[The sixth Workshop on Markov processes and related topics, Anhui Normal University, 21-26 July 2008]

Transportation-Information inequalities for Markov chains

Liming Wu

Wuhan University and Université Blaise Pascal

Based on a series of works :

[1] A. Guillin, Ch. Léonard, L. Wu and N. Yao, [I], to appear in PTRF

[2] A. Guillin, Ch. Léonard, F.Y. Wang and L. Wu, [II] relations with other functional inequalities. Preprint 08

[3] A. Guillin, A. Joulin, Ch. Léonard, and L. Wu, [III] jumps case. Preprint 08

[4] F. Gao and L. Wu, Gibbs measures, Preprint 07

[5] L. Wu, discrete time case. Preprint 08.

1. Three objects

- (E, d) is a separable, complete metric space with Borel field $\mathcal{B}$.

Given two probability measures ν, μ on $(E, \mathcal{B})$,

their **L^p-Wasserstein distance** is defined by

$$W_{p,d}(\nu, \mu) := \inf_{\pi} \left(\iint_{E \times E} d(x, y)^p \pi(dx, dy) \right)^{1/p} \tag{1}$$

where the infimum is taken over all probability measures π on $E \times E$ such that its marginal distributions are respectively ν and μ, i.e.,

$$\pi(A \times E) = \nu(A), \pi(E \times B) = \mu(B), \ \forall A, B \in \mathcal{B}.$$

Such π is called *coupling of* (ν, μ).

Definition 1 *the relative entropy (or the Kullback information) of ν w.r.t. μ is defined by*

$$H(\nu|\mu) := \begin{cases} \int \frac{d\nu}{d\mu} \log \frac{d\nu}{d\mu} d\mu, & \text{if } \nu \ll \mu; \\ +\infty, & \text{otherwise.} \end{cases} \tag{2}$$

For $0 \le f \in L^1(\mu)$, the entropy of f w.r.t. μ is defined as

$$Ent_\mu(f) := \int f \log f d\mu - \mu(f) \log \mu(f) \in [0, +\infty]. \tag{3}$$

Remarks 1 *$\nu \to H(\nu|\mu)$ is the rate function in the LDP of*

$$L_n := \frac{1}{n} \sum_{k=1}^{n} \delta_{X_k}$$

where (X_k) is i.i.d.r.v. of law μ.

Definition 2 *Given the Dirichlet form $\mathcal{E}$ with domain $\mathbb{D}(\mathcal{E})$ on $L^2(\mu)$, the Fisher-Donsker-Varadhan information of ν with respect to μ is defined by*

$$I(\nu|\mu) := \begin{cases} \mathcal{E}(\sqrt{f}, \sqrt{f}), & \text{if } \nu = f\mu, \sqrt{f} \in \mathbb{D}(\mathcal{E}) \\ +\infty, & \text{otherwise.} \end{cases} \tag{4}$$

Remarks 2 $\nu \mapsto I(\nu|\mu)$ is exactly the Donsker-Varadhan entropy i.e. the rate function governing the large deviation principle of the empirical measure

$$L_t := \frac{1}{t}\int_0^t \delta_{X_s} ds$$

for large time t, where (X_t) is the reversible Markov process associated with $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$.

This was proved by Donsker and Varadhan (CPAM75, 76, 83) under some conditions of absolute continuity and regularity of $P_t(x, dy)$, and established in full generality by L. Wu (JFA00).

Example 1 *$\mu = e^{-V(x)}dx/Z$ (Z is the normalization constant) with $V \in C^1$ on a complete connected Riemannian manifold $E = M$, the diffusion (X_t) generated by $\mathcal{L} = \Delta - \nabla V \cdot \nabla$ (Δ, ∇ are respectively the Laplacian and the gradient on M) is μ-reversible and the corresponding Dirichlet form is given by*

$$\mathcal{E}_\mu(g,g) = \int_M |\nabla g|^2 \, d\mu, \; g \in \mathbb{D}(\mathcal{E}_\mu) = H^1(\mathcal{X}, \mu)$$

If $\nu = f\mu$ with $0 < f \in C^1(M)$, then

$$I(\nu|\mu) = \int_M |\nabla \sqrt{f}|^2 \, d\mu = \frac{1}{4} \int_M \frac{|\nabla f|^2}{f} \, d\mu. \tag{5}$$

