Heat Flow and Concentration of Measure on Directed Graphs with a Lower Ricci Curvature Bound

Ryunosuke Ozawa¹ · Yohei Sakurai² · Taiki Yamada³

Received: 11 March 2021 / Accepted: 3 March 2022 / Published online: 28 March 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
In a previous work (Ozawa et al. Calc. Var. Partial Diff. Equ. 59(4), 39 2020), the authors introduced a Lin-Lu-Yau type Ricci curvature for directed graphs referring to the formulation of the Chung Laplacian. The aim of this note is to provide a von Renesse-Sturm type characterization of our lower Ricci curvature bound via a gradient estimate for the heat semigroup, and a transportation inequality along the heat flow. As an application, we will conclude a concentration of measure inequality for directed graphs of positive Ricci curvature.

Keywords Directed graph · Ricci curvature · Gradient estimate · Heat flow · Concentration of measure · Functional inequality

Mathematics Subject Classification (2010) Primary 05C20 · 05C12 · 05C81 · 53C21 · 53C23

Yohei Sakurai
ysakurai@rimath.saitama-u.ac.jp

Ryunosuke Ozawa
rozawa@nda.ac.jp

Taiki Yamada
taiki.yamada@riko.shimane-u.ac.jp

¹ Department of Mathematics, National Defense Academy of Japan, 1-10-20 Hashirimizu, Yokosuka, 239-8686, Japan

² Department of Mathematics, Saitama University, 255 Shimo-Okubo, Sakura-ku, Saitama-City, Saitama, 338-8570, Japan

³ Interdisciplinary Faculty of Science and Engineering, Shimane University, 1060 Nishikawatsu-cho, Matsue, Shimane, 690-8504, Japan
1 Introduction

1.1 Main Results

On a smooth Riemannian manifold, von Renesse-Sturm [14] have characterized a lower Ricci curvature bound in terms of a gradient estimate for the heat semigroup, and a transportation inequality along the heat flow. Münch-Wojciechowski [11] have studied its discrete analogue, and produced a characterization of a lower bound of the Ricci curvature introduced by Lin-Lu-Yau [10] on (undirected) graphs (see also [12]).

In [13], the authors have introduced a Lin-Lu-Yau type Ricci curvature for directed graphs inspired by the formulation of the Chung Laplacian ([3], [4]). In the present note, we aim to extend the results of Münch-Wojciechowski [11] to our directed setting. Let us introduce our main results (see Section 2 for the precise meaning of the notations). Let \((V, \mu)\) be a simple, strongly connected, finite weighted directed graph, where \(V\) is the vertex set, and \(\mu : V \times V \to [0, \infty)\) is the (non-symmetric) edge weight. We denote by \(d : V \times V \to [0, \infty)\) the (non-symmetric) graph distance function on \(V\), and by \(W\) the Wasserstein distance function. For \(x, y \in V\) with \(x \neq y\), let \(\kappa(x, y)\) stand for the Ricci curvature introduced in [13]. For \(f : V \to \mathbb{R}\), let \(\text{Lip} f\) denote its Lipschitz constant. We denote by \(L\) the Chung Laplacian, and by \(P_t\) the heat semigroup

\[
P_t \coloneqq e^{-tL}.
\]  

(1.1)

Furthermore, let \(p^t_x\) be the associated heat kernel measure at \(x\).

In [13], the authors have derived a representation formula for the Ricci curvature in terms of the Chung Laplacian (see Theorem 2.4 below). With the help of such a formula, we first prove the following characterization theorem in our directed setting:

**Theorem 1.1** Let \((V, \mu)\) denote a simple, strongly connected, finite weighted directed graph. For \(K \in \mathbb{R}\), the following are equivalent:

1. \(\inf_{x \neq y} \kappa(x, y) \geq K\);
2. for all \(f : V \to \mathbb{R}\) and \(t > 0\),

\[
\text{Lip} P_t f \leq e^{-Kt} \text{Lip} f;
\]

(1.2)

3. for all \(x, y \in V\) and \(t > 0\),

\[
W(p^t_x, p^t_y) \leq e^{-Kt} d(x, y).
\]

(1.3)

Theorem 1.1 has been obtained by Münch-Wojciechowski [11] in the undirected case (see [11, Theorem 3.8]). Notice that in the undirected case, they have proven it not only for finite graphs but also for infinite graphs.

Having Theorem 1.1 at hand, we will investigate the concentration of measure phenomena with respect to the Perron measure \(\mu\). Due to the lack of symmetry of the distance function, we need to introduce the following notion: For \(x, y \in V\), we set

\[
\mathcal{D}(x, y) := \max\{d(x, y), d(y, x)\}, \quad \mathcal{D}_x := \sup_{y \in \mathcal{N}_x} \mathcal{D}(x, y),
\]

where \(\mathcal{N}_x\) is the neighborhood of \(x\). Note that \(\mathcal{D}_x \geq 1\) in general. In the undirected case, the distance function \(d\) is symmetric, and hence we possess \(\mathcal{D}_x = 1\). Taking the asymmetry into account, we conclude the following concentration inequality for directed graphs:
Theorem 1.2 Let $(V, \mu)$ denote a simple, strongly connected, finite weighted directed graph. For $K > 0$ and $\Lambda \geq 1$, we assume $\inf_{x \neq y} \kappa(x, y) \geq K$ and $\sup_{x \in V} D_x \leq \Lambda$. Then for every 1-Lipschitz function $f : V \to \mathbb{R}$ we have

$$m(\{f \geq m(f) + r\}) \leq e^{-\frac{Kr^2}{\Lambda^2}},$$

where $m(f)$ is the mean of $f$ defined as

$$m(f) := \sum_{x \in V} f(x)m(x).$$

In the undirected case, Jost-Münch-Rose [8] have shown Theorem 1.2 based on the method of [15] (see [8, Theorem 3.1]). Also, Fathi-Shu [6] have (implicitly) shown a similar result for reversible Markov chains via functional inequalities (see [6, Theorems 1.13 and 2.4], and cf. [5]).

