegative off-diagonal entries such that each row sums to zero. We require that  $Q$  is reversible under the probability distribution  $\pi$ , namely  $\pi(x)Q(x, y) = \pi(y)Q(y, x)$  for all  $x, y \in \Omega$ . It is well-known that if  $Q$  is irreducible, then the continuous-time chain defined by  $p_t := e^{tQ}$  converges to the stationary distribution, namely  $p_t(x, y) \rightarrow \pi(y)$  as  $t \rightarrow \infty$  for all  $x, y \in \Omega$ . The mixing time defined by

$$t_{\text{mix}}(\epsilon) := \inf\{t \geq 0 : \max_{x \in \Omega} \|p_t(x, \cdot) - \pi\|_{\text{TV}} \leq \epsilon\}$$

---

\*This work is part of an undergraduate thesis done under the supervision of Prof. Pierre Youssef.

measures the rate at which this convergence occurs, where  $\|\mu - \nu\|_{\text{TV}} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|$  is the total variation distance and  $\epsilon > 0$  is the threshold.

We denote by  $\mathbb{E}_\pi[\cdot]$ ,  $\text{Var}_\pi(\cdot)$  and  $\text{Ent}_\pi(\cdot)$  the standard expectation, variance and entropy over the probability space  $(\Omega, \mathcal{P}(\Omega), \pi)$ , namely

$$\begin{aligned}\mathbb{E}_\pi[f] &:= \sum_{x \in \Omega} \pi(x) f(x); \\ \text{Var}_\pi(f) &:= \mathbb{E}_\pi[f^2] - (\mathbb{E}_\pi[f])^2; \\ \text{Ent}_\pi(f) &:= \mathbb{E}_\pi[f \log f] - \mathbb{E}_\pi[f] \log \mathbb{E}_\pi[f].\end{aligned}$$

Variance and entropy capture the global variation of  $f$  under  $\pi$ , while the local variation associated with  $Q$  can be measured by the Dirichlet form

$$\mathcal{E}(f, g) := \langle -Qf, g \rangle_\pi = - \sum_{x \in \Omega} (Qf)(x) g(x) \pi(x)$$

for  $f, g : \Omega \rightarrow \mathbb{R}$ . Establishing bounds on global variations using average local variations provides powerful control over the mixing behavior of the chain. Specifically, the Poincaré constant  $\lambda(Q)$ , the modified log-Sobolev constant  $\alpha(Q)$  and the log-Sobolev constant  $\rho(Q)$  are defined respectively as the supremum of  $\lambda, \alpha, \rho \geq 0$  such that

$$\begin{aligned}\mathcal{E}(f, f) &\geq \lambda \text{Var}_\pi(f); \\ \mathcal{E}(f, \log f) &\geq \alpha \text{Ent}_\pi(f); \\ \mathcal{E}(\sqrt{f}, \sqrt{f}) &\geq \rho \text{Ent}_\pi(f),\end{aligned}$$

for all  $f : \Omega \rightarrow (0, \infty)$ . The three inequalities are called the Poincaré inequality, the modified log-Sobolev inequality (MLSI) and the log-Sobolev inequality (LSI) respectively. It can be shown that the mixing time is of order  $O(\frac{1}{\lambda(Q)} \log \pi_{\min}^{-1})$ ,  $O(\frac{1}{\alpha(Q)} \log \log \pi_{\min}^{-1})$  and  $O(\frac{1}{\rho(Q)} \log \log \pi_{\min}^{-1})$ , where  $\pi_{\min} := \min_{x \in \Omega} \pi(x)$  and the three constants satisfy  $2\lambda(Q) \geq \alpha(Q) \geq 4\rho(Q)$  (see [3, 2]).

## 1.2 Decomposition Methods and Fuzzy Decomposition

Powerful as these functional inequalities are, proving sharp lower bounds for these constants remains difficult in practice. In fact, proving the optimal LSI and MLSI for even a two-state chain is highly nontrivial [5]. It is thus desirable to find a reduction method for establishing such functional inequalities. A natural way to proceed is to partition the state space  $\Omega$  into disjoint subsets. This naturally induces several restriction chains, one on each block, as well as a projection chain between blocks. The restriction and projection chains could be simpler in structure and easier to establish their corresponding functional inequalities. The task then is to recover the functional-inequality constant for the original chain from those of the restriction chains and the projection chain. A review of such decomposition methods can be found in [6].