Information for discrete time Markov chains

Definition 3 *Given a symmetric Markov kernel $P(x, dy)$ on $L^2(\mu)$, Donsker-Varadhan information of ν with respect to (P, μ) is defined by*

$$I(\nu|P,\mu) := \begin{cases} \sup_{1 \le u \text{ bounded}} \int \log \frac{u}{Pu} d\nu, & \text{if } \nu \ll \mu, \sqrt{f} \in \mathbb{D}(\mathcal{E}) \\ +\infty, & \text{otherwise.} \end{cases} \tag{6}$$

Let (X_n) be the Markov chain with transition kernel P, then $\nu \longrightarrow I(\nu|P, \mu)$ is the rate function governing the large deviations of $L_n := \frac{1}{n}\sum_{k=1}^{n} \delta_{X_k}$.

2. Transportation-entropy inequalities $W_pH(C)$

$$W_p(\nu,\mu)^2 \le 2CH(\nu|\mu),\ \forall \nu \in M_1(E), \qquad (W_pH(C))$$

Theorem 1 $\mu \in W_1H(C)$ *iff for every Lipschitzian continuous function* $f : E \longrightarrow \mathbb{R}$ *with* $\|f\|_{Lip} = 1$, $f \in L^1(\mu)$ *and*

$$\mathbb{E}^\mu e^{\lambda(f-\mu(f))} \le e^{C\lambda^2/2},\ \forall \lambda \ge 0; \qquad (7)$$

(Bobkov-Götze criterion, JFA99), iff for every f *with* $\|f\|_{Lip} = 1$,

$$\mathbb{P}\left(\sqrt{n}(L_n(f) - \mu(f)) > r\right) \le e^{-r^2/2C},\ \forall r > 0, n \ge 1.$$

(Gozlan-Léonard's criterion, PTRF 08). So the best constant C *in* $W_1H(C)$ *could be called "Gaussian constant* $C_G(\mu)$*" of* μ.

Theorem 2 *(Djellout-Guillin-Wu, AOP04) A given probability measure* μ *on* (E, d) *satisfies* $W_1H(C)$ *on* (E, d) *if and only if*

$$\exists\delta > 0 : \iint e^{\delta d^2(x,y)} d\mu(x) d\mu(y) < +\infty. \tag{8}$$

In the latter case,

$$C = C(\delta) := \frac{1}{2\delta}\left(1 + 2\log \mathbb{E}e^{\delta d(\xi,\xi')^2}\right) \tag{9}$$

(estimate due to Bolley-Villani 05 and Gozlan 06)

About $W_2H(C)$: Talagrand's transportation inequality

Theorem 3 *(Talagrand, GFA96) Let μ be $\mathcal{N}(0, I)$ on $\mathbb{R}^d$. Then*

$$\mu \in W_2H(C),\ C = 1\text{(sharp)}.$$

Theorem 4 *(Otto-Villani, JFA00) On a complete connected Riemannian manifold, if μ satisfies log-Sobolev inequality, i.e.*

$$H(\nu|\mu) \le 2CI(\nu|\mu) \qquad (HI(C))$$

then $\mu \in W_2H(C)$. If $\mu \in W_2H(C)$, then μ satisfies the Poincaré inequality

$$Var_\mu(f) \le C\mathcal{E}(f, f). \qquad \text{(P(C))}$$

In summary,

$$HI(C) \implies W_2H(C) \implies P(C).$$

Remarks 3 • $P(C) \not\Rightarrow W_2H(C)$:

counter-example $\mu = e^{-|x|}/2dx$ on $\mathbb{R}$.

• $W_2H(C) \not\Rightarrow HI(C)$:

first counter-example given by Cattiaux-Guillin (JPAM 06)

Further reading:

— F.Y. Wang, W_2H on path spaces JFA 02

— Djellout-Guillin-Wu, W_2H for paths of dissipative diffusions w.r.t. L^2-metric, AOP 04

— J. Shao and S. Fang, W_2H on loop groups,

— K. Marton, W_2H for Gibbs measures

— L. Wu, W_1H for Gibbs measures, AOP 06

— F. Gao and L. Wu, $W_pI(C)$ for Gibbs measures.

Central idea: (1) $H(\nu|\mu)$ is the rate function for i.i.d. sequence (X_n) of common law μ.