1.2 Organization

In Section 2, we will review basics of directed graphs. In Section 3, we prove Theorem 1.1. In Sections 4 and 5, we prove Theorem 1.2 in two different ways. In Section 4, we do it by following the argument of [15] as in [8]. In Section 5, we examine functional inequalities such as transportation-information inequality and transportation-entropy inequality, and apply them to another proof of Theorem 1.2 as in [6].

2 Preliminaries

We review basics on directed graphs. We refer to [13].

2.1 Directed Graphs

Let $(G, \mu)$ be a finite weighted directed graph, namely, $G = (V, E)$ is a finite directed graph, and $\mu : V \times V \to [0, \infty)$ is a function such that $\mu(x, y) > 0$ if and only if $x \to y$, where $x \to y$ means $(x, y) \in E$. The function $\mu$ is called the edge weight, and we write $\mu(x, y)$ by $\mu_{xy}$. Note that $(G, \mu)$ is undirected if and only if $\mu_{xy} = \mu_{yx}$ for all $x, y \in V$, and simple if and only if $\mu_{xx} = 0$ for all $x \in V$. It is also called unweighted if $\mu_{xy} = 1$ whenever $x \to y$. The weighted directed graph can be written as $(V, \mu)$ since the full information of $E$ is included in $\mu$.

For $x \in V$, its outer neighborhood $N_x$, inner one $\overset{\leftarrow}{N}_x$, and neighborhood $\mathcal{N}_x$ are defined as

$$N_x := \{y \in V \mid x \to y\}, \quad \overset{\leftarrow}{N}_x := \{y \in V \mid y \to x\}, \quad \mathcal{N}_x := N_x \cup \overset{\leftarrow}{N}_x,$$

respectively.

A sequence $\{x_i\}_{i=0}^l$ of vertices is called a directed path from $x$ to $y$ if $x_0 = x$, $x_l = y$ and $x_i \to x_{i+1}$ for all $i = 0, \ldots, l - 1$, where $l$ is called its length. $(V, \mu)$ is said to be strongly connected if for any $x, y \in V$, there is a directed path from $x$ to $y$. For strongly connected $(V, \mu)$, the (non-symmetric) distance function $d : V \times V \to [0, \infty)$ is defined as follows:

$$d(x, y) = \text{the minimum of the lengths of directed paths from } x \text{ to } y.$$

For $f : V \to \mathbb{R}$, its Lipschitz constant is defined by

$$\text{Lip } f := \sup_{x \neq y} \nabla_{xy} f,$$
where $\nabla_{xy}$ is the gradient operator defined as

$$\nabla_{xy} f := \frac{f(y) - f(x)}{d(x, y)}.$$ 

For $L > 0$, $f$ is said to be $L$-Lipschitz if

$$\text{Lip } f \leq L.$$ 

Let $\text{Lip}_L(V)$ stand for the set of all $L$-Lipschitz functions.

### 2.2 Chung Laplacian

Hereafter, let $(V, \mu)$ be a simple, strongly connected, finite weighted directed graph. In this subsection, we review the formulation of the Chung Laplacian introduced in [3], [4]. The transition probability kernel $P : V \times V \to [0, 1]$ is defined as

$$P(x, y) := \frac{\mu_{xy}}{\mu(x)},$$

where

$$\mu(x) := \sum_{y \in V} \mu_{xy}.$$ 

Since $(V, \mu)$ is finite and strongly connected, the Perron-Frobenius theorem ensures that there is a unique (up to scaling) positive function $m : V \to (0, \infty)$ such that

$$m(x) = \sum_{y \in V} m(y) P(y, x). \quad (2.1)$$ 

A probability measure $m : V \to (0, 1]$ satisfying (2.1) is called the Perron measure (or the stationary probability measure).

Let $m$ be the Perron measure. For a non-empty subset $\Omega \subset V$, we set

$$m(\Omega) := \sum_{x \in \Omega} m(x).$$

The reverse transition probability kernel $\bar{P} : V \times V \to [0, 1]$, and the mean transition probability kernel $\mathcal{P} : V \times V \to [0, 1]$ are defined as

$$\bar{P}(x, y) := \frac{m(y)}{m(x)} P(y, x), \quad \mathcal{P} := \frac{1}{2}(P + \bar{P}).$$

Let $\mathcal{F}$ be the set of all functions on $V$. The Chung Laplacian $\mathcal{L} : \mathcal{F} \to \mathcal{F}$ is given by

$$\mathcal{L} f(x) := f(x) - \sum_{y \in V} \mathcal{P}(x, y) f(y).$$

We will also use the negative Laplacian $\Delta : \mathcal{F} \to \mathcal{F}$ defined by

$$\Delta := -\mathcal{L}.$$ 

We define a function $m : V \times V \to [0, \infty)$ by

$$m(x, y) := \frac{1}{2}(m(x) P(x, y) + m(y) P(y, x)) = m(x) \mathcal{P}(x, y).$$

We write $m(x, y)$ by $m_{xy}$. The following properties hold: (1) $m_{xy} = m_{yx}$; (2) $m_{xy} > 0$ if and only if $y \in N_x$ (or equivalently, $x \in N_y$); (3) $\mathcal{P}(x, y) = m_{xy}/m(x)$. 