The main contribution of this paper is a generalization of a partition method developed by Lu and Yau in the context of interacting particle systems [7]. A succinct discussion on this method is given by Hermon and Salez in [8], which builds on the usual notion of partition. Specifically, let  $I$  be a finite set whose elements are thought of as classes to partition  $\Omega$  into. Traditionally, a partition on  $\Omega$  is given by pairwise disjoint nonempty subsets  $\Omega_i \subseteq \Omega$  such that  $\Omega = \cup_{i \in I} \Omega_i$ . We refer to this as exact partition in this paper. In some cases, however, it may be natural to relax the condition of exact membership and interpret each state as belonging to a number of classes with different degrees of membership. This motivates the definition of *fuzzy decomposition*, which we describe below in the language of fuzzy sets and fuzzy relations [9].

Given a set  $X$ , a fuzzy set on  $X$  is defined as a map  $\mu : X \rightarrow [0, 1]$ . A subset  $A \subseteq X$  can be thought of as the collection of elements in  $X$  satisfying some property  $P$ . A fuzzy set  $\mu$  on  $X$  is meant to model the degree to which each  $x \in X$  satisfies some property  $P$ . The closer  $\mu(x)$  is to one (resp. zero), the more (resp. less) we consider  $x$  to satisfy  $P$ . A fuzzy relation between sets  $X$  and  $Y$  is a fuzzy set on  $X \times Y$ . The underlying mapping  $X \times Y \rightarrow [0, 1]$  quantifies the degree to which this relation is satisfied.

We define a fuzzy decomposition (or a fuzzy partition) of  $\Omega$  with classes  $I$  as a fuzzy relation  $a : \Omega \times I \rightarrow [0, 1]$  on  $\Omega \times I$  such that (1)  $\sum_{i \in I} a(x, i) = 1$  for all  $x \in \Omega$  and (2) for all  $i \in I$  we have  $a(x, i) > 0$  for some  $x \in \Omega$ . We write  $a_i(x) := a(x, i)$  and interpret this as the degree to which  $x$  belongs

to class  $i$ . The first condition means that for each state  $x$ , the degrees of membership across all classes sum up to unity, while the second condition rules out empty classes. For each  $i \in I$  we refer to the set  $\Lambda_i := \{x \in \Omega : a_i(x) > 0\}$  as class  $i$ , consisting of states with positive membership value. The traditional notion of partition is recovered by restricting the range of  $a$  to  $\{0, 1\}$  and set  $\Omega_i$  to  $\Lambda_i$ . The main result of the paper generalizes the treatment in [8] to the case of fuzzy decompositions.

### 1.3 Main Result

Suppose we are given a reversible chain  $(Q, \pi)$  on  $\Omega$  and a fuzzy decomposition  $a$  on  $\Omega$  with classes  $I$ , where

$$\Lambda_i := \{x \in \Omega \mid a_i(x) > 0\}$$

is referred to as class  $i$ . The stationary measure  $\pi$  yields a probability measure  $\hat{\pi}$  on  $I$  defined by

$$\hat{\pi}(i) := \sum_{x \in \Lambda_i} a_i(x) \pi(x), \quad i \in I$$

as well as a probability measure on each class  $\Lambda_i$  defined by

$$\pi_i(x) := \frac{a_i(x) \pi(x)}{\hat{\pi}(i)}, \quad x \in \Lambda_i.$$

Define the projection chain  $\hat{Q}$  of  $Q$  as follows: for  $i, j \in I$ ,  $i \neq j$  let

$$\hat{Q}(i, j) := \frac{1}{\hat{\pi}(i)} \sum_{x \in \Lambda_i} \sum_{y \in \Lambda_j \setminus \{x\}} a_i(x) a_j(y) \pi(x) Q(x, y),$$

and set  $\hat{Q}(i, i)$  such that each row of  $\hat{Q}$  sums to 0. For each  $i \in I$ , the restriction chain  $Q_i$  of  $Q$  on  $\Lambda_i$  is defined as follows:  $Q_i(x, y) := a_i(y) Q(x, y)$  if  $x \neq y$  and set  $Q_i(x, x)$  such that each row sums to 0.