(2) In the dependent stationary case of common law μ, if $I(\nu)$ is the rate function for LD of L_n, then

$$\mathbb{P}\left(W_1(L_n,\mu) > r\right) = e^{-n\inf\{I(\nu); W_1(\nu,\mu)>r\}+o(n)} \le e^{-n\alpha(r)+o(n)}$$

if the following transportation inequality holds:

$$\alpha(W_1(\nu,\mu)) \le I(\nu).$$

3. Transportation-information inequality W_pI

$$W_p(\nu,\mu)^2 \le 2CI(\nu|\mu),\ \forall \nu \in M_1(E), \qquad (W_pH(C))$$

Theorem 5 (Guillin-Léonard-Wu-Yao 06) *Let $c > 0$ and let (X_t) be a μ-reversible and ergodic Markov process associated with $(\mathcal{E}, \mathbb{D}(\mathcal{E}))$ such that*

$$\int d^2(x, x_0)\, d\mu(x) < +\infty.$$

Let

$$P_t^u f(x) := \mathbb{E}_x f(X_t) \exp\left(\int_0^t u(X_s) ds\right)$$

the Feynmann-Kac semigroup, whose generator is $\mathcal{L} + u$. The statements below are equivalent:

(i) The following $W_1 I(C)$ inequality holds true:

$$W_1^2(\nu,\mu) \le 2C I(\nu|\mu), \ \forall \nu \in M_1(\mathcal{X}); \qquad (W_1 I(C))$$

(ii) For all Lipschitz function u with $\|u\|_{\mathrm{Lip}} \le 1$, $\mu(u) = 0$ and all $\lambda \ge 0$,

$$\limsup_{t\to+\infty} \frac{1}{t} \log \mathbb{E}_\mu \exp\left(\lambda \int_0^t u(X_s)\,ds\right) \le C\lambda^2/2;$$

(iii) For all Lipschitz function u, $r > 0$ and $\beta \in M_1(\mathcal{X})$ such that $d\beta/d\mu \in L^2(\mu)$,

$$\mathbb{P}_\beta\left(\frac{1}{t}\int_0^t u(X_s)\,ds \ge \mu(u) + r\right) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-\frac{r^2}{2C\|u\|_{\mathrm{Lip}}^2}\right).$$

The constant C in $W_1 I(C)$ can be again interpreted as the Gaussian constant $C_G((P_t),\mu)$ for (X_t).

Relations between W_2I, Poincaré and log-Sobolev inequalities

Proposition 1 (Guillin-Léonard-Wu-Yao 06) *Let $\mathcal{X}$ be a complete connected Riemannian manifold and $\mu = e^{-V(x)}dx/Z$ where dx is the Riemannian volume measure, $V \in C^2(\mathcal{X})$ and $Z = \int_{\mathcal{X}} e^{-V} dx < +\infty$. Let $\mathbb{D}(\mathcal{E})$ be the space $H^1(\mathcal{X}, \mu)$ of those functions $g \in L^2(\mathcal{X}, \mu)$ such that $\nabla g \in L^2(TM, \mu)$ in the sense of distribution and consider the Dirichlet form,*

$$\mathcal{E}_{\nabla}(g, g) := \int_{\mathcal{X}} |\nabla g|^2 \, d\mu, \; g \in \mathbb{D}(\mathcal{E})$$

and the associated Fisher-Donsker-Varadhan information $I(\nu|\mu)$, see (5).

(a) If the log-Sobolev inequality below

$$H(\nu|\mu) \leq 2C \, I(\nu|\mu), \; \forall \nu$$

is satisfied, then μ *satisfies* $W_2I(2C^2)$.

(b) If $W_2I(C)$ *holds, then the Poincaré inequality holds with constant* $C_P \le \sqrt{2C}$.

(c) Assume that the Bakry-Emery curvature

$$\mathrm{Ric} + \mathrm{Hess}\mathrm{V} \ge \mathrm{K}$$

where Ric *is the Ricci curvature and* $\mathrm{Hess}V$ *is the Hessian of* V. *If* $W_2I(C)$ *holds with* $\sqrt{C/2}K \le 1$ *(this is possible by Part (a) and Bakry-Emery's criterion), then the log-Sobolev inequality*

$$H(\nu|\mu) \le 2(\sqrt{2C} - CK/2)\, I(\nu|\mu),\ \forall \nu$$

Proposition 2 *(Guillin-Léonard-Wang-Wu 07) In the same framework, we have for* $p = 1$ *or* 2,

$$W_pI(C) \implies W_pH(C).$$

4. $W_1I(C)$ for μ-symmetric Markov chain (X_n) with transition P

$\boldsymbol{W_1I(C)}$:

$$\boldsymbol{W_1(\nu,\mu)^2 \le 2CI(\nu|P,\mu).}$$

Two questions:

Q1. What is the probabilistic meaning of $\boldsymbol{W_1I(C)}$?