© Springer
The inner product on $F$ is defined by
\[(f_0, f_1) := \sum_{x \in V} f_0(x) f_1(x) m(x) = m(f_0 f_1).\]

Also, the $\Gamma$-operator is defined as
\[\Gamma(f_0, f_1) := \frac{1}{2} (\Delta(f_0 f_1) - f_0 \Delta f_1 - f_1 \Delta f_0), \quad \Gamma(f) := \Gamma(f, f).\]

By direct computations, we see the following (cf. [13, Proposition 7.1]):
\[\Gamma(f_0, f_1)(x) = \frac{1}{2} \sum_{y \in V} (f_0(y) - f_0(x))(f_1(y) - f_1(x)) P(x, y).\]

We possess the following integration by parts formula (see e.g., [13, Proposition 2.4]):

**Proposition 2.1** Let $\Omega \subset V$ be a non-empty subset. Then for all $f_0, f_1 : V \to \mathbb{R},$
\[\sum_{x \in \Omega} \mathcal{L}f_0(x) f_1(x) m(x) = \frac{1}{2} \sum_{x, y \in \Omega} (f_0(y) - f_0(x))(f_1(y) - f_1(x)) m_{xy} - \sum_{x \in \Omega} \sum_{y \in V \setminus \Omega} (f_0(y) - f_0(x)) f_1(x) m_{xy}.
\]

In particular,
\[(\mathcal{L}f_0, f_1) = \sum_{x \in V} (\Gamma(f_0, f_1)) = (f_0, \mathcal{L}f_1).
\]

Let $P_t$ be the heat semigroup defined as Eq. 1.1 such that $P_tf$ solves
\[\begin{cases}
\partial_t u = \Delta u, \\
u|_{t=0} = f.
\end{cases}
\]

From Proposition 2.1 we can derive
\[(P_t f_0, f_1) = (f_0, P_t f_1). \tag{2.2}\]

The heat kernel measure $p^t_x$ is defined as
\[p^t_x := \frac{m}{m(x)} P_t \delta_x
\]
for the Dirac measure $\delta_x$ at $x$. The equality Eq. 2.2 yields
\[P_t f(x) = \sum_{y \in V} p^t_x(y) f(y). \tag{2.3}\]

Furthermore, we see $P_t 1_V = 1_V$ by Eq. 2.2, and hence $p^t_x$ is a probability measure.

**2.3 Optimal Transport Theory**

We next recall the basics of the optimal transport theory (cf. [16]). For two probability measures $\nu_0, \nu_1$ on $V$, a probability measure $\pi : V \times V \to [0, \infty)$ is called a coupling of $(\nu_0, \nu_1)$ if
\[\sum_{y \in V} \pi(x, y) = \nu_0(x), \quad \sum_{x \in V} \pi(x, y) = \nu_1(y).
\]
Let \( \Pi(v_0, v_1) \) stand for the set of all couplings of \((v_0, v_1)\). The Wasserstein distance from \(v_0\) to \(v_1\) is defined as

\[
W(v_0, v_1) := \inf_{\pi \in \Pi(v_0, v_1)} \sum_{x, y \in V} d(x, y)\pi(x, y),
\]

which is a (non-symmetric) distance function on the set of all probability measures on \(V\).

The following Kantorovich-Rubinstein duality formula is well-known (cf. [16, Theorem 5.10 and Particular Cases 5.4 and 5.16]):

**Proposition 2.2** For any two probability measures \(v_0, v_1\) on \(V\), we have

\[
W(v_0, v_1) = \sup_{f \in \text{Lip}_1(V)} \sum_{x \in V} f(x)(v_1(x) - v_0(x)).
\]

**Remark 2.3** In the discussion of [16, Particular Cases 5.4], the symmetry for cost functions is not required, and hence Proposition 2.2 holds in our directed setting.

### 2.4 Ricci Curvature

In this subsection, we recall the formulation of the Ricci curvature introduced in [13]. For \(\varepsilon \in [0, 1]\), and for \(x, y \in V\) with \(x \neq y\), we set

\[
\kappa_\varepsilon(x, y) := 1 - \frac{W(v_\varepsilon^x, v_\varepsilon^y)}{d(x, y)},
\]

where \(v_\varepsilon^x : V \to [0, 1]\) is a probability measure defined by

\[
v_\varepsilon^x(z) = (1 - \varepsilon)\delta_x(z) + \varepsilon \mathcal{P}(x, z).
\]

The authors [13] have introduced the Ricci curvature as follows (see [13, Definition 3.6]):

\[
\kappa(x, y) := \lim_{\varepsilon \to 0} \frac{\kappa_\varepsilon(x, y)}{\varepsilon},
\]

which is well-defined (see [13, Lemmas 3.2 and 3.4, and Definition 3.6]). In the undirected case, this is nothing but the Lin-Lu-Yau Ricci curvature in [10].