It is clear from their definition that  $\hat{Q}$  is reversible under  $\hat{\pi}$  and that  $Q_i$  is reversible under  $\pi_i$  for all  $i \in I$ . The interpretations of  $\hat{Q}$  and  $Q_i$  are natural:  $\hat{Q}(i, j)$  represents the transition probability from class  $i$  to class  $j$  as a normalized sum of transition probabilities from states in class  $i$  to states in class  $j$ , weighted by proportions of class membership. The transition probability  $Q_i(x, y)$  is simply the original transition probability  $Q(x, y)$  discounted by  $a_i(y)$ , namely how strongly  $y$  belongs to class  $i$ .

Now suppose that for  $(i, j) \in I^2$  with  $\hat{Q}(i, j) > 0$  we are given couplings  $\kappa_{ij}$  of  $\pi_i$  and  $\pi_j$ , meaning that  $\kappa_{ij}$  is a probability distribution on  $\Lambda_i \times \Lambda_j$  such that for all  $x \in \Lambda_i$  and  $y \in \Lambda_j$ ,

$$\sum_{x \in \Lambda_i} \kappa_{ij}(x, y) = \pi_j(y) \quad \text{and} \quad \sum_{j \in \Lambda_j} \kappa_{ij}(x, y) = \pi_i(x).$$

We define the quality of these couplings to be

$$\chi := \min_{\substack{(x, y, i, j) \text{ s.t.} \\ \kappa_{ij}(x, y) > 0, \\ x \neq y}} \frac{a_i(x) a_j(y) \pi(x) Q(x, y)}{\hat{\pi}(i) \hat{Q}(i, j) \kappa_{ij}(x, y)}.$$

This quantity allows us to estimate the Poincare, log-Sobolev and modified log-Sobolev constants of  $Q$  from those of  $Q_i$  and  $\hat{Q}$  in the following manner.

**Theorem 1.1.** Let  $(\Omega, \mathcal{P}(\Omega), \pi)$  be a finite probability space. Suppose we are given a fuzzy partition on  $\Omega$  with classes  $I$  which admits couplings  $\kappa_{ij}$  between  $\pi_i$  and  $\pi_j$  for all  $(i, j) \in I^2$  such that  $\hat{Q}(i, j) > 0$ , with quality  $\chi$ . Then

$$\lambda(Q) \geq \min\{\chi \lambda(\hat{Q}), \min_{i \in I} \lambda(Q_i)\};$$

$$\alpha(Q) \geq \min\{\chi \alpha(\hat{Q}), \min_{i \in I} \alpha(Q_i)\};$$

$$\rho(Q) \geq \min\{\chi \rho(\hat{Q}), \min_{i \in I} \rho(Q_i)\}.$$

Let us note that Theorem 1.1 is a strict generalization of its non-fuzzy version in [7, 8]. This is because in order for such decomposition methods to give nontrivial bounds, the projection chain and all the restriction chains must be irreducible, and at the same time the existence of couplings with nonzero quality is needed. For some Markov chains, it is not always possible to achieve all these requirements simultaneously. For instance, the non-fuzzy version of Theorem 1.1 cannot be applied to Example 2.1 in general due to the asymmetry of the state space.