Q2. Criteria for $\boldsymbol{W_1I(C)}$?

Q1. Probabilistic meaning

Theorem 6 (Wu 08) *Let $C > 0$ and let (X_n) be a μ-reversible Markov chain with transition P. The statements below are equivalent:*

(i) The following $W_1I(C)$ inequality holds true:

$$W_1^2(\nu, \mu) \le 2CI(\nu|\mu), \ \forall \nu; \qquad (W_1I(C))$$

(ii) For all bounded Lipschitz function u with $\|u\|_{\mathrm{Lip}} \le 1, \ \mu(u) = 0$ and all $\lambda \ge 0$,

$$\limsup_{n \to +\infty} \frac{1}{n} \log \mathbb{E}_\mu \exp\left(\lambda \sum_{k=1}^{n} u(X_k)\right) \le C\lambda^2/2;$$

(iii) For all Lipschitz function u with $\|u\|_{\mathrm{Lip}} = 1$, $r > 0$ and $\beta \in M_1(\mathcal{X})$such that $d\beta/d\mu \in L^2(\mu)$,

$$\mathbb{P}_\beta\left(\tilde{L}_n(u) > \mu(u) + r\right) \le \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-\frac{nr^2}{2C}\right)$$

where $\tilde{L}_n(u) := \frac{1}{n}\left(\frac{u(X_0)+u(X_n)}{2} + \sum_{k=1}^{n-1} u(X_k)\right)$ is the trapeze type empirical mean.

The constant C in $W_1I(C)$ can be again interpreted as the Gaussian constant $C_G(P,\mu)$ for (X_n).

Proof : $(i) \implies (iii)$. By Lei (Bernoulli 07), for any $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}_\beta\left(\tilde{L}_n(u) > \mu(u) + r + \varepsilon\right)\\
&\leq \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-n \inf\{I(\nu|P,\mu); \nu(u) - \mu(u) > r\}\right)\\
&\leq \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-\frac{nr^2}{2C}\right)
\end{aligned}$$

because $r < \nu(u) - \mu(u) \leq W_1(\nu, \mu) \leq \sqrt{2CI(\nu|P,\mu)}$.

$(ii) \implies (i)$: by large deviations in Wu (JFA 00).

5. Three criteria

5.1. Poincaré is equivalent to $W_1 I(C)$ in the trivial metric case

Fact: if $d(x,y) = 1_{x \neq y}$, $W_1(\nu,\mu) = \|\nu - \mu\|_{TV}/2$.

Theorem 7 *Let $((X_n)_{n\geq 0}, \mathbb{P}_\mu)$ be a μ-symmetric ergodic Markov chain with transition probability P.*

1. *The Poincaré inequality*

$$Var_\mu(g) \leq C_P \langle g, (I-P)g \rangle_\mu, \ \forall g \in L^2(\mu) \tag{10}$$

implies

$$\|\nu - \mu\|_{TV}^2 \leq 4 C_P \, I(\nu|P,\mu), \ \forall \nu \in M_1(\mathcal{X}). \tag{11}$$

In particular for $u \in b\mathcal{B}$, for every initial probability measure $\beta \ll \mu$ with $d\beta/d\mu \in L^2(\mu)$ and with $\mu(u) = 0$ and for all $r, \varepsilon > 0$ and

$n \in \mathbb{N}^*$,

$$\mathbb{P}_\beta\left(\tilde{L}_n(u) \geq \mu(u) + r\right) \leq \left\|\frac{d\beta}{d\mu}\right\|_2 \exp\left(-\frac{nr^2}{c_P \delta(u)^2}\right) \qquad (12)$$

where $\delta(u) := \sup_{x,y\in\mathcal{X}} |u(x) - u(y)|$ *is the oscillation of* u.