We have the following representation formula, which has been established by Münch-Wojciechowski [11] in the undirected case (see [13, Theorem 3.10] and [11, Theorem 2.1]):

**Theorem 2.4** [11, 13]

\[
\kappa(x, y) = \inf_{f \in \mathcal{F}_{xy}} \nabla_{xy} \mathcal{L} f,
\]

where

\[
\mathcal{F}_{xy} := \{ f \in \text{Lip}_1(V) \mid \nabla_{xy} f = 1 \}.
\]

### 3 Heat Flow

In this section, we give a proof of Theorem 1.1.

#### 3.1 Curvature Bound and Gradient Estimate

In this subsection, we show the equivalence of (1.1) and (1.1) in Theorem 1.1.
Proposition 3.1 For $K \in \mathbb{R}$, the following are equivalent:

1. $\inf_{x \neq y} \kappa(x, y) \geq K$;
2. for all $f : V \to \mathbb{R}$ and $t > 0$, we have Eq. 1.2.

Proof We begin with the implication from (3.1) to (3.1). For a fixed $x, y \in V$, and $f : V \to \mathbb{R}$, it is enough to prove

$$\nabla_{xy} P_t f \leq e^{-Kt} \text{Lip } f$$

for every $t > 0$. If $f$ is constant, then this immediately follows from $P_t 1_V = 1_V$. Thus we may assume that $f$ is non-constant; in particular, $\text{Lip } f > 0$. We will prove it by contradiction. Suppose that there exists $t > 0$ such that

$$e^{Kt} \nabla_{xy} P_t f > \text{Lip } f.$$ 

We here notice that

$$e^{Kt} \nabla_{xy} P_t f |_{t=0} = \nabla_{xy} f \leq \text{Lip } f.$$ 

In this case, there is $t_0 > 0$ such that

$$e^{Kt_0} \nabla_{xy} P_{t_0} f > \text{Lip } f, \quad \partial_t (e^{Kt} \nabla_{xy} P_t f)|_{t=t_0} > 0.$$ 

Due to Theorem 2.4,

$$0 < \frac{e^{-Kt_0}}{\nabla_{xy} P_{t_0} f} \partial_t (e^{Kt} \nabla_{xy} P_t f)|_{t=t_0} = \left( K - \nabla_{xy} L \frac{P_{t_0} f}{\nabla_{xy} P_{t_0} f} \right) \leq (K - \kappa(x, y)),$$ 

and hence $\kappa(x, y) < K$. This contradicts with the curvature bound.

We next show the opposite direction. By Eq. 1.2, for any $f \in \mathcal{F}_{xy}$,

$$\nabla_{xy} L f = -\partial_t \nabla_{xy} P_t f |_{t=0} = \lim_{t \to 0} \frac{1}{t} (\nabla_{xy} f - \nabla_{xy} P_t f) \geq \lim_{t \to 0} \frac{1}{t} (1 - e^{-Kt}) = K,$$ 

here we used $\text{Lip } f = 1$. Theorem 2.4 leads us to the desired conclusion. \hfill \Box

3.2 Gradient Estimate and Transportation Inequality

Next, we prove the equivalence of (1.1) and (1.1), and conclude Theorem 1.1.

Proposition 3.2 For $K \in \mathbb{R}$, the following are equivalent:

1. For all $f : V \to \mathbb{R}$ and $t > 0$, we have Eq. 1.2;
2. for all $x, y \in V$ and $t > 0$, we have Eq. 1.3.

Proof We start with the implication from (3.2) to (3.2). By Eqs. 2.3 and 1.2, it holds that

$$\sum_{z \in V} f(z) (p^t_{y}(z) - p^t_{x}(z)) = P_t f(y) - P_t f(x) \leq \text{Lip } P_t f d(x, y) \leq e^{-Kt} \text{Lip } f d(x, y) \leq e^{-Kt} d(x, y)$$

for every $f \in \text{Lip}_1(V)$. With the help of Proposition 2.2, we arrive at Eq. 1.3.

We now consider the opposite one. We fix $f : V \to \mathbb{R}$. If $f$ is constant, then the desired assertion is trivial by $P_t 1_V = 1_V$. If $f$ is non-constant, then $\text{Lip } f$ is positive, and hence we
can define \( g \in \text{Lip}_1(V) \) by
\[
g := \frac{f}{\text{Lip} f}.
\]
Therefore, Eq. 2.3, Proposition 2.2 and Eq. 1.3 imply
\[
P_t f(y) - P_t f(x) = \text{Lip} f (P_t g(y) - P_t g(x)) = \text{Lip} f \sum_{z \in V} g(z) (p^t_y(z) - p^t_x(z))
\leq \text{Lip} f W(p^t_x, p^t_y) \leq e^{-Kt} \text{Lip} f d(x, y).
\]
By dividing the both sides by \( d(x, y) \), we complete the proof. \( \square \)

We are now in a position to conclude Theorem 1.1.