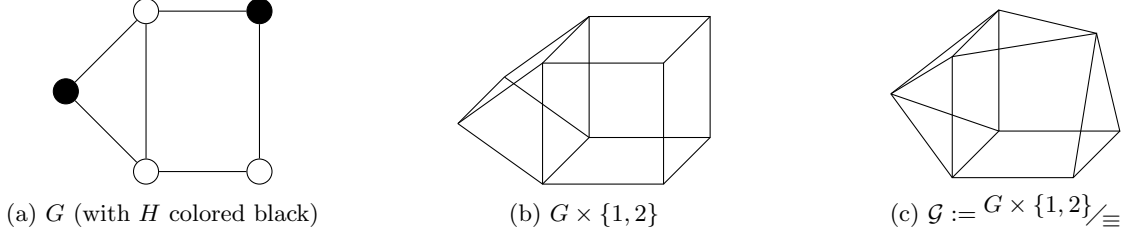


Figure 1: Illustration of the construction of  $\mathcal{G}$

## 2 Related Works and Examples

A number of decomposition methods have been developed for Markov Chain analysis. In the method proposed by Madras and Randall [10], decomposition takes the form of a set cover  $\Omega = \bigcup_{i \in I} A_i$  where  $A_i \subseteq \Omega$  are not necessarily pairwise disjoint. A different version of projection chain and restriction chains is considered. In particular, the projection chain depends only on the probabilities of  $A_i$  and the maximum overlap, defined as the maximum cardinality of  $\{i \in I : x \in A_i\}$  across  $x \in \Omega$ . The method by Jerrum et al. [6] works with exact partition  $\Omega = \bigcup_{i \in I} \Omega_i$  of the state space. While their version of projection chain and restriction chain coincide with the one considered in this paper, they do not appeal to couplings of the distributions  $\pi_i$ . Instead, the quantity

$$\gamma := \max_{i \in I} \max_{x \in \Omega_i} P(x, \Omega \setminus \Omega_i)$$

describing the maximum probability of exiting the current block is considered, where  $P$  is the transition probability matrix of the chain. However, unlike Theorem 1.1, these two settings are either insensitive to or unable to capture the degree of membership that may be desirable in some applications.

The following is an example of how Theorem 1.1 may apply.

**Example 2.1.** Let  $G$  be a finite set on which an undirected connected graph is defined. We consider a graph consisting of two copies of  $G$  but with some points glued together. More precisely, let  $H$  be a subset of  $G$  such that if  $x \in H$  then  $N_G(x) \cap H = \emptyset$ , where  $N_G(x)$  is the set of neighbors of  $x$  in  $G$ . Define an equivalence relation  $\equiv$  on  $G \times \{1, 2\}$  by  $(x, i) \equiv (y, j)$  if and only if  $(x, i) = (y, j)$  or  $x = y \in H$ . Let  $\mathcal{G} := G \times \{1, 2\} / \equiv$ . We identify the elements of  $H$  with their equivalent classes; namely, since  $(h, 1) \equiv (h, 2)$  for  $h \in H$  we abuse the notation and write  $H = \{(h, i) : h \in H, i \in \{1, 2\}\}$ . We also write  $G_i = \{(x, i) \in \mathcal{G} \mid x \in G \setminus H\}$ . Thus  $\mathcal{G} = G_1 \cup G_2 \cup H$ . The edges of  $\mathcal{G}$  are defined by the edges of the two copies of  $G$  while connecting  $(x, 1)$  and  $(x, 2)$  for all  $x \in G \setminus H$ . Write  $v \sim w$  if  $v$  and  $w$  are neighbors in  $\mathcal{G}$ . See Figure 2.1 for an illustration of this construction for  $G = \{a, b, c, d, e\}$  with edges  $a \sim b, b \sim c, c \sim d, d \sim e, e \sim b, e \sim a$  with  $H = \{a, c\}$ .

Now we consider the simple random walk on  $\mathcal{G}$ . Let  $d_G(x)$  and  $d_{\mathcal{G}}(x)$  denote the degree of  $x$  in  $G$  and  $\mathcal{G}$  respectively, and let

$$D_G := \sum_{x \in G} d_G(x), \quad D_{\mathcal{G}} := \sum_{x \in \mathcal{G}} d_{\mathcal{G}}(x).$$

Then the chain in question is  $Q(x, y) = 1/d_{\mathcal{G}}(x)$  if  $x \sim y$  and  $\pi(x) = d_{\mathcal{G}}(x)/D_{\mathcal{G}}$ , where

$$d_{\mathcal{G}}(x) = \begin{cases} 2d_G(x) & \text{if } x \in H, \\ d_G(x) + 1 & \text{otherwise.} \end{cases}$$

Thus  $D_{\mathcal{G}} = 2D_G + 2\#(G \setminus H)$  where  $\#$  denotes the cardinality of a set. Assign weight vectors of  $x \in \mathcal{G}$  to be  $a(x) = (1/2, 1/2)$  if  $x \in H$ ,  $a(x) = (1, 0)$  if  $x \in G_1$  and  $a(x) = (0, 1)$  if  $x \in G_2$ . Thus  $\Lambda_i = G_i \cup H$  and  $\hat{\pi}(i) = 1/2$  by symmetry ( $i = 1, 2$ ). Note that  $\pi_1(x) = \pi_2(x) = \pi(x)$  for  $x \in H$  and  $\pi_1(x) = \pi_2(y) = 2\pi(x)$  if  $x \in G_1$  and  $y \in G_2$  correspond to the same element in  $G$ . Hence the probability measure  $\kappa := \kappa_{12}$  on  $\Lambda_1 \times \Lambda_2$  defined by