2. *Conversely in the symmetric case, if* $\alpha(\|\nu-\mu\|_{TV}) \leq I(\nu|P,\mu),\ \forall\nu$, *for some nonnegative nondecreasing left-continuous function* $\alpha : \mathbb{R}^+ \longrightarrow [0, +\infty]$ *with* $\alpha(1) > 0$, *then the Poincaré inequality (10) holds with*

$$c_P \leq \frac{1}{1 - e^{-\alpha(1)}}. \qquad (13)$$

5.2. Unbounded metric: spectral gap in the space of Lipschitzian functions

The carré-du-champs operator associated with $\mathcal{L} = P - I$ is

$$\Gamma(g,h)(x) = \frac{1}{2}\int (g(y) - g(x))(h(y) - h(x))P(x,dy)$$

Consider the following condition relating P with the metric d:

$$\sup_{g:\|g\|_{Lip}=1} \sup_{x\in\mathcal{X}} \sqrt{\Gamma(g)}(x) \le M. \tag{14}$$

Notice that (14) is satisfied if

$$\frac{1}{2}\int d^2(x,y)P(x,dy) \le M^2,\ \forall x. \tag{15}$$

Let $C_{Lip}(\mathcal{X})$ (resp. $C_{Lip,0}(\mathcal{X})$) be the space of all d-Lipschitzian functions g (resp. with $\mu(g) = 0$) on $\mathcal{X}$.

Theorem 8 *Assume (14) and* $\int d^2(x,x_0)d\mu(x) < +\infty$ *and* P *is* μ*-symmetric. Suppose that* P *admits a spectral gap in* $C_{Lip}(\mathcal{X})$*, i.e., for any* $g \in C_{Lip,0}(\mathcal{X})$*, there is* $G \in C_{Lip,0}(\mathcal{X})$ *solving the Poisson equation* $(I-P)G = g,\ \mu - a.s.$ *and satisfying*

$$\|G\|_{Lip} \le c_{P,Lip}\|g\|_{Lip} \tag{16}$$

where $c_{P,Lip} > 0$ *is the best constant (here the index* P *refers to Poincaré). Then* μ *satisfies the Poincaré inequality with* $c_P \le c_{P,Lip}$ *and it satisfies* $W_1 I$ *below*

$$W_1(\nu,\mu)^2 \le 4(Mc_{P,Lip})^2 I(\nu),\ \forall \nu \in \mathcal{M}_1(\mathcal{X}). \tag{17}$$

The following result, inspired of Djellout-Guillin-Wu (AOP 04), provides sharp constant.

Proposition 3 *In the framework of Theorem 8 but without condition (14), assume that for some constant $c_H(P) > 0$,*

$$W_1^2(\nu, P(x,\cdot)) \le 2c_H(P)H(\nu|P(x,\cdot)),\ \forall x \in \mathcal{X}, \nu \in M_1(\mathcal{X}) \tag{18}$$

Then

$$W_1^2(\nu, \mu) \le 2(c_{P,Lip})^2 c_H(P) I(\nu|P, \mu). \tag{19}$$

Remarks 4 The (inverse) Lipschitzian spectral gap constant $c_{P,Lip}$ can be estimated easily by

$$c_{P,Lip} \le \sum_{n=0}^{\infty} \|P^n\|_{Lip}$$

where

$$\|P\|_{Lip} = \sup_{g:\|g\|_{Lip}=1} \|Pg\|_{Lip} = \sup_{x \ne y} \frac{W_1(P(x,\cdot), P(y,\cdot))}{d(x,y)}.$$

Ollivier called

$$\kappa(x,y) = 1 - \frac{W_1(P(x,\cdot),P(y,\cdot))}{d(x,y)}$$

(Ricci) curvature of the Markov chain. If $\kappa(x,y) \geq \kappa > 0$, then $\|P\|_{Lip} \leq 1-\kappa$ and then $c_{P,Lip} \leq 1/\kappa$. This last estimate is far from being sharp in general: for the Markov chain on $\{0,1\}$ given by $P(0,1) = P(1,0) = 1$, we have $\kappa = 0$ whereas $c_{P,Lip} = 1/2$.