**Proof of Theorem 1.1** Theorem 1.1 is a direct consequence of Propositions 3.1 and 3.2. \( \square \)

### 3.3 Characterization via Heat Flow

In the undirected case, Münch-Wojciechowski [11] have formulated not only the characterization of lower Ricci curvature bound but also that of Ricci curvature itself (see [11, Theorem 5.8]). In our directed setting, we also have the following characterization of Münch-Wojciechowski type:

**Theorem 3.3**
\[
\kappa(x, y) = \lim_{t \to 0} \frac{1}{t} \left( 1 - \frac{W(p^t_x, p^t_y)}{d(x, y)} \right).
\]

We can prove Theorem 3.3 by the same argument as in the proof of [11, Theorem 5.8]. We omit the proof.

### 4 Concentration of Measure

In this section, we will prove Theorem 1.2 along the line of the proof of [8, Theorem 3.1].

#### 4.1 Laplace Functional Estimates

For \( \lambda \geq 0 \), the Laplace functional (or moment generating functional) is defined as follows (see e.g., [9, Subsection 1.6], [2, Subsection 2.1]):
\[
E(\lambda) := \sup_{f \in \text{Lip}_1(V)} m(e^{\lambda f}),
\]
where the supremum is taken over all \( f \in \text{Lip}_1(V) \) with \( m(f) = 0 \). In this subsection, we give an upper bound of the Laplace functional under the same setting as in Theorem 1.2. To do so, we prepare the following lemma (cf. [15]):

**Lemma 4.1** For all \( \lambda \geq 0 \) and \( f : V \to \mathbb{R} \) we have
\[
m(\Gamma(f, e^{\lambda f})) \leq \lambda (e^{\lambda f}, \Gamma(f)) \).
\]
Proof We see
\[ m(\Gamma(f, e^{\lambda f})) = \frac{1}{2} \sum_{x, y \in V} (f(y) - f(x))(e^{\lambda f(y)} - e^{\lambda f(x)})m_{xy} \]
\[ = \sum_{f(y) > f(x)} (f(y) - f(x))(e^{\lambda f(y)} - e^{\lambda f(x)})m_{xy}. \]

We now recall the following elementary inequality (see e.g., [9, Corollary 5.8]): For all \( s > t \),
\[ \frac{e^s - e^t}{s - t} \leq \frac{e^s + e^t}{2}. \]

If \( f(y) > f(x) \), then
\[ \frac{e^{\lambda f(y)} - e^{\lambda f(x)}}{f(y) - f(x)} \leq \lambda \frac{e^{\lambda f(y)} + e^{\lambda f(x)}}{2}. \]

It follows that
\[ m(\Gamma(f, e^{\lambda f})) \leq \frac{\lambda}{2} \sum_{f(y) > f(x)} \left( e^{\lambda f(y)} + e^{\lambda f(x)} \right) (f(y) - f(x))^2 m_{xy} \]
\[ = \frac{\lambda}{2} \sum_{x, y \in V} e^{\lambda f(y)} (f(y) - f(x))^2 m_{xy} = \lambda(e^{\lambda f}, \Gamma(f)). \]

This proves the lemma. \( \Box \)

We now state the desired assertion:

**Proposition 4.2** For \( K > 0 \) and \( \Lambda \geq 1 \), we assume \( \inf_{x \neq y} \kappa(x, y) \geq K \) and \( \sup_{x \in V} D_x \leq \Lambda \). Then for every \( \lambda \geq 0 \),
\[ E(\lambda) \leq e^{\frac{\lambda^2 \Lambda^2}{4K}}. \]

**Proof** Let \( f \in \text{Lip}_1(V) \) with \( m(f) = 0 \). Using Proposition 2.1 and Lemma 4.1, we obtain
\[ \partial_t m(e^{\lambda P_t f}) = -\lambda (\mathcal{L} P_t f, e^{\lambda P_t f}) = -\lambda m(\Gamma(P_t f, e^{\lambda P_t f})) \geq -\lambda^2 (e^{\lambda P_t f}, \Gamma(P_t f)). \]

By virtue of Theorem 1.1,
\[ \Gamma(P_t f) = \frac{1}{2} \sum_{x, y \in V} (P_t f(y) - P_t f(x))^2 m_{xy} \]
\[ \leq \frac{1}{2} (\text{Lip} P_t f)^2 \sum_{x, y \in V} D(x, y)^2 m_{xy} \leq \frac{1}{2} \Lambda^2 e^{-2Kt} (\text{Lip} f)^2 \leq \frac{1}{2} \Lambda^2 e^{-2Kt}. \]

By combining the above inequalities,
\[ \partial_t m(e^{\lambda P_t f}) \geq -\frac{\lambda^2 \Lambda^2}{2} e^{-2Kt} m(e^{\lambda P_t f}), \]
and hence
\[ \log m(e^{\lambda P_t f}) - \log m(e^{\lambda f}) \geq \frac{\lambda^2 \Lambda^2}{4K} \left( e^{-2Kt} - 1 \right). \]

By letting \( t \to \infty \) we arrive at
\[ m(e^{\lambda f}) \leq e^{\frac{\lambda^2 \Lambda^2}{4K}}. \]

Here we used \( m(e^{\lambda P_t f}) \to 1 \) as \( t \to \infty \), which is a consequence of \( m(f) = 0 \) and the fact that \( m \) is a probability measure. This completes the proof. \( \Box \)
4.2 Concentration Inequalities

Let us recall the following Chernoff bounding method (see e.g., [9, Proposition 1.14], [2, Subsection 2.1]):

**Proposition 4.3** Let $c > 0$. If

$$E(\lambda) \leq e^{\frac{\lambda^2}{2c}}$$

for all $\lambda \geq 0$, then

$$m(\{ f \geq m(f) + r \}) \leq e^{-\frac{c r^2}{2}}.$$

Now, one can now derive Theorem 1.2.