$$\kappa(x, y) = \begin{cases} \pi(x) & \text{if } x = y \in H \\ 2\pi(x) & \text{if } x = (v, 1) \text{ and } y = (v, 2) \text{ for some } v \in G \setminus H \\ 0 & \text{otherwise} \end{cases}$$

is a coupling of  $\pi_1$  and  $\pi_2$ . Again, by symmetry we set  $\kappa_{21}(y, x) = \kappa(x, y)$ . Note that  $\pi(x)Q(x, y) = \mathbb{1}_{x \sim y}/D_G$ , so the projection chain is

$$\hat{Q}(2, 1) = \hat{Q}(1, 2) = \frac{2}{D_G} \left( \sum_{x \in H} d_G(x) + \#(G \setminus H) \right).$$

Recall that the quality of the couplings are

$$\chi = \min \frac{a_1(x)a_2(y)\pi(x)Q(x, y)}{\hat{\pi}(1)\hat{Q}(1, 2)\kappa(x, y)}$$

where the minimum is taken for  $x \neq y$  with  $\kappa(x, y) > 0$  and therefore  $x = (v, 1)$ ,  $y = (v, 2)$  for some  $v \in G \setminus H$ . It follows that

$$\chi = \min_{x \in G_1 \cup G_2} \frac{1/D_G}{\frac{1}{2}\hat{Q}(1, 2) \cdot 2d_G(x)/D_G} = \frac{1}{\hat{Q}(1, 2)} \frac{1}{\max_{v \in G \setminus H} d_G(v) + 1}.$$

The restriction chains  $Q_i(x, y)$  on  $G$  coincide with  $Q(x, y)$  when  $y \in G \setminus H$  and  $Q_i(x, y) = \frac{1}{2}Q(x, y)$  if  $y \in H$ . We conclude that

$$\lambda(Q) \geq \min \left\{ \frac{2}{\max_{v \in G \setminus H} d_G(v) + 1}, \lambda(Q_i) \right\}$$

and similarly for  $\alpha(Q)$  and  $\rho(Q)$ . A lower bound on  $\lambda(Q)$  can be established if one can solve for  $\lambda(Q_i)$ , for example, by recursively apply decomposition techniques or spectral methods, depending on the structure of the original graph.

### 3 Proof

A proof of Theorem 1.1 is presented in this section, which directly generalizes the treatment in [8]. First, we note that by reversibility, the Dirichlet forms associated with the three constants can be written as

$$\mathcal{L}_\pi f := \frac{1}{2} \sum_{(x, y) \in \Omega^2} \pi(x)Q(x, y)\Psi(f(x), f(y)), \text{ where}$$

$$\Psi(u, v) = \begin{cases} (u - v)^2 & \text{for Poincare;} \\ (u - v)(\log u - \log v) & \text{for modified log-Sobolev;} \\ (\sqrt{u} - \sqrt{v})^2 & \text{for log-Sobolev.} \end{cases}$$

For notational convenience we also write

$$\mathcal{R}_\pi f := \begin{cases} \text{Var}_\pi f & \text{for Poincare} \\ \text{Ent}_\pi f & \text{for modified log-Sobolev and log-Sobolev.} \end{cases}$$

The idea is to decompose the variance, entropy and Dirichlet forms according to the partition. To this end, for  $i \in I$  we define

$$\hat{f}(i) := \mathbb{E}_{\pi_i} [f|_{\Lambda_i}] = \sum_{x \in \Lambda_i} f(x)\pi_i(x).$$

Thus

$$\begin{aligned} \text{Var} f &= \frac{1}{2} \sum_{x, y \in \Omega} \pi(x)\pi(y)(f(x) - f(y))^2 \\ &= \frac{1}{2} \sum_{i, j \in I} \sum_{x, y \in \Omega} a_i(x)a_j(y)\pi(x)\pi(y)[(f(x) - \hat{f}(i)) + (\hat{f}(i) - \hat{f}(j)) + (\hat{f}(j) - f(y))]^2. \end{aligned}$$