Lyapunov function criterion for $W_1 I$:

Theorem 9 *Assume that P is μ-symmetric and satisfies Poincaré inequality (10) with best constant $c_P < \infty$. If the Lyapunov condition*

(H) *There exist a measurable function $U : \mathcal{X} \to [1, +\infty)$ and a non-negative function ϕ and a constant $b > 0$ such that $PU(x) < +\infty,\ \mu - a.e.$ and*

$$\log \frac{U}{PU} \geq \delta d^2(x, x_0) - b,\ \ \mu\text{-}a.s.$$

holds. Then

$$W_1(\nu, \mu)^2 \leq 2\tilde{C} I(\nu),\ \forall \nu;\ \tilde{C} := \frac{2}{\delta}[1 + (1 + b)c_P \mu(d^2(x, x_0))] \quad (20)$$

6. Two examples

6.1. Two points model We begin with the simplest Markov chain on $\mathcal{X} = \{0,1\}$ equipped with the trivial metric d, with transition matrix $P = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$, where $a, b \in (0,1]$. Notice that P is symmetric w.r.t. μ given by

$$\mu(0) = \frac{b}{a+b} =: q, \ \mu(1) = \frac{a}{a+b} =: p.$$

Though this model is simple, but its study is abundant: for the Dirichlet form

$$\mathcal{E}_P(g,g) = \langle g, (I-P)g \rangle_m u = \frac{ab}{a+b}(g(1) - g(0))^2$$

associated with the continuous time Markov process generated by $\mathcal{L} = P - I$,

1. the best log-Sobolev constant is known, see Saloff-Coste and al.;

2. the best constant $C_H(p)$ in $W_1(\nu,\mu)^2 \le 2CH(\nu|\mu)$ is obtained recently by Bobkov, Houdré and Tetali (JIM 08) :

$$C_H(p) = \frac{p-q}{2(\log p - \log q)} \; (:= 1/4 \;\; if \;\; p = q); \tag{21}$$

3. the best rate $\kappa > 0$ in the exponential entropy convergence

$$H(\nu P_t|\mu) \le e^{-\kappa t} H(\nu|\mu)$$

is unknown, only some accurate estimates are known, see M.F. Chen (07).

4. the best constant C in $W_1(\nu|\mu)^2 \le 2CI_c(\nu)$ where

$$I_c(\nu) = \langle (I-P)\sqrt{\frac{d\nu}{d\mu}}, \sqrt{\frac{d\nu}{d\mu}} \rangle_\mu$$

is known: $C = 1/[2(a+b)]$ (Guillin-Léonard-Wu-Yao 08).

$$c_{P,Lip} = c_P = (a+b)^{-1}$$

(however its curvature $\kappa = 1 - |1 - (a+b)|$).

$$C_G(P,\mu) \leq \frac{\max\{c_H(a), c_H(b)\}}{(a+b)^2} \tag{22}$$

which becomes **equality** if $a+b=1$ (i.e., i.i.d. case, for $c_H(a) = c_H(b) = C_H(\mu)$). By the calculation of the asymptotic variance $V(g)$ we have

$$C_G(P,\mu) \geq \left(\frac{2}{a+b} - 1\right)\frac{ab}{(a+b)^2}.$$

Notice also a curious phenomena: if $a = b = 1$, $c_H(P) = 0$ and then $C_G(P,\mu) = 0$.

We do not know the exact expression of $C_G(P,\mu)$.

6.2. Complete graph Let $\mathcal{X}$ be the complete graph of N (≥ 3) vertices, i.e., any two vertices are connected by an edge and then the graph distance is given by $d(x, y) = 1_{x \neq y}$ (the trivial metric). The probability transition matrix is given by $P(x, y) = \frac{1}{N-1}$ for all $y \neq x$. It is symmetric w.r.t. the uniform measure $\mu(x) = 1/N$ (for each $x \in \mathcal{X}$). It is easy to see that

$$(I - P)G = \frac{N}{N-1}(G - \mu(G)).$$

Thus $c_P = \frac{N-1}{N} = c_{P,Lip}$. By Theorem 7, we have

$$C_G(P, \mu) \leq \frac{N-1}{2N}.$$

Proposition 3 yields the better

$$C_G(P, \mu) \leq \frac{1}{4}\left(\frac{N-1}{N}\right)^2. \tag{23}$$

This is sharp for large N. Indeed

$$C_G(P,\mu) \geq \begin{cases} \frac{N-2}{4N}, & \text{if } N\ even; \\ \frac{N-2}{4N}(1-\frac{1}{N^2}), & \text{if } N\ odd. \end{cases}$$

Thanks!