**Proof of Theorem 1.2** Proposition 4.2 together with Proposition 4.3 with $c = 2K/\Lambda^2$ implies Theorem 1.2.

5 Functional Inequalities

Here we discuss several functional inequalities, and give another proof of Theorem 1.2.

5.1 Transportation-Information Inequality

In this subsection, we examine a transportation-information inequality. We first show the following lemma (cf. [6, Lemma 5.1]):

**Lemma 5.1** For $K > 0$ and $\Lambda \geq 1$, we assume $\inf_{x \neq y} \kappa(x, y) \geq K$ and $\sup_{x \in V} D_x \leq \Lambda$. Then for every probability density $\rho : V \to [0, \infty)$ (i.e., $m(\rho) = 1$) we have

$$W(m, \rho m) \leq \frac{\Lambda}{2K} \sum_{x, y \in V} |\rho(y) - \rho(x)| m_{xy}.$$

**Proof** Proposition 2.2 can be written as

$$W(m, \rho m) = \sup_{f \in \text{Lip}_1(V)} \int_0^\infty \frac{d}{dt} (P_t f, \rho) dt = \sup_{f \in \text{Lip}_1(V)} \int_0^\infty m(\Gamma(P_t f, \rho)) dt.$$

Let us fix $f \in \text{Lip}_1(V)$ with $m(f) = 0$. Theorem 1.1 tells us that

$$\int_0^\infty m(\Gamma(P_t f, \rho)) dt \leq \frac{1}{2} \int_0^\infty \sum_{x, y \in V} (P_t f(y) - P_t f(x))(\rho(y) - \rho(x)) m_{xy} dt$$

$$\leq \frac{1}{2} \int_0^\infty \text{Lip} P_t f dt \sum_{x, y \in V} D(x, y)|\rho(y) - \rho(x)| m_{xy}$$

$$\leq \frac{\Lambda}{2K} \sum_{x, y \in V} |\rho(y) - \rho(x)| m_{xy}.$$

This proves the lemma.
For a probability density \( \rho : V \to [0, \infty) \), the Fisher information is defined by
\[
\mathcal{I}(\rho) := 4m(\Gamma(\sqrt{\rho})) = 2 \sum_{x, y \in V} \left( \sqrt{\rho(y)} - \sqrt{\rho(x)} \right)^2 m_{xy}.
\]

We now state our desired inequality (cf. [6, Theorem 1.13]).

**Theorem 5.2** For \( K > 0 \) and \( \Lambda \geq 1 \), we assume \( \inf_{x \neq y} \kappa(x, y) \geq K \) and \( \sup_{x \in V} D_x \leq \Lambda \). Then for every probability density \( \rho : V \to [0, \infty) \) we have
\[
W(\rho, \rho m)^2 \leq \frac{\Lambda^2}{2K^2} \mathcal{I}(\rho) \left( 1 - \frac{1}{8} \mathcal{I}(\rho) \right) \leq \frac{\Lambda^2}{2K^2} \mathcal{I}(\rho).
\]

**Proof** Fathi-Shu [6] have proved a similar result for reversible Markov chains (see [6, Theorem 1.13]). We will prove it along the line of their argument. Since \( \rho \) is a probability density, we see
\[
\sum_{x, y \in V} \left( \sqrt{\rho(y)} + \sqrt{\rho(x)} \right)^2 m_{xy} = \sum_{x, y \in V} \left( 2\rho(y) + 2\rho(x) - \left( \sqrt{\rho(y)} - \sqrt{\rho(x)} \right)^2 \right) m_{xy}
= 4 - \frac{1}{2} \mathcal{I}(\rho).
\]

From Lemma 5.1 we deduce
\[
W(\rho, \rho m) \leq \frac{\Lambda}{2K} \sum_{x, y \in V} \| \rho(y) - \rho(x) \| m_{xy}
= \frac{\Lambda}{2K} \sum_{x, y \in V} \left| \sqrt{\rho(y)} - \sqrt{\rho(x)} \right| \left( \sqrt{\rho(y)} + \sqrt{\rho(x)} \right) m_{xy}
\leq \frac{\Lambda}{2K} \sqrt{\mathcal{I}(\rho)} \sqrt{\frac{1}{2} \sum_{x, y \in V} \left( \sqrt{\rho(y)} + \sqrt{\rho(x)} \right)^2 m_{xy}} \leq \frac{\Lambda}{2K} \sqrt{\mathcal{I}(\rho)} \sqrt{2 - \frac{1}{4} \mathcal{I}(\rho)}.
\]

This completes the proof. \( \square \)

### 5.2 Transportation-Entropy Inequality

We next investigate a transportation-entropy inequality. For a probability density \( \rho : V \to [0, \infty) \), its relative entropy is defined by
\[
\mathcal{E}(\rho) := m(\rho \log \rho).
\]

We notice the following characterization (see e.g., [9, (5.13)]):
\[
\mathcal{E}(\rho) = \sup_g \{ g, \rho \}, \quad (5.1)
\]
where the supremum is taken over all \( g : V \to \mathbb{R} \) with \( m(e^g) \leq 1 \).