When expanding the square, the cross terms involving  $(f(x) - \hat{f}(i))$  or  $(\hat{f}(j) - f(y))$  are 0 by the definition of  $\hat{f}$ . Thus,  $\text{Var}_\pi f$  is equal to

$$\begin{aligned} & \frac{1}{2} \sum_{i,j \in I} \sum_{x,y \in \Omega} a_i(x) a_j(y) \pi(x) \pi(y) [(f(x) - \hat{f}(i))^2 + (\hat{f}(i) - \hat{f}(j))^2 + (\hat{f}(j) - f(y))^2] \\ &= \sum_{i \in I} \sum_{x \in \Omega} a_i(x) \pi(x) (f(x) - \hat{f}(i))^2 + \frac{1}{2} \sum_{i,j \in I} \hat{\pi}(i) \hat{\pi}(j) (\hat{f}(i) - \hat{f}(j))^2 \\ &= \sum_{i \in I} \hat{\pi}(i) \text{Var}_{\pi_i} f + \text{Var}_{\hat{\pi}} \hat{f}. \end{aligned}$$

Similarly, the entropy of  $f : \Omega \rightarrow (0, \infty)$  can be written as

$$\begin{aligned} \text{Ent}_\pi f &= \mathbb{E}_\pi [f(\log f - \log \mathbb{E}_\pi f)] = \sum_{x \in \Omega} \sum_{i \in I} a_i(x) \pi(x) f(x) [\log f(x) - \log \mathbb{E}_\pi f] \\ &= \sum_{i \in I} \hat{\pi}(i) \sum_{x \in \Lambda_i} \pi_i(x) f(x) [\log f(x) - \log \mathbb{E}_\pi f] \\ &= \sum_{i \in I} \hat{\pi}(i) \text{Ent}_{\pi_i} f + \sum_{i \in I} \hat{\pi}(i) \sum_{x \in \Lambda_i} \pi_i(x) f(x) [\log \mathbb{E}_{\pi_i} f - \log \mathbb{E}_\pi f] \\ &= \sum_{i \in I} \hat{\pi}(i) \text{Ent}_{\pi_i} f + \sum_{i \in I} \hat{\pi}(i) \hat{f}(i) [\log \hat{f}(i) - \log \mathbb{E}_{\hat{\pi}} \hat{f}] \\ &= \sum_{i \in I} \hat{\pi}(i) \text{Ent}_{\pi_i} f + \text{Ent}_{\hat{\pi}} \hat{f}. \end{aligned}$$

The Dirichlet forms are

$$\begin{aligned} \mathcal{L}_\pi f &= \frac{1}{2} \sum_{i,j \in I} \sum_{x,y \in \Omega} a_i(x) a_j(y) \pi(x) Q(x,y) \Psi(f(x), f(y)) \\ &= \frac{1}{2} \sum_{i \in I} \sum_{x,y \in \Omega} a_i(x) \pi(x) a_i(y) Q(x,y) \Psi(f(x), f(y)) + \\ &\quad \frac{1}{2} \sum_{i \in I} \sum_{j \neq i} \sum_{x,y \in \Omega} a_i(x) a_j(y) \pi(x) Q(x,y) \Psi(f(x), f(y)) \end{aligned}$$

The first term is

$$\frac{1}{2} \sum_{i \in I} \hat{\pi}(i) \sum_{x,y \in \Lambda_i} \pi_i(x) Q_i(x,y) \Psi(f(x), f(y)) = \sum_{i \in I} \hat{\pi}(i) \mathcal{L}_{\pi_i} f.$$

By Jensen's inequality and the convexity of  $\Psi$  in all three cases,

$$\sum_{x \in \Lambda_i} \sum_{y \in \Lambda_j} \kappa_{ij}(x,y) \Psi(f(x), f(y)) \geq \Psi(\hat{f}(i), \hat{f}(j)).$$

For any  $(x, y, i, j) \in \Lambda_i \times \Lambda_j \times I^2$  with  $x \neq y$  and  $\hat{Q}(i, j) > 0$ , by the definition of  $\chi$  we have

$$a_i(x) a_j(y) \pi(x) Q(x,y) \Psi(f(x), f(y)) \geq \chi \hat{\pi}(i) \hat{Q}(i, j) \kappa_{ij}(x,y) \Psi(f(x), f(y)).$$