We verify the following Bobkov-Götze type criterion due to the lack of symmetry of the distance function (cf. [1, Theorem 1.3], [9, Proposition 6.1]):

**Lemma 5.3** For \( c > 0 \), the following are equivalent:

\( \square \) Springer
For every probability density \( \rho : V \to [0, \infty) \) we have
\[
W(m, \rho m)^2 \leq \frac{2}{c} \mathcal{E}(\rho); \tag{5.2}
\]

(2) for all \( \lambda \geq 0 \),
\[
E(\lambda) \leq e^{\frac{\lambda^2}{2c}}. \tag{5.3}
\]

Proof Let us show the implication from (5.3) to (5.3). Fix \( f \in \text{Lip}_1(V) \) with \( m(f) = 0 \), and set
\[
g_f := \lambda f - \frac{\lambda^2}{2c}, \quad \rho_f := \frac{e^{g_f}}{m(e^{g_f})}. \tag{5.4}
\]
By Proposition 2.2,
\[
(f, \rho_f) \leq W(m, \rho_f m) \leq \sqrt{\frac{2}{c} \mathcal{E}(\rho_f)} \leq \frac{\lambda}{2c} + \frac{1}{\lambda} \mathcal{E}(\rho_f),
\]
and hence \( (g_f, \rho_f) \leq \mathcal{E}(\rho_f) \). On the other hand, straightforward computations imply
\[
\mathcal{E}(\rho_f) = (g_f, \rho_f) - \log m(e^{g_f}).
\]
Therefore, \( m(e^{g_f}) \leq 1 \), which is equivalent to
\[
m(e^{\lambda f}) \leq e^{\frac{\lambda^2}{2c}}.
\]
We have shown the desired estimate.

We prove the opposite one. Fix a probability density \( \rho \). Proposition 2.2 can be written as
\[
W(m, \rho m) = \sup_{f \in \text{Lip}_1(V)} (f, \rho), \tag{5.5}
\]
where the supremum is taken over all \( f \in \text{Lip}_1(V) \) with \( m(f) = 0 \). We also fix such a \( f \), and define \( g_f \) as Eq. 5.4. From Eq. 5.3 we derive \( m(e^{g_f}) \leq 1 \). In view of Eq. 5.1,
\[
\lambda(f, \rho) - \frac{\lambda^2}{2c} = (g_f, \rho) \leq \mathcal{E}(\rho)
\]
By letting \( \lambda \to \sqrt{2c \mathcal{E}(\rho)} \), we arrive at
\[
(f, \rho) \leq \sqrt{\frac{2}{c} \mathcal{E}(\rho)}. \tag{5.6}
\]
Combining Eqs. 5.5 and 5.6, we obtain Eq. 5.2. We complete the proof. \( \square \)

Proposition 4.2 together with Lemma 5.3 yields the following transportation-entropy inequality:

**Theorem 5.4** For \( K > 0 \) and \( \Lambda \geq 1 \), we assume \( \inf_{x \neq y} \kappa(x, y) \geq K \) and \( \sup_{x \in V} D_x \leq \Lambda \). Then for every probability density \( \rho : V \to [0, \infty) \) we have
\[
W(m, \rho m)^2 \leq \frac{2\Lambda^2}{K} \mathcal{E}(\rho).
\]
5.3 Relation Between Functional Inequalities

We finally mention the relation between the transportation-information inequality and the transportation-entropy inequality. In order to do so, we prepare the following (cf. [6, Lemma 2.3]):

**Lemma 5.5** For every $f : V \to \mathbb{R}$ we have

$$m(\Gamma(e^f)) \leq (e^{2f}, \Gamma(f)).$$

**Proof** It holds that

$$m(\Gamma(e^f)) = \frac{1}{2} \sum_{x,y \in V} (e^{f(y)} - e^{f(x)})^2 m_{xy} = \sum_{f(y) > f(x)} (e^{f(y)} - e^{f(x)})^2 m_{xy}.$$

If $f(y) > f(x)$, then we see

$$\frac{e^{f(y)} - e^{f(x)}}{f(y) - f(x)} \leq \frac{e^{f(y)} + e^{f(x)}}{2}.$$

We also notice that

$$\left(\frac{e^{f(y)} + e^{f(x)}}{2}\right)^2 = \frac{e^{2f(y)} + e^{2f(x)}}{2} - \left(\frac{e^{f(y)} - e^{f(x)}}{2}\right)^2 \leq \frac{e^{2f(y)} + e^{2f(x)}}{2}.$$

It follows that

$$m(\Gamma(e^f)) \leq \sum_{f(y) > f(x)} \left(\frac{e^{2f(y)} + e^{2f(x)}}{2}\right) (f(y) - f(x))^2 m_{xy}$$

$$= \frac{1}{2} \sum_{x,y \in V} e^{2f(y)} (f(y) - f(x))^2 m_{xy} = (e^{2f}, \Gamma(f)).$$

This proves the lemma.

We possess the following relation (cf. [6, Theorem 2.4], [7, Theorem 2.1]):

**Proposition 5.6** For $\Lambda \geq 1$, we assume $\sup_{x \in V} D_x \leq \Lambda$. Let $c > 0$, and let $\rho : V \to [0, \infty)$ be a probability density. If

$$W(\rho, \rho^m)^2 \leq \frac{1}{c^2} \mathcal{I}(\rho),$$

then

$$W(\rho, \rho^m)^2 \leq \frac{\sqrt{2}\Lambda}{c} \mathcal{E}(\rho).$$

**Proof** Fix $f \in \text{Lip}_1(V)$ with $m(f) = 0$, and set

$$\rho_{f,\lambda} := \frac{e^{\lambda f}}{m(e^{\lambda f})}.$$
From Proposition 2.2, Eq. 5.7, and Lemma 5.5, it follows that
\[
\frac{d}{d\lambda} \log m(e^{\lambda f}) = (f, \rho_{f,\lambda} m) \leq W(m, \rho_{f,\lambda} m) \leq \sqrt{\frac{4}{c^2} \frac{1}{m(e^{\lambda f})} m(\Gamma(e^{\lambda f}))} \leq \sqrt{\frac{\lambda^2}{c^2} \frac{1}{m(e^{\lambda f})} (e^{\lambda f}, \Gamma(\frac{\lambda f}{2}))} = \sqrt{\frac{\lambda^2}{c^2} \frac{1}{m(e^{\lambda f})} (e^{\lambda f}, \Gamma(f))}.
\]

Now, we have
\[
2\Gamma(f)(x) = \sum_{y \in V} (f(y) - f(x))^2 P(x, y) \leq \sum_{y \in V} D(x, y)^2 P(x, y) \leq \Lambda^2,
\]
and hence
\[
\frac{d}{d\lambda} \log m(e^{\lambda f}) \leq \frac{\Lambda \lambda}{\sqrt{2} c}.
\]
Integrating both sides, we obtain
\[
E(\lambda) \leq e^{\frac{\Lambda \lambda^2}{2 c}}.
\]
Thanks to Lemma 5.3, we arrive at the desired inequality.

We are now in a position to provide another proof of Theorem 1.2.

\textbf{Another proof of Theorem 1.2} Theorem 5.2 together with Lemma 5.3 and Proposition 5.6 with \(c = \sqrt{2} K / \Lambda\) implies the same conclusion as in Proposition 4.2. Thus, we conclude Theorem 1.2 due to Proposition 4.3.

\textbf{Acknowledgements} The authors are grateful to the anonymous referees for valuable comments. The first named author was supported in part by JSPS KAKENHI (19K14532). The first and second named authors were supported in part by JSPS Grant-in-Aid for Scientific Research on Innovative Areas “Discrete Geometric Analysis for Materials Design” (17H06460). The third named author was supported in part by JSPS KAKENHI (19K23411).

\textbf{References}

1. Bobkov, S.G., Götze, F.: Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. J. Funct. Anal. 163(1), 1–28 (1999)
2. Boucheron, S., Lugosi, G., Massart, P.: Concentration Inequalities. A Nonasymptotic Theory of Independence, With a Foreword by Michel Ledoux. Oxford University Press, Oxford (2013)
3. Chung, F.: Laplacians and the Cheeger inequality for directed graphs. Ann. Comb. 9(1), 1–19 (2005)
4. Chung, F.: The diameter and Laplacian eigenvalues of directed graphs. Electron. J. Combin. 13(1), 6 (2006). Note 4
5. Elden, R., Lee, J.R., Lehec, J.: Transport-Entropy Inequalities and Curvature in Discrete-Space Markov Chains, A Journey Through Discrete Mathematics, pp. 391–406. Springer, Cham (2017)
6. Fathi, M., Shu, Y.: Curvature and transport inequalities for Markov chains in discrete spaces. Bernoulli 24(1), 672–698 (2018)
7. Guillin, A., Léonard, C., Wang, F.-Y., Wu, L.: Transportation-information inequalities for Markov processes (II): Relations with other functional inequalities preprint. arXiv:0902.2101
8. Jost, J., Münch, F., Rose, C.: Liouville property and non-negative Ollivier curvature on graphs, preprint arXiv:1903.10796
9. Ledoux, M.: The Concentration of Measure Phenomenon, Mathematical Surveys and Monographs, vol. 89. American Mathematical Society, Providence (2001)
10. Lin, Y., Lu, L., Yau, S.-T.: Ricci curvature of graphs. Tohoku Math. J. (2) 63(4), 605–627 (2011)
11. M"unch, F., Wojciechowski, R.K.: Ollivier Ricci curvature for general graph Laplacians: Heat equation, Laplacian comparison, non-explosion and diameter bounds. Adv. Math. \textbf{356}(106759), 45 (2019)
12. Ollivier, Y.: Ricci curvature of Markov chains on metric spaces. J. Funct. Anal. \textbf{256}(3), 810–864 (2009)
13. Ozawa, R., Sakurai, Y., Yamada, T.: Geometric and spectral properties of directed graphs under a lower Ricci curvature bound. Calc. Var. Partial Diff. Equ. \textbf{59}(4), 39 (2020). Paper No. 142
14. von Renesse, M.-K., Sturm, K.-T.: Transport inequalities, gradient estimates, entropy, and Ricci curvature. Comm. Pure Appl. Math. \textbf{58}(7), 923–940 (2005)
15. Schmuckenschl"ager, M.: Curvature of nonlocal Markov generators, Convex geometric analysis (Berkeley, CA, 1996), 189–197, Math. Sci. Res. Inst Publ., vol. 34. Cambridge Univ. Press, Cambridge (1999)
16. Villani, C.: Optimal Transport: Old and New. Springer, Berlin (2009)

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.