This holds true even for  $x = y$  or for  $(i, j)$  such that  $\hat{Q}(i, j) \leq 0$ , so

$$\begin{aligned} & \sum_{x,y \in \Omega} a_i(x) a_j(y) \pi(x) Q(x,y) \Psi(f(x), f(y)) \geq \chi \hat{\pi}(i) \hat{Q}(i, j) \Psi(\hat{f}(i), \hat{f}(j)) \\ \implies & \frac{1}{2} \sum_{i \in I} \sum_{j \neq i} \sum_{x,y \in \Omega} a_i(x) a_j(y) \pi(x) Q(x,y) \Psi(f(x), f(y)) \end{aligned}$$

and thus

$$\frac{1}{2} \sum_{i \in I} \sum_{j \neq i} \sum_{x,y \in \Omega} a_i(x) a_j(y) \pi(x) Q(x,y) \Psi(f(x), f(y)) \geq \frac{1}{2} \sum_{i,j \in I} \chi \hat{\pi}(i) \hat{Q}(i, j) \Psi(\hat{f}(i), \hat{f}(j)) = \chi \mathcal{L}_{\hat{\pi}} \hat{f}$$

In summary,

$$\begin{aligned}\mathcal{R}_\pi f &= \sum_{i \in I} \hat{\pi}(i) \mathcal{R}_{\pi_i} f + \mathcal{R}_{\hat{\pi}} \hat{f} \\ \mathcal{L}_\pi f &\geq \sum_{i \in I} \hat{\pi}(i) \mathcal{L}_{\pi_i} f + \chi \mathcal{L}_{\hat{\pi}} \hat{f}.\end{aligned}$$

For all  $\hat{\tau}, \tau_i \geq 0$  such that  $\mathcal{L}_{\hat{\pi}}(\hat{f}) \geq \hat{\tau} \mathcal{R}_{\hat{\pi}}(\hat{f})$  and  $\mathcal{L}_{\pi_i}(f) \geq \tau_i \mathcal{R}_{\pi_i}(f)$  for each  $i \in I$ , we have

$$\mathcal{L}_\pi(f) \geq \sum_{i \in I} \hat{\pi}(i) \tau_i \mathcal{R}_{\pi_i}(f) + \chi \hat{\tau} \mathcal{R}_{\hat{\pi}}(\hat{f}) \geq \min\{\chi \hat{\tau}, \min_{i \in I} \tau_i\} \mathcal{R}_\pi(f),$$

and the conclusion readily follows.  $\square$

## References

- [1] R. van Handel, “Probability in high dimension,” *Lecture Notes (Princeton University)*, 2014.
- [2] S. G. Bobkov and P. Tetali, “Modified logarithmic sobolev inequalities in discrete settings,” *Journal of Theoretical Probability*, vol. 19, pp. 289–336, 2006.
- [3] P. Diaconis and L. Saloff-Coste, “Logarithmic sobolev inequalities for finite markov chains,” *The Annals of Applied Probability*, vol. 6, no. 3, pp. 695–750, 1996.
- [4] D. A. Levin and Y. Peres, “Markov chains and mixing times, second edition,” *American Mathematical Society*, 2017.
- [5] M. Ledoux, “Concentration of measure and logarithmic sobolev inequalities,” in *Seminaire de probabilites XXXIII*, pp. 120–216, Springer, 2006.
- [6] M. Jerrum, J.-B. Son, P. Tetali, and E. Vigoda, “Elementary bounds on poincaré and log-sobolev constants for decomposable markov chains,” *The Annals of Applied Probability*, vol. 14, pp. 1741–1765, 2004.
- [7] S. L. Lu and H.-T. Yau, “Spectral gap and logarithmic sobolev inequality for kawasaki and glauber dynamics,” *Communications in Mathematical Physics*, vol. 156, no. 2, pp. 399–433, 1993.
- [8] J. Hermon and J. Salez, “Modified log-sobolev inequalities for strong-rayleigh measures,” *The Annals of Applied Probability*, vol. 33, no. 2, pp. 1501–1514, 2023.
- [9] R. Lowen, *Fuzzy Set Theory : Basic Concepts, Techniques and Bibliography*. Springer Netherlands, 2010.
- [10] N. Madras and D. Randall, “Markov chain decomposition for convergence rate analysis,” *The Annals of Applied Probability*, pp. 581–606, 2